M.Sc. Agnieszka Poulain
Department of Chemistry, Faculty of Crystallography, Adam Mickiewicz University in Poznań, Poland
Lorraine University, Cristallographie, Résonance Magnétique et Modelisations, CRM2, UMR UHP-CNRS 7036, Institut Jean Barriol, BP 70239, 54506 Vandoeuvre-les-Nancy, France

Summary of the thesis dissertation:

Experimental charge density distribution in the crystals of 4-nitroimidazole derivatives

Theoretical introduction

The standard resolution crystal structures of the several nitroimidazole derivatives have been widely investigated in our laboratory in order to classify and characterize their intermolecular interactions, such as strong and weak hydrogen bonds, halogen bonds, dipolar interactions, and layer association. However this lower resolution data modelled in the Independent Atom Model (IAM) completely neglect the asphericity of an atom. The charge density is localized around the nuclei, as the core scattering is assumed to dominate the total scattering of an atom, and the crystal structure is seen as superposition of isolated spherical atomic densities. However this simple model is not suitable for light atoms and excludes the possibility of charge transfer.

To overcome these shortcomings a more complex model is usually used for the high resolution data, namely the Hansen-Coppens multipolar model [1], based on the following formula:

$$\rho_{tot}(r) = \rho_{core}(r) + P_{val}\kappa^3\rho_{val}(\kappa, r) + \sum_{l=0}^{l_{max}} \kappa'^3 R_l(\kappa', r) \sum_{m=0}^{+l} P_{lm} d_{lm\pm}(\theta, \varphi)$$

where P_{val} and $P_{lm's}$ are the valence and multipole populations and κ and κ' are the contraction/expansion coefficients for spherical and aspherical valence density, respectively. The two first terms are the spherically averaged core and valence electron densities of an atom calculated from Hartree-Fock (HF) wave functions and the last term corresponds to expansion/contraction of the non-spherical valence density, with a Slater-type radial function:

$$R_l(r) = \frac{\xi_l^{n_l+3}}{(n_l+2)!} r^{n_l} e^{-\xi_l r}, n_l \geq 1$$

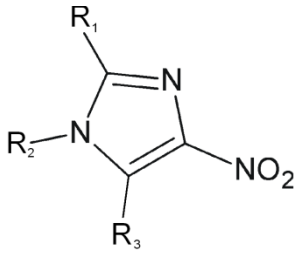
Each number *l* corresponds to different multipolar level: there are three dipoles, five quadrupoles, seven octapoles and nine hexadecapoles in the current version of MoPro program, which was used for multipolar calculations [3].

According to Bader's theory of Atoms-In-Molecules [2] analysis of the charge density distribution at each point of the unit cell allows for examination of the chemical structure of the molecule and intermolecular interactions, especially the weak ones. The topology of electron density distribution, which is a scalar function, can be analyzed by looking at its gradient vector field, the critical points (extremes of the electron density where the gradient of electron density vanishes), Laplacian of electron density (sum of the three principal axes of curvature of diagonal elements in Hessian matrix composed of partial second derivatives of the electron density). Finally the AIM enables to obtain the properties by integration of the corresponding property density (such as charge, volume, multipolar moments or energy) over the atomic basin (with a *zero flux surface*), or calculation of the interaction energy densities and electrostatic potential maps.

Investigated molecules

The aim of the thesis dissertation was first to collect the high resolution data for several 4-nitroimidzole derivatives, then to re-determine their molecular structures using IAM and next Hansen-Coppens model and finally the AIM topological analysis of the intramolecular changes due to different substituents, as well as the analysis of the intermolecular interactions.

A good quality crystals were obtained for five structures (see scheme below): 1-(2'-aminophenyl)-2-methyl-4-nitroimidazole (**I**), 2-methyl-4-nitro-1-phenyl-1H-imidazole-5-carbonitrile (**II**), solid solution of 1-(4-chlorophenyl)-2-methyl-4-nitro-1H-imidazole-5-carbonitrile and 5-bromo-1-(4'-chlorophenyl)-2-methyl-4-nitro-1H-imidazole (**IIIa**), 1-(4'-chlorophenyl)-4-nitro-5-methyl-imidazole (**IV**) and 2-chloro-1-methyl-4-nitro-1*H*-imidazole (**V**).



