
Refinement, modeling and validation in BUSTER

Clemens Vonnheim
Global Phasing Ltd.

DLS/CCP4 2020

Quality of final (complex) structure depends on:

- ☐ Good **crystals** (co-crystallisation or soaking)
- ☐ Well designed **data collection**
- ☐ Correct **data processing**

- ☐ Careful **refinement** of (initial) protein model
- ☐ Assessment of ligand density:
“Is there any ligand bound?”

- ☐ Ligand/cofactor **restraints-dictionary** based on correct chemistry
- ☐ Correct **fit** of ligand to density
- ☐ Final **refinement** of ligand-protein complex

- ☐ **A mistake in any one of these can lead to poor ligand-protein complex structures**

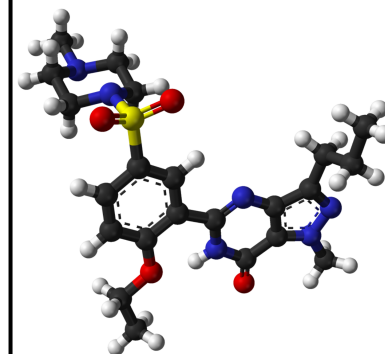
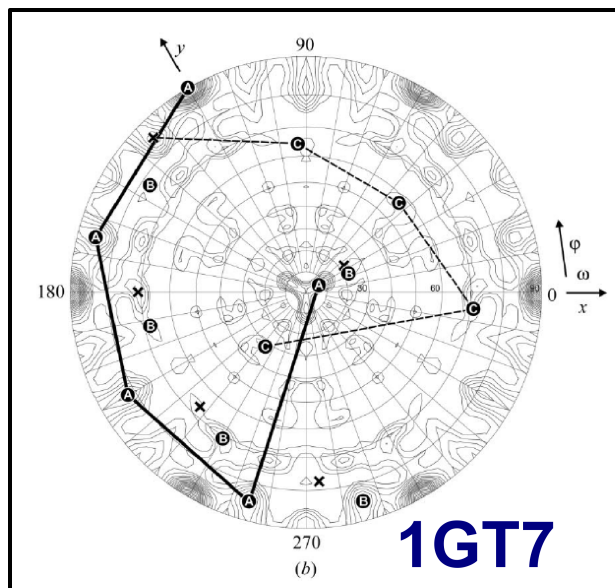
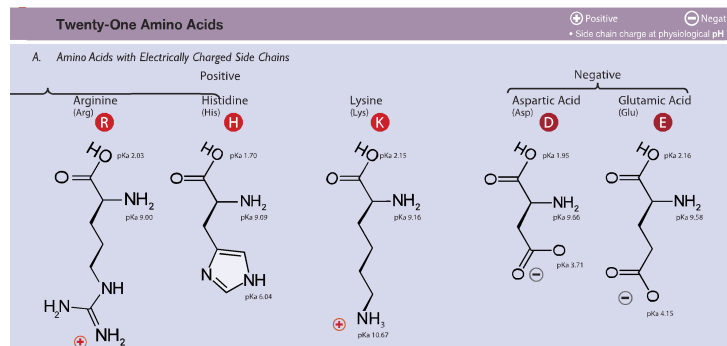
All steps need to be done (in fully automated mode) for large number of structures



We need validation at every step along the way.

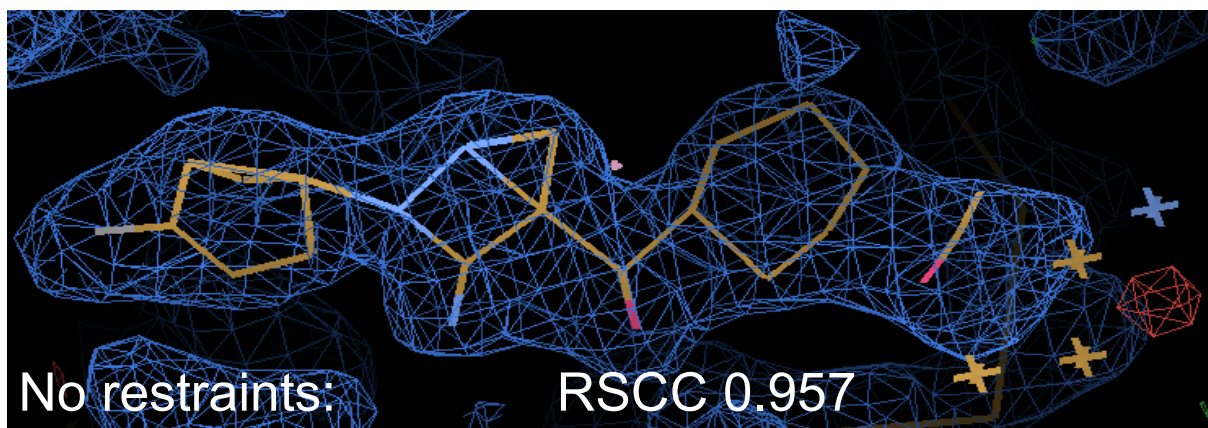
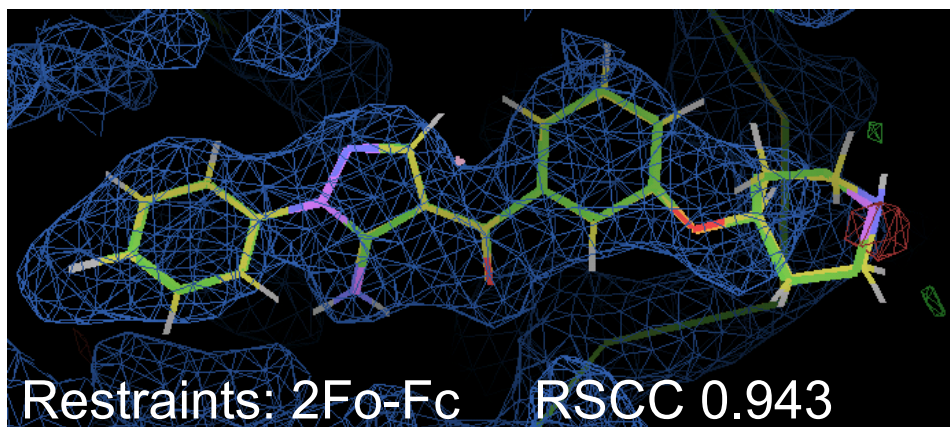
BUSTER: Refinement

- ❑ Combines prior knowledge and observed (X-Ray) data
- ❑ Chemistry (bonds, angles, ...)
 - Protein (Engh & Huber)
 - Compound/Ligand
 - ...
- ❑ Similarity
 - Within the crystal (NCS)
 - To other structures (“targeting”)
- ❑ Occupancy, disorder (alternate conformations), ...



Prior knowledge is essential

2bal kinase 2.1Å resolution.

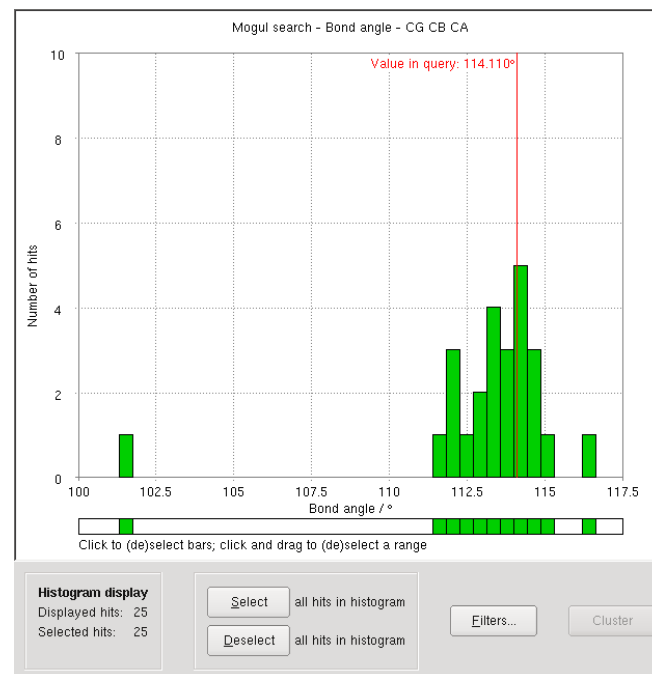
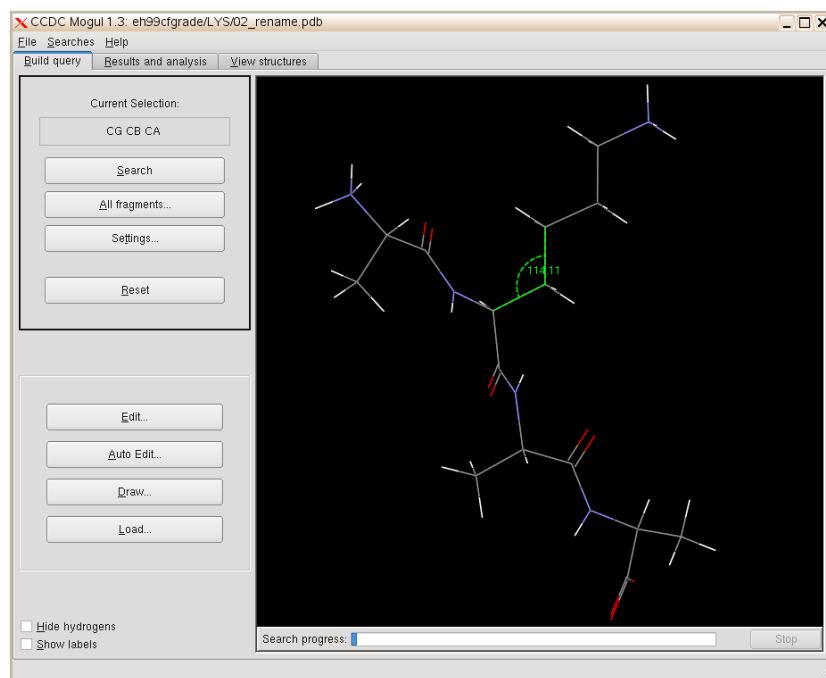


RSCC = real-space correlation coefficient

RSCC is a poor validation criteria here: it tells us that ligand model and map agree, but not if we have a good ligand model

Grade: ligand dictionaries based on CSD

- ❑ Use CCDC **mogul** program to survey CSD



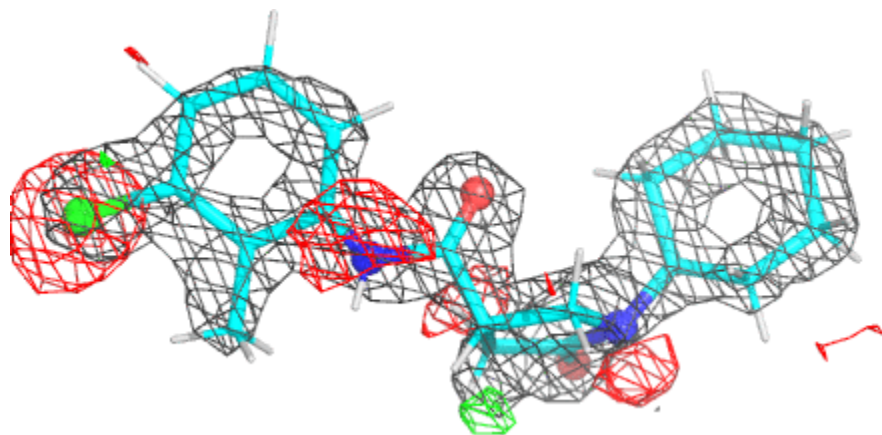
- ❑ Use mogul as **source of information for restraints** (not only in validation)
- ❑ Also AceDRG (Long et al, 2017: COD), eLBOW (Moriarty et al, 2011: CSD)

Grade: ligand dictionaries based on CSD

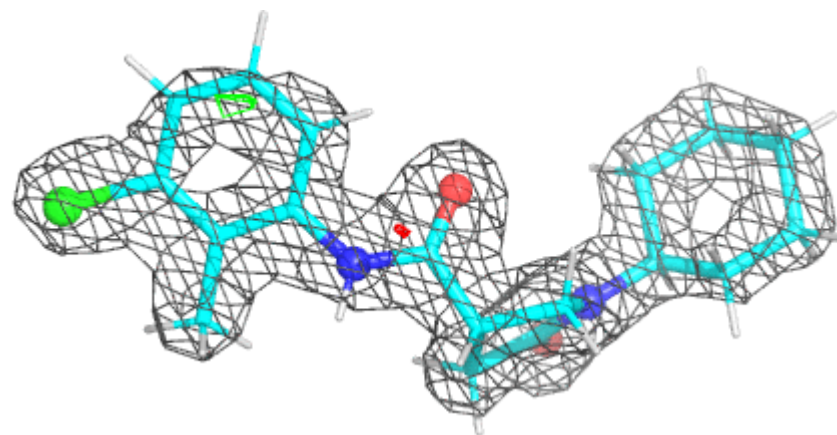
- ❑ Dictionaries based on
 - Cambridge Structural Database (CSD) = small molecule crystal structures
 - QM



<http://grade.globalphasing.org/>



2h7p (residue 468)



After BUSTER refinement with
grade dictionary

Structure

Meeting Report

Outcome of the First wwPDB/CCDC/D3R Ligand Validation Workshop

Paul D. Adams,¹ Kathleen Aertgeerts,² Cary Bauer,³ Jeffrey A. Bell,⁴ Helen M. Berman,^{5,6} Talapady N. Bhat,⁷ Jeff M. Blaney,⁸ Evan Bolton,⁹ Gerard Bricogne,¹⁰ David Brown,^{11,12} Stephen K. Burley,^{5,6,13,*} David A. Case,⁶ Kirk L. Clark,¹⁴ Tom Darden,¹⁵ Paul Emsley,¹⁶ Victoria A. Feher,^{17,*} Zukang Feng,^{5,6} Colin R. Groom,^{18,*} Seth F. Harris,⁸ Jorg Hendle,¹⁹ Thomas Holder,⁴ Andrzej Joachimiak,²⁰ Gerard J. Kleywegt,²¹

502 Structure 24, April 5, 2016 ©2016 Elsevier Ltd All rights reserved

Figure

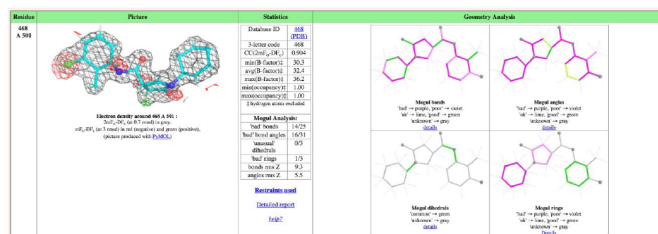


Figure 1. Example highlighting the value of presenting ligand electron density model fit and geometrical analysis from CCDC Mogul from the Global Phasing Buster Report (PDB ID: 2H7P, later superseded by entry 4T2T (He et al., 2006); CCD ID: 468).

- ☐ REMARK 3 header
- ☐ buster-report tool
- ☐ Part of BUSTER distribution

July 2019: Buster-report now part of wwPDB validation suite

06/11/2019

Improvements to visualization of ligand validation and electron density maps in the wwPDB validation report

Our recent update to the wwPDB validation reports provides much clearer validation information for ligands.

We now include 2-dimensional diagrams of ligands, highlighting geometric validation criteria and, for structures determined by crystallography, 3-dimensional views of electron density.

We also provide calculated electron density map coefficients which were used to generate the analysis in the validation reports.

Ligand Validation

We have collaborated with [Global Phasing Ltd](#) to integrate the ligand visualization from [buster-report](#) into the wwPDB validation report, as recommended by the [wwPDB/CCDC/D3R Ligand Validation Workshop](#). The ligand visualization will be available for ligands that have been designated as "Ligand of Interest" by the depositor and ligands with a molecular weight greater than 250 Daltons that have outliers.

The following ligand instance of NAP was chosen intentionally as a representative of sub-optimal quality in both the ligand model and its agreement with the X-ray data.

