PAIREF: Introduction to paired refinement

Petr Kolenko[1,2]

[1] Czech Technical University in Prague
[2] Institute of Biotechnology AS CR, v.v.i.

23rd March, 2021

Content

1. What is my resolution?
2. Paired refinement with *PAIREF*
3. What about the influence on the electron density map?
4. Tutorial
5. Discussion & things to be done in PAIREF

Resolution: What does it mean and what is my resolution?

6GG1

Structure of PROSS-edited human interleukin 24

Kolenko, P., Zahradnik, J., Kolarova, L., Schneider, B.

(2019) FEBS J **286**: 3858-3873

Released	2019-05-15
Method	X-RAY DIFFRACTION 1.3 Å
Organisms	Homo sapiens
Macromolecule	Interleukin-24 (protein)
Unique Ligands	NI, SO4

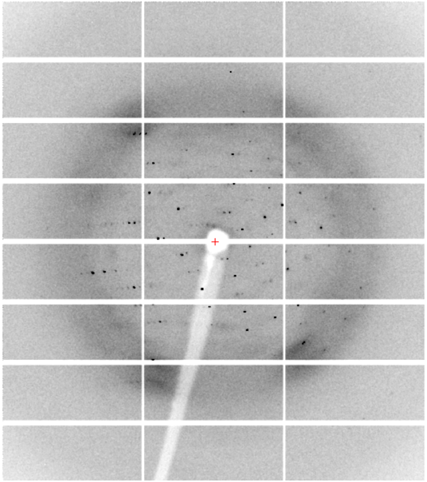
www.rcsb.org

Resolution (Online Dictionary of Crystallography)

is defined as the minimum plane spacing given by Bragg's law for a particular set of X-ray diffraction intensities *used in structure refinement*.

`https://dictionary.iucr.org/Resolution`

Why do we have to limit the resolution? (make the cut off?)



Aim of the data processing:

- Include as much experimental information as possible.
- Avoid noisy or even bad data.

Cut in resolution is important!

How to make the right decision?

Data quality indicators:

$$R_{\mathrm{merge}} = \frac{\sum_{\mathbf{h}} \sum_i |I_{\mathbf{h}i} - <I_{\mathbf{h}}>|}{\sum_{\mathbf{h}} \sum_i <I_{\mathbf{h}}>}$$

$$R_{\mathrm{meas}} = R_{\mathrm{r.i.m.}} = \frac{\sum_{\mathbf{h}} \sum_i \sqrt{\frac{n_{\mathbf{h}}}{n_{\mathbf{h}}-1}} |I_{\mathbf{h}i} - <I_{\mathbf{h}}>|}{\sum_{\mathbf{h}} \sum_i <I_{\mathbf{h}}>}$$

$$\mathrm{CC}_{1/2} = \frac{\sum_{m=1}^{N/2} (I_m - \bar{I})(J_m - \bar{J})}{\sqrt{\sum_{m=1}^{N/2} (I_m - \bar{I})^2} \sqrt{\sum_{m=1}^{N/2} (J_m - \bar{J})}}$$

Further indicators: $< I/\sigma >$, $R_{\mathrm{p.i.m.}}$, completeness, ...

How to make the right decision?

Data quality indicators:

$$R_{\mathrm{merge}} = \frac{\sum_{\mathbf{h}} \sum_i |I_{\mathbf{h}i} - <I_{\mathbf{h}}>|}{\sum_{\mathbf{h}} \sum_i <I_{\mathbf{h}}>}$$

$$R_{\mathrm{meas}} = R_{\mathrm{r.i.m.}} = \frac{\sum_{\mathbf{h}} \sum_i \sqrt{\frac{n_{\mathbf{h}}}{n_{\mathbf{h}}-1}} |I_{\mathbf{h}i} - <I_{\mathbf{h}}>|}{\sum_{\mathbf{h}} \sum_i <I_{\mathbf{h}}>}$$

$$\mathrm{CC}_{1/2} = \frac{\sum_{m=1}^{N/2} (I_m - \bar{I})(J_m - \bar{J})}{\sqrt{\sum_{m=1}^{N/2} (I_m - \bar{I})^2} \sqrt{\sum_{m=1}^{N/2} (J_m - \bar{J})}}$$

Further indicators: $< I/\sigma >$, $R_{\mathrm{p.i.m.}}$, completeness, . . .

