Structural bioinformatics

Map2mod—a server for evaluation of crystallographic models and their agreement with electron density maps

Olga Kirillova and Wladek Minor

Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA, USA

Received on February 1, 2006; revised on March 21, 2006; accepted on March 16, 2006
Advance Access publication April 27, 2006
Associate Editor: Anna Tramontano

ABSTRACT

Summary: Here we report on recent developments of the map2mod server. It has been designed for validation of protein models created by X-ray data interpretation. It can also be used during the refinement process since it is able to indicate problem regions in the model. Apart from evaluation of model quality, it has an option to remove atoms of side chains, which are not consistent with the maps as well as improperly placed water molecules. There are two additional options: checking the B-factors of atoms in the provided model and comparison of R and Rfree values obtained as the result of refinement with the averages characteristic for the data resolution shell.

Availability: The map2mod server is freely available for all user community (http://jurand.med.virginia.edu/m2m/map2model.html).

Contact: olga.wladek@iwonka.med.virginia.edu

1 INTRODUCTION

The rapid progress of structural genomics initiatives underscored the importance of highly accurate molecular structures for rational drug design and for structure-based functional studies. Indeed, it provides valuable help in the development of effective therapeutic agents and drugs (Glen and Allen, 2003). Building of protein structure models based on an electron density map interpretation makes crystallography different from other techniques directed at the same aim. Utilization of information from the experimental data provides higher accuracy of the model to be built by crystallographic methods in comparison with others.

The tool described here has been developed to complement both existing validation techniques; structure-based, implemented in, for example, ADIT (Yang et al., 2004), Whatchek (Hoofi et al., 1996, http://www.cmbi.kun.nl/swift/whatcheck/), Molprobity (Simon et al., 2003), PROCHECK (Laskowski et al., 1993) and statistical, implemented in, for example, SFCHECK (Vaguine et al., 1999) and OOPS (Kleywegt and Jones, 1996). We tried to develop a procedure applicable for assessment of both the local and global correctness of the model in a fast and automatic way. To check whether the model is in agreement with the interpreted data, we took the same way, which is usually taken by a crystallographer in the manual inspection of the model on a graphical display of manual model building programs like O (Jones et al., 1991), Coot (Emsley and Cowtan, 2004), XtalView (McRee, 1999), etc. But doing this automatically rather than manually allows to reduce the time of evaluation considerably. Selecting the first option of the server, which searches for atoms out of density, the user will get a list of outlier atoms within seconds.

In case the data are poor or are of low resolution, the density for side chains is not always clearly seen. Omitting parts of the model located in poor density regions eventually results in decreased phase error and allows to avoid model bias in the next round of the interpretation. Implemented in the second option of the server procedure are cut of side chain atoms, which are out of experimental density or within negative density. This hopefully can reduce the number of manual rebuilding steps necessary for improving the map by excluding the likely wrongly placed atoms from the phases calculation.

The last stage of model building usually consists in ligands and water placement. The third option of the map2mod tool provides the user with the ability to remove waters, which do not match the density in a faster and more effective way than when it is performed manually with the help of the graphical programs.

Another way to detect problematic regions of the structure consists of considering atoms with considerably higher (or lower) than average B-factor values (Radivojac et al., 2004). The fourth option of the tool is designed to search for atoms with outlying B-factor values. In the fifth option, the user can check whether the obtained R and Rfree values are greater or less than the averages characteristic for that resolution shell.

We suggest the use of the first two options at the beginning, especially if molecular replacement technique has been applied and the last three at the end just before deposition.

2 PROGRAM DESCRIPTION AND ORGANIZATION

Use of the server starts with filling out the form. The user has to upload the relevant coordinates and maps or structure factor files and (optionally) set up the parameters of his/her choice for the evaluation/reduction test. Then he/she has to choose the option for data processing. Although the interface is self-explanatory there is a manual on the server, which can be helpful if something is not entirely clear. After the data are uploaded and files formats checked, the main evaluation programs are executed.

In the first option, if maps are provided, the program reads one of the uploaded maps and the model file, and for each atom of the model it searches the points in the map located within a specified (in the interface) distance (we call it radius) from the atom under consideration with a density value above the specified cut-off. This procedure runs first with the 2Fo-Fc map, then with the Fo-Fc map with a positive cut-off and finally with the Fo-Fc map with a negative cut-off.
a negative cut-off. The program returns a list of atoms found to be out of the density (do not have neighboring map points) or having positive or negative Fo-Fc density nearby. If a structure factor file is uploaded, the maps are calculated on the server and then processed as described above.

