GΦL
Global Phasing Limited

Refinement:
The BUSTER perspective

Clemens Vonrhein
Global Phasing Ltd.

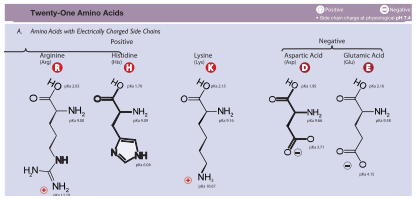
DLS/CCP4 2024

GΦL
Global Phasing Limited

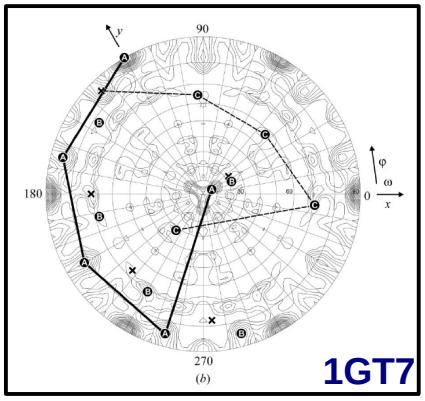
Refinement

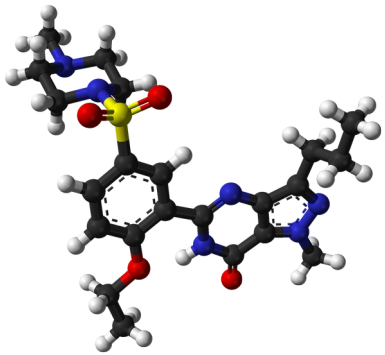
- Combines prior knowledge and observed (X-Ray) data

- Chemistry (bonds, angles, ...)
  - Protein (Engh & Huber)
  - Compound/Ligand
  - ...



- Similarity
  - Within the crystal (NCS)
  - To other structures ("targeting")
- Occupancy, disorder (alternate conformations), ...

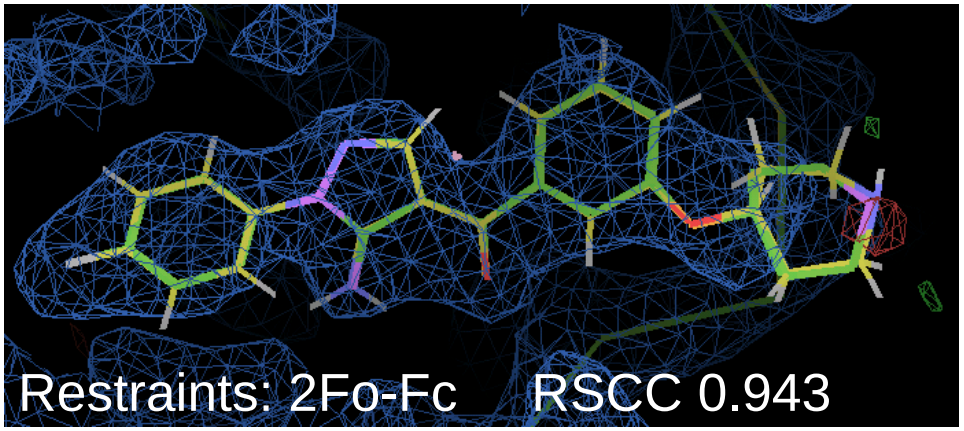




GΦL
Global Phasing Limited

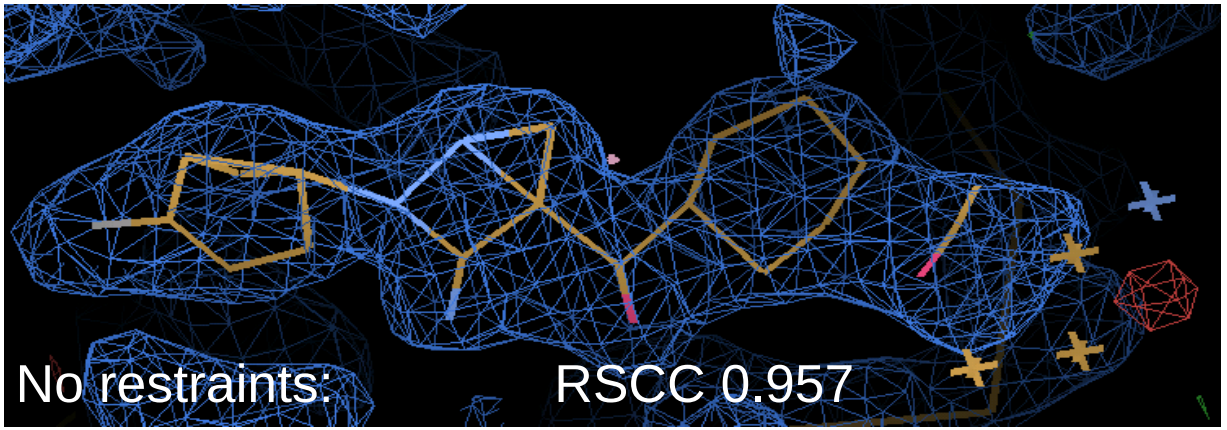
Prior knowledge is essential

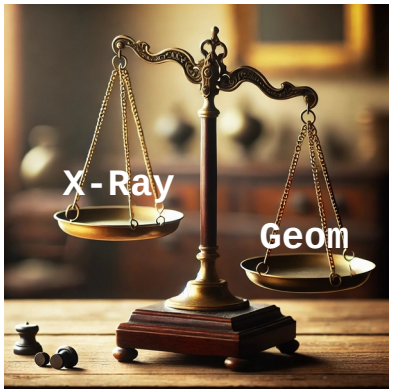
2bal kinase 2.1Å resolution.



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 that also follows prior chemical knowledge

→ pick your quality criteria carefully!

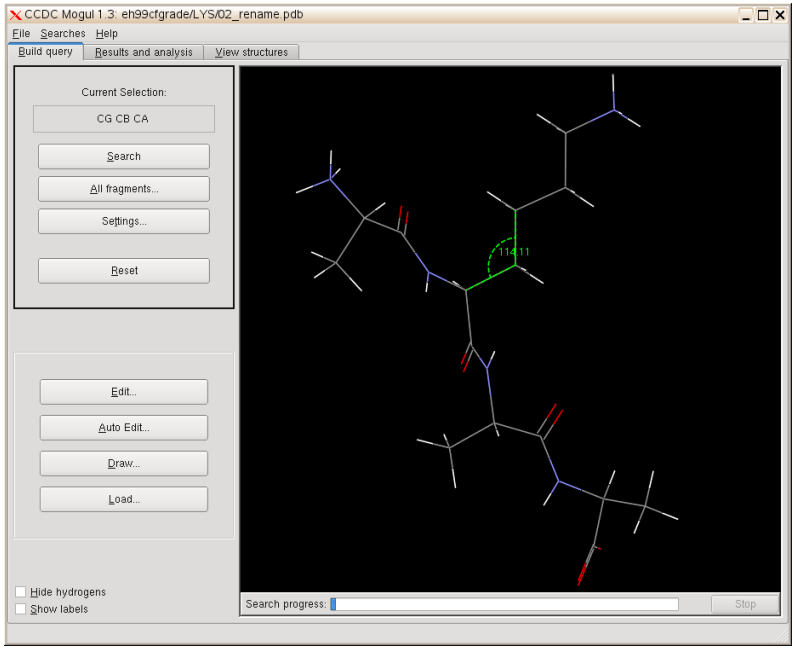


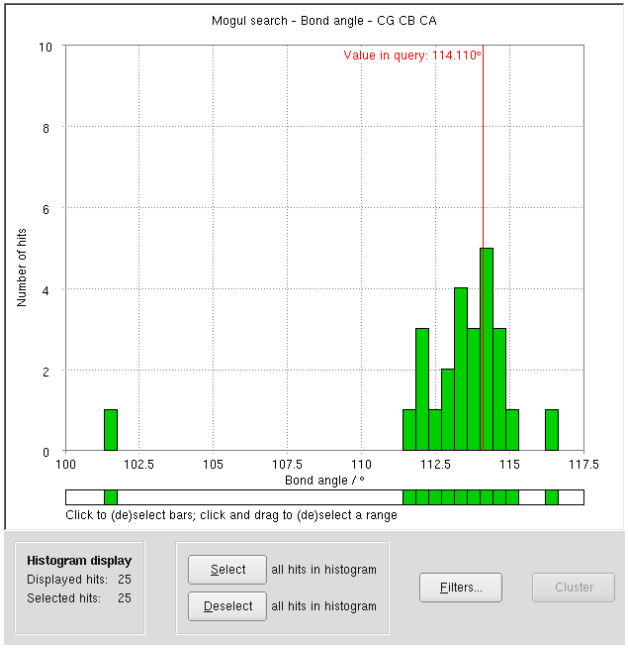


RSCC = real-space correlation coefficient

Ligand dictionaries based on CSD
grade.globalphasing.org

- Grade (2011): use CCDC **mogul** program to survey **CSD**





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

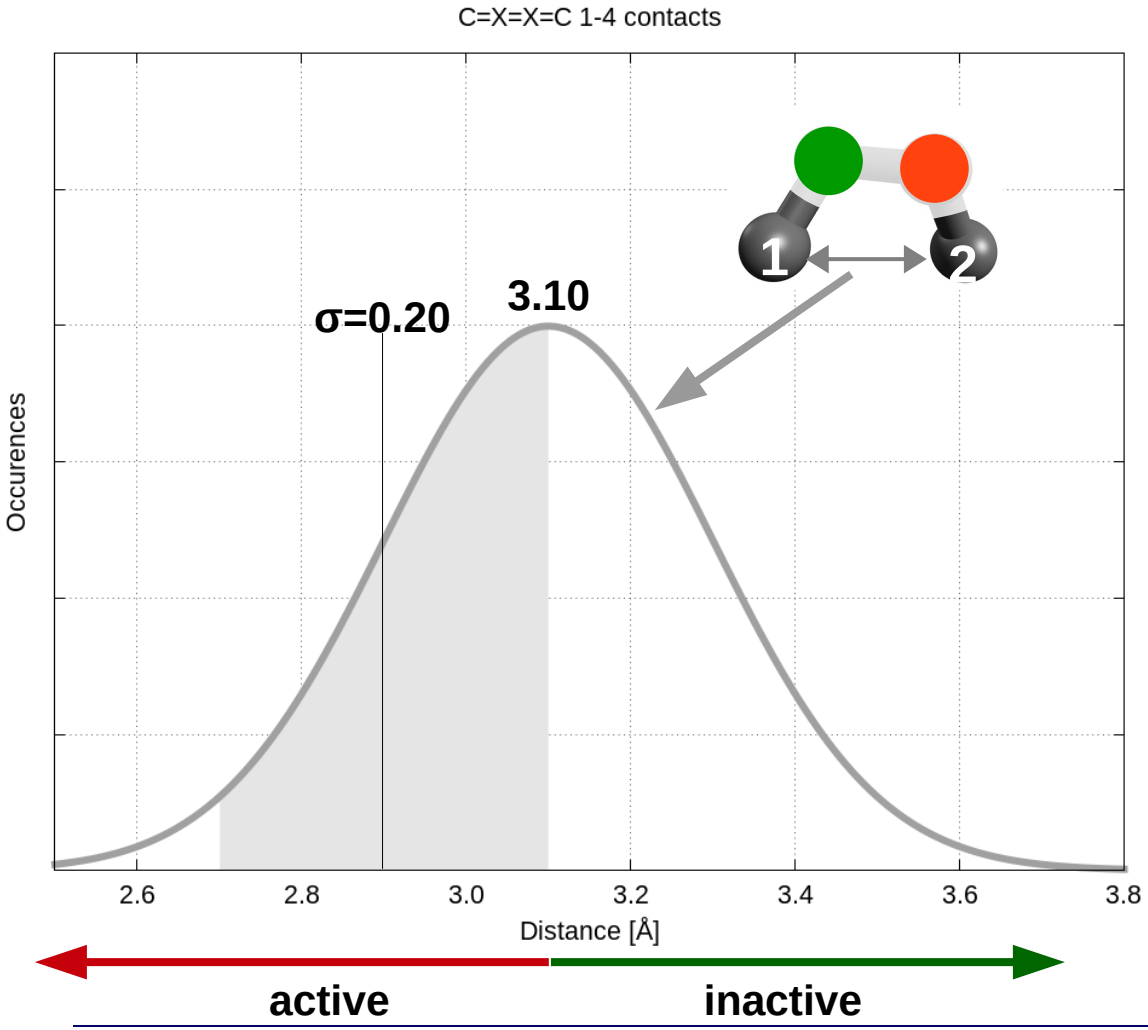
Grade2 (rewrite, July 2021):
https://www.globalphasing.com/buster/manual/grade2/manual/

GΦL
Global Phasing Limited

Restraints - not just for ligands

- Standard geometry restraints for bonded atoms
  - BOND (1-2)
  - ANGLE (1-3)
  - TORSION (1-4)
  - CHIRAL
  - PLANE
- Non-bonded atoms
  - anti-bumping restraints
- B-factor restraints
  - bonded atoms
  - isotropic/anisotropic B-factor
- Similarity restraints
  - NCS
  - targeting
- *Conformational restraints*
  - *Ramachandran*
  - *rotamer*

Contact restraints for 1-4 connected atoms - V1



only first and last atoms of 1-4 (torsion) contact taken into account

BUSTER pre-20190923:

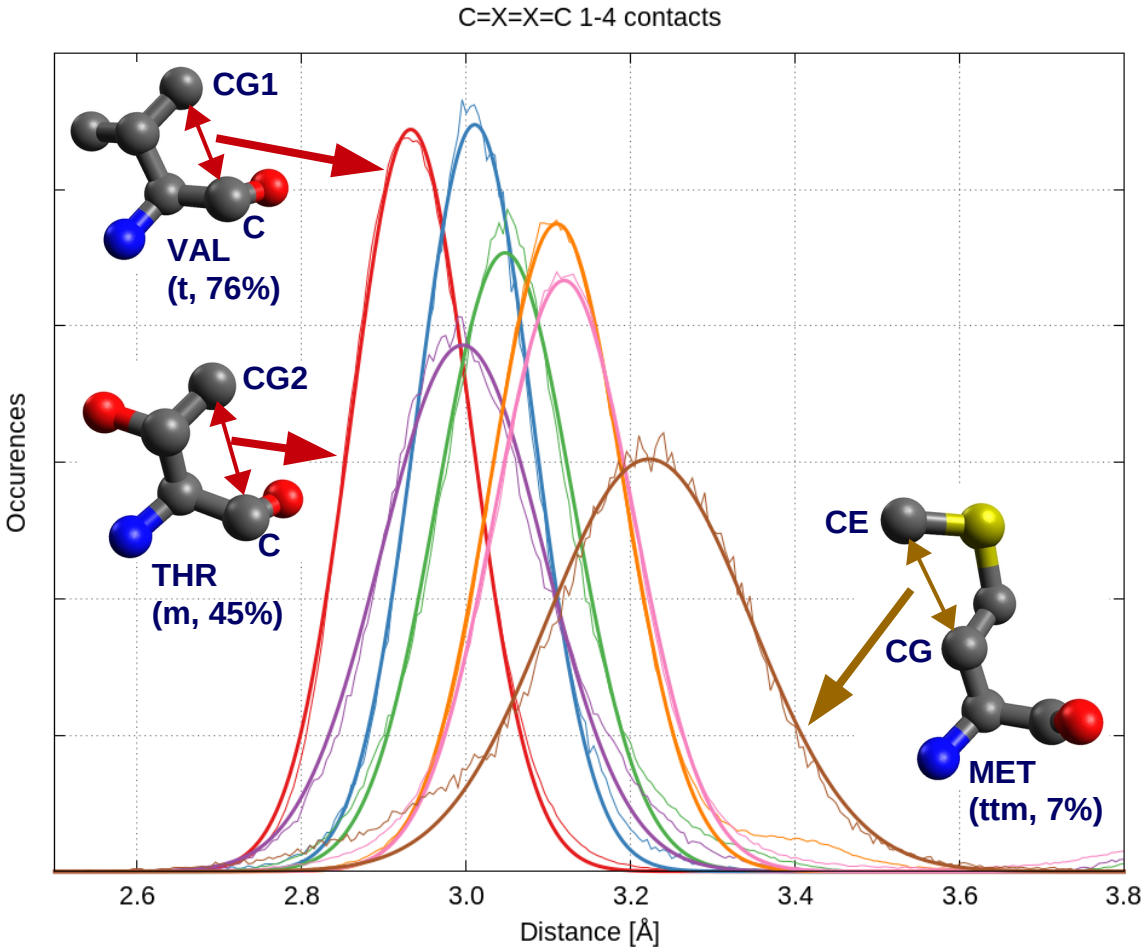
$vdW_1+vdW_2+Dinc_1+Dinc_2$

Element	Dinc
C	-0.15
N	-0.10
O	-0.10
X	-0.15

fixed sigma (0.20)

