An introduction to BUSTER

Oliver S. Smart
Global Phasing Ltd

ACA workshop

24th July 2010

Origins of BUSTER (1989-)

- Proper implementation of *ab initio* phasing by entropy maximisation and likelihood ranking (as proposed in 1984), a topic for a Transactions Symposium in 1994.
- Most general implementation of Bertaut's linearisation formula for computing moments of trigonometric structure factors from random-atom structures under very general hypotheses (including NCS).
- Powerful tree search (sampling by error-correcting codes, …).
- Likelihood computation, analysis and optimisation.
- Possible incorporation of external phase information (elliptic Rice likelihood function) and of systematic resolution of phase ambiguity by "mode permutation".
- In 1994: connection to the TNT refinement package

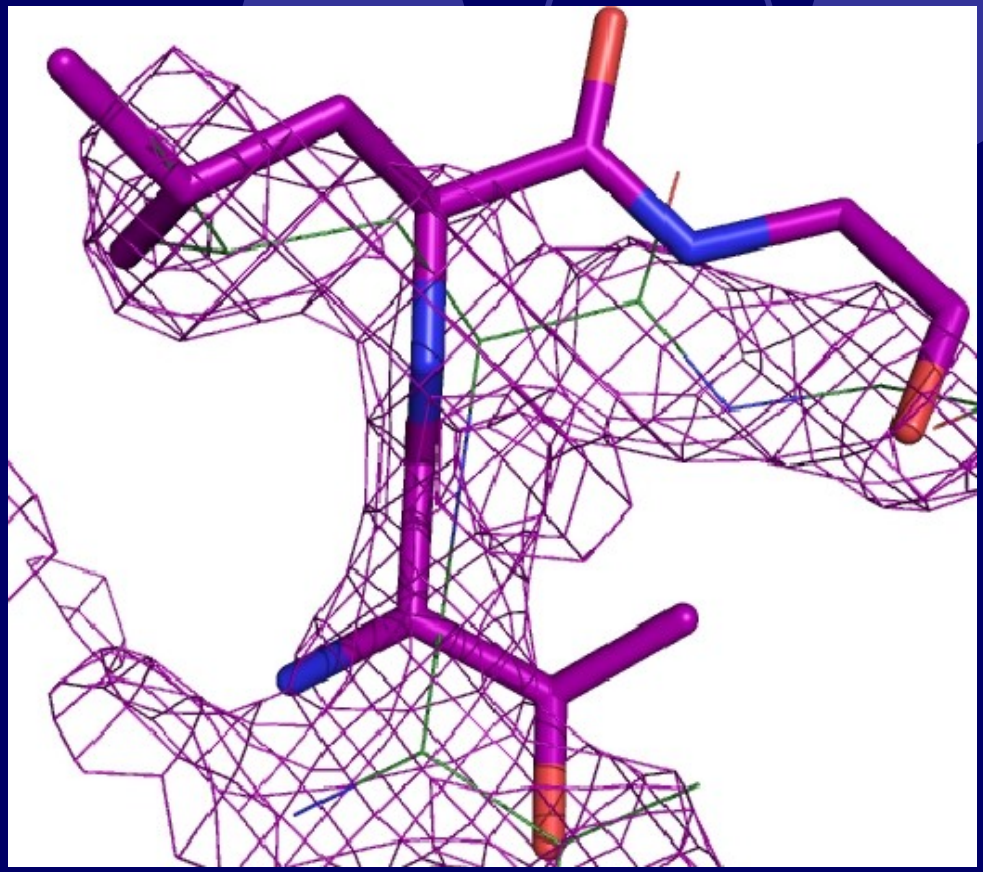
The first BUSTER-TNT (1994)

- BUSTER-TNT used TNT, with BUSTER replacing the **least squares X-ray term with Maximum Likelihood**.
- Produced “superior” maps from a given model and dataset.
- Was ineffective at moving model into map.

Improving BUSTER-TNT 2004-2007

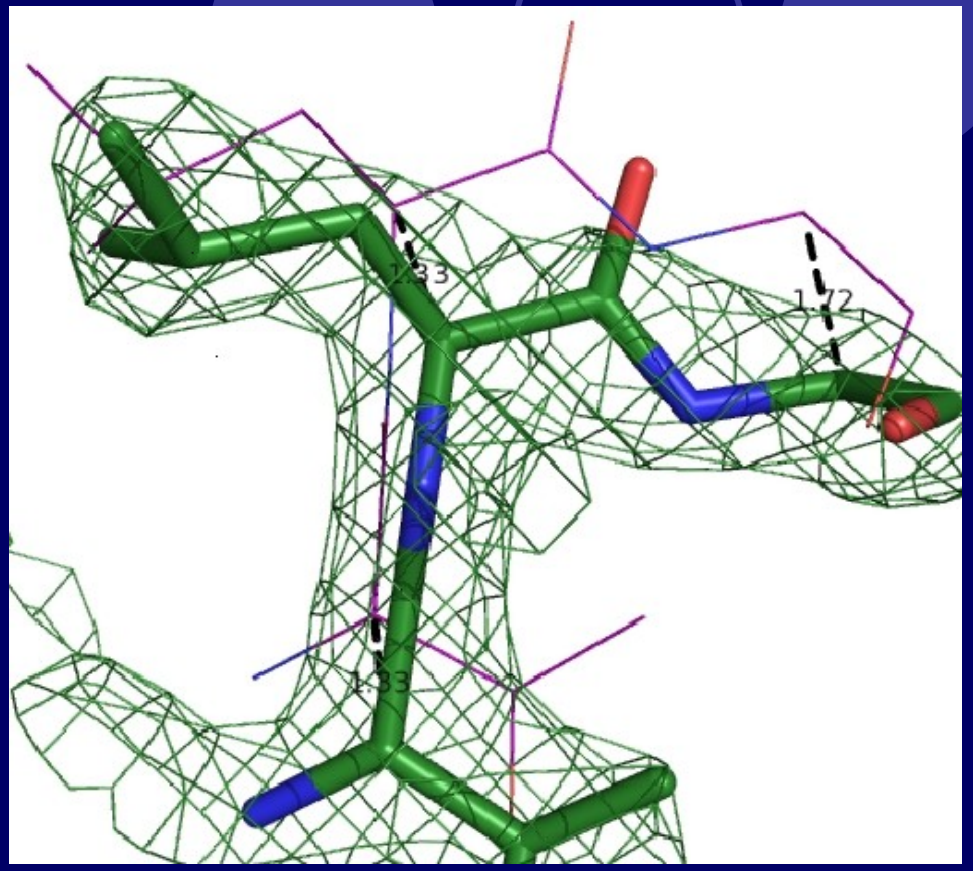
- We drew on the experience of rewriting SHARP (1998-2002), which produced speedups of 20-200, as well as much improved results thanks to a better optimiser.
- Improve:
  - Much better X-ray scaling ...
  - Much better optimizer
  - Command line interface in place of old style gui