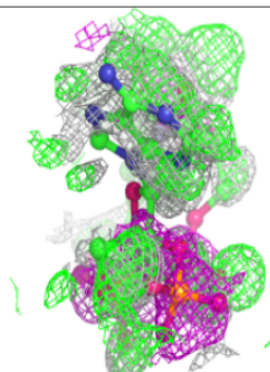
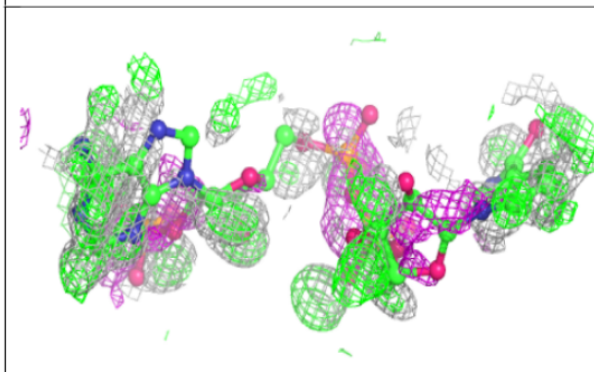
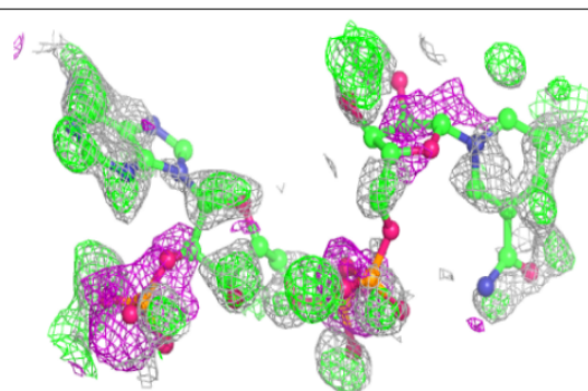
Geometric analysis provided by CCDC Mogul will be highlighted on a 2D diagram of the ligand, as shown below.

RCSB PDB News Image

In addition to geometric validation for ligands, for X-ray diffraction PDB entries the wwPDB validation report also presents images displaying the ligand and the surrounding electron density map.

Electron density around NAP A 1270:

$2mF_o - DF_c$ (at 0.7 rmsd) in gray
 $mF_o - DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

**Take care in restraint dictionary generation**

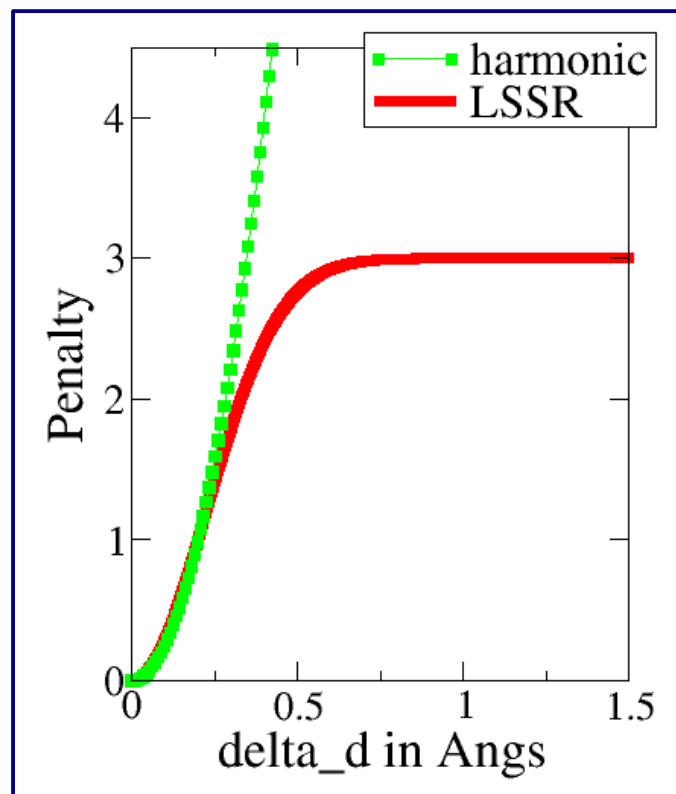
Local Structural Similarity Restraints (LSSR)

- ❑ Conventional superposition-based restraints are laborious to use.
- ❑ LSSR is a much easier to use approach to NCS:
 - “involves” local contact distances
 - “softer” than superposition-based methods – violations entail only a fixed cost
- ❑ method in BUSTER fully automated (detection and application): **-autoncs**

❑ ***Always use NCS restraints***

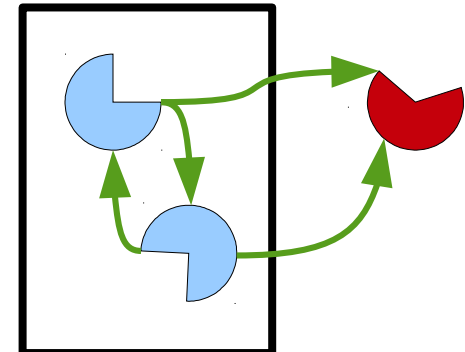
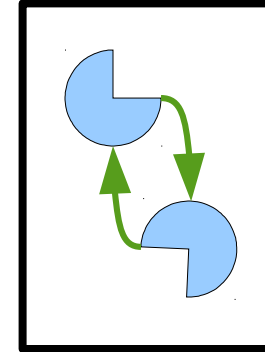
Smart et al (2008). Abstr. Annu. Meet. Am. Crystallogr. Assoc., Abstract TP139, p. 117.

Smart et al (2012). Acta D68, 368–380



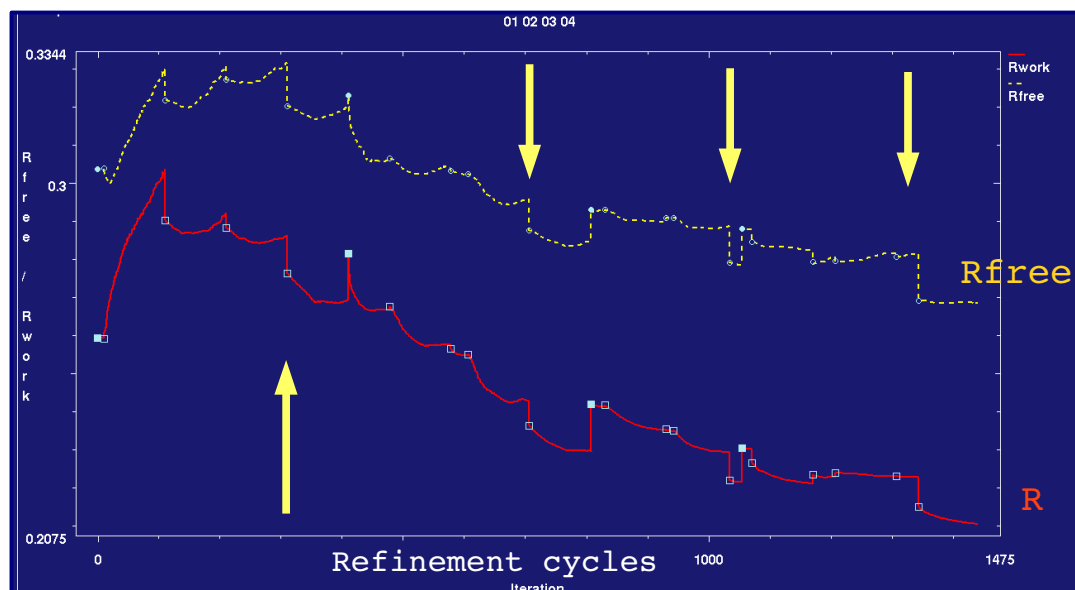
LSSR Target Restraints

- ❑ NCS restraints: between chains within the same model
- ❑ Often, the chain being refined is similar to a structure that has already been solved
- ❑ For example:
 - ligand complex with higher resolution apo
 - two crystal forms of the same protein
 - partial datasets from non-isomorphous crystals
 - following radiation damage
- ❑ The already solved structure becomes the “target”
- ❑ Apply LSSR restraints to the fixed target structure supplied as pdb file (`-target some.pdb`)



BUSTER - void correction

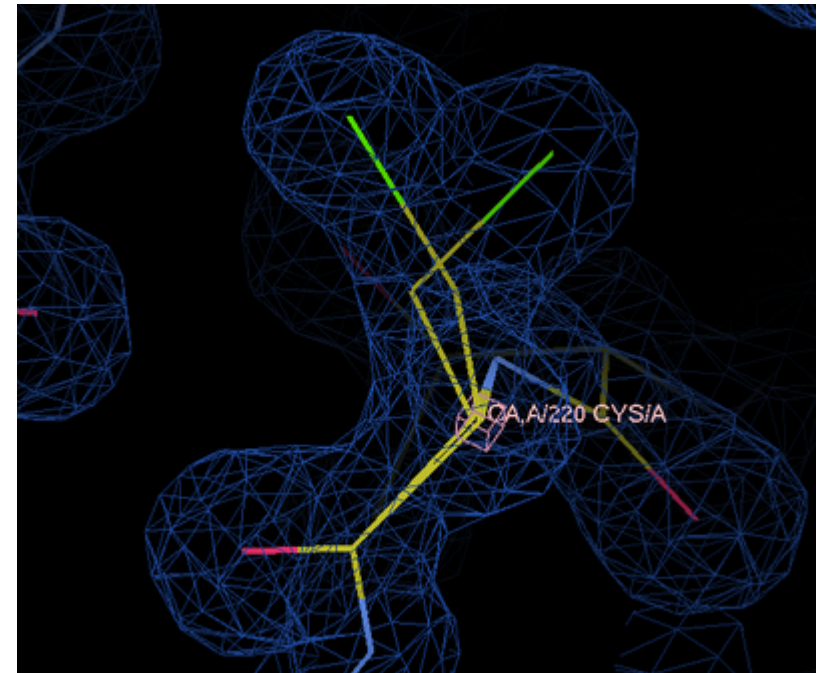
- ❑ Structures often have voids/cavities filled neither by ordered or disordered solvent
 - “filled” by vacuum
 - mask-based bulk solvent correction needs to take this into account
- ❑ Before the last big cycle, BUSTER will try and detect such voids and exclude them automatically from the bulk solvent area
- ❑ This can often bring a 'jump' in R/Rfree at the last big cycle
- ❑ But main effect is to clean-up difference (mFo-DFc) map ... especially if using rms-based cutoff to decide on significance of difference map



Alternate conformations

Alternate conformation

- ❑ Atoms with altConf “A” see atoms with no altConf or with altConf “A”
- ❑ Atoms with different altConf don’t see one another
- ❑ Set occupancies you want refined to something not equal to 1.0, and then run
 - `pdb2occ -p model.pdb -o occ.gelly`
 - `refine -Gelly occ.gelly ...`
- ❑ If $\text{occ}(A) + \text{occ}(B) = 1$ in input, `pdb2occ` will restrain occupancy sum (to 1)



BUSTER – additional restraints

- E.g. using powerful and flexible utility restraints (Gelly) in BUSTER
- Example: $\text{Mg}^{2+} \cdot (\text{H}_2\text{O})_6$

4U1U

- Display Files ▾
- Download Files ▾
- Usage Note ?
- Download Citation ▾

Entry 4U1U supersedes and combines 4PE9 4PEA 4PEB 4PEC

Primary Citation

Synergy of streptogramin antibiotics occurs independently of their effects on translation.

Noeske, J. , Huang, J. , Olivier, N.B. , Giacobbe, R.A. , Zambrowski, M. , Cate, J.H. 

Journal: (2014) Antimicrob.Agents Chemother. **58**: 5269-5279

PubMed: 24957822 

PubMedCentral: PMC4135883 

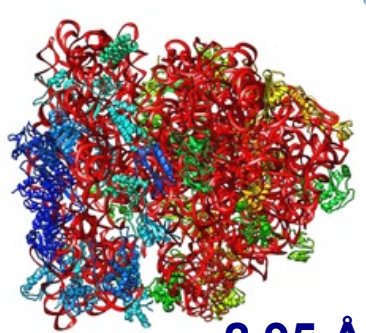
DOI: 10.1128/AAC.03389-14 

Search Related Articles in PubMed 

PubMed Abstract:

Streptogramin antibiotics are divided into types A and B, which in combination can act synergistically. We compared the molecular interactions of the streptogramin combinations Synercid (type A, dalbapristin; type B, quinupristin) and NXL 103 (type A, flopristin; type B, linopristin)... [[Read More & Search PubMed Abstracts](#)]

◀ Biological Assembly 1 ? ▶



2.95 Å

 3D View: Jmol [More Images](#)

Biological assembly 1 assigned by authors

Downloadable viewers:

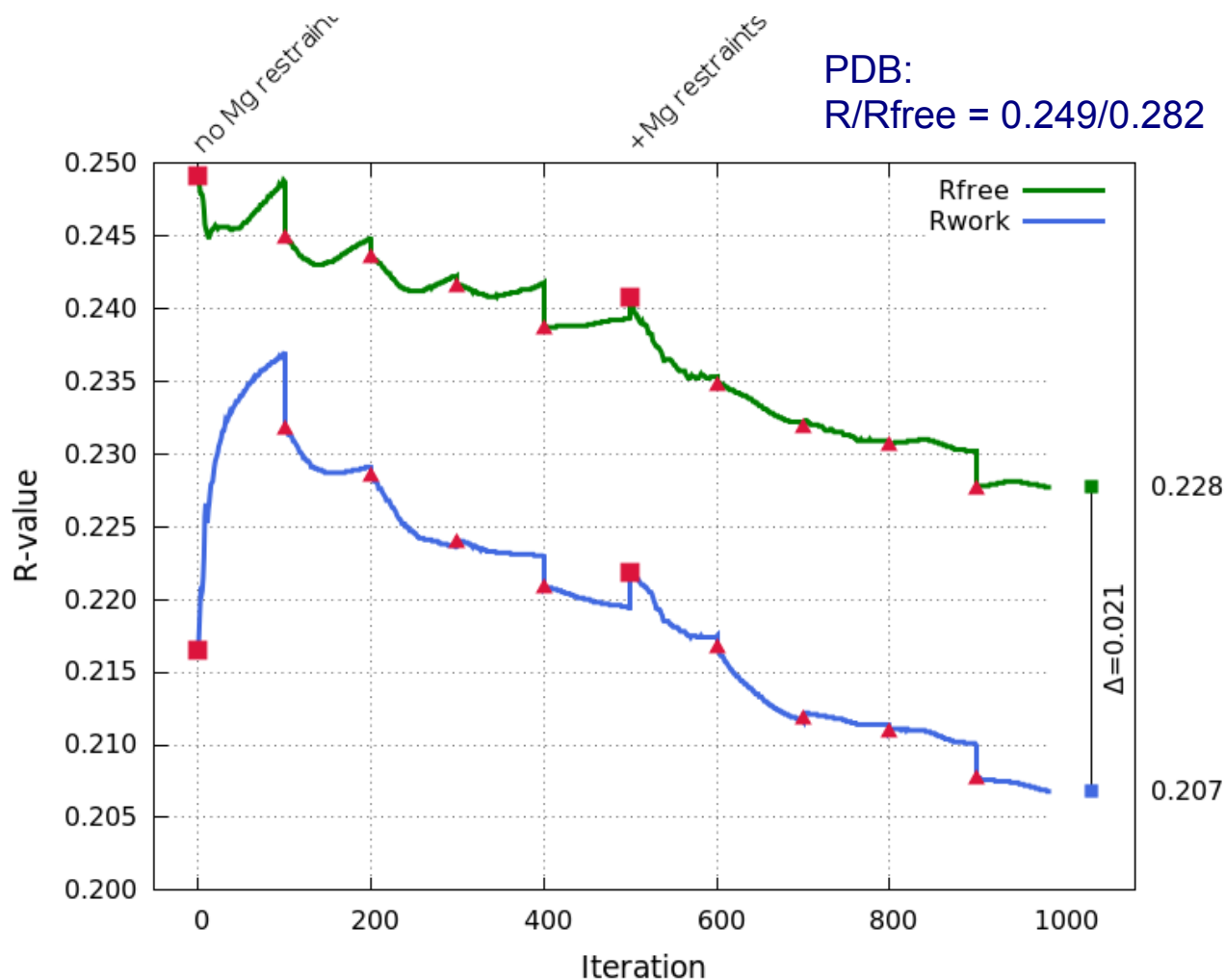
[Simple Viewer](#) [Protein Workshop](#)

⚡ Molecular Description

Hide

- **1.1 Mio reflections**
- **290 000 atoms in asymmetric unit**
- **11 000 amino acids, 9000 RNA, 500 Mg, 1700 HOH**

4U1U: BUSTER refinement



CAVEAT:

Comparing R/Rfree from different programs difficult: they use different methods (e.g. bulk solvent model) and scaling.

Sometimes program A gives lower R/Rfree, sometimes program B.

Goal is to get best model - fitting data, chemistry and prior knowledge.

Good to have choice of different approaches and parametrisation!