GΦL
Global Phasing Limited

Contact restraints V2



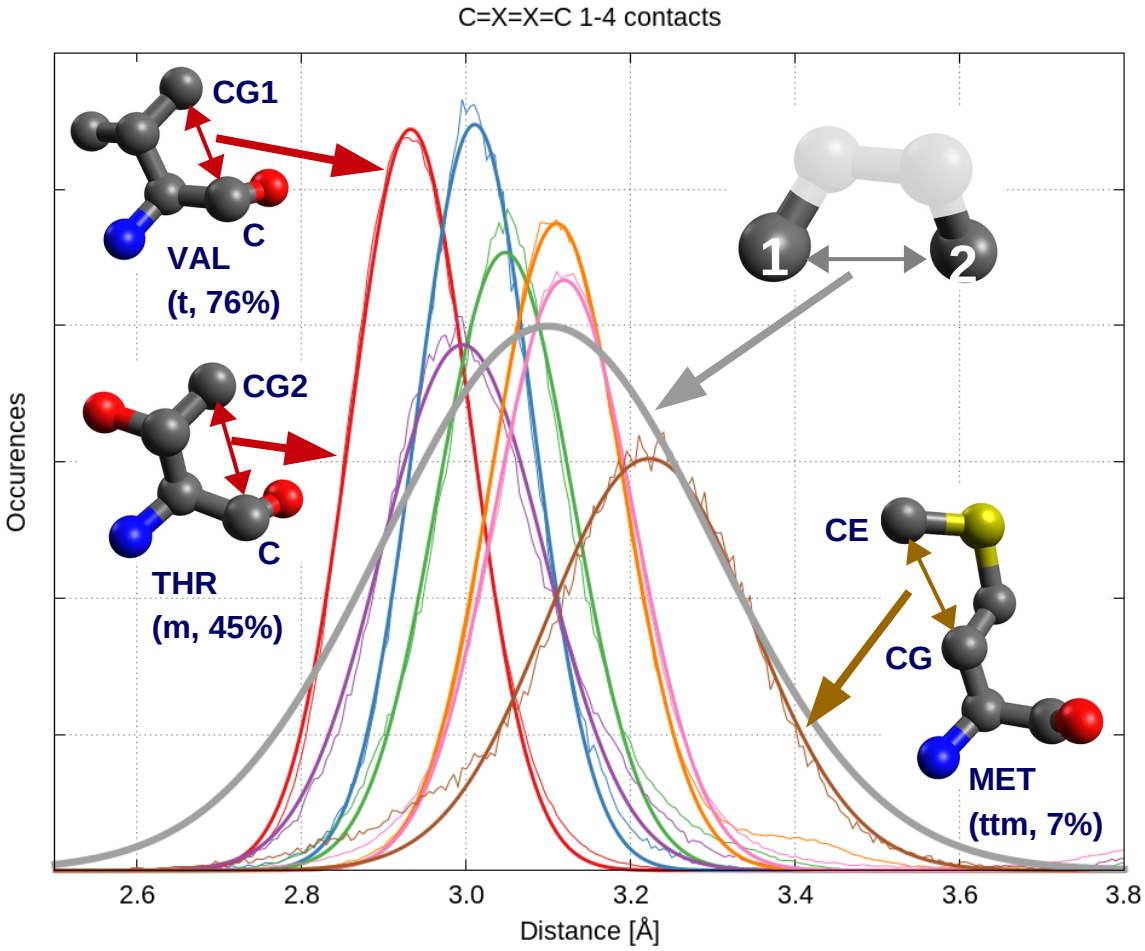

```
C=CH1=CH1=CH3     (2.93 σ=0.073, N=2191809)
C=CH1=CH1=CH2     (3.01 σ=0.071, N= 121226)
C=CH1=CH2=C       (3.05 σ=0.082, N= 519830)
C=CH1=CH2=CR5     (3.00 σ=0.099, N= 232154)
CH1=CH2=CH1=CH3   (3.11 σ=0.080, N=1402344)
CH3=CH1=CH2=CH3   (3.12 σ=0.085, N= 910621)
CH2=CH2=S=CH3     (3.22 σ=0.125, N= 121100)
```

BUSTER ≥ 20190923:

- 1-(2,3)-4 atom-typing DB
- > 30000 PDB structures (2015-2018)
- Rfree <= 25%
- better than 2 Å
- collect minimum 1-4 distance for a given atom
- use CCP4 atom typing of all 4 atoms for classification

GΦL
Global Phasing Limited

Contact restraints V1 vs V2




```
C=CH1=CH1=CH3     (2.93 σ=0.073, N=2191809)
C=CH1=CH1=CH2     (3.01 σ=0.071, N= 121226)
C=CH1=CH2=C       (3.05 σ=0.082, N= 519830)
C=CH1=CH2=CR5     (3.00 σ=0.099, N= 232154)
CH1=CH2=CH1=CH3   (3.11 σ=0.080, N=1402344)
CH3=CH1=CH2=CH3   (3.12 σ=0.085, N= 910621)
CH2=CH2=S=CH3     (3.22 σ=0.125, N= 121100)
C=X=X=C           (3.10 σ=0.200            )
```

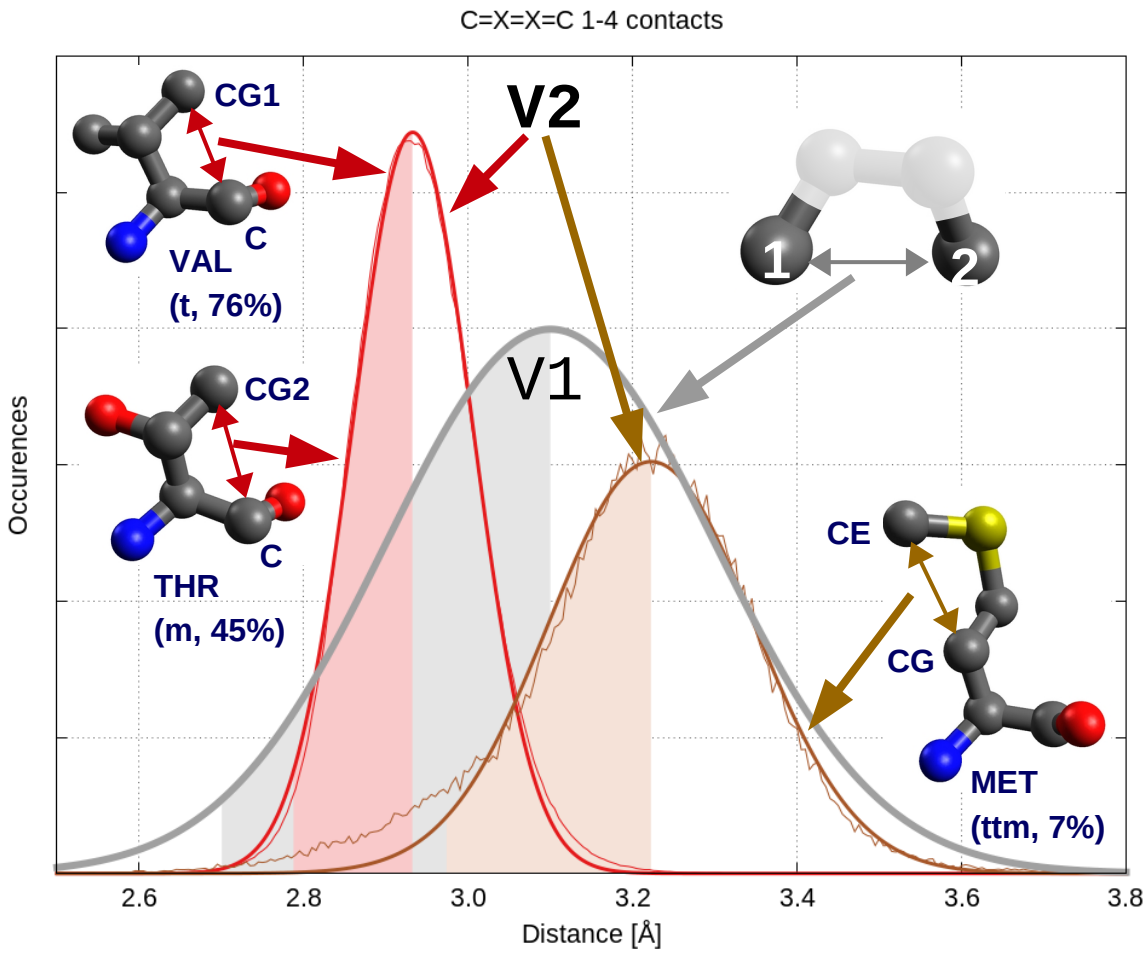
BUSTER pre-20190923:

$vdW_1+vdW_2+Dinc_1+Dinc_2$

Element	Dinc
C	-0.15
N	-0.10
O	-0.10
X	-0.15

GΦL
Global Phasing Limited

Contact restraints V1 vs V2



BUSTER pre-20190923:

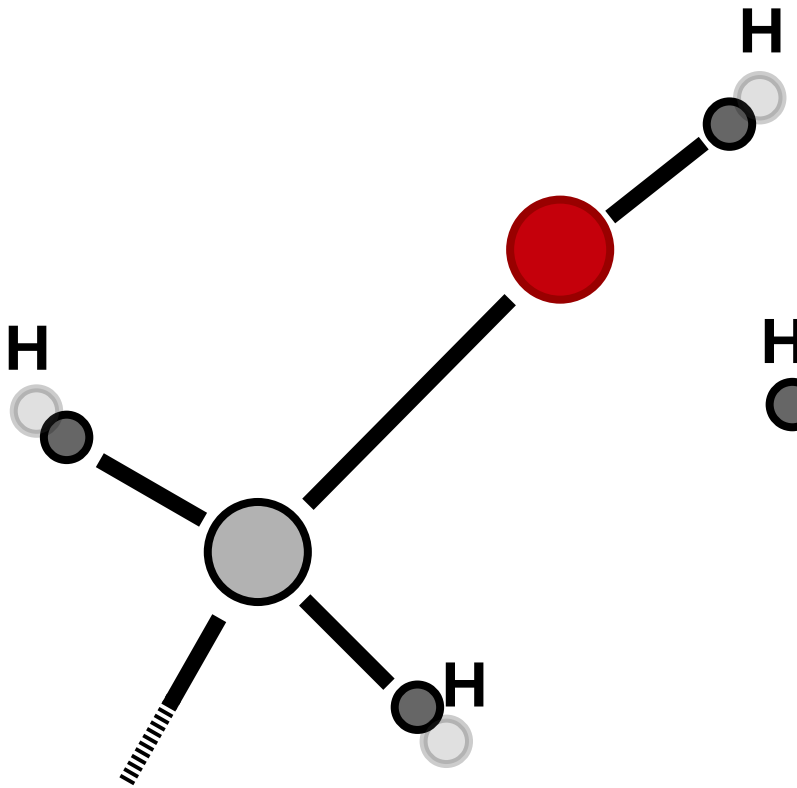
$vdW_1+vdW_2+Dinc_1+Dinc_2$

Element	Dinc
C	-0.15
N	-0.10
O	-0.10
X	-0.15

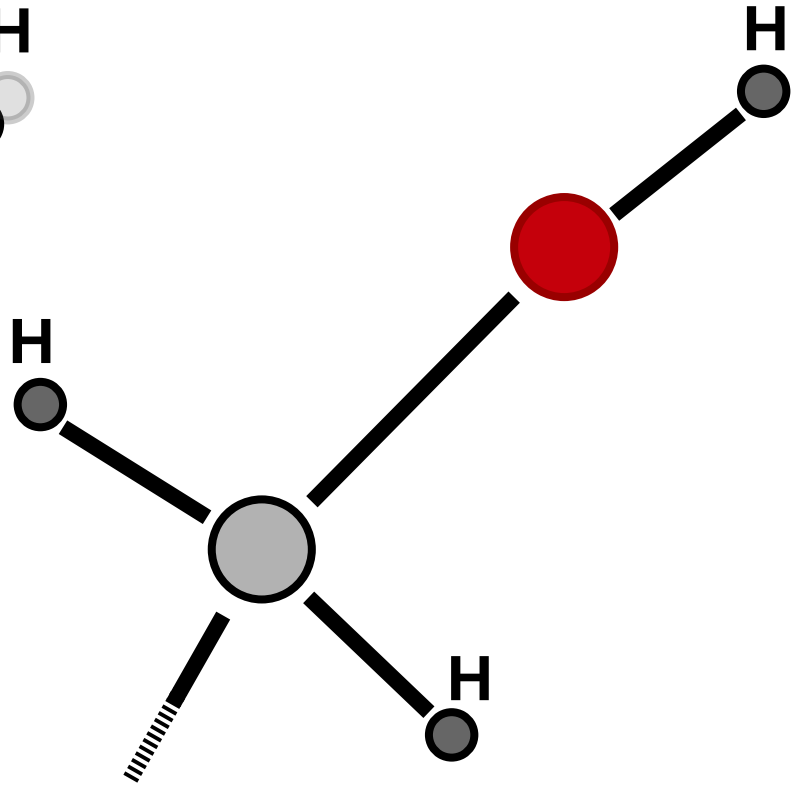
GΦL
Global Phasing Limited

Hydrogen handling in X-Ray refinement

e^--cloud position



nuclear position

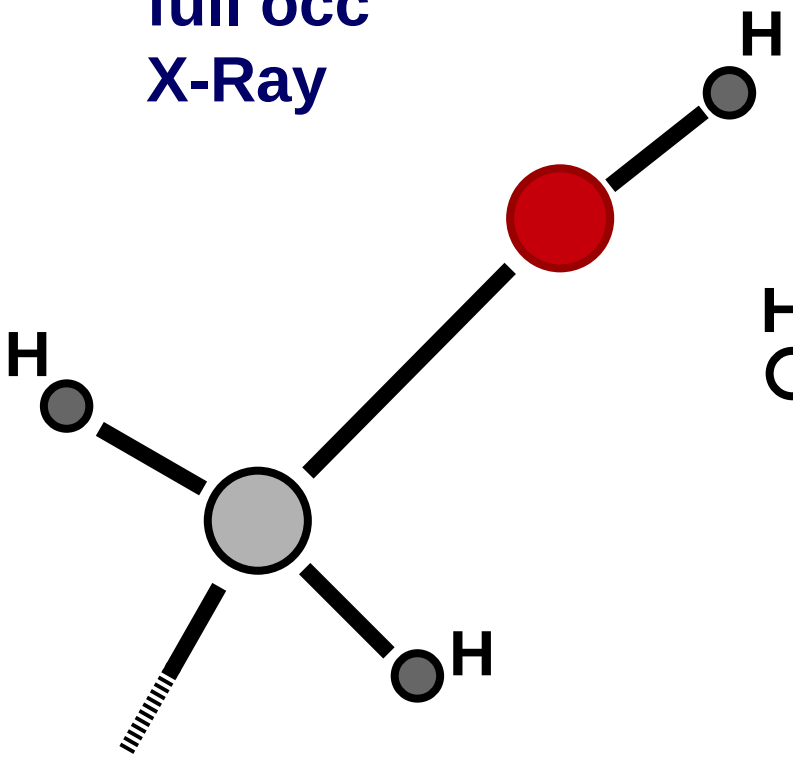


See e.g.: Williams et al (2018). Protein Science, 27, 293-315.