Criteria used in past:

- $< I/\sigma >_{\mathrm{HighRes}}$ higher than 2
- R_{merge} lower than 0.15
- $R_{\mathrm{merge,HighRes}}$ lower than 0.6
- R_{meas} lower than 0.15
- $\mathrm{CC}_{1/2,\mathrm{HighRes}}$ higher than 0.3
- . . .

How to make the right decision?

Data quality indicators:

$$R_{\mathrm{merge}} = \frac{\sum_{\mathbf{h}} \sum_i |I_{\mathbf{h}i} - <I_{\mathbf{h}}>|}{\sum_{\mathbf{h}} \sum_i <I_{\mathbf{h}}>}$$

$$R_{\mathrm{meas}} = R_{\mathrm{r.i.m.}} = \frac{\sum_{\mathbf{h}} \sum_i \sqrt{\frac{n_{\mathbf{h}}}{n_{\mathbf{h}}-1}} |I_{\mathbf{h}i} - <I_{\mathbf{h}}>|}{\sum_{\mathbf{h}} \sum_i <I_{\mathbf{h}}>}$$

$$\mathrm{CC}_{1/2} = \frac{\sum_{m=1}^{N/2} (I_m - \bar{I})(J_m - \bar{J})}{\sqrt{\sum_{m=1}^{N/2} (I_m - \bar{I})^2} \sqrt{\sum_{m=1}^{N/2} (J_m - \bar{J})}}$$

Further indicators: $< I/\sigma >$, $R_{\mathrm{p.i.m.}}$, completeness, . . .

Criteria used in past:

- $< I/\sigma >_{\mathrm{HighRes}}$ higher than 2
- R_{merge} lower than 0.15
- $R_{\mathrm{merge,HighRes}}$ lower than 0.6
- R_{meas} lower than 0.15
- $\mathrm{CC}_{1/2,\mathrm{HighRes}}$ higher than 0.3
- . . .

Is application of these critera useful?

How to make the right decision?

Data quality indicators:

$$R_{\mathrm{merge}} = \frac{\sum_{\mathbf{h}} \sum_i |I_{\mathbf{h}i} - <I_{\mathbf{h}}>|}{\sum_{\mathbf{h}} \sum_i <I_{\mathbf{h}}>}$$

$$R_{\mathrm{meas}} = R_{\mathrm{r.i.m.}} = \frac{\sum_{\mathbf{h}} \sum_i \sqrt{\frac{n_{\mathbf{h}}}{n_{\mathbf{h}}-1}} |I_{\mathbf{h}i} - <I_{\mathbf{h}}>|}{\sum_{\mathbf{h}} \sum_i <I_{\mathbf{h}}>}$$

$$\mathrm{CC}_{1/2} = \frac{\sum_{m=1}^{N/2} (I_m - \bar{I})(J_m - \bar{J})}{\sqrt{\sum_{m=1}^{N/2} (I_m - \bar{I})^2} \sqrt{\sum_{m=1}^{N/2} (J_m - \bar{J})}}$$

Further indicators: $< I/\sigma >$, $R_{\mathrm{p.i.m.}}$, completeness, . . .

Criteria used in past:

- $< I/\sigma >_{\mathrm{HighRes}}$ higher than 2
- R_{merge} lower than 0.15
- $R_{\mathrm{merge,HighRes}}$ lower than 0.6
- R_{meas} lower than 0.15
- $\mathrm{CC}_{1/2,\mathrm{HighRes}}$ higher than 0.3
- . . .

Is application of these critera useful?

Yes, but only as a starting point.

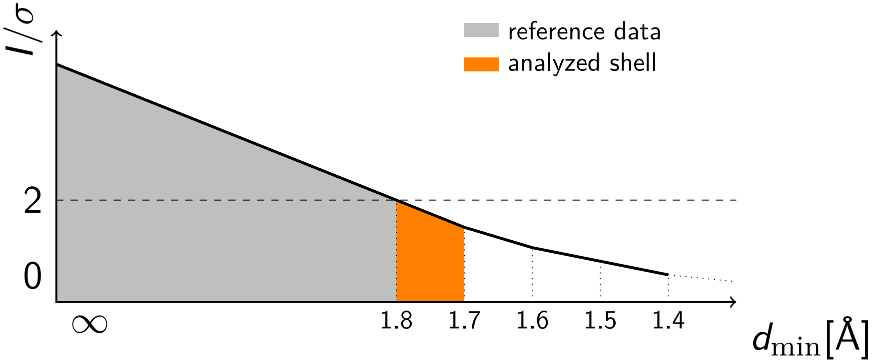
Paired refinement - the (current) optimal way

Paired refinement requires reference data!

