Paired refinement: Impact of reintegration and rescaling

Martin Malý[1,2], Tereza Skálová[2], Leona Švecová[1,2], Jan Dohnálek[2], Jan Bláha[3], Ondřej Vaněk[3], Karl Harlos[4], 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 CAS, v.v.i., BIOCEV, Vestec, Czech Republic
[3] Department of Biochemistry, Faculty of Science, Charles University, Czech Republic
[4] Division of Structural Biology, Wellcome Centre for Human Genetics, University of Oxford, United Kingdom
malymar9@fjfi.cvut.cz

Abstract

Paired refinement is a procedure which analyses the impact of data from high resolution on the quality of a refined structure model. Subsequent R-value analysis leads to the determination of the optimal high resolution cutoff. In our particular case of data from human receptor NKR-P1 extracellular domain, we could observe differences in the paired refinement results in the case when the data were reintegrated separately.

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

Introduction

X-ray diffraction data are usually weak in their high resolution shells – the intensity to noise ratio decreases with the increasing angular distance of a reflection from the beam position. Therefore, the resolution cutoff is determined from indicators of diffraction data quality and applied during data processing. Ideally, the data should be cut off at such resolution that the next shell of data does not contain any useful signal which could improve the quality of the structure model [1]. However, the appropriateness of the diffraction limit choice should be checked at the end of the structure refinement using so called paired refinement procedure [2] in order not to discard useful data.

In our study, we applied three different approaches of diffraction data processing and scaling to analyze their influence on the consequent paired refinement results. The calculations were executed using the diffraction data from a crystal of the human receptor NKR-P1 extracellular domain [3] (PDB ID 5MGR [4]).

Material and methods

Diffraction data of the structure (PDB ID 5MGR [4]) were collected on beamline I03, Diamond Light Source, Oxfordshire (Table 1). The deposited structure [4] was solved at 1.8 Å resolution and contains 2,447 non-H atoms (including 225 solvent atoms), the values of R_{work}, R_{free}, and R_{all} are 16.7%, 20.2%, and 16.8%, respectively.

Subsequently, paired refinement was performed to analyze the high resolution cutoffs of 1.7 Å, 1.6 Å, and 1.5 Å. The procedure was executed with diffraction data processed by three different approaches using the same set of free reflection labels:

1. *No reintegration, no rescaling*: Data were reprocessed in XDS [5] and rescaled in AIMLESS [6] only once up to 1.5 Å (Table 2). The setting of the diffraction limits (1.7 Å, 1.6 Å, and 1.5 Å) was controlled by the parameter of the refinement software REFMAC5 [7].
2. *No reintegration, rescaling only*: Data were reprocessed up to resolution 1.5 Å and then separately rescaled using a diffraction limit of 1.7 Å, 1.6 Å, and 1.5 Å.
3. *Reintegration and rescaling*: Data were separately both reprocessed and rescaled using a diffraction limit of 1.7 Å, 1.6 Å, and 1.5 Å.

Table 1: Diffraction experiment – data collection parameters (structure PDB ID 5MGR [4]).

Detector	DECTRIS PILATUS3 6M
Wavelength [Å]	0.97625
Temperature [K]	100
Distance crystal-detector [mm]	340
Exposure time per frame [s]	0.02
Oscillation angle/range [°]	0.1 / 720