Before …



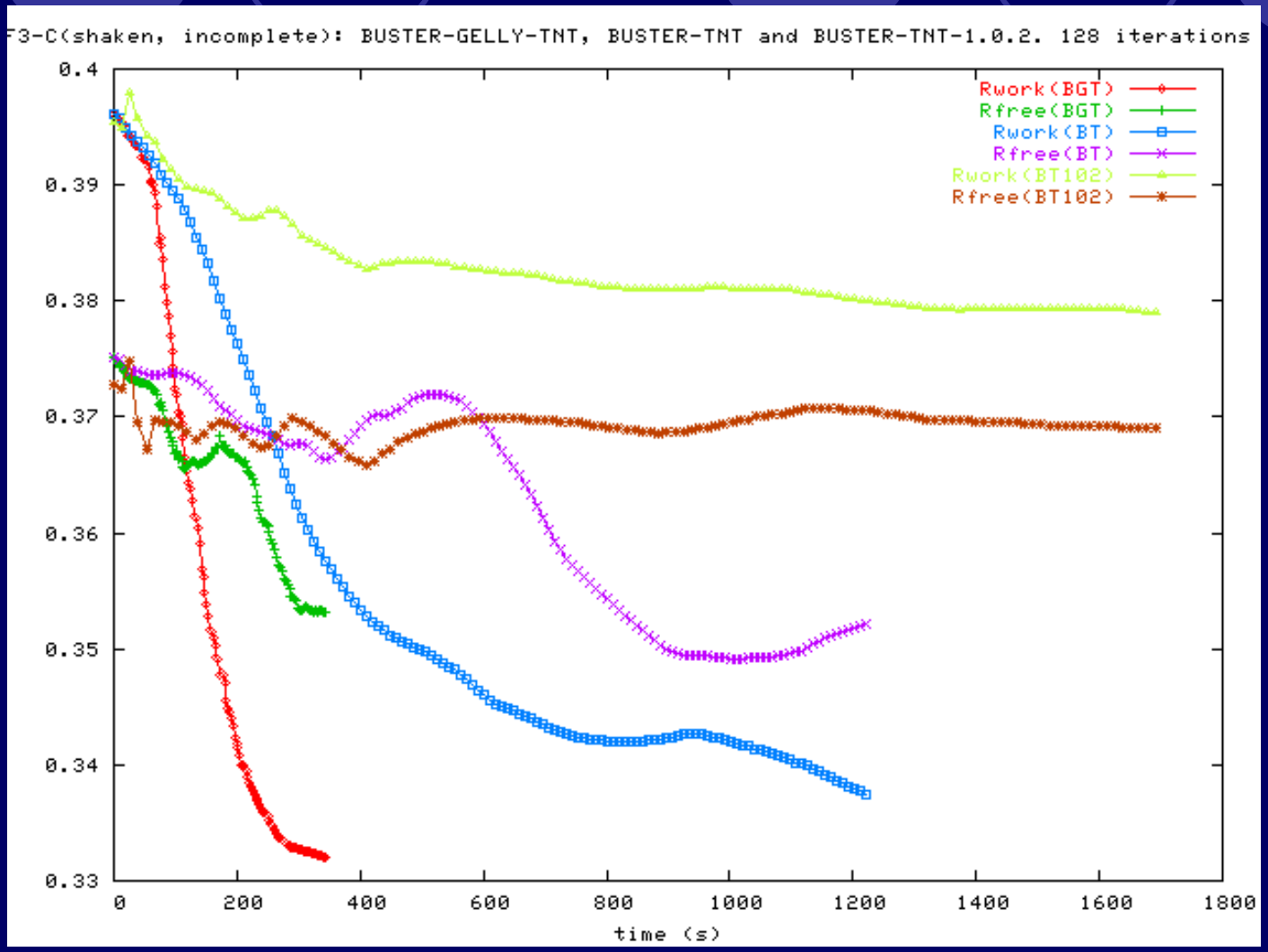
- Initial model in wrong place
- Density showing up in right place
- Model not being moved into density

… and After



- Model is now being moved into density

The new optimiser highlights the importance of observations/parameter ratio



BUSTER improvements 2008…

Need to improve the economics of observations vs. parameters:

1. Better NCS
   - New method for ncs: LSSR
   - -autoncs option

1. target restraints
2. TLS
3. Improved protein geometry function

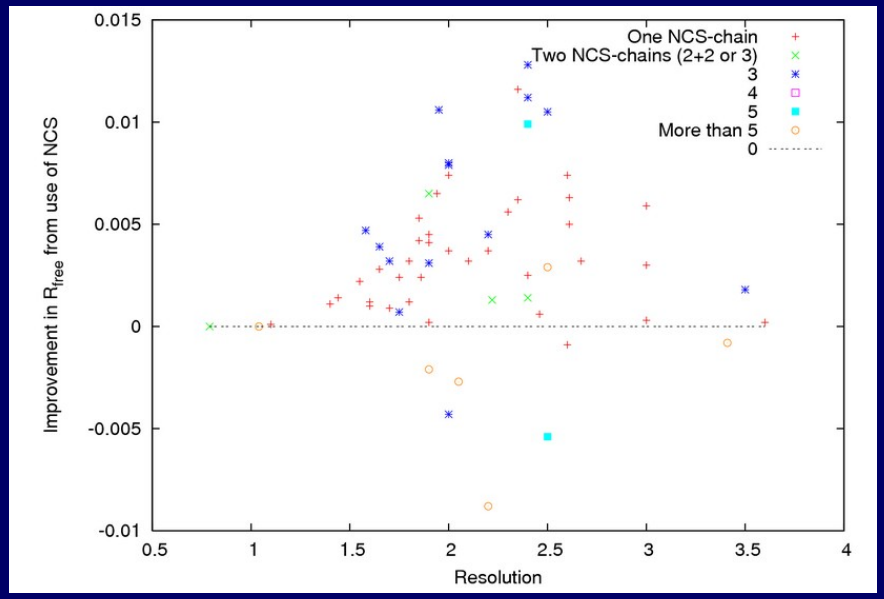
Buster improvement 1:

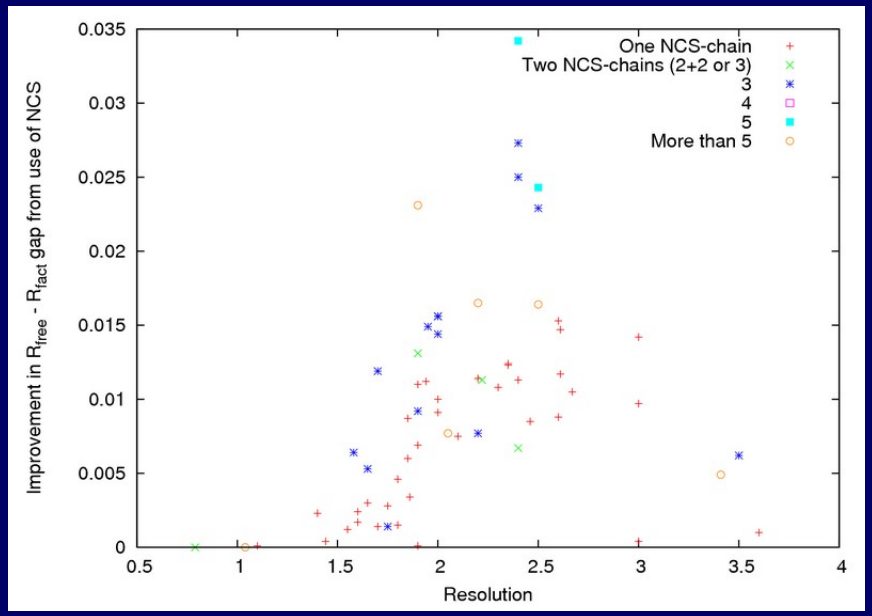
BETTER NCS *2008*

Better NCS by Local Structural Similarity Restraints (LSSR)

- Conventional superposition-based NCS is laborious to use.
- Developed LSSR a much easier to use approach to NCS in 2007, released 2008
- LSSR involves local contact distances
- Is much softer than superposition-based methods – violations entail only a fixed cost
- *-autoncs* method in BUSTER fully automated detection and application

Application of autoBUSTER – autoncs to re refinement of 60 pdb structures





- Conclusion: as can be expected using NCS produces slightly better Rfree and markedly better Rfree-Rwork gap
- -autoncs circumvents the “to use or not to use” dilemma

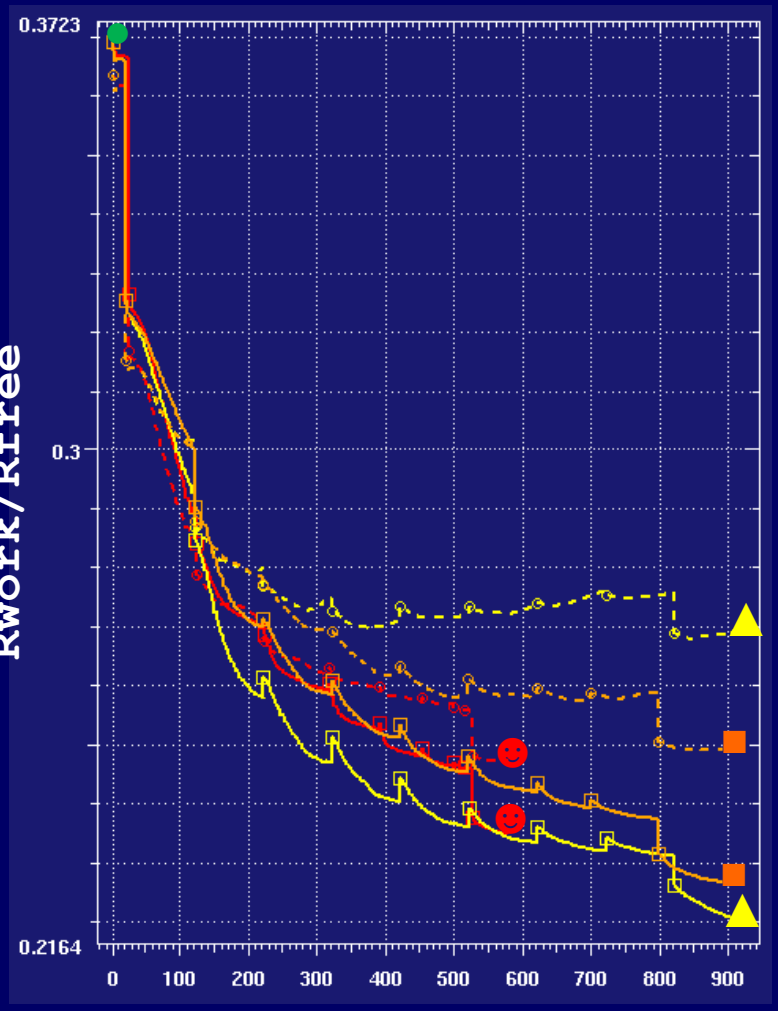
Buster improvement 2:

TARGET RESTRAINTS *2008*

LSSR Target Restraints

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

LSSR Target example: glutamate receptors *2e4y* 3.40Å resolution MR solution from *2e4u* 2.35Å resolution



structure	Rwork/ Rfree	Ramach % core	molprob score
MR solution	*0.354* *0.351*	91.7%	3.03
AB control no NCS no target	0.220 0.269	87.4%	3.22
AB -autoncs	0.227 0.249	91.0%	3.06
AB -target 2e4u -autoncs	0.235 0.247	92.2%	2.95

www.globalphasing.com/buster/wiki

Buster improvement 3:

TLS *2009*

Buster improvement 4:

IMPROVED GEOMETRY FUNCTION *2009*

2009 Improved Geometry Function

- Looked at exactly what TNT and refmac did for every geometry term
- Where better adopted refmac
- Improved things where possible
  - BOND & ANGLE EH99 in place of EH91
  - TORSION sinusoidal term now used
  - CONTACTS ideal distances including 1-4
- Result is an improved geometry function

Pietro C5 example: 2008 autoBUSTER worsens geometry quality indices

Job	*old autoBUSTER*	*Refmac*	*old autoBUSTER*	*Refmac*	*2008 autoBUSTER*
Molprobity Clashscore	172 0th percentile	43 55th percentile	131 3rd percentile	41 61st percentile	**155** **1st percentile**
Rotamer outliers	42.0%	19.5%	36.8%	19.7%	**41.8%**
Ramachandran outliers	12.8%	6.0%	9.0%	5.6%	**11.4%**
Ramachandran favoured	66.0%	79.0%	70.3%	80.3%	**68%**
MolProbity Score	4.80 0th percentile	3.84 24th percentile	4.61 1st percentile	3.81 25th percentile	**4.74** **1st percentile**
REMARK 500 bad angle	?	?	?	22	**37**