I: $R_1 = CH_3$; R_2 = *o*-NH_2-Ph; R_3 = H; **II**: $R_1 = CH_3$; R_2 = Ph; R_3 = CN; **IIIa**: $R_1 = CH_3$; R_2 = *p*-Cl-Ph; R_3 = CN or Br; **IV**: R_1 = H; R_2 = *p*-Cl-Ph; $R_3 = CH_3$; **V**: R_1 = Cl; $R_2 = CH_3$; R_3 = H.

It appeared that the high resolution data refinement is not a trivial task and beside the good data quality it requires a careful analysis at each step of the procedure and appropriate response to the emerging problems.

This work, that basically was intended to examine and prioritize the weak interactions, such as H-bonds, dipolar interactions, halogen bonds and van der Waals contacts, as well as the influence of substituents on the imidazole ring, appeared to bring many interesting points of considerations and new perspectives for the future work.

Free R factor calculations [5-6]

In the first place the R_{free} factor calculations were performed for molecule **I** to estimate if dissimilarities of the charge density between the two symmetry independent molecules are reliable or results from noise and uncertainties, and for molecules **I** and **II** to find the optimal restraints level. It was one of the first attempts to use R_{free} calculations in the experimental charge density modeling of small organic molecules.

The main idea of the R_{free} calculations is the following: 5% (1/20) of the reflections are used as a test set and the remaining 95% in the least squares refinement. The free *R* factors are averaged over 20 individual free *R* factors obtained from 20 different refinements. The two series of refinements are performed, with a different level of restraints/constraints imposed. In the first one all the P_{val}, P_{lm} and κ are constrained to be identical for the equivalent atoms or corresponding atoms in the two symmetry independent molecules. Varying restraints weights $W=1/\sigma_r^2$ are applied to the symmetry of atoms. The wR^2F factor decreases when weaker restraints are applied and concerning wR^2F_{free}, there is a minimum for moderately restrained refinement at the U-shape curve. The totally unconstrained refinement has R_{free} factor close to the minimum, while the constrained refinement yields higher *free R* factor value.

In the second series of calculations, the symmetry restraints are fixed at the optimal value and additional refinements are performed with varying levels of restraints imposed on P_{val}, P_{lm} and κ similarity. Trends similar to the previous refinement are observed, but a less pronounced minimum for the free *R* factor, still lower than the minimum of the first series of *R*-free tests. The combination of the two types of restraints on the charge density yields a better refinement while the one with chemically equivalent atoms constrained to have the same charge density parameters is not relevant for the current study as its wR^2F_{free} value is higher.

For the remaining molecules the weak restraints on the chemical equivalency and local symmetry were applied, with the σ value similar that was found optimal for molecules **I** and **II**, indeed for the similar molecules. The R_{free} factor calculations were not repeated neither for structure **IIIa** nor for **IV**-**V**, as it is a time consuming procedure, and according to our findings the optimal model is always the one with soft restraints imposed, very closed to unconstrained model.

Dipole moment [7]

A deep investigation was carried on for the theoretical and experimental data in order to find the best conditions to calculate the dipole moment with the magnitude and direction similar to the one obtained from theory. The selected molecule **II** was especially suitable for such an analysis, as the positive and negative fragments are well separated at both sides of the molecule, and the H-atoms are located at the periphery. These two factors may significantly influence the μ value.

It was proved, that for all tested models (multipolar Hansen-Coppens, virtual atom and kappa) there is a set of restraints and constraints for the theoretical and experimental data (not always the same for corresponding models MM_{theo} and MM_{exp}, *etc*.) that should be applied to obtain the dipole moment vector closest to the theoretical value.

For the multipolar models the best results from theoretical data were obtained for the multipolar expansion fixed at a dipolar level for H-atoms, an octupolar level for non-H atoms and hexadecapoles used only for atoms involved in electron rich bonds like C≡N, while the κ_{hyd} either constrained to 1.16 or refined did not influence the μ value. For the experimental data this best theoretical model was supplemented by the ADPs (Anisotropic Displacement Parematers) for the H-atoms estimated by SHADE and their κ and κ' parameters restrained respectively to 1.16 and 1.25 (σ_r = 0.01), which were found to be the best values after the theoretical structure factors refinement.