Paired refinement - the (current) optimal way

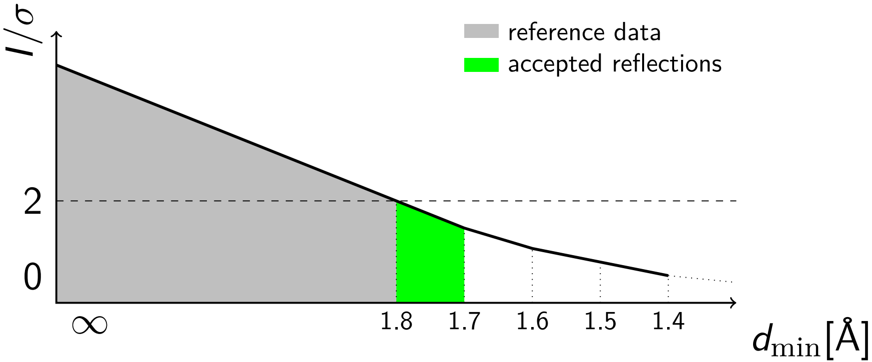
Paired refinement requires reference data!

Does addition of new reflections improve my model?



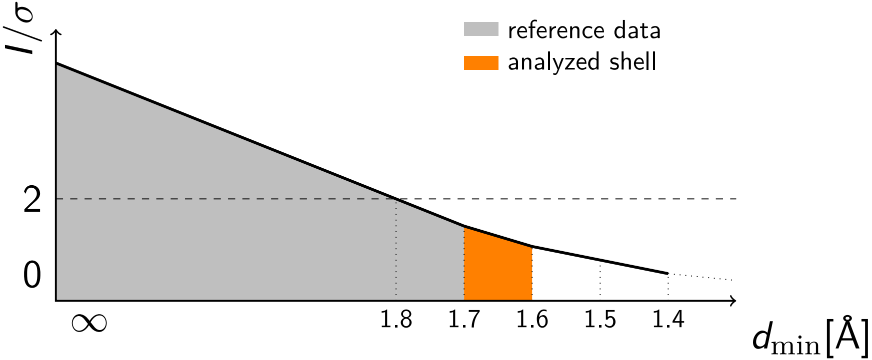
Paired refinement - the (current) optimal way

Does addition of new reflections improve my model? **Yes!**



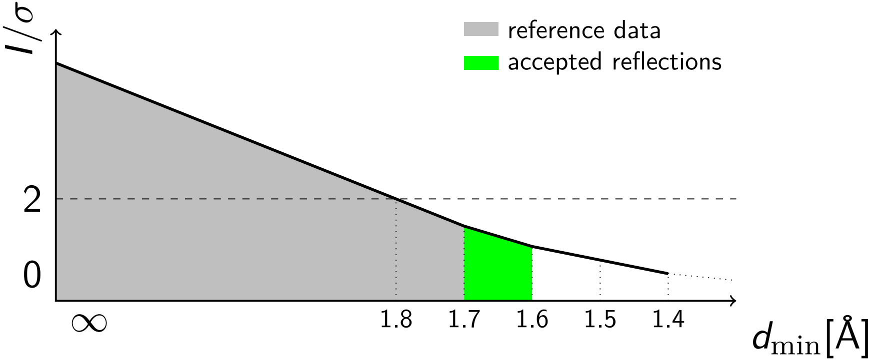
Paired refinement - the (current) optimal way

Does addition of new reflections improve my model?



