L12

THE EUROPEAN XFEL AND ITS POTENTIALS IN STRUCTURAL BIOLOGY

Janos Hajdu

Uppsala University and The European XFEL

At the beginning of 2017, the European X-ray Free-Electron-Laser (XFEL) in Hamburg will begin user operations. Free-electron lasers are the most brilliant sources of X-rays to date, exceeding the peak brilliance of conventional synchrotrons by a factor of 10 billion, and improving. In the duration of a single flash, the beam focused to a micron-sized spot has the same power density as all the sunlight hitting the Earth, focused to a millimetre square. The interaction of an intense X-ray pulse with matter is profoundly different from that of an optical pulse. A necessary goal of research with these machines is to explore photon-material interactions in strong X-ray fields. The aim in biology is to step beyond conventional damage limits and develop the science and technology required to enable high-resolution imaging of both crystalline and non-crystalline biological objects at high resolution. Eligible targets include single virus particles, organelles, cells, nanocrystals, engineered nanoclusters and isolated macromolecules. The talk will summarise developments at the European XFEL and provide an overview of some of the biological results from the Linac Coherent Light Source (LCLS), the first hard X-ray free-electron laser. One of the aims of the talk us explore possibilities for interested Czech scientists to participate in revolutionary new experiments at the European XFEL.

Friday, March 20, Session IV

L13

STRUCTURAL BIOINFORMATICS - A BRIDGE BETWEEN STRUCTURAL BIOLOGY AND BIOINFORMATICS

J. Koča[1,2], R. Svobodová Vařeková [1,2], D. Sehnal [1,2], C.-M. Ionescu[1], S. Geidl[1,2], L. Pravda[1,2], D. Jaiswal[1], V. Horský [2], M. Wimmerová [1,2]

[1]CEITEC - Central European Institute of Technology, Masaryk University, Brno, Czech Republic
[2]National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno, Czech Republic

Structural biology of today is a well-defined field of science. It is not so for bioinformatics, which is understood from very narrow classical view (informatics of the genome) on the one hand, to very wide concept of informatics of any biology related information. In all cases, bioinformatics becomes an important field of science as the amount of bio-related information, especially from Next Generation Sequencing (NGS), is increasing dramatically, and, for the time being, there is no software tool available that would be able to extract all the biological information hidden in the data.

In contrary, structural bioinformatics is relatively well defined part of bioinformatics (see, for example [1-2]), which is related to the analysis and prediction of the three-dimensional structure of biological macromolecules.The term *structural* has the same meaning as in *structural* biology, and structural bioinformatics can be seen as a part of *computational structural* biology. Even if the grow of 3D structural data is much lower compared to NGS, also here the increase is exponential and calls for new approaches to extract structurally and/or biologically relevant information.

In our group, we have developed several software tools that are able to help in solving such a task. These are MotiveQuery [3] for quick finding and extraction of biomacromolecular fragments, SiteBinder [4] for fast and accurate comparison of these fragments, MotiveValidator [5] and ValidatorDB [6] for validation of ligands and non-standard residues, and AtomicChargeCalculator [7] for calculation of partial atomic charges. Last but not least, we have developed also MOLE [8], a software tool for detection and characterization of channels and pores in biomacromolecules. All the software tools are accessible from the link http://ncbr.muni.cz/WebChemistry.

1. Bourne, P.E., and Gu, J. (2009) Structural Bioinformatics (2nd edition), John Wiley & Sons, New York, ISBN 978-0-470-18105-8.
2. Bourne, P.E., and Weissig, H. (2003) Structural Bioinformatics, Wiley ISBN 0-471-20199-5.
3. Sehnal, D., Pravda, L., Svobodová Vařeková, R., Ionescu, C.-M., and Koča, J.: MotiveQuery: Web application for fast detection of biomacromolecular fragments in whole Protein Data Bank. (prepared for submission).
4. Sehnal, D., Svobodová Vařeková, R., Huber, H. J., Geidl, S., Ionescu, C. M., Wimmerová, M., and Koča, J. SiteBinder: An improved approach for comparing multiple protein structural motifs. J. Chem. Inf. Model. (2012), 52(2), 343–359.
5. Svobodová Vařeková, R., Jaiswal, D., Sehnal, D., Ionescu, C.-M., Geidl, S., Pravda, L., Horský, V., Wimmerová, M., and Koča, J.: MotiveValidator: interactive web-based vali-