In the case of the virtual atom models the restriction of κ_{hyd} was crucial to limit the influence of peripheral hydrogen atoms on the dipole moment magnitude and direction. A particularly good agreement was obtained for κ_{hyd} =1.13 for theoretical data, but for the experimental one the κ_{hyd} had to be lowered to 1.10 and the direction of μ vector in this second case was much worse than expected. Moreover one should keep in mind that the doubling of virtual atoms on the nitro N=O bonds was essential to achieve featureless residual maps only for the theoretical data models and it simultaneously caused a significant decrease of the molecular dipole moment. Also the ADPs of the H-atoms from SHADE may be not appropriate for the models other than the multipolar one, as they allow predicting the direction of μ but not its magnitude.

The simplest kappa model applied for the theoretical data brought not the best but still acceptable results, while for the experimental data the correct direction could not be achieved, no matter the restraints used.

Joined experimental (from X-ray diffraction) and theoretical (from DFT calculations) data refinement were performed only for one molecule (**II**) of the series, in the course of best model finding for the reliable dipole moment calculations. A new parameter (κ-core) had to be introduced only for the theoretical data modeling of non-H atoms in order to take into account a significant depletion of the charge density at the positions of the atoms, due to the different wave functions used in CRYSTAL09 (atomic Gaussian-type orbitals, DFT) and for the multipolar refinement (Clementi & Roetti – Slater type expansion, Hartree-Fock). It would be then interesting to examine the remaining imidazole derivatives to verify this assumption, and to review if the restraints obtained for the best μ value from theoretical and experimental data for **II** (that show the sharpest separation of the negative and positive charge) are the same for the rest of the molecules in this series.

Solid solution / Disorder

Structure **IIIa** appeared to be the only one in the series that revealed a different unit cell packing after high resolution structure re-determination, with an unexpected residual density

peak (3.5e) located at ca. 1.94 Å from $C_{imidazole}$ atom, when the structure was refined with the full occupancy of C≡N group. It appeared that the solid solution (see scheme) was accidentally obtained after the re-crystallization process, as the bromine substituted molecule was the previous step of in the chemical reaction. This special kind of disorder can be modelled with a special caution, by means of the transfer of the multipolar parameters from analogous molecules. It should be underlined, that the quality of measured data was excellent (R_{int} = 2.6%), refinement in the IAM model was stable even if the thermal motion parameters were refined together with the coordinates and site occupancy factor. Only multipolar refinement had to be constrained or strongly restrained in the problematic fragment of the structure.

Even if the problem of disorder was solved using the transferability of the multipolar parameters from known structures for this particular molecule, the further investigation is still necessary to answer the important questions:

a) is the bromine contamination a reason of different crystal packing in **III** and **IIIa**?

b) what is the structure of pure Br form of this imidazole derivative?

c) what is the phase diagram illustrating the unit cell parameters dependence on the Br/CN form concentration?

These questions were left without an answer due to time shortage, as there is a need of synthesis of the pure components of **IIIa** before any trials of crystallization.

Anharmonicity [6]

There was a need to introduce the ANMs (Anharmonic Nuclear Motions) parameters for correct modelling of some atoms in three out of five analysed molecules: for the two nitrogen atoms of the amino groups and one nitro group in **I** and for the two chlorine atoms in structures **IIIa** and **V**. In each case the high residual density peaks arranged in the ‘shashlik-like’ pattern or the significant, unreasonable distortion of the deformation density were the symptoms of such a phenomenon. Any proper model with featureless residual maps and symmetric deformation density were unreachable without these additional parameters included in the process of refinement, even if only few ANMs components were statistically significant.