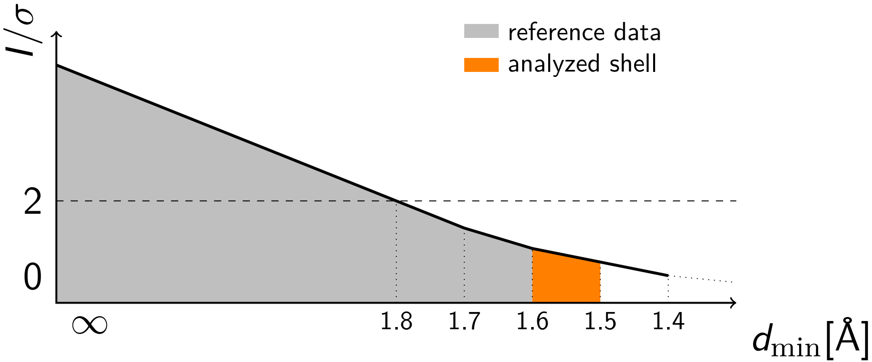
Paired refinement - the (current) optimal way

Does addition of new reflections improve my model? **Yes!**



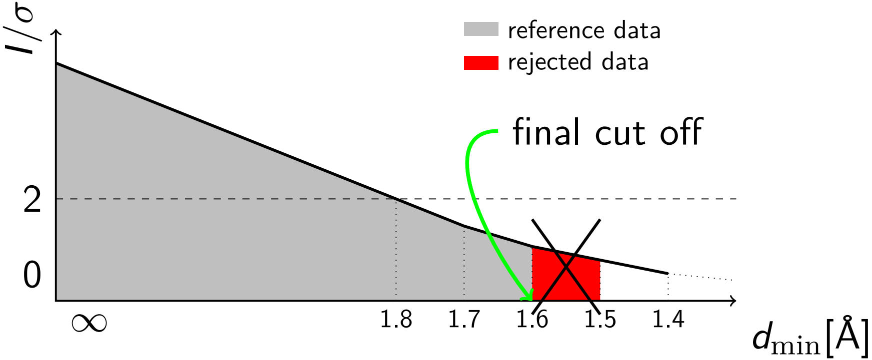
Paired refinement - the (current) optimal way

Does addition of new reflections improve my model?



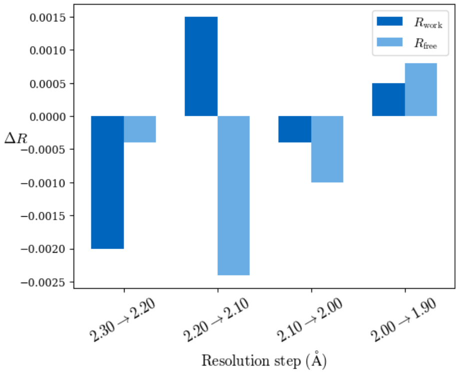
Paired refinement - the (current) optimal way

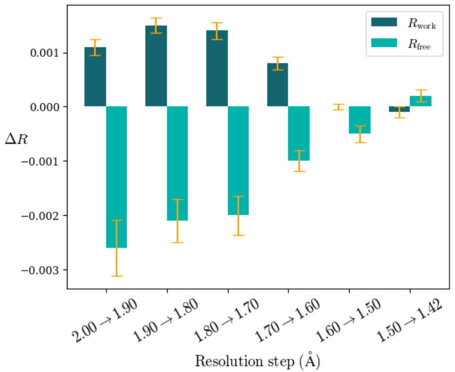
Does addition of new reflections improve my model? **No!**



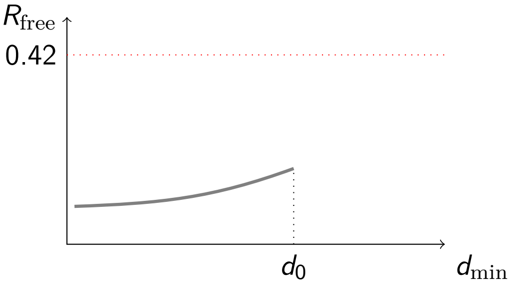
How do I know that the model was improved?

The R-values must be calculated always using the reflections from the reference data!

