Recent developments in BUSTER: autoNCS, targetting, and improved ligand restraints

Oliver S. Smart

Tom Womack

Global Phasing Ltd

CCP4 Study Weekend

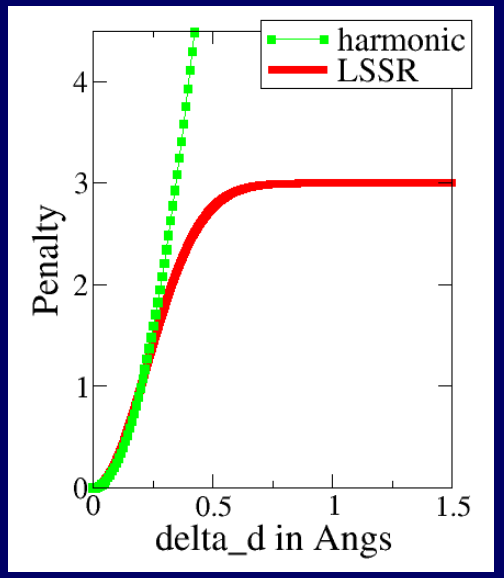
6th January 2011

Acknowledgements

- CCDC
- Martin Field, Alexei Vagin, Garib Murshudov, openbabel developers
- BUSTER users
- Support:
  - Members of the Global Phasing Consortium,
  - Vizier Project FP6
- GΦL
  - Tom Womack
  - Maria Brandl
  - Clemens Vonrhein, Claus Flensburg, Wlodek Paciorek,, Peter Keller, Andrew Sharff, Gerard Bricogne

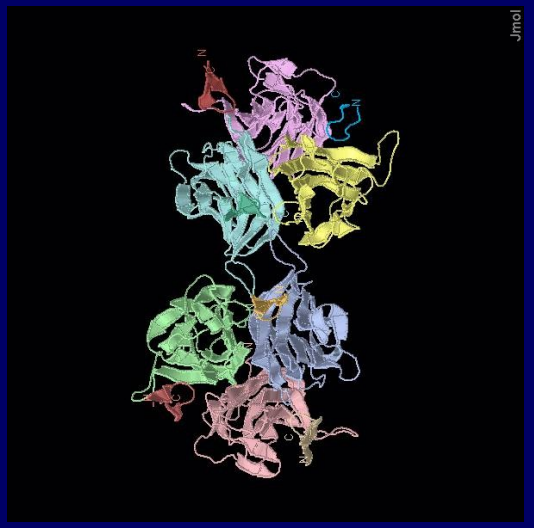
Better NCS by Local Structural Similarity Restraints (LSSR)

- Conventional superposition-based NCS is laborious to use.
- Developed LSSR a much easier to use approach to NCS
- 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
  - presented at ACA 2008 and 2009
  - released 7/2008 to companies,
  - released 7/2009 to academics



