Validation of PDB models of potential drug design targets for SARS-CoV-2 coronavirus

Mariusz Jaskólski

Structural biology in general, and protein crystallography in particular, has played a paradigm changing role in the approach to drug development. Currently, modern drugs are developed using rational approaches based on support from experimental information about the structure of the targeted macromolecules, either by modeling of small molecules that would fit the hotspots of the targets, or by assembling such molecules through fragment screening. Historically, the first spectacular success of rational drug design, achieved in the situation of a looming global pandemic, was the development of retroviral protease inhibitors as drugs for the treatment of HIV infections. Since then, structural biology has more than once reacted promptly, furnishing reliable drug targets for emerging threats, especially those with potential global impact, such as SARS, MERS, Ebola or Zika.

It is extremely encouraging that in the current pandemic situation created by the outbreak of the new SARS-CoV-2 coronavirus (hereinafter CoV-2), many structural biologists and other molecular biology specialists have responded instantaneously, producing in quick succession a number of structural models of the proteins encoded by CoV-2 and making them immediately available as drug design targets through deposition in the Protein Data Bank (PDB), often without any embargo period and before publication.

However, haste in pursuit of scientific discovery, while fully justified in a global emergency situation, also has its negative consequences, such as mistakes and errors of different severity, with the ultimate danger of creating false or irreproducible results. Instead of helping the combined scientific mobilization, such errors have actually the opposite effect by diverting attention, manpower and resources in the wrong direction, confounding the subject, corrupting the faith in scientific credibility, and in the long run - undermining the public trust in science. There are many examples of such errors generated by haste, lack of supervision or proper validation of the results. A historically well known case is part of the race to discover the structure of retroviral protease, where a model with incorrectly folded C terminus thwarted proper understanding of the virus maturation process. It is therefore important that in any battle to quickly find a biostructural path to urgently needed therapy, there is a second line of research, where the first-line results are carefully verified and validated.

We and others have contributed a number of structure re-examination campaigns, usually devoted to concrete and timely biomedical challenges, tracking suboptimal models in the PDB and trying to minimize their devastating ripple effect in the literature. Whenever we found that a PDB structure could or should be improved or corrected, our priority has always been to re-deposit the revised models in the PDB jointly with the original authors.

In the present project I propose to carefully analyzed the PDB models of CoV-2 proteins and present the results in a dedicated, publicly available webserver with the aim of helping the biomedical community to establish a validated starting data base for COVID-19 drug development. In case a given model requires serious re-interpretation, the revised structure will be ultimately redeposited in the PDB, always after contacting the original authors and, if possible, getting access to the original diffraction data. Particularly close attention will be paid to the sensitive area of protein-ligand complexes, which potentially might contain drug lead compounds that can be used directly for COVID-19 drug optimization.