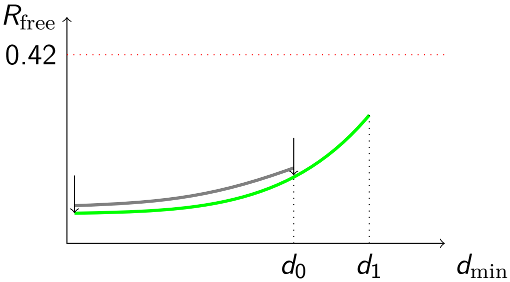

Differences of overall R-values


Differences of overall R-values averaged over free sets

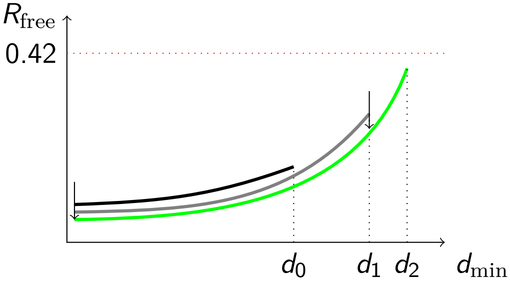
R-values in resolution bins (monitor FREE reflections!)



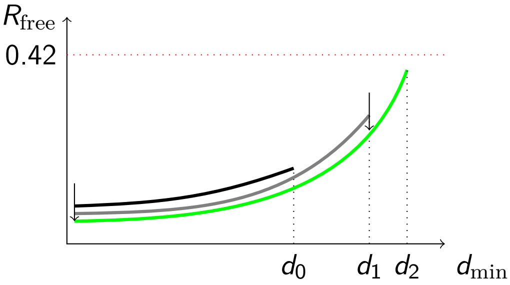
R-values in resolution bins (monitor FREE reflections!)

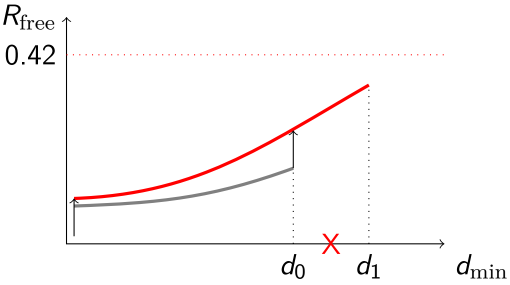


R-values in resolution bins (monitor FREE reflections!)

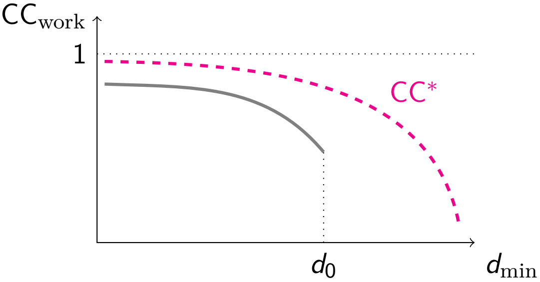


R-values in resolution bins (monitor FREE reflections!)

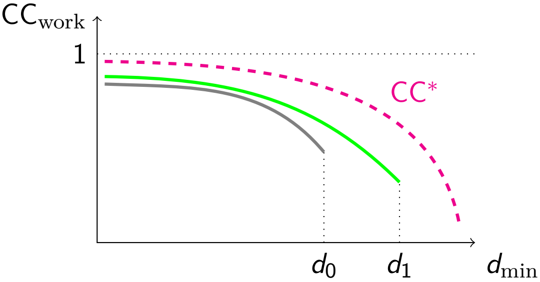




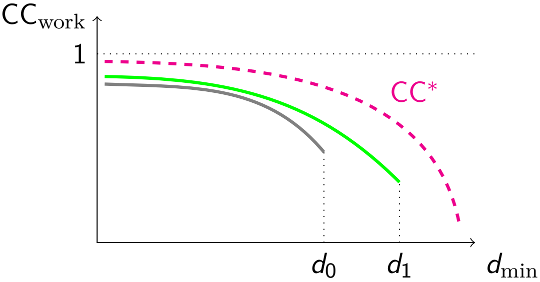
CC_{work} in resolution bins (monitor WORK reflections!)

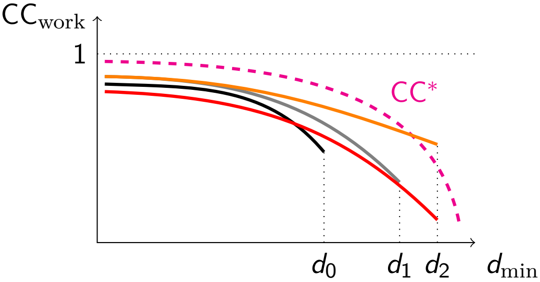


CC_{work} in resolution bins (monitor WORK reflections!)

