Maura Malińska
Faculty of Chemistry, University of Warsaw

Supervisors: prof. dr hab. Krzysztof Woźniak,
prof. dr hab. Andrzej Kutner

Structure and Charge Density Studies of Pharmaceutical Substances in the Solid State

Abstract of PhD Thesis

Electrostatic forces are one of the most important factors contributing to the formation of crystal structures and protein-ligand complexes. Experimental and theoretical charge density data provide insight into the electronic nature of the molecular crystal structure and, thus, enable not only qualitative but also quantitative analysis of intermolecular interactions, in particular their strength. This fact should be used in the ligand design process as a more general approach than those based on force field methods.

In my PhD thesis, I have used for the first time the aspherical atom databank approach, Hirshfeld surface analysis and Quantum-Theory-of-Atoms-in-Molecules as a foundation for the analysis of interactions between ligands and proteins. Comparative studies regarding the electrostatic interaction energy calculation and electrostatic potential analysis can explain the binding preferences of small molecules with proteins. Analysis of numerous complexes with the same receptor can lead to the design of ligands of proteins with the lowest electrostatic interaction energy.

My PhD project consists of two parts. The first part includes the experimental charge density studies of protein kinase inhibitors and analysis of their interaction with proteins. The experimental charge density of sunitinib was compared to the analogous representation derived on the basis of the aspherical databank, which supported once again its applicability to reconstruction of charge distributions. Additionally, the experimentally obtained charge density distribution of sunitinib was transferred to the complex of sunitinib with protein kinases. The final outcomes of electrostatic interaction energies were comparable. The second part of my Thesis is based on the analysis of 24 complexes between Vitamin D analogs and a Vitamin D Receptor.

Additionally, I have established experimental charge densities of two biologically active compounds WP1066 and sunitinib malate. Both of them constitute a basis for characterization of interaction character and strength. First one, WP1066, is a novel type STAT3 inhibitor developed to inhibit its activation. The experimental charge density distribution of WP1066 using

high resolution X-ray diffraction data collected at 90K allows quantitative evaluation of the nature of the C-Br...N≡C halogen bonds in the molecular crystal. Moreover, the occurrence of both halogen bonds and hydrogen bonds in this structure gave the opportunity for comparison of the interaction energies and details of the charge density distribution. Topological analysis of the total charge density was performed for the halogen bond and the hydrogen bonds. The $\rho(\vec{r})$ values at the critical points of these interactions are: 0.059(1) eÅ^{-3}, 0.77(2) eÅ^{-5} and 0.19(4) eÅ^{-3} and 1.45(6) eÅ^{-5}, respectively. Electron density is anisotropically distributed around the bromine atoms in the crystal studied, therefore the bromine atoms exhibit electrophilic character along the axis of the C-X bond and nucleophilic character perpendicular to this axis. The interaction energy of the halogen bond is comparable to the strength of C-H...O interactions and it is ten times weaker than the energy of the strong hydrogen bond found in this structure. Charge transfer (0.33 e) occurs between molecules due to the halogen bond interaction character. This phenomenon enhance the strength of the interaction between particular dimers from in the WP1066 crystal structure.

The second X-ray high resolution measurement was done for sunitinib malate. The strongest hydrogen bonds in the crystal lattice can be interpreted as charge-assisted hydrogen bonds and resonance-assisted hydrogen bonds. A set of selected kinase-protein complexes was explored using the UBDB-derived charge density, as this allows analysis of the electrostatic properties of VEGRF2, CDK2, G2K, KIT and IT kinases complexes. The Hirshfeld surface analyses revealed similar patterns for interactions occurring between sunitinib molecules in the sunitinib malate crystal and in the protein binding pockets. However, as expected, the packing of sunitinib in the sunitinib malate crystal is more condensed. The analysis of interactions by means of interatomic distances and van der Waals radii is insufficient. The QTAIM approach was helpful in detecting the most conserved interactions between the ligand and the residues forming the binding pocket and revealed a characteristic pattern of interactions that includes more interactions than those pinpointed in literature - three hydrogen bonds, two weak C-H...O interactions and four C-H...π contacts with the corresponding residues. These residues are the gatekeeper residues and following four residues of the hinge region, and also either leucine or isoleucine from the Gly-rich loop. As indicated by the simple empirical formula proposed by Espinosa & Lecomte, the head of the sunitinib molecule interacts with similar strength among the studied structures, only in the case of the CDK2 complex is this interaction significantly stronger.