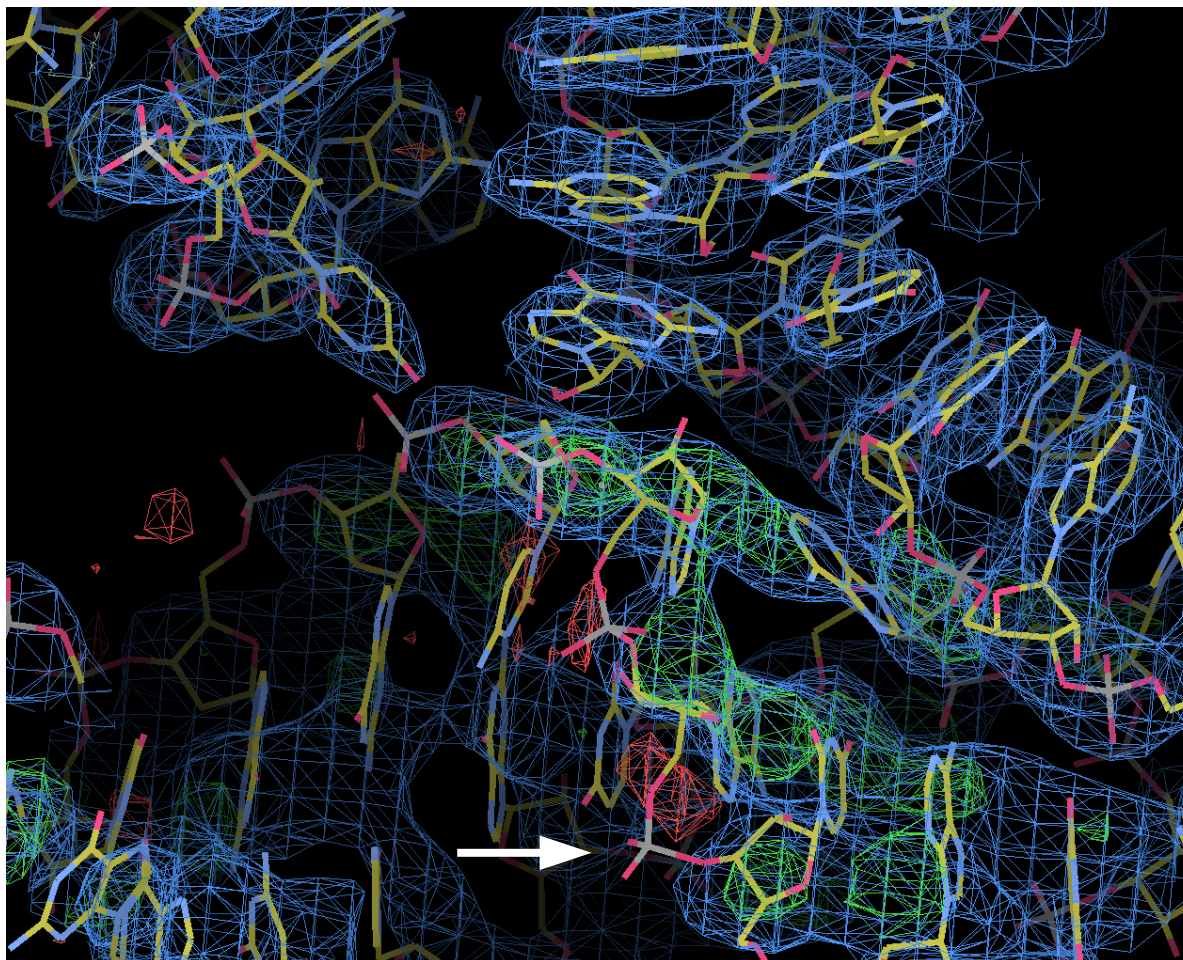
BUSTER: 4U1U

Region-1:

deposited model

Initial map

2Fo-Fc @ 1.5rms
Fo-Fc @ 3.5rms



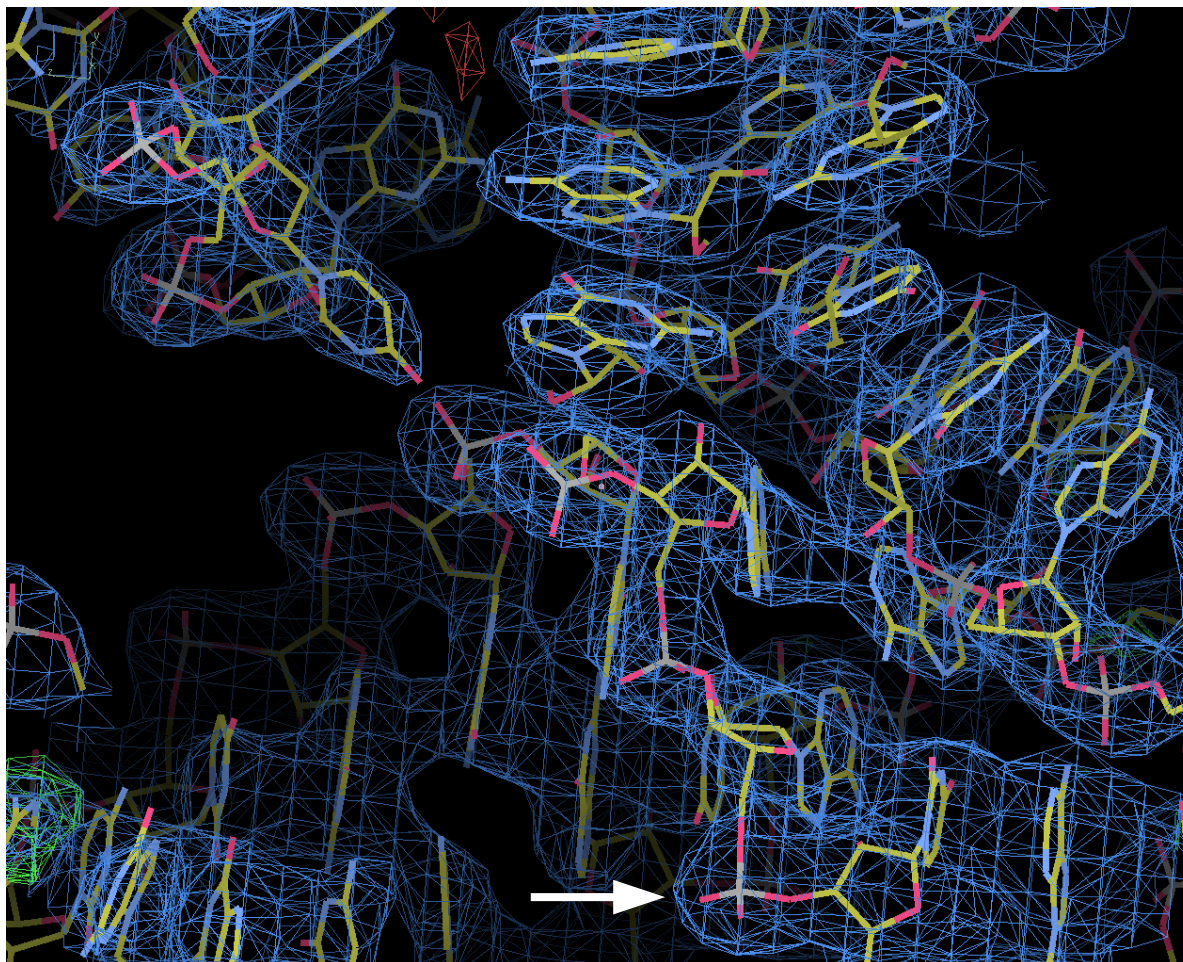
BUSTER: 4U1U

Region-1:

BUSTER model

Final map

2Fo-Fc @ 1.5rms
Fo-Fc @ 3.5rms



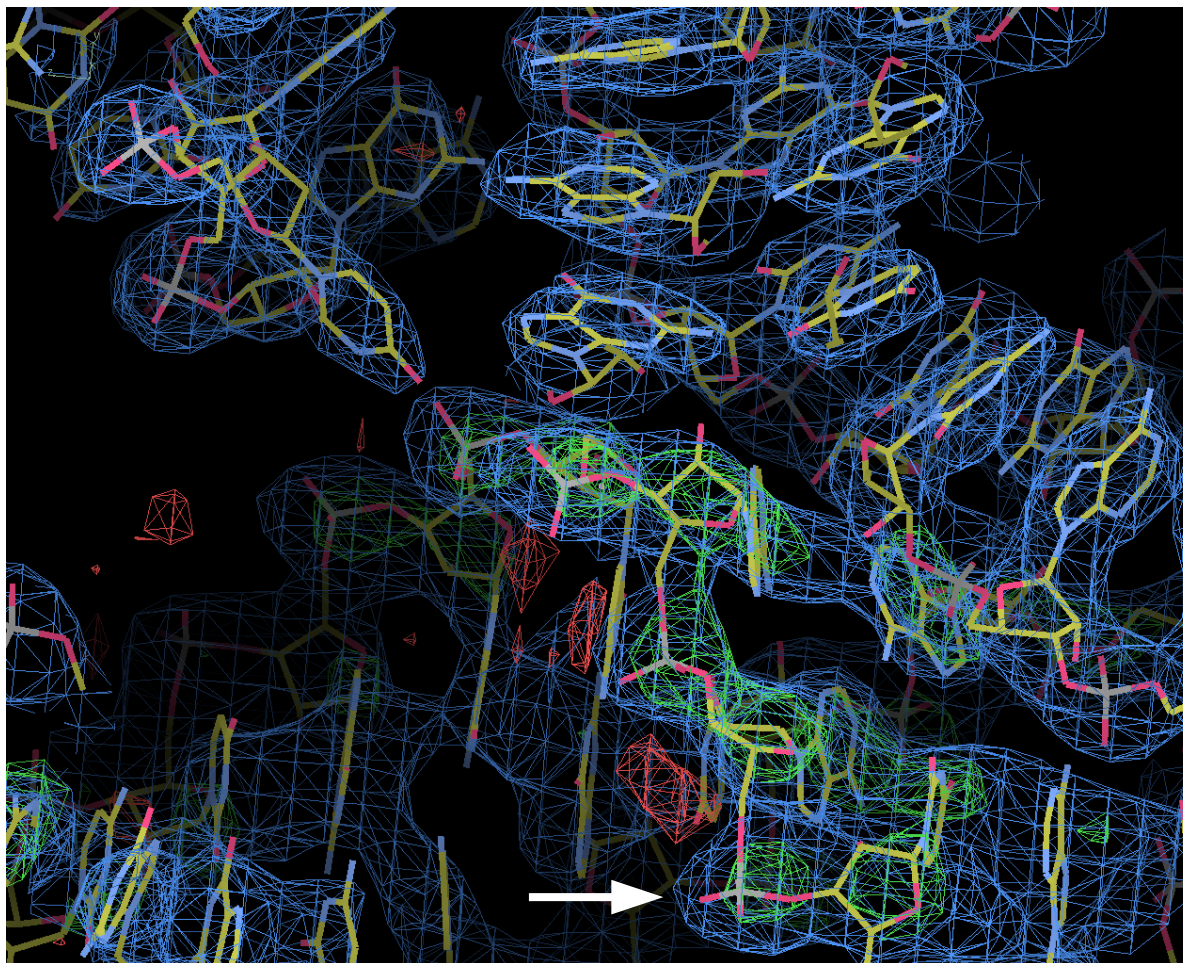
BUSTER: 4U1U

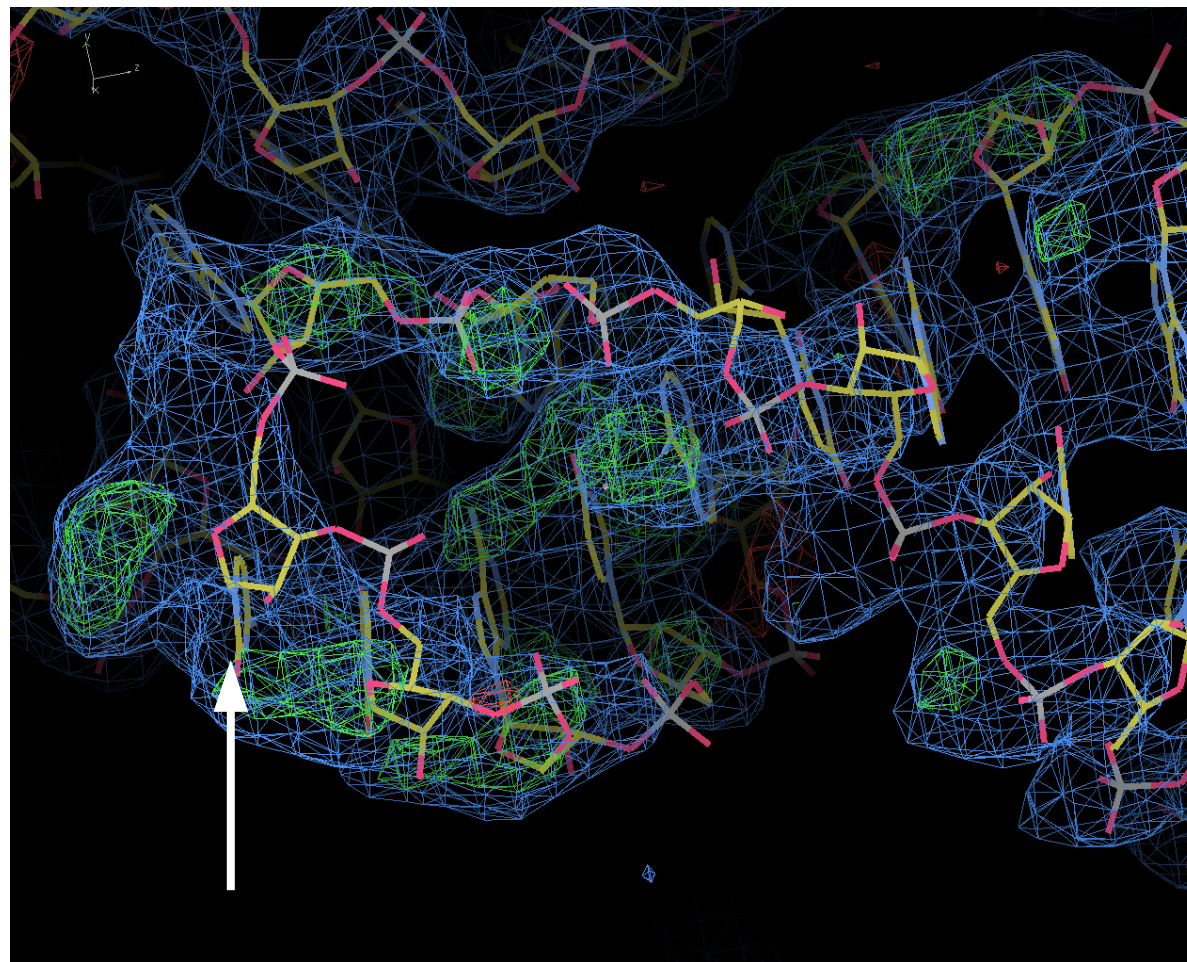
Region-1:

BUSTER model

Initial map

**2Fo-Fc @ 1.5rms
Fo-Fc @ 3.5rms**



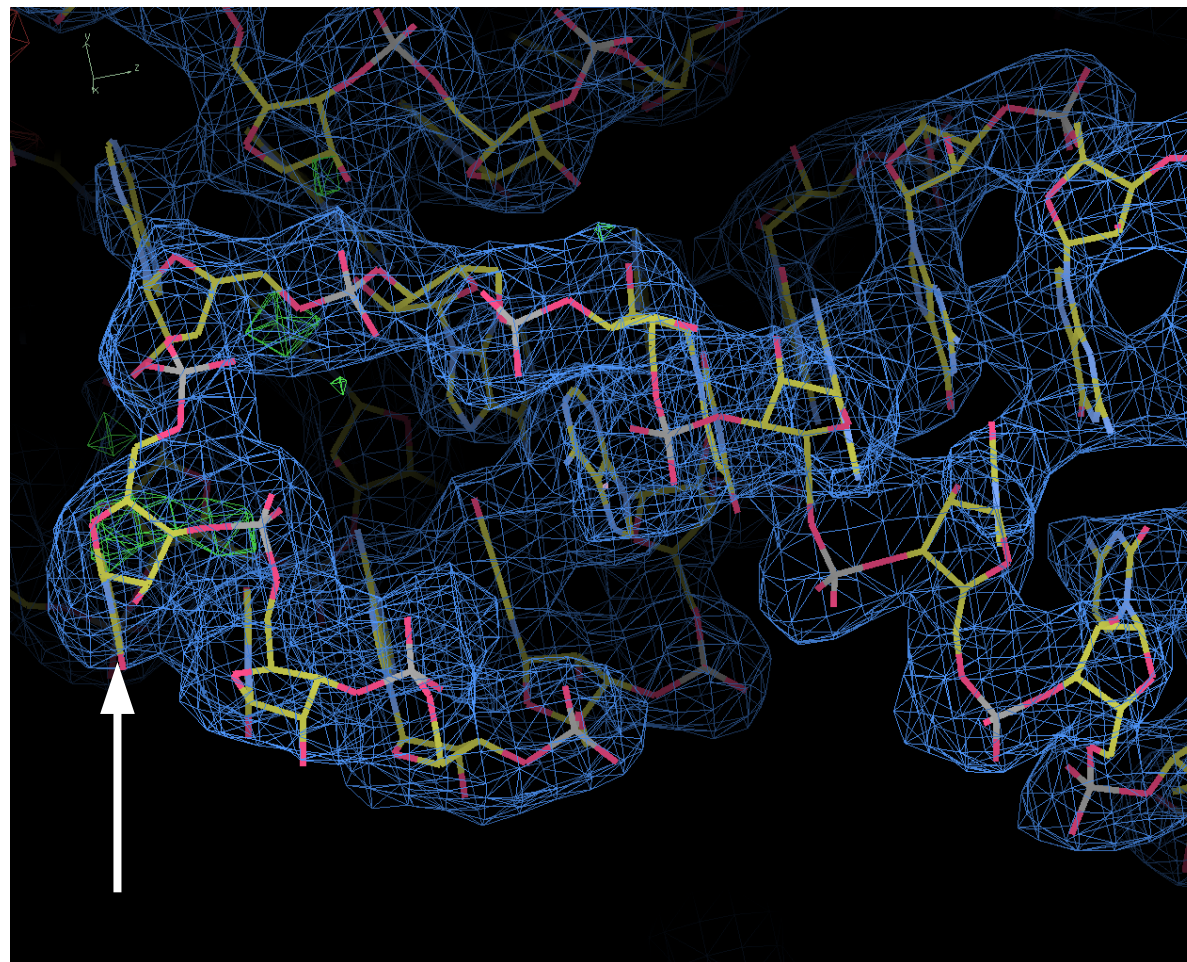


Region-2:

deposited model

Initial map

**2Fo-Fc @ 1.5rms
Fo-Fc @ 3.5rms**

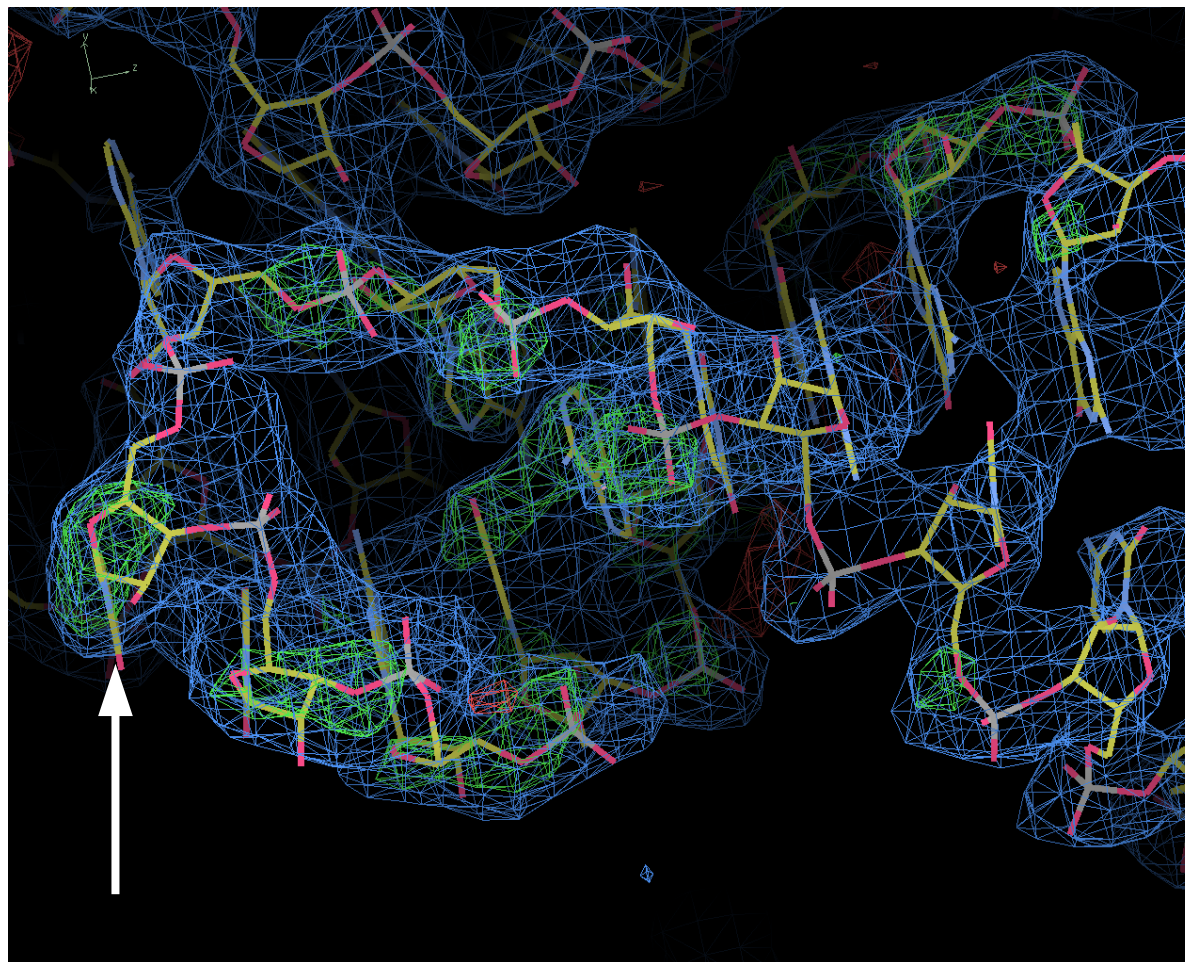


Region-2:

BUSTER model

Final map

2Fo-Fc @ 1.5rms
Fo-Fc @ 3.5rms



Region-2:

BUSTER model

Initial map

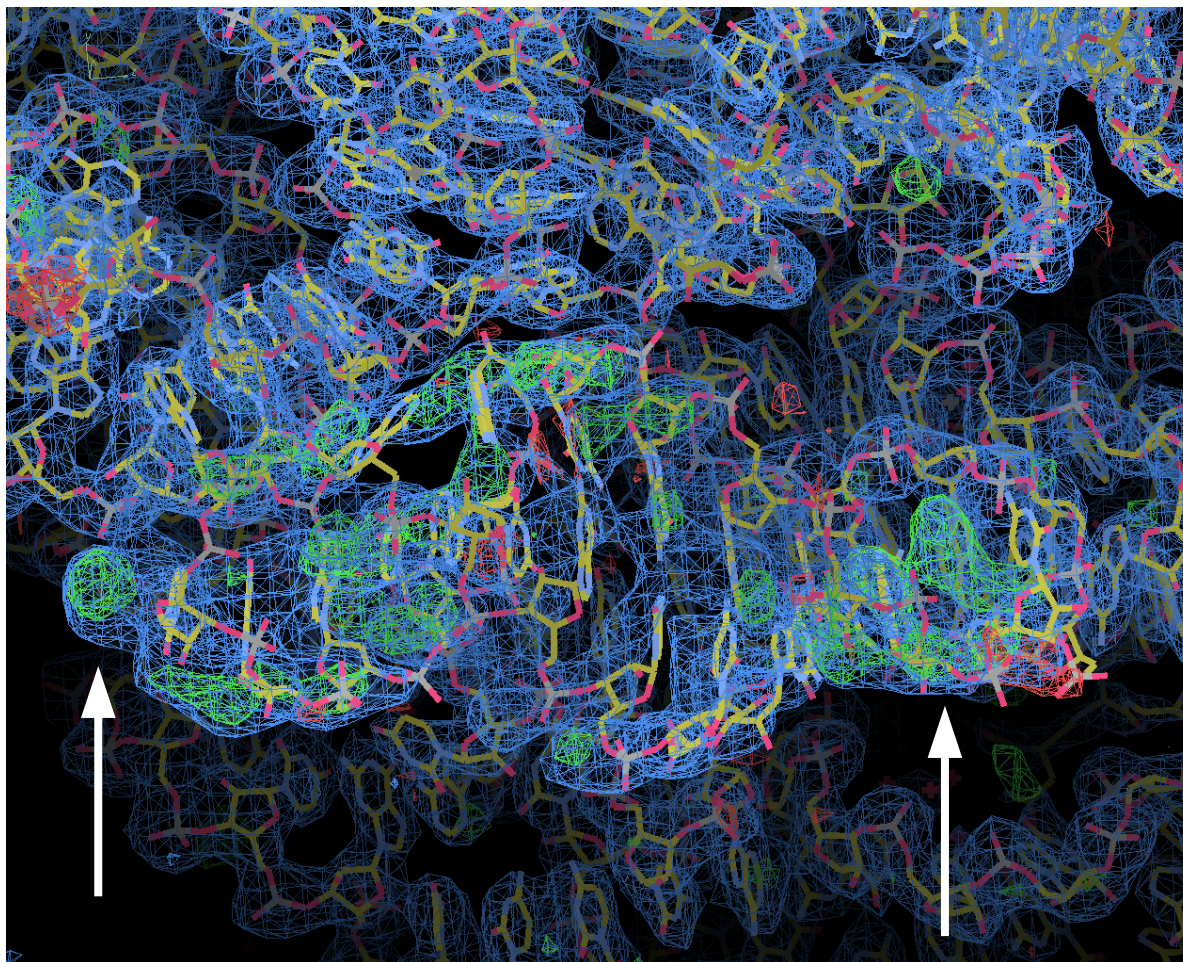
2Fo-Fc @ 1.5rms
Fo-Fc @ 3.5rms

Region-2:

deposited model

Initial map

2Fo-Fc @ 1.5rms
Fo-Fc @ 3.5rms

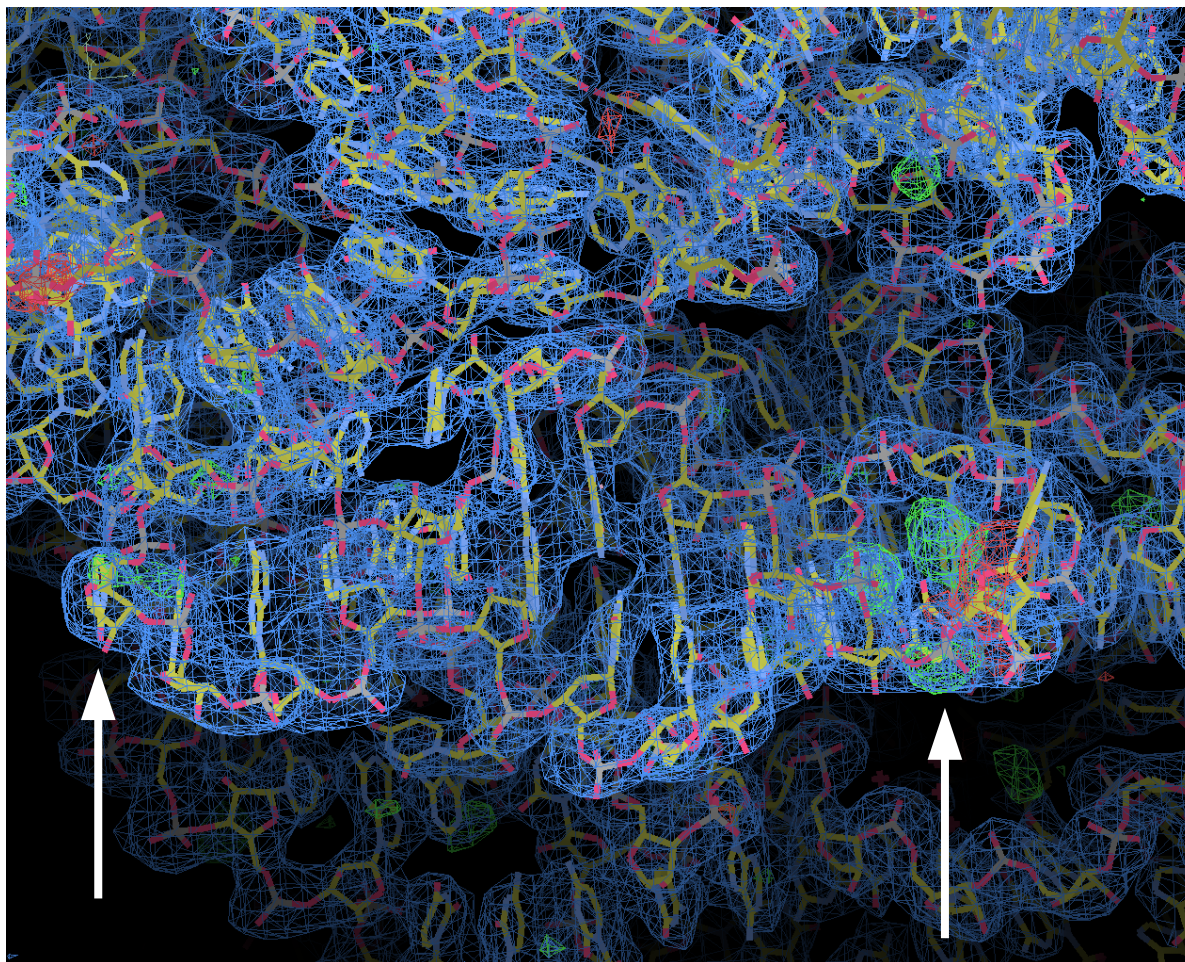


Region-2:

BUSTER model

Final map

2Fo-Fc @ 1.5rms
Fo-Fc @ 3.5rms

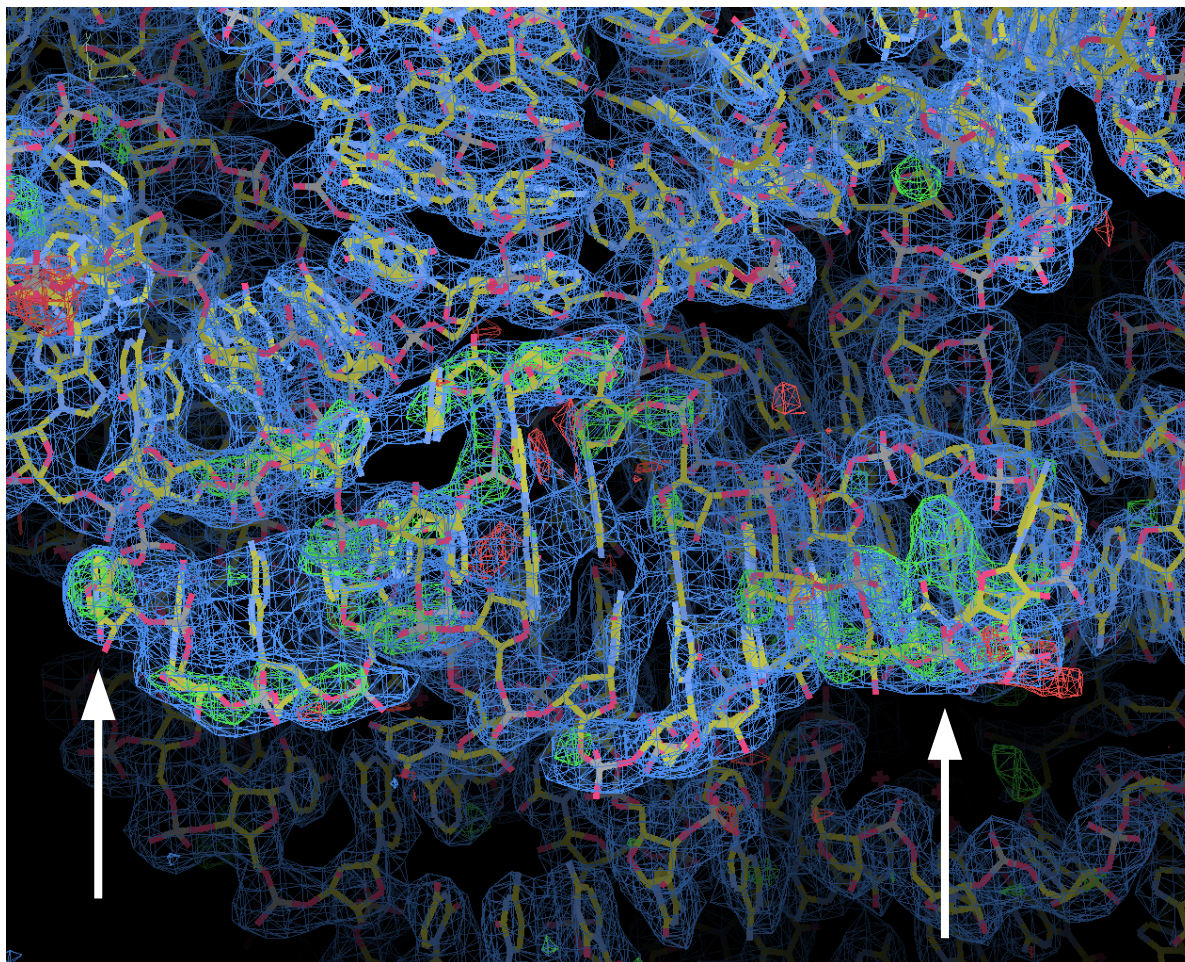


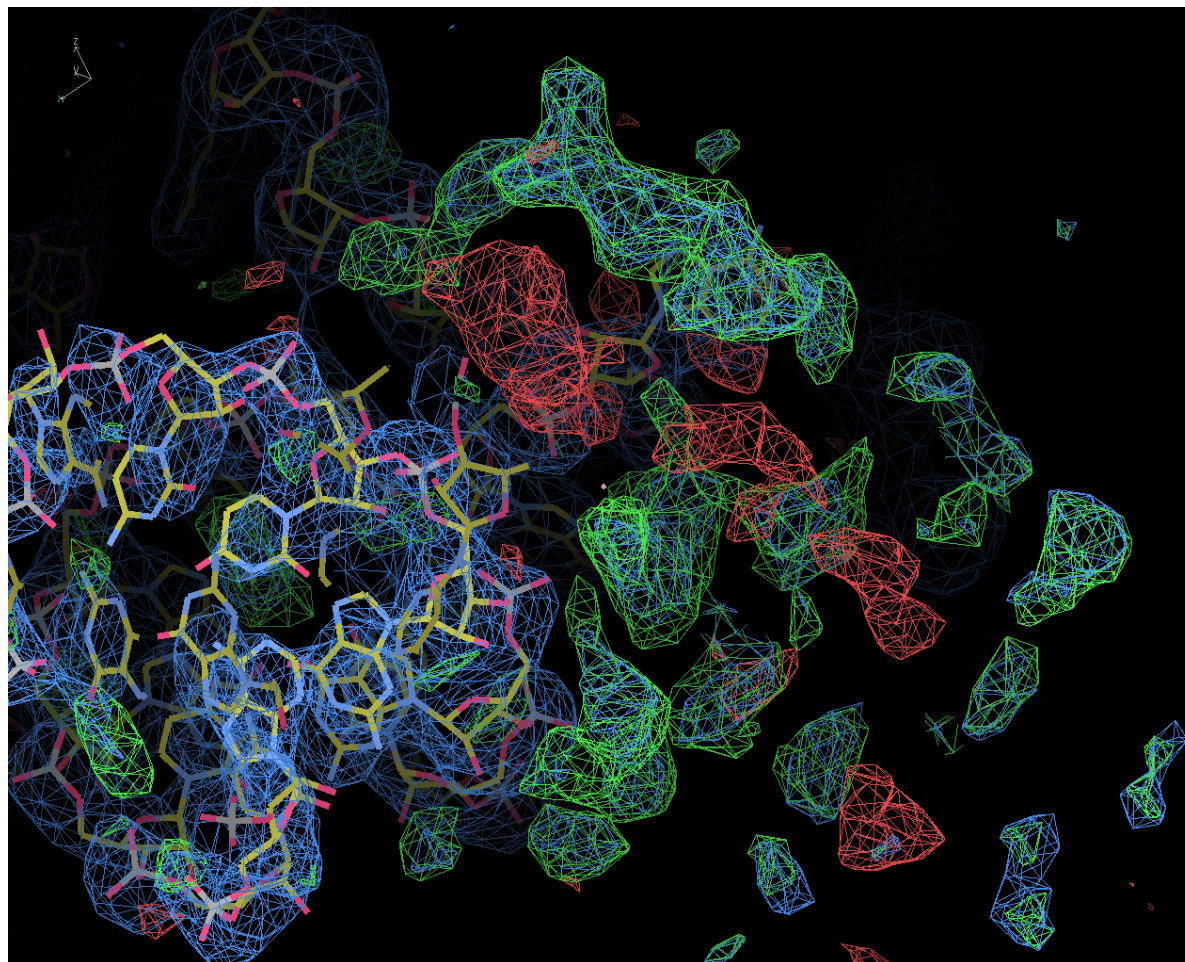
Region-2:

BUSTER model

Initial map

2Fo-Fc @ 1.5rms
Fo-Fc @ 3.5rms





Region-3:

deposited model

Initial map

2Fo-Fc @ 1.5rms

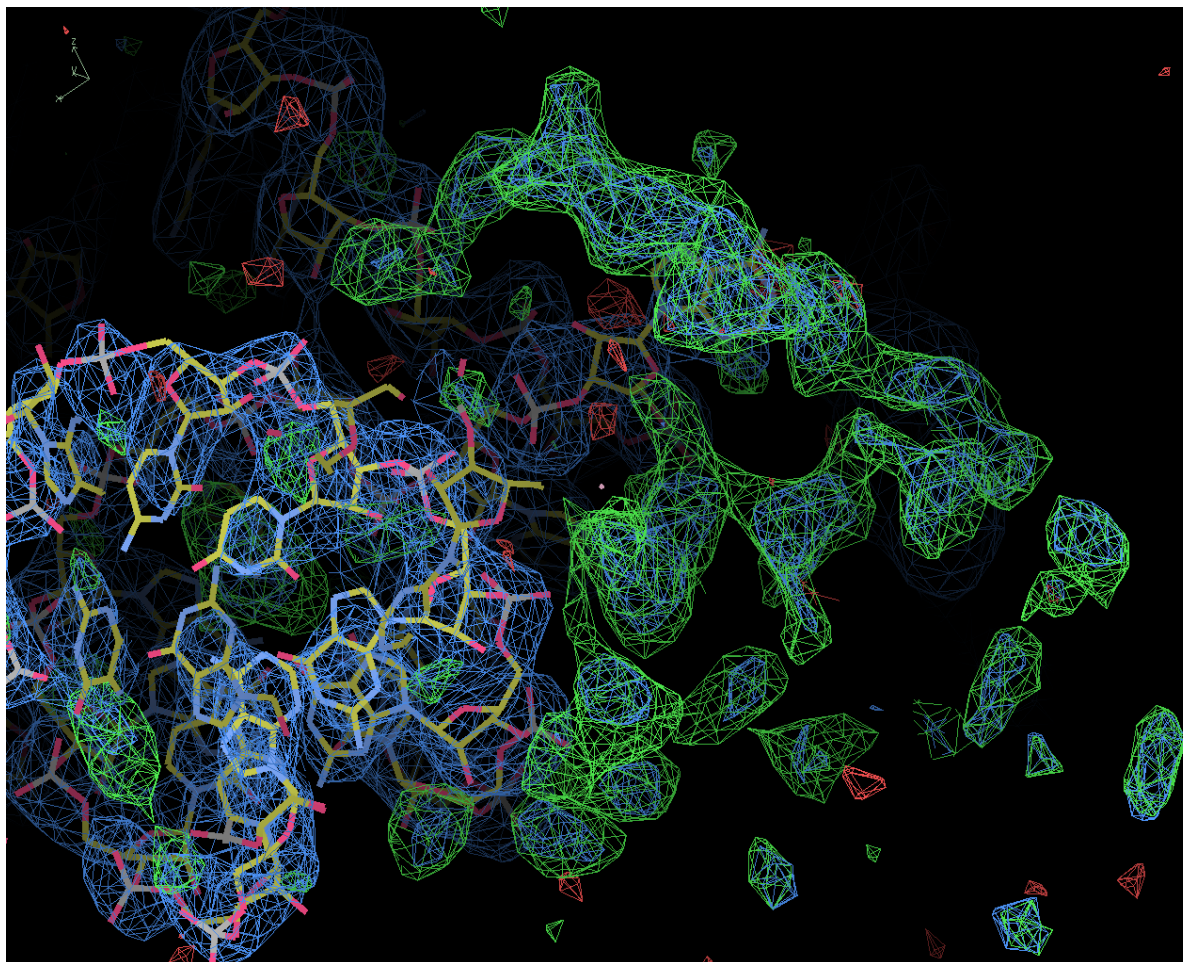
Fo-Fc @ 3.5rms

Region-3:

BUSTER model

Final map

2Fo-Fc @ 1.5rms
Fo-Fc @ 3.5rms



Some notes about rotamers and outliers

- ❑ MolProbity rotamer analysis often used as a quality criteria
- ❑ outliers are considered problematic und highly unlikely (the stated goal is <1% outliers)

6D0E (deposited)

2%	A	10	LEU:1.00:0.4:280.2:36.0::Allowed:mp
2%	A	19	LEU:1.00:0.4:273.1:36.4::Allowed:mp
2%	A	24	LEU:1.00:0.4:260.6:43.2::Allowed:mp
2%	A	45	LYS:1.00:0.5:199.5:130.2:301.8:173.9:Allowed:ttmt
	A	50	ILE:1.00:0.0:249.5:205.1::OUTLIER:OUTLIER
20%	A	82	AVAL:0.33:2.1:311.2:::Favored:m
	A	97	LEU:1.00:0.0:235.3:15.2::OUTLIER:OUTLIER
59%	B	23	LEU:1.00:0.4:247.2:198.3::Allowed:mt
?	B	41	ARG:1.00:0.4:142.9:262.7:191.0:181.3:Allowed:tmt170
2%	B	45	LYS:1.00:0.3:203.2:129.9:301.1:170.3:Allowed:ttmt
8%	B	50	ILE:1.00:0.4:215.5:155.0::Allowed:tt
20%	B	82	VAL:1.00:1.2:312.5:::Allowed:m
	B	97	LEU:1.00:0.0:236.7:13.8::OUTLIER:OUTLIER

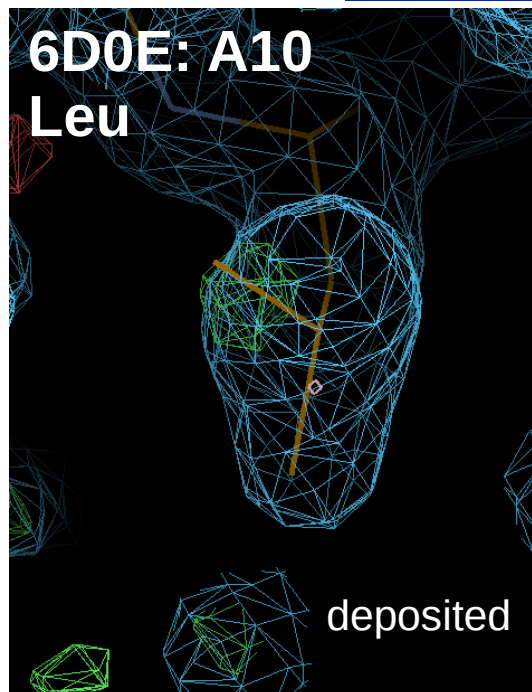
BUSTER

A	10	LEU:1.00:0.0:259.1:21.3:::OUTLIER:OUTLIER
A	19	LEU:1.00:0.1:309.3:303.1:::OUTLIER:OUTLIER
A	24	LEU:1.00:0.2:297.4:296.7:::OUTLIER:OUTLIER
A	45	LYS:1.00:0.1:197.6:122.2:308.4:165.6:OUTLIER:OUTLIER
A	50	ILE:1.00:0.0:239.0:174.6:::OUTLIER:OUTLIER
A	82	AVAL:0.33:0.2:320.6:::OUTLIER:OUTLIER
A	97	LEU:1.00:0.0:238.9:5.1:::OUTLIER:OUTLIER
B	23	LEU:1.00:0.0:180.7:193.4:::OUTLIER:OUTLIER
B	41	ARG:1.00:0.1:138.1:266.5:172.2:188.0:OUTLIER:OUTLIER
B	45	LYS:1.00:0.1:197.5:123.2:307.8:166.2:OUTLIER:OUTLIER
B	50	ILE:1.00:0.0:238.2:171.6:::OUTLIER:OUTLIER
B	82	VAL:1.00:0.1:326.1:::OUTLIER:OUTLIER
B	97	LEU:1.00:0.0:238.8:5.0:::OUTLIER:OUTLIER

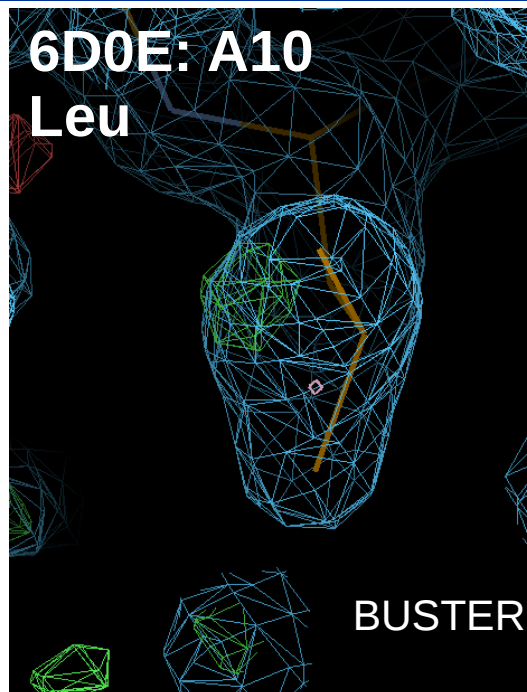
Unlikely – but not
classified as “outlier”

- ❑ Do we have a problem with rotamers in BUSTER refinements?

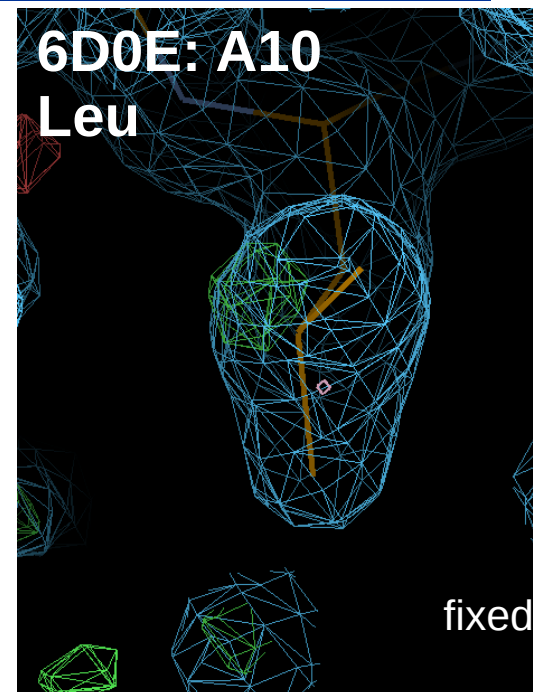
Rotamer has to fit density (for X-Ray crystallographic models)