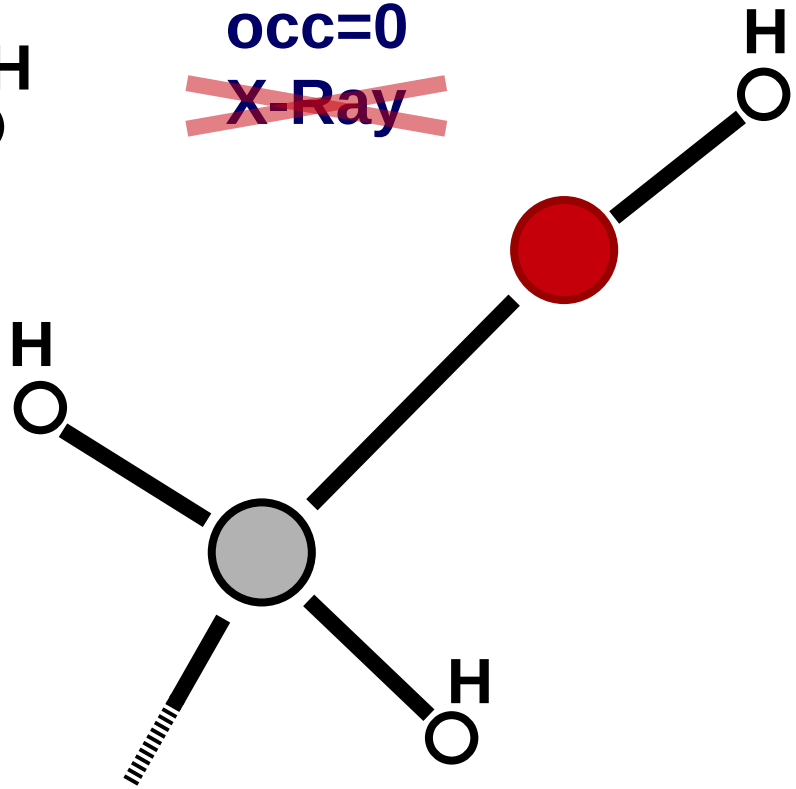
GΦL
Global Phasing Limited

Hydrogen placement has consequences

e^--cloud position
full occ
X-Ray



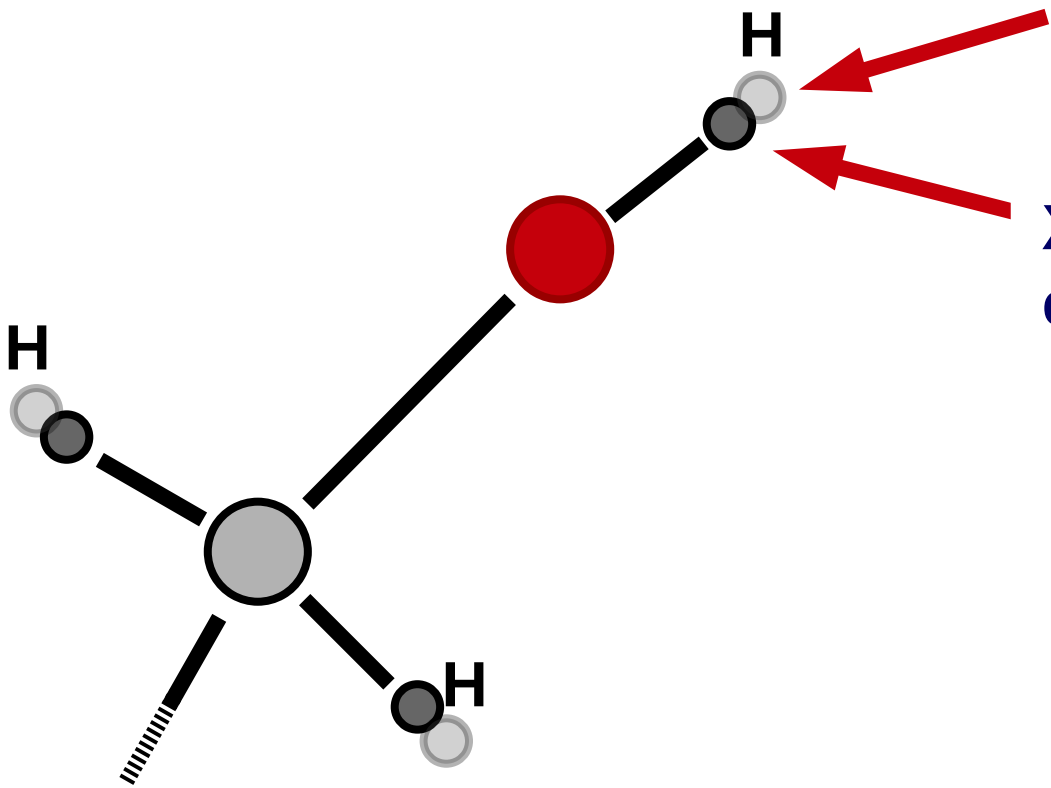
nuclear position
occ=0
~~**X-Ray**~~



BUSTER: hydrogens refined as any other atom - not “riding hydrogen” model

GΦL
Global Phasing Limited

Mixed-mode (hybrid) hydrogen model



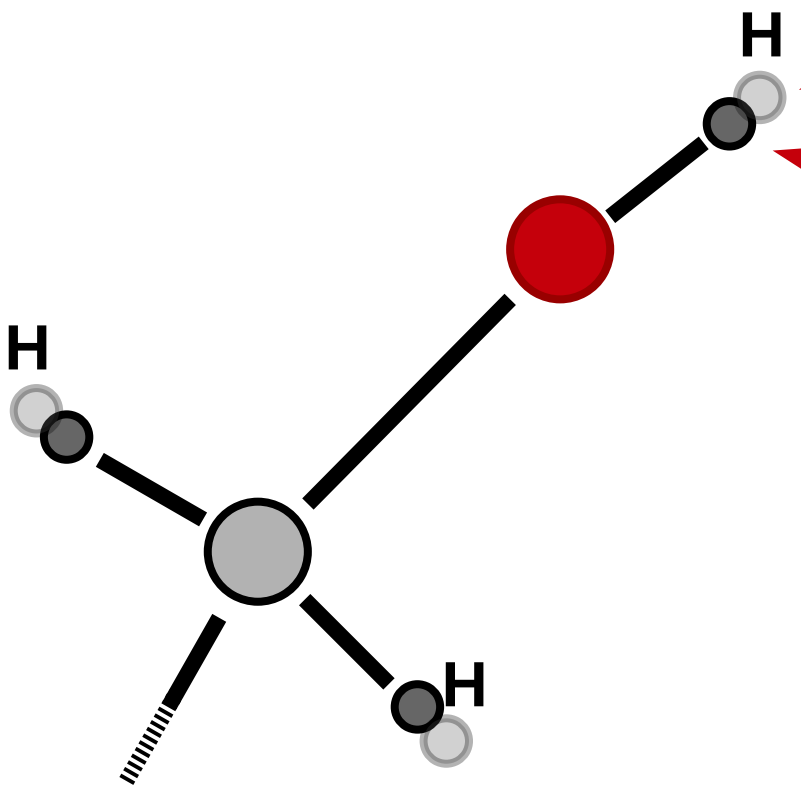
geometry:
nuclear position

X-Ray:
e^--cloud position

BUSTER: refine -M HydrogenHybridModel ...

GΦL
Global Phasing Limited

Mixed-mode (hybrid) hydrogen model



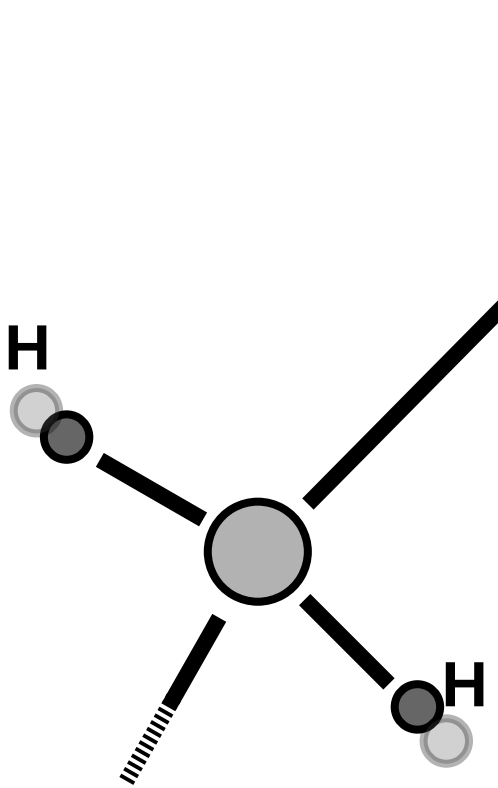
geometry:
nuclear position

X-Ray:
e^--cloud position

only real option here

GΦL
Global Phasing Limited

Mixed-mode (hybrid) hydrogen model



geometry:
nuclear position

X-Ray:
e^--cloud position

- Explicitely refined hydrogens
- Hydrogens in output model at nuclear position
- Validation needs to take hydrogen model into account

See also discussions in: Liebschner et al (2018). Acta D74, 800-813.

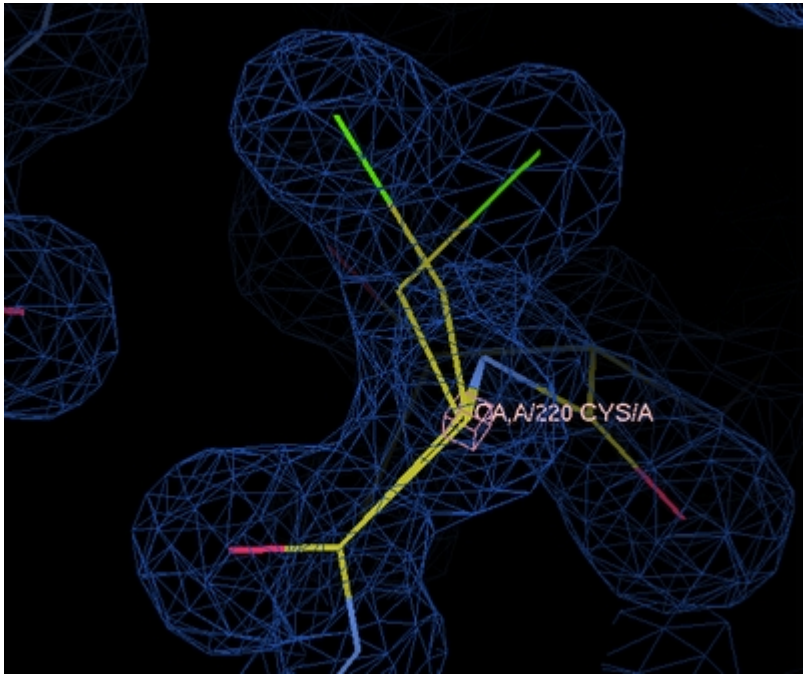
GΦL
Global Phasing Limited

Alternate conformations

Alternate conformation

- Atoms with altConf “A” see atoms without altConf or with same altConf “A”
- Atoms with different altConfs don’t see each another

- Set occupancies you want to refine to something not equal to 1.0, and then run
  - **pdb2occ** `–p model.pdb –o occ.gelly`
  - **refine** `–Gelly occ.gelly ...`
- If occ(A)+occ(B)=1 in input, **pdb2occ** will restrain occupancy sum (to 1)



Global Phasing Limited

Some notes about rotamers and outliers

- During validation, MolProbity rotamer analysis often used as a quality criteria for model
- 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

