MAX PERUTZ LABS VIENNA

Faculty of Science
Charles University
Albertov 6, 128 00 Prague 2
Czech Republic

Vienna, August 31st 2020

Review of the PhD thesis "Modeling of Chemical Properties of Nano- and Biostructures" by Mgr. David Jakubec

The non-covalent binding between nucleic acids and proteins, and especially DNA and proteins, is one of the most fundamental interaction types defining the inner workings of biological systems at the molecular level. The PhD thesis by Mgr. David Jakubec presents the results of an extensive, multifaceted analysis of the fundamental aspects of DNA/protein interactions using computational approaches. Specifically, the candidate and his coworkers focus on understanding the physicochemical principles behind the molecular recognition in DNA-protein complexes, an extremely challenging and timely problem. The thesis is composed of multiple chapters providing an extensive summary and discussion of a number of already published articles on the topic.

In the introductory chapter, the candidate presents a relatively succinct, but expert overview of the fundamentals of the structural organization of proteins and nucleic acids together with a summary of the most widely used methods to study them. A major focus is given to DNA-protein interactions, the main topic of the present thesis. The heart of the thesis, however, is contained in its Chapters 1 and 2, and it gives an overview of the methods and the results of the already published manuscripts authored by the candidate. Specifically, Chapter 1 focuses on the candidate's studies dealing with the pairwise interactions between the basic building blocks of proteins and nucleic acids i.e. amino acids and nucleotides, and their influence on establishing specificity in protein-DNA interactions. The candidate and his coworkers provide multiple lines of evidence forging a link between the energetics of amino-acid/nucleotide interactions and the geometric interaction preferences as observed in high-resolution structures of protein-DNA complexes. In the process, they also provide a detailed comparative analysis of the computational methods used to evaluate the energetics of amino-acid/nucleotide interactions. Finally, they link the physicochemical properties of amino acids at the protein-DNA interfaces with the shape of DNA molecules involved. On the other hand, Chapter 2 focuses on molecular dynamics (MD) simulations of select DNA-protein complexes and critically addresses their ability to quantitatively address their binding equilibria using both equilibrium and non-equilibrium approaches. More methodological in spirit, this chapter provides an important set of benchmarks and pairwise comparisons and will likely serve as an important reference for future developments of MD force fields and algorithms.

Finally, the Discussion section is dedicated to a somewhat separate set of studies dealing with the evolutionary aspects of protein structure organization. Specifically, the candidate and his coworkers demonstrate that the evolution of individual protein domains in multidomain proteins exhibits in many cases an important degree of coupling, challenging a long-standing paradigm that protein structures evolve in a modular fashion. The Discussion section is capped off by a presentation of a computational tool (3DPatch) allowing the users to interactively explore the evolutionary conservation in protein structures. Parenthetically, while it is clear that the bulk of the thesis work, including the discussion of the results obtained, is presented in the already published articles, this reviewer was somewhat surprised by the choice of the candidate to dedicate the Discussion section belonging to the whole thesis to the presentation of additional articles and not a panoptic, overarching overview of the thesis work as a whole. This comment aside, I find the work presented in the Discussion section to be very stimulating and relevant.

With regards to the defense, there are several questions that the candidate should address in his presentation.

1. The questions of the impact of the solvent (water) on amino-acid/nucleotide interaction energies has been dealt with by the author in a separate publication (Jakubec et al. PLOS One, 2016). However, there the focus has primarily been on implicit solvation and the dielectric permittivity of water as the principal contributor. However, the water effect on the thermodynamics of amino-acid/nucleotide binding can also be elicited in other ways e.g. via the entropic means and, in this sense, may require an explicit representation to be adequately studied. The candidate should in the course of the defense comment more on the impact on solvent from this perspective and also link it with his own analysis of the thermodynamics of DNA-protein binding via MD simulations in explicit solvent.

2. In evolutionary processes, nucleic acids and proteins are perturbed via mutations and evolve at the level of primary sequences, while in most cases, the effects are manifested at the level of secondary and tertiary structure. Just this fact alone could suggest that there must be important, stable information contained in the interaction preferences of individual amino acids and nucleotides, which transmits itself to the level of secondary and tertiary structure in relatively robust, predicable ways. In other words, an important component of specificity in interaction between secondary and tertiary structures of biopolymers must be robustly encoded in the interaction preferences of individual building blocks. A major part of the candidate's work went into exploring this possibility, but it would be valuable to provide a strong summary statement along these lines. Specifically, the candidate should provide a succinct assessment of how much he thinks that when it comes to nucleic-acid/protein binding "the interaction code is in the sequence".

3. I very much agree with the findings presented in Jakubec et al., PLOS One, e0203085, 2018, that individual domains in multidomain proteins exhibit could exhibit coordinated evolutionary developments. Namely, I find it very reasonable that the functional constraints that brought two domains together, in addition to the known coupling when it comes to their stabilities, allosteric communication or folding pathways, must also reflect themselves when it comes to coupled evolution. This, however, refers primarily to the evolution of two domain sequences while they are connected into the same polypeptide. On the other hand, when one speaks of domain modularity, in my opinion, one refers more to the fact that a given domain can be (and has been) transplanted into different contexts without too much evolutionary difficulty/cost. This, however, the authors' shuffling experiment does not address directly as it deals with already existing domains and their sequences and cannot predict whether a transplantation of given domain into different contexts would soon and in an easy way, in an evolutionary sense, result in an acceptable multidomain protein. The candidate should comment on this distinction.

Overall, David Jakubec' thesis contains the results of an impressive, wide-reaching, technically mature and conceptually groundbreaking work. Importantly, most of the results of the thesis have already been published in respected journals in the field with the candidate being the first author in 6 of them. In my opinion, the frontiers opened up by the candidate, and especially the question of linking the specificity and structure of nucleic-acid/protein complexes with the intrinsic binding preferences of the individual nucleotides and amino acids, represents a so-far neglected, yet extremely important and fertile area of research, which will only grow over time. **Altogether, the doctoral thesis by David Jakubec represents a convincing and extensive body of work, which in both scope and quality satisfies the requirements for a PhD degree and I, therefore, recommend the candidate for a PhD defense. Finally, I grade the thesis itself with the best possible grade.**

FINAL GRADE: Excellent (1)

Sincerely,

Bojan Zagrovic, PhD

Professor of Molecular Biophysics

Department of Structural and Computational Biology
Max Perutz Labs, University of Vienna
Campus Vienna Biocenter 5, A-1030 Vienna, Austria
email: bojan.zagrovic@univie.ac.at
Tel: +43-1-4277-52271; Fax: +43-1-4277-9522