dation of ligand and residue structure in biomolecular complexes. Nucleic Acids Res. (2014), 42, W227-33.

6. Sehnal D., Svobodová Vařeková R., Pravda L., Ionescu C.-M., Geidl S., Horský V., Jaiswal D., Wimmerová M., and Koča J.: ValidatorDB – database of up-to-date and comprehensive validation results for ligands and non-standard residues from the Protein Data Bank. Nucleic Acids Res. (2014), 43(D1), D369–D375.
7. Ionescu, C.-M., Sehnal, D., Svobodová Vařeková, R., Geidl, S., Falginella, F.L., Pant, P., Pravda, L., Mishra, S.K., and Koča, J.: AtomicChargeCalculator: Interactive Web-based calculation of atomic charges in large biomolecular complexes and drug like molecules (prepared for submission).
8. Berka, K., Hanák, O., Sehnal, D., Banáš, P., Navrátilová, V., Jaiswal, D., Ionescu, C.-M., Svobodová Vařeková, R., Koča, J., and Otyepka, M. MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. Nucleic Acids Res. (2012), 40(W1), W222–W227.

L14

ENHANCER-PROMOTER INTERACTION FACILITATED BY TRANSIENT G-QUADRUPLEXES FORMING BETWEEN THEM

Hedi Hegyi

Structural Biology Program, CEITEC, Masaryk University, Kamenice 5, Brno, CZ-628 00, Czech Republic

G-quadruplexes are nucleic acid sequences that are rich in guanine and are capable of forming a four-stranded structure through Hoogsteen hydrogen bonding. G-rich regions capable of forming G-quadruplex (G4) structures are highly concentrated near promoters and transcription start sites suggesting a role in gene regulation. They are less often found on the template strand than on the non-template strand where they either inhibit or enhance transcription, respectively. However, their potential role in enhancers and in other distal regulatory elements has not been assessed yet. Here we show that DNAse hypersensitive (DHS) cis regions with regulatory roles are also enriched in Gs and their G-content correlate with that of their respective promoters. Besides local G4s, the distal cis regions have the capability to form G-quadruplexes together with the promoters, each contributing half of a G4 only. This model is supported more for the non-template strand and we hypothesised that the G4 forming capability of the promoter and the enhancer non-template strand could facilitate their binding together and making the DHS regions accessible for the transcription factory. As G4s are targets of intense research and cancer drug development, we anticipate that our model will induce experimentalists to verify it in the lab and contribute to a better understanding of such a basic phenomenon as gene expression regulation.

L15

HYDRATION OF AMINO ACID RESIDUES IN PROTEINS: WHAT CAN WE LEARN FROM DATA MINING?

Lada Biedermannová, Jiří Černý, Bohdan Schneider

Inst. of Biotechnology, Laboratory of Biomolecular Recognition,
Vídeňská 1083, 142 20 Prague, Czech Republic
lada.biedermannova@ibt.cas.cz

It is nowadays accepted that water is not only a passive medium, but a key determinant of protein structure, dynamics and function, and that protein-water interactions govern various processes, including protein folding, enzymatic catalysis, and molecular recognition. Water does not simply fill up the available space around proteins, but occupies specific sites and forms localized clusters, determined by its hydrogen-bonding capabilities. Distribution of water around amino acid residues in proteins has been the subject of several studies, but only on a small number of crystal structures [1], or without consideration of the amino acid conformation, i.e. secondary structure and side-chain rotameric state [2].