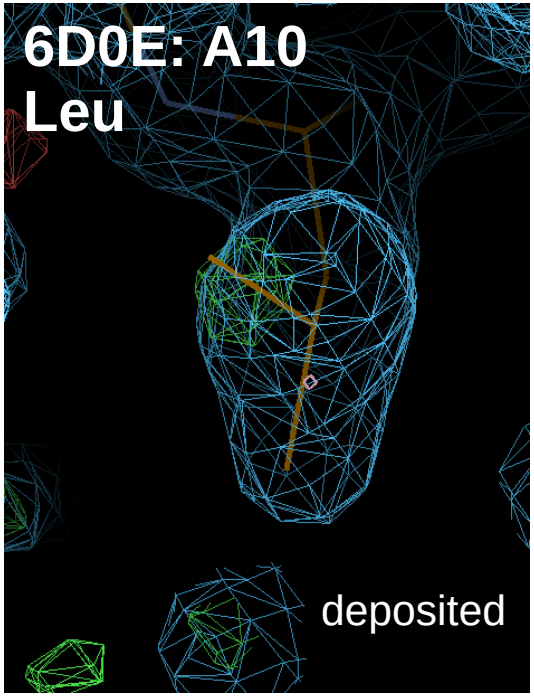
```
A   10  LEU:1.00:0.0:259.1:21.3:::OUTLIER:OUTLIER
A   19  LEU:1.00:0.1:309.3:303.1:::OUTLIER:OUTLIER
A   24  LEU:1.00:0.2:297.4:296.7:::OUTLIER:OUTLIER
A   45  LYS:1.00:0.1:197.6:122.2:308.4:165.6:OUTLIER:OUTLIER
A   50  ILE:1.00:0.0:239.0:174.6:::OUTLIER:OUTLIER
A   82 AVAL:0.33:0.2:320.6::::OUTLIER:OUTLIER
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?

GΦL
Global Phasing Limited

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

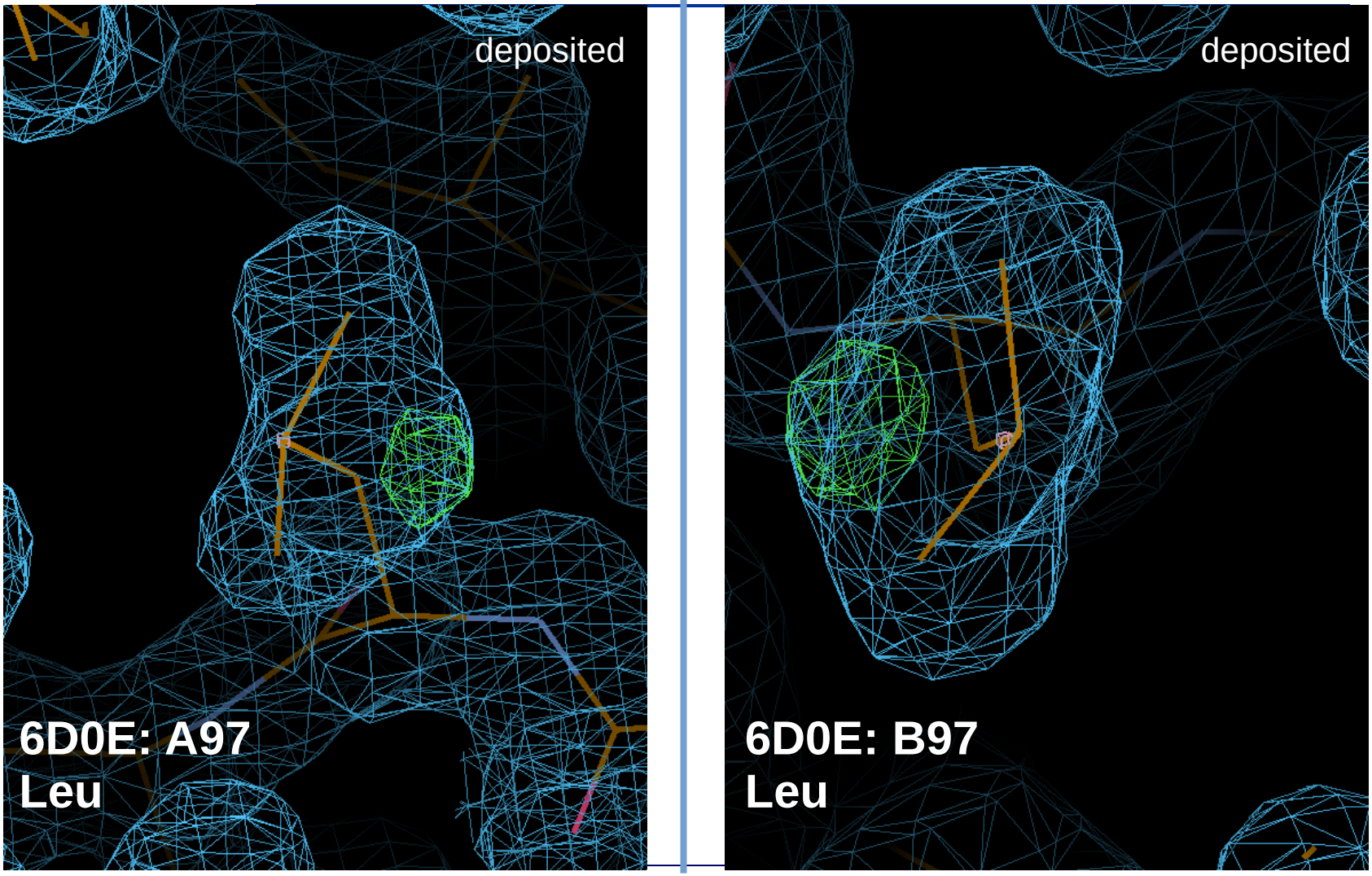


mp (2%): Allowed | **OUTLIER** | **mt (59%): favoured**

- 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

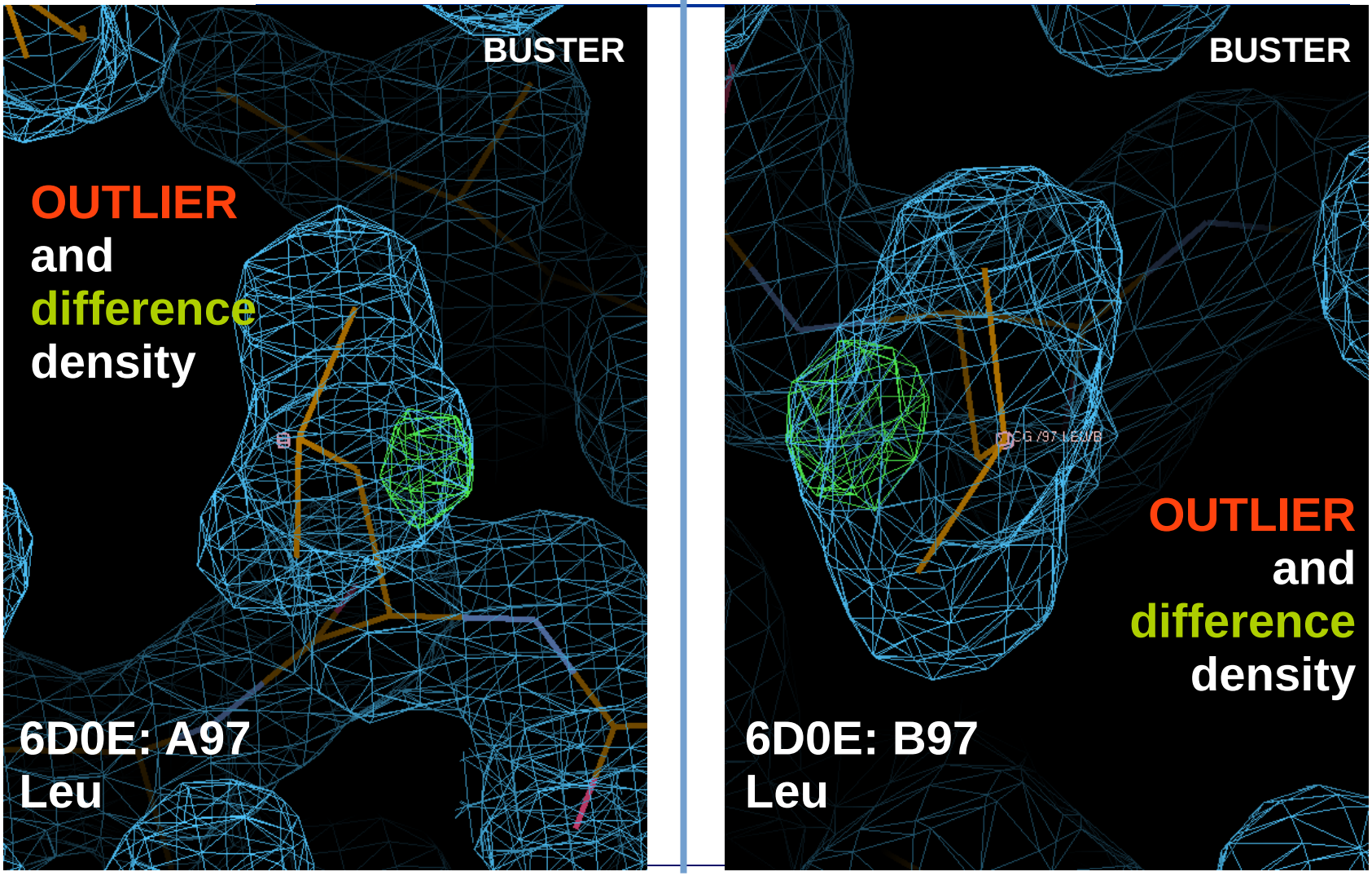
GΦL
Global Phasing Limited

Rotamer outliers are useful markers



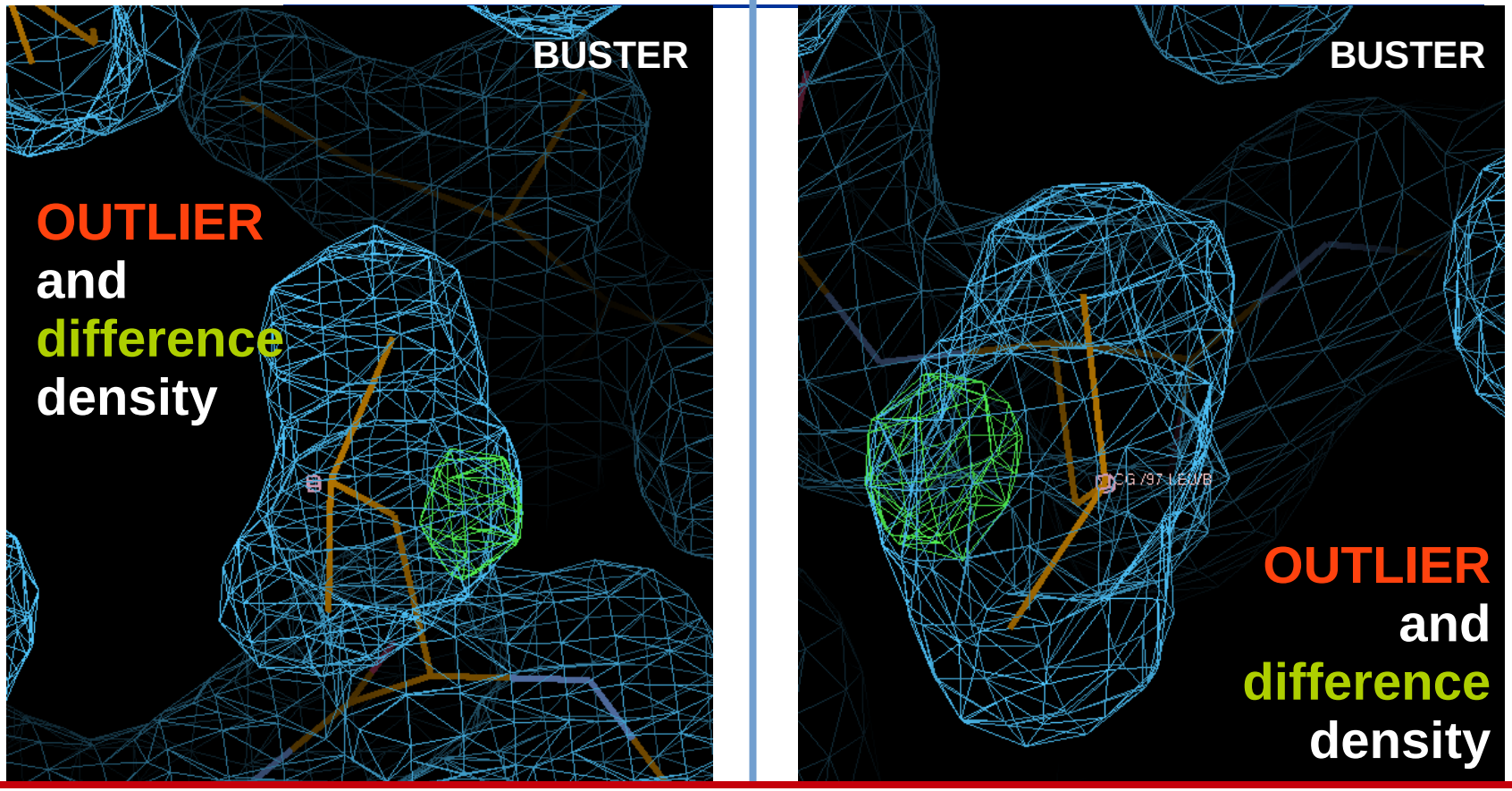
GΦL
Global Phasing Limited

Rotamer outliers are useful markers



GΦL
Global Phasing Limited

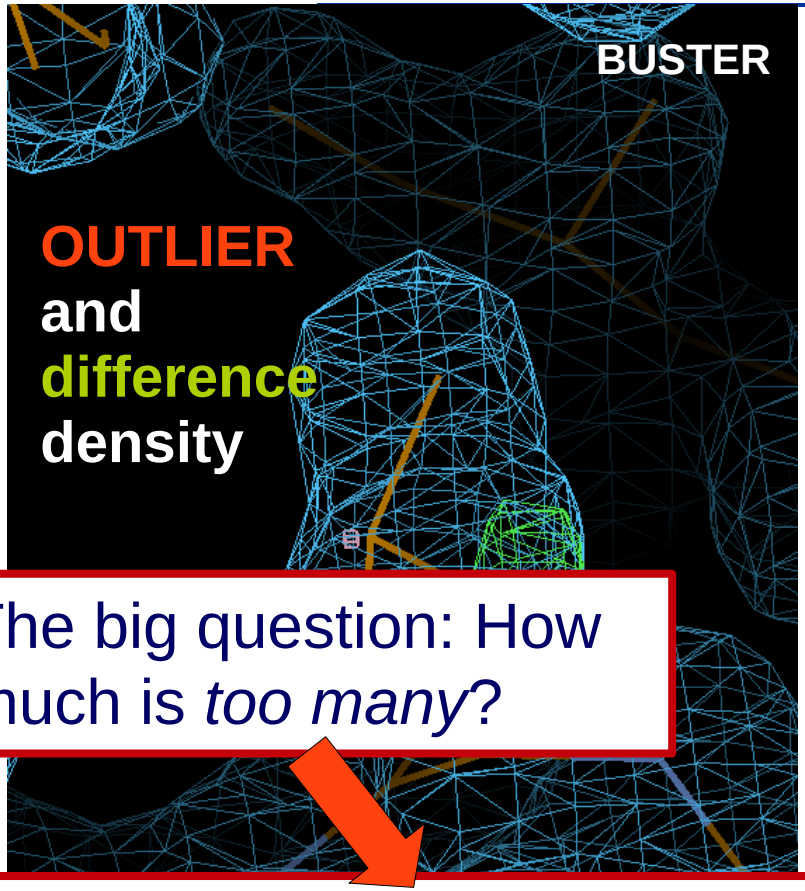
Rotamer outliers are useful markers

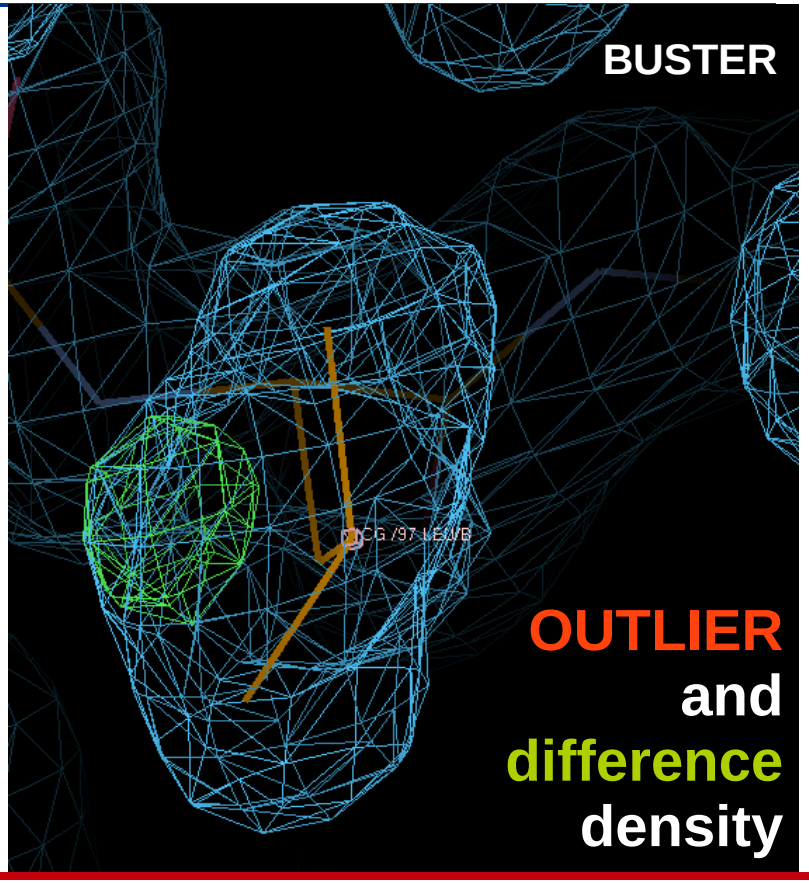


Do not use *too many* additional restraints to keep geometry under control if underlying problem is incorrect atomic model - as visible in density!

GΦL
Global Phasing Limited

Rotamer outliers are useful markers





The big question: How much is *too many*?

Do not use *too many* additional restraints to keep geometry under control if underlying problem is incorrect atomic model - as visible in density!

GΦL
Global Phasing Limited

BUSTER: Is there any any ligand bound?

Acta Cryst. (2005). A**61**, C248

