High Throughput Calculations as an Elevator on the Way from Chemical Structure to Novel Materials

I.D. Yushina, G.I. Makarov, Y.V. Matveychuk, E.V. Bartashevich
Nanotechnology Research & Education Centre, South Ural State University, Chelyabinsk, Russia
E-mail: iushinaid@susu.ru

Abstract. Development of new materials and drugs presently requires trial chemical methods that are time consuming and that depends on the chance more than we would have liked. We aim to describe, how the high throughput calculations can be applied for computational predictions of the organic crystalline structure and its properties. In this work we tried to illustrate the examples of techniques that were applied to accelerate the design of iodine-contained organic materials with nonlinear optical properties, to explain the plasticity of maleate amino acids crystals by means of analysis of hydrogen bonds orientation, to clarify the biological activity of traditional antibiotics using molecular dynamics modeling.

1. Introduction
Nowadays, the high throughput calculations have penetrated and infiltrated all main sections of chemistry. Advances in the field of physical chemistry are unthinkable without solving the problems of structure and properties predicting, such as optical, elastic, electrical, thermodynamic, thermochemical properties. In turn, the properties predictions are based on virtual models of the molecules, crystals, polymers and solids, as well as on our understanding how fragments of multicomponent systems are bound to each other.

Nowadays due to the diffusion of supercomputer technologies in our current knowledge, the new disciplines are born at the close junctions of various fundamental sciences. At least two prominent examples can confirm this fact: the formation of quantum crystallography [1, 2] and the relevance of skills that chemoinformatics [3] forms. Quantum crystallography is a research area exploiting the fact that parameters of quantum-mechanically valid electronic model of a crystal can be derived from the accurately measured set of X–ray coherent diffraction structure factors [1]. On the one hand, the chemistry describes how atoms and molecules interact via chemical bonds to form new chemical compounds; on the other hand, the informatics involves the practice of information processing and the engineering of information systems. Thus, the chemoinformatics is aimed to reproduce the data into information and information into knowledge for making the better and faster decisions in material science and drug design.

The purpose of our present work is to illustrate two important aspects: 1) how the understanding the nature of chemical bonds helps us in the structure modeling and the properties predictions; 2) how the digital tools of computational chemistry inextricably linked with the modern approaches of the drug design.
Pursuing the goals, firstly, we consider how the concept of halogen bonding and the methods of evolutionary algorithms [4] could be applied to accelerate design of organic materials with nonlinear optical properties. Secondly, we describe an example in which quantitative analysis of the hydrogen bonds orientation helped to explain the plasticity of amino acid maleates crystals. Thirdly, we touch the case study based on molecular dynamics modeling that returned the attention and interest to the traditional antibiotics.

2. Computational design strategy for NLO properties of iodine–containing organic crystals

One of our specific tasks was to develop and deploy methodology for modeling of novel materials based on the modification of the molecular and crystal structure of iodine–derived compounds bound by the iodine–iodine halogen bonds [5, 6] and the other electrostatically driven interactions. Our interest to iodine–containing materials is justified by the high polarizability of iodine atoms and their ability to build the multiple non–covalent interactions, forming infinite chains and nets in crystals.

Recently the computational design strategy for iodine–containing organic crystals was proposed [7]. In general, the strategy is based on lattice energy minimization for the crystal packing predictions using evolutionary algorithms implemented in USPEX program [8]. The protocol of structure and properties prediction for iodine–containing organic crystals [9] is illustrated on Fig. 1A

Suggested protocol allowed us to predict the unknown polymorphs of iodine-derived benzenes, with the higher efficiency of second harmonic generation. We obtained the justification for NLO properties impact on hyperpolarizabilities of crystals with iodine halogen bonding [10]. Next, we illustrate the results of our modeling and NLO properties predictions for 2–iodo–3–hydroxypyridin, which crystalline structure was described earlier [11]. Here we deepen our understanding of the role of halogen and hydrogen bonding as well as focus on several computational methodology observations in wider context.