Pietro C5 example: 2009 autoBUSTER improves geometry quality indices

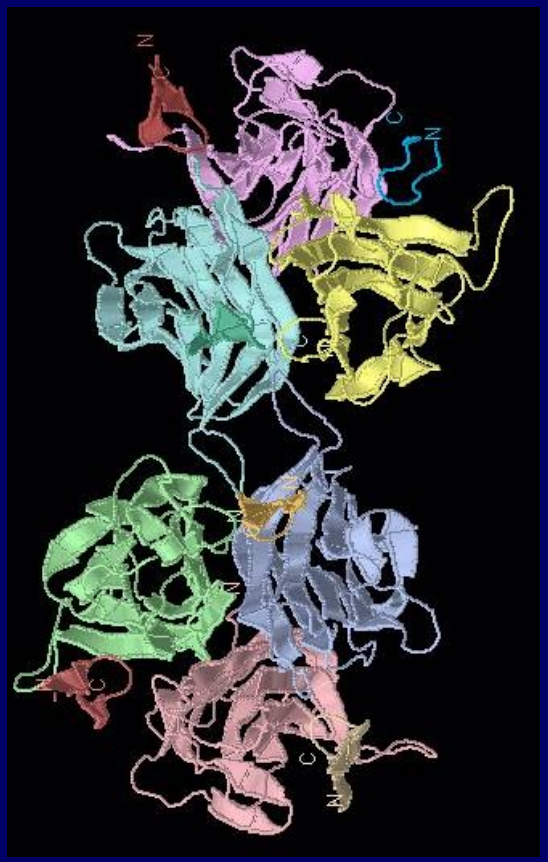
Job	*old autoBUSTER*	*Refmac*	*old autoBUSTER*	*Refmac*	*2009 autoBUSTER*
Molprobity Clashscore	172 0th percentile	43 55th percentile	131 3rd percentile	41 61st percentile	**18** **96th percentile**
Rotamer outliers	42.0%	19.5%	36.8%	19.7%	**17.9%**
Ramachandran outliers	12.8%	6.0%	9.0%	5.6%	**5.6%**
Ramachandran favoured	66.0%	79.0%	70.3%	80.3%	**81.9%**
MolProbity Score	4.80 0th percentile	3.84 24th percentile	4.61 1st percentile	3.81 25th percentile	**3.42** **51st percentile**
REMARK 500 bad angle	?	?	?	22	**1**

Buster improvements:

AN EXAMPLE SHOWS WHAT COMBINATION CAN DO

tutorial example: 1osg

- complex between BAFF with the peptide bhpBR3
- 3.0Å resolution
- Two protein trimers each binds cyclic peptide
- Originally refined with refmac including (weak) rmsD NCS
- Tutorial wiki example *www.globalphasing.com/buster/wiki*

