Extension of the aspherical pseudoatom databank towards nucleic acids and its application in structural, charge density, and energy studies

Katarzyna Natalia Jarzembska

PhD Thesis

Advisor: Dr Paulina Maria Dominiak

Supervisor: Prof. Krzysztof Woźniak

Laboratory of Crystallochemistry

Department of Chemistry

University of Warsaw

Warsaw 2012

I dedicate this work to my beloved ***Mother****,*
for all the love, devotion, support and encouragement
that she has been giving me.

I would like to express my deep gratitude to my supervisors
Prof. Krzysztof Woźniak *and* ***Dr Paulina Dominiak***
for introducing me to the secrets of science and its milieu, and the enormous confidence they put in me throughout my PhD research.

My sincere thanks go to the entire research group for the interesting and inspiring discussions we had, especially to Sławomir Domagała, Krzysztof Durka, Anna Goral and Anna Hoser.

I would also like to thank the co-authors of publications, and manuscripts in preparation, for their valuable contribution.

Furthermore, special thanks to my Father for his invaluable comments and suggestions, and my Family for their support.

Finally, I would like to express my warm thanks to my fiancé, Radek, for his love, patience, exceptional help and support.

The support received from the Foundation for Polish Science with the POMOST/2010-2/3 project co-financed by the European Union, Regional Development Fund and from the MASTER subsidy is greatly acknowledged. I would also like to thank Professor Claude Lecomte for arranging the helium temperature measurements.

Table of Contents

Abbreviations

ADP	–	Atomic Displacement Parameter
B3LYP	–	Becke 3-parametr Lee-Yang-Parr functional
BCP	–	Bond Critical Point
BP	–	Bond Path
IAM	–	Independent Atom Model
TAAM	–	Transferable Aspherical Atom Model
UBDB	–	University at Buffalo DataBank
UBDB2011	–	new version of UBDB
EPMM	–	Exact Potential / Multipole Method
ESP	–	Electrostatic Potential
QTAIM	–	Quantum Theory of Atoms in Molecules
HOMA	–	Harmonic Oscillator Model of Aromaticity
NICS	–	Nucleus Independent Chemical Shift
CSD	–	Cambridge Structural Database
DNA	–	Deoxyribonucleic Acid
RNA	–	Rybonucleic Acid
ORTEP	–	Oak Ridge Thermal Ellipsoid Plot
VSCC	–	Valence Shell Charge Concentration
TDS	–	Thermal Diffuse Scattering
DFT	–	Density Functional Theory
HB	–	Hydrogen Bond
MO-LCAO	–	Molecular Orbitals – Linear Combination of Atomic Orbitals method
BSSE	–	Basis Set Superposition Error
NAB	–	Nucleic Acid Base
PDB	–	Protein Databank
NDB	–	Nucleic acid Databank
e.s.d.	–	Estimated Standard Deviation
RMSD	–	Root Mean Square Deviation
TLS	–	Translation-Libration-Screw formalism
CCD	–	Coupled Charge Device
SHADE	–	Simple Hydrogen Atomic Displacement Estimator

E_{int}	–	interaction energy
E_{coh}	–	cohesive energy
E_{surf}	–	surface free energy
E_{intl}	–	interlayer energy
E_{bulk}	–	energy of bulk
E_{mol}	–	energy of molecule
E_{slab}	–	energy of slab
E_{es}	–	electrostatic interaction energy

a	–	scalar quantity (*e.g.* F, h)
$\mathbf{a}$	–	vector quantity (*e.g.* $\mathbf{r}$, $\mathbf{H}$)
$\|\cdot\|$	–	vector norm (*e.g.* $\|\mathbf{r}\|$)
$\lvert\cdot\rvert$	–	complex number absolute value (*e.g.* $\lvert F\rvert$)
$\mathbf{H}$	–	reciprocal lattice vector ($h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$)
ϱ	–	electron density
$\nabla^2\varrho$	–	Laplacian of electron density
V	–	potential energy density
G	–	kinetic energy density
F	–	structure factor
f_k	–	atomic scattering factor of the k-th atom
Q_k	–	AIM atom charge of the k-th atom
i	–	imaginary unit

Compounds

CSD codes of new structures are given in parentheses

U (**U**) / A / T / C / G	–	Uracil / Adenine / Thymine / Cytosine / Guanine
1mT / 15dmU	–	1-methylthymine
(CCDC 873162)[a]		
1mU	–	1-methyuracil
(CCDC 873157)[a]		
2tU	–	2-thiouracil
(CCDC 873158)[a]		
4tU	–	4-thiouracil
(CCDC 873159)[a]		
24dtU	–	2,4-ditiouracil
(CCDC 873163)[a]		
5m2tU	–	5-methyl-2-thiouracil
6m2tU	–	6-methyl-2-thiouracil
(CCDC 873151)[a]		
(CCDC 888248)[b]		
(CCDC 888247)[c]		
1m4tU	–	1-methyl-4-thiouracil
5fU	–	5-fluorouracil
(CCDC 873160)[a]		
9mA	–	9-methyladenine
(CCDC 888250)[a]		
9mA:1mT	–	9-methyladenine:1-methylthymine co-crystal
(CCDC 888249)[b]		
9mA$_c$	–	**9mA** molecule present in the **9mA:1mT** crystal structure
1mT$_c$	–	**1mT** molecule present in the **9mA:1mT** crystal structure
AmB	–	amphotericn B
AmB-I	–	*N*-iodoacetylamphotericn B
NA	–	Influenza Neuraminidase
ZMR	–	zanamivir
SIA	–	sialic acid

[a] Measured at 100 K. [b] Measured at 90 K. [c] Measured at 10 K.

"This report, by its very length, defends itself against the risk of being read"

Winston Churchill

"The existing scientific concepts cover always only a very limited part of reality, and the other part that has not yet been understood is infinite"

Werner Heisenberg

I. Introduction

The turn of the XX century abounded in many ground-breaking scientific ideas, spectacular discoveries and great inventions. At that time people started to interpret the surrounding World in an entirely new way, using complex, more abstract theories, and all that in a much wider context. The more they were able to measure, or detect, the more ideas emerged, and inevitably, also, far more new questions appeared. It seemed no limits any longer existed.

That was also the time when X-ray crystallography came into being. The first step towards it was Wilhelm Röntgen's discovery of rays, which he later called X-rays, for which he was awarded the Nobel Prize in 1901. Next, Max von Laue discovered diffraction of X-rays in crystals. Shortly after this success, first theories, which constitute the basis for crystal structure determination, had been formulated. Brilliant work of Max von Laue and the Braggs (William Henry and his son William Lawrence), who introduced the diffraction law, and also, built the first X-ray spectrometer, were soon appreciated, followed by further Nobel Prize awards in 1914 and 1915, respectively. As of that time, scientists had been learning how to collect diffraction data efficiently, how to use it, and also what kind of information might be extracted from such measurements. Inspired by first experiments, encouraged by observations, and fascinated by the beautiful phenomenon of crystal diffraction, scientists have been exploring the nature of crystals, constantly improving experimental methods, and developing the theory behind it. X-ray crystallography has become a powerful tool for crystal structure determination and charge density distribution analysis. Nowadays, this method not only constitutes a good analytical tool, but is also a starting point for further energetic studies. Furthermore, charge density distribution investigations may shed light on some fine

effects present in the crystal lattice, such as crystal field influence, atomic polarization, charge transfer, subtle disorder, *etc.* X-ray crystallography provides the basis for crystal engineering, and is widely applied for the purpose of pharmacy, medicine, technology and others. The new results of scientific research, pertaining to this method, are still being appreciated and considered valuable. Even last year a Nobel Prize was awarded to Daniel Shechtman, who introduced the idea of quasi crystals, and alike in 2009, when the structural solution of ribosome was rewarded (Ada E. Yonath, Venkatraman Ramakrishnan, Thomas A. Steitz).

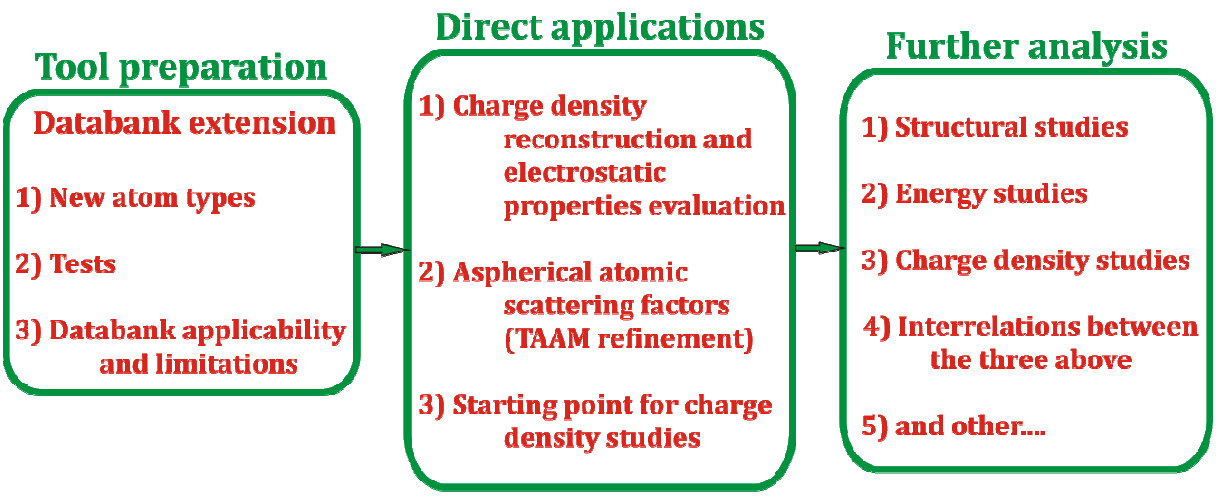
However, till today there are many questions still awaiting answers, and things that need to be improved. This stimulates the ongoing development in this field, and also, makes crystallography so fascinating. All this makes up for the magical value and huge potential of X-ray methods, which enchanted me so much.

I.1. Preface

A few decades ago, charge density distribution studies in crystal phase constituted a forefront scientific subject in the field of crystallography. It had already been clear that approximate experimental charge density could be obtained as a product of Fourier transform of the structure factors with known phases. Thanks to direct methods the 'phase problem', *i.e.* extracting the phase information from the registered amplitudes of the diffracted X-ray beam, had been overcome. Nevertheless, the interpretation of such density distribution still posed a problem. Thus, its proper modelling and description had become a scientific challenge of that time. The use of conventional spherical-atom models, widely applied for X-ray structure determination, was not sufficient for that purpose. Significant discrepancies were observed between molecular charge density distributions reconstructed with the aid of these models and the information contained in the collected experimental reflection intensities. Most significant residual features appeared around middle points of covalent bonds, or in the regions of the expected lone electron pairs. Indeed, the spherical atom model does not mimic the various electron density deformations resulting from interatomic interactions. There were also substantial differences noted for bond lengths and atomic displacement parameters (ADPs) between X-ray and neutron data collected for the same species, especially pronounced with regards to the positions of hydrogen atoms. The reason for that once again lies in the specificity of the spherical atom model, and the least square procedure used to fit the experimental electron density in structural refinement. Therefore, scientists have put a lot of effort both into the experiment technique development and proper data analysis. With the improvement of X-ray data collection techniques, several atomic density models and corresponding atomic scattering factors were suggested, among which the most widely applied is the Hansen-Coppens formalism. It is worth noting, though, that it is possible to obtain experimental charge density data only, if the crystal is of high quality (most preferably no twinning, no disorder, *etc.*) and diffracts well up to subatomic resolution of 0.5 Å. These conditions, as practice shows, are quite demanding and rarely fulfilled for crystals formed by rather labile macromolecules. In the case of large biological systems, it is additionally impossible to run *ab initio* calculations, which is an option for smaller compounds. And this is why the idea of aspherical pseudoatom databank has arisen…

I.2. Objectives

My PhD project, in its first order of priority, was devoted to the extension of the aspherical pseudoatom databank, UBDB (University at Buffalo Databank), towards nucleic acid modelling. Once the bank is prepared and thoroughly tested, it is used in three different ways, so as to cover the scope of its main applications, *i.e.* charge density reconstruction (and on that basis estimation of electrostatic properties), a source of aspherical atomic scattering factors, and finally – a starting point for experimental charge density distribution evaluation. For the first purpose, I have used the complex of influenza neuraminidase with a set of inhibitors, whereas for the last two, I have chosen a series of uracil derivatives, which reflects the leading idea of the project, *i.e.* first steps in the direction of reconstructing nucleic acid chain charge density distribution. I have found the latter compounds extremely interesting, and particularly suitable for studies going far beyond pure structural and charge density analysis. Therefore, the second goal of my PhD project was to explore and find the interplay between structure (geometry), charge density distribution, crystal architecture energetic features, and crystal morphology. The idea of my PhD thesis is schematically illustrated below (Scheme 1.1).



Scheme 1.1. The scope of my PhD thesis.

I.3. Thesis structure and content

My PhD thesis consists of two main parts. The first one deals with the databank development. The aim of this stage of my project was to find more accurate atom type definitions, to model the remaining atom types present in the nucleic acid chains and other biological molecules of interest, and to recalculate the parameters of the old atom types.

The next step was to verify and validate the obtained aspherical atom parameters. Thus, this part contains a detailed description of the databank construction and all the introduced modifications with respect to its initial version. The atom type definition changes (the addition of new atomic keys), and the increased number of atom types, required substantial interference into the code of the related *LSDB* program. *LSDB* assigns the parameters stored in the databank to a given molecular geometry intended to be further analysed. *LSDB* provides a number of useful options, some of which were modified or added, and therefore have been for the first time accurately described. As the electrostatic properties are most sensitive to the quality of aspherical charge density parameters, I have decided to validate the databank on the basis of the derived electrostatic interaction energy values. As a test set, the systems containing nucleic acid base dimers, or mixed amino acid- nucleic acid base complexes were chosen. The electrostatic interaction energy results obtained on the basis of the databank were compared to that coming from *ab-initio* or DFT calculations. I verified the accuracy, applications and limitations of our method. The first part of my PhD thesis can be briefly summed up as follows:

- New atom types definitions and *LSDB* modifications;
- UBDB databank extension with atom types present in nucleic acid bases and other biologically relevant molecules;
- Pseudoatom charge density parameter tests on a set of nucleic acid base complexes and mixed amino acid- nucleic acid base dimers;
- Formulating the limitations of the databank applicability.

The second part of my thesis consists of practical applications of the newly extended databank. However, UBDB constitutes here usually a starting point for further investigations. The idea of this part was rather to examine the electrostatic interactions in the influenza neuraminidase active site, and to explore the rules governing the crystal

lattice of uracil based systems. I aimed to find the interplay between charge density distribution and the interaction energy, between the molecular geometry and energy features, but also to relate crystal architecture motif characteristics to the crystal morphology and stability. Some of the results were additionally analysed in a wider context of nucleic acids.

As the leading motif of this part of my thesis is the UBDB databank application, it is divided into three sections. The analysis of neuraminidase electrostatics opens up the undertaken set of scientific problems. This study shows the potential of the databank to solve some pharmaceutical and biological queries. My contribution to the wide inhibitor-neuraminidase binding property study was to analyse the role of water molecules in the active site. Therefore, the emphasis is put on this particular investigation.

The second major application of the databank is its use as a source of aspherical atomic scattering factors in the so-called TAAM (Transferable Aspherical Atom Model) refinement. I have selected a series of uracil derivatives as a set of test compounds. I have carefully analysed the quality of molecular geometries obtained both from IAM and TAAM refinement in respect to the optimised structures and also to some available charge density and neutron diffraction results. Apart from this analysis, I have subjected all the structure sets to energy investigations. The whole analysis has resulted in a comprehensive structural and energy study of the selected systems.

The third section of this part of my PhD thesis is devoted to experimental charge density analysis. The databank was used as the first approximation of the experimental charge density distribution, but also served as a sulphur atom model. The analysed compounds concern a single uracil derivative, 6-methyl-2-thiouracil, and the Hoogsteen-Watson-Crick type dimer, a co-crystal of 9-methyl adenine and 1-methyl thymine. These two provided a wide range of different scientific problems to be thoroughly investigated. I compared the UBDB derived molecular charge densities with the real, experimentally obtained values. Multi-temperature studies enabled more proper de-convolution of thermal motion and static charge density. A lot of my attention was also paid to the modelling of sulphur atom. Furthermore, crystal architecture energetic features were extensively studied and related to charge density findings and crystal morphology. Finally, the co-crystal structure was analysed in the context of other purine-pyrimidine

crystals, as well as, in respect to the RNA base-pairing. The second part of my PhD thesis can be shortly recapitulated as follows:

- Electrostatics derived on the basis of UBDB: neuraminidase active site analysis;
- TAAM refinements: uracil derivatives, geometry and energy studies;
- UBDB as a starting point for charge density studies:

  (1) 6-methyl-2-thiouracil, charge density distribution modelling, tautomerism analysis, energetic investigations, crystal morphology aspects;

  (2) 9-methyladenine-1-methylthymine complex, charge density distribution modelling, energetic investigations, RNA context, UBDB-derived charge density comparison with experimental charge density model.

"I don't know anything, but I do know that everything is interesting if you go into it deeply enough"

Richard Feynman

II. Theoretical Background and Methods

To introduce the subject matter of my PhD thesis, it is required to recall the major facts regarding diffraction experiments, charge density evaluation and related topological analysis. Additionally, the 'pseudoatom databanks – idea, construction and applications' part provides the indispensable information about the existing pseudoatom databanks and their applications, as a great majority of my studies is devoted to one of them- the theoretical databank of transferable aspherical pseudoatoms – UBDB. Periodic energetic computations and other theoretical methods also constitute an integral part of the conducted analyses. Thus, the final section of this chapter includes a brief description of the main principles of quantum chemical methods used for energy elucidation and geometry optimisation. More specific information regarding investigations presented within my PhD thesis shall be provided during the actual analysis of the results.

II.1. Experimental charge density distribution evaluation and analysis

My PhD study was based on X-ray diffraction by crystals, whereas the acquired data was an inspiration for the presented scientific ideas and conclusions. Therefore, the first part of this chapter deals with the diffraction phenomenon, relevant data collection, structural refinement, charge density modelling, and finally – data interpretation. It also includes experimental requirements and information on model limitations. The last part, as an additional item, contains a description of the Bader theory of 'Atoms in Molecules'.

II.1.1. Diffraction experiment geometry

The diffraction of X-rays by a crystal can be well explained using the Ewald construction.[1] This model illustrates intelligibly the diffraction experiment geometry and indicates the link between the X-ray beam trajectory and the crystal lattice specificity. In the Ewald approach the primary and reflected beams are represented by $\mathbf{s}_0$ and $\mathbf{s}$ unit vectors, respectively. These vectors are related to the reciprocal lattice vector $\mathbf{H}$ (*i.e.*, $h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$) by the formula II.1 presented below:

$$\mathbf{H} = \frac{\mathbf{s} - \mathbf{s}_0}{\lambda} \tag{II.1}$$

where λ stands for radiation wavelength. Graphical representation of the Ewald construction is shown in Figure 2.1. The drawn sphere of the radius equal to $1/\lambda$ is called the *Ewald sphere*. This figure indicates the conditions, which need to be fulfilled to enable the occurrence of diffraction signals. Diffraction occurs when a reciprocal lattice vector, anchored at the point $\boldsymbol{O}$, intersects with the Ewald sphere. Consequently, in order to experimentally register reflections coming from different reciprocal lattice nodes, i.e. to make them fulfil the diffraction condition, either the crystal needs to be put into oscillation (or rotated), or the monochromated radiation should be replaced by a polychromatic beam. Polychromatic beam is utilised in the Laue method, which is, however, very complex to interpret. Here, the attention is focused on the goniometric method, which was applied in the course of my studies. The crystal lattice is coupled with the reciprocal lattice, and thus its motion drives the simultaneous rotation of the reciprocal lattice nodes and allows for occurrence of numerous diffraction processes. The Ewald construction has been already applied to explain various diffraction accompanying phenomena, but also helped in designing new robust data collection techniques (*e.g.*, Weissenberg method, Buerger precession method, and all modern

goniometer methods). What is worth mentioning, simple geometrical considerations confirm the equivalence of the Ewald construction with the well-known *Wulff-Braggs equation* in the following form:

$$\frac{\sin\theta}{\lambda} = \frac{\|\mathbf{H}\|}{2} \tag{II.2}$$

where θ is the half of the diffraction angle (sometimes called the Bragg angle) and $\|\mathbf{H}\|$ is the length of the vector **H**.

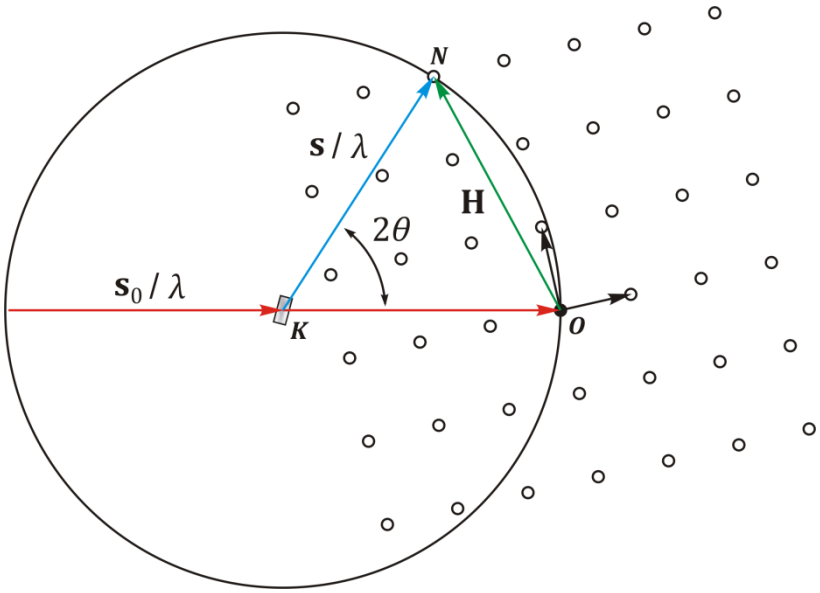


Figure 2.1. X-ray diffraction experiment geometry (2D Ewald sphere projection).

II.1.2. Structure factor

The Ewald construction provides a concise image of the X-ray diffraction geometry, being an illustration for the Bragg equation. However, it is the electron density which induces the X-ray diffraction phenomenon. Thus, to draw conclusions regarding crystal structure solution and further charge density distribution evaluation, one needs to understand the nature of the deflected beam intensity distribution. According to the kinematical theory, the integrated intensity of the diffracted beam, I, is proportional to the square module of the complex structural quantity called the structure factor, F:[2]

$$I(\mathbf{H}) \sim |F(\mathbf{H})|^2 \tag{II.3}$$

where **H** is the reciprocal lattice vector describing the diffracted beam direction. The above introduced structure factor describes the 'scattering power' of the electron density within the unit cell. The kinematical theory states that structure factors can be therefore derived as a Fourier transform of the electron density distribution of such a unit cell:

$$F_{\mathrm{a}}(\mathbf{H}) = \int \varrho(\mathbf{r})\exp(2\pi\mathrm{i}\mathbf{H}\cdot\mathbf{r})\mathrm{d}\mathbf{r} \tag{II.4}$$

where the subscript "a" hereafter denotes the absolute scale. To compute the quoted integral some approximations are being made. The typical electron density model consists of 'independently' vibrating atom-centred contributions, that is:

$$\varrho(\mathbf{r}) = \sum_k \int \varrho_k(\mathbf{r}-\mathbf{r}_k-\mathbf{u}_k)P_k(\mathbf{u}_k)\mathrm{d}\mathbf{u}_k \tag{II.5}$$

where ϱ_k is the electron density distribution of the k-th atomic contribution, $\mathbf{r}_k$ is the atom location is space, $\mathbf{u}_k$ is the corresponding thermal displacement vector, and finally P_k constitutes the probability density function describing atomic thermal smearing. Substituting the above expression into the structure factor definition, the following form is readily obtained:

$$F_{\mathrm{a}}(\mathbf{H}) = \sum_k f_k T_k \exp(2\pi\mathrm{i}\mathbf{H}\cdot\mathbf{r}_k) \tag{II.6}$$

where

$$f_k = \int \varrho_k(\mathbf{r})\exp(2\pi\mathrm{i}\mathbf{H}\cdot\mathbf{r})\mathrm{d}\mathbf{r} \tag{II.7}$$

$$T_k = \int P_k(\mathbf{u}_k)\exp(2\pi\mathrm{i}\mathbf{H}\cdot\mathbf{u}_k)\mathrm{d}\mathbf{u}_k \tag{II.8}$$

The quantities f_k and T_k are called the atomic scattering factor and the atomic thermal factor, respectively. The first one plays a crucial role in the structural analysis, as it is directly related to the atomic electron density features (*e.g.* 'shape'). The second one reflects atomic thermal vibrations and shall be discussed in the following paragraph.

II.1.3. Thermal smearing

The last integral in the previous paragraph, II.8, expresses the way in which electrons are 'smeared' over the space due to atomic vibrations. The P function is usually considered within harmonic approximation, which assumes that atoms are submerged in the potential of the Gaussian-type shape distribution.[3] Its general anisotropic form is described by the formula:

$$P(\mathbf{u}) = [(2\pi)^3\det\boldsymbol{\sigma}]^{-1/2}\exp\left(-\frac{\mathbf{u}^{\mathrm{T}}\boldsymbol{\sigma}^{-1}\mathbf{u}}{2}\right) \tag{II.9}$$

where $\boldsymbol{\sigma}$ is a symmetric matrix of mean-square atomic displacements ($\sigma_{ij} = \langle u_i u_j\rangle$). The corresponding thermal atomic factor is then as follows:

$$T(\mathbf{H}) = \exp(-2\pi^2 \mathbf{H}^{\mathrm{T}} \boldsymbol{\sigma} \mathbf{H}) = \exp(-\mathbf{H}^{\mathrm{T}} \boldsymbol{\beta} \mathbf{H}) \tag{II.10}$$

where

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{12} & \beta_{22} & \beta_{23} \\ \beta_{13} & \beta_{23} & \beta_{33} \end{bmatrix} \tag{II.11}$$

The matrix $\boldsymbol{\beta}$ represents anisotropic atomic vibration and contains 6 independent parameters, β_{ij}. These parameters are related to the frequently quoted U_{ij}'s by the formula:

$$\beta_{ij} = 2\pi^2 U_{ij} a_i^* a_j^* \tag{II.12}$$

where a_i^* stands for the i-th reciprocal space vector length. More sophisticated atomic vibration description can be achieved via the statistical approaches such as the Gram-Charlier formalism, which introduce the anharmonic motion terms.

Finally, some more attention should be devoted to the formula II.5, which can be rewritten to the equivalent form:

$$\varrho(\mathbf{r}) = \sum_k \varrho_k(\mathbf{r} - \mathbf{r}_k) \star P_k(\mathbf{u}_k) \tag{II.13}$$

where $\star$ denotes the convolution operation. The expression above shows the convolution of the static charge density with the probability density function. The fundamental idea of deriving the static electron density from the X-ray diffraction data requires its proper de-convolution from the thermal motion. This is going to be discussed further in the chapter.

II.1.4. Fourier synthesis

Having a set of structure factors, it is, of course, possible to perform an inverse Fourier transform, and on that basis obtain electron density distribution within a unit cell. Such a set of structure factors is discrete, as only Bragg reflections are taken into account, so the integral can be replaced by the following sum:

$$\varrho(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{H}} |F_{\mathrm{a}}(\mathbf{H})| \exp(\mathrm{i}\varphi_{\mathbf{H}}^{\mathrm{c}}) \exp(-2\pi \mathrm{i} \mathbf{H} \cdot \mathbf{r}) \tag{II.14}$$

where V is the unit cell volume and $\varphi_{\mathbf{H}}^{\mathrm{c}}$ stands for the calculated (superscript "c") reflection phase. The given formula II.14 describes the so-called dynamic electronic density, as it includes thermal motion. Similarly, the dynamic electron density can be derived from the sum of the 'experimental' $|F^{\mathrm{o}}|$ values (superscript "o" stands for 'observed'). It is worth noting here, that such 'measured' structure factors are never

given in the absolute scale. Therefore, the introduction of the scale factor, $k_S > 0$, is required. The scale factor, which always scales the model not the data, is defined as:

$$F^c(\mathbf{H}) = k_S \cdot F_a(\mathbf{H}) \tag{II.15}$$

When $|F^o|$ is related to $|F^c|$, this results in the so-called residual density:

$$\varrho_{res}(\mathbf{r}) = \varrho^o(\mathbf{r}) - \varrho^c(\mathbf{r}) = \\ = \frac{1}{V}\sum_{\mathbf{H}} k_S^{-1}(|F^o(\mathbf{H})| - |F^c(\mathbf{H})|)\exp(i\varphi_{\mathbf{H}}^c)\exp(-2\pi i\mathbf{H}\cdot\mathbf{r}) \tag{II.16}$$

Obviously, due to the well-known phase problem, the phases can be taken only from the currently investigated model. The initial phases are obtained on the basis of direct method assumptions or within the charge flipping approach. All in all, such a residual density is a fairly good indicator of the model quality, as it shows to what extent the model fits the data.

II.1.5. Spherical electron density models

In order to reflect a real charge density distribution of a given system, one needs to find a proper, rather flexible and physically reasonable, charge density model capable of describing electron density distribution within a crystal as precisely as possible. However, for the purpose of this thesis, it is desired to start the story from the simplest models, which are, however, still quite often used. Thus, the Independent Atom Model (IAM) is the first selected one. This model considers molecular charge density as the sum over vibrating *spherical* and *non-interacting* (*i.e.*, isolated) atoms. The atomic scattering factor within IAM is calculated as follows:

$$f_k(\|\mathbf{H}\|) = 4\pi\int_0^\infty r^2\varrho_k(r)\frac{\sin(2\pi\|\mathbf{H}\|r)}{2\pi\|\mathbf{H}\|r}\mathrm{d}r \tag{II.17}$$

In the above definition the ϱ_k is the spherical electron density calculated for an isolated k-th atom. Such integral is called the Fourier-Bessel integral due to the presence of the zero-order Bessel function of the first type: $j_0(x) = x^{-1}\sin x$, and it is usually calculated numerically. As clearly seen, the value of f_k for $\|\mathbf{H}\| = 0$ amounts to the number of electrons of the particular k-th atom:

$$f_k(\|\mathbf{H}\|) = N_k \tag{II.18}$$

and is approaching the zero value at the infinity (Figure 2.2). This is the reason why the high-resolution (*i.e.*, with a larger θ diffraction angle value) reflections are on average

much weaker than the low-resolution ones. Above expression also explains why, in general, crystals containing some heavy atoms diffract better.

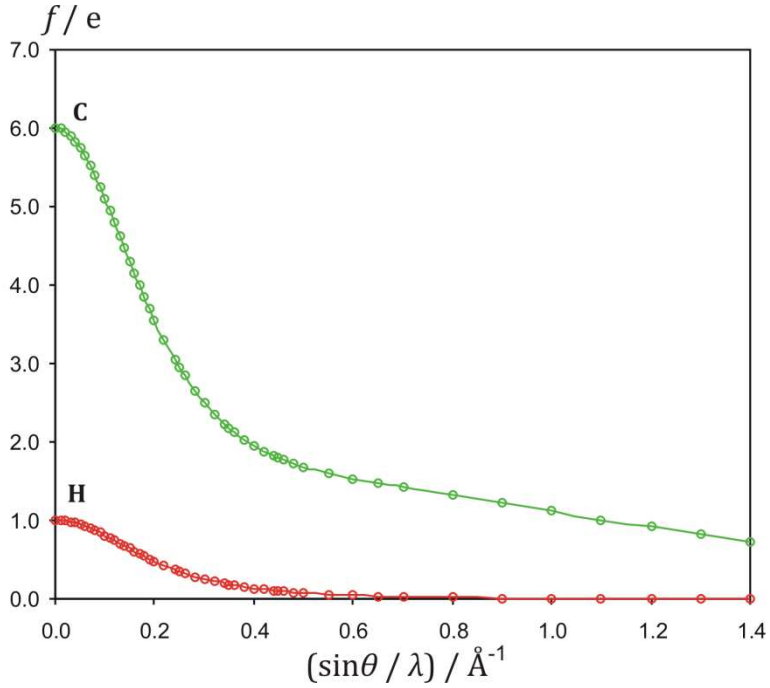


Figure 2.2. Typical spherical atomic scattering factor functions.

The IAM atomic scattering factors have been computed numerically according to the II.17 formula (using Hartree-Fock or Thomas-Fermi wave functions) and tabulated in the International Tables for Crystallography.[4] However, a useful, more approximate, representation of those functions has also been provided, and is shown below:

$$f\left(\frac{\sin\theta}{\lambda}\right) = c + \sum_{i=1}^{4} a_i \exp\left(-b_i \frac{\sin^2\theta}{\lambda^2}\right) \qquad \text{(II.19)}$$

where the a_i, b_i and c coefficients were fitted by means of least-square method and tabulated. This concise expression (II.19) allows for storing the atomic scattering factor information more compactly by means of the mentioned 9 coefficients per an atom type. Despite the great success of IAM, there are many serious drawbacks of this method. The most significant is the fact that it completely ignores the charge transfer between the atoms. One of the first solutions to this problem was proposed by Coppens *et al.*[5] within the developed κ-formalism. They introduced additional parameters to the description of the spherical atomic electron density:

$$\varrho_k(r) = P_{ck}\varrho_{ck}(r) + \kappa_k^3 P_{vk}\varrho_{vk}(\kappa_k r) \qquad \text{(II.20)}$$

where ϱ_{ck} and $\varrho_{vk}(\kappa_k r)$ are spherical one-electron normalized core and valence electron densities (for k-th atomic contribution), P_{ck} and P_{vk} are the population factors, and κj stays for the valence shell contraction-expansion parameter. That way, one can account for the charge transfer by adjusting the P_{vk} and κ_k parameters. At the early days of the crystallographic tool development, the κ-formalism was quite successful and provided useful chemical information about molecules in crystals.[5]

Another serious problem of IAM is its inability to describe chemical bonding features and lone electron pairs. Such a limitation causes systematic errors visible in the residual density maps evaluated for the best fit of the model to the data. It is clear though, that the experimentally-derived structure factor values contain the whole information (within a given method accuracy, and mathematical approximations) about the electron density distribution in a given crystal. Nevertheless, IAM constitute only a set of overlapping non-interacting atoms composing the studied chemical system. Therefore, charge density distribution asphericity is not accounted for by the model. Indeed, the residual density maxima are located in-between atomic positions and at places where electron lone pairs are found, as illustrated in Figure 2.3. It shows the need for a more sophisticated scattering formalism which would allow modelling of fine features of charge density distribution.

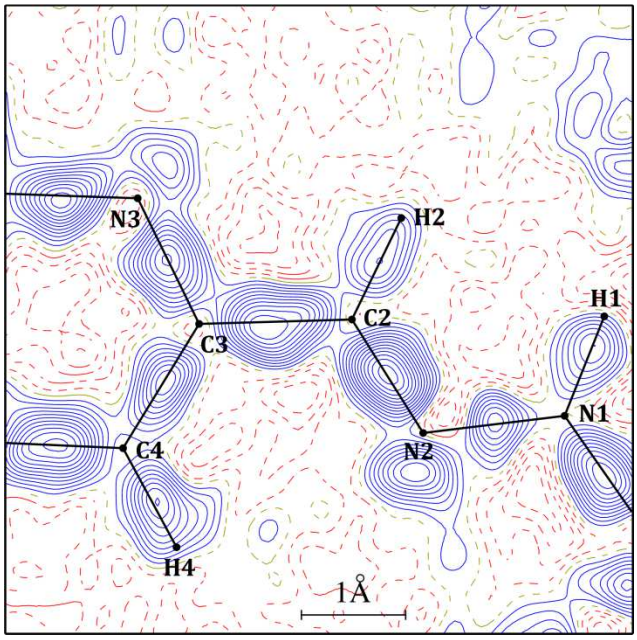


Figure 2.3. Exemplary Fourier residual density map obtained for the final IAM model (0.05 e·A^{-3} contours, blue – positive, red – negative).

II.1.6. Hansen-Coppens model

In reality, the studies of atom aspherical behaviour in the solid state date back to the year 1921, when W. L. Bragg was investigating the case of diamond.[6] Only by taking into account atomic deformation, he could explain the registered 'forbidden' 222 reflection. Similar effect was observed for the silicon structure. These examples show the importance of considering the atom asphericity when modelling a unit cell charge density distribution. The concepts presented by Kurki-Suonio,[7] Stewart,[8] Hirshfeld[9] and Coppens[10] resulted in the vast development of atom-centred formalisms based on the description of deformation terms via various functions. The most well-known approaches are the ones introduced by Hirshfeld, Stewart, and Hansen & Coppens. However, for the purpose of this thesis, only the last one shall be described in details.

In the Hansen-Coppens formalism,[10] the total atomic electron density (of the k-th atom) is the sum of three components:

$$\varrho_k(\mathbf{r}) = P_{ck}\varrho_{ck}(r) + \kappa_k^3 P_{vk}\varrho_{vk}(\kappa_k r) + \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} P_{lmk}\kappa_k'^3 R_{lk}(\kappa_k' r) d_{lmk}(\mathbf{r}/r) \qquad \text{(II.21)}$$

where ϱ_{ck} and ϱ_{vk} are the spherical one-electron normalized core and valence electron densities, respectively (the same as in the κ-formalism). The third term contains the sum of the angular functions (d_{lmk}), which model aspherical deformations. The angular functions d_{lmk} constitute real spherical harmonic functions normalized to the electron density:

$$\int |d_{lmk}| \mathrm{d}\theta \mathrm{d}\varphi = \begin{cases} 1 & \text{for} \quad l = 0 \\ 2 & \text{for} \quad l > 0 \end{cases} \qquad \text{(II.22)}$$

The direct form of the deformation functions is as follows:

$$d_{lmk} = N_{lmk} P_l^{|m|}(\cos\theta)\Phi_{mk}(\varphi) \qquad \text{(II.23)}$$

where $P_l^{|m|}$ constitute the associated Legendre polynomials, and Φ_{mk} is described as:

$$\Phi_{mk}(\varphi) = \begin{cases} \cos(m\varphi) & \text{for} \quad m > 0 \\ 1 & \text{for} \quad m = 0 \\ \sin(m\varphi) & \text{for} \quad m < 0 \end{cases} \qquad \text{(II.24)}$$

The coefficients P_{vk} and P_{lmk} in the formula II.21 stand for multipole population of the valence and deformation density multipoles, respectively. In turn, R_{lk} is the radial function which is defined as:

$$R_{lk}(r) = \frac{\zeta_{lk}^{n_{lk}+3}}{(n_{lk}+2)!} \exp(-\zeta_{lk} r) \qquad \text{(II.25)}$$

where ζ_{lk} and n_{lk} parameters are assigned to each element type separately. κ and κ' are scaling parameters, which control the expansion and contraction of the valence spherical and deformation densities, respectively. The deformation functions are visualized in Figure 2.4.

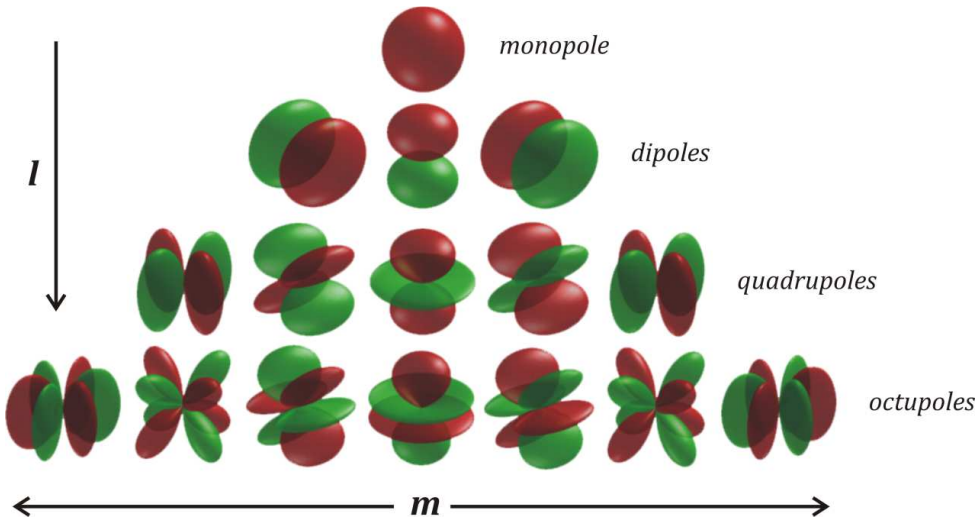


Figure 2.4. Visualization of various multipolar functions (picture created by *Sarxos*, made available for free download through Internet).

Substitution of such atomic electron density into the definition of an atomic scattering factor results in the following expression:

$$f_k(\mathbf{H}) = P_{ck} f_{ck}(\|\mathbf{H}\|) + P_{vk} f_{vk}(\|\mathbf{H}\|/\kappa_k) + 4\pi \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} P_{lmk} \mathrm{i}^l \langle j_{lk}(\|\mathbf{H}\|/\kappa'_k)\rangle d_{lmk}(\mathbf{H}/\|\mathbf{H}\|) \quad \text{(II.26)}$$

where $\langle j_l \rangle$ is the Fourier-Bessel transform defined as:

$$\langle j_{lk}(x)\rangle = \int_0^\infty r^2 R_{lk}(r) j_l(2\pi x r)\mathrm{d}r \quad \text{(II.27)}$$

Such functions can be evaluated numerically, or (preferably) analytically. When radial functions (or spherical electron densities) are expressed as a linear combination of Slater functions, the Fourier-Bessel integrals can be either computed with the aid of hypergeometric series as shown by Stewart,[11] or using the simple recursive procedure as presented by Avery & Watson and Su & Coppens.[12]

II.1.9. Structure and charge density refinement basics

When a properly designed model and the set of measured structure factors, the aim is to adjust the model parameters so, as to most accurately reflect the experimental data. A measure of the fit adequacy can be described in terms of the least square method, as the following residual function:

$$S[F] = \sum_{\mathbf{H}} (|F^{\mathrm{o}}| - |F^{\mathrm{c}}|)^2 \tag{II.28}$$

which is clearly the sum of squared differences between observed and calculated structure factors. To achieve the best fit, the above function needs to be minimised. Expanding such a residual function into the Taylor series truncated after all linear terms (*i.e.*, neglecting second derivatives), gives the set of normal equations in the matrix form:

$$\mathbf{A}\Delta\mathbf{X} = \mathbf{D} \tag{II.29}$$

where the appropriate matrices are constructed as follows:

$$a_{ij} = \sum_{\mathbf{H}} \left(\frac{\partial |F^{\mathrm{c}}|}{\partial x_i}\right)\left(\frac{\partial |F^{\mathrm{c}}|}{\partial x_j}\right) \tag{II.30}$$

$$d_i = \sum_{\mathbf{H}} (|F^{\mathrm{o}}| - |F^{\mathrm{c}}|)\left(\frac{\partial |F^{\mathrm{c}}|}{\partial x_i}\right) \tag{II.31}$$

Solution of these normal equations provides the parameter shift values $\Delta\mathbf{X}$:

$$\Delta\mathbf{X} = \mathbf{A}^{-1}\mathbf{D} \tag{II.32}$$

It is worth noting, that due to the non-linearity of the procedure only parameter shifts can be obtained. This is in contrast to the linear least-squares method where the best parameters can be evaluated directly. The whole process requires an iterative scheme, where the model is adjusted in small (linear) steps, hopefully towards improvement. The procedure is then repeated until convergence occurs. The matrix inversion in each step is most often performed via the Gauss-Jordan elimination or diagonalization techniques. However, there are many methods which pass around direct matrix inversion used in order to get the solution of normal equations (*e.g.* conjugate gradients).

The refinement of a crystal structure can be facilitated by using different residual functions. The one presented above is the simplest example of the so-called "refinement in respect to *F*". Nevertheless, as the intensity is proportional to the squared module of a structure factor, one could construct another residual function, such as, for example:

$$wS[F^2] = \sum_{\mathbf{H}} w_{\mathbf{H}}(|F^{\mathrm{o}}|^2 - |F^{\mathrm{c}}|^2)^2 \qquad \text{(II.33)}$$

where the reflection weights are additionally introduced. Minimization of such a residual function is based on F^2. The question whether the refinement should be performed on F or F^2 is still under discussion in the literature. Currently, more refinements are based on F^2. In the case of the studies presented in the thesis, the multipolar refinements were based on F. The tests performed on the ground of F^2 lead basically to the same results. The next issue is the choice of appropriate weight functions. When chosen properly, they should increase the convergence of the refinement procedure. Up to now, different weighting schemes were proposed and the choice is still not quite straightforward. For the purpose of this thesis, the simplest case was selected, *i.e.* $w_{\mathbf{H}} = 1/\sigma_{\mathbf{H}}^2$ (the $\sigma_{\mathbf{H}}$ is the estimated standard deviation of the intensity / structure factor). These weights are called 'statistical', and serve satisfactorily well in the case of multipolar refinements. Detailed description of the refinement procedure and its specificity is well presented in the publication by Watkin.[13]

During the charge density refinement, a set of additional parameters is refined together with classical ones (as summarized in Table 2.1). However, due to the complexity of the model and possible parameter correlations (especially when being far from the minimum) the convergence is usually very poor. Therefore, for the purpose of charge density studies, a very careful step-by-step approach is applied. Sometimes it is also desired to first get close to the real minimum by applying the so-called block refinement, as discussed in the literature (especially when many different effects are considered together, such as anharmonicity, extinction, multipole model *etc.*). For each case presented in this thesis, the refinement strategy is given in details in the appropriate section. The P_{00k} parameters were never refined, as they are strongly correlated with the $P_{\mathrm{v}k}$ ones.

Table 2.1. Summary of least-square variables.

Classical variables	*Charge density variables*
Scale factor (k_{S})	
Atomic coordinates (x, y, z)	*Valence shell parameters*
Harmonic thermal parameters (U_{iso}, U_{ij})	Population parameters (P_{v})
Extinction corrections (y_{exti})	Valence shell scaling factor (κ)
Occupancy parameters (η)	
Anharmonic parameters (C_{ijk}, D_{ijkl} *etc.*)	*Deformation functions*
Overall temperature factor (k_{T})	Deformation parameters (P_{lm})
Rigid body rotation angles (φ, χ, ω)	Contraction-expansion factors (κ')

II.1.10. Experimental intensity distribution

A reasonable interpretation of the refined model is very much dependent on the measured data quality. In general, the intensity of the scattered X-ray beam is described by the following formula:

$$I(\mathbf{H}) = I_0 \left(\frac{q_e^2}{m_e^2 c^2}\right)\left(\frac{\lambda^3 V_c}{V^2}\right) LPA(1+\alpha)|F(\mathbf{H})|^2 \qquad \text{(II.34)}$$

where I_0 is the intensity of the incident beam, V_c is the crystal volume, L and P are the so-called Lorentz and polarization factors, A is the transmission factor and α represents the thermal diffuse scattering (TDS) effect. F is the structure factor, already corrected for extinction, if necessary (Appendix A, Comment A1). The Lorentz and polarization corrections depend on the applied data collection strategy and the used instrumentation, and are nowadays automatically estimated by data reduction software. The first one accounts for various speeds of reflections passing through the Ewald sphere during the single exposure. In turn, polarization factor is related to the used goniometer geometry and radiation source, and corrects the scattered intensity for the polarisation effect in specific directions, caused by crystal.

The A factor, sometimes called the absorption correction, is calculated on the basis of the Lambert-Beer law:

$$I = I_0 \cdot \exp(-\mu x) \qquad \text{(II.35)}$$

stating that the beam intensity is attenuated by the filter proportionally to its thickness, x, by the adequate linear absorption coefficient, μ. Thus, the transmission factor in the case of the crystal looks as follows:

$$A = \frac{1}{V_c}\int \exp(-\mu[x(\mathbf{r}) + y(\mathbf{r})])\mathrm{d}\mathbf{r} \qquad \text{(II.36)}$$

The integration goes across the whole crystal volume and $x(\mathbf{r})$ and $y(\mathbf{r})$ are the incident and diffracted beams path lengths for the particular point $\mathbf{r}$ in some crystal orientation. The absorption correction can be calculated analytically or numerically. The numerical evaluation of this correction via a simple 3-dimensional Gaussian integration on the crystal grid was proposed by Coppens,[14] and it is the most frequently used method. Another popular approach that, in turn, does not employ the specified crystal shape was developed by Blessing.[15] This method is based on the least-squares fitting of spherical harmonics populations to the transmission surface by minimizing the differences in intensities of symmetry equivalent reflections.

The last important factor in the case of analysing the intensity distribution is the (1+α) factor. It corrects the diffracted beam intensity for the TDS effect resulting from the inelastic X-ray scattering. The TDS phenomenon causes the increase of certain reflection intensities, and is usually observed with high-resolution data.[16] Additionally, reflection profiles are significantly broadened. Adequate considering of TDS is extremely difficult, as there is no general method available. Most of the existing approaches require some external information about a given structure such as crystal elasticity constants. In addition, none of the methods is implemented in commercial data reduction software. Thermal diffuse scattering is significantly diminished at very low temperatures (say around 10 K). The effect of TDS is, though, most pronounced when heavy atoms (*e.g.* metal centres) are present in a structure. Nevertheless, in this thesis no systems containing heavy elements were analysed, and so TDS was considered as negligible.

II.1.11. Experimental conditions and instrumentation

In order to collect high quality (charge density) data, particular attention should be paid to all the phenomena and corrections presented above. When the corrections are large it decreases the measurement reliability. Therefore, below I present a few general recommendations which may increase the quality of the collected data:

- The size of a crystal should be as small as possible to reduce potential extinction and absorption problems. On the other hand, a crystal needs to diffract up to high diffraction angles, so the proper de-convolution of electron density features from the thermal motion can be achieved. Usually, the high-resolution term means that the resolution greater than 1.0 $Å^{-1}$ is achieved. The crystal sizes range from very small (50 μm linear size) for synchrotron sources to larger ones suitable for the beam size of a diffractometer (around 0.3 mm linear size).
- The most powerful X-ray sources are desired for performing very accurate charge density measurements. Although synchrotron radiation facilities provide high intensity X-ray beam with adjustable wavelength (particular ranges), additional problems may occur, *e.g.* some beam instability, problematic intensity corrections, *etc*. In the case of standard diffractometers, a rotating anode is a solution which significantly increases the beam intensity when compared to a sealed tube.
- An X-ray measurement should be carried out at the lowest possible temperature. In the case of nitrogen cooled samples, it is possible to go down to around 90 K,

whereas with liquid helium cooling setup temperatures of 10 K may be reached. Low temperature reduces thermal vibration, eliminates TDS, and thus increases the resolution limit, up to which the data can be collected.

- Reflection intensities need to be measured very accurately and precisely. This can be achieved with the aid of modern area CCD detectors. They additionally allow for a very fast data collection. However, such registered reflection positions are not as accurately determined as by point detectors. This can be, to some extent, overcome by thin slicing of 3D reflection profiles, such as every 0.1° or 0.25°. New detectors with super-fast readout based on the CMOS technology, which are being developed currently, may constitute a solution to these problems. However, at the moment there is no appropriate software to analyse CMOS data.
- Proper modelling of electron density features requires the best possible data quality. Therefore, a proper method of outlier rejection together with the high redundancy of the data set should be assured. Very robust methods of outlier treatment (and data scaling and merging) were, for example, developed by Blessing and implemented in the *SORTAV* program.[15] The issue of high data set redundancy depends on the scientist that needs to design a proper data collection strategy concerning measurement time and a goniometer setup. The redundancy at least around six is a suitable value to perform a proper statistical treatment of data.

II.1.12. Charge density refinement issues and limitations

Another important issue which is worth discussing concerns the analysis of high-resolution charge density quality data, and thus the restrictions and limitations of the electron density multipole refinement. In principle, the refinement of population parameters of the deformation functions should lead to the improved fit descriptors, thermal parameters etc. However, it is not always the case. Some precautions and recommendations are summarized below:

- As previously mentioned, the X-ray diffraction data should be characterized by high redundancy, but should also be complete. This is important regarding the Fourier series. The electron density distribution may be accurately reconstructed only when an infinite number of Bragg peaks is considered. This is, however, unachievable in practice. The incompleteness of the Fourier series leads to a set of the so-called truncation errors. These are associated with the limited resolution and

experimental restrictions of the goniometer geometry, and wavelength. In consequence, a broadened electron density distribution, especially near the nucleus position, is observed. Regardless the resolution, there is also the matter of missing data points across the whole range of the measured data. For example, absence of the low-resolution intense reflections usually leads to some systematic errors visible on the residual density maps. As a result, Fourier-space fitting techniques (*e.g.* considered least squares) cannot work properly as the number of highly contributing observables is not sufficient. The described effects are more significantly pronounced for structures containing heavy atoms (*e.g.* metal centres). Anyway, they should always be kept in mind, and so the collected data must be carefully analysed and interpreted at all times.

- Low-temperature techniques are crucial in the case of the multipole refinement. The proper de-convolution of the electron density features and thermal motion very much depends on the quality of the data, which usually improves with the lower experiment temperatures. However, the de-convolution is also strongly influenced by the used thermal motion model. Hydrogen atoms are most problematic considering position and thermal motion estimation on the basis of X-ray data. Fortunately, a neutron diffraction experiment, when available, provides more accurate hydrogen atom geometrical and thermal motion parameters. However, in most cases such measurement is not achievable (lack of beamtime, insufficient crystal size), and other approaches have to be employed. The most popular ones are based on the estimation of hydrogen atom anisotropic parameters via the TLS (Translation-Libration-Screw rigid body motion formalism) analysis[17] combined with either (*a*) theoretical determination of U_{ij}'s, or (*b*) library approach based on available neutron data. The latter one, implemented in the *SHADE* server,[18] has been utilized in this thesis when modelling the charge density sets. It was verified by Hoser *et al.*[19] that the proper charge density refinement of the X-ray data supplemented solely with the *SHADE* application leads to results comparable with neutron diffraction studies.
- Finally, it should be kept in mind that the Hansen-Coppens multipole model is not flexible enough to describe all features of the electron density simultaneously. Its behaviour was extensively studied and discussed in the literature. The conclusions were the following: (*a*) differences between theory and experiment are probably

devoted to the lack of flexibility of radial functions (there is usually only one radial function per atom); (*b*) better multipolar fit is obtained with dynamic structure factors than with the static ones. Thus, it is clear that the multipole model cannot describe properly the density in the bonding region and near the nuclei simultaneously (due to the truncation errors as well). To fix these problems. several models are being developed, such as wavefunction fitting[20] (quantum-mechanical constrained static density), κ_c-refinement[21] (additional scaling parameter for the core density), or stockholder radial function derivation.[22] Despite their success, they are not well tested. It is still under discussion whether the experimental data refinement is capable of recovering very subtle charge density effects. Anyhow, one needs to bear in mind that the model of electron density has to be reasonable, regardless the fit. The improper fit may be caused by other effects such as extinction, absorption, twinning, crystal imperfections and many other, which need to be carefully considered at all times.

II.1.13. Charge density topological analysis

(i) Atomic basins and critical points. The electron density distribution can be analysed in various ways. One of the possibilities is to apply the so-called Quantum Theory of Atoms In Molecules (QTAIM) developed by Bader and co-workers.[23] Unlike many other theories of chemical bonding, QTAIM involves a discrete boundary partitioning scheme based on a topology of electron density. Another unique feature of the theory is that the electron density distribution can be analysed directly, regardless its source – the theory or experiment. However, it has to be kept in mind that having solely the electron density, it is only possible to evaluate one-electron properties on that basis.

At each point of space the gradient of the electron density function may be computed:

$$\nabla\varrho(\mathbf{r}) = \frac{\partial\varrho}{\partial x}\mathbf{e}_x + \frac{\partial\varrho}{\partial y}\mathbf{e}_y + \frac{\partial\varrho}{\partial z}\mathbf{e}_z \qquad \text{(II.37)}$$

where $\mathbf{e}_i$ stands for the basis vector of the Cartesian system in *i*-th direction. The gradient path indicates the existence of the so-called atomic basins (Figure 2.5).

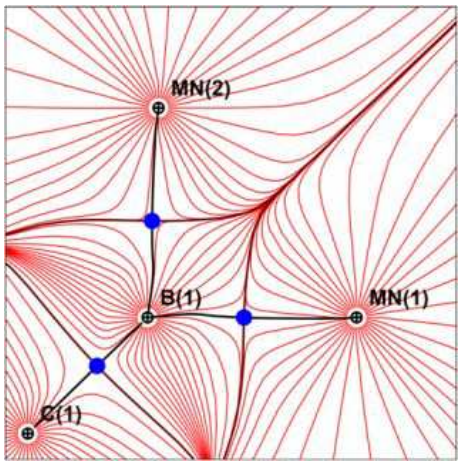


Figure 2.5. Examplary atomic basins of boron, carbon and manganesium atoms derived on the basis of the QTAIM approach (picture taken from the literature).[24]

The atomic basins are separated by the Zero Flux Surfaces (ZFS's) for which the following condition applies:

$$\nabla\varrho(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) = 0 \tag{II.38}$$

where $\mathbf{n}(\mathbf{r})$ is the normal to the surface. Points in space where the gradient of the electron density is equal to zero:

$$\nabla\varrho(\mathbf{r}) = \mathbf{0} \tag{II.39}$$

are called the critical points (mathematically speaking – stationary points). These can be classified by means of the Hessian matrix:

$$H(\mathbf{r}) = \begin{bmatrix} \frac{\partial^2\varrho}{\partial x^2} & \frac{\partial^2\varrho}{\partial x\partial y} & \frac{\partial^2\varrho}{\partial x\partial z} \\ \frac{\partial^2\varrho}{\partial x\partial y} & \frac{\partial^2\varrho}{\partial y^2} & \frac{\partial^2\varrho}{\partial y\partial z} \\ \frac{\partial^2\varrho}{\partial x\partial z} & \frac{\partial^2\varrho}{\partial y\partial z} & \frac{\partial^2\varrho}{\partial z^2} \end{bmatrix} \tag{II.40}$$

The $H(\mathbf{r})$ eigenvalues (λ_1, λ_2, λ_3) enable to classify the critical points by means of ranks and signatures. The rank is the number of non-zero eigenvalues, and signature is defined as a sign of the sum of all eigenvalues:

$$\sigma = \sum_{i=1}^{3} \mathrm{sgn}(\lambda_i) \tag{II.41}$$

In the case of 3-dimensional systems, only well-defined cases possible which are summarized in Table 2.2. The example distribution of critical points is shown in Figure 2.6 (molecular graph).

Figure 2.6. Exemplary molecular graph resulting from the QTAIM analysis.

Bond critical points (BCPs) appear in-between atoms, which are thus considered to be interacting. The number of each type of critical points reflects the correctness of the topology of the electron density. This relationship is described by the Poincaré-Hopf rule as:

$$n_{(\mathrm{N})\mathrm{NA}} - n_{\mathrm{BCP}} + n_{\mathrm{RCP}} - n_{\mathrm{CCP}} = 1 \tag{II.42}$$

for molecular, and:

$$n_{(\mathrm{N})\mathrm{NA}} - n_{\mathrm{BCP}} + n_{\mathrm{RCP}} - n_{\mathrm{CCP}} = 0 \tag{II.43}$$

for periodic systems, where n_{CP} denotes the number of each type of critical points, *i.e.* BCP – bond critical points, BCR – ring critical points, and CCP – cage critical points, whereas $n_{(\mathrm{N})\mathrm{NA}}$ is the number of (non-)nuclear attractors.

Table 2.2. Electron density critical point types.

CP type	*Acronym*	sgn(λ_1)	sgn(λ_2)	sgn(λ_1)	(rank,signature)
(Non-)Nuclear Attractor	(N)NA	–	–	–	(3,–3)
Bond Critical Point	BCP	–	–	+	(3,–1)
Ring Critical Point	RCP	–	+	+	(3,+1)
Cage Critical Point	CCP	+	+	+	(3,+3)

(ii) Bonding analysis. Instead of providing a simple image of the bonding situation in the considered molecular system, QTAIM gives a full spectrum of quantitative information. The trace of the Hessian matrix is of course the Laplacian of the electron density at a given point:

$$\nabla^2\varrho(\mathbf{r}) = \mathrm{tr}H(\mathbf{r}) = \frac{\partial^2\varrho}{\partial x^2} + \frac{\partial^2\varrho}{\partial y^2} + \frac{\partial^2\varrho}{\partial z^2} \tag{II.44}$$

The electron density value and Laplacian calculated at BCPs help to explain the nature of these particular interactions. The existing possibilities are summarized in Table 2.3. Some scientists analyse Laplacian values, other negative Laplacian values. As in my

thesis I utilise the Laplacian value without changing its sign, I shall present all remarks in this convention. Thus, electron sharing characteristic for the formation of the so-called covalent bond is described by a rather large electron density value at the bond critical point, and a very negative value of the corresponding $\nabla^2\varrho$. On the other hand, the interactions with positive values of Laplacian and small electron density are considered as the closed-shell ones. The latter case is especially common for the ionic interaction type, however, not only. Such situation occurs also in the case of metal…metal bonds.

Table 2.3. Classification of bonding types in terms of charge density properties.

Property	*Shared shell (covalent and polar bonds)*	*Closed shell (ionic bonds, weak interactions)*
$\varrho(\mathbf{r}_{\mathrm{BCP}})$	large	small
λ_i	$\lambda_{1,2}$ dominant	λ_3 dominant
$\nabla^2\varrho(\mathbf{r}_{\mathrm{BCP}})$	< 0	> 0

The spatial distribution of Laplacian of the electron density gives another useful descriptor enabling the system quantification. It appears that the local minima of $\nabla^2\varrho$ indicate regions of increased charge concentration. The (3,-3) points of the Laplacian distribution are called charge concentrations (CCs) and are divided into bonding (BCCs) and non-bonding ones. The latter ones, when attributed to the atomic valence shell, represent well the electron lone pairs from the classical theory. The analysis of CCs is particularly useful in the case of metal complexes, where it shows the distribution of the electron cloud and explains its features.

Furthermore, there is another useful parameter based on the Hessian matrix. When one is able to calculate Hessian matrix eigenvalues at BCP, the property called ellipticity ε, defined as follows:

$$\varepsilon = \frac{\lambda_1}{\lambda_2} - 1 \qquad \text{(II.45)}$$

describes how the bond electron density is deformed from spherical shape in perpendicular directions. Such deformation coincides well with the single and double bond character in the classical theory. The larger is the ellipticity, the larger the double character of a given bond. Of course, all intermediate values can be also observed and then they indicate some electron delocalisation.

Finally, ZFS's defining the subsystems can be used for the integration of various properties. The most natural and simple, while being very useful, is the atomic charge defined as:

$$Q_k = Z_k - \int_{\Omega_k} \varrho(\mathbf{r})\mathrm{d}\mathbf{r} \tag{II.46}$$

where Z_k is the atomic number of the k-th atom, and Ω_k defines the corresponding atomic basin. QTAIM atomic charges usually reflect very well the charge transfer between atoms. Obviously, not only the electron density may be integrated, but also other functions, such as:

$$\boldsymbol{\mu}_k = \int_{\Omega_k} \mathbf{r}\varrho(\mathbf{r})\mathrm{d}\mathbf{r} \tag{II.47}$$

providing the atomic dipole moment $\boldsymbol{\mu}_k$. The total molecular dipole moment is then calculated as:[25]

$$\boldsymbol{\mu} = \sum_k (Q_k \mathbf{r}_k + \boldsymbol{\mu}_k) \tag{II.48}$$

(iii) Energetic features derived from experiment. Charge density distribution naturally allows deriving electrostatic properties of molecules. This can be for example electrostatic potential, *ESP*, which I am going to discuss while analysing charge density datasets. Electrostatic potential can be evaluated at a given point directly from the electron density distribution on the basis of the following formula:

$$ESP(\mathbf{r}) = \sum_k \frac{Z_k}{\|\mathbf{r} - \mathbf{r}_k\|} - \int \frac{\varrho(\mathbf{r})}{\|\mathbf{r} - \mathbf{r}'\|} d\mathbf{r}' \tag{II.49}$$

where Z_k is the nuclear charge of the k-th atom.

Additionally, on the basis of QTAIM, the kinetic energy density, abbreviated as G, respectively, can be estimated. Abramov[26] proposed an approximate formula linking the charge density topological features with the kinetic electronic energy density at a given point r given by the following equation:

$$G(\mathbf{r}) = \frac{3}{10}(3\pi^2)^{2/3}\varrho^{5/3}(\mathbf{r}) + \frac{1}{72}\frac{\|\nabla\varrho(\mathbf{r})\|^2}{\varrho(\mathbf{r})} + \frac{1}{6}\nabla^2\varrho(\mathbf{r}) \tag{II.50}$$

However, a good agreement between the above expression and Hartree–Fock calculations of G is obtained only in the medium-range region, *i.e.* for distances in the range from 0.5 Å to 2.1 Å from the atomic nuclei. Therefore, Espinosa *et al.*[27] considered

the above expression valuable for closed-shell interactions such as hydrogen bonds. As the electron density, independently within each separate atomic QTAIM basin, obeys the virial theorem ($-V/G = 2$, where V denotes potential energy density), basing on the II.50 formula potential energy density, V, can be derived. In the Espinosa's approach, the weak interaction energy of a particular intermolecular HB contact is considered as being equal to approximately half of the estimated potential energy density value at the corresponding BCP:

$$V(\mathbf{r}_{\mathrm{BCP}}) = \frac{\hbar^2}{4m_{\mathrm{e}}}\nabla^2\varrho(\mathbf{r}_{\mathrm{BCP}}) - 2G(\mathbf{r}_{\mathrm{BCP}}) \tag{II.51}$$

$$E_{\mathrm{HB}} = \frac{1}{2}V(\mathbf{r}_{\mathrm{BCP}}) \tag{II.52}$$

where E_{HB} is the hydrogen bond energy, and the scale factor is given in cubic atomic units. This constitutes a natural linkage between the charge density distribution features and computational energetic considerations, which shall be presented later in the result analysis part.

II.2. Pseudoatom databanks – idea, construction and applications

II.2.1. Idea behind pseudoatom databanks

Experimental determination of charge density distribution is a difficult and complex task. Additionally, it is not always possible to find a crystal suitable for such studies. Diffraction data quality is often not good enough to obtain reliable charge density results. The confidence in experimental charge density might be compromised either by experimental errors, multipole pseudoatom model limitations, or by lack of accurate phases and large uncertainties of hydrogen atom positions and thermal motion. This is especially the case with macromolecules, these being of great importance from biological point of view, but generally hard to model, also by the means of computational chemistry. Thus, a question appeared whether it was possible to reconstruct charge density of large systems (of known geometry) on the basis of the available data for small molecules, providing the starting point for evaluating electrostatic properties and wider analysis.

The other concern was pertained to the subject of atomic scattering factors. It was clear that IAM was no longer sufficient. Evidently, introducing asphericity to an atom model, which better reflects the reality, resulted in much better determined molecular geometries, close to the neutron measurement results. This was supported by numerous charge density studies. When having reasonable high-resolution data, it is possible to more accurately deconvolute the static density and the atomic thermal motion. Would it also be possible if such data were not available?

The breakthrough came in 1991, when Brock *et al.*,[28] introduced the new idea of transferability of atomic multipolar parameters (pseudoatom charge density models) between different molecules. Indeed, it appeared that atoms with similar chemical environment usually do not differ much in terms of charge density description, when switching from one moiety to another. These observations initiated the creation of databanks of aspherical atom models. A common concept was to store the charge density information of an atom type in the Hansen-Coppens formalism,[10] i.e. by means of multipolar population values and κ and κ' parameters. As this is not sufficient, the atom type definition should also contain atom chemical characterisation (element type, closest chemical neighbourhood *etc.*) and the assigned unambiguous local coordinate system. Furthermore such a databank needs to be accompanied by a program properly

transferring the adequate pseudoatom parameters to a chosen structure. The remaining flexibility in the databank construction concerns first of all the method of multipolar parameter evaluation and the exact atom type definitions.

II.2.2. Application of the existing pseudoatom databanks

Up to now, there are three well established databanks available. These are the experimental databank **ELMAM/ELMAM2**,[29] the theoretical **Invariom** database,[30] and the middle option- the University at Buffalo Pseudoatom Databank (**UBDB**).[31]

First of all such pseudoatom databases offer a possibility to perform structure refinement with the use of **aspherical atomic scattering factors** computed from the Transferable Aspherical Atom Model (**TAAM**).[32] This constitutes an improvement over the extensively applied Independent Atom Model (IAM) refinement, which does not include atomic charge density deformations due to bond formation, or lone pairs. It is, therefore, possible to model electron density distribution, and then, to deconvolute thermal motion more accurately for typical X-ray data ($\sin\theta/\lambda \leq 0.7$ Å^{-1}).[29c, 32b, 33] Such Atomic Displacement Parameter (ADP) values are consequently closer to those obtained from multipole refinements of high resolution X-ray data.[32a, 33-34] It was also already shown that the TAAM refinement significantly improves the discrepancy *R*-factors, molecular geometry[32a, 33-35] and precision of the Flack parameter determination[35a] when compared to the standard method based on IAM.

Secondly, the aforementioned databanks can be employed **to reconstruct the electron density distribution** of macromolecules, and further, to estimate the electrostatic properties of such complex systems.[36] According to Bąk *et al.*,[32a] electrostatic interaction energies (E_{es}'s) computed on the basis of each database model are closer to the results obtained theoretically for isolated molecules than to those derived from periodic calculations. The smallest differences in the E_{es} values with respect to *ab-initio* results and the highest correlations were found for the UBDB database.

Moreover, the fast progress in the quality of X-ray measurements of biomolecules has entailed the idea of more precise structure refinement also in the case of biological systems. Therefore, currently efforts are being made to incorporate the UBDB databank, which would serve as a source of aspherical atomic factors, into the *PHENIX* code[37] widely used by protein crystallographers.

Finally, in this paragraph, it is worth mentioning that pseudoatom databanks can be applied as a **starting point in charge density** studies.[38] A multipolar refinement final result might be affected by initial parameters, though it should not be. Therefore, a databank is a good source of a reliable first charge density model, which then helps to overcome the problem of the refinement method inadequacy.

II.2.3. Description and comparison of the existing pseudoatom databanks

This section gives a brief description of the construction method of each databank, to emphasize the similarities, and differences among them. Such database overview would be helpful to understand the specificity of UBDB databank, development of which constitutes a substantial part of my PhD thesis, in a wider context.

(i) ELMAM/ELMAM2.[29] Both ELMAM-type pseudoatom libraries are based on experimentally evaluated charge density parameters. In the case of the ELMAM database approach, high quality and homogeneous (in terms of data collection conditions and refinement strategy) amino-acid, and small peptide charge density models serve as a source of multipolar parameters. These values are averaged over a family of statistically similar atom types among the selected structures. The peptide charge-density refinements are performed according to a common general strategy. During such a multipolar refinement, local atomic site symmetry constrains are applied, and X–H bond lengths are restrained to the standard neutron values (or the corresponding values when neutron data is available). For non-hydrogen atoms multipolar parameters are refined up to octupoles, whereas in the case of hydrogen atoms only monopoles and bond directed dipoles are refined. The κ and κ' parameters are refined separately for all atom types. All multipolar refinements are performed in the *MOPRO* program .[35b] Among the differences between ELMAM and its new version, ELMAM2, is the scope of the initially selected molecules and the multipolar expansion limit. In the ELMAM2 database approach the charge density parameter source was extended from protein atom types to all common organic molecules for which reliable high-resolution studies are possible. The other difference concerns the hydrogen atom treatment. Here, apart from monopoles and bond directed dipoles, also bond directed quadrupolar parameters are refined. Additionally, for non-hydrogen atoms, not heavier than fluorine atom, multipolar parameters are expanded up to octupoles, and up to hexadecapoles

otherwise. Additionally, some changes were applied to the local coordinate systems being assigned to atom types, and the multipolar parameter averaging procedure.

(ii) Invariom database. [30] The Invariom database is purely theoretical. In order to obtain a particular atom type, first the smallest possible molecule which mimics the desired chemical environment is chosen. Then its geometry is optimised (so the X–H bond lengths) and charge density evaluated at the B3LYP/D95++(3df,3pd) level of theory. Subsequently, theoretical structure factors (cubic cell with a = 30 Å, space group $P\bar{1}$) are calculated[39] and a multipole refinement with the *XDLSM* module from the *XD* suite[40] is performed. At the multipolar refinement stage local atomic site symmetry constrains are applied. Multipole parameters are refined up to hexadecapolar level for all atoms, however, in the case of hydrogen atoms these are restricted only to bond directed multipoles. The κ parameters are refined separately for all atom types, while κ' parameters are refined solely for non-hydrogen pseudoatoms. In contrast to ELMAM, multipolar refinements for the purpose of the Invariom databank are performed in the *XDLSM* program.

(iii) University at Buffalo Databank. [31] University at Buffalo Databank (UBDB) is a theoretical databank based on experimental geometries with the neutron-normalised X–H distances. For that purpose sufficient quality experimental molecular geometries are selected from the Cambridge Structural Database (CSD) resources.[41] Molecular wave functions are calculated at the B3LYP/6-31G** level of theory. Complex static valence-only structure factors in the range $0.0 \leq \sin\theta/\lambda \leq 1.1$ Å^{-1} are derived, similarly to these of Invarioms, by analytic Fourier transform of the molecular charge densities for reciprocal lattice points corresponding to a pseudo-cubic cell with 30 Å edges.[42] Subsequently, they are fitted with the Hansen-Coppens pseudoatom formalism,[10] using the *XD* program suite.[40] For the purpose of the refinement, theoretically derived structure factor phases are used, which increase the adequacy of the model fitted to the data of non-centrosymmetric structures. During the multipolar refinement procedure local atomic site symmetry constrains are applied. Derived multipole parameters together with κ and κ' values are averaged over a family of chemically similar atoms with an eye on their statistical consistence, leading to a particular atom type stored in the databank. Multipoles are refined up to hexadecapolar level in the case of non-hydrogen atoms. For hydrogen atoms solely monopoles and bond directed dipoles and

quadrupoles are refined. The κ and κ' parameters are refined separately for all atom types.

(iv) Databank comparison. In general, all the databanks behave similarly well as a source of aspherical scattering factors, constituting a significant improvement in reference to the IAM method. Additionally, Johnas *et al.*[43] compared UBDB and Invariom databanks together and did not find any significant differences between the derived electron density distributions (bond critical point properties *etc.*). Electrostatic properties, however, are much more sensitive to any changes in the reconstructed molecular charge densities, so any differences are more pronounced. Consequently, the accuracy of pseudoatom-based electrostatics of macromolecular systems is much more problematic. Bąk *et al.*[32a] showed that for peptides the smallest differences in the E_{es} values with respect to *ab-initio* results and the highest correlations were found for the UBDB database. Nevertheless, the deviations from the theoretical values reached about 15 kJ·mol^{-1}.

Another aspect concerns direct consequences of a databank construction method. The advantages and disadvantages of experimental versus theoretical databases have been discussed in details by Pichon-Pesme *et al.*[29c] and Volkov *et al.*[31b, 33] The absence of experimental errors and thermal motion convoluted with electron density constitute the major advantage of theoretically derived databanks. What is more, theoretical pseudoatom libraries might be easily extended by unlimited number of new atom types, as this requires simple single point calculations. At the same time, the experimental databases are limited to the available charge density data of organic molecules. On the other hand, ELMAM-type databanks were meant to address the influence of crystal filed on charge density parameters. This is surely impossible in the case of Invarioms and doubted in the case of UBDB, where the experimental molecular geometries might provide some information on the crystal field effect. It remains, however, debatable, if the charge density parameters averaged over a family of similar atoms may still keep the crystal field influence, and if so – to what extent?

II.2.4. Practical remarks

This paragraph is devoted to practical aspects regarding pseudoatom databank applications. I found it relevant to briefly describe the TAAM refinement bacjground, as it is tightly related to the scope of my PhD project. For the same purpose, it is desirable

to present methods which enable electrostatic property evaluation on the basis of molecular charge densities reconstructed with the aid of pseudoatom libraries. Consequently, I included a note on the Exact Potential and Multipole Method (EPMM) used for electrostatic interaction energy estimation, being applied during the course of my studies, both for the UBDB databank testing and neuraminidase active site analysis.

(i) Transferable Aspherical Atom Model refinement. As mentioned before, Transferable Aspherical Atom Model (TAAM) refinement method is based on aspherical atom charge density models kept in the Hansen-Coppens formalism.[10] Such models imitate real atoms more reliably than the spherical approximation approach (IAM), and so may serve as more appropriate atomic scattering factors.

The overall procedure consists of three main steps. Firstly, the IAM method is applied in the low-resolution experimental data refinement ($0.0 \leq \sin\theta/\lambda \leq 0.7$ Å^{-1}) to obtain initial molecular geometry and atomic displacement parameters for the purpose of the subsequent TAAM approach. In the IAM refinement, the scale factor, atom coordinates, anisotropic atomic displacement parameters of non-hydrogen atoms and isotropic atomic displacement parameters of hydrogen atoms, are being refined.

Then, with the obtained molecular geometry, adequate pseudoatom multipolar parameters can be either transferred from an independent source such as a database, or calculated theoretically. The transfer of the pseudoatom parameters from the particular aspherical atom type library, and the local coordinate system assignment (characteristic for the database) can be accomplished with the Invariomtool in the case of Invariom database,[44] by using the *LSDB* program in the case of UBDB,[31b, 45] and applying the *MOPRO* suite[35b] designed for ELMAM and ELMAM2. To ensure the molecular electroneutrality (or in general impose a given charge), the monopole populations of the pseudoatoms are scaled *a posteriori* by distributing the same fraction of charge difference over all the atoms in the molecule, as it designed for ELMAM and Invariom databases, or by employing the Faerman & Price method[46] in the case of the ELMAM2 and UBDB databases.

A structural refinement performed with these transferred aspherical atomic scattering factors constitutes the last step. During a standard TAAM refinement procedure, an overall scale factor, atomic positions, and ADPs are refined with reference to experimental data, while the multipolar populations and kappa values are fixed. In general, such a refinement might be performed within the *XD* package (*XDLSM*

module),[40] *JANA* program,[47] or *MOPRO* suite.[35b] However, it is much easier to use either *MOPRO* or *XD*, as programs applied for the databank transfer provide suitable input files. TAAM refinement in *JANA* requires an input file modification from either the *XDLSM* or *MOPRO* format to the appropriate one.

(ii) Exact Potential Multipole Method.[48] The Exact Potential Multipole Method (EPMM) approach was designed for the purpose of relatively effective and accurate electrostatic interaction energy evaluation on the basis of molecular charge density model. It combines the very fine Exact Potential method with the much more approximate approach based on atomic moment expansion. Within the first method the electrostatic energy between two unperturbed molecular charge distributions is defined by the exact potential (EP):

$$E_{\mathrm{elst}}^{\mathrm{EP}} = \sum_{a\in\mathrm{A}}\sum_{b\in\mathrm{B}} \frac{Z_a Z_a}{r_{ab}} + \int \varrho_{\mathrm{A}}(\mathbf{r}_{\mathrm{A}}) V_{\mathrm{A}}^{\mathrm{nuc}}(\mathbf{r}_{\mathrm{A}}) \mathrm{d}\mathbf{r}_{\mathrm{A}} + \int \varrho_{\mathrm{B}}(\mathbf{r}_{\mathrm{B}}) V_{\mathrm{B}}^{\mathrm{nuc}}(\mathbf{r}_{\mathrm{B}}) \mathrm{d}\mathbf{r}_{\mathrm{B}} + \\ + \iint \frac{\varrho_{\mathrm{A}}(\mathbf{r}_{\mathrm{A}}) \varrho_{\mathrm{B}}(\mathbf{r}_{\mathrm{B}})}{\|\mathbf{r}_{\mathrm{A}} - \mathbf{r}_{\mathrm{B}}\|} \mathrm{d}\mathbf{r}_{\mathrm{A}} \mathrm{d}\mathbf{r}_{\mathrm{B}} \tag{II.53}$$

where ϱ_{A} and ϱ_{B} are the molecular electron densities, and $V_{\mathrm{A}}^{\mathrm{nuc}}$ and $V_{\mathrm{A}}^{\mathrm{nuc}}$ are the nuclear potentials of molecules A and B, respectively.

The second method is based on atomic moments obtained from different types of electron density partitioning, allowing also Hansen-Coppens multipolar model, using a Buckingham-type (multipole, MM) approximation:[49]

$$E_{\mathrm{elst}}^{\mathrm{MM}} = \sum_{a\in\mathrm{A}}\sum_{b\in\mathrm{B}} T q_a q_b + T_\alpha \left(q_a \mu_{\alpha,b} q_b \mu_{\alpha,a}\right) \\ + T_{\alpha\beta}\left(\frac{1}{3} q_a \Theta_{\alpha\beta,b} + \frac{1}{3} q_b \Theta_{\alpha\beta,a} - \mu_{\alpha,a}\mu_{\alpha,b}\right) \\ + T_{\alpha\beta\gamma}\left(\frac{1}{15} q_a \Omega_{\alpha\beta\gamma,b} - \frac{1}{15} q_b \Omega_{\alpha\beta\gamma,a} - \frac{1}{3}\mu_{\alpha,a}\Theta_{\alpha\beta,b} \right. \\ \left. + \frac{1}{3}\mu_{\alpha,b}\Theta_{\alpha\beta,a}\right) + \cdots \tag{II.54}$$

where T, T_α, $T_{\alpha\beta}$ *etc.* are the interactions tensors, and q, μ_α, $\Theta_{\alpha\beta}$ *etc.* are the atomic multipole moments.

In the Exact Potential / Multipole Method, the exact Coulomb integral for pseudoatom–pseudoatom interactions in the inner sphere (for the overlapping electron densities, short distance atom-atom interactions) is evaluated numerically, whereas to estimate the interactions in the outer region (long distance atom-atom, or even

molecule-molecule interactions) the Buckingham-type multipole approximation is applied. Although this multipole approximation is not valid for overlapping charge distributions, it has been already shown that it works well for longer distances. The unquestionable advantage of this method is its low computational cost, and relatively short calculation time.

When having a multipolar charge density model of a studied system, corresponding E_{es} (EP) can be evaluated on the basis of the psedudoatom charge distributions. The E_{es} (EP) between the two pseudoatom charge distributions is evaluated by numerical (quadrature) integration over both pseudoatoms. The 6D integral in the last term of II.53 can be much simplified when rewritten as:

$$\begin{aligned}\iint \frac{\varrho_a(\mathbf{r}_a)\varrho_b(\mathbf{r}_b)}{\|\mathbf{r}_a-\mathbf{r}_b\|}\mathrm{d}\mathbf{r}_a\mathrm{d}\mathbf{r}_b &= \int \varrho_a(\mathbf{r}_a)V_b^{\mathrm{elec}}(\mathbf{r}_a)\mathrm{d}\mathbf{r}_a \\ &= \int \varrho_b(\mathbf{r}_b)V_a^{\mathrm{elec}}(\mathbf{r}_b)\mathrm{d}\mathbf{r}_b\end{aligned} \tag{II.55}$$

where V_a^{elec} and V_b^{elec} are the electronic potentials of pseudoatoms a and b, respectively. This leads to integrals analogous to the ones in the second and third terms of equation II.53, which can be evaluated in a single integration step.[50] Calculating of both integrals in the II.55 formula is considered as a good test of the method accuracy.

The Exact Potential / Multipole Method (EPMM) is implemented in the *XDPROP* module of the *XD* package.[40] It combines numerical evaluation of the exact Coulomb integral in the inner region (typically ≤ 4.5 Å) with the Buckingham-type multipole approximation for the long-range interatomic interactions.[49]

II.3. Theoretical calculations

The aim of this chapter is to present the foundations of quantum chemical theory and calculations. Additionally, I decided to provide some basic information about the *SPDFG* and *CRYSTAL* programs, which I used during the course of my studies.

II.3.1. Principles

Investigating, and understanding the structure of materials at the subatomic level requires the use of mathematically rigorous framework of quantum mechanics. The way a quantum state of a physical system changes in time may be described by Schrödinger equation. In its most general form in the non-relativistic approach to quantum mechanics, the equasion is as follows:

$$\hat{H}\Psi(\mathbf{r},t) = \mathrm{i}\hbar\frac{\partial\Psi(\mathbf{r},t)}{\partial t} \tag{II.56}$$

where $\hat{H}$ is the Hamilton operator, and Ψ is a wavefunction. Assuming that the system does not change with time, the above formula may be converted to the time-independent form:

$$\hat{H}\Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \tag{II.57}$$

in which E is the energy of a system for a given Ψ. In case of molecules, the Hamilton operator is the sum of kinetic and potential energy operators (in atomic units):

$$\begin{aligned}\hat{H} &= T_\mathrm{e} + V_\mathrm{ee} + V_\mathrm{en} + V_\mathrm{nn} = \\ &= -\frac{1}{2}\sum_k \nabla_k^2 + \sum_k\sum_{m>k}\frac{1}{\|\mathbf{r}_k - \mathbf{r}_m\|} - \sum_k\sum_a\frac{Z_a}{\|\mathbf{R}_a - \mathbf{r}_k\|} + \\ &+ \sum_a\sum_{b>a}\frac{Z_aZ_b}{\|\mathbf{R}_a - \mathbf{R}_b\|}\end{aligned} \tag{II.58}$$

where T_e is the kinetic energy operator for electrons, and V_ee, V_en and V_nn are the electron-electron, electron-nuclei and nuclei-nuclei potential energy operators, respectively. Solution of the Schrödinger equation provides detailed information regarding properties of molecules and solids. However, to date, only few exact solutions of Schrödinger equation are known. As it is usually a very complex and difficult task to solve the equation, some simplifications need to be allowed for. The most widely applied is the Born-Oppenheimer approximation, which assumes that the mass ratio of nuclei to electrons is large enough for the chemical system to be treated as a set of static nuclei and moving electrons. This approximation provides the foundation for such concepts as

"molecular structure", or "electronic structure". Further approximations usually concern the form of the wavefunction (the basis set used) describing the analysed chemical/physical system, *e.g.*, Hartree-Fock single Slater determinant approximation. One should also keep in mind that, typically, relativistic effects are completely neglected. The aspects which need to be additionally taken into account are the electron correlation energy and dispersive interaction. The initial Hartree-Fock method does not include the electron correlation energy and dispersion interaction, thus a set of the so-called post-Hartree-Fock methods has been developed (*e.g.*, configuration interaction, coupled clusters, or Møller-Plesset perturbation theory). The main goal is to evaluate the most accurate ground state electronic structure (described by the wavefunction) and the total energy of interacting electrons for a given nuclear configuration.

II.3.2. Density Functional Theory

On the basis of electronic wavefunction, it is possible to define various useful chemical and physical properties of a quantum system. Among the most important ones is the electron density, which is defined as follows:

$$\varrho(\mathbf{r}) = N \sum_{\sigma_1} \int |\Psi(\mathbf{r}_1, \sigma_1, \dots, \mathbf{r}_N, \sigma_N)|^2 \mathrm{d}\boldsymbol{\tau}_2 \dots \mathrm{d}\boldsymbol{\tau}_\mathrm{N} \tag{II.59}$$

for an N-electron system, where the integration goes over all electron coordinates but one, and all spin coordinates. In this sense, the electron density distribution can be interpreted as a probability density of finding an electron at a given point $\mathbf{r}$ in space (strictly speaking $P(\mathbf{r}) = \varrho(\mathbf{r})/\mathrm{N}$). Naturally, the electron density should integrate to the total number of electrons in the studied chemical system:

$$\int \varrho(\mathbf{r}) \mathrm{d}\mathbf{r} = N \tag{II.60}$$

It is worth noting, that the electron density itself contains virtually the same physical information about the system, as the initial wavefunction. This constitutes prerequisite for the so-called Density Functional Theory (DFT) group of methods. The latter are based on the theorems formulated by Hohenberg and Kohn, who developed the foundations of the DFT.[51] The first theorem states that the external potential v_ext is a unique functional of the ground state electron density. The second theorem tells about the variation principle and introduces the Hohenberg-Kohn functional, E_v^HK:

$$E_v^{\mathrm{HK}}[\varrho] = F_{\mathrm{HK}}[\varrho] + \int v_{\mathrm{ext}}(\mathbf{r})\varrho(\mathbf{r})\mathrm{d}\mathbf{r} \tag{II.61}$$

where F_{HK} is described as:

$$F_{\mathrm{HK}}[\varrho] = \min_{\Psi\to\varrho}\langle\Psi|T_{\mathrm{e}} + V_{\mathrm{ee}}|\Psi\rangle \tag{II.62}$$

Hohenberg and Kohn concluded that E_v^{HK} reaches its minimum only for the ideal ground state density. The aim is then to find the form of the Hohenberg-Kohn functional. Although it is not possible to give the exact solution to the problem, lots of approximations are present in the literature. These are usually based on the fit of the exchange-correlation part of E_v^{HK} to the large number of molecular densities calculated by other methods, such as post-Hartree-Fock ones (*e.g.* coupled clusters). One of the most widely used semiempirical functional is the hybrid exchange-correlation Becke three-parameter Lee-Yang-Parr functional (B3LYP).[52] This hybrid functional has become very popular due to its simplicity and usefulness for predicting molecular properties and geometries of organic moieties.

The method of practical application of DFT was proposed by Kohn and Sham.[53] Similarly, as in the case of the Hartree-Fock method, they considered the single determinant as a probe wavefunction. A derivation based on the Lagrange multipliers leads to a set of Kohn-Sham equations, analogous to the ones obtained by Fock. A widely used form of this method is realized by means of the so-called MO-LCAO approach (Molecular Orbitals – Linear Combination of Atomic Orbitals), where every molecular orbital is described as follows:

$$\varphi_i(\mathbf{r}) = \sum_{\mu} c_{\mu i}\chi_{\mu}(\mathbf{r} - \mathbf{R}_{\mathrm{A}}) \tag{II.63}$$

where c_μ is the linear coefficient for the μ-th atomic orbital. When the Kohn-Sham equations are solved using the MO-LCAO approach, the electron density can be simply calculated as a sum over the occupied molecular orbitals:

$$\varrho(\mathbf{r}) = \sum_{i} |\varphi_i(\mathbf{r})|^2 \tag{II.64}$$

In many cases the results of DFT calculations for solid-state systems agree quite satisfactorily with experimental data. Additionally, the DFT computational costs are relatively low when compared to traditional methods, such as approach and its descendants based on the complex many-electron wavefunction. However, main

problems regarding DFT concern the choice of the exchange-correlation functional and consideration of the dispersive interaction energy component.

II.3.3. Electrostatic interaction energy derived from wavefunctions

This paragraph shall be devoted to the *SPDFG* program,[48a] which I used to provide reference *ab-initio* interaction energy values (or close to *ab initio* results, as in my case the vawefunction was calculated using the DFT(B3LYP) approach, which inherently contains some empirical contribution). The *SPDFG* program evaluates the electrostatic interaction energy (E_{es}) between two interacting molecules, for which WFN-type files, containing molecular wavefunctions, had been made available. Therefore, its applicability is narrowed to chemical systems of the size adequate for the *GAUSSIAN* calculations.[54] On the other hand, the accuracy of the interaction electrostatic energy results depends strongly strongly on the level of theory, and the options applied at the stage of wave function computation. Nevertheless, E_{es}'s for Slater-type wave functions evaluated with the *SPDFG* program were found to be in excellent agreement with those obtained with Morokuma energy decomposition scheme used in *GAMESS-US*.[55]

The key point of the program is that it employs the numerical Rys quadrature method[56] for the evaluation of one- and two-electron Coulomb integrals. The method is based on a set of orthogonal (Rys) polynomials, which guarantees a simple general formula for integrals over basis functions, χ, of arbitrarily high angular momentum:

$$\int \chi_i(\mathbf{r}_1)\chi_j(\mathbf{r}_1)\frac{1}{r_{12}}\chi_k^*(\mathbf{r}_2)\chi_l^*(\mathbf{r}_2)d\mathbf{r}_1 d\mathbf{r}_2 = \sum_{\alpha=1}^{N} I_x(u_\alpha)I_y(u_\alpha)I_z^*(u_\alpha)W_\alpha \qquad \text{(II.65)}$$

in which u_α and W_α are the roots and weights of the Nth-order Rys polynomial, and I_x, I_y, and I_z^* are simple two dimensional integrals, evaluated using efficient and compact recurrence formulas. Additionally, the program is paralleled using the message-passing interface and can handle basis functions of any angular momentum.

II.3.4. Solid state calculations

(i) CRYSTAL code. Another program I used for the purpose of my study was the *CRYSTAL* package.[57] *CRYSTAL* is the quantum chemistry computational package designed especially for solid state calculations. The basic theory behind the program has been developed by Pisani, Dovesi, Roetti and Saunders. The program is focused on solid

state matter problems, with particular attention dedicated to periodic systems, such as crystals (3D), slabs (2D) *etc.* Two methods using the one determinant approximation are implemented in the code, *i.e.*, Hartee-Fock and Kohn-Sham approaches. *CRYSTAL* takes into account space group symmetry, and it always works in the primitive cell (which is the smallest one) to facilitate all computations. Unlike many other programs (*e.g. VASP* or *CASTEP*), *CRYSTAL* does not use plane waves, but a simple approach well-known from molecular computations, based on the Gaussian-type-orbitals (GTOs) centred at each atom site. Linear combinations of GTOs lead to local functions of the following form:

$$\chi_{\mu}^{\mathbf{g}}(\mathbf{r}) = \sum_{j}^{n_{\mathrm{c}}} d_j G\left(\alpha_j, \mathbf{r} - \mathbf{R}_{\mu} - \mathbf{g}\right) \tag{II.66}$$

where G is the normalized primitive GTO function, $\mathbf{R}_{\mu}$ is the position of the μ-th atomic centre, $\mathbf{g}$ is the lattice vector, and d_j and α_j are the linear coefficients and exponents, respectively. Such functions are then used to construct Bloch functions (BFs), distributed throughout the entire lattice:

$$\Phi_{\mu}(\mathbf{r}, \mathbf{k}) = \sum_{\mathbf{g}}^{\infty} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{g}} \chi_{\mu}^{\mathbf{g}}(\mathbf{r}) \tag{II.67}$$

where $\mathbf{k}$ is the wave vector. Finally, the crystalline orbitals (COs) (analogues of molecular orbitals) constitute linear combinations of BFs:

$$\Psi(\mathbf{r}, \mathbf{k}) = \sum_{\mu} c_{\mu i} \Phi_{\mu}(\mathbf{r}, \mathbf{k}) \tag{II.68}$$

On the basis of Bloch functions, the Hamiltionian matrix is easily factorised into diagonal blocks of finite size (the number of BFs in the unit cell), each corresponding to a point in the reciprocal space. The Schrödinger equation is then solved via the Hartree-Fock or Kohn-Sham approach, independently for each **k**-point. All properties, such as electron density, or electrostatic potential, can be further evaluated, similarly as in the case of single molecules. Additionally, *CRYSTAL* is able to compute the static theoretical structure factors by simply performing the Fourier transform of the unit cell electron density. Such evaluated theoretical structure factors are useful for expirement-theory comaparative charge density studies (see also Appendix A, Comment A4).

In the course of my PhD project, I utilised the *CRYSTAL* package so, as to calculate the cohesive energies of the studied crystals, structural motif interaction energy, surface free energy, and finally for the derivation of theoretical structure factors. Calculation

details are provided in Appendix A, Comment A3, A4 and A5. It's good practice to apply two main corrections, especially in the case of energetic investigations. The first one is the so-called Basis Set Superposition Error (BSSE) correction (counterpoise correction developed by Boyd and Bernardi).[58] The BSSE occurs when calculations of molecular properties (or crystal *etc.*) are based on finite basis sets. The second correction is the dispersion correction. It is a well-known fact that DFT methods cannot be used to estimate the dispersion energy component correctly. Consequently, energy calculations need correcting. Nowadays many methods dealing with this problem are being developed. Among them, there are the density functionals with the dispersion energy correction built in, and specially designed 'external' dispersion corrections. One of the very robust and computationally cheap way to account for dispersion in the DFT method, is to use the semi-empirical approach proposed by Grimme.[59] This correction method is implemented in the *CRYSTAL* code, and thus shall be shortly described.

(ii) Grimme dispersion correction. Grimme original method of the dispersive energy estimation is given by the following formula:[59-60]

$$E_{\mathrm{disp}} = -s_6 \sum_i \sum_{j>i} \frac{C_6^{ij}}{R_{ij}^6} f_{\mathrm{dmp}}(R_{ij}) \qquad \text{(II.69)}$$

In this expression, s_6 is the global scaling factor dependent on the density functional used, C_6^{ij} is the dispersion coefficient for the *i*-th and *j*-th pair of atoms, and R_{ij} is the interatomic distance. In the Grimme approach the essential part is the dumping function which generally is described as:

$$f_{\mathrm{dmp}}(R_{ij}) = \left(1 + \exp\left[-d\left(R_{ij}/R_{ij}^0 - 1\right)\right]\right)^{-1} \qquad \text{(II.70)}$$

where R_{ij}^0 is the sum of the van der Waals radii for the mentioned pair of atoms, and *d* is a scaling factor. The role of the damping function is to prevent near-singularities in cases of short interatomic distances. Extensive computational studies of various solids using periodic *CRYSTAL* approach, lead Civalleri *et al.*[61] to find out that typical values proposed by Grimme can be adjusted to better describe crystals. Civalleri *et al.* suggested to simply scale the van der Walls radii by 1.05 and 1.30 for heavy atoms and hydrogen atoms, respectively.

"Always do your best. What you plant now,
you will harvest later"

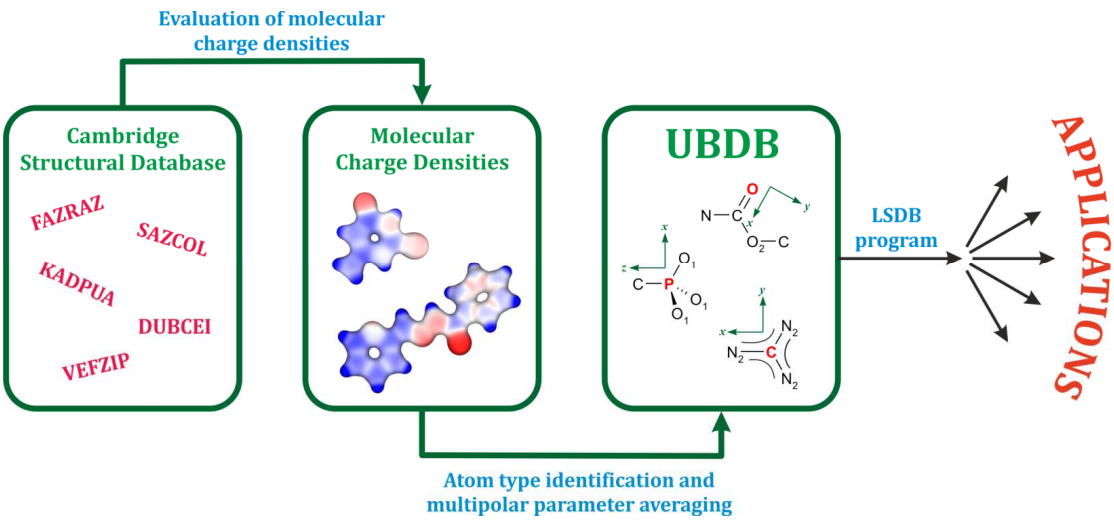
Og Mandino

III. Tool Development and Tests

The first challenge in my PhD project was extending the existing University at Buffalo Databank[31] so, as to provide tools for modelling nucleic acid electrostatic properties, and generally, to increase the scope of systems, which can be analysed using the databank. Such proceeding constituted a natural continuation of earlier research conducted at Prof. Coppens' laboratory at Buffalo University by Volkov *et al.*,[31b] and subsequently Dominiak *et al.*[31a] This task involved modification of atom type definitions, and also, changing substantially the *LSDB* program code. Once the new atom types were added, and the old ones recalculated, appropriate tests were carried out to verify the applicability of the derived pseudoatom models. This chapter includes a detailed description of each step made, together with the acquired results. I have also decided to add a section to include an *LSDB* manual, as such manual has not been publicized, to date. My work presented here, the new version of the databank, the UBDB2011, became the starting point for further research, described in the second part of my PhD thesis.

III.1. UBDB databank extension towards nucleic acid modelling

The accepted general idea of the UBDB databank construction is illustrated in Scheme 3.1. As shown, the first step includes the selection of sufficient quality molecular geometries from the Cambridge Structural Database.[41] The selected molecular structures need to fulfil the criteria of no disorder, exhibit an *R*-factor lower than 5%, and, certainly, they should contain the atom types of interest (that is atoms with a particular chemical environment). Then a set of theoretical calculations is performed on the chosen structures in order to evaluate the molecular charge density distribution.



Scheme 3.1. Schematic illustration of the databank construction procedure.

Subsequently, atom types are recognised, and their multipolar parameters are compared and averaged. If the two atoms, concerned as one atom type, differ by more than one standard deviation, a new atom type is introduced. UBDB contains atom type definitions together with multipolar, and expansion- contraction, κ and κ', parameters.

III.1.1. Selected CSD structures

First of all, I supplemented a previously used set of high quality experimental molecular geometries according to the Cambridge Structural Database (CSD)[41] to cover most of atom types that I might need. These structures included not only atoms present in proteins and nucleic acid chains, but also atoms indispensable for modelling amphotericin B and its iodoacetyl derivative, hydrazones, sunitinib, or boronic acids. Most of the corresponding REFCODEs (around 600 structures) are listed in the

Appendix B (Table B4). I enriched this set with some molecular geometries, which I had been working with, such as uracil derivatives, or the analysed hydrazone molecules.

The chosen CSD structures required further preparation as X–H bonds (where X is a non-hydrogen atom) were systematically shorter than in reality, due to the IAM refinement specificity. Therefore, I obtained hydrogen atom positions by extending X–H distances to their standard neutron diffraction values with the use of the new version of the *LSDB* program. Hydrogen atoms are of great importance for estimating any electrostatic properties of an organic molecular system (and not only electrostatics),[62] especially interaction energy values, which are among principle applications of the developed databank. Thus, I defined all the different C–H, N–H, O–H, S–H and P–H distances according to the newest article of Allen and Bruno.[63]

Such prepared molecular geometries were subjected to further theoretical calculations. What is worth mentioning, as the UBDB2011 version of the databank contains modified atomic keys describing the atom type's closest chemical environment more precisely with respect to its parent version, I carried out all of the calculations for both previously used molecular geometries and the newly added ones. All the old atom type entries and pseudoatom parameters stored in the databank were verified, modified if necessary, and recalculated, whereas the new entries were added.

III.1.2. Theoretical calculation details

I performed all the theoretical calculations following the procedure of constructing the previous version of the databank.[31a] This included:

(*a*) single-point calculations on the described set of small molecules performed with the *GAUSSIAN03* program[54] using density functional theory (DFT) with a 6-31G** basis set[64] and B3LYP hybrid exchange-correlation functional[52, 65] (molecular charges were set to the formal values) ⇨ molecular wavefunctions;

(*b*) analytic Fourier transform of the molecular charge densities (based on previously calculated wavefunctions) for reciprocal lattice points corresponding to a pseudo-cubic cell with 30 Å edges[42] ⇨ theoretical valence-only structure factors in the range $0.0 < \sin\theta/\lambda < 1.1$ Å^{-1}.

III.1.3. Aspherical atom refinement of the theoretical structure factors

The resulting theoretical structure factors were subsequently fitted with the Hansen-Coppens pseudoatom formalism[10] using the *XD* program suite.[40] According to the previously described procedure, I refined both radial screening factors, *i.e.* κ and κ', independently for each atom, with the exception for the chemically equivalent hydrogen atoms. The latter shared the same κ and κ' parameters. The multipolar expansion was truncated at the hexadecapolar level (l_{max} = 4) for the non-hydrogen atoms. In the case of hydrogen atoms I refined only bond-directed functions up to quadrupoles (l_{max} = 2). In order to reduce the number of least-squares variables, I imposed local-symmetry constraints in some atom cases. For non-charged molecules I applied electroneutrality constraint, whereas other molecules were scaled to their formal charges and these were kept constant during all refinements. To avoid the phase ambiguity problem the phases were fixed at theoretically calculated values. Clementi and Roetti atomic wave functions[66] served as a basis for atomic scattering factors, and were modified by the κ parameter in the course of the refinement. For the purpose of the deformation functions I used single-ζ exponents corresponding to weighted averages over the *s*- and *p*-shell values suggested by Clementi and Raimondi,[67] with powers of *r* (n_l) given in Table 3.1.

Table 3.1. Slater radial function coefficients, n_l and ζ, applied in the refinements of theoretical structure factors

Element	n_l	ζ / a_0^{-1}
B	2, 2, 3, 4	2.463
C	2, 2, 3, 4	3.176
N	2, 2, 3, 4	3.839
O	2, 2, 3, 4	4.466
F	2, 2, 3, 4	5.186
P	6, 6, 6, 6	3.459
S	2, 2, 6, 8[a]	3.851
Cl	4, 4, 4, 4	4.259
I[b]	8, 8, 8, 8	4.848
H	1, 2	2.000

[a] As suggested by Dominiak and Coppens[68]
[b] Su-Coppnes-Macchi function coefficients[69]

Finally, I averaged the multipolar parameters, *i.e.* P_v – the valence electron population (in the model used equivalent to the spherical monopole population), P_{lm} – populations of the deformation functions, κ and κ' – the expansion-contraction parameters, over each atom type. These obtained values constitute the prime information stored in the databank.