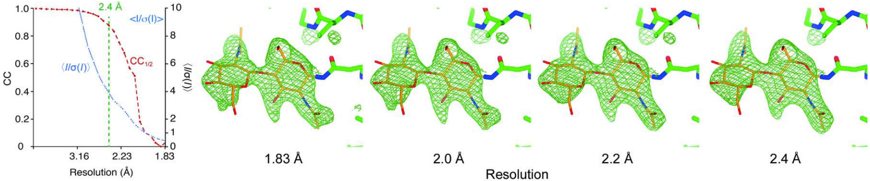


$\mathrm{CC_{work}}$ in resolution bins (monitor WORK reflections!)

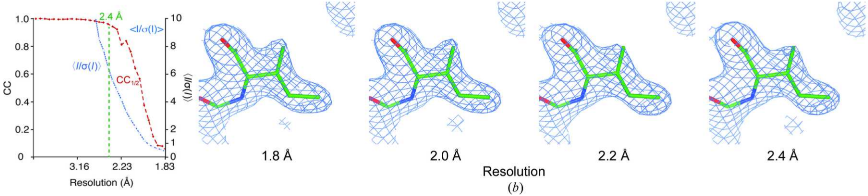




Are there any changes in the electron density?



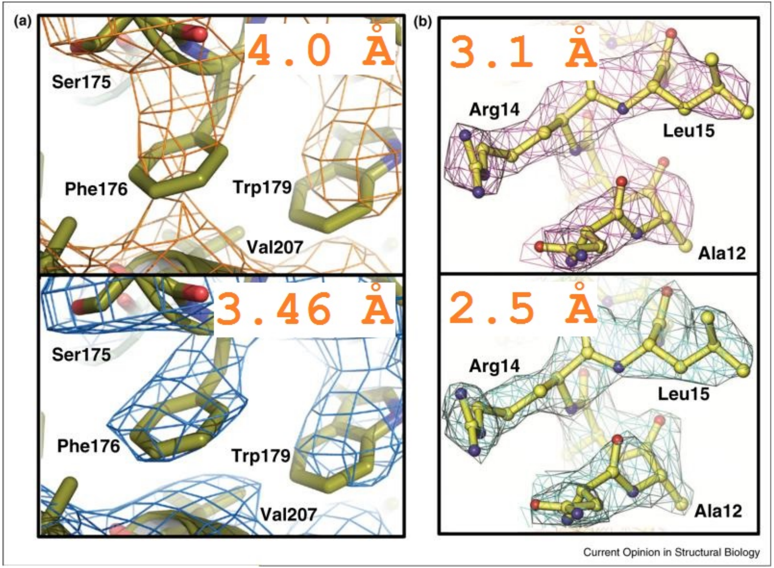
Resolution
(*a*)



Resolution
(*b*)

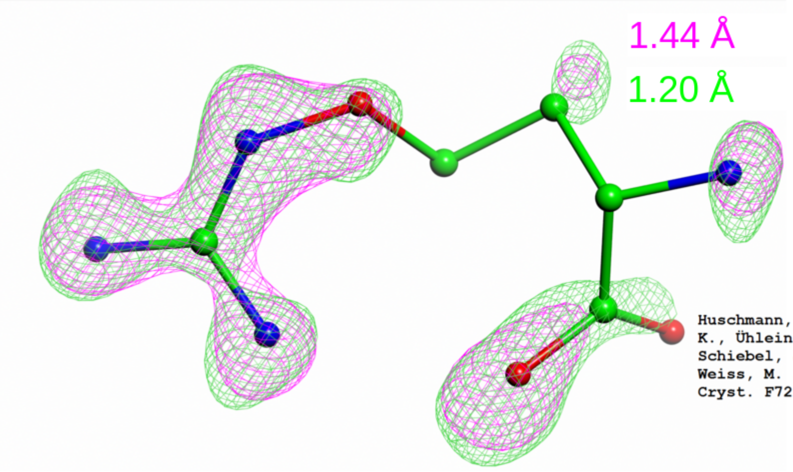
P.R. Evans, G.N. Murshudov. (2013). How good are my data and what is the resolution? *Acta Cryst D***69**, 1204-1214.

Are there any changes in the electron density?