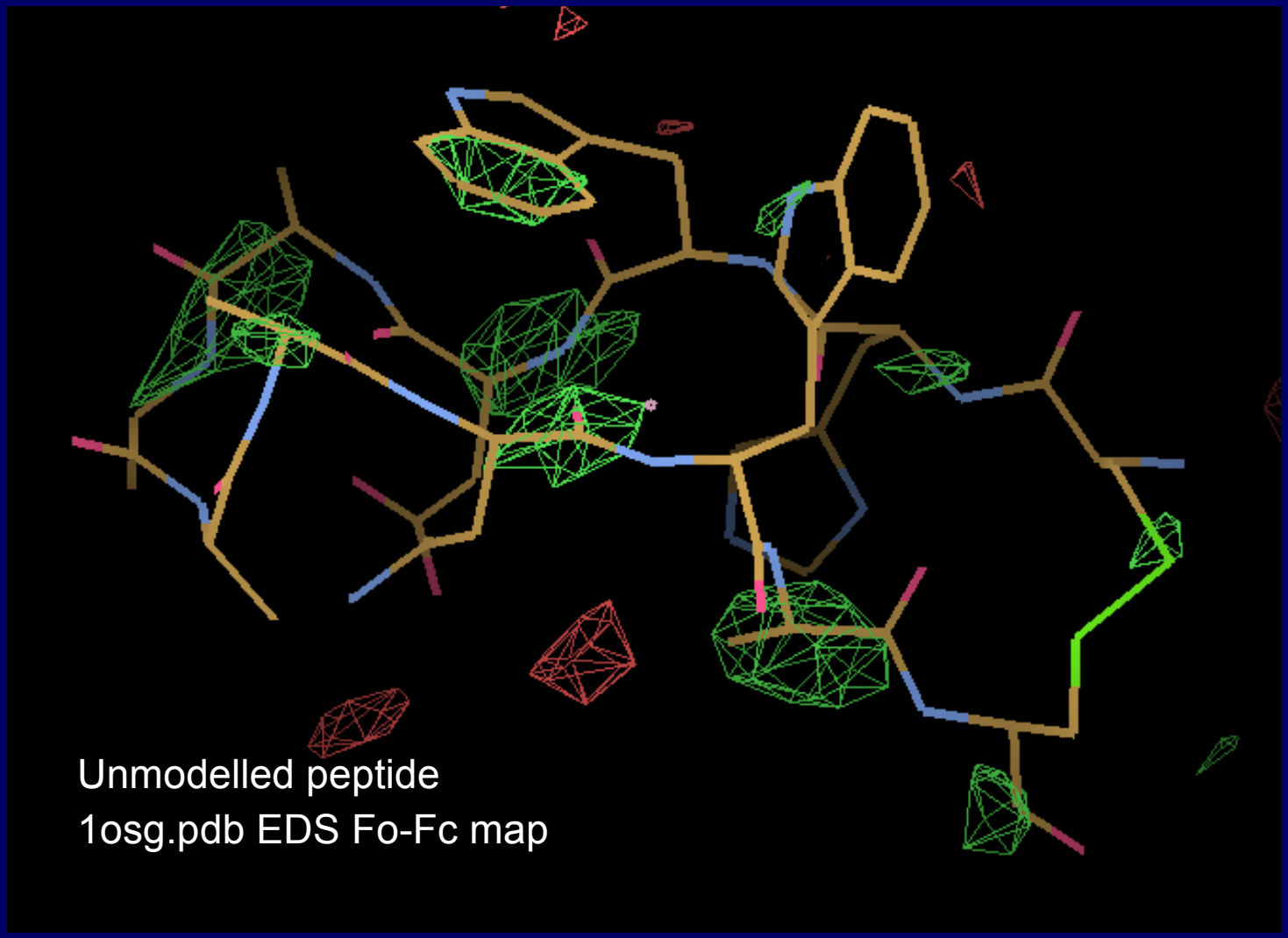


autoBUSTER refinement of 1osg

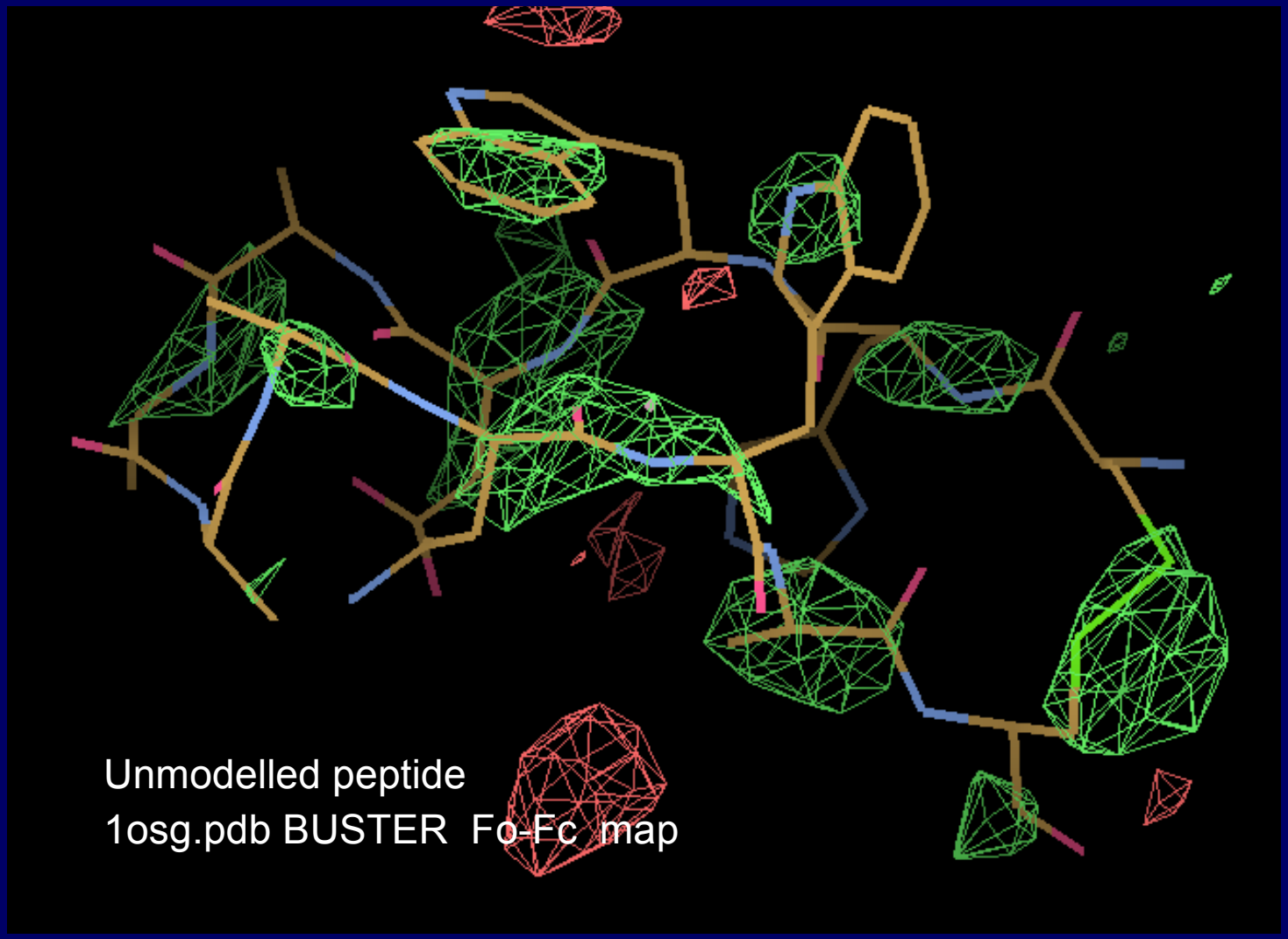
structure	BUSTER R_{work} R_{free}	Gap R_{free} –R_{work}	Molprobity Ramach. favored	Molprobity score
1osg pdb (refined B's)	0.185 0.243	5.8%	94.5%	2.81
autoBUSTER control no NCS	*0.169* *0.249*	*7.8%*	*95.4%*	*2.59*
autoBUSTER - autoncs	0.181 0.223	4.3%	96.4%	2.23
autoBUSTER - autoncs TLS	0.169 0.211	4.1%	96.5%	2.23

- autoBUSTER with automated LSSR NCS results in 2% drop in Rfree, better gap and better geometry
- TLS produces further improvements.

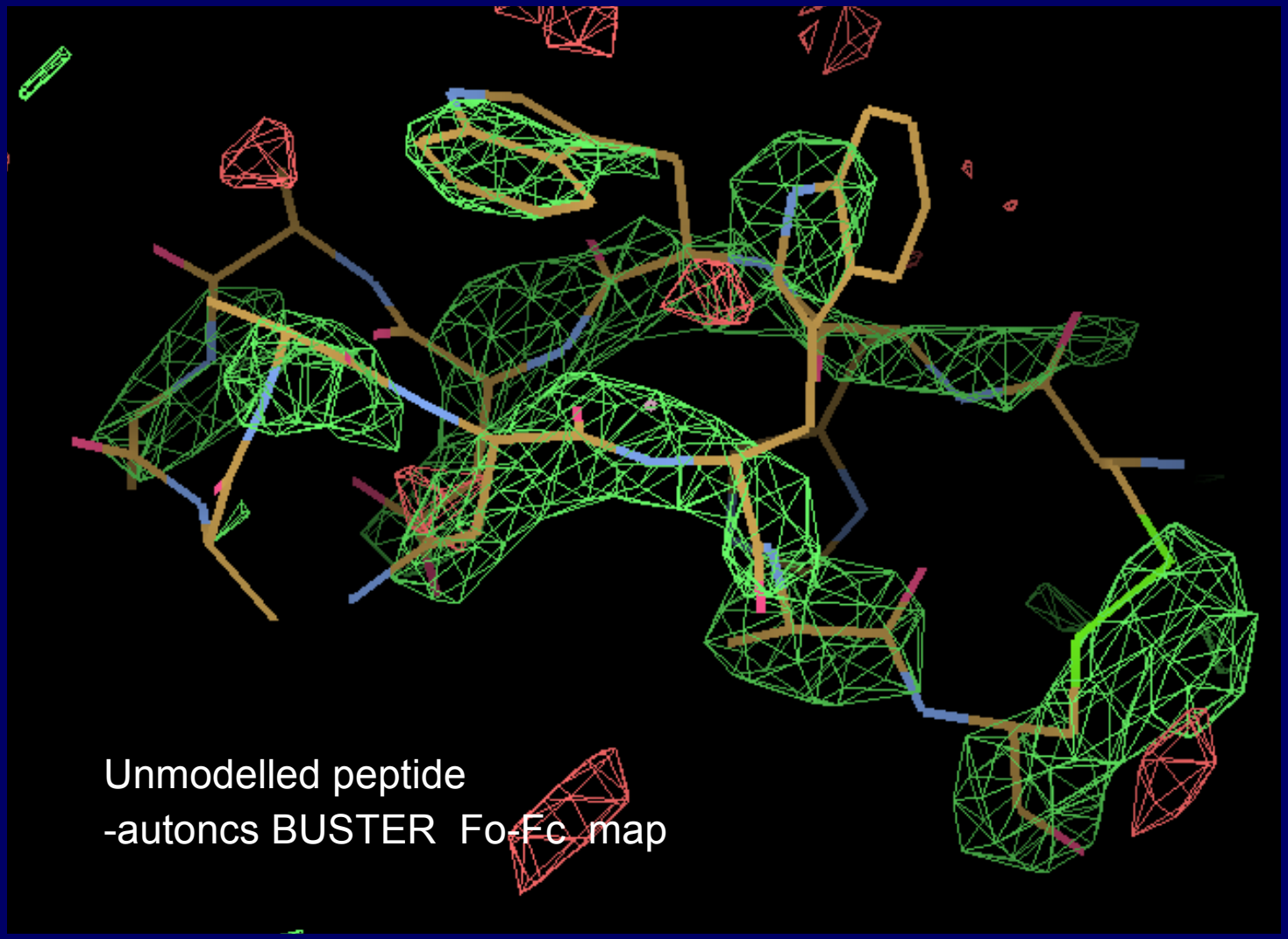
There is an extra peptide at a crystal contact: not clear from EDS



