Improvement of model quality by rejection of non-isomorphous frames

Greta Assmann[1], Kay Diederichs[1]
[1]Molecular Bioinformatics, Department of Biology and Konstanz Research School Chemical Biology, University of Konstanz

ABSTRACT

Crystallographic data sets consisting of numerous frames usually possess random error, but additionally may display systematic differences arising from radiation damage or the experimental setup. It is crucial to combine only those frames to one complete data set, which are similar and display no systematic differences, which are in fact isomorphous. In this approach a method using $CC_{1/2}$ [1,2] was used to identify non-isomorphous frames from a reference data set [3], reject those and improve the data statistics such as $CC_{1/2}$ and internal correlation of the merged data sets. Moreover, correlation with the previously published model [4] was improved after rejection. CC1/2 therefore correctly predicts non-isomorphism and the agreement of data and model.

INTRODUCTION

What is non-isomorphism?

A protein crystal is exposed to radiation and the diffraction pattern is used to build the atomic model (Fig.1). During data collection lots of frames from different crystals are measured for reconstruction of the model. Because of radiation damage and different experimental conditions systematic errors, thus **non-isomorphism**, can occur. Only (similar) isomorphous frames should be merged for calculations.

How to find these similar datasets?

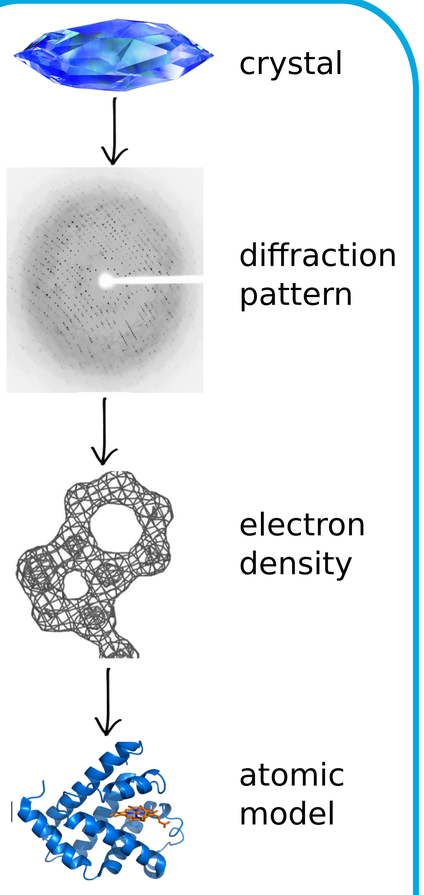


Fig.1: Molecular X-ray crystallography

METHOD

The $\Delta CC_{1/2}$ method uses $CC_{1/2}$ [1] to identify isomorphous frames. The $CC_{1/2}$ of all frames is calculated which is denoted as $\mathbf{CC_{1/2\_overall}}$. For every frame $CC_{1/2}$ is again calculated, but one specific frame is excluded during the calculations resulting in $\mathbf{CC_{1/2\_i}}$. Finally, the difference is taken:

$$\mathbf{\Delta CC_{1/2} = CC_{1/2\_overall} - CC_{1/2\_i}}$$

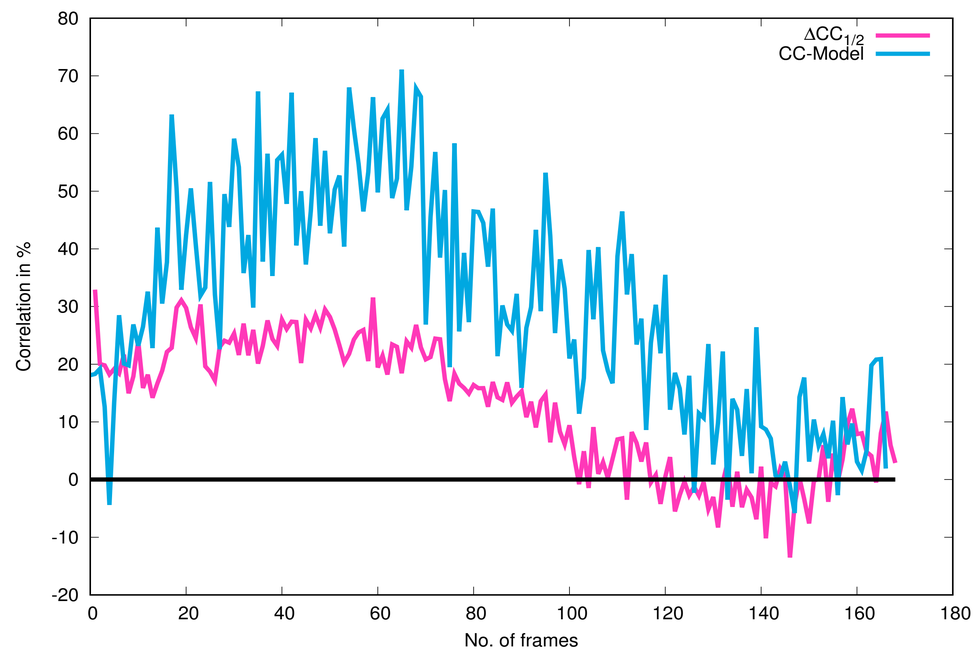
Frames with a negative $\Delta CC_{1/2}$ therefore display impairment of the data set and *vice versa*.

Acknowledgements:
We gratefully acknowledge support be the Konstanz Research School Chemical Biology (KoRS CB)

RESULTS

167 frames of one out of 5 crystals of the SBDG [3] reference set SNX17 were analysed with XDSGUI (graphical user interface of XDS) (Fig.2) and non-isomorphous frames were identified according to $\Delta CC_{1/2}$ (~>100). Exactly the same tendency can be observed for the correlation of every single frame with the previously published model [4], as non-isomorphous data sets showing negative $\Delta CC_{1/2}$ also show reduced correlation with the model. Rejection of those frames rescaling and merging with 4 other data sets led to improved $CC_{1/2}$, R_{meas} and internal correlation between the merged 5 data sets.

Fig.2:Identification of non-isomorphous frames according to $\Delta CC_{1/2}$ and validation by correlation of the single frames with the model (CC-Model)



CONCLUSION

What is new?
Frames, which are systematically different, can be identified and separated.
Why is this so fascinating?
Separation of these (somehow) different frames leads to improved data 'quality'.
Why is this necessary?
If data 'quality' is not high enough, the structure can not be solved**. Improved 'quality' leads to a solved structure.**

References
[1] Meyer P. A. et al. (2016). Nat. Commun. 7, 10882.
[2] Karplus P. A. , Diederichs K. (2012). Science 336, 1030-1.033
[3] Assmann G.,Brehm W., Diederichs K. (2016). J.Appl.Cryst. 49, 1021-1028.
[4] Stiegler A. L., Zhang R., Liu W., Boggon T. J. (2014). J.Biol.Chem. 289, 25362-25373.