III.1.4. Atom type definitions

Construction of a databank requires selecting atoms, which are similar enough to be averaged and, therefore, suitable to represent a particular atom type. Such a set of atom types should constitute the smallest possible number of pseudoatoms accurately reproducing the charge density distribution of many molecules. With the latter in mind, the following general criteria for the definition of an atom type had been defined:

(1) element type,

(2) the number of attached atoms (number of the nearest neighbours, group planarity),

(3) nearest neighbour type,

(4) aromaticity (ring planarity),

(5) local symmetry.

In the new version of databank I subdivided point (4) into the number of aromatic rings to which a given atom belongs and a summary ring member number. Consequently, atoms, being planar rings' members, are given a new aromaticity flag consisting of two key words: RING and MEMB in the databank file (Appendix B, Table B3). The first one stands for a number of planar rings, while the second describes the total number of the selected ring members. Such a definition distinguishes atoms belonging to 5- and 6-membered rings. This improves multipolar parameter adjustment to the geometry of a given atom surroundings, and so increases consistency among the atom types. Moreover, 5-membered rings are more often heterocyclic than 6-membered ones, which affects the multipolar populations. This effect is also reduced by distinguishing atoms belonging to rings of different sizes. Nevertheless, as a result atom belonging to 5-membered rings are more often characterised by higher multipolar parameter standard deviations than atoms from 6-membered rings (obviously, when both compared samples are of similar diversity and number). C378 carbon atom type (Figure 3.1), for instance, showed standard deviation equal to 0.27. After splitting into atoms C378 (MEMB = 6) and C379 (MEMB = 5), the standard deviation of the new C378 atom type decreased to 0.11, while for C379 that amounted to 0.039. Additionally, these new atomic keys enable a better description of atoms which join fused aromatic rings together. This is of particular importance for modelling of DNA and RNA base fragments.

Another important thing is to characterize the atom type chemical environment. Generally, any neighbouring atom (point (3)) is described only by element type and hybridisation. Hybridisation states of neighbouring atoms are derived solely on the basis

of the number of atoms attached. This information is usually kept in the local symmetry key (SYMM) within the atom type definition. However, some atoms need to be treated more specifically. Therefore nitrogen atoms are additionally split into $sp^3(4)$, $sp^3(3)$, $sp^2(3)$, and $sp^2(2)$ types (hybridisation state followed by the number of closest neighbours in parentheses). Such an approach takes into account also the planarity of the nitrogen containing group and is included in the so-called exception key, *i.e.* the EXCEPT GROUP. Whenever required, the described criteria are also modified by providing more precise neighbour atom definition, defining if it belongs to a planar ring or not, or by including the effect of the next-nearest neighbours. An example of the first case is represented by statistically distinguishable carbon atoms C378 and C381 (Figure 3.1a), while the second occurs for oxygen atoms such as O101, O102 and O103 (Figure 3.1b). Whereas oxygen atoms O102 and O103 are statistically equal, O101 is significantly different. However, the addition of each 'exception group' should be preceded by a statistical study to judge if the new atom type is distinguishable, and then the EXCEPT GROUP key should be formulated in the most precise and unambiguous way. Introducing of each 'exception group' includes *LSDB* program modification, as any new entry needs to be carefully defined prior to incorporating into the code. I have added more than twenty exception keys, pertaining to all atom types present in the databank to date (*e.g.* PH-F, PH-Cl, N-c3conf, aromaturea, aromatketone, aromatester, aromatamide, mixed nitrogens, urea, amido, cPo3 *etc.*; see the databank file in the Appendix B, Table B3), and those that might possibly be needed in the foreseeable future.

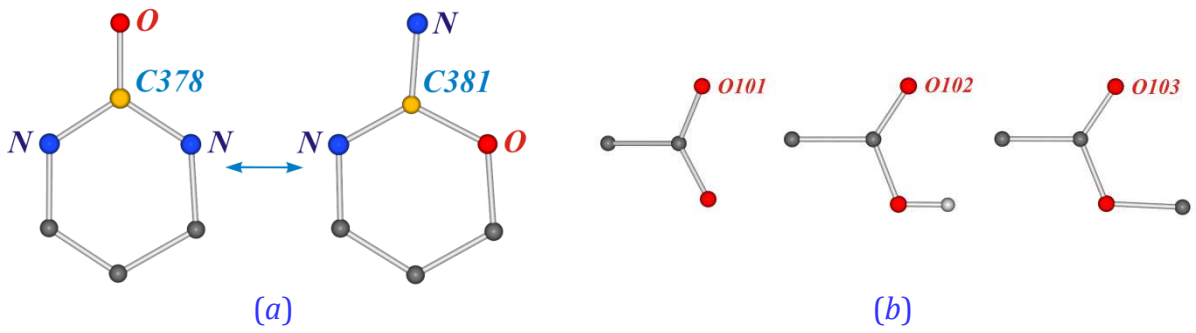


Figure 3.1. Examples of more precise neighbour atom definition requirement: (*a*) carbon atoms C378 and C381; (*b*) Oxygen atom types: carboxylate O101, carboxylic O102 and ester O103. Atom colour coding: oxygen – red, nitrogen – blue, carbon – dark grey, hydrogen – light grey.

On the other hand, sometimes, a simpler definition of an atom type is used. This takes place, when the number of occurrences of an intended atom type in the molecular sample is too small to obtain statistically meaningful average values of the deformation-density parameters, or when it is difficult to define a certain atom type. I marked these atoms in the *db2011.db* file with a TEMP flag (see also Appendix B, Table B3). When the number of the averaged atoms is equal to, or lower than 5, the TEMP flag is followed by the number indicating the sum of the averaged atoms. These atom type definitions shall be improved together with the next version of the databank. A standard entry in the databank is presented below (Scheme 3.2). As an example I have once again selected the C378 atom type.

```
ATOM 6 COMMENT C378 n-=C(=o)=-n  AROMATIC
NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m (PLAN t)
TYPE 1 ATOM 7 NUMB 2
TYPE 2 ATOM 8 NUMB 1
EXCEPT GROUP aromaturea
 PVAL   4.152  KAPPA   0.993  KPRIM   0.830       SIGPV   0.011
 PLMS 1  1  -0.111  PLMS 2  0  -0.310  PLMS 2  2  0.138  PLMS 2 -2  0.011
 PLMS 3  3  -0.472  PLMS 4  0  0.086  PLMS 4  2  -0.026  PLMS 4  4  -0.114
```

Scheme 3.2. Exemplary UBDB2011 entry.

The first line contains chemical element type information and a brief atom type description following the word COMMENT. This includes the atom type number in the databank file. The example atom type is coded as C378, where C stands for carbon (element type), 3 is the number of first neighbours, and the other digits are ordinal ones. The second line consists of the following key words:

NEIG – number of neighbouring atoms,

TYPE – number of different chemical elements within neighbours,

RING – number of planar rings, to which a given atom belongs,

MEMB – a total number of ring members,

SYMM – atom local symmetry (always m if an atom belongs to a planar ring),

and optionally PLAN which is either T (*true*) or F (*false*) and describes the planarity of an atomic group (if not provided, it is automatically set to F). The next two lines contain information about first neighbours. Here, the key word TYPE indicates the serial number

of a particular neighbour type, ATOM brings the information about the element type of this neighbour type, while NUMB shows how many neighbours of this type (*i.e.* how many 'nitrogens', how many 'carbons' *etc.*) are attached to the parent atom (here, the carbon atom being described). Finally, the EXCEPT GROUP, as previously mentioned, is the additional key applied, if an atom definition requires more precision. Here 'aromaturea' means that this particular carbon atom is attached to two nitrogen atoms which belong to the aromatic ring, while the oxygen atom does not. It is essential in that case, because both atoms C378 and C381 are characterised by the same basic keys, while they are not identical, but in fact substantially different (Figure 3.1).

The last lines of the atom type entry provide charge density information stored by means of multipole population values and κ and κ' contraction-expansion parameters. Additionally, SIGPV, being a standard deviation of the mean valence population value, gives the estimation of the atom type 'quality', *i.e.* the diversity among the averaged atoms considered as one atom type.

III.1.5. Local coordinate system assignment

A proper averaging of the multipole populations of equivalent atom types and also pseudoatom parameter transfer onto a given molecule are essential both for databank construction purposes and for its applicability. Consistency in selecting the local coordinate system is crucial here. Therefore, the spherical harmonic functions centred at atoms should be expressed in a common local frame, uniquely defined for every atom type. The *LSDB* program provides fully automatic definition of unique local coordinate systems based on the coordination environment of each atom. It is especially useful in the case of charge density analysis of any large molecules, where the manual assignment of coordinate systems becomes prohibitively cumbersome. It is also essential in the building of the pseudoatom databank based on a large number of small molecules.

Local coordinate system associated with an atom is oriented such as to allow local symmetry constraints. When there is more than one possibility of selecting a particular local coordinate axis within a given local symmetry, an additional procedure is applied. It is based on a set of criteria defining the analysed atom environment, i.e. atomic number, hybridisation state, valency and, if it is still inconclusive, distance to the central atom. Neighbouring atom priority is defined as follows:[31a]

(1) atoms with the highest atomic number are selected, and if there is only one such atom, it is chosen as atom 1.

(2) if step (1) is inconclusive, atoms with the highest atomic number are ordered according to their hybridization state (sp first, then sp^2 and sp^3) and then by their valency (number of attached atoms; the smallest number first); the first atom from the list is chosen as atom 1.

(3) if the preceding procedure is not conclusive, *i.e.*, there is more than one atom at the top of the list with the same element type, hybridization, and valency, atoms are sorted according to their distance to the central atom, and the one with the shorter distance is chosen as atom 1.

After atom 1 is selected, atom 2 is chosen from the remaining neighbouring atoms, adhering to the rules defined above. In some cases, it is necessary to add a dummy atom to make use of the symmetry, allowing minimization of the number of multipole parameters. The local symmetry assignment to atoms having two neighbours is based on criterion (3), however, when there is only one hydrogen atom, the selected first local axis is always directed along the other bond. In the case of atoms belonging to one planar ring, the *X*-axis is always oriented towards the centre of the ring and mirror plane symmetry (*m*) is imposed even if higher symmetry is possible. I decided that the same symmetry is kept when an atom is common for two fused planar rings. Here, the *X*-axis is then directed towards the second atom of the kind, the *Y*-axis towards an atom from the 6-membered ring keeping right angle and the m point group symmetry (Figure 3.2).

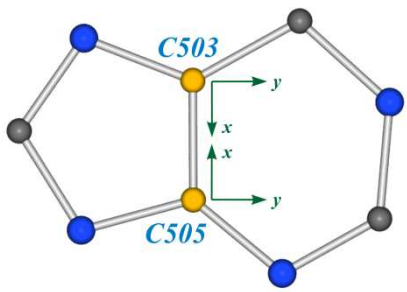


Figure 3.2. Local coordinate system schema for atoms belonging to two planar rings, C503 and C505. Atom colour coding: nitrogen – blue, carbon – dark grey.

Usually, a right-handed coordinate system is defined, except in the case of chiral atoms, for which both right-and left-handed systems are allowed. Chirality is defined locally, only by the character of the nearest neighbours. To illustrate the local symmetry

assignment in different cases, I have selected example atoms of *no*, *m*, *mm*2 and, 3*m* symmetry present in the new version of the databank (Figure 3.3).

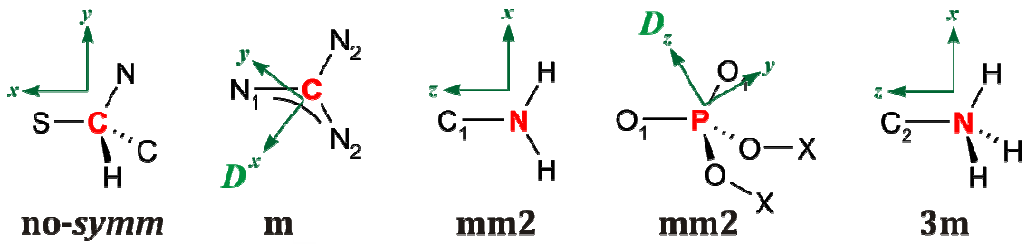


Figure 3.3. Local coordinate systems assigned for atoms with different symmetry types; local symmetry recognised by *LSDB* in some cases is lower than the real one.

III.1.5. Atom types in the databank

Currently, the databank contains 218 atom types, among which there are 19 hydrogen atom types, 111 carbon atom types, 31 nitrogen atom types, 35 oxygen atom types, 8 sulphur atom types, 8 phosphorus atom types, 2 chlorine atom types and 2 fluorine atoms, as well as the recently added boron and iodine atoms. It includes all atom types encountered in peptides, proteins, nucleic acid bases and some other biologically interesting molecules (atoms indispensable for modelling amphotericin B and its iodoacetyl derivative, hydrazones, sunitinib *etc.*), or present in boronic acids. As it was noted before, I recomputed all the old atom types, modified some of them, while other added. All atom types incorporated in the databank are listed in the Appendix B (Tables B1 & B2), while the atom type charge density distribution parameters are collected in the *db2011.db* file (which is also available from the Appendix B, Table B3). Sample atom types stored in the UBDB2011 databank[70] are presented in Figure 3.4.

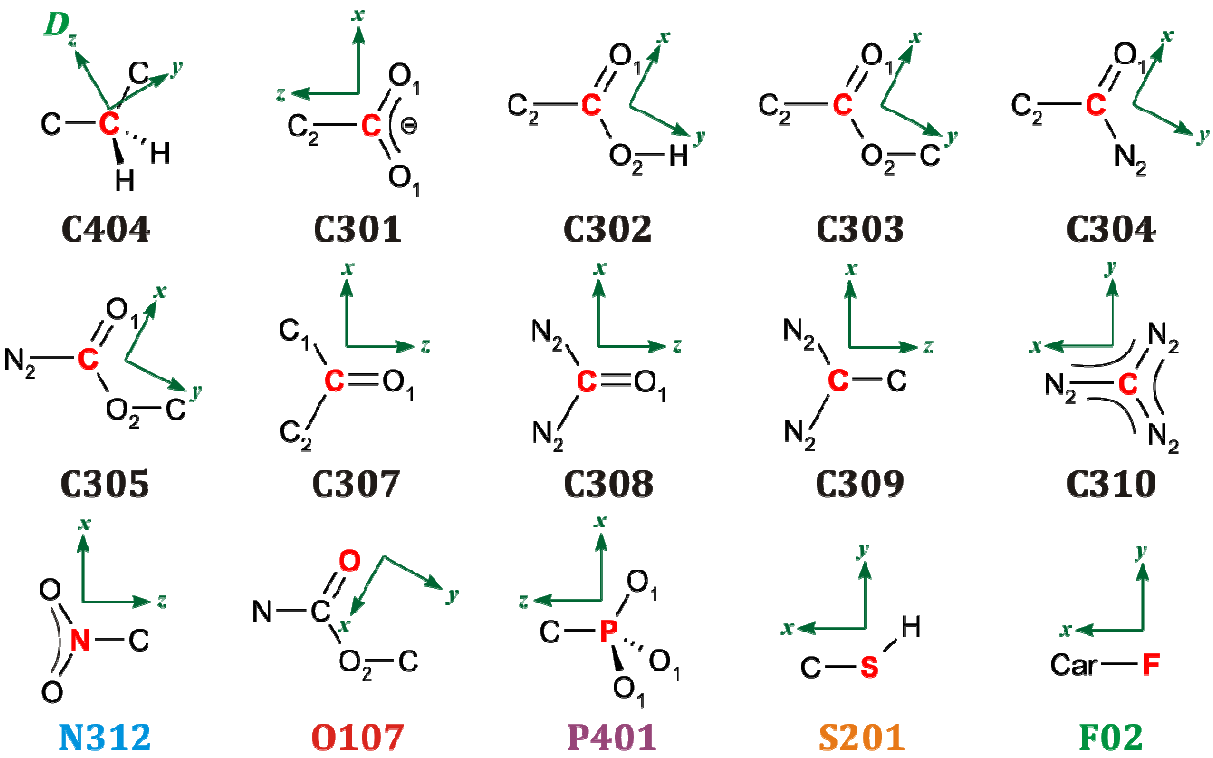


Figure 3.4. Example atom types stored in the UBDB2011 databank.

III.2. LSDB program

The *LSDB* program constitutes a crucial tool both for applying the databank and for adding and/or modifying the databank entries. Therefore, my project related to the extension of the UBDB databank and introduction of new atom type definition keys required substantial work on the *LSDB* Fortran code. Therefore, I decided to dedicate this section entirely to the program itself, as there is no other proper manual available. I am also going to present the changes made, and new features, which are now available.

(i) Modifications. In general, the core of the program is preserved with respect to its earlier version (Volkov *et al.* 2001-2006).[31, 33] The modifications I introduced, concern mainly the atom type definition module, the aromatic ring definition, and the local coordinate system determination in the case of atoms at the intersection of fused aromatic rings. Additionally, I incorporated new X–H bond lengths into the code, according to the latest article of Allen & Bruno.[63] I also corrected the module responsible for kappa parameter assignment to the chemically equivalent atoms, i.e. to the same atom types regarding the databank entries. Furthermore, now *LSDB* automatically writes the *MOPRO*-type input files. I found it especially handy in the second part of my thesis, where I tested different refinement strategies, which may be obtained using the *MOPRO*[35b] much easier than with the *XD* package.[40]

(ii) Input/output files. The *LSDB* program was originally designed to read and produce files suitable for the *XD* package, *i.e.* the *xd.mas* and *xd.inp* files. It also reads the crystallographic information from *SHELX*-type files (**.res* and **.ins*), and on that basis assigns atom types, local coordinate systems, *etc.*, which are being written to the *xd_lsdb.mas* and *xd_lsdb.inp* files. The other option, concerning the use of the *PLATON*-type files, was basically introduced for visualising the local coordinate systems. *LSDB* produces also supplementary output files containing structural information and suggested restraints for the lengths of different X–H bond types.

For many years, the *XD* package has been the most popular suite of programs for evaluating charge density distributions of crystals within the Hansen-Coppens formalism.[10] However, in the meanwhile the *MOPRO* suite of programs had undergone a dynamic development, and thus has become a very promising alternative for *XD*. *MOPRO*'s very big advantage is its constant progress, whereas the *XD* package is not being developed and updated a lot, nowadays. Accordingly, as previously mentioned, I implemented a new procedure into *LSDB* for automatic writing of the input files suitable

for the *MOPRO* package. These are the *mopro_lsdb.par* file with some refinement instructions and crystallographic data, and also, the *con_mopro.inp* and *res_mopro.inp* files. The latter two contain information about possible restraints and constraints which may be applied during the refinement procedure. The *res_mopro.inp* file includes X–H bond length restraints (according to Allen & Bruno, 2012) and standard restraints of hydrogen atom thermal motion isotropic parameters, depending on the parent atom they are attached to. In turn, the *con_mopro.inp* file stores the constraints regarding atomic symmetry and similarity, which are based on the UBDB atom type definitions. Example *MOPRO* input files, created for the 6-methyl-2-thiouracil structure by *LSDB*, are available in the Appendix B (Table B5).

What is worth noting, the LSDB program to function correctly additionally requires the *lsdb.inp* and the databank files. The *lsdb.inp* file defines the requested input and output file types and contains the instructions concerning the information one wants to obtain using *LSDB*. A template of an input file is created automatically when executing the *LSDB* program with the '-h' option. The exemplary input file is shown in Scheme 3.3.

The *lsdb.inp* file may be divided into parts, including four distinguishable sets of instructions. These are marked with different colours in the scheme. Generally, the input file consists of instruction lines which contain a keyword followed by possible options to be selected, whereas there is a line above each keyword containing its brief description. To choose a certain option one should sign it with a star symbol '*'.

The first block highlighted in light blue contains the information of the input and output files. Naturally, the FORMAT and INPUT keywords pertain to the previously described input file formats and the actual name of the files intended to be read. In turn, VERBOSE indicates the amount of information to be printed out in the *lsdb.out* file. When 1 is chosen the content of *lsdb.out* will contain less details, while when 3 – more details.

The *LSDB* program allows to change the tolerances for bond lengths (regarding the sum of atomic covalence radii), and for atomic group and ring planarity (TOLS option). It is also possible to tune up the covalence radius of a particular atom, as shown on the example of hydrogen and helium atoms (RADII H 0.1 and He 0.2). The new atomic covalence radius is defined after the RADII keyword, and follows the chosen atom symbol. NCELLS is used when only a part of a molecule is in the asymmetric unit.

```
!
! Example of LSDB input file (lsdb.inp) :
!
FORMAT  *shelx xd platon
! for SHELX :
INPUT  shelx.ins rename
! for XD :
!INPUT  xd.mas xd.inp
! etc :
!INPUT  file.inp
! printout level (higher number gives more printout) :
VERBOSE  1
! tolerances for bonds / planar rings / planar groups
TOLS  bond 0.4 ring 0.1 group 0.1
! modify covalent radii
RADII  H 0.1 He 0.2
! number of cells in each direction for complete fragment search
! comment it out or set to -1 to turn off search
NCELLS  -1
! how to handle disordered part
DISORD  extendH  db
! generation of keys
KEYGEN  *XYZUIJ *all ordered  *MULT *all ordered
! use simple model for determination of chemical equivalency
SIMPLE
! use/do_not_use chemical/kappa contraints for H/non-H atoms
CHEMCON  *hydrogen   *other
! extend H-atoms to neutron/standard distance
Extend_H
! search for 3-8 membered rings
RINGS  3  8  *special
! assign pseudoatom parameters from UBDB2011
DATABANK  ubdb2011.dat
! select method for scaling of Pv's
SCALE  *sigma diff fract
! number of fragments
NFRAG 2
! fragment specifications
FRAG 1 1-12
FRAG 2 13 -34
! net charges of fragments
CHFRAG 0.0 0.0
```

Scheme 3.3. Exemplary *LSDB* input file.

Then such a molecule needs to be completed and the number following the NCELLS keyword denotes a number of elementary cells to be explored (in each direction: *X*, *Y*, *Z*) for fragment searching and fragment growing purposes. Atomic disorder constitutes another problem requiring special treatment. Unfortunately the existing DISORD keyword is a temporary one, suitable only for a narrow range of disorder cases. So far the problem of disorder is not universally solved in *LSDB*, and thus shall be addressed in the nearest future. All of the described options regarding structure perception constitute the second instructions block.

The green section of the instruction file includes keywords related to atomic specificity and the output files. Within the KEYGEN option one may change the content of the output xd.mas file, or more precisely speaking – the refinement description block. When XYZUIJ is activated, the program writes '1' numbers in the atom coordinate and thermal motion commend block. The MULT option assigns the refineable multipolar parameters according to the symmetry of a particular atom type as in the databank. The SIMPLE instruction key should be always on, as it is closely related to the CHEMCON action. CHEMCON indicates atoms to be treated equally in the refinement procedure, if they represent the same atom type in the databank. If RINGS is activated with option the 'special', it means that the aromatic ring atoms are treated as mirror symmetry ones (just like for the purpose of the databank entry). If it is switched off, planar ring members are treated as non-aromatic atoms with all its consequences (therefore, different local symmetries are allowed). Extend_H fixes X–H at their neutron-normalised values. In turn, the DATABANK key specifies the databank file intended to be used.

The last part, marked in light violet colour in Scheme 3.3, deals basically with charge scaling. Pseudoatom paramaters taken directly from the databank may lead to summary charge values slightly different, than the desired formal charges of studied molecules. Therefore, a scaling procedure is required. For that purpose the SCALE keyword is included. It allows for applying three different scaling methods. The 'sigma' option denotes Faerman and Price algorithm, which employs weighting scheme based on standard deviations of atom type valence population parameters (P_v).[46, 71] Each atom P_v is scaled individually, being multiplied by a weighting factor equal to:

$$w_k = \frac{P_\mathrm{f} - P_\mathrm{UBDB}}{P_\sigma}\sigma(P_{\mathrm{v}k}) \qquad \text{(III.1)}$$

where P_f is the sum of valence electrons of atoms constituting a given molecule or system accounted for its formal charge (specified in the *lsdb.inp* file within the CHARGE option), PUBDB is the summary P_v of all these atoms denoted by the assigned UBDB atom types, P_σ is the summary P_v standard deviation taken from the adequate UBDB entries, and finally $\sigma(P_v)$ is the P_v *e.s.d.* value of the particular k-th atom, to which the w_k weight corresponds. In consequence, atom types characterised by greater standard deviations, and thus being more diverse among the initial atom group used to prepare the UBDB atom entry, are more affected by the scaling procedure. In other words, their valence populations change more in respect to their pure databank P_v values, than for atom types described by lower standard deviation values. This is in agreement with common sense – when an atom type is characterised by higher standard deviation it means that it is, in general, more variable depending on its chemical environment.

The other two possibilities, *i.e.* difference 'diff' and fraction 'fract' options, are much more simply defined. Both procedures are not atom-type-dependent, and thus scale every atom equally:

$$\Delta_{\mathrm{diff}} = \frac{P_f - P_{\mathrm{UBDB}}}{N} \qquad \text{(III.2)}$$

$$\Delta_{\mathrm{fract}} = \frac{P_f}{P_{\mathrm{UBDB}}} \qquad \text{(III.3)}$$

where N is the number of atoms in a given fragment. In the first approach, the calculated Δ_{diff} value is added to every UBDB atomic valence population, while in the second method each UBDB atomic valence population is multiplied by Δ_{fract}. Nonetheless, the most widely used and recommended scaling procedure is the Faerman and Price approach. *LSDB* allows also for dividing the studied system into fragments, which can be specified with the NFRAG and FRAG keywords. NFRAG denotes the total number of fragments, whereas FRAG assigns particular atoms to a fragment. Additionally, as mentioned earlier, the CHARGE option is designed to define molecular fragment charges.

(iii) Further code development. The most important things, which need to be implemented, or improved, concern:

- handling *R*-centred lattice types,
- disorder treatment,
- providing fully universal ring search procedure.

(*) The LSDB program for Windows and Linux operating systems is available under the following website address: crystal@chem.uw.edu.pl.

III.3. Databank verification

On the basis of atom type multipole parameters and the corresponding κ and κ' values, one can model charge density distribution of a macromolecular system, for which high-resolution X-ray data is unavailable. Having such a reconstructed charge density, a number of its electrostatic properties can be computed, *e.g.* bond properties, molecular electrostatic moments, and electrostatic potential. Such an analysis may constitute a good foundation for deriving essential information of the binding properties of studied molecules, and electrostatic interactions present in biological systems, or pharmaceutical complexes. Therefore, I found it desirable to check, whether it is justified to use a particular pseudoatom databank for such a purpose, and, if so, what are the method applicability limits. The three main pseudoatom databanks were already tested on peptide molecules and compared by Bąk *et al.* Nevertheless, this choice of the test molecules did not cover the newly added atom types. Prior to that, in the paper by Czyżnikowska *et al.*,[72] I had contributed with providing estimations of electrostatic interaction energies of Watson-Crick nucleic acid base pairs, obtained by using a very preliminary extended version of the present UBDB2011 databank. Here, the DF-DFT-SAPT method served as a point of reference for my results. These early results were very promising.

Electrostatic interaction energy is one of the electrostatic properties, which are most sensitive to any imperfections in the modelled charge density distribution. Thus, I decided to use this property for verifying the UBDB2011. As previously emphasised, the databank was extended with an eye to model atom types present in nucleic acid chains, RNA and DNA, and in some other species of interest. The majority of newly calculated parameters concerned atoms belonging to pyrimidine and purine bases. Therefore, I tested the obtained pseudoatom charge density models on a set of dimers formed by nucleic acid bases, or containing nucleic acid base and amino acid fragments, for which *ab-initio* computations, or other advanced quantum chemical methods were applicable.

III.3.1. Test dimers and method details

For the purpose of the investigation, I chose a set of over 150 nucleic acid base dimers, *i.e.* adenine-thymine (A:T), guanine-cytosine (G:C) and also guanine-adenine (G:A) complexes (Figure 3.5), which had been earlier designed for energy studies

published by Czyżnikowska *et al.*[72-73] Additionally, I considered a set of nucleic acid bases (NABs) interacting with amino acid (AA) fragments, both neutral and charged (Figure 3.5*d*, Appendix B, FigureB12).[73] The monomer geometries used in this study were optimised using the *GAMESS-US* package[74] at the MP2/aug-cc-pVDZ level of theory imposing *m* point group symmetry.[75] The relative position and orientation of either guanine and cytosine, or adenine and thymine, were uniquely defined by the six base-pair parameters (shear, stagger, stretch, opening, buckle and propeller). The dimers were generated with the aid of the 3DNA program based on the experimentally determined values of the base-pair parameters taken from crystallographic data.[72, 76] In the case of both A:T and G:C complexes most representative 'crystallographic' mutual configurations were chosen. They were subsequently used for potential energy surface scans with respect to a given base-pair parameter, with the remaining five kept fixed. These scans are mentioned in Figure 3.7*a*,*b* presenting energetic results. The ranges of the scanned parameters were selected regarding the experimental data and were divided into equal intervals. Altogether I tested 55 A:T and 60 G:C dimers, 48 G:A geometries (stacked complexes appearing in B-DNA crystals) and 13 nucleic acid base-amino acid side chain (NAB:AA) benchmark complexes.[72-73, 76] Such a sample covered all of the most important interaction types, from hydrogen bonding, through π-stacking contacts, to interactions involving charged species. Therefore, it was possible to deeply explore the applicability of UBDB2011.[70]

To assign appropriate pseudoatom multipolar parameters to atoms belonging to selected systems, I used the modified *LSDB* program. Apart from a direct shift of the parameters from the databank to a structure, this procedure required charge scaling. This is due to the fact that the straightforward transfer of valence populations results in the total molecular charge magnitude different than neutrality or other formal charge value characterising the examined molecule. Therefore, to ensure electroneutrality (or any other formal charge), I *a posteriori* adjusted all monomers to their net charges by scaling the pseudoatoms according to the Faerman and Price scaling algorithm[46] implemented in the *LSDB* (more details in §III.2). Subsequently, I employed the Exact Potential and Multipole Method (EPMM),[48] implemented in the *XDPROP* module of the *XD* package,[40] to compute electrostatic interaction energies from the derived densities. It combines numerical evaluation of the exact Coulomb integral in the inner region

(≤ 4.5 Å) with the Buckingham-type multipole approximation[49] for the long-range interatomic interactions as described in the second chapter in §II.2.4.

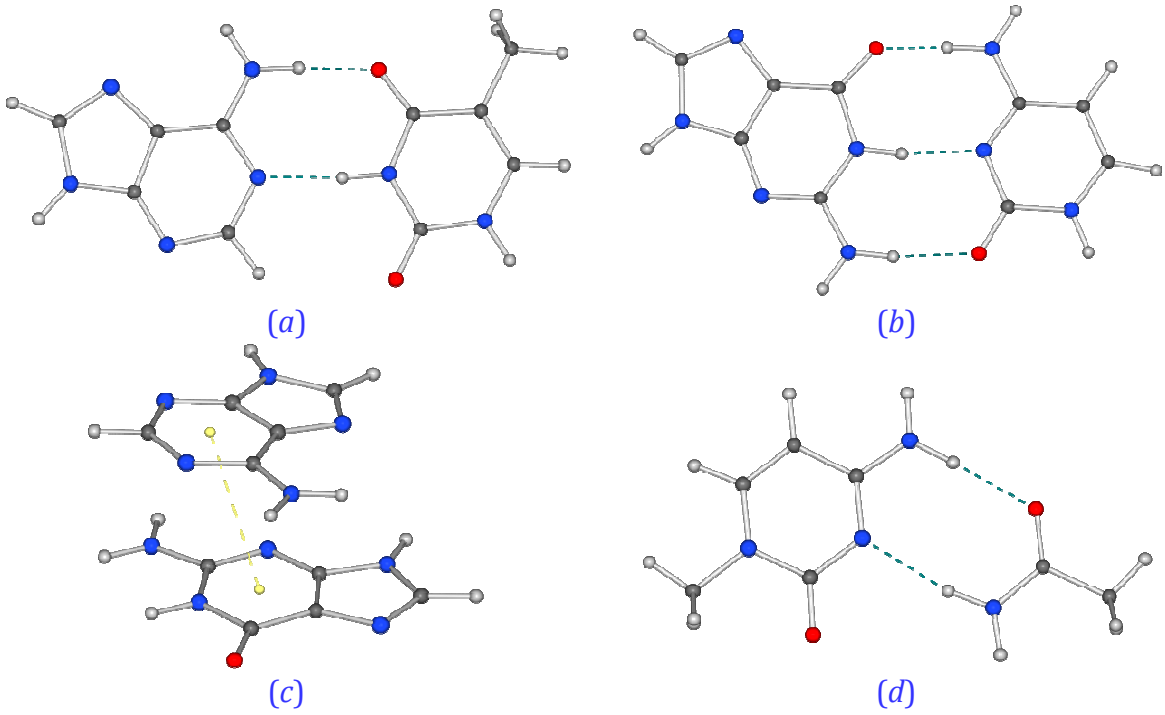


Figure 3.5. Selected chemical systems: (*a*) A:T; (*b*) G:C; (*c*) G:A; (*d*) uracil interacting with methylamide residue (NAB:AA). Atom colour coding: nitrogen – blue, oxygen – red, carbon – dark grey, hydrogen – light grey.

I compared such electrostatic interaction energy (E_{es}) values with the corresponding reference results, which I obtained directly from the molecular wave functions at the same level of theory (B3LYP)[52, 65] with two different all-electron basis sets: 6-31G** and aug-cc-pVDZ.[64, 75b] In order to evaluate E_{es} from monomer charge density distributions expressed in terms of Gaussian-type basis functions, I applied the *SPDFG* program. This program requires WFN type input files. These were calculated within *GAUSSIAN03* package.[54] I conducted all the *GAUSSIAN03* calculations with the SCF=Tight option, which requests tight self-consistent field convergence criteria. The *SPDFG* program E_{es} results might be taken as a reference point as it was already shown that they are in excellent agreement with those obtained with Morokuma-Ziegler energy decomposition scheme[77] implemented in *GAMESS-US*.[55]

As I mentioned previously, the chosen systems cover two most important interaction types present in the DNA and RNA molecules, *i.e.* hydrogen bonding and π…π stacking interactions. However, so far, it is only possible to derive electrostatic component of the total interaction energy on the basis of molecular charge density

distribution obtained from the databank. The correlations between the combined UBDB2011+EPMM approach and the *SPDFG* data are quite rewarding, $R^2 > 0.90$ in the case of A:T, G:C and NAB:AA complexes and $R^2 = 0.75$ for G:A (Table 3.1, Appendix B).

Table 3.1. Determination coefficient (R^2) between UBDB2011+EPMM and SPDFG E_{es} values.

Complex type	R^2
A:T	0.91
G:C	0.97
G:A	0.75
NAB:AA	0.99

Table 3.2. Root Mean Square Deviations (RMSDs) between UBDB2011+EPMM and SDPFG/6-31G** or SPDFG/aug-cc-pVDZ E_{es} values, respectively.

Complex type	RMSD / kcal·mol^{-1}	
	(1)	(2)
A:T	2.2	4.3
G:C	5.0	1.4
G:A	3.5	–
NAB:AA	1.5	–

(1) B3LYP/SPDFG/6-31G**;
(2) B3LYP/SPDFG/aug-cc-pVDZ

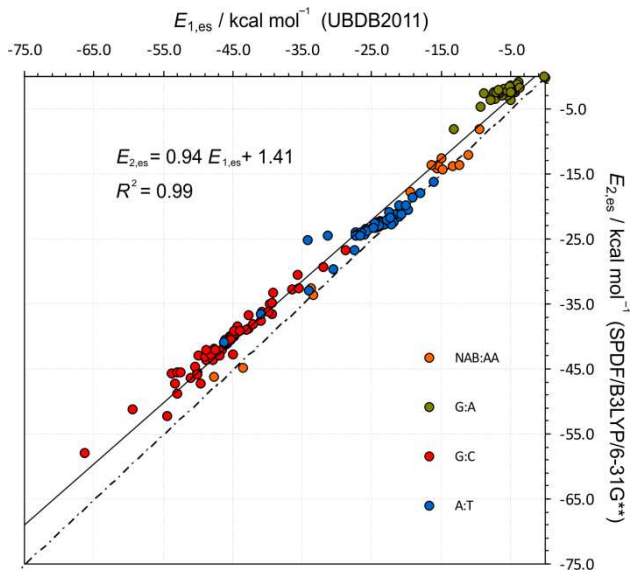


Figure 3.6. Correlation between SPDFG/B3LYP/6-31G** E_{es} results and UBDB2011+EPMM method. Total RMSD = 3.7 kcal·mol^{-1}. Solid line stands for linear least-squares fit ($R^2 = 0.99$). Dashed-dotted line stands for $E_{1,es} = E_{2,es}$ diagonal.

The overall Root Mean Square Deviation (RMSD) for all the data sets taken together amounts to 3.7 kcal·mol^{-1}, while the linear constant is close to unity and the correlation

coefficient reaches the value of 0.99. The distribution of energy points between the UBDB2011+EPMM and SPDFG methods for the analysed data series is shown in Figure 3.6. RMSD between the SPDFG/B3LYP/6-31G** values of electrostatic energy and the UBDB2011+EPMM amounts to 2.2 kcal·mol^{-1} for the set comprising 55 A:T Watson-Crick base pairs, and to about 5.0 kcal·mol^{-1} if we consider the set of G:C complexes (Table 3.2). What is interesting, G:C UBDB2011+EPMM results are systematically closer to the SPDFG/B3LYP/aug-cc-pVDZ derived energy values (RMSD = 1.4 kcal·mol^{-1}) than to the corresponding SPDFG/B3LYP/6-31G** results. Such an effect might be caused by the superposition of errors in determining the electrostatic interaction energy by the UBDB2011+EPMM method, which is related to the molecular mutual geometry and interacting atom types. The overestimation of the E_{es} might be resulting from a number of atoms in close contact calculated by the Exact Potential (EP) method on the basis of not perfectly accurate pseudoatom models. There are also particular random mutual configurations of the interacting nucleic acid bases for which the UBDB-derived E_{es}'s significantly differ from the *SPDFG* values. This may result from some specific orientations and covering of the multipoles which affect the proper E_{es} estimation (*e.g.* peak at 12th position, Figure 3.7*a*). This effect is most emphasized in the presence of charged species containing oxygen atoms. Therefore, it should be noted that multipole model might not be flexible enough to properly describe some of the electron density distribution features, *e.g.* electron lone pairs, which influences the derived energy values. On the other hand, it is worth stressing that the RMSD values between UBDB2011+EPMM and *SPDFG* are comparable to the electrostatic interaction energy differences derived by *SPDFG* with the use of 6-31G** and aug-cc-pVDZ basis sets (3.9 kcal·mol^{-1} and 4.6 kcal·mol^{-1} for A:T and G:C complexes, respectively). Even though Figures 3.6 and 3.7 clearly show that the UBDB2011+EPMM method tends to overestimate the interaction energy when compared to the *ab-initio* results (assuming the DFT method applied for the wave function evaluation may be considered as an *ab initio* approach), the overall performance of the UBDB2011 databank is rather satisfactory in predicting electrostatic energy of hydrogen-bonded nucleic acid base pairs. I found that both methods preserve the general trends in electrostatic interaction energy values according to mutual geometry variation of the analysed complexes (Figure 3.7). This is a very important result in the context of comparative studies of interacting macromolecular systems.

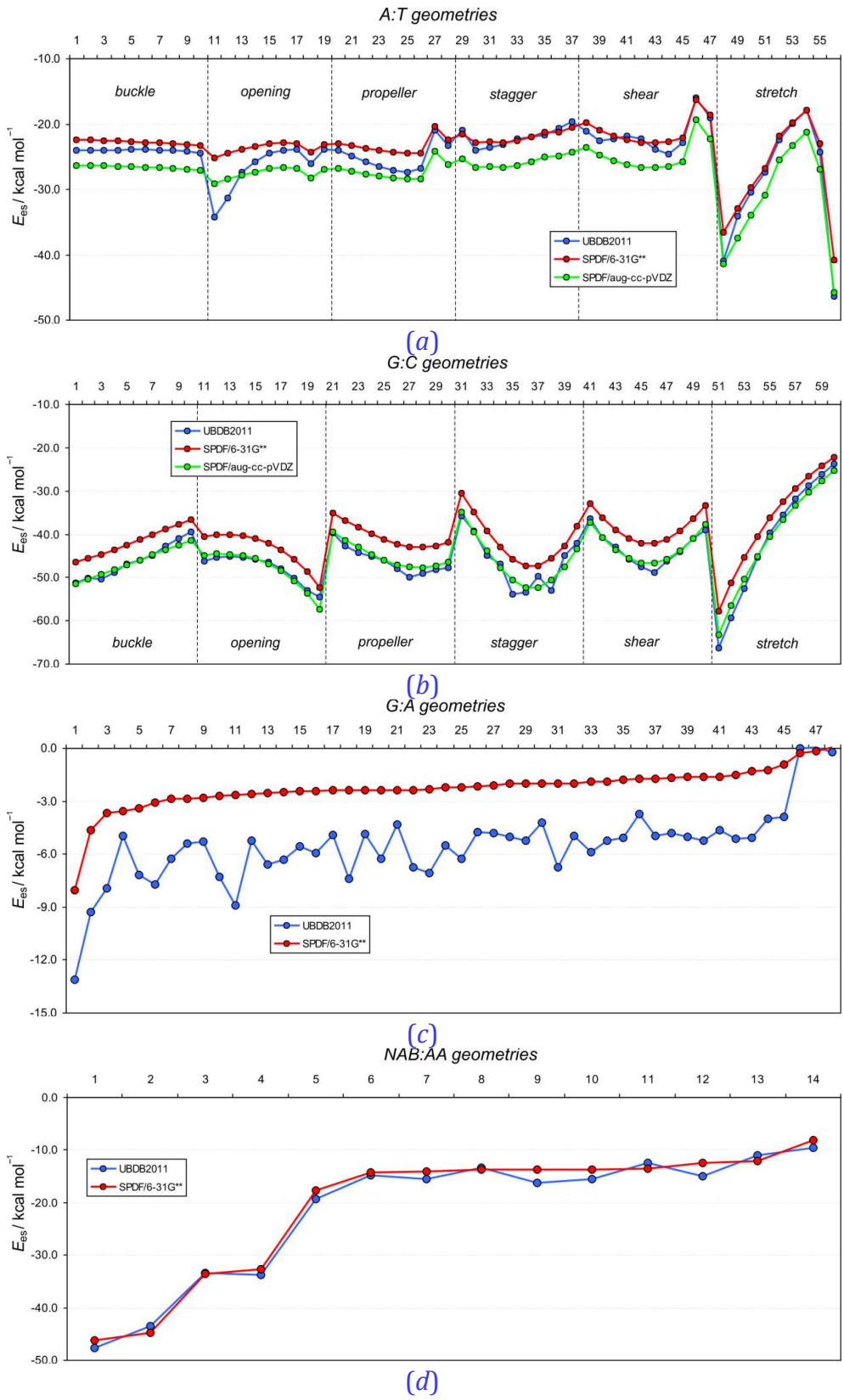


Figure 3.7. Energy trends. SPDFG and UBDB+EPMM comparison: (*a*) A:T; (*b*) G:C; (*c*) G:A; (*d*) NAB:AA.

The study also showed that in the case of G:A complexes the energetic variability is to some extent compatible for both methods (UBDB2011+EPMM and SPDFG). However, the magnitude of the UBDB2011+EPMM derived energy value and its deviation from the theoretical results, are comparable. This means, that such low E_{es} values obtained for the G:A dimers, characterised by more pronounced dispersive interaction contribution, are statistically meaningless. The UBDB2011+EPMM method works better for stronger electrostatic interactions around 50 – 70 kcal·mol^{-1}. When I took into account a small group of the amino acid species, it occurred that the UBDB2011+EPMM method exhibits the same accuracy as the previous version of the databank (R^2 = 0.99, RMSD = 1.5 kcal·mol^{-1}).[31b] Nevertheless, it still not clear to what extent such energetic discrepancies accumulate and affect the electrostatic interaction energies derived for larger molecular systems containing a number of amino acid or nucleic base pair residues.

To summarise this section, my results confirm sufficiently good quality of the pseudoatom parameters of the newly added atom types. The UBDB2011+EPMM method seems to be applicable for quantitative electrostatic interaction energy evaluation in the case of macromolecules but only if accuracy of 4 – 5 kcal·mol^{-1} will not discredit the results. If, for instance, the interaction energy values of a protein with different inhibitors are similar (*i.e.* differ by about 5 kcal·mol^{-1} or less), no meaningful conclusions can be drawn.

III.3.2. Remarks and limitations

Most of the atom types are satisfactorily transferable and described by means of Hansen-Coppens formalism, especially those consisting light chemical elements. However, in some cases, it is very difficult to define an atom type properly or even impossible to do it in a simple way. There are several factors which influence the atom type description, much of which depend on the CSD sample. If an atom is usually present in a group of similar molecules, its definition is more consistent and parameter deviations smaller. Among others, this was the case when I mentioned the specificity of 5-membered rings, being usually much more diverse (more often heterocycles) than 6-membred ones. This in turn decreases the quality of a simply defined atom type. When a particular atom type is averaged over structurally various and sometimes quite exotic molecules, its definition is statistically less accurate or it requires a special treatment.

Some of the atoms are more sensitive to their chemical environment, *i.e.* easier polarized, than others. This is especially true for carbon atoms bonded to nitrogen atoms. Here, the valency and nitrogen atom neighbours are of great importance.

What is quite straightforward for a chemist, aromatic systems or those containing coupled or alternated double bonds, can be particularly problematic, as they have labile π-electrons. Such molecules mark out the limits of atomic transferability. This is well visible in Figure 3.8. The simple example of pentachlorophenolate shows that having equal first, second and even third neighbours, does not guarantee the same set of multipole parameters. There is a significant alternation of P_v parameter values and all the multipole populations on the carbon atoms with chlorine substituents. The same situation is observed for the chlorine atoms themselves (P_v(*o*-Cl) = 7.23, P_v(*m*-Cl) = 7.19, P_v(*p*-Cl) = 7.31). This is a result of a chemically known phenomenon explained on the basis of resonance electronic structures. Therefore, even though the chemical environments of aromatic carbon atoms are practically equal, their multipole populations differ significantly as atoms in meta position usually exhibit more negative charges than *para* and *ortho* equivalents. To precisely describe those atom types a very complex atom type definitions should be applied. Figure 3.9 presents a well-described atom deposited in the UBDB2011 databank in contrast to the previously mentioned example.

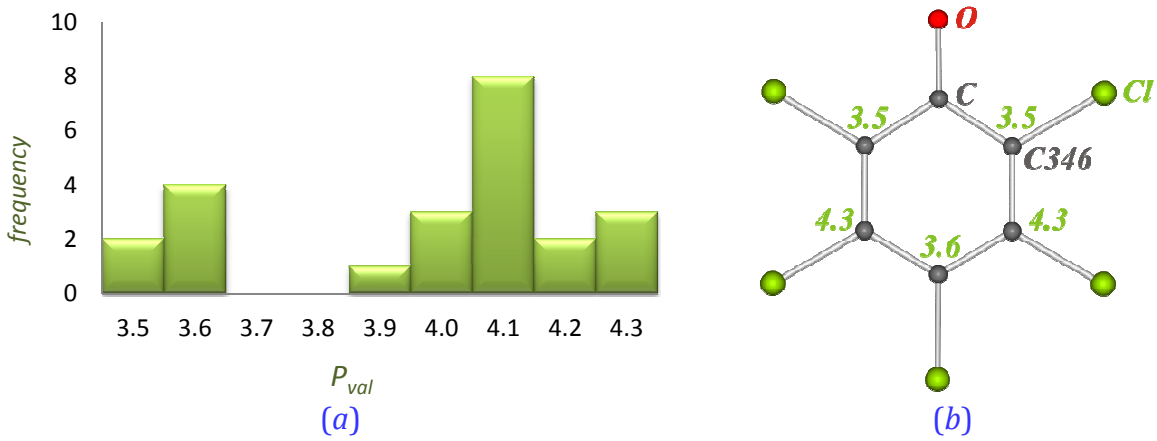
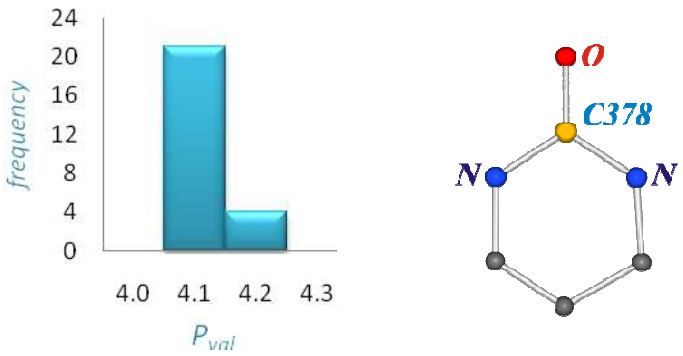


Figure 3.8. (*a*) C346 atom type P_v values distribution ($\sigma(P_v)$ = 0.27); (*b*) P_v alternation depending on carbon atom position in the aromatic ring of pentachlorophenolate. Atom colour coding: oxygen – red, carbon – dark grey, chlorine – green.



(*a*) (*b*)

Figure 3.9. (*a*) C378 atom type P_v values distribution ($\sigma(P_v)$ = 0.011). (*b*) Schematic representation of the C378 atom type. Atom colour coding: nitrogen – blue, oxygen – red, carbon – dark grey.

Aromatic systems may be very sensitive to substituent effect and therefore very difficult to model. However, the differences in multipole parameters are usually less pronounced than in the case of pentachlorophenolate. The effect of charge density distribution influenced by some further neighbours can be observed in the case of 5- and 6-membered rings.

Heterocyclic character of a ring, or different substituents among the calculated structures cause the ambiguity in atom type description, reflected in higher standard deviations. Similarly, the carbon atom types C311 – C313, being modelled on the basis of a set of atoms more sensitive to electronic effects due to the vicinity of double bonds, need improved atom type definitions.

The sample consistency and atom chemical environment are not the only problem which may be encountered. Equally important is the flexibility of the Hansen-Coppens formalism, and also satisfactorily derived atomic scattering parameters stored in the scattering tables assigned to particular chemical elements. The level of theory employed for the purpose of deriving molecular wave functions when constructing the databank, and Fourier truncation error are also limiting factors. Therefore, heavier chemical elements and/or those which are easily polarized may pose some problems in the future when further extending the databank. Here, especially in the case of possible addition of some metal atoms or ions, different approaches should be concerned to model these atoms properly as for example the modification of atomic core description proposed by Fischer *et al.*[21] Some problems are already visible in the case of sulphur atom which is generally difficult to be modelled and hardly ever the model reflects the reality, i.e. the

experimental data. This is going to be one of the subjects undertaken in the second part of my PhD thesis.

According to the above remarks, I would like to illustrate the influence of the ambiguous atom type definitions on the electrostatic properties on the example of amphotericin B (**AmB**) and its iodoacetyl derivative (**AmB-I**).[62c] During one of the side projects, I worked with a modification of a well-known antifungal agent, *N*-iodoacetylamphoterin B molecule (**AmB-I**). This compound contains 143 atoms, and thus it is at the edge of the DFT approach reasonable applicability. To investigate and compare the chemical character and nature of possible interactions concerning **AmB** and **AmB-I** molecules, I modelled their electrostatic potential.

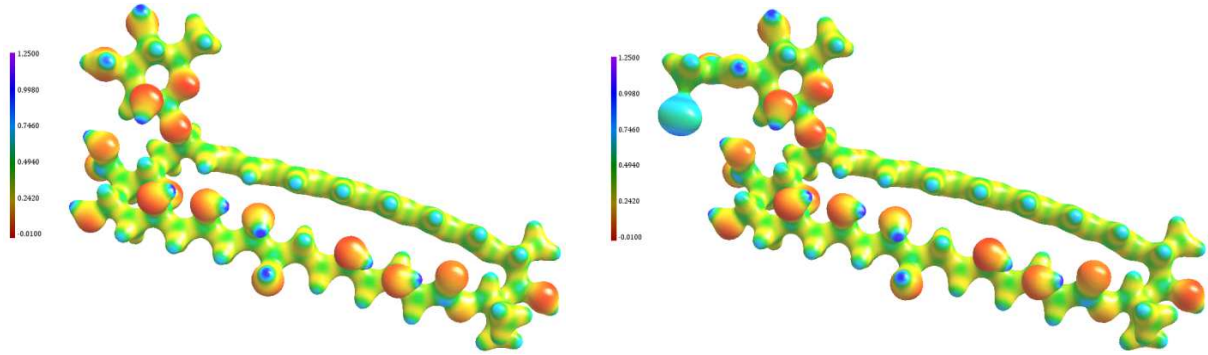
I tested three different ways of evaluating this electrostatic property. The first method employed Gaussian calculation to prepare charge density and electrostatic potential CUBE files at the DFT(B3LYP)/6-31G**(for iodine: 6-311G**) level of theory (140 grid points) (see Figure 3.10). The second method employed the UBDB2011 pseudoatom databank to construct charge density distribution for the studied molecules (see Figure 3.10*c,d*). Some atoms, *e.g.* iodine atom, were additionally modelled to supplement the latest version of the UBDB2011 databank.[70] To ensure electroneutrality, molecules were adjusted *a posteriori* to their net charges by scaling the pseudoatoms according to the Faerman and Price scaling algorithm implemented in *LSDB*.[46] Charge density distribution and electrostatic potential grids were generated by the *XDPROP* program from the *XD* package. In the third approach, **AmB-I** charge density distribution was calculated in a way it is done when preparing new entries for the UBDB databank (see Figure 3.10*e*). Therefore, the charge density evaluation was based on *GAUSSIAN* wave functions and the subsequently computed theoretical complex static valence-only structure factors. Such data was fitted with the Hansen-Coppens pseudoatom formalism in the *XD* program suite. Derived multipole parameters, together with κ and κ', were used to produce both grid files in *XD*. It is worth noting that in all three cases crystallographic geometry and conformation of the *N*-iodoacetylamphotericin B derivative served as a prototype of **AmB**, the crystal structure of which is known. All X–H distances were defined according to the newest article of Allen and Bruno[63] as hydrogen atoms are of great importance for estimating any electrostatic properties of an organic molecular system.

The **AmB** and **AmB-I** *GAUSSIAN* electrostatic potential maps derived at the (DFT)B3LYP/6-31G** (6-311G** in the case of iodine atom) level of theory are alike (Figure 3.10*a,b*). The only discrepancy appears in the *N*-fragment. The iodoacetyl substituent introduced into the **AmB** moiety influences the electrostatic potential of the N atom, shifting its potential towards more positive values.

In all other models, main qualitative features are preserved, however, the polyene fragment modelled by the UBDB2011 databank is more polarised (more negative electrostatic potential values closer to the headgroup fragment) when compared to the Gaussian results. It may result from ambiguous definitions of the atom types belonging to the alternated double bond chain. These atom types were already mentioned before, *i.e.* C111- C113. This, in turn, may lead to problems with scaling the molecular charge due to the Faerman & Price algorithm specificity (§II.3). When a certain atom is characterised by higher valence population *e.s.d.* value, then its atom charge is more labile during the scaling. Polyene chain polarisation may occur in a crystal lattice due to crystal field, or in solution, however, it is less likely to appear in the case of an isolated molecule. Consequently, it is rather an artefact of the method. This might be improved by dividing the molecule into fragments, which are then scaled separately.

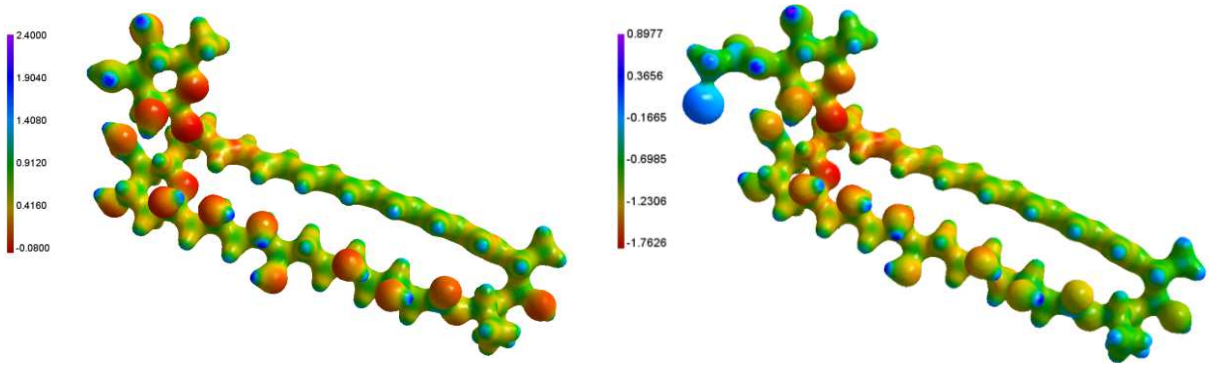
The results, which deviated most significantly from the ones obtained by use of the Gaussian package, were observed for the electrostatic potential of the **AmB-I** molecule obtained on the basis of the charge density distribution reconstructed with the aid of UBDB2011. Here, the total electrostatic potential got shifted to more negative values for the whole molecule. The reason for such results comes probably not only from the presence of polyene chain, which is easily polarisable, but also because of the presence of the iodine atom. The whole **AmB-I** molecule is noticeably specific, when compared to some small molecules containing an iodomethyl group used to create the databank entries. In general, P_v values for such an iodine atom type oscillate around 7.2. However, in the case of the third approach, it is equal to about 6.8. Next, during the molecule charge scaling procedure, the charge of the iodine atom was subjected to equilibration using statistically worse described carbon atoms from the conjugated double bond chain yielding less accurate charge density models. Furthermore the modelled iodine atom is characterised by a very low value of the κ' parameter (around 0.5). This was observed for the whole small-molecule dataset, when used for this atom type modelling. The low κ' value significantly spreads the iodine atom deformation density, and may affect its

surroundings, leading to peculiar molecular charge distribution, and, consequently, electrostatic potential. This may the result from using a more sophisticated basis set for the iodine atom, when compared to all other atoms, inappropriate radial functions, or deficiencies of the multipolar model and charge density refinement.



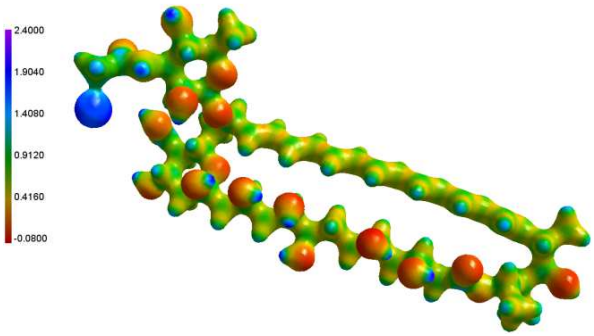
(*a*) (*b*)

Electrostatic potential mapped on the electron density isosurface of (*a*) **AmB** and (*b*) **AmB-I**, contoured at a value of 0.1 e·a_0^{-3} (≈ 0.67 e·Å^{-3}); single point B3LYP/6-31G** (6-311G** for iodine atom) calculation.[54]



(*c*) Electrostatic potential mapped on the electron density isosurface of **AmB** based on pseudoatoms from UBDB2011, contoured at a value of 0.65 e·Å^{-3}.

(*d*) Electrostatic potential mapped on the electron density isosurface of **AmB-I** based on pseudoatoms from UBDB2011, contoured at a value of 0.65 e·Å^{-3}.



(*e*) Electrostatic potential mapped on the electron density isosurface of **AmB-I** based on theoretical structure factors, contoured at a value of 0.65 e·Å^{-3}.

Figure 3.10. Visualisation of the electrostatic potential of **AmB** and **AmB-I** molecules derived with different theoretical methods.[78]

It seems that the UBDB databank is not accurate and flexible enough for proper description of iodine atoms and the polyene fragment. I also tested using separate scaling of different fragments of **AmB-I** molecule, but no mprovement was observed. On the whole, the described effects narrow the applicability of the databank to electrostatic energy estimation for more complicated, or exotic systems. However, it can be still satisfactorily applied as a source of aspherical atom scattering factors in the TAAM refinement.

III.1.8. Summary of the first part of my PhD

I have recalculated and extended the UBDB databank with atoms required to model RNA and DNA molecules. It currently contains over 200 atom types present in the most relevant biomolecules. I have also implemented new atomic keys which allow for distinguishing the geometry of multipole populations centred at atoms belonging to 5- and 6-membered planar rings, and properly defined atoms common for two fused planar rings, crucial in the case of purine bases. I have modified the LSDB program so as to provide adequate atom type definitions, local coordinate systems, and updated X–H bond distances. Additionally, I have corrected the CHEMCON module responsible for assigning common multipolar and contraction-expansion parameters to atoms of the same kind. Currently, *LSDB* writes also automatically input files suitable for the *MOPRO* suite of programs.

My test of the UBDB2011 pseudoatom databank applicability showed that the UBDB2011+EPMM method satisfactorily reproduces electrostatic interaction energies for a set of nucleic acid base complexes with respect to *ab initio* results ($R^2 > 0.9$, RMSD = 3.7 kcal·mol^{-1}). Correlations are high while energy trends are preserved. UBDB2011 can be therefore applied to estimate electrostatic interaction energies of macromolecular systems. However, one should be aware of the method's limitations. The databank does not describe conformational variety and does not take into account the crystal field influence and other subtle effects. It is also quite restricted by the accuracy of atom type entries (basis set used, DFT method, Fourier truncation error, Hansen-Coppens model limitations, atom type definition adequacy, etc.). Consequently, when applied for electrostatic interaction energy evaluation UBDB2011+EPMM results should be interpreted rather qualitatively than quantitatively. Meaningful comparisons may only be drawn when the electrostatic interaction energy difference for two chemical systems

exceeds 5 kcal·mol^{-1}. For more precise interaction energy computations and deriving other electrostatic properties (usually slightly less sensitive to the charge density model than electrostatic energy values), a new more sophisticated and powerful model is required, *e.g.* Koritsanszky *et al.*[79]

In view of the above, considering the currently achievable experimental data resolution and its quality, the presented approach is still accurate enough for the purpose of structure refinement. Despite its restrictions, the UBDB2011 constitutes a good source of aspherical atomic scattering factors, which may be used for TAAM refinement to enhance the quality of the final molecular structure. This is among the subjects undertaken in the second part of my PhD thesis.

"Minds are like parachutes. They only function when they are open"

Sir James Dewar

IV. Results and Discussion

IV.1. UBDB databank in practice

Once the tools had been prepared and extensively tested, they were ready to use. This part of my PhD project deals with basic applications of the databank, however, the presented analysis goes far beyond that. The UBDB2011 databank served as a starting point for investigations concerning electrostatics of influenza neuraminidase active site, as well as, for exploring energetic features of crystal lattices of a series of uracil modifications, and finally, for fine charge density and energy studies of a set of nucleic acid base derivatives. My investigations were an attempt to find the interrelations between molecular geometry, crystal lattice energetic features, charge density distribution, and crystal morphology. Additionally, results in this part of my thesis showed the databank capability to imitate molecular charge density distribution.

IV.2. Electrostatics on the basis of the UBDB databank

The UBDB databank was designed specifically to reconstruct molecular charge densities on the basis of pseudoatom parameters. This is especially important for macromolecular studies. Biological systems are usually out of the capability of fine quantum chemical methods. They are also very often difficult to be crystallised. Huge molecules have many labile fragments, and so the quality of most of the crystals is below expectations for experimental charge density studies. Therefore, to investigate energetic features of such systems some approximate approaches are needed.

The following chapter deals with the influenza neuraminidase inhibition aspects. The aim of this part of my thesis was, among others, to show an exemplary energetic study of macromolecular system on the basis of the pseudoatom databank. The whole project consisted of a comprehensive analysis of the electrostatic interaction energy between the protein, influenza neuraminidase, and a series of various inhibitors. This was a subject of the paper published by Dominiak *et al.*[80] I was directly involved in the investigations of water molecule contribution to the protein-ligand complex electrostatic stabilisation energy. Therefore, here, I will briefly describe influenza neuraminidase active site and present main conclusions regarding electrostatics and specificity of the inhibitor binding. Nevertheless, I shall focus particular attention on the solvent aspects.

IV.2.1. Influenza neuraminidase inhibition

Biologically active influenza neuraminidase (NA) is a homotetramer with about 470 residues falling on each monomer (Figure 4.1). A single subunit includes a globular head domain placed at the end of a long stalk. Despite the low sequence identity between head fragments of influenza neuraminidase type A and type B, which is approximately 30%, and so among subtypes of influenza neuraminidase type A – about 50%,[81] all known structures of the head fragment form the same fold.[82] They consist of six four-stranded antiparallel β sheets arranged in a right-handed propeller motif. The loops between β-strands arranged on one surface of the propeller form a hollow cavity, where the enzymatically active site is located.

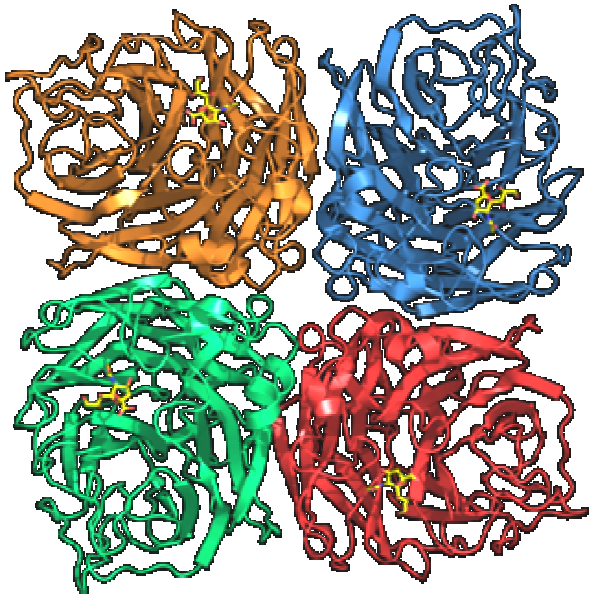


Figure 4.1. Schematic illustration of influenza virus neuraminidase. Each monomer of the homotetramer is differently coloured. Sialic acid molecules indicate the active sites of monomers.

A high affinity calcium-binding site is formed in the vicinity of the active site, whereas a low-affinity Ca^{2+} site is located along the fourfold axis of the tetramer and most probably contributes to the tetramer stabilisation. The active site is created by 18 highly conserved residues.[83] This region superimposes very closely for all NA strains (both for the main-chain and side-chain atoms). These 18 aminoacid residues are then divided into two groups, depending on their function in the substrate binding process. Eight of them, *i.e.* Arg118, Asp151, Arg152, Arg224, Glu276, Arg292, Arg371 and Tyr406, interact directly with the sialic acid fragment of a substrate and are called functional (shown in red, Figure 4.2). The remaining ten residues (Glu119, Arg156, Ser179, Trp178, Asn/Asp198, Ile222, Glu227, Glu277, Asn294 and Glu425) stabilise the active site and are called structural (shown in blue, Figure 4.2). Such a conservation of the active site of the influenza virus neuraminidase has become an attractive target for broad spectrum of anti-influenza drug design. Over the last two decades several potent and specific inhibitors have been developed.[84] The availability of crystal structures of inhibitor-neuraminidase complexes[82, 85] allows for a detailed analysis of the structural foundation of the inhibition process. Many computational methods have already been applied to predict inhibition efficiency.[86] The differences in inhibitor molecule recognition are, though, not apparent from the structure itself, as all inhibitors bind in the same pocket and interact with the same residues in a similar fashion.

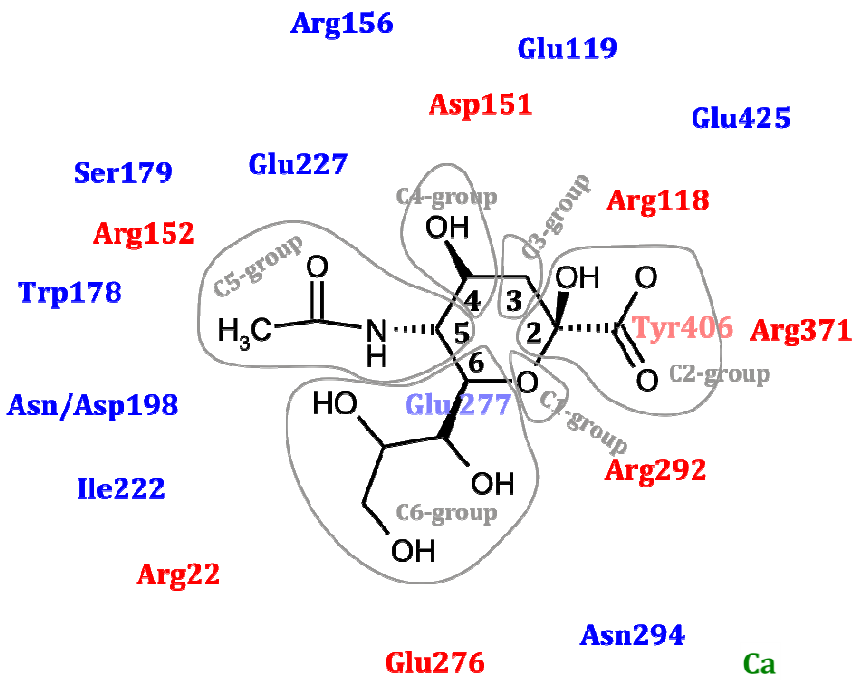


Figure 4.2. Schematic representation of the active site of the influenza virus neuraminidase and its interaction with sialic acid (SIA) fragments. Conserved functional residues are shown in red, conserved structural residues – in blue, and a calcium cation – in green. The six distinctive regions of SIA are shown in grey envelopes. The numbering scheme is as in 1mwe structure (N9 NA) (Appendix C, Tabl3 C1).

IV.2.2. Importance of electrostatics in biological systems

The importance of electrostatic interactions in the analysis of structure-function correlations of biological molecules has quite recently been re-emphasized.[87] Although electrostatic interactions contribute only a part of the interaction energy between macromolecules, unlike dispersion forces, they are highly directional, and therefore dominate the nature of molecular packing in crystals and in biological complexes. Influencing the relative orientation of the enzyme and substrate,[88] they may govern molecular recognition and substrate specificity. Accordingly, the electrostatic component of the interaction energy between influenza neuraminidase and its inhibitors has received extensive attention.[85f, 86e, 87-89]

IV.2.3. Studied systems and methods

In the reported study, a wide range of complexes of influenza neuraminidases with inhibitor molecules (sialic acid derivatives and others) have been analyzed using the reconstructed charge density distributions on the basis of the transferable aspherical-

atom databank, UBDB. 36 crystal structures of type A NA-inhibitor complexes (nine N2, three N6 and twenty four N9) and 11 crystal structures of type B NA-inhibitor complexes were retrieved from the Protein Data Bank.[90] All the studied inhibitors are schematically illustrated in Figure 4.3. The PDB codes of 70 complexes, the protein subtypes, mutated residues, inhibitor codes and protein and inhibitor charges together with inhibition constants (K_i), when available, or IC50 indexes are given in the Appendix C (Table C1).

The use of the UBDB databank allows for quite reliable estimation of total electrostatic interaction energy, but also for deriving contributions coming from single amino acid residues and inhibitor functional groups. The analysis employed the previously described EPMM approach (§II.2.4). Here the boundary of the inner region was set up to 4.2 Å.

IV.2.4. Summary of the results for waterless structures

The main investigations of the influenza neuraminidase complexes with different inhibitors were based on the available crystallographic structures, where all the solvent molecules were excluded. Water molecules were not taken into account because of the high uncertainty in the hydrogen atom positions. However, it should be pointed out that a comprehensive analysis of the enzyme-substrate interaction should include the contributions from solvation and desolvation of the interacting species as well as the accompanying entropic factors.

This section recalls the main conclusions coming out of the NA inhibition study. Figure 4.4 shows the total electrostatic interaction energy between particular protein types and various ligands. As the majority of the active site residues are charged, NA forms strong electrostatic interactions with its inhibitors. Our results show that the total electrostatic interaction energies of inhibitor molecules interacting with neuraminidase (amino-acid residues plus proximal Ca^{2+} ion; E_{es}) range from –164 to –970 kJ·mol^{-1}, with an average value of –600 kJ·mol^{-1}. The most negative E_{es} values are observed for fully deprotonated phosphonate analogs of SIA (AXP and EQP) in complexes with particular NA types (–937, –916 and –970 kJ·mol^{-1} for AXP-N2, EQP-N2 and EQP-B, respectively). Stronger than average electrostatic interactions are observed for almost all 2,3-dihydropyranosidic ring-based inhibitors (*e.g.* –949 kJ·mol^{-1} and –918 kJ·mol^{-1} for the 4AM-N9 and 49A-N9 complexes, respectively) and for G39 and BCZ.

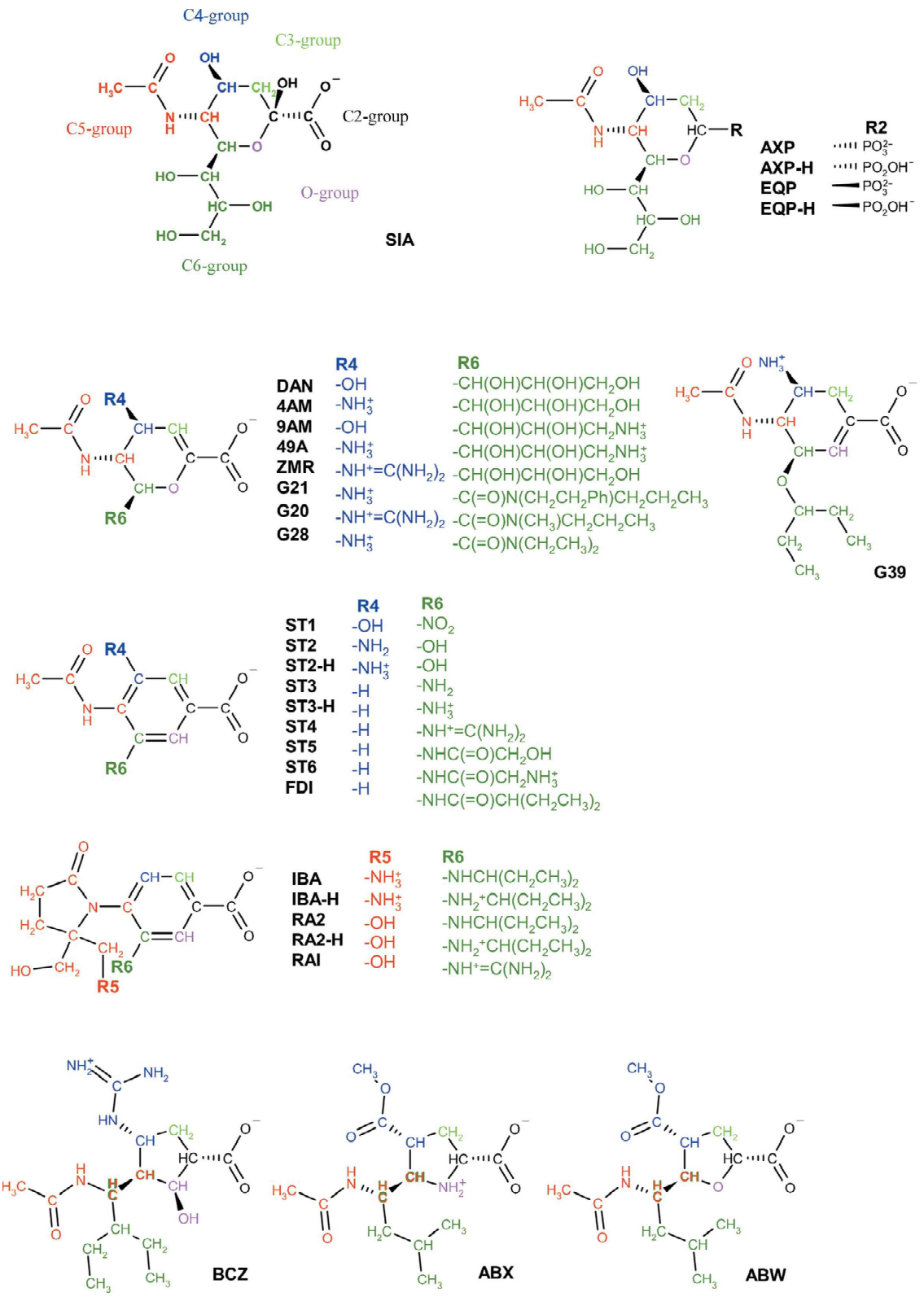


Figure 4.3. Structures of inhibitors. The colours define six distinctive regions of a given inhibitor: the C2-group (black), C3-group (light green), C4-group (blue), C5-group (red), C6-group (green) and O-group (magenta).[80]

The least negative values of E_{es} are found for NA interacting with inhibitors with aromatic central rings (*e.g.* –164 kJ·mol^{-1} for the ST3-N2 complex and –349 kJ·mol^{-1} for the ST1–N2 complex). Some NA-type dependent differences in the electrostatic energy values are also observed. The absolute values of E_{es} for EQP/EQP-H interacting with N9 NA, for example, are 20 – 35% smaller than for N2 or B NA complexes. This observation qualitatively agrees with experimental results showing a ten times weaker inhibitory potency of EQP against N9 when compared with other NA strains. Previously, it had been attributed to a possible difference in the dynamic properties of N9 NA in solution.[85p] On the other hand, complexes of SIA or DAN show systematically more pronounced negative values of E_{es} for type B NA than for type A NA. Nevertheless, there is no experimental evidence supporting any substantial difference in the inhibitory activities of DAN (or SIA) against type A and B neuraminidase. As no structural information on other inhibitors interacting with different types of neuraminidase is available in the PDB, no general conclusions on the dependence of the electrostatic interaction energy on NA type can be drawn. Further details concerning different aspects and results of the analysis are available in the Dominiak *et al.* publication.[80]

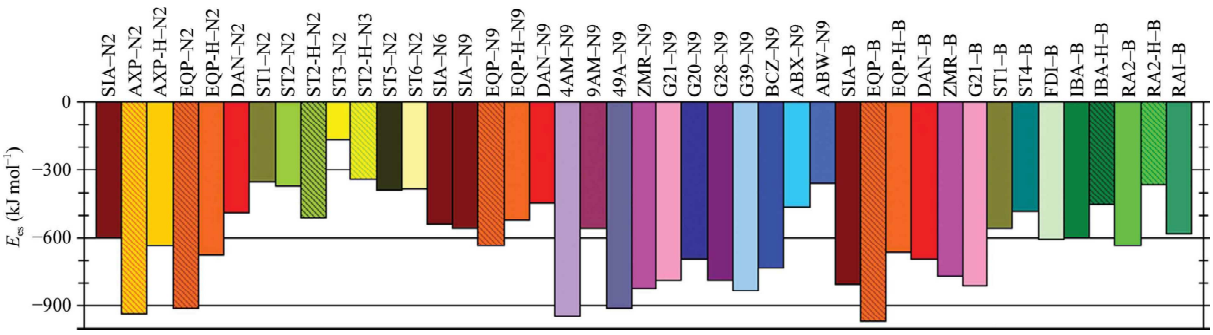


Figure 4.4. Electrostatic interaction energies (E_{es}) of inhibitor–neuraminidase complexes.

Having a careful look solely at the active site, in general, the study showed that the conserved residues in the active site contribute as much as 85% of the electrostatic interaction energy of inhibitor-neuraminidase complexes. In the case of inhibitor molecules, there can be selected six distinct molecular regions as shown in Figure 4.2 on the example of sialic acid. Naturally, the acidic, and usually charged, C2-groups make the largest contributions to the electrostatic interactions. The interactions of the carboxylic group in the non-aromatic inhibitors with the conserved residues sum up to about –300 kJ·mol^{-1}. For the same fragment of the aromatic inhibitors, the interaction energy is 30 –

60% lower. However, the protein-ligand binding is by 50% stronger in the case of inhibitors without the central aromatic ring, but with the fully deprotonated phosphoric group. Additionally, positively charged substituents of the C4-group further increase the electrostatic interactions with NA active site by –250 kJ·mol^{-1}. Interestingly, both amino and guanidine fragments make similar total contributions, although they interact differently with each of the conserved residues. Glycerol side chains at C6 contribute *ca.* –120 kJ·mol^{-1}. The same holds true when one of the hydroxyl substituents of glycerol is replaced by an amino group. The electrostatic contribution from non-glycerol-like C6-groups is much smaller. The C5 groups, which all contain the acetamide fragment, contribute –90 kJ·mol^{-1} to the electrostatic interactions with conserved residues. The contributions from the remaining inhibitor groups, C3 and O, are negligible.

Another carefully analysed problem was the Arg292 ⇨ Lys mutation. Generally, the resistance of the mutant is greater to inhibitors in which the glycerol moiety at the C6-position has been replaced by hydrophobic chains (G20, G28, G39 or BCZ) (Appendix C, Table C1). Indeed, although the mutation reduces the electrostatic interactions of the enzyme with the acidic group at C2 for all inhibitors that have been studied (SIA, DAN, 4AM, ZMR, G20, G28, G39 and BCZ), it enhances the interactions with the glycerol group at C6 when present.

IV.2.3. Water molecules in the active site

As stated before, in general water molecules were not taken into account. However, as some of the water molecules in the active site can significantly contribute to the protein-ligand interactions, I carried out additional calculations for the four selected complexes, *i.e.* of N9 with ZMR, SIA, DAN and 4AM (PDB codes 1nnc, 1mwe, 1f8b and 1f8c, respectively), including three frequently considered in the literature water molecules (labelled **W3**, **W4** and **W5**)[89c, 91] (Figure 4.5). I took the O-atom positions directly from the PDB files and set the H–O bond lengths and the H–O–H angles to the values characteristic for ice crystals, *i.e.* 0.965 Å and 108°, respectively. I oriented the water molecules such as to maximize the electrostatic interactions. For that purpose, I rotated them around the oxygen atoms in the planes defined by the water O-atoms and the water hydrogen atom acceptor atoms of amino acid residues, verifying the resulting electrostatic interaction energies of each orientation. The final geometries are illustrated in Figure 4.5.

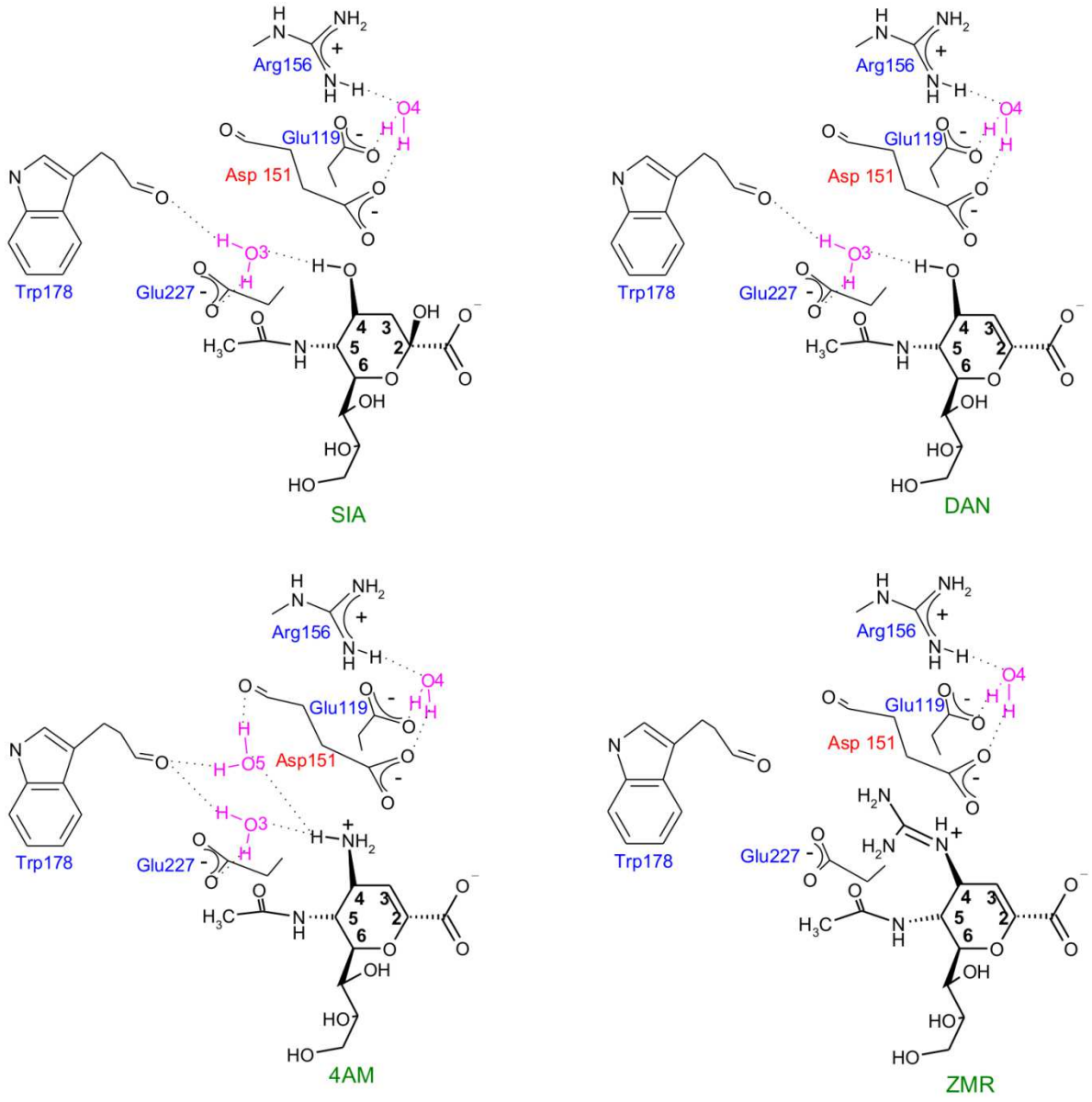


Figure 4.5. Schematic representation of the interactions of the C4 group of the SIA, DAN, 4AM and ZMR inhibitors with N9 neuraminidase and the surrounding water molecules.

The four complexes of N9 with the 4AM, DAN, SIA and ZMR inhibitors differ mainly due to the inhibitor substituents at the C4 position. In order to explore the influence of water molecules on the active site, I considered three well ordered water molecules in different bonding environments (**W3**, **W4** and **W5**) located in the active site of neuraminidase, adjacent to the mentioned C4 inhibitor group (Figure 4.5). The **W3** H-atoms participate in hydrogen bonding to the carbonyl O atom of Trp178 and carboxylic group of Glu227, whereas **W3** accepts a hydrogen atom from the C4 group of the inhibitors. **W4** forms hydrogen bonds with the carboxylic groups of Glu119 and Asp151, while one of the **W4** lone electron pair points to an NH_2^+ hydrogen atom of the guanidine

group of Arg156. **W5** hydrogen atoms bond to the carbonyl O atoms of Trp178 and Asp151, while the **W5** oxygen atom is the acceptor of hydrogen bond created with the NH_3^+ of the C4 group. The number of water molecules in the complexes varies with the character of the C4 functional group and its volume. Only the 4AM-N9 complex shows all three water-bonding arrangements, whereas both **W3** and **W4** are present in DAN-N9 and SIA-N9 and **W4** is the only water molecule present in the ZMR-N9 complex.

Table 4.1. Electrostatic energies of interactions between water molecules (**W3**, **W4**, or **W5**), the protein (**P**, all aminoacid residues) and the inhibitor (**I**) (energy values are given in kJ·mol^{-1}).

Complex	**P-W3**	**I-W3**	**P-W4**	**I-W4**	**P-W5**	**I-W5**	**W3-W5**
N9–SIA	−73	−34	−174	−17	–	–	–
N9–DAN	−74	−29	−190	−18	–	–	–
N9–4AM	−74	−35	−171	+22	−68	−67	+13
N9–ZMR	–	–	−167	5	–	–	–

The **W4** water is observed in all complexes. This is mainly because it interacts more strongly with the protein than **W3** and **W5**, and rather weakly with the ligand molecule (which eliminates the extent to what inhibitor substituent modifications affect the water molecule binding). The **W4**-protein electrostatic interaction energy is in the range from −167 to −190 kJ·mol^{-1}, while it is about −74 kJ·mol^{-1} for **W3**, and −68 kJ·mol^{-1} for **W5** (Table 4.1). On the other hand, the **W4** interaction with the inhibitor is the weakest and depends on the nature of the substituent at C4. In turn, the **W3** water molecule is present in complexes of inhibitors with a small substituent at C4, *i.e.* hydroxyl or amino groups (SIA, DAN and 4AM). In the all three cases, **W3** interacts similarly with the molecular surroundings (−74 and −33 kJ·mol^{-1} on average for **W3**-protein and **W3**-inhibitor interactions, respectively). The **W5** molecule, which is only found in the complex in which a small and positively charged group is present at C4, exhibits the highest attractive interaction with the inhibitor (−67 kJ·mol^{-1}) among the analyzed water molecules.

As pointed out previously,[86d, 91c] the expulsion of water **W3** (and probably **W5**) from the active site of neuraminidase by the guanidine group of ZMR increases the binding interaction through an entropy gain. This is consistent with the fact that ZMR is a slow binding inhibitor.[92] The electrostatic energy lost by water dissociation upon binding ZMR in the active site equals about 130 kJ·mol^{-1} for **W3** and **W5**. This value is still lower than the gain in the electrostatic protein–inhibitor interaction energy upon

the replacement of DAN (or SIA) with the ZMR inhibitor, which is 263 kJ·mol^{-1} (or 166 kJ·mol^{-1}). This suggests that the ZMR complex not only interacts more strongly with the protein molecule than other inhibitors, but also that this interaction is strong enough to outweigh the energy lost on water dissociation. While the electrostatic gain of water binding is not negligible, its effect is balanced by an entropy gain on dissociation. To evaluate the net effect of water binding, both factors must be taken into account. Therefore it is difficult to explain the nature of binding between N9 and 4AM inhibitor molecule solely on the basis of electrostatic factors. When analysing electrostatic interaction energy, it turns out that this ligand is a better inhibitor than zanamivir, whereas its measured activity is weaker.

IV.2.4. Remarks

All the above results clearly show that the electrostatic interaction is a non-negligible component of the interaction energy governing the strength of neuraminidase inhibitors and must be understood if more potent inhibitors are to be designed. What is important, UBDB databank enables considering the asphericity of the atoms. Unlike point-charge models, it accounts for the directionality of the atom-atom interactions. The model is applicable to microscopic analysis of structure-function correlation in biological molecules.[87]

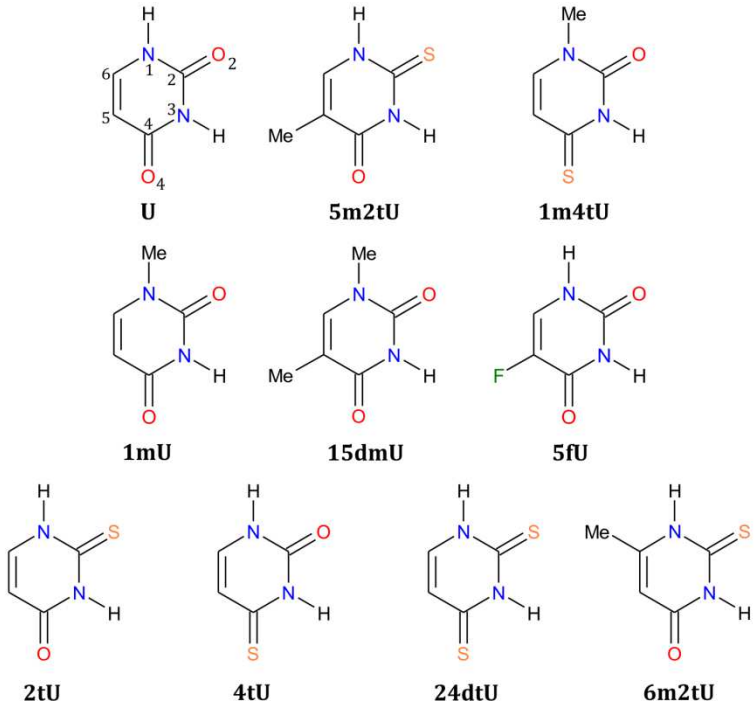
IV.3. UBDB as a source of aspherical atomic scattering factors

In this chapter, I shall present the second main application of the UBDB databank. This time pseudoatom multipolar parameters stored in UBDB2011 serve as a source of aspherical atomic scattering factors. As the UBDB2011 databank is an extended version towards nucleic acid molecule modelling, I have chosen for the purpose of tests, and further investigations, a set of uracil derivatives (nucleic acid bases and their modifications). The study contains a detailed geometrical analysis, which illustrates the quality of crystal structures obtained from the TAAM refinement procedure. Nevertheless, I paid careful attention to the energetic aspects of a series of uracil derivatives, and to relations linking molecules with the crystal architectures they form. The whole study constituted the subject of this year's article.[93] This contribution was the first complex attempt to characterise a cross-sectional set of uracil derivatives both, energetically and geometrically (*i.e.* quantitatively and qualitatively).

IV.3.1. Selected compounds- uracil derivatives

Purine and pyrimidine bases, being nucleic acid components as well as constituting the integral parts of other biologically relevant molecules, have been of constant scientific interest ever since they were discovered.[72-73, 76, 94] The biological and physicochemical properties of these compounds are strongly related to the ability of molecules to form hydrogen bonding and also to participate in π-stacking interactions.[94k, 95] It occurs that a bonding potential characteristic can be relatively easily influenced by various substituents introduced into aromatic rings, or when oxygen atoms are replaced by heavier analogues, *i.e.* sulphur or selenium atoms.[96] Some of such nucleic acid base derivatives were already found in biological organisms, while others were obtained artificially. They generally resemble the natural bases, although they may substantially affect the properties of the nucleic acids when built into their molecules.[94b, 94c, 94f, 97] Therefore, particular purine and pyrimidine modifications exhibit important biological activity and also some of them have found pharmaceutical applications as clinically useful drugs.[94h, 94j, 97a, 98]

For the purpose of the study, I have chosen a series of uracil derivatives shown in Scheme 4.1. These were supplemented by uracil (**U**),[99] 1–methyl–4–thiouracil (**1m4tU**)[100] and 5–methyl–2–thiouracil (**5m2tU**)[101] structures taken from the Cambridge Structural Database (CSD).[41]



Scheme 4.1. Studied uracil derivatives with their abbreviations (general numbering scheme is included).

Apart from their important medical applications and biological activity, especially of **5fU**,[94b] commonly applied in anti-cancer chemotherapy, and numerous sulphur-substituted pyrimidines,[94d, 102] uracil modifications possess interesting photochemical properties. Consequently, they have been extensively investigated by means of spectroscopy,[96a, 96b, 96e, 103] computations,[96c, 96d, 104] and biological activity examination.[94c, 94d, 94f, 97b, 102, 105] Nevertheless, the majority of the selected compounds had been structurally evaluated several decades ago.[94e, 99-100, 106] Only **5fU** and **5m2tU** have caught crystallographic attention recently, so there is some newer crystallographic data available,[101, 104a] and thus – more accurate molecular geometries. Additionally, neutron experiments were carried out in the case of **15dmU** (**1mT**)[107] and **1mU**,[108] while charge density studies were performed for **2tU**[94i] and **1mU**,[109] and so, the corresponding data are also available. Exclusively the **4tU** crystallographic structure has not been deposited in the CSD yet, although it was a subject of quite a few studies.[94c, 97b, 102, 104b, 105, 110] This is probably due to the known structure of its nucleoside.[110e] On the other hand, **6m2tU**, an analogue of a well-known antihyroid drug, had been considered with less care in the literature than the other studied systems.[94e, 111]

IV.3.2. Materials and methods

This section contains a brief description of growing crystals and further X-ray data collection and structural refinement (both IAM and TAAM). Additionally, I am going to present the details of the calculations performed for the purpose of my investigations.

(i) Crystallisation. All of the selected substances dissolve in water and, upon evaporation at room temperature, form crystals suitable for X-ray examination. Crystallizations were also adjusted in a set of different solvents such as methanol, ethanol, propanol, butanol, acetone and chloroform, at various evaporation surfaces. Nevertheless, it occurred that the best crystals were grown from water and ethanol solutions. Therefore, for the purpose of this study, I examined the crystals obtained from water solution. (*) Crystallisations adjusted by my MSc student, Marcin Kubsik.

(ii) X-ray data collection. Single-crystal X-ray measurements of **24dtU** and **4tU** were carried out on Bruker AXS Kappa APEX II Ultra diffractometer equipped with TXS rotating anode (Mo-K_α radiation, λ = 0.71073 Å), multi-layer optics, and Oxford Cryosystems nitrogen gas-flow device (700 Series Cryostream). Measurements of all other studied uracil derivatives, *i.e.* **15dmU**, **1mU**, **2tU**, **5fU** and **6m2tU**, were performed on an Agilent Technologies KM4CCD κ-axis diffractometer (recently upgraded with Opal area detector) equipped with a Mo-K_α sealed tube, graphite monochro mator, and Oxford Cryosystems nitrogen gas-flow device (600 Series Cryostream). In all cases, single crystals of suitable sizes were mounted on a cryo-loop using *Paratone N* oil and positioned at 50 mm distance from the CCD camera. Data collection strategies, based on ω scans, were optimized and monitored applying the appropriate algorithms implemented within the *APEX2*[112] or *CRYSALIS*[113] suits of programs, respectively. The unit cell parameter determination and raw diffraction image integration were performed with the diffractometer software (*APEX2* or *CRYSALIS*). Data sets were corrected for Lorentz, polarization and oblique incidence effects. The multi-scan absorption correction, frame-to-frame scaling and merging of reflections were carried out with the *SORTAV* program.[15] Final data collection parameters are listed in the Appendix D (Comment D2).

(iii) Structure solution and refinement. As the first approximation, all structures were solved by direct methods using *SHELXS-97*[114] and refined with *SHELXL-97* within the IAM (Independent Atom Model) model. The refinement was based on F^2 for all reflections, except for those with high negative F^2 values. Weighted R-factors and all

goodness-of-fit values are based on F^2. Conventional R-factors are based on F with F set to zero for negative F^2. The $|F^o|^2 > 2\sigma(|F^o|^2)$ criterion was applied for calculating R-factors, but was not made relevant to the choice of reflections intended for the refinement. The values of R-factors based on F^2 are about twice as large as the corresponding ones based on F. Scattering factors were taken from the International Tables for Crystallography.[4] All non-H atoms were refined anisotropically. Final statistics of spherical refinement together with comments specific for each structure are provided in the Appendix D (Comment D2,Table D3).

The second step concerned the Transferable Aspherical Atom Model (TAAM) refinements. I performed all those refinements in the *MOPRO* package[35b] combined with the new version of the University at Buffalo Data Bank (UBDB2011),[31a, 70] The idea of TAAM refinements has been already described in §II.2.4. Refinements in the *MOPRO* program were based on F, taking into account solely the reflections fulfilling the $|F^o|^2 \geq 3\sigma(|F^o|^2)$ condition, and applying statistical weights. Initial atomic coordinates x, y, and z, and anisotropic displacement parameters (U_{ij}) for each atom were taken from the spherical refinement stage, whereas initial multipolar and contraction-expansion parameters were transferred from UBDB2011 using the *LSDB* program, which also assigns optimal local coordinate systems for each atom, scales the fragment charges, and standardizes the X–H bond lengths (X = non-hydrogen atom) to average neutron distances. *MOPRO* allows for the application of specific restraints during the refinement. Therefore, in all the refinements carried out, the hydrogen atom U_{iso} parameters (*i.e.* isotropic thermal parameter) were restrained to the value of $y \cdot U_{eq}^{X}$ (y = 1.2 and 1.5 for X = C and X = N, respectively) with σ = 0.01 (where an appropriate restraint weight is equal to $1/\sigma^2$). The overall refinement strategy was as follows:

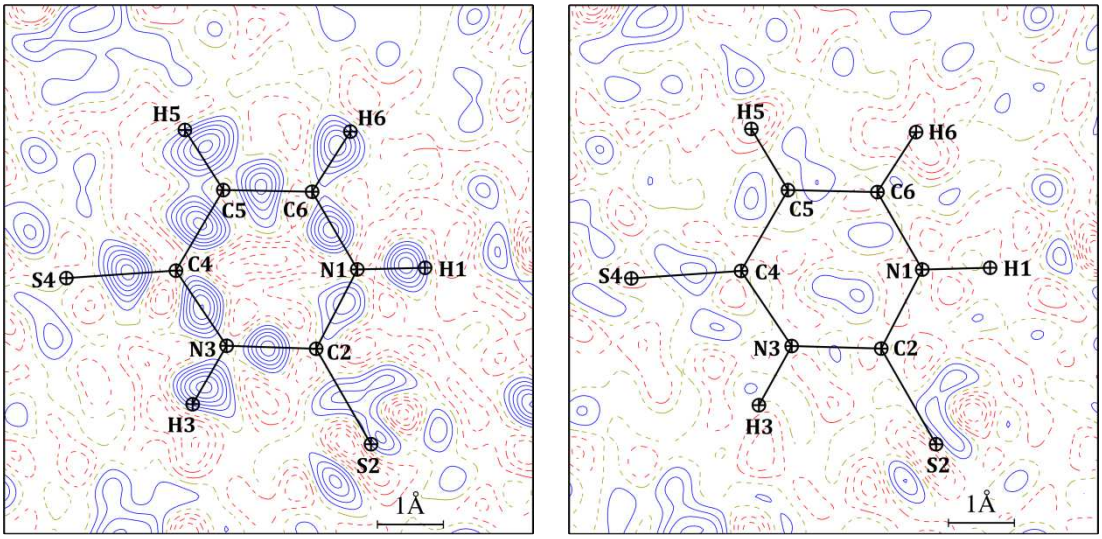
(1) scale factor;

(2) scale factor and atomic coordinates;

(3) scale factor, atomic coordinates and thermal parameters.

Simultaneous refinement of all parameters, just after UBDB2011 transfer, was in some cases not possible. Several models with different X–H bond length restraints and sulphur atom description were tested. However, no significant changes were observed, neither in R-factor values, nor residual density plots. Direct comparison of the resulting geometries (bond lengths and angles) shows differences within the estimated standard deviation. The modified sulphur scattering factors proposed by Dominiak and

Coppens[68] did not improve any of the tested models, thus the standard one was used, as included in the *MOPRO* library (ζ = 3.851 a_0^{-1}; n_l = 4, 4, 6, 8 for l = 1, 2, 3, 4, respectively). Test refinements with freely refined hydrogen atom positions lead in some cases to a large (> 0.02 Å on average) underestimation of the X–H bond length. Therefore, in the final model, which I used for the purpose of further analyses, X–H bonds were restrained to a standardized neutron distance with σ = 0.001 (for all compounds). Additionally, each atom was assigned with core and spherical-valence scattering factors derived from Su and Coppens wave functions.[69] It is worth noting that in comparison with IAM refinement, the residual electron density is significantly lowered in the case of each studied compound (Figure 4.6). UBDB2011 transfer clearly makes the residual density increasingly featureless.



(*a*) (*b*)

Figure 4.6. Example residual electron density map (up to sin θ/λ = 0.70 Å^{-1}) for **24dtU** (*a*) before and (*b*) after UBDB2011 parameters transfer (both with subsequent refinement of positional and thermal parameters; 6-membered ring plane; 0.05 e·Å^{-3} contours, blue solid lines – positive, red dashed lines – negative).

(iv) Energy computation details. I performed the energy computations with the *CRYSTAL09*[57] program package at the DFT(B3LYP) level of theory.[52, 65] The 6-31G** molecular all-electron basis set[64] was found to be sufficient for the purpose of the conducted computations.[115] Both the Grimme dispersion correction and correction for basis set superposition error (BSSE) were applied.[59, 61] Ghost atoms were selected up to 5 Å distance from the considered molecule in a crystal lattice, and were used for the BSSE estimation. The evaluation of Coulomb and exchange series was controlled by five thresholds, set arbitrarily to the values of 10^{-7}, 10^{-7}, 10^{-7}, 10^{-7}, 10^{-25}. Shrinking factor

was equal to 8, which refers to 170 **k**-points in the irreducible Brillouin zone and assures the full convergence of the total energy. The cohesive energy (E_{coh}) was calculated following the procedure described in the literature[61] and also briefly recalled in the Appendix A (Comment A3):

$$E_{\mathrm{coh}} = \frac{1}{Z} E_{\mathrm{bulk}} - E_{\mathrm{mol}} \tag{IV.1}$$

where E_{bulk} is the total energy of a system (calculated per unit cell) and E_{mol} is the energy of an isolated molecule extracted from the bulk (with the same geometry as in the crystal phase). Z stands for the number of molecules in the unit cell. In order to test the impact of differently derived molecular geometries on the resulting cohesive energy values, three different data set types served as an input for the *CRYSTAL* periodic calculations. In the first approach, I took coordinates directly from the IAM refinement stage. In the second approach I employed crystal geometry obtained in the TAAM procedure. Finally, I optimised the crystal structure in the *CRYSTAL* program at the DFT(B3LYP)/6-31G** level of theory and then I used it for the energy computations (referred to as OPT-3D). I kept cell parameters fixed during the optimisation process, and let the atom positions vary. Such proceedings were applied due to the used basis set, as it may cause underestimation of the optimised cell dimensions, and thus lead to false molecular geometry.