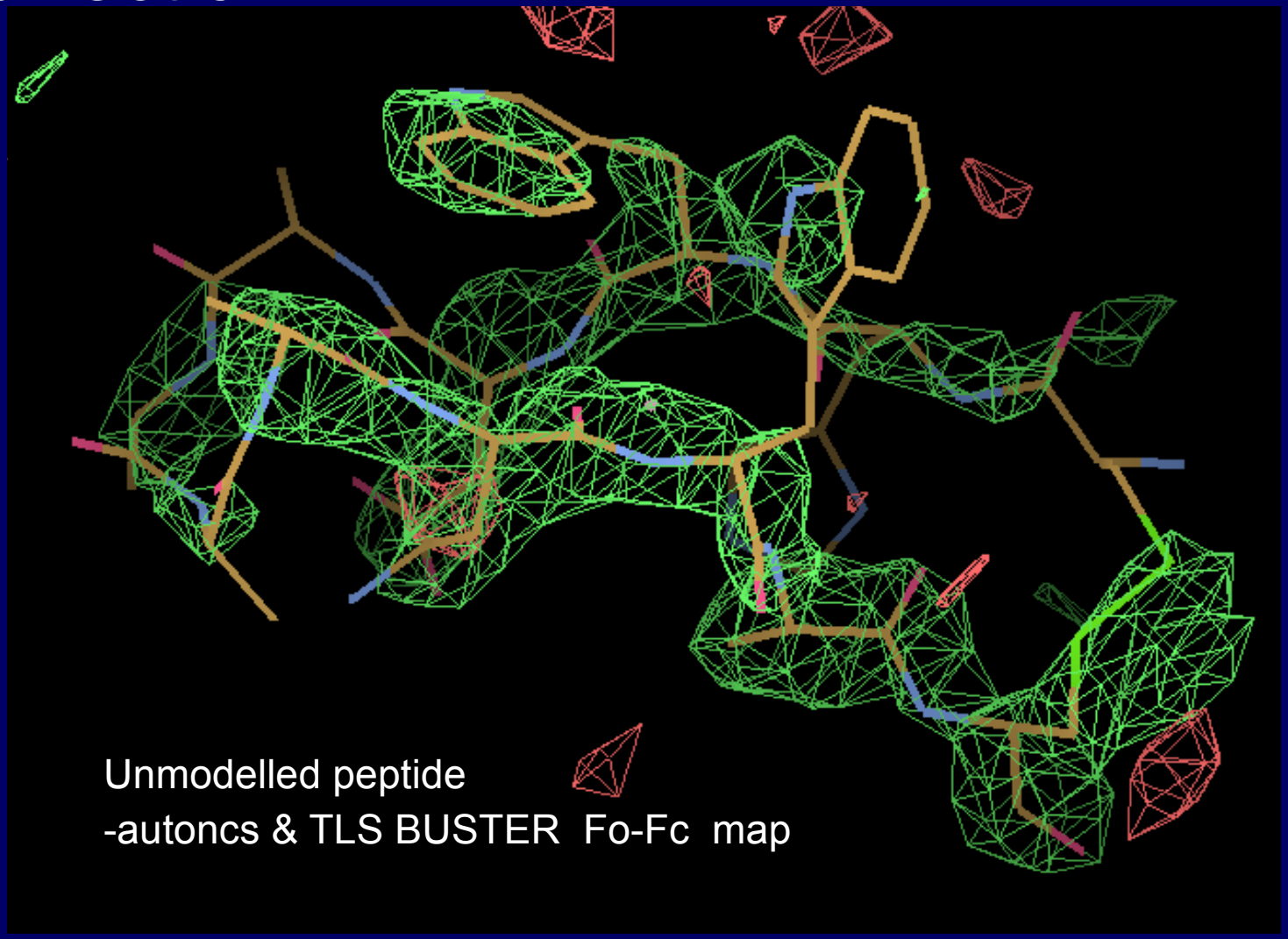
Extra peptide: initial BUSTER map



Extra peptide: -autoncs could just interpret



Extra peptide: clear from difference map with -autoncs & TLS despite bulk solvent correction

