Role of FFT grid in paired refinement

Martin Malý[1,2], Kay Diederichs[3] Jan Dohnálek[2], Leona Švecová[1,2], Petr Kolenko[1,2]

[1]Department of Solid State Engineering, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Czech Republic
[2]Institute of Biotechnology AS CR, BIOCEV, Vestec, Czech Republic
[3]University of Konstanz, Germany
martin.maly@fjfi.cvut.cz

Abstract

Paired refinement is the current standard for high-resolution cutoff determination in macromolecular crystallography. As the protocol depends on several parameters, this work focuses on the impact of a choice of an FFT-grid spacing in the structure refinement. Five test data sets were analysed. A slight influence of the investigated parameter on results was shown, however, no general trend could be observed. Therefore, to enable such analysis for the users, program *PAIREF* supports the option to run the protocol with a fixed grid over the whole analysed resolution range.

Keywords: Macromolecular crystallography; X-ray structure analysis; Diffraction limit; Paired refinement; Fourier transform

Introduction

The fundamental mathematical operation in the physical description of the diffraction phenomenon on macromolecular crystals is the Fourier transform (FT). After the experiment, amplitudes of *structure factor* F_{obs}(hkl) are determined from the diffraction data and after the solution of their phases, *electron density* is calculated using the inverse FT. Subsequently, a structure model can be built into the electron density map and refined.

Especially during the model refinement, agreement between the model and measured data is quantified – generally with the R-value: $R = \frac{\sum ||F_{\mathrm{obs}}| - |F_{\mathrm{calc}}||}{\sum |F_{\mathrm{obs}}|}$. Hence, this formula requires knowledge of F_{calc}(hkl), the Fourier coefficients based on a structure model. They are most commonly calculated using the fast Fourier transform algorithm (FFT) from an approximation of the electron density equivalent to the model. To perform this, the electron density must be sampled at grid points; the distance between such points depend on the highest resolution d_{min} and usually equals to $d_{\mathrm{min}}/3$ [1].

As the signal-to-noise ratio decreases with the increasing diffraction angle, a high-resolution cutoff is usually applied to discard the noisy data. The current optimal approach for the estimation of the highest resolution d_{min} is the *paired refinement* protocol [2]: Initially, an conservative cutoff is chosen during data processing. While a structure model is refined, formerly discarded data from higher resolution shells are added by the step-by-step process to the refinement. Then statistics (usually R-values) relating to the original model and the model refined at higher resolution can be compared. For validity of the results, the statistics must be calculated using the same data.

Recently, we released program *PAIREF* providing automation of the procedure [3]. In this work, we investigate whether the choice of the spacing of the FFT grid can have a significant influence on results from paired refinement.

Materials and methods

For our analysis of the impact of the FFT grid, the following five data sets were selected: simulated data for hen egg-white lysozyme (SIM, PDB entry 1H87) [4] and experimental data for complex of cysteine dioxygenase from *R. norvegicus* (CDO, PDB entry 3ELN) [2], complex of endothiapepsin from *C. parasitica* (EP, PDB entry 4Y4G) [5], interferon gamma from *P. olivaceus* (POLI, PDB entry 6F1E) [6] and bilirubin oxidase from *M. verrucaria* (BO, PDB entry 6I3J) [7]. These data were previously reprocessed and used for demonstration of the standard paired refinement protocol using the program *PAIREF* [3]. During the paired refinement of the structure models, different FFT-grid spacing was used in each resolution step in the former work [3]; the higher the resolution, the finer the grid was applied.

Hence, to get rid of the influence of this parameter, analogous paired refinement calculations with program *PAIREF* were carried out while keeping the FFT grid constant. This was achieved with the keyword `SHANNON FACTOR` of *REFMAC*5 [8] that was set to the value of $1.5 \times d_{\text{current}}/d_{\text{min}}$ where d_{current} denotes the current examined high-resolution limit during paired refinement. Thus, the finest grid relating to the highest resolution was used for all the refinement runs in each of the five cases. Other refinement parameters were left unchanged. *REFMAC*5 was used in version 5.8.0258.

Results and discussion

Comparison of results from paired refinement using a different (already published [3]) and a constant FFT-grid spacing is shown in Fig. 1. All the results are available from http://doi.org/10.5281/zenodo.4159436.