I used the same sets of coordinates for interlayer interaction computations. All of the calculation parameters were set identically as for the cohesive energy estimation. The only difference was introduced to ghost atom definition. Here, an additional upper and lower molecular layer were used as ghost function sets in order to obtain BSSE. The interlayer interaction energy (E_{intl}) calculation formula is analogous to that for cohesive energy:

$$E_{\mathrm{intl}} = \frac{1}{n} E_{\mathrm{bulk}} - E_{\mathrm{slab}} \tag{IV.2}$$

where E_{slab} is the energy of a molecular slab extracted from the bulk, while n indicates the slab number per unit cell. Therefore the resulting stabilisation energy falls on a part of a slab belonging to the unit cell. *CRYSTAL* automatically assigns the slab group symmetry and cuts out the repeatable fragment. In the studied cases these contain four molecules, except for **2tU** (two molecules). To allow a direct comparison, I standardised the **2tU** slab interaction energy to a four molecule fragment, *i.e.* the energy result was multiplied by two.

Additionally, I calculated selected dimer interactions for the earlier optimised geometries in the *CRYSTAL* package, taking into account both BSSE and dispersive corrections. Furthermore, I carried out two other geometry optimisations at the DFT(B3LYP)/6-31G** level of theory in order to evaluate the deformation energy values. These were molecule geometry optimisation in a 2D slab (only the molecular coordinates were varied), and for an isolated molecule.

(v) Atomic charges and NICS values. To supplement the analysis, I decided to compute atomic charges, which may provide some information of the relative basicity/acidicity of atoms. This, in turn, may contribute to the knowledge of preferable intermolecular contact formation in the crystal lattice. Therefore, I carried out single point DFT calculations using the *GAUSSIAN09* suite of programs.[54] I applied the B3LYP hybrid exchange correlation functional together with the aug-cc-pVDZ basis set.[75b] Atomic partial charges were determined using the Merz-Kollman-Singh[116] fit to electrostatic potentials and are available from the Appendix D (Table D9).

Finally, I thought of examining the aromaticity indices, *HOMA* and *NICS*, for the studied compounds. *HOMA* should reflect geometry deviations among molecules, while *NICS* may indicate the electronic similarities and differences within the investigated series. I evaluated NICS (Nucleus-Independent Chemical Shift)[117] aromatic indices , both at the ring centroid and 1 Å above the plane of the aromatic ring, at the same level of theory and basis set as the ones used for atomic partial charges. I derived both atomic charges and *NICS* indices on the basis of the molecular geometries optimised in a rigid crystal lattice (OPT-3D).

IV.3.3. Basic structural features

Most of the selected compounds have already been a subject of crystallographic discussion. Therefore in this section I do not put the attention on the structural details, but I rather focus on the common features characteristic for the chosen series. Basic structural and atom charge relations will point out problems requiring a closer look, and thus deeper analysis.

All of the studied uracil derivatives crystallise in centrosymmetric crystal space groups, most commonly forming $P2_1/c$ crystal lattice type. Among the measured crystal structures solely **5fU** and **2tU** adopt $P\bar{1}$ crystal settings, while **1mU** constitutes the only orthorhombic *Ibam* space group. The higher space group symmetry in the case of **1mU**

results from the particular location of molecules on a mirror plane. Although all the other molecules of the studied uracil derivatives may also belong to the *m* point group symmetry themselves, in a crystal lattice this symmetry is broken. Molecular geometries and ADP representations resulting from the TAAM refinement are shown in Figure 4.7. In all the cases, except for **5fU**, there is one molecule in the asymmetric part of the unit cell. Structural details are available from Appendix D (Comment D2).

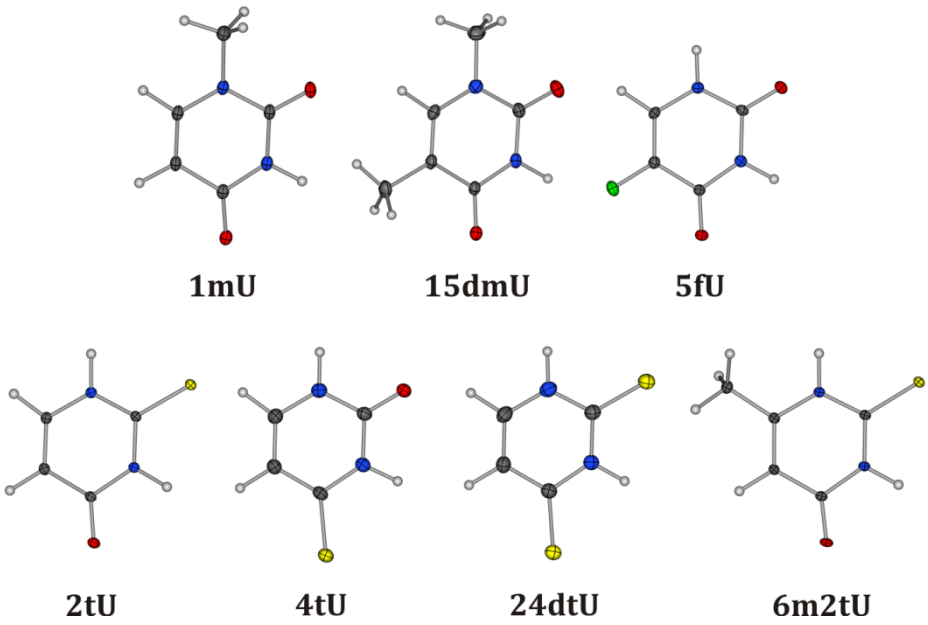


Figure 4.7. Atom representations and estimation of atomic thermal motion as ADPs for all studied structures after the TAAM refinement (colour coding: C – grey, H – white, N – blue, O – red, S – yellow, F – green; colours remain the same for all figures). Thermal ellipsoids are drawn at the 50% probability. In the case of **5fU** the other three symmetry independent molecules are omitted as they are much alike to the one shown.

Larger Atomic Displacement Parameter (ADP) values, in the case of 4-thio modifications, reflect a presumably lower crystal quality. Additionally, the S4 sulphur atom introduced in the pyrimidine ring changes the crystal colour from transparent for non-sulphur compounds to dark yellow, or orangish for **4tU** and **24dtU**, respectively. This is due to a considerable red-shift of the *n*-*n** UV absorption band caused by the sulphur atom, especially emphasised in the case of the S4 substituent.[96a, 96b] In turn, **15dmU**, **1mU**, **5fU** and also 2-thio derivatives, form crystals showing well-defined faces, being usually transparent stable prisms.

An important feature of uracil based compounds is their tendency to appear in different tautomeric forms. In the solid state, they preferably adopt the oxo-thione forms. The percentage contribution of various tautomers in solution or vacuum is

explained with the aid of spectroscopy and computations, to a less, or greater extent employing the solid state structure. It is therefore worth stressing the different tendencies observed for C–O and C–S distances, depending on oxygen or sulphur atom position (Table 4.2). C–S bonds are naturally longer than C–O, among which C4–O4 always exceeds the length of C2–O2 by over 0.01 Å. The opposite effect is observed for the sulphur atom, which is against some of the earlier reports.[118] Nevertheless, the discrepancy in C–S between C2–S2 and C4–S4 is less pronounced than in the case of C–O bond lengths, and is not statistically significant.

Table 4.2. Selected covalent bond lengths for studied uracil derivatives after TAAM refinement.

Compound	$d_{C2-O2/S2}$ / Å	$d_{C4-O4/S4}$ / Å	d_{N1-C2} / Å	d_{N1-C6} / Å
1mT	1.2256(5)	1.2370(5)	1.3874(5)	1.3746(5)
1mU	1.2214(9)	1.234(1)	1.375(1)	1.364(1)
6m2tU	1.676(1)	1.233(1)	1.358(1)	1.379(2)
4tU	1.227(2)	1.675(2)	1.363(2)	1.360(2)
2tU	1.6839(3)	1.2318(4)	1.3521(4)	1.3707(4)
24dtU	1.674(1)	1.670(1)	1.351(2)	1.359(2)
5fU [a]	1.226(2)	1.233(2)	1.360(2)	1.364(2)
	1.221(2)	1.234(2)	1.369(2)	1.366(2)
	1.220(2)	1.233(2)	1.368(2)	1.367(2)
	1.225(2)	1.229(2)	1.368(2)	1.367(2)
Compound	d_{N3-C2} / Å	d_{N3-C4} / Å	d_{C4-C5} / Å	d_{C5-C6} / Å
1mT	1.3803(5)	1.3805(5)	1.4381(5)	1.3600(6)
1mU	1.374(1)	1.379(1)	1.439(1)	1.353(1)
6m2tU	1.359(2)	1.389(1)	1.438(2)	1.362(2)
4tU	1.372(2)	1.371(2)	1.423(2)	1.353(2)
2tU	1.3558(4)	1.3951(4)	1.4443(4)	1.3523(4)
24dtU	1.359(2)	1.375(2)	1.419(2)	1.358(2)
5fU [a]	1.387(2)	1.375(2)	1.441(2)	1.352(2)
	1.381(2)	1.377(2)	1.439(2)	1.347(2)
	1.382(2)	1.372(2)	1.448(2)	1.351(2)
	1.383(2)	1.376(2)	1.445(2)	1.348(2)

[a] For **5fU** labelling is slightly different and it was adjusted for the purpose of this table; each row represents symmetry independent molecule.

IV.3.4. Molecular geometry evaluation and verification

For the purpose of a reliable crystallographic and energy characterisation of any compound, a well determined crystal structure, and thus molecular geometry, is essential. Therefore, at the very beginning of my studies, I tested the quality of the derived geometries. I wanted to check to what extent the TAAM refinement improves the final molecular geometry in comparison to the standard IAM model, taking the optimised structures or charge-density-derived ones as a reference point.

Molecular structures of the studied uracil derivatives were first obtained in a conventional way via X-ray data refinement, employing the Independent Atom Model. Such evaluated atom positions and ADPs were used in the TAAM refinement procedure, as I described in the §IV.3.2*(iii)* section, leading to new molecular geometries. Finally, TAAM geometries constituted a starting point for the subsequent geometry optimisations performed in a fixed experimentally determined crystal lattice (§IV.3.2*(iv)*). I compared together these three data sets and I also related them to the available neutron or charge density results reported in the literature. Direct comparison was made between bond lengths resulting from different methods. Root mean square deviation (RMSD) values of X–H and heavy atom bond distances, as well as the overall RMSD, for each of the examined systems, are grouped together in Table 4.3.

Table 4.3. Bond length RMSD values for X-ray geometries evaluated with IAM and TAAM models referred to the *CRYSTAL*-optimised geometry (OPT-3D). Additionally, RMSD values obtained for the optimised and TAAM geometries related to neutron (N) or charge density (CD) results, where available, are summarized. X–H, non-H and total keywords denote RMSD values calculated only for H-atoms, only for non-H atoms and all atoms, respectively.

Compound	RMSD / Å					
	IAM *vs.* OPT-3D			TAAM *vs.* OPT-3D		
	X–H	non-H	total	X–H	non-H	total
1mU	0.108	0.007	0.077	0.010	0.008	0.009
15dmU	0.121	0.005	0.081	0.013	0.005	0.010
5fU	0.143	0.009	0.083	0.006	0.011	0.009
2tU	0.143	0.005	0.083	0.004	0.006	0.005
4tU	0.143	0.004	0.083	0.003	0.008	0.007
6m2tU	0.149	0.006	0.094	0.012	0.006	0.008
24dtU	0.143	0.011	0.083	0.002	0.008	0.007
	TAAM *vs.* N/CD			OPT-3D *vs.* N/CD		
1mU [a]	0.004	0.004	0.005	0.008	0.005	0.006
15dmU [b]	0.024	0.007	0.017	0.037	0.011	0.026
2tU [c]	0.002	0.002	0.002	0.003	0.005	0.004

[a] 1mU – neutron data at 123 K; [b] 15dmU – neutron data at 297 K;
[c] 2tU – charge density data at 90 K

The results clearly show a huge underestimation of X–H bonds in the case of the refinement based on IAM. There is, however, not much of a difference for heavy atom–heavy atom across all the methods. What is more, TAAM results resemble neutron or charge density geometries to a greater extent than the corresponding optimisation or IAM results do. The discrepancy between TAAM and OPT-3D is, among the others, caused by different temperatures to which the evaluated geometries correspond (100 K

and 0 K, respectively). Neutron and charge density data were also collected at various temperatures. For this reason the **2tU** geometry derived from charge density data, collected at 90 K, is closest to the TAAM result, while the 297 K **15dmU** neutron data differs most from both OPT-3D and TAAM geometries.

In the case of the energy studies, hydrogen atom positions are extremely important. They specify not only the X–H length but also the bond directionality. The relevance of this problem was already a subject of the detailed analysis of charge density data.[19] Both TAAM and optimisation procedures allow for adjustment of these two parameters. Standard IAM refinement is usually not accurate enough to determine the hydrogen atom positions. Manual extending of the X–H bond lengths to the standardized neutron distances for IAM geometries may also not be fully satisfactory, especially when the initial X–H bond direction is not the proper one. I encountered this problem when analysing amphoterin B crystal structure, but it was also crucial to draw the proper conclusions in a comprehensive analysis of a series of fluorinated diboronic acids.[62a, 62b] All that may then hamper or even falsify the subsequent analyses.

Consequently, in order to analyse the nature of intermolecular interactions, and so the energy and geometry related properties, I used TAAM geometries, considered as the most accurate. All of them are based on crystal data collected at 100 K (literature data is not temperature consistent and therefore not comparable). On the other hand, the OPT-3D geometries (what guaranteed the result consistency) were employed for the purpose of some of the theoretical analyses, concerning computing atomic charges, *NICS* indices as well as some energy comparisons involving the chosen CSD supplementary compounds (for which there are no TAAM data).

IV.3.5. Crystal packing characterization

To fully understand further energetic considerations of crystal structure motifs, it is essential to have a wider view of the overall arrangement of molecular fragments in the solid state. Although several studies concerning crystal packing of different uracil derivatives are present in the literature, a direct comparison of all structures has not been performed. Therefore, in this study, I carried out a comprehensive structural analysis supplemented with Hirshfeld surfaces.

A summary of short intermolecular distances in all structures is presented in Tables 4.4 and 4.5 All of these values were obtained after the TAAM refinements. Thus, in

contrast to most of other structural investigations, I report the distances and angles with a higher precision on average (*i.e.* the estimated standard deviations are smaller), than usual.

Table 4.4. Geometrical parameters for selected hydrogen bonds and other intermolecular interactions obtained after TAAM refinements. D and A denote interaction donor and acceptor, respectively.

Compound	*Interaction*	d_{D-H} / Å	$d_{H...A}$ / Å	$d_{D...A}$ / Å	$\theta_{D-H...A}$ / Å
1mU	N3–H3...O4$^{\#1}$	1.030(2)	1.775(2)	2.805(1)	178.4(2)
	C7–H7A...O4$^{\#2}$	1.077(2)	2.336(2)	3.413(1)	177.9(1)
	C6–H6...O2$^{\#3}$	1.083(2)	2.367(7)	3.106(1)	124.1(5)
	C5–H5...O2$^{\#3}$	1.083(2)	2.624(8)	3.219(1)	113.9(6)
15dmU	N3–H3...O4$^{\#4}$	1.029(1)	1.794(2)	2.8218(4)	176.2(2)
	C6–H6...O2$^{\#5}$	1.083(1)	2.045(2)	3.1079(5)	166.6(2)
	C7–H7C...O4$^{\#6}$	1.076(1)	2.448(3)	3.4513(6)	154.6(3)
	C8–H8C...O4$^{\#7}$	1.077(1)	2.643(6)	3.2267(5)	113.4(4)
	C8–H8B...C4$_{\pi}^{\#8}$	1.077(1)	2.841(3)	3.8800(6)	162.0(2)
5fU	N1–H1...O5$^{\#9}$	1.030(1)	1.800(3)	2.817(2)	168.6(3)
	N3–H3...O8$^{\#10}$	1.030(1)	1.793(3)	2.817(2)	172.7(3)
	N7–H7...O1$^{\#11}$	1.030(1)	1.747(2)	2.776(2)	176.2(2)
	N9–H9...O6$^{\#12}$	1.030(1)	1.768(2)	2.796(2)	175.1(2)
	N13–H13...O7	1.030(1)	1.819(4)	2.838(2)	169.4(4)
	N15–H15...O4$^{\#12}$	1.030(1)	1.802(3)	2.823(2)	170.3(4)
	N19–H19...O3$^{\#13}$	1.030(1)	1.780(2)	2.809(2)	179.0(1)
	N21–H21...O2$^{\#14}$	1.030(1)	1.796(4)	2.814(2)	169.1(5)
	C6–H6...O4$^{\#15}$	1.083(1)	2.110(4)	3.168(2)	165.0(4)
	C12–H12...O8$^{\#16}$	1.083(1)	2.095(4)	3.136(2)	160.3(4)
	C18–H18...O2$^{\#14}$	1.083(1)	2.188(5)	3.206(2)	155.5(5)
	C24–H24...O6$^{\#17}$	1.083(1)	2.167(7)	3.174(2)	153.7(7)
2tU	N1–H1...S2$^{\#18}$	1.029(1)	2.292(2)	3.2991(3)	165.7(2)
	N3–H3...O4$^{\#19}$	1.029(1)	1.795(1)	2.8202(4)	173.6(2)
	C6–H6...S2$^{\#20}$	1.082(1)	2.771(4)	3.6363(3)	136.8(3)
	C5–H5...O4$^{\#21}$	1.083(1)	2.263(2)	3.3377(4)	171.4(1)
4tU	N1–H1...S4$^{\#22}$	1.030(1)	2.273(3)	3.289(1)	168.8(3)
	N3–H3...O2$^{\#19}$	1.030(1)	1.771(4)	2.792(2)	170.5(5)
	C6–H6...S4$^{\#23}$	1.083(1)	2.961(1)	3.706(2)	126.3(8)
	C5–H5...O2$^{\#24}$	1.083(1)	2.275(3)	3.342(2)	168.3(3)
6m2tU	N3–H3...S2$^{\#11}$	1.030(1)	2.385(4)	3.387(1)	164.0(4)
	N1–H1...O4$^{\#25}$	1.030(1)	1.779(5)	2.761(1)	158.2(7)
	C7–H7C...O4$^{\#26}$	1.077(1)	3.000(6)	3.669(2)	120.7(8)
24dtU	N3–H3...S4$^{\#27}$	1.030(1)	2.280(4)	3.284(1)	164.7(4)
	N1–H1...S2$^{\#28}$	1.030(1)	2.269(4)	3.271(1)	163.8(4)
	C5–H5...S2$^{\#29}$	1.082(1)	2.760(4)	3.804(1)	162.0(3)
	C6–H6...S4$^{\#26}$	1.083(1)	2.78(1)	3.460(1)	121.1(7)

Symmetry transformations: (#1) $-x, -y+2, z$; (#2) $-x+0.5, y-0.5, -x+2$; (#3) $x+0.5, -y+1.5, -z+2$; (#4) $-x+1, -y, -z+2$; (#5) $-x+2, y+0.5, -z+1.5$; (#6) $x+1, y, z$; (#7) $x, -y+0.5, z-0.5$; (#8) $x, -y+0.5, z+0.5$; (#9) $x-1, y, z$; (#10) $x, y-1, z$; (#11) $-x, -y+1, -z+1$; (#12) $-x+1, -y+1, -z+2$; (#13) $-x+1, -y+1, -z+1$; (#14) $x, y+1, z$; (#15) $-x, -y+1, -z+2$; (#16) $-x, -y+2, -z+1$; (#17) $x, y, z-1$; (#18) $-x, -y, -z$; (#19) $-x+1, -y+1, -z+1$; (#20) $x+1, y-1, z$; (#21) $-x+2, -y, -z+1$; (#22) $x+1, -y+0.5, z+0.5$; (#23) $-x, y-0.5, -z+0.5$; (#24) $x-1, -y+0.5, z-0.5$; (#25) $x+1, -y+1.5, z-0.5$ (#26) $x, -y+1.5, z-0.5$; (#27) $-x+1, -y+1, -z+2$; (#28) $-x+2, -y+1, -z+1$; (#29) $-x+1, y+0.5, -z+1.5$.

Table 4.5. Geometrical parameters for selected short intermolecular contacts obtained after TAAM refinements. X and Y denote pairs of directly interacting atoms.

Compound	*Interaction*	X...Y / Å	*Compound*	*Interaction*	X...Y / Å
1mU	$C6_{\pi}...C6_{\pi}^{\#1}$	3.2220(4)	**5fU**	$C10_{\pi}...C10_{\pi}^{\#9}$	3.290(3)
	$O2_{\pi}...C2_{\pi}^{\#2}$	3.1485(2)		$N7_{\pi}...F3_{\pi}$	2.913(1)
	$O4_{\pi}...C4_{\pi}^{\#3}$	3.1852(2)		$C16_{\pi}...O3_{\pi}$	3.165(2)
15dmU	$C4_{\pi}...C6_{\pi}^{\#4}$	3.3723(5)		$C24_{\pi}...C10_{\pi}^{\#10}$	3.332(2)
	$C5_{\pi}...C6_{\pi}^{\#4}$	3.3350(5)		$C4_{\pi}...C4_{\pi}^{\#11}$	3.263(3)
2tU	$C5_{\pi}...C2_{\pi}^{\#5}$	3.3189(5)		$C14_{\pi}...O7_{\pi}^{\#12}$	3.043(2)
	$C5_{\pi}...O4_{\pi}^{\#6}$	3.2170(4)		$C4_{\pi}...O7_{\pi}^{\#12}$	3.080(2)
4tU	$C2_{\pi}...C4_{\pi}^{\#5}$	3.393(3)		$F1...F2^{\#9}$	3.052(1)
	$C2_{\pi}...C5_{\pi}^{\#7}$	3.368(3)		$F1...F3^{\#13}$	3.055(1)
6m2tU	$C4_{\pi}...N1_{\pi}^{\#8}$	3.226(2)		$F2...F4^{\#14}$	3.058(1)
				$F3...F4^{\#12}$	3.091(1)

Symmetry transformations: (#1) $-x+0.5, -y+1.5, z-0.5$; (#2) $-x, y, z-0.5$; (#3) $x, -y+2, z-0.5$; (#4) $x, -y+0.5, z+0.5$; (#5) $x+1, y, z$; (#6) $-x+1, -y, -z+1$; (#7) $x, -y+0.5, z-0.5$; (#8) $x-1, y, z$; (#9) $-x, -y+1, -z+2$; (#10) $x, y, z-1$; (#11) $-x, -y, -z+1$; (#12) $-x+1, -y+1, -z+1$; (#13) $x, y-1, z$; (#14) $-x, -y+2, -z+1$.

This has two origins: (1) the more accurate description of the electron density distribution within the crystal, and also, (2) X–H distances standardization to neutron values (which in fact shows up as approximately 0.001 Å e.s.d.s). Of course, one has to remember that these numbers originate from numerical procedures, and 'real' errors have to be considered as least of order of magnitude larger. To supplement the analysis, I calculated partial *ESP* charges at the DFT(B3LYP)/aug-cc-pVDZ level of theory as described in §IV.3.2*(v)* section. The atomic charge values are stored in the Appendix D (Table D9). To some extent they illustrate the substituent effect, as well as shedding some light on the atom character, and therefore might be helpful in defining the mutual atom basic and acidic properties.

I supported the analysis of all crystal structures by a brief comparison with **U**, **1m4tU** and **5m2tU** compounds taken from the CSD. Each of them exhibits quite different crystal packing motifs, and will be treated separately. Although, all other compounds, except for **4tU**, were reported before, some of the studies are rather outdated and incomplete, so there is a need to revise the crystal structures.

It is clearly seen that strong hydrogen bonds of the N–H...O or N–H...S type play a dominant role in crystal packing of these uracil derivatives. Furthermore N–H...O hydrogen bonds tend to form rigid dimeric moieties, whereas the N–H...S contacts are more flexible. Oxygen atoms, being more negative than sulphur equivalents, are expected to create more advantageous, stronger and shorter hydrogen bonds. This

statement is in accordance with the *ESP* atomic charges (Appendix D, Table D9). In the case of sulphur substituent they should be longer and presumably less directional as sulphur is less acidic than oxygen, and also, more polarisable.

As a consequence of the above observations, the $R_2^2(8)$ ring units based on N–H…O or N–H…S, hereafter abbreviated as **A** and **B**, respectively, constitute the most commonly created synthons. These hydrogen bonds supported by other weak interactions form sheet motifs (except for the **24dtU**) in crystal structures. Within such supramolecular layers each molecule is surrounded by six closest neighbours. The mutual arrangement of the surrounding molecules depends on uracil ring substituents. Naturally, there are several ways to face the supramolecular characterisation problem of the selected series. Hydrogen bond net, as well as stacking interactions, may constitute a good starting point for the subsequent discussion. However, instead of the classical approach based on the detailed analysis of crystal packing motifs, I decided to utilize Hirshfeld surface fingerprint plots (Figure 4.8), which link qualitative and quantitative points of view.[119]

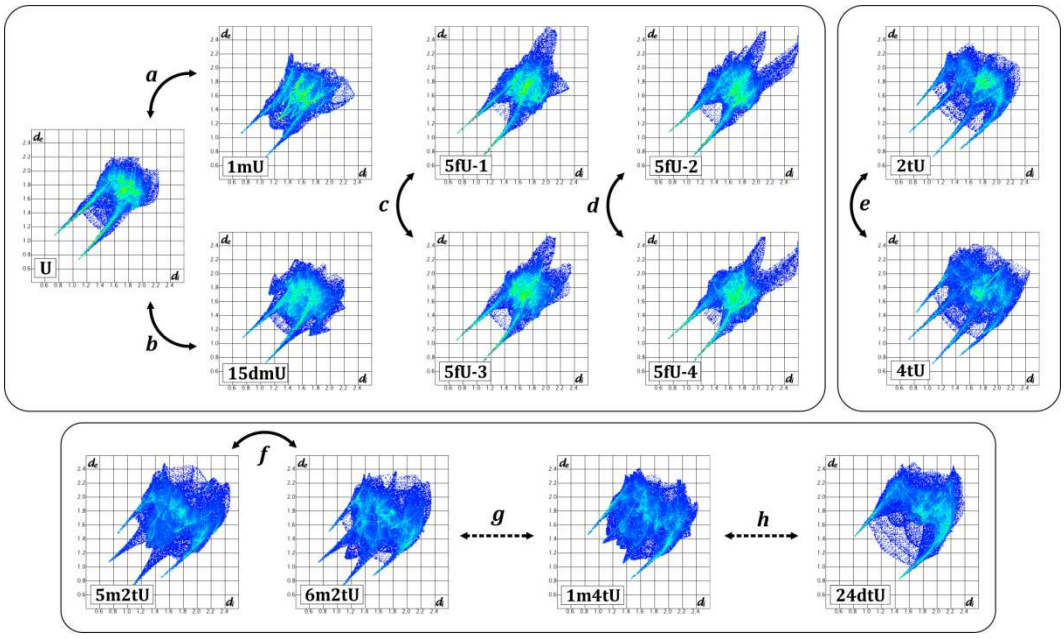


Figure 4.8. Hirshfeld surface fingerprint plots for the studied uracil derivatives. Small letters and arrows denote specific relations between given structures and are described in the text.[119b]

Following the fingerprint plots illustrated in Figure 4.8, all structures can be grouped into a few categories, *i.e.* (**I**) **U**, **1mU**, **15dmU** and **5fU**; (**II**) **2tU** and **4tU**; (**III**) **5m2tU**, **6m2tU**, **1m4tU** and **24dtU**.

In group **I** the structures of **1mU** and **15dmU** are very similar to the parent uracil. They all exhibit strong intermolecular dimeric motifs based on N–H…O hydrogen bonds (synthon of type **A**, Figure 4.9). In the case of **U** such dimers are further connected via hydrogen bonding into one very flat layer (Figure 4.9*a*), and thus motifs of large 6-membered rings are also present (type **C**). Layers are bound together by π-stacking interactions in the crystal lattice forming simple 3D structure. The distance between the interacting atoms from the adjacent layers is quite typical for π-stacking contacts (about 3 Å). This is also true for the other studied compounds (Table 4.5).

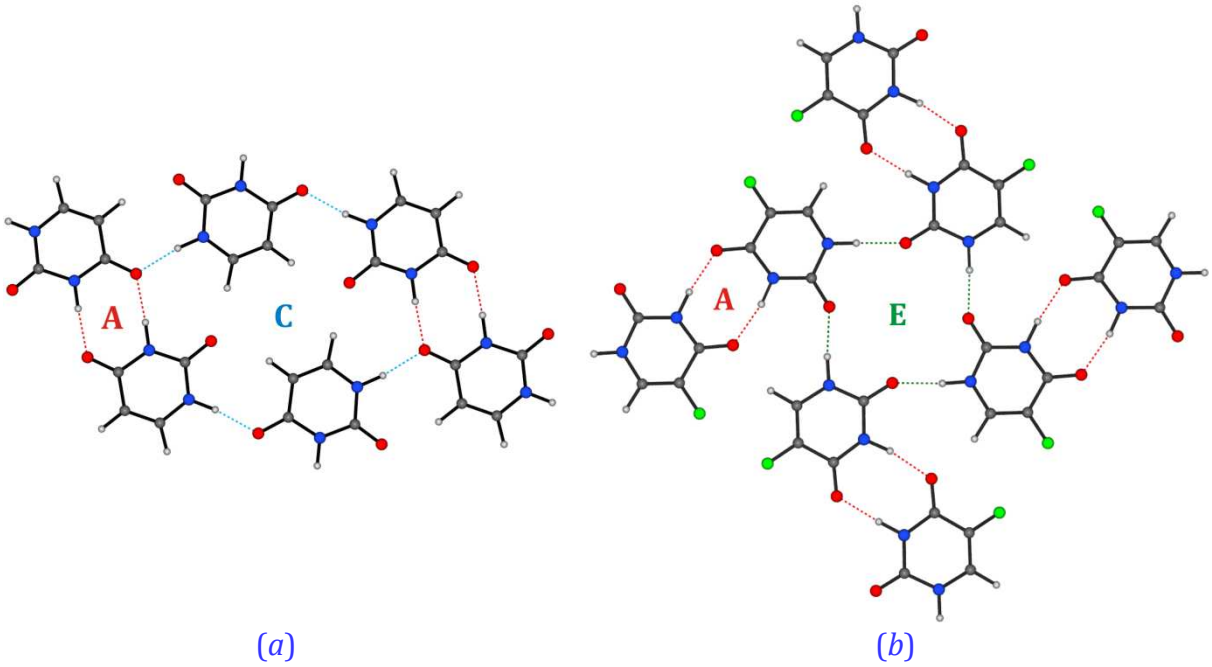


Figure 4.9. Structural motifs formed within layers for (*a*) **U** (**A** and **C**) and (*b*) **5fU** (**A** and **E**) crystal structures. Some weak intermolecular contacts are omitted for clarity.

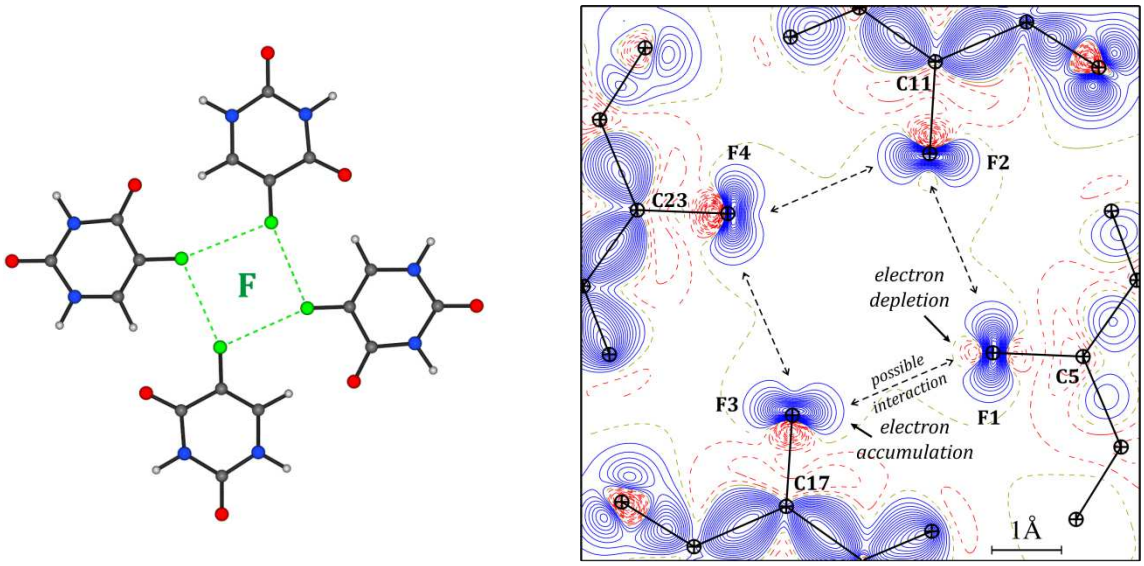
A slightly more distorted picture emerges from the analysis of **15dmU** and **1mU** crystal structures (relations ***a*** and ***b***). The presence of methyl groups prevents the formation of strongly bound 2D sheets, which are nevertheless still very flat. What is worth noting here is the fact that the neighbouring layers in **1mU** and **15dmU** are more mutually shifted than in the other cases. Additionally, we can observe that methyl groups tend to be arranged together.

In the case of **5fU**, where there are four symmetry-independent molecules in the unit cell, dimeric motifs are also formed (type **A**, Figure 4.9*b*). However, they are differently arranged due to the presence of a highly electronegative fluorine atom. Richness of possible contacts results in forming large tetrameric moieties (motif **E**). It

has been already noted in the literature that all fluorine atoms are brought together and it was stated that this seems to be quite unusual.[120] According to recent studies on halogen bonding,[121] the spatial arrangement of four fluorine atoms may rather indicate the presence of favourable F…F interactions (motif **F**, Figure 5*a*). However, the F…F contact distances, being equal to about 2.999 – 3.044 Å for the optimised geometries (OPT-3D), are slightly longer than the sum of fluorine atom van der Waals radii. Contacts shorter than 2.95 Å could be straightforwardly considered as F…F halogen-halogen interaction.[122] The studied case seems to be more complex than that. These contacts are well visible on fingerprint plots near the diagonal area and cover approximately of 7% of Hirshfeld surface for each molecule in the asymmetric unit (Appendix D, Figure D5*k*).

The TAAM procedure allows for quite reliable reconstruction of the electron density distribution in a crystal structure, though this method is not accurate enough to model crystal field effects.[32a] Figure 4.10*b* shows a static deformation density which I derived for the **5fU** tetrameric motif. Due to the presence of slightly bent 'lump-hole' patterns (*i.e.* regions of accumulated density pointing towards ones with depleted density), it is possible to rationalize the existence of some F…F interactions. The regions of depleted and accumulated electron density seem to be slightly exaggerated in comparison with the recent unconstrained multipolar refinements. Nevertheless, this is generally in accordance with the previously reported halogen bonding experimental charge-density studies.[121a, 121b] The analysis shows that F…F interactions should have some impact on the crystal stability, even if the molecular arrangement is basically forced by other stronger hydrogen bond contact formation. Due to the F…F distance, these interactions are at the edge of being considered as halogen-halogen bonding. However, the increased contact lengths may result from the anti-cooperative character of the cyclic fluorine interactions, as this is the case for hydrogen bonds forming cycles.

Small distortions caused by F…F and hydrogen bond interactions make layers deviate slightly from the perfect planarity. Similarly, as in the other compounds, layers are held together by π-stacking interactions. Interestingly, it seems there are two types of **5fU** molecules in the crystal lattice (relations ***c*** and ***d***) having compared their fingerprint plots.



(*a*) (*b*)

Figure 4.10. (*a*) Tetrameric motif showing F…F contacts between four symmetry non-equivalent/independent molecules in the solid state. (*b*) UBDB-derived static deformation density map for tetrameric motif (0.05 e·Å^{-3} contours, blue solid lines – positive, red dashed lines – negative) (symmetry codes are omitted for clarity; the whole motif is slightly bent out of the plane).

Within group **II**, the fingerprint plots for compounds **2tU** and **4tU** are very much alike (relation ***e***). Nevertheless, these structures deviate from each other quite significantly, when it comes to crystal packing comparison. Despite the fact that **2tU** and **4tU** structures exhibit dimeric motifs based on N–H…O interactions (**A** and **A'**, Figure 4.11), sulphur atoms at the 2- and 4-positions cause different electronic effects within molecules. In both cases the hydrogen atom which is involved in the strong N–H…O hydrogen bond, is the one in-between oxo and thio functional groups (namely H3 atom), as it is more acidic than the other one (such a tendency is also observed for other structures). Consequently, different dimeric motifs are preferably formed, which leads to a distinct molecular packing in the solid state. In the **2tU** crystal structure the two **A** type dimeric motifs are bound by N–H…S interactions (**B**), thus building a chain pattern. Such chains are connected via secondary C–H…O and C–H…S contacts forming flat molecular layers. In the **4tU** structure the remaining hydrogen bonding acceptor and donor centres (sulphur atom and N–H group), are located at the opposite sides of the molecule. This leads to the formation of a complex 2D net spanning the whole layer by N–H…S hydrogen bonds. This is a perfect example of how small changes in molecular connectivity may affect the whole crystal architecture. Explanation of the fingerprint plot similarity is that the overall neighbouring shell, and thus the summary of different

contact types, is quite alike in both cases. Interestingly, **A'** and **C'** motifs found in 4tU (Figure 4.11*b*) are very similar to the ones observed in the uracil structure.

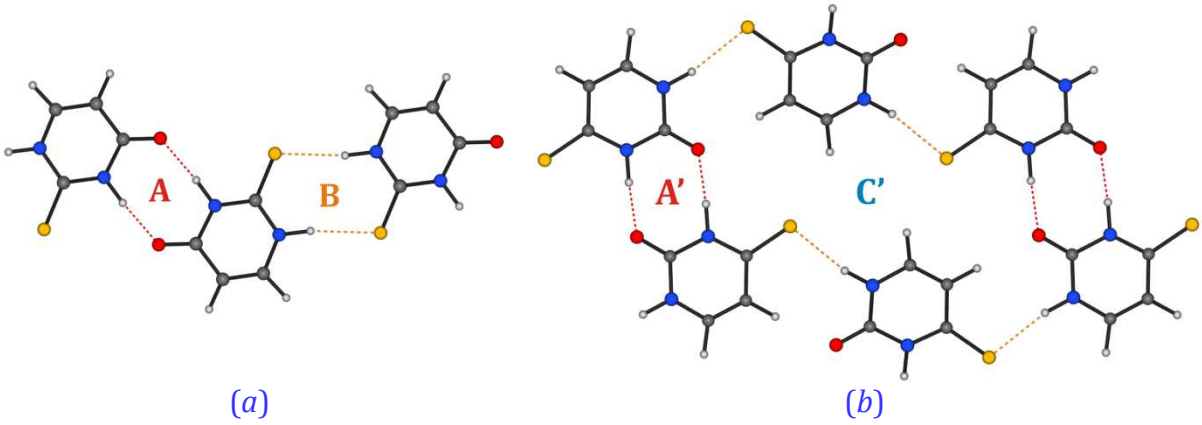


Figure 4.11. Differences in N–H…O dimer orientation in (*a*) **2tU** (motif **A**) and (*b*) **4tU** (motif **A'**) structures. Some weak intermolecular contacts are omitted for clarity.

Finally, group **III** shows an interesting complex situation of possible structural resemblance of various crystal structures. In particular, **5m2tU** and **6m2tU** are very much similar (relation ***f***). They both form chain motifs based on N–H…O interactions, which are joined together via dimeric motifs of the N–H…S flexible hydrogen bonds (Figure 4.12*a*). Such layers interact with each other by stacking interactions. However, the main difference between these structures is the position of the methyl group. In **6m2tU** the CH_3 group does not participate in any significant interlayer interaction, whereas in 5m2tU it is located in such a way that weak C–H…π and C–H…S contacts are formed. This distorts the layer structure to far from being flat. Nonetheless, both structures exhibit equivalent topology in the solid state.

Other molecular behaviour is observed for **1m4tU** and **24dtU**. Although the **1m4tU** fingerprint plot shows resemblance to the one of **6m2tU** (relation ***g***), some similarities to the **24dtU** plot can also be found (relation ***h***). In comparison with **6m2tU** the **1m4tU** fingerprint plot lacks 'spikes' near the diagonal, which are mostly attributed to N–H…O contacts. Indeed, when looking carefully at the **1m4tU** crystal structure, the only O…H type interactions, weak C–H…O contacts, are present within and between the layers. Definitely the N–H…S and C–H…S hydrogen bonds play a major role here. The flexibility of these contacts results in layers, which significantly deviate from planarity (Figure 4.12*b*). Additionally, the methyl groups of the neighbouring molecules form a similar pattern to the one observed in **1mU** or **15dmU** (Figure 8). We recall that the neighbouring planes are much more shifted in **1mU** and **15dmU** than in **1m4tU**.

Consequently, the closest molecular surrounding within the layers of **1m4tU** is analogous to that of **1mU** (Figure 4.13), however the adjacent layers are mutually arranged rather characteristically for the sulphur derivatives. This all accounts for the quite blurry picture of the **1m4tU** Hirshfeld surface fingerprint plot.

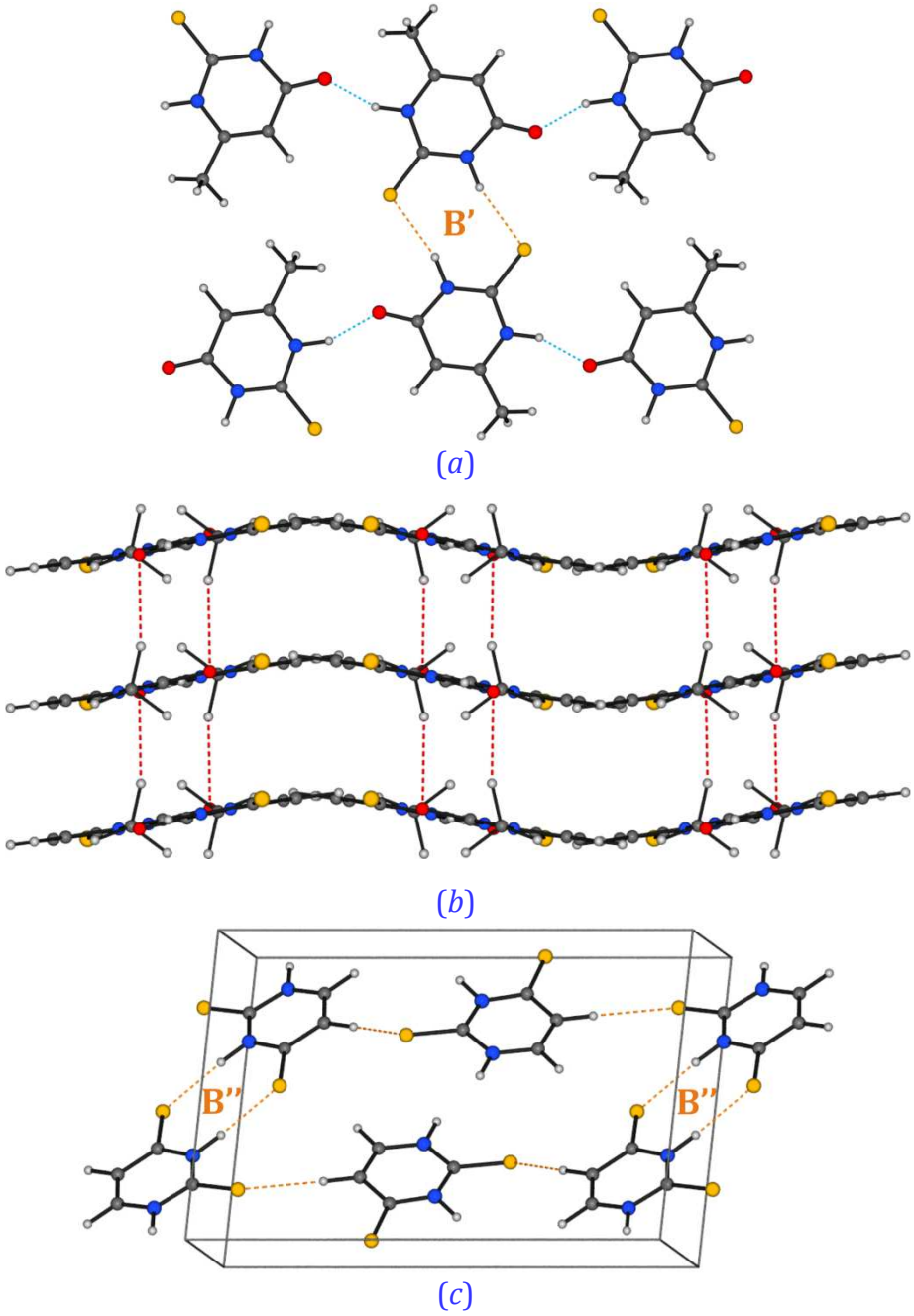
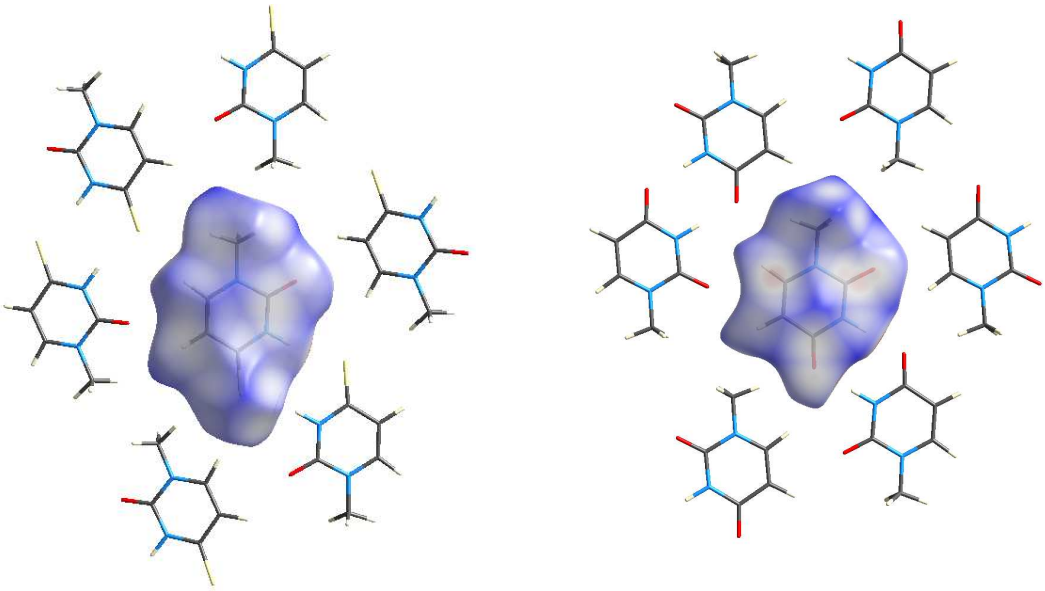


Figure 4.12. (*a*) Chain motifs present in the **6m2tU** crystal structure. (*b*) Layers of **1m4tU** molecules bound by C–H…O interactions (view along [201] direction). (*c*) Fragment of the 3D net present in **24dtU**. Some weak intermolecular contacts are omitted for clarity.



(*a*) (*b*)

Figure 4.13. Hirshfeld surfaces with mapped d_{norm} property and neighboring molecules for (*a*) **1m4tU** and (*b*) **1mU** projected in the hydrogen-bonded layer plane.

Further 'evolution' of the sulphur containing compound is **24dtU**, which does not include any oxygen atoms. In this view, the pattern observed in the fingerprint plot becomes clear (no 'spikes' near diagonal). The structure is held by N–H…S and C–H…S interactions forming a 3D net, where no specific supramolecular planes can be distinguished (Figure 4.12*c*). Despite that, **24dtU** is very similar to **2tU**, when it comes to the analysis of single molecular motifs. Indeed, the chains are connected via $R_2^2(8)$ synthons. The only difference is the overall geometry of **A** and **B** synthons in the case of **2tU** and solely **B** type units in the case of **24dU**. The different 3D structures of both compounds once again result from the less defined geometries of **B** type synthons.

Recapitulating, there is no comprehensive way to compare all the studied uracil derivatives on the basis of one criterion, neither taking into account solely Hirshfeld surface fingerprint plot, nor layer architecture, or others. Full analysis would constitute a multi-parameter space exploration. An excellent example emerges from the symmetry analysis. Considerably different **5fU** and **2tU** crystal structures exhibit layers forming motifs of only one type of molecules, while in other systems, molecules and their mirror equivalents are present. This is clearly reflected in their space group symmetry assignment (triclinic *vs.* monoclinic or orthorhombic).

IV.3.6. Energetic study

Descriptive analysis of the crystal lattice, based on molecular motifs and intermolecular distances, gives only qualitative information about formed molecular architectures. Hirshfeld surface studies and atom charge determination may enrich drawn conclusions with some roughly estimated interaction types, their percentage contributions, and indicate some of the bonding preferences. However, among the theoretical methods, solely computational energetic analysis of the different aspects of the studied systems quantifies these intermolecular contacts and thus sheds light on the interaction nature and crystal stability. Moreover, cohesive energy evaluation shows also the differences between IAM, TAAM and OPT-3D geometries, pointing out the great importance of hydrogen atom positions. Therefore, in this section I am going to present a comprehensive energetic study of the selected chemical systems, including cohesive energy estimation, crystal architecture motif interactions, and molecule geometry deformation energy when switching from vacuum to crystal structure.

(i) Cohesive and interaction energy analysis. Cohesive energy values obtained from *CRYSTAL* calculations for the whole series of the selected uracil derivatives on the basis of different geometries are summed up in Table 4.6, whereas the corresponding trends are visualised in Figure 4.14*a*. As clearly seen, the energies computed for the TAAM atomic coordinates behave much better than the corresponding results for the standard IAM data, when compared to the cohesive energy values derived for the optimised geometries. The energy variability is preserved between TAAM and OPT-3D. In the case of IAM, due to the improper X–H bond distances and directions, the cohesive energy trend is distorted and significantly deviates from both TAAM and OPT-3D energy results. Energy differences reach up to over 20 $kJ \cdot mol^{-1}$. Similarly, in the case of interlayer interaction energies evaluated for the measured structures, TAAM and OPT-3D results go together, while IAM results are more distant (Table 4.7, Figure 4.14*b*). Here, however, the trends are consistent among the differently derived interaction energies as generally π-stacking contacts contribute less significantly to the total energy value than hydrogen bonding and they do not involve the X–H distances to such an extent. Additionally, opposite to the obtained cohesive energy which is least advantageous in the case of IAM geometry, interlayer stabilisation energy is overestimated in the latter case (Figure 4*b*). These observations support the TAAM refinement as a powerful tool in order to get reliable molecular geometries. It is worth

stressing that IAM geometries of **U** and **5m2tU** with the X–H standardised to neutron average distances also behave quite well, however slightly worse than the TAAM itself. For **1m4tU** all the cohesive energy results are close one to another. In the case of slab stabilisation energy determination both **U** and **1m4tU** with the extended X–H lengths deviate from the OPT-3D results by about 6 kJ·mol^{-1}, whereas the energy values obtained for TAAM geometries differ not more than 2 kJ·mol^{-1}.

Table 4.6. Cohesive energy values (E_{coh}) calculated for differently derived geometries of the studied molecules.

Compound	E_{coh} / kJ·mol^{-1}		
	TAAM	IAM	OPT-3D
U [a]	−140.5	−121.1	−143.6
5fU	−141.0	−123.7	−143.3
6m2tU	−128.5	−116.4	−131.0
4tU	−125.8	−108.9	−127.8
2tU	−125.5	−109.2	−126.9
5m2tU [a]	−121.2	−114.6	−123.9
15dmU	−122.1	−116.5	−122.6
1mU	−118.7	−113.7	−120.3
24dtU	−108.7	−91.8	−109.3
1m4tU [a]	−103.1	−99.8	−104.4

[a] TAAM is replaced by CSD-taken structure with neutron-normalised X–H distances.

Table 4.7. Interlayer interaction energies (E_{intl}) calculated for different molecule geometries. R_{intl} denotes distance between planes describing the adjacent molecular slabs.

Compound	*Crystal plane*	R_{intl} / Å	E_{intl} / kJ·mol^{-1}		
			TAAM	IAM	OPT-3D
15dmU	(102)	3.301	−204.2	−215.2	−203.0
6m2tU	(102)	3.194	−185.2	−198.0	−184.9
1mU	(001)	3.098	−185.9	−194.8	−184.2
1m4tU [a]	$(10\bar{2})$	3.297	−188.6	-200.7	−182.7
5fU	(110)	3.288	−172.6	−182.0	−174.2
2tU	$(11\bar{2})$	3.128	−163.6	−169.3	−164.8
4tU	$(\bar{1}02)$	3.256	−159.2	−166.1	−159.5
U [a]	(001)	3.136	−145.1	−150.4	−139.6

[a] TAAM is replaced by CSD-taken structure with elongated X–H distances.

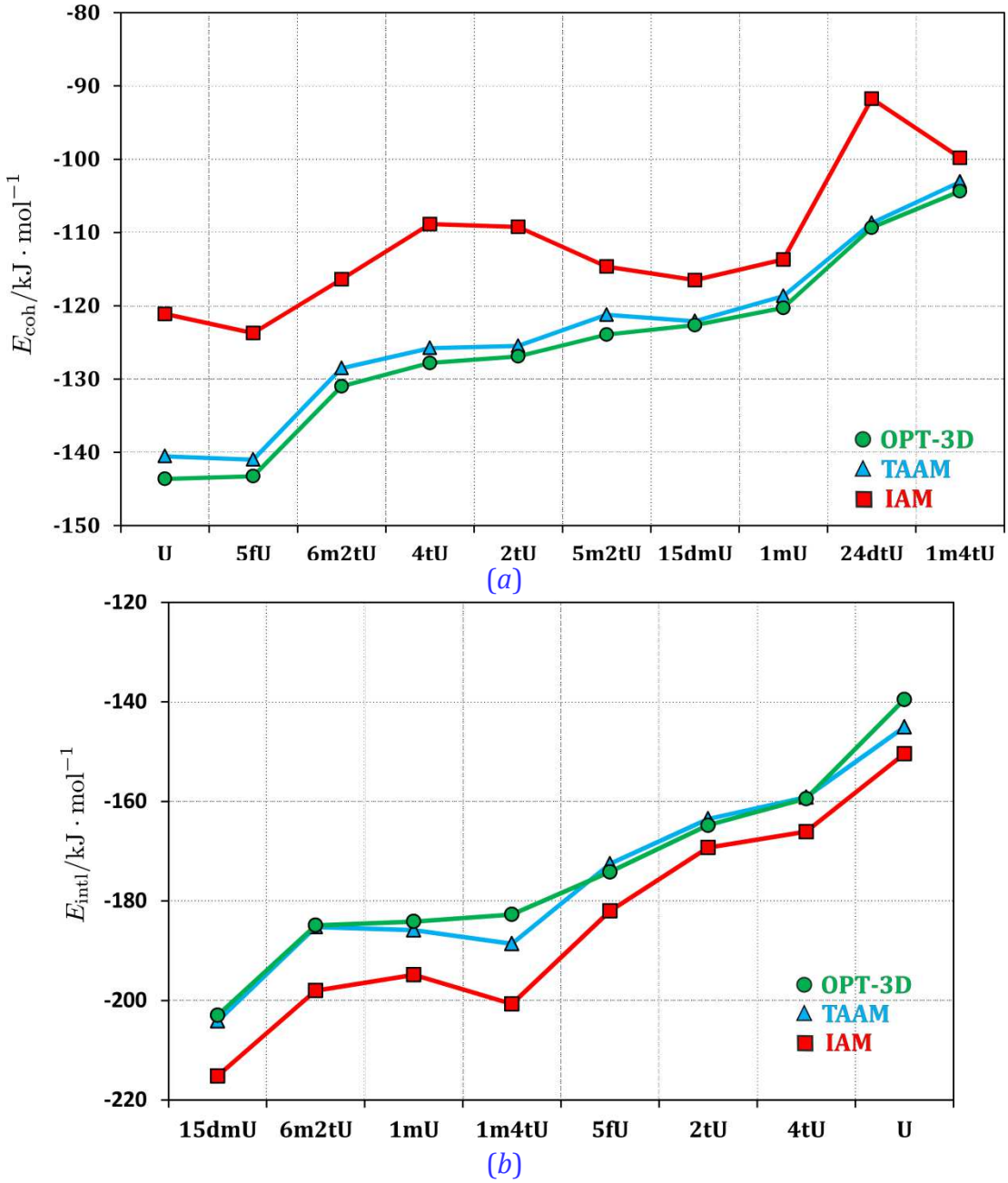


Figure 4.14. (*a*) Plot illustrating the discrepancies in the cohesive energy values (E_{coh}) obtained on the basis of differently evaluated molecular geometries. (*b*) Interlayer interaction energy values (E_{intl}) corresponding to TAAM, IAM and OPT-3D geometries, computed for the studied systems with the well-defined layer architectures.

Having carefully looked at the crystallographic structures and the corresponding stabilisation energy results, we may notice that the crystal lattice is thermodynamically most stable when the hydrogen bonding net within the 2D motifs is most saturated and all the proton donor and acceptor atoms are accessible. It seems that the dispersive interlayer contacts are of secondary importance. The methyl fragment at the 1-position causes not only a steric hindrance, but also blocks the formation of one hydrogen bond,

leading to the crystal networks of higher cohesive energy. This is why **1mU** and **15dmU** are less stabilised than **U** and **5fU**. A sulphur atom, being a weaker Lewis base than the oxygen analogue, also decreases the lattice stability as indicated by direct comparison of **U** and the sulphur derivatives **2tU**, **4tU**, and **24dtU**. **24dtU** containing two sulphur atoms is characterised by the highest cohesive energy among the mentioned group of compounds. The only less, and in fact least stabilised, crystal lattice is formed in the case of **1m4tU**. This is due to the superposition of the two effects, *i.e.* sulphur instead of an oxygen atom at the 4-position in the aromatic ring and a methyl group attached at N1-position. **1m4tU** adopts the same molecular organisation within the 2D sheets as **1mU**, but different arrangement of the neighbouring layers, which is more similar to that of the sulphur derivatives, as previously indicated by fingerprint plot resemblance. Its crystal lattice is almost 40 $kJ \cdot mol^{-1}$ less stabilised than that of the parent uracil, and 16 $kJ \cdot mol^{-1}$ less than that of **1mU**.

Nevertheless, weaker hydrogen bonding contacts might be compensated by stronger dispersive interactions. Table 4.7 contains all the details regarding layer description and interlayer interaction energy values. Figure 4.15 shows that the interlayer interaction energy is rather inversely proportional to the cohesive energy value, which suggests that its contribution is most significant when the creation of directional and strong hydrogen bonds is somewhat hampered. Furthermore, it is not correlated with the interlayer distances (assigned as a distance between the two closest corresponding crystal planes). This is not surprising, as the weaker bounded **A** or **B** type dimers tend to construct undulated 2D sheet motifs, which increases the effectiveness of intermolecular contacts. The presence of a methyl group also reduces the real interatomic distances between the two layers. Therefore structures containing methyl groups are characterised by favoured interlayer interaction. **15dmU** (two methyl substituents) and further **6m2tU**, **1mU** and **1m4tU** exhibit significantly lower slab interaction energy values than other uracil derivatives, which indicates a better mutual stabilisation of the neighbouring motifs. This is also reflected in the crystal morphology.[94e, 103b, 103c, 106a, 107] **5fU** is situated in the middle of the energy list as here the molecular planes are not perfectly flat due to the additional F…F and other short contacts listed in Table 4.5, and also, previously reported.[104a, 120] As mentioned earlier, a more polarisable and bulky sulphur atom usually introduces some flexibility to the

hydrogen bond contacts, causing layer motif deviation from planarity. This makes such molecular slabs better stabilised, than the flat planes of parent uracil molecules.

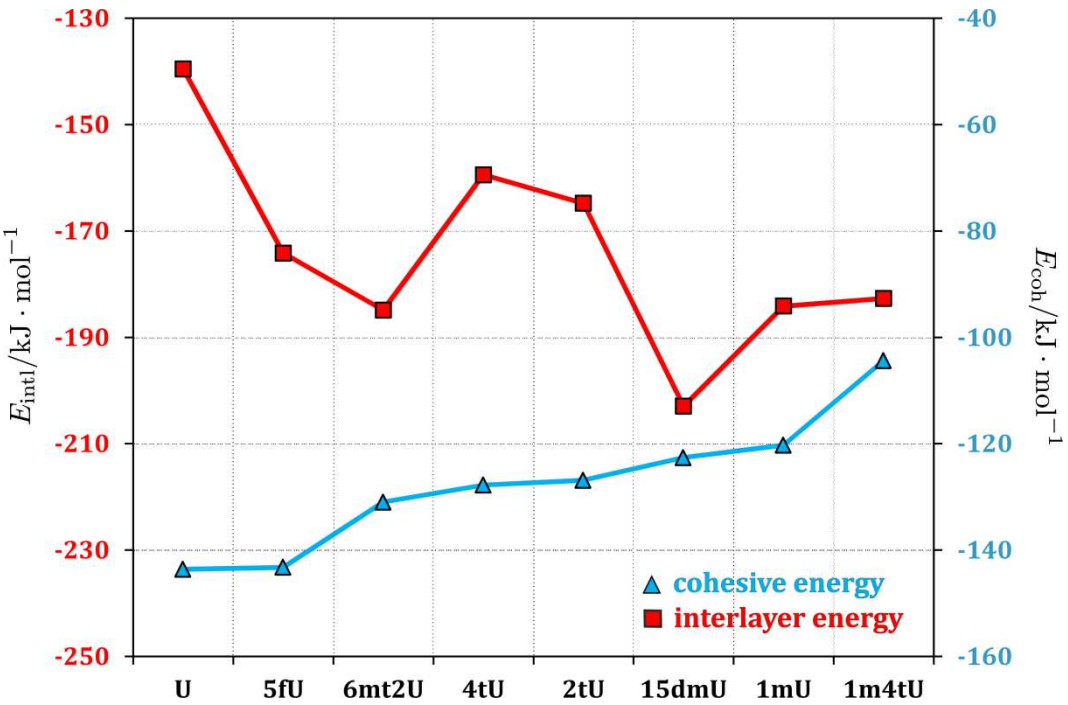


Figure 4.15. Cohesive energy (E_{coh}) and the corresponding interlayer energy (E_{intl}) values calculated on the basis of the optimised geometries (OPT-3D) of the studied systems.

In order to investigate the relations between the two types of the competitive, or rather complementary, hydrogen bonding and stacking interactions, I performed molecule geometry optimisations in three different surroundings. Firstly, molecules were optimised in vacuum, then in the 2D sheet motif, and finally in a crystal lattice (see the §IV.3.2*(iv)* section). In all the cases molecule energies (E^{opt}) were computed for the resulting geometries. The deformation energies, *i.e.* the energy differences between each geometry pair ($E_{i/k}^{opt} = E_i^{opt} - E_k^{opt}$ for i/k pair) are brought together in Table 4.8 and plotted in Figure 4.16. The total deformation energy generally amounts to around 10 kJ·mol^{-1}. Slightly lower absolute values were obtained for **24dtU** and N1-substituted methyl derivatives, *i.e.* **15dmU**, **1mU** and **1m4tU**. The latter four form less stabilised crystal lattices, where the dispersive interactions are relatively more pronounced than hydrogen bond contacts, when compared to the other studied systems. Consequently, **15dmU**, **1mU** and **1m4tU** molecules are not so strongly deformed due to hydrogen bonding formations as they are involved solely in one hydrogen bonded motif of type **A**

and type **B**, respectively. **24dtU** is slightly more affected by the crystal interactions due to two main motifs of the **B** type.

Table 4.8. Deformation energy values: $\Delta E^{\mathrm{opt}}_{\mathrm{mol/bulk}} = E^{\mathrm{opt}}_{\mathrm{mol}} - E^{\mathrm{opt}}_{\mathrm{bulk}}$ (mol/bulk), $\Delta E^{\mathrm{opt}}_{\mathrm{slab/bulk}} = E^{\mathrm{opt}}_{\mathrm{slab}} - E^{\mathrm{opt}}_{\mathrm{bulk}}$ (slab/bulk) and $\Delta E^{\mathrm{opt}}_{\mathrm{mol/slab}} = E^{\mathrm{opt}}_{\mathrm{mol}} - E^{\mathrm{opt}}_{\mathrm{slab}}$ (mol/slab).

Compound	ΔE / kJ·mol^{-1}		
	mol/bulk	*slab/bulk*	*mol/slab*
4tU	−11.2	−3.0	−8.2
U	−11.1	−0.4	−10.8
6mt2U	−10.4	−3.8	−6.6
5fU	−10.2	−1.6	−8.6
2tU	−10.1	−1.9	−8.1
24dtU	−8.3	–	–
5m2tU	-7.7	–	–
1mU	−7.2	−1.7	−5.5
15dmU	−6.2	−1.8	−4.4
1m4tU	-5.0	-1.4	-3.6

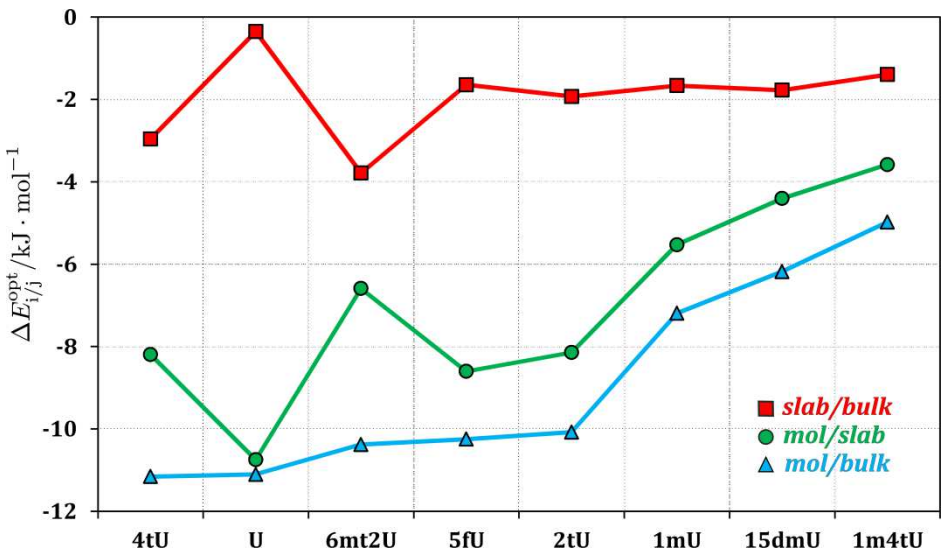


Figure 4.16. Deformation energy values plotted for the selected molecules: $\Delta E^{\mathrm{opt}}_{\mathrm{mol/bulk}} = E^{\mathrm{opt}}_{\mathrm{mol}} - E^{\mathrm{opt}}_{\mathrm{bulk}}$ (mol/bulk), $\Delta E^{\mathrm{opt}}_{\mathrm{slab/bulk}} = E^{\mathrm{opt}}_{\mathrm{slab}} - E^{\mathrm{opt}}_{\mathrm{bulk}}$ (slab/bulk) and $\Delta E^{\mathrm{opt}}_{\mathrm{mol/slab}} = E^{\mathrm{opt}}_{\mathrm{mol}} - E^{\mathrm{opt}}_{\mathrm{slab}}$ (mol/slab).

Such a tendency is generally observed when switching from a molecule in vacuum (1D) to a molecule in a slab (2D motif), and then, to a crystal lattice (3D system). The deformation energy between an isolated molecule and a molecule in a slab increases with the number and strength of interatomic contacts. For that reason $\Delta E^{\mathrm{opt}}_{\mathrm{mol/slab}}$ amplitudes in the case of **U**, **5fU** are highest, while for **1mU**, **15dmU** and **1m4tU** lowest, when compared to the rest of the systems. The deformation energy magnitudes, however, cannot be directly related to the cohesive or interlayer energy results, and

also, compared among the studied compounds. Some absolute scale must be found. For that purpose bulk to slab ratio, which refers the strength of the whole stabilization energy (that is mainly hydrogen bond contacts; cohesive energy) to the dispersive interaction contribution to the crystal lattice (interlayer stabilisation energy per one molecule), was compared with the difference of 'slab-to-bulk' ($\Delta E^{\mathrm{opt}}_{\mathrm{slab/bulk}}$) and 'mol-to-slab' ($\Delta E^{\mathrm{opt}}_{\mathrm{mol/bulk}}$) deformation energy results (Figure 4.17). These results correlate quite well with the determination coefficient, R^2, equal to about 0.90. The greater the change of $\Delta E^{\mathrm{opt}}_{\mathrm{mol/bulk}}$, the relatively smaller change of $\Delta E^{\mathrm{opt}}_{\mathrm{slab/bulk}}$, which also goes along with the decreasing contribution of the dispersive interactions in the crystal lattice. An extreme result is observed for **U** lattice where the interlayer interactions are negligible as compared to the strong hydrogen bond contacts within the 2D motifs.

4tU slightly, while **6m2tU** strongly deviate from the above rule. They both participate in quite a rich interaction net within the 2D sheet motif. **6m2tU** exhibits relatively low cohesive energy, but also, quite significant interlayer stabilisation interactions, which is unique among the studied crystal architectures. In turn, **4tU** has in general a somewhat more emphasised molecule deformation due to the crystal field. It might be caused by the **C'** pattern formation, leading to a quite undulated 2D motif, presumably involving some additional tensions.

The calculated energy results provide information about the thermodynamic crystal stability. Nevertheless, crystal formation is a complex process involving also kinetic effects. For instance, theoretically less stable **24dtU** crystals grow easier than those of the parent uracil. This might be caused by relatively low interlayer stabilisation energy of **U**, hampering the crystal formation along the *Z* direction, or some other circumstances like its higher solvent (*e.g.* water) affinity. Thus, there is no simple relation between the crystal stability and formation.

Anyway, the example of uracil shows that apart from the low cohesive energy value, the interlayer interaction energy contribution might be of great importance. In consequence, structures characterised by balanced cohesive energy to interlayer energy ratio form nice crystals, with well-defined faces, *e.g.* **15dmU** or **6m2tU**. This should also go along with the crystal mechanical resistance.

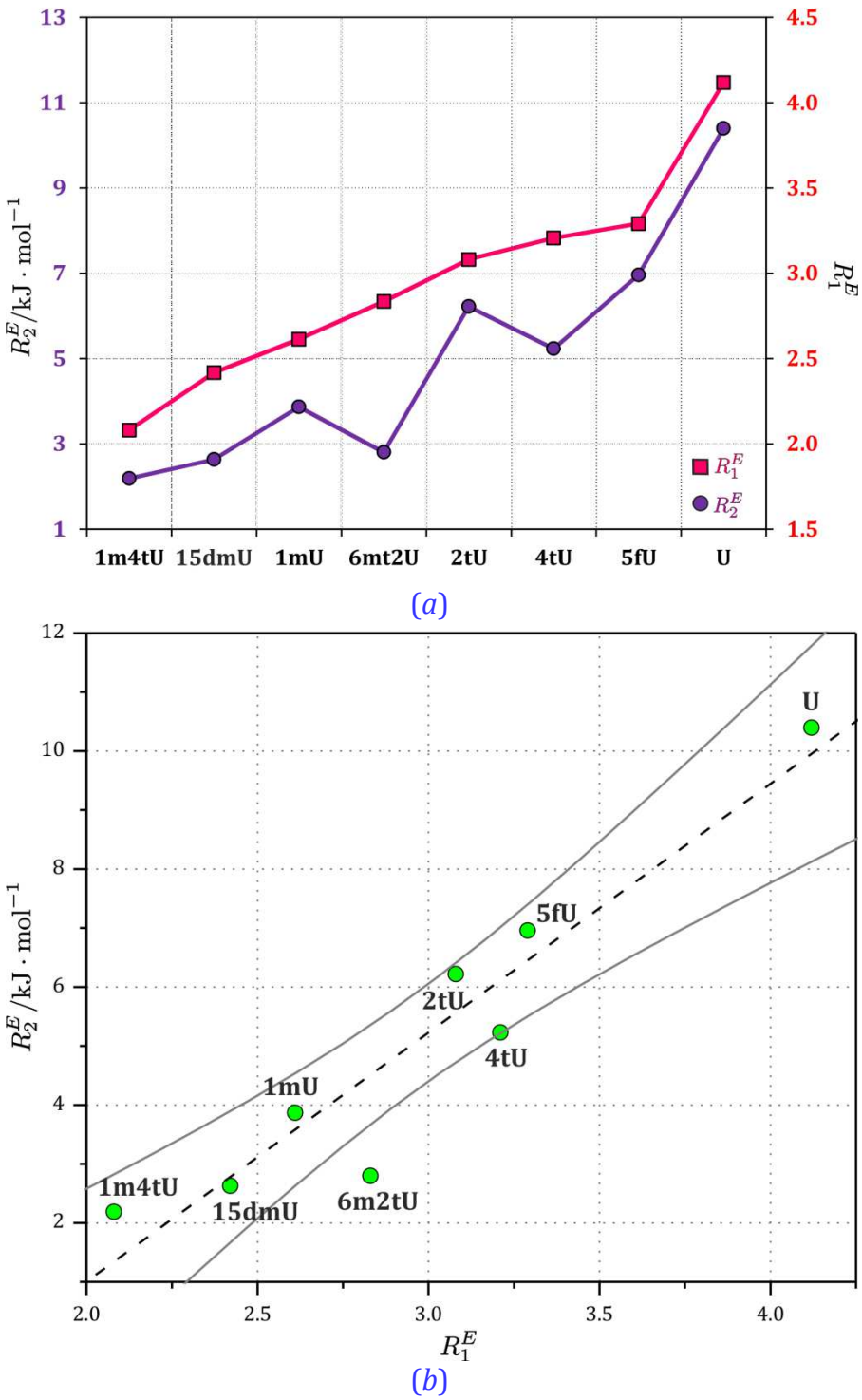


Figure 4.17. (a) Cohesive energy to interlayer energy ratio ($R_1^E = E_{\mathrm{coh}}/E_{\mathrm{intl}}$; both evaluated for OPT-3D geometries) related to the deformation energy difference ($R_2^E = \Delta E^{\mathrm{opt}}_{\mathrm{slab/bulk}} - \Delta E^{\mathrm{opt}}_{\mathrm{mol/slab}}$). ($b$) Linear correlation between the two factors (black dashed line; $R_2^E = A \cdot R_1^E + B$, where A = 4.2 ± 0.6 kJ·mol^{-1}, B = −7.4 ± 1.7 kJ·mol^{-1}; R^2 = 0.90 is the determination coefficient) shown together with the 95% confidence limits (grey solid lines).

However, generally once the crystal is formed it should be more stable when the cohesive energy is lower. Melting point temperature might constitute such a stability index. Therefore, we measured the adequate melting point temperatures and they are shown in Table 4.9. Indeed, uracil exhibits the highest melting point temperature being one of the two best stabilised studied uracil derivatives, according to the *CRYSTAL* results. **6m2tU** and some other compounds seem to follow the trend somehow, however, the temperature differences are not proportional to the energy mismatch. There are also significant exceptions, as the low melting point temperature of the theoretically very stable fluorine derivative, and also, the extremely low relative melting point temperature of **1mU**. It may result from the different crystal lattice stabilities at various temperatures as suggested by Hulme *et al.*[104a, 120] Yet no straightforward conclusions may be drawn.

Table 4.9. Melting point temperatures (T_m) of studied systems.

Compound	**U**	**6m2tU**	**2tU**	**4tU**	**15dmU**	**24dtU**	**5fU**	**1mU**
T_m / °C	334 [a]	330	308	290 [a]	284	278 [a]	277	235

[a] Sample decomposition was observed.

(ii) Motif energy characterisation. As a supplementation of the above energy analysis, I evaluated dimer interaction energies following procedures described in the §IV.3.2*(iv)* section. The results obtained for the most significant dimers from the crystal lattice of each of the studied compounds are shown in Table 4.10. In the case of **5fU** the stabilisation energies for the two tetramer types there were also computed, *i.e.* **E** and **F** motifs (Figures 4.9*b* and 4.10*a*). The strength of all of these interactions depends on the intermolecular distances, mutual molecular arrangement in space (*i.e.* geometry), X–H bond lengths, atom basicity or acidicity, and many others, which makes the problem very complex. As can be seen, motif **A** is usually similarly stabilised with the interaction energy close to −60 kJ·mol^{-1}. It is just slightly higher in the case of **2tU**. **A'** type dimer of **4tU** is about 10 kJ·mol^{-1} less stabilised than the already mentioned motifs **A**. **B** type synthons exist in a wider range of various geometries than **A**, and therefore, are bonded with significantly different strengths. They are characterised with the stabilisation energy reaching −70 kJ·mol^{-1} as well as equal to a half of this number. I verified the importance of the X–H bond length for **A** and **B** motif stabilisation taking **2tU** structure as an example. It appears that even small differences in the N–H distance (*ca.* 0.1 Å) cause significant changes to the total energy, which may reach up to approximately 14

kJ· mol^{-1} per one molecule (Appendix D, Figure D8). This result, being in a close relation with the lattice energy calculations, shows again the great importance of proper X–H standardization.

Single 'dispersively' interacting dimers are characterised by less stabilising energy values of up to −25 kJ·mol^{-1}. In the case of **U** and **24dtU** there are no significant contributions coming from stacking type of dimer motifs. The interaction energies do not exceed −13 kJ·mol^{-1} in strength.

Table 4.10. Selected dimer and tetramer interaction energies.

Motif	*Compound*							
	U	**1mU**	**15dmU**	**5fU**	**2tU**	**4tU**	**24dtU**	**6m2tU**
A / A'	−61.6	−60.5	−60.0	−62.7	−57.6	−51.5		
B / B'					−69.7		−34.3	−33.9
other **H** [a]	−62.3 [c]			−27.0 [d]	−33.6	−52.5 [e]	−72.7 [f]	−52.9
other π [b]		−24.6	−19.7	−14.9	−15.5	−21.1		-24.8
E				−141.9				
F				−138.1				

[a] other (usually weaker) hydrogen-bonded motifs such as C–H…O *etc.*; [b] stacking-type interactions (*e.g.* between layers); [c] dimeric motif present in motif **C**, different than **A**; [d] dimeric motif present in **E** (average value); [e] dimeric motif present in motif **C'**, different than **B**; [f] **B''** motif present only in **24dtU**; energy values given in kJ·mol^{-1}.

Interestingly, both tetramer motifs in the case of **5fU** are similarly stabilised, with the total interaction energy corresponding to four strong hydrogen bonds. It is perfectly true for the **E** tetramer, whereas for the **F** pattern it results from the superposition of the C6–H6…O4 contact strength and F…F interactions. An average dimer interaction of the adjacent **5fU** molecules, being a constructing brick of **F**, amounts to about −27 kJ·mol^{-1}. In turn, the single diagonal **5fU**:**5fU** interaction, not really considered as a F…F contact (more than 4 Å distant), is rather negligible (about −3 kJ·mol^{-1}), but still slightly attractive. This indicates that the F…F interaction at about 3 Å distance should be stronger, which seems to be supported by the rough deformation density map presented earlier (Figure 4.10). The comparison of the estimated C–H…O contact strengths with the tetramer **F** stabilisation energy suggested that the F…F interaction energy should amount to about −10 kJ·mol^{-1}. The above results together with the previous observations indicate the existence of the weak halogen-halogen interaction type in the case of **5fU**. However, substituting the fluorine atom with its heavier analogues, *i.e.* chlorine, bromine or iodine, leads to a substantial arrangement change due to their increased size, atom charge and higher polarisability.[123]

Some observations concerning S…S contacts in **2tU** are also present in the literature.[94i] Therefore, I performed adequate dimer energy elucidations. Nevertheless, I did not find any attractive interactions between two sulphur atoms. Their strength usually amounts to positive values of a few kJ·mol^{-1} in magnitude. In the case of the shortest S…S contact, 3.671 Å distant (in the optimised structure), this energy reaches +3.5 kJ·mol^{-1}. One should note that DFT method suffers in the area of dispersion energy estimation. In this case, the Grimme correction was applied, which improves the results but it is still not of the highest accuracy and may bias some of the obtained energy values. All in all, the calculations carried out do not support the weak bonding formation between the two sulphur atoms.

IV.3.7. Aromaticity

The last investigation, closing my analysis of the uracil derivatives, concerns aromaticity indices. Such a study may help to better understand substituent effects and electronic properties of the examined systems.

Among the wide variety of different methods of characterising the so-called molecule aromatic properties, I chose *HOMA* and *NICS* for the subsequent discussion.[117, 124] *HOMA* is based solely on geometrical parameters, therefore it should somehow reflect the geometry changes within the studied systems. It is defined as:

$$HOMA = 1 - \frac{\alpha}{n}\sum_{i=1}^{n}\left(d_{\mathrm{opt}} - d_i\right)^2 \qquad \text{(IV.3)}$$

where n is the number of bonds taken into account, α is a normalization constant (chosen to give *HOMA* = 0 for a model non-aromatic system, and *HOMA* = 1 for the system with all bonds equal to the optimal value; α values are equal to 257.7 Å^{-2} and 93.52 Å^{-2} for C–C and C–N bonds, respectively), d_{opt} stands for the optimal bond lengths for an ideal aromatic system of the kind (1.388 Å and 1.334 Å for C–C and C–N bonds, respectively), and d_i are real bond lengths in the examined ring.

On the other hand, *NICS* deals with π-electron ring current formation when a molecule is exposed to an external magnetic field, a phenomenon that is associated with magnetic susceptibility changes, and ^{1}H NMR chemical shifts. This index is purely theoretical and is defined as a negative value of the absolute magnetic shielding at ring centres.[117a] There has been also some *NICS* index modifications introduced in the literature such as *NICS(1)* and δ_{zz} values to better dissect σ and π contributions.[117b]

Negative values of *NICS* denote aromatic, while positive antiaromatic character. *NICS* type indices should provide some information about the ring electronic distribution changes, when affected by different aromatic ring substituents. *NICS* and *HOMA* index values derived for the studied systems are shown in Table 4.11.

Table 4.11. *HOMA* and *NICS*-type indices of aromaticity derived for uracil derivatives. Estimated standard deviations for *HOMA(TAAM)* values were calculated using the error propagation formula.

Compound	*NICS* / ppm	δ_{zz} / ppm	*NICS(1)* / ppm	$\delta(1)_{zz}$ / ppm	*HOMA(TAAM)*	*HOMA(OPT-3D)*
5fU	−2.3	−14.6	−1.8	−2.7	0.69(1) / 0.72(1) [a]	0.702 / 0.718 [a]
15dmU	−1.1	−12.2	−2.0	−3.6	0.721(3)	0.670
1mU	−0.9	−12.0	−1.8	−3.6	0.739(6)	0.705
U	−0.3	−11.2	−1.3	−2.5	– [b]	0.738
5m2tU	+0.8	−11.0	−1.3	−3.4	– [b]	0.692
1m4tU	+1.0	−11.0	−1.3	−3.0	– [b]	0.793
6m2tU	+1.0	−10.9	−1.0	−2.9	0.766(6)	0.781
2tU	+1.3	−10.1	−0.8	−2.9	0.717(2)	0.723
4tU	+1.6	−9.9	−0.5	−1.3	0.827(9)	0.825
24dtU	+3.7	−7.8	+0.4	−2.0	0.870(8)	0.825

[a] Average values calculated for two geometrically similar molecules, each; [b] experimental TAAM values are not available.

In the case of the examined uracil derivatives *NICS* and *HOMA* results are behaving somewhat differently. *HOMA* generally indicates the medium aromatic character of the compounds, far weaker than the basic pyrimidine ring (0.99). According to the *HOMA* values obtained for both TAAM and OPT-3D geometries, 4-thio derivatives seem to be the most aromatic ones. In turn, the S2 atom does not increase the aromaticity when compared to **U**. It is worth noting here that the *HOMA* value for uracil is greater than the average *HOMA* index (0.670) calculated for a set of different crystallographic structures containing this moiety, as reported elsewhere.[125] TAAM and OPT-3D geometries result in very close *HOMA* index magnitudes (as in the earlier investigation of bond lengths). However, optimisation of an isolated molecule leads to a significant geometry change, which is reflected in the essentially decreased *HOMA* values, *e.g.* to about 0.5 for **U**, depending on the geometry optimisation theory level and basis set used. Additionally, the *HOMA* index, being based on geometry parameters, confirms clearly the presence of the two types of **5fU** molecules in a crystal lattice, as pointed out previously by Hirshfeld fingerprint plot analysis.

When analysing *NICS*-type results, the situation is more complex. The aromaticity trend is significantly different than in the *HOMA* index case. Again S4 affects the

aromaticity a lot, but in the opposite way. Here, **24dtU** and **4tU** derivatives are the most antiaromatic ones, whereas least aromatic systems such as **5fU** and **15dmU**, according to *HOMA* indication, exhibit the strongest aromatic properties. Yet, all the *NICS* results are still close to zero, suggesting that in general uracil series is far away from a 'real' aromaticity (*NICS* < −7). Due to the ring substituents *NICS* values are more affected by σ type of bonds, which is especially pronounced in the case of the attached sulphur atom. Therefore, *NICS(1)* should be more informative. The *NICS(1)* aromaticity sequence is as follows (in order of the decreasing aromaticity): **15dmU** > **5fU** ≈ **1mU** > **5m2tU** ≈ **1m4tU** ≈ **U** > **6m2tU** > **2tU** > **4tU** > **24dtU**. Nevertheless, two contradictory effects might be observed: (1) the aromaticity increasing influence of the methyl group and (2) the inverse impact of the sulphur atom. The methyl derivatives tend to be more aromatic, whereas mixed sulphur and methyl ones similarly, and thio derivatives significantly less aromatic than uracil. The introduction of a sulphur atom at 4-position has a much greater effect on the electron distribution than S2, which is also visible in the UV spectra and which affects other properties such as biological activity.[96a, 126] As an additional observation, neither *HOMA* nor *NICS* index values go along with the ring planarity indicated by geometry studies (Appendix D, Comment D6).

IV.3.8. Summary

Molecular properties of the studied compounds, involving their biological activity, and in particular crystal lattice formation, have two origins. The uracil molecule, and so its examined modifications, contains hydrogen donor and acceptor centres eager to form hydrogen bonding. On the other hand, pyrimidine rings exhibit some aromatic properties and therefore may interact via delocalised π-electrons. Indeed, these two types of interactions govern the crystal lattice formation. The capability of various uracil modifications to participate in different type of contacts might be substantially influenced by aromatic ring substituents, relatively easily affecting the molecular charge distribution. It seems that the general tendency is to first saturate all the hydrogen bond possibilities, with the formation sequence depending on the mutual basicity and acidicity of the proton acceptor and donor atoms. Hydrogen bond network, supported by other weak interatomic contacts, most often leads to 2D molecular sheets formation. **24dtU** is the only structure, which does not form a layered architecture, however it is still rich in hydrogen bonding type contacts. Such molecular layers are held together

with the adjacent 2D motifs via π-stacking contacts. A methyl group usually increase the strength of the interlayer interaction, but hampers the stabilisation of the molecular slab itself. The most thermodynamically stable crystals are formed by **U** and **5fU**, both characterised by particularly strong hydrogen bond interactions within the molecular sheets. The cohesive energy, however, does not reflect the ease of a particular crystal formation. As indicated by our studies, crystals usually grow more easily when the interactions in different crystallographic directions are both advantageous and balanced. This might be the reason for which **15dmU** or **6m2tU** form well-defined prismatic crystals from water solution easily, while **U** does not. The presence of S4 substituent decreases substantially the overall strength of hydrogen bonding affecting the whole crystal thermodynamic stability, and so the crystal quality, to a greater extent than S2. This is also well visible in *NICS* and *HOMA* results, and indirectly explains so diverse biological properties of 2-thio- and 4-thio-uracil derivatives.

The contribution of hydrogen bond type interactions and weaker dispersive contacts is clearly seen from the molecule deformation energy analysis, when a molecule is transferred from vacuum to slab, and then further to crystal lattice. Energy of a uracil derivative molecule in vacuum and in a 3D crystal lattice may differ quite substantial, which is of up to −10 $kJ \cdot mol^{-1}$ (at this particular level of theory and with 6-31G** basis set). It is worth stressing here that for the purpose of some computational studies just the optimised molecule or molecular dimer model geometries are considered. In the face of the above results, *i.e.* the significant deformation energy values for strongly hydrogen bonded crystal structures, such an approach may not be sufficient.

Additionally, my investigations show the importance of properly determined molecular geometries, with an emphasis on hydrogen atoms positions, for the purpose of a quantitative study. The TAAM results with quite tightly constrained H–X distances, correlate very well with the geometries optimised in a crystal lattice. Therefore, TAAM refinement is a better alternative to the IAM approach.

IV.4. UBDB databank as a starting point for charge density studies

In this chapter, I shall present my results of charge density studies, of two chemical systems containing uracil derivatives. As the first compound, I selected 6-methyl-2-thiouracil (**6m2tU**). This choice permitted a natural continuation of investigations described in the previous chapter. For the subsequent investigation, I selected a co-crystal constructed by a pair of complementary nucleic acid bases, *i.e.* 9-methyladenine and 1-methyl thymine (**9mA:1mT**). In both cases, the UBDB databank served as a starting charge density model in the multipole refinement. However, my studies are not restricted solely to pure charge density distribution analyses. I also explored the interrelations between geometry, energy features of crystal architecture, charge density distribution characteristics, and crystal morphology. In the case of **9mA:1mT**, I additionally dedicated a part of the study to the similarities and differences between nucleic acid base pair mutual orientation, and geometries found in the available co-crystal structures compared to those present in DNA and RNA macromolecular systems.

IV.4a. 6-methyl-2-thiouracil study

Among the pyrimidine bases, thiouracil derivatives constitute a valuable class of compounds due to their biological and physicochemical properties.[94c, 94d, 94f, 94g, 96a, 96b, 97a, 105] A number of uracil mono- and dithio- modifications have already found medical applications.[94h, 94j, 97a, 98h] Thanks to their close relation to RNA/DNA components, they usually act as antivirial, antithyroid or antitumor agents. For instance, 6-propyl-2-thiouracil (PTU) is a well-known antihyroid drug.[127] Its methyl analogue, 6-methyl-2-thiouracil (**6m2tU**), also exhibits some therapeutic activity, however, it is especially characteristic for its capability of forming high-quality well-shaped crystals (from aqueous solutions), suitable for high-resolution X-ray measurements and for morphological investigations. This property opens up possibilities of detailed charge density distribution studies and crystal architecture energy aspect exploration.

Therefore, my analysis is devoted to meeting the interrelations between charge density distribution and interaction energy of selected molecular motifs present in the crystal network of **6m2tU**, and further to exploring the structural energetic features in the context of crystal morphology.

IV.4a.1. Materials and methods

(i) Crystal synthesis and data collection. Charge-density-quality single crystals of **6m2tU** were grown from water solution by slow evaporation of solvent at room temperature. It is worth noting, as has already been reported in the literature,[94e] that **6m2tU** tends to form non-merohedrally twinned crystals. Therefore, crystals appropriate for the subsequent analyses were very carefully chosen.

(ii) X-ray data collection. For the purpose of my study I carried out two high-resolution X-ray diffraction experiments at different temperatures, *i.e.* at 90 K and 10 K, on two different crystals taken from the same crystallization batch. Due to quality difference between the two data sets, great majority of my investigations is based on the data collected at 90 K. Details on the two measurements together with data processing are given below.

Single-crystal X-ray measurement of **6m2tU** at 90 K was carried out on a Bruker AXS Kappa APEX II Ultra diffractometer equipped with a TXS rotating anode (Mo-K_α radiation, λ = 0.71073 Å), multi-layer optics, and Oxford Cryosystems nitrogen gas-flow

device (700 Series Cryostream). A single crystal of a suitable size was attached to the MiTeGen MicroMount tool using the *Paratone N* oil, and mounted on a goniometer head at 60 mm distance from an APEX II CCD detector. Data collection strategy was constructed so as to scan over the whole reciprocal space with a reasonable reflection redundancy vs. data collection time, and served as an input for the *APEX2* suit of programs.[112] Solely ω-type scans were taken into account with a rotation step of 0.25°. The designed strategy, which resulted in 10800 frames, was as follows (Euler geometry parameters; χ = −55°, $\Delta\omega$ = 150°): (1) four scans with 2θ = −30° (φ = 0°, 90°, 180° and 270°) and 10 s of exposure time; (2) four scans with 2θ = −60° (φ = 0°, 90°, 180° and 270°) and 30 s of exposure time; (3) eight scans with 2θ = −90° (φ = 45°, 135°, 225°, 315°, 0°, 90°, 180° and 270°) and 60 s of exposure time; (4) two scans with 2θ = −20° (φ = 45° and 135°) and 10 s of exposure time.

In turn, the 10 K experiment was carried out with a locally-modified Nonius Kappa CCD diffractometer equipped with the molybdenum X-ray sealed tube, graphite monochromator, and an orange closed-circuit cryostat. Full details of this setup were previously reported by Fertey *et al.*[128] A single crystal of **6m2tU** was attached to a glass rod, mounted on a specially designed goniometer head and placed in the cryostat 40 mm from an APEX II CCD detector. The data collection approach, which resulted in 7004 frames, constituted a combination of ω and φ scans with a rotation step of 0.25°. Due to mechanical restrictions of the inner minigoniometer the χ angle was kept fixed at 43.373°. Final strategy was as follows (all ω scans were made as long as possible, preserving the safe distance to the helium chamber): (1) eight ω scans with 2θ = −30° (φ = −250°, −200°, −160°, −110°, −70°, −20°, 20° and 70°) and 15 s of exposure time; (2) one full φ scan with 2θ = −30° (ω = 260°) and 15 s of exposure time; (3) eight ω scans with 2θ = −90° (φ = −250°, −160°, −115°, −70°, −30°, −25°, 20° and 65°) and 60 s of exposure time; (4) one full φ scan with 2θ = −90° (ω = 200°) and 15 s of exposure time.

I performed data processing similarly in both cases. The determination of unit cell parameters and the integration of raw diffraction images were conducted with the APEX2 program package. However, for the purpose of the 10 K dataset, the orientation matrix was adjusted every 5th frame, while for the 90 K data a standard procedure was applied. Datasets were corrected for Lorentz, polarization and oblique incidence effects. The multi-scan absorption correction, frame-to-frame scaling and merging of reflections

were carried out with the *SORTAV* program.[15] The final data collection and reduction parameters are presented in Table 4.12.

Table 4.12. Selected crystal setting information together with the data collection parameters for both 90 K and 10 K measurement temperatures.

	90 K data set	*10 K data set*
Formula	$C_5H_6N_2O_S$	
Molecular mass, M_r / a.u.	142.2	
Crystal system	monoclinic	
Space group	$P2_1/c$	
Z	4	
$F(000)$	296	
a / Å	4.3315(2)	4.3135(7)
b/ Å	14.4396(5)	14.432(3)
c/ Å	9.5617(4)	9.5548(18)
α / °	90	90
β / °	90.8638(7)	91.170(6)
γ / °	90	90
Volume, V / Å^3	597.97(4)	594.69(18)
d_{calc} / g·cm^{-3}	1.580	1.588
Completeness	93.7%	89.7%
$(\sin\theta / \lambda)_{max}$ / Å^{-1}	1.19	1.25
θ range / °	2.56 – 58.00	2.56 – 62.58
Absorption coefficient, μ / mm^{-1}	0.445	0.447
Index ranges	$-10 \le h \le 7$ $-34 \le k \le 34$ $-19 \le l \le 22$	$-10 \le h \le 10$ $-34 \le k \le 35$ $-19 \le l \le 21$
No. of reflections collected / unique	45681/ 6473	39616/ 8699
R_{int}	3.94%	4.98%

Nevertheless, it is worth stressing once again, that the 10 K measurement is of generally lower quality than the 90 K one. This is due to the helium temperature experimental setup mechanical restrictions and its slight misalignments. Thus, the 10 K dataset lacks the lowest resolution reflections and contains much more problematic background signal, as well as a worse defined crystal orientation matrix. The illustration of the setup and exemplary problems which may occur during the 10 K measurement are provided in the Appendix E (Figures E1 and E2).

(iii) Structure determination and multipole refinements. Crystal structure of **6m2tU** was solved by a charge-flipping method with the *SUPERFLIP* program,[129] and refined on F^2 in the *JANA* package,[47] within the Independent Atom Model (IAM). Atomic

scattering factors, in their analytical form, were taken from the International Tables for Crystallography.[4] All non-H atoms were refined anisotropically. Hydrogen atoms were placed geometrically within the riding model for their isotropic ADPs ($d_{\mathrm{C-H}}$ = 0.96 Å, $U_{\mathrm{iso}}^{\mathrm{H}} = 1.2 \cdot U_{\mathrm{eq}}^{\mathrm{C}}$; $d_{\mathrm{N-H}}$ = 0.87 Å, $U_{\mathrm{iso}}^{\mathrm{H}} = 1.5 \cdot U_{\mathrm{eq}}^{\mathrm{N}}$). The orientation of the methyl group was determined from the adequate Fourier map.

These results served as a starting point for the subsequent multipole refinements. I evaluated the charge density model using the *MOPRO* suite[35b] combined with the newly prepared version of the University at Buffalo Data Bank (UBDB2011),[70] based on the Hansen-Coppens multipole formalism. Such a proceeding was additionally facilitated by specific modifications I made in the *LSDB* code, which currently automatically provides the input files desired for the *MOPRO* package. As a result all deformation parameters were defined in respect to their local Cartesian coordinate systems assigned by the *LSDB* program. It is also worth noting that local symmetries were imposed on atoms during the refinement. Otherwise, the energetic results for the derived geometries are significantly deviated from the corresponding ones obtained for the theoretically optimised structures (see Appendix E, Tables E16 & E17). Similar issue has recently been raised in the literature by Zarychta *et al.*[130] and Paul *et al.*[131] The authors indicated the importance of properly defined restraints and constraints within the multipole refinement.

(iv) Charge density refinement technical details. All refinements (including the preliminary ones) were based on *F*, and only the reflections fulfilling $|F^{\mathrm{o}}|^2 \geq 3\sigma(|F^{\mathrm{o}}|^2)$ condition were taken into account, as this was found to provide reliable results. In all the cases statistical weights were used (*i.e.* for *i*-th reflection $w_i = 1/\sigma_i^2$). Initial atomic coordinates, *x*, *y*, and *z*, and anisotropic displacement parameters (U_{ij}'s) for each atom were taken from the spherical refinement stage, whereas initial multipolar and contraction-expansion parameters were transferred from UBDB2011 with the aid of the *LSDB* program. Additionally, the X–H bond lengths (X = non-hydrogen atom) were standardised to neutron-normalised distances.[63] The *MOPRO* program allows for application of specific restraints during the refinement. Therefore, in the initial stage, the hydrogen atom U_{iso} parameters (*i.e.* isotropic thermal parameters) were restrained to the value of $y \cdot U_{\mathrm{eq}}^{\mathrm{X}}$ (y = 1.2 and 1.5 for X = C and X = N, respectively) with σ = 0.01 (where an appropriate restraint weight is equal to $1/\sigma^2$). In the final stage of refinements, the derivation of anisotropic hydrogen atom ADPs was carried out using

the *SHADE2* server. During the refinement procedure the contraction-expansion parameters for all atoms, κ and κ', were kept fixed at the UBDB2011 values. The X–H bond lengths were restrained to neutron-normalised distances with σ = 0.001. This approach provided earlier results comparable with the corresponding theoretical periodic computations and neutron studies for a series of uracil derivates (chapter IV.3).[93] Test refinements with freely variable hydrogen atom positions were leading to significant shortening of X–H distances, and thus were abandoned. The multipole expansion was truncated at the hexadecapole (l_{max} = 4) and quadrupole (l_{max} = 2) levels for all non-hydrogen and hydrogen atoms, respectively. In the case of hydrogen atoms only bond-directed multipoles were refined. The general refinement strategy was as follows:

(1) scale factor (which was also refined in almost all other stages);
(2) atomic coordinates;
(3) atomic coordinates and ADPs;
(4) *SHADE2* estimation of anisotropic hydrogen atom ADPs (which was also updated in-between other stages);
(5) multipole parameters in a stepwise manner;
(6) all parameters simultaneously.

According to the reasoning presented below, in the final model, I kept the deformation terms describing the sulphur atom fixed at the values transferred from the UBDB, whereas I refined the sulphur atom valence population.

(v) High-resolution data interpretation. The interpretation of the collected high resolution X-ray data occurred not to be straightforward. A proper evaluation of the charge density model required a lot of attention. A simple multipole refinement described above (with all κ and κ' parameters being fixed at the UBDB-transferred values), led to perhaps quite a satisfying overall fit ($\Delta\varrho_{res}$ = −0.27/+0.35 e·Å^{-3}), however, it did not converge to the physically reliable model. The charge density attributed to the sulphur atom valence shell was cumulated somehow above and below the aromatic ring plane, suggesting a quite peculiar electron lone pair distribution (which was not dictated by any strong intermolecular interaction in the crystal lattice) (Appendix E, Figure E4). Furthermore, the theoretical charge density data, which I obtained after the multipole refinement in respect to the periodically-computed structure factors, did not support the existence of such unexpectedly distributed electron density around the sulphur atom

(for details see Appendix E, Figure E5). Instead, both lone pairs at the sulphur atom exhibited their maxima in the aromatic ring plane, forming a characteristic for a sulphur atom deformation density pattern.

In order to find the origin of this effect, I carried out several additional investigations. The first attempt to refine the sulphur atom related features concerned the adjustment of the corresponding contraction-expansion parameters. Unfortunately, these proceedings neither improved the physical quality of the model, nor reduced the charge density residues ($\Delta\varrho_{\mathrm{res}}$ = −0.31/+0.42 e·Å^{-3}). Furthermore, to improve the model of the sulphur atom scattering factor, I also performed an extensive scan over different sulphur atom radial functions, however, with no satisfactory results.

Therefore, the other possible reason, which would explain the observed residues and unnatural charge density distribution, might be an improper description of the sulphur atom thermal motion. This supposition was partially backed up by the high-order (*i.e.* using only the reflections with $\sin\theta/\lambda \geq 0.8$ Å^{-1}) Fourier residual density maps (Appendix E, Figure E6*a*) for the 90 K dataset. The observed residual charge density pattern is characteristic for ananharmonic thermal motion of atoms, what has recently been reported and deeply analysed in the literature.[132] A departure from the harmonic model can be introduced via the Gram-Charlier series:[133]

$$T(\mathbf{H}) = \left[1 - \mathrm{i}\frac{4\pi^3}{3}\sum_j\sum_k\sum_l C_{jkl}h_jh_kh_l + \right. \\ \left. + \frac{2\pi^4}{3}\sum_j\sum_k\sum_l\sum_m D_{jklm}h_jh_kh_lh_m\right]T_0(\mathbf{H}) \qquad \text{(IV.4)}$$

where $T_0(\mathbf{H})$ is the harmonic temperature factor, C_{jkl} and D_{jklm} are the 3rd and 4th order tensor refineable coefficients, and h_k represents the k-th component of the scattering vector **H**.

The anharmonic refinement provided a substantial improvement of the overall fit to the experimental data. The charge density residual peaks were remarkably decreased ($\Delta\varrho_{\mathrm{res}}$= −0.17/+0.23 e·Å^{-3}) (Appendix E, Figure E6*b*). To verify the legitimacy of such a model, I deeply analysed the 10 K dataset. The mentioned characteristic residual pattern, being a consequence of atomic anharmonic thermal motion, should practically vanish at such low temperature. Despite the lower quality of the 10 K dataset (the high-order data is complete, though), essentially the same residual pattern as for the 90 K

dataset was observed (Appendix E, Figure E6*c,d*). Although in the latter case this effect seems to be slightly less pronounced (keeping in mind lower data quality), the results suggested that this is not anharmonicity, but rather some other effect unaccounted by the model (present in both measured crystals).

Such findings led me to a detailed search for some fine structural effects, regarding possible twinning or some subtle disorder. A very careful analysis of the 90 K dataset excluded the existence of twinning (the reciprocal space is very clear with no typical signs of twinning). Therefore, ultimately, the investigated residual density features might result from some disorder present in the crystal lattice of **6m2tU**. Unfortunately, it was not possible to find and refine any reasonable model of disorder. Nonetheless, this statement could be confirmed by the normal probability plots, which clearly suggest that some systematic error exists in both datasets (for all models) (see Appendix E, Figure E11). Since, for the time being, I could not prove the above hypothesis, I decided to proceed further with the sulphur atom described by the deformation density populations kept at the UBDB-transferred values, whereas with the atomic scattering factor modified according to the literature (ζ = 3.851 a_0^{-1}; n_l = 2, 4, 6, 8). The valence population of the sulphur atom was refined. This choice was made on the basis of the closer resemblance of such deformation density plots to the previously mentioned theoretical data results (see Appendix). In this model, the so far unexplained residual density peaks located in the vicinity of the sulphur atom are very significant ($\Delta\varrho_{res}$= −0.28/+0.64 e·Å^{-3}) (90 K data set, see Appendix E, Figures E8 & E9). However, the specific positions of the residual density maxima on both sides of the sulphur atom, above and below the aromatic ring plane, implied the idea of some contribution coming from a different 6m2tU tautomeric form. This would be a kind of disorder related to a proton transfer, and therefore quite difficult to detect. I am going to discuss this hypothesis, being a possible explanation of the observed phenomenon, further in the chapter.