Periodic Kohn-Sham calculations were carried out in CRYSTAL14 [12] with PBE0 exchange–correlation functional with D2 dispersion correction and localized atomic basis sets. The vibrational frequencies were computed via coupled perturbed Kohn–Sham analytical approach

Figure 1. A: The protocol of structure and properties prediction for iodine–containing organic crystals. B: 2–Iodo–3–hydroxypyridine molecules interact via L–shaped I...I halogen bonds.
Crystal structure of 2–iodo–3–hydroxypyridin has two significant structure–determining motives: the N...H—O hydrogen bond that organizes molecules in anti–parallel way and is directed mainly in a crystallographic axis direction or the I...I halogen bond that forms L–shaped zigzags along c crystallographic axis direction (Fig. 1B).

Thus, the main task of crystal packing prediction and reproduction in this case was to preserve these structural features.

Besides analysis of sustainability of main packing motives, it is essential to control space groups that are obtained after generation of modelled structures and their optimization. The study of second order NLO properties, which are commonly expressed as the ability for second harmonic generation, one usually takes into consideration only noncentrosymmetric space groups in which second order hyperpolarizabilities \( \beta \) have nonzero values.

As we have previously showed [10], we succeeded in reproducing the crystal packing of experimental 2–iodo–3–hydroxypyridin structure and obtained values of \( \chi(2)_Z \) — the only nonzero direction in tensor of second order susceptibilities \( \chi(2) \) according to symmetry restrictions of Pna2\(_1\) space group. Moreover, due to the usage of generic algorithms is was possible to obtain model structure in Cc space group that due to absence of symmetry limitations has also nonzero component \( \chi(2)_X \) in addition to \( \chi(2)_Z \), thus resulting higher overall values of NLO activity.

Combination of hydrogen and halogen bonding features allows analyzing the influence of both of them on NLO properties.

The effect of hydrogen bonding on hyperpolarizability values is mostly manifested in the direction of a crystallographic axis. Only in a direction there is a clear trend: the shorter is a axis value, the shorter is N...H—O distance and the higher is the value of total hyperpolarizability. Other crystallographic directions are less sensitive to the changes of N...H—O distance, although the values of hyperpolarizability change drastically.

Thus, in case of 2–iodo–3–hydroxypyridine hydrogen bonding has greater impact on the stabilization of experimental and generated structures while halogen bonding is more significant for the emergence and increasing of NLO properties.

3. Effect of the H–bonds orientation the features of crystal structure and elastic properties

The numerical description of crystal mechanical properties was offered for a long time in the form of stiffness and compliance fourth rank tensors. However, the solutions describing the anisotropy of crystalline materials and following directly from these tensors were represented only during last decade. These are general elastic-anisotropy index \( A^* \) [13], universal elastic anisotropy index \( A^U \) [14], log–Euclidean anisotropy index \( A^L \) [15]. The influence of the anisotropy of intermolecular interactions on crystal mechanical properties was studied by Desiraju et al. [16]. However, the complex comparison of H–bonds orientation with elastic moduli anisotropy in combination with \( A^U \) calculation was presented in our previous work [17]. In this approach the quantitative analysis of H–bonds orientation and stiffness tensor anisotropy was made for L–alanine crystal (A) [18], L–isoleucinium hydrogen maleate hemihydrate (IL) and L–leucinium hydrogen maleate (L) [19]. The structure relaxation and the stiffness tensor calculation were carried out on B3LYP/6–31G(d,p) level of theory. The universal elastic anisotropy indexes \( A^U \) and the spatial dependences of elastic moduli can be calculated using ELATE software [20].

The anisotropy of Young’s and shear moduli is approximately identical for the IL and A crystals, as well as the differences of minimum and maximum values of shear modulus. The corresponding values for L crystal are qualitatively bigger. The calculations of \( A^U \) have quantitatively confirmed this fact. The values of the universal elastic anisotropy index are
1.38 (IL), 3.76 (L) and 2.36 (A). As the $A^U$ value is one of key indicators of mechanical flexibility/brittleness of a crystal, we can assume that the L crystal can show flexibility in some directions, but other samples do not demonstrate such property. This assumption has been confirmed experimentally for L and IL earlier [19].

The comparison of the stiffness tensor properties with the structural features of these crystals has allowed establishing the main reason of differences in mechanical properties. The main structure–forming interactions in crystals of amino acids and their salts are intermolecular H–bonds.