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

Clemens Vonrhein, Gerard Bricogne, *Global Phasing Ltd., Cambridge, UK.* E-mail: vonrhein@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., Vonrhein 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

Global Phasing Limited

BUSTER: Is there any any ligand bound?

GΦL **LigandDetectionModes**

Favorite? Search 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 actualy 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/

GΦL
Global Phasing Limited

Difference Fourier maps - simple & useful

- **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

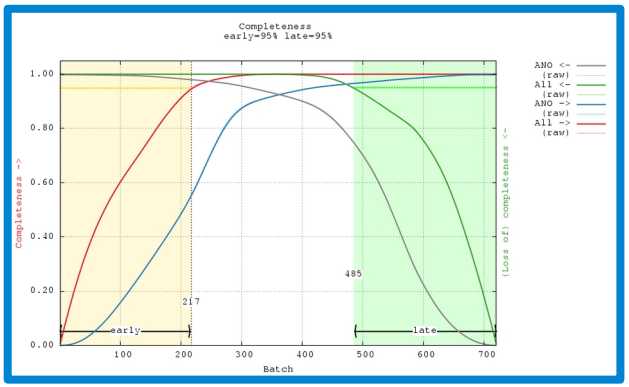
```
diff_fourier -h
```

GΦL
Global Phasing Limited

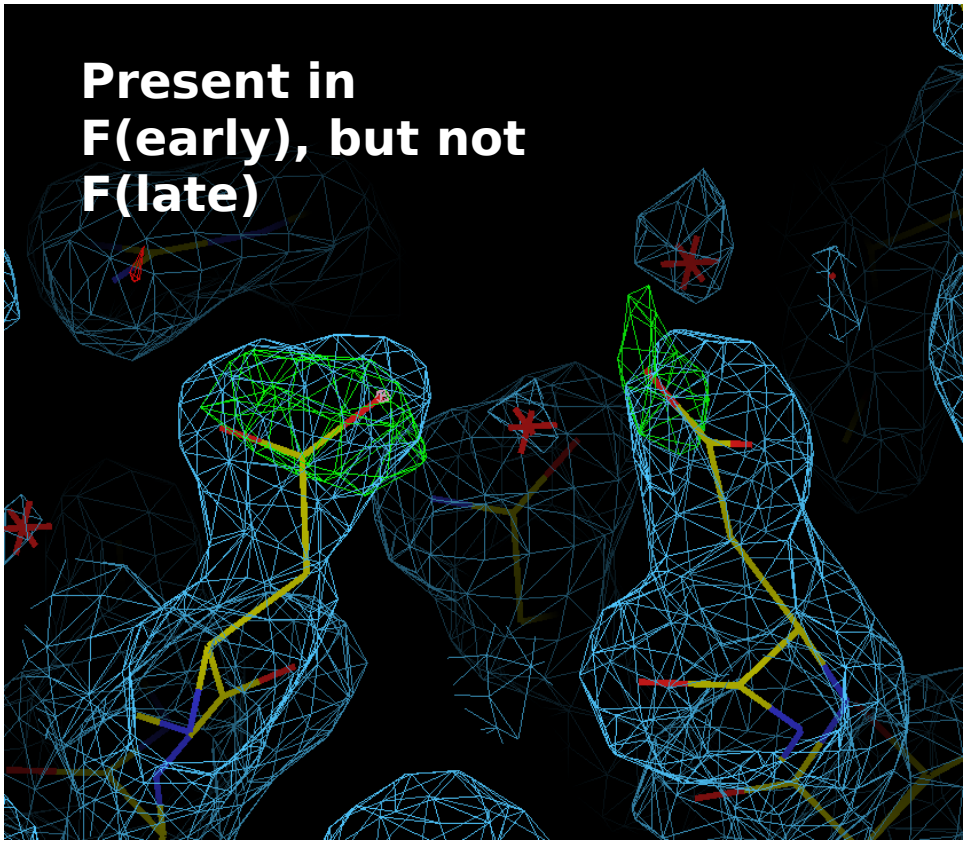
Difference Fourier maps: F(early)-F(late) detecting/describing radiation damage

Typical decarboxylation of ASP/GLU residues

Normal mFo-Dfc maps will show negative peaks.



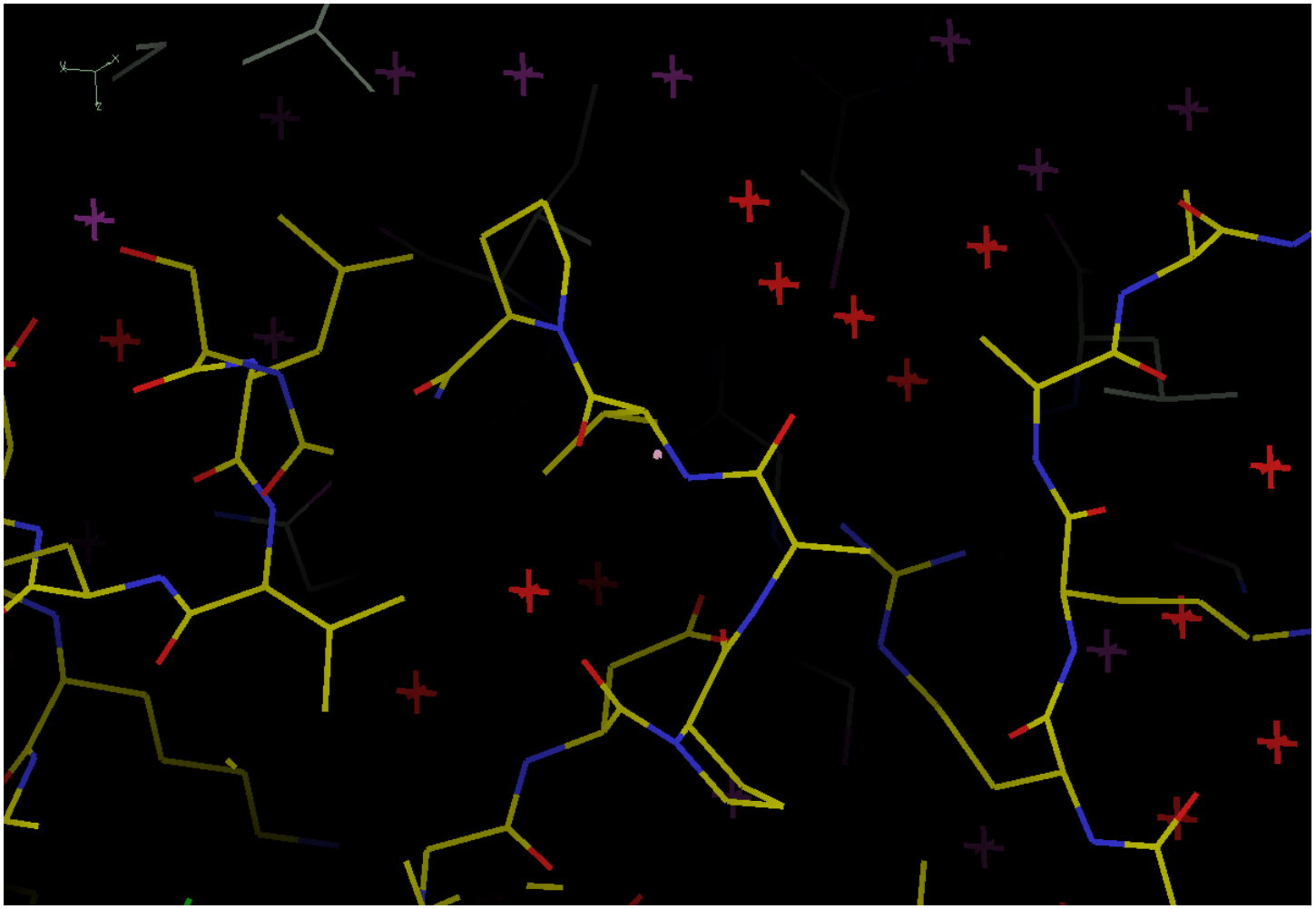
low dose, high-multiplicity



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.

GΦL
Global Phasing Limited

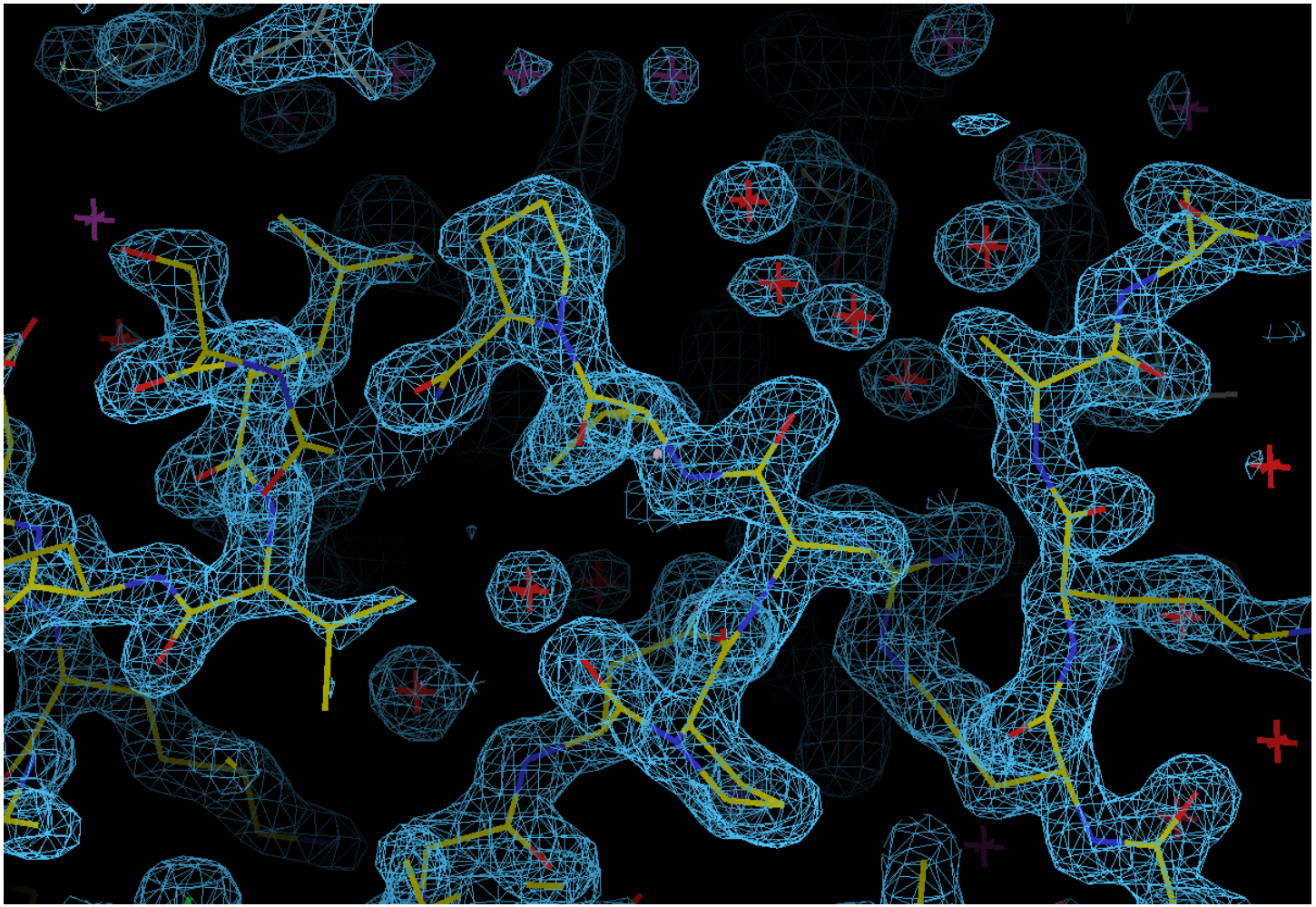
F(early)-F(late): water in crystal contact



4LER

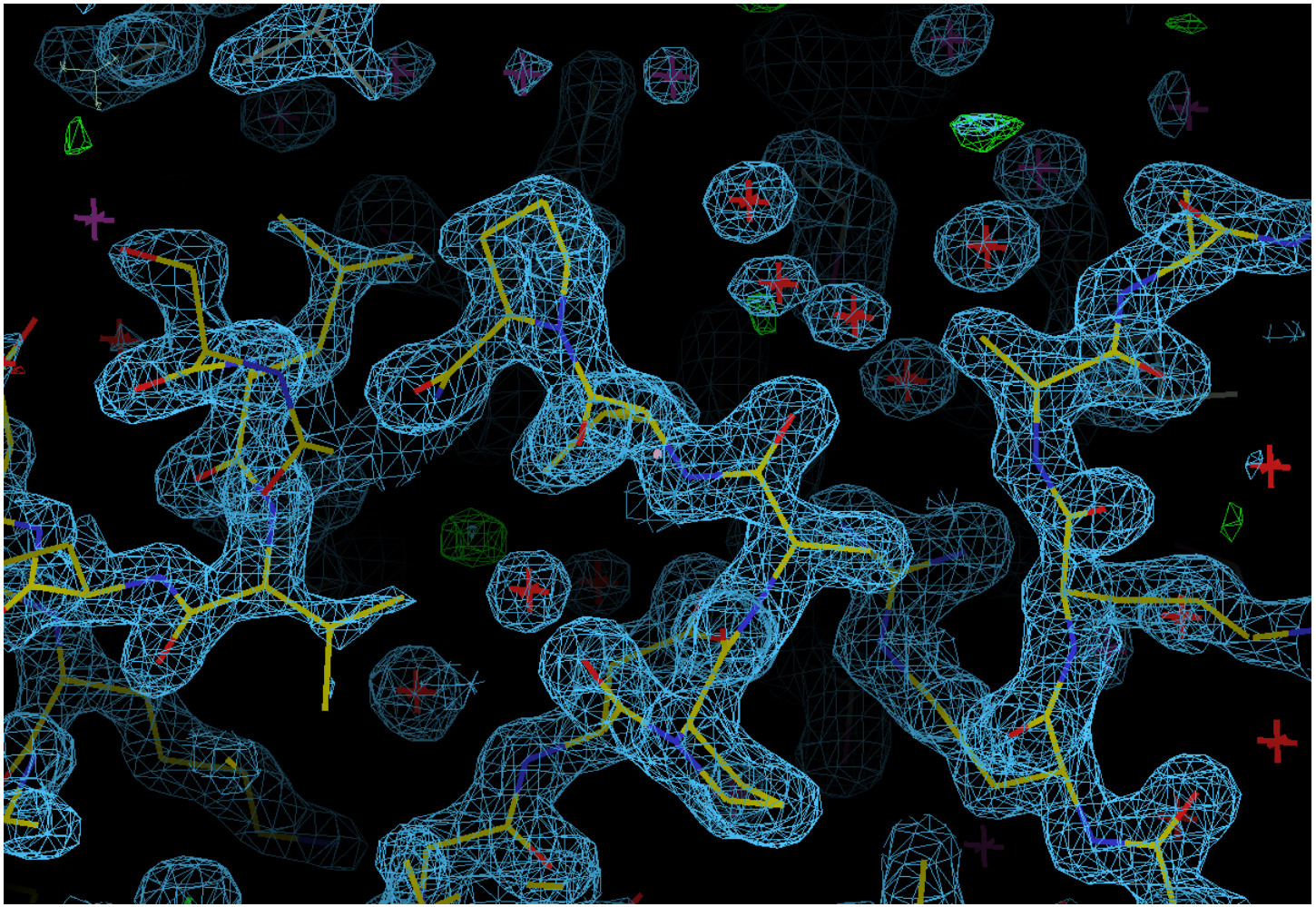
GΦL
Global Phasing Limited

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



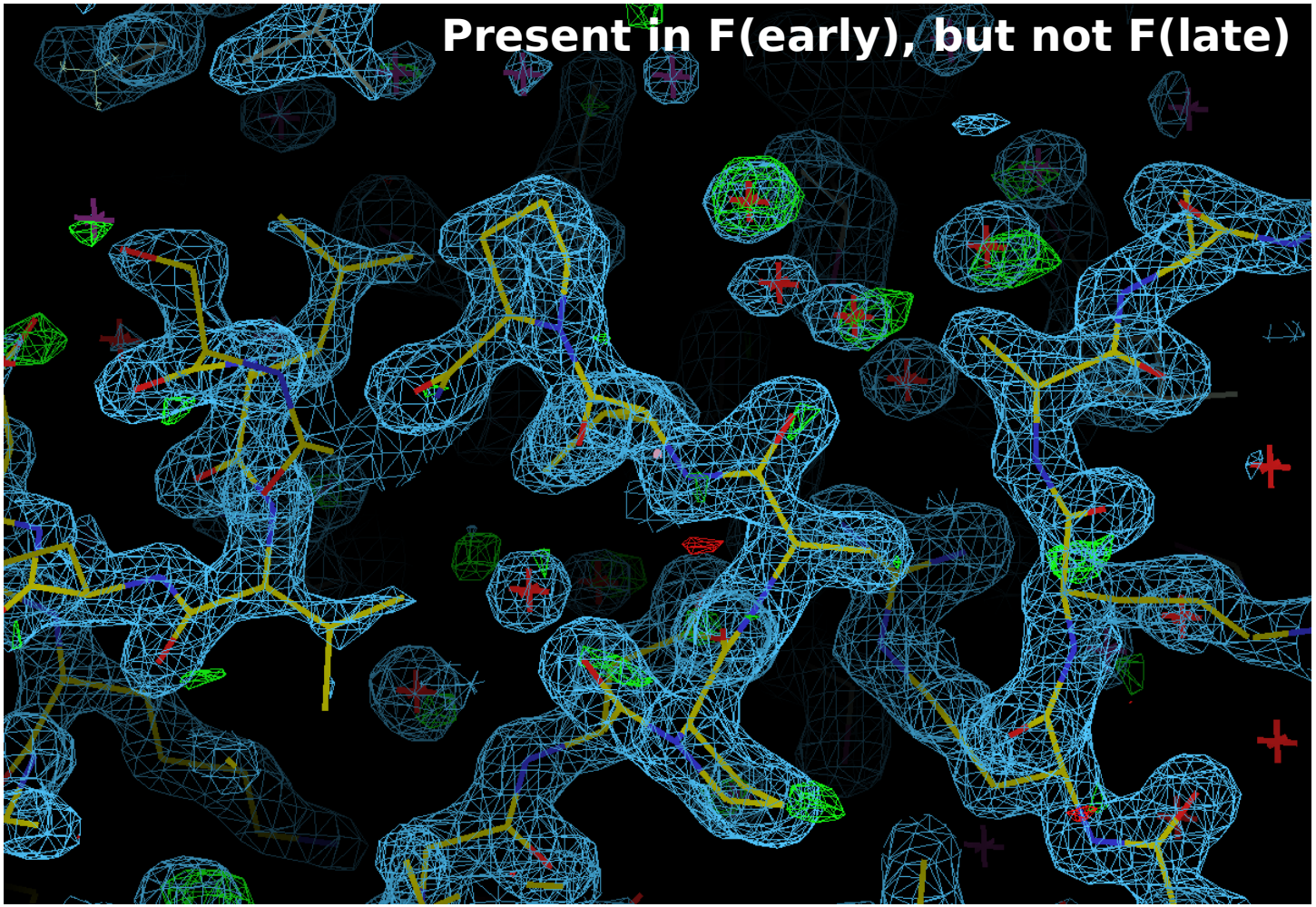
GΦL
Global Phasing Limited

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



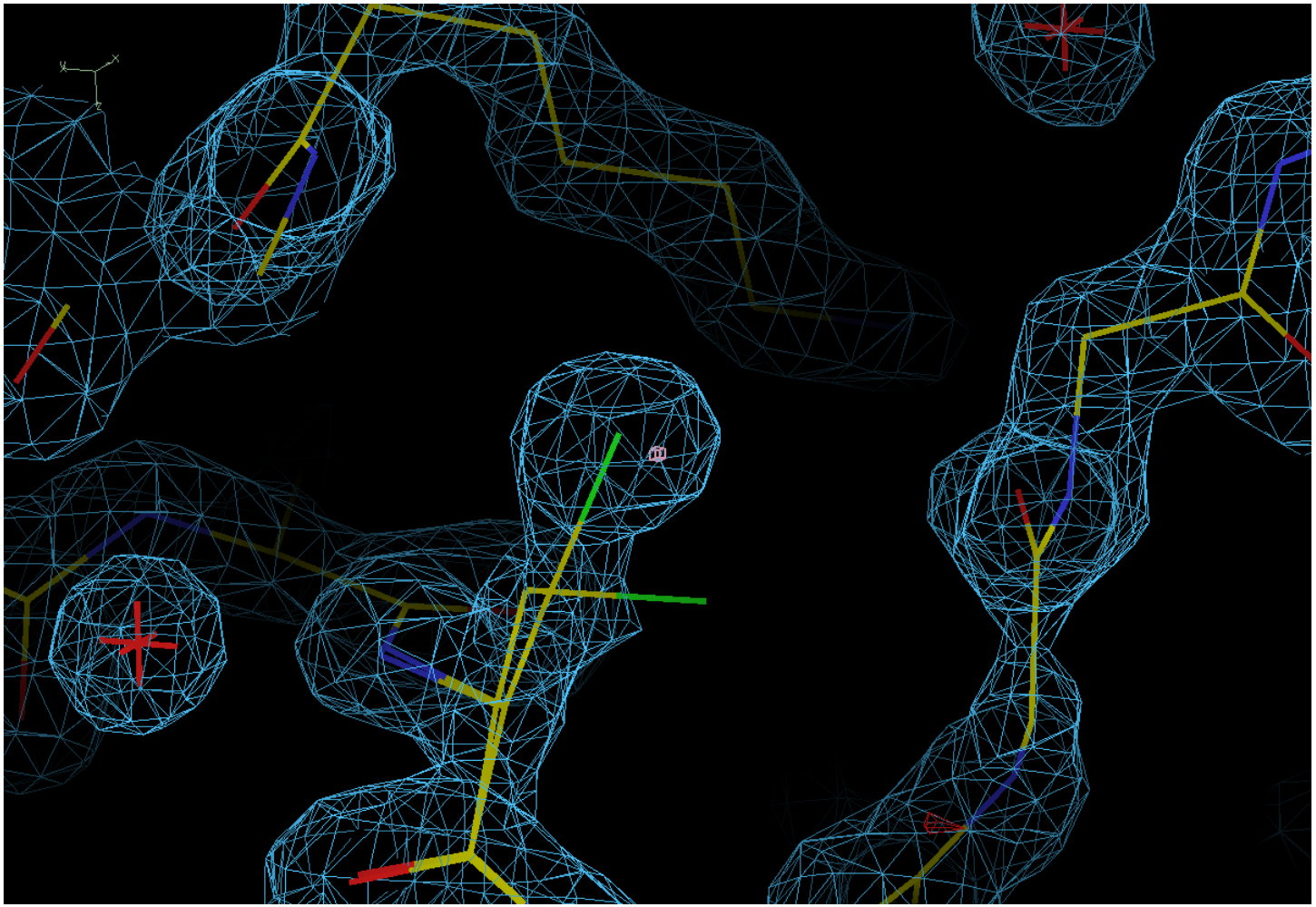
GΦL
Global Phasing Limited