The structural models evaluated on the basis of different charge density refinements lead to slightly modified molecular geometries. A comprehensive energetic comparison of the obtained results is available in the Appendix E (Comment E12).

IV.4a.2. Structural remarks

I have already presented a detailed analysis of **6m2tU** molecular motifs present in the solid state, which was a subject of the previous chapter, covering structural and energetic features of a series of uracil derivatives (IV.3). Therefore, here, I am going to recall solely some basic structural aspects of the **6m2tU** crystal lattice.

6m2tU crystallises in the monoclinic $P2_1/c$ space group with one molecule in the asymmetric unit. Its crystal architecture consists of molecular hydrogen-bonded chains forming sheet-type motifs via N–H…S contacts. Such molecular layers are arranged parallel to one another in space. The three main dimeric units that can be distinguished in the 6m2tU crystal structure include: (i) **C** – motif based on the strong N–H…O hydrogen bond; (ii) **D** – dimeric entity bound by N–H…S interactions; (iii) **S** – π-stacking motif of two adjacent parallel located molecules (Figure 4.18).

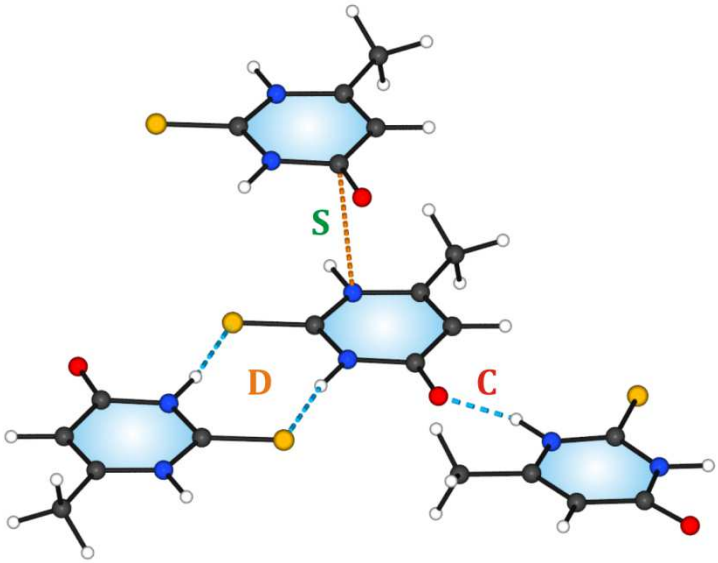


Figure 4.18. Main dimeric motifs present in the **6m2tU** crystal structure (**C**, **D** and **S**).

IV.4a.3. Charge density distribution and dimer interaction energy analysis

Despite some problems highlighted in §IV.4a.1, the electron density distribution within the **6m2tU** crystal was quite satisfactorily derived from the experimental data. The corresponding Fourier residual density maps together with the charge density parameters are summed up in the Appendix E.

Subsequently, I analysed the final charge density model within the Quantum Theory of Atoms in Molecules (QTAIM), developed over the years by Bader and co-workers.[134] As described in §II.1.13, the QTAIM involves a discrete boundary partitioning scheme

based on a topology of the electron density distribution. This theory provides a method of electron density distribution characterisation by means of various descriptors, such as bond critical points (BCPs), or integrated properties (atomic charges, dipole moments etc.). Afterwards, Abramov's investigations,[26] and further Espinosa's remarks,[27, 135] lead to a method linking the charge density distribution features with the energy of a hydrogen bond. As described in §II.1.13, in the Espinosa's approach, the weak interaction energy of a particular intermolecular contact is considered as being equal to approximately a half of the estimated potential energy density value at the BCP (the proportionality factor is in volume atomic units) (eq. II.52). Therefore, my analysis was supplemented by the energetic considerations regarding hydrogen bond contacts in line with Espinosa's approach.

Tables 4.13 and 4.14 summarize the detected bond paths (BPs) and bond critical points (BCPs), whereas the **6m2tU** molecular graph is presented in Figure 4.19. BPs and BCPs were found between all atoms in the molecule, showing that the overall topology of the electron density is consistent. In most of the cases (*i.e.* for C–C, C–N, C–O, C–H and N–H bonds) the values of the electron density and Laplacian are within typical ranges for such bond types.

Table 4.13. Selected QTAIM parameters of covalent bonds at BCPs (R – distance between bonded atoms; R_1 and R_2 – distances from 1st and 2nd atom to the bond critical point, respectively; ϱ – electron density; ε – bond ellipticity).

Bond	R / Å	R_1 / Å	R_2 / Å	$\varrho(\mathbf{r}_{BCP})$ / e·Å^{-3}	$\nabla^2\varrho(\mathbf{r}_{BCP})$ / e·Å^{-5}	ε
S2–C2	1.675	0.678	0.997	1.48	−6.7	0.06
O4–C4	1.237	0.809	0.428	2.71	−23.7	0.07
N1–C2	1.359	0.790	0.569	2.29	−23.6	0.10
N1–C6	1.375	0.803	0.572	2.14	−21.2	0.12
N3–C4	1.387	0.805	0.582	2.11	−20.1	0.11
N3–C2	1.359	0.801	0.558	2.23	−22.2	0.12
C4–C5	1.432	0.743	0.689	1.96	−17.5	0.16
C5–C6	1.361	0.644	0.718	2.19	−19.4	0.22
C6–C7	1.493	0.801	0.693	1.78	−14.2	0.09
N1–H1	1.030	0.755	0.275	2.34	−41.3	0.02
N3–H3	1.030	0.755	0.275	2.29	−34.3	0.02
C5–H5	1.083	0.693	0.390	1.91	−21.6	0.04
C7–H7A	1.077	0.703	0.374	1.86	−19.8	0.05
C7–H7B	1.077	0.704	0.373	1.84	−19.6	0.05
C7–H7C	1.077	0.709	0.368	1.81	−18.7	0.05

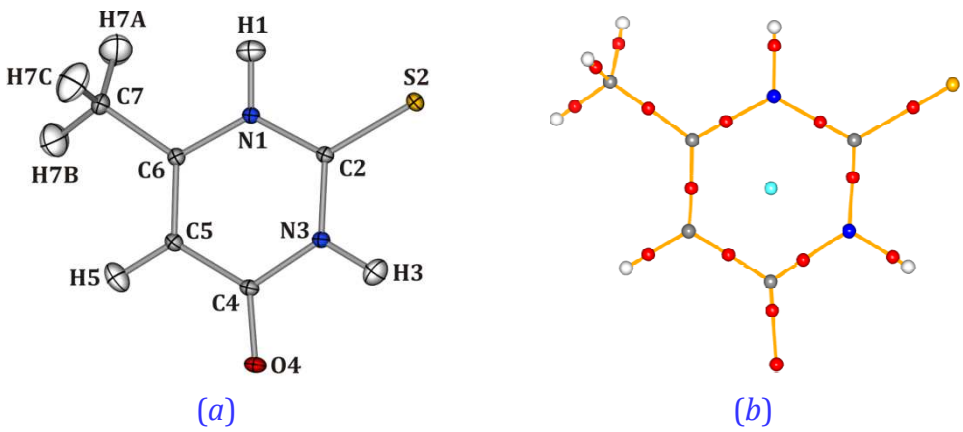


Figure 4.19. (*a*) Labelling of atoms and estimation of their thermal motion parameters, ADPs (50% probability level), for **6m2tU** at 90 K after final multipole refinement. (*b*) Experimental molecular graph for **6m2tU** (small red spheres – BCPs, small magenta sphere – RCP, orange solid lines – BPs).

As expected, the electron density value at the C–S bond critical point is much lower than the C–O bond, and generally the lowest among the analysed covalent bonds. Additionally, the related Laplacian value is less negative than that of all other bonds. Significant differences can be also observed between N–H and C–H bond types. In the case of both N–H bonds, the BCPs are shifted about 0.1 Å towards the hydrogen atoms, when referred to the analogue value of the C–H contacts. Such a trend has three origins: (i) a nitrogen atom, owning more electrons than a carbon atom, is more expanded (Table 4.16); (ii) the hydrogen atoms, H1 and H3, being involved in rather strong hydrogen bond contacts (iii) the electronegativity difference between nitrogen and carbon atoms. Finally, bond ellipticity indices show that all pyrimidine ring bonds can be treated somewhat in-between single and double, among which the C5–C6 bond has the largest double bond character (ε = 0.22).

In the crystalline state every molecule interacts with its surrounding. Therefore, a number of distinct intermolecular BPs and BCPs can be detected. Consequently, the three main structural dimeric motifs described previously (*i.e.* **C**, **D** and **S**, Figure 4.18) were supplemented by five additional ones. I selected them on the basis of the observed BPs. These motifs are presented in Figure 4.20. The numerical parameters for the possible interactions, which I found significant or worth exploring, are summarized in Table 4.14. Three new motifs, *i.e.* **M1**, **M2** and **M3**, involve interactions between the methyl group and the sulphur atom. On the basis of the charge density property values at BCPs, it can be assumed that **M1** is the most favourable motif among the group due to

the observed extra H5…S2 and N1…O4 bond paths and their features. Indeed, the **M1** dimer is characterised by a quite significant interaction energy value according to the *CRYSTAL* single point calculation (Table 4.15). Surprisingly, it is twice as well stabilised as the π-stacking **S** motif. The charge density study clearly shows that the methyl group is engaged in a variety of both, in- and off-layer, interactions. This goes along with my previous remarks, when analysing a set of uracil derivatives. Finally, the two motifs, H1 and H2, were chosen to illustrate the existing weak H…H close contacts topologically comparable with the observed S…H interactions. H1 is the in-layer contact, whereas H2 is assigned to the off-layer interaction.

Table 4.14. Selected QTAIM parameters for weak intermolecular interactions at the positions of respective BCPs (G – kinetic energy density; V – potential energy density).

Interaction	*Motif*	R / Å	R_1 / Å	R_2 / Å	$\varrho(\mathbf{r}_{BCP})$ / e·Å^{-3}	$\nabla^2\varrho(\mathbf{r}_{BCP})$/ e·Å^{-5}	$G(\mathbf{r}_{BCP})$ / kJ·mol^{-1}·a$_0^{-3}$	$V(\mathbf{r}_{BCP})$ / kJ·mol^{-1}·a$_0^{-3}$
O4…H1$^{\#1}$	**C**	1.764	1.150	0.617	0.27	1.7	66.9	−86.5
O4…H7A$^{\#1}$	**C**	2.414	1.394	1.031	0.07	1.1	23.6	−17.0
H5…S2$^{\#1}$	**C**	3.286	1.239	2.056	0.03	0.2	4.7	−3.4
S2…H3$^{\#2}$	**D**	2.378	1.581	0.799	0.15	0.9	29.2	−35.1
S3…N3$^{\#3}$	**S**	3.500	1.882	1.619	0.04	0.5	11.1	−8.1
N1…C4$^{\#3}$	**S**	3.219	1.639	1.583	0.04	0.5	10.7	−8.0
H7A…C5$^{\#3}$	**S**	2.946	1.223	1.726	0.03	0.4	8.8	−5.8
S2…H5$^{\#4}$	**M1**	3.060	1.889	1.249	0.04	0.5	11.3	−7.9
N1…O4$^{\#4}$	**M1**	3.389	1.698	1.691	0.03	0.5	9.3	-6.2
S2…H7B$^{\#4}$	**M1**	3.238	2.019	1.229	0.02	0.3	6.3	−4.1
S2…H7A$^{\#5}$	**M2**	3.198	1.957	1.265	0.03	0.4	7.4	−5.1
S2…H7C$^{\#6}$	**M3**	2.996	1.881	1.117	0.04	0.4	9.0	−6.9
H7B…H7B$^{\#7}$	**H1**	2.332	1.166	1.166	0.03	0.6	11.5	−7.4
H7B…H7B$^{\#8}$	**H2**	2.695	1.348	1.348	0.02	0.4	7.5	−4.5

Symmetry transformations: (#1) x+1, −y+0.5, z−1/2; (#2) −x+2, −y+1, −z; (#3) x−1, y, z; (#4) x, −y+0.5, z+0.5; (#5) −x+1, y+0.5, −z+0.5; (#6) −x+2, y+0.5, −z+0.5; (#7) −x+2, −y, −z; (#8) −x+1, −y, −z.

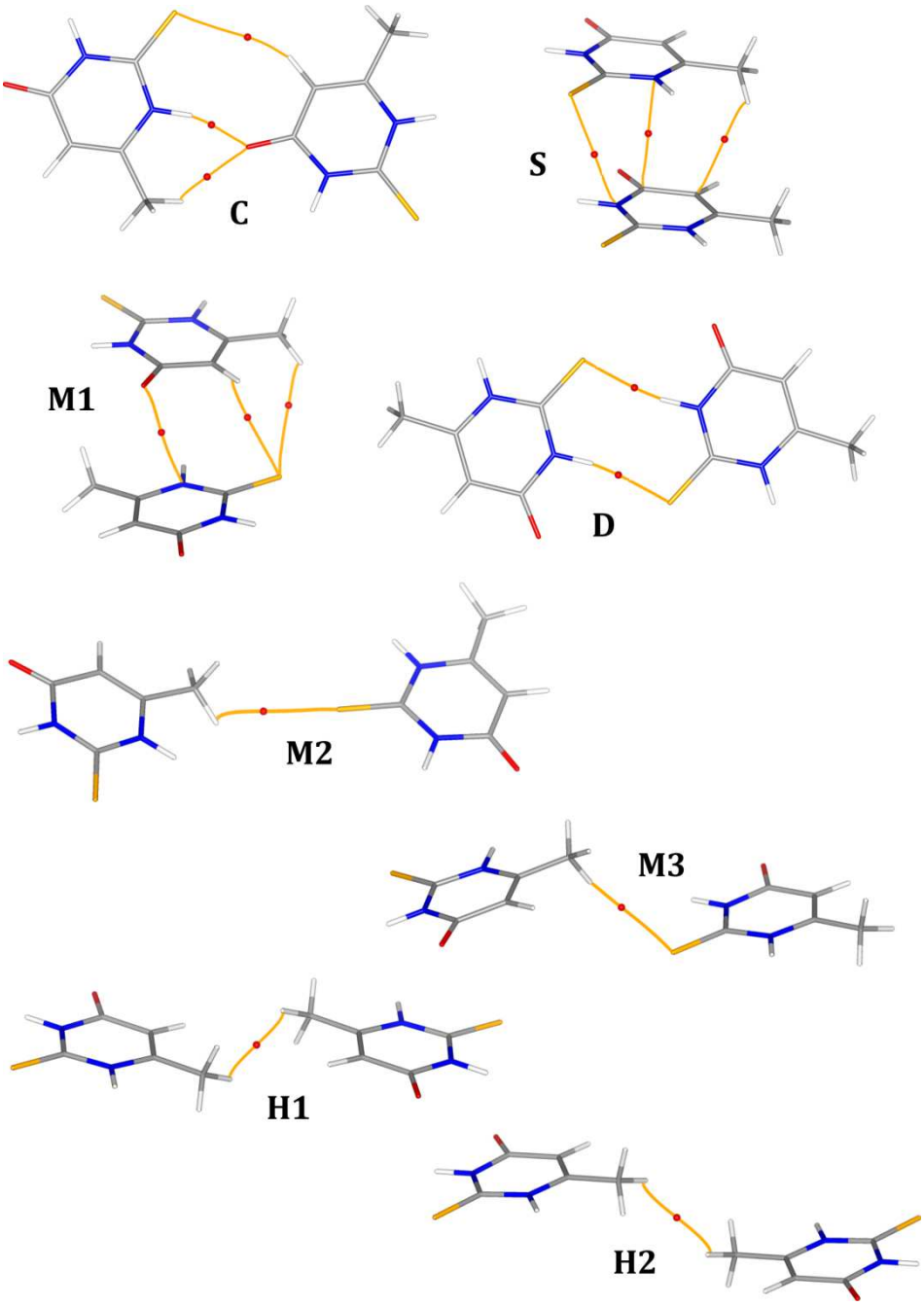


Figure 4.20. Bond paths (BPs) and respective bond critical points (BCPs) for all structural motifs present in the **6m2tU** crystal structure (**C**: N–H…O chain; **D**: N–H…S dimer; **S**: π-stacking; **M1**, **M2**, **M3**: interaction with methyl groups; **H1**, **H2**: H…H contacts).

Hydrogen-bonded chains (formed by **C** units) are characterised by three BPs, referring to the N–H…O, C–H…O and S…H–C interactions. Table 4.14 clearly shows that the N–H…O hydrogen bond is the strongest one. The particular strength of the N–H…O contact is additionally confirmed by the deformation density and Laplacian maps, which show the polarization of the oxygen atom towards the H1 atom of the neighbouring

molecule (Figure 4.21). A significant polarization is also visible for the H1 hydrogen atom in the Laplacian map, when the hydrogen quadrupolar parameters are refined. The two remaining interactions, i.e. C–H…O and S…H–C, also cooperatively contribute to a particular stabilisation of the **C** motif. This all makes the supramolecular chains very favourable while forming the crystal lattice. The total interaction energy of the **6m2tU** molecules building the **C** unit amounts to about −52 kJ·mol^{-1} (Table 4.15), which corresponds to the energy of two medium-strength hydrogen bond contacts.

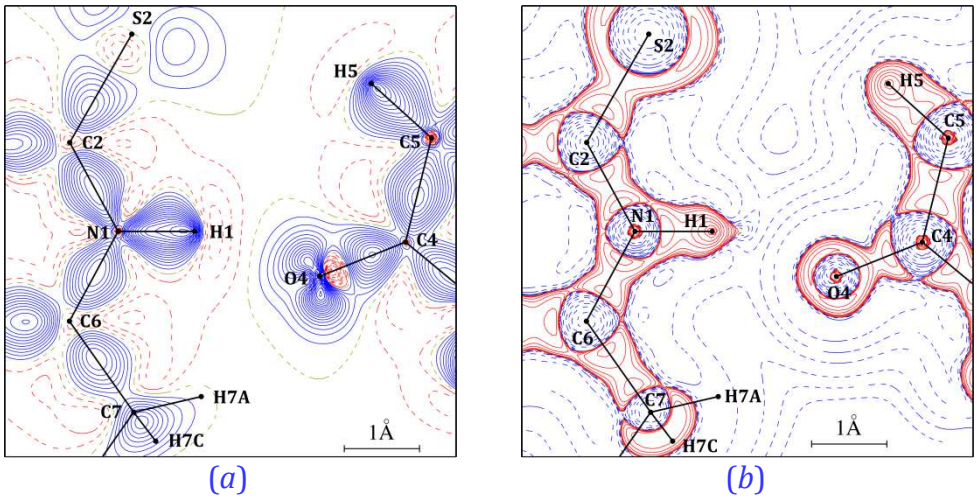


Figure 4.21. (*a*) Deformation density map (N1-H1-O4 plane, contours at 0.05 e·Å^{-1}, blue solid lines – positive contours, red dashed lines – negative contours). (*b*) Laplacian of electron density map (contours in logarithmic scale).

Table 4.15. Selected dimer motif intermolecular interaction energies calculated at the DFT(B3LYP)/pVTZ level of theory, accounted for BSSE and dispersion ($E_{\mathrm{int,I}}$ and $E^{\mathrm{opt}}_{\mathrm{int,I}}$ are the interaction energies of dimers extracted from the X-ray determined crystal structure and the optimized structure of **6m2tU**, respectively). Last column includes the corresponding hydrogen bond energy values (and other short contacts; $E_{\mathrm{top,I}}$) derived from the experimental potential energy density according to Espinosa's approach.

Motif	$E_{\mathrm{int,I}}$ / kJ·mol^{-1}	$E^{\mathrm{opt}}_{\mathrm{int,I}}$ / kJ·mol^{-1}	$E_{\mathrm{top,I}}$ / kJ·mol^{-1}
C	−52.2	−52.4	−53.5
D	−32.5	−34.0	−35.1[a]
S	−12.5	−12.1	−10.9
M1	−24.7	−25.4	−6.0
M2	−8.5	−8.5	−2.6
M3	−8.3	−8.6	−3.5
H1	0.9	–	−3.7
H2	4.1	–	−2.3

[a] About −35 kJ·mol^{-1} for all tested charge density models of sulphur atom.

The **C** motif energy value obtained from the single point *CRYSTAL* calculation is well reproduced by Espinosa's estimation grounded on the potential energy density at the hydrogen bond critical point. As electrostatic interaction energy is additive (as is dispersion energy to some extent) and that is also the main component of hydrogen-bonding interaction, the partial contributions coming from all three contact types were simply added leading to a final value of −53.5 kJ·mol^{-1}. The energy density results indicate a quite remarkable influence of the C−H…O contact on the **C** motif stability. These remarks together with the observations made on the **M**-type dimers suggest that the methyl group interactions with the adjacent molecules are crucial, both for the molecular layer stability and for the strength of interlayer contacts. This is naturally a result of the overall **6m2tU** crystal packing, where the relatively bulky methyl substituent substantially influences the preferred mutual orientation adopted by **6m2tU** molecules.

The second most stabilised motif is the **D** dimer. It is formed by two equal S…H close contacts. Once again, the dimer interaction energy calculated via the *CRYSTAL* package is very well reflected by the charge density derived hydrogen bond energy. Furthermore, this energy remains very much the same, no matter what sulphur atom model is applied. As can be noticed on the basis of Table 4.15, Espinosa's approach is indeed much more appropriate in the case of stronger intermolecular interactions than weaker and less electrostatic ones. This observation is also illustrated in Figure 4.22.

A deeper insight into electronic charge distribution can be also achieved by the analysis of atomic charges. These are presented in Table 4.16 along with the respective atomic basin volumes. Oxygen and nitrogen atoms are, as expected, negatively charged to a high degree. There is just a small charge difference between the N1 and N3 atoms (*ca.* 0.06 e), which can be disregarded. Carbon atom ring members reveal a rather predictable pattern, *i.e.* the atoms owning more electronegative neighbours are more positively charged. The C7 carbon atom (from the methyl group) is the only one bearing a negative charge, which is anyway rather small. All hydrogen atoms are positively charged. Among them, H1 and H3, involved in the N−H…O and N−H…S hydrogen bonds, are characterised by the most positive charges (about +0.5 e).

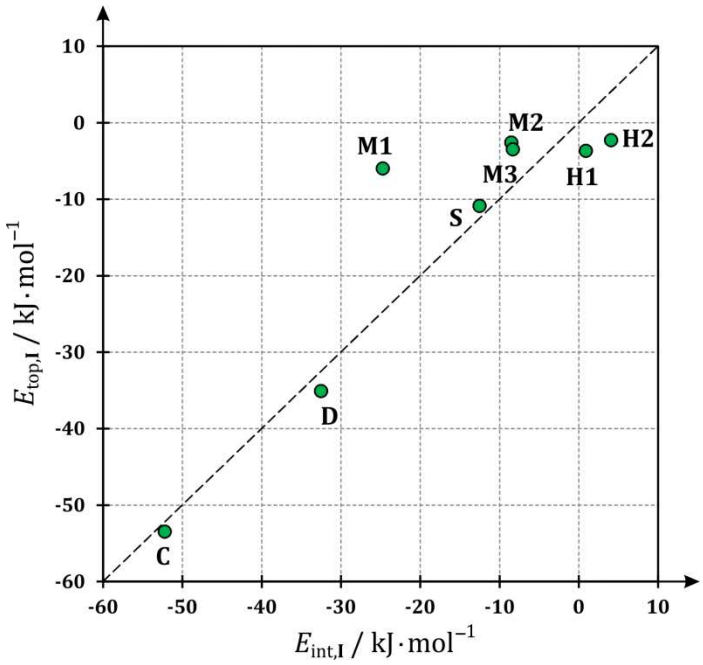


Figure 4.22. Correlation plot of $E_{top,I}$ (sum of interaction energies obtained via the Espinosa approach) *vs.* $E_{int,I}$ (*CRYSTAL*-computed total interaction energy). Dashed line, $E_{top,I} = E_{int,I}$, serves as a reference line.

Even though the sulphur atom exhibits a rather small negative charge (*ca.* −0.27 e), it forms many different interatomic contacts in the solid state. This is mainly due to its size (about 33 Å^3), and consequently its high polarizability. As a result, seven diverse intermolecular BCPs and the respective BPs, leading to the sulphur atom, were found. Two of them are engaged in the **D** and **S** motifs. Nevertheless, a detailed analysis of the sulphur atom is a bit hampered by the unrefined multipolar populations. The transferred UBDB parameters impose a certain fixed distribution of the lone electron pairs (and consequently Valence Shell Charge Concentrations (VSCCs)) on the sulphur atom. One of VSCCs is involved in the interaction with the H3 hydrogen atom, which stabilises the **D** motif. The second one points towards the H5 atom. The remaining atoms interacting with the sulphur atom are engaged in less significant intermolecular contacts. On the whole, the sulphur atom seems to substantially affect the crystal stability and so the formed crystal architecture. At the end it is worth noting that the presented topological features correspond very well with the respective values obtained for the theoretical charge density model (*i.e.* the one derived on the basis of the computed structure factors). For details see the Appendix E, Comment E12, Tables E13 & E14.

Table 4.16. AIM charges (Q_{AIM}) and volumes (V_{AIM}) for the **6m2tU** molecule present in the crystal lattice.

Atom	Q_{AIM} / e	V_{AIM} / Å^3
S2	−0.273	32.815
O4	−1.161	16.622
N1	−1.001	12.154
N3	−1.063	13.318
C4	+1.330	5.546
C2	+0.558	7.756
C5	+0.027	11.019
C6	+0.278	8.677
C7	−0.100	10.935
H1	+0.510	2.234
H3	+0.415	3.153
H5	+0.116	6.497
H7A	+0.112	5.761
H7B	+0.140	6.109
H7C	+0.118	6.605

IV.4a.4. Electrostatic potential analysis

Multipole model allows for reconstructing the electron density distribution of the crystal structure, and on that basis, for deriving various electrostatic properties on that basis, *e.g.* electrostatic potential (*ESP*). In turn, such electrostatic potential mapped onto the molecular surface may provide some additional insight into the nature of intermolecular contacts inside the crystalline state. As suggested by Spackman *et al.*,[136] the problem of overlapping molecular domains can be overcome when Hirshfeld surfaces are employed. Therefore, I have computed the electrostatic potential solely for the central molecule and mapped it onto the Hirshfeld surface. Figure 4.23*a* illustrates the Hirshfeld surface along the **C**- and **D**-type dimeric motif bond paths. The nature of these interactions is well visible as the complementary *ESP* region contacts. In the case of **D**, the electropositive region of H3 hydrogen atoms interacts with the electronegative region of the S2 sulphur atom. Essentially the same pattern is observed for the **C** motif. *ESP* distribution clearly shows why **6m2tU** forms layers, where molecules are bound by strong, electrostatic in nature, hydrogen bonds. Secondary organization is obtained by the π-stacking interactions, being more dispersive in nature. Naturally, in the solid state the molecular electron density (and thus *ESP*) is influenced by the surrounding molecules. The *ESP* computed by taking into account contributions of the neighbouring moieties is presented in Figure 4.23*c*. Interactions with other molecules change the *ESP* dramatically. The influence of the neighbouring molecules counterbalances the

electrostatic potential of a single **6m2tU** moiety, leading to a more uniform electrostatic potential distribution.

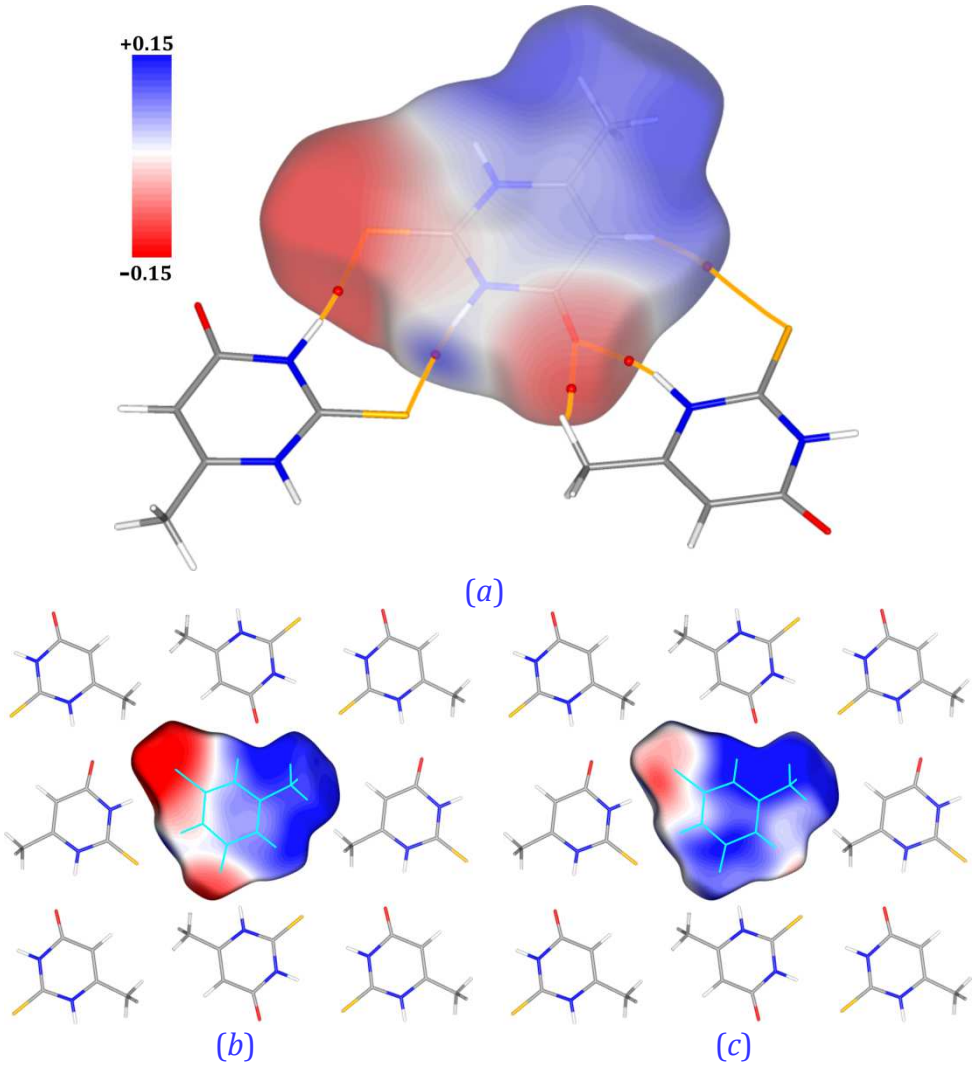


Figure 4.23. (*a*) Molecular electrostatic potential mapped on the Hirshfeld surface (semi-transparent representation) together with two neighbouring molecules forming **C** and **D** motifs (units are e·Å^{-1}; small red spheres – BCPs, yellow lines – BPs). (*b*) Layer fragment with Hirshfeld surface coloured with molecular *ESP*. (*c*) Layer fragment with Hirshfeld surface coloured with *ESP* coming from a central molecule together with its surrounding in a crystal lattice.

IV.4a.5. Periodic quantum chemical computations

In general, **6m2tU** forms two types of crystals from water solution. These are crystal prisms with well-defined facets (Figure 4.24), and thin elongated plates not suitable for high quality X-ray experiments. What is more, as already noted, **6m2tU** crystals exhibit some tendency to twin. The observed crystal morphology, growth, or twinning, are all complex problems involving kinetic and thermodynamic processes, and obviously

solvation effects. Nevertheless, a detailed analysis of the crystal architecture features may shed some light on the above-mentioned aspects.

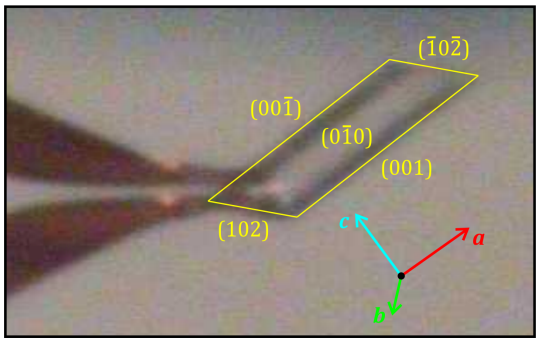


Figure 4.24. 6m2tU monocrystal with the indexed crystal faces.

Therefore, taking into account the layered crystal network of **6m2tU**, the interlayer interactions become worth exploring. As clearly seen in Figure 4.25, the previously analysed slabs parallel to the (102) crystal plane, are not the only easily distinguished ones (IV.3).[93] The regularity of **6m2tU** crystal architecture and the simplicity of molecules allow for detecting differently oriented molecular layers. These various molecular layers also interact with one another in different manners, involving either mostly dispersive contacts, or some stronger hydrogen bond interactions. For the purpose of the analysis, I have chosen molecular slabs related to the measured crystal face indices and to the ones previously mentioned in the literature.[94e]

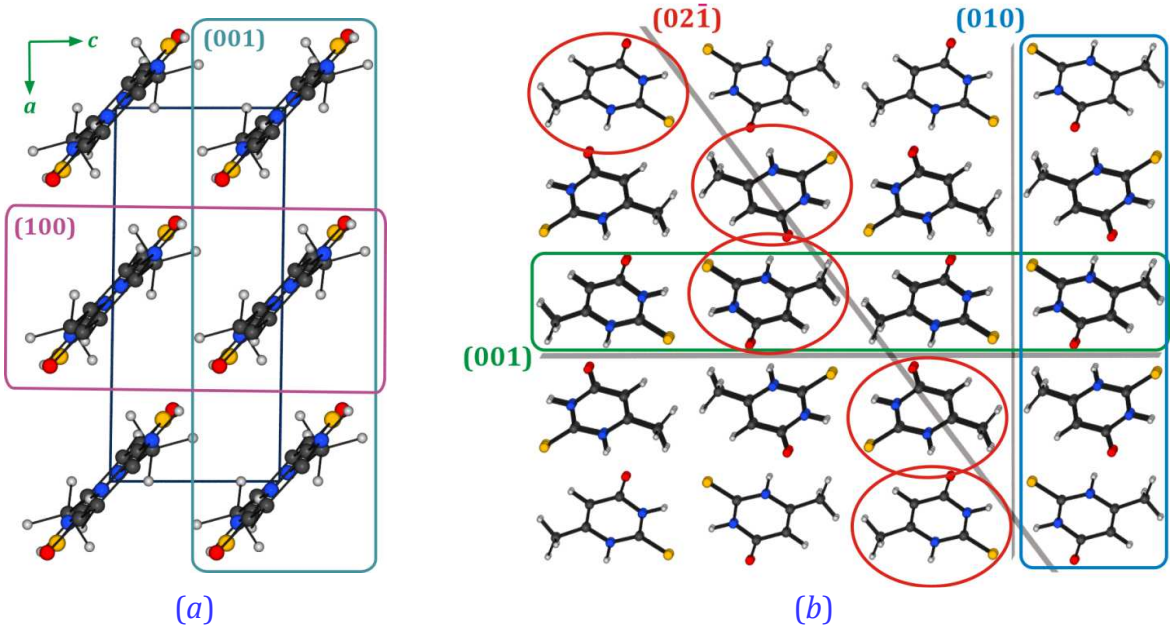


Figure 4.25. (*a*) Projection of the **6m2tU** crystal packing along the *Y*-axis; (*b*) Selected molecular slabs parallel to different crystal planes illustrated along the *Z*-axis.

To characterise each of the slab types, I considered two energetic parameters, *i.e.* interlayer interaction energy and surface free energy. The interlayer interaction energy was calculated at the DFT(B3LYP)/pVTZ level of theory following the procedure described in §IV.3.2*(iv)* and in Appendix A (Comment A3), whereas the evaluated surface free energy constitutes a rough estimation provided on the basis of the classical approach applied to metal or metal oxide surfaces (Appendix A, Comment A4).[62a] The obtained energetic values for the selected molecular slabs are put together in Table 4.17, and all the computational details are available from the Appendix A, E.

Table 4.17. Interlayer interaction energy (E_{intl}) and estimated surface free energy (E_{surf}) of the selected molecular slabs. Miller indices indicate the crystal plane family being parallel to a chosen slab (see Figure 4.25).

Slab	*Miller index*	E_{intl} / kJ·mol^{-1}	E_{surf} / J·m^{-2}
SB1	(102)	−48	0.104
SB2	(010)	−33	0.145
SB3	(001)	−102	0.296
SB4	(100)	−83	0.220
SB5	(110)	−93	0.230
SB6	$(02\bar{1})$	−74	0.255

In general, the calculated interlayer energies, together with the surface free energy magnitudes, should explain to some extent the crystal shape and its growth tendencies. The strength of an inter-slab contact provides some information about the growth rate in the related crystallographic direction. In turn, the free energy estimated per molecular area somewhat reflects the relative stability of the molecular surfaces being compared. The greater the surface free energy is, the less stable the molecular surface and, thus, the corresponding crystal faces less likely formed. Of course, one should keep in mind the other factors governing crystal formation. Solvent molecules usually significantly change the properties of pure substance crystal faces.

Having carefully looked at the **6m2tU** crystal morphology and its face indices, one can notice that the crystal is elongated in the [100] direction, exhibiting a well-developed {010}-type crystal faces, more rectangular medium {001}-type ones, and the smallest facets assigned to the {102} form. There are also other less significant crystal faces visible, *e.g.* the $\{02\bar{1}\}$ -type ones.

Among the studied crystal faces, {001} are characterised by the lowest interlayer interaction energy of the corresponding molecular slabs and the highest surface free

energy value (Figure 4.25*a*, Table 4.17). The {100} set of facets seems to be very similar in nature, however, more energetically balanced. The related molecular slab types, SB3 and **SB4**, respectively, interact via the N–H…O hydrogen bonds, what leads to the high interlayer interaction energy values. This may suggest that the crystal growth is faster along the [001] or [100] direction, and the formed surfaces of the adequate facets are limited in size. Indeed, depending on the solvent, either [100] or [001] is parallel to the elongation direction of the crystal. In water solution a **6m2tU** crystal is elongated along the *X*-axis. Nevertheless, it is exposed with the {102} facet types in that direction, instead of {100}, due to their better stability. According to the literature reports,[94e] in the 1:1 ratio mixture of glacial acidic acid and water, the obtained crystals are extended along the [001] direction. Therefore, this is a complementary situation to the one observed in water. Molecules of **6m2tU** are easily attached to such molecular surfaces contributing to the crystal elongation process. However, when such a surface is formed, it might attract also water molecules. Water molecules may in turn improve the surface stability and slow down the crystal growing in the corresponding direction. Such solvent influence may also cause some defects in the resulting crystal lattice or stimulate twinning.

On the other hand, in line with our expectations, the most developed {010}-type crystal facet corresponds to the molecular slabs (**SB2**) of the greatest interlayer interaction energy (*i.e.* the least interlayer interaction strength) and the smallest surface free energy. The molecular layers of the **SB2** type interact via the N–H…S hydrogen bonds and other less significant contacts.

The remaining crystal faces listed in Table 4.17 constitute examples of more rough molecular surfaces, which are less likely formed. This happens probably because these surfaces have less preferable energetic indices. It might also result from the specificity of crystallisation process, its kinetics, and from possible unfavourable consequences for other, more stable, crystal faces.

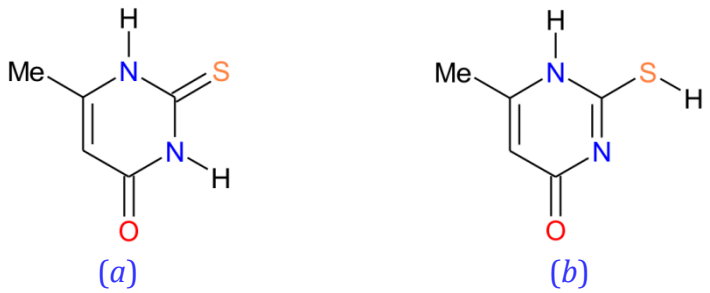
Despite many different effects affecting the crystal growth, our study shows that the simple analysis of energetic features of crystal network motifs may also lead to meaningful conclusions. After all, crystal thermodynamic properties, solvation effects, and crystallization kinetics in general, are all factors responsible for the final product, *i.e.* crystal and its features. Much of the information is also reflected in the resulting crystal

architecture. Generally, balanced and advantageous stabilisation energy contributes to the overall thermodynamic and mechanical crystal stability.

IV.4a.6. Tautomer analysis

Difficulties with the interpretation of the residual density around the sulphur atom (for datasets collected at both temperatures) led me to investigate the phenomenon of tautomerism in the solid state. In general, nucleic acid bases may exist, with different dose of probability, in several tautomeric forms. Numerous computational and spectroscopic studies have already been devoted to this problem, as it is believed to be tightly related to the base-pairing mutations.[96a, 96b, 137] In the case of uracil derivatives the energy difference between the normal and rare tautomers is greater than in the case of guanine or cytosine. For mono-substituted uracil modifications the oxo-thione form is the most favourable one, especially in the solid state (Scheme 4.2). On the other hand, according to the recent computational and spectroscopic studies of thio-uracil derivatives, single molecules might be quite easily photo-induced and transformed from the thione to thiol form.[96a, 96b]

Depending on the chemical environment, the tautomeric transformation process might be either an intramolecular or intermolecular proton transfer. Nowadays a lot of research deals with the proton transfer phenomenon taking place in the crystal lattice.[138] Is it then possible that **6m2tU** exists in more than one tautomeric form in its crystal, *e.g.* the ones shown in Scheme 4.2? To approach this problem, I performed a set of computational investigations together with a careful analysis of the charge density features.



Scheme 4.2. Two possible tautomeric forms of **6m2tU** related by the proton transfer between nitrogen and sulphur atoms: (*a*) tautomer **I** (oxo-thione form) and (*b*) tautomer **II** (oxo-thiol form).

The two high residual density peaks $\Delta\varrho_{max}$ = +0.64 e·Å^{-3}) observed in the vicinity of sulphur atom are presented in Figure 4.26*a*. Their location out of the plane of the **6m2tU** aromatic ring is not supported by the UBDB-transferred sulphur atom deformation parameters for the C=S fragment (Figure 4.26*b*). Nevertheless, if the second tautomer (tautomer **II**) is taken into consideration, one can find that the arrangement of the unexplained charge density resembles the electron lone pair distribution of the C–S–H group located in the molecular plane (Figure 4,26*c,d*). Therefore, to supply this assumption, I modelled a hypothetical geometry of tautomer **II** in the crystal lattice. For that purpose, an artificial crystal network, analogous to the one of tautomer **I** (the oxo-thione form) and based on tautomer **II**, was created and optimised in the *CRYSTAL* package (for details see Appendix E, Comment E15). Having the potential molecular geometry, I could generate the deformation density maps of tautomer **II** with the aid of the UBDB and ELMAM databases (Figues 4.26*c,d*).[29d, 70] Indeed, the shape and positions of the C–S–H sulphur atom valence charge density fitted the earlier unexplained residual density features. This seems to be an evidence of the presence of both tautomers in the crystal lattice. Unfortunately, it was not possible to refine those fine features of the electron density distribution. Thus, I could estimate only the percentage of the possible second tautomer content at approximately 10% (calculated as $\Delta\varrho_{max}$ × 100% / $P_{v,S}$).

The computational results revealed about 90 kJ·mol^{-1} energy difference between the molecule of the oxo-thione form and that of the hypothetically formed tautomer **II** (Appendix E, Comment E15, Tables E16 & E17). Despite the substantially less beneficial molecular energy of tautomer **II** in respect to tautomer **I**, the cohesive energy of the hypothetical crystal structure constructed of pure tautomer **II** seems to be more advantageous, which is of about 30 kJ·mol^{-1} (tautomer **I**: E_{coh} = −132 kJ·mol^{-1}, tautomer **II**: E_{coh} = −162 kJ·mol^{-1}). Nevertheless, considering the balance of the molecular and cohesive energy, the oxo-thione form still remains the more beneficial one. The energy results obtained for the selected dimer motifs show much more negative values for the hydrogen-bonded oxo-thiol species. This surely contributes to the overall crystal lattice stability value. The corresponding dimer interaction energy values are grouped in Table 4.18. These results reflect stronger electrostatic interactions and thus an increased stability of **C** and **D** units for tautomer **II**, when compared to the oxo-thione form of **6m2tU** (Table 4.15). This is caused by a higher charge separation. In consequence, the potential of the N3 atom is more negative, while for the S–H fragment it is more positive,

as illustrated in Figure 4.27. On the other hand, more pronounced charge partition causes lower values of inter-layer contacts.

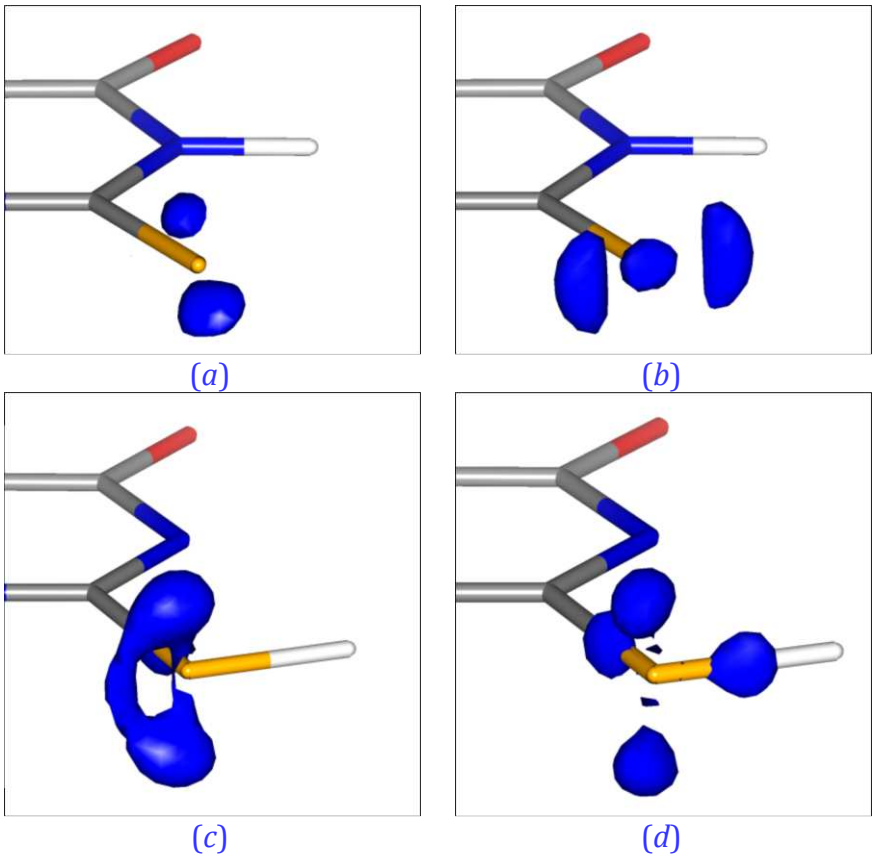


Figure 4.26. (*a*) Residual electron density near the sulphur atom (isosurface drawn at +0.3 e·Å^{-3}). (*b*) Standard UBDB-reconstructed deformation density for the sulphur atom (+0.2 e·Å^{-3}). (*c*) Deformation density taken from the ELMAM library (+0.2 e·Å^{-3}) for computed tautomer **II**. (*d*) Deformation density taken from the UBDB databank (+0.2 e·Å^{-3}) for computed tautomer **II**.

Table 4.18. Selected dimer motif intermolecular interaction energies calculated for the optimised hypothetical tautomer **II** crystal structure (analogue to that of tautomer **I**) at the DFT(B3LYP)/pVTZ level of theory, accounted for BSSE and dispersion,

Motif	$E^{\mathrm{opt}}_{\mathrm{int,II}}$ / kJ·mol^{-1}
C	−61.1
D	−82.4
S	−1.8
M1	−20.4
M2	+1.0
M3	−3.0
H1	–
H2	–

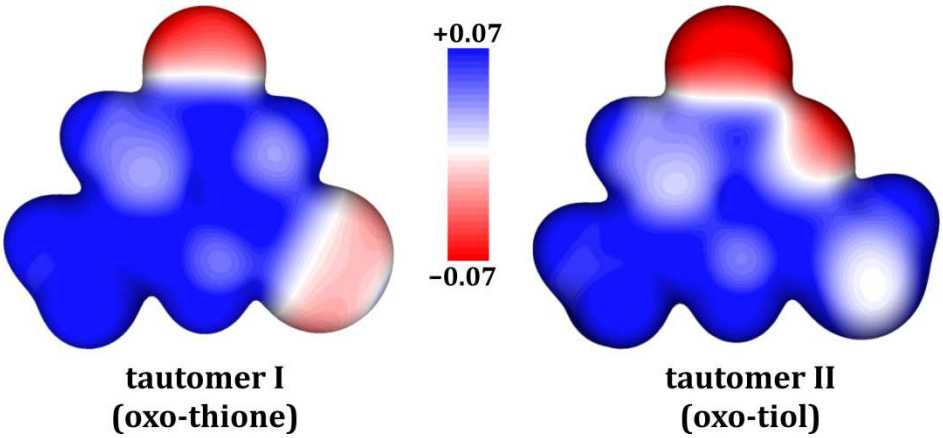


Figure 4.27. Electrostatic potential ($e \cdot a_0^{-1}$) mapped on the electron density isosurface of tautomer **I** and tautomer **II**, computed at the DFT(B3LYP)/aug-cc-pVDZ level of theory in the *GAUSSIAN* package.[54]

Although the above observations and results seem to be quite convincing, full explanation of the charge density features, which would assess the provided hypothesis, requires some additional research. As it is difficult to grow **6m2tU** crystals to the size suitable for the neutron diffraction studies, solid state NMR and/or some photocrystallographic studies could be an option and may lead to the final answer. Such investigations are scheduled for the nearest future.

IV.4a.7. Summary

The study provides a detailed charge density distribution analysis together with comprehensive energetic investigations. Furthermore, the findings constitute a link between crystal network features, electronic structure, molecular motif interaction energy and crystal morphology. My results show the importance of careful analysis of obtained charge density data and attentive charge density distribution model evaluation. A proper interpretation of anharmonic motion parameters resulting from the multipole refinement might be especially problematic. Therefore, I shall emphasise the need of a very low temperature (down to 10 K) X-ray diffraction experiment, which in my case allowed for excluding the supposition of anharmonic behaviour of the sulphur atom. On the other hand, once the model was properly derived, a number of interesting relations could have been found. My results revealed accordance between the topologically estimated hydrogen bond energy and the computational values. It also occurred that the methyl group and sulphur atom significantly affect crystal packing and contribute to a variety of intermolecular contacts. Nevertheless, the strongest

interactions observed in the crystal lattice are of electrostatic type. It is well reflected in the generated molecular electrostatic potential surfaces of the complementary character. Additionally, **6m2tU** exhibits a layered architecture which can be quite easily associated with its crystal morphology. Simple analysis of the inter-slab interaction energies and surface free energy values led to the rough explanation of the crystal elongation direction and preferably created facets, such as the well-developed {010} crystal form. Finally, the complex analysis of the residual charge density features, the reciprocal space, and different charge density models suggested the existence of the second tautomer in the crystal lattice, the oxo-thiol form. Unfortunately, it was impossible to reliably refine such a model. However, the percentage contribution of each tautomer might be estimated as 1:9 for the benefit of the 90 $kJ \cdot mol^{-1}$ more stable oxo-thione form. What is interesting, despite the higher molecular energy of the oxo-thiol tautomer, it may form a more advantageous crystal architecture due to much more pronounced electrostatic interactions, as indicated by the electrostatic potential and dimer interaction energy values. Nevertheless, to finally confirm or reject the hypothesis, a set of further investigations need to be conducted.

At the end, it is worth stressing that the databank was not only used to evaluate a starting model for further charge density studies, but it was also employed to reproduce the problematic sulphur atom charge density distribution.

IV.4b. 9-methyladenine:1-methylthymine charge density study

Nucleic acid chains play, undoubtedly, a crucial role in biology, coding and storing most of the information pertaining to living organisms. Consequently, they have been attracting much of researchers' attention, ever since they were discovered.[98b, 139] Numerous reports concern base pairing phenomenon in DNA and more complex inter-residue interactions present in RNA macromolecules.[72, 94k-p, 98b, 139g, 139h, 140] Also the number of published crystal X-ray and NMR structures is increasing each year. Additionally, a lot of computational effort has been devoted to the third and fourth order structure of nucleic acids, their stability and folding, and also to the role of stacking type interactions.[72-73, 76, 94l-n, 140b, 141]

Nonetheless, there are still few high quality crystal and co-crystal structures of nucleic acid bases and their modifications. The available crystallographic data usually comes from mid-twentieth century, and thus is of lower accuracy and/or lacks the determined hydrogen atom positions.[106a, 142] Therefore, I had already undertaken the subject of uracil derivatives (§IV.3),[93] whereas here, I focused my attention on hydrogen bonded nucleic acid base pair crystals. In consequence, this part of my PhD thesis is dedicated to the 9-methyladenine:1-methylthymine co-crystal structure (**9mA:1mT**).

The **9mA:1mT** structure was determined by Hoogsteen in 1959[142g] and further described in 1963.[106a] The aim was to synthesize nucleic acid base pair which would reflect the ones present in DNA chains. The most desirable crystal was found to be the one composed of derivatives of thymine and adenine in which the respective 1 and 9 nitrogen positions are blocked by the methyl substituent. At that time, however, structure determination itself was regarded as a great success. Therefore, these first works presented mainly structural features of the studied system and some cursory reference to DNA structure.

Thus, my contribution offers not only a more accurate structure determination, but also includes the first charge density analysis of the crystal containing nucleic acid base pair (Hübschle *et al.* presented some similar charge density data at Sagamore XVth conference in 2006, but these were never published in the form of article). This, in turn, opens up the possibilities of the topological studies and deeper investigations of the intermolecular interaction types and their nature. A detailed charge density distribution exploration is supplemented by comprehensive energetic considerations. I compared

and analysed the relations between the co-crystal structure and the crystal networks of its single components by means of molecular motifs, intermolecular contacts and their stabilisation energy. My aim was also to explain the specific base pairing in a crystal in respect to the preferable configurations found in RNA or DNA structures. I examined both hydrogen bond aspects and stacking interactions.

IV.4b.1. Materials and methods

(i) Crystal synthesis and data collection. Crystallization of 9-methyladenine (**9mA**) was adjusted in a water solution and lead to regularly shaped transparent crystals. In order to receive 9-methyladenine with 1-methylthymine co-crystals, equal 0.0001 molar amounts of the components were mixed together in 5 mL volume of distilled water. Subsequently, the obtained suspension was heated to 80°C and sustained until the substrates dissolved entirely. Such a solution was left for slow evaporation of solvent, which resulted in transparent prismatic crystals. (*) **9mA:1mT** crystal provided by Dr Roman Gajda.

As I had already measured and analysed 100 K 1mT (**15dmU**) structure, two other X-ray experiments – a standard one of **9mA**, and a high-resolution one of **9mA:1mT**, were carried out. The details on all the measurements and data processing are given in the Appendix F (Comment F1).

(ii) Structure determination and multipole refinement. Both crystal structures, *i.e.* **9mA** and **9mA:1mT**, were initially solved by a charge-flipping method with the *SUPERFLIP* program,[129] and refined on F^2 in the *JANA* package,[47] within the Independent Atom Model (IAM). Atomic scattering factors, in their analytical form, were taken from the International Tables for Crystallography.[4] All non-H atoms were refined anisotropically. Hydrogen atoms were placed geometrically within the riding model for their isotropic ADPs ($d_{\mathrm{C-H}}$ = 0.96 Å, $U_{\mathrm{iso}}^{\mathrm{H}} = 1.2 \cdot U_{\mathrm{eq}}^{\mathrm{C}}$; $d_{\mathrm{N-H}}$ = 0.87 Å, $U_{\mathrm{iso}}^{\mathrm{H}} = 1.5 \cdot U_{\mathrm{eq}}^{\mathrm{N}}$). Orientations of methyl groups were determined from Fourier maps. Final statistics of spherical refinement together with the comments specific for each structure are presented in the Appendix F (Comment F4).

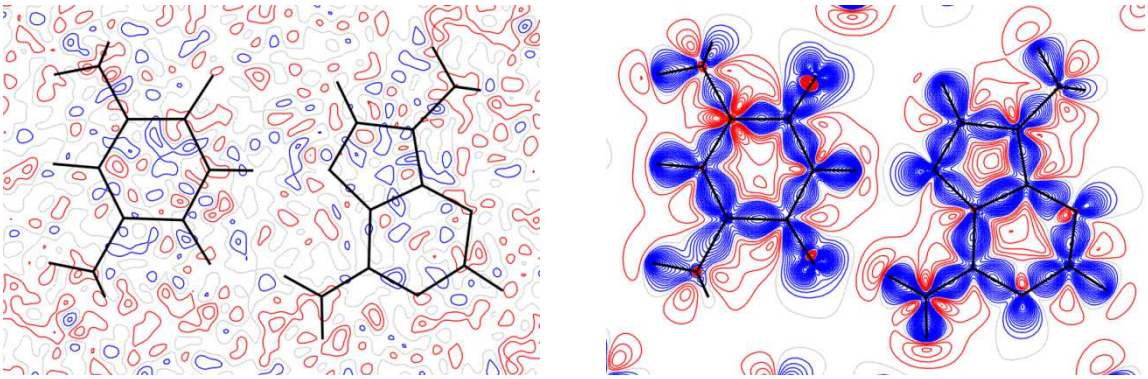
In the case of **9mA**, I subsequently conducted a Transferable Aspherical Atom Model (TAAM) refinement, according to the procedure described earlier for uracil derivatives. In turn, for the **9mA:1mT** crystal structure, similarly to the case of **6m2tU**, these preliminary results served as a starting point for further multipole refinement. Charge density model was evaluated within the Hansen-Coppens formalism[10] with the aid of

the *MOPRO* suite[35b] combined with the new version of the University at Buffalo Databank (UBDB2011).[45] All deformation parameters were defined in respect to their local Cartesian coordinate systems assigned by the *LSDB* program.

Charge density refinement was based on *F*, and only the reflections fulfilling the $|F^{o}|^2 \geq 3\sigma(|F^{o}|^2)$ condition were taken into account as this was found to provide reliable results. During the refinement, statistical weights were applied (*i.e.* for *i*-th reflection $w_i = 1/\sigma_i^2$). As stated before, the initial atomic coordinates, *x*, *y*, and *z*, and anisotropic displacement parameters (U_{ij}'s) for each atom were taken from the spherical refinement stage, whereas the initial multipolar and contraction-expansion parameters were transferred from UBDB2011 via the *LSDB* program. However, the positional, thermal and multipolar parameters of almost all atoms were constrained to specific values (*e.g.* y_{O2} = 0.75) or zeroed (*e.g.* $U_{12,O4} = U_{23,O4} = 0$ or $P_{20,C7} = 0$), because of the fact that both molecules (**1mT$_c$** and **9mA$_c$**) occupy a special position, being located at a mirror plane. In the case of methyl groups, the default 3*m* symmetry was decreased to *m*, consistent with the space group set of operators. In the initial stage of the refinement, the hydrogen atom U_{iso} parameters (i.e. isotropic thermal parameters) were restrained to the value of $y \cdot U_{eq}^{X}$ (y = 1.2 and 1.5 for X = C and X = N, respectively) with σ = 0.01 (where an appropriate restraint weight is equal to $1/\sigma^2$). In the final stage, the derivation of anisotropic hydrogen atom ADPs was carried out using the SHADE server. Each molecule, *i.e.* either **9mA$_c$** or **1mT$_c$**, was treated separately. The **9mA$_c$** fragment was found more problematic and, consequently, less accurately fitted by means of the TLS motion formalism. Consequently, it has to be kept in mind that the obtained hydrogen atom ADPs for **9mA$_c$** moiety are less reliable and could affect the static electron density distribution. The X–H bond lengths were restrained to neutron-normalised distances[63] with σ = 0.001. The multipole expansion was truncated at the hexadecapole (l_{max} = 4) and quadrupole (l_{max} = 2) levels for all non-hydrogen and hydrogen atoms, respectively. The contraction-expansion parameters, κ and κ', for all atoms were kept fixed at the UBDB2011 values during the refinement. The general refinement strategy was analogous to the one applied for **6m2tU** (points 1-6, §IV.4a.1*(iv)*).

The resulting residual density is very flat and featureless ($\Delta\varrho_{res}$ = ±0.15 e·Å^{-3}) (Figure 4.28*a*). This observation is additionally confirmed by the fractal residual density analysis (Appendix F, Figure F3). The obtained atom parameters show no deviations

from the reasonable values. All details regarding the final multipole model are located in Appendix F (Comment F4), whereas the deformation density map is illustrated in Figure 4.28*b*.



(*a*) (*b*)

Figure 4.28. (*a*) Fourier residual density map (dimer plane, 0.05 e·Å^{-3} contours), $\Delta\varrho_{res}$ = ±0.15 e·Å^{-3}. (*b*) Dimer deformation charge density map (dimer plane, 0.05 e·Å^{-3} contours, blue – positive, red – negative values).

IV.4b.1. Structural remarks

As the co-crystal structure has already been a subject of crystallographic consideration, here I shall briefly recall just the most important facts. **9mA:1mT** crystallises in the monoclinic $P2_1/m$ space group, where the two molecules, constituting the asymmetric unit, are located at a mirror plane (Figure 4.29*a*). 1-methylthymine and 9-methyadenine bases do not form a standard Watson-Crick (WC) hydrogen bonded pair. Instead, according to the Leontis and Westhof (LW)[139g, 143] classification, they create a *cis*-Hoogsteen-Watson-Crick (HWc) AT motif (Figure 4.29*b*). **9mA:1mT** structural details and bond lengths are available from the Appendix F (Table F6), where for the comparative purposes I also recalled the corresponding values for the mono-component structures of **1mT** and **9mA**.

Similarly to the previously analysed uracil derivative crystal structures, **9mA:1mT** exhibits layered crystal architecture. In general, hydrogen bonded species form molecular tapes, which held by weaker intermolecular contacts, construct flat layers parallel to the (010) crystal plane. Such molecular slabs interact with the adjacent layers mainly via π-stacking contact types. All short contacts are summed up in Table 4.19 and Table 4.20 together with the analogue characterisation of both mono-component crystal structures of **1mT** and **9mA**.

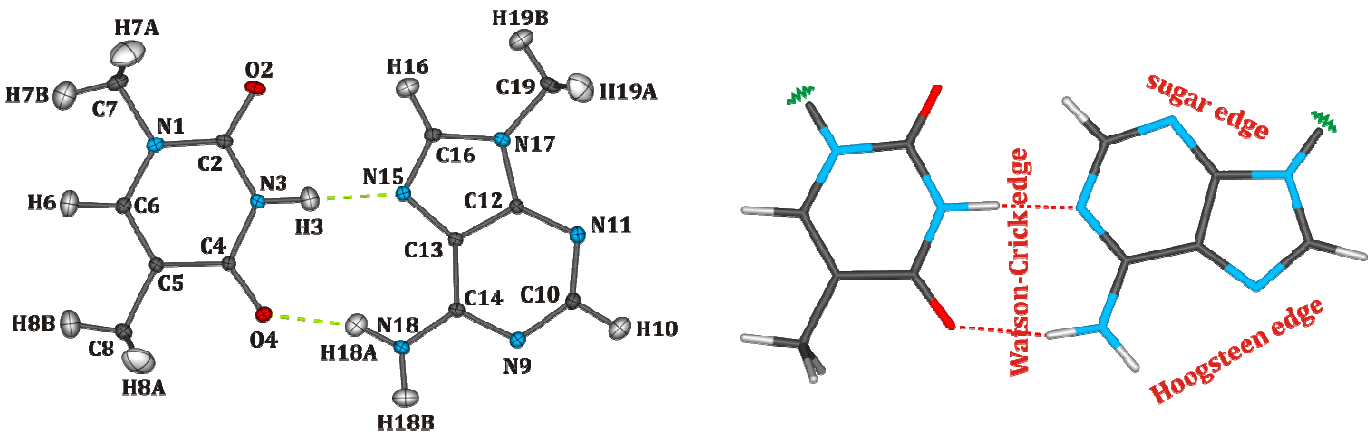


Figure 4.29. (*a*) ORTEP representation of the 1-methymine and 9-methyladenine dimer from **9mA:1mT**; thermal ellipsoids illustrated at 50% level of probability; the numbering of **9mA** fragment changed for comparative purposes (*b*) *cis*-Watson-Crick base-pairing **AT** motif with the indication of different adenine moiety interaction edges (LW).

Table 4.19. Geometrical parameters for selected hydrogen bonds and other intermolecular interactions obtained after TAAM refinements or charge density refinement in the case of **9mA:1mT** dimer. D and A denote interaction donor and acceptor atoms, respectively.

Compound	*Interaction*	d_{D-H} / Å	$d_{H...A}$ / Å	$d_{D...A}$ / Å	$\theta_{D-H...A}$ / Å
1mT	N3–H3...O4$^{\#1}$	1.029(1)	1.794(2)	2.8218(4)	176.2(2)
	C6–H6...O2$^{\#2}$	1.083(1)	2.045(2)	3.1079(5)	166.6(2)
	C7–H7C...O4$^{\#3}$	1.076(1)	2.448(3)	3.4513(6)	154.6(3)
	C8–H8C...O4$^{\#4}$	1.077(1)	2.643(6)	3.2267(5)	113.4(4)
9mA:1mT	N3–H3...N15	1.030(1)	1.874(1)	2.9042(4)	178.7(1)
	N18–H18A...O4	1.010(1)	1.852(2)	2.8250(4)	160.8(3)
	N18–H18B...O2$^{\#6}$	1.010(1)	1.938(5)	2.8672(5)	151.7(5)
	C19–H19A...N9$^{\#7}$	1.077(1)	2.699(5)	3.5761(2)	138.3(4)
	C6–H6...N11$^{\#8}$	1.083(1)	2.713(2)	3.7866(4)	171.2(2)
9mA	N18–H18A...N9$^{\#9}$	1.010(1)	1.968(5)	2.951(6)	163.7(4)
	N18–H18B...N15$^{\#10}$	1.010(1)	2.046(4)	3.052(4)	174.0(2)
	C19–H19B...N11$^{\#11}$	1.077(1)	2.536(8)	3.349(3)	131.6(6)
	C19–H19A...N9$^{\#12}$	1.077(1)	2.655(5)	3.648(2)	152.9(4)

Symmetry transformations: (#1) $-x+1, -y, -z+2$; (#2) $-x+2, y+0.5, -z+1.5$; (#3) $x+1, y, z$; (#4) $x, -y+0.5, z-0.5$; (#5) $x, -y+0.5, z+0.5$; (#6) $x+1, y, z$; (#7) $-x+2, y+0.5, -z$; (#8) $x, y, z+1$; (#9) $x-0.5, -y+0.5, z-0.5$; (#10) $x+0.5, -y+0.5, z+0.5$; (#11) $x-0.5, -y+1.5, z-0.5$; (#12) $-x+0.5, y+0.5, -z+0.5$

Among the hydrogen bond contacts present within the molecular layers, there can be distinguished three most significant ones, exhibiting a typical geometry for these kinds of interactions. These are the two hydrogen bonds linking the **9mA$_c$** and **1mT$_c$** base pair (Figure 4.29*a*), *i.e.* N3–H3...N15 and N18–H18A...O4 (Table 4.19), and the third one, which combines the **1mT$_c$** and **9mA$_c$** moieties coming from the neighbouring

base dimers (N18–H18…O2), leading to the mentioned tape motif formation. As denoted by Table 4.19, there is also a presumably weaker, more distant contact present in the plane between two **9mA$_c$** molecules (C19–H19…N9), and an interlayer interaction between the **9mA$_c$** methyl group and the N11 nitrogen atom electron lone pair from the above located **9mA$_c$** molecule. Considering the other interlayer interactions preset in the three studied systems which are shown in Table 4.20, all of them are within the range of the typical π…π contacts. Unlike in **9mA:1mT**, in the pure **1mT** (**15dmU**) structure the molecular layers are a little bit wavy, and in the case of **9mA** they are even more undulated, however, in all the cases the average interlayer distance does not exceed 3.31 Å.

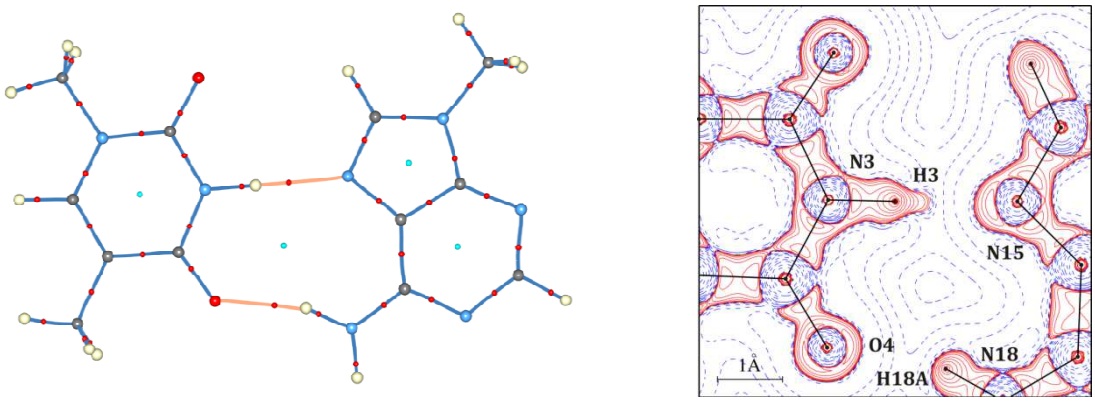
Table 4.20. Geometrical parameters for selected short intermolecular contacts obtained after TAAM refinements (**1mT**, **9mA**), or charge density refinement (**1mT:9mA**). X and Y denote pairs directly interacting atom pairs.

Compound	*Interaction*	X…Y / Å	*Compound*	*Interaction*	X…Y / Å
1mT	$C4_{\pi}…C6_{\pi}^{\#1}$	3.3723(5)	**9mA : 1mT**	$C10_{\pi}…C12_{\pi}^{\#3}$	3.1890(1)
	$C5_{\pi}…C6_{\pi}^{\#1}$	3.3350(5)		$N3_{\pi}…C4_{\pi}^{\#4}$	3.2430(1)
9mA	$C12_{\pi}…C16_{\pi}^{\#2}$	3.396(4)		O2…H8A$^{\#4}$	2.756(2)

Symmetry transformations: (#1) *x*, −*y*+0.5, *z*+0.5; (#2) −*x*, −*y*+1, −*z*; (#3) −*x*+2, *y*+0.5, −*z*; (#4) −*x*+2, −*y*+2, −*z*+1;

IV.4b.2. 9mA:1mT molecular topology analysis

The obtained charge density model reproduced very well the experimental results and appeared to be a reliable fit, as indicated by both residual and deformation density maps (Figure 4.28). Therefore, I subsequently analysed the evaluated charge density distribution within the QTAIM. To begin the investigations, I detected bond paths (BPs) and bond critical points (BCPs) in both molecular fragments, as this usually gives an overview of the topology quality. The obtained results are summarised in Table 4.21, whereas the **9mA:1mT** molecular graph is presented in Figure 4.30*a*. BPs and BCPs were found between all atoms in the molecule, what supported the consistence of the overall topology of the electron density. Generally, the values of the electron density and Laplacian are within typical ranges of the adequate bond types, considering the available data of similar compounds, *e.g.* earlier presented charge density studies of **6m2tU**, and literature 2-thiouracil (**2tU**), or cytosine (C) data.[94i] This is especially true for the aromatic ring C–C and C–N bonds, which exhibit comparable topological characteristics, having a mixed nature between the single and double bond.



(*a*) (*b*)

Figure 4.30. (*a*) Bond paths of intramolecular and intermolecular contacts with the corresponding BCPs shown. (*b*) Laplacian of charge density in the plane of a dimer (logarithmic contours).

Their ellipticity parameter oscillates around 0.20, however, C–C bonds usually exhibit greater ε values. Among the latter type of bonds, the C4–C5 and C12–C13 (the common bond of the fused aromatic rings in the **9mA** fragment) reveal the largest double bond character (ε = 0.27 and ε = 0.29, respectively). The C5–C6 bonds in the analysed **1mT** fragment and in **6m2tU**, having almost exactly the same chemical neighbourhood, are also very similarly described in topological terms. Though, in the absence of the methyl substituent, this bond gains some more double character as it is visible for the **2tU** charge density distribution.

Table 4.21. Selected QTAIM parameters of covalent bonds at BCPs for **9mA:1mT** (R – distance between bonded atoms; R_1 and R_2 – distances from 1st and 2nd atom to the bond critical point, respectively; ε – bond ellipticity).

Molecule	*Bond*	R / Å	R_1 / Å	R_2 / Å	$\varrho(\mathbf{r}_{BCP})$ / e·Å^{-3}	$\nabla^2\varrho(\mathbf{r}_{BCP})$ / e·Å^{-5}	ε
1mT	N1–C2	1.378	0.797	0.582	2.25	−25.4	0.21
	N1–C6	1.375	0.788	0.587	2.13	−17.4	0.14
	N1–C7	1.467	0.838	0.629	1.72	−12.8	0.13
	N3–C2	1.374	0.800	0.573	2.25	−24.3	0.18
	N3–C4	1.377	0.805	0.572	2.13	−21.4	0.11
	N3–H3	1.030	0.747	0.283	2.31	−38.0	0.03
	C2–O2	1.229	0.424	0.805	2.83	−25.6	0.07
	C4–O4	1.234	0.418	0.817	2.78	−19.8	0.17
	C4–C5	1.438	0.733	0.706	2.02	−18.2	0.27
	C5–C6	1.358	0.717	0.641	2.26	−21.7	0.22
	C5–C8	1.492	0.782	0.710	1.67	−10.6	0.01
	C6–H6	1.083	0.721	0.362	2.00	−26.6	0.05
	C7–H7A	1.077	0.702	0.375	1.81	−17.4	0.14
	C7–H7B	1.077	0.696	0.382	1.77	−15.7	0.20
	C8–H8A	1.077	0.686	0.391	1.98	−22.6	0.04
	C8–H8B	1.077	0.687	0.392	1.77	−16.9	0.06

	N9–C10	1.343	0.745	0.598	2.36	−21.4	0.17
	N9–C14	1.354	0.773	0.581	2.32	−21.5	0.08
	N11–C10	1.333	0.762	0.571	2.40	−24.2	0.17
	N11–C12	1.342	0.750	0.593	2.35	−21.2	0.10
	N15–C13	1.385	0.810	0.575	2.08	−16.4	0.17
	N15–C16	1.319	0.798	0.521	2.44	−24.3	0.17
	N17–C12	1.373	0.851	0.523	2.12	−22.4	0.21
9mA	N17–C16	1.370	0.800	0.570	2.16	−20.0	0.18
	N17–C19	1.451	0.852	0.599	1.74	−12.1	0.02
	N18–C14	1.333	0.796	0.538	2.41	−29.1	0.20
	C12–C13	1.391	0.725	0.666	2.20	−22.5	0.29
	C13–C14	1.413	0.717	0.697	2.09	−20.3	0.19
	N18–H18A	1.010	0.729	0.281	2.33	−34.1	0.05
	N18–H18B	1.010	0.731	0.279	2.39	−40.7	0.08
	C10–H10	1.083	0.747	0.336	1.87	−22.0	0.09
	C16–H16	1.083	0.722	0.361	1.93	−23.5	0.10
	C19–H19A	1.077	0.724	0.353	1.86	−20.2	0.04
	C19–H19B	1.077	0.718	0.359	1.73	−15.5	0.16

Similar trends to the ones observed between N–H and C–H bond types in the **6m2tU** charge density distribution, are also found in **9mA:1mT**. In the case of all N–H bonds, the BCPs are always shifted more than 0.05 Å towards the hydrogen atoms, when compared to analogical values of the C–H contacts. Apart from a greater atomic volume of nitrogen atom as compared to a carbon atom (however, here aromatic ring atoms should be compared separately to the atoms from methyl or amine groups) (Table 4.22) and electronegativity difference, the reason of such an effect may be related to the specificity of the hydrogen atoms attached to nitrogen atoms. These hydrogen atoms are all involved in rather strong hydrogen bond contacts, and thus the N–H are characterised by the relatively high magnitudes of electron density at BCP (over 2.30 e·Å^{-3}) and distinctively low Laplacian values (lower than −34.0 e·Å^{-5}). When the nature of hydrogen atom contacts is significantly different, it may also lead to visible discrepancies, even among the C–H bonds composing the same methyl fragment. Indeed, the C19 methyl group constitutes such a case, where C19–H19A bond is described by substantially lower Laplacian and ellipticity values than the C19–H19B analogue. This results from the fact that H19A (lower Laplacian value) is mediating the C19–HA…N9 interlayer interaction (Table 4.19 and Table 4.23), whereas H19B is engaged solely in the H…H type short contact. However, some other discrepancies may also come from the earlier noted less accurate thermal motion description for the **9mA$_c$** moiety, or constitute an artefact of the atomic symmetry imposed during refinement.

A deeper insight into the electronic charge distribution can also be obtained by analysis of atomic charges. These are presented in Table 4.22 along with the respective

atomic basin volumes. Oxygen and nitrogen atoms are, as expected, very much negatively charged. Both oxygen atoms, forming comparable hydrogen bonds, posses equal charges. However, more pronounced differences are observed among nitrogen atoms. The N1 and N3 atoms differ significantly from each other (*ca.* 0.3 e), also in respect to the case of **6m2tU**, due to the presence of the *N*-methyl substituent in the **1mT** moiety. In the **9mA** fragment, N15 is most negatively charged and most bulky when compared to the other aromatic ring nitrogen atoms. This is due to the strong hydrogen bond the N15 nitrogen atom participates in. As indicated by Figure 4.30*b*, N15 is quite significantly polarised with its lone electronic pair shifted towards the H3 atom. In turn, all hydrogen atoms are positively charged, among which, H3, involved in the presumably strongest N3–H3…N15 hydrogen bond, is characterised by the most positive charge (about +0.46 e). This assumption is also confirmed by the charge density Laplacian map (Figure 4.30*b*), showing the characteristic deformation of the H3 hydrogen atom Laplacian contour. Finally, carbon atom ring members reveal a rather predictable pattern, *i.e.* the atoms owing more electronegative neighbours are more positively charged. The C7 and C19 carbon atoms (from the methyl groups) are the only ones bearing the negative charge, which is rather small.

Table 4.22. AIM charges (Q_{AIM}) and volumes (V_{AIM}) of atoms from **9mA:1mT**.

1mT$_c$			**9mA$_c$**		
Atom	Q_{AIM} / e	V_{AIM} / Å^3	*Atom*	Q_{AIM} / e	V_{AIM} / Å^3
O2	−1.241	17.679	N9	−0.907	15.376
O4	−1.241	16.915	N11	−0.851	14.909
N1	−0.732	10.694	N15	−1.050	15.627
N3	−1.074	11.871	N17	−0.961	11.099
C2	+1.655	4.688	N18	−1.088	16.689
C4	+1.369	4.943	C10	+0.624	8.144
C5	−0.324	10.157	C12	+0.707	6.430
C6	+0.375	10.247	C13	+0.343	9.014
C7	−0.063	11.945	C14	+0.883	7.686
C8	−0.090	10.157	C16	+0.677	9.850
H3	+0.458	2.432	C19	−0.055	11.789
H6	+0.171	6.347	H10	+0.157	6.137
H7A	+0.194	5.519	H16	+0.114	7.490
H7B	+0.142	6.924	H18A	+0.472	2.606
H8A	+0.036	6.893	H18B	+0.469	3.190
H8B	+0.138	5.434	H19A	+0.142	7.242
			H19B	+0.185	6.275

IV.4b.3. Topological and energy study of 9mA:1mT

Another aspect of the charge density distribution in the crystalline state concerns the interactions of the studied molecules with their chemical surrounding. Therefore, I detected a number of distinct intermolecular BPs and BCPs. The numerical parameters for the possible interactions which I found significant or worth exploring are summarized in Table 4.23. The intermolecular BPs and BCPs are illustrated in Figure 4.31. Additionally, similarly as in the case of **6m2tU**, I supplemented my analysis with the energetic considerations employing the Espinosa *et al.* approach.[27, 135] This idea is derived from the Abramov equation[26] based on the virial theorem, and constitutes a natural link between the charge density features and interatomic interactions in the distance range from 0.5 Å to 2.1 Å from the atomic nuclei. According to the method's assumptions, the weak interaction energy of a particular intermolecular contact is considered as being equal to approximately half of the estimated potential energy density value at BCP. As hydrogen bonds are mostly electrostatic interactions, they are particularly well described by the presented formula, as already shown in literature[27, 135] and also supported by my previous studies of **6m2tU**. Thus, the application of this approach together with the further computational analysis will contribute to the quantification of the previously discussed hydrogen bonds.

Table 4.23. Selected QTAIM parameters of weak intermolecular interactions at BCPs.

Interaction	*Motif*	R / Å	R_1 / Å	R_2 / Å	$\varrho(\mathbf{r}_{BCP})$ / e·Å^{-3}	$\nabla^2\varrho(\mathbf{r}_{BCP})$ / e·Å^{-5}	$G(\mathbf{r}_{BCP})$ / kJ·mol^{-1}·a$_0^{-3}$	$V(\mathbf{r}_{BCP})$ / kJ·mol^{-1}·a$_0^{-3}$
H3…N15	**AT**	1.874	0.670	1.205	0.28	1.2	59.1	−85.3
H18A…O4	**AT**	1.852	0.643	1.209	0.18	2.3	60.6	−57.6
H6…N11$^{\#1}$	**iAT**	2.713	1.155	1.561	0.05	0.4	9.0	−7.7
H8B…H10$^{\#1}$	**iAT**	1.889	0.944	0.945	0.09	0.6	16.8	−17.3
O2…H18B$^{\#2}$	**D**	1.938	1.234	0.706	0.16	1.6	44.5	−44.7
H16…N9$^{\#2}$	**AA1**	2.916	1.251	1.671	0.03	0.3	6.8	−4.9
H19B…H10$^{\#2}$	**AA1**	2.205	1.081	1.154	0.03	0.5	10.6	−7.0
H19A…N9$^{\#3}$	**AA2**	2.699	1.113	1.587	0.04	0.6	11.7	−8.3
C10…C12$^{\#3}$	**AA2**	3.189	1.615	1.586	0.05	0.5	11.8	−9.1
N3…C5$^{\#4}$	**TT1**	3.318	1.613	1.745	0.05	0.6	12.3	−9.2
O4…C2$^{\#4}$	**TT1**	3.277	1.617	1.661	0.04	0.4	9.0	−6.5
O2…H8A$^{\#4}$	**TT1**	2.756	1.568	1.189	0.04	0.4	8.9	−6.8
O2…H7A$^{\#5}$	**TT2**	2.626	1.077	1.552	0.04	0.5	9.9	−7.1
H7A…H7A$^{\#5}$	**TT2**	2.380	1.190	1.190	0.03	0.3	6.9	−4.5

Symmetry transformations: (#1) $x, y, z+1$; (#2) $x+1, y, z$; (#3) $-x+2, y+0.5, -z$; (#4) $-x+2, -y+2, -z+1$; (#5) $x-1, y+0.5, -z+1$;

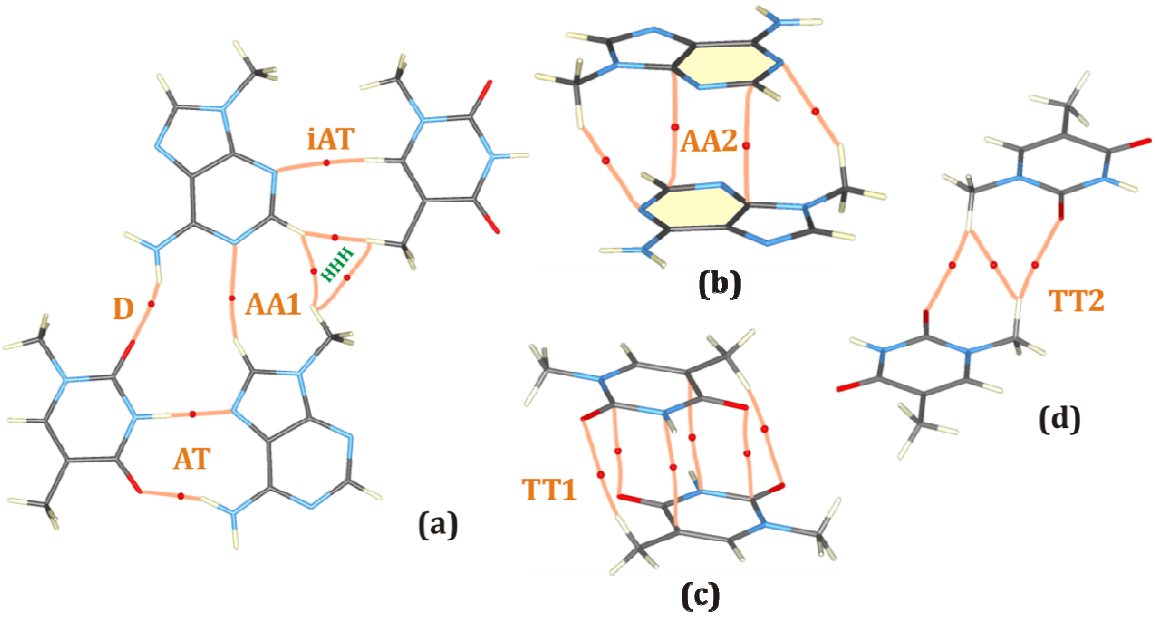


Figure 4.31. Selected dimers extracted from the **9mA:1mT** co-crystal structure with the illustrated bond paths and BCPs.

According to the topological characteristics, there can be distinguished four most significant dimers located in the same (010) crystal plane. These are the Watson-Crick-Hoogsteen base pair (AT) and its complementary interaction between the 'sugar edge' side of **9mA$_c$** and the C–H part of **1mT$_c$** (**iAT**), the hydrogen bond contact **D** between the O2 oxygen atom and the **9mA$_c$** amino group, and finally the only relevant in-plane interaction between the same type of nucleic acid bases – **AA1**. As indicated by Table 4.23, the typical hydrogen bonds are created within the **AT** and **D** motifs. As previously expected, the strongest one among them is the N3–H3…N15 contact. The potential energy density at the BCP of this hydrogen bond, being equal to −85.3 kJ·mol^{-1}·a$_0^{-3}$, is very similar to that of the very advantageous N1–H1…O4 contact observed in **6m2tU**. This topological value corresponds with the interaction energy equal to about −43 kJ·mol^{-1} in the Espinosa's approach. The other **AT** motif interatomic contact, O4…N18–H18A, contributes to the **AT** dimer interaction with the *ca.* −29 kJ·mol^{-1} energy value and constitutes the second strongest interaction in the co-crystal structure. The summary interaction energy of the **AT** unit estimated on the basis of the topological analysis agrees well with the DFT computed value (Table 4.24), leading to around −70 kJ·mol^{-1} (DFT(B3LYP)/pVTZ level of theory, methodology analogous to the one presented in the case of other uracil derivatives, for details see Appendix A, Comment A3). Interestingly, the studied case seems to be a more stabilising base pair orientation

than that of the standard Watson-Crick adenine-thymine dimer (−54 kJ·mol^{-1} on average) found in the DNA chain.[72] Such an observation can also result from a more effective mutual orientation of the corresponding purine and pyrimidine bases in a co-crystal lattice, where they are freer to adopt more favourable adjacent positions, than in a stiff DNA chain. However, the optimised WC-type A:T dimer (BD0037b constituted a starting geometry, Appendix B, Table B7)[72] is also characterised by slightly less advantageous interaction energy value in respect to the optimised **AT** motif (about 4 kJ·mol^{-1} difference). Additional methyl groups in **AT** should not affect the final energetic result significantly. I shall discuss this problem in more details, and wider context later.

Table 4.24. Selected motif interaction energies (E_{int}). Energy calculations corrected for BSSE and dispersion calculated at the DFT(B3LYP)/pVTZ level of theory in the supermolecular approach.

Compound	*Dimer code*	E_{int} / kJ·mol^{-1}
9mA:1mT	**AT**	−68.8 / −71.5[a] / −73.6[c] / −70.8[d]
	iAT	−13.1 / −12.5[a]
	D	−17.3 / −22.4[a]
	AA1	−14.1
	AA2	−32.4
	TT1	−38.9
	TT2	−19.7
	AT·AT	−174.3
	AT·iAT(π)	−167.3
	H·H·H	−26.7
A:T	BD0026b	−51.2 / −52.6[b]
A:T	BD0037b	−55.7 / −57.0[b] / −69.4[c]
A:T	BD0002b	−58.8 / −59.9[b]
A:T	BD0052e	−43.8 / −45.2[b]

[a] Espinosa *et al.* HB interaction energy values. [b] Structures with neutron-normalised X–H distances. [c] Optimised dimers. [d] 9mA:1mT dimer optimised in the crystal lattice (DFT(B3LYP)/pVTZ, Appendix A).

The third most significant hydrogen bond was detected in the **D** motif. The interaction energy of the **D** dimer is also quite satisfactorily reproduced by the Espinosa method. Finally, the interaction energies of the other two selected motifs, *i.e.* **iAT** and **AA1** dimers, do not exceed −15 kJ·mol^{-1}.

Another interesting feature of the analysed (010) molecular plane, are the surprisingly close H…H contacts (**HHH** in Figure 4.31*a*). Especially peculiar is the H8B…H10 interaction characterised by the remarkably high amplitude of the energy density values and the short interatomic distance equal to 1.889 Å. The interaction energy of the **HHH** motif (two **9mA$_c$** fragments and one **1mT$_c$**) is stabilising, however,

slightly less advantageous than the sum of the component energies of the single **AA1** and **iAT** dimers. On the other hand, the topologically estimated interaction energy of the **iAT** motif is comparable to the theoretically computed one only if the significant H8B…H10 potential energy density, having more impact than the H6…N11 contact, is taken into account.

Apart from the intermolecular contacts between the moieties belonging to the (010) crystal and molecular plane, several inter-slab interactions may be selected. It occurs that the most stabilised interlayer interactions are formed between molecules of the same type. The corresponding dimers **AA2**, **TT1**, and **TT2** are shown in Figure 4.31*b,c,d*, respectively. In the case of **1mT$_c$** dimers, the more advantageous one is **TT1**, where three distinct types of bond paths were detected. Beside the significant π-stacking/dispersive contacts, the interaction between the O2 oxygen and the C8 methyl group is observed. O2 also interacts with the C7 methyl substituent of another thymine moiety, which is the main interaction encountered for the **TT2** unit. Nevertheless, the **TT1** dimer is twice stronger stabilised than **TT2**. In turn, in the case of **AA2** there are no oxygen mediated interactions as in the **TT** motifs, instead the C19 methyl group contact with the N9 nitrogen atom was detected. Nevertheless, both **9mA$_c$** molecules interact remarkably, mostly due to a good π-stacking covering between the 6-membered aromatic rings (3.585 Å distance between 6-membered ring centroids).

Additionally, it was found that the stabilisation energy of the four molecule pattern constituting two dimeric 9-methyladenine:1-methylthymine pairs, located either above each other (**AT·iAT**(π)), or in the same plane (**AT·AT**, here joined via **D**, **AA1** and **HHH** BPs), does not differ much. However, it should be noted that within the **AT·iAT(π)** motif there is one most advantageous **AT** interaction and the second weaker **iAT** dimer, as these two are placed one above another. Despite the latter, the energy difference equals 7 kJ·mol^{-1} for the benefit of the in-plane motif. These results show the particular importance and strength of the inter-layer interactions.

IV.4b.4. Electrostatic potential analysis

In order to better apprehend the charge distribution and the nature of intermolecular interaction, the electrostatic potential (*ESP*) was mapped on Hirshfeld surfaces of **9mA$_c$** and **1mT$_c$** fragments. Figure 4.32*a* shows the electrostatic potential derived solely on the basis of the molecular charge density models. As clearly seen, the

electrostatic potential of both moieties is complementary, especially in the area of the existing hydrogen bonds. The electrostatic potential distribution of **9mA$_c$** is generally more uniform and characterised by more negative potential values than that of **1mT$_c$**. **1mT$_c$** is divided into the methyl substituent part with more positive potential, and the oxygen containing fragment of lower potential values. The appearance of the *ESP* generated for **9mA$_c$** and **1mT$_c$** suggests that the adjacent **9mA** molecules, located one above another, are less mutually shifted than the **1mT$_c$** ones, and thus interact more effectively in a π-stacking manner. In turn, **1mT$_c$** exhibits more beneficial oxygen atom - methyl group contacts, which agrees with the previous observations.

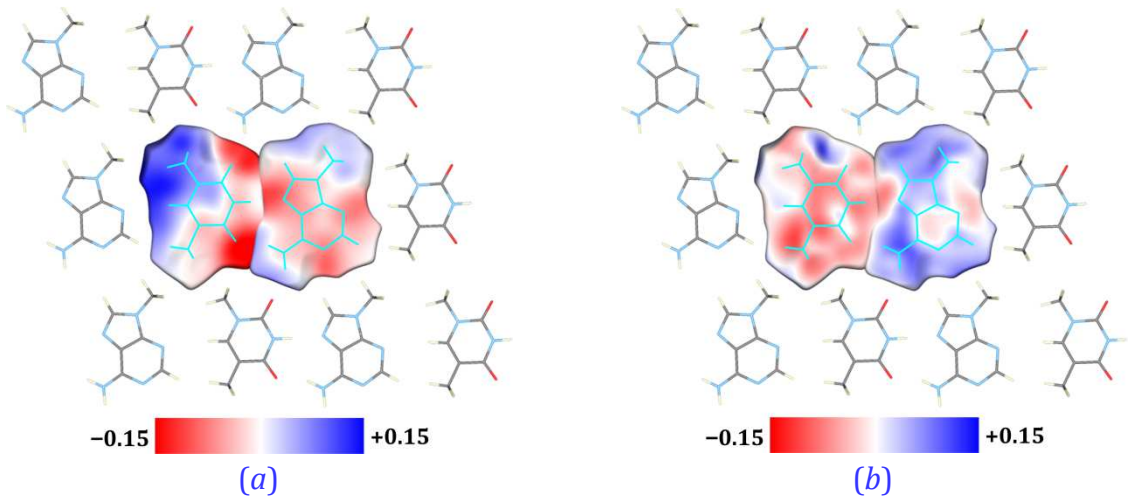


Figure 4.32. Electrostatic potentials mapped (units are e·Å^{-1}) on the Hirshfeld surfaces of **9mA** and **1mT** fragments in the co-crystal lattice (*a*) using molecular charge density models, (*b*) taking into account all contributions from the crystal to the Hirshfeld surface.

Nevertheless, the *ESP* features are changed diametrically, when the contributions coming from all surrounding molecules are also taken into account. This is illustrated in Figure 4.32*b*. The picture shows some kind of charge overlap, where the earlier more negative **9mA$_c$** potential becomes more positive, while the opposite effect is observed for the **1mT$_c$** fragment. However, no significant charge difference between **9mA$_c$** and **1mT$_c$** is observed for the final charge density model (even though the neutral charge was imposed on the whole dimer during the refinement). One can see that lower *ESP* values are found for the HB regions and for the areas of aromatic rings involved in more significant interactions (*e.g.* the particularly strong **TT1** motif stabilisation, the interaction between **9mA** pyrimidine rings).

IV.4b.5. Comparison of 9mA:1mT, 9mA and 1mT crystal architectures

In this section, I would like to compare the intermolecular interactions and crystal packing in the co-crystal structure with the corresponding features of its component structures. For that purpose the previously evaluated TAAM structure of **1mT** is used together with the newly established **9mA** one. All the three systems, *i.e.* **9mA:1mT**, **1mT** and **9mA**, form layered crystal architectures, where the molecular planes are based on hydrogen bonded motifs, whereas the interlayer interactions are more dispersive in nature.

A simple method, which provides some qualitative and semi-quantitative knowledge about the interaction types and their percentage contribution, is the Hirshfeld surface analysis (Appendices D, Comment D4, and F, Figure F7-9).[119] The resulting fingerprint plots and Hirshfeld surfaces generated for every component in each structure are presented in Figure 4.33 and Figure 4.34, respectively.

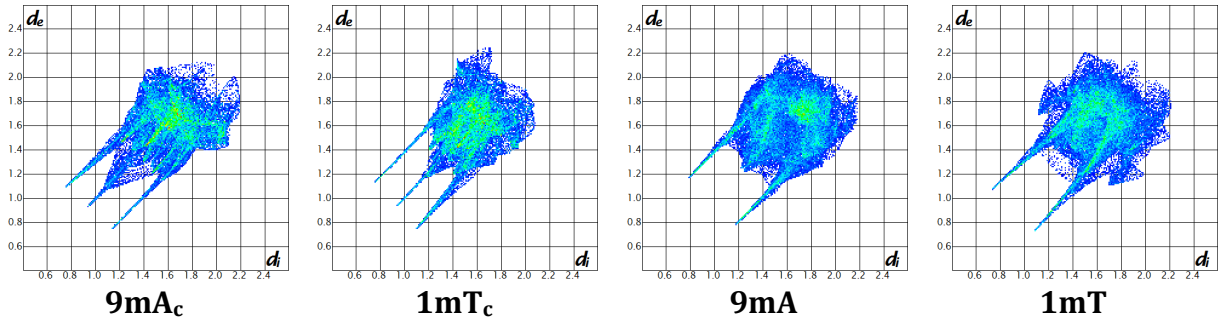


Figure 4.33. Fingerprint plots of **9mA** and **1mT** from the co-crystal structure (**9mA$_c$** and **1mT$_c$**) and the analogue illustrations of the contacts formed by these species in their mono-component crystals.[119b]

The obtained fingerprint plots suggest that crystal packing in **9mA:1mT** is more compact (points are more condensed and shifted towards shorter distances) than in both mono-component structures. What is also well visible, it is the presence of an additional middle spike in the case of the co-crystal structure. This stands for the very short H…H contacts as the ones already mentioned in the intermolecular BCP analysis. This feature is, however, not observed for **9mA** and **1mT** crystal structures. Nevertheless, the percentage contribution of all the H…H contacts in the studied structures is comparable and outnumbers the other interactions (around 40% of all contacts falling on one molecule, Figure 4.34). On the other hand, the lateral spikes reflect the hydrogen bond contact types, *i.e.* both H…N and H…O for **9mA:1mT**, whereas

either symmetrically arranged N…H or O…H contacts in **9mA** and **1mT** crystals, respectively. These contacts are second in terms of percentage contribution, and are crucial for crystal stability (Figure 4.34). The importance of hydrogen bonds is also illustrated in Figure 4.35, where bright red spots indicate mostly N–H…N or N–H…O interactions. These projections show the molecular layers and provide an idea of the molecular surrounding within the molecular slabs. Generally, in each case the hydrogen bond donor and acceptor centres are saturated (see also Appendix F, Figure F7). The H8B…H10 is also well seen as a small bright red spot on the **9mA$_c$** Hirshfeld surface in Figure 4.35*a*.

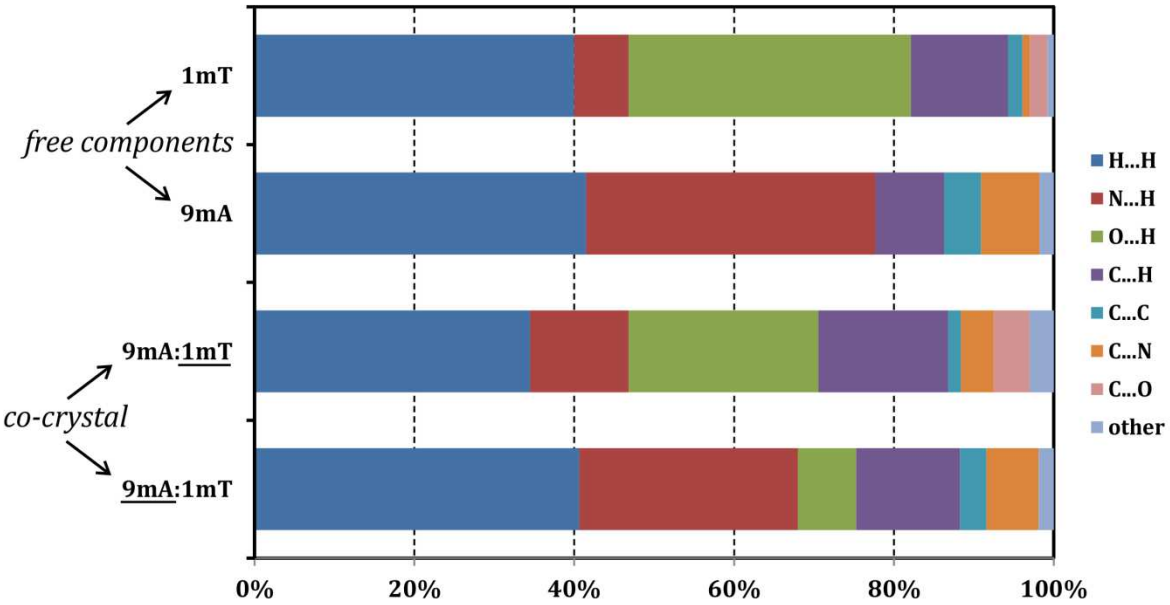


Figure 4.34. Percentage contributions of different interatomic contact types to the Hirshfeld surfaces of **1mT**, **9mA** from the mono-component structures, and **1mT**, **9mA** from the co-crystal structure.

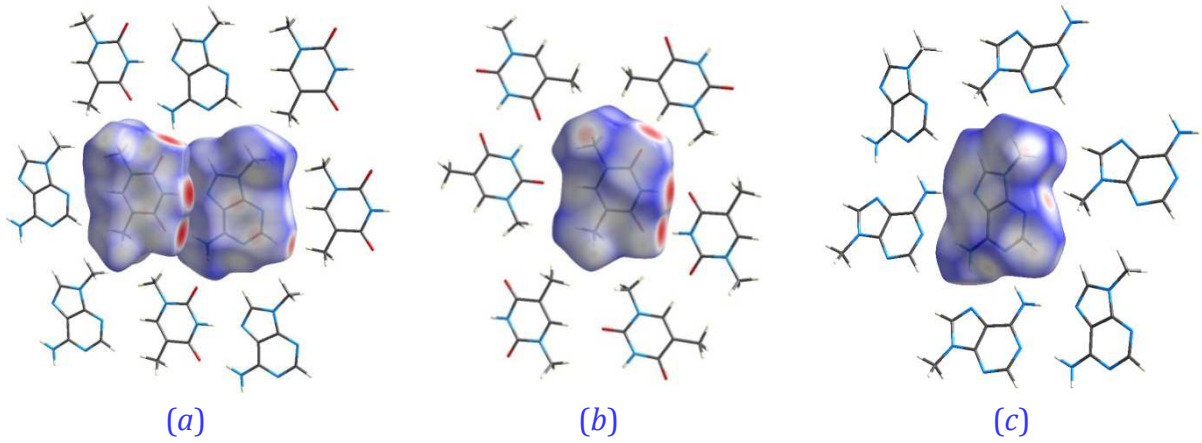


Figure 4.35. d_{norm} mapped on Hirshfeld surfaces of (*a*) **9mA:1mT**, (*b*) **9mA**, (*c*) **1mT**.

According to the cohesive energy calculation results, the **9mA** crystal is characterised by the most advantageous energy per one molecule (Table 4.25). In turn, the co-crystal cohesive energy falling on one species is equal to about the average value of the cohesive energies for **1mT** and **9mA** crystals. Therefore, it seems that **9mA** is at least similarly, or probably more effectively packed in its mono-component structure. In the **9mA** crystal, purine base moiety forms two major in-plane complexes with the adjacent molecules, *i.e.* WHt and HWt (according to the LW notation), and also the third interaction via the 'sugar edge' side with the *N*-methyl group of the neighbouring 9-methyladenine.

Table 4.25. Crystal cohesive energy (E_{coh}) values (per asymmetric unit) and the interlayer interaction energies (E_{intl}) (per molecule) (DFT(B3LYP)/pVTZ, Appendix A)

Crystal	E_{coh} / kJ·mol^{-1}	*Crystal plane*	R_{intl} / Å	E_{intl} / kJ·mol^{-1}
9mA:1mT	−256.0	(010)	3.173	–
1mT	−123.5	(102)	3.301	−51.3
9mA	−133.3	(10$\bar{1}$)	3.145	−67.9

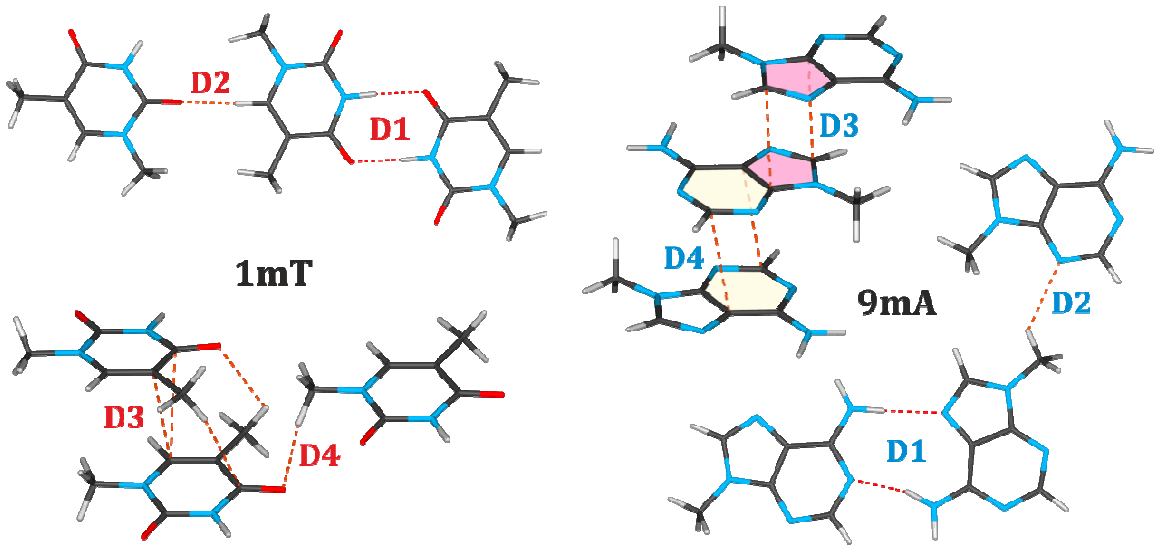


Figure 4.36. Selected dimers extracted from the **9mA:1mT** crystal structure with the illustrated bond paths and BCPs.