For SIM and EP, the impact of an FFT grid on trends of R-values during paired refinement is negligible (Fig. 1ac). Moreover for SIM, differences in coordinates compared to the original structure model (*i.e.* the base of the data set generation) are nearly identical (Fig. 1b). However, ADPs are little closer to the original model while using different grids but their trend is similar.

Differences in results are observed for the structures solved in medium-to-lower resolution: POLI (Fig. 1d) and BO (Fig. 1e). Diffraction data from POLI exhibited severe anisotropy. Whereas the strictly increasing trend of R_{work}-values was observed for BO while using different grids, the situation is just the opposite for the fixed grid. Besides, the magnitudes of differences in R_{free} became larger. Despite this disagreement, the decision on high-resolution cutoff would remain the same as in the previous work [3]: R_{free} decreases or stays constant until 2.0 Å for POLI; R_{free} increases when any resolution shell above 2.59 Å is involved in refinement.

In contrary for CDO, although the influence of grid setting on R-values is rather small (Fig. 1f), it could affect the cutoff choice. After the addition of shell 1.60–1.50 Å, R_{free} increases slightly while using different grids but a significant drop is observed while using a constant grid.

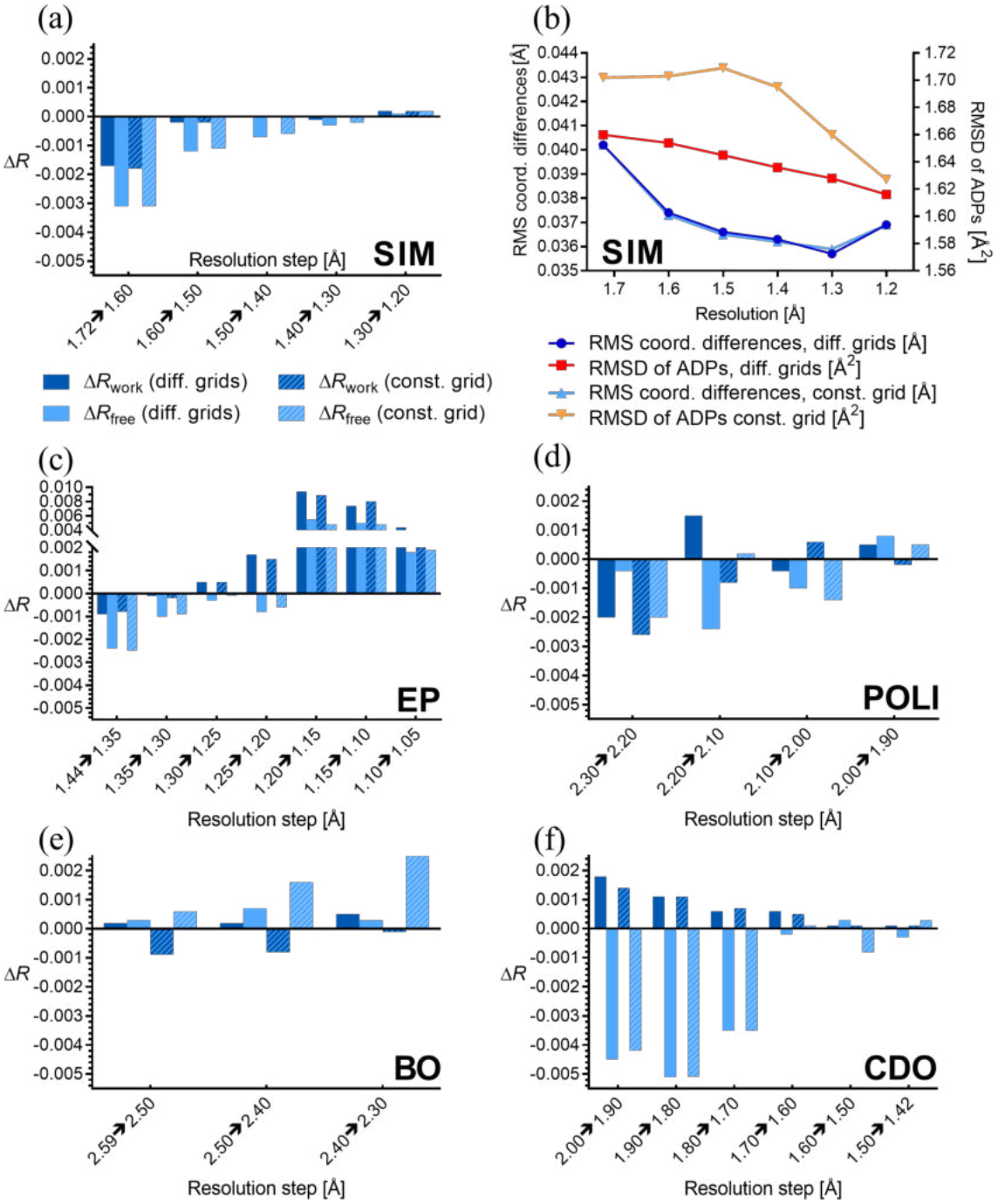

Figure 1: Comparison of results from paired refinement using different FFT grids and a constant grid for five test cases. (a,c–f) Differences in the overall R-values during paired refinement for SIM (a), EP (c), POLI (d), BO (e) and CDO (f). For each incremental step of resolution for X→Y, the R-values were calculated at resolution X. (b) SIM: Root mean squared differences (RMSD) in coordinates and ADPs against the original structure.

Conclusion

Only a marginal impact of the FFT-grid spacing in *REFMAC*5 on the choice of a high-resolution cutoff in paired refinement was observed in the five selected cases. However, both approaches can possibly lead to different results, different R-values. Therefore, both approaches have been implemented to program *PAIREF*: with the usage of different FFT grids that vary with resolution (default option) and a constant FFT grid for all refinement runs using the `--constant-grid` option.