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



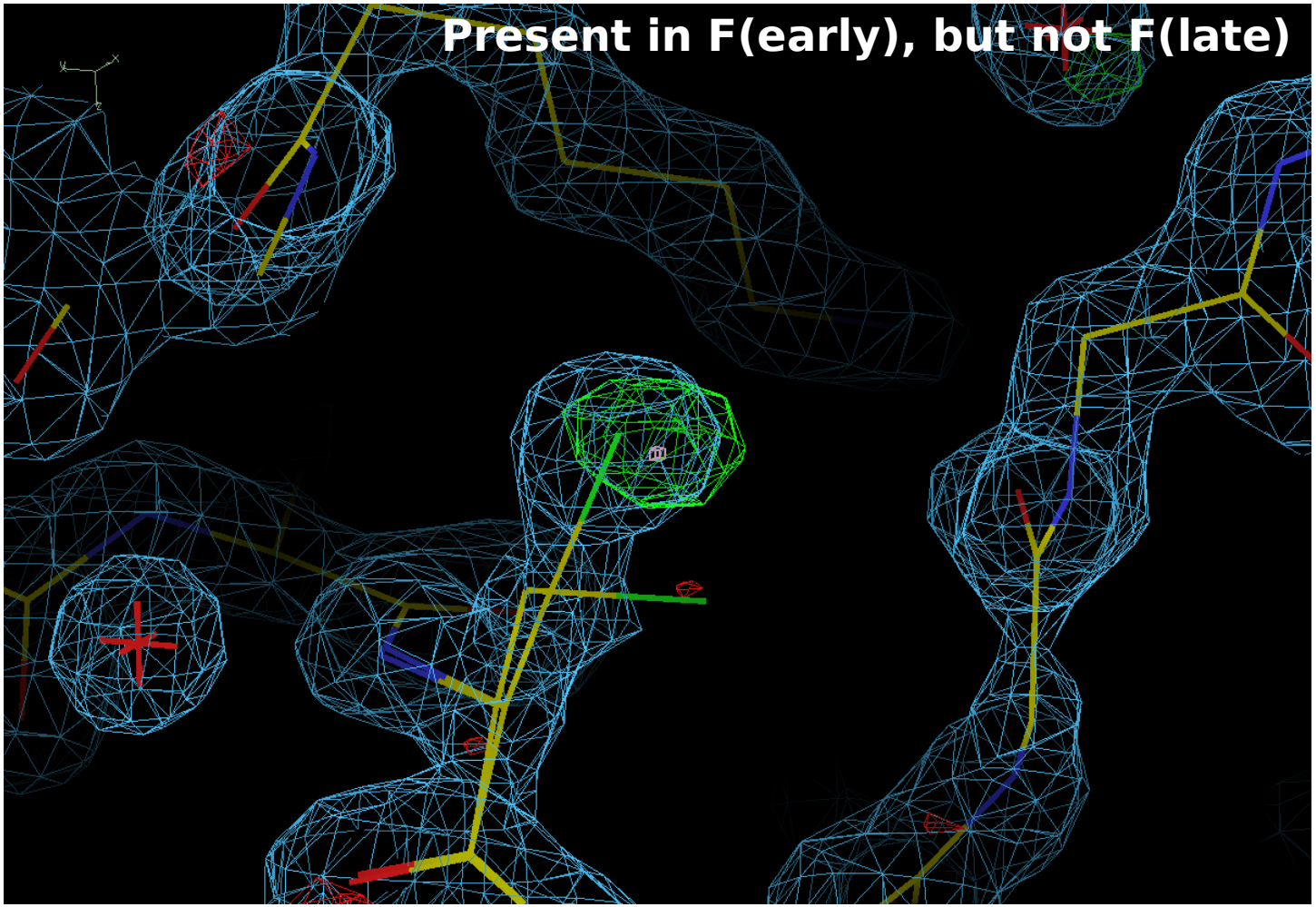
GΦL
Global Phasing Limited

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



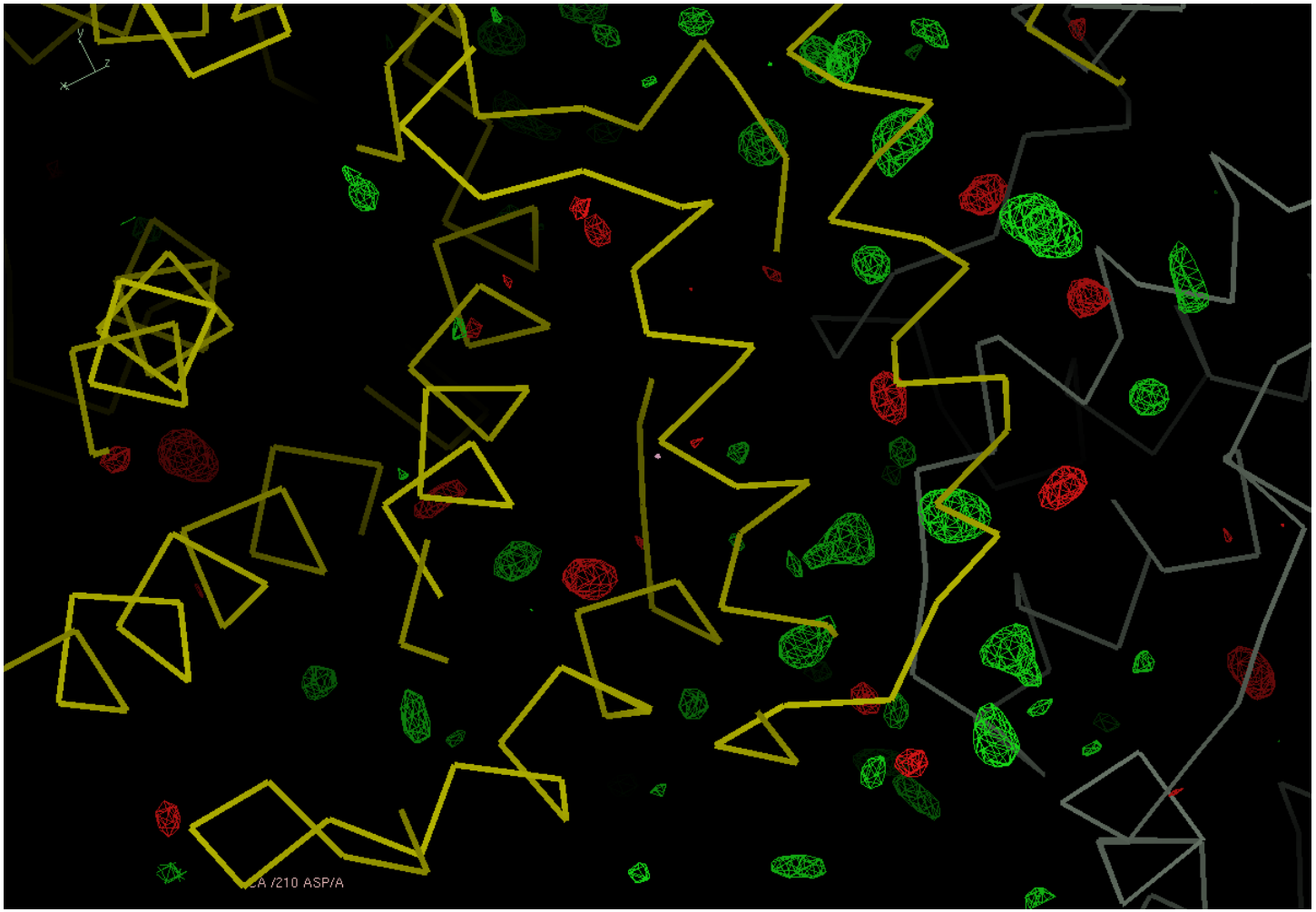
GΦL
Global Phasing Limited

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



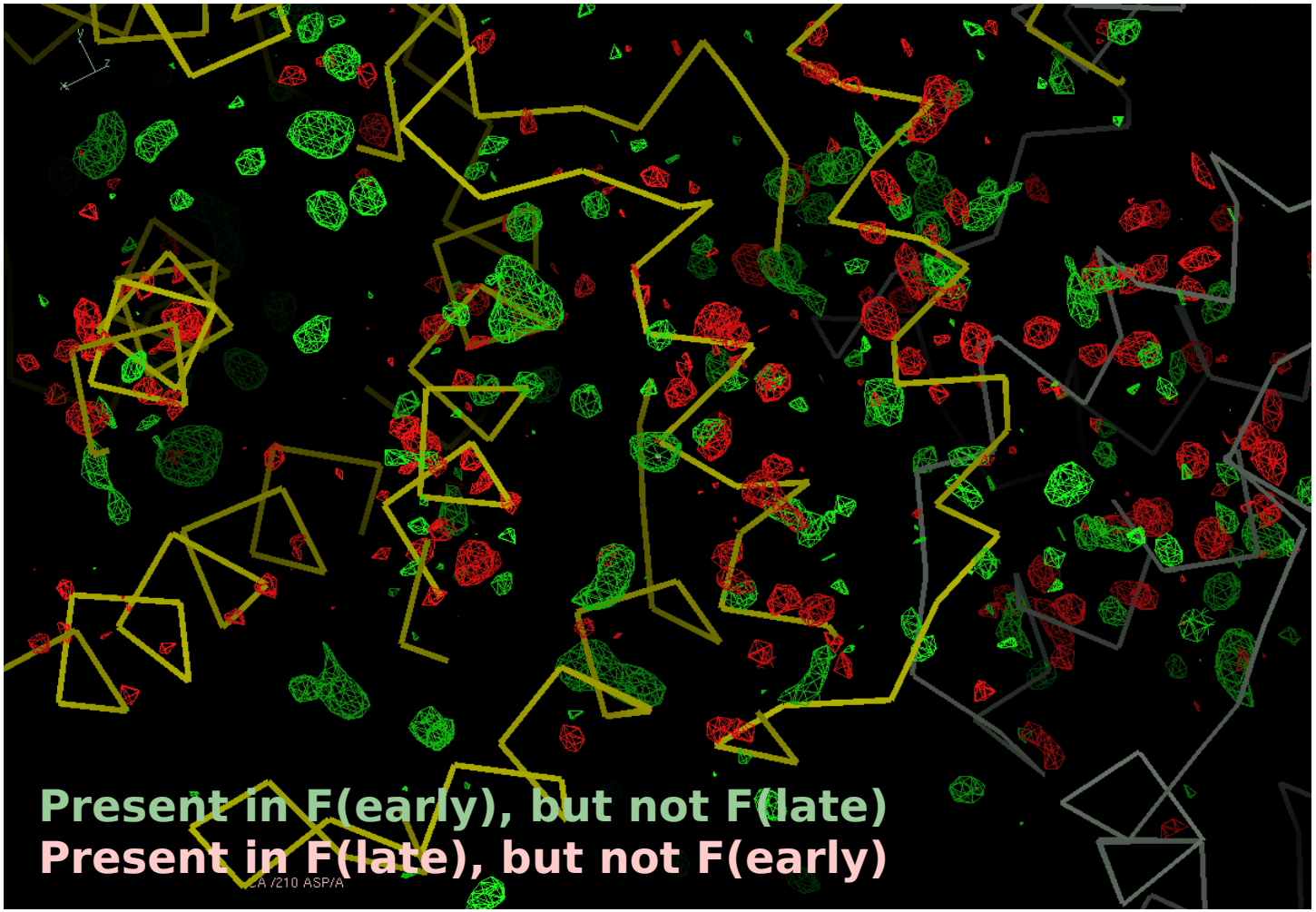
GΦL
Global Phasing Limited

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



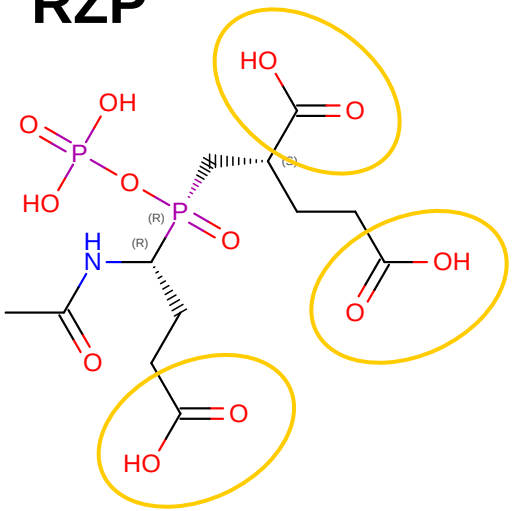
GΦL
Global Phasing Limited

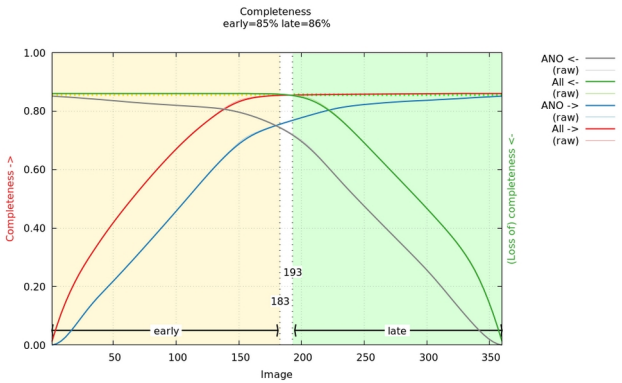
F(early)-F(late): large-scale movements



Radiation damage - decarboxylation

RZP





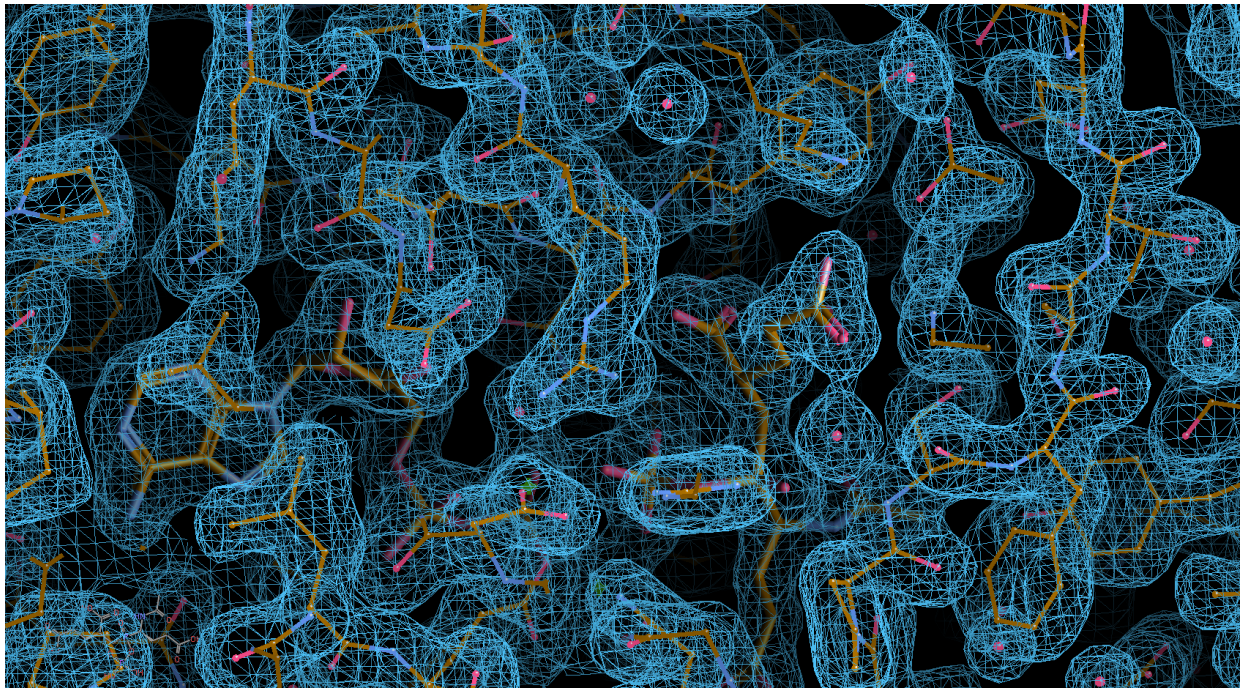
6VZU

Structural basis for polyglutamate chain initiation and elongation by TTLL family enzymes.

Mahalingan, K.K., Keith Keenan, E., Strickland, M., Li, Y., Liu, Y., Ball, H.L., Tanner, M.E., Tjandra, N., Roll-Mecak, A.

(2020) Nat Struct Mol Biol **27**: 802-813

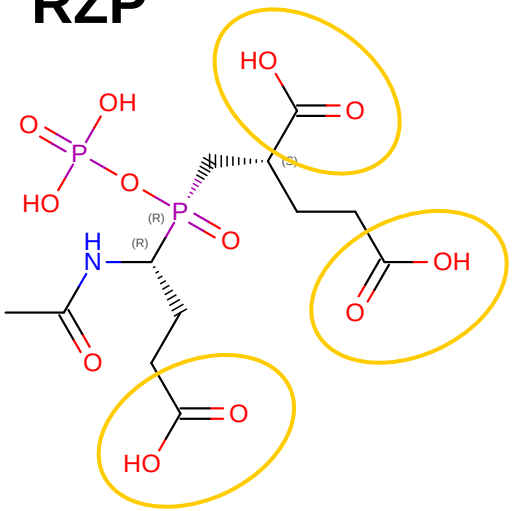
Radiation damage - decarboxylation



2mFo-DFc @ 1.0 rms, 2 Å resolution.

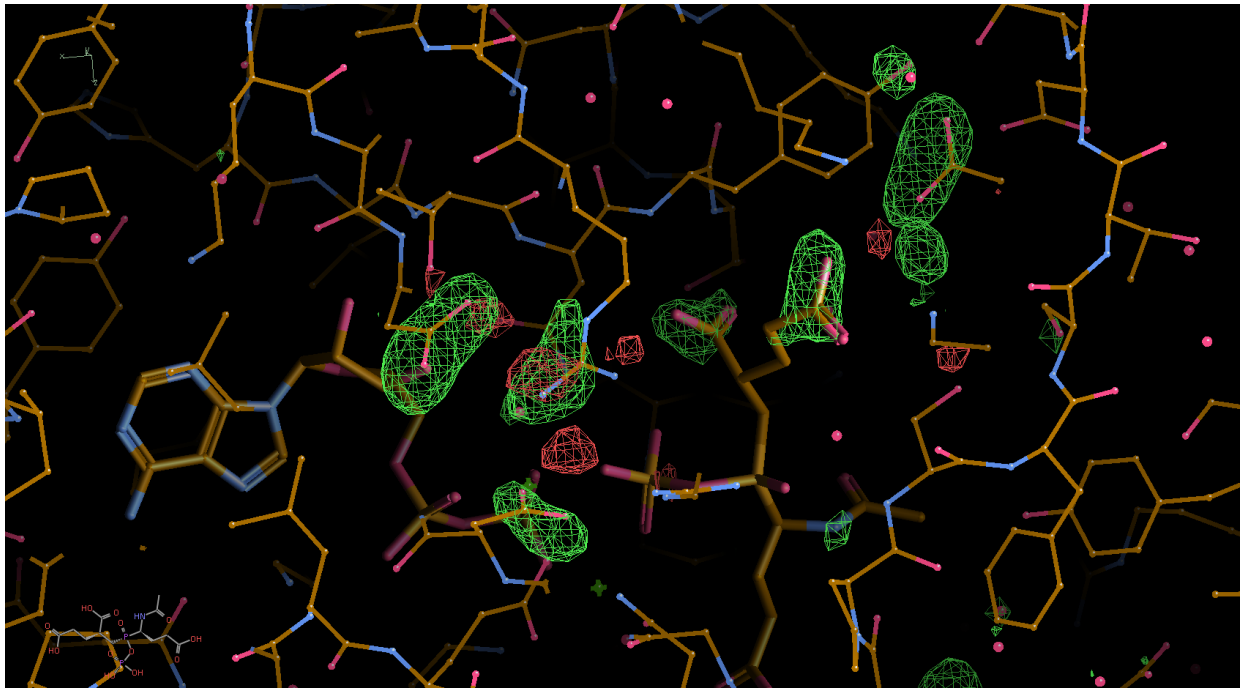
6VZU

RZP

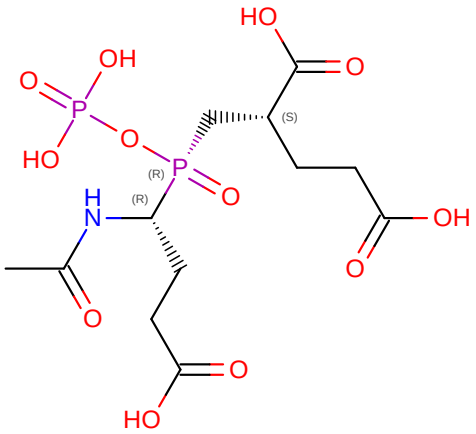


GΦL
Global Phasing Limited

Radiation damage - decarboxylation



RZP



F(early)-F(late) @ 6.0 rms

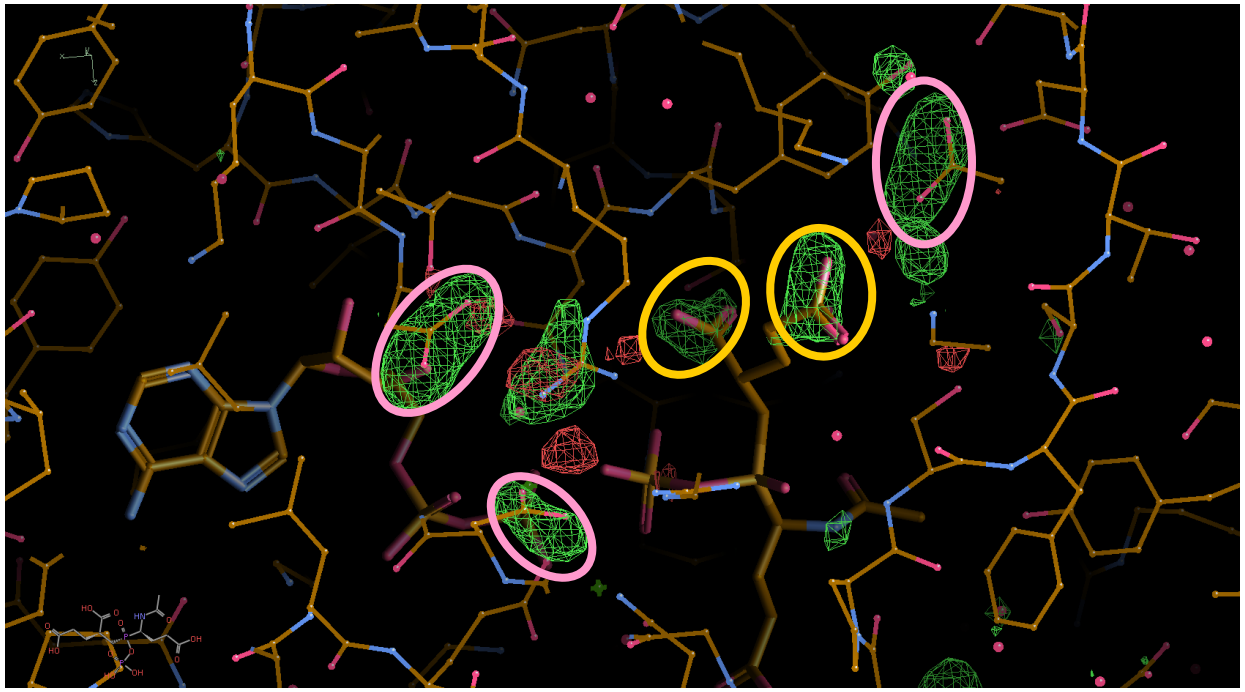
6VZU

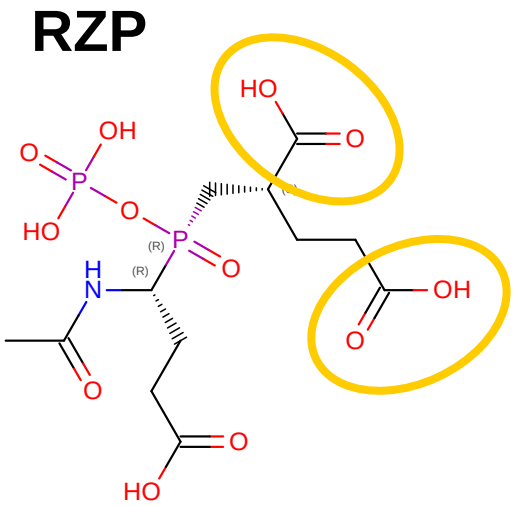
Showing loss of density due to
- reduction in occupancy
- increase in temperature factor
- loss of atoms/groups
- movement of atoms/groups

- **early and late data** automatically generated by **autoPROC**