Additionally, each **9mA** participates in two different types of π-stacking interactions, *i.e.* via pyrimidine rings (**D4**) and imidazole rings (**D3**) shown in Figure 4.36. As indicated by Table 4.26, the **D4** dimer is a little better stabilised than **D3**. **D4** is analogous to the **AA2** interaction in the co-crystal structure and also described by almost equal interaction energy. It seems that **9mA** in the mono-component crystal interacts stronger in-plane while comparably off-plane in respect to the co-crystal. On

the other hand, **1mT** is presumably more beneficially stabilised in the co-crystal lattice. Nevertheless, it is clearly visible that the main interaction in the co-crystal structure between **1mT** and **9mA** species is the strongest one among all the other analyzed interactions. Thus, this could be the driving force for the co-crystal formation. However, as the stability difference of the studied crystals is not obvious, from a mixture of both substrates one can obtain, with some probability, all kinds of products. The crystallization process depends on the ratio of both compounds and crystallization kinetics.

Table 4.26. Selected dimer interaction energies (E_{int}), notation according to Figure 4.36. Superscript indicates the symmetry transformation of the second dimer component.

Compound	*Dimer code*	E_{int} / kJ·mol^{-1}
1mT	**D1**$^{\#1}$	−56.6
	D2$^{\#2}$	−19.4
	D3$^{\#3}$	−19.4
	D4$^{\#4}$	−7.6
9mA	**D1**$^{\#5}$	−53.0
	D2$^{\#2}$	−10.7
	D3$^{\#4}$	−27.8
	D4$^{\#3}$	−32.1

Symmetry transformations: (#1) $-x+1, -y, -z+2$; (#2) $-x+2, y+0.5, -z+1.5$; (#3) $x, -y+0.5, z-0.5$; (#4) $x-1, y, z$; (#5) $x-0.5, -y+0.5, z-0.5$.

IV.4b.6. Mutual orientation of nucleic acid base derivatives

This section is dedicated to a brief analysis of various nucleic acid base (NAB) pair motifs present either in co-crystal structures or in the available macromolecular structures of nucleic acid chain fragments. First of all, I would like to recall again the LW classification of purine-pyrimidine mutual orientation types, which is very helpful in the description of different NAB dimers of interest. As mentioned in §IV.4b.1, Leontis and Westhof[139g] proposed a base pair classification based on planar edge-to-edge hydrogen bonding interactions between NABs, involving one of three distinct edges: the Watson–Crick (W) edge, the Hoogsteen (H) edge and the Sugar (S) edge (illustrated in Figure 4.29*b* and Figure 4.37). This classification is applicable to base pairs connected by at least two hydrogen bonds. To fully define the relative orientation of the two bases, a line is drawn parallel to and between the two connecting HBs. The interaction of the bases is called trans if the glycosidic bonds of the interacting nucleotides lie on opposite sides of

the line, whereas it is called *cis* when the glycosidic bonds of the nucleotides are on the same side of the line (Figure 4.36). In the case of free nucleic acid base pairs, *i.e.* not connected further to sugar species, I shall classify their mutual orientation on the basis of the location of adequate *N*-groups (literally N–CH_3, the ones being employed in the glycosidic bonds in RNA/DNA chains).

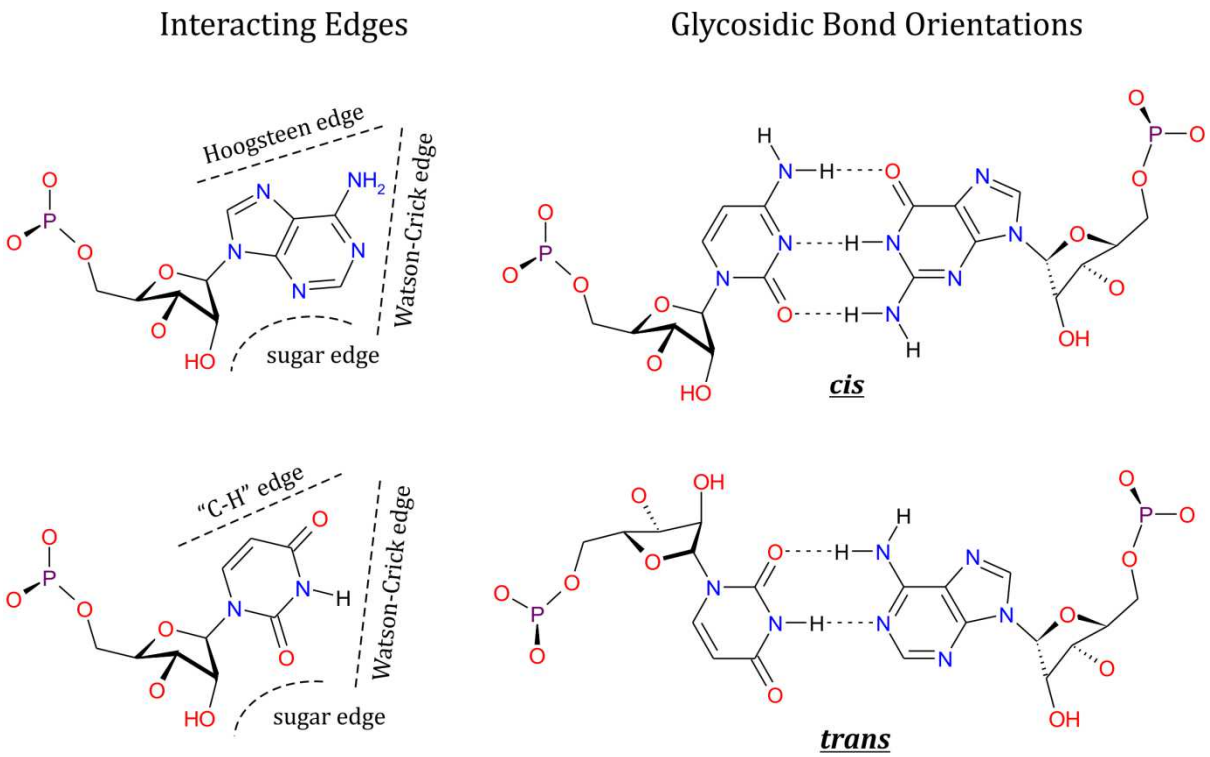


Figure 4.37. Leontis and Westhof nucleic acid base pair classification.[143]

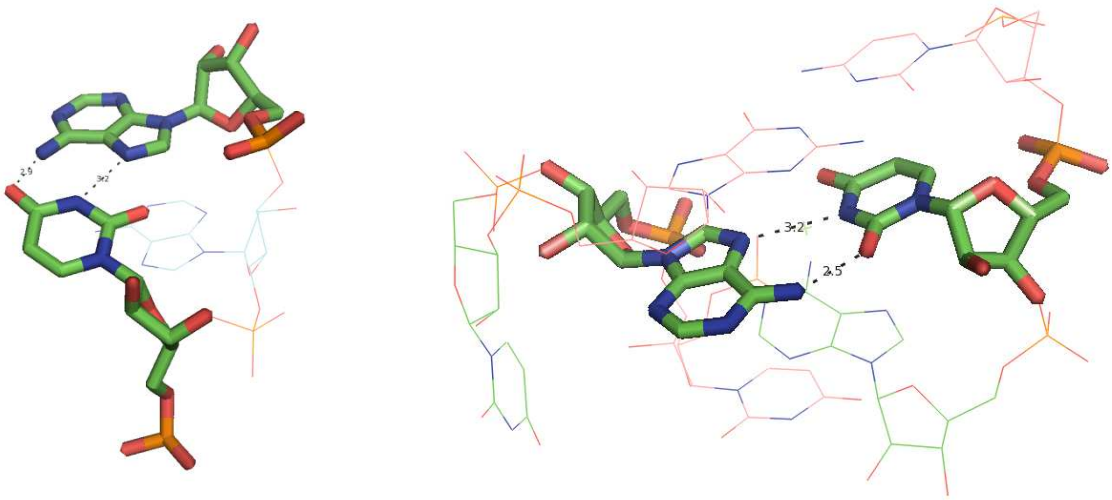
A careful look at the available co-crystal structures of adenine-thymine/uracil derivative co-crystals and the analogue guanine-cytosine structures (Appendix F, Figure F13),[142a, 142c-f] leads to a specific conclusion. It occurs that cytosine and guanine derivatives are very eager to form Watson-Crick pairs, as they do in biological systems. In turn, adenine and uracil derivatives rather do not create standard dimers known from DNA double helices. Instead, they usually bind in the mixed HW manner, which is also the case of the studied **9mA:1mT** co-crystal. Indeed, according to the literature, the G:C WWc motif present in DNA is more stable than the A:T analogue considering hydrogen bonds (three instead of two), but also less free to move in order to satisfy stacking interactions with the adjacent base pairs.[140b] A:T base pairs are less firmly bound, and thus more likely to be separated, which is the key property of the so-called TATA boxes. However, it would be desirable to check the relative stability of WWc and HWc A:T/U

dimers and investigate if, when and why HWc or HWt motifs are formed in biological macromolecular structures.

DNA stiff backbones and well-established double helix structure do not easily allow the formation of more exotic in nature base pair connections. Therefore, in DNA mostly WWc G:C and A:T interactions are observed. On the other hand, it is interesting to check which nucleic acid base motifs are encountered in more flexible RNA chains, and with what probability. It was found that the WWc configuration constitute only 72.3% of the overall number of inter-base interactions in the known RNA structures.[143] The contribution of various trans base pair orientations covers the next 20.3% of the analysed contacts. Furthermore, in the case of A:U dimers both HWc and HWt configurations are observed, whereas the latter one is the second most popular binding type, after WWc, among this group. Following Yang *et al.*,[143] I have chosen the indicated 41 datasets as representative RNA structures. According to the authors, within the overall number of 590 A:U species, 443 form WWc base pairs, 98 are oriented in the HWt manner, while 19 create HWc dimers analogous to the AT motif found in the **9mA:1mT** co-crystal. Similar statistics performed for the G:C complexes are substantially different. Here, among 1583 G:C base pairs, the vast majority (1532) is in the WWc configuration, just 10 G:Cs form the WWt motif, which is the second most numerous group, other are negligible. Such findings support the WWc G:C configuration as being the most stable one and much more favourably formed than the remaining base pair contacts. Therefore, also in small molecular crystal structures of guanine and cytosine derivatives, WW dimers are encountered most often. A:T or A:U type base pairs are more diverse. Regarding the earlier mentioned energetic results (§IV.4b.3), it occurred that WHc seems to be slightly more energetically advantageous than WWc. Therefore, I decided to more thoroughly investigate geometries and energetic features of A:U/T type complexes in order to find explanation for the made observations.

I have chosen a set of 18 A:U HWc complexes mentioned in the literature and 5 exemplary HWt ones. The selected dimers were cut out of the RNA strands and supplemented by methyl groups replacing the glicosidic bond (so A:U dimer is **9mA:1mU** dimer in the energetic calculations). All X–H bonds were fixed at neutron-normalised distances. I have additionally performed some structure optimisation for comparative purposes. The selected dimer details and base pair mutual orientation parameters are available in the Appendix F (Table F14).

The HWc base pair binding type occurs most often at the intersection of two RNA double helices, or at places where mono-strand structural loops are formed. It is also found in the remarkable configuration, where two bases separated by another residue interact with each other, *e.g.* RR0056 (A431 and U429; **cAU15**) (Figure 4.38*a*). Similarly to HWc A:U dimers, an HWt A:U motif is created in the irregular unpaired fragments of RNA double helix strands (Figure 4.38*b*).



(*a*) (*b*)

Figure 4.38. (*a*) HWc-type A:U base pair separated by one non-interacting residue (**cAU15**, RR0056); (*b*) HWt A:U configuration in the irregular RNA double helix region (**tAU5**, UR0019).

According to the optimisation results, the most stable configuration of HWc and HWt A:U complexes is the one where both molecules lie in one molecular plane with very complementarily located HB acceptor and donor atoms ($d_{N3...N15}$ = 2.845 Å and $d_{O4...N18}$ = 2.965 Å for the optimised HWc A:U; corresponding values for the optimised **9mA:1mT**: and $d_{N3...N15}$ = 2.924 Å and $d_{O4...N18}$ = 2.810 Å). Such a favourable orientation of purine and pyrimidine species is present in the studied co-crystal structure and characterised by very stabilising interaction energy. In RNA chains, the most advantageous configuration is less easily achieved due to numerous restraints. However, as shown in Table 4.27, some of the HWc interactions are very significant and exceed the average DNA A:T contact strengths. The 'natural' A:U dimer geometries may be affected by structure resolution, which is given in Appendix F (Table F14). Generally, less stabilised A:U motifs are taken from lower quality structures. It may suggest that these data falsify to some extent the energy information. Anyhow, it was again found that the

directionality of HBs is more important for A:U/T binding stability than interatomic distances.

Table 4.27. Selected base pair motif interaction energies.

Dimer code	E_{int} / kJ·mol^{-1}	*Dimer code*	E_{int} / kJ·mol^{-1}
cAU1	+16.2	**cAU21**	−41.4
cAU2	−65.2	**cAU22**[c]	−22.4
cAU3	−67.6	**tAU1**	−47.9
cAU4	−55.4	**tAU2**	−64.9
cAU5	−58.0	**tAU3**	−60.5
cAU6	−48.6	**tAU4**	−59.2
cAU7	−65.7	**tAU5**	−17.1 / −71.7[a]
cAU8	−51.8		
cAU9	−50.6	**sAA1**	−22.4
cAU10	−61.1	**sAA2**	−26.2
cAU11	−61.1	**sAA3**	−25.3
cAU12	−60.4	**sAT4**	−25.3
cAU13	−48.0	**sAT5**	−22.6
cAU14	−50.1	**sGU6**	−34.8
cAU15[c]	−8.9 / −73.3[a] / −70.8[b]	**sAU8**	−24.4
cAU16	−49.1	**kAU**[d]	−69.6

[a] Isolated dimer optimised. [b] Dimer from optimised trinucleotide fragment optimised. [c] Dimer constructed by two bases separated by the third residue. [d]Co-crystal **AT** dimer with 1-methylthymine molecule converted to 1-methyluracil

Exemplary superpositions of naturally formed A:U dimer geometries and the optimised ones are illustrated in Figure 4.39. As can be seen, the naturally encountered base pairs may be quite significantly affected by the RNA environment, and thus they may be visibly shifted from their optimal orientation. In turn, when a trinucleotide structure is optimised, the perfect planar orientation of adenine and uracil bases is not achieved. However, the resulting optimised trinucleotide structure still provides a very advantageous HWc A:U geometry, characterised by the total interaction energy close to −70 kJ·mol^{-1} (trinucleotide containing the **cAU15** dimer, Table 4.27, Figure 4.39*d*). As indicated by Figure 4.39*c*, for the highly beneficial **cAU12** dimer geometry, both structures, the natural and optimised one, are very much alike. Generally, the analysed HWc A:U motif orientations are more stabilising in the case of higher-resolution RNA data. There is also no clear trend determining the mutual base pair geometry on the basis of RNA structural motif leading to the HW A:U contact formation.

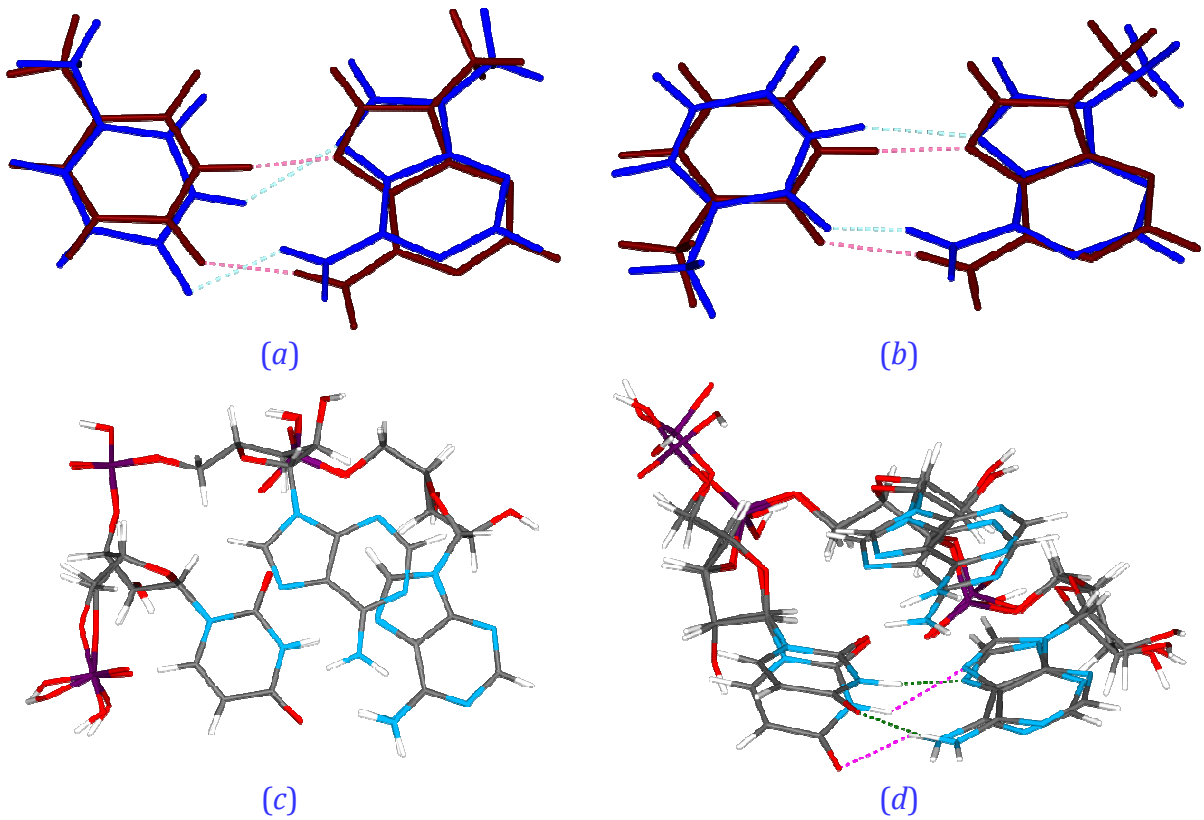


Figure 4.39. (*a*) RR0056 **cAU15**, optimised structure marked in dark red; (*b*) **tAU5**, optimised structure marked in dark red. Superposition of trinucleotide structures: the optimised one and the one taken from RNA (*c*) RR003, **cAU12** base pair, (*d*) RR0056 **cAU15** base pair (green intermolecular contacts indicate the optimised structure).

To fully understand base pair formation, also water molecule impact should be taken into account together with possible competitive interactions between nucleic acid bases and sugar fragments, phosphate species, or with metal ions. On the basis of the above results, I have reasons to conclude that structural restrictions of RNA backbones influence significantly the type of A:U configurations being preferably created. Standard Watson-Crick double helix is formed in the first priority in the case of RNA molecules, and thus the analysed HW motif occurs in structurally irregular regions. However, when HWc or HWt are formed, they are usually similarly stabilised to the WWc dimers. HW dimers might be more sensitive to the base pair mutual orientation changes than WW dimers, which could also support their rare presence in the available RNA structures. This, however, should be further investigated.

Table 4.27 shows also some π-stacking interactions between adenine and thymine species found in nucleic acid chains. Stacking interactions between A:A and A:T fragments are comparable in strength and reach the energy values in-between −20 kJ·mol^{-1} and −30 kJ·mol^{-1} (Table 4.27). Adenine moieties interact more frequently via 6-

membered rings. The strength of contacts between two NABs located one above another is usually increased when oxygen atom-methyl group interactions are allowed. Nevertheless, such contacts are only possible in the small molecule co-crystal structures. In general, to draw further conclusions more energetic and geometrical statistics should be carried out.

IV.4b.7. Experimental charge density distribution vs. the UBDB model

At the end of the **9mA:1mT** co-crystal analysis, I shall devote some attention to the charge density distribution comparison between the UBDB model and the experimental one. **9mA:1mT** constitutes an ideal subject for such investigations, as it covers the main goal of the newly extended UBDB databank, *i.e.* charge density evaluation for DNA/RNA chain fragments. For that purpose, I used the final geometry obtained in the experimental charge density refinement, and applied the UBDB atom type densities in two different ways. The first UBDB model ($M1_{UBDB}$) was based purely on pseudoatom parameters stored in the databank entries, and each molecule, $\mathbf{9mA_c}$ and $\mathbf{1mT_c}$, was separately neutralised. In the second model ($M2_{UBDB}$), solely deformation multipole populations and kappa parameters were taken from UBDB2011, whereas P_v were set at the experimental charge density values. The resulting difference density maps between the experimental charge density model (M_{CD}) and the two UBDB-derived ones are illustrated in Figure 4.40.

As clearly seen, the highest discrepancies between the UBDB-based models and M_{CD} are visible for oxygen and nitrogen atoms, especially in the HB region. The appearance of the N18 and C7 group areas results from different atom local symmetries applied in the compared models. In the case of the experimental charge density distribution refinement, the symmetry of NH_2 and CH_3 groups was set to *m*, while in the case of the databank-based models it was higher, being *mm*2 and 3*m*, respectively. The effect of the too restricted symmetry of the databank atoms is also visible in the Fourier residual density map (Figure 4.40*c*), exhibiting basically the same residual features as these visible in Figure 4.40*b*.

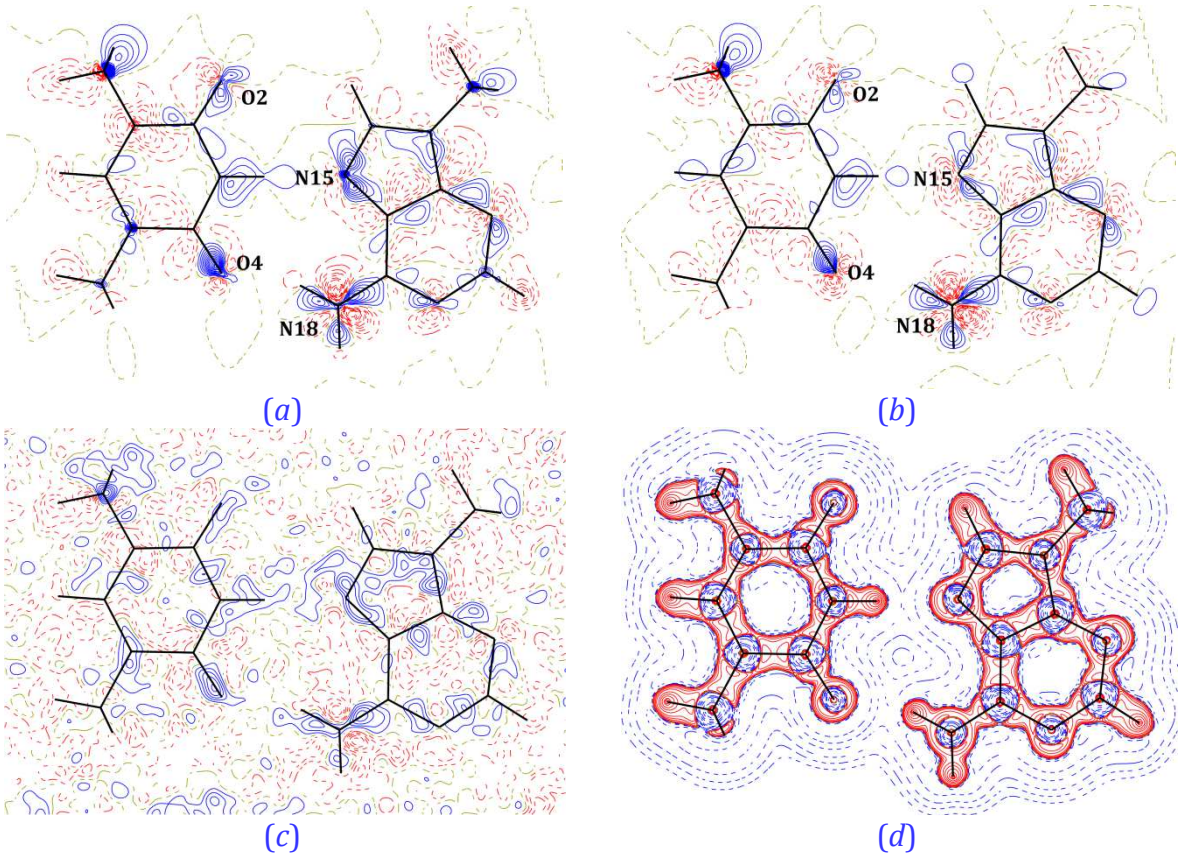


Figure 4.40. (*a*) Difference density map between experimental charge density model (M_{CD}) and $M1_{UBDB}$ (dimer plane, 0.05 e·Å^{-3} contours), $\Delta\varrho_{def}^{min/max}$ = −1.35/+1.47 e·Å^{-3}. (*b*) Difference density map M_{CD}-$M2_{UBDB}$ (dimer plane, 0.05 e·Å^{-3} contours), $\Delta\varrho_{def}^{min/max}$ = −1.54/+1.11 e·Å^{-3}.(*c*) Fourier residual density map (dimer plane, 0.05 e·Å^{-3} contours, $M1_{UBDB}$), $\Delta\varrho_{res}$ = ±0.33 e·Å^{-3}. (*d*) Laplacian of charge density for $M1_{UBDB}$ in the plane of a dimer (logarithmic contours, $M1_{UBDB}$).

Anyhow, the most significant charge density distribution differences, reaching almost ±2 e·Å^{-3} for $M1_{UBDB}$, are observed mainly at atom positions. Changing the UBDB P_v parameters to the ones obtained for the experimental data, decreases the residual peaks located at atoms, however, the bonding residual features remain very much the same. The maximum and minimum residual density values are still over ±0.5 e·Å^{-3} for $M2_{UBDB}$. The fractal dimension plots for both UBDB models indicate the symmetrical values of maximal and minimal residual peaks (Appendix F, Figure F12), showing the sensibility of these models.

Oxygen and nitrogen atoms are in general significantly difficult to be modelled theoretically, especially when averaged (as in the case of UBDB), because of lone electron pairs they possess and their sensitivity to chemical environment. Hydrogen bonding specificity is another aspect which is not well reproduced by the UBDB model,

due to significant atomic polarization not taken into account when modelling charge density distribution of isolated molecules. The charge density Laplacian map clearly shows no polarisation effects accounted by the model (Figure 4.40*d*). In turn, the charge density peaks and holes centred at atom positions may also result from inadequate experimental data resolution, the problem raised in the second chapter while discussing experimental and methodological limitations. In the theoretical model, there is also no problem with the static charge density evaluation, whereas it is more problematic in the case of experimental measurement at higher temperature, *e.g.* 90 K, where atomic thermal motion is significant, which may also cause some inadequacies.

Despite all these shortages of the models, the experimental topological parameters are quite well reconstructed by the UBDB databank (Appendix F, Tables F10 & F11). Intramolecular bonds are satisfactorily described, but obviously the subtle details characteristic for M_{CD} are not reproduced. The middle-strength intermolecular interaction properties are particularly well reflected by the UBDB-derived models. As indicated by Table 4.28, also the HB potential energy density at BCPs is rather in accordance with the experimental values for the H18A…O4 and O2…H18B HBs. It is though noticeably underestimated for the strong N3…N15 contact. However, all in all, the corresponding dimer interaction energies are acceptably predicted on the basis of the databank, *i.e.* E_{int} = −65.9 kJ·mol^{-1} (for **AT**), E_{int} = −12.8 kJ·mol^{-1} (for **iAT**), and E_{int} = −24.2 kJ·mol^{-1} (for **D**). Such findings additionally support the databank applicability to estimate electrostatic interaction energy values.

Table 4.28. Potential energy density (*V*) for selected intermolecular contacts at BCPs derived on the basis of $M1_{UBDB}$ model.

Interaction	*Motif*	$V(\mathbf{r}_{BCP})$ / kJ·mol^{-1}·a$_0^{-3}$
H3…N15	**AT**	−67.3
H18A…O4	**AT**	−64.5
H6…N11$^{\#1}$	**iAT**	−7.2
H8B…H10$^{\#1}$	**iAT**	−18.3
O2…H18B$^{\#2}$	**D**	−48.3

Symmetry transformations: (#1) *x*, *y*, *z*+1; (#2) *x*+1, *y*, *z*;

IV.4b.8. Summary

The **9mA:1mT** study, similarly to the **6m2tU** one, provides a detailed charge density distribution analysis supported by comprehensive energetic investigations. I have thoroughly explored the nature of intermolecular interactions existing in the co-crystal structure in respect to those specific for the **9mA** and **1mT** mono-component crystals. The findings constitute a link between crystal network features, electronic structure and molecular motif interaction energy. Once again it is clearly seen that basing on the Espinosa's approach it is possible to estimate quite reliably HB interaction energy value.

The co-crystal structure is on average more stable than **1mT**, but slightly less advantageous than **9mA**. This is in agreement with the experimental observations. To obtain co-crystals from aqueous solution, the substrates should be mixed accurately in 1:1 molar ratio, as otherwise undesired mono-component crystals are easily grown. All the three systems form molecular layers governed by HB interactions whereas interacting mostly dispersively with each other. The inter-layer contacts are relatively strong, though. Very close H…H distances constitute the very distictive feature of the **9mA:1mT** crystal lattice, in respect to the **9mA** and **1mT** crystals. Another interesting characteristic of **9mA:1mT** crystals is the mutual orientation of the two bases in the major structural dimer. They are connected in the HWc manner instead of creating the most frequently appearing in DNA Watson-Crick base pair motif. Furthermore, it occurred that this particular A:U/T orientation is slightly more stable than the analogous WWc one. Nevertheless, in RNA chains, these being more flexible than DNA molecules, the WHc A:U base-pairing remains rather rarely encountered, which is probably the effect of the rigidity of nucleic acid chain backbones. The A:U HWc configuration is usually observed for the inter-strand regions, or when mono-strand loop motifs are created. In general, the purine-pyrimidine interaction strength is most sensitive to the directionality of the formed hydrogen bonds.

Finally, I have also checked the applicability of the UBDB databank to modelling interacting nucleic acid base pairs. It occurred that the charge density is quite well reproduced by the pseudoatom models, however, high discrepancies are observed at the atom positions and also at more spread regions located in the vicinity of the polarised atoms, especially of oxygen and nitrogen. Nevertheless, generally, the topological features of the intermolecular interactions agree with the experimental ones. The

overall HB interaction energy value is thus satisfactorily estimated on the basis of the UBDB-derived charge density distribution.

"Don't judge each day by the harvest you reap
but by the seeds that you plant"

Robert Louis Stevenson

V. Conclusions

V.1. Final remarks

My PhD project consists of two clearly distinguishable parts. The first part includes modifications and extension of the aspherical pseudoatom databank, the UBDB (University at Buffalo Databank), and the related *LSDB* program. It also contains extensive tests of the databank applicability to the estimation of electrostatic properties of molecules.

The second part of my Thesis is based on the use of the extended UBDB for the following scientific purposes:

(1) reconstruction of the charge density distribution of macromolecules, which may constitute foundation for further estimation of electrostatic properties;

(2) as a source of aspherical atomic scattering factors, which could be used in the structure refinement (so-called Transferable Aspherical Atom Model (TAAM) refinement);

(3) as a starting model in the experimental charge density studies.

The study is not just limited to databank-related scientific issues, but it also addresses subjects regarding the interrelations between geometry, energy, charge density and crystal morphology. Selected biochemical aspects were additionally investigated.

Consequently, my study resulted in a new modified version of the UBDB databank extended with atoms required for modelling RNA and DNA molecules. The enhanced bank, the UBDB2011, now contains over 200 atom types present in the most relevant biochemical systems. Multipolar parameters stored in the databank were evaluated on the basis of over 600 organic molecules. During my work on the databank, I also modified the UBDB-related *LSDB* program, so as to include adequate atom type

definitions, local coordinate systems, and updated X–H bond distances according to Allen *et al*. The *LSDB* now also writes automatically input files suitable for the *MOPRO* suite of programs, which is an alternative for the well-known *XD* package.

Once the UBDB2011 version of the databank was ready, I extensively tested its applicability to estimating the electrostatic properties of chemical systems. The verification was based on the electrostatic interaction energy evaluation for a set of nucleic acid base and amino acid complexes, for which it was possible to run the reference *ab initio* calculations. I showed that the UBDB2011+EPMM (Exact Potential Multipole Method (EPMM)) method satisfactorily reproduces electrostatic interaction energies for a set of nucleic acid base complexes with respect to *ab initio* and/or DFT results (R^2 > 0.9, RMSD = 3.7 kcal·mol^{-1}). Correlations were found to be high, while energy trends were preserved. What is important, the UBDB databank enables taking the asphericity of atoms into consideration, therefore, unlike point-charge models, it addresses the directionality of atom–atom interactions.

However, the UBDB+EPMM approach has a number of limitations. The databank parameters do not reproduce conformational variety, and do not include the crystal field influence, and some other subtle effects. The UBDB data is also constrained by the accuracy of atom type definitions and the charge density parameter computation method (basis set used, DFT method, Fourier truncation error, Hansen-Coppens model limitations, atom type definition adequacy, *etc.*). Consequently, when applied for electrostatic interaction energy evaluation, the UBDB2011+EPMM accuracy is of the order of 5 kcal·mol^{-1}. Furthermore, the databank is not appropriate for describing charge density distribution of non-standard molecules, or systems containing alternated double bond fragments, which I showed using the example of Amphotericin B, and its iodoacetyl derivative.

The prepared and tested UDBD2011 new databank opened up the possibilities for further studies. Thus, as mentioned in the preface, the second part of my Thesis can be divided into three separate sections. The first one is devoted to the electrostatic interaction investigations of influenza neuraminidase complexes with a series of inhibitors, with particular attention paid to the role of water molecules in the active site. The bank was used for charge density distribution reconstruction. In the second section, I focused on a set of 10 uracil derivatives. Here, the databank was applied as a source of atomic scattering factors in the TAAM refinement. Finally, I dedicated my attention to

charge density studies, where I managed to obtain high resolution data, *i.e.* of 6-methyl-2-thiouracil (**6m2tU**), and a co-crystal of a complementary nucleic acid base pair, 1-methylthymine and 9-methyladenine (**9mA:1mT**).

The studies of the neuraminidase-inhibitor complex revealed the databank potential in the analysis of biochemical systems and drug activities. Thanks to the UBDB+EPMM method, it was possible to estimate the partial contributions to the total electrostatic interaction energy coming from single fragments of ligand molecules, or from any selected aminoacid residue of the protein part. This way it is easy to identify the active site region (strongest interacting residues), or crucial atomic groups, limiting or stimulating the protein-inhibitor binding. My contribution was, though, directly related to the water molecule role in the neuraminidase active site. I succeeded to estimate the solvent influence on the complex stability. The results helped to explain some differences in the inhibition activity of selected ligands, which in certain cases were not correlated directly with the strength of electrostatic binding energy of ligand- protein complex. On the whole, such electrostatic interaction energy studies are most justified in the case of systems, where electrostatic interactions outweigh other interaction types. Influenza neuraminidase was a perfect case for such considerations.

Subsequent the analysis of 10 uracil derivatives (*i.e.*, uracil, 1-methyluracil, 1,5-dimethyluracil, 2-thiouracil, 4-thiouracil, 2,4-dithiouracil, 1-methyl-4-thiouracil, 2-thio-5-methylouracil, 2-thio-6-methyluracil, and 5-fluorouracil) revealed that the UBDB2011 databank works also very well as a source of atomic aspherical scattering factors. It occurred that TAAM refinement provides significantly more accurate molecular geometries than the standard IAM (Independent Atom Model) refinement. The energy and geometry studies showed that TAAM-derived molecular geometries are closer to those obtained from the experimental charge density studies and neutron measurement, but also to the periodically optimised structures within the *CRYSTAL* package. The TAAM cohesive energy and motif interaction energy trends are also in perfect agreement with the ones observed for optimised structures, whereas the IAM results differ significantly. My studies provided seven high quality structures, among which the 4-thio-uracil structure had not been earlier deposited in the CSD. All the investigated systems, apart from 2,4-dithiouracil, form layered architectures. The molecules within the layers are held by hydrogen bonds, whereas the molecular layers interact with one another via more dispersive in nature, however, significant contacts. The cohesive energy difference

between the most thermodynamically stable 5-fluorouracil, and the least advantageous crystal lattice of 1-methyl-40-thiouracil, amounts to 40 kJ·mol^{-1}. The analysis of structural motifs revealed the methyl group stimulation of stronger interlayer contacts, whereas the S4 substituent usually affects the stability and quality of the formed crystal. These two factors most substantially influence the aromaticity index values, *HOMA* and *NICS*. *HOMA* and *NICS* do not agree, though. The calculated deformation molecule energy between the isolated molecule and the molecule in a crystal lattice, depends on the strength and number of intermolecular interactions, in which this particular molecule participates, and can reach about 10 kJ·mol^{-1}. I have additionally analyzed the interesting motif created by the fluorine atoms in the 5-fluorouracil (**5fU**) crystal lattice. Four **5fU** molecules create a tetrameric motif with all the four fluorine atoms pointing into each other. On the basis of the databank, I have reconstructed the electron density of molecular fragments contributing to such a pattern. The derived deformation electron density indicated the potential stabilizing character of the F…F contact, of the 'lump to hole' type.

The analysis of charge density distribution of nucleic acid base crystals constituted the last subject undertaken in my PhD Thesis. The project resulted in two high-resolution datasets, *i.e.*, for 6-methyl-2-thiouracil and for the co-crystal structure of 1-methylthymine and 9-methyladenine. The **6m2tU** studies showed that the databank can not only be applied as a starting charge density model, but can also be helpful in handling disorder in a crystal lattice. In my case, the UBDB2011 was used to model the problematic sulphur atom, and indicated the hypothesis of some content of the oxo-thiol **6m2tU** tautomer exiting in the crystal lattice. This could explain systematic errors, which I observed during the charge density data analysis, and peculiar residual peaks visible in the vicinity of the sulphur atom. The study showed also that one has to be very careful with the modelling and data interpretation. On the other hand, regular and well-diffracting **6m2tU** crystals allowed me to find the interrelations between charge density, energy and crystal morphology. It occurred that there is a connection between the crystal architecture features, and the macroscopic shape of the crystal. I have also found relations between crystal thermodynamic characteristics and crystal face stability, and crystal formation. Finally, good agreement was observed between the calculated hydrogen bond energy with that derived from the charge density in the Espinosa's approach.

The **9mA:1mT** co-crystal constituted the last subject of my studies. Interestingly, the **9mA:1mT** base pair does not form the standard Watson-Crick configuration in the crystal lattice, but the Hoogsteen-Watson-Crick base pair motif. It was also found out, that generally adenine derivatives when forming crystals with uracil species, tend to bind in this particular manner. At the same time, this kind of purine-pyrimidine orientation is rarely encountered even in the relatively labile RNA structures. Nonetheless, computational analysis results support slightly better stabilisation of the cHW-type dimers, than the standard cWW ones. The topological investigations gave very similar results to those obtained for the analogous molecules. What is more, the databank parameters reconstructed reasonably well the experimental charge density evakuated for **9mA:1mT**, while the model satisfactorily reproduced the intermolecular interaction region. The **9mA:1mT** is yet another crystal structure of layered architecture characterised by quite strong hydrogen bonds and effective π-stacking interactions. An interesting feature of the crystal network is the presence of close H…H contacts, which is not the case in the mono-component **1mT** and **9mA** crystals.

To recapitulate, my PhD thesis provides extended and improved tools for charge density modelling, which can be successfully applied to electrostatic property study of macromolecules, for deriving more accurate structure geometries, or used in charge density analysis. Furthermore, my studies indicate the interrelations between geometry, structure motifs, lattice energy and crystal morphology. Most of the presented results have already been subjects of scientific articles (to date, I am a co-author of 11 publications). Last manuscripts, regarding these topics, are currently under preparation.

V.2. Future perspectives

The further development in the directions indicated by the presented studies concern the databank application to macromolecular systems, which are of biological importance. The UBDB databank and the idea of TAAM refinement should be popularised, as only the IAM refinement is widely applied, at present. Currently, the databank is being combined with the *PHENIX* package to enable the application of the TAAM procedure also to proteins. Investigations show that there should be some complementary methods provided to supplement the electrostatics derived on the basis of the databank, among which, also, solvent effects should be taken into account. On the other hand, one could consider alternative ways of accurate charge density modelling, free from the limitations of the UBDB databank, and also, the multipolar model.

A comprehensive geometry, energy, and environment analysis of the non-standard nucleic acid base interactions, and also of other important factors, such as interactions with sugar moieties, phosphate groups, water or ion species seems interesting, too. This could give a better insight into the RNA binding properties, and may also lead to some pharmaceutical outcome.

The problem of uracil derivative tautomerism constitutes another topic, which is worth exploring. Some photocrystallographic investigations and solid state NMR could be carried out. This should give a more complete answer to the crystal lattice aspects of **6m2tU**. Similar studies can be performed in the case of 4-thioderivatives, which can be photoinduced and transformed to another tautometic form relatively easily. Low quality of **24dtU** and **4tU** crystals may, in fact, result from some subtle disorder in crystal structures, which can be related to tautomeric processes. This is, of course, highly hypothetical, but definitely worth exploring, as, naturally, many other things not mentioned here, yet…

VI. References

[1] P. P. Ewald, *Acta Cryst.* **1969**, *A25*, 103.
[2] L. V. Azaroff, *Elements of X-ray Crystallography*, McGraw-Hill, New York, **1968**.
[3] K. N. Trueblood, H.-B. Burgi, H. Burzlaff, J. D. Dunitz, C. M. Gramaccioli, H. H. Schulz, U. Schmueli, S. C. Abrahams, *Acta Cryst.* **1996**, 770.
[4] H. Fuess, International Union of Crystallography, Chester, **2006**.
[5] P. Coppens, T. N. G. Row, P. Leung, P. J. Becker, Y. W. Yang, E. D. Stevens, *Acta Cryst.* **1979**, *A35*, 63.
[6] W. H. Bragg, *Proc. Phys. Soc. London* **1921**, *33*, 301.
[7] K. Kurki-Suonio, *Acta Cryst.* **1968**, 379.
[8] R. F. Stewart, *Acta Cryst.* **1975**, *A32*, 565.
[9] F. L. Hirshfeld, *Acta Cryst.* **1971**, *B27*, 769.
[10] N. K. Hansen, P. Coppens, *Acta Cryst.* **1978**, *A34*, 909.
[11] R. F. Stewart, *J. Chem. Phys.* **1968**, *51*, 4569.
[12] a)J. Avery, K. J. Watson, *Acta Cryst.* **1977**, *A33*, 679; b)Z. Su, P. Coppens, *J. Appl. Cryst.* **1990**, *23*, 71.
[13] D. Watkin, *J. Appl. Cryst.* **2008**, *41*, 491.
[14] P. Coppens, L. Leiserowitz, D. Rabinovich, *Acta Cryst.* **1965**, *18*, 1035.
[15] R. H. Blessing, *Acta Cryst.* **1995**, *A51*, 33.
[16] R. B. Helmholdt, A. Vos, **1977**, *A33*, 38.
[17] V. Schomaker, K. N. Trueblood, *Acta Cryst.* **1968**, *B24*, 63.
[18] a) A. Ø. Madsen, *J. Appl. Cryst.* **2006**, *39*, 757; b) P. Munshi, A. Ø. Madsen, M. A. Spackman, S. Larsen, R. Destro, *Acta Cryst.* **2008**, *A64*, 465.
[19] A. A. Hoser, P. M. Dominiak, K. Woźniak, *Acta Cryst.* **2009**, *A65*, 300.
[20] D. Jayatilaka, D. J. Grimwood, *Acta Cryst.* **2001**, *A57*, 76.
[21] A. Fischer, D. Tiana, W. Scherer, K. Batke, G. Eickerling, H. Svendsen, N. Bindzus, B. B. Iversen, *Journal of Physical Chemistry A* **2011**, *115*, 13061.
[22] D. Jayatilaka, B. Dittrich, *Acta Cryst.* **2008**, *A64*, 383.
[23] R. F. W. Bader, *Atoms in Molecules - A Quantum Theory*, Oxford University Press, Oxford, **1990**.
[24] U. Flierler, M. Burzler, D. Leusser, J. Henn, H. Ott, H. Braunschweig, D. Stalke, *Angew. Chem. Int. Ed.* **2008**, *47*, 4321.
[25] P. Coppens, A. Volkov, Y. Abramov, T. Koritsanszky, *Acta Cryst.* **1999**, *A55*, 965.
[26] Y. A. Abramov, *Acta Cryst.* **1997**, *A53*, 264.
[27] E. Espinosa, E. Molins, C. Lecomte, *Chem. Phys. Lett.* **1998**, *285*, 170.
[28] C. P. Brock, J. D. Dunitz, F. L. Hirshfeld, *Acta Cryst.* **1991**, *B47*, 789.
[29] a) S. Domagała, C. Jelsch, *Journal of Applied Crystallography* **2008**, *41*, 1140; b) V. Pichon-Pesme, Lecomte, C., Lachekar , H., *J. Phys. Chem.* **1995**, *99*, 6242; c)V . Pichon-Pesme, C. Jelsch, B. Guillot, C. Lecomte, *Acta Crystallographica Section A* **2004**, *60*, 204; d) S. Domagała, B. Fournier, D. Liebschner, B. Guillot, C. Jelsch, *Acta Crystallographica Section A* **2012**, *68*, 337.
[30] a) B. Dittrich, C. B. Hübschle, P. Luger, M. A. Spackman, *Acta Crystallographica Section D-Biological Crystallography* **2006**, *62*, 1325; b) B. Dittrich, T. Koritsanszky, P. Luger, *Angewandte Chemie-International Edition* **2004**, *43*, 2718.
[31] a) P. M. Dominiak, A. Volkov, X. Li, M. Messerschmidt, P. Coppens, *Journal of Chemical Theory and Computation* **2007**, *3*, 232; b) A. Volkov, X. Li, T. Koritsanszky, P. Coppens, *Journal of Physical Chemistry A* **2004**, *108*, 4283.

[32] a) J. M. Bąk, S. Domagała, C. Hübschle, C. Jelsch, B. Dittrich, P. M. Dominiak, *Acta Crystallographica Section A* **2011**, *67*, 141; b) B. Dittrich, J. J. McKinnon, J. E. Warren, *Acta Crystallographica Section B-Structural Science* **2008**, *64*, 750.
[33] A. Volkov, M. Messerschmidt, P. Coppens, *Acta Crystallographica Section D-Biological Crystallography* **2007**, *63*, 160.
[34] B. Dittrich, T. Koritsanszky, A. Volkov, S. Mebs, P. Luger, *Angewandte Chemie-International Edition* **2007**, *46*, 2935.
[35] a) B. Dittrich, M. Strumpel, M. Schafer, M. A. Spackman, T. Koritsanszky, *Acta Crystallographica Section A* **2006**, *62*, 217; b) C. Jelsch, B. Guillot, A. Lagoutte, C. Lecomte, *Journal of Applied Crystallography* **2005**, *38*, 38.
[36] a) A. Volkov, H. F. King, P. Coppens, L. J. Farrugia, *Acta Crystallographica Section A* **2006**, *62*, 400; b) P. M. Dominiak, A. Volkov, A. P. Dominiak, K. N. Jarzembska, P. Coppens, *Acta Cryst.* **2009**, *D65*, 485; c) X. Li, A. V. Volkov, K. Szalewicz, P. Coppens, *Acta Crystallographica Section D-Biological Crystallography* **2006**, *62*, 639; d) C. Lecomte, B. Guillot, C. Jelsch, A. Podjarny, *International Journal of Quantum Chemistry* **2005**, *101*, 624; e) B. Zarychta, V. Pichon-Pesme, B. Guillot, C. Lecomte, C. Jelsch, *Acta Crystallographica Section A* **2007**, *63*, 108; f) J. J. Holstein, C. B. Hübschle, B. Dittrich, *Crystengcomm* **2012**, *14*, 2520.
[37] a) P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Crystallographica Section D-Biological Crystallography* **2010**, *66*, 213; b) P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, N. Echols, J. J. Headd, L.-W. Hung, S. Jain, G. J. Kapral, R. W. G. Kunstleve, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Methods* **2011**, *55*, 94.
[38] J. M. Bąk, P. M. Dominiak, C. C. Wilson, K. Woźniak, *Acta Cryst.* **2009**, *A65*, 490.
[39] G. S. Chandler, M. A. Spackman, *Acta Crystallographica Section A* **1978**, *34*, 341.
[40] A. Volkov, P. Macchi, L. J. Farrugia, C. Gatti, P. Mallinson, T. Richter, T. Koritsanszky, **2006**.
[41] F. H. Allen, *Acta Cryst.* **2002**, *B58*, 380.
[42] T. Koritsanszky, A. Volkov, P. Coppens, *Acta Crystallographica Section A* **2002**, *58*, 464.
[43] S. K. J. Johnas, B. Dittrich, A. Meents, M. Messerschmidt, E. F. Weckert, *Acta Crystallographica Section D-Biological Crystallography* **2009**, *65*, 284.
[44] C. B. Hübschle, P. Luger, B. Dittrich, *Journal of Applied Crystallography* **2007**, *40*, 623.
[45] K. N. Jarzembska, P. M. Dominiak, *Acta Crystallographica Section A* **2012**, *68*, 139.
[46] C. H. Faerman, S. L. Price, *Journal of the American Chemical Society* **1990**, *112*, 4915.
[47] V. Petricek, M. Dušek, L. Palatinus, Institute of Physics, Praha, Czech Republic, **2006**.
[48] a) A. Volkov, H. F. King, P. Coppens, *Journal of Chemical Theory and Computation* **2006**, *2*, 81; b) A. Volkov, T. Koritsanszky, P. Coppens, *Chemical Physics Letters* **2004**, *391*, 170.
[49] A. D. Buckingham, *Adv. Chem. Phys.* **1967**, *12*, 107.
[50] G. T. Velde, F. M. Bickelhaupt, E. J. Baerends, C. F. Guerra, S. J. A. Van Gisbergen, J. G. Snijders, T. Ziegler, *Journal of Computational Chemistry* **2001**, *22*, 931.
[51] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864.
[52] a) A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098; b) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
[53] W. Kohn, L. Sham, *Phys. Rev. A* **1965**, *140*, 1133.

[54] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, C.02 ed., Gaussian, Inc., Wallingford CT, **2004**.

[55] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. J. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *J.Comput.Chem.* **1993**, *14*, 1347.

[56] J. Rys, M. Dupuis, H. F. King, *Journal of Computational Chemistry* **1983**, *4*, 154.

[57] R. Dovesi, V. R. Saunders, R. Roetti, R. Orlando, C. M. Zicovich-Wilson, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison, I. J. Bush, P. D'Arco, M. Llunell, University of Torino, Torino, **2009**.

[58] S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553.

[59] S. Grimme, *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2011**, *1*, 211.

[60] S. Grimme, S. Ehrlich, L. Goerigk, *Journal of Computational Chemistry* **2011**, *32*, 1456.

[61] B. Civalleri, C. M. Zicovich-Wilson, L. Valenzano, P. Ugliengo, *Crystengcomm* **2008**, *10*, 405.

[62] a) A. A. Hoser, K. N. Jarzembska, Ł. Dobrzycki, M. J. Gutmann, K. Woźniak, *Cryst. Growth Des.* **2012**; b) K. Durka, K. N. Jarzembska, R. Kamiński, S. Luliński, J. Serwatowski, K. Woźniak, *Cryst. Growth Des.* **2012**; c) K. N. Jarzembska, D. Kamiński, A. A. Hoser, M. Malińska, B. Senczyna, K. Woźniak, M. Gagos, *Crystal Growth & Design* **2012**, *12*, 2336.

[63] F. H. Allen, I. J. Bruno, *Acta Cryst.* **2010**, *B66*, 380.

[64] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650.

[65] J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822.

[66] a) C. Roetti, E. Clementi, *Journal of Chemical Physics* **1974**, *60*, 3342; b) C. Roetti, E. Clementi, *Journal of Chemical Physics* **1974**, *60*, 4725.

[67] E. Clementi, D. L. Raimondi, *J. Chem. Phys.* **1963**, *38*, 2686.

[68] P. M. Dominiak, P. Coppens, *Acta Crystallographica Section A* **2006**, *62*, 224.

[69] a) P. Macchi, P. Coppens, *Acta Cryst.* **2001**, *A57*, 656; b) Z. Su, P. Coppens, *Acta Cryst.* **1998**, *A54*, 357; c) Z. Su, P. Coppens, *Acta Cryst.* **1998**, *A54*, 646.

[70] K. N. Jarzembska, P. M. Dominiak, *Acta crystallographica. Section A, Foundations of crystallography* **2012**, *68*, 139.

[71] A. G. Tsarkov, V. G. Tsirelson, *Phys. Status Solidi B* **1991**, *167*, 417.

[72] Ż. Czyżnikowska, R. W. Góra, R. Zaleśny, P. Lipkowski, K. N. Jarzembska, P. M. Dominiak, J. Leszczyński, *Journal of Physical Chemistry B* **2010**, *114*, 9629.

[73] Ż. Czyżnikowska, P. Lipkowski, R. W. Góra, R. Zaleśny, A. C. Cheng, *Journal of Physical Chemistry B* **2009**, *113*, 11511.

[74] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *Journal of Computational Chemistry* **1993**, *14*, 1347.

[75] a) C. Møller, M. S. Plesset, *Phys. Rev.* **1934**, *46*, 618; b) T. H. Dunning, *J. Chem. Phys.* **1989**, *90*, 1007.

[76] Ż. Czyżnikowska, *Journal of Molecular Structure-Theochem* **2009**, *895*, 161.

[77] a) K. Morokuma, *Journal of Chemical Physics* **1971**, *55*, 1236; b) T. Ziegler, A. Rauk, E. J. Baerends, *Theoretica Chimica Acta* **1977**, *43*, 261.

[78] C. B. Hübschle, B. Dittrich, *Journal of Applied Crystallography* **2011**, *44*, 238.

[79] *Modern Charge-Density Analysis*, Springer, **2012**.

[80] P. M. Dominiak, A. Volkov, A. P. Dominiak, K. N. Jarzembska, P. Coppens, *Acta Crystallographica Section D-Biological Crystallography* **2009**, *65*, 485.
[81] P. M. Colman, *Protein Sci.* **1994**, *3*, 1687.
[82] a) P. Bossart-Whitaker, M. Carson, Y. S. Babu, C. D. Smith, W. G. Laver, G. M. Air, *J. Mol. Biol.* **1993**, *232*, 1069; b) W. P. Burmeister, R. W. Ruigrok, S. Cusack, *EMBO J.* **1992**, *11*, 49; c) J. N. Varghese, P. M. Colman, *J.Mol.Biol.* **1991**, *221*, 473.
[83] P. M. Colman, P. A. Hoyne, M. C. Lawrence, *J. Vir.* **1993**, *67*, 2972.
[84] a) H. Liu, A. Warshel, *Biochemistry FIELD Full Journal Title:Biochemistry* **2007**, *46*, 6011; b) K. Chow, H. H. Y. Tong, S. Lum, A. H. L. Chow, *Journal of Pharmaceutical Sciences* **2008**, *97*, 2855; c) C. Sangma, S. Hannongbua, *Current Computer-Aided Drug Design* **2007**, *3*, 113.
[85] a) W. P. Burmeister, B. Henrissat, C. Bosso, S. Cusack, R. W. Ruigrok, *Structure* **1993**, *1*, 19; b) J. B. Finley, V. R. Atigadda, F. Duarte, J. J. Zhao, W. J. Brouillette, G. M. Air, M. J. M. B. Luo, 293, , *J. Mol. Biol.* **1999**, *293*, 1107; c) M. J. Jedrzejas, S. Singh, W. J. Brouillette, W. G. Laver, G. M. Air, M. Luo, *Biochemistry* **1995**, *34*, 3144; d) M. J. Jedrzejas, S. Singh, W. J. Brouillette, G. M. Air, M. Luo, *Proteins* **1995**, *23*, 264; e) B. S. Lommer, S. M. Ali, S. N. Bajpai, W. J. Brouillette, G. M. Air, M. Luo, *Acta Cryst.* **2004**, *D60*, 1017; f) B. J. Smith, P. M. Colman, M. von Itzstein, B. Danylec, J. N. Varghese, 10, . *Protein. Sci.* **2001**, *10*, 689; g) B. J. Smith, J. L. McKimm-Breshkin, M. McDonald, R. T. Fernley, J. N. Varghese, P. M. Colman, *J. Med. Chem.* **2002**, *45*, 2207; h) E. A. Sudbeck, M. J. Jedrzejas, S. Singh, W. J. Brouillette, G. M. Air, W. G. Laver, Y. S. Babu, S. Bantia, P. Chand, N. Chu, J. A. Montgomery, D. A. Walsh, M. Luo, *J. Mol. Biol.* **1997**, *267*, 584; i) N. R. Taylor, A. Cleasby, O. Singh, T. Skarzynski, A. J. Wonacott, P. W. Smith, S. L. Sollis, P. D. Howes, P. C. Cherry, R. Bethell, P. Colman, J. J. M. C. Varghese, *J. Med. Chem.* **1998**, *41*, 798; j) J. N. Varghese, P. M. Colman, A. van Donkelaar, T. J. Blick, A. Sahasrabudhe, J. L. McKimm-Breschkin, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 11808; k) J. N. Varghese, V. C. Epa, P. M. Colman, *Protein Sci.* **1995**, *4*, 1081; l) J. N. Varghese, J. L. McKimm-Breschkin, J. B. Caldwell, A. A. Kortt, P. M. Colman, *Proteins* **1992**, *14*, 327; m) J. N. Varghese, P. W. Smith, S. L. Sollis, T. J. Blick, A. Sahasrabudhe, J. L. McKimm-Breschkin, P. M. Colman, *Structure* **1998**, *6*, 735; n) N. Vasanthan, I. D. Shin, L. Huang, S. Nojima, A. E. Tonelli, *Macromolecules* **1997**, *30*, 3014; o) G. T. Wang, S. Wang, R. Gentles, T. Sowin, C. J. Maring, D. J. Kempf, W. M. Kati, V. Stoll, K. D. Stewart, G. Laver, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 125; p) C. L. White, M. N. Janakiraman, W. G. Laver, C. Philippon, A. Vasella, G. M. Air, M. Luo, *J. Mol. Biol.* **1995**, *245*, 623.
[86] a) M. Fornabaio, P. Cozzini, A. Mozzarelli, D. J. Abraham, G. E. Kellogg, *J. Med. Chem.* **2003**, *46*, 4487; b) T. Steindl, T. Langer, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1849; c) R. P. Verma, C. Hansch, *Bioorg. Med. Chem. Lett.* **2006**, *14*, 982; d) M. von Itzstein, J. C. Dyason, S. W. Oliver, H. F. White, W. Y. Wu, G. B. Kok, M. S. Pegg, *J. Med. Chem.* **1996**, *39*, 388; e) T. Wang, R. C. Wade, *J. Med. Chem.* **2001**, *44*, 961; f) J. Zhang, K.-Q. Yu, W. Zhu, H. Jiang, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3009; g) M. Zheng, K. Yu, H. Liu, X. Luo, K. Chen, W. Zhu, H. Jiang, *J. Comput. Aided. Mol. Des.* **2006**, *20*, 549.
[87] A. Warshel, P. K. Sharma, M. Kato, W. W. Parson, *Biochim. Biophys. Acta, Proteins Proteomics FIELD Full Journal Title:Biochimica et Biophysica Acta, Proteins and Proteomics* **2006**, *1764*, 1647.
[88] K. M. Masukawa, P. A. Kollman, I. D. Kuntz, *J. Med. Chem.* **2003**, *46*, 5628.
[89] a) K. A. Armstrong, B. Tidor, A. C. Cheng, *J. Med. Chem.* **2006**, *49*, 2470; b) P. Bonnet, R. A. Bryce, *Protein Sci.* **2004**, *13*, 946; c) P. Bonnet, R. A. Bryce, *J. Mol. Graph. Model.* **2005**, *24*, 147; d) X. Yi, Z. Guo, F. M. Chu, *Bioorg. Med. Chem.* **2003**, *11*, 1465.
[90] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235.
[91] a) C. J. Maring, V. S. Stoll, C. Zhao, M. Sun, A. C. Krueger, K. D. Stewart, D. L. Madigan, W. M. Kati, Y. Xu, R. J. Carrick, D. A. Montgomery, A. Kempf-Grote, K. C. Marsh, A. Molla, K. R. Steffy, H. L. Sham, W. G. Laver, Y.-g. Gu, D. J. Kempf, W. E. Kohlbrenner, *J. Med. Chem.* **2005**, *48*, 3980; b) N. R. Taylor, M. v. Itzstein, *J. Med. Chem.* **1994**, *37*, 616; c)I. D. Wall, A. R. Leach, D. W. Salt, M. G. Ford, J. W. Essex, *J. Med. Chem.* **1999**, *42*, 5142.

[92] M. S. Pegg, M. Vonitzstein, *Biochemistry and Molecular Biology International* **1994**, *32*, 851.
[93] K. N. Jarzembska, M. Kubsik, R. Kamiński, K. Woźniak, P. M. Dominiak, *Crystal Growth & Design* **2012**, *12*, 2508.
[94] a) S. B. Greenbaum, W. L. Holmes, *Journal of the American Chemical Society* **1954**, *76*, 2899; b) C. Heidelberger, N. K. Chaudhuri, P. Danneberg, D. Mooren, L. Griesbach, R. Duschinsky, R. J. Schnitzer, E. Pleven, J. Scheiner, *Nature* **1957**, *179*, 663; c) M. N. Lipsett, *Biochemical and Biophysical Research Communications* **1965**, *20*, 224; d) W. H. Miller, R. O. Roblin, E. B. Astwood, *Journal of the American Chemical Society* **1945**, *67*, 2201; e) G. S. Parry, F. Strachan, *Acta Crystallographica* **1958**, *11*, 303; f) P. Ajitkumar, J. D. Cherayil, *Microbiological Reviews* **1988**, *52*, 103; g) Bergstro.De, N. J. Leonard, *Biochemistry* **1972**, *11*, 1; h) J. Crooks, *Thyroid and antythyroid drugs*, Elsevier, Amsterdam, **1972**; i) P. Munshi, T. N. G. Row, *Acta Crystallographica Section B-Structural Science* **2006**, *62*, 612; j) R. A. Nugent, S. T. Schlachter, M. J. Murphy, *J. Med. Chem.* **1998**, *41*, 3739; k) J. Sponer, J. Leszczyński, P. Hobza, *Journal of Biomolecular Structure & Dynamics* **1996**, *14*, 117; l) J. Sponer, J. Florian, P. Hobza, J. Leszczyński, *Journal of Biomolecular Structure & Dynamics* **1996**, *13*, 827; m) J. Sponer, J. Leszczyński, P. Hobza, *Journal of Computational Chemistry* **1996**, *17*, 841; n) J. Sponer, J. Leszczyński, P. Hobza, *Journal of Physical Chemistry* **1996**, *100*, 5590; o) J. Sponer, J. Leszczyński, P. Hobza, *Journal of Physical Chemistry* **1996**, *100*, 1965; p) J. Sponer, J. Leszczyński, V. Vetterl, P. Hobza, *Journal of Biomolecular Structure & Dynamics* **1996**, *13*, 695.
[95] a) S. R. Whittleton, K. C. Hunter, S. D. Wetmore, *Journal of Physical Chemistry A* **2004**, *108*, 7709; b) F. Peral, D. Troitino, *Journal of Molecular Structure-Theochem* **2010**, *944*, 1.
[96] a) A. Khvorostov, L. Łapiński, H. Rostkowska, M. J. Nowak, *Photochemistry and Photobiology* **2005**, *81*, 1205; b) A. Khvorostov, L. Łapiński, H. Rostkowska, M. J. Nowak, *Journal of Physical Chemistry A* **2005**, *109*, 7700; c) J. Leszczyński, K. Lammertsma, *Journal of Physical Chemistry* **1991**, *95*, 3128; d) J. Leszczyński, J. Sponer, *Journal of Molecular Structure-Theochem* **1996**, *388*, 237; e) H. G. Mautner, W. D. Kumler, *Journal of the American Chemical Society* **1956**, *78*, 97.
[97] a) E. D. Clercq, J. Balzarini, *Farmaco* **1995**, *50*, 737; b) E. B. Ziff, J. R. Fresco, *Journal of the American Chemical Society* **1968**, *90*, 7338.
[98] a) E. Coats, W. R. Glave, C. Hansch, *Journal of Medicinal Chemistry* **1970**, *13*, 913; b)A. Copik, J. Suwiński, K. Walczak, J. Bronikowska, Z. Czuba, W. C. Krol, *Nucleosides Nucleotides & Nucleic Acids* **2002**, *21*, 377; c) B. M. Jimenez, P. Kranz, C. S. Lee, A. M. Gero, W. J. Osullivan, *Biochemical Pharmacology* **1989**, *38*, 3785; d) T. S. Rao, R. F. Rando, J. H. Huffman, G. R. Revankar, *Nucleosides & Nucleotides* **1995**, *14*, 1997; e) K. A. Anikienko, E. A. Bychikhin, V. K. Kurochkin, V. S. Reznik, V. D. Akamsin, I. V. Galyametdinova, *Doklady. Biochemistry and biophysics* **2001**, *376*, 39; f) J. Balzarini, E. Declercq, P. Herdewijn, M. J. Robins, *Molecular Pharmacology* **1985**, *27*, 578; g) E. I. Besyadetskaya, V. G. Zubenko, L. V. Lozyuk, *Khimiko-Farmatsevticheskii Zhurnal* **1980**, *14*, 42; h) H. Tanaka, H. Takashima, M. Ubasawa, *J. Med. Chem.* **1995**, *38*, 2860; i) W. Tjarks, D. Gabel, *J. Med. Chem.* **1991**, *34*, 315.
[99] R. F. Stewart, *Acta Crystallographica* **1967**, *23*, 1102.
[100] S. W. Hawkinson, *Acta Crystallographica Section B-Structural Science* **1975**, *31*, 2153.
[101] D. Matkovic-Calogovic, E. Besic, K. Sankovic, *Acta Crystallographica Section C-Crystal Structure Communications* **2002**, *58*, O568.
[102] M. Bretner, K. Felczak, J. M. Dzik, B. Gołoś, W. Rode, A. Drabikowska, J. Poznański, K. Krawiec, A. Piasek, D. Shugar, T. Kulikowski, *Nucleosides & Nucleotides* **1997**, *16*, 1295.
[103] a) M. Graindourze, T. Grootaers, J. Smets, T. Zeegershuyskens, G. Maes, *Journal of Molecular Structure* **1990**, *237*, 389; b) T. P. Lewis, H. T. Miles, E. D. Becker, *Journal of Physical Chemistry* **1984**, *88*, 3253; c) J. S. Novros, L. B. Clark, *Journal of Physical Chemistry* **1986**, *90*, 5666; d) R. A. Yadav, P. N. S. Yadav, J. S. Yadav, *Proceedings of the Indian Academy of Sciences-Chemical Sciences* **1988**, *100*, 69.

[104] a) A. T. Hulme, S. L. Price, D. A. Tocher, *Journal of the American Chemical Society* **2005**, *127*, 1116; b) A. Psoda, Z. Kazimierczuk, D. Shugar, *Journal of the American Chemical Society* **1974**, *96*, 6832.
[105] M. N. Lipsett, Peterkof.A, *Proceedings of the National Academy of Sciences of the United States of America* **1966**, *55*, 1169.
[106] a) K. Hoogsteen, *Acta Crystallographica* **1963**, *16*, 28; b) K. Hoogsteen, *Acta Crystallographica* **1963**, *16*, 907.
[107] A. Kvick, T. F. Koetzle, R. Thomas, *Journal of Chemical Physics* **1974**, *61*, 2711.
[108] R. K. McMullan, B. M. Craven, *Acta Crystallographica Section B-Structural Science* **1989**, *45*, 270.
[109] W. T. Klooster, S. Swaminathan, R. Nanni, B. M. Craven, *Acta Crystallographica Section B-Structural Science* **1992**, *48*, 217.
[110] a) G. Laus, A. G. Muller, H. Schottenberger, K. Wurst, M. R. Buchmeiser, K.-H. Ongania, *Monatsh. Chem.* **2006**, *137*, 69; b) R. K. Kumar, D. R. Davis, *Nucleic Acids Research* **1997**, *25*, 1272; c) L. Dolak, W. T. Sokolski, S. Mizsak, D. W. Stroman, O. K. Sebek, *Antimicrobial Agents and Chemotherapy* **1977**, *11*, 569; d) M. Geller, Pohorill.A, A. Jaworski, *Biochimica Et Biophysica Acta* **1973**, *331*, 1; e) A. Jarmula, R. Anulewicz, A. Leś, M. K. Cyrański, L. Adamowicz, M. Bretner, K. Felczak, T. Kulikowski, T. M. Krygowski, W. Rode, *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1998**, *1382*, 277.
[111] C. Delage, A. Hnaifi, M. Goursolle, A. Carpy, *Comptes Rendus De L Academie Des Sciences Serie Ii* **1986**, *302*, 219.
[112] 2010.3-0 ed., Bruker AXS Inc., Madison, WI, USA, **2010**.
[113] 171.33.66 ed., Oxford Diffraction Ltd., **2007**.
[114] G. M. Sheldrick, *Acta Cryst.* **2008**, *A64*, 112.
[115] L. Maschio, B. Civalleri, P. Ugliengo, A. Gavezzotti, *Journal of Physical Chemistry A* **2011**, *115*, 11179.
[116] B. H. Besler, K. M. Merz, P. A. Kollman, *Journal of Computational Chemistry* **1990**, *11*, 431.
[117] a) P. V. Schleyer, C. Maerker, A. Dransfeld, H. J. Jiao, N. Hommes, *Journal of the American Chemical Society* **1996**, *118*, 6317; b) C. Corminboeuf, T. Heine, G. Seifert, P. V. Schleyer, J. Weber, *Physical Chemistry Chemical Physics* **2004**, *6*, 273; c) A. Stanger, *Journal of Organic Chemistry* **2006**, *71*, 883.
[118] E. Shefter, H. G. Mautner, *Journal of the American Chemical Society* **1967**, *89*, 1249.
[119] a) M. A. Spackman, D. Jayatilaka, *CrystEngComm* **2009**, *11*, 19; b) M. A. Spackman, J. J. McKinnon, *Crystengcomm* **2002**, 378.
[120] L. Fallon, *Acta Crystallographica Section B-Structural Science* **1973**, *29*, 2549.
[121] a) V. R. Hathwar, R. G. Gonnade, P. Munshi, M. M. Bhadbhade, T. N. G. Row, *Crystal Growth & Design* **2011**, *11*, 1855; b) T. T. T. Bui, S. Dahaoui, C. Lecomte, G. R. Desiraju, E. Espinosa, *Angewandte Chemie-International Edition* **2009**, *48*, 3838; c) R. J. Baker, P. E. Colavita, D. M. Murphy, J. A. Platts, J. D. Wallis, *Journal of Physical Chemistry A* **2012**, *116*, 1435.
[122] a) L. Pauling, Cornell University Press: Ithaca, NY, **1960**; b) C. F. Matta, N. Castillo, R. J. Boyd, *Journal of Physical Chemistry A* **2005**, *109*, 3669.
[123] a) H. Sternglanz, C. E. Bugg, *Biochimica Et Biophysica Acta* **1975**, *378*, 1; b) H. Sternglanz, G. R. Freeman, C. E. Bugg, *Acta Crystallographica Section B-Structural Science* **1975**, *31*, 1393; c) S. A. Barnett, A. T. Hulme, N. Issa, T. C. Lewis, L. S. Price, D. A. Tocher, S. L. Price, *New Journal of Chemistry* **2008**, *32*, 1761; d) G. Portalone, *Acta Crystallographica Section E-Structure Reports Online* **2008**, *64*, O365.
[124] a) M. K. Cyrański, *Chemical Reviews* **2005**, *105*, 3773; b) T. M. Krygowski, M. K. Cyrański, *Chemical Reviews* **2001**, *101*, 1385; c) T. M. Krygowski, *Journal of Chemical Information and Computer Sciences* **1993**, *33*, 70; d) J. Kruszewski, T. M. Krygowsk. *Bulletin De L Academie Polonaise Des Sciences-Serie Des Sciences Chimiques* **1972**, *20*, 907.
[125] M. K. Cyrański, M. Gilski, M. Jaskólski, T. M. Krygowski, *Journal of Organic Chemistry* **2003**, *68*, 8607.

[126] a) E. B. Astwood, A. Bissell, A. M. Hughes, *Endocrinology* **1945**, *36*, 72; b) R. K. Robins, *Journal of Medicinal Chemistry* **1964**, *7*, 186.
[127] H. Nakamura, J. Y. Noh, K. Itoh, S. Fukata, A. Miyauchi, N. Hamada, *J. Clin. Endocrinol. Metab.* **2007**, *92*, 2157.
[128] P. Fertey, R. Argoud, P. Bordet, J. Reymann, C. Palin, C. Bouchard, R. Bruyere, E. Wenger, C. Lecomte, *Journal of Applied Crystallography* **2007**, *40*, 526.
[129] L. Palatinus, G. Chapuis, *J. Appl. Cryst.* **2007**, *40*, 786.
[130] B. Zarychta, J. Zaleski, J. Kyziol, Z. Daszkiewicz, C. Jelsch, *Acta Crystallographica Section B-Structural Science* **2011**, *67*, 379.
[131] A. Paul, M. Kubicki, C. Jelsch, P. Durand, C. Lecomte, *Acta Crystallographica Section B-Structural Science* **2011**, *67*, 365.
[132] V. V. Zhurov, E. A. Zhurova, A. A. Pinkerton, *J. Appl. Cryst.* **2008**, *41*, 340.
[133] H. O. Sorensen, R. F. Stewart, G. J. McIntyre, S. Larsen, *Acta Cryst.* **2003**, *A59*, 540.
[134] a) R. F. W. Bader, *Atoms in Molecules. A Quantum Theory*, Oxford University Press, New York, NY, **1994**; b) R. F. W. Bader, *J. Phys. Chem. A* **1998**, *102*, 7314.
[135] E. Espinosa, C. Lecomte, E. Molins, *Chem. Phys. Lett.* **1999**, *300*, 745.
[136] M. A. Spackman, J. J. McKinnon, D. Jayatilaka, *Crystengcomm* **2008**, *10*, 377.
[137] a) A. Leś, L. Adamowicz, *Journal of the American Chemical Society* **1990**, *112*, 1504; b) A. Leś, L. Adamowicz, *Journal of Physical Chemistry* **1990**, *94*, 7021.
[138] a) L. H. Thomas, N. Blagden, M. J. Gutmann, A. A. Kallay, A. Parkin, C. C. Seaton, C. C. Wilson, *Cryst. Growth Des.* **2010**, *10*, 2770; b) A. Parkin, C. C. Seaton, N. Blagden, C. C. Wilson, *Cryst. Growth Des.* **2007**, *7*, 531; c) A. Parkin, K. Woźniak, C. C. Wilson, *Cryst. Growth Des.* **2007**, *7*, 1393.
[139] a) J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 964; b) J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737; c) C. R. Cantor, P. L. Wollenzien, J. E. Hearst, *Nucleic Acids Research* **1980**, *8*, 1855; d) T. R. Cech, *Biochemical Society Transactions* **1993**, *21*, 229; e) T. R. Cech, *Nature* **1993**, *365*, 204; f) T. R. Cech, *Gene* **1993**, *135*, 33; g) N. B. Leontis, E. Westhof, *Current Opinion in Structural Biology* **2003**, *13*, 300; h) A. Lescoute, N. B. Leontis, C. Massire, E. Westhof, *Nucleic Acids Research* **2005**, *33*, 2395.
[140] a) C. A. Hunter, X. J. Lu, *Journal of Biomolecular Structure & Dynamics* **1997**, *14*, 747; b) C. A. Hunter, X. J. Lu, *Journal of Molecular Biology* **1997**, *265*, 603; c) X. J. Lu, M. A. ElHassan, C. A. Hunter, *Journal of Molecular Biology* **1997**, *273*, 681.
[141] a) J. A. Cruz, M.-F. Blanchet, M. Boniecki, J. M. Bujnicki, S.-J. Chen, S. Cao, R. Das, F. Ding, N. V. Dokholyan, S. C. Flores, L. Huang, C. A. Lavender, V. Lisi, F. Major, K. Mikolajczak, D. J. Patel, A. Philips, T. Puton, J. Santalucia, F. Sijenyi, T. Hermann, K. Rother, M. Rother, A. Serganov, M. Skorupski, T. Soltysinski, P. Sripakdeevong, I. Tuszynska, K. M. Weeks, C. Waldsich, M. Wildauer, N. B. Leontis, E. Westhof, *Rna-a Publication of the Rna Society* **2012**, *18*, 610; b) C. K. Kennaway, J. E. Taylor, C. F. Song, W. Potrzebowski, W. Nicholson, J. H. White, A. Swiderska, A. Obarska-Kosińska, P. Callow, L. P. Cooper, G. A. Roberts, J.-B. Artero, J. M. Bujnicki, J. Trinick, G. G. Kneale, D. T. F. Dryden, *Genes & Development* **2012**, *26*, 92; c)A. Philips, K. Milanowska, G. Lach, M. Boniecki, K. Rother, J. M. Bujnicki, *Bioinformatics* **2012**, *28*, 198.
[142] a) S. H. Kim, A. Rich, *Science* **1967**, *158*, 1046; b) S. H. Kim, A. Rich, *Journal of Molecular Biology* **1969**, *42*, 87; c) E. J. Obrien, *Acta Crystallographica* **1967**, *23*, 92; d) H. M. Sobell, K. Tomita, A. Rich, *Proceedings of the National Academy of Sciences of the United States of America* **1963**, *49*, 885; e) L. Katz, K. I. Tomita, A. Rich, *Acta Crystallographica* **1966**, *21*, 754; f) A. E. V. Haschemeyer, H. M. Sobell, *Proceedings of the National Academy of Sciences of the United States of America* **1963**, *50*, 872; g) K. Hoogsteen, *Acta Crystallographica* **1959**, *12*, 822.
[143] H. W. Yang, F. Jossinet, N. Leontis, L. Chen, J. Westbrook, H. Berman, E. Westhof, *Nucleic Acids Research* **2003**, *31*, 3450.

VII. APPENDIX A

Supplementary Materials
Introduction Part

Comment A1. Extinction correction.

The kinematical theory proposed for the first time by Laue cannot fully describe all phenomena occurring when a crystal is exposed to X-ray beam. Extinction constitutes one of such effects. The existence of extinction was first introduced by Bragg, James and Bosanquet[1] to describe abnormally large absorption coefficient. They applied the kinematical theory to evaluate structure factor values, and subsequently interpreted the observed peculiarities by the crystal imperfections. About one year later, Darwin came out with his dynamical theory of X-ray diffraction suitable for imperfect crystal.[2] The theory explains the phenomena of extinction by considering the so-called 'crystal mosaicity'. Darwin also introduced the terms of primary and secondary extinction. The primary extinction is the effect resulting from the existence of perfect domains which are not thin enough. Thus, their reflectivity becomes closer to that predicted by the dynamical theory. In such a case, a special correction factor needs to be applied. This effect is particularly visible for low-resolution reflections, for which $|F^{o}| < |F^{c}|$. The less pronounced effect is observed if the crystal 'mosaicity' is more emphasised. The origin of the secondary extinction is that if the planes of two successive domains are nearly parallel, the part of the incident intensity is reflected off by the first domain before it reaches the second one. This effect is then indistinguishable from the absorption and so it is modelled using the artificial absorption coefficient.

Darwin theory is though quite crude and it is not really applicable for most of the crystals. In order to take into account the primary extinction, several theories were

developed in which an appropriate correction, y_{exti}, to the 'kinematical' total reflection intensity is applied:

$$|F_{\text{e}}|^2 = y_{\text{exti}} \cdot |F|^2 = y_{\text{exti}} \cdot k_{\text{S}}^2 \cdot |F|^2 \tag{A.1}$$

where $|F_{\text{e}}|^2$ stands for the extinction-corrected intensity. The first theory, based on the Darwin transfer (differential) equations was proposed by Zachariasen,[3] and implemented as a least-square procedure by Coppens and Hamilton.[4] The theory was later supplemented and extended for isotropic and anisotropic cases by Becker & Coppens.[5] It is used nowadays in the case of a majority of experimental electron density studies. Fortunately, throughout this thesis there was no need to apply such corrections.

Additional references

1 Bragg, W. L.; James, R. W.; Bosanquet, C. H. The intensity of reflection of X-ray by rock-salt. *Phil. Mag.* **1921**, *42*, 1-32

2 Darwin, C. G. The reflection of X-rays from imperfect crystals. *Phil. Mag.* **1922**, *43*, 800-829

3 Zachariasen, W. H. Theory of X-ray diffraction in crystals. New York, **1945**, John Wiley.

4 Coppens, P.; Hamilton, W. C. Anisotropic extinction corrections in the Zachariasen approximation. *Acta Cryst.* **1970**, *A26*, 71-83.

5 Becker, P.; Coppens, P. Extinction within the limit of validity of the Darwin transfer equations. I. General formalisms for primary and secondary extinction and their application to spherical crystals. *Acta Cryst.* **1974**, *A30*, 129-147.

Comment A2. Anomalous scattering.

Another important effect that needs to be considered while performing the structural analysis is called anomalous scattering. It is a result of a resonance between the primary X-ray beam frequency and internal modes of electron vibrations within atoms. It can be introduced analytically into the model by adding anomalous dispersion terms to the atomic scattering factor:

$$f_k^{\text{a}} = f_k + f_k' + \text{i}f_k'' \tag{A.2}$$

where f_k' and $f_k''$are the anomalous dispersion corrections. Substituting this into the structure factor expression gives the following result:

$$F = \sum_k (f_k + f_k' + \mathrm{i}f_k'')\, T_k \exp(2\pi\mathrm{i}\mathbf{H}\cdot\mathbf{r}_k) =$$

$$\begin{aligned} &= \sum_k (f_k + f_k')\, T_k \exp(2\pi\mathrm{i}\mathbf{H}\cdot\mathbf{r}_k) + \mathrm{i}\sum_k f_k''\, T_k \exp(2\pi\mathrm{i}\mathbf{H}\cdot\mathbf{r}_k) \\ &= \sum_k (f_k + f_k')\, T_k \cos(2\pi\mathbf{H}\cdot\mathbf{r}_k) + i\sum_k (f_k + f_k')\, T_k \sin(2\pi\mathbf{H}\cdot\mathbf{r}_k) \\ &+ \mathrm{i}\sum_k f_k''\, T_k \cos(2\pi\mathbf{H}\cdot\mathbf{r}_k) - \sum_k f_k''\, T_k \sin(2\pi\mathbf{H}\cdot\mathbf{r}_k) \end{aligned} \tag{A.3}$$

If now:

$$\begin{aligned} A &= \sum_k (f_k + f_k')\, T_k \cos(2\pi\mathbf{H}\cdot\mathbf{r}_k) \quad B = \sum_k (f_k + f_k')\, T_k \sin(2\pi\mathbf{H}\cdot\mathbf{r}_k) \\ a &= \sum_k f_k''\, T_k \cos(2\pi\mathbf{H}\cdot\mathbf{r}_k) \quad b = \sum_k f_k''\, T_k \sin(2\pi\mathbf{H}\cdot\mathbf{r}_k) \end{aligned} \tag{A.4}$$

the structure factor can be expressed by a simple form:

$$F = (A - b) + \mathrm{i}(B + a) \tag{A.5}$$

It is now clearly visible how the various anomalous corrections affect the structure factor magnitude. All these corrections depend on the X-ray radiation frequency and atom types present in the crystal structure. They are usually larger for heavier atoms (*e.g.* Mg or K), but also lighter atoms such as oxygen possess quite significant anomalous signal. All such corrections are tabulated in the International Tables for Crystallography,[4] and there are also some theoretical models of calculating the anomalous scattering coefficients. In addition, due to the presence of the anomalous signal the Friedel law is not fulfilled:

$$|F(\mathbf{H})|^2 \neq |F(-\mathbf{H})|^2 \tag{A.6}$$

This makes sense only for non-centrosymmetric structures. This observation is a fundament for different methods of determining the so-called absolute structure. Different approaches were successfully developed by Rogers,[1] Flack,[2] van Hooft,[3] Parsons[4] and others. Throughout this work, there were no non-centrosymmetric crystal structures studied there was no need for a absolute structure determination. Nevertheless, all considered models (if not theoretically derived) were corrected for anomalous scattering. Similarly, all Fourier synthesis maps were computed with the anomalous dispersion 'removed' according to the method proposed by Ibers and Hamilton.[5]

Additional references

1 Rogers, D. On the application of Hamilton's ratio test to the assignment of absolute configuration and an alternative test. *Acta Cryst.* **1981**, *A37*, 734-741.

2 Flack, H. D.; Bernardinelli, G. Reporting and evaluating absolute-structure and absolute-configuration determinations. *J. Appl. Cryst.* **2000**, *33*, 1143-1148.

3 Hooft, R. W. W.; Straver, L. H.; Spek, A. L. Determination of absolute structure using Bayesian statistics on Bijvoet differences. *J. Appl. Cryst.* **2008**, *41*, 96-103.

4 Parsons, S.; Flack, H. D. Precise absolute-structure determination in light-atom crystals. *Acta Cryst.* **2004**, *A60*, s61.

5 Ibers, J. A.; Hamilton, W. C. Dispersion corrections and crystal structure refinements. *Acta Cryst.* **1964**, *17*, 781-782.

Comment A3. Computation of cohesive, interlayer and dimer interaction energies.

(i) Cohesive energy. For the purpose of the presented research, energy computations were performed with the *CRYSTAL* program package at the DFT(B3LYP) level of theory. Either 6-31G**, or pVTZ (specified in the text) molecular all-electron basis set was used. Both Grimme dispersion correction and correction for basis set superposition error (BSSE) were applied. Ghost atoms were selected up to 5 Å distance from the considered molecule in the crystal lattice, and were utilised for the BSSE estimation. The evaluation of Coulomb and exchange series was controlled by five thresholds, set arbitrarily to the values of 10^{-7}, 10^{-7}, 10^{-7}, 10^{-7}, 10^{-25}. Shrinking factor was equal to 8, which in most of the cases referred to 170 **k**-points in the irreducible Brillouin zone (exceptions: **1mU** (95 **k**-points), **2tU** and **5fU** (260 **k**-points)) and assures the full convergence of the total energy. The cohesive energy (E_{coh}) was calculated following the procedure described in the literature:

$$E_{\mathrm{coh}} = \frac{1}{Z} E_{\mathrm{bulk}} - E_{\mathrm{mol}} \tag{A.7}$$

where E_{bulk} is the total energy of a system (calculated per unit cell) and E_{mol} is the energy of an isolated molecule extracted from the bulk (with the same geometry as in the crystal phase). Z stands for the number of molecules in the unit cell.

(ii) Interlayer interaction energy. All of the calculation parameters were set identically for the purpose of crystal interlayer interaction computations. The only difference was introduced to ghost atom definition. Here, an additional upper and lower molecular layer were used as ghost function sets in order to obtain BSSE. The interlayer interaction energy (E_{intl}) calculation formula is analogous to that for cohesive energy:

$$E_{\text{intl}} = \frac{1}{n} E_{\text{bulk}} - E_{\text{slab}} \tag{A.8}$$

where E_{slab} is the energy of a molecular slab extracted from the bulk, while n indicates the slab number per unit cell. Therefore the resulting stabilisation energy falls on a part of a slab belonging to the unit cell. *CRYSTAL* automatically assigns the slab group symmetry and cuts out the repeatable fragment.

(iii) Dimer interaction energy. The analysed structural motif stabilisation energies (*e.g.*, dimer, tetramer interaction energy) were also calculated in the *CRYSTAL* package, taking into account both BSSE and dispersive corrections. Stabilisation energy values were evaluated by means of the supermolecular approach.

(iv) Geometry optimisation. All the used optimised geometries were obtained with the aid of *CRYSTAL* package. In the case of crystal structure 3D optimisations, atom positions were varied, while cell parameters kept fixed during the optimisation procedure, if not stated otherwise.

Comment A4. Surface free energy estimation.

An estimation of the surface free energy, E_{surf}, was made on the basis of some approaches regarding metal oxide and metal surfaces. There are two basic surface free energy definitions, *i.e.* for a wide gap insulator, and for metals and small band gap semiconductors. In the case of insulators, one may apply the following formula:

$$E_{\text{surf}} = \frac{E_n - n \cdot E_{\text{bulk}}}{2A} \tag{A.9}$$

where E_n is the energy of the n-layer slab, n denotes the number of atomic layers of the type in bulk, E_{bulk} is the bulk energy, and A stands for the slab area (surface). This expression is applicable, when the convergence of E_{bulk} and E_n with respect to **k**-space sampling is satisfactory enough. This means that E_s converges rapidly and the surface energy is well defined. In metals, or small band gap semiconductors, a more numerically stable definition of the surface energy is used. In these cases the following formula should be applied:

$$E_{\text{surf}} = \frac{E_n - n \cdot (E_n - E_{n-1})}{2A} \tag{A.10}$$

Here, E_{bulk} has been replaced by $E_n - E_{n-1}$ energy difference. n should be chosen so as to fulfil the convergence criterion. The difference between E_n and E_{n-1} should reach the bulk energy value. In this case the bulk energy has to be effectively computed under

exactly the same numerical conditions as the surface layers, leading to cancellation of any systematic error.

I used this approach in the case of a molecular crystal, *i.e.,* for **6m2tU**. Molecular surfaces are not as clearly defined (and are much more developed) as the atomic ones, characteristic for the above-mentioned systems. However, similar approach may also shed some light on the relative surface free energies of different slabs extracted from the bulk. This, in turn, may contribute to the understanding of crystal facet stability and formation. It is assumed that the higher the surface energy value, the less stable is the surface itself.

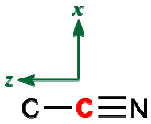
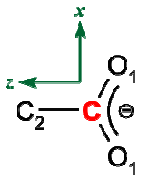
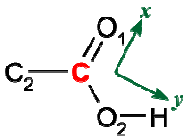
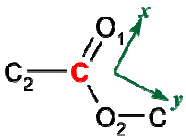
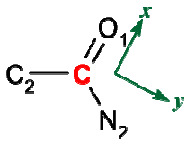
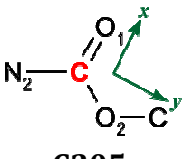
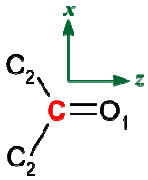
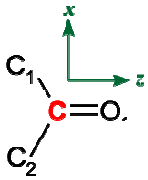
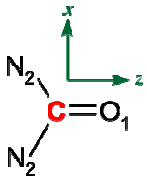
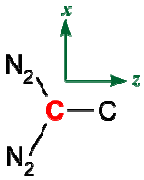
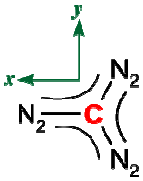
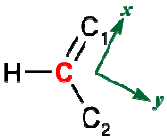
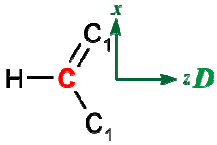
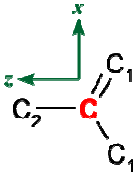
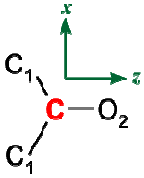
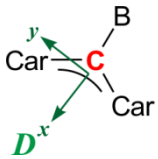
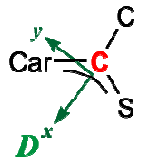
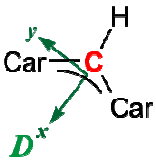
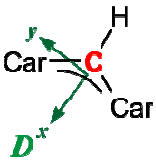
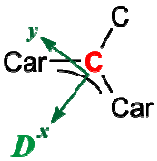
Comment A5. Theoretical charge density evaluation.

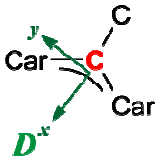
In order to obtain theoretical structure factors, firstly single-point periodic calculations need to be performed using the *CRYSTAL* code. Upon the convergence on energy, the periodic wave function is obtained and subsequently used to generate the symmetry-unique set of static theoretical structure factors up to the resolution of 1.2 $Å^{-3}$. A standard charge density refinement is then carried out with respect to all reflections. During the refinement, all multipole and contraction-expansion parameters assigned to all atoms, are iteratively varied. According to standard procedures, during the refinement atomic coordinates are kept fixed, thermal motion is set to zero for all atoms, unit weights are applied, whereas scale factor is not refined.

VIII. APPENDIX B

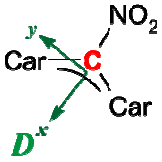
Supplementary Materials
UBDB Part

Table B1. Atom types in the UBDB2011 databank. The *D* symbol represents dummy atom required for coordinate system definition. The first letter in the atom type code goes for a chemical element symbol while the first numeral indicates its valency (*i.e.*, number of attached neighbours). To distinguish atoms joining fused rings together the first numeral is set to 5.