mp (2%): Allowed



OUTLIER



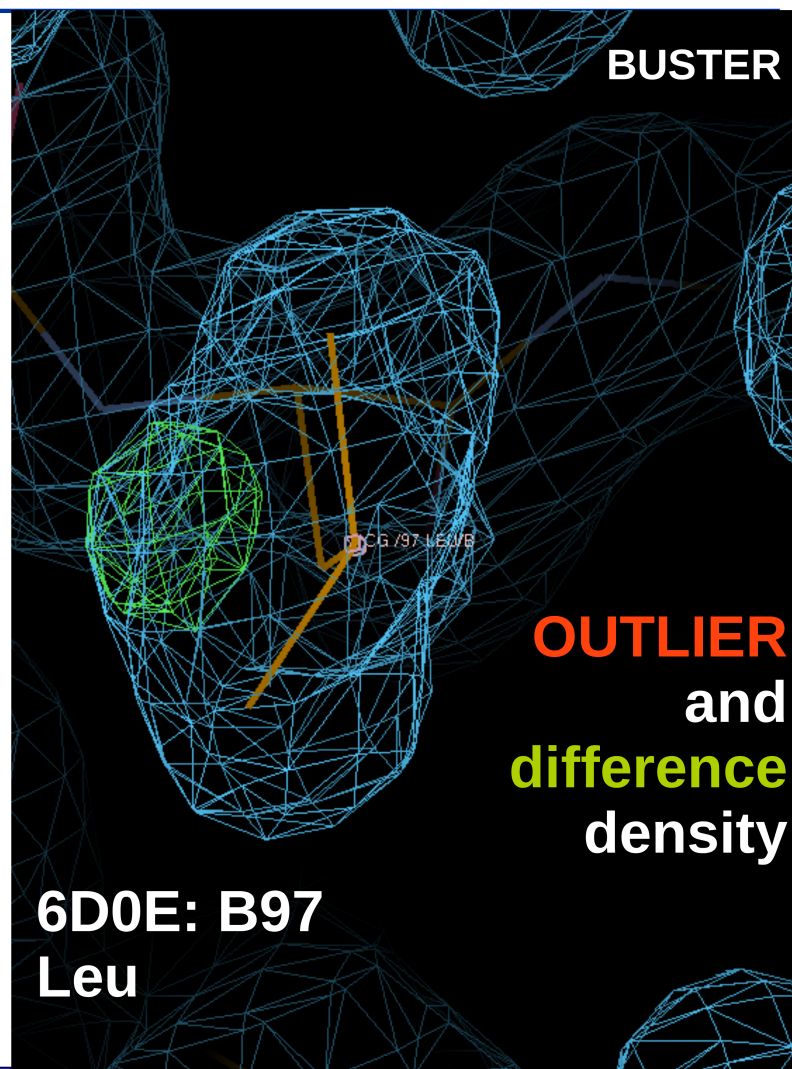
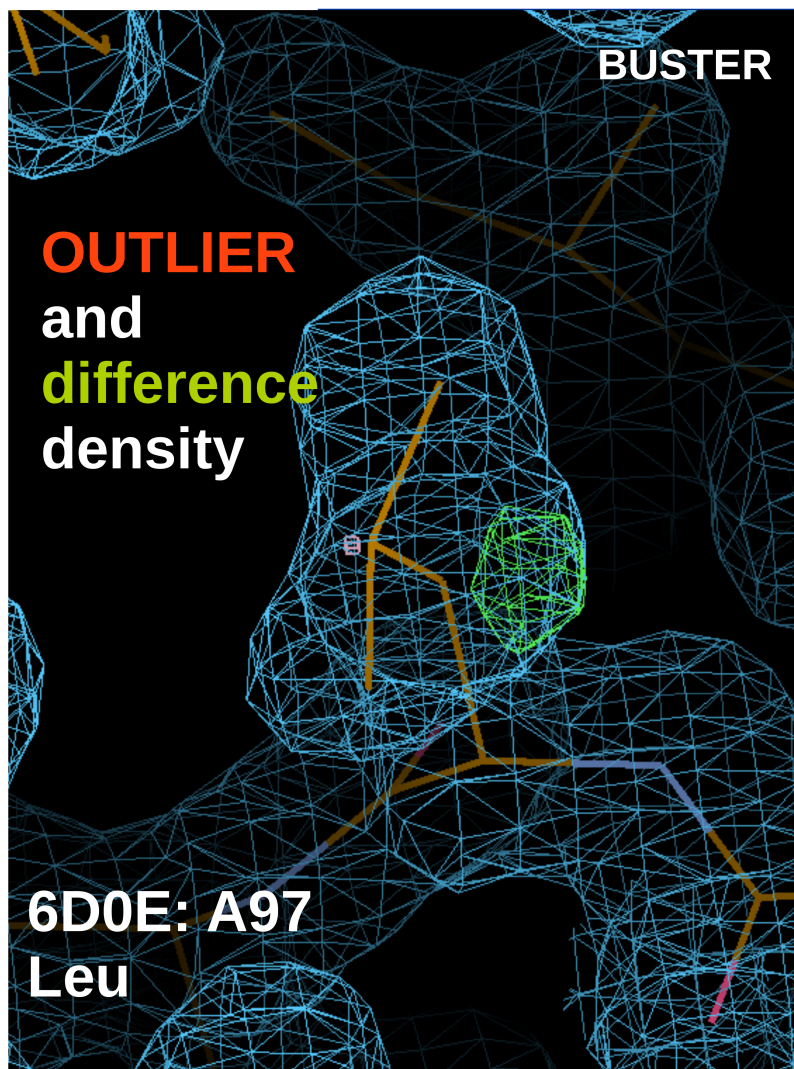
mt (59%): favoured

- small, but significant difference between mt (59%) and mp (2%)
- BUSTER refines model away from poor/wrong rotamer and gives clear indications:
 - positive difference density
 - rotamer now highlighted as outlier
- enforcing rotamer restraints can mask incorrect side-chain conformations

Rotamer outliers are useful markers



Rotamer outliers are useful markers



BUSTER: Is there any any ligand bound?

Acta Cryst. (2005). **A61**, C248

Automated Structure Refinement for High-throughput Ligand Detection with BUSTER-TNT

Clemens Vornrhein, Gerard Bricogne, *Global Phasing Ltd., Cambridge, UK*. E-mail: vornrhein@GlobalPhasing.com

The use of crystallography for the discovery of lead compounds often involves a large number of experiments with different soaking or co-crystallization trials. The subsequent refinement and analysis of the resulting datasets can be time-consuming and tedious. Since the crystallographic parameters (resolution, space group, cell parameters) are quite often similar, this task is ideally suited for automation.

We present a method (**autoBUSTER**) that automates the refinement, solvent model update, ligand detection and analysis. Centered around the BUSTER-TNT program [1,2], it requires a minimal amount of user input. Although it can be used at any resolution and for any kind of macromolecular structure, it is tuned to the refinement of protein structures at better than 2.8 Å resolution.

The knowledge of any (possibly) bound ligand can be given (a) explicitly by supplying a PDB file of dummy atoms that describes the assumed binding site, or (b) by letting the system automatically analyze the residual density of difference Fourier maps. A unique feature of BUSTER-TNT is used, where the various masks describing the known fragment, the bulk solvent and the missing part can be given independently from each other. The results show that this can greatly enhance the capability of uniquely defining any bound ligand.

[1] Bricogne G., Irwin J., *Macromolecular Refinement: Proceedings of the CCP4 Study Weekend*, Warrington: Daresbury Laboratory, 1996, 85-92. [2] Blanc E., Roversi P., Vornrhein C., Flensburg C., Lea S. M., Bricogne G., *Acta Cryst.*, 2004, **D60**, 2210-2221.

- ☐ BUSTER's '-L' feature tries to take an **unbiased look** at data/model to decide **if, where and how** something ***might*** have bound.
- ☐ No prior knowledge (bias?) required/used.
- ☐ Used since 2003, with first academic release in 2009

Liebschner et al (2017): "Polder maps: improving OMIT maps by excluding bulk solvent.", *Acta D* 73, 148

BUSTER: Is there any any ligand bound?

GΦL
LigandDetectionModes

☐ Favorite?
🏠 🗑️ 🔑 ⭐ Attachments 🖨️

(You are *AnonymousGnome*)

Content:

- [Introduction](#)
- [Unknown location](#)
- [Known location](#)
- [Caveats](#)
- [Summary](#)

Introduction

Apart from the standard procedures for detecting ligands (difference Fourier maps), BUSTER has one particular feature that needs a bit further explanation - to explain what it can do, what it can't do and what potential bias it might introduce.

This feature is triggered by the `-L` and `-Lpdb` flags to the refine command:

```
% refine -p some.pdb -m some.mtz -L ...
- or -
% refine -p some.pdb -m some.mtz -Lpdb bindingsite.pdb ...
```

It will treat a certain region of the model differently during the last big cycle of refinement: that region will neither contain an atomic model nor a contribution from the bulk solvent. However, if there is some electron density present in this region, it should show up in difference (Fo-Fc) maps as strong positive density. The interpretation, what this electron density might represent (atomic model, bulk solvent or a mixture) is up to the user.

Here we're going to explain the typical usage of this BUSTER feature, their assumptions and caveats.

Unknown location

The least biased assumption is that there might be some ligand bound, but its location is unknown. In that case one would use the

- ❑ Unbiased towards “expected” binding (just because the crystal was soaked/co-crystallised with compound doesn’t mean it actually is bound)
- ❑ Unbiased towards binding site (there can be unexpected or new binding sites)
- ❑ Unbiased towards binding pose

BUSTER wiki

www.globalphasing.com/buster/wiki/

Difference Fourier maps

- **Anomalous:**

- distinguish ion from solvent
- help difficult sequencing (Cys/Met marker)
- help placing of compound
- automatically generated and analysed in BUSTER

- **Early-Late:**

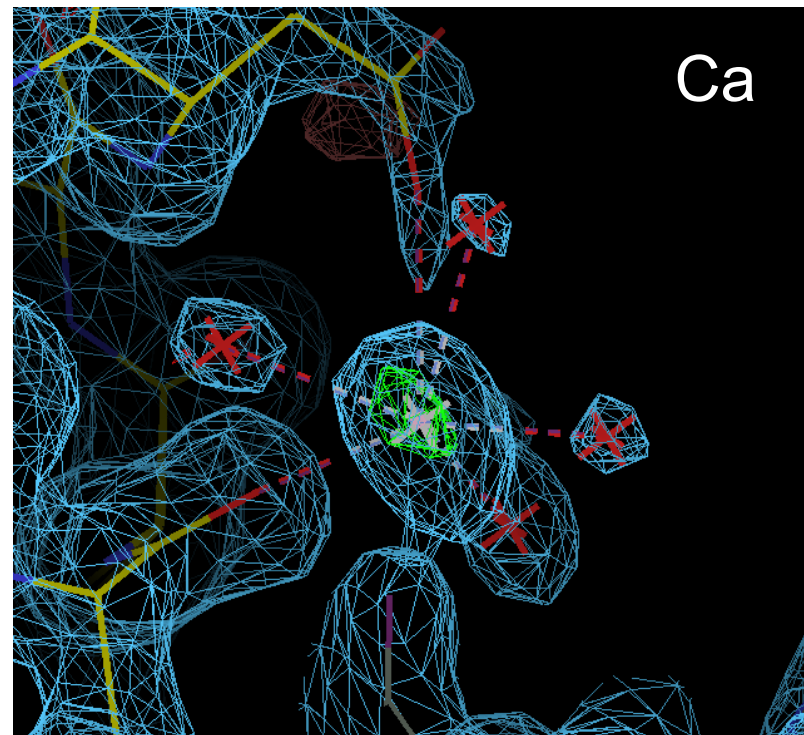
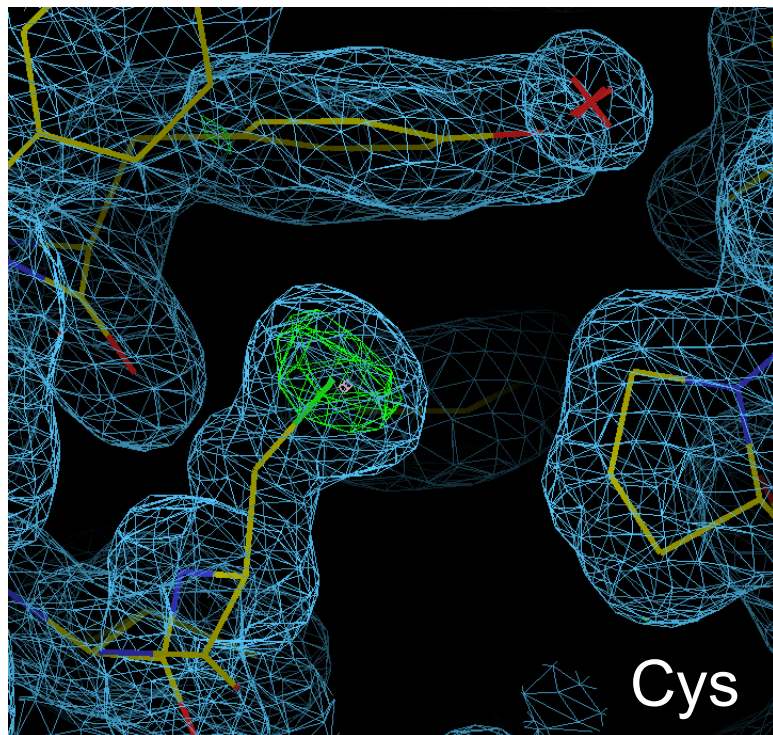
- show radiation damage effects
- automatically prepared in autoPROC - then generated and analysed in BUSTER

- **Fo-Fo:**

- (Compound - Apo) to show compound

```
diff_fourier -h
```

Difference Fourier maps 1: ANO



Crystal structure of a DNA methyltransferase 1 associated protein 1 (DMP1) from Homo sapiens at 1.45 Å resolution, JCSG (2013)

4IEJ: anomalous Fourier map at 3.5 rms

Wavelength = 0.9795 Å

$f'(S)=0.23$

$f'(Ca)=0.56$

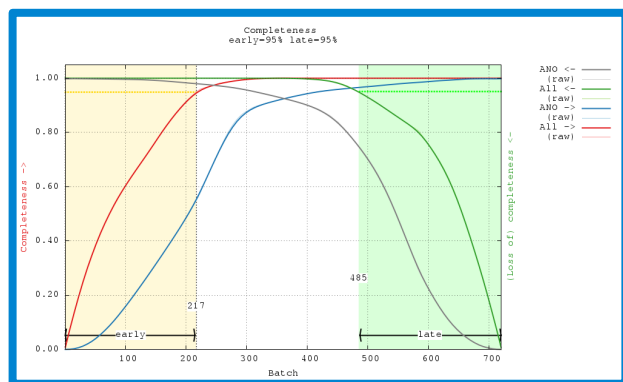
Low dose (rate), high multiplicity, fine-slicing

... and don't lose anomalous data columns in output reflection data

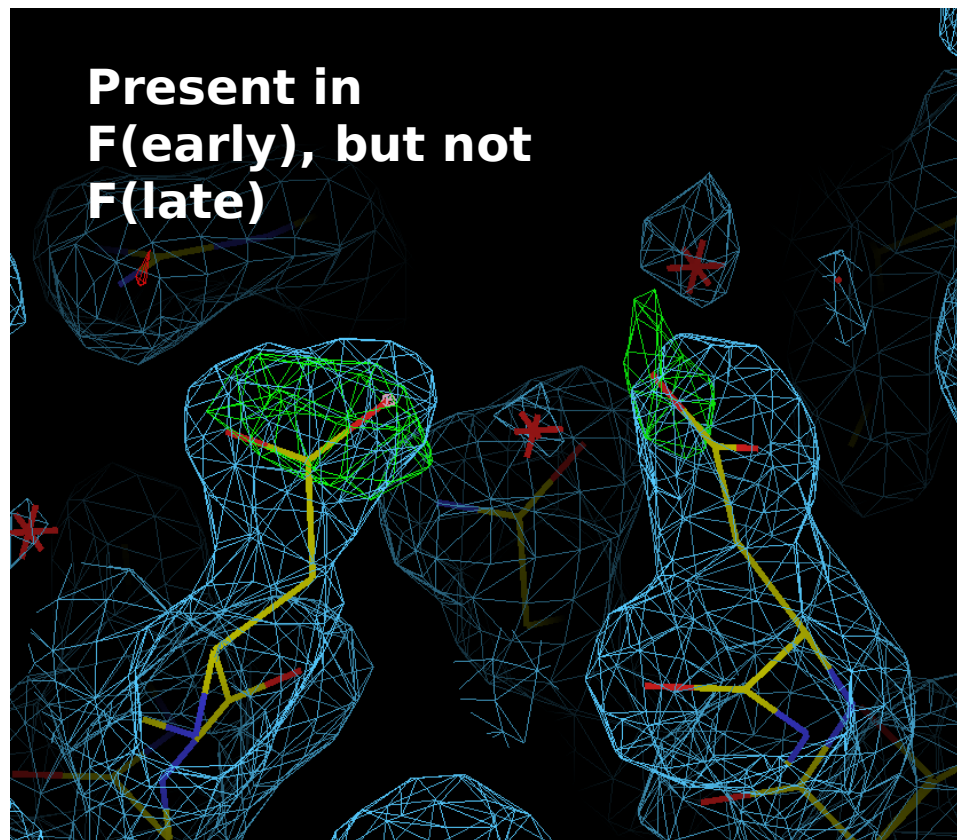
Difference Fourier maps: $F(\text{early}) - F(\text{late})$ detecting/describing radiation damage

Typical decarboxylation of ASP/GLU residues

Normal mFo-DFc maps will show negative peaks.