Low multi-temperature data collections

Several low temperature measurements were performed to verify the necessity of the anharmonic modeling of some atoms in molecule **I**: single crystal high resolution measurements at 10 K and 100 K at Agilent Technologies devices; powder diffraction measurement between 23 K and 298 K at Panalytical X’Pert Pro diffractometer and finally for

the first time the high resolution measurements at 35 K and 70 K at four circle diffractometer with Nonius Kappa CCD detector with helium-bath orange cryostat device [4].

First of all it appeared, that there is a necessity of the ANMs for the correct modeling of five atoms (one nitro group and the two nitrogen atoms of amino groups) at 75 K and 100 K data, while for the lower temperature data (10 K and 35 K) the deformation and residual electron density maps did not show the typical distortion and 'shashlik-like' pattern of the residual peaks. From the powder diffraction measurement it was found that there is a minimum at about 75 K for *a*, *b* and *c* parameters that suggests the isomorphic phase transition, as the forbidden in $P2_1/c$ group reflections are not observed.

In order to validate the consistency and credibility of the models at different temperatures, the CPs topological parameters of the corresponding covalent bonds in the phenyl rings and anharmonic fragments were compared. It appeared, that for 70 K data the interatomic distances are constantly the longest. The Laplacian values for the anharmonic/harmonic fragments were found systematically larger for 70-100 K data, that can be connected with their different treatment.

It was concluded, that the unexpectedly long interatomic distances at 70 K may be caused by the difficulty in the correct unit cell finding from the measurement using the four circle diffractometer with orange cryostat. In fact the recalculation of the bond distances using the unit cell parameters from the powder diffraction experiment, that should provide the most accurate results, brings much better agreement than the bond lengths from single-crystal measurements. Therefore the 'correct' D12 distances for 35-100K data (no powder diffraction data at 10 K) are longer at higher temperature, as the result of higher degree of the precision of the molecular geometry at lower temperatures.

Atomic charges

The multipolar (N_{val} - P_{val}) and integrated charges were analyzed for molecules **I**-**V**. The nitro group that is usually involved in the strong or weak H-bonds (shortest X-H···O=N contacts in **I**, **IV** and **V**) is the most negative in **I**, **IIIa** and **V** for multipolar and integrated charge definition, while for **II** and **IV** is less negative. The charges of the phenyl – imidazole parts of the molecules depend on the attached substituents. The charges in molecule **II** are well separated: the positive charge of the phenyl ring without additional substituents (multipolar charge 0.65 |e|, integrated charge 0.91 |e|) is neutralized by the negative charge of the nitroimidazole ring with the two strongly electronegative nitro and cyano groups. For the remaining molecules the difference is smaller, as there are chlorine atoms in *para* positions (**IIIa** and **IV**) or amino group in *meta* position (**I**). The charges and volumes of the two cyano groups (**II** and **IIIa**) are in a good agreement, as the parameters from molecule **II** were transferred to obtain a reasonable model of molecule **IIIa**. The multipolar charges of chlorine atoms are continuously slightly positive (0.00-0.19 |e|) while insignificantly negative for integrated charges of **IIIa** and **V** (-0.05 |e| and -0.12 |e|, respectively, and 0.12|e| in **IV**, which has the strongest halogen bonds).

Topological analysis of covalent bonds

The total electron density at Critical Point (ρ_{tot} at CP) plotted versus the interatomic distance for the certain type of bonds occupy quite narrow areas no matter the type of atoms involved in the bond, except for the short triple CN bond of the cyano group. For the homoatomic carbon-carbon bonds the total density at CPs is systematically the lowest for single $C_{imidazole}$-C_{methyl} bonds, increases for $C_{imidazole}$-C_{cyano} and is the highest for aromatic C_{phenyl}-C_{phenyl} and $C_{imidazole}$-$C_{imidazole}$ bonds. Similar trend is found for C-N covalent bonds: the single C_{phenyl}-$N_{imidazole}$ and $C_{imidazole}$-C_{nitro} bonds have the lowest ρ_{tot} values, while the double $C_{imidazole}$-$N_{imidazole}$ and triple C_{cyano}-N_{cyano} bonds have much higher ρ_{tot} values. The total electron density at the N=O CPs is systematically almost as high as for the triple cyano group, while at C-Cl CPs show the lowest value among all non-H atoms bonds.

