

Work Package 7, Milestone 7.3: The workflow and i/o of the Refmac web service defined

Introduction

Remote access to the REFMAC engine will be provided within the next-generation Web Service of the automated model-building suite ARP/wARP. Here we present basic concepts of the updated service and the description of the file formats used for handling input and output data.

Details

A number of widely used file formats exist to store experimental data in structural biology. Moreover, their content is not necessarily strictly defined, which makes them uneasy to deal with. Scientists at EMBL-HH have twenty years of experience in automated processing of such a data. Moreover, since 2004 they provide to tens of thousands users in the community a web-service for the model-building suite ARP/wARP, which includes REFMAC engine for reciprocal space refinement. Therefore, to assure a robust and reliable provision of the REFMAC web-service for cryo-EM models refinement it will be built based on the next-generation ARP/wARP web infrastructure.

The data processing will commence with parsing and validation of the user-provided density map, model, sequence, and optionally restraints. The latter can include NMR restraints such as residual dipolar couplings and pseudocontact shifts. The details of the data container (individual files or compressed archives) will be defined at a later stage of the project. The web-service will accept maps in the MRC and CCP4 formats that are commonly used by electron microscopists. NMR restraints are provide as text lists of the measured quantities for each bond vector or atom. The input model should be provided in the PDB or mmCIF format. The use of the latter is essential owing the large sizes of macromolecules studied by cryo-EM. Finally, sequence information that enables model validation and rebuilding (if required) will be accepted in the standard FASTA/PIR formats. The web-service may also accept additional restraints (e.g base-pairing in nucleic acids) as text that may be crucial for refining structures against low-resolution maps.

The workflow diagram presents the main steps of the web-service for the integration of cryo-EM. At first, the input data is validated for consistency and any issues are fixed if appropriate and reported to the user. Next, the map and the coordinates are converted to a minimum pseudo-crystallographic asymmetric unit-cell that in some cases may substantially reduce computation time. The map in a new coordinate frame is then used to calculate structure factors in an MTZ format, which is required by

Milestone M7.3



REFMAC. Finally, a REFMAC-based reciprocal-space refinement or a combined iterative ARP/wARP-REFMAC model building is performed depending on the user choice and the characteristics of the data. At the last step, refined model is back-transformed to the original reference frame and provided to the user in PDB and mmCIF file formats.

At variance with cryo-EM, the use of NMR restraints does not require substantial deviations from the standard REFMAC workflow, except for the adoption of a specific library with a new set of hydrogen atoms.

About ARP/wARP

The current version of the ARP/wARP suite and the corresponding web-services are available at

<http://www.embl-hamburg.de/ARP/>

or

<http://www.arp-warp.org/>