The lack of molecular layers in A crystals assumes the small probability of the elastic properties manifestation caused mainly by the shift of these layers. The calculated anisotropy of elastic moduli of A crystals is generally small and corresponds to such assumption. The different situation is observed with layered L and IL crystals. The smallest resistance to shift in these crystals is observed in the direction parallel to layers, what is caused by lack of strong interactions between layers. However, in IL crystals the possibility of layers shift is complicated by action of other factors, mainly steric. That is, the layered structure is not a sufficient condition for possibility of crystal flexibility manifestation in certain directions. The calculation of stiffness tensor in this case gives the chance to reduce the series of the crystals capable to have such unique properties.

4. Molecular dynamics simulation method in the study of bacterial ribosome function

The modeling of biopolymers by the method of classical molecular dynamics (MD) enables to investigate their conformational mobility, to estimate the stability of intramolecular non–covalent interactions, for example, hydrogen bonds, and to describe their interaction with low–molecular ligands, ions and water molecules. The MD method consists of the evolution of molecular systems during time; in the method of classical MD, the movements of molecules is described by classical mechanics model, so that molecular systems are represented as a set of material points interacting with each other. The potential energy of the system is given by the sum of the potentials describing some or other covalent and non–covalent interactions. We use GROMACS [21] software to perform MD simulations. Biopolymers were modeled with the parm99sb force field [22], while small ligands are modeled with the GAFF [23].

An important example of the application of classical MD is the simulations of a bacterial ribosome [24]. MD simulations enables to investigate consistent conformational alterations in ribosomal RNA [25], which is especially important in the context of multiple evidence of the developed allosteric relationships regulating the function of the ribosome [26]. Methods of MD enable to simulate the interaction of antibiotics with a ribosome, explaining the differences in their activity [27] and the mechanism of their action.

The ribosome is a very large ribonucleoprotein complex, so that its simulation requires a huge expenditure of computational resources. However, when studying the binding of low molecular ligands to the ribosome, it is possible to model a sufficiently large fragment of the ribosome, completely including the binding site and its nearest vicinity. Such an approach preserves the local conformational mobility of biopolymers and excludes the global one, allowing to considerably economize required computational resources. We use cubic fragment of core area of large subunit of E. coli ribosome. This fragment includes all amino acid and nucleotide residues were set apart, as well as K$^+$ and Mg$^{2+}$ in which at least one atom fell into a cubic area with 7 nm edge including the entire nascent peptide exit tunnel and peptidyl transferase center in such a way that the center of this region was located at the nascent peptide exit tunnel aligned along an imaginary applicate axis. During simulation this area is centered in a cubic cell with 8.8 nm edge filled with water molecules so that the system edges were covered by a 0.9 nm thick solvent layer.
Figure 2. A: Conformation of TylPhe phenylalanyl substituent modeled by MD methods; the hydrogen bonds are shown by dashes. B: Interaction between chloramphenicol and E. coli ribosome, staying in canonical A,A/P,P–state, modeled by MD methods. Hydrogen bond is shown by black dashes.

In our previous work [28] we have modeled the structure of the complex of bacterial ribosome with TylPhe antibiotic obtained from a tylosin. This compound is superior in activity to the tylosin, despite the fact that it is unable to covalently bind to 23S rRNA. We established the conformation of the phenylalanyl substituent, which explains this effect (Fig. 2A). Moreover, the molecular dynamics method allows the structure of antibiotic complexes to be simulated with a certain functional state of the ribosome: for example, recently we modeled the structure of the chloramphenicol complex with the E.coli ribosome stalled in the canonical A,A/P,P–state, which explains the results of biochemical studies the binding of chloramphenicol [29] (Fig. 2B).

5. Conclusion

Thus, in this way we have illustrated how the global challenges in material science and drug design can be solved with the help of high throughput calculations and supercomputer technologies applied as an elevator in modern structural chemistry. We have demonstrated the examples how the following problems can be solved: to expand the possibilities in the organic crystal structure modeling; to deepen the understanding of the chemical bonds and non-covalent interactions nature; to develop methods for the nonlinear optical and elastic properties predictions. Such studies based on the time-consuming calculations [30] are of direct importance for the development and improvement of novel materials and antibacterial drugs necessary to maintain the modern quality of life.

Acknowledgments

This work was supported by the Ministry of Education and Science of the Russian Federation, grant 4.1157.2017/4.6, and by the Government of the Russian Federation, Act 211, contract No. 02.A03.21.0011. Quantum–chemical calculations have been performed using computational possibilities of TORNADO supercomputer, South Ural State University.