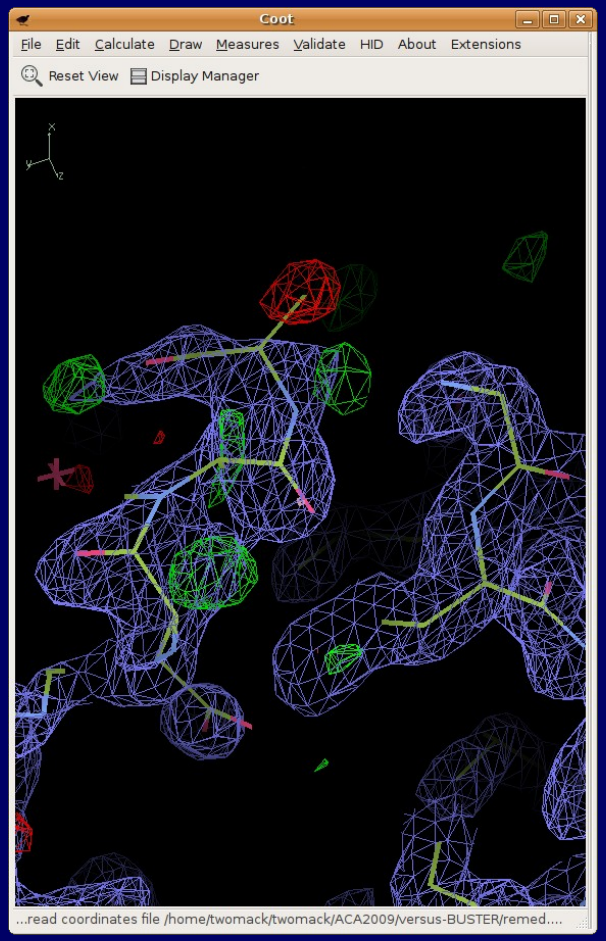


Second upgrade: results

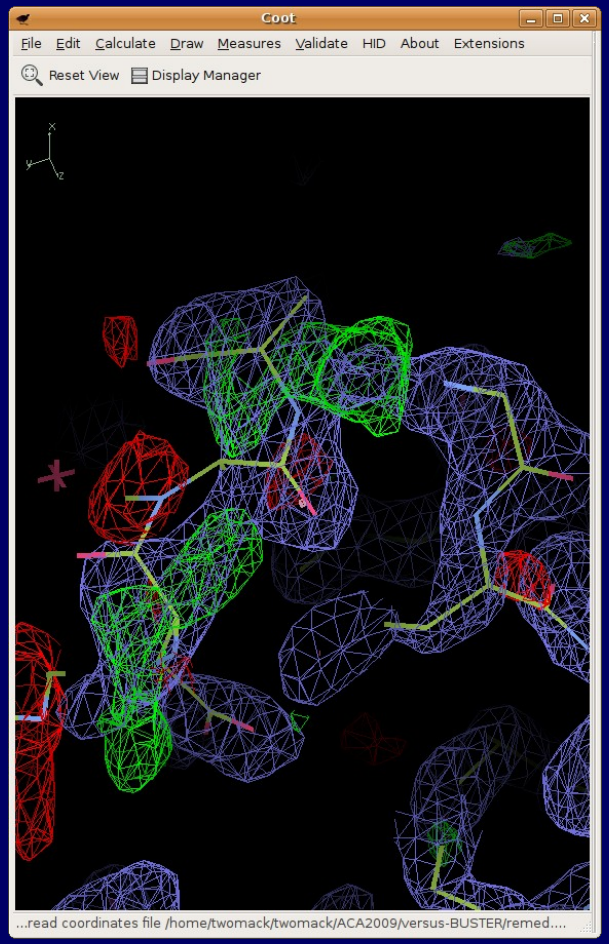
- Better R-values
  - Lower Rfree
  - Lower Rfree- Rwork gap
- Better model quality
  - Better Molprobity scores (even without explicit hydrogens)
  - Better Ramachandran plot statistics
- Better maps and difference maps
- Examples:
  - 2wfw (out of Tom Womack's weekly PDBRuns)
  - The RNA Polymerase II – TFIIB complex structure

2wfw: released 12 May 2009, 1.6Å resolution

Electron Density Server

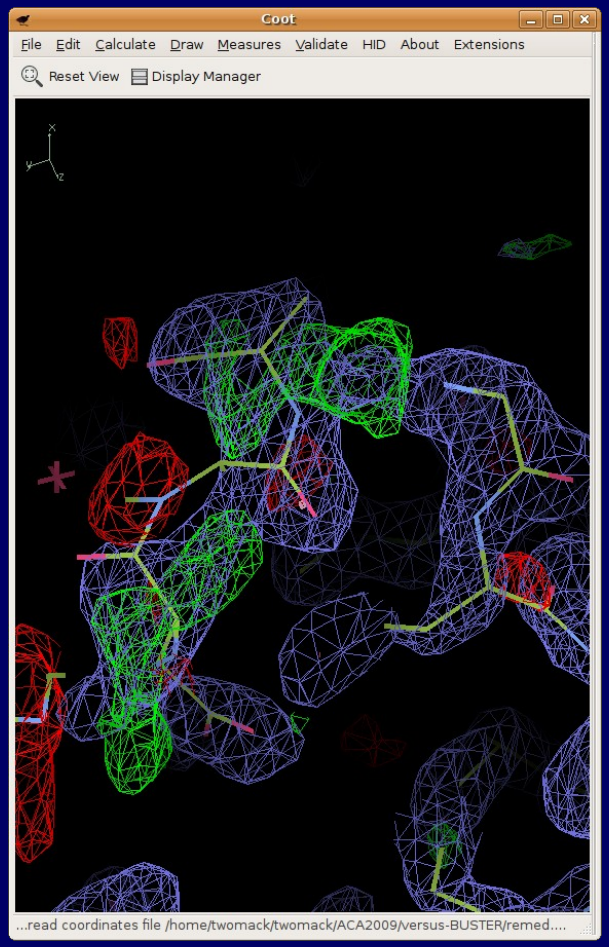


Initial BUSTER maps

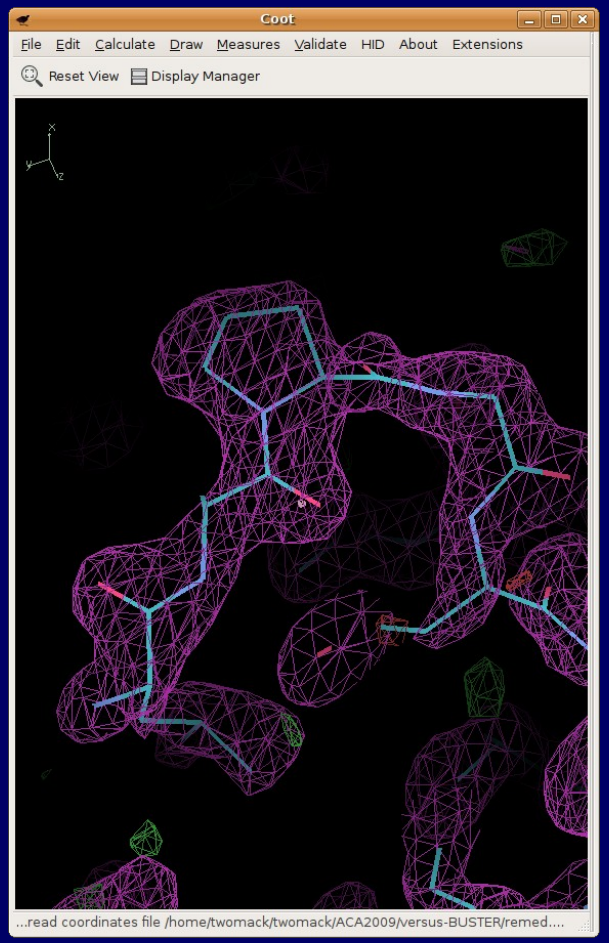


2wfw: released 12 May 2009, 1.6Å resolution

Initial BUSTER maps



Rebuilt & re-refined model



2wfw: released 12 May 2009, 1.6Å resolution

	As deposited	After default refinement	After remediation
Rwork	0.210	0.228	0.193
Rfree	0.246	0.263	0.215
Rotamer outliers %	4.70	3.69	1.78
Ramach. outliers %	3.32	3.88	0.29
Ramach. favoured %	94.5	93.35	98.3
Molprobity score	2.55	2.27	1.28
Molprobity percentile	12	26	97

American Crystallographic Association, Inc

2010 Annual Meeting
Chicago, IL July 24-29, 2010

Programme	Titles	Presenters	Topics
Poster Session II			

MONDAY, JULY 26, 05:30pm-07:30pm
River Exhibition Hall

poster M-228 Monday
Come and see if your pdb entry has been selected as interesting!