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*

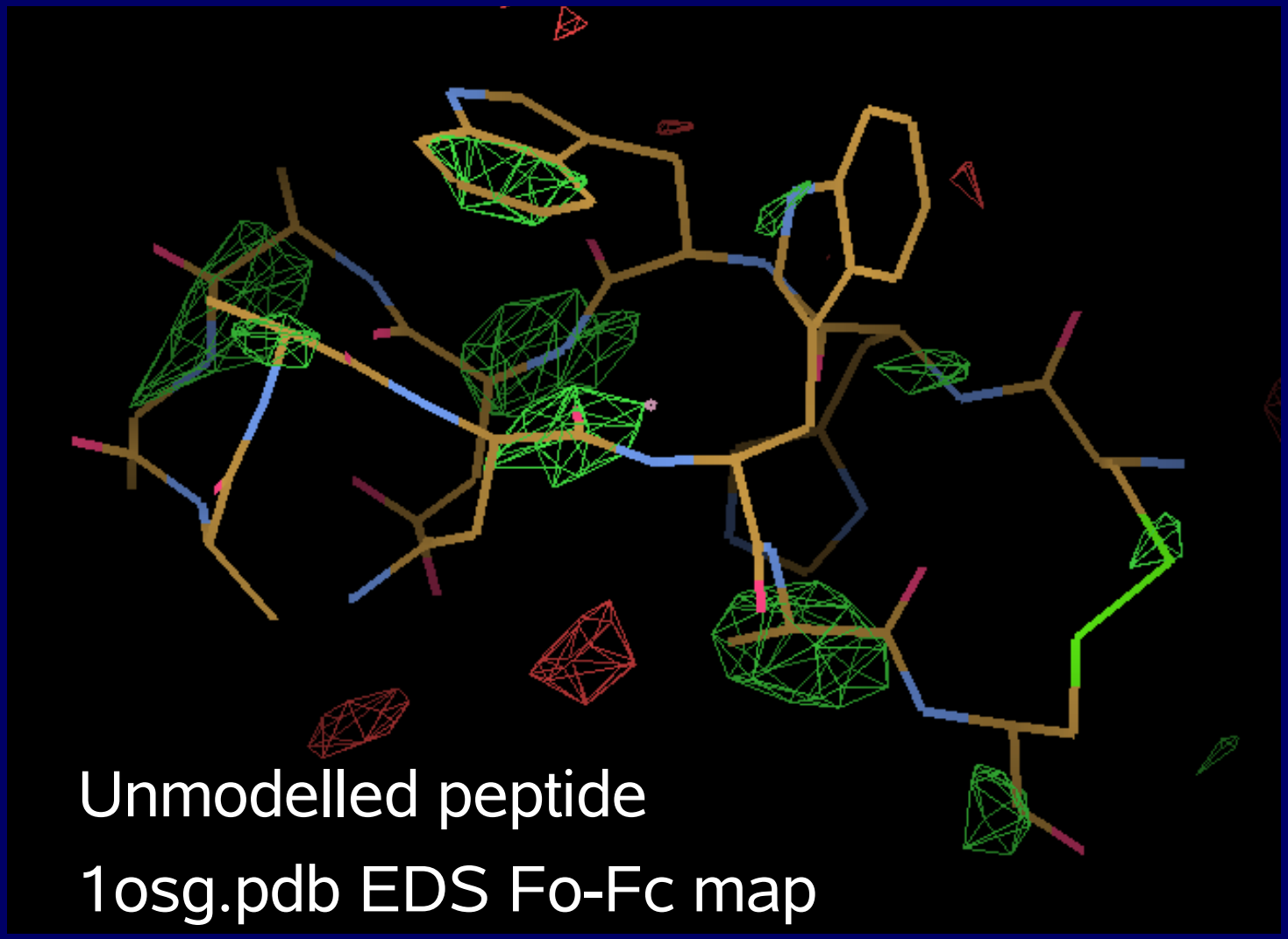


autoBUSTER refinement of 1osg

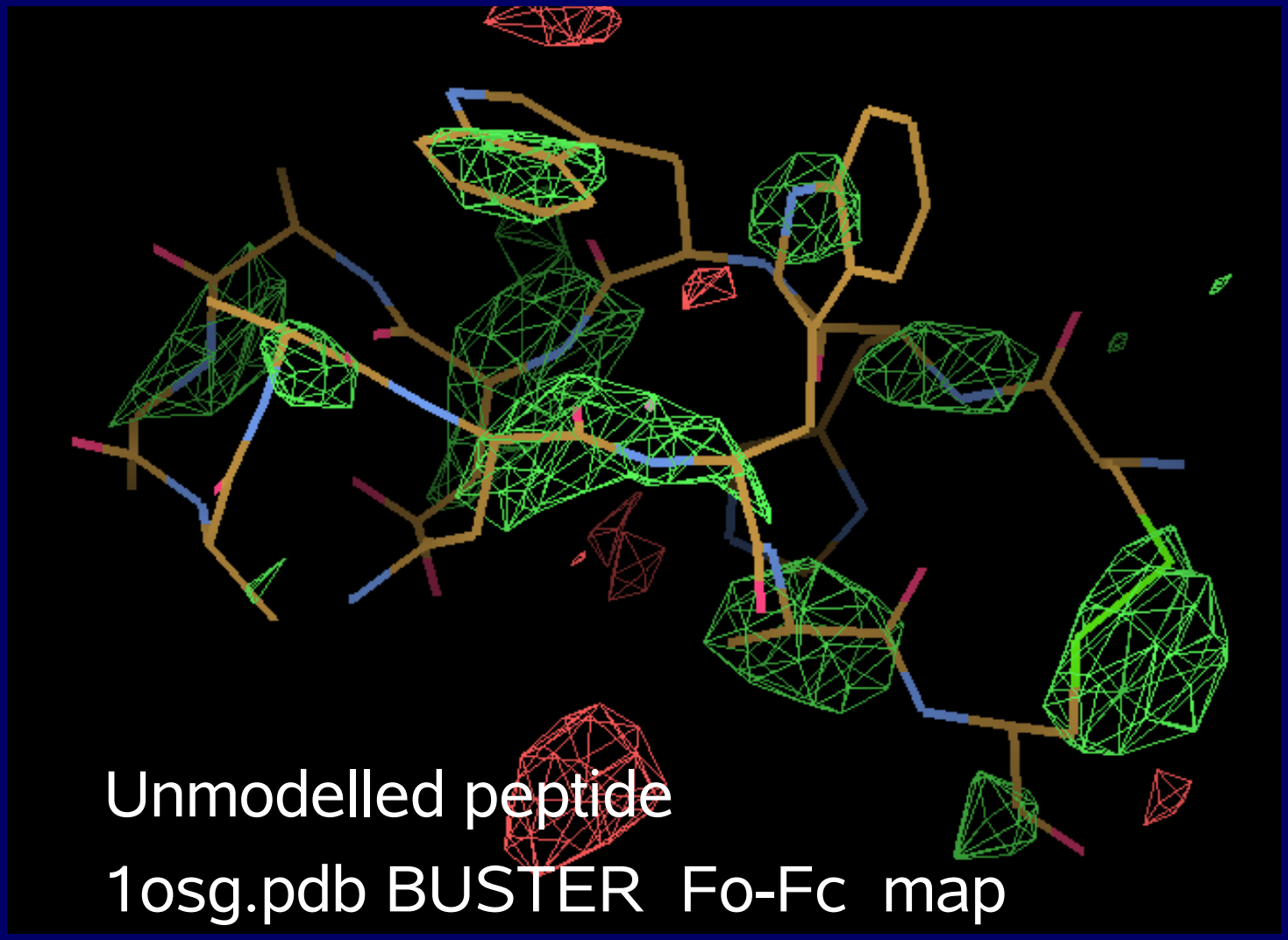
structure	BUSTER R_{work} R_{free}	Gap $R_{free} - R_{work}$	Molprobity Ramach. favored	Molprobity score
1osg pdb (refined B's)	0.185 0.243	5.8%	94.5%	2.81
autoBUSTER control no NCS	*0.169 0.249*	*7.8%*	*95.4%*	*2.59*
autoBUSTER - autoncs	0.181 0.223	4.3%	96.4%	2.23
autoBUSTER - autoncs TLS	0.169 0.211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