A more general approach which also takes into account long range interactions is the combination of the aspherical atom databank approach and the Exact Potential Multipole Method, which enables the calculation of pure electrostatic interaction energies. Considering all the 24 selected residues

forming the binding pocket, the energetic results fall into a small range of interaction energies for the sunitinib head, where only the VEGFR2 kinase is characterized by a significantly lower interaction energy. The results of the additional Exact Potential Multipole Method calculation suggest that sunitinib interacts with VEGFR2 in its unprotonated form. Taking into account also the tail of sunitinib, the energetic results are more diverse, however, still comparable. It can be explained by the conformational flexibility of sunitinib, especially in the side chain region. A sunitinib molecule can adjust its conformation to enhance the electrostatic interaction with kinases. For instance, in the CDK2 structure it adopts a less favorable conformation so as to saturate the hydrogen bond donors and acceptors of Asp86 and Leu134.

In contrast to the ATP molecule, sunitinib is small and fits to the binding pocket irrespective of the activation state of protein kinases (small pocket for inactive kinases: VEGRF2, KIT, and bigger pocket for the active ones: CDK2, G2K, and IT proteins). The interaction energy with the highly-conserved DFG triad, responsible for closing the binding pocket, is similar for all the studied complexes. The KIT kinase protein structures in the inactive (with sunitinib) and active (with ADP) form were analyzed. Both ligands interact similarly with the residues of the hinge region forming hydrogen bonds of comparable interaction energy. Moreover the total energies for their own binding pockets are in the same range. In the KIT kinases, the reversible conformational change DFGout to DFGin involves modification of their electrostatic potential. They exhibit more negative and more positive values of electrostatic potential mapped on van der Waals surfaces for the inactive and active form, respectively. Those features are in agreement with the electrostatic potentials of sunitinib and the ATP molecule. These results explain the preference of the inactive form of KIT over the active one for the sunitinib molecule.

Analysis of the crystal structure of Vitamin D analogs bearing similar molecular structures with a complex of a Vitamin D Receptor enabled the design of new agonists. The electrostatic interaction energies available after the reconstruction of charge density with the aid of the UBDB gave an opportunity to understand possible interactions between the natural ligand, Vitamin D analogs and the receptor. The structural analysis of the available human/rat Vitamin D Receptor - ligand complexes, together with the alignment techniques, revealed that there are 29 residues which contact the ligand. Trp286, which is specific to VDR among the representative of the Nuclear Receptor Family, plays the crucial role of positioning the ligand. Trp286 forms an intramolecular hydrogen bond network with Ser275, which in turn is hydrogen bonded to Met272. Trp286 forms dispersive interaction, mostly C-H...π, with an average strength of -4 kcal$\cdot$mol^{-1} indicating its very strong character. The electrostatic component of the interaction energy is not the most important component of the interaction. Ligands curve around the helix H3 of the Vitamin D Receptor, with its A-ring interacting

with the H5 helix and the 24-/25-hydroxyl group with residues of helices H7 and H11. The ligand binding pocket is primarily composed of hydrophobic residues, however there are 6 hydrogen bonds, which is characteristic for all ligand studied. The 1-hydroxyl forms two hydrogen bonds with Ser237 (H3) and Arg274 (H5), whereas the 3-hydroxyl forms two hydrogen bonds with Ser278 (H5) and Tyr143. They electrostatic interaction energies strongly contribute to the total interaction energy with binding pocket, with an average strength of -8 kcal·mol^{-1}, -19 kcal·mol^{-1}, -11 kcal·mol^{-1} and -12 kcal·mol^{-1}, respectively. A water channel is also observed, with water molecules hydrogen bonded to Arg274 leading to the solvent. The analogs with a long chain at the 2- position disturb the water channel, therefore only the 2- methyl derivatives contribute to the formation of favorable interactions. The aliphatic chain adopts an extended conformation and is surrounded by hydrophobic residues. The 25-hydroxyl group is hydrogen bonded to His305 (which lies in the loop connecting H6 and H7) and His397 (which lies in helix H11) with electrostatic interaction energies of -13 and -11 kcal·mol^{-1}.

The Ees for all studied Vitamin D analogs are in the range -53 kcal·mol^{-1} to -111 kcal·mol^{-1} depending on the change in molecular structure of the ligand. The Vitamin D analogs with the lowest electrostatic interaction energies are OCC, KH1, O1C and EIM (ligand name as the PDBIDs). The performed analysis has led to four newly proposed ligands which should have the lowest Ees. The geometries of complexes of the proposed ligand with VDR were obtained by the docking procedure (Autodock4.3) followed by Ees calculations. The final results for TB1 and TB2 (more promising) are -153 and -120 kcal·mol^{-1}.

To summaries, my PhD thesis provides one of the first attempts to use the charge density modeling as the tool in drug design process. It can be applied to any drug or drug candidate with available single crystals and/or known crystal structure of its complex with the drug target (protein, DNA, macromolecules). The QTAIM analysis of the charge density can cast more light on interactions in ligand-protein complexes. It is however crucial to further test and develop this approach, especially from the software side i.e. automation, which enable the analysis of numerous ligands. Many of the presented results have already been published (see my publication list). Other manuscripts based on the work in this thesis are currently in the process of being published or are under preparation.