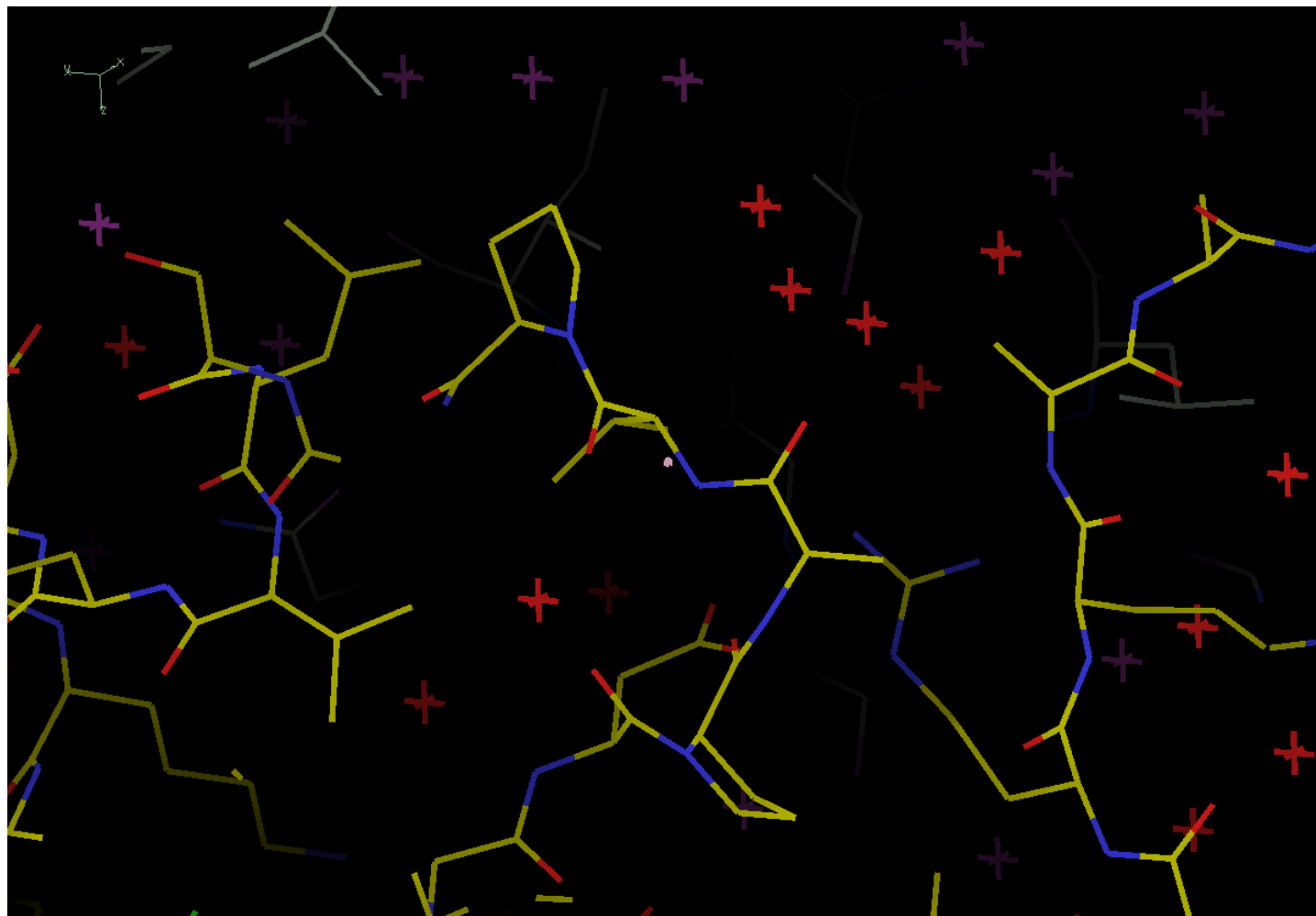


low dose, high-multiplicity



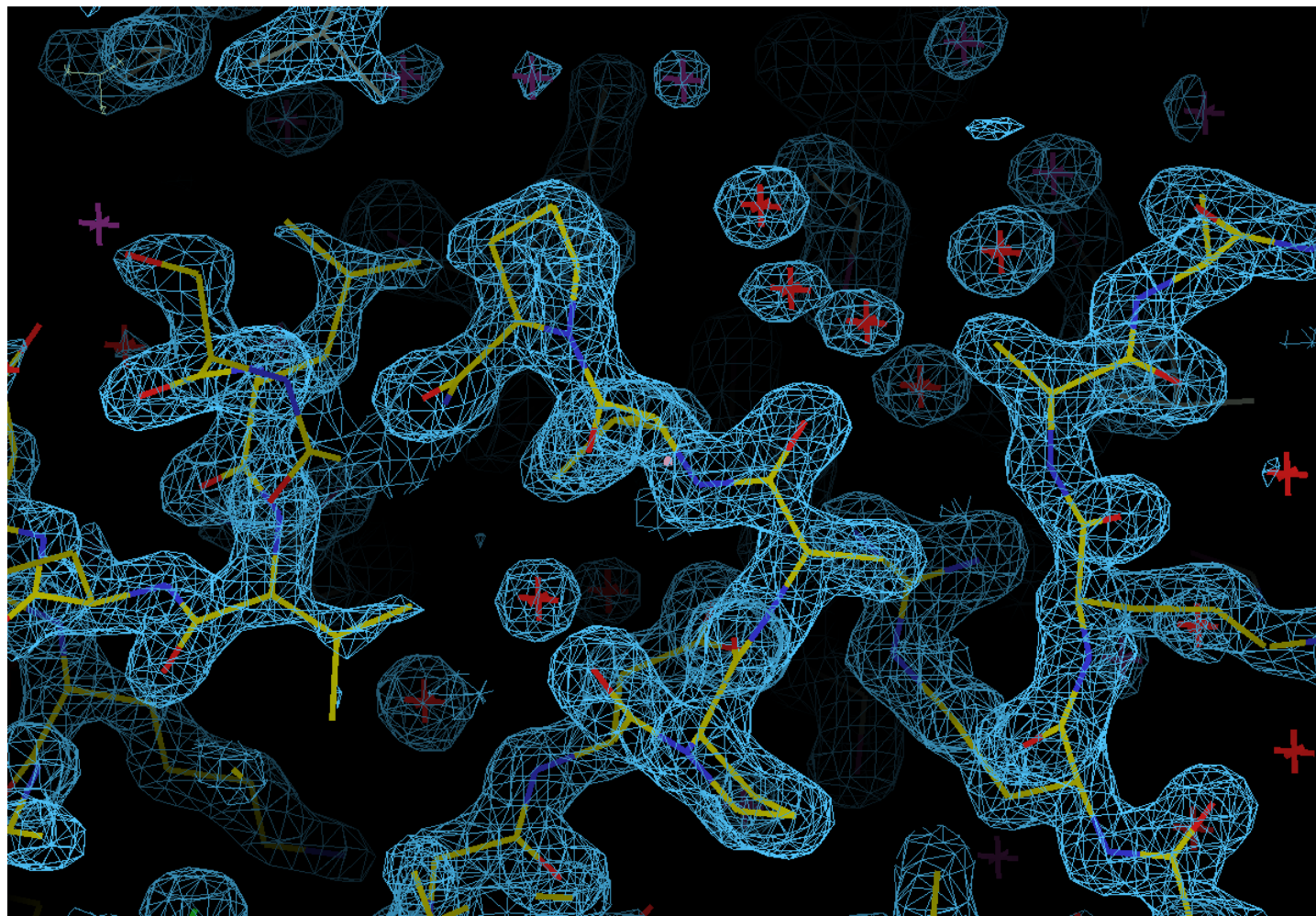
A novel mode of Gleevec binding is revealed by the structure of spleen tyrosine kinase. Atwell, S. et al. (2004) J.Biol.Chem. 279, 55827-55832.

F(early)-F(late): water in crystal contact



4LER

F(early)-F(late): water in crystal contact

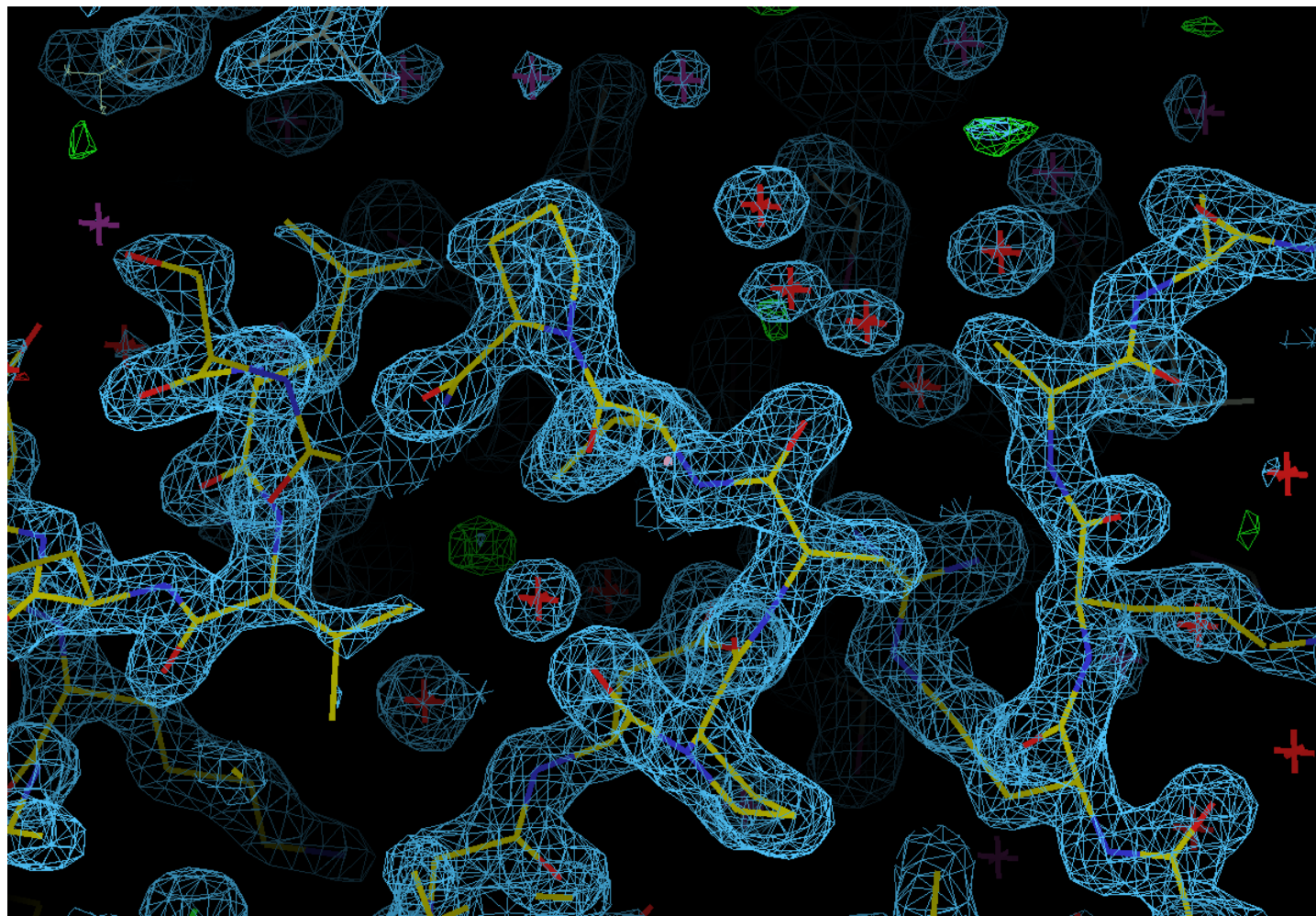


4LER

2Fo- F_c

1.5 rms

F(early)-F(late): water in crystal contact



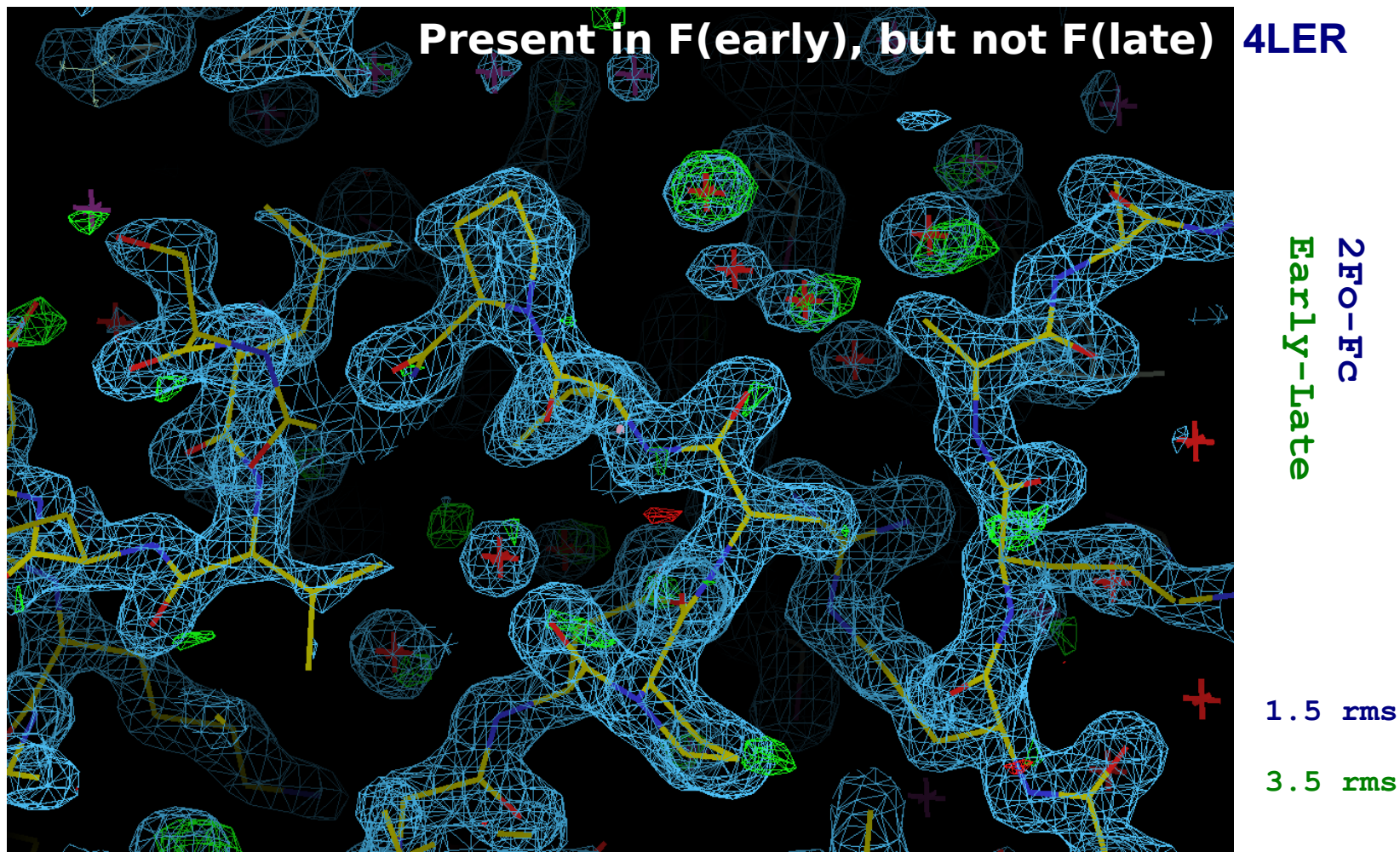
4LER

2Fo-FC
Fo-FC

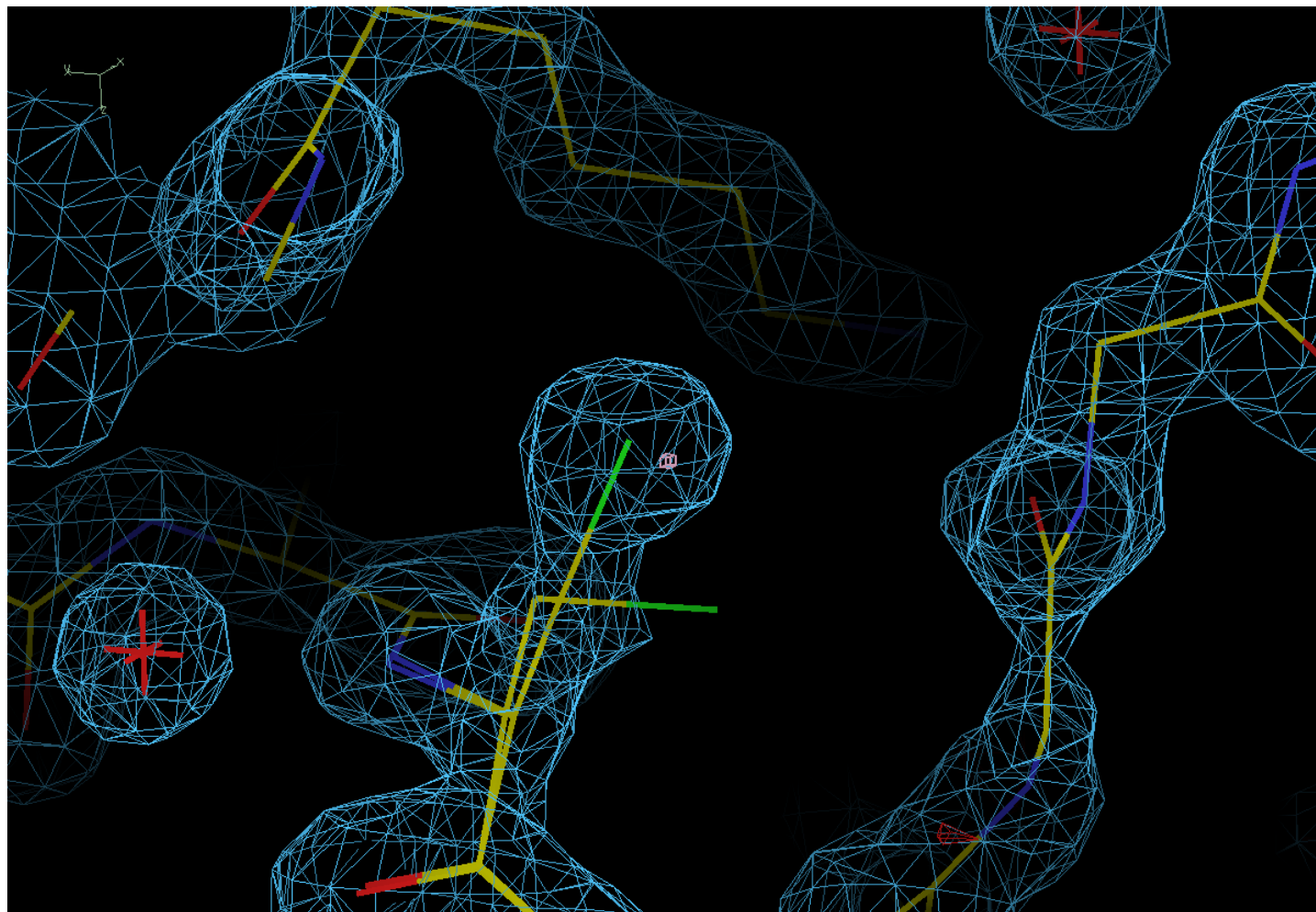
1.5 rms

3.5 rms

F(early)-F(late): water in crystal contact



F(early)-F(late): alternate conformation?