The Laplacian values are slightly more spread for given type of bond and there is an opposite direction of the $\nabla^2\rho$ increase for C-C/C-N bonds compared with ρ_{tot}, *i.e.* higher negative values of the Laplacians at CPs are connected with higher positive values of the total electron density at CPs. Therefore the C-C and C-N bond types depending on the Laplacian values are arranged in the following order: a) C_{phenyl}-C_{methyl} > $C_{imidazole}$-C_{cyano} > C_{phenyl}-C_{phenyl} > $C_{imidazole}$-$C_{imidazole}$; b) C_{phenyl}-$N_{imidazole}$ > $C_{imidazole}$-N_{nitro} > $C_{imidazole}$-$N_{imidazole}$ > C_{cyano}-N_{cyano}. This tendency is also followed by C-Cl bonds but not by N=O bond, for which the ρ_{tot} was one of the highest, while the Laplacian has rather low negative values.

The main curvature (λ_3) dependence on the interatomic distance (D12) shows more contracted areas for the corresponding CPs. C-C and C-N CPs almost overlap (except the cyano group), while the remaining bonds do not differ by more than 10 e/Å^5 for chosen bond type. Among all covalent bonds summarized the highest λ_3 values are found for N=O and C≡N bonds.

The bonds involving halogen atoms show systematically the lowest values of the three above mentioned descriptors. The N-H bonds show systematically higher values of the density, negative Laplacian and the main curvature at CP than C-H bonds, as N-H bonds are relatively shorter and stronger.

Intermolecular interactions analysis

Several dependencies of the topological descriptors on the interatomic distances and Laplacian values were analyzed for the nitroimidazole in the analyzed series. The whole interactions were divided into three distinguished batches for easier classification: the H-bond types of contacts (D-H···A, batch 1), the H-H stabilizing interaction (batch 2), non-H atoms contacts ($\pi\cdot\cdot\pi$ together with halogen bonds and dipolar interactions and van der Walls contacts, batch 3).

No clear tendency was observed on the ρ_{tot}(D12) plot, even for selected type of contacts. The N-H···N/O relatively strong hydrogen bonds approachC-H···O/N contacts in terms of similar ρ_{tot} value, which in turn are classified as the weak H-bonds. There is a continuum of C-

H···O/N contacts, which fall in the range of weak H-bonds, region of overlap between H-bonds and vdW interactions (2.75-2.85 Å) and vdW interactions (> 2.85 Å). The C-H···Cl/C_π contacts (except one C-H···C_π) lie in the region of overlap, but mostly in the region of vdW interactions. The H-H stabilizing contacts are found below D12 = 2.45 Å, while some H-H contacts at 2.7 Å should be considered as vdW interactions. The non-H atom contacts are the longest with D12 = 3.7-3.8 Å being Cl···C and Cl··Cl.

Laplacian at the critical point plotted versus interatomic distance is divided into three batches mentioned before. The best exponential fitting is found for H-H contacts, with the correlation of 96%, and then still acceptable fitting level is reached for H-bond type of contacts (88%). The worst agreement for contacts not involving H-atoms can be explained by the different nature of interactions covered by this criterion – separated curves for halogen O···Cl contacts (86%) and C···Oπ··π contacts (76%) show much better agreement, while for the remaining ones there are too few points to give any firm conclusions.

Similar exponential lines were plotted for the λ_3(D12) dependence. The tendency to fit the regression line for H-H contacts (95%) and H-bond type of contacts (87%) is nearly the same as for Laplacian plots. Once again the exponential fitting of non-H atom contacts is the worst (54%), with some better fitting for separated C···O(73%) and O···Cl (75%).

The dependences of kinetic and potential energy densities on the interatomic distances is exponential, as pronounced as for the previous descriptors (for kinetic part: R^2 for H-bond types 86%, for H-H contacts 95% and for non-H atom contacts 61%), while potential energy density plots show more fuzzy areas of points representing the CPs, with the R^2 equal 81%, 92% and 54%, respectively. The dependence of the total energy density (H = V + G) on the main curvature λ_3 shows very good agreement for the H-bond type of contacts (R^2 = 93%) and H-H contacts (R^2 = 92%), while the non-H atoms interactions the fitting is much less effective (R^2 = 64%).