We therefore decided to study the hydration patterns for all twenty standard amino acids in their main conformational states (rotamers), using large number of high-resolution protein crystal structures. Specifically, we used a set of 2,818 PDB structures of monomeric proteins with resolution better than 1.8 Å, maximum R-factor value of 0.22 and mutual sequence identity of the protein chains of 50% or less. The contacts of each amino acid residue with waters within 3.2 Å were detected. Residue conformations were clustered separately in each class defined by residue type, secondary structure (alpha helix/beta sheet) and chi1 rotameric state (as gauche+, gauche-, or trans) using quality threshold algorithm. Clusters of residues with the associated water molecules were then subjected to the method of density representation [3] in order to identify the preferred location of hydration sites. Briefly, atom positions of water oxygens were transformed using a Fourier transform technique to the corresponding electron densities. Density peaks of water molecules were then detected

and positions of hydration sites, their occupancies, and B-factors refined using standard crystallographic procedures.

The result of our study is a detailed atlas of the structure of protein hydration, containing the hydration sites, i.e. most populated positions of waters around each amino acid residue type in each of its main conformational states. Analysis of these hydration sites revealed frequent occurrence of positions where water interacts simultaneously with the side-chain and main-chain of the amino acid residue. By comparing the hydration of various conformers we also observed strong dependence of the positions of hydration sites on the amino acid conformation. Thus, our analysis revealed the spatial distribution of the preferred water positions in the first hydration layer of proteins. The hydrated amino acid rotamers obtained from our study can be used in many areas of structural biology, from molecular replacement and crystallographic refinement, to the improvement of accuracy of *ab initio* protein structure prediction methods.

The study was conducted at the Institute of Biotechnology AS CR (RVO:86652036) and was supported by project BIOCEV CZ.1.05/1.1.00/02.0109 from the ERDF and by the Czech Science Foundation (GA CR) grant no. P205/12/P729.

1. N. Thanki, J. M. Thornton, J. M. Goodfellow, J. Mol. Biol. 1988, 202, 637.
2. D. Matsuoka and M. Nakasako, J. Phys. Chem. B 2009, 113, 11274.
3. B. Schneider, H. M. Berman, Biophys. J. 1995, 69, 2661.

L16

LARGE-SCALE QUANTITATIVE ASSESSMENT OF BINDING PREFERENCES IN PROTEIN – NUCLEIC ACID COMPLEXES

D. Jakubec, J. Vondrášek

Institute of Organic Chemistry and Biochemistry AS CR, Flemingovo nám. 2, 166 10 Praha 6, Czech Republic
david.jakubec@uochb.cas.cz, jiri.vondrasek@uochb.cas.cz

Our work was focused on elucidating the selection rules governing the specificity of protein – DNA interactions. From a non-redundant set of protein – DNA complexes, we isolated all nucleotide – amino acid dimers. This provided a set of contacts mapping the positions occupied by each amino acid around each nucleotide thorough the entire Protein Data Bank. Considering the lack of distinguishing marks on the DNA sugar-phosphate backbone, we further reduced the problem of DNA sequence recognition to interactions between the DNA bases and amino acid side chains.

The distributions of each of the 20 standard amino acids around each DNA base served as a basis for our calculations. In these distributions, we recognised spatially defined clusters as areas where each particular amino acid was more prone to occur. Furthermore, for each cluster a single contact was selected as a representative based on statistical scoring.

The interaction energy of each cluster representative was calculated by a series of commonly used empirical force fields. These interaction energies, representing statistically significant contacts, were mapped against the energy distributions corresponding to the clusters they belonged to, as well as to the interaction energy profiles of the entire distributions. The validity of our results is supported by *ab initio* calculations performed on the same set of structures.

We were able to find that for certain DNA base – amino acid pairs, significantly stabilising interaction energies could be achieved only within a rather limiting set of mutual orientations of the interacting partners. An online repository providing access to the results in graphical form was established.