autoBUSTER re-refinement reliably reveals interesting features in newly-released PDB structures

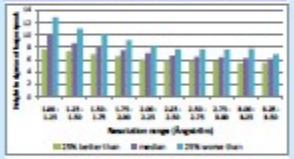
T. Womack, C. Flensburg, P. Keller, W. Paciorek, A. Sharff, O.S. Smart, C. Vonrhein, G. Bricogne*,
Global Phasing Ltd, Sheraton House, Castle Park, Cambridge CB3 0AX, UK
www.globalphasing.com

GΦL

Weekly PDB runs

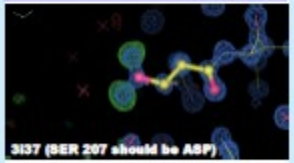
- About 150 new structures, with structure factors, get released by the PDB every week
- autoBUSTER can take these structures as they come and re-refine them; ~ 1 CPU-hour per structure
- autoNCS via LSSR[1] used throughout
- The maps produced are generally clearer than those calculated by the Uppsala EDS[2], and often show interesting features
- Automatic analysis of difference maps can flag these features
- **Other work in this area**: Robbie Joosten (CMBI) has[3] used refmac, concentrating on TLS decompositions; no equivalent of autoNCS

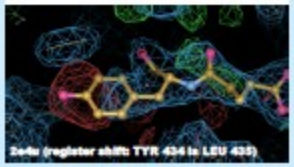
Difference map peaks are ubiquitous



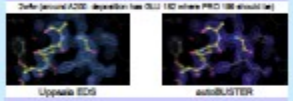
- Half of all structures have difference map peaks higher than 7σ
- The upper quartile corresponds roughly to omitting a water; higher levels are more interesting
  - Right-tailed distribution; some structures have lots of errors, few structures with lots of errors have no blatant ones
- Errors are much clearer at high resolution, so illustrations here are from high-res structures

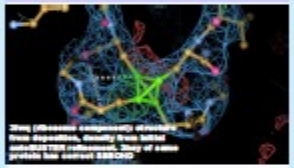
Sequence errors are not uncommon



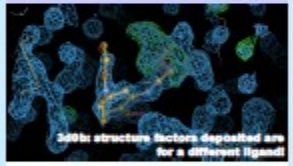


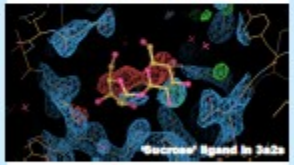
autoBUSTER density is clearer for wrongly-built loops

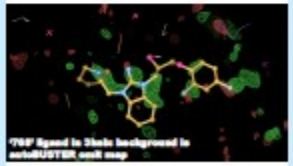


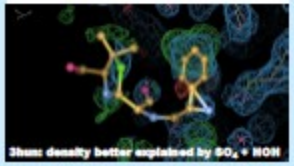


"But I'm only interested in the ligand"



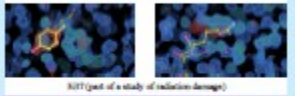






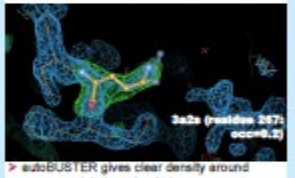
"But that's not what I'm studying"

- Crystallography is famously non-local
- autoBUSTER is good at dealing with model bias: is the same true of the protocol you're using for your metric of interest?

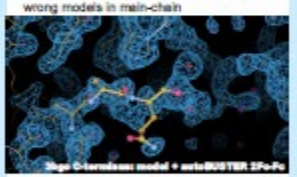


Problems on the main chain

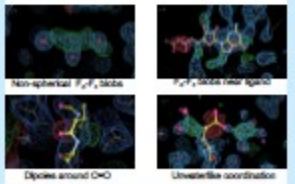
- Occupancies are not routinely refined; coot building can leave single model at occ<1



- autoBUSTER gives clear density around wrong models in main-chain



We are developing tools to identify and fix issues automatically



Acknowledgments

- Robbe Joosten and Gert Vriend, CMBI
- RCSB staff, particularly Rachel Kramer Green, for their responses to dozens of the presenter's emails
- Support:
  - Members of the Global Phasing Consortium
  - Victor project EC FP6 LSHG-CT-2004-511960

Acknowledgements

- GΦL
  - Tom Womack, Andrew Sharff, Clemens Vonrhein, Claus Flensburg, Wlodek Paciorek, Maria Brandl, Peter Keller, Gerard Bricogne
- BUSTER users
- Support:
  - Members of the Global Phasing Consortium,
  - Vizier Project FP6

Vol 462 | 19 November 2009 | doi:10.1038/nature08548

nature

ARTICLES

RNA polymerase II–TFIIB structure and mechanism of transcription initiation

Dirk Kostrewa[1]*, Mirijam E. Zeller[2]*, Karim-Jean Armache[1]*†, Martin Seizl[1], Kristin Leike[1], Michael Thomm[2] & Patrick Cramer[1]

To initiate gene transcription, RNA polymerase II (Pol II) requires the transcription factor IIB (B). Here we present the crystal structure of the complete Pol II–B complex at 4.3 Å resolution, and complementary functional data. The results indicate the mechanism of transcription initiation, including the transition to RNA elongation. Promoter DNA is positioned over the Pol II active centre cleft with the 'B-core' domain that binds the wall at the end of the cleft. DNA is then opened with the help of the 'B-linker' that binds the Pol II rudder and clamp coiled-coil at the edge of the cleft. The DNA template strand slips into the cleft and is scanned for the transcription start site with the help of the 'B-reader' that approaches the active site. Synthesis of the RNA chain and rewinding of upstream DNA displace the B-reader and B-linker, respectively, to trigger B release and elongation complex formation.

Diffraction data were collected in 0.75° increments at the protein crystallography beamline ID 29 at ESRF. Diffraction data were processed with XDS and scaled with XSCALE[4]. The structure was solved by molecular replacement with PHASER[5] using the first 12-subunit Pol II without nucleic acids of the "EC I" complex[6]. The structure was refined with PHENIX[7] and with BUSTER[8] using TLS, group B-factors per residue, tight Engh-Huber geometrical restraints, and additional hydrogen-bond distance restraints for secondary structural elements (DK, unpublished data) against the observed data that were sharpened[9] with a B-factor of -80 Å^2. BUSTER was used initially either with a coarse mask encompassing the whole Pol II cleft describing the expected region for the missing TFIIB (B) atoms, or by specifying the expected number of missing B atoms. In later refinement stages, no expectation of missing atoms was given. BUSTER produced clearer electron density maps, lower R-factor and Free-R-factor with better geometry. Data processing and refinement statistics are shown Supplementary Table 1.