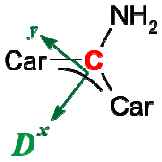
CARBON				
				
C201	**C301**	**C302**	**C303**	**C304**
				
C305	**C306**	**C307**	**C308**	**C309**
				
C310	**C311-T**	**C312-T**	**C313-T**	**C314**
				
C328-6M	**C329-5M-T3**	**C330-6M**	**C331-5M**	**C332-6M**



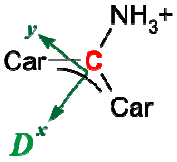
C333-5M



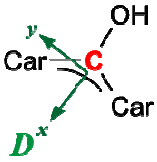
C334-6M



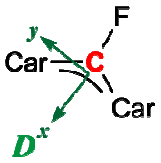
C338-6M



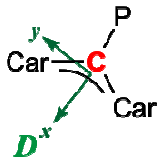
C340-6M



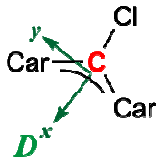
C341-6M



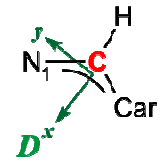
C344-6M-T2



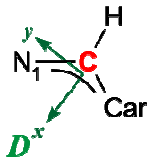
C345-6M



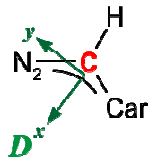
C346-6M-T



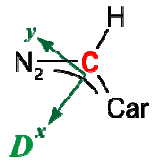
C350-6M



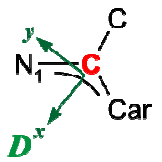
C351-5M



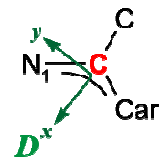
C352-6M



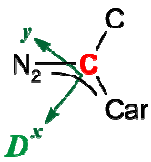
C353-5M



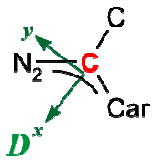
C354-6M



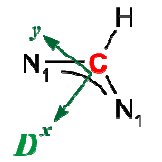
C355-5M



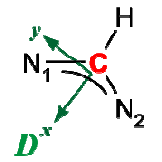
C356-6M



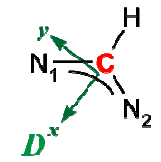
C357-5M



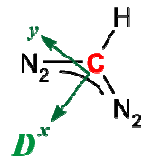
C358-6M



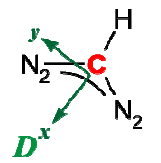
C360-6M



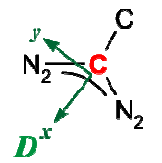
C361-5M



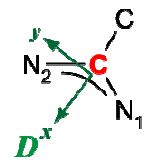
C362-6M



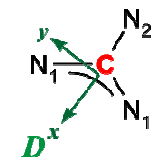
C363-5M



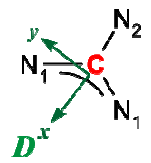
C366-5M-T3



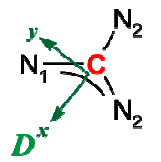
C367-6M



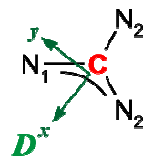
C368-6M



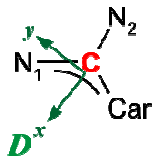
C369-5M



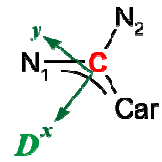
C370-6M



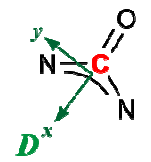
C371-5M



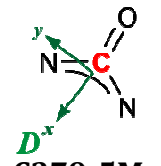
C373-6M



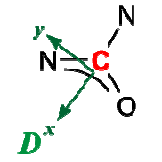
C374-6M



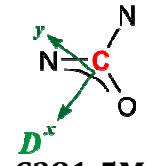
C378-6M



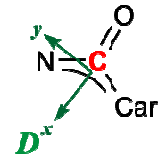
C379-5M



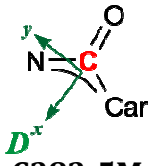
C380-6M



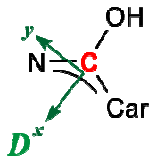
C381-5M

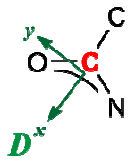


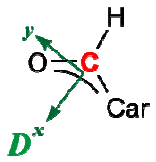
C382-6M

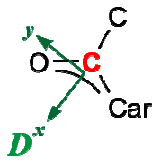


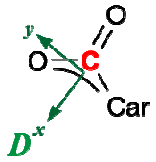
C383-5M

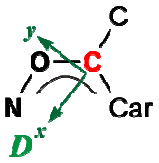

C385-5M

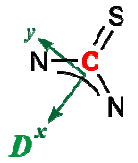

C386-5M

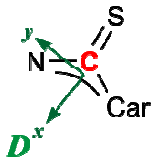

C388-5M

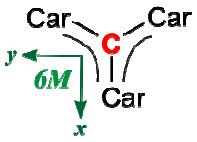

C389-5M

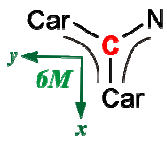

C390-6M

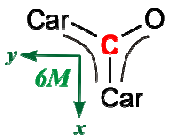

C391-5M

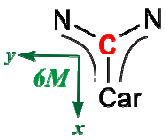

C392-6M-T3

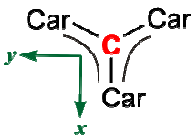

C393-6M-T2

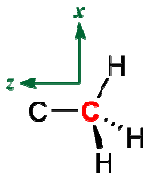

C501-11M

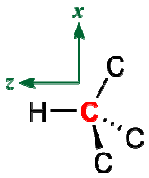

C503-11M

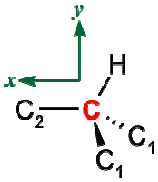

C504-11M

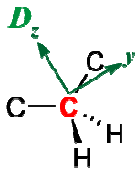

C505-11M

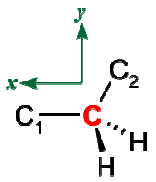

C507-12M

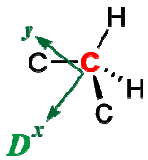

C401

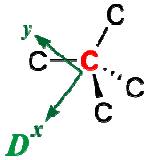

C402

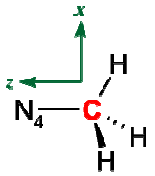

C403*

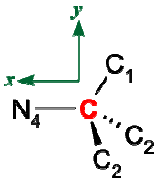

C404

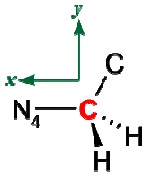

C405

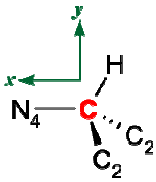

C406-5M

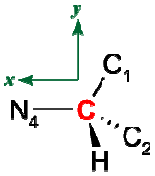

C407-5M-T2

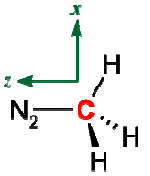

C410

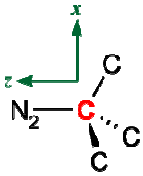

C411

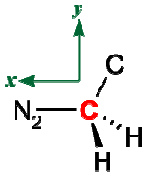

C412

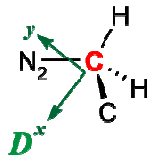

C413*

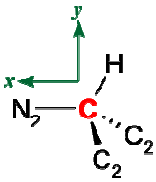

C414

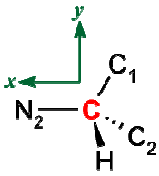

C416

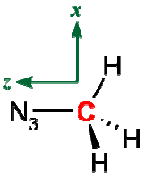

C417

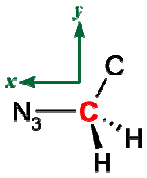

C418

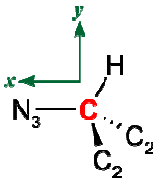

C419-5M

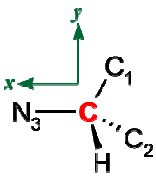

C420*

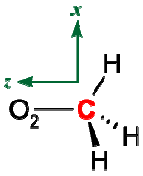

C421


C422

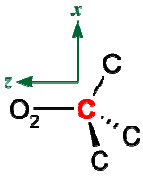

C423


C424*

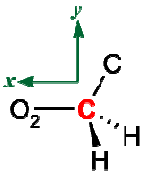

C425



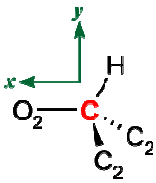
C427



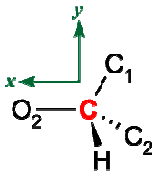
C428



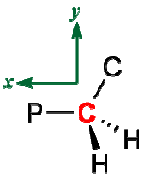
C429



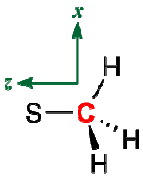
C430*



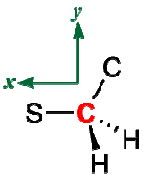
C431



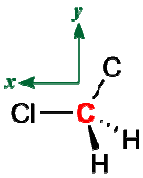
C432



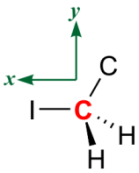
C433



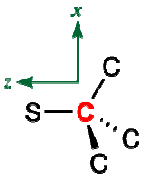
C434



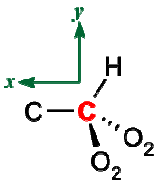
C437



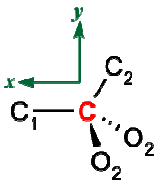
C438



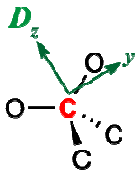
C439



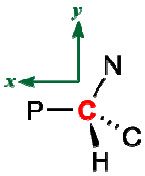
C440



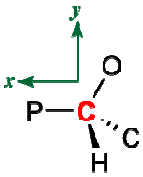
C443



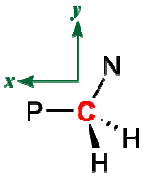
C444-T



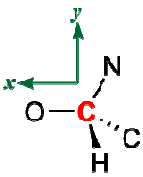
C446



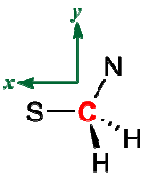
C447



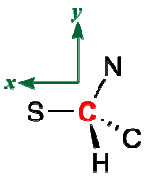
C448



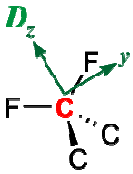
C450



C452

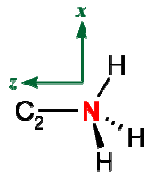


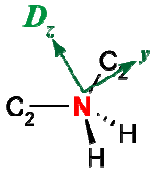
C453

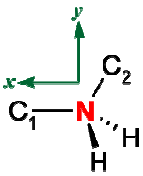


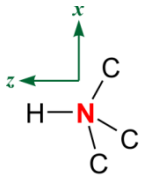
C455

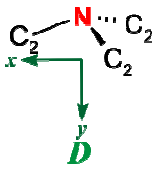
NITROGEN

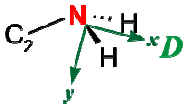

N401

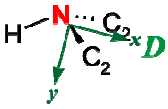

N402

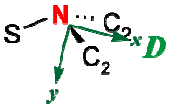

N403

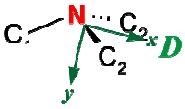

N404

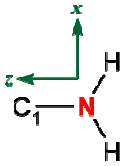

N301

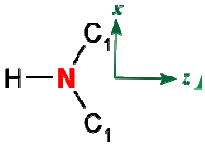

N302

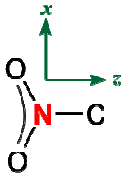

N303

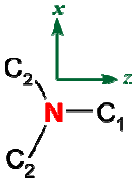

N304

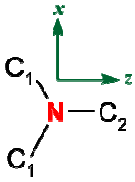

N306

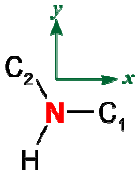

N310

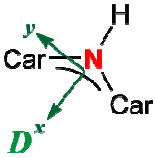

N311

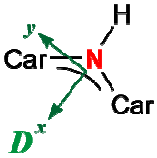

N312

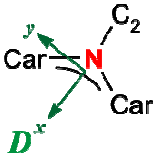

N313

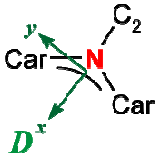

N314

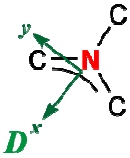

N315

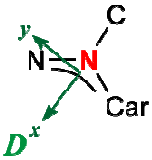

N320-6M

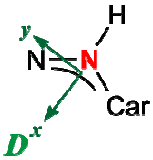

N321-5M

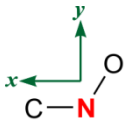

N322-6M

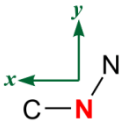

N323-5M

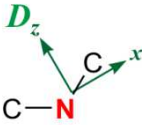

N325-6M

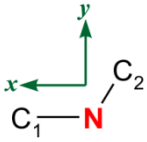

N326-6M-T

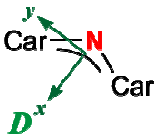

N328-5M

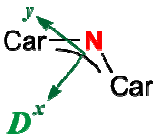

N202

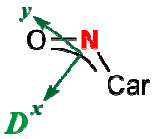

N203

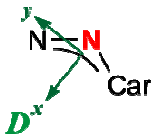

N204

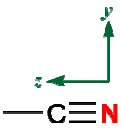

N205


N210-6M

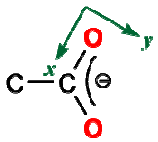

N211-5M

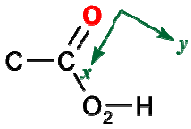

N212-5M

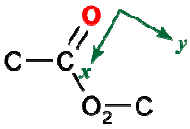

N213-5M

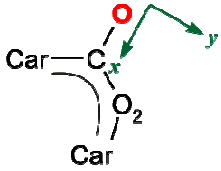

N101

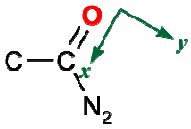
OXYGEN

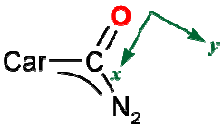

O101

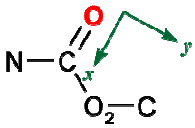

O102

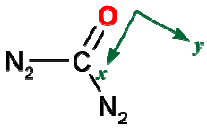

O103

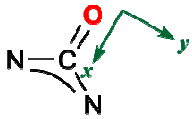

O104

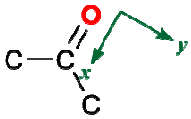

O105

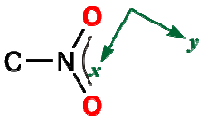

O106

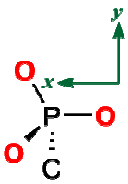

O107

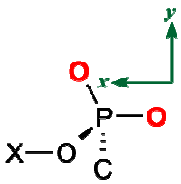

O108

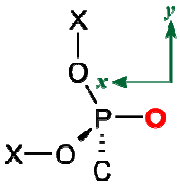

O109

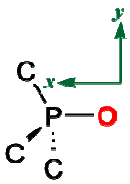

O110

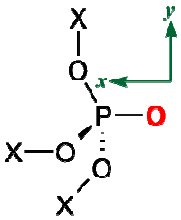

O113

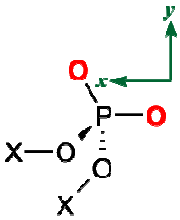

O114

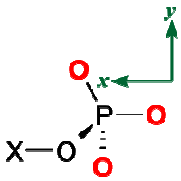

O115

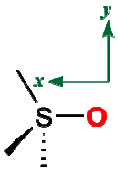

O116

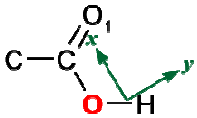

O117

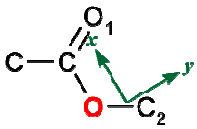

O119

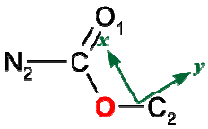

O120

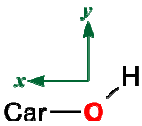

O121

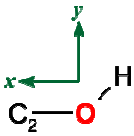

O122

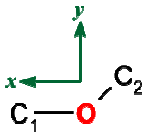

O201

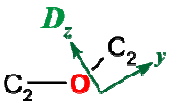

O202

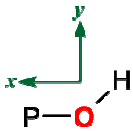

O203

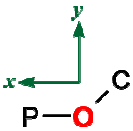

O204

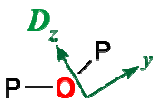

O205

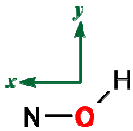

O206

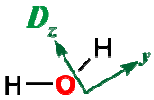

O207

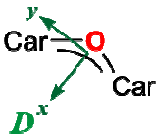

O208

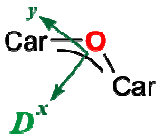

O209

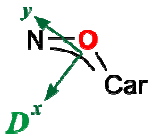

O210

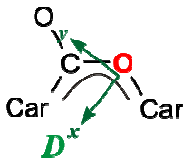

O211


O001

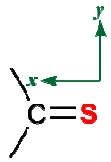

O230-6M

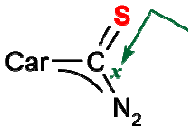

O231-5M

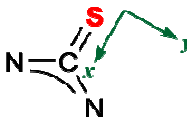

O232-5M

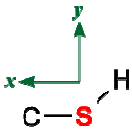

O233-6M

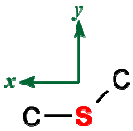
SULPHUR

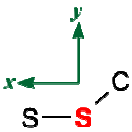

S101-T3

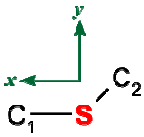

S103-T2

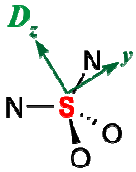

S104-T2


S201

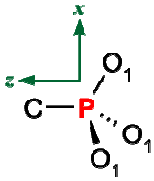

S202

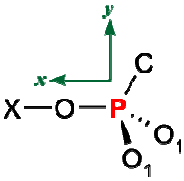

S203

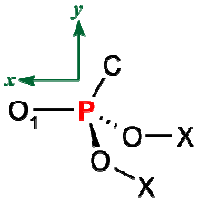

S204

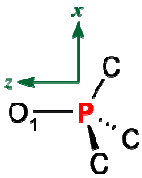

S404-T3

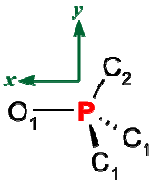
PHOSPHORUS

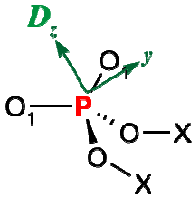

P401

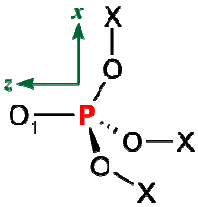

P402

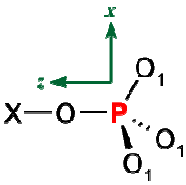

P403

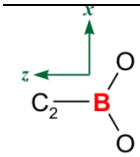
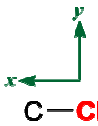
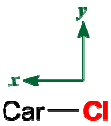
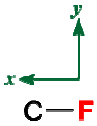
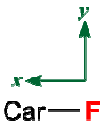
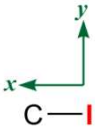

P404-T2


P405


P407


P408


P409

BORON	CHLORINE		FLUORINE		IODINE
					
B01	Cl01-T	Cl02-T	F01	F02	I01-T

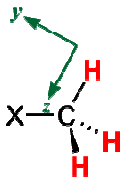
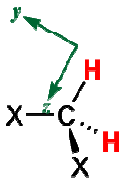
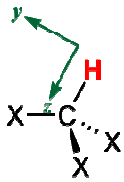
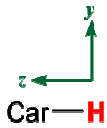
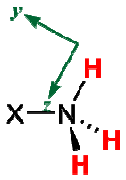
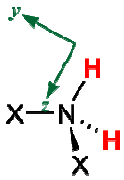
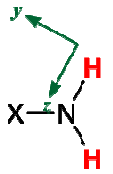
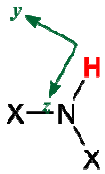
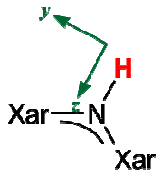
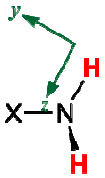
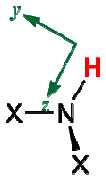
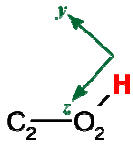
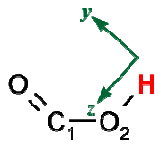
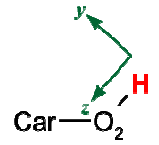
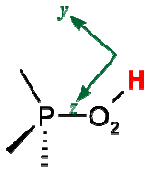
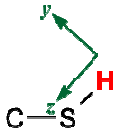
HYDROGEN				
				
H100	**H101**	**H102**	**H103**	**H104**
				
H105	**H106**	**H107**	**H108**	**H109**
				
H110	**H111**	**H112**	**H113**	**H114**
				
H115	**H116**	**H117**	**H001**	

Table B2. Labelling of the nearest and next-nearest neighbour atom types.[a]

C	any Csp^2 or Csp^3 carbon
C_1	Csp^2 aromatic and non-aromatic
C_2	Csp^3
C_{ar}	Csp^2 in planar ring
N_1	$Nsp^2(2)$
N_2	$Nsp^2(3)$
N_3	$Nsp^3(3)$
N_4	$Nsp^3(4)$
O_1	$Osp^2(1)$
O_2	$Osp^3(2)$
X	any non-hydrogen atom; in the case of CPO_3 and CPO_4 groups carbon or hydrogen atoms

* – there can be both two C_2 atoms or two C_1 atoms
nM – indicates number of neighbours
T – simplified atom definition

[a] numbers in parentheses represent number of bonds to the atoms

Table B3. The UBDB2011 databank entries.

```
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
! PseudoAtom Multipolar DataBank
!
! Created by Katarzyna N. Jarzembska, Paulina M. Dominiak, Xue Li & Anatoliy Volkov
!
! For atom types with TEMP word in the COMMENT line there was not enough atoms
! to average (fewer than 5) or the entry needs to be better described.
!
! PROBLEMs: CHIRAL atoms, TEMP atoms
! when using S or P atom types - don't forget to change SCAT in .mas file
!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
!  H Y D R O G E N
!
!
ATOM 1 COMMENT H100 H-nc3     DATE Fri Apr 22 15:44:52 2011
 PRNT 7 NEIG 4 HYDR 1 SYMM cyl
 PVAL    0.806   KAPPA    1.167   KPRIM    1.248           SIGPV    0.017
 PLMS 1  0   0.157   PLMS 2  0   0.069
!
ATOM 1 COMMENT H101 x-cH3     DATE Fri Apr 22 15:44:52 2011
 PRNT 6 NEIG 4 HYDR 3 SYMM cyl
 PVAL    1.089   KAPPA    1.094   KPRIM    1.156           SIGPV    0.039
 PLMS 1  0   0.185   PLMS 2  0   0.085
!
ATOM 1 COMMENT H102 x-cH2-x     DATE Fri Apr 22 15:45:11 2011
 PRNT 6 NEIG 4 HYDR 2 SYMM cyl
 PVAL    1.062   KAPPA    1.099   KPRIM    1.144           SIGPV    0.045
 PLMS 1  0   0.179   PLMS 2  0   0.086
!
ATOM 1 COMMENT H103 H-cx3     DATE Fri Apr 22 15:45:29 2011
 PRNT 6 NEIG 4 HYDR 1 SYMM cyl
 PVAL    1.010   KAPPA    1.118   KPRIM    1.160           SIGPV    0.050
 PLMS 1  0   0.166   PLMS 2  0   0.079
!
ATOM 1 COMMENT H104 H-c (aromatic) or >c=c<     DATE Fri Apr 22 15:45:36 2011
 PRNT 6 NEIG 3 HYDR 1 SYMM cyl
 PVAL    1.034   KAPPA    1.110   KPRIM    1.158           SIGPV    0.049
 PLMS 1  0   0.182   PLMS 2  0   0.086
!
ATOM 1 COMMENT H105 x-nH3     DATE Fri Apr 22 15:45:53 2011
 PRNT 7 NEIG 4 HYDR 3 SYMM cyl
 PVAL    0.772   KAPPA    1.178   KPRIM    1.316           SIGPV    0.037
 PLMS 1  0   0.141   PLMS 2  0   0.061
!
ATOM 1 COMMENT H106 x-nH2-x     DATE Fri Apr 22 15:45:55 2011
 PRNT 7 NEIG 4 HYDR 2 SYMM cyl
 PVAL    0.782   KAPPA    1.178   KPRIM    1.314           SIGPV    0.026
 PLMS 1  0   0.142   PLMS 2  0   0.064
!
ATOM 1 COMMENT H107 x-nH2     DATE Fri Apr 22 15:45:55 2011
 PRNT 7 NEIG 3 HYDR 2 SYMM cyl
 EXCEPT GROUP planar amine
 PVAL    0.903   KAPPA    1.155   KPRIM    1.368           SIGPV    0.033
 PLMS 1  0   0.152   PLMS 2  0   0.066
!
ATOM 1 COMMENT H108 x-nH-x     DATE Fri Apr 22 15:45:57 2011
 PRNT 7 NEIG 3 HYDR 1 SYMM cyl
 EXCEPT GROUP planar amine
 PVAL    0.912   KAPPA    1.152   KPRIM    1.404           SIGPV    0.030
 PLMS 1  0   0.136   PLMS 2  0   0.057
!
```

```
ATOM 1 COMMENT H109 x-=nH=-x     DATE Fri Apr 22 15:45:58 2011
 PRNT 7 NEIG 3 HYDR 1 SYMM cyl
 EXCEPT GROUP aromatic
 PVAL    0.851   KAPPA    1.166   KPRIM    1.383           SIGPV    0.050
 PLMS 1  0   0.141   PLMS 2  0   0.059
!
ATOM 1 COMMENT H110 x-nH2     DATE Fri Apr 22 15:45:59 2011
 PRNT 7 NEIG 3 HYDR 2 SYMM cyl
 EXCEPT GROUP pyramidal amine
 PVAL    0.962   KAPPA    1.151   KPRIM    1.324           SIGPV    0.069
 PLMS 1  0   0.165   PLMS 2  0   0.075
!
ATOM 1 COMMENT H111 x-nH-x     DATE Fri Apr 22 15:46:00 2011
 PRNT 7 NEIG 3 HYDR 1 SYMM cyl
 EXCEPT GROUP pyramidal amine
 PVAL    0.917   KAPPA    1.150   KPRIM    1.399           SIGPV    0.057
 PLMS 1  0   0.136   PLMS 2  0   0.060
!
ATOM 1 COMMENT H112 x-oH -> for -csp3-oH     DATE Fri Apr 22 15:46:00 2011
 PRNT 8 NEIG 2 HYDR 1 SYMM cyl
 EXCEPT GROUP alcohol
 PVAL    0.881   KAPPA    1.162   KPRIM    1.396           SIGPV    0.027
 PLMS 1  0   0.145   PLMS 2  0   0.072
!
ATOM 1 COMMENT H113 x-oH -> for -c(=o)-oH     DATE Fri Apr 22 15:46:02 2011
 PRNT 8 NEIG 2 HYDR 1 SYMM cyl
 EXCEPT GROUP carboxylic
 PVAL    0.804   KAPPA    1.159   KPRIM    1.488           SIGPV    0.032
 PLMS 1  0   0.120   PLMS 2  0   0.053
!
ATOM 1 COMMENT H114 x-oH -> for -csp2-oH     DATE Fri Apr 22 15:46:02 2011
 PRNT 8 NEIG 2 HYDR 1 SYMM cyl
 EXCEPT GROUP phenol
 PVAL    0.817   KAPPA    1.177   KPRIM    1.463           SIGPV    0.070
 PLMS 1  0   0.131   PLMS 2  0   0.059
!
ATOM 1 COMMENT H115 x-oH -> for x3--p-oH     DATE Fri Apr 22 15:46:02 2011
 PRNT 8 NEIG 2 HYDR 1 SYMM cyl
 EXCEPT GROUP phosp... acid
 PVAL    0.828   KAPPA    1.160   KPRIM    1.382           SIGPV    0.060
 PLMS 1  0   0.129   PLMS 2  0   0.061
!
ATOM 1 COMMENT H116 c-sH     DATE Fri Apr 22 15:46:03 2011
 PRNT 16 NEIG 2 HYDR 1 SYMM cyl
 PVAL    0.873   KAPPA    1.168   KPRIM    1.420           SIGPV    0.009
 PLMS 1  0   0.089   PLMS 2  0   0.033
!
ATOM 1 COMMENT H117 b-oH     DATE Fri Apr 22 15:46:03 2011
 PRNT 8 NEIG 2 HYDR 1 SYMM cyl
 EXCEPT GROUP B-O-H
 PVAL    0.847   KAPPA    1.168   KPRIM    1.420           SIGPV    0.009
 PLMS 1  0   0.114   PLMS 2  0   0.052
!
ATOM 1 COMMENT H001 in water     DATE Fri Apr 22 15:46:03 2011
 PRNT 8 NEIG 2 HYDR 2 SYMM cyl
 PVAL    0.828   KAPPA    1.173   KPRIM    1.373           SIGPV    0.003
 PLMS 1  0   0.133   PLMS 2  0   0.049
!
```

```
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
!  C A R B O N
!
! **** 2 NEIGHBOURS ****
!
ATOM 6 COMMENT C201 c-C=-n  nitrile    DATE Fri Apr 22 15:46:03 2011
 NEIG 2 TYPE 2 RING 0 MEMB 0 SYMM cyl
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 7 NUMB 1
 PVAL    4.484   KAPPA    0.994   KPRIM    0.780           SIGPV    0.073
 PLMS 1  0  -0.350   PLMS 2  0   0.558   PLMS 3  0  -0.026   PLMS 4  0   0.135
!
! **** 3 NEIGHBOURS ****
!
ATOM 6 COMMENT C301 c-Coo[-]  carboxylate     DATE Fri Apr 22 15:46:03 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM mm2
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 8 NUMB 2
 EXCEPT GROUP carboxylate
 PVAL    4.039   KAPPA    0.990   KPRIM    0.872           SIGPV    0.037
 PLMS 1  0  -0.081   PLMS 2  2   0.275   PLMS 3  0   0.313   PLMS 3  2  -0.232
 PLMS 4  0   0.040   PLMS 4  4   0.043
!
ATOM 6 COMMENT C302 c-C(=o)-oh  carboxylic     DATE Fri Apr 22 15:46:03 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 8 NUMB 2
 EXCEPT GROUP carboxylic
 PVAL    4.089   KAPPA    0.993   KPRIM    0.864           SIGPV    0.031
 PLMS 1  1   0.119   PLMS 1 -1  -0.043   PLMS 2  0  -0.300   PLMS 2  2   0.111
 PLMS 2 -2   0.042   PLMS 3 -1  -0.006   PLMS 3  3   0.418   PLMS 4  0   0.068
 PLMS 4  2  -0.014   PLMS 4  4  -0.059   PLMS 4 -4   0.055
!
ATOM 6 COMMENT C303 c-C(=o)-o-c  ester     DATE Fri Apr 22 15:46:03 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 8 NUMB 2
 EXCEPT GROUP ester
 PVAL    4.086   KAPPA    0.995   KPRIM    0.858           SIGPV    0.049
 PLMS 1  1   0.128   PLMS 1 -1  -0.054   PLMS 2  0  -0.301   PLMS 2  2   0.128
 PLMS 2 -2   0.059   PLMS 3  3   0.422   PLMS 3 -3  -0.022   PLMS 4  0   0.073
 PLMS 4  2  -0.013   PLMS 4  4  -0.051   PLMS 4 -4   0.055
!
ATOM 6 COMMENT C304 c-C(=o)-n  amide     DATE Fri Apr 22 15:46:05 2011
 NEIG 3 TYPE 3 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 7 NUMB 1
 TYPE 3 ATOM 8 NUMB 1
 PVAL    4.112   KAPPA    0.994   KPRIM    0.856           SIGPV    0.051
 PLMS 1  1   0.075   PLMS 2  0  -0.292   PLMS 2  2   0.088   PLMS 3  3   0.419
 PLMS 4  0   0.062   PLMS 4  2  -0.011   PLMS 4  4  -0.049   PLMS 4 -4   0.036
!
ATOM 6 COMMENT C305 n-C(=o)-o-c  carbamate ester     DATE Fri Apr 22 15:46:06 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 7 NUMB 1
 TYPE 2 ATOM 8 NUMB 2
 EXCEPT GROUP carbamate ester
 PVAL    4.085   KAPPA    0.996   KPRIM    0.842           SIGPV    0.028
 PLMS 1  1   0.140   PLMS 1 -1  -0.041   PLMS 2  0  -0.290   PLMS 2  2   0.135
 PLMS 2 -2   0.043   PLMS 3  1  -0.012   PLMS 3 -1  -0.010   PLMS 3  3   0.467
 PLMS 4  0   0.068   PLMS 4  2  -0.018   PLMS 4 -2   0.012   PLMS 4  4  -0.112
!
```

```
ATOM 6 COMMENT C306 c2-C=o    DATE Fri Apr 22 15:46:06 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM mm2
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 8 NUMB 1
 EXCEPT GROUP ketone
 PVAL    4.085   KAPPA    0.995   KPRIM    0.867           SIGPV    0.074
 PLMS 1  0   0.082   PLMS 2  0   0.245   PLMS 2  2   0.216   PLMS 3  0   0.318
 PLMS 3  2  -0.250   PLMS 4  0   0.020   PLMS 4  2   0.057   PLMS 4  4   0.025
!
ATOM 6 COMMENT C307 c2-C=o    DATE Fri Apr 22 15:46:06 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 8 NUMB 1
 EXCEPT GROUP ketone
 PVAL    4.063   KAPPA    0.997   KPRIM    0.873           SIGPV    0.053
 PLMS 1  1   0.085   PLMS 2  0  -0.312   PLMS 2  2   0.085   PLMS 3  1   0.011
 PLMS 3  3   0.369   PLMS 4  0   0.064   PLMS 4  2  -0.015
!
ATOM 6 COMMENT C308 n2-C=o     DATE Fri Apr 22 15:46:06 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM mm2
 TYPE 1 ATOM 7 NUMB 2
 TYPE 2 ATOM 8 NUMB 1
 PVAL    4.153   KAPPA    0.990   KPRIM    0.835           SIGPV    0.031
 PLMS 1  0   0.081   PLMS 2  0   0.231   PLMS 2  2   0.213   PLMS 3  0   0.424
 PLMS 3  2  -0.297   PLMS 4  0  -0.048   PLMS 4  2   0.110   PLMS 4  4   0.017
!
ATOM 6 COMMENT C309 n--Cc--n  planar    DATE Fri Apr 22 15:46:06 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM mm2 PLAN t
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 7 NUMB 2
 PVAL    4.139   KAPPA    0.995   KPRIM    0.864           SIGPV    0.065
 PLMS 2  0   0.136   PLMS 2  2   0.255   PLMS 3  0   0.359   PLMS 3  2  -0.255
 PLMS 4  2   0.036   PLMS 4  4   0.038
!
ATOM 6 COMMENT C310 n--Cn--n  guanidine     DATE Fri Apr 22 15:46:06 2011
 NEIG 3 TYPE 1 RING 0 MEMB 0 SYMM 3m PLAN t
 TYPE 1 ATOM 7 NUMB 3
 PVAL    4.185   KAPPA    0.992   KPRIM    0.841           SIGPV    0.069
 PLMS 2  0  -0.306   PLMS 3  3   0.512   PLMS 4  0   0.054
!
ATOM 6 COMMENT C311 c=Ch-c  TEMP    DATE Fri Apr 22 15:46:06 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 1 ATOM 6 NUMB 2
 PVAL    4.022   KAPPA    0.997   KPRIM    0.876           SIGPV    0.101
 PLMS 1  1   0.101   PLMS 1 -1   0.021   PLMS 2  0  -0.185   PLMS 2  2   0.076
 PLMS 3  1   0.064   PLMS 3  3   0.296   PLMS 4  0   0.022   PLMS 4  2   0.034
!
ATOM 6 COMMENT C312 c=Ch-c  TEMP    DATE Fri Apr 22 15:46:07 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM mm2
 TYPE 1 ATOM 1 NUMB 1
 TYPE 1 ATOM 6 NUMB 2
 PVAL    3.906   KAPPA    1.001   KPRIM    0.893           SIGPV    0.111
 PLMS 1  0   0.058   PLMS 2  0   0.039   PLMS 2  2   0.164   PLMS 3  0  -0.248
 PLMS 3  2   0.154   PLMS 4  2   0.021
!
ATOM 6 COMMENT C313 c=Cc2-c1  TEMP    DATE Fri Apr 22 15:46:07 2011
 NEIG 3 TYPE 1 RING 0 MEMB 0 SYMM mm2
 TYPE 1 ATOM 6 NUMB 3
 PVAL    4.014   KAPPA    0.998   KPRIM    0.898           SIGPV    0.195
 PLMS 2  0   0.096   PLMS 2  2   0.143   PLMS 3  0   0.239   PLMS 3  2  -0.169
!
```

```
ATOM 6 COMMENT C314 c=Co-c    DATE Fri Apr 22 15:46:07 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM mm2
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 8 NUMB 1
 PVAL    3.730   KAPPA    1.012   KPRIM    0.970           SIGPV    0.043
 PLMS 1  0  -0.052   PLMS 2  2   0.096   PLMS 3  0   0.232   PLMS 3  2  -0.094
 PLMS 4  0   0.071   PLMS 4  2  -0.032
!
!------ set of x-=Cx=-x aromatic -------------------------------------------
!
!
ATOM 6 COMMENT C328 c-Cc-b             DATE Tue Apr 07 11:10:35 2009
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 5 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 PVAL    4.106   KAPPA    0.995   KPRIM    0.885           SIGPV    0.052
 PLMS 1  1  -0.023   PLMS 2  0  -0.213   PLMS 2  2   0.032   PLMS 3  1   0.009
 PLMS 3  3  -0.279   PLMS 4  0   0.015   PLMS 4  2  -0.014   PLMS 4  4  -0.023
!
ATOM 6 COMMENT C329 c-Cc-s    TEMP3    DATE Tue Apr 07 11:10:35 2009
 NEIG 3 TYPE 2 RING 1 MEMB 5 SYMM m PLAN t
 TYPE 1 ATOM 16 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 PVAL    4.010   KAPPA    1.002   KPRIM    0.866           SIGPV    0.033
 PLMS 1  1   0.095   PLMS 2  0  -0.153   PLMS 2  2  -0.094   PLMS 3  1   0.007
 PLMS 3  3  -0.277   PLMS 4  0   0.020   PLMS 4  2  -0.011   PLMS 4  4  -0.014
!
ATOM 6 COMMENT C330 c-=Ch=-c  AROMATIC     DATE Fri Apr 22 15:46:07 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 PVAL    3.907   KAPPA    1.004   KPRIM    0.909           SIGPV    0.049
 PLMS 1  1   0.056   PLMS 2  0  -0.180   PLMS 2  2  -0.029   PLMS 3  1   0.030
 PLMS 3  3  -0.264   PLMS 4  0   0.020   PLMS 4  2  -0.019
!
ATOM 6 COMMENT C331 c-=Ch=-c  AROMATIC     DATE Fri Apr 22 15:46:21 2011
 NEIG 3 TYPE 2 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 PVAL    3.882   KAPPA    1.006   KPRIM    0.912           SIGPV    0.096
 PLMS 1  1   0.058   PLMS 2  0  -0.145   PLMS 3  1   0.025   PLMS 3 -1   0.019
 PLMS 3  3  -0.246   PLMS 4  0   0.023   PLMS 4  2  -0.017
!
ATOM 6 COMMENT C332 c-=Cc=-c  AROMATIC     DATE Fri Apr 22 15:46:21 2011
 NEIG 3 TYPE 1 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 6 NUMB 3
 PVAL    4.031   KAPPA    0.999   KPRIM    0.887           SIGPV    0.093
 PLMS 1  1   0.047   PLMS 2  0  -0.179   PLMS 2  2  -0.033   PLMS 3  1   0.026
 PLMS 3  3  -0.289   PLMS 4  0   0.018   PLMS 4  2  -0.013
!
ATOM 6 COMMENT C333 c-=Cc=-c  AROMATIC     DATE Fri Apr 22 15:46:24 2011
 NEIG 3 TYPE 1 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 6 NUMB 3
 PVAL    3.971   KAPPA    1.003   KPRIM    0.899           SIGPV    0.082
 PLMS 1 -1   0.025   PLMS 2  0  -0.128   PLMS 3  1   0.020   PLMS 3 -1   0.019
 PLMS 3  3  -0.255   PLMS 4  0   0.020   PLMS 4  4  -0.012
!
ATOM 6 COMMENT C334 c-=Cn=-c  AROMATIC nitro     DATE Fri Apr 22 15:46:24 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 7 NUMB 1
 EXCEPT GROUP nitrophenyl
 PVAL    3.841   KAPPA    1.014   KPRIM    0.924           SIGPV    0.089
 PLMS 1  1   0.128   PLMS 2  0  -0.090   PLMS 2  2  -0.128   PLMS 3  1   0.026
 PLMS 3  3  -0.250   PLMS 4  0   0.009   PLMS 4  2  -0.031   PLMS 4 -4  -0.004
!
```

```
ATOM 6 COMMENT C338 c-=Cn=-c  AROMATIC amino     DATE Fri Apr 22 15:46:24 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 7 NUMB 1
 EXCEPT GROUP aminophenyl
 PVAL    4.088   KAPPA    0.998   KPRIM    0.872           SIGPV    0.069
 PLMS 1  1   0.082   PLMS 2  0  -0.187   PLMS 2  2  -0.074   PLMS 3  1   0.041
 PLMS 3  3  -0.338   PLMS 4  0   0.028   PLMS 4  2  -0.018   PLMS 4  4   0.028
!
ATOM 6 COMMENT C340 c-=Cn=-c AROMATIC prot amino     DATE Fri Apr 22 15:46:25 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 7 NUMB 1
 EXCEPT GROUP protonated aminophenyl
 PVAL    3.936   KAPPA    1.004   KPRIM    0.917           SIGPV    0.056
 PLMS 1  1   0.172   PLMS 2  0  -0.063   PLMS 2  2  -0.173   PLMS 3  1   0.031
 PLMS 3  3  -0.249   PLMS 4  2  -0.020   PLMS 4  4   0.013
!
ATOM 6 COMMENT C341 c-=Co=-c  AROMATIC -oh     DATE Fri Apr 22 15:46:25 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 8 NUMB 1
 EXCEPT GROUP phenolic
 PVAL    4.049   KAPPA    0.998   KPRIM    0.864           SIGPV    0.060
 PLMS 1  1   0.119   PLMS 2  0  -0.190   PLMS 2  2  -0.106   PLMS 3  1   0.045
 PLMS 3  3  -0.357   PLMS 4  0   0.028   PLMS 4  2  -0.020   PLMS 4  4   0.047
 PLMS 4 -4   0.015
!
ATOM 6 COMMENT C344 c-=Cf=-c  AROMATIC  TEMP2     DATE Fri Apr 22 15:46:25 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 9 NUMB 1
 PVAL    3.981   KAPPA    1.001   KPRIM    0.904           SIGPV    0.001
 PLMS 1  1   0.165   PLMS 1 -1   0.010   PLMS 2  0  -0.124   PLMS 2  2  -0.185
 PLMS 2 -2   0.005   PLMS 3  1   0.047   PLMS 3  3  -0.303   PLMS 3 -3   0.006
 PLMS 4  0   0.010   PLMS 4  2  -0.019   PLMS 4 -4  -0.018
!
ATOM 6 COMMENT C345 c-=Cp=-c  AROMATIC     DATE Fri Apr 22 15:46:25 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 15 NUMB 1
 PVAL    3.961   KAPPA    1.006   KPRIM    0.889           SIGPV    0.035
 PLMS 1  1   0.050   PLMS 2  0  -0.165   PLMS 2  2  -0.048   PLMS 3  3  -0.251
 PLMS 4  0   0.011   PLMS 4  4  -0.010
!
ATOM 6 COMMENT C346 c-=Ccl=-c  AROMATIC    TEMP    DATE Fri Apr 22 15:46:25 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM  6 NUMB 2
 TYPE 1 ATOM 17 NUMB 1
 PVAL    4.083   KAPPA    0.997   KPRIM    0.874           SIGPV    0.027
 PLMS 1  1   0.117   PLMS 2  0  -0.139   PLMS 2  2  -0.140   PLMS 3  1   0.033
 PLMS 3  3  -0.305   PLMS 4  0   0.014   PLMS 4  2  -0.019   PLMS 4  4   0.013
!
ATOM 6 COMMENT C350  nsp2(2)-=Ch=-c  AROMATIC     DATE Fri Apr 22 15:46:26 2011
 NEIG 3 TYPE 3 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 7 NUMB 1
 EXCEPT GROUP non-substituted nitrogen
 PVAL    3.977   KAPPA    1.003   KPRIM    0.886           SIGPV    0.035
 PLMS 1  1   0.067   PLMS 2  0  -0.220   PLMS 2  2  -0.021   PLMS 3  1   0.018
 PLMS 3 -1  -0.016   PLMS 3  3  -0.331   PLMS 4  0   0.040   PLMS 4  2  -0.013
 PLMS 4 -2  -0.020   PLMS 4  4  -0.007   PLMS 4 -4  -0.024
!
```

```
ATOM 6 COMMENT C351  nsp2(2)-=Ch=-c  AROMATIC     DATE Fri Apr 22 15:46:26 2011
 NEIG 3 TYPE 3 RING 1 MEMB 5 SYMM m PLAN t
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 7 NUMB 1
 EXCEPT GROUP non-substituted nitrogen
 PVAL    3.987   KAPPA    1.004   KPRIM    0.877           SIGPV    0.043
 PLMS 1  1   0.076   PLMS 1 -1  -0.031   PLMS 2  0  -0.154   PLMS 2  2   0.011
 PLMS 3 -1  -0.030   PLMS 3  3  -0.308   PLMS 3 -3   0.030   PLMS 4  0   0.046
 PLMS 4  2  -0.023   PLMS 4 -2  -0.017   PLMS 4  4  -0.038   PLMS 4 -4  -0.024
!
ATOM 6 COMMENT C352 nsp2(3)-=Ch=-c  AROMATIC     DATE Fri Apr 22 15:46:26 2011
 NEIG 3 TYPE 3 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 7 NUMB 1
 EXCEPT GROUP substituted nitrogen
 PVAL    3.913   KAPPA    1.007   KPRIM    0.899           SIGPV    0.106
 PLMS 1 -1  -0.084   PLMS 2  0  -0.214   PLMS 2 -2  -0.083   PLMS 3  1   0.017
 PLMS 3 -1  -0.039   PLMS 3  3  -0.292   PLMS 3 -3   0.024   PLMS 4  0   0.038
 PLMS 4  2  -0.013   PLMS 4 -2  -0.021   PLMS 4 -4  -0.026
!
ATOM 6 COMMENT C353 nsp2(3)-=Ch=-c  AROMATIC     DATE Fri Apr 22 15:46:27 2011
 NEIG 3 TYPE 3 RING 1 MEMB 5 SYMM m PLAN t
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 7 NUMB 1
 EXCEPT GROUP substituted nitrogen
 PVAL    3.891   KAPPA    1.010   KPRIM    0.910           SIGPV    0.061
 PLMS 1  1   0.030   PLMS 1 -1  -0.078   PLMS 2  0  -0.124   PLMS 2  2   0.037
 PLMS 2 -2  -0.088   PLMS 3 -1  -0.037   PLMS 3  3  -0.255   PLMS 3 -3   0.028
 PLMS 4  0   0.038   PLMS 4  2  -0.017   PLMS 4 -2  -0.018   PLMS 4  4  -0.021
 PLMS 4 -4  -0.018
!
ATOM 6 COMMENT C354 nsp2(2)-=Cc=-c  AROMATIC    DATE Fri Apr 22 15:46:27 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 7 NUMB 1
 EXCEPT GROUP non-substituted nitrogen
 PVAL    4.083   KAPPA    0.998   KPRIM    0.869           SIGPV    0.101
 PLMS 1  1   0.055   PLMS 2  0  -0.235   PLMS 2 -2  -0.024   PLMS 3  1   0.014
 PLMS 3 -1  -0.028   PLMS 3  3  -0.352   PLMS 4  0   0.037   PLMS 4  2  -0.011
 PLMS 4 -2  -0.020   PLMS 4 -4  -0.037
!
ATOM 6 COMMENT C355 nsp2(2)-=Cc=-c  AROMATIC    DATE Fri Apr 22 15:46:27 2011
 NEIG 3 TYPE 2 RING 1 MEMB 5 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 7 NUMB 1
 EXCEPT GROUP non-substituted nitrogen
 PVAL    4.129   KAPPA    0.999   KPRIM    0.858           SIGPV    0.057
 PLMS 1  1   0.058   PLMS 1 -1  -0.035   PLMS 2  0  -0.180   PLMS 2  2   0.018
 PLMS 2 -2  -0.030   PLMS 3  1   0.009   PLMS 3 -1  -0.031   PLMS 3  3  -0.343
 PLMS 3 -3   0.032   PLMS 4  0   0.047   PLMS 4  2  -0.024   PLMS 4 -2  -0.020
 PLMS 4  4  -0.050   PLMS 4 -4  -0.022
!
ATOM 6 COMMENT C356 nsp2(3)-=Cc=-c  AROMATIC     DATE Fri Apr 22 15:46:27 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 7 NUMB 1
 EXCEPT GROUP substituted nitrogen
 PVAL    4.125   KAPPA    0.995   KPRIM    0.868           SIGPV    0.107
 PLMS 1 -1  -0.079   PLMS 2  0  -0.257   PLMS 2 -2  -0.081   PLMS 3  1   0.014
 PLMS 3  3  -0.341   PLMS 4  0   0.042   PLMS 4 -2  -0.024   PLMS 4  4  -0.028
 PLMS 4 -4  -0.028
!
```

```
ATOM 6 COMMENT C357 nsp2(3)-=Cc=-c  AROMATIC     DATE Fri Apr 22 15:46:27 2011
 NEIG 3 TYPE 2 RING 1 MEMB 5 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 7 NUMB 1
 EXCEPT GROUP substituted nitrogen
 PVAL    4.041   KAPPA    1.003   KPRIM    0.890           SIGPV    0.078
 PLMS 1 -1  -0.084   PLMS 2  0  -0.130   PLMS 2  2   0.045   PLMS 2 -2  -0.101
 PLMS 3  1   0.012   PLMS 3 -1  -0.041   PLMS 3  3  -0.287   PLMS 3 -3   0.037
 PLMS 4  0   0.035   PLMS 4  2  -0.018   PLMS 4 -2  -0.027   PLMS 4  4  -0.041
 PLMS 4 -4  -0.018
!
ATOM 6 COMMENT C358 nsp2(2)-=Ch=-nsp2(2)  AROMATIC     DATE Fri Apr 22 15:46:27 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 7 NUMB 2
 EXCEPT GROUP non-substituted nitrogens
 PVAL    4.037   KAPPA    1.000   KPRIM    0.884           SIGPV    0.037
 PLMS 1  1   0.055   PLMS 2  0  -0.229   PLMS 2  2  -0.015   PLMS 3  1   0.009
 PLMS 3  3  -0.360   PLMS 4  0   0.050   PLMS 4  2  -0.009   PLMS 4  4   0.009
!
ATOM 6 COMMENT C360 nsp2(2)-=Ch=-nsp2(3)  AROMATIC     DATE Fri Apr 22 15:46:27 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 7 NUMB 2
 EXCEPT GROUP mixed nitrogens
 PVAL    4.077   KAPPA    1.002   KPRIM    0.875           SIGPV    0.030
 PLMS 1  1   0.022   PLMS 1 -1   0.078   PLMS 2  0  -0.214   PLMS 2 -2   0.093
 PLMS 3 -1   0.019   PLMS 3  3  -0.361   PLMS 3 -3  -0.085   PLMS 4  0   0.059
 PLMS 4  2  -0.007   PLMS 4  4  -0.014   PLMS 4 -4   0.017
!
ATOM 6 COMMENT C361 nsp2(2)-=Ch=-nsp2(3)  AROMATIC     DATE Fri Apr 22 15:46:28 2011
 NEIG 3 TYPE 2 RING 1 MEMB 5 SYMM m PLAN t
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 7 NUMB 2
 EXCEPT GROUP mixed nitrogens
 PVAL    4.023   KAPPA    1.007   KPRIM    0.882           SIGPV    0.052
 PLMS 1  1   0.052   PLMS 1 -1   0.078   PLMS 2  0  -0.176   PLMS 2  2   0.032
 PLMS 2 -2   0.084   PLMS 3  1  -0.009   PLMS 3 -1   0.017   PLMS 3  3  -0.340
 PLMS 3 -3  -0.020   PLMS 4  0   0.059   PLMS 4  2  -0.020   PLMS 4  4  -0.049
!
ATOM 6 COMMENT C363 nsp2(3)-=Ch=-nsp2(3)  AROMATIC     DATE Fri Apr 22 15:46:28 2011
 NEIG 3 TYPE 2 RING 1 MEMB 5 SYMM m PLAN t
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 7 NUMB 2
 EXCEPT GROUP substituted nitrogens
 PVAL    3.832   KAPPA    1.018   KPRIM    0.908           SIGPV    0.076
 PLMS 2  0  -0.162   PLMS 2  2   0.054   PLMS 3  1  -0.011   PLMS 3  3  -0.290
 PLMS 4  0   0.058   PLMS 4  2  -0.014   PLMS 4  4  -0.021
!
ATOM 6 COMMENT C366 nsp2(3)-=Cc=-nsp2(3)  AROMATIC  TEMP3     DATE Fri Apr 22 15:46:28 2011
 NEIG 3 TYPE 2 RING 1 MEMB 5 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 7 NUMB 2
 EXCEPT GROUP substituted nitrogens
 PVAL    4.111   KAPPA    1.001   KPRIM    0.849           SIGPV    0.040
 PLMS 1  1  -0.033   PLMS 2  0  -0.241   PLMS 2  2   0.053   PLMS 3  1  -0.010
 PLMS 3  3  -0.395   PLMS 4  0   0.078   PLMS 4  2  -0.016   PLMS 4  4  -0.060
!
ATOM 6 COMMENT C367 nsp2(2)-=Cc=-nsp2(3)  AROMATIC  TEMP2     DATE Fri Apr 22 15:46:28 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 7 NUMB 2
 EXCEPT GROUP mixed nitrogens
 PVAL    4.096   KAPPA    0.998   KPRIM    0.862           SIGPV    0.081
 PLMS 1 -1   0.088   PLMS 2  0  -0.255   PLMS 2 -2   0.086   PLMS 3 -1   0.009
 PLMS 3  3  -0.392   PLMS 3 -3  -0.025   PLMS 4  0   0.052   PLMS 4  4  -0.024
 PLMS 4 -4   0.004
```

```
ATOM 6 COMMENT C368 nsp2(2)-=Cnsp2(3)=-nsp2(2)  AROMATIC    DATE Fri Apr 22 15:46:28 2011
 NEIG 3 TYPE 1 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 7 NUMB 3
 EXCEPT GROUP 3N-sp2(232)
 PVAL    4.146   KAPPA    0.994   KPRIM    0.844           SIGPV    0.051
 PLMS 1 -1   0.010   PLMS 2  0  -0.281   PLMS 2 -2   0.011   PLMS 3 -1   0.005
 PLMS 3  3  -0.450   PLMS 4  0   0.070   PLMS 4  4   0.048
!
ATOM 6 COMMENT C369 nsp2(2)-=Cnsp2(3)=-nsp2(2)  AROMATIC     DATE Fri Apr 22 15:46:28 2011
 NEIG 3 TYPE 1 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 7 NUMB 3
 EXCEPT GROUP 3N-sp2(232)
 PVAL    4.094   KAPPA    1.010   KPRIM    0.849           SIGPV    0.077
 PLMS 1  1   0.161   PLMS 1 -1   0.048   PLMS 2  0  -0.182   PLMS 2  2  -0.077
 PLMS 2 -2   0.025   PLMS 3 -1   0.028   PLMS 3  3  -0.418   PLMS 3 -3  -0.034
 PLMS 4  0   0.075   PLMS 4  2  -0.044   PLMS 4  4  -0.060   PLMS 4 -4  -0.012
!
ATOM 6 COMMENT C370 nsp2(2)-=Cnsp2(3)=-nsp2(3)  AROMATIC     DATE Fri Apr 22 15:46:28 2011
 NEIG 3 TYPE 1 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 7 NUMB 3
 EXCEPT GROUP 3N-sp2(233)
 PVAL    4.191   KAPPA    0.991   KPRIM    0.835           SIGPV    0.038
 PLMS 1 -1   0.059   PLMS 2  0  -0.296   PLMS 2 -2   0.077   PLMS 3 -1   0.009
 PLMS 3  3  -0.476   PLMS 3 -3  -0.062   PLMS 4  0   0.069   PLMS 4  2   0.011
 PLMS 4  4   0.030   PLMS 4 -4   0.037
!
ATOM 6 COMMENT C371 nsp2(2)-=Cnsp2(3)=-nsp2(3)  AROMATIC     DATE Fri Apr 22 15:46:28 2011
 NEIG 3 TYPE 1 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 7 NUMB 3
 EXCEPT GROUP 3N-sp2(233)
 PVAL    4.184   KAPPA    1.001   KPRIM    0.831           SIGPV    0.098
 PLMS 1  1   0.066   PLMS 1 -1   0.061   PLMS 2  0  -0.230   PLMS 2 -2   0.071
 PLMS 3  3  -0.468   PLMS 3 -3  -0.035   PLMS 4  0   0.074   PLMS 4 -2  -0.004
 PLMS 4  4  -0.037
!
ATOM 6 COMMENT C373 nsp2(2)-=Cnsp2(3)=-csp2(3)  AROMATIC     DATE Fri Apr 22 15:46:28 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 7 NUMB 2
 EXCEPT GROUP n2n3c3
 PVAL    4.148   KAPPA    0.995   KPRIM    0.852           SIGPV    0.030
 PLMS 1  1   0.049   PLMS 2  0  -0.275   PLMS 3 -1  -0.015   PLMS 3  3  -0.421
 PLMS 4  0   0.048   PLMS 4 -2  -0.012   PLMS 4  4   0.039   PLMS 4 -4  -0.031
!
ATOM 6 COMMENT C374 nsp2(3)-=Cnsp2(3)=-csp2(3)  AROMATIC     DATE Fri Apr 22 15:46:29 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 7 NUMB 2
 EXCEPT GROUP n3n3c3
 PVAL    4.171   KAPPA    0.995   KPRIM    0.843           SIGPV    0.031
 PLMS 1 -1  -0.066   PLMS 2  0  -0.294   PLMS 2  2   0.021   PLMS 2 -2  -0.073
 PLMS 3  1   0.007   PLMS 3 -1  -0.022   PLMS 3  3  -0.434   PLMS 3 -3   0.044
 PLMS 4  0   0.060   PLMS 4 -2  -0.015   PLMS 4 -4  -0.045
!
ATOM 6 COMMENT C378 n-=C(=o)=-n  AROMATIC     DATE Fri Apr 22 15:46:29 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 7 NUMB 2
 TYPE 2 ATOM 8 NUMB 1
 EXCEPT GROUP aromaturea
 PVAL    4.152   KAPPA    0.993   KPRIM    0.830           SIGPV    0.011
 PLMS 1  1  -0.111   PLMS 2  0  -0.310   PLMS 2  2   0.138   PLMS 2 -2   0.011
 PLMS 3  3  -0.472   PLMS 4  0   0.086   PLMS 4  2  -0.026   PLMS 4  4  -0.114
!
```

```
ATOM 6 COMMENT C379 n-=C(=o)=-n  AROMATIC    TEMP3    DATE Fri Apr 22 15:46:29 2011
 NEIG 3 TYPE 2 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 7 NUMB 2
 TYPE 2 ATOM 8 NUMB 1
 EXCEPT GROUP aromaturea
 PVAL    4.144   KAPPA    0.992   KPRIM    0.853           SIGPV    0.039
 PLMS 1  1  -0.070   PLMS 1 -1   0.020   PLMS 2  0  -0.282   PLMS 2  2   0.118
 PLMS 2 -2   0.024   PLMS 3  1  -0.019   PLMS 3 -1   0.011   PLMS 3  3  -0.427
 PLMS 4  0   0.079   PLMS 4  4  -0.127
!
ATOM 6 COMMENT C380 n-=C(-n)=-o  AROMATIC    DATE Fri Apr 22 15:46:29 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 7 NUMB 2
 TYPE 2 ATOM 8 NUMB 1
 EXCEPT GROUP pirofurano
 PVAL    4.071   KAPPA    1.000   KPRIM    0.847           SIGPV    0.026
 PLMS 1 -1  -0.124   PLMS 2  0  -0.252   PLMS 2  2   0.036   PLMS 2 -2  -0.129
 PLMS 3 -1  -0.015   PLMS 3  3  -0.431   PLMS 3 -3   0.091   PLMS 4  0   0.061
 PLMS 4 -4  -0.057
!
ATOM 6 COMMENT C381 n-=C(-n)=-o  AROMATIC    DATE Fri Apr 22 15:46:29 2011
 NEIG 3 TYPE 2 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 7 NUMB 2
 TYPE 2 ATOM 8 NUMB 1
 EXCEPT GROUP pirofurano
 PVAL    4.136   KAPPA    1.001   KPRIM    0.834           SIGPV    0.043
 PLMS 1 -1  -0.131   PLMS 2  0  -0.229   PLMS 2  2   0.043   PLMS 2 -2  -0.144
 PLMS 3 -1  -0.021   PLMS 3  3  -0.462   PLMS 3 -3   0.070   PLMS 4  0   0.062
 PLMS 4  4  -0.027   PLMS 4 -4  -0.035
!
ATOM 6 COMMENT C382 n-=C(=o)=-c  AROMATIC     DATE Fri Apr 22 15:46:29 2011
 NEIG 3 TYPE 3 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 7 NUMB 1
 TYPE 3 ATOM 8 NUMB 1
 EXCEPT GROUP amido
 PVAL    4.150   KAPPA    0.993   KPRIM    0.844           SIGPV    0.033
 PLMS 1  1  -0.083   PLMS 1 -1  -0.049   PLMS 2  0  -0.309   PLMS 2  2   0.112
 PLMS 2 -2  -0.055   PLMS 3  3  -0.431   PLMS 3 -3   0.029   PLMS 4  0   0.075
 PLMS 4  2  -0.016   PLMS 4  4  -0.063   PLMS 4 -4  -0.038
!
ATOM 6 COMMENT C383 n-=C(=o)=-c  AROMATIC     DATE Fri Apr 22 15:46:30 2011
 NEIG 3 TYPE 3 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 7 NUMB 1
 TYPE 3 ATOM 8 NUMB 1
 EXCEPT GROUP amido
 PVAL    4.129   KAPPA    0.994   KPRIM    0.848           SIGPV    0.027
 PLMS 1  1  -0.071   PLMS 2  0  -0.295   PLMS 2  2   0.129   PLMS 3  3  -0.408
 PLMS 4  0   0.068   PLMS 4  2  -0.033   PLMS 4  4  -0.094   PLMS 4 -4  -0.035
!
ATOM 6 COMMENT C385 n-=C(=oh)=-c  AROMATIC    DATE Fri Apr 22 15:46:30 2011
 NEIG 3 TYPE 3 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 7 NUMB 1
 TYPE 3 ATOM 8 NUMB 1
 EXCEPT GROUP amidoH
 PVAL    4.178   KAPPA    1.000   KPRIM    0.854           SIGPV    0.025
 PLMS 1  1   0.153   PLMS 1 -1   0.053   PLMS 2  0  -0.220   PLMS 2  2  -0.050
 PLMS 2 -2   0.015   PLMS 3  1   0.036   PLMS 3  3  -0.437   PLMS 4  0   0.056
 PLMS 4  2  -0.043   PLMS 4  4   0.028   PLMS 4 -4  -0.032
!
```

```
ATOM 6 COMMENT C386 o-=Cc=-n  AROMATIC   DATE Fri Apr 22 15:46:30 2011
 NEIG 3 TYPE 3 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 7 NUMB 1
 TYPE 3 ATOM 8 NUMB 1
 EXCEPT GROUP o-Cc-n
 PVAL    4.121   KAPPA    1.003   KPRIM    0.857           SIGPV    0.052
 PLMS 1 -1  -0.140   PLMS 2  0  -0.167   PLMS 2  2   0.063   PLMS 2 -2  -0.155
 PLMS 3  1  -0.024   PLMS 3 -1  -0.028   PLMS 3  3  -0.378   PLMS 3 -3   0.060
 PLMS 4  0   0.053   PLMS 4  2  -0.016   PLMS 4  4  -0.076
!
ATOM 6 COMMENT C388 c-=Ch=-o  AROMATIC    DATE Fri Apr 22 15:46:30 2011
 NEIG 3 TYPE 3 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 8 NUMB 1
 PVAL    3.887   KAPPA    1.009   KPRIM    0.900           SIGPV    0.055
 PLMS 1  1   0.044   PLMS 1 -1  -0.118   PLMS 2  0  -0.088   PLMS 2  2   0.023
 PLMS 2 -2  -0.140   PLMS 3 -1  -0.058   PLMS 3  3  -0.279   PLMS 3 -3   0.065
 PLMS 4  0   0.030   PLMS 4  2  -0.021   PLMS 4 -2  -0.033   PLMS 4  4  -0.060
 PLMS 4 -4  -0.043
!
ATOM 6 COMMENT C389 o-=Cc=-c  AROMATIC    DATE Fri Apr 22 15:46:30 2011
 NEIG 3 TYPE 2 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 8 NUMB 1
 EXCEPT GROUP furano
 PVAL    3.996   KAPPA    1.006   KPRIM    0.905           SIGPV    0.062
 PLMS 1 -1  -0.112   PLMS 2  0  -0.093   PLMS 2  2   0.046   PLMS 2 -2  -0.147
 PLMS 3 -1  -0.054   PLMS 3  3  -0.277   PLMS 3 -3   0.067   PLMS 4  0   0.026
 PLMS 4  2  -0.012   PLMS 4 -2  -0.032   PLMS 4  4  -0.074   PLMS 4 -4  -0.034
!
ATOM 6 COMMENT C390 o-=Co=-c  AROMATIC    DATE Fri Apr 22 15:46:30 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 8 NUMB 2
 EXCEPT GROUP anhydride
 PVAL    4.112   KAPPA    0.991   KPRIM    0.845           SIGPV    0.025
 PLMS 1  1  -0.131   PLMS 1 -1  -0.127   PLMS 2  0  -0.277   PLMS 2  2   0.145
 PLMS 2 -2  -0.169   PLMS 3  1   0.008   PLMS 3 -1  -0.019   PLMS 3  3  -0.374
 PLMS 3 -3   0.076   PLMS 4  0   0.064   PLMS 4  2  -0.030   PLMS 4  4  -0.064
 PLMS 4 -4  -0.027
!
ATOM 6 COMMENT C391 (n)o-=Cc=-c AROMATIC    DATE Fri Apr 22 15:46:30 2011
 NEIG 3 TYPE 2 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 8 NUMB 1
 EXCEPT GROUP furanoN
 PVAL    4.137   KAPPA    0.999   KPRIM    0.856           SIGPV    0.031
 PLMS 1  1   0.035   PLMS 1 -1  -0.101   PLMS 2  0  -0.179   PLMS 2  2   0.028
 PLMS 2 -2  -0.134   PLMS 3 -1  -0.042   PLMS 3  3  -0.359   PLMS 3 -3   0.088
 PLMS 4  0   0.044   PLMS 4  2  -0.040   PLMS 4 -2  -0.037   PLMS 4  4  -0.104
 PLMS 4 -4  -0.043
!
ATOM 6 COMMENT C392 n-=C(=s)=-n  AROMATIC  TEMP3    DATE Fri Apr 22 15:46:30 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 7 NUMB 2
 TYPE 2 ATOM 16 NUMB 1
 PVAL    4.024   KAPPA    1.001   KPRIM    0.887           SIGPV    0.012
 PLMS 1  1  -0.089   PLMS 2  0  -0.212   PLMS 2  2   0.070   PLMS 2 -2   0.029
 PLMS 3  1  -0.063   PLMS 3  3  -0.318   PLMS 3 -3  -0.010   PLMS 4  0   0.036
 PLMS 4  2   0.017   PLMS 4 -2  -0.012   PLMS 4  4  -0.085   PLMS 4 -4  -0.011
!
```

```
ATOM 6 COMMENT C393 n-=C(=s)=-c  AROMATIC  TEMP2    DATE Fri Apr 22 15:46:30 2011
 NEIG 3 TYPE 3 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 7 NUMB 1
 TYPE 3 ATOM 16 NUMB 1
 EXCEPT GROUP THIO
 PVAL    4.005   KAPPA    1.003   KPRIM    0.877           SIGPV      0.01
 PLMS 1  1  -0.032   PLMS 2  0  -0.264   PLMS 2  2   0.030   PLMS 2 -2   0.058
 PLMS 3  1  -0.052   PLMS 3 -1  -0.014   PLMS 3  3  -0.331   PLMS 3 -3  -0.019
 PLMS 4  0   0.036   PLMS 4  2   0.022   PLMS 4 -2  -0.009   PLMS 4  4  -0.064
 PLMS 4 -4  -0.015
!
ATOM 6 COMMENT C501 (c-Cc-c 2-rings)    DATE Fri Apr 22 15:46:30 2011
 NEIG 3 TYPE 1 RING 2 MEMB 11 SYMM m   PLAN t
 TYPE 1 ATOM 6 NUMB 3
 PVAL    4.015   KAPPA    1.000   KPRIM    0.886           SIGPV    0.063
 PLMS 2  0  -0.173   PLMS 2 -2  -0.041   PLMS 3 -1   0.016   PLMS 3  3   0.270
 PLMS 3 -3   0.048   PLMS 4  0   0.021   PLMS 4  4   0.022
!
ATOM 6 COMMENT C503 (c-Cc-n 2-rings)    DATE Fri Apr 22 15:46:30 2011
 NEIG 3 TYPE 2 RING 2 MEMB 11 SYMM m   PLAN t
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 7 NUMB 1
EXCEPT GROUP C(2cyc)-n5
 PVAL    3.957   KAPPA    1.007   KPRIM    0.893           SIGPV    0.097
 PLMS 1  1   0.038   PLMS 2  0  -0.131   PLMS 2 -2  -0.053   PLMS 3 -1   0.033
 PLMS 3  3   0.284   PLMS 3 -3   0.068   PLMS 4  0   0.028   PLMS 4  2   0.017
 PLMS 4 -2  -0.011   PLMS 4  4   0.032   PLMS 4 -4  -0.035
!
ATOM 6 COMMENT C504 (c-Cc-o 2-rings)    TEMP3    DATE Fri Apr 22 15:46:31 2011
 NEIG 3 TYPE 2 RING 2 MEMB 11 SYMM m   PLAN t
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 8 NUMB 1
 PVAL    4.040   KAPPA    1.004   KPRIM    0.875           SIGPV    0.007
 PLMS 1  1   0.088   PLMS 1 -1   0.122   PLMS 2  0  -0.119   PLMS 2  2   0.093
 PLMS 2 -2  -0.140   PLMS 3  1  -0.014   PLMS 3 -1   0.050   PLMS 3  3   0.264
 PLMS 3 -3   0.124   PLMS 4  0   0.025   PLMS 4  2   0.010   PLMS 4 -2  -0.028
 PLMS 4  4   0.036   PLMS 4 -4  -0.038
!
ATOM 6 COMMENT C505 (n-Cc-n 2-rings)    DATE Fri Apr 22 15:46:31 2011
 NEIG 3 TYPE 2 RING 2 MEMB 11 SYMM m   PLAN t
 TYPE 1 ATOM 6 NUMB 1
 TYPE 1 ATOM 7 NUMB 2
 PVAL    4.118   KAPPA    1.000   KPRIM    0.846           SIGPV    0.036
 PLMS 1  1   0.083   PLMS 1 -1   0.069   PLMS 2  0  -0.220   PLMS 2  2   0.038
 PLMS 2 -2  -0.065   PLMS 3  1   0.015   PLMS 3 -1   0.007   PLMS 3  3   0.359
 PLMS 3 -3   0.145   PLMS 4  0   0.071   PLMS 4  2   0.033   PLMS 4 -2   0.026
!
ATOM 6 COMMENT C507 (c-Cc-c 2-rings)  TEMP4     DATE Fri Apr 22 15:46:31 2011
 NEIG 3 TYPE 1 RING 2 MEMB 12 SYMM m   PLAN t
 TYPE 1 ATOM 6 NUMB 3
 PVAL    3.983   KAPPA    1.004   KPRIM    0.903           SIGPV    0.057
 PLMS 2  0  -0.170   PLMS 3 -1   0.011   PLMS 3  3   0.279   PLMS 4  0   0.012
 PLMS 4 -2  -0.012   PLMS 4  4   0.008
!
! **** 4 NEIGHBOURS ****
!
!------ only c and h -------------------------------------------------------
!
ATOM 6 COMMENT C401 c-Ch3     DATE Fri Apr 22 15:46:31 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM 3m
 TYPE 1 ATOM 1 NUMB 3
 TYPE 2 ATOM 6 NUMB 1
 PVAL    3.588   KAPPA    1.020   KPRIM    0.927           SIGPV    0.126
 PLMS 3  0   0.239   PLMS 3  3   0.188   PLMS 4  0   0.054   PLMS 4  3  -0.080
!
```

```
ATOM 6 COMMENT C402 h-Cc3     DATE Fri Apr 22 15:46:36 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM 3m   PLAN f
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 3
 PVAL    3.946   KAPPA    0.999   KPRIM    0.885           SIGPV    0.135
 PLMS 1  0  -0.028   PLMS 3  0   0.281   PLMS 3  3   0.231   PLMS 4  0   0.063
 PLMS 4  3  -0.109
!
ATOM 6 COMMENT C403 c-c1Cc1-h     DATE Fri Apr 22 15:46:37 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 3
 PVAL    3.844   KAPPA    1.006   KPRIM    0.913           SIGPV    0.091
 PLMS 3  1  -0.143   PLMS 3 -1  -0.196   PLMS 3  3   0.175   PLMS 3 -3  -0.056
 PLMS 4  0   0.027   PLMS 4  2  -0.019   PLMS 4 -2   0.079   PLMS 4  4   0.037
 PLMS 4 -4   0.033
!
ATOM 6 COMMENT C404 c-Ch2-c     DATE Fri Apr 22 15:46:38 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM mm2
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 2
 PVAL    3.762   KAPPA    1.008   KPRIM    0.908           SIGPV    0.104
 PLMS 1  0   0.028   PLMS 3  2   0.322   PLMS 4  0  -0.086   PLMS 4  4   0.062
!
ATOM 6 COMMENT C405 c1-Ch2-c2     DATE Fri Apr 22 15:46:41 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 2
 PVAL    3.691   KAPPA    1.014   KPRIM    0.922           SIGPV    0.085
 PLMS 3  1  -0.127   PLMS 3 -1  -0.196   PLMS 3  3   0.179   PLMS 3 -3  -0.033
 PLMS 4  0   0.016   PLMS 4  2  -0.036   PLMS 4 -2   0.079   PLMS 4  4   0.024
 PLMS 4 -4   0.027
!
ATOM 6 COMMENT C406 c-Ch2-c  cycloene     DATE Fri Apr 22 15:46:42 2011
 NEIG 4 TYPE 2 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 2
 PVAL    3.683   KAPPA    1.014   KPRIM    0.922           SIGPV    0.066
 PLMS 1  1   0.027   PLMS 2  2   0.017   PLMS 3  1  -0.231   PLMS 3  3  -0.171
 PLMS 4  0   0.019   PLMS 4  2   0.075   PLMS 4  4  -0.036
!
ATOM 6 COMMENT C407 c-Cc2-c  cycloene  TEMP2    DATE Fri Apr 22 15:46:43 2011
 NEIG 4 TYPE 1 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 6 NUMB 4
 PVAL    3.980   KAPPA    0.997   KPRIM    0.903           SIGPV    0.002
 PLMS 1  1   0.052   PLMS 2  0  -0.040   PLMS 2 -2   0.010   PLMS 3  1  -0.266
 PLMS 3  3  -0.203   PLMS 3 -3   0.015   PLMS 4  0   0.023   PLMS 4  2   0.095
 PLMS 4  4  -0.057
!
!------ one hetero atom ----------------------------------------------------
!
ATOM 6 COMMENT C410 n-Ch3 n4     DATE Fri Apr 22 15:46:43 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM 3m
 TYPE 1 ATOM 1 NUMB 3
 TYPE 2 ATOM 7 NUMB 1
 EXCEPT GROUP ammonium
 PVAL    3.526   KAPPA    1.026   KPRIM    0.973           SIGPV    0.158
 PLMS 1  0  -0.064   PLMS 2  0  -0.151   PLMS 3  0   0.210   PLMS 3  3   0.183
 PLMS 4  0   0.070   PLMS 4  3  -0.066
!
```

```
ATOM 6 COMMENT C411 n-c1Cc1-c n4  TEMP2    DATE Fri Apr 22 15:46:43 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM m   PLAN f
 TYPE 1 ATOM 6 NUMB 3
 TYPE 2 ATOM 7 NUMB 1
 EXCEPT GROUP ammonium
 PVAL    3.859   KAPPA    1.007   KPRIM    0.948           SIGPV    0.008
 PLMS 1  1  -0.123   PLMS 1 -1   0.046   PLMS 2  0   0.073   PLMS 2  2  -0.175
 PLMS 3  1  -0.116   PLMS 3 -1  -0.187   PLMS 3  3   0.126   PLMS 3 -3  -0.056
 PLMS 4  0   0.030   PLMS 4  2  -0.021   PLMS 4 -2   0.065   PLMS 4  4   0.051
 PLMS 4 -4   0.016
!
ATOM 6 COMMENT C412 n-Ch2-c n4     DATE Fri Apr 22 15:46:43 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 7 NUMB 1
 EXCEPT GROUP ammonium
 PVAL    3.687   KAPPA    1.016   KPRIM    0.951           SIGPV    0.089
 PLMS 1  1  -0.087   PLMS 2  0   0.082   PLMS 2  2  -0.128   PLMS 3  1  -0.119
 PLMS 3 -1  -0.202   PLMS 3  3   0.160   PLMS 3 -3  -0.034   PLMS 4  0   0.026
 PLMS 4  2  -0.046   PLMS 4 -2   0.076   PLMS 4  4   0.041   PLMS 4 -4   0.021
!
ATOM 6 COMMENT C413 n-c1Cc1-h n4    DATE Fri Apr 22 15:46:43 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m   PLAN f
 TYPE 1 ATOM 1 NUMB 1
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 7 NUMB 1
 EXCEPT GROUP ammonium
 PVAL    3.818   KAPPA    1.007   KPRIM    0.949           SIGPV    0.056
 PLMS 1  1  -0.107   PLMS 1 -1  -0.014   PLMS 2  0   0.085   PLMS 2  2  -0.133
 PLMS 3  1  -0.125   PLMS 3 -1  -0.194   PLMS 3  3   0.146   PLMS 3 -3  -0.067
 PLMS 4  0   0.030   PLMS 4  2  -0.019   PLMS 4 -2   0.067   PLMS 4  4   0.049
 PLMS 4 -4   0.030
!
ATOM 6 COMMENT C414 n-c2Ch-c1 n4 CHIRAL    DATE Fri Apr 22 15:46:43 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM no
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 TYPE 3 ATOM 7 NUMB 1
 EXCEPT GROUP ammonium
 PVAL    3.775   KAPPA    1.011   KPRIM    0.945           SIGPV    0.053
 PLMS 1  1  -0.111   PLMS 2  0   0.072   PLMS 2  2  -0.155   PLMS 3  1  -0.117
 PLMS 3 -1  -0.198   PLMS 3  3   0.148   PLMS 3 -3  -0.044   PLMS 4  0   0.031
 PLMS 4  2  -0.038   PLMS 4 -2   0.076   PLMS 4  4   0.046   PLMS 4 -4   0.020
!
ATOM 6 COMMENT C416 n-Ch3 n3planar     DATE Fri Apr 22 15:46:44 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM 3m
 TYPE 1 ATOM 1 NUMB 3
 TYPE 2 ATOM 7 NUMB 1
 EXCEPT GROUP planar amine
 PVAL    3.584   KAPPA    1.023   KPRIM    0.949           SIGPV    0.088
 PLMS 1  0  -0.038   PLMS 2  0  -0.098   PLMS 3  0   0.251   PLMS 3  3   0.186
 PLMS 4  0   0.069   PLMS 4  3  -0.073
!
ATOM 6 COMMENT C417 n-Cc3 n3planar     DATE Fri Apr 22 15:46:45 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM 3m
 TYPE 1 ATOM 6 NUMB 3
 TYPE 2 ATOM 7 NUMB 1
 EXCEPT GROUP planar amine
 PVAL    4.012   KAPPA    0.995   KPRIM    0.896           SIGPV    0.090
 PLMS 1  0  -0.066   PLMS 2  0  -0.094   PLMS 3  0   0.289   PLMS 3  3   0.236
 PLMS 4  0   0.093   PLMS 4  3  -0.103
!
```

```
ATOM 6 COMMENT C418 n-Ch2-c n3planar     DATE Fri Apr 22 15:46:45 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 7 NUMB 1
 EXCEPT GROUP planar amine
 PVAL    3.747   KAPPA    1.012   KPRIM    0.926           SIGPV    0.078
 PLMS 1  1  -0.049   PLMS 1 -1   0.016   PLMS 2  0   0.052   PLMS 2  2  -0.077
 PLMS 3  1  -0.147   PLMS 3 -1  -0.210   PLMS 3  3   0.196   PLMS 3 -3  -0.043
 PLMS 4  0   0.025   PLMS 4  2  -0.052   PLMS 4 -2   0.083   PLMS 4  4   0.048
 PLMS 4 -4   0.022
!
ATOM 6 COMMENT C419 n-Ch2-c n3planar in ring    DATE Fri Apr 22 15:46:46 2011
 NEIG 4 TYPE 3 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 7 NUMB 1
 EXCEPT GROUP planar amine
 PVAL    3.736   KAPPA    1.014   KPRIM    0.925           SIGPV    0.053
 PLMS 1 -1  -0.044   PLMS 2  0   0.061   PLMS 2  2   0.046   PLMS 2 -2  -0.056
 PLMS 3  1  -0.267   PLMS 3  3  -0.187   PLMS 3 -3   0.025   PLMS 4  0   0.023
 PLMS 4  2   0.094   PLMS 4 -2  -0.011   PLMS 4  4  -0.056   PLMS 4 -4  -0.005
!
ATOM 6 COMMENT C420 n-cCc-h n3planar     DATE Fri Apr 22 15:46:46 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 TYPE 3 ATOM 7 NUMB 1
 EXCEPT GROUP planar amine
 PVAL    3.882   KAPPA    1.004   KPRIM    0.915           SIGPV    0.064
 PLMS 1  1  -0.055   PLMS 2  0   0.046   PLMS 2  2  -0.081   PLMS 3  1  -0.164
 PLMS 3 -1  -0.211   PLMS 3  3   0.187   PLMS 3 -3  -0.068   PLMS 4  0   0.035
 PLMS 4  2  -0.032   PLMS 4 -2   0.084   PLMS 4  4   0.064   PLMS 4 -4   0.034
!
ATOM 6 COMMENT C421 n-c2Ch-c1 n3planar  CHIRAL     DATE Fri Apr 22 15:46:46 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM no
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 TYPE 3 ATOM 7 NUMB 1
 EXCEPT GROUP planar amine
 PVAL    3.868   KAPPA    1.007   KPRIM    0.914           SIGPV    0.041
 PLMS 1  1  -0.037   PLMS 2  0   0.050   PLMS 2  2  -0.047   PLMS 3  1  -0.155
 PLMS 3 -1  -0.214   PLMS 3  3   0.202   PLMS 3 -3  -0.037   PLMS 4  0   0.039
 PLMS 4  2  -0.053   PLMS 4 -2   0.087   PLMS 4  4   0.052   PLMS 4 -4   0.032
!
ATOM 6 COMMENT C422 n-Ch3 n3pyramidal     DATE Fri Apr 22 15:46:46 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM 3m
 TYPE 1 ATOM 1 NUMB 3
 TYPE 2 ATOM 7 NUMB 1
 EXCEPT GROUP pyramidal amine
 PVAL    3.626   KAPPA    1.022   KPRIM    0.929           SIGPV    0.084
 PLMS 1  0  -0.025   PLMS 2  0  -0.086   PLMS 3  0   0.271   PLMS 3  3   0.203
 PLMS 4  0   0.080   PLMS 4  3  -0.086
!
ATOM 6 COMMENT C423 n-Ch2-c n3pyramidal     DATE Fri Apr 22 15:46:47 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 7 NUMB 1
 EXCEPT GROUP pyramidal amine
 PVAL    3.796   KAPPA    1.009   KPRIM    0.906           SIGPV    0.063
 PLMS 1  1  -0.019   PLMS 2  0   0.033   PLMS 2  2  -0.044   PLMS 3  1  -0.167
 PLMS 3 -1  -0.216   PLMS 3  3   0.212   PLMS 3 -3  -0.053   PLMS 4  0   0.030
 PLMS 4  2  -0.050   PLMS 4 -2   0.096   PLMS 4  4   0.055   PLMS 4 -4   0.028
!
```

```
ATOM 6 COMMENT C424 n-cCc-h n3pyramidal     DATE Fri Apr 22 15:46:47 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 TYPE 3 ATOM 7 NUMB 1
 EXCEPT GROUP pyramidal amine
 PVAL    3.947   KAPPA    1.002   KPRIM    0.892           SIGPV    0.075
 PLMS 1  1  -0.029   PLMS 2  0   0.025   PLMS 2  2  -0.050   PLMS 3  1  -0.172
 PLMS 3 -1  -0.223   PLMS 3  3   0.217   PLMS 3 -3  -0.065   PLMS 4  0   0.038
 PLMS 4  2  -0.044   PLMS 4 -2   0.107   PLMS 4  4   0.069   PLMS 4 -4   0.043
!
ATOM 6 COMMENT C425 n-c2Ch-c1 n3pyramidal  CHIRAL     DATE Fri Apr 22 15:46:48 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM no
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 TYPE 3 ATOM 7 NUMB 1
 EXCEPT GROUP pyramidal amine
 PVAL    3.881   KAPPA    1.005   KPRIM    0.904           SIGPV    0.045
 PLMS 2  0   0.026   PLMS 2  2  -0.048   PLMS 3  1  -0.160   PLMS 3 -1  -0.214
 PLMS 3  3   0.208   PLMS 3 -3  -0.047   PLMS 4  0   0.032   PLMS 4  2  -0.052
 PLMS 4 -2   0.098   PLMS 4  4   0.061   PLMS 4 -4   0.026
!
ATOM 6 COMMENT C427 o-Ch3     DATE Fri Apr 22 15:46:48 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM 3m   PLAN f
 TYPE 1 ATOM 1 NUMB 3
 TYPE 2 ATOM 8 NUMB 1
 PVAL    3.596   KAPPA    1.020   KPRIM    0.947           SIGPV    0.067
 PLMS 1  0  -0.065   PLMS 2  0  -0.129   PLMS 3  0   0.264   PLMS 3  3   0.200
 PLMS 4  0   0.075   PLMS 4  3  -0.080
!
ATOM 6 COMMENT C428 o-Cc3    DATE Fri Apr 22 15:46:49 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM 3m   PLAN f
 TYPE 1 ATOM 6 NUMB 3
 TYPE 2 ATOM 8 NUMB 1
 PVAL    3.962   KAPPA    0.995   KPRIM    0.901           SIGPV    0.032
 PLMS 1  0  -0.080   PLMS 2  0  -0.136   PLMS 3  0   0.284   PLMS 3  3   0.237
 PLMS 4  0   0.104   PLMS 4  3  -0.096
!
ATOM 6 COMMENT C429 o-Ch2-c     DATE Fri Apr 22 15:46:49 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 8 NUMB 1
 PVAL    3.744   KAPPA    1.010   KPRIM    0.917           SIGPV    0.075
 PLMS 1  1  -0.066   PLMS 2  0   0.061   PLMS 2  2  -0.092   PLMS 3  1  -0.164
 PLMS 3 -1  -0.217   PLMS 3  3   0.210   PLMS 3 -3  -0.055   PLMS 4  0   0.030
 PLMS 4  2  -0.046   PLMS 4 -2   0.094   PLMS 4  4   0.054   PLMS 4 -4   0.032
!
ATOM 6 COMMENT C430 o-cCc-h     DATE Fri Apr 22 15:46:51 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 TYPE 3 ATOM 8 NUMB 1
 PVAL    3.872   KAPPA    1.004   KPRIM    0.905           SIGPV    0.072
 PLMS 1  1  -0.069   PLMS 2  0   0.041   PLMS 2  2  -0.094   PLMS 3  1  -0.165
 PLMS 3 -1  -0.222   PLMS 3  3   0.219   PLMS 3 -3  -0.051   PLMS 4  0   0.040
 PLMS 4  2  -0.051   PLMS 4 -2   0.100   PLMS 4  4   0.061   PLMS 4 -4   0.038
!
ATOM 6 COMMENT C431 o-c2Ch-c1 CHIRAL    DATE Fri Apr 22 15:46:53 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM no
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 TYPE 3 ATOM 8 NUMB 1
 PVAL    3.860   KAPPA    1.005   KPRIM    0.909           SIGPV    0.069
 PLMS 1  1  -0.065   PLMS 2  0   0.064   PLMS 2  2  -0.083   PLMS 3  1  -0.170
 PLMS 3 -1  -0.220   PLMS 3  3   0.207   PLMS 3 -3  -0.059   PLMS 4  0   0.034
 PLMS 4  2  -0.052   PLMS 4 -2   0.101   PLMS 4  4   0.059   PLMS 4 -4   0.027
```

```
ATOM 6 COMMENT C432 p-Ch2-c     DATE Fri Apr 22 15:46:53 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m PLAN f
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 15 NUMB 1
 PVAL    3.724   KAPPA    1.016   KPRIM    0.935           SIGPV    0.106
 PLMS 1  1   0.042   PLMS 2  2   0.038   PLMS 3  1  -0.102   PLMS 3 -1  -0.188
 PLMS 3  3   0.153   PLMS 3 -3  -0.025   PLMS 4  0   0.009   PLMS 4  2  -0.025
 PLMS 4 -2   0.073   PLMS 4 -4   0.024
!
ATOM 6 COMMENT C433 s-Ch3     DATE Fri Apr 22 15:46:54 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM 3m PLAN f
 TYPE 1 ATOM 1 NUMB 3
 TYPE 2 ATOM 16 NUMB 1
 PVAL    3.490   KAPPA    1.027   KPRIM    0.947           SIGPV    0.084
 PLMS 2  0  -0.071   PLMS 3  0   0.189   PLMS 3  3   0.176   PLMS 4  0   0.032
 PLMS 4  3  -0.068
!
ATOM 6 COMMENT C434 s-Ch2-c     DATE Fri Apr 22 15:46:54 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m PLAN f
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 16 NUMB 1
 PVAL    3.742   KAPPA    1.012   KPRIM    0.906           SIGPV    0.085
 PLMS 1  1  -0.041   PLMS 2  0   0.043   PLMS 2  2  -0.072   PLMS 3  1  -0.112
 PLMS 3 -1  -0.219   PLMS 3  3   0.171   PLMS 3 -3  -0.037   PLMS 4  0   0.019
 PLMS 4  2  -0.031   PLMS 4 -2   0.090   PLMS 4  4   0.025   PLMS 4 -4   0.036
!
ATOM 6 COMMENT C437 cl-Ch2-c     DATE Fri Apr 22 15:46:54 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m PLAN f
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 17 NUMB 1
 PVAL    3.990   KAPPA    0.997   KPRIM    0.882           SIGPV    0.072
 PLMS 1  0  -0.058   PLMS 2  0  -0.105   PLMS 3  0   0.225   PLMS 3  3   0.238
 PLMS 4  0   0.058   PLMS 4  3  -0.097
!
ATOM 6 COMMENT C438 i-Ch2-c     DATE Tue Apr 26 16:34:34 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m PLAN f
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 53 NUMB 1
 PVAL    3.817   KAPPA    1.004   KPRIM    0.945           SIGPV    0.093
 PLMS 1  1  -0.016   PLMS 2  0   0.060   PLMS 2  2  -0.085   PLMS 3  1  -0.104
 PLMS 3 -1  -0.199   PLMS 3  3   0.147   PLMS 3 -3  -0.036   PLMS 4  0   0.009
 PLMS 4  2  -0.024   PLMS 4 -2   0.081   PLMS 4  4   0.010   PLMS 4 -4   0.029
!
ATOM 6 COMMENT C439 c-cCs-c  DATE Fri Apr 22 15:46:54 2011
NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM 3m PLAN f
 TYPE 1 ATOM 6 NUMB 3
 TYPE 2 ATOM 16 NUMB 1
 PVAL    3.990   KAPPA    0.997   KPRIM    0.882           SIGPV    0.072
 PLMS 1  0  -0.058   PLMS 2  0  -0.105   PLMS 3  0   0.225   PLMS 3  3   0.238
 PLMS 4  0   0.058   PLMS 4  3  -0.097
!
!------ two hetero atoms ----------------------------------------------------
!
ATOM 6 COMMENT C440 c-oCo-h     DATE Fri Apr 22 15:46:54 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 8 NUMB 2
 PVAL    3.881   KAPPA    1.005   KPRIM    0.900           SIGPV    0.121
 PLMS 1  1   0.065   PLMS 1 -1   0.070   PLMS 2  0  -0.107   PLMS 2 -2  -0.060
 PLMS 3  1  -0.187   PLMS 3 -1  -0.254   PLMS 3  3   0.241   PLMS 3 -3  -0.059
 PLMS 4  0   0.041   PLMS 4  2  -0.058   PLMS 4 -2   0.153   PLMS 4  4   0.046
 PLMS 4 -4   0.040
```

```
ATOM 6 COMMENT C443 c1-oCo-c2    DATE Fri Apr 22 15:46:54 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 2 ATOM 6 NUMB 2
 TYPE 3 ATOM 8 NUMB 2
 PVAL    3.891   KAPPA    1.002   KPRIM    0.899           SIGPV    0.059
 PLMS 1 -1   0.056   PLMS 2  0  -0.082   PLMS 2 -2  -0.053   PLMS 3  1  -0.165
 PLMS 3 -1  -0.264   PLMS 3  3   0.227   PLMS 3 -3  -0.056   PLMS 4  0   0.038
 PLMS 4  2  -0.071   PLMS 4 -2   0.164   PLMS 4  4   0.041   PLMS 4 -4   0.030
!
ATOM 6 COMMENT C444 o-cCc-o    TEMP    DATE Fri Apr 22 15:46:54 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM mm2
 TYPE 2 ATOM 6 NUMB 2
 TYPE 3 ATOM 8 NUMB 2
 PVAL    3.916   KAPPA    1.001   KPRIM    0.891           SIGPV    0.138
 PLMS 1  0  -0.072   PLMS 2  2  -0.117   PLMS 3  2   0.415   PLMS 4  0  -0.158
 PLMS 4  2   0.048   PLMS 4  4   0.101
!
ATOM 6 COMMENT C446 p-cCh-n  CHIRAL    DATE Fri Apr 22 15:46:55 2011
 NEIG 4 TYPE 4 RING 0 MEMB 0 SYMM no
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 7 NUMB 1
 TYPE 4 ATOM 15 NUMB 1
 PVAL    3.813   KAPPA    1.012   KPRIM    0.906           SIGPV    0.042
 PLMS 2  0   0.038   PLMS 2 -2   0.035   PLMS 3  1  -0.125   PLMS 3 -1  -0.208
 PLMS 3  3   0.172   PLMS 3 -3  -0.080   PLMS 4  0   0.028   PLMS 4 -2   0.091
 PLMS 4  4   0.037   PLMS 4 -4   0.045
!
ATOM 6 COMMENT C447 p-cCh-o  CHIRAL    DATE Fri Apr 22 15:46:55 2011
 NEIG 4 TYPE 4 RING 0 MEMB 0 SYMM no
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 8 NUMB 1
 TYPE 4 ATOM 15 NUMB 1
 PVAL    3.729   KAPPA    1.015   KPRIM    0.921           SIGPV    0.080
 PLMS 1 -1  -0.063   PLMS 2  0   0.063   PLMS 2  2   0.061   PLMS 2 -2   0.047
 PLMS 3  0  -0.033   PLMS 3  1  -0.117   PLMS 3 -1  -0.219   PLMS 3  3   0.179
 PLMS 3 -3  -0.061   PLMS 4  0   0.049   PLMS 4 -2   0.095   PLMS 4  4   0.035
 PLMS 4 -4   0.051
!
ATOM 6 COMMENT C448 p-hCh-n    DATE Fri Apr 22 15:46:55 2011
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 7 NUMB 1
 TYPE 3 ATOM 15 NUMB 1
 PVAL    3.713   KAPPA    1.017   KPRIM    0.968           SIGPV    0.060
 PLMS 1  1   0.097   PLMS 1 -1  -0.050   PLMS 2  0   0.050   PLMS 2  2   0.121
 PLMS 2 -2   0.078   PLMS 3  1  -0.107   PLMS 3 -1  -0.164   PLMS 3  3   0.144
 PLMS 3 -3  -0.031   PLMS 4  0   0.010   PLMS 4  2  -0.006   PLMS 4 -2   0.078
 PLMS 4 -4   0.039
!
ATOM 6 COMMENT C450 o-cCh-n  CHIRAL    DATE Fri Apr 22 15:46:55 2011
 NEIG 4 TYPE 4 RING 0 MEMB 0 SYMM no
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 7 NUMB 1
 TYPE 4 ATOM 8 NUMB 1
 PVAL    3.837   KAPPA    1.009   KPRIM    0.915           SIGPV    0.066
 PLMS 1  1  -0.046   PLMS 1 -1  -0.070   PLMS 2  0   0.105   PLMS 2 -2   0.041
 PLMS 3  1  -0.177   PLMS 3 -1  -0.225   PLMS 3  3   0.210   PLMS 3 -3  -0.066
 PLMS 4  0   0.047   PLMS 4  2  -0.034   PLMS 4 -2   0.105   PLMS 4  4   0.061
 PLMS 4 -4   0.042
!
```

```
ATOM 6 COMMENT C452 h-sCn-h    DATE Tue Apr 07 11:10:35 2009
 NEIG 4 TYPE 3 RING 0 MEMB 0 SYMM m PLAN f
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 7 NUMB 1
 TYPE 3 ATOM 16 NUMB 1
 PVAL    3.783   KAPPA    1.013   KPRIM    0.918           SIGPV    0.053
 PLMS 1  1  -0.015   PLMS 2  0   0.072   PLMS 2  2  -0.036   PLMS 2 -2   0.038
 PLMS 3  1  -0.118   PLMS 3 -1  -0.233   PLMS 3  3   0.174   PLMS 3 -3  -0.053
 PLMS 4  0   0.026   PLMS 4  2  -0.019   PLMS 4 -2   0.096   PLMS 4  4   0.035
 PLMS 4 -4   0.047
!
ATOM 6 COMMENT C453 c-sCn-h    DATE Tue Apr 07 11:10:35 2009
 NEIG 4 TYPE 4 RING 0 MEMB 0 SYMM no PLAN f
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 1
 TYPE 3 ATOM 16 NUMB 1
 TYPE 4 ATOM 7 NUMB 1
 PVAL    3.904   KAPPA    1.004   KPRIM    0.891           SIGPV    0.054
 PLMS 1  1  -0.038   PLMS 2  0   0.067   PLMS 2  2  -0.053   PLMS 2 -2   0.033
 PLMS 3  1  -0.120   PLMS 3 -1  -0.254   PLMS 3  3   0.178   PLMS 4  0   0.029
 PLMS 4 -2   0.107   PLMS 4  4   0.035   PLMS 4 -4   0.045
!
ATOM 6 COMMENT C455 f-cCc-f    DATE Tue Apr 07 11:10:35 2009
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM mm2 PLAN f
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 9 NUMB 2
 PVAL    3.788   KAPPA    1.009   KPRIM    0.930           SIGPV    0.047
 PLMS 1  0  -0.114   PLMS 2  2  -0.125   PLMS 3  2   0.382   PLMS 4  0  -0.121
 PLMS 4  4   0.091
!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
!  N I T R O G E N
!
!
!-------- set of N-x4 --------------------------------------------------------
!
ATOM 7 COMMENT N401 c-Nh3     DATE Fri Apr 22 15:46:56 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM 3m PLAN f
 TYPE 1 ATOM 1 NUMB 3
 TYPE 2 ATOM 6 NUMB 1
 PVAL    5.104   KAPPA    0.990   KPRIM    0.811           SIGPV    0.056
 PLMS 1  0   0.042   PLMS 3  0   0.268   PLMS 3  3   0.212   PLMS 4  0   0.048
 PLMS 4  3  -0.095
!
ATOM 7 COMMENT N402 c-Nh2-c     DATE Fri Apr 22 15:46:56 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM mm2 PLAN f
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 2
 PVAL    5.150   KAPPA    0.986   KPRIM    0.800           SIGPV    0.051
 PLMS 1  0   0.032   PLMS 2  2   0.011   PLMS 3  2   0.342   PLMS 4  0  -0.089
 PLMS 4  2  -0.014   PLMS 4  4   0.063
!
ATOM 7 COMMENT N403 c1-Nh2-c2     DATE Fri Apr 22 15:46:56 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM m PLAN f
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 2
 PVAL    5.166   KAPPA    0.985   KPRIM    0.802           SIGPV    0.019
 PLMS 1  1   0.020   PLMS 1 -1   0.030   PLMS 2 -2  -0.007   PLMS 3  1  -0.150
 PLMS 3 -1  -0.220   PLMS 3  3   0.205   PLMS 3 -3  -0.043   PLMS 4  0   0.014
 PLMS 4  2  -0.039   PLMS 4 -2   0.101   PLMS 4  4   0.028   PLMS 4 -4   0.029
!
```

```
ATOM 7 COMMENT N404 c3-Nh     DATE Fri Apr 22 15:46:56 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM 3m PLAN f
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 3
 PVAL    5.198   KAPPA    0.981   KPRIM    0.784           SIGPV    0.038
 PLMS 1  0  -0.024   PLMS 2  0  -0.019   PLMS 3  0   0.277   PLMS 3  3   0.236
 PLMS 4  0   0.088   PLMS 4  3  -0.094
!
!-------- set of N-x3 pyramidal ---------------------------------------------
!
ATOM 7 COMMENT N301 c-Nc-c pyramidal     DATE Fri Apr 22 15:46:57 2011
 NEIG 3 TYPE 1 RING 0 MEMB 0 SYMM 3m PLAN f
 TYPE 1 ATOM 6 NUMB 3
 PVAL    5.103   KAPPA    0.986   KPRIM    1.043           SIGPV    0.019
 PLMS 1  0  -0.107   PLMS 2  0   0.141   PLMS 3  0  -0.059   PLMS 3  3   0.107
 PLMS 4  0  -0.004   PLMS 4  3   0.037
!
ATOM 7 COMMENT N302 c-Nh-h pyramidal     DATE Fri Apr 22 15:46:57 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM m PLAN f
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 1
 PVAL    4.917   KAPPA    0.999   KPRIM    1.015           SIGPV    0.054
 PLMS 1  1  -0.033   PLMS 1 -1  -0.067   PLMS 2  0  -0.044   PLMS 2  2  -0.052
 PLMS 2 -2   0.076   PLMS 3  1   0.128   PLMS 3 -1  -0.026   PLMS 3 -3   0.024
 PLMS 4  2   0.029   PLMS 4 -2   0.011   PLMS 4  4  -0.003
!
ATOM 7 COMMENT N303 h-Nc-c pyramidal    DATE Fri Apr 22 15:46:57 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM m PLAN f
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 PVAL    5.061   KAPPA    0.989   KPRIM    1.028           SIGPV    0.019
 PLMS 1  1  -0.066   PLMS 1 -1  -0.086   PLMS 2  0  -0.073   PLMS 2  2  -0.044
 PLMS 2 -2   0.089   PLMS 3  1   0.109   PLMS 3 -1  -0.036   PLMS 3  3   0.052
 PLMS 3 -3   0.018   PLMS 4  2   0.030   PLMS 4 -2   0.018   PLMS 4  4  -0.005
 PLMS 4 -4   0.012
!
ATOM 7 COMMENT N304 c-Ns-c    DATE Fri Apr 22 15:46:57 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM m PLAN f
 TYPE 1 ATOM 16 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 PVAL    5.053   KAPPA    0.994   KPRIM    1.026           SIGPV    0.036
 PLMS 1  1  -0.026   PLMS 1 -1  -0.068   PLMS 2  0  -0.056   PLMS 2  2  -0.070
 PLMS 2 -2   0.098   PLMS 3  1   0.116   PLMS 3 -1  -0.021   PLMS 3  3   0.021
 PLMS 3 -3   0.021   PLMS 4  0   0.007   PLMS 4  2   0.021   PLMS 4 -2   0.023
 PLMS 4  4  -0.011   PLMS 4 -4  -0.008
!
ATOM 7 COMMENT N306 c-Nc-c    DATE Fri Apr 22 15:46:57 2011
 NEIG 3 TYPE 1 RING 0 MEMB 0 SYMM m PLAN f
 TYPE 1 ATOM 6 NUMB 3
 PVAL    5.020   KAPPA    0.993   KPRIM    1.056           SIGPV    0.037
 PLMS 1  1  -0.016   PLMS 1 -1  -0.038   PLMS 2  0  -0.047   PLMS 2  2  -0.065
 PLMS 2 -2   0.040   PLMS 3  1   0.112   PLMS 3 -1  -0.018   PLMS 3 -3   0.028
 PLMS 4  2   0.018   PLMS 4 -2   0.011
!
!-------- set of N-x3 planar ------------------------------------------------
!
ATOM 7 COMMENT N310 c--Nh-h planar     DATE Fri Apr 22 15:46:57 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM mm2 PLAN t
 TYPE 1 ATOM 1 NUMB 2
 TYPE 2 ATOM 6 NUMB 1
 PVAL    4.894   KAPPA    1.003   KPRIM    1.039           SIGPV    0.056
 PLMS 2  0  -0.050   PLMS 2  2  -0.066   PLMS 3  0   0.102   PLMS 3  2  -0.094
 PLMS 4  0  -0.006   PLMS 4  4   0.009
!
```

```
ATOM 7 COMMENT N311 h-Nc--c planar     DATE Fri Apr 22 15:46:58 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM mm2 PLAN t
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 PVAL    4.948   KAPPA    0.999   KPRIM    1.076           SIGPV    0.053
 PLMS 2  0  -0.042   PLMS 2  2  -0.066   PLMS 3  0  -0.104   PLMS 3  2   0.068
 PLMS 4  0   0.013   PLMS 4  4   0.006
!
ATOM 7 COMMENT N312 c-No--o nitro     DATE Fri Apr 22 15:46:59 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM mm2 PLAN t
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 8 NUMB 2
 PVAL    5.022   KAPPA    0.987   KPRIM    0.763           SIGPV    0.023
 PLMS 2  0   0.075   PLMS 2  2   0.202   PLMS 3  0   0.348   PLMS 3  2  -0.268
 PLMS 4  0   0.028   PLMS 4  2   0.023   PLMS 4  4   0.058
!
ATOM 7 COMMENT N313 c--Nc1-c1 c1=csp3 planar     DATE Fri Apr 22 15:46:59 2011
 NEIG 3 TYPE 1 RING 0 MEMB 0 SYMM mm2 PLAN t
 TYPE 1 ATOM 6 NUMB 3
 EXCEPT GROUP csp2sp3sp3
 PVAL    5.037   KAPPA    0.993   KPRIM    1.042           SIGPV    0.044
 PLMS 2  0  -0.034   PLMS 2  2  -0.071   PLMS 3  0   0.097   PLMS 3  2  -0.098
 PLMS 4  0  -0.006   PLMS 4  2   0.007   PLMS 4  4   0.009
!
ATOM 7 COMMENT N314 c--Nc1-c1 c1=csp2 planar    DATE Fri Apr 22 15:46:59 2011
 NEIG 3 TYPE 1 RING 0 MEMB 0 SYMM mm2 PLAN t
 TYPE 1 ATOM 6 NUMB 3
 EXCEPT GROUP csp3sp2sp2
 PVAL    5.014   KAPPA    0.994   KPRIM    1.057           SIGPV    0.047
 PLMS 2  0  -0.048   PLMS 2  2  -0.069   PLMS 3  0   0.113   PLMS 3  2  -0.067
 PLMS 4  0   0.011   PLMS 4  4   0.006
!
ATOM 7 COMMENT N315 c1--Nh-c2 planar     DATE Fri Apr 22 15:47:00 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM m PLAN t
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 PVAL    5.010   KAPPA    0.996   KPRIM    1.034           SIGPV    0.049
 PLMS 2  0   0.087   PLMS 3  1   0.016   PLMS 3  3   0.129   PLMS 4  0   0.012
 PLMS 4  2   0.005   PLMS 4  4  -0.006   PLMS 4 -4  -0.006
!
!-------- set of N-x2 planar ------------------------------------------------
!
ATOM 7 COMMENT N202 c-N-oh    DATE Fri Apr 22 15:47:00 2011
 NEIG 2 TYPE 2 RING 0 MEMB 0 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 8 NUMB 1
 EXCEPT GROUP N-OH
 PVAL    4.982   KAPPA    0.991   KPRIM    1.020           SIGPV    0.012
 PLMS 1  1  -0.076   PLMS 1 -1  -0.168   PLMS 2  0  -0.065   PLMS 2  2   0.066
 PLMS 2 -2   0.132   PLMS 3  1   0.007   PLMS 3 -1  -0.048   PLMS 3  3   0.104
 PLMS 4  0   0.005   PLMS 4  2   0.022   PLMS 4 -2   0.007   PLMS 4 -4   0.048
!
ATOM 7 COMMENT N203 c-N-n  TEMP2    DATE Fri Apr 22 15:47:00 2011
 NEIG 2 TYPE 2 RING 0 MEMB 0 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 7 NUMB 1
 PVAL    5.043   KAPPA    0.989   KPRIM    1.055           SIGPV    0.036
 PLMS 1  1  -0.080   PLMS 1 -1  -0.124   PLMS 2  0  -0.041   PLMS 2  2   0.026
 PLMS 2 -2   0.099   PLMS 3  1   0.009   PLMS 3 -1  -0.040   PLMS 3  3   0.080
 PLMS 4  0   0.004   PLMS 4  2   0.016   PLMS 4 -2   0.009   PLMS 4  4   0.013
 PLMS 4 -4   0.033
!
```

```
ATOM 7 COMMENT N204 c-N-c    DATE Fri Apr 22 15:47:00 2011
 NEIG 2 TYPE 1 RING 0 MEMB 0 SYMM mm2 PLAN t
 TYPE 1 ATOM 6 NUMB 2
 PVAL    5.059   KAPPA    0.989   KPRIM    1.035           SIGPV    0.035
 PLMS 1  0  -0.129   PLMS 2  0   0.079   PLMS 3  0  -0.046   PLMS 3  2  -0.066
 PLMS 4  0  -0.015   PLMS 4  2  -0.013
!
ATOM 7 COMMENT N205 c-N-c    DATE Fri Apr 22 15:47:00 2011
 NEIG 2 TYPE 1 RING 0 MEMB 0 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 2
 PVAL    5.072   KAPPA    0.987   KPRIM    1.055           SIGPV    0.044
 PLMS 1  1  -0.064   PLMS 1 -1  -0.110   PLMS 2  0  -0.060   PLMS 2 -2   0.068
 PLMS 3 -1  -0.027   PLMS 3  3   0.073   PLMS 4  2   0.011   PLMS 4  4   0.010
 PLMS 4 -4   0.023
!
!-------- set of N-x planar ------------------------------------------------
!
ATOM 7 COMMENT N101 -c=-N nitrile    DATE Fri Apr 22 15:47:00 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM cyl PLAN t
 TYPE 1 ATOM 6 NUMB 1
 PVAL    4.771   KAPPA    0.993   KPRIM    1.066           SIGPV    0.124
 PLMS 1  0  -0.121   PLMS 2  0   0.157   PLMS 3  0   0.015   PLMS 4  0  -0.034
!
!-------- set of N aromatic ------------------------------------------------
!
ATOM 7 COMMENT N320 c-=Nh=-c AROMATIC     DATE Fri Apr 22 15:47:00 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 PVAL    4.929   KAPPA    1.001   KPRIM    1.031           SIGPV    0.059
 PLMS 2  0   0.046   PLMS 3  1   0.016   PLMS 3  3  -0.137   PLMS 4  0   0.005
 PLMS 4  4   0.010
!
ATOM 7 COMMENT N321 c-=Nh=-c AROMATIC     DATE Fri Apr 22 15:47:01 2011
 NEIG 3 TYPE 2 RING 1 MEMB 5 SYMM m PLAN t
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 2
 PVAL    5.026   KAPPA    0.995   KPRIM    0.947           SIGPV    0.075
 PLMS 1  1   0.026   PLMS 3  1   0.018   PLMS 3 -1   0.009   PLMS 3  3  -0.172
 PLMS 4  2  -0.012
!
ATOM 7 COMMENT N322 c1-=Nc=-c1 AROMATIC     DATE Fri Apr 22 15:47:01 2011
 NEIG 3 TYPE 1 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 3
 EXCEPT GROUP csp3sp2sp2
 PVAL    5.034   KAPPA    0.994   KPRIM    0.961           SIGPV    0.091
 PLMS 3  1   0.019   PLMS 3  3  -0.171
!
ATOM 7 COMMENT N323 c1-=Nc=-c1 AROMATIC     DATE Fri Apr 22 15:47:02 2011
 NEIG 3 TYPE 1 RING 1 MEMB 5 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 3
 EXCEPT GROUP csp3sp2sp2
 PVAL    5.107   KAPPA    0.990   KPRIM    0.877           SIGPV    0.088
 PLMS 2  2   0.013   PLMS 3  1   0.022   PLMS 3 -1   0.007   PLMS 3  3  -0.209
 PLMS 4  2  -0.013   PLMS 4  4  -0.014
!
ATOM 7 COMMENT N325 c-=Nc=-c  AROMATIC    DATE Fri Apr 22 15:47:03 2011
 NEIG 3 TYPE 1 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 3
 PVAL    5.140   KAPPA    0.988   KPRIM    0.838           SIGPV    0.070
 PLMS 3  1   0.027   PLMS 3  3  -0.249   PLMS 4  2  -0.013
!
```

```
ATOM 7 COMMENT N326 c-=Nc=-n  AROMATIC  TEMP2    DATE Fri Apr 22 15:47:03 2011
 NEIG 3 TYPE 2 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 2
 TYPE 2 ATOM 7 NUMB 1
 PVAL    4.993   KAPPA    0.998   KPRIM    1.027           SIGPV    0.019
 PLMS 1  1  -0.007   PLMS 1 -1  -0.021   PLMS 2  0   0.079   PLMS 2  2   0.017
 PLMS 2 -2  -0.019   PLMS 3 -1  -0.014   PLMS 3  3  -0.146   PLMS 3 -3   0.011
 PLMS 4  0   0.012   PLMS 4 -2  -0.009   PLMS 4 -4  -0.021
!
ATOM 7 COMMENT N328 c-=Nh=-n  AROMATIC    DATE Fri Apr 22 15:47:03 2011
 NEIG 3 TYPE 3 RING 1 MEMB 5 SYMM m PLAN t
 TYPE 1 ATOM 1 NUMB 1
 TYPE 1 ATOM 6 NUMB 1
 TYPE 1 ATOM 7 NUMB 1
 PVAL    5.005   KAPPA    0.999   KPRIM    0.993           SIGPV    0.012
 PLMS 1 -1   0.007   PLMS 2  0   0.055   PLMS 2  2   0.022   PLMS 3  1   0.006
 PLMS 3 -1  -0.028   PLMS 3  3  -0.165   PLMS 3 -3   0.016   PLMS 4  0   0.009
 PLMS 4  2  -0.004   PLMS 4 -2  -0.017   PLMS 4  4  -0.013   PLMS 4 -4  -0.017
!
ATOM 7 COMMENT N210 c-=N=-c AROMATIC     DATE Fri Apr 22 15:47:03 2011
 NEIG 2 TYPE 1 RING 1 MEMB 6 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 2
 PVAL    5.006   KAPPA    0.990   KPRIM    1.045           SIGPV    0.058
 PLMS 1  1  -0.135   PLMS 2  0  -0.077   PLMS 2  2   0.077   PLMS 3  1  -0.022
 PLMS 3 -1   0.006   PLMS 3  3  -0.074   PLMS 4  0  -0.004   PLMS 4  4  -0.021
!
ATOM 7 COMMENT N211 c-=N=-c AROMATIC     DATE Fri Apr 22 15:47:04 2011
 NEIG 2 TYPE 1 RING 1 MEMB 5 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 2
 PVAL    5.035   KAPPA    0.987   KPRIM    1.051           SIGPV    0.049
 PLMS 1  1  -0.130   PLMS 2  0  -0.075   PLMS 2  2   0.081   PLMS 2 -2   0.012
 PLMS 3  1  -0.024   PLMS 3 -1   0.010   PLMS 3  3  -0.068   PLMS 4  0  -0.003
 PLMS 4 -2   0.006   PLMS 4  4  -0.023
!
ATOM 7 COMMENT N212 c-=N=-o  AROMATIC    DATE Fri Apr 22 15:47:04 2011
 NEIG 2 TYPE 2 RING 1 MEMB 5 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 1
 TYPE 1 ATOM 8 NUMB 1
 PVAL    5.006   KAPPA    0.990   KPRIM    1.077           SIGPV    0.040
 PLMS 1  1  -0.161   PLMS 2  0   0.020   PLMS 2  2   0.089   PLMS 2 -2  -0.125
 PLMS 3  1  -0.026   PLMS 3 -1  -0.019   PLMS 3  3  -0.073   PLMS 3 -3   0.004
 PLMS 4  0   0.004   PLMS 4 -2  -0.018   PLMS 4  4  -0.042   PLMS 4 -4  -0.008
!
ATOM 7 COMMENT N213 c-=N=-n  AROMATIC    DATE Fri Apr 22 15:47:04 2011
 NEIG 2 TYPE 2 RING 1 MEMB 5 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 1
 TYPE 1 ATOM 7 NUMB 1
 PVAL    5.034   KAPPA    0.989   KPRIM    1.058           SIGPV    0.039
 PLMS 1  1  -0.150   PLMS 2  0  -0.049   PLMS 2  2   0.093   PLMS 2 -2  -0.064
 PLMS 3  1  -0.030   PLMS 3 -1  -0.017   PLMS 3  3  -0.079   PLMS 3 -3   0.006
 PLMS 4 -2  -0.016   PLMS 4  4  -0.035   PLMS 4 -4  -0.009
!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
!  O X Y G E N
!
!
!-------- set of O--x --------------------------------------------------------
!
ATOM 8 COMMENT O101 -c-cOO[-] carboxylate     DATE Fri Apr 22 15:47:05 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 EXCEPT GROUP carboxylate
 PVAL    6.321   KAPPA    0.977   KPRIM    1.162           SIGPV    0.051
 PLMS 1  1  -0.088   PLMS 1 -1   0.009   PLMS 2  0  -0.042   PLMS 2  2  -0.074
 PLMS 2 -2   0.008   PLMS 3  1  -0.007   PLMS 3  3   0.018   PLMS 3 -3   0.003
 PLMS 4  4   0.008
```

```
ATOM 8 COMMENT O102 c-c(=O)-oh carboxylic     DATE Fri Apr 22 15:47:05 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 EXCEPT GROUP carboxylic
 PVAL    6.123   KAPPA    0.988   KPRIM    1.157           SIGPV    0.042
 PLMS 1  1  -0.098   PLMS 2  0  -0.073   PLMS 2  2  -0.072   PLMS 3  1  -0.006
 PLMS 3  3   0.014   PLMS 4  0  -0.005   PLMS 4  2   0.004   PLMS 4  4   0.005
!
ATOM 8 COMMENT O103 c-c(=O)-o-c ester     DATE Fri Apr 22 15:47:06 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 EXCEPT GROUP ester
 PVAL    6.111   KAPPA    0.988   KPRIM    1.159           SIGPV    0.030
 PLMS 1  1  -0.098   PLMS 1 -1  -0.005   PLMS 2  0  -0.076   PLMS 2  2  -0.069
 PLMS 2 -2  -0.003   PLMS 3  1  -0.005   PLMS 3  3   0.016   PLMS 4  0  -0.006
 PLMS 4  2   0.004   PLMS 4  4   0.003
!
ATOM 8 COMMENT O104 c-c(=O)-o-c aromatester    DATE Fri Apr 22 15:47:07 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 EXCEPT GROUP aromatester
 PVAL    6.096   KAPPA    0.989   KPRIM    1.159           SIGPV    0.018
 PLMS 1  1  -0.100   PLMS 1 -1  -0.004   PLMS 2  0  -0.066   PLMS 2  2  -0.071
 PLMS 2 -2  -0.002   PLMS 3  1  -0.005   PLMS 3 -1  -0.003   PLMS 3  3   0.008
 PLMS 4  0  -0.004   PLMS 4  2   0.002   PLMS 4  4   0.007
!
ATOM 8 COMMENT O105 c-c(=O)-n amide     DATE Fri Apr 22 15:47:07 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 EXCEPT GROUP amide
 PVAL    6.178   KAPPA    0.985   KPRIM    1.151           SIGPV    0.032
 PLMS 1  1  -0.101   PLMS 1 -1  -0.005   PLMS 2  0  -0.062   PLMS 2  2  -0.082
 PLMS 2 -2  -0.002   PLMS 3  1  -0.005   PLMS 3  3   0.019   PLMS 4  0  -0.003
 PLMS 4  4   0.006
!
ATOM 8 COMMENT O106 c-c(=O)-n aromatamide    DATE Fri Apr 22 15:47:08 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 EXCEPT GROUP aromatamide
 PVAL    6.133   KAPPA    0.987   KPRIM    1.134           SIGPV    0.043
 PLMS 1  1  -0.103   PLMS 1 -1  -0.005   PLMS 2  0  -0.068   PLMS 2  2  -0.090
 PLMS 3  1  -0.003   PLMS 3 -1   0.002   PLMS 3  3   0.014   PLMS 4  0  -0.004
 PLMS 4  2   0.002   PLMS 4  4   0.006
!
ATOM 8 COMMENT O199 x-c(=O)-x     TEMP     DATE Fri Apr 22 15:47:08 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 PVAL    6.071   KAPPA    0.990   KPRIM    1.166           SIGPV    0.036
 PLMS 1  1  -0.100   PLMS 1 -1  -0.004   PLMS 2  0  -0.067   PLMS 2  2  -0.067
 PLMS 3  1  -0.003   PLMS 3 -1  -0.002   PLMS 3  3   0.016   PLMS 4  0  -0.005
 PLMS 4  2   0.003   PLMS 4  4   0.003
!
ATOM 8 COMMENT O107 n-c(=O)-o-c(4) carbamate ester    DATE Fri Apr 22 15:47:08 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 EXCEPT GROUP carbamate ester
 PVAL    6.127   KAPPA    0.988   KPRIM    1.178           SIGPV    0.034
 PLMS 1  1  -0.100   PLMS 2  0  -0.043   PLMS 2  2  -0.070   PLMS 3  1  -0.006
 PLMS 3  3   0.014   PLMS 4  0  -0.004   PLMS 4  2   0.002   PLMS 4  4   0.005
!
ATOM 8 COMMENT O108  n-c(=O)-n urea     DATE Fri Apr 22 15:47:09 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 EXCEPT GROUP urea
 PVAL    6.164   KAPPA    0.986   KPRIM    1.160           SIGPV    0.054
 PLMS 1  1  -0.101   PLMS 2  0  -0.040   PLMS 2  2  -0.080   PLMS 3  1  -0.005
 PLMS 3  3   0.013   PLMS 4  0  -0.003   PLMS 4  4   0.009
```

```
ATOM 8 COMMENT O109  n-c(=O)-n aromaturea     DATE Fri Apr 22 15:47:09 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 EXCEPT GROUP aromaturea
 PVAL    6.152   KAPPA    0.986   KPRIM    1.166           SIGPV    0.063
 PLMS 1  1  -0.102   PLMS 2  0  -0.048   PLMS 2  2  -0.081   PLMS 3  1  -0.004
 PLMS 3  3   0.012   PLMS 4  0  -0.003   PLMS 4  4   0.008
!
ATOM 8 COMMENT O110  >c=O    DATE Fri Apr 22 15:47:09 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 EXCEPT GROUP ketone
 PVAL    6.114   KAPPA    0.989   KPRIM    1.128           SIGPV    0.045
 PLMS 1  1  -0.100   PLMS 2  0  -0.118   PLMS 2  2  -0.084   PLMS 3  3   0.020
 PLMS 4  0  -0.008   PLMS 4  2   0.005   PLMS 4  4   0.004
!
ATOM 8 COMMENT O113  n--O nitro2     DATE Fri Apr 22 15:47:09 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 7 NUMB 1
 EXCEPT GROUP NO2
 PVAL    6.124   KAPPA    0.990   KPRIM    1.130           SIGPV    0.033
 PLMS 1  1  -0.120   PLMS 2  0  -0.048   PLMS 2  2  -0.155   PLMS 3  1  -0.009
 PLMS 3  3   0.017   PLMS 3 -3   0.006   PLMS 4  2  -0.002   PLMS 4  4   0.013
!
ATOM 8 COMMENT O114  p--O in c-pO3     DATE Fri Apr 22 15:47:10 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 15 NUMB 1
 EXCEPT GROUP cPo3
 PVAL    6.451   KAPPA    0.969   KPRIM    1.170           SIGPV    0.064
 PLMS 1  1  -0.063   PLMS 2  0   0.016   PLMS 2  2  -0.036   PLMS 3  1  -0.007
 PLMS 3  3   0.010   PLMS 4  0   0.003   PLMS 4  2  -0.004   PLMS 4  4   0.006
!
ATOM 8 COMMENT O115  p--O in c-pO2-o-x     DATE Fri Apr 22 15:47:10 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 15 NUMB 1
 EXCEPT GROUP cPo2(o-x)
 PVAL    6.234   KAPPA    0.980   KPRIM    1.213           SIGPV    0.035
 PLMS 1  1  -0.072   PLMS 1 -1   0.007   PLMS 2  0   0.019   PLMS 2  2  -0.031
 PLMS 2 -2   0.012   PLMS 3  1  -0.003   PLMS 3  3   0.003   PLMS 4  4   0.004
!
ATOM 8 COMMENT O116  p--O in c-pO-(o-x)2     DATE Fri Apr 22 15:47:10 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 15 NUMB 1
 EXCEPT GROUP cPo(o-x)2
 PVAL    6.079   KAPPA    0.988   KPRIM    1.197           SIGPV    0.027
 PLMS 1  1  -0.084   PLMS 2  0   0.017   PLMS 2  2  -0.028   PLMS 3  1  -0.002
 PLMS 4  4   0.002
!
ATOM 8 COMMENT O117  p--O in c3-pO    DATE Fri Apr 22 15:47:10 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 15 NUMB 1
 EXCEPT GROUP C3PO
 PVAL    6.097   KAPPA    0.987   KPRIM    1.167           SIGPV    0.018
 PLMS 1  1  -0.089   PLMS 2  0   0.022   PLMS 2  2  -0.039   PLMS 3  1  -0.003
 PLMS 3  3   0.003   PLMS 4  4   0.002
!
ATOM 8 COMMENT O119  p--O in O-p(ox)3    DATE Fri Apr 22 15:47:11 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 15 NUMB 1
 EXCEPT GROUP O-p(ox)3
 PVAL    6.109   KAPPA    0.986   KPRIM    1.213           SIGPV    0.077
 PLMS 1  1  -0.080   PLMS 2  0   0.016   PLMS 2  2  -0.034   PLMS 4  0   0.002
 PLMS 4  2  -0.003   PLMS 4  4   0.004
!
```

```
ATOM 8 COMMENT O120  p--O in o-pO-(ox)2     DATE Fri Apr 22 15:47:10 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 15 NUMB 1
 PVAL    6.261   KAPPA    0.978   KPRIM    1.206           SIGPV    0.054
 PLMS 1  1  -0.073   PLMS 1 -1   0.005   PLMS 2  0   0.021   PLMS 2  2  -0.035
 PLMS 2 -2   0.011   PLMS 3  3   0.005   PLMS 4  0   0.002   PLMS 4  2  -0.002
 PLMS 4  4   0.005
!
ATOM 8 COMMENT O121  p--O in o2-pO-ox     DATE Fri Apr 22 15:47:11 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 15 NUMB 1
 EXCEPT GROUP O3-p(ox)
 PVAL    6.422   KAPPA    0.971   KPRIM    1.225           SIGPV    0.086
 PLMS 1  1  -0.060   PLMS 2  0   0.018   PLMS 2  2  -0.033   PLMS 3  1  -0.003
 PLMS 3  3   0.006   PLMS 4  0   0.003   PLMS 4  2  -0.004   PLMS 4  4   0.006
!
ATOM 8 COMMENT O122 s=O    DATE Fri Apr 22 15:47:11 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 16 NUMB 1
 PVAL    6.309   KAPPA    0.980   KPRIM    1.314           SIGPV    0.092
 PLMS 1  1  -0.064   PLMS 2  0   0.033   PLMS 2  2  -0.052   PLMS 3  1  -0.004
 PLMS 3  3   0.005   PLMS 4  0   0.003   PLMS 4  2  -0.004   PLMS 4  4   0.005
!
!-------- set of x--O--x ----------------------------------------------------
!
ATOM 8 COMMENT O201 c-c(=o)-O-h carboxylic     DATE Fri Apr 22 15:47:12 2011
 NEIG 2 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 1
 EXCEPT GROUP carboxylic
 PVAL    6.167   KAPPA    0.989   KPRIM    1.191           SIGPV    0.030
 PLMS 1  1  -0.049   PLMS 1 -1  -0.074   PLMS 2  0   0.062   PLMS 2  2  -0.015
 PLMS 2 -2   0.063   PLMS 3 -1  -0.015   PLMS 3  3   0.039   PLMS 3 -3  -0.008
 PLMS 4  0   0.007   PLMS 4 -2   0.004   PLMS 4  4   0.012   PLMS 4 -4   0.012
!
ATOM 8 COMMENT O202 c-c(=o)-O-c(4) ester     DATE Fri Apr 22 15:47:12 2011
 NEIG 2 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 2
 EXCEPT GROUP ester
 PVAL    6.188   KAPPA    0.988   KPRIM    1.178           SIGPV    0.026
 PLMS 1  1  -0.047   PLMS 1 -1  -0.078   PLMS 2  0   0.063   PLMS 2  2  -0.021
 PLMS 2 -2   0.064   PLMS 3 -1  -0.016   PLMS 3  3   0.041   PLMS 4  0   0.007
 PLMS 4 -2   0.004   PLMS 4  4   0.008   PLMS 4 -4   0.011
!
ATOM 8 COMMENT O203 n-c(=o)-O-c(4) carbamate ester     DATE Fri Apr 22 15:47:13 2011
 NEIG 2 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 2
 EXCEPT GROUP carbamate ester
 PVAL    6.191   KAPPA    0.988   KPRIM    1.186           SIGPV    0.027
 PLMS 1  1  -0.052   PLMS 1 -1  -0.073   PLMS 2  0   0.075   PLMS 2  2  -0.028
 PLMS 2 -2   0.060   PLMS 3  1  -0.002   PLMS 3 -1  -0.014   PLMS 3  3   0.038
 PLMS 4  0   0.007   PLMS 4 -2   0.004   PLMS 4  4   0.009   PLMS 4 -4   0.011
!
ATOM 8 COMMENT O204 c-O-h phenol     DATE Fri Apr 22 15:47:13 2011
 NEIG 2 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 1
 EXCEPT GROUP phenol
 PVAL    6.172   KAPPA    0.988   KPRIM    1.184           SIGPV    0.050
 PLMS 1  1  -0.060   PLMS 1 -1  -0.080   PLMS 2  0   0.089   PLMS 2  2  -0.026
 PLMS 2 -2   0.051   PLMS 3  1  -0.006   PLMS 3 -1  -0.015   PLMS 3  3   0.040
 PLMS 3 -3  -0.010   PLMS 4  0   0.007   PLMS 4 -2   0.003   PLMS 4  4   0.012
 PLMS 4 -4   0.011
!
!
```

```
ATOM 8 COMMENT O205 c-O-h alcohol     DATE Fri Apr 22 15:47:13 2011
 NEIG 2 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 1
 EXCEPT GROUP alcohol
 PVAL    6.237   KAPPA    0.984   KPRIM    1.123           SIGPV    0.028
 PLMS 1  1  -0.060   PLMS 1 -1  -0.093   PLMS 2  0   0.127   PLMS 2  2  -0.028
 PLMS 2 -2   0.053   PLMS 3  1  -0.009   PLMS 3 -1  -0.017   PLMS 3  3   0.045
 PLMS 3 -3  -0.014   PLMS 4  0   0.011   PLMS 4 -2   0.003   PLMS 4  4   0.015
 PLMS 4 -4   0.013
!
ATOM 8 COMMENT O206 c1-O-c2 ether m     DATE Fri Apr 22 15:47:13 2011
 NEIG 2 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 2
 PVAL    6.212   KAPPA    0.987   KPRIM    1.186           SIGPV    0.028
 PLMS 1  1  -0.056   PLMS 1 -1  -0.082   PLMS 2  0   0.092   PLMS 2  2  -0.031
 PLMS 2 -2   0.056   PLMS 3  1  -0.005   PLMS 3 -1  -0.013   PLMS 3  3   0.038
 PLMS 4  0   0.008   PLMS 4 -2   0.004   PLMS 4  4   0.009   PLMS 4 -4   0.012
!
ATOM 8 COMMENT O207 c-O-c ether mm2     DATE Fri Apr 22 15:47:15 2011
 NEIG 2 TYPE 1 RING 0 MEMB 0 SYMM mm2
 TYPE 1 ATOM 6 NUMB 2
 PVAL    6.248   KAPPA    0.984   KPRIM    1.154           SIGPV    0.024
 PLMS 1  0  -0.100   PLMS 2  2   0.132   PLMS 3  0  -0.025   PLMS 3  2  -0.039
 PLMS 4  0  -0.012   PLMS 4  2  -0.016   PLMS 4  4   0.007
!
ATOM 8 COMMENT O208 p-O-h     DATE Fri Apr 22 15:47:16 2011
 NEIG 2 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 1 ATOM 15 NUMB 1
 PVAL    6.247   KAPPA    0.986   KPRIM    1.167           SIGPV    0.040
 PLMS 1  1  -0.017   PLMS 1 -1  -0.074   PLMS 2  0   0.083   PLMS 2  2   0.017
 PLMS 2 -2   0.069   PLMS 3 -1  -0.015   PLMS 3  3   0.033   PLMS 4  0   0.007
 PLMS 4  2   0.003   PLMS 4 -2   0.005   PLMS 4  4   0.005   PLMS 4 -4   0.015
!
ATOM 8 COMMENT O209 c-O-p     DATE Fri Apr 22 15:47:17 2011
 NEIG 2 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 1 ATOM 15 NUMB 1
 PVAL    6.245   KAPPA    0.987   KPRIM    1.184           SIGPV    0.036
 PLMS 1  1  -0.054   PLMS 1 -1  -0.049   PLMS 2  0   0.081   PLMS 2  2  -0.058
 PLMS 2 -2   0.030   PLMS 3  1  -0.013   PLMS 3 -1  -0.004   PLMS 3  3   0.030
 PLMS 3 -3  -0.004   PLMS 4  0   0.007   PLMS 4  2  -0.005   PLMS 4  4   0.009
 PLMS 4 -4   0.008
!
ATOM 8 COMMENT O210 p-O-p     DATE Fri Apr 22 15:47:18 2011
 NEIG 2 TYPE 1 RING 0 MEMB 0 SYMM mm2
 TYPE 1 ATOM 15 NUMB 2
 PVAL    6.260   KAPPA    0.989   KPRIM    1.308           SIGPV    0.054
 PLMS 1  0  -0.035   PLMS 2  0   0.014   PLMS 2  2   0.057   PLMS 3  0  -0.011
 PLMS 3  2  -0.012   PLMS 4  0  -0.003   PLMS 4  2  -0.003   PLMS 4  4   0.004
!
ATOM 8 COMMENT O211 n-O-h    DATE Fri Apr 22 15:47:18 2011
 NEIG 2 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 1 ATOM 7 NUMB 1
 PVAL    6.215   KAPPA    0.987   KPRIM    1.142           SIGPV    0.032
 PLMS 1  1  -0.072   PLMS 1 -1  -0.086   PLMS 2  0   0.127   PLMS 2  2  -0.053
 PLMS 2 -2   0.050   PLMS 3  1  -0.012   PLMS 3 -1  -0.014   PLMS 3  3   0.040
 PLMS 3 -3  -0.027   PLMS 4  0   0.011   PLMS 4  2  -0.004   PLMS 4 -2   0.002
 PLMS 4  4   0.026   PLMS 4 -4   0.008
!
```

```
ATOM 8 COMMENT O001 in water     DATE Fri Apr 22 15:47:18 2011
 NEIG 2 TYPE 1 RING 0 MEMB 0 SYMM mm2
 TYPE 1 ATOM 1 NUMB 2
 PVAL    6.345   KAPPA    0.979   KPRIM    1.116           SIGPV    0.006
 PLMS 1  0  -0.126   PLMS 2  0  -0.017   PLMS 2  2   0.134   PLMS 3  0  -0.016
 PLMS 3  2  -0.031   PLMS 4  0  -0.032   PLMS 4  2  -0.032   PLMS 4  4   0.009
!
!-------- set of O aromatic --------------------------------------------------
!
ATOM 8 COMMENT O230 c-=O=-c  AROMATIC  TEMP2    DATE Fri Apr 22 15:47:18 2011
 NEIG 2 TYPE 1 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 6 NUMB 2
 PVAL    6.175   KAPPA    0.990   KPRIM    1.215           SIGPV    0.011
 PLMS 1  1  -0.086   PLMS 2  0   0.049   PLMS 2  2   0.063   PLMS 3  1  -0.009
 PLMS 3  3  -0.034   PLMS 3 -3   0.005   PLMS 4  0   0.005   PLMS 4  2   0.006
 PLMS 4  4  -0.012
!
ATOM 8 COMMENT O231 c-=O=-c  AROMATIC    DATE Fri Apr 22 15:47:18 2011
 NEIG 2 TYPE 1 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 6 NUMB 2
 PVAL    6.154   KAPPA    0.990   KPRIM    1.241           SIGPV    0.038
 PLMS 1  1  -0.082   PLMS 1 -1  -0.005   PLMS 2  0   0.056   PLMS 2  2   0.060
 PLMS 2 -2  -0.005   PLMS 3  1  -0.010   PLMS 3  3  -0.030   PLMS 4  0   0.005
 PLMS 4  4  -0.018   PLMS 4 -4   0.002
!
ATOM 8 COMMENT O232 c-=O=-n  AROMATIC    DATE Fri Apr 22 15:47:18 2011
 NEIG 2 TYPE 2 RING 1 MEMB 5 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 1 ATOM 7 NUMB 1
 PVAL    6.119   KAPPA    0.990   KPRIM    1.179           SIGPV    0.031
 PLMS 1  1  -0.103   PLMS 1 -1   0.006   PLMS 2  0   0.070   PLMS 2  2   0.073
 PLMS 2 -2  -0.027   PLMS 3  1  -0.017   PLMS 3 -1  -0.005   PLMS 3  3  -0.045
 PLMS 3 -3   0.006   PLMS 4  0   0.007   PLMS 4 -2  -0.004   PLMS 4  4  -0.021
 PLMS 4 -4  -0.003
!
ATOM 8 COMMENT O233 c-c(=o)-O-c(3) ester     DATE Fri Apr 22 15:47:18 2011
 NEIG 2 TYPE 1 RING 1 MEMB 6 SYMM m
 TYPE 1 ATOM 6 NUMB 2
 EXCEPT GROUP ester
 PVAL    6.167   KAPPA    0.990   KPRIM    1.228           SIGPV    0.021
 PLMS 1  1  -0.080   PLMS 1 -1  -0.015   PLMS 2  0   0.042   PLMS 2  2   0.061
 PLMS 2 -2  -0.017   PLMS 3  1  -0.007   PLMS 3  3  -0.031   PLMS 3 -3   0.002
 PLMS 4  0   0.006   PLMS 4  2   0.005   PLMS 4  4  -0.013
!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
!  S U L P H U R
!
!
!-------- set of S--x --------------------------------------------------------
!
ATOM 16 COMMENT S101 >c=S thio m  TEMP3    DATE Fri Apr 22 15:47:18 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 PVAL    6.406   KAPPA    0.978   KPRIM    0.981           SIGPV    0.048
 PLMS 1  1  -0.055   PLMS 2  2  -0.171   PLMS 3  1  -0.050   PLMS 3  3   0.112
 PLMS 3 -3  -0.006   PLMS 4  0   0.011   PLMS 4  2  -0.038   PLMS 4  4   0.066
!
ATOM 16 COMMENT S103 c-c(=S)-n thioPh m TEMP2    DATE Fri Apr 22 15:47:19 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 EXCEPT GROUP aromathioamide
 PVAL    6.329   KAPPA    0.984   KPRIM    0.997           SIGPV    0.0
 PLMS 1  1  -0.048   PLMS 2  0  -0.040   PLMS 2  2  -0.160   PLMS 3  1  -0.049
 PLMS 3 -1  -0.006   PLMS 3  3   0.121   PLMS 4  0   0.021   PLMS 4  2  -0.033
 PLMS 4  4   0.071
!
```

```
ATOM 16 COMMENT S104 n-c(=S)-n thioPh m  TEMP2    DATE Fri Apr 22 15:47:19 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 EXCEPT GROUP PHthiourea
 PVAL    6.362   KAPPA    0.983   KPRIM    1.016           SIGPV    0.010
 PLMS 1  1  -0.053   PLMS 2  2  -0.156   PLMS 3  1  -0.044   PLMS 3 -1  -0.006
 PLMS 3  3   0.097   PLMS 4  0   0.024   PLMS 4  2  -0.038   PLMS 4  4   0.061
!
!-------- set of x--S--x ----------------------------------------------------
!
ATOM 16 COMMENT S201 c-S-h thiol    DATE Fri Apr 22 15:47:19 2011
 NEIG 2 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 1 NUMB 1
 TYPE 2 ATOM 6 NUMB 1
 PVAL    6.374   KAPPA    0.982   KPRIM    1.044           SIGPV    0.026
 PLMS 1  1  -0.034   PLMS 1 -1  -0.038   PLMS 2  0   0.152   PLMS 2  2  -0.008
 PLMS 2 -2   0.050   PLMS 3  1  -0.090   PLMS 3 -1  -0.102   PLMS 3  3   0.174
 PLMS 3 -3  -0.126   PLMS 4  0   0.076   PLMS 4  2  -0.006   PLMS 4 -2   0.012
 PLMS 4  4   0.122   PLMS 4 -4   0.034
!
ATOM 16 COMMENT S202 c-S-c thioether mm2    DATE Fri Apr 22 15:47:19 2011
 NEIG 2 TYPE 1 RING 0 MEMB 0 SYMM mm2 PLAN t
 TYPE 1 ATOM 6 NUMB 2
 PVAL    6.378   KAPPA    0.981   KPRIM    1.030           SIGPV    0.050
 PLMS 1  0  -0.052   PLMS 2  0  -0.035   PLMS 2  2   0.155   PLMS 3  0  -0.102
 PLMS 3  2  -0.236   PLMS 4  0  -0.083   PLMS 4  2  -0.115   PLMS 4  4   0.043
!
ATOM 16 COMMENT S203 c-S-s     DATE Fri Apr 22 15:47:19 2011
 NEIG 2 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 16 NUMB 1
 PVAL    6.309   KAPPA    0.986   KPRIM    0.999           SIGPV    0.033
 PLMS 1  1  -0.030   PLMS 1 -1  -0.048   PLMS 2  0   0.183   PLMS 2  2   0.018
 PLMS 2 -2   0.072   PLMS 3  1  -0.087   PLMS 3 -1  -0.064   PLMS 3  3   0.203
 PLMS 3 -3  -0.060   PLMS 4  0   0.064   PLMS 4  2  -0.029   PLMS 4 -2   0.014
 PLMS 4  4   0.094   PLMS 4 -4   0.050
!
ATOM 16 COMMENT S204 c-S-c thioether m     DATE Fri Apr 22 15:47:19 2011
 NEIG 2 TYPE 1 RING 0 MEMB 0 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 2
 PVAL    6.369   KAPPA    0.981   KPRIM    1.054           SIGPV    0.009
 PLMS 1  1  -0.035   PLMS 1 -1  -0.034   PLMS 2  0   0.126   PLMS 2 -2   0.048
 PLMS 3  1  -0.059   PLMS 3 -1  -0.097   PLMS 3  3   0.175   PLMS 3 -3  -0.102
 PLMS 4  0   0.065   PLMS 4  2   0.007   PLMS 4 -2   0.011   PLMS 4  4   0.106
!
!-------- set of S--x4 -----------------------------------------------------
!
ATOM 16 COMMENT S404 o-nSn-o   TEMP3    DATE Fri Apr 22 15:47:19 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM mm2 PLAN f
 TYPE 1 ATOM 7 NUMB 2
 TYPE 1 ATOM 8 NUMB 2
 PVAL    5.911   KAPPA    0.965   KPRIM    1.039           SIGPV    0.029
 PLMS 1  0   0.004   PLMS 2  0  -0.017   PLMS 2  2   0.115   PLMS 3  0  -0.085
 PLMS 3  2   0.599   PLMS 4  0  -0.218   PLMS 4  2  -0.124   PLMS 4  4   0.211
!
```

```
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
!  P H O S P H O R U S
!
!
!-------- set of P-x4 --------------------------------------------------------
!
ATOM 15 COMMENT P401 c-Po3    DATE Fri Apr 22 15:47:19 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM 3m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 8 NUMB 3
 EXCEPT GROUP cPo3
 PVAL    5.312   KAPPA    0.932   KPRIM    1.059   SIGPV    0.042
 PLMS 1  0  -0.097   PLMS 2  0  -0.171   PLMS 3  0   0.410   PLMS 3  3   0.449
 PLMS 4  0   0.143   PLMS 4  3  -0.166
!
ATOM 15 COMMENT P402 x-o-Po2-c     DATE Fri Apr 22 15:47:19 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 8 NUMB 3
 EXCEPT GROUP cPo2(o-x)
PVAL    5.297   KAPPA    0.942   KPRIM    1.038   SIGPV    0.031
 PLMS 1  1  -0.124   PLMS 1 -1  -0.116   PLMS 2  0   0.284   PLMS 2 -2   0.098
 PLMS 3  1  -0.368   PLMS 3 -1  -0.439   PLMS 3  3   0.266   PLMS 3 -3  -0.108
 PLMS 4  0   0.108   PLMS 4  2  -0.070   PLMS 4 -2   0.169   PLMS 4  4   0.118
 PLMS 4 -4   0.062
!
ATOM 15 COMMENT P403 o-P(o-x)2-c     DATE Fri Apr 22 15:47:19 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 TYPE 2 ATOM 8 NUMB 3
 EXCEPT GROUP cPo(o-x)2
 PVAL    5.285   KAPPA    0.947   KPRIM    1.037   SIGPV    0.010
 PLMS 1  1   0.251   PLMS 1 -1   0.058   PLMS 2  0  -0.161   PLMS 2  2   0.217
 PLMS 2 -2  -0.045   PLMS 3  1  -0.380   PLMS 3 -1  -0.377   PLMS 3  3   0.467
 PLMS 3 -3  -0.078   PLMS 4  0   0.033   PLMS 4  2  -0.050   PLMS 4 -2   0.261
 PLMS 4  4   0.075   PLMS 4 -4   0.084
!
ATOM 15 COMMENT P404 o-Pc3  TEMP2    DATE Fri Apr 22 15:47:19 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM 3m
 TYPE 1 ATOM 6 NUMB 3
 TYPE 2 ATOM 8 NUMB 1
 PVAL    5.409   KAPPA    0.942   KPRIM    1.034           SIGPV    0.006
 PLMS 1  0   0.151   PLMS 2  0   0.227   PLMS 3  0   0.620   PLMS 3  3   0.365
 PLMS 4  0   0.083   PLMS 4  3  -0.222
!
ATOM 15 COMMENT P405 o-Pc3    DATE Fri Apr 22 15:47:19 2011
 NEIG 4 TYPE 2 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 3
 TYPE 2 ATOM 8 NUMB 1
 PVAL    5.417   KAPPA    0.944   KPRIM    1.021           SIGPV    0.025
 PLMS 1  1   0.166   PLMS 2  0  -0.107   PLMS 2  2   0.216   PLMS 3  1  -0.385
 PLMS 3 -1  -0.380   PLMS 3  3   0.463   PLMS 3 -3  -0.092   PLMS 4  0   0.043
 PLMS 4  2  -0.044   PLMS 4 -2   0.217   PLMS 4  4   0.056   PLMS 4 -4   0.073
!
ATOM 15 COMMENT P407 o2-P(ox)2     DATE Fri Apr 22 15:47:19 2011
 NEIG 4 TYPE 1 RING 0 MEMB 0 SYMM mm2
 TYPE 1 ATOM 8 NUMB 4
 PVAL    5.208   KAPPA    0.945   KPRIM    1.045           SIGPV    0.066
 PLMS 1  0   0.214   PLMS 2  0  -0.051   PLMS 2  2   0.306   PLMS 3  0  -0.133
 PLMS 3  2   0.648   PLMS 4  0  -0.196   PLMS 4  2  -0.060   PLMS 4  4   0.171
!
```

```
ATOM 15 COMMENT P408 o-P(ox)3    DATE Fri Apr 22 15:47:19 2011
 NEIG 4 TYPE 1 RING 0 MEMB 0 SYMM 3m
 TYPE 1 ATOM 8 NUMB 4
 PVAL    5.177   KAPPA    0.949   KPRIM    1.040           SIGPV    0.059
 PLMS 1  0   0.246   PLMS 2  0   0.296   PLMS 3  0   0.638   PLMS 3  3   0.376
 PLMS 4  0   0.097   PLMS 4  3  -0.287
!
ATOM 15 COMMENT P409 o3-P(ox)    DATE Fri Apr 22 15:47:20 2011
 NEIG 4 TYPE 1 RING 0 MEMB 0 SYMM 3m
 TYPE 1 ATOM 8 NUMB 4
 EXCEPT GROUP o3-P(ox)
 PVAL    5.250   KAPPA    0.935   KPRIM    1.078           SIGPV    0.035
 PLMS 1  0  -0.130   PLMS 2  0  -0.234   PLMS 3  0   0.373   PLMS 3  3   0.436
 PLMS 4  0   0.137   PLMS 4  3  -0.150
!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
!  B O R O N
!
!
!-----------------------------------------------------------------------------
!
ATOM 5 COMMENT B01 c-B-o2     DATE Fri Apr 22 15:47:20 2011
 NEIG 3 TYPE 2 RING 0 MEMB 0 SYMM mm2
 TYPE 1 ATOM 6 NUMB 1
 TYPE 1 ATOM 8 NUMB 2
 PVAL    2.881   KAPPA    1.030   KPRIM    0.999           SIGPV    0.047
 PLMS 1  1  -0.029   PLMS 2  0   0.140   PLMS 2  2   0.240   PLMS 3  0   0.320
 PLMS 4  0  -0.066   PLMS 4  2   0.064   PLMS 4  4   0.011
!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
!  I O D I N E
!
!
!-----------------------------------------------------------------------------
!
ATOM 53 COMMENT I01 c-I     DATE Tue Apr 26 16:34:36 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 PVAL    7.186   KAPPA    1.003   KPRIM    0.490           SIGPV    0.024
 PLMS 1  1  -0.066   PLMS 2  0   0.109   PLMS 2  2  -0.180   PLMS 3  1  -0.046
 PLMS 3  3   0.058   PLMS 4  0   0.014   PLMS 4  2  -0.017   PLMS 4  4   0.022
!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
!  C H L O R I N E
!
!
!-----------------------------------------------------------------------------
!
ATOM 17 COMMENT CL01 c-Cl*     DATE Thu Apr 21 18:14:09 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 PVAL    7.243   KAPPA    0.992   KPRIM    0.938           SIGPV    0.038
 PLMS 1  1  -0.034   PLMS 2  0   0.081   PLMS 2  2  -0.157   PLMS 3  1  -0.050
 PLMS 3  3  -0.063   PLMS 4  0   0.009   PLMS 4  2  -0.014   PLMS 4  4   0.016
!
ATOM 17 COMMENT CL02 c-Cl*     DATE Thu Apr 21 18:14:09 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 EXCEPT GROUP Ph-Cl
 PVAL    7.242   KAPPA    0.992   KPRIM    0.943           SIGPV    0.008
 PLMS 1  1  -0.046   PLMS 2  0   0.077   PLMS 2  2  -0.155   PLMS 3  1  -0.011
 PLMS 3  3   0.072   PLMS 4  0   0.009   PLMS 4  2  -0.015   PLMS 4  4   0.019
!
```

```
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
!  F L U O R I N E
!
!
!-----------------------------------------------------------------------------
!
ATOM 9 COMMENT F01 c-F     DATE Fri Apr 22 15:47:20 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m
 TYPE 1 ATOM 6 NUMB 1
 PVAL    7.185   KAPPA    0.991   KPRIM    1.229           SIGPV    0.031
 PLMS 1  1  -0.064   PLMS 2  0   0.058   PLMS 2  2  -0.088   PLMS 2 -2   0.005
 PLMS 3  1  -0.011   PLMS 3  3   0.010   PLMS 4  0   0.004   PLMS 4  2  -0.006
 PLMS 4  4   0.006
!
ATOM 9 COMMENT F02 c-F     DATE Fri Apr 22 15:47:20 2011
 NEIG 1 TYPE 1 RING 0 MEMB 0 SYMM m PLAN t
 TYPE 1 ATOM 6 NUMB 1
 EXCEPT GROUP Ph-F
 PVAL    7.191   KAPPA    0.991   KPRIM    1.200           SIGPV    0.023
 PLMS 1  1  -0.064   PLMS 2  0   0.053   PLMS 2  2  -0.096   PLMS 3  1  -0.008
 PLMS 3  3   0.008   PLMS 4  0   0.005   PLMS 4  2  -0.006   PLMS 4  4   0.008
!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
```