- **F(early)-F(late)** maps automatically generated and analysed by **BUSTER**

GΦL
Global Phasing Limited

Radiation damage - decarboxylation





F(early)-F(late) @ 6.0 rms

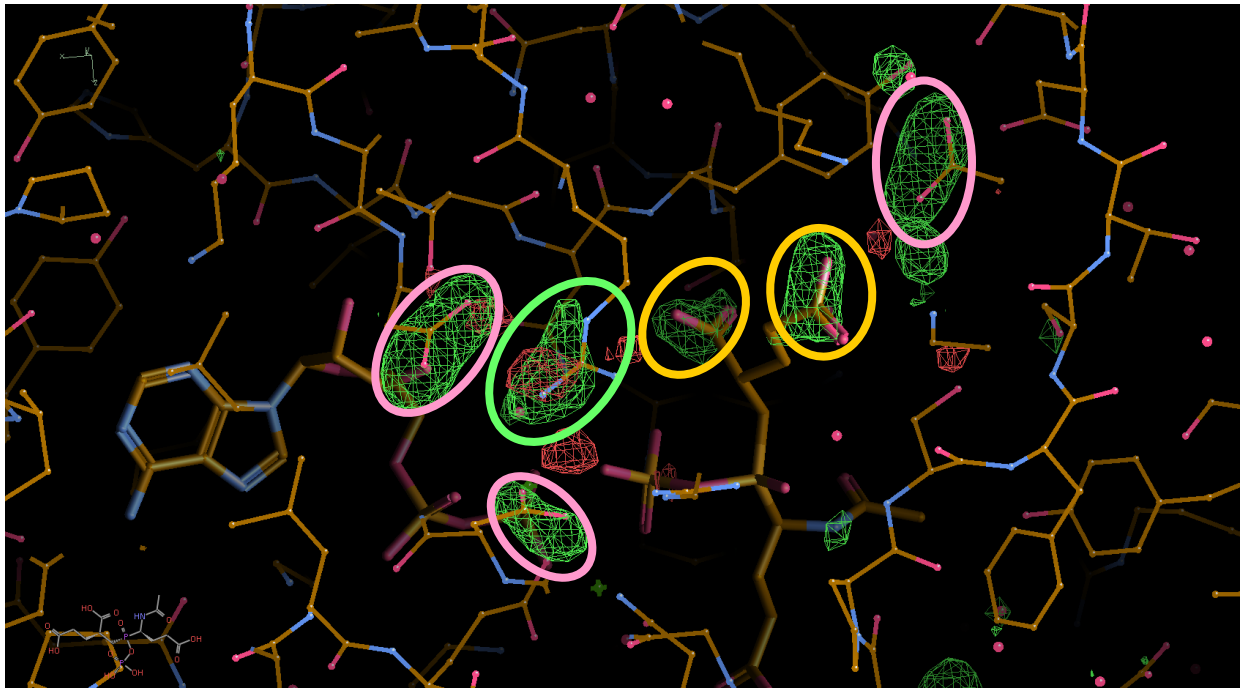
Showing loss of density due to
- reduction in occupancy
- increase in temperature factor
- loss of atoms/groups
- movement of atoms/groups

6VZU

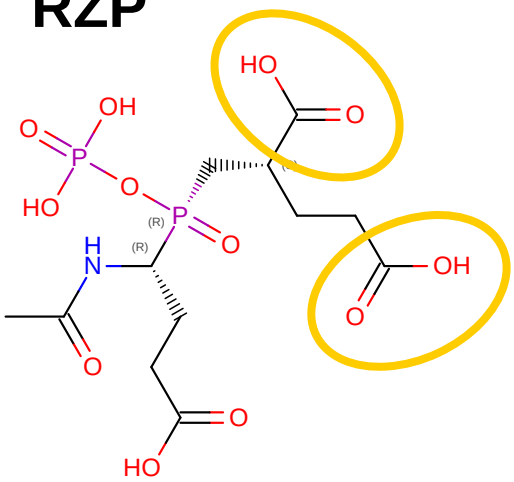
de-carboxylation of ASP/GLU and compound

GΦL
Global Phasing Limited

Radiation damage - decarboxylation



RZP



F(early)-F(late) @ 6.0 rms

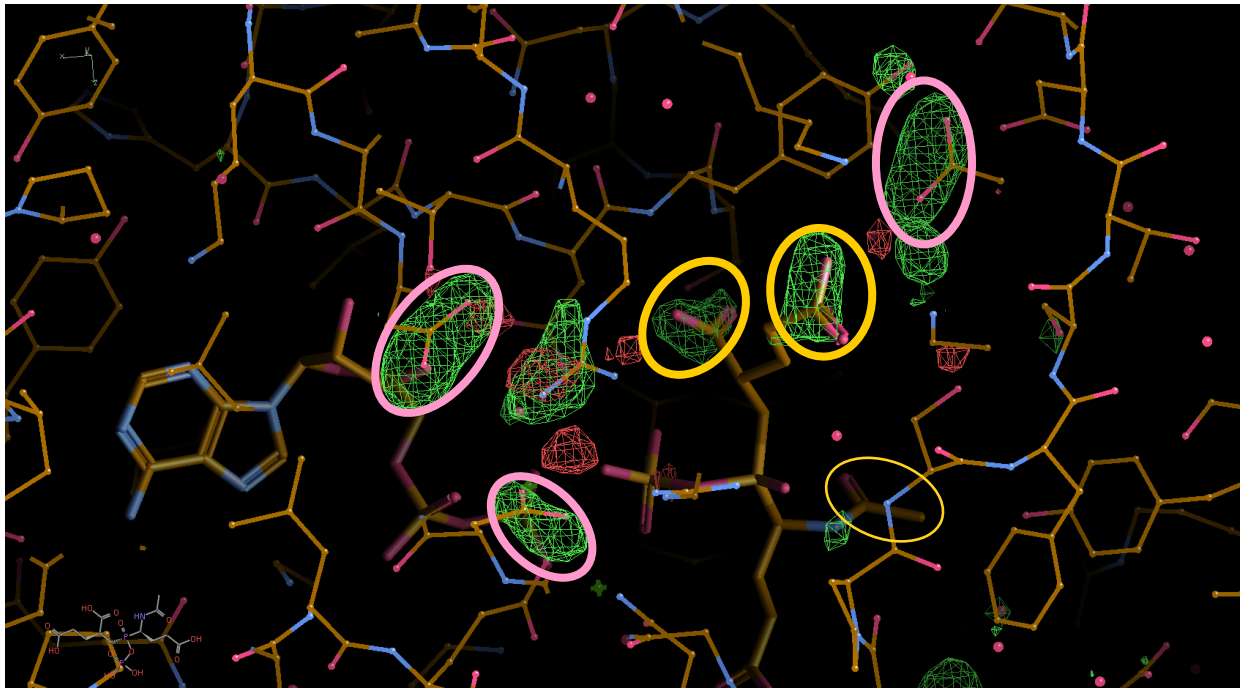
Showing loss of density due to
- reduction in occupancy
- increase in temperature factor
- loss of atoms/groups
- movement of atoms/groups

6VZU

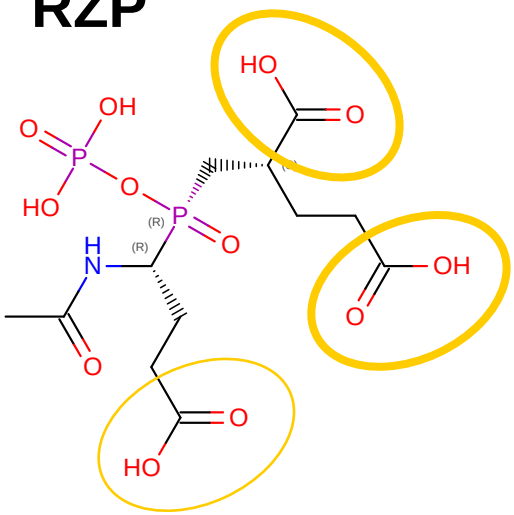
de-carboxylation of ASP/GLU and compound

GΦL
Global Phasing Limited

Radiation damage - decarboxylation



RZP



F(early)-F(late) @ 6.0 rms

Showing loss of density due to
- reduction in occupancy
- increase in temperature factor
- loss of atoms/groups
- movement of atoms/groups

6VZU

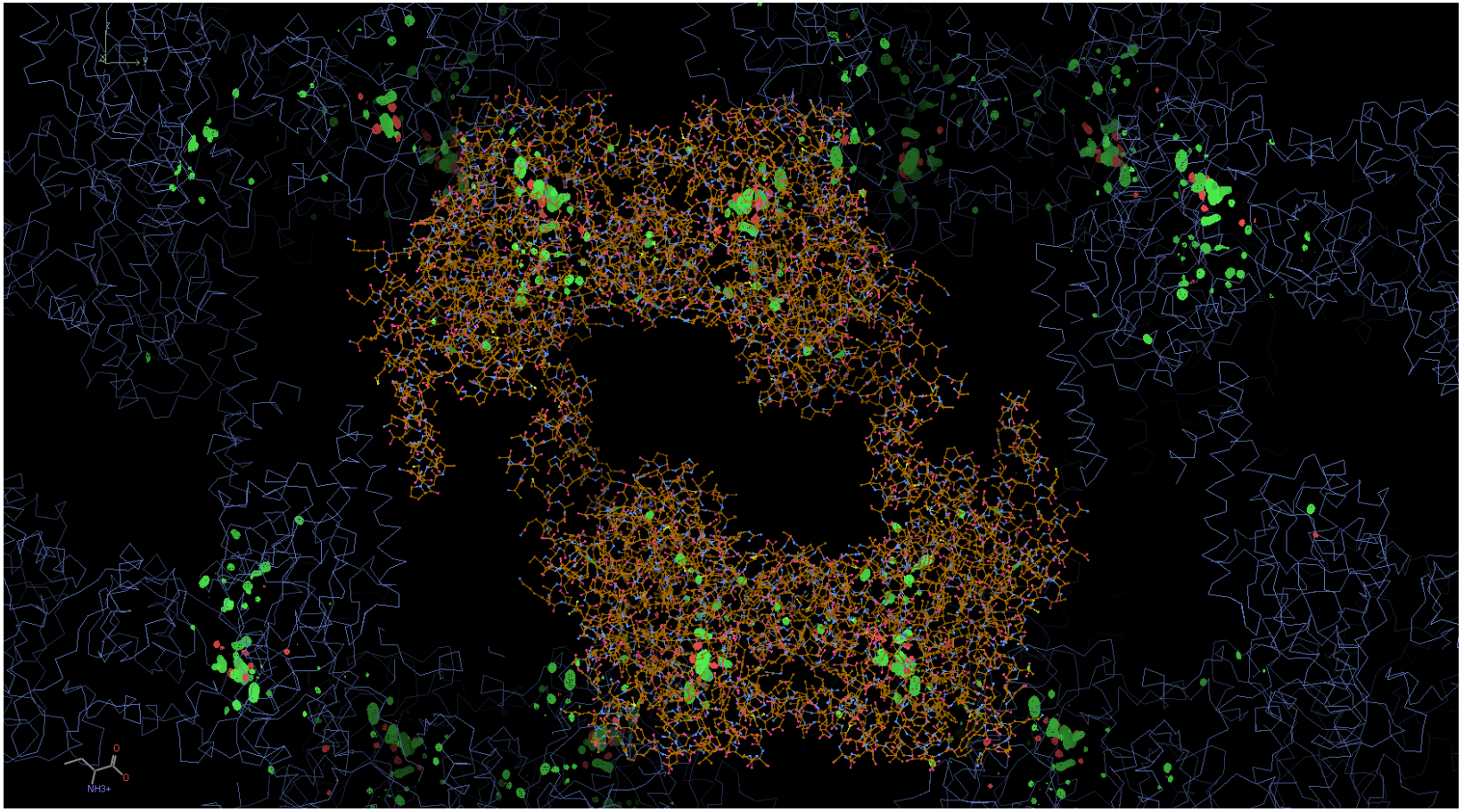
de-carboxylation of ASP/GLU and compound

not every carboxy group suffers radiation damage to the same extent

GΦL
Global Phasing Limited

Radiation damage - decarboxylation

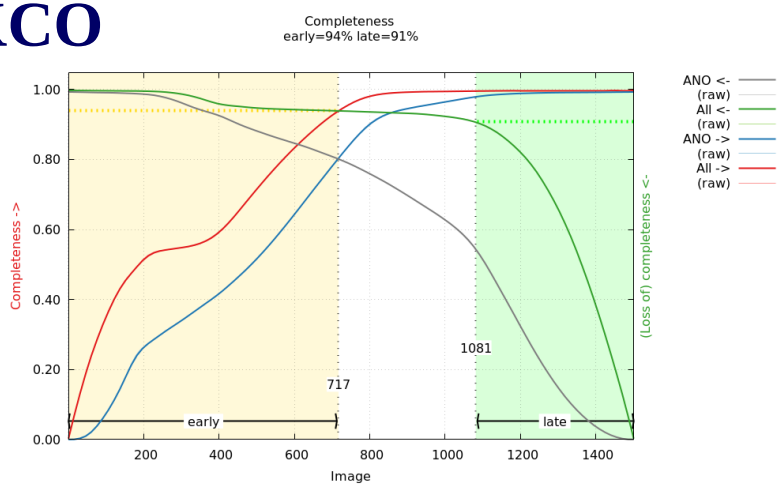
6VZU

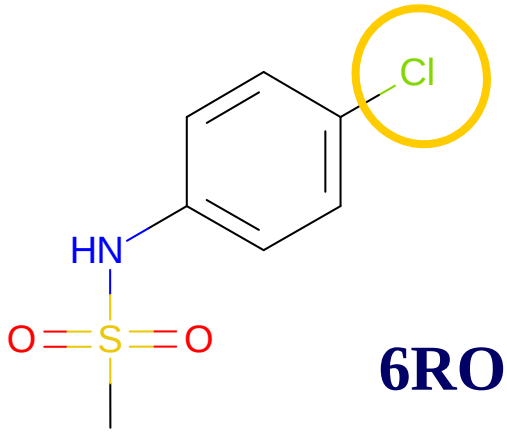


F(early)-F(late) @ 6.0 rms showing hotspots and “connected” regions

Radiation Damage - improved model parametrisation

5KCO