In the second option, an atom is removed if it is out of 2Fo-Fc density or within negative difference density. In the current version only side chain atoms (except for C\(^{\alpha}\)'s) are removed. The program lists on the screen which atoms have been removed. If an atom, assigned to be deleted, breaks the connection with the rest of the side chain, the rest is removed as well. The modified file with the model can be downloaded from the server and/or sent by email to the user (if the email address is provided).

The third option is similar to the previous one except that it includes removal of only water molecules, which do not fit the density. The result can also be either downloaded from the server or sent by email.

On choosing the fourth option, the user has to provide a threshold value for B-factor either in sigma units or as an absolute value. Atoms of the uploaded model, which have B-factors above or below (depending on the selected options) the threshold are printed on the screen.

Comparison of \(R\) and \(R_{\text{free}}\) values obtained for the structure under evaluation with the average characteristic for the Protein Data Bank (Berman et al., 2000, http://rutsrs.csb.org/pdb) is performed in the fifth option. The user has to provide \(R\) and \(R_{\text{free}}\) values and the highest resolution used in the model building. If these data are present in the uploaded coordinate file, one can select the option to pick them up from the file automatically. The \(R\) and \(R_{\text{free}}\) can be present as percentages or in fractional format. For the fourth and fifth options, neither maps nor structure factors files are required.

The program is designed to work with CCP4 (CCP4, 1994) format files: the model has to be provided in the pdb format, a detailed description of which can be found on the CCP4 Program Documentation web page, the maps have to be in a binary format readable by CCP4 program suite, the structure factors file must be in mtz format.

To prepare the maps in the suitable format run the fht program (Ten Eyck, 1973) from the CCP4 suite or in case you use refmac5 (Murshudov et al., 1997) for refinement, the maps can be generated if in the refmac5 interface the option ‘generate weighted difference maps files’ is set to ON. Since the larger the file, the more time is spent for its uploading, we do not recommend extending the map to cover the model, as the extension is performed on the server.

If your original structure factor file is not in mtz format, the program f2mtz (Kjeldgaard, http://www.ccp4.ac.uk/dist/html/f2mtz.html), included in the ccp4 suite, can be helpful for the conversion.

For the first three options, the program has six parameters: three for setting the density cutoffs and three for radii. For computational processing a map is sampled on a grid. In other words, it is represented by a set of points, each with a distinct value of the density (\(p\)). The cut-off (\(p^*\)) is the value where grid points are taken into further consideration have \(p > p^*\) for the 2Fo-Fc and Fo-Fc maps with \(p^* > 0\) and when \(p < p^*\) for the Fo-Fc map with \(p^* < 0\). The radius is the maximum allowed distance between a map point and an atom of the model. The latter is covered by the map (is within the map), if it has at least one point of the map with density above the cut-off within the distance specified by the radius.

The default values can be used in general but experienced users can adjust the parameters to the data. That can be especially useful for low (~3 Å or lower) resolution, since in that case the maps are smoother (e.g. have a smaller difference between the minimum and the maximum of the density). In addition, taking into account that the map grid is usually calculated as the resolution divided by 3, in the low resolution case the grid spacing will be 1 Å or greater. For a 1 Å grid if an atom is settled equidistantly between grid points the distance to the nearest point is \(\sqrt{3}/2 \approx 0.866\). If the radius is set to less than this value for resolution of 3 Å or lower, important information can be lost because of a too sparse grid rather than low density in the region.

The algorithm is fairly efficient, as most of the time is spent uploading the data. The processing itself takes only a few seconds.

For the first three options, the evaluation result consists of three sets: all atoms of the model, which do not have neighbors in the 2Fo-Fc density; all atoms of the model, which have neighbors in the Fo-Fc positive density, and all atoms of the model, which have neighbors in the Fo-Fc negative density.

In the fourth option the average B-factor value and the deviation are calculated and printed on the screen followed by atoms, which have a B-factor greater or/and less than the threshold.

In the fifth option the output consists of the average \(R\) and \(R_{\text{free}}\) values for the provided resolution, calculated based on overall pdb statistics and the difference between these averages and the values characteristic for the structure under processing.

**ACKNOWLEDGEMENTS**

The authors would like to thank M. Zimmerman for useful comments on Perl scripting, M. Chruszcz for notes and advices during the testing of the server and anonymous referees for critical reading of the manuscript. The research has been supported by NIGMS grant P-50-GM62414.

**Conflict of Interest** none declared.

**REFERENCES**