References

[1] Tsirelson V 2018 Journal of Computational Chemistry 39 1029–1037
[2] Massa L, Huang L and Karle J 1995 International Journal of Quantum Chemistry 56 371–384
[3] Madzhidov T, Baskin I, Antipin I and Varnek A 2015 Introduction to chemoinformatics (series of benefits) vol 10–2 (Kazan) pp 198–200
[4] Zhu Q, Oganov A R, Glass C W and Stokes H T 2012 Acta Crystallographica Section B 68 215–226
[5] Desiraju G, Ho P, Kloo L, Legon A, Marquardt R, Metrangolo P, Politzer P, Resnati G and Rissanen K 2013 *Pure Appl. Chem.* **85** 1711–1713

[6] Bartashevich E and Tsirelson V 2014 *Russian Chemical Reviews* **83** 1181–1203

[7] Glass C W, Oganov A R and Hansen N 2006 *Computer Physics Communications* **175** 713 – 720

[8] Masunov A E, Torres K, Dyakov A A, Yushina I D and Bartashevich E V 2018 *The Journal of Physical Chemistry C* **122** 22622–22631

[9] Tkatchenko A, DiStasio R A, Car R and Scheffler M 2012 *Phys. Rev. Lett.* **108** 236402

[10] Yushina I D, Masunov A E, Lopez D, Dyakov A A and Bartashevich E V 2018 *Crystal Growth & Design* **18** 5069–5079

[11] Saha B K, Nangia A and Nicoud J F 2006 *Crystal Growth & Design* **6** 1278–1281

[12] Dovesi R, Orlando R, Erba A, Zicovich-Wilson C M, Civalleri B, Casassa S, Maschio L, Ferrabone M, De La Pierre M, D’Arco P, Noël Y, Caus M, Rérat M and Kirtman B 2014 *International Journal of Quantum Chemistry* **114** 1287–1317

[13] Ledbetter H and Miglioria A 2006 *Journal of Applied Physics* **100** 063516

[14] Ranganathan S I and Ostoja-Starzewski M 2008 *Phys. Rev. Lett.* **101** 055504

[15] Kube C M 2016 *AIP Advances* **6** 095209

[16] Mishra M K, Ramamurty U and Desiraju G R 2016 *Current Opinion in Solid State and Materials Science* **20** 361 – 370

[17] Matveychuk Y V, Bartashevich E V and Tsirelson V G 2018 *Crystal Growth & Design* **18** 3366–3375

[18] Simpson Jnr H J and Marsh R E 1966 *Acta Crystallographica* **20** 550–555

[19] Arkhipov S G, Rychkov D A, Pugachev A M and Boldyreva E V 2015 *Acta Crystallographica Section C* **71** 584–592

[20] Gaillac R and Coudert F X ELATE: Elastic tensor analysis. http://progs.coudert.name/elate

[21] van der Spoel D, Lindahl E, Hess B and Kutzner C 2008 *J. Chem. Theory Comput.* **4** 435–447

[22] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A and Simmerling C 2006 *Proteins: Structure, Function, and Bioinformatics* **65** 712–725

[23] Wang J, Wolf R M, Caldwell J W, Kollman P A and Case D A 2004 *J. Comput. Chem.* **25** 1157–1174

[24] Makarov G I, Makarova T M, Sumbatyan N V and Bogdanov A A 2016 *Biochemistry (Moscow)* **81** 1579–1588

[25] Makarov G I, Golovin A V, Sumbatyan N V and Bogdanov A A 2015 *Biochemistry (Moscow)* **80** 1047–1056

[26] Makarova T and Bogdanov A 2017 *Biochemistry (Moscow)* **82** 1557–1571

[27] Shishkina A V, Makarova T M, Tereshchenkov A G, Makarov G I, Korshunova G A and Bogdanov A A 2015 *Biochemistry (Moscow)* **80** 1500–1507

[28] Makarov G I, Sumbatyan N V and Bogdanov A A 2017 *Biochemistry (Moscow)* **82** 925–932

[29] Makarov G and Makarova T 2018 *Biochim. Biophys. Acta* **1862** 2940 – 2947

[30] Kostenetskii P and Semenikhina P 2018 SUSU supercomputer resources for industry and fundamental science 2018 Global Smart Industry Conference (GloSIC) (IEEE) pp 1–7