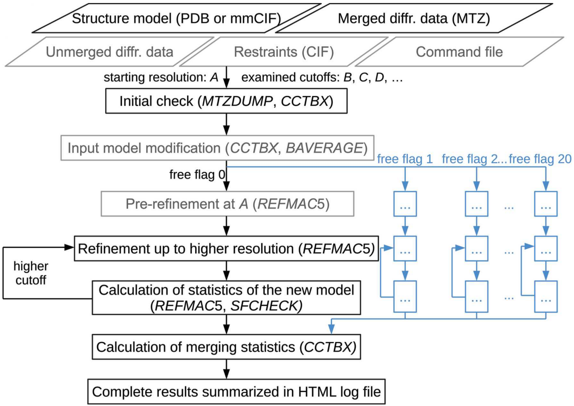
Karplus, P. A. & Diederichs, K. (2015). Curr. Opin. Struct. Biol. 34, 60–68.

Is there anything practical?



Huschmann, F. U., Linnik, J., Sparta, K., Ühlein, M., Wang, X., Metz, A., Schiebel, J., Heine, A., Klebe, G., Weiss, M. S. & Mueller, U. (2016). Acta Cryst. F72, 346–355.

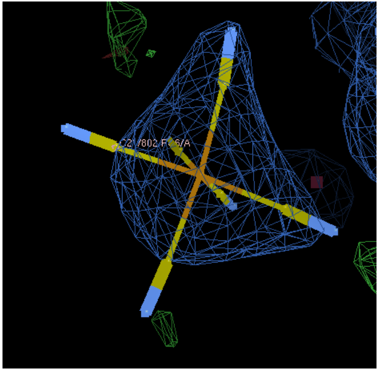
Algorithm of *PAIREF*



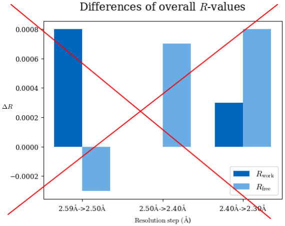
Strength of PAIREF

- For both CCP4 and PHENIX users.
- Number of plots and tables for purpose of complex decision making.
- Customizable selection of shells.
- Fully customizable refinement scenario.
- Work with ligands.
- **Complete cross-validation procedure**.
- Refinement stability control.

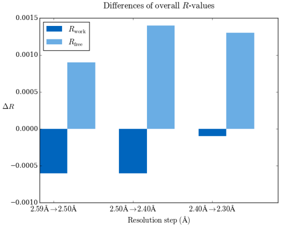
What PAIREF (currently) cannot do for you? - **Model validation!**



Application of external harmonics was crucial for correct geometry of the refinement output.



(without)



(with)

Can we start the tutorial?

Most of the questions can be answered during the tutorial, but is there any urgent question right now?

Martin, the screen is yours . . .

Things to be done in PAIREF

- Monitoring of the structural changes during paired refinement.
- Paired refinement and anisotropic data?
  - What is my reference data?
  - How to scale the anisotropic data on my own?
- Parallelization of processes the complete cross-validation procedure.
- Improvement of the optical resolution report.
- *Low resolution paired refinement?*

Any idea? Please, send it to petr.kolenko@fjfi.cvut.cz, or martin.maly@fjfi.cvut.cz. ;-)

External sources

References:

- M. Malý, K. Diederichs, J. Dohnálek, P. Kolenko. (2020). Paired refinement under the control of PAIREF. *IUCrJ*, **7**, 681-692.
- P.A. Karplus, K. Diederichs. (2012). Linking Crystallographic Model and Data Quality. *Science*, **336**, 1030-1033.
- P.A. Karplus, K. Diederichs. (2015). Assessing and maximizing data quality in macromolecular crystallography. *Curr. Opin. Struct. Biol.*, **34**, 60-68.
- P.R. Evans, G.N. Murshudov. (2013). How good are my data and what is the resolution? *Acta Cryst D***69**, 1204-1214.
- K. Diederichs, P.A. Karplus. (2013). Better models by discarding data? *Acta Cryst D***69**, 1215-1222.

Acknowledgement

- Martin Malý
- Kay Diederichs
- Jan Dohnálek

Some people, who helped us without actually knowing that:

- Andrew Karplus
- James Holton
- Manfred Weiss
- Jan Wollenhaupt
- . . .

PAIREF

`pairef.fjfi.cvut.cz`