Table B4. A list of CSD structures (Cambridge Structural Databank REFCODEs) used for the databank construction. Additional number following a REFCODE suggests that the initial chemical system was decomposed into molecules and then calculated. Some of the molecules were partially used or excluded when modelling particular atom types.

ABEBUE
ABELEZ
ABIMUT
ABIZAM
ABUGOT
ABUQOD
ACANAC11
ACEZIR99
ACGLUA11
ACHIST20
ACIMDC99
ACUZAZ99
ACYHXA01
ADAGOC
ADALAT
ADENCH99
ADIMIK
ADPOSD99
ADUNAO
ADUWIF
ADUXAY
AEPHOS02
AFALIC
AFCYDP99
AFIFOK
CATXOK98
CATXOK99
CEVYEH95
CEVYEH96
CEVYEH97
CEVYEH98
CEVYEH99
CEYCIS99
CIBTUD
CIBVEP
CIDHON
CIXWEL
COHNES99
COJMUJ
COMXAD97
CUJQED
CYTIAC
CYTIDI11
CYTOSC99
DAFTEK99
DAQHOS98
DAQHOS99
DAQYAV99
DASYEC99
DAYWEF
GLYCPH98
GLYCPH99
GOJPAW98
GOJPAW99
GOJPUQ99
GOPZIU98
GOPZIU99
GUNQUB97
GUNQUB99
HAFHIF97
HAFHIF98
HAFHIF99
HAMZIF
HEGFIJ95
HEGFIJ96
HEGFIJ97
HEGFIJ98
HEGFIJ99
HIDXIB
HIFQES
HISAPH97
HISAPH98
HISAPH99
HISTAN
HISTCM12
LAVHAR
LAXTIO98
LAXTIO99
LAXYOY98
LAXYOY99
LESJEY
LESJOI
LIDYEC
LIQWEN
LOMTEM98
LOMTEM99
LOPXUJ
LOPYAQ
LTYRGG98
LTYRGG99
LUFBOD99
MABZAR99
MADCMP99
MAJGAG99
MALAMS98
MALAMS99
MAPNUN
MEDMEO
MHPTEC98
MHPTEC99
QUBVUE99
QULVEY98
QULVEY99
RADGAF98
RADGAF99
RADKOX98
RADKOX99
RAFBAB
RAJXUW97
RAJXUW98
RAJXUW99
RAXKAD96
RAXKAD97
RAXKAD98
RAXKAD99
RBFRPZ10
RISCUR
RIWLAK99
RODFAR99
RUWFEU
SADMEQ99
SADMUG99
SAHBUZ99
SARZAM
SASBIX98
VUXKII
VUXKOO
VUXLAB
VUXLEF
VUYMAD98
VUYMAD99
WALWOW
WEBFIS
WEJBIW
WIPCAZ98
WIPCAZ99
WOBXAM
WOCHOL
WOCSAI99
WOMYEC
WOQZOR94
WOQZOR95
WOQZOR96
WOQZOR97
WOQZOR98
WOQZOR99
WOTZEK99
WOVXOU01
WUXBIA99
XANOPT99

AGOCII99	DEBHEX01	HISTPA10	MIPWEN98	SASBIX99	XATXEV98
AHUZEI	DHPMAD98	HISTPA10	MIQNEF98	SATCEW99	XATXEV99
AHUZIM	DHPMAD99	HIYDIC	MIQNEF99	SAVSOX99	XAZWOL99
AJACAP	DIHRAN	HIYJUU99	MIQNIJ98	SAYWAQ99	XEGLAW98
AJACIX	DIUREA05	HOPJUR	MIQNIJ99	SAZCOL	XEGLAW99
AKAXAL99	DIYHIC	HUHBIV98	MMCPUR96	SCGMPT99	XIHSEM96
AKIFEF	DOGBIK99	HUSCAZ98	MMCPUR97	SEDBUZ99	XIHSEM97
AKIWOG	DOQSUX01	HUSCAZ99	MMCPUR98	SEGGER98	XIHSEM98
AKOWAY	DOVTAJ	HYPRCX	MMCPUR99	SEGGER99	XIHSEM99
ALAZAO	DOWDEY98	IFIDUW97	MOBYIL98	SEJSIJ	XORNOH
ALEVES	DOWDEY99	IFIDUW98	MOBYIL99	SEKWIO99	XOWGOF98
ALPHAGLY	DOYNOU	IFIDUW99	MOPMOT96	SERPOP01	XOWGOF99
APAPYR	DOZMIO	IGAYUK99	MOPMOT98	SEWGEG	XOYFIA
APURAE99	DUBCEI	IJICIN97	MOPMOT99	SIYLOB	XUCHAE
AQEWOI	DUKSOR98	IJICIN98	MOSRUH99	SUFNUC98	XUJWUU
ARACYP	DUKSOR99	IJICIN99	MUGKAA01	SUFNUC99	YAJZUF
ARFCYT99	DUXWIC	ILIDAI	MUHCOH98	SUMPAR99	YALLON99
ASTROM99	EKOKEU	IWILAB	MUHCOH99	SURREC	YANQEK99
ASULOP98	ELEYID98	IWUGIQ	MUNVAS98	TAHJIV99	YAYDAD
ASULOP99	ELEYID99	IYAYIQ	MUNVAS99	TEDZUY	YEMFOL99
ATESIB	ELINOC	JAFJAB98	MURYED99	TEJDAN96	YEQCIG98
ATOXEM	ERAHAG98	JAFJAB99	MURYIH99	TEJDAN97	YIRNOC
ATUTAK99	ERAHAG99	JALQET	MUYYUA	TEJDAN98	YOPSUR
AVAGIN	ESAMIU	JAPPEV99	NAGGEH	TEJDAN99	YURNAA
AWOSOU	EVENIC99	JATWUW	NAGLEN98	TEJMOK98	ZAYJOY
AXIMEZ	EVOTUE	JAXSEH	NAGLEN99	TEJMOK99	ZAYPAQ99
AZAJAM	EXOFOM95	JAYNIG99	NALCYS02	TEJNEB95	ZEJMOQ99
AZURUI99	EXOFOM96	JEDLEK99	NAPRLA02	TEJNEB96	ZETKUE
BACTIJ	EXOFOM97	JEMSUP	NARGIX99	TEJNEB97	ZIKFOO99
BAKFUO	EXOFOM98	JIBDIH98	NATMOK	TEJNEB98	ZILHOR02
BALMOR	EXOFOM99	JIBDIH99	NEFGAG98	TEJNEB99	ZOZWAM
BAMDIC	FAGNEG99	JIYREO98	NEFGAG99	TEKNAY97	ZZZSBA97
BANLEH	FAPKEN99	JIYREO99	NEQNIG	TEKNAY98	ZZZSPS99
BARFIJ	FAXPOJ99	JIYWET96	NETGIC97	TEKNAY99	ZZZUEE04
BATSUK	FAZRAZ	JIYWET97	NETGIC98	TERMIM99	ARUTAI
BATWOI	FAZRED99	JIYWET98	NETGIC99	TERNOT99	BALBAS
BAVBOP	FECQAF	JIYWET99	NEZNAH	TIRPAL	BALBOG
BCBANN01	FEDBOF	JOWWIB	NITRBE01	TIXPIZ	BASBIH
BEBWEK	FEGSEP	JUHGUO99	NOREPH01	TODNUV97	BASCAA
BEDDOE	FESPAU99	KACTUE99	NOWCAD99	TODNUV98	BEWHUG
BEKBOJ98	FETZEJ97	KADPUA98	NUTGEO99	TODNUV99	BOSLEB
BEKBOJ99	FETZEJ98	KADPUA99	NUXXEJ	TONQAO99	BRUCLB
BEMDAZ97	FETZEJ99	KAKLUE	OBIDAE	TPHCUR97	CEPRUK98
BEMDAZ98	FICWAQ99	KAYWIR01	OGEYAA	TPHCUR98	CEPRUK99
BEMDAZ99	FICYOG	KECGOO99	OKESOM	TPHCUR99	DIKFEJ
BENTEU	FIKCEI98	KEHZAY98	PABMUA98	TUBERC01	DIKGAG
BEPRAP99	FIKCEI99	KEHZAY99	PABMUA99	TUDMOU99	DIXJIE
BERFUZ	FIMPEW	KEPHUI	PAGXOL99	TUGWUN	DPHOCE03
BERXAY	FIQKAS99	KEZREM	PAKWON	TUGXAU	EMINET

BEWQOK	**FISKIC99**	**KGULAM**	**PEXTAN98**	**UHAWUV**	**EZOROA**
BIBXUG	**FOKRUS**	**KICJEL99**	**PEXTAN99**	**UNOCAB**	**FOLKUM**
BIDMIK97	**FONFIY99**	**KIHVAY97**	**PHOGLY04**	**VALTUX97**	**FUCNAS**
BIDMIK98	**FORHAV**	**KIHVAY98**	**PINPAD01**	**VALTUX98**	**GAWMUN**
BIDMIK99	**FOYXIA99**	**KIHVAY99**	**PIPVAL**	**VANREH**	**GIHVIC**
BIJSIX96	**FUFZAH**	**KIHVEC**	**POJQEK98**	**VANRIL**	**HEPPAT98**
BIJSIX97	**FUGQON01**	**KIMNID99**	**POJQEK99**	**VARNIM**	**HEPPAT99**
BIJSIX98	**FUXBIJ01**	**KITREK**	**PONTIV99**	**VEFZIP**	**HERKAR**
BIJSIX99	**GAFNAC**	**KOCHOZ**	**POSTAS99**	**VEGKIB97**	**HESTIA**
BISDIQ	**GAGSIR05**	**KOFKUL98**	**POTHRE01**	**VEGKIB98**	**INEGOX**
BOPTZO	**GAKGEE99**	**KOFKUL99**	**POVKIU**	**VEGKIB99**	**JAHBEA**
BOQJEW98	**GARVOK99**	**KOFLOG**	**QACLAH**	**VIGNII97**	**JEBNUA**
BOQJEW99	**GEDLEH98**	**KOKBER98**	**QAKWAB99**	**VIGNII98**	**LEZHUT**
BOXGIE99	**GEDLEH99**	**KOMHAV01**	**QANJUL98**	**VIGNII99**	**MOSXAU**
BOXKEE99	**GEDPIP98**	**KOVCON**	**QANJUL99**	**VIOLME98**	**ODIBUZ**
BOXWOA	**GEDPIP99**	**KUTHOW**	**QAXDEY99**	**VIOLME99**	**POLZEV**
BUHJAP	**GEJBIH99**	**LABRUB98**	**QAXDIC99**	**VIRYEA**	**QOWQEY**
BUWNOW	**GEJKUC97**	**LABRUB99**	**QECCIK**	**VISVIC**	**RAQREH**
CABSIH10	**GEJKUC98**	**LABVIT98**	**QEKXUZ**	**VOCKIH**	**RAVVEQ**
CAINSP95	**GEJKUC99**	**LABVIT99**	**QELCUF**	**VOJJIN98**	**REKWEK**
CAINSP96	**GERDOW98**	**LADTIT**	**QEVPEM**	**VOJJIN99**	**SAMBOX**
CAINSP97	**GERDOW99**	**LALNIN03**	**QIJHUM97**	**VOWGAP**	**TEHHOD**
CAINSP98	**GIMKAO**	**LAPFIR98**	**QIJHUM98**	**VOWTEG98**	**UFUMOY**
CAKRIQ	**GIQYEK**	**LAPFIR99**	**QIJHUM99**	**VOWTEG99**	**WAVXUN**
CAMVES98	**GLPHEA98**	**LARGPH97**	**QINGAV**	**VUXKAA98**	**WIXPID**
CAMVES99	**GLPHEA99**	**LARGPH98**	**QUBVUE97**	**VUXKAA99**	**YAKXOY**
CANCAW	**GLUTAM02**	**LARGPH99**	**QUBVUE98**	**VUXKEE**	**YAMXUF**

Table B5. *MOPRO* input files produced for the **6m2tU** structure by LSDB.

(a) mopro_lsdb.par (parameter file)

```
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!               CREATED BY PROGRAM << LSDB >>
!                modified by K.N.Jarzembska
!                       Nancy, 2012
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
CELL    4.3315   14.4396    9.5617  90.0000  90.8638  90.0000  0.71073

SYMM   2 P CENTRO
  X, Y, Z
 -X,1/2+Y,1/2-Z

SCALE  1  1.0

FMULT  1.0

UOVER  0.0

SOLVT  0.0  50.0

EXTIN  0.0  GAUSSIAN  ISOT  TYP1

DUMMY   1
DUMY     0.960668    0.273768    0.007100  D1     0

KAPPA  11
  0.983000  1.016000
  0.987000  1.134000
  1.001000  1.031000
  0.993000  0.844000
  1.001000  0.887000
  1.004000  0.909000
  0.995000  0.868000
  1.020000  0.927000
  1.166000  1.383000
  1.094000  1.156000
  1.110000  1.158000

ATOMS   15

ATOM   1 S2        MOL   1    0.692451    0.440945    0.165475 1.000  1 S
XY   C2           N1          HEX       K1       V0  M0  Q0
UANI     0.012460  0.008080  0.010620  0.000785  0.003690 -0.000994
 6.36200 0.00000  -.053  0.000  0.000  0.000  0.000  0.000  -.156  0.000
  0.000  -.044  -.006  0.000  0.000  0.097  0.000
  0.024  0.000  0.000  -.038  0.000  0.000  0.000  0.061  0.000

ATOM   2 O4        MOL   1    1.364950    0.307720   -0.191350 1.000  1 O
XY   C4           N3          HEX       K2       V0  M0  Q0
UANI     0.015720  0.012260  0.013550 -0.000200  0.008500 -0.000180
 6.13300 0.00000  -.103  -.005  0.000  -.068  0.000  0.000  -.090  0.000
  0.000  -.003  0.002  0.000  0.000  0.014  0.000
  -.004  0.000  0.000  0.002  0.000  0.000  0.000  0.006  0.000

ATOM   3 N1        MOL   1    0.757040    0.261600    0.110150 1.000  1 N
XY   D1           C2          HEX       K3       V0  M0  Q0
UANI     0.010510  0.007550  0.009440 -0.000170  0.003390 -0.000080
 4.92900 0.00000  0.000  0.000  0.000  0.046  0.000  0.000  0.000  0.000
  0.000  0.016  0.000  0.000  0.000  -.137  0.000
  0.005  0.000  0.000  0.000  0.000  0.000  0.000  0.010  0.000

ATOM   4 N3        MOL   1    1.048770    0.359810   -0.023210 1.000  1 N
```

```
XY   D1           C2            HEX     K3       V0  M0  Q0
UANI     0.011020  0.008140  0.010180 -0.000750  0.004160 -0.000540
 4.92900 0.00000  0.000  0.000  0.000  0.046  0.000  0.000  0.000  0.000
  0.000  0.016  0.000  0.000  0.000  -.137  0.000
  0.005  0.000  0.000  0.000  0.000  0.000  0.000  0.010  0.000

ATOM   5 C4        MOL   1    1.176390    0.288310   -0.100050 1.000  1 C
XY   D1           N3            HEX     K4       V0  M0  Q0
UANI     0.010250  0.009430  0.009100 -0.000140  0.003450 -0.000470
 4.15000 0.00000  -.083  -.049  0.000  -.309  0.000  0.000  0.112  -.055
  0.000  0.000  0.000  0.000  0.000  -.431  0.029
  0.075  0.000  0.000  -.016  0.000  0.000  0.000  -.063  -.038

ATOM   6 C2        MOL   1    0.839850    0.350110    0.080450 1.000  1 C
XY   D1           N1            HEX     K5       V0  M0  Q0
UANI     0.009340  0.007830  0.008620 -0.000050  0.002340 -0.000440
 4.02400 0.00000  -.089  0.000  0.000  -.212  0.000  0.000  0.070  0.029
  0.000  -.063  0.000  0.000  0.000  -.318  -.010
  0.036  0.000  0.000  0.017  -.012  0.000  0.000  -.085  -.011

ATOM   7 C5        MOL   1    1.073940    0.197070   -0.064750 1.000  1 C
XY   D1           C6            HEX     K6       V0  M0  Q0
UANI     0.011690  0.008600  0.010140  0.000140  0.003310 -0.001010
 3.90700 0.00000  0.056  0.000  0.000  -.180  0.000  0.000  -.029  0.000
  0.000  0.030  0.000  0.000  0.000  -.264  0.000
  0.020  0.000  0.000  -.019  0.000  0.000  0.000  0.000  0.000

ATOM   8 C6        MOL   1    0.868020    0.185710    0.040010 1.000  1 C
XY   D1           N1            HEX     K7       V0  M0  Q0
UANI     0.010170  0.007680  0.009530 -0.000040  0.002190 -0.000190
 4.12500 0.00000  0.000  -.079  0.000  -.257  0.000  0.000  0.000  -.081
  0.000  0.014  0.000  0.000  0.000  -.341  0.000
  0.042  0.000  0.000  0.000  -.024  0.000  0.000  -.028  -.028

ATOM   9 C7        MOL   1    0.753990    0.093140    0.085940 1.000  1 C
ZX   C6           H7B           HEX     K8       V0  M0  Q0
UANI     0.014790  0.008310  0.014760 -0.000510  0.003860  0.000460
 3.58800 0.00000  0.000  0.000  0.000  0.000  0.000  0.000  0.000  0.000
  0.239  0.000  0.000  0.000  0.000  0.188  0.000
  0.054  0.000  0.000  0.000  0.000  -.080  0.000  0.000  0.000

ATOM  10 H1        MOL   1    0.598731    0.250802    0.187785 1.000  1 H
ZY   N1           C2            QUA     K9       V0  M0  Q0
UISO     0.027500
 0.85100 0.00000  0.000  0.000  0.141  0.059  0.000  0.000  0.000  0.000

ATOM  11 H3        MOL   1    1.123024    0.425521   -0.047708 1.000  1 H
ZY   N3           C2            QUA     K9       V0  M0  Q0
UISO     0.030600
 0.85100 0.00000  0.000  0.000  0.141  0.059  0.000  0.000  0.000  0.000

ATOM  12 H7A       MOL   1    0.515164    0.095400    0.115504 1.000  1 H
ZY   C7           C6            QUA     K10      V0  M0  Q0
UISO     0.023500
 1.08900 0.00000  0.000  0.000  0.185  0.085  0.000  0.000  0.000  0.000

ATOM  13 H7B       MOL   1    0.774010    0.043976    0.001859 1.000  1 H
ZY   C7           C6            QUA     K10      V0  M0  Q0
UISO     0.050200
 1.08900 0.00000  0.000  0.000  0.185  0.085  0.000  0.000  0.000  0.000

ATOM  14 H7C       MOL   1    0.884898    0.068119    0.174831 1.000  1 H
ZY   C7           C6            QUA     K10      V0  M0  Q0
UISO     0.041200
 2.00000 0.00000  0.000  0.000  0.185  0.085  0.000  0.000  0.000  0.000

ATOM  15 H5        MOL   1    1.160997    0.139402   -0.124898 1.000  1 H
```

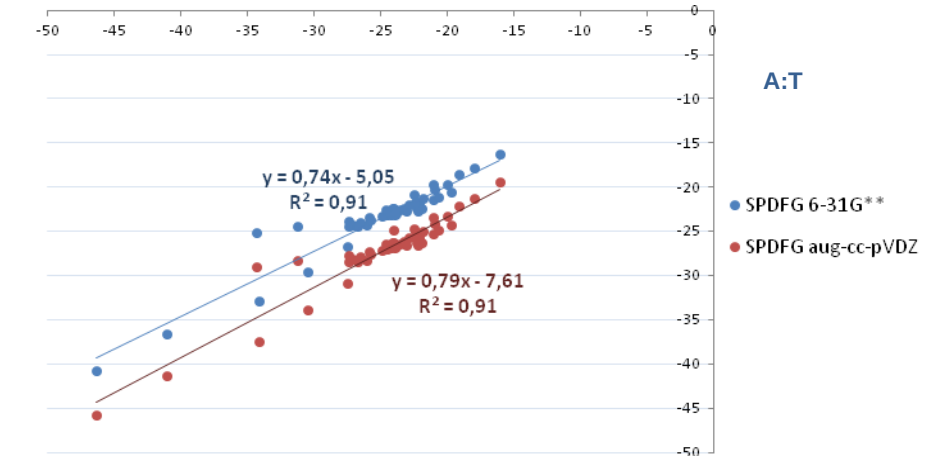
```
ZY   C5           C6            QUA     K11      V0  M0  Q0
UISO     0.032200
 1.03400 0.00000  0.000  0.000  0.182  0.086  0.000  0.000  0.000  0.000

ANHAR    0    3         !  #atoms #order
! 1         ! atom number
! 0. 0. 0. 0. 0. 0. 0. 0.
! 0. 0.

STOP
```

(b) res_mopro.inp (restrain file)

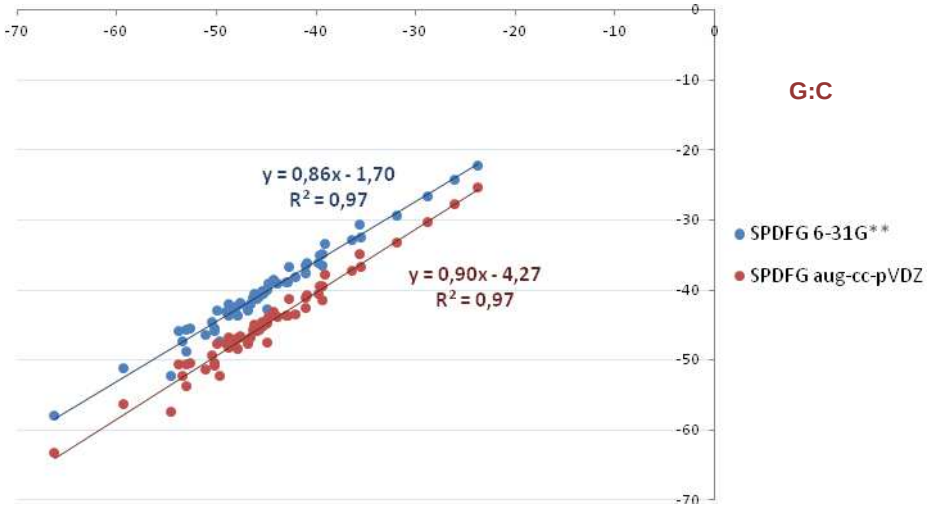
```
!
!     X-H distances
!
DISTAN    N1   1   H1   1  1.030 0.001
DISTAN    N3   1   H3   1  1.030 0.001
DISTAN    C7   1  H7A   1  1.077 0.001
DISTAN    C7   1  H7B   1  1.077 0.001
DISTAN    C7   1  H7C   1  1.077 0.001
DISTAN    C5   1   H5   1  1.083 0.001
!
! isotropic H atoms ADPs
!
URATIO       N1   1   H1   1    1.5 0.01
URATIO       N3   1   H3   1    1.5 0.01
URATIO       C7   1  H7A   1    1.2 0.01
URATIO       C7   1  H7B   1    1.2 0.01
URATIO       C7   1  H7C   1    1.2 0.01
URATIO       C5   1   H5   1    1.2 0.01
!
! end
```

(c) con_mopro.inp (constrain file)

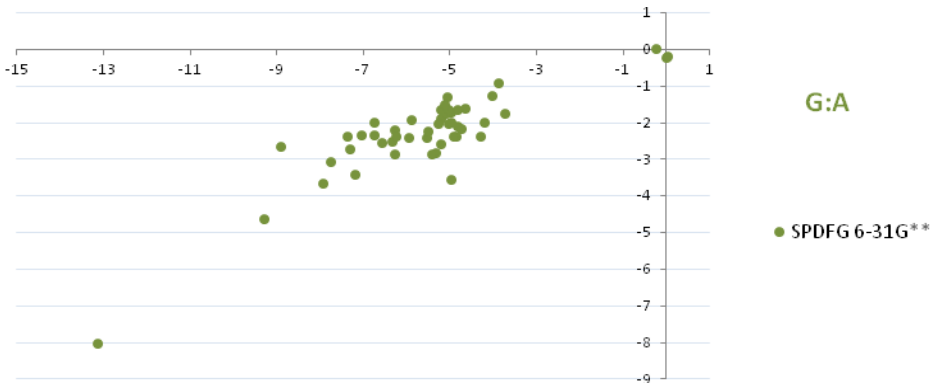
```
!
! local symmetry
!
SYMPLM  mz   S2         1
SYMPLM  mz   O4         1
SYMPLM  mz   N1         1
SYMPLM  mz   N3         1
SYMPLM  mz   C4         1
SYMPLM  mz   C2         1
SYMPLM  mz   C5         1
SYMPLM  mz   C6         1
SYMPLM  3m   C7         1
SYMPLM  cy   H1         1
SYMPLM  cy   H3         1
SYMPLM  cy   H7A        1
SYMPLM  cy   H7B        1
SYMPLM  cy   H7C        1
SYMPLM  cy   H5         1
!
! similarity (switched-off)
!
! CONPVM  N3        1 N1        1
! CONPVM  H3        1 H1        1
! CONPVM  H7B       1 H7A       1
! CONPVM  H7C       1 H7A       1
!
! end
```



(*a*)



(*b*)



(*c*)

Figure B6. Correlation between SPDFG/B3LYP/6-31G** or SPDFG/B3LYP/aug-cc-pVDZ E_{es} results and UBDB2011+EPMM method. (*a*) A:T (Table B7) (*b*) G:C (Table B8) (*c*) G:A (Table B9); *y* – SPDFG, *x* – UBDB+EPMM, all energy values are given in kcal·mol^{-1}.

Table B7. Electrostatic energy values for A:T complexes.

A:T scan	E_{es} / kcal·mol^{-1}		
BD0037	UBDB2011	SPDFG/B3LYP/6-31G**	SPDFG/B3LYP/aug-cc-pVDZ
buckle10	-24.5	-23.2	-27.0
buckle1	-24.1	-22.4	-26.3
buckle2	-24.0	-22.5	-26.3
buckle3	-24.0	-22.5	-26.4
buckle4	-24.0	-22.6	-26.4
buckle5	-23.9	-22.7	-26.5
buckle6	-23.9	-22.8	-26.6
buckle7	-24.0	-22.9	-26.7
buckle8	-24.0	-23.0	-26.9
buckle9	-24.2	-23.1	-26.9
opening10	-23.9	-23.1	-26.9
opening1	-34.3	-25.2	-29.1
opening2	-31.3	-24.5	-28.4
opening3	-27.4	-23.9	-27.8
opening4	-25.8	-23.5	-27.3
opening5	-24.4	-22.9	-26.8
opening6	-24.0	-22.9	-26.7
opening7	-23.9	-23.0	-26.8
opening9	-26.0	-24.3	-28.3
propeller10	-23.2	-22.5	-26.2
propeller1	-24.1	-23.0	-26.8
propeller2	-24.9	-23.3	-27.2
propeller3	-25.7	-23.7	-27.6
propeller4	-26.5	-24.0	-27.9
propeller5	-27.1	-24.3	-28.2
propeller6	-27.3	-24.5	-28.4
propeller8	-26.7	-24.4	-28.4
propeller9	-20.9	-20.3	-24.1
shear10	-19.7	-20.6	-24.4
shear1	-21.0	-21.5	-25.3
shear2	-24.0	-22.8	-26.6
shear3	-23.5	-22.7	-26.5
shear4	-23.1	-22.8	-26.6
shear6	-22.3	-22.5	-26.3
shear7	-22.0	-22.0	-25.8
shear8	-21.7	-21.3	-25.1
shear9	-20.7	-21.2	-24.9
stagger10	-19.1	-18.6	-22.2
stagger1	-21.1	-19.8	-23.5
stagger2	-22.5	-20.9	-24.7
stagger3	-22.2	-21.8	-25.6
stagger4	-21.8	-22.5	-26.3
stagger5	-22.2	-22.8	-26.6
stagger6	-23.9	-22.9	-26.7
stagger7	-24.6	-22.6	-26.4

stagger8	-22.8	-22.1	-25.8
stagger9	-16.0	-16.3	-19.4
stretch10	-46.3	-40.8	-45.8
stretch1	-41.0	-36.6	-41.4
stetch2	-34.1	-32.9	-37.5
stretch3	-30.5	-29.6	-34.0
stretch4	-27.4	-26.7	-30.9
stretch5	-22.4	-21.8	-25.5
stretch7	-20.0	-19.8	-23.3
stretch8	-17.9	-17.9	-21.3
stretch9	-24.8	-23.4	-27.2

Table B8. Electrostatic energy values for G:C complexes.

G:C scan	E_{es} / kcal·mol^{-1}		
BD0002i	UBDB2011	SPDFG/B3LYP/6-31G**	SPDFG/B3LYP/aug-cc-pVDZ
buckle10	-39.4	-36.5	-41.5
buckle1	-51.1	-46.5	-51.4
buckle2	-50.1	-45.6	-50.3
buckle3	-50.4	-44.6	-49.3
buckle4	-48.8	-43.6	-48.2
buckle5	-46.9	-42.5	-47.1
buckle6	-45.9	-41.3	-45.9
buckle7	-44.8	-40.1	-44.8
buckle8	-42.8	-38.9	-43.6
buckle9	-41.0	-37.6	-42.5
opening10	-54.5	-52.3	-57.4
opening1	-46.2	-40.5	-44.9
opening2	-45.4	-40.1	-44.6
opening3	-45.1	-40.0	-44.6
opening4	-45.3	-40.2	-44.8
opening5	-45.7	-41.0	-45.6
opening6	-46.5	-42.0	-46.8
opening7	-47.9	-43.7	-48.4
opening8	-50.1	-45.9	-50.7
opening9	-53.1	-48.7	-53.7
propeller10	-47.6	-41.8	-46.5
propeller1	-39.7	-35.0	-39.5
propeller2	-42.7	-36.8	-41.3
propeller3	-44.3	-38.4	-43.0
propeller4	-45.1	-39.9	-44.6
propeller5	-46.1	-41.3	-45.9
propeller6	-48.0	-42.4	-47.0
propeller7	-50.0	-42.9	-47.6
propeller8	-49.0	-43.0	-47.7
propeller9	-48.2	-42.7	-47.4
shear10	-42.1	-38.1	-43.4
shear1	-35.6	-30.6	-34.9

shear2	-39.3	-34.9	-39.4
shear3	-44.8	-39.1	-43.7
shear4	-46.9	-42.9	-47.7
shear5	-53.8	-45.8	-50.6
shear6	-53.4	-47.3	-52.3
shear7	-49.6	-47.2	-52.2
shear8	-53.0	-45.6	-50.5
shear9	-44.9	-42.7	-47.5
stagger10	-39.1	-33.3	-37.7
stagger1	-36.4	-32.8	-37.3
stagger2	-40.9	-36.2	-40.7
stagger3	-43.0	-38.9	-43.6
stagger4	-45.8	-41.0	-45.6
stagger5	-47.6	-42.1	-46.7
stagger6	-48.8	-42.1	-46.7
stagger7	-46.3	-41.1	-45.7
stagger8	-43.9	-39.2	-43.7
stagger9	-41.0	-36.5	-41.0
stretch10	-23.7	-22.2	-25.3
stretch1	-66.3	-57.9	-63.3
stetch2	-59.4	-51.2	-56.4
stretch3	-52.6	-45.4	-50.3
stetch4	-45.2	-40.5	-45.2
stretch5	-39.8	-36.2	-40.6
stretch6	-35.5	-32.5	-36.7
stretch7	-31.9	-29.4	-33.2
stretch8	-28.8	-26.6	-30.3
stretch9	-26.0	-24.3	-27.6

Table B9. Electrostatic energy values for G:A complexes.

G:A	E_{es} / kcal·mol^{-1}	
	UBDB2011	SPDFG/B3LYP/6-31G**
ga_BD0001a	-4.3	-2.4
ga_BD0002a	-5.0	-1.7
ga_BD0002b	-5.0	-3.6
ga_BD0002c	-5.9	-2.4
ga_BD0004a	-9.3	-4.6
ga_BD0005a	-5.2	-1.9
ga_BD0014a	-5.2	-2.6
ga_BD0018a	-6.3	-2.5
ga_BD0019a	-7.7	-3.1
ga_BD0024a	0.0	-0.2
ga_BD0024b	0.0	-0.2
ga_BD0029a	-5.3	-2.0
ga_BD0032a	-4.6	-1.6
ga_BD0037a	-5.3	-2.8

ga_BD0041a	-5.0	-2.0
ga_BD0052a	-6.3	-2.2
ga_BD0052b	-3.9	-0.9
ga_BD0052c	-7.3	-2.7
ga_BD0054a	-5.0	-2.0
ga_BD0069a	-7.2	-3.4
ga_BD0070a	-5.4	-2.9
ga_BD0080a	-4.8	-1.7
ga_BD0082a	-6.3	-2.4
ga_BD0084a	-13.1	-8.0
ga_BD0087a	-0.2	0.0
ga_BD0090a	-4.8	-2.1
ga_BD0090b	-4.7	-2.2
ga_BDF068a	-5.5	-2.4
ga_BDJ008a	-4.2	-2.0
ga_BDJ025a	-5.2	-1.6
ga_BDJ025b	-6.3	-2.9
ga_BDJ036a	-5.1	-1.3
ga_BDJ037a	-5.9	-1.9
ga_BDJ060a	-6.6	-2.5
ga_BDJ060b	-4.0	-1.3
ga_BDJ081b	-8.9	-2.7
ga_BDL002a	-7.4	-2.4
ga_BDL005a	-5.1	-1.8
ga_BDL012a	-7.9	-3.7
ga_BDL020a	-5.0	-1.6
ga_BDL028a	-6.7	-2.4
ga_BDL029a	-7.1	-2.3
ga_BDL042a	-3.7	-1.7
ga_BDL084a	-5.5	-2.2
ga_DD0081a	-5.1	-1.5
ga_UDI030a	-4.8	-2.4
ga_UDJ060a	-6.7	-2.0
ga_UDM010a	-4.9	-2.4

Table B10. Electrostatic energy values for NAB:AA complexes.**

NAB* + aminoacid fragment	E_{es} / kcal·mol^{-1}	
	UBDB2011	SPDFG/B3LYP/6-31G**
Ksmgua_3_3_4_6	-47.7	-46.3
Nsmade_5_2_3_14	-15.5	-14.1
Nsmade_5_8_3_15	-15.4	-13.6
Nsmcyt_5_4_3_11	-19.4	-17.8
Nsmgua_5_10_3_13	-14.8	-14.3
Nsmgua_5_3_6_6	-9.5	-8.2
Nsmura_5_8_3_10	-15.0	-12.5

Nsmura_5_9_3_10	-16.3	-13.7
Rsmcyt_10_4_8_6	-33.4	-33.6
Rsmgua_7_3_8_6	-33.7	-32.6
Ssmade_3_2_2_14	-11.1	-12.1
Ssmade_3_8_2_15	-13.4	-13.7
Ssmcyt_3_4_2_11	-12.4	-13.6

* NAB – nucleic acid base; ** Geometries from private communication (Ż. Czyżnikowska)

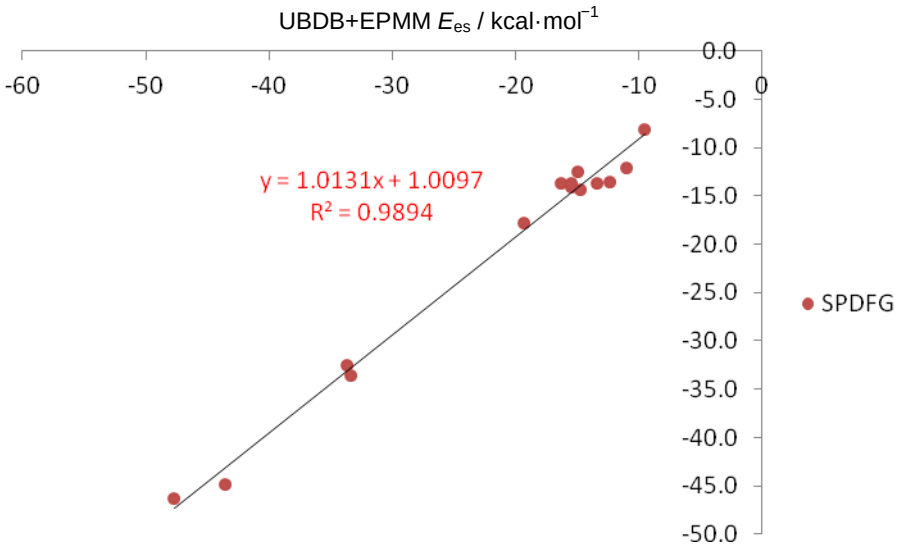


Figure B11. Correlation between SPDFG/B3LYP/6-31G** E_{es} results and UBDB2011+EPMM method for a set of NABs interacting with amino acid side chains. All energy values are given in kcal·mol^{-1}.

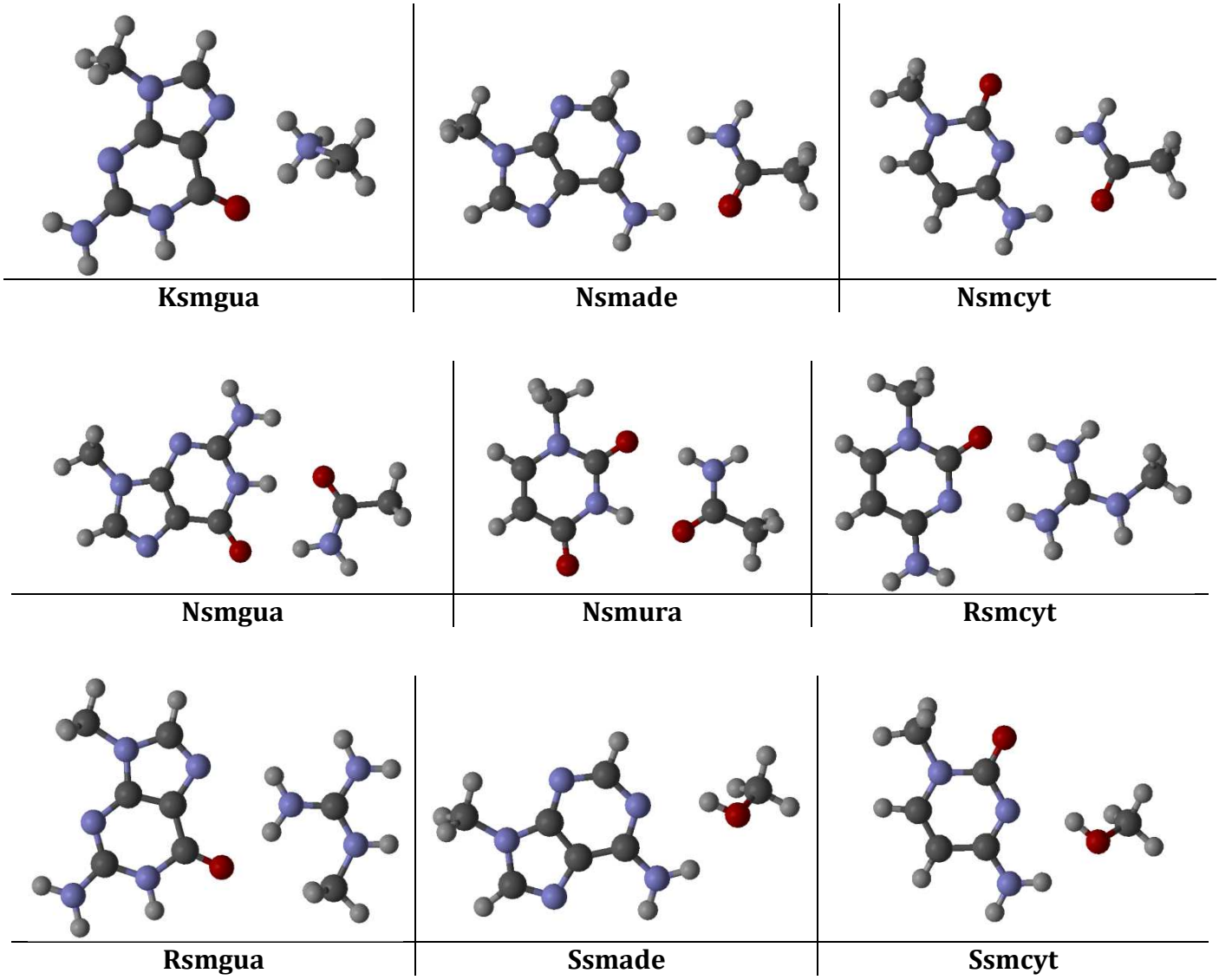


Figure B12. Schema of NABs interacting with aminoacid residue fragments.

IX. APPENDIX C

Supplementary Materials
Neuraminidase Part

Table C1. Influenza neuraminidase:inhibitor complexes.

No.		*PDB code*[a]	*Mutation*[b]	*Protein charge*[c]	*Inhibitor*[d]	*Inhibitor charge*	pK_i^{exp}	pIC_{50}^{exp}
1	N2[e]	2BAT	Asp339Asn	+2	SIA	−1	≈ 3.00[1]	
2	N2	1INW		+2	AXP	−2	–	like SIA2
3	N2	1INW		+2	AXP-H	−1	–	
4	N2	1INX		+1	EQP	−2	–	≈ 4.00[3]
5	N2	1INX		+1	EQP-H	−1	–	
6	N2	1IVF		+1	DAN	−1	5.40[4] 5.82[5]	
7	N2	1IVD		+1	ST1	−1	–	3.12[6]
8	N2	1IVC		+1	ST2	−1	–	< 2.00[7]
9	N2	1IVC		+1	ST2-H	0	–	
10	N2	1IVE		+1	ST3	−1	–	< 2.00[8]
11	N2	1IVE		+1	ST3-H	0	–	
12	N2	1ING:A		+1	ST5	−1	–	2.40[9]
13	N2	1ING:B		+1	ST5	−1	–	
14	N2	1INH:A		+1	ST6	0	–	2.30[9]
15	N2	1INH:B		+1	ST6	0	–	
16	N6[f]	1W1X:A		+2	SIA	−1	–	
17	N6	1W1X:B		+3	SIA	−1	–	
18	N6	1W1X:C		+3	SIA	−1	–	
19	N6	1W1X:D		+2	SIA	−1	–	
20	N6	1W20:A		+2	SIA	−1	–	
21	N6	1W20:B		+2	SIA	−1	–	
22	N6	1W20:C		+2	SIA	−1	–	
23	N6	1W20:D		+2	SIA	−1	–	

24	N6	1W21:A		+2	SIA	−1	–	
25	N6	1W21:B		+2	SIA	−1	–	
26	N6	1W21:C		+2	SIA	−1	–	
27	N6	1W21:D		+2	SIA	−1	–	
28	N9[g]	1MWE		−1	SIA	−1	4.26[1]	
29	N9	2QWB	Arg292Lys[h]	−1	SIA	−1	2.74[11]	
30	N9	1INY	Ser370Leu	−1	EQP	−2	–	≈ 3.00[12]
31	N9	1INY	Ser370Leu	−1	EQP-H	−1	–	
32	N9	1F8B		−1	DAN	−1	5.30[13] 5.58[14]	
33	N9	1NNB		−1	DAN	−1	–	
34	N9	2QWC	Arg292Lys[h]	−1	DAN	−1	3.55[15]	
35	N9	1F8C		−1	4AM	0	6.83[16] 7.40[17] 7.41[18]	
36	N9	2QWD	Arg292Lys[h]	−1	4AM	0	4.85[19]	
37	N9	1F8D		−1	9AM	0	3.40[20]	
38	N9	1F8E		−1	49A	+1	4.82[21]	
39	N9	1NNC		−1	ZMR	0	8.71[22] 8.89[23] 9.19[18]	
40	N9	2QWE	Arg292Lys[h]	−1	ZMR	0	7.48[24]	
41	N9	1BJI		0	G21	0	9.19[18]	
42	N9 N9	2QWF 2QWI	Arg292Lys[h]	0 0	G20 G20	0 0	5.67[25] 8.36[26], 9.28[18]	
43	N9	2QWG	Arg292Lys[h]	0	G28	0	–	3.64[27]
44	N9	2QWJ		−1	G28	0	9.41[18]	6.64[28]
45	N9	2QWH	Arg292Lys[h]	−1	G39	0	–	< 4.52[29] 4.89[30]
46 47	N9	2QWK		0	G39	0	9.96[31]	8.70[32] 9.10[33]
48	N9	1L7F		0	BCZ	0	10.82[34]	
49	N9	1L7G	Glu119Gly[h]	+1	BCZ	0	–	
50	N9	1L7H	Arg292Lys[h]	0	BCZ	0	–	
51	N9	1XOE		−1	ABX	0	6.82[35] 7.43[36]	7.39[37]
52	N9	1XOG		−1	ABW	−1	–	6.39[38]
53	B/B[i]	1NSC:A		+4	SIA	−1	–	< 2.00[39] 3.00[40]
54	B/B	1NSC:B		+4	SIA	−1	–	

55	B/L[j]	1INV		+3	EQP	−2	–	~4.00[41]
56	B/L	1INV		+3	EQP-H	−1	–	
57	B/B	1NSD:A		+4	DAN	−1	–	4.82[9]
58	B/B	1NSD:B		+4	DAN	−1	–	
59	B/B	1A4G:A		+4	ZMR	0	8.85[42]	
60	B/B	1A4G:B		+4	ZMR	0	–	
61	B/B	1A4Q:A		+4	G21	0	–	5.07[42]
62	B/B	1A4Q:B		+4	G21	0	–	
63	B/L	1IVB		+3	ST1	−1	–	3.12[43]
64	B/L	1INF	Arg382Lys	+3	ST4	0	–	5.00[9]
65	B/L	1B9S		+4	FDI	−1	–	< 3.18[44,45]
66	B/L	1VCJ		+3	IBA	0	–	4.59[46]
67	B/L	1VCJ		+3	IBA-H	+1	–	
68	B/L	1B9V		+4	RA2	−1	–	3.57[47] 3.65[48] 3.98[49]
69	B/L	1B9V		+4	RA2-H	0	–	
70	B/L	1B9T		+3	RAI	0	–	5.10[50,49]

[a] Chain identificator is denoted after colon. [b] Residue numbers like in original PDB structures. [c] The charge of polipeptide chain plus calcium ion. [d] Inhibitors and their protonation state used in calculations are shown in Figure 4.3 in the main text. [f] A/Tokyo/3/67 (H2N2). [e] A/duck/England/56(H11N6). [g] A/tern.Australia/G70C/75 (H11N9). [h] Mutation in active site. [i] B/Beijing/1/87. [j] B/Lee/40.

Additional references

1 von Itzstein, M.; Dyason, J. C.; Oliver, S. W.; White, H. F.; Wu, W. Y.; Kok, G. B.; Pegg, M. S. A study of the active site of influenza virus sialidase: an approach to the rational design of novel anti-influenza drugs. *J. Med. Chem.* **1996**, *39*, 388-391.

2 White, C. L.; Janakiraman, M. N.; Laver, W. G.; Philippon, C.; Vasella, A.; Air, G. M.; Luo, M. A sialic acid-derived phosphonate analog inhibits different strains of influenza virus neuraminidase with different efficiencies. *J. Mol. Biol.* **1995**, *245*, 623–634.

3 White, C. L.; Janakiraman, M. N.; Laver, W. G.; Philippon, C.; Vasella, A.; Air, G. M.; Luo, M. A sialic acid-derived phosphonate analog inhibits different strains of influenza virus neuraminidase with different efficiencies. *J. Mol. Biol.* **1995**, *245*, 623–634.

4 Taylor, N. R.; von Itzstein, M. Molecular Modeling Studies on Ligand Binding to Sialidase from Influenza Virus and the Mechanism of Catalysis. *J. Med. Chem.* **1994**, *37*, 616–624.

5 Stoll, V.; Stewart, K. D.; Maring, C. J.; Muchmore, S.; Giranda, V.; Gu, Y. G.; Wang, G.; Chen, Y.; Sun, M.; Zhao, C.; Kennedy, A. L.; Madigan, D. L.; Xu, Y.; Saldivar, A.; Kati, W.; Laver, G.; Sowin, T.; Sham, H. L.; Greer, J.; Kempf, D. Influenza neuraminidase inhibitors: structure-based design of a novel inhibitor series. *Biochemistry* **2003**, *42*, 718-727.

6 Jedrzejas, M. J.; Singh, S.; Brouillette, W. J.; Laver, W. G.; Air, G. M.; Luo, M. Structures of aromatic inhibitors of influenza virus neuraminidase. *Biochemistry* **1995**, *34*, 3144–3151.

[7] Jedrzejas, M. J.; Singh, S.; Brouillette, W. J.; Laver, W. G.; Air, G. M.; Luo, M. Structures of aromatic inhibitors of influenza virus neuraminidase. *Biochemistry* **1995**, *34*, 3144–3151.

[8] Jedrzejas, M. J.; Singh, S.; Brouillette, W. J.; Laver, W. G.; Air, G. M.; Luo, M. Structures of aromatic inhibitors of influenza virus neuraminidase. *Biochemistry* **1995**, *34*, 3144–3151.

[9] Singh, S.; Jedrzejas, M. J.; Air, G. M.; Luo, M.; Laver, W. G.; Brouillette, W. J. Structure-based inhibitors of influenza virus sialidase. A benzoic acid lead with novel interaction. *J. Med. Chem.* **1995**, *38*, 3217-3225.

[10] Varghese, J. N.; Smith, P. W.; Sollis, S. L.; Blick, T. J.; Sahasrabudhe, A.; McKimm-Breschkin, J. L.; Colman, P. M. Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. *Structure* **1998**, *6*, 735–746.

[11] Varghese, J. N.; Smith, P. W.; Sollis, S. L.; Blick, T. J.; Sahasrabudhe, A.; McKimm-Breschkin, J. L.; Colman, P. M. Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. *Structure* **1998**, *6*, 735–746.

[12] White, C. L.; Janakiraman, M. N.; Laver, W. G.; Philippon, C.; Vasella, A.; Air, G. M.; Luo, M. A sialic acid-derived phosphonate analog inhibits different strains of influenza virus neuraminidase with different efficiencies. *J. Mol. Biol.* **1995**, *245*, 623–634.

[13] Mann, M. C.; Islam, T.; Dyason, J. C.; Florio, P.; Trower, C. J.; Thomson, R. J.; von Itzstein, M. Unsaturated N-acetyl-D-glucosaminuronic acid glycosides as inhibitors of influenza virus sialidase. *Glycoconj. J.* **2006**, *23*, 127–133.

[14] McKimm-Breschkin, J.L.; Sahasrabudhe, A.; Blick, T.J.; McDonald, M.; Colman, P. M.; Hart, G. J.; Bethell, R. C.; Varghese, J. N. Mutations in a conserved residue in the influenza virus neuraminidase active site decreases sensitivity to Neu5Ac2en-derived inhibitors. *J. Vir.* **1998**, *72*, 2456–2462.

[15] McKimm-Breschkin, J.L.; Sahasrabudhe, A.; Blick, T.J.; McDonald, M.; Colman, P. M.; Hart, G. J.; Bethell, R. C.; Varghese, J. N. Mutations in a conserved residue in the influenza virus neuraminidase active site decreases sensitivity to Neu5Ac2en-derived inhibitors. *J. Vir.* **1998**, *72*, 2456–2462.

[16] McKimm-Breschkin, J.L.; Sahasrabudhe, A.; Blick, T.J.; McDonald, M.; Colman, P. M.; Hart, G. J.; Bethell, R. C.; Varghese, J. N. Mutations in a conserved residue in the influenza virus neuraminidase active site decreases sensitivity to Neu5Ac2en-derived inhibitors. *J. Vir.* **1998**, *72*, 2456–2462.

[17] Smith, B. J.; Colman, P. M.; von Itzstein, M.; Danylec, B.; Varghese, J. N. Analysis of inhibitor binding in influenza virus neuraminidase. *Protein. Sci.* **2001**, *10*, 689–696.

[18] Smith, P. W.; Sollis, S. L.; Howes, P. D; Cherry, P. C.; Cobley, K. N.; Taylor, H.; Whittington, A. R.; Scicinski, J.; Bethell, R. C.; Taylor, N.; Skarzynski, T.; Cleasby, A.; Singh, O.; Wonacott, A.; Varghese, J.; Colman, P. Novel inhibitors of influenza sialidases related to GG167 structure-activity, crystallographic and Molecular dynamics studies with 4H-pyran-2-carboxylic acid 6-carboxamides. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2931–2936.

[19] McKimm-Breschkin, J.L.; Sahasrabudhe, A.; Blick, T.J.; McDonald, M.; Colman, P. M.; Hart, G. J.; Bethell, R. C.; Varghese, J. N. Mutations in a conserved residue in the influenza virus neuraminidase active site decreases sensitivity to Neu5Ac2en-derived inhibitors. *J. Vir.* **1998**, *72*, 2456–2462.

[20] Smith, B. J.; Colman, P. M.; von Itzstein, M.; Danylec, B.; Varghese, J. N. Analysis of inhibitor binding in influenza virus neuraminidase. *Protein. Sci.* **2001**, *10*, 689–696.

[21] Smith, B. J.; Colman, P. M.; von Itzstein, M.; Danylec, B.; Varghese, J. N. Analysis of inhibitor binding in influenza virus neuraminidase. *Protein. Sci.* **2001**, *10*, 689–696.

[22] McKimm-Breschkin, J.L.; Sahasrabudhe, A.; Blick, T.J.; McDonald, M.; Colman, P. M.; Hart, G. J.; Bethell, R. C.; Varghese, J. N. Mutations in a conserved residue in the influenza virus neuraminidase active site decreases sensitivity to Neu5Ac2en-derived inhibitors. *J. Vir.* **1998**, *72*, 2456–2462.

[23] Kati, W. M.; Montgomery, D.; Carrick, R.; Gubareva, L.; Maring, C.; McDaniel, K.; Steffy, K.; Molla, A.; Hayden, F.; Kempf, D.; Kohlbrenner, W. In vitro characterization of A-315675, a highly potent inhibitor of A and B strain influenza virus neuraminidases and influenza virus replication. *Antimicrob. Agents Chemother.* **2002**, *46*, 1014–1021.

[24] McKimm-Breschkin, J.L.; Sahasrabudhe, A.; Blick, T.J.; McDonald, M.; Colman, P. M.; Hart, G. J.; Bethell, R. C.; Varghese, J. N. Mutations in a conserved residue in the influenza virus neuraminidase active site decreases sensitivity to Neu5Ac2en-derived inhibitors. *J. Vir.* **1998**, *72*, 2456–2462.

[25] McKimm-Breschkin, J.L.; Sahasrabudhe, A.; Blick, T.J.; McDonald, M.; Colman, P. M.; Hart, G. J.; Bethell, R. C.; Varghese, J. N. Mutations in a conserved residue in the influenza virus neuraminidase active site decreases sensitivity to Neu5Ac2en-derived inhibitors. *J. Vir.* **1998**, *72*, 2456–2462.

[26] McKimm-Breschkin, J.L.; Sahasrabudhe, A.; Blick, T.J.; McDonald, M.; Colman, P. M.; Hart, G. J.; Bethell, R. C.; Varghese, J. N. Mutations in a conserved residue in the influenza virus neuraminidase active site decreases sensitivity to Neu5Ac2en-derived inhibitors. *J. Vir.* **1998**, *72*, 2456–2462.

[27] Varghese, J. N.; Smith, P. W.; Sollis, S. L.; Blick, T. J.; Sahasrabudhe, A.; McKimm-Breschkin, J. L.; Colman, P. M. Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. *Structure* **1998**, *6*, 735–746.

[28] Varghese, J. N.; Smith, P. W.; Sollis, S. L.; Blick, T. J.; Sahasrabudhe, A.; McKimm-Breschkin, J. L.; Colman, P. M. Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. *Structure* **1998**, *6*, 735–746.

[29] Yen, H. L.; Herlocher, L. M.; Hoffmann, E.; Matrosovich, M. N.; Monto, A. S.; Webster, R. G.; Govorkova, E. A. Neuraminidase inhibitor-resistant influenza viruses may differ substantially in fitness and transmissibility. *Antimicrob. Agents Chemother.* **2005**, *49*, 4075–4084.

[30] Varghese, J. N.; Smith, P. W.; Sollis, S. L.; Blick, T. J.; Sahasrabudhe, A.; McKimm-Breschkin, J. L.; Colman, P. M. Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. *Structure* **1998**, *6*, 735–746.

[31] Kati, W. M.; Montgomery, D.; Carrick, R.; Gubareva, L.; Maring, C.; McDaniel, K.; Steffy, K.; Molla, A.; Hayden, F.; Kempf, D.; Kohlbrenner, W. In vitro characterization of A-315675, a highly potent inhibitor of A and B strain influenza virus neuraminidases and influenza virus replication. *Antimicrob. Agents Chemother.* **2002**, *46*, 1014–1021.

[32] Varghese, J. N.; Smith, P. W.; Sollis, S. L.; Blick, T. J.; Sahasrabudhe, A.; McKimm-Breschkin, J. L.; Colman, P. M. Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. *Structure* **1998**, *6*, 735–746.

[33] Yen, H. L.; Herlocher, L. M.; Hoffmann, E.; Matrosovich, M. N.; Monto, A. S.; Webster, R. G.; Govorkova, E. A. Neuraminidase inhibitor-resistant influenza viruses may differ substantially in fitness and transmissibility. *Antimicrob. Agents Chemother.* **2005**, *49*, 4075–4084.

[34] Kati, W. M.; Montgomery, D.; Carrick, R.; Gubareva, L.; Maring, C.; McDaniel, K.; Steffy, K.; Molla, A.; Hayden, F.; Kempf, D.; Kohlbrenner, W. In vitro characterization of A-315675, a highly potent inhibitor of A and B strain influenza virus neuraminidases and influenza virus replication. *Antimicrob. Agents Chemother.* **2002**, *46*, 1014–1021.

[35] Maring, C. J.; Stoll, V. S.; Zhao, C.; Sun, M.; Krueger, A. C.; Stewart, K. D.; Madigan, D. L.; Kati, W. M.; Xu, Y.; Carrick, R. J.; Montgomery, D. A.; Kempf-Grote, A.; Marsh, K. C.; Molla, A.; Steffy, K. R.; Sham, H. L.; Laver, W. G.; Gu, Y. G.; Kempf, D. J.; Kohlbrenner, W. E. Structure-based characterization and optimization of novel hydrophobic binding interactions in a series of pyrrolidine influenza neuraminidase inhibitors. *J. Med. Chem.* **2005**, *48*, 3980–3990.

[36] Stoll, V.; Stewart, K. D.; Maring, C. J.; Muchmore, S.; Giranda, V.; Gu, Y. G.; Wang, G.; Chen, Y.; Sun, M.; Zhao, C.; Kennedy, A. L.; Madigan, D. L.; Xu, Y.; Saldivar, A.; Kati, W.; Laver, G.; Sowin, T.; Sham, H. L.; Greer, J.; Kempf, D. Influenza neuraminidase inhibitors: structure-based design of a novel inhibitor series. *Biochemistry* **2003**, *42*, 718-727.

[37] Wang, G. T.; Wang, S.; Gentles, R.; Sowin, T.; Maring, C. J.; Kempf, D. J.; Kati, W. M.; Stoll, V.; Stewart, K. D.; Laver, G. Design, synthesis, and structural analysis of inhibitors of influenza neuraminidase containing a 2,3-disubstituted tetrahydrofuran-5-carboxylic acid core. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 125–128.

[38] Wang, G. T.; Wang, S.; Gentles, R.; Sowin, T.; Maring, C. J.; Kempf, D. J.; Kati, W. M.; Stoll, V.; Stewart, K. D.; Laver, G. Design, synthesis, and structural analysis of inhibitors of influenza neuraminidase containing a 2,3-disubstituted tetrahydrofuran-5-carboxylic acid core. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 125–128.

[39] Jędrzejas, M. J.; Singh, S.; Brouillette, W. J.; Laver, W. G.; Air, G. M.; Luo, M. Structures of aromatic inhibitors of influenza virus neuraminidase. *Biochemistry* **1995**, *34*, 3144–3151.

[40] Taylor, N. R.; Cleasby, A.; Singh, O.; Skarzynski, T.; Wonacott, A. J.; Smith, P. W.; Sollis, S. L; Howes, P. D.; Cherry, P. C.; Bethell, R.; Colman, P.; Varghese, J. Dihydropyrancarboxamides related to zanamivir: a new series of inhibitors of influenza virus sialidases. 2. Crystallographic and molecular modeling study of complexes of 4-amino-4H-pyran-6-carboxamides and sialidase from influenza virus types A and B. *J. Med. Chem.* **1998**, *41*, 798–807.

[41] White, C. L.; Janakiraman, M. N.; Laver, W. G.; Philippon, C.; Vasella, A.; Air, G. M.; Luo, M. A sialic acid-derived phosphonate analog inhibits different strains of influenza virus neuraminidase with different efficiencies. *J. Mol. Biol.* **1995**, *245*, 623–634.

[42] Smith, P. W.; Sollis, S. L.; Howes, P. D.; Cherry, P. C.; Starkey, I. D.; Cobley, K. N.; Weston, H.; Scicinski, J.; Merritt, A.; Whittington, A.; Wyatt, P.; Taylor, N.; Green, D.; Bethell, R.; Madar, S.; Fenton, R. J.; Morley, P. J.; Pateman, T.; Beresford, A. Dihydropyrancarboxamides related to zanamivir: a new series of inhibitors of influenza virus sialidases. 1. Discovery, synthesis, biological activity, and structure-activity relationships of 4-guanidino-and 4-amino-4H-pyran-6-carboxamides. *J. Med. Chem.* **1998**, *41*, 787–797.

[43] Jedrzejas, M. J.; Singh, S.; Brouillette, W. J.; Laver, W. G.; Air, G. M.; Luo, M. Structures of aromatic inhibitors of influenza virus neuraminidase. *Biochemistry* **1995**, *34*, 3144–3151.

[44] Finley, J. B.; Atigadda, V. R.; Duarte, F.; Zhao, J. J.; Brouillette, W. J.; Air, G. M.; Luo, M. Novel aromatic inhibitors of influenza virus neuraminidase make selective interactions with conserved residues and water molecules in the active site. *J. Mol. Biol.* **1999**, *293*, 1107–1119.

[45] Atigadda, V. R.; Brouillette, W. J.; Duarte, F.; Babu, Y. S.; Bantia, S.; Chand, P.; Chu, N.; Montgomery, J. A.; Walsh, D. A.; Sudbeck, E.; Finley, J.; Air, G. M.; Luo, M.; Laver, G. W. Hydrophobic benzoic acids as inhibitors of influenza neuraminidase. *Bioorg. Med. Chem.* **1999**, *7*, 2487–2497.

[46] Lommer, B. S.; Ali, S. M.; Bajpai, S. N.; Brouillette, W. J. ; Air, G. M.; Luo, M. A benzoic acid inhibitor induces a novel conformational change in the active site of Influenza B virus neuraminidase. *Acta Cryst.* **2004**, *D60*, 1017–1023.

[47] Lommer, B. S.; Ali, S. M.; Bajpai, S. N.; Brouillette, W. J. ; Air, G. M.; Luo, M. A benzoic acid inhibitor induces a novel conformational change in the active site of Influenza B virus neuraminidase. *Acta Cryst.* **2004**, *D60*, 1017–1023.

[48] Finley, J. B.; Atigadda, V. R.; Duarte, F.; Zhao, J. J.; Brouillette, W. J.; Air, G. M.; Luo, M. Novel aromatic inhibitors of influenza virus neuraminidase make selective interactions with conserved residues and water molecules in the active site. *J. Mol. Biol.* **1999**, *293*, 1107–1119.

[49] Atigadda, V. R.; Brouillette, W. J.; Duarte, F.; Ali, S. M.; Babu, Y. S.; Bantia, S.; Chand, P.; Chu, N.; Montgomery, J. A.; Walsh, D. A.; Sudbeck, E. A.; Finley, J.; Luo, M.; Air, G. M.; Laver, G. W. Potent inhibition of influenza sialidase by a benzoic acid containing a 2-pyrrolidinone substituent. *J. Med. Chem.* **1999**, *42*, 2332–2343.

[50] Finley, J. B.; Atigadda, V. R.; Duarte, F.; Zhao, J. J.; Brouillette, W. J.; Air, G. M.; Luo, M. Novel aromatic inhibitors of influenza virus neuraminidase make selective interactions with conserved residues and water molecules in the active site. *J. Mol. Biol.* **1999**, *293*, 1107–1119.

X. APPENDIX D

Supplementary Materials
Uracil derivative Part

Comment D1. Materials and crystallisation details.

Studied uracil derivatives were purchased from Sigma-Aldrich as 97-99% pure crystalline powders. All of these powders quickly dissolve in water and, upon evaporation at room temperature, form crystals suitable for X-ray examination. Crystallizations were adjusted in a set of different solvents such as methanol, ethanol, propanol, butanol, acetone and chloroform, at various evaporation surfaces. It occurred that the best crystals were grown from water and ethanol solutions. For the purpose of this study crystals obtained from water solution were examined.

Comment D2. X-ray measurement and structural analysis details for uracil derivatives.

1mU: formula: $C_5H_6N_2O_2$; molecular mass: 126.12 a.u.; appearance: colourless, unspecified shape; measurement temperature: T = 100(1) K; crystal system: orthorhombic; space group: *Ibam* (no. 72); unit cell dimensions: a = 13.1594(5) Å, b = 13.2174(9) Å, c = 6.1966(10) Å, $\alpha = \beta = \gamma = 90°$, V = 1077.79(19) Å^3; Z = 8; Z' = 1/2; density: d_{calc} = 1.554 g·cm^{-3}; $F(000)$ = 528; size: 0.10 × 0.14 × 0.27 mm^3; θ range for data collection: 2.18 – 36.02°; $(\sin\theta / \lambda)_{max}$ = 0.83 Å^{-1}; index ranges: $-21 \le h \le 18$, $-21 \le k \le 10$, $-8 \le l \le 10$; reflections collected / unique: 6354 / 1318 (R_{int} = 3.25%); completeness: > 96%; absorption coefficient: μ = 0.123 mm^{-1}; absorption correction method: multi-scan; refinement method: full-matrix least-squares on F^2; reflections with $|F^o|^2 > 2\sigma(|F^o|^2)$ / parameters / restraints: 1013 / 60 / 0; final R-factors ($|F^o|^2 > 2\sigma(|F^o|^2)$): $R[F]$ = 4.10%, $wR[F^2]$ = 4.10%; R-factors (all data): $R[F]$ = 11.68%, $wR[F^2]$ = 12.87%; goodness-of-fit (all data): $GooF$ = 1.06; largest residual density peak / hole: +0.51 / –0.33 e·Å^{-3}.

Comments: Hydrogen atoms bound to the ring are placed geometrically within riding model for their isotropic ADPs (d_{N-H} = 0.88 Å, d_{C-H} = 0.95 Å, $U_{iso}^{H} = 1.2 \cdot U_{eq}^{X}$). Methyl group hydrogen atoms positions were refined with riding ADPs (d_{C-H} = 0.98 Å, $U_{iso}^{H} = 1.5 \cdot U_{eq}^{C}$).