Finally the supplementary dependency plots were prepared for the Laplacian values at the CPs, as this indicator is the second derivative, very sensitive to the small changes in the total electron density at CPs, and its dependency on the interatomic distance may not properly reflect its nature. The $\nabla^2(\rho_{tot})$ correlation is not evident even if the total number of interactions is divided in the three batches. The R^2 factor for regression lines ranges from 71% to 76%. On the other side the phenomenological linear dependencies of the Laplacian on the main curvature and of the total energy densities (kinetic, potential and total) at CPs are observed, with strong correlations between the analyzed descriptors found for all observed interactions at once (R^2 for $\nabla^2(\lambda_3)$ is 98.5% and for ∇^2(G/V/H) ranges from 95.3% to 99.5%).

To conclude the presented interactions classification, one can notice, that even for weak interactions the exponential fitting is still reasonable, with the R^2 values worse than found in the literature, what can be explained by mostly weak interactions occurring between the nitroimidazole molecules. Each pair of interacting atom types should have its own fitting line.

In general the strong and weak H-bonds present the largest values of ρ_{tot}, $\nabla^2\rho$, λ_3 and energy densities at CPs, while the range of these descriptors for the remaining non-H atoms interactions, with the longest D12 distances, fall in the same range as weak van der Waals interactions.

The summary of the interactions is given in table below. The types of contacts are arranged from the strongest to the weakest, but in each batch there is a continuum of interactions, so the limits are rather fuzzy concept and the region of overlap may be found. The shortest and most energetic interactions are the strong and weak hydrogen bonds, which also have the highest values of the total electron density and Laplacian at the critical points. Then the H-H stabilizing and C≡N···C≡N dipolar contacts fall in the range of the middle strength contacts, with overlapping values of the main topological descriptors. Finally the last and weakest but still significant interactions are the π···π and halogen bonds, which in the analyzed series approach to the van der Waals contacts.

Table 1. Interactions summary.

interaction type	D12 [Å]	ρ_{tot} (e/Å^3)	$\nabla^2\rho$ (e/Å^5)	λ_3 (e/Å^5)	$G(\mathbf{r}_{CP})$ kJ/mol·au^3	$V(\mathbf{r}_{CP})$ kJ/mol·au^3	$H(\mathbf{r}_{CP})$ kJ/mol·au^3
strong H-bonds	2.03-2.30	0.046-0.070	1.07-2.47	1.44-2.96	21.2-47.8	-13.4 to -28.1	7.8-19.7
weak H-bonds	2.34-2.79	0.020-0.061	0.39-1.06	0.49-1.55	7.5-22.9	-4.4 to -17.0	3.1-6.8
H-H stabilizing	2.10-2.45	0.023-0.042	0.42-0.75	0.57-0.99	8.1-14.8	-5.0 to -9.3	3.1-5.4
C≡N···C≡N	3.16-3.22	0.055-0.056	0.60-0.70	0.78-0.83	13.5-14.9	-10.3 to -11.3	3.0-3.6
π···π	3.08-3.70	0.013-0.056	0.23-0.77	0.30-0.97	4.4-16.4	-2.5 to -12.0	1.9-4.4
halogen bonds	3.12-3.75	0.015-0.049	0.28-0.60	0.35-0.91	5.5-13.0	-3.2 to -9.6	2.0-3.5
vdW contacts	2.81-3.37	0.009-0.051	0.20-0.58	0.25-0.86	3.9-12.7	-2.2 to -9.6	1.6-3.2
H-H vdW contacts	2.69-2.73	0.011-0.018	0.19-0.25	0.26-0.34	3.7-5.0	-2.1 to -3.1	1.6-1.9

The set of chosen molecules from the imidazole family could be enlarged by some other structures (having interaction types of different strength), for which the standard resolution data were collected in our laboratory, but crystals suitable for the high resolution measurements were not obtained yet. Changing the crystallization techniques could bring the good quality crystals and finally increase the number of points at the fitting curves drawn for the CPs descriptors.