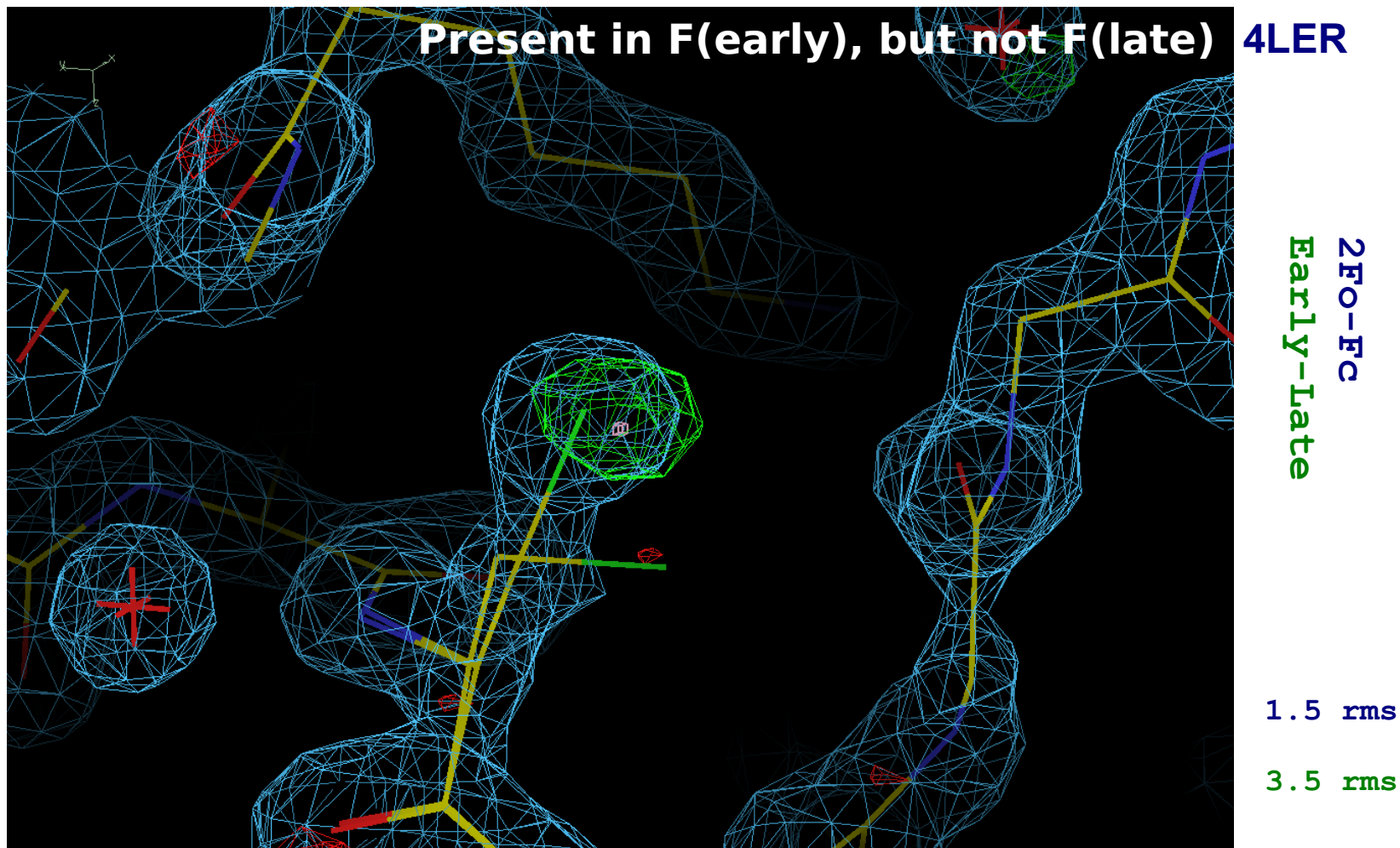
4LER

2Fo-Fc
Fo-Fc

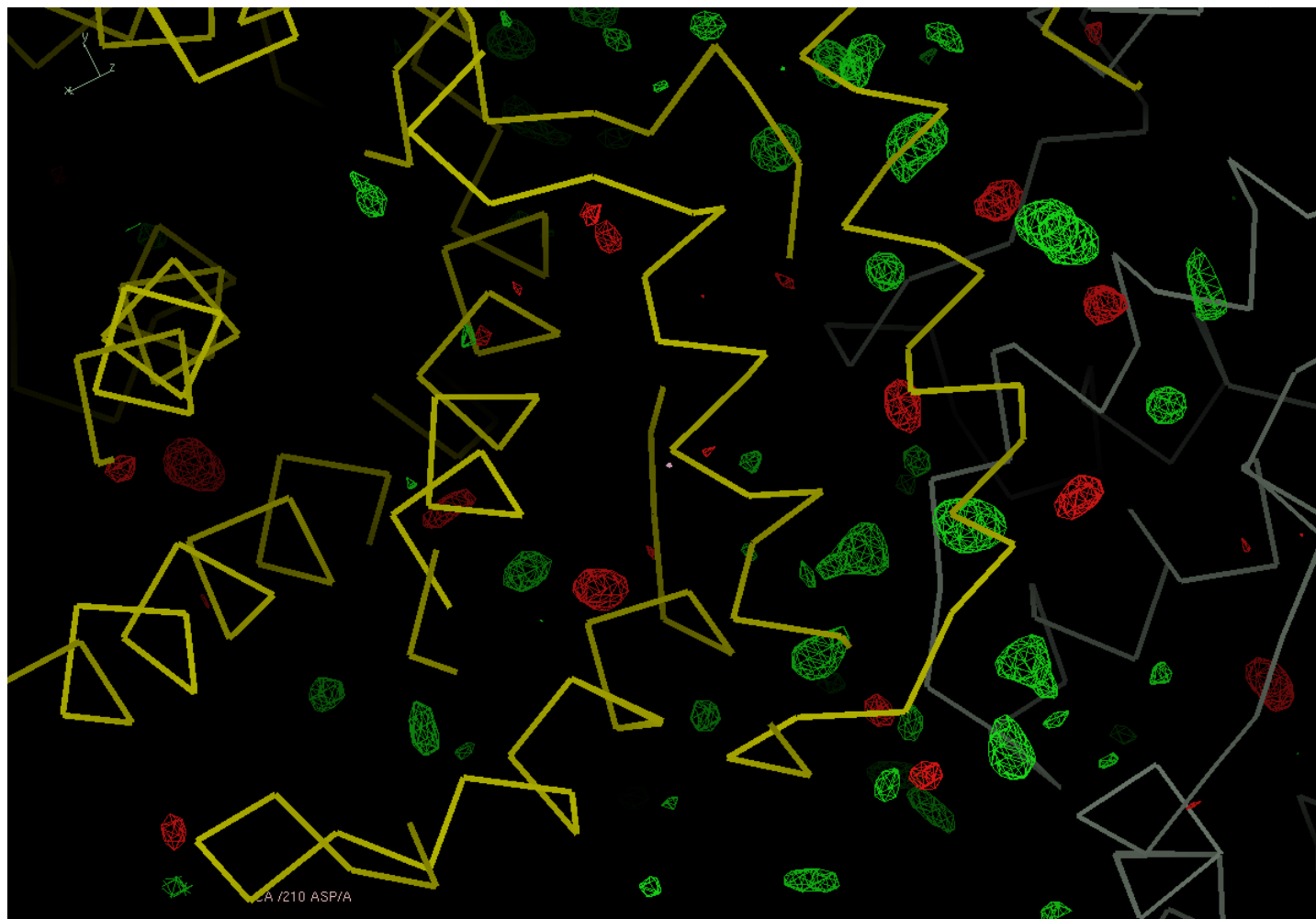
1.5 rms

3.5 rms

F(early)-F(late): alternate conformation?



F(early)-F(late): large-scale movements

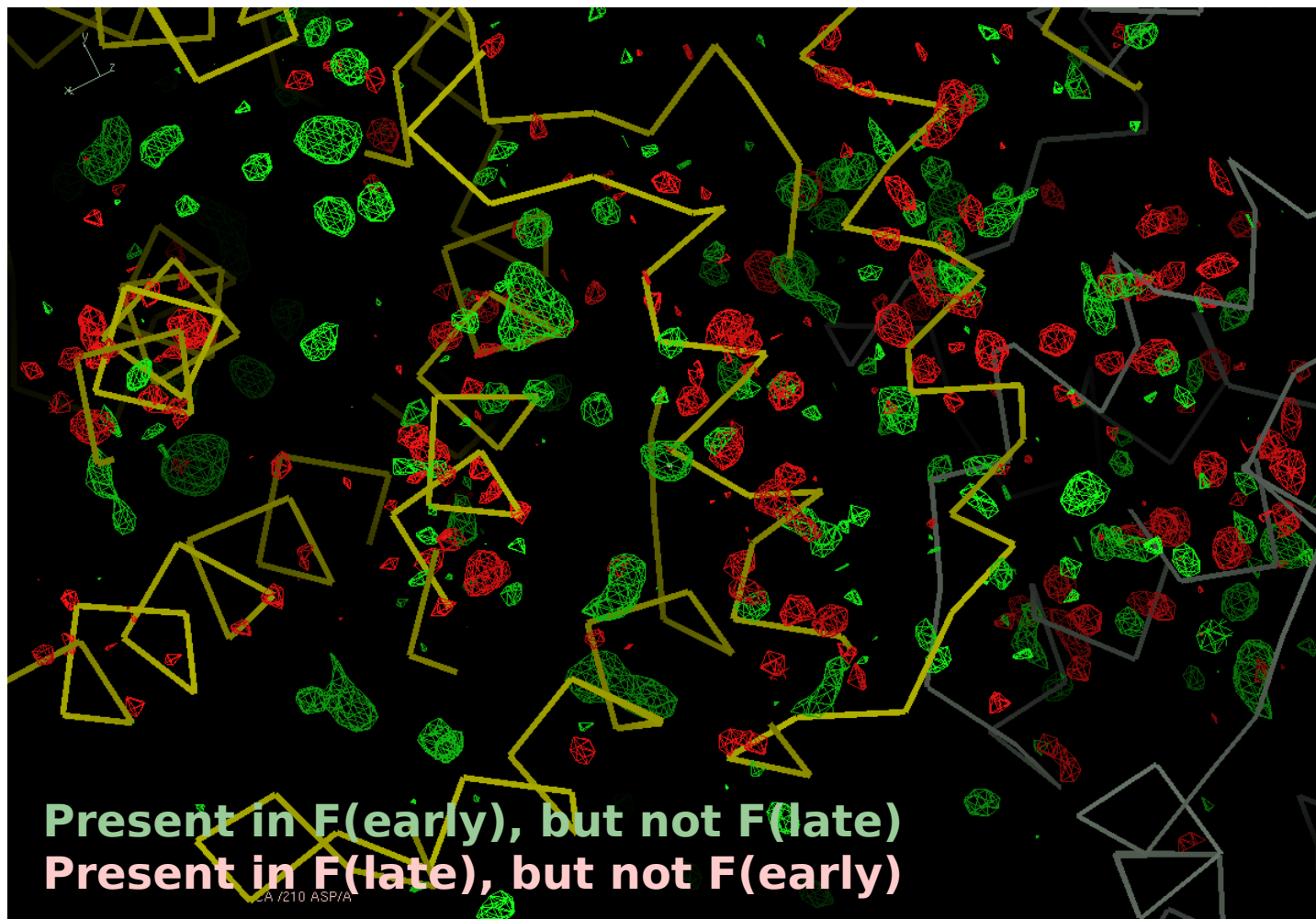


4LER

Fo-Fc

3.5 rms

F(early)-F(late): large-scale movements

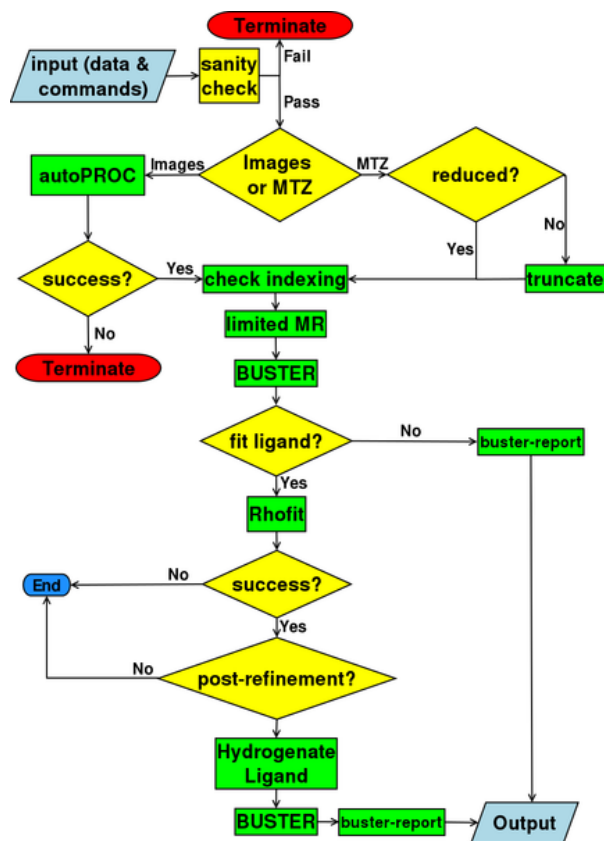


4LER

Early-Late

3.5 rms

Pipedream - combining all steps



Expert system that links data processing (**autoPROC**), initial refinement & ligand detection (**BUSTER**), ligand fitting (**Rhofit**), final refinement (**BUSTER**) and analysis (**buster-report**) - including pepflip/SideAide option.

Used in large-scale operations within commercial and academic environments.

For a given crystal form:

- 1) Apo model
- 2) Dataset (directory with images)
- 3) Ligand dictionary (CIF)

=> refined complex structure

<https://www.globalphasing.com/buster/wiki/index.cgi?PipedreamMainpage>

-
- ❑ **Careful processing of data** can be crucial (not everything is Lysozyme)
 - A **good collection strategy** (that adapts to crystal quality, SG, orientation, detector, beamline, goniostat) will result in better data, better density and better models
 - ❑ Collect all **prior knowledge**
 - Geometric restraints (good ligand dictionaries)
 - Similarity restraints (NCS, targeting)
 - Occupancy
 - Correct formfactors (fluorescence scan)
 - ❑ Refine to **convergence**
 - Otherwise difference maps can be hard to interpret
 - And statistics like R/Rfree are (fairly) meaningless
 - ❑ What happened to crystal during data collection is still (very) important at this stage
 - **Radiation damage**: F(early)-F(late) maps
 - **Anisotropy** (STARANISO)
 - ❑ Often the defaults (quick click) work well ... but to get the correct structure interpretation the **best/correct parametrisation and refinement is important**
-

Acknowledgements

- ❑ Global Phasing, Cambridge (UK):
 - Gérard Bricogne, Leigh Carter, Claus Flensburg, Rasmus Fogh, Peter Keller, Wlodek Paciorek, Andrew Sharff, Ian Tickle
 - Oliver Smart, Thomas Womack, Eric Blanc, Gwyndaf Evans, Pietro Roversi, John Irwin, Eric de la Fortelle, Marc Schiltz
- ❑ Vizier Project FP6
- ❑ JCSG
- ❑ W. Kabsch, K. Diederichs
- ❑ P. Evans
- ❑ CCP4
- ❑ CCDC
- ❑ Martin Field, Alexei Vagin, Garib Murshudov, openbabel developers
- ❑ R. Joosten & PDB_REDO team
- ❑ Global Phasing Consortium members
- ❑ ... many, many users & collaborators

<https://www.globalphasing.com/>

<https://www.globalphasing.com/buster/wiki/> (Tutorials etc)