15dmU: formula: $C_6H_8N_2O_2$; molecular mass: 140.14 a.u.; appearance: colourless, unspecified shape; measurement temperature: T = 100(1) K; crystal system: monoclinic; space group: $P2_1/c$ (no. 14); unit cell dimensions: a = 7.1769(5) Å, b = 12.1325(7) Å, c = 7.5571(6) Å, $\alpha = \gamma = 90°$, β = 91.743(7)°, V = 657.72(8) Å^3; Z = 4; Z' = 1; density: d_{calc} = 1.415 g·cm^{-3}; $F(000)$ = 296; size: 0.31 × 0.24 × 0.16 mm^3; θ range for data collection: 3.30 – 45.77°; $(\sin\theta / \lambda)_{max}$ = 1.01 Å^{-1}; index ranges: $-14 \le h \le 14$, $-24 \le k \le 19$, $-15 \le l \le 12$; reflections collected / unique: 11143 / 5489 (R_{int} = 2.34%); completeness: > 97%; absorption coefficient: μ = 0.108 mm^{-1}; absorption correction method: multi-scan; refinement method: full-matrix least-squares on F^2; reflections with $|F^o|^2 > 2\sigma(|F^o|^2)$ / parameters / restraints: 3929 / 93 / 0; final R-factors ($|F^o|^2 > 2\sigma(|F^o|^2)$): $R[F]$ = 3.98%, $wR[F^2]$ = 5.72%; R-factors (all data): $R[F]$ = 11.60%, $wR[F^2]$ = 13.29%; goodness-of-fit (all data): $GooF$ = 1.09; largest residual density peak / hole: +0.61 / –0.31 e·Å^{-3}. Comments: Hydrogen atoms bound to the ring are placed geometrically within riding model for their isotropic ADPs (d_{N-H} = 0.88 Å, d_{C-H} = 0.95 Å, $U_{iso}^{H} = 1.2 \cdot U_{eq}^{X}$). Orientations of methyl groups were determined from Fourier maps. Methyl groups hydrogen atoms are with ridden ADPs (d_{C-H} = 0.98 Å, $U_{iso}^{H} = 1.5 \cdot U_{eq}^{C}$).

5fU: formula: $C_4H_3FN_2O_2$; molecular mass: 130.08 a.u.; appearance: colourless, unspecified shape; measurement temperature: T = 100(1) K;; crystal system: triclinic; space group: $P\bar{1}$ (no. 2); unit cell dimensions: a = 8.6224(4) Å, b = 9.1703(5) Å, c = 12.5769(8) Å, α = 99.340(5)°, β = 100.163(4)°, γ = 90.303(4)°, V = 965.29(9) Å^3; Z = 8; Z' = 4; density: d_{calc} = 1.790 g·cm^{-3}; $F(000)$ = 528; size: 0.19 × 0.09 × 0.07 mm^3; θ range for data collection: 3.26 – 30.68°; $(\sin\theta / \lambda)_{max}$ = 0.72 Å^{-1}; index ranges: $-12 \le h \le 11$, $-12 \le k \le 12$, $-18 \le l \le 11$; reflections collected / unique: 15162 / 3896 (R_{int} = 4.29%); completeness: > 72%; absorption coefficient: μ = 0.169 mm^{-1}; absorption correction method: multi-scan; refinement method: full-matrix least-squares on F^2; reflections with $|F^o|^2 > 2\sigma(|F^o|^2)$ / parameters / restraints: 3004 / 325 / 0; final R-factors ($|F^o|^2 > 2\sigma(|F^o|^2)$): $R[F]$ = 4.52%, $wR[F^2]$ = 6.51%; R-factors (all data): $R[F]$ = 11.00%, $wR[F^2]$ = 12.67%; goodness-of-fit (all data): $GooF$ = 1.06; largest residual density peak / hole:

+0.33 / −0.27 e·Å^{-3}. Comments: Hydrogen atoms are placed geometrically within riding model for their isotropic ADPs (d_{N-H} = 0.88 Å, d_{C-H} = 0.95 Å, $U_{iso}^{H} = 1.2 \cdot U_{eq}^{X}$).

2tU: formula: $C_4H_4N_2OS$; molecular mass: 128.15 a.u.; appearance: pale yellow, unspecified shape; measurement temperature: T = 100(1) K;; crystal system: triclinic; space group: $P\bar{1}$ (no. 2); unit cell dimensions: a = 4.2484(3) Å, b = 5.9704(3) Å, c = 10.5660(6) Å, α = 105.734(5)°, β = 94.894(5)°, γ = 91.898(5)°, V = 256.58(3) Å^3; Z = 1; Z' = 1; density: d_{calc} = 1.659 g·cm^{-3}; F(000) = 132; size: 0.40 × 0.11 × 0.06 mm^3; θ range for data collection: 3.55 – 36.48°; $(\sin\theta / \lambda)_{max}$ = 0.84 Å^{-1}; index ranges: $-7 \le h \le 7$, $-9 \le k \le 9$, $-17 \le l \le 17$; reflections collected / unique: 14783 / 2505 (R_{int} = 2.20%); completeness: > 99%; absorption coefficient: μ = 0.508 mm^{-1}; absorption correction method: multi-scan; refinement method: full-matrix least-squares on F^2; reflections with $|F^o|^2 > 2\sigma(|F^o|^2)$ / parameters / restraints: 2382 / 73 / 0; final R-factors ($|F^o|^2 > 2\sigma(|F^o|^2)$): $R[F]$ = 2.06%, $wR[F^2]$ = 2.17%; R-factors (all data): $R[F]$ = 5.96%, $wR[F^2]$ = 6.04%; goodness-of-fit (all data): $GooF$ = 1.09; largest residual density peak / hole: +0.67 / −0.29 e·Å^{-3}. Comments: Hydrogen atoms are placed geometrically within riding model for their isotropic ADPs (d_{N-H} = 0.88 Å, d_{C-H} = 0.95 Å, $U_{iso}^{H} = 1.2 \cdot U_{eq}^{X}$).

4tU: formula: $C_4H_4N_2OS$; molecular mass: 128.15 a.u.; appearance: yellowish, unspecified shape; measurement temperature: T = 100(1) K;; crystal system: monoclinic; space group: $P2_1/c$ (no. 14); unit cell dimensions: a = 4.2588(12) Å, b = 12.419(4) Å, c = 9.844(3) Å, α = γ = 90°, β = 91.289(9)°, V = 520.5(3) Å^3; Z = 4; Z' = 1; density: d_{calc} = 1.635 g·cm^{-3}; F(000) = 264; size: 0.19 × 0.10 × 0.07 mm^3; θ range for data collection: 2.64 – 30.02°; $(\sin\theta / \lambda)_{max}$ = 0.70 Å^{-1}; index ranges: $-5 \le h \le 5$, $-17 \le k \le 17$, $-13 \le l \le 13$; reflections collected / unique: 10911 / 1479 (R_{int} = 6.29%); completeness: > 97%; absorption coefficient: μ = 0.501 mm^{-1}; absorption correction method: multi-scan; refinement method: full-matrix least-squares on F^2; reflections with $|F^o|^2 > 2\sigma(|F^o|^2)$ / parameters / restraints: 1031 / 73 / 0; final R-factors ($|F^o|^2 > 2\sigma(|F^o|^2)$): $R[F]$ = 3.70%, $wR[F^2]$ = 6.35%; R-factors (all data): $R[F]$ = 8.91%, $wR[F^2]$ = 9.81%; goodness-of-fit (all data): $GooF$ = 1.06; largest residual density peak / hole: +0.37 / −0.32 e·Å^{-3}. Comments: Hydrogen atoms are placed geometrically within riding model for their isotropic ADPs (d_{N-H} = 0.88 Å, d_{C-H} = 0.95 Å, $U_{iso}^{H} = 1.2 \cdot U_{eq}^{X}$).

24dtU: formula: $C_4H_4N_2S_2$; molecular mass: 144.21 a.u.; appearance: orange, unspecified shape; measurement temperature: T = 100(1) K; crystal system: monoclinic;

space group: $P2_1/c$ (no. 14); unit cell dimensions: a = 4.1926(4) Å, b = 14.5355(11) Å, c = 9.7860(9) Å, α = γ = 90°, β = 99.772(5)°, V = 587.72(9) Å^3; Z = 4; Z' = 1; density: d_{calc} = 1.630 g·cm^{-3}; F(000) = 296; size: 0.07 × 0.18 × 0.18 mm^3; θ range for data collection: 2.54 – 29.64°; $(\sin\theta / \lambda)_{max}$ = 0.70 Å^{-1}; index ranges: $-5 \le h \le 5$, $-20 \le k \le 19$, $-13 \le l \le 13$; reflections collected / unique: 16082 / 1610 (R_{int} = 5.14%); completeness: > 99%; absorption coefficient: μ = 0.784 mm^{-1}; absorption correction method: multi-scan; refinement method: full-matrix least-squares on F^2; reflections with $|F^o|^2 > 2\sigma(|F^o|^2)$ / parameters / restraints: 1329 / 73 / 0; final R-factors ($|F^o|^2 > 2\sigma(|F^o|^2)$): $R[F]$ = 2.98%, $wR[F^2]$ = 3.99%; R-factors (all data): $R[F]$ = 6.69%, $wR[F^2]$ = 7.15%; goodness-of-fit (all data): $GooF$ = 1.05; largest residual density peak / hole: +0.35 / –0.35 e·Å^{-3}. Comments: Hydrogen atoms are placed geometrically within riding model for their isotropic ADPs (d_{N-H} = 0.88 Å, d_{C-H} = 0.95 Å, $U_{iso}^{H} = 1.2 \cdot U_{eq}^{X}$).

6m2tU: formula: $C_5H_6N_2OS$; molecular mass: 142.18 a.u.; appearance: colourless, unspecified shape; measurement temperature: T = 100(1) K;; crystal system: monoclinic; space group: $P2_1/c$ (no. 14); unit cell dimensions: a = 4.34050(10) Å, b = 14.4580(5) Å, c = 9.5743(3) Å, α = γ = 90°, β = 90.758(3)°, V = 600.78(3) Å^3; Z = 4; Z' = 1; density: d_{calc} = 1.572 g·cm^{-3}; F(000) = 296; size: 0.05 × 0.05 × 0.18 mm^3; θ range for data collection: 2.55 – 31.73°; $(\sin\theta / \lambda)_{max}$ = 0.74 Å^{-1}; index ranges: $-6 \le h \le 6$, $-21 \le k \le 20$, $-13 \le l \le 13$; reflections collected / unique: 5672 / 1961 (R_{int} = 3.34%); completeness: > 95%; absorption coefficient: μ = 0.443 mm^{-1}; absorption correction method: multi-scan; refinement method: full-matrix least-squares on F^2; reflections with $|F^o|^2 > 2\sigma(|F^o|^2)$ / parameters / restraints: 1434 / 83 / 0; final R-factors ($|F^o|^2 > 2\sigma(|F^o|^2)$): $R[F]$ = 3.65%, $wR[F^2]$ = 5.95%; R-factors (all data): $R[F]$ = 10.53%, $wR[F^2]$ = 13.33%; goodness-of-fit (all data): $GooF$ = 1.15; largest residual density peak / hole: +0.53 / –0.46 e·Å^{-3}. Comments: Hydrogen atoms bound to the ring are placed geometrically within riding model for their isotropic ADPs (d_{N-H} = 0.88 Å, d_{C-H} = 0.95 Å, $U_{iso}^{H} = 1.2 \cdot U_{eq}^{X}$). Methyl group hydrogen atoms is with ridden ADPs (d_{C-H} = 0.98 Å, $U_{iso}^{H} = 1.5 \cdot U_{eq}^{C}$).

Table D3. *R*-factors and related parameters after TAAM refinements (all calculated using reflections with $|F^o|^2 \geq 3\sigma(|F^o|^2)$).

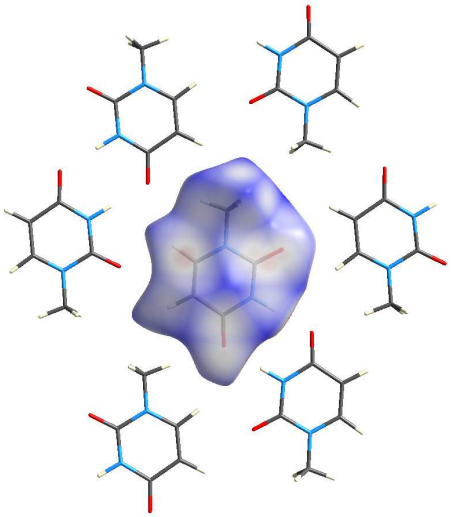
Descriptor	**1mU**	**15dmU**	**5fU**	**2tU**	**4tU**	**24dtU**	**6m2tU**
$R[F]$	2.40%	2.81%	2.96%	1.20%	2.80%	2.28%	2.41%
$wR[F]$	3.04%	3.08%	3.41%	1.85%	2.77%	2.47%	2.25%
$R[F^2]$	4.50%	4.22%	4.80%	1.98%	3.81%	3.03%	3.12%
$wR[F^2]$	5.95%	6.19%	6.65%	3.67%	5.40%	4.86%	4.54%
$\Delta\varrho_{max}$ / e·Å^{-3}	+0.30	+0.23	+0.39	+0.08	+0.31	+0.23	+0.24
$\Delta\varrho_{min}$ / e·Å^{-3}	−0.28	−0.31	−0.39	−0.11	−0.21	−0.34	−0.25
GooF	1.55	1.39	0.99	1.31	1.16	1.21	0.97

Comment D4. Hirshfeld surfaces and fingerprint plots.

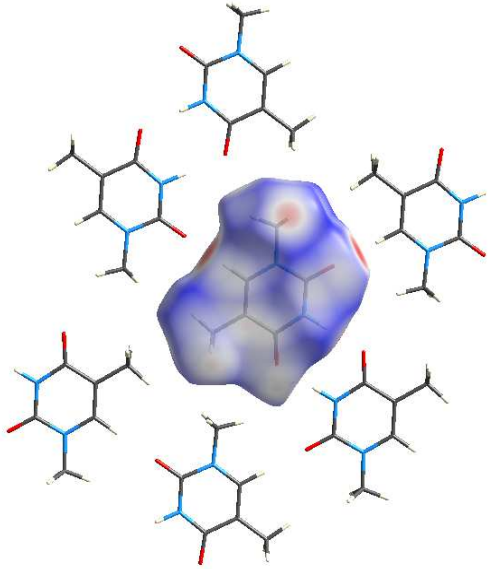
To compare intermolecular interactions in the investigated solvates we used fingerprint plots generated from the so-called Hirshfeld surfaces. A following weighing function was applied to define Hirshfeld surface:

$$w_{\mathrm{M}}(\mathbf{r}) = \frac{\sum_{k\in \mathrm{M}} \rho_k(\mathbf{r}-\mathbf{r}_k)}{\sum_{k\in \mathrm{C}} \rho_k(\mathbf{r}-\mathbf{r}_k)}$$

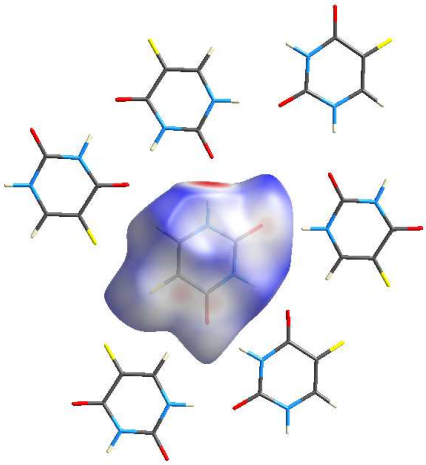
where M is molecule, C is the whole crystal, ρ_k is the spherically averaged atomic electron density of the k-th atom (centred at point $\mathbf{r}_k$). This weighing function defines the Hirshfeld surface for molecule M while $w_{\mathrm{M}}(\mathbf{r}) = 0.5$ for every $\mathbf{r}$ point at the surface. Within the model, the promolecule electron density dominates over the procrystal electron density. Consequently, it is possible to map different properties on such derived surfaces: these describing the shape of the surface (*e.g.* curvedness) and also those related to distances: d_e – external distance from the Hirshfeld surface to an atom belonging to the closest molecule outside the surface, d_i – internal distance from the surface to an atom inside the surface and d_{norm}, which combines both d_e and d_i, each normalised by the van der Waals (vdW) radius for particular atoms involved in the close proximity to the surface. When de and di are calculated for each point of the Hirshfeld surface, a 2D fingerprint plot (d_e *vs.* d_i) can be created. As on Hirshfeld surfaces the closest contacts between a surface point and a particular atom both inside and outside the surface can be illustrated, one can easily compute relative contributions to the Hirshfeld surface area for the various close intermolecular interactions. All of interactions sum up to 100%, so a percentage contribution of each particular interaction type can be estimated. The presented figures of Hirshfeld surface and fingerprint plots (Figure D5) were prepared in *CRYSTALEXPLORER* program which employs standardised hydrogen atom positions.



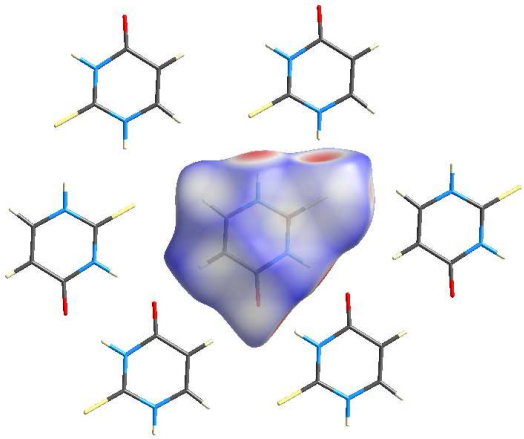
(*a*)



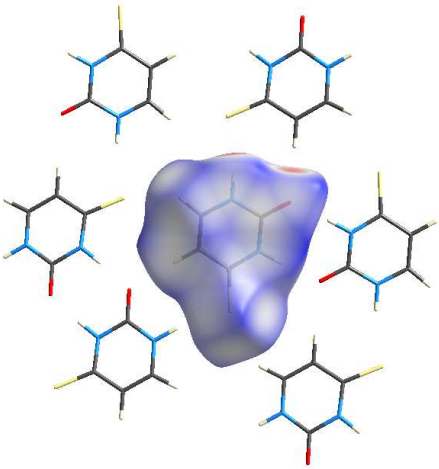
(*b*)



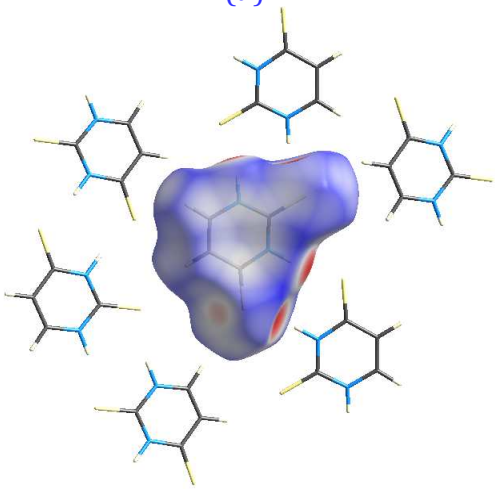
(*c*)



(*d*)



(*e*)



(*f*)

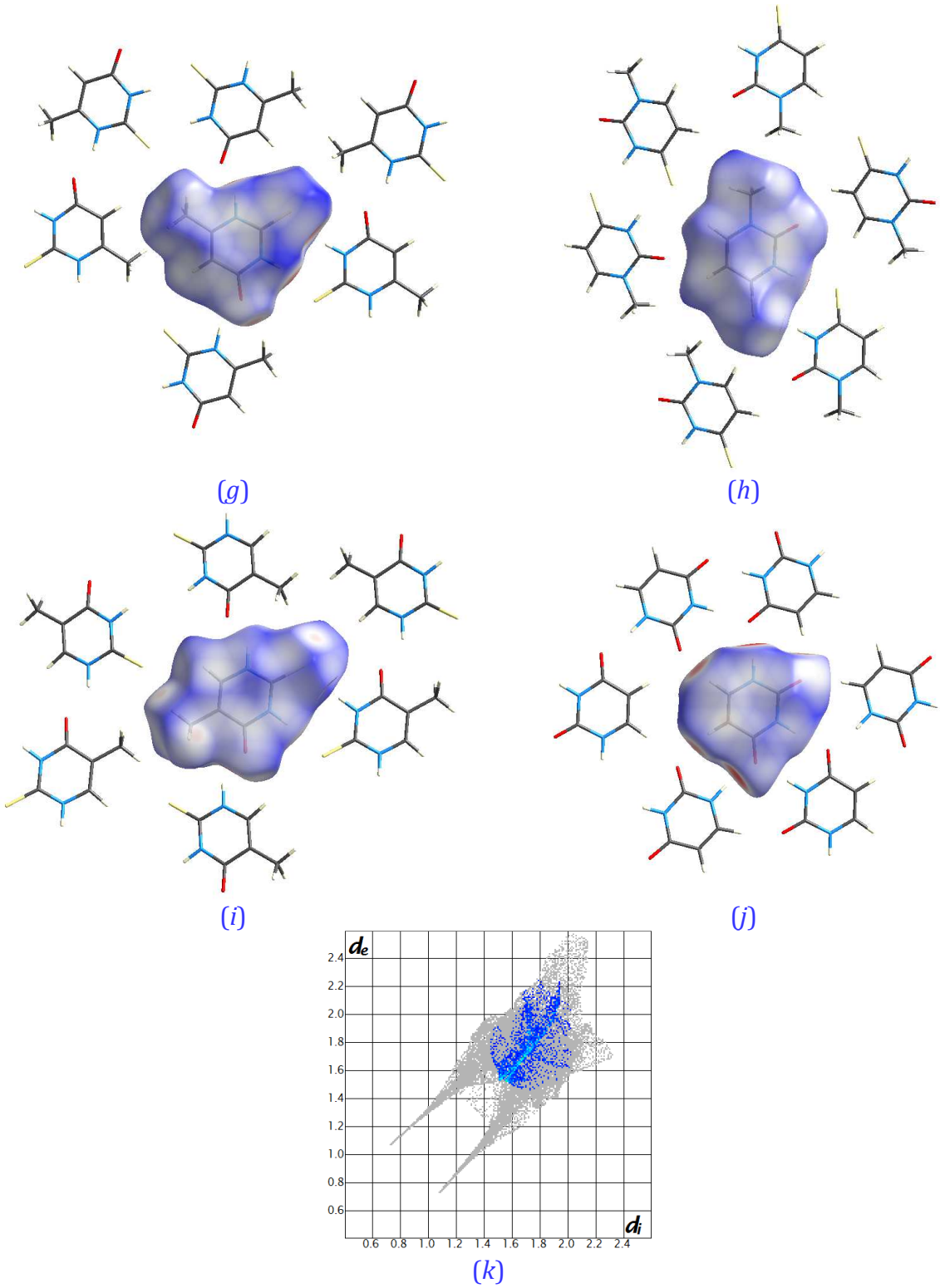


Figure D5. (*a-j*) Exemplary Hirsfeld surface plots with mapped dnorm property for **1mU**, **14dmU**, **5fU** (one molecule for clarity), **2tU**, **4tU**, **24dtU**, **6m2tU**, **1m4tU**, **5m2tU**, **U** compounds, sequentially (together with the neighbouring molecules). (*k*) Hirshfeld surface fingerprint plot showing relative percentage contributions (7.7%) of F…F interactions to the Hirshfeld surface coverage for **5fU** crystal structure (one molecule for clarity).

Comment D6. Uracil ring deviation from planarity.

A sequence showing how the studied uracil derivatives deviate from planarity (defined as the maximal distance between atoms from aromatic rings and the least-square fitted plane): **1mU** > **U** > **6m2tU** > **24dtU** > **5m2tU** > **5fU** > **15dmU** > **2tU** > **1m4tU** > **4tU**.

Comment D7. Dimer energy calculations.

To gain a deeper understanding of the lattice energy results, single-point energy calculations were carried out for two types of dimers present in the structure of 2tU: bound by (*a*) two N–H…O hydrogen bonds (**A** motif) and by (*b*) two N–H…S interactions (**B** motif). In both cases we ran a geometry-constrained scan (*i.e.*, only one parameter was varied) of the N–H bond length ranging from 0.95 Å to 1.15 Å with a step of 0.01 Å. For each point the total energy of a dimer was evaluated at the B3LYP/aug-cc-pVDZ level of theory (using the *GAUSSIAN* program). Both energy curves are illustrated in Figure D8 and they clearly show a direct relationship between the energy of a dimer and N–H bond length. The only difference between N–H…O and N–H…S interactions are the locations of the corresponding minima. In the case of sulphur-bound hydrogen bond the minimum is found at slightly shorter N–H distance. In conclusion, even small differences in the N–H distance (*ca.* 0.1 Å) cause significant changes of the total energy, reaching the values of up to 14 – 16 kJ·mol^{-1} per one molecule. Presumably, the energy change would be less pronounced if one could analyze the dependence of the counterpoise-corrected interaction energy. Nevertheless, this result, being in a close relation to the cohesive energy calculations, shows the importance of the N–H standardization.

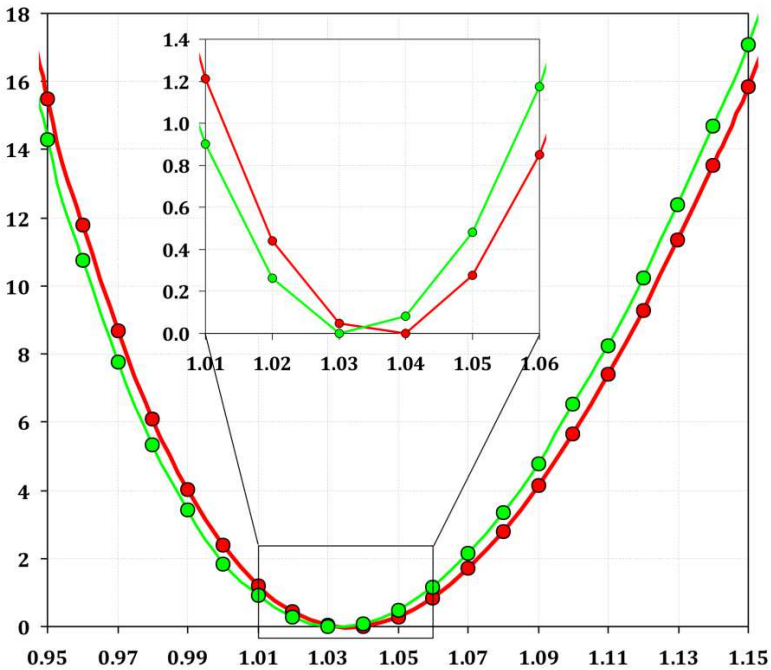


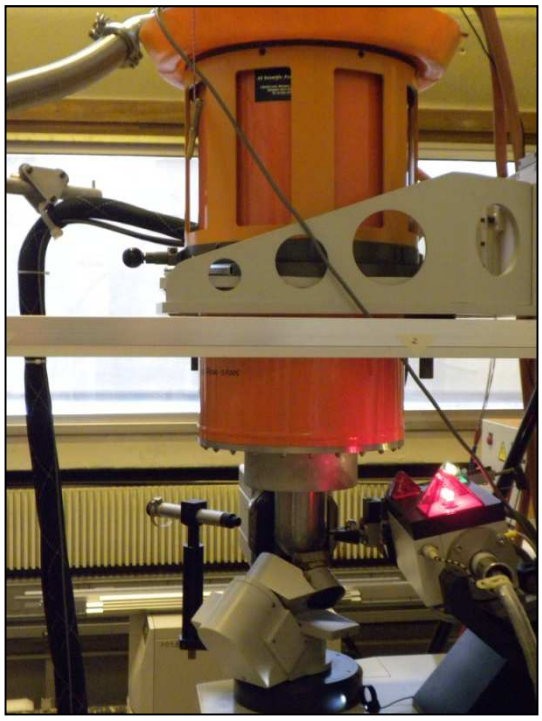
Figure D8. Constrained energy scan for N–H bond distance for two types of dimeric moieties (**A** and **B** motifs) present in the **2tU** crystal structure. Red line – N–H…O bound dimer, green line – N–H…S bound.

Table D9. Atom charges fitted to the electrostatic potential derived from charge density grid calculated at the DFT(B3LYP)/aug-cc-pVDZ level of theory.

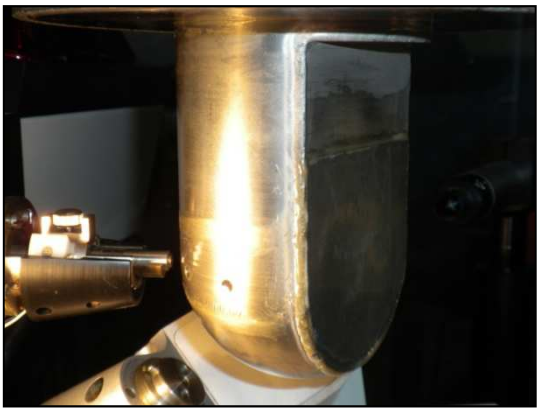
Atom	*Compound*									
	U	**15dmU**	**1mU**	**5fU**	**2tU**	**4tU**	**24dtU**	**6m2tU**	**1m4tU**	**5m2tU**
N1	−0.428	−0.006	−0.014	−0.450	+0.078	−0.384	+0.160	−0.094	+0.157	+0.134
C2	+0.722	+0.662	+0.713	+0.736	−0.121	+0.598	−0.232	−0.053	+0.472	−0.174
N3	−0.496	−0.621	−0.645	−0.548	−0.013	−0.194	+0.397	−0.086	−0.088	+0.006
C4	+0.749	+0.662	+0.807	+0.636	+0.597	+0.085	−0.105	+0.626	−0.030	+0.424
C5	−0.478	−0.021	−0.505	+0.058	−0.347	−0.112	−0.006	−0.516	−0.053	+0.136
C6	+0.116	−0.248	−0.038	−0.060	−0.098	−0.018	−0.175	+0.357	−0.207	−0.291
O2/S2	−0.582	−0.575	−0.580	−0.586	−0.322	−0.525	−0.300	−0.324	−0.533	−0.317
O4/S4	−0.594	−0.572	−0.599	−0.538	−0.561	−0.308	−0.297	−0.570	−0.317	−0.526
H1	+0.326	–	–	+0.357	+0.202	+0.346	+0.191	+0.224	–	+0.198
H3	+0.324	+0.374	+0.370	+0.359	+0.213	+0.252	+0.088	+0.228	+0.220	+0.224
H5	+0.202	–	+0.216	−0.158	+0.195	+0.116	+0.105	+0.219	+0.116	–
H6	+0.140	+0.210	+0.179	+0.194	+0.178	+0.153	+0.173	–	+0.182	+0.207

XI. APPENDIX E

Supplementary Materials
6-methyl-2-thiouracil Part



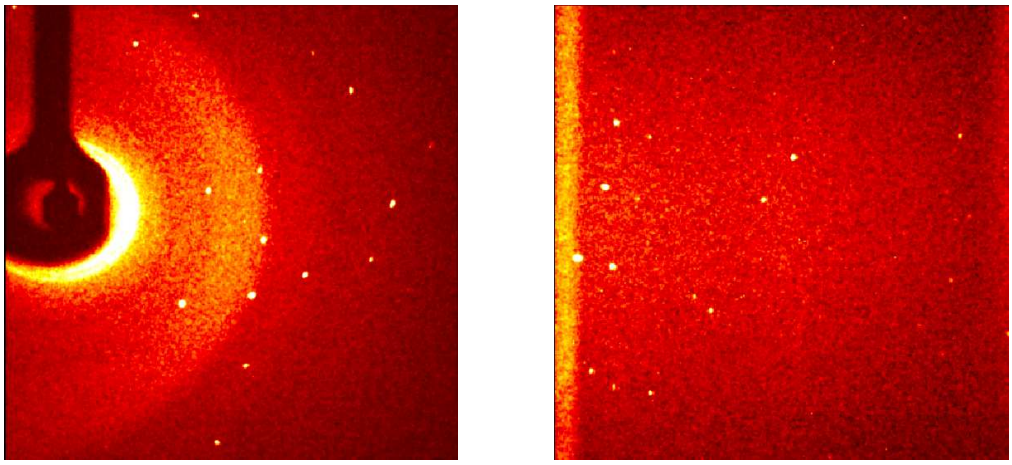
(*a*)



(*c*)

(*b*)

Figure E1. Pictures of the He-temperature closed-circuit setup. (*a*) Instrumentation in general. (*b*,*c*) Hellium chamber with the beryllium window and master magnet visible.



(*a*) (*b*)

Figure E2. Example of low- (*a*) and high-resolution (*b*) frames collected at the He-setup. Problems with the background estimation are clearly visible.

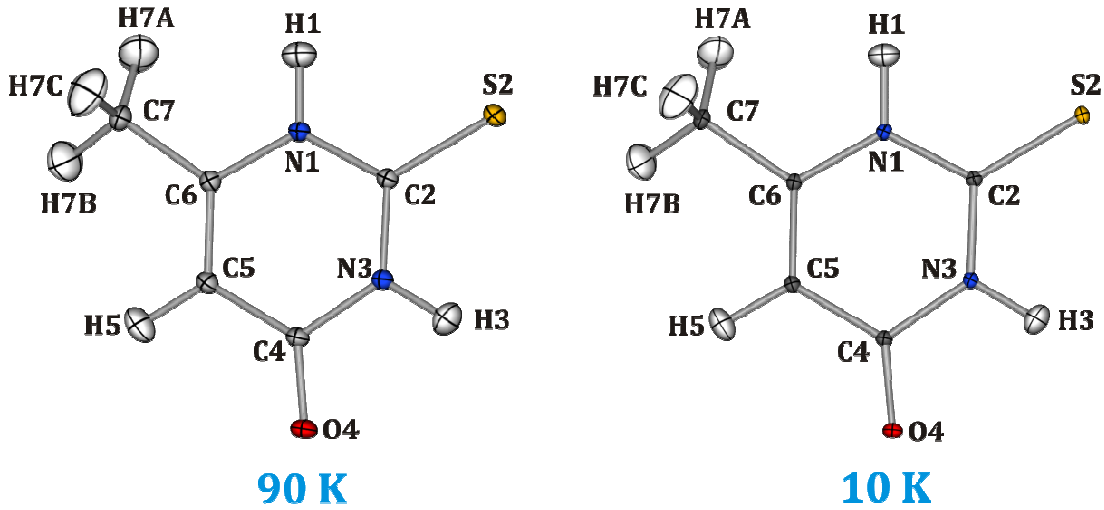


Figure E3. Labelling of atoms and estimation of their thermal motion parameters, ADPs (50% probability level), for 6m2tU at 90 K after final multipole refinement, and for the 10 K dataset refined with the 90 K case final strategy.

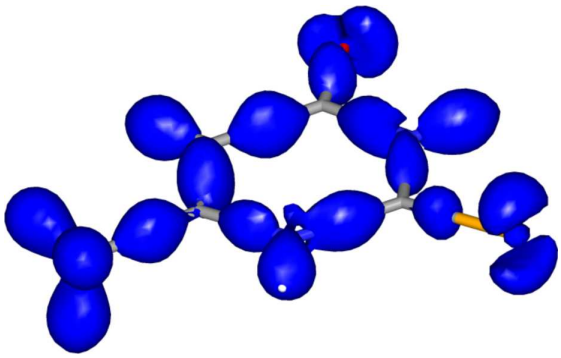


Figure E4. Static deformation density isosurface (drawn at 0.25 e·Å^{-3}) for the model with freely refined sulphur atom deformation parameters (90 K data set).

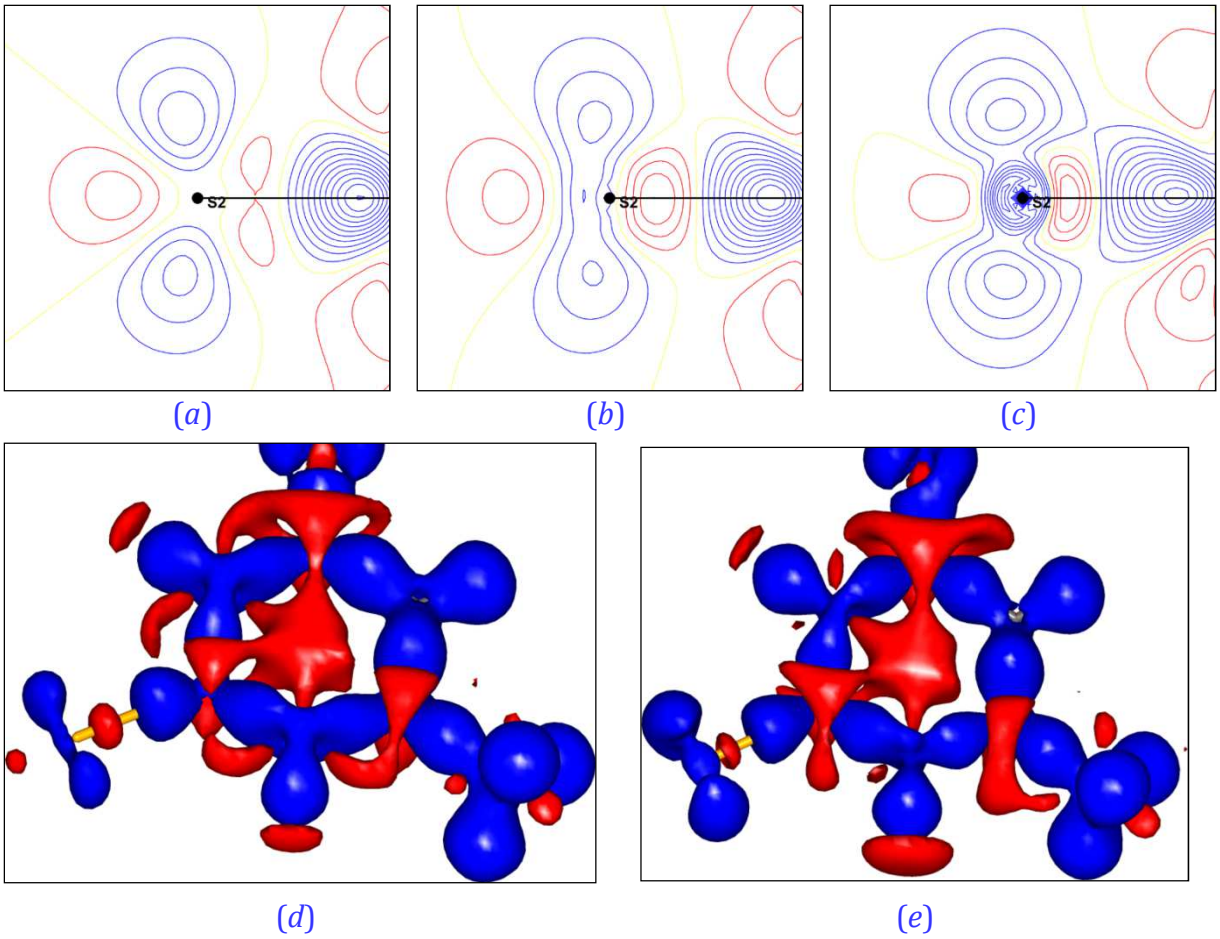


Figure E5. Static deformation density maps in the vicinity of the sulphur atom for (*a*) the theoretical model (*i.e.*, based on theoretical scattering factors) with standard sulphur scattering factor (ζ = 3.851 a_0^{-1}; n_l = 4, 4, 4, 4), (*b*) the theoretical model with the modified sulphur scattering factor (ζ = 3.851 a_0^{-1}; n_l = 2, 4, 6, 8), and (*c*) the final experimental model. 3D static deformation density maps for (*d*) the theoretical model with the modified sulphur atom scattering factor and for (*e*) the final experimental data are also presented (isosurfaces drawn at ±0.11 e·Å^{-3}, blue – positive, red – negative).

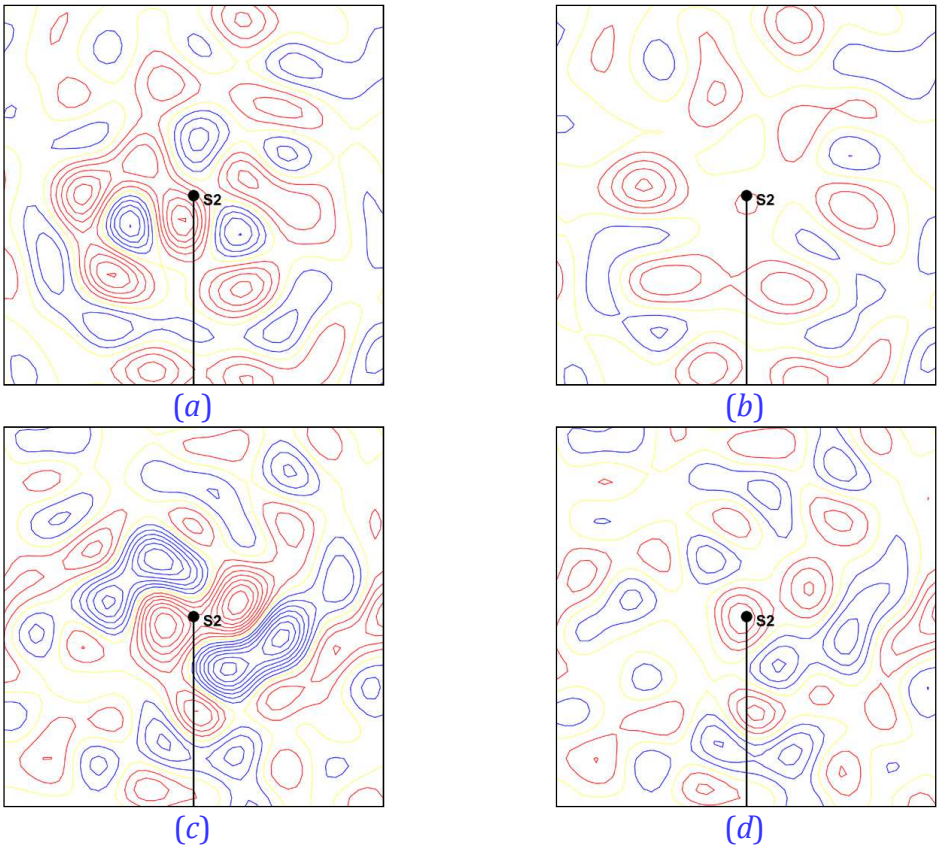


Figure E6. High order (sin $\theta/\lambda \geq 0.8$ Å^{-1}, $|F^{o}|^2 \geq 3\sigma(|F^{o}|^2)$) Fourier residual density maps before (*a*,*c*) and after (*b*,*d*) the sulphur atom anharmonic motion refinement; (*a*,*b*) electron density maps for 90 K and (*c*,*d*) for 10 K dataset; (*b*,*d*) models include the refined sulphur atom deformation and 4th order anharmonic thermal Gram-Charlier parameters (*ZX* plane, definition of initial local axes: X_i = S2→C2, Y_i = S2→N1; 0.02 e·Å^{-3} contours for 90 K and 0.05 e·Å^{-3} for 10 K; blue solid lines – positive, red solid lines – negative).

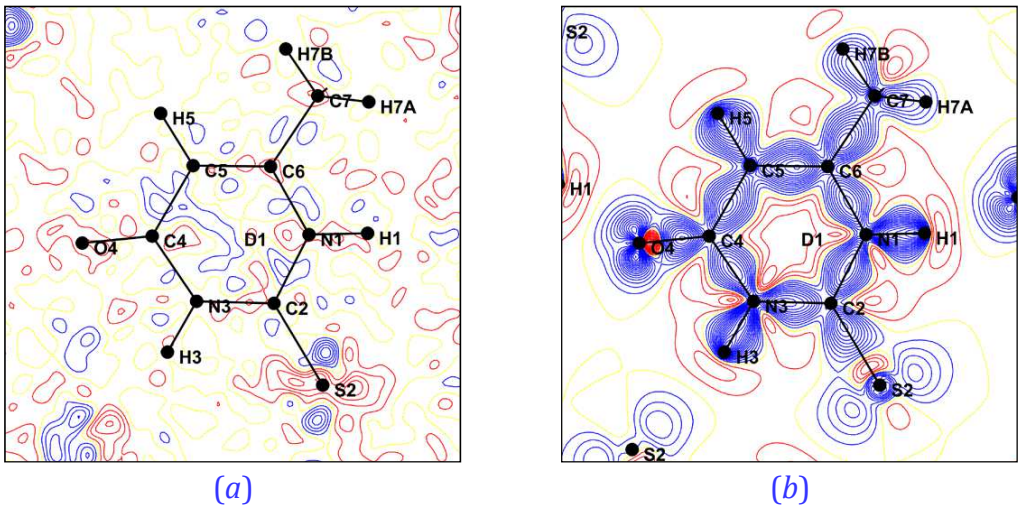
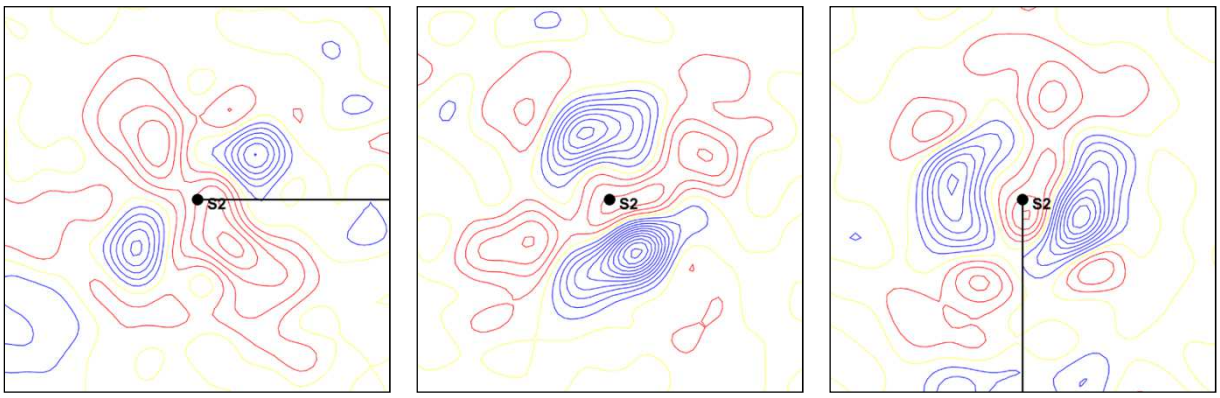
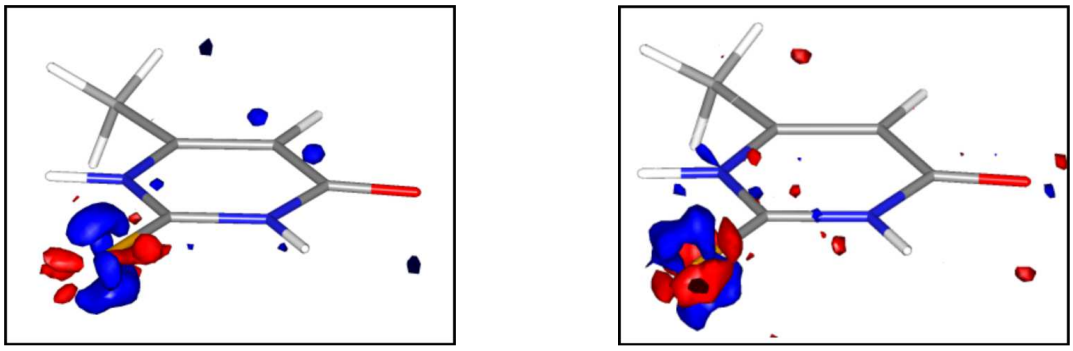


Figure E7. 90 K dataset final charge density model: (*a*) Residual density map in the plane of the phenyl ring. (*b*) Static deformation density map in the plane of the phenyl ring. Contours at 0.05 e·Å^{-1}, blue solid lines – positive, red solid lines – negative.



(*a*) (*b*) (*c*)

Figure E8. 90 K dataset final charge density model. Residual density maps of the region in the vicinity of the sulphur atom along 3 perpendicular directions: (*a*) *XY*, (*b*) *YZ* and (*c*) *ZX* plane (contours at 0.05 e·Å^{-1}, blue solid lines – positive, red solid lines – negative; definition of initial local axes: X_i = S2→C2, Y_i = S2→N3).



(*a*) (*b*)

Figure E9. 3D residual density maps for the (*a*) 90 K and (*b*) 10 K data sets (isosurfaces drawn at 0.15 e·Å^{-3} and 0.25 e·Å^{-3} for 90 K and 10 K, respectively; blue – positive, red – negative, final multipole models).

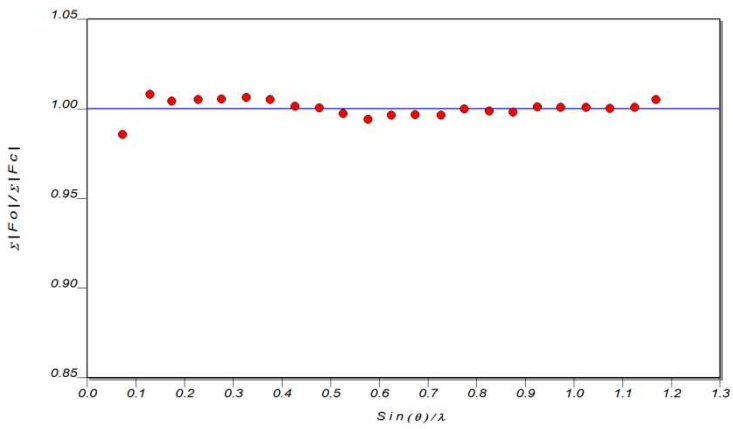
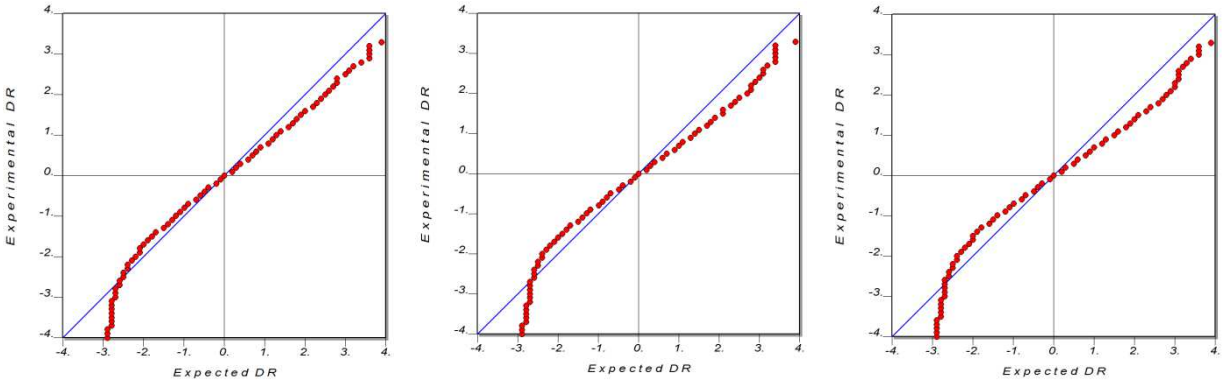


Figure E10. Scale plot for the final multipole model for data collected at 90 K.



(*a*) (*b*) (*c*)

Figure E11. Normal probability plots for (*a*) final model, (*b*) harmonic model with refined sulphur atom parameters, and (*c*) anharmonic model; all models concern the 90 K dataset. Slope < 1 indicates a systematic error.

Comment E12. Theoretical charge density evaluation for **6m2tU**.

Single-point periodic calculations were performed using the *CRYSTAL* code at the DFT(B3LYP)/pVTZ level of theory. The calculation was based on the 3D optimized geometry with the unit cell parameters fixed at the 90 K experimental values. Upon the convergence on energy, the periodic wave function was obtained and used to generate the symmetry-unique set of static theoretical structure factors up to the resolution of 1.2 $Å^{-3}$. Subsequently, a standard charge density refinement was carried out with respect to all reflections. During the refinement, all multipole and contraction-expansion parameters assigned to all atoms, were iteratively varied. According to standard procedures, during the refinement coordinates were kept fixed, thermal motion was set to zero for all atoms, unit weights were applied, whereas scale factor was not refined. The results of the topological analysis of the 'theoretical charge density model' are summed up in Table E13 and E14.

Table E13. Selected QTAIM parameters of covalent bonds at BCPs (R – distance between bonded atoms; R_1 and R_2 – distances from 1st and 2nd atom to the bond critical point, respectively; ρ – electron density; ε – bond ellipticity) calculated for the theoretical charge density model.

Bond	R / Å	R_1 / Å	R_2 / Å	$\rho(\mathbf{r}_{BCP})$ / e·Å^{-3}	$\nabla^2\rho(\mathbf{r}_{BCP})$ / e·Å^{-5}	ε
S2–C2	1.690	0.724	0.966	1.32	−4.3	0.02
O4–C4	1.237	0.787	0.451	2.70	−27.8	0.06
N1–C2	1.360	0.786	0.573	2.19	−20.3	0.14
N1–C6	1.374	0.789	0.585	2.09	−17.1	0.11
N3–C4	1.390	0.795	0.596	2.04	−16.5	0.14
N3–C2	1.357	0.787	0.570	2.19	−20.4	0.14
C4–C5	1.427	0.745	0.682	1.97	−14.8	0.15
C5–C6	1.358	0.648	0.709	2.21	−18.0	0.20
C6–C7	1.493	0.794	0.699	1.70	−10.8	0.04
N1–H1	1.030	0.760	0.270	2.17	−34.5	0.05
N3–H3	1.030	0.755	0.275	2.18	−33.0	0.04
C5–H5	1.083	0.694	0.389	1.91	−21.0	0.06
C7–H7A	1.077	0.684	0.393	1.88	−20.0	0.00
C7–H7B	1.077	0.679	0.398	1.87	−19.6	0.01
C7–H7C	1.077	0.684	0.393	1.83	−18.6	0.01

Table E14. Selected QTAIM parameters for weak intermolecular interactions at the positions of respective BCPs (G – kinetic energy density; V – potential energy density) calculated for the theoretical charge density model. Differences in bond paths are shown in bold.

Interaction	*Motif*	R / Å	R_1 / Å	R_2 / Å	$\rho(\mathbf{r}_{BCP})$ / e·Å^{-3}	$\nabla^2\rho(\mathbf{r}_{BCP})$ / e·Å^{-5}	$G(\mathbf{r}_{BCP})$ / kJ·mol^{-1}·a$_0^{-3}$	$V(\mathbf{r}_{BCP})$ / kJ·mol^{-1}·a$_0^{-3}$
O4…H1$^{\#1}$	**C**	1.765	0.632	1.134	0.28	2.2	77.5	−94.5
O4…H7A$^{\#1}$	**C**	2.379	1.006	1.374	0.08	1.2	25.4	−19.1
H5…S2$^{\#1}$	**C**	3.266	2.027	1.243	0.03	0.2	5.2	−3.8
S2…H3$^{\#2}$	**D**	2.359	1.560	0.800	0.16	0.8	29.8	−36.9
S3…N3$^{\#3}$	**S**	3.504	1.639	1.865	0.04	0.5	10.6	−7.5
N1…C4$^{\#3}$	**S**	3.220	1.592	1.629	0.04	0.5	11.1	−8.0
H7A…C5$^{\#3}$	**S**	2.963	1.724	1.253	0.03	0.5	9.2	−6.1
S2…H5$^{\#4}$	**M1**	3.087	1.872	1.269	0.04	0.5	10.8	−7.8
N1…**C4**$^{\#4}$	**M1**	3.368	1.680	1.710	0.03	0.4	9.0	−6.1
S2…H7B$^{\#4}$	**M1**	3.208	1.960	1.249	0.03	0.3	6.6	−4.6
S2…H7A$^{\#5}$	**M2**	3.211	1.906	1.306	0.02	0.3	5.7	−3.8
S2…H7C$^{\#6}$	**M3**	2.995	1.831	1.172	0.05	0.4	9.2	−7.7
H7B…H7B$^{\#7}$	**H1**	2.337	1.169	1.169	0.04	0.7	13.6	−9.1
H7B…**H7A**$^{\#8}$	**H2**	2.637	1.281	1.386	0.02	0.4	6.9	−4.3

Comment E15. 6m2tU charge density and structural models.

Differently evaluated structural models lead to slightly modified molecular geometries. These were compared by means of cohesive energy, molecular energy, and the total energy of the elementary cell contents, derived within the *CRYSTAL09* program. The crystal lattice itself was also optimised in three ways at the DFT(B3LYP) level of theory, *i.e.* with the (i) 6-31G** all-electron basis set applied and fixed unit cell parameters (**OPT3**), (ii) pVTZ basis set and fixed unit cell parameters (**OPT2**), and finally with the (iii) pVTZ basis set and variable unit cell parameters (**OPT1**). All the described results together with those obtained after crude IAM refinement and various TAAM refinements are compared in Tables E14 and E15. Generally, the most significant discrepancies among the studied cases are caused by the accuracy of the determined hydrogen atom positions and X–H bond lengths, as observed previously. The basis set change from 6-31G** to pVTZ results in up to 2 $kJ \cdot mol^{-1}$ difference on the cohesive energy value, which is comparable to the energetic change coming from the slightly different unit cell parameters (obtained at various temperatures, or after optimisation). This is, however, in the range of the method accuracy. Furthermore, all the different charge density models lead to very similar crystal structures, as reflected in the consistent energetic results. Much difference in the obtained molecular energy is observed solely for the **6m2tU** geometry resulting from the charge density refinement with no symmetry imposed on the multipole populations. Additionally, for the comparative purposes, similar energy set was calculated for a different tautomer of **6m2tU** (tautomer **II**), and for the analogue to the **oxo-thione 6m2tU** hypothetical crystal lattice it might form.

Table E16. Studied **6m2tU** geometries.

Model	*Description*
Remark: *Optimised crystal structures of the oxo–thione tautomeric form of* ***6m2tU***	
OPT1	*CRYSTAL*-optimised crystal lattice of **6m2tU**; charge density model used as a starting geometry; DFT(B3LYP)/pVTZ level of theory; atomic positions and cell parameters optimised.
OPT2	*CRYSTAL*-optimised crystal lattice of **6m2tU**; charge density model used as a starting geometry; DFT(B3LYP)/pVTZ level of theory; atomic positions optimised, cell parameters kept fixed.
OPT3	*CRYSTAL*-optimised crystal lattice of **6m2tU**; charge density model used as a starting geometry; DFT(B3LYP)/6-31G** level of theory; atomic positions optimised, cell parameters kept fixed.
Remark: *Experimental structures of* ***6m2tU****: various charge density models*	
F	Final **6m2tU** crystal structure obtained after charge density refinement described and further analysed in the manuscript.
E1[a]	Crystal structure obtained after charge density refinement with the sulphur atom multipolar parameters refined (but not κ and κ'), harmonic atom motion model.
E2[a]	Crystal structure obtained from the 10 K dataset after charge density refinement with the sulphur atom multipolar parameters refined (but not κ and κ'), harmonic atom motion model.
E3[a]	Crystal structure obtained after charge density refinement with the sulphur atom multipolar and anharmonic motion parameters refined (but not κ and κ').
E4[a]	Crystal structure obtained after charge density refinement with the sulphur atom multipolar and contraction-expansion parameters refined, harmonic atom motion model.
E5[a]	Crystal structure obtained after charge density refinement with the sulphur atom multipolar and anharmonic motion parameters refined (but not κ and κ'), with no symmetry constraints applied to multipolar populations.
Remark: *Various structural refinement types*	
I0	Crystal structure obtained after standard IAM refinement carried out in the *JANA* program.
I1	Crystal structure obtained after standard IAM refinement carried out in the *JANA* program; the hydrogen atom positions refined analogically as in the TAAM procedure.
I2	Crystal structure obtained after standard IAM refinement carried out in the *JANA* program; the X–H bond lengths set to the neutron-normalized distances.
T1	Crystal structure obtained after standard TAAM procedure.
T2	Crystal structure obtained after the TAAM procedure with the anisotropic thermal motion of hydrogen atoms modelled within the *SHADE* server.
T3	Crystal structure obtained after the TAAM procedure with the freely refined hydrogen atom positions.
Remark: *Tautomer* ***II*** *structures*	
T-II-1	Crystal structure of the possible 100% second tautomer as refined with the *JANA* program and *MOPRO* refinement. *JANA* suggested the presence of the possible 15% content of the second tautomer in the crystal structure (the refinement did not fully converged and, thus, is not fully reliable).
T-II-2	*CRYSTAL*-optimised crystal lattice of the **6m2tU** second tautomer; DFT(B3LYP)/pVTZ level of theory; atomic positions optimised, cell parameters kept fixed.

[a] If not stated otherwise, κ and κ' are always fixed at the UBDB2011 values

Table E17. Energy comparison of differently refined data (E_{mol} – energy of the molecule, E_{bulk} – energy of the unit cell extracted from the bulk, E_{coh} – cohesive energy, ΔE_i – differences of respective *i*-th energy *vs.* reference values of **OPT1**).

Model	E_{mol} / kJ·mol^{-1}	E_{bulk} / kJ·mol^{-1}	E_{coh} / kJ·mol^{-1}	ΔE_{mol} / kJ·mol^{-1}	ΔE_{bulk} / kJ·mol^{-1}	ΔE_{coh} / kJ·mol^{-1}
OPT1[a]	−2039964.76	−8160430.53	−132.58	–[b]	–[b]	–[b]
OPT2	−2039964.30	−8160430.32	−132.97	+0.47	+0.21	−0.39
OPT3	−2039963.94	−8160427.43	−132.60	+0.82	+3.10	−0.02
F	−2039963.22	−8160423.30	−132.34	+1.54	+7.23	+0.24
E1	−2039962.89	−8160422.24	−132.39	+1.89	+8.29	+0.19
E2[c]	−2039963.05	−8160420.45	−131.73	+1.71	+10.08	+0.85
E4	−2039963.12	−8160422.54	−132.26	+1.64	+7.99	+0.32
E4	−2039963.14	−8160422.70	−132.27	+1.66	+7.83	+0.31
E5	−2039956.25	−8160406.48	−135.04	+8.51	+24.04	−2.46
T-II-1	−2039828.00	−8159942.24	−147.36	+136.77	+488.29	−14.78
T-II-2	−2039872.83	−8160180.24	−162.01	+91.93	+250.29	−29.43
I0	−2039710.93	−8159360.31	−119.28	+253.83	+1070.21	+13.30
I1	−2039958.74	−8160407.77	−132.92	+6.02	+22.76	−0.33
I2	−2039962.61	−8160421.28	−132.42	+2.15	+9.25	+0.16
T1	−2039963.43	−8160420.16	−131.34	+1.33	+10.36	+1.24
T2	−2039962.74	−8160418.90	−131.72	+2.02	+11.63	+0.87
T3	−2039958.09	−8160380.08	−126.79	+6.67	+50.45	+5.79

[a] Optimised unit cell parameters: a = 4.3288 Å, b = 14.5110 Å; c = 9.5238 Å; β = 91.616°; [b] Reference state; [c] Unit cell parameters determined for the 10 K crystal structure: a = 4.3135(7) Å, b = 14.432(3) Å, c = 9.5548(18) Å, β = 91.170(6)°.

XII. APPENDIX F

Supplementary Materials
9mA:1mT Co-crystal Part

Comment F1. X-ray data collection.

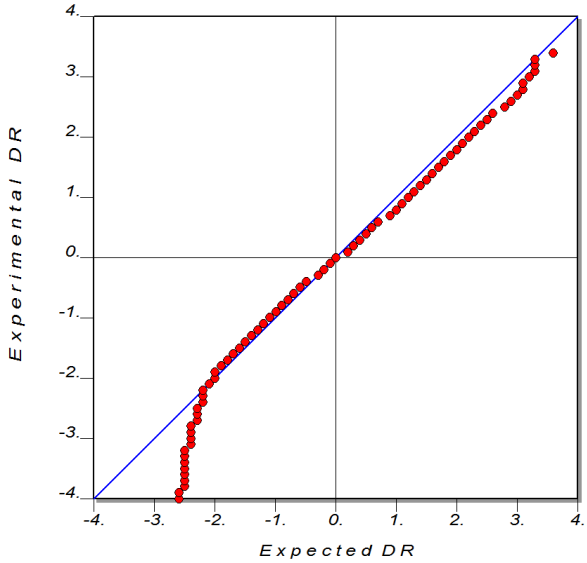
Single-crystal X-ray measurement of **9mA:1mT** at 90 K was carried out on a Bruker AXS Kappa APEX II Ultra diffractometer equipped with a TXS rotating anode (Mo-K_α radiation, λ = 0.71073 Å), multi-layer optics, and Oxford Cryosystems nitrogen gas-flow device (700 Series Cryostream). A single crystal of a suitable size was attached to the MiTeGen MicroMount tool using the *Paratone N* oil, and mounted on a goniometer head at 60 mm distance from an APEX II CCD detector. Data collection strategy was constructed so as to scan over the whole reciprocal space with a reasonable reflection redundancy *vs.* data collection time, and served as an input for the *APEX2* suit of programs. Solely ω-type scans were taken into account with a rotation step of 0.25°. The designed strategy resulted in 9600 frames and was as follows (Euler geometry parameters; χ = −55°, $\Delta\omega$ = 150°): (1) four scans with 2θ = −30° (φ = 0°, 90°, 180° and 270°) and 12 s of exposure time; (2) four scans with 2θ = −60° (φ = 0°, 90°, 180° and 270°) and 40 s of exposure time; (3) eight scans with 2θ = −80° (φ = 45°, 135°, 225°, 315°, 0°, 90°, 180° and 270°) and 120 s of exposure time.

The determination of unit cell parameters and the integration of raw diffraction images were performed with the *APEX2* program package. Dataset was corrected for Lorentz, polarization and oblique incidence effects. The multi-scan absorption correction, frame-to-frame scaling and merging of reflections were carried out with the *SORTAV* program. Final data collection and reduction parameters are presented in Table F2.

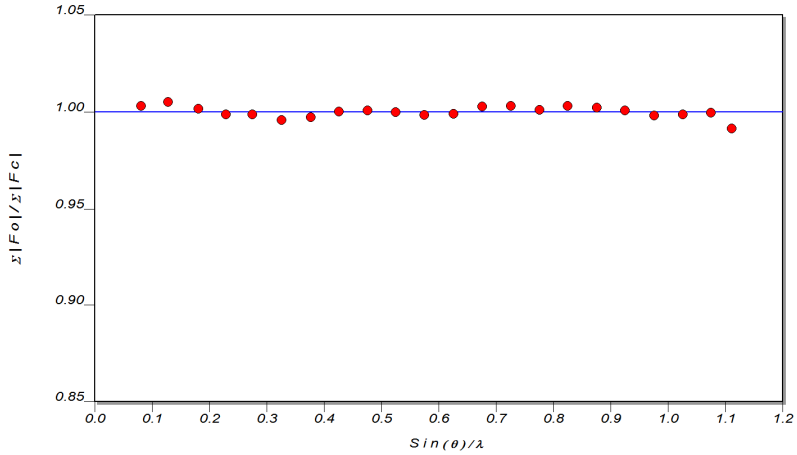
The measurement of **9mA** was performed on Agilent Technologies KM4CCD κ-axis diffractometer (recently upgraded with Opal area detector) equipped with a Mo-K_α sealed tube, graphite monochromator, and Oxford Cryosystems nitrogen gas-flow device (600 Series Cryostream). A single crystal of suitable size was mounted on a cryo-loop using *Paratone N* oil and positioned at 50 mm distance from the CCD camera. Data collection strategy, based on ω scans (rotation step of 0.5°), was optimized and monitored with the aid of appropriate algorithms implemented within the *CRYSALIS* suite of programs. The unit cell parameter determination and raw diffraction image integration were performed with the *CRYSALIS* software. Data set was corrected for Lorentz, polarization and oblique incidence effects. The multi-scan absorption correction, frame-to-frame scaling and merging of reflections were carried out with *SORTAV* program. Final strategy, which resulted in 1632 frames, was as follows (κ geometry parameters): (1) four scans with 2θ = 30° ($\Delta\omega$ = 102°, κ = −79°; φ = 0°, 90°, 180° and 270°) and 30 s of exposure time; (2) four scans with 2θ = 30° ($\Delta\omega$ = 102°, κ = 79°; φ = 0°, 90°, 180° and 270°) and 30 s of exposure time.

Table F2. Selected crystal setting information together with the data collection parameters; **9mA:1mT** and **9mA** measurements.

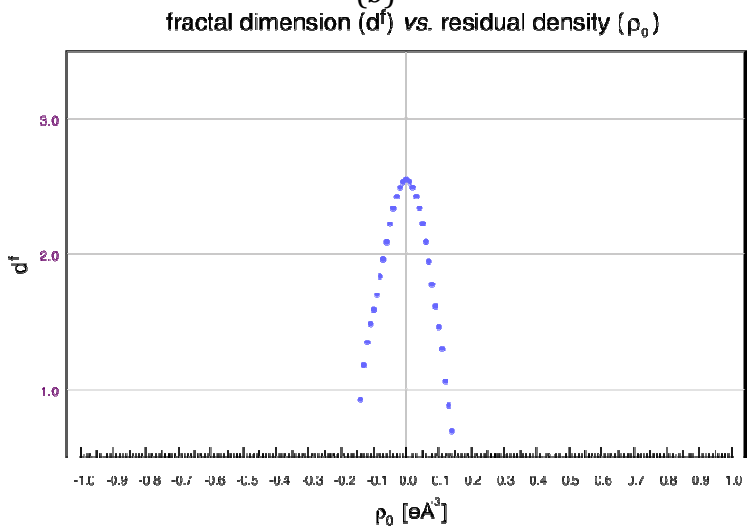
	9mA:1mT	**9mA**
Formula	$C_{12}H_{15}N_7O_2$	$C_6H_7N_5$
Molecular mass,	289.30	149.16
Temperature, T / K	90 K	100 K
Crystal system	monoclinic	monoclinic
Space group	$P2_1/m$	$P2_1/n$
Z	2	4
$F(000)$	304	312
a / Å	8.2688(3)	7.4877(5)
b/ Å	6.3458(2)	12.2678(6)
c/ Å	12.8427(4)	7.6797(5)
α / °	90	90
β / °	106.7855(7)	112.110(7)
γ / °	90	90
Volume, V / Å^3	645.17(4)	653.56(8)
d_{calc} / g·cm^{-3}	1.489	1.515
Completeness	97.6%	95.1%
$(\sin\theta / \lambda)_{max}$ / Å^{-1}	1.12	0.78
θ range / °	2.63 – 53.03	3.24 – 33.72
Absorption	0.108	0.105
Index ranges	$-18 \le h \le 18$	$-11 \le h \le 11$
No. of reflections	44916 / 7890	14104 / 2477
R_{int}	3.18%	5.03%



(*a*)



(*b*)



(*c*)

Figure F3. (*a*) Normal probability plots for the final model. (*b*) Scale plot for the final multipole model. (*c*) Fractal dimension plot.

Comment F4. Structural (**9mA**) and charge density (**9mA:1mT**) refinement details.

9mA: formula: $C_6H_7N_5$; molecular mass: 149.16 a.u.; appearance: colourless, unspecified shape; measurement temperature: T = 100(1) K; crystal system: orthorhombic; space group: $P2_1/n$ (no. 14); unit cell dimensions: a = 7.4877(5) Å, b = 12.2678(6) Å, c = 7.6797(5) Å, $\alpha = \gamma = 90°$, β = 112.110(7)°, V = 653.56(8) Å^3; Z = 4; density: d_{calc} = 1.515 g·cm^{-3}; $F(000)$ = 312; size: 0.30 × 0.21 × 0.18 mm^3; θ range for data collection: 3.24 – 33.72°; $(\sin\theta / \lambda)_{max}$ = 0.78 Å^{-1}; index ranges: $-11 \le h \le 11$, $-18 \le k \le 19$, $-11 \le l \le 11$; reflections collected / unique: 14104 / 2477 (R_{int} = 5.03%); completeness: > 95%; absorption coefficient: μ = 0.105 mm^{-1}; absorption correction method: multi-scan; refinement method: full-matrix least-squares on F; reflections with $|F^o|^2 \ge 3\sigma(|F^o|^2)$ / parameters / restraints: 100 / 0 / 0; final R-factors ($|F^o|^2 \ge 3\sigma(|F^o|^2)$): $R[F]$ = 4.13%, $wR[F^2]$ = 4.88%; R-factors (all data): $R[F]$ = 8.11%, $wR[F^2]$ = 10.98%; goodness-of-fit (all data): $GooF$ = 3.23; largest residual density peak / hole: +0.53 / –0.72 e·Å^{-3}. Comments: Hydrogen atoms are placed geometrically within riding model for their isotropic ADPs (d_{N-H} = 0.87 Å, $U_{iso}^{H} = 1.5 \cdot U_{eq}^{N}$; d_{C-H} = 0.96 Å, $U_{iso}^{H} = 1.2 \cdot U_{eq}^{C}$). Orientation of methyl group was determined from the Fourier map.

9mA:1mT$_{CD}$: formula: $C_{12}H_{15}N_7O_2$; molecular mass: 289.30 a.u.; appearance: colourless, unspecified shape; measurement temperature: T = 90(1) K; crystal system: monoclinic; space group: $P2_1/m$ (no. 11); unit cell dimensions: a = 8.2688(3) Å, b = 6.3458(2) Å, c = 12.8427(4) Å, $\alpha = \gamma = 90°$, β = 106.7855(7)°, V = 645.17(4) Å^3; Z = 2; density: d_{calc} = 1.489 g·cm^{-3}; $F(000)$ = 304; size: 0.27 × 0.12 × 0.07 mm^3; θ range for data collection: 2.63 – 53.03°; $(\sin\theta / \lambda)_{max}$ = 1.12 Å^{-1}; index ranges: $-18 \le h \le 18$, $-7 \le k \le 13$, $-28 \le l \le 28$; reflections collected / unique: 44916 / 7890 (R_{int} = 3.18%); completeness: > 97%; absorption coefficient: μ = 0.108 mm^{-1}; absorption correction method: multi-scan; refinement method: full-matrix least-squares on F; reflections with $|F^o|^2 \ge 3\sigma(|F^o|^2)$ / parameters / restraints: 6351 / 505 / 12; final R-factors ($|F^o|^2 \ge 3\sigma(|F^o|^2)$): $R[F]$ = 1.51%, $wR[F^2]$ = 2.07%; goodness-of-fit: $GooF$ = 1.07; largest residual density peak / hole: +0.15 / –0.15 e·Å^{-3}.

Table F5. *R*-factors and related parameters after TAAM refinement (all calculated using reflections with $|F^o|^2 \geq 3\sigma(|F^o|^2)$).

Descriptor	**9mA**
$R[F]$	2.68%
$wR[F]$	2.75%
$R[F^2]$	3.35%
$wR[F^2]$	5.85%
$\Delta\rho_{max}$ / e·Å^{-3}	+0.17
$\Delta\rho_{min}$ / e·Å^{-3}	−0.30
GooF	1.26

Table F6. Bond length list for the structures of **1mT**$_{TAAM}$, **9mA**$_{TAAM}$ and co-crystal **9mA:1mT**$_{CD}$. d_{NAB} denotes nucleic acid mono-component structure, whereas d_{dim} stands for co-crystal structure.

Fragment	*Bond type*	d_{NAB} / Å	d_{dim}/ Å
1mT	N1–C2	1.3874(5)	1.3780(3)
	N1–C6	1.3746(5)	1.3745(3)
	N1–C7	1.4706(6)	1.4671(3)
	N3–C2	1.3803(5)	1.3738(3)
	N3–C4	1.3805(5)	1.3769(3)
	N3–H3	1.029(1)	1.030(1)
	C2–O2	1.2256(5)	1.2286(4)
	C4–O4	1.2370(5)	1.2343(3)
	C4–C5	1.4381(5)	1.4383(3)
	C5–C6	1.3600(6)	1.3580(3)
	C5–C8	1.4969(6)	1.4921(3)
	C6–H6	1.083(1)	1.083(1)
	C7–H7A	1.077(1)	1.077(1)
	C7–H7B	1.077(1)	1.077(1)
	C7–H7C	1.076(1)	1.077(1)
	C8–H8A	1.077(1)	1.077(1)
	C8–H8B	1.077(1)	1.077(1)
	C8–H8C	1.077(1)	1.077(1)
9mA	N9–C10	1.340(1)	1.3425(4)
	N9–C14	1.356(1)	1.3535(3)
	N11–C10	1.326(1)	1.3325(4)
	N11–C12	1.341(1)	1.3420(3)
	N15–C13	1.383(1)	1.3848(3)
	N15–C16	1.317(1)	1.3191(3)
	N17–C12	1.365(1)	1.3733(3)
	N17–C16	1.362(1)	1.3700(3)
	N17–C19	1.457(1)	1.4507(3)
	N18–C14	1.334(1)	1.3333(3)
	N18–H18A	1.010(1)	1.010(1)
	N18–H18B	1.010(1)	1.010(1)
	C12–C13	1.414(1)	1.3909(3)
	C13–C14	1.414(1)	1.4132(3)
	C10–H10	1.083(1)	1.083(1)
	C16–H16	1.083(1)	1.083(1)
	C19–H19A	1.077(1)	1.077(1)
	C19–H19B	1.077(1)	1.077(1)
	C19–H19C	1.077(1)	1.077(1)

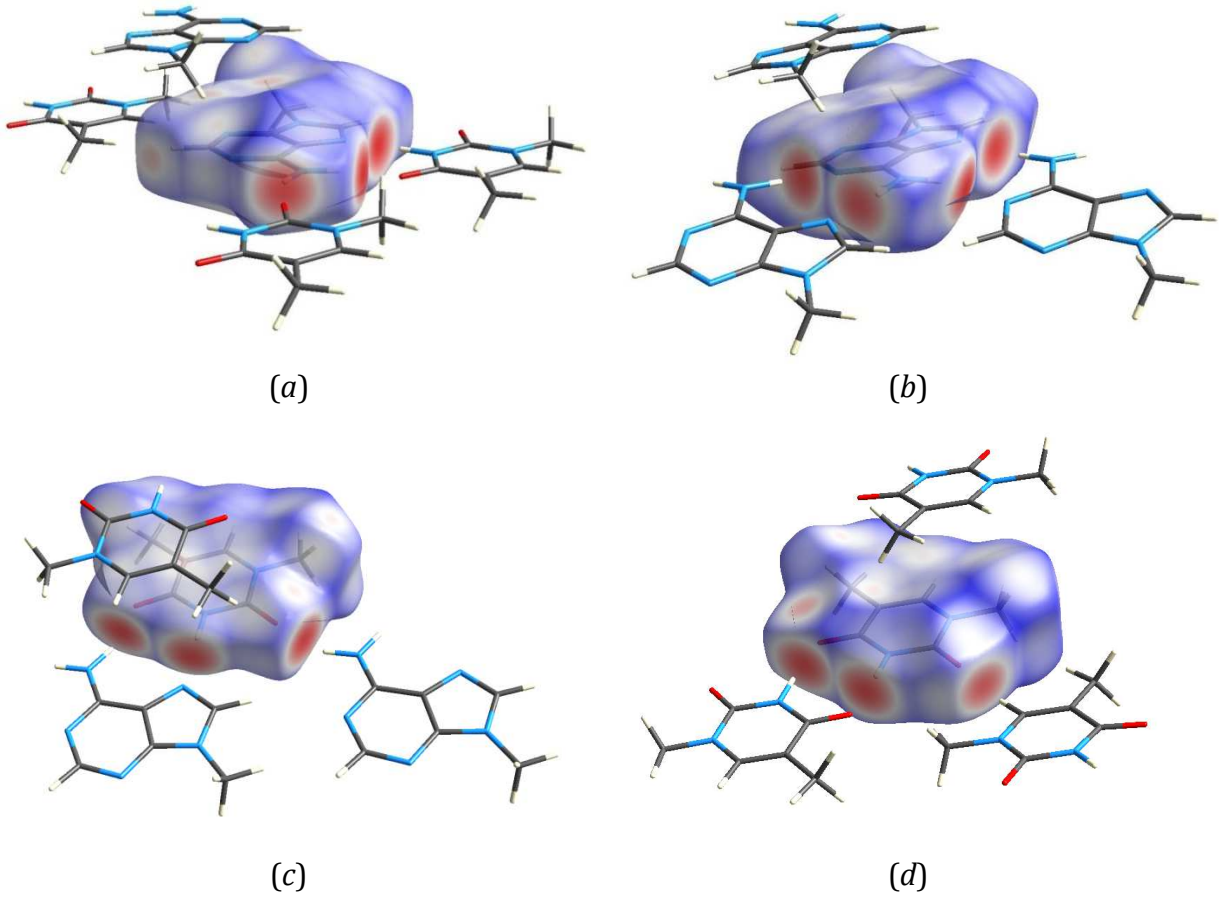


Figure F7. The comparison of the intermolecular contacts formed by 9-methyladenine (*a*) in the co-crystal structure and (*b*) in the mono-component structure, and the comparison of intermolecular contacts formed by 1-methylthymine (*c*) in the co-crystal structure and (*d*) in the mono-component structure.

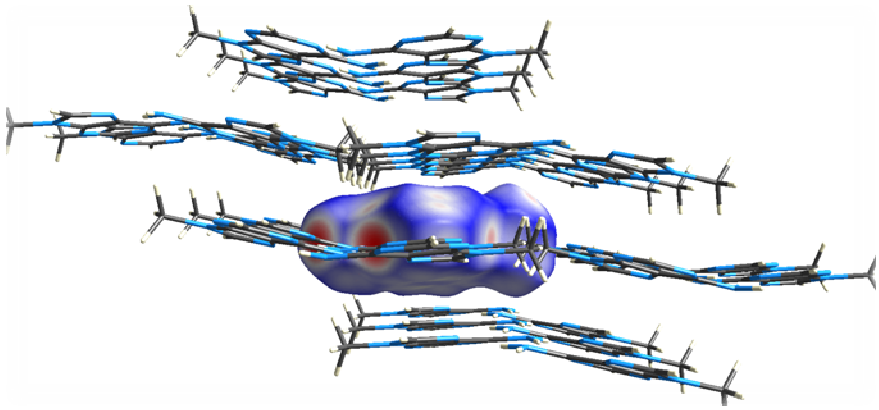


Figure F8. Hirshfeld surface of 9-methyladenine in the **9mA** crystal. Projection illustrates **9mA** layered architecture.

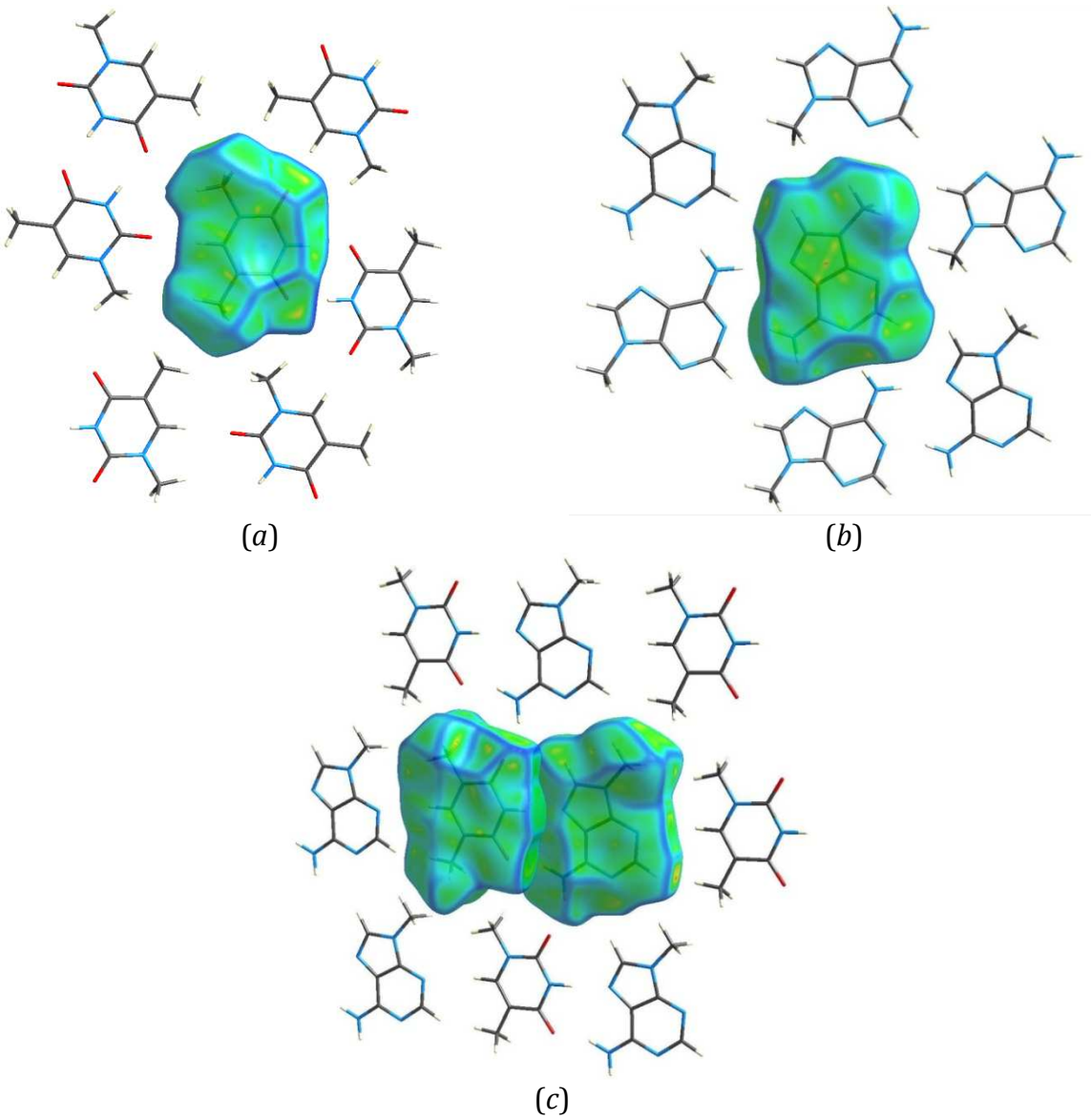


Figure F9. Curvedness mapped on Hirshfeld surfaces of (*a*) 1-methylthymine in the **1mT** crystal structure (*b*) 9-methyladenine in **9mA**, and (*c*) a dimer in the co-crystal structure.

Table F10. Selected QTAIM parameters of cobalent bonds at BCPs (R – distance between bonded atoms; R_1 and R_2 – distances from 1st and 2nd atom to the bond critical point, respectively; ϱ – electron density; ε – bond ellipticity) calculated for the $M1_{UBDB}$ charge density model.

Bond	R / Å	R_1 / Å	R_2 / Å	$\varrho(\mathbf{r}_{BCP})$ / e·Å^{-3}	$\nabla^2\varrho(\mathbf{r}_{BCP})$ / e·Å^{-5}	ε
O2–C2	1.229	0.799	0.430	2.80	-25.5	0.03
O4–C4	1.234	0.795	0.439	2.75	-27.5	0.03
N1–C2	1.378	0.818	0.561	2.20	-23.5	0.16
N1–C6	1.375	0.810	0.565	2.09	-18.6	0.09
N1–C7	1.467	0.845	0.622	1.70	-10.6	0.06
N3–C2	1.374	0.802	0.572	2.18	-22.2	0.16

N3–C4	1.377	0.808	0.569	2.11	-20.2	0.13
N3–H3	1.030	0.738	0.292	2.18	-29.2	0.06
C4–C5	1.438	0.722	0.717	2.09	-18.5	0.15
C5–C6	1.358	0.658	0.701	2.27	-21.2	0.17
C5–C8	1.492	0.754	0.738	1.69	-12.2	0.01
C6–H6	1.083	0.701	0.382	1.98	-22.5	0.03
C7–H7A	1.077	0.685	0.392	1.95	-22.2	0.03
C7–H7B	1.076	0.685	0.391	1.96	-22.4	0.04
C8–H8A	1.077	0.673	0.404	1.90	-20.9	0.01
C8–H8B	1.077	0.673	0.404	1.90	-20.9	0.00
N9–C10	1.343	0.759	0.584	2.28	-20.4	0.12
N9–C14	1.354	0.763	0.590	2.29	-20.8	0.09
N11–C10	1.333	0.762	0.572	2.33	-21.8	0.11
N11–C12	1.342	0.765	0.582	2.28	-20.2	0.07
N15–C13	1.385	0.779	0.606	2.00	-13.1	0.05
N15–C16	1.319	0.746	0.573	2.45	-23.6	0.17
N17–C12	1.373	0.806	0.574	2.06	-15.4	0.06
N17–C16	1.370	0.805	0.566	2.13	-19.3	0.17
N17–C19	1.451	0.844	0.607	1.79	-12.5	0.05
N18–C14	1.333	0.753	0.580	2.21	-19.3	0.11
N18–H18A	1.010	0.716	0.294	2.20	-26.2	0.00
N18–H18B	1.010	0.716	0.294	2.20	-26.2	0.00
C10–H10	1.083	0.694	0.390	1.98	-23.3	0.05
C12–C13	1.391	0.714	0.681	2.22	-20.9	0.20
C13–C14	1.413	0.694	0.719	2.07	-17.4	0.15
C16–H16	1.083	0.698	0.385	1.95	-21.8	0.04
C19–H19A	1.077	0.685	0.392	1.95	-22.3	0.04
C19–H19B	1.077	0.686	0.391	1.95	-22.3	0.04

Table F11. Selected QTAIM parameters for weak intermolecular interactions at the positions of respective BCPs (G – kinetic energy density; V – potential energy density) calculated for the $M1_{UBDB}$ charge density model. Differences in bond paths are shown in bold. Symmetry transformations (of the 2nd atom) are given as ORTEP codes.

Interaction	*Symmetry code*	R / Å	R_1 / Å	R_2 / Å	$\varrho(\mathbf{r}_{BCP})$ / e·Å^{-3}	$\nabla^2\varrho(\mathbf{r}_{BCP})$/ e·Å^{-5}	$G(\mathbf{r}_{BCP})$ / kJ·mol^{-1}·a$_0^{-3}$	$V(\mathbf{r}_{BCP})$ / kJ·mol^{-1}·a$_0^{-3}$
C2…O4	77651	3.277	1.670	1.608	0.03	0.41	8.6	-6.0
O2…H7A	66651	2.626	1.525	1.107	0.05	0.51	11.1	-8.2
O2…H8A	77651	2.756	1.583	1.173	0.04	0.40	8.6	-6.2
O2…H18B	45501	1.938	1.226	0.712	0.16	1.95	50.7	-48.3
O4…C2	77651	3.277	1.608	1.670	0.03	0.41	8.6	-6.0
O4…C7	65501	3.404	1.614	1.790	0.02	0.45	8.7	-5.2
O4…H18A	56552	1.852	1.192	0.661	0.20	2.27	63.1	-64.5
N3…C4	77651	3.243	1.634	1.622	0.05	0.53	11.7	-9.1
H3…N15	56552	1.874	0.642	1.233	0.21	2.10	62.3	-67.3
C4…N3	77651	3.243	1.622	1.634	0.05	0.53	11.7	-9.1
H6…N11	55601	2.713	1.128	1.586	0.04	0.43	9.4	-7.2
H7A…O2	67651	2.626	1.107	1.525	0.05	0.51	11.1	-8.2
C7…O4	45501	3.404	1.790	1.614	0.02	0.45	8.7	-5.2
C7…N18	77651	3.558	1.827	1.741	0.02	0.34	6.6	-4.1
H7A…H7A	65602	2.380	1.190	1.190	0.03	0.39	8.0	-5.5
H8A…O2	77651	2.756	1.173	1.583	0.04	0.40	8.6	-6.2
C8…N15	76651	3.540	1.797	1.751	0.02	0.35	6.9	-4.3

H8B…H10	55601	1.889	0.944	0.946	0.10	0.62	17.6	-18.3
N9…H16	65501	2.916	1.692	1.229	0.03	0.31	6.4	-4.4
N9…H19A	77551	2.699	1.588	1.113	0.04	0.54	11.4	-8.1
N11…H6	55401	2.713	1.586	1.128	0.04	0.43	9.4	-7.2
C10…C12	77551	3.189	1.583	1.610	0.06	0.47	11.2	-9.6
N15…C8	76651	3.540	1.751	1.797	0.02	0.35	6.9	-4.3
H18B…O2	65501	1.938	0.712	1.226	0.16	1.95	50.7	-48.3
N18…C7	77651	3.558	1.741	1.827	0.02	0.34	6.6	-4.1
H10…H8B	55401	1.889	0.946	0.944	0.10	0.62	17.6	-18.3
H10…H19B	65501	2.205	1.104	1.102	0.05	0.52	11.3	-8.4
C12…N11	77551	3.384	1.610	1.953	0.06	0.47	11.2	-9.6
H16…N9	45501	2.916	1.229	1.692	0.03	0.31	6.4	-4.4
H19A…N9	77551	2.699	1.113	1.588	0.04	0.54	11.4	-8.1
H19…BH10	45501	2.205	1.102	1.104	0.05	0.52	11.3	-8.4
C19…H7B	55401	3.204	1.876	1.336	0.02	0.25	4.9	-3.1

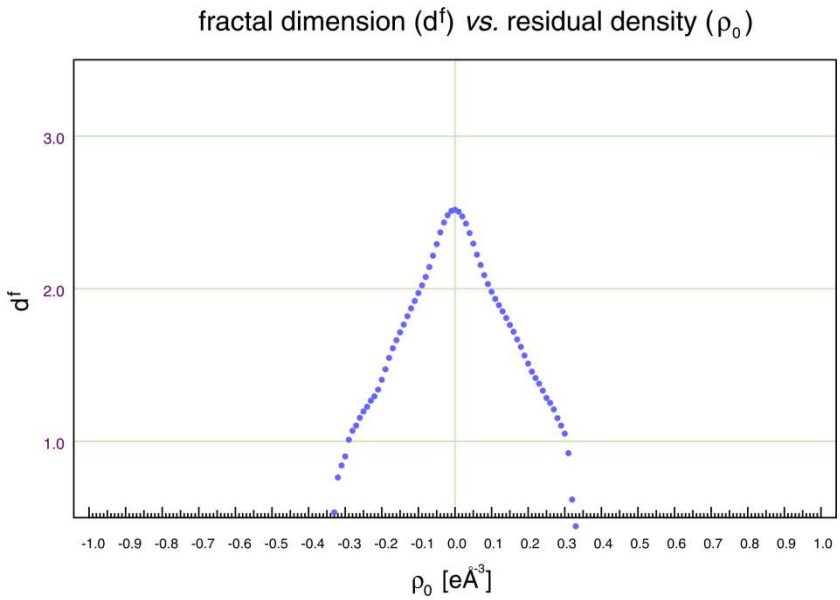
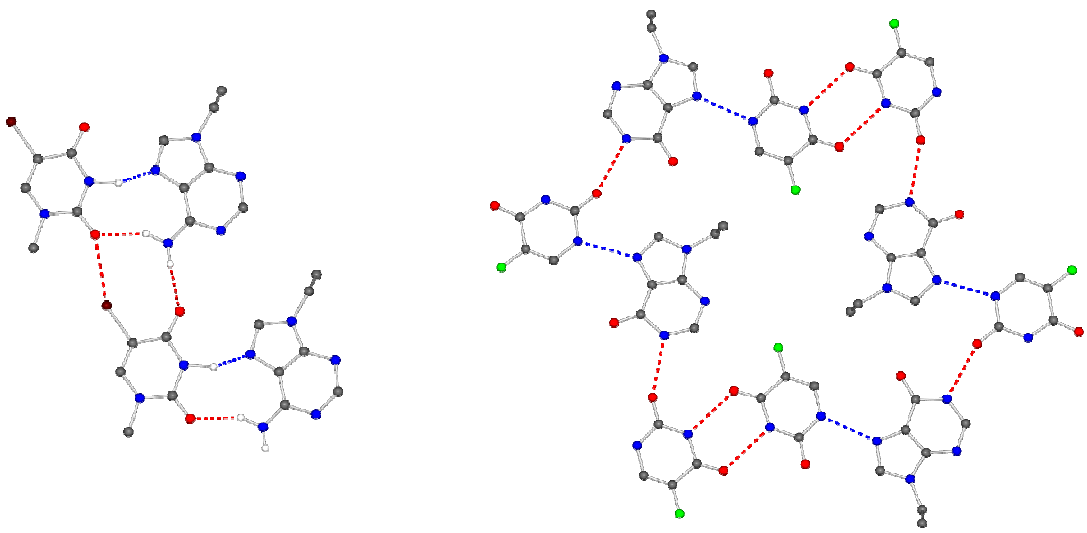


Figure F12. Fractal dimension plot for **9mA:1mT** ($M1_{UBDB}$).



(*a*) (*b*)

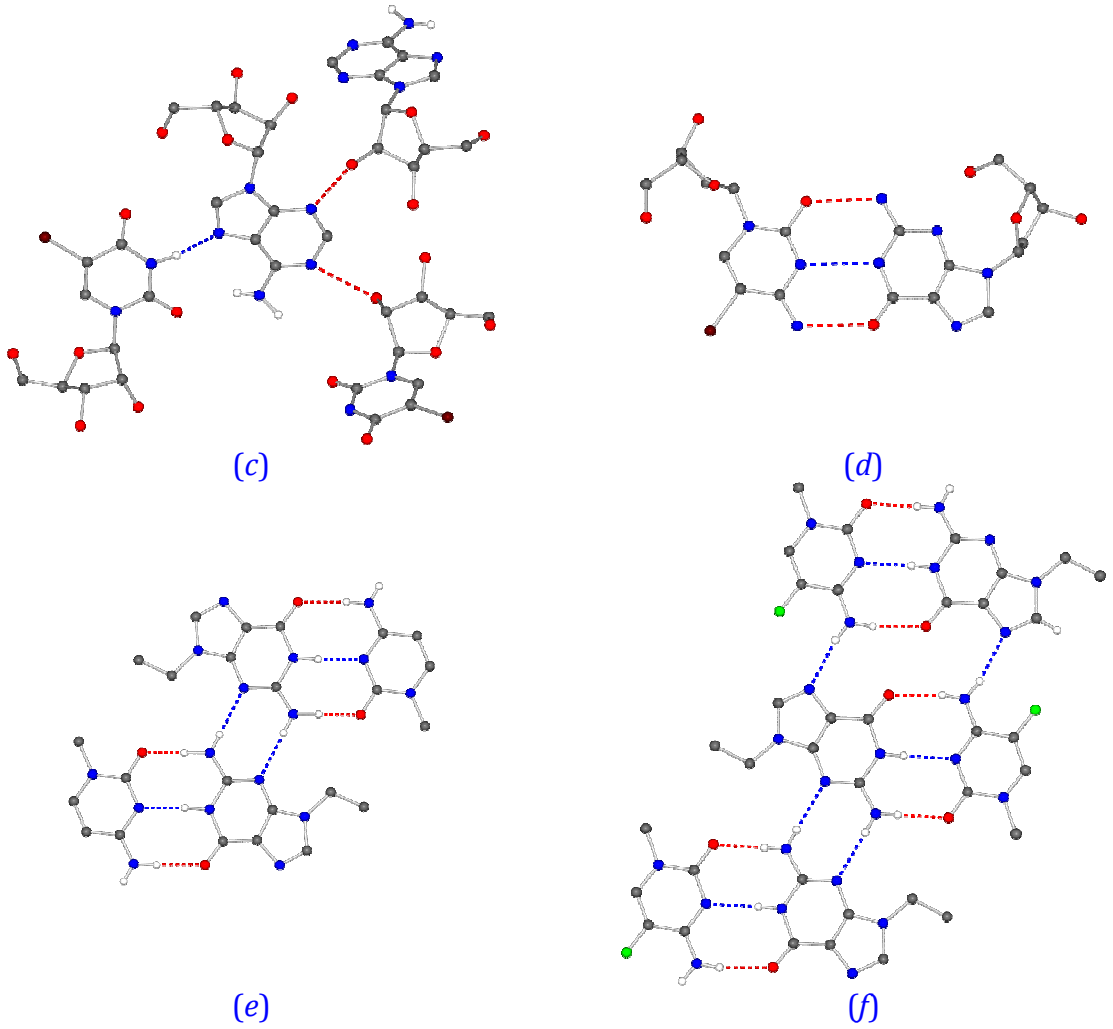


Figure F13. Literature co-crystal structures formed by the complementary nucleic base derivatives: (*a*) **5BrU:9eA**,[1] (*b*) **5FU:9eHyp**,[2] (*c*) **5BrUridine:Adenosine**,[3] (*d*) **5BrCytosine:Deoxyguanosine**,[4] (*e*) **1mC:9eG**,[5] (*f*) **1m5FC:9eG**.[5]

Additional references

1 Katz, L.; Tomita, K.-I.; Rich, A. The crystal Structure of the intermoleeular complex 9-ethyladenine: 1-methyl-5-bromouracil. *Acta Cryst.* **1966**, *21*, 754-764.

2 Kim, S.-H.; Rich, A. Crystal structure of the 1:1 complex of 5-fluorouracil and 9-ethylhypoxanthine. *Science* **1967**, *158*, 1046-1048.

3 Haschemeyer, A. E. V.; Sobell, H. M. The crystal structure of a hydrogen bonded complex of adenosine and 5-bromouridine. *Acta Cryst.* **1965**, *18*, 525-532.

4 Haschemeyer, A. E. V.; Sobell, H. M. The crystal structure of a hydrogen bonded complex of deoxyguanosine and 5-bromodeoxycytidine. *Acta Cryst.* **1965**, *19*, 125-130.

5 O'brien, E. J. Crystal structures of two complexes containing guanine and cytosine derivatives. *Acta Cryst.* **1967**, *23*, 92-106.

Table F14. Details of the selected adenine-uracil dimers taken out from RNA molecules.

d_{max} / Å	*PDB code*	*NDB code*	*Our code*	*Pair number*	*Pair name*	*t* / Å			*Θ* / °		
						shear	*stretch*	*stagger*	*buckle*	*propeller*	*opening*
2.90	1KH6	UR0018	**cAU1**	14	A_5BU38:A7_A	0.177	2.578	0.344	7.680	-10.249	-67.396
2.25	1HR2	UR0012	**cAU2**	25	A_U135:A187_A	-0.526	3.818	0.118	-2.616	-12.924	-64.839
		UR0012	**cAU3**	110	B_U135:A187_B	-0.865	3.745	0.094	2.407	-15.032	-72.977
2.57	1MMS	RR0009	**cAU4**	16	C_U1060:A1088_C	-0.630	3.398	0.653	-2.629	5.058	-67.855
		RR0009	**cAU5**	36	D_U1060:A1088_D	-0.666	3.422	0.421	-1.228	5.799	-68.728
2.40	1JJ2	RR0033	**cAU6**	53	0_U63:A70_0	-0.706	3.818	-0.315	3.313	-18.895	-66.471
		RR0033	**cAU7**	63	0_A45:U115_0	0.372	-3.510	0.401	-8.929	11.586	62.167
		RR0033	**cAU8**	383	0_U1831:A857_0	-0.510	3.506	-1.334	16.321	-4.439	-74.785
		RR0033	**cAU9**	489	0_U1109:A1247_0	0.025	3.526	1.318	-21.413	26.544	-57.862
		RR0033	**cAU10**	692	0_A1684:U1722_0	0.470	-3.555	-1.008	19.261	4.811	72.015
		RR0033	**cAU11**	747	0_U1890:A2011_0	-0.115	3.484	1.134	-28.302	-4.661	-60.370
		RR0033	**cAU12**	1034	0_U2791:A2793_0	-0.559	3.487	-0.009	7.710	-0.618	-71.828
3.00	1N32	RR0056	**cAU13**	8	A_A915:U13_A	0.402	-3.262	0.592	-6.709	11.928	71.032
		RR0056	**cAU14**	188	A_U405:A499_A	-0.710	3.608	0.517	2.119	-34.073	-72.650
		RR0056	**cAU15**	196	A_A431:U429_A	-0.102	-2.869	-1.480	21.023	-4.942	71.935
		RR0056	**cAU16**	254	A_U820:A873_A	-0.610	3.535	0.730	-26.246	-11.296	-60.805
3.00	1FJG	RR0012	**cAU21**	189	A_A499:U405_A	1.061	-3.927	-0.687	-0.469	24.892	68.591
noLW		RR0012	**cAU22**	208	A_A431:U429_A	0.185	-2.973	-0.946	14.696	-6.520	69.316
2.40	1JJ2	RR0033	**tAU1**	975	0_A2702:U2693_0	-4.455	-2.701	-0.790	-12.022	-11.910	-92.232
1.11	483D	UR0007	**tAU2**	7	A_U2656:A2665_A	4.142	-1.900	-0.857	6.748	-20.941	-102.271
1.93	1EH2	tr0001	**tAU3**	16	A_U8:A14_A	4.137	-1.910	0.766	-4.615	12.117	-103.091
3.00	1KXK	UR0019	**tAU4**	1	A_U4:A66_A	4.230	-2.149	-0.280	5.879	-1.911	-101.729
		UR0019	**tAU5**	11	A_U16:A55_A	4.259	-2.248	0.574	-10.343	1.395	-81.193

Table F15. Details of the selected stacking type dimers of nucleic acid bases.

Dimer code	*PDB code*	*NDB code*	d_{max} / Å	*Pair name*	E_{int} / kJ·mol^{-1}
AA1	1JJ2	RR0033	2.40	A_2702-2703_A	−22.4
AA2	458D	BD0011	2.30	DA16_DA17	−26.2
AA3	431D	BD0006	1.15	DA16_DA17	−25.3
AT4	458D	BD0011	2.30	DA18_DT19	−25.3
AT5	1M77	AD0026	1.25	DA5_DT6	−22.6
GU6	3I2U	NA0054	2.80	G21_U61	-34.8
AU8	3NKB	NA0604	1.92	A57_U56	−24.4