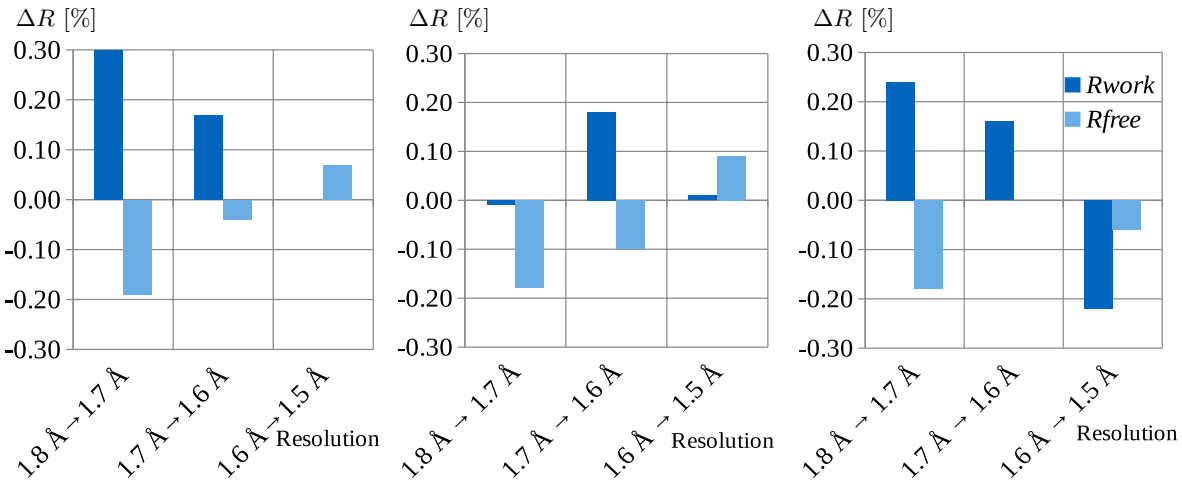
Results and discussion

Addition of data from the resolution shells 1.8–1.7 Å and 1.7–1.6 Å led to an improvement of the structure model in all three approaches as the R_{free}-values did not increase (Figure 1). The R-values changes related to the different approaches correspond to each other qualitatively, not quantitatively.

The use of data from the shell 1.6–1.5 Å caused an increase of the R_{free}-values in the cases without reintegration though it was possible to observe a decrease of this value in the case of the *"rescaling and reintegration"* approach. However, this drop was quite small and the data from the shell 1.6–1.5 Å contain relatively low signal compared to noise as can be seen from the values of the data statistics (*e.g.* $CC_{1/2} = 0.11$). Thus, usability of the data from this shell is very questionable. Therefore, we suggest to apply the high resolution cutoff at 1.6 Å resolution.

Table 2: Data collection statistics (structure PDB ID 5MGR [4]).

Space group	$P3_121$				
Unit-cell parameters [Å, °]	$a = b = 68.24,\ c = 127.19,\ \alpha = \beta = 90,\ \gamma = 120$				
Resolution range [Å]	50.0–1.5	1.9–1.8	1.8–1.7	1.7–1.6	1.6–1.5
No. observed reflections	1,737,128	194,115	206,086	165,775	82,156
No. unique reflections	103,336	9,272	11,636	14,586	15,245
Redundancy	17	21	18	11	5
Completeness [%]	96.5	100	100	99.6	80.7
Mean $I/\sigma(I)$	20.9	5.8	2.9	1.1	0.4
R_{meas}	0.06	0.55	0.97	1.82	3.37
$CC_{1/2}$	1.00	0.97	0.87	0.51	0.11



(a) *No reintegration, no rescaling.* (b) *No reintegration, rescaling only.* (c) *Reintegration and rescaling.*

Figure 1: R-factor analysis of the paired refinement results. The pair of bars for each step of resolution from $A \rightarrow B$ represents changes of R_{work} and R_{free} for a model refined using data at resolution B. Both R-values were calculated at resolution A in order to be comparable.

Conclusion

High resolution cutoff is an important decision which has a significant impact on the quality of structures. In our particular case, we performed the paired refinement procedure using data processed by three different approaches to study the impact of reintegration and rescaling of diffraction data on the selection of the resolution cutoff.

Involving the rescaling step in the procedure did not cause any remarkable changes in the paired refinement results. However, a significant difference in the R-values behaviour for the model refined against the data covering the highest resolution shell was observed when the reintegration was included in the procedure. Rescaling probably can be incorporated in the paired refinement procedure. Inclusion of reintegration at the moment cannot be recommended before the observed effects are fully understood. To their explanation, a further analysis covering more data sets is necessary.

References

[1] P. R. Evans. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Cryst. D***67**: 282–292, 2011.

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

[3] J. Bláha *et al. Protein Expr. and Purif.* **140**: 36–43, 2017.

[4] T. Skálová *et al.* PDB ID 5MGR. Human receptor NKR-P1 in glycosylated form, extracellular domain. [2018-08-30], `https://www.rcsb.org/structure/5mgr`.

[5] W. Kabsch. XDS. *Acta Cryst. D***66**: 125–132, 2010.

[6] P. R. Evans, G. N. Murshudov. How good are my data and what is the resolution? *Acta Cryst. D***69**: 1204–1214, 2013.

[7] 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

This work was supported by the Czech science foundation (project 18-10687S), the Grant Agency of the CTU in Prague (grant No. SGS16/246/OHK4/3T/14), and the institutional support of IBT CAS (RVO: 86652036).