References

[1] J. Drenth, M. Jeroen. Refinement of the Model Structure. In *Principles of Protein X-ray Crystallography, 3rd ed.*, chap. 13. Springer, New York, 2010.

[2] P. A. Karplus, K. Diederichs. Linking Crystallographic Model and Data Quality. *Science* **336**: 1030–1033, 2012.

[3] M. Malý, K. Diederichs, J. Dohnálek, P. Kolenko. Paired refinement under the control of *PAIREF*. *IUCrJ* **7**: 681–692, 2020.

[4] J. M. Holton, S. Classen, K. A. Frankel, J. A. Tainer. The R-factor gap in macromolecular crystallography: an untapped potential for insights on accurate structures. *FEBS J* **281**: 4046–4060, 2014.

[5] F. Huschmann, J. Linnik, K. Röwer, M. Ühlein, X. Wang, A. Metz, J. Schiebel, A. Heine, G. Klebe, M. Weiss, U. Mueller. Structures of endothiapepsin–fragment complexes from crystallographic fragment screening using a novel, diverse and affordable 96-compound fragment library. *Acta Cryst. D***72**: 346–355, 2016.

[6] J. Zahradník, L. Kolářová, H. Pařízková, P. Kolenko, B. Schneider. Interferons type II and their receptors R1 and R2 in fish species: Evolution, structure, and function. *Fish Shellfish Immunol.* **79**: 140–152, 2018.

[7] T. Koval', L. Švecová, L. H. Østergaard, T. Skalova, J. Dušková, J. Hašek, P. Kolenko, K. Fejfarová, J. Stránský, M. Trundová, J. Dohnálek. *Sci. Rep.* **9**: 13700, 2019.

[8] G. N. Murshudov, A. A. Vagin, E. J. Dodson. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst. D***53**:240–255, 1997.

Acknowledgement

We thank Jan Wollenhaupt and Manfred S. Weiss for providing the EP data set. This work was supported by the MEYS CR – projects CAAS (grant No. CZ.02.1.01/0.0/0.0/16_019/0000778 to the to Faculty of Nuclear Sciences and Physical Engineering, CTU in Prague) and BIOCEV (grant No. CZ.1.05/1.1.00/02.0109 to the IBT AS CR), from the ERDF fund; the Grant Agency of the CTU in Prague (grant No. SGS19/189/OHK4/3T/14), and the institutional support of IBT AS CR (RVO: 86652036).