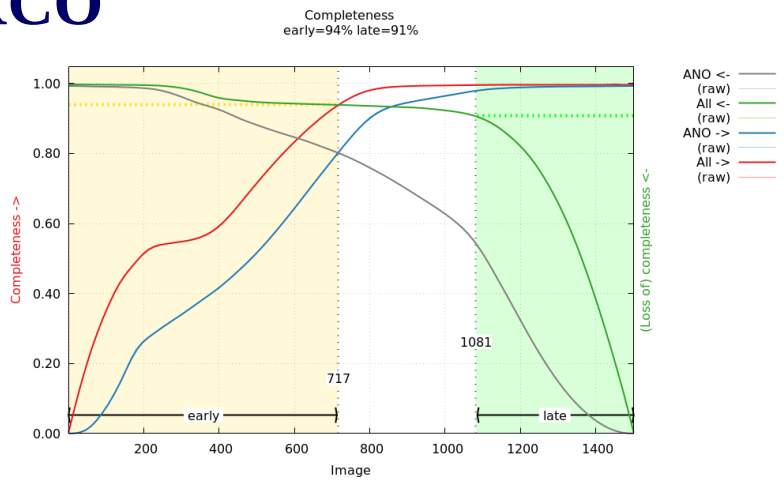


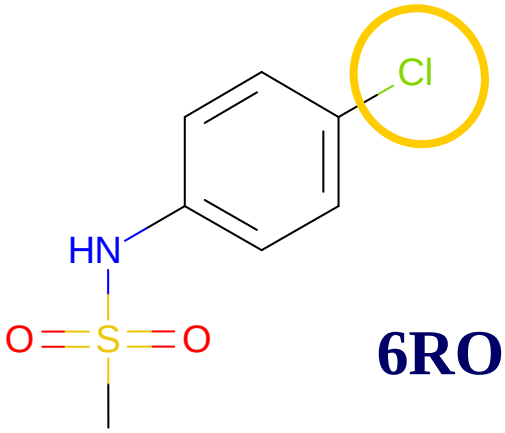
6RO

5KCO Tempel et al (2016).

Radiation Damage - improved model parametrisation

5KCO



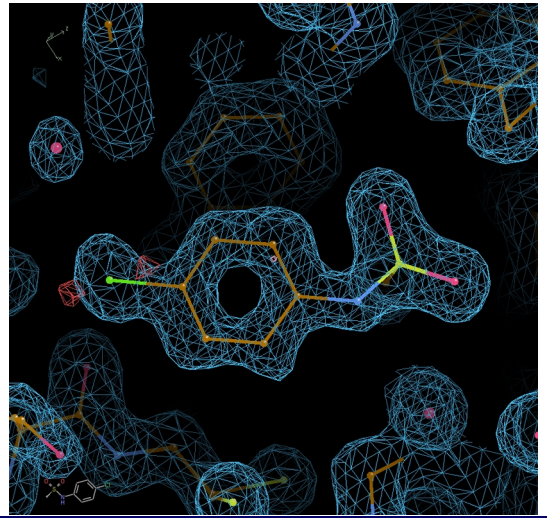


6RO

2mFo-DFc @ 1.0 rms

mFo-DFc @ 3.5 rms

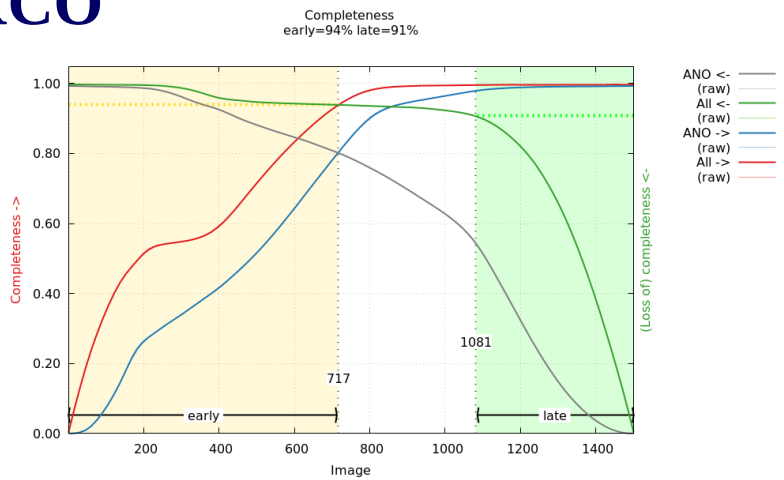
very clean

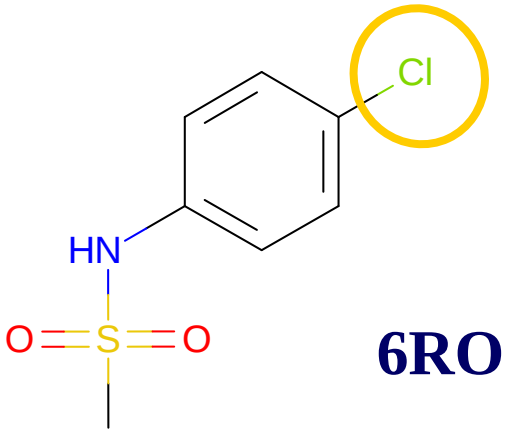


GΦL
Global Phasing Limited

Radiation Damage - improved model parametrisation

5KCO



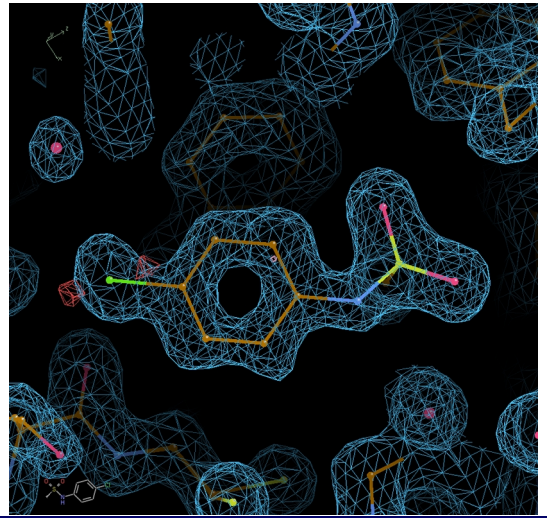


6RO

2mFo-DFc @ 1.0 rms

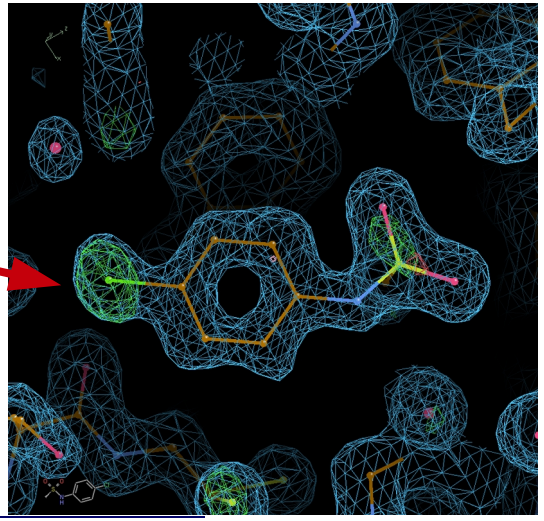
mFo-DFc @ 3.5 rms

very clean



F(early)-F(late) @ 4.0 rms

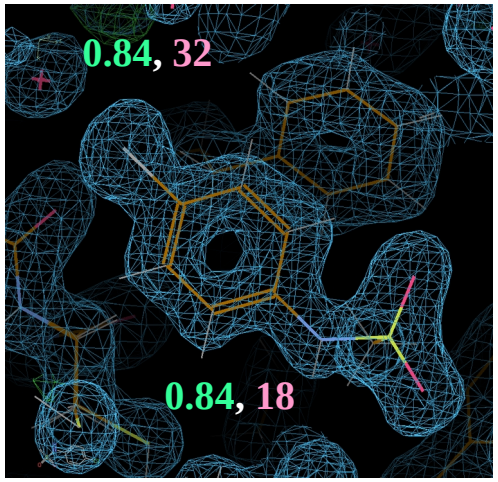
strong peak on chlorine



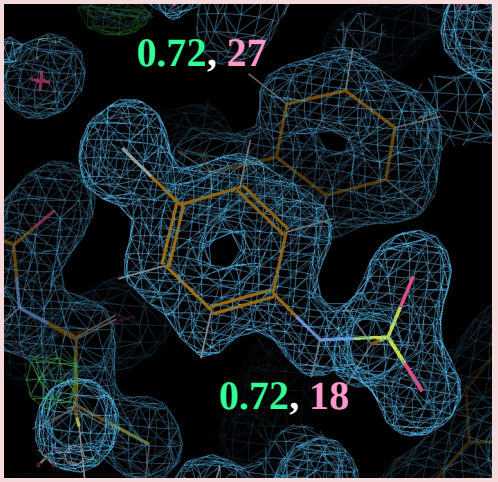
GΦL
Global Phasing Limited

Re-refined
2mFo-DFc @ 1.0 rms
mFo-Dfc @ 3.5 rms

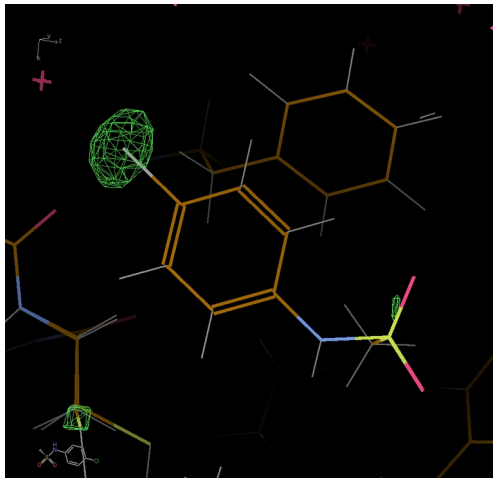
Deposited



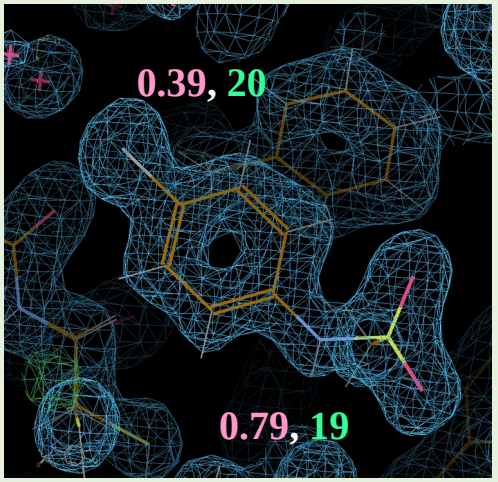
Re-processed single OCC



F(early)-F(late) @ 5.0 rms



Re-processed split OCC



GΦL
Global Phasing Limited

Summary

- **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**

GΦL
Global Phasing Limited

Acknowledgements

- Global Phasing, Cambridge (UK):
  - Gérard Bricogne, Leigh Carter, Claus Flensburg, Rasmus Fogh, Peter Keller, Wlodek Paciorek, Andrew Sharff, Oliver Smart, Ian Tickle
  - Thomas Womack, Eric Blanc, Gwyndaf Evans, Pietro Roversi, John Irwin, Eric de la Fortelle, Marc Schiltz

- CCDC
- CCP4

- Global Phasing Consortium members
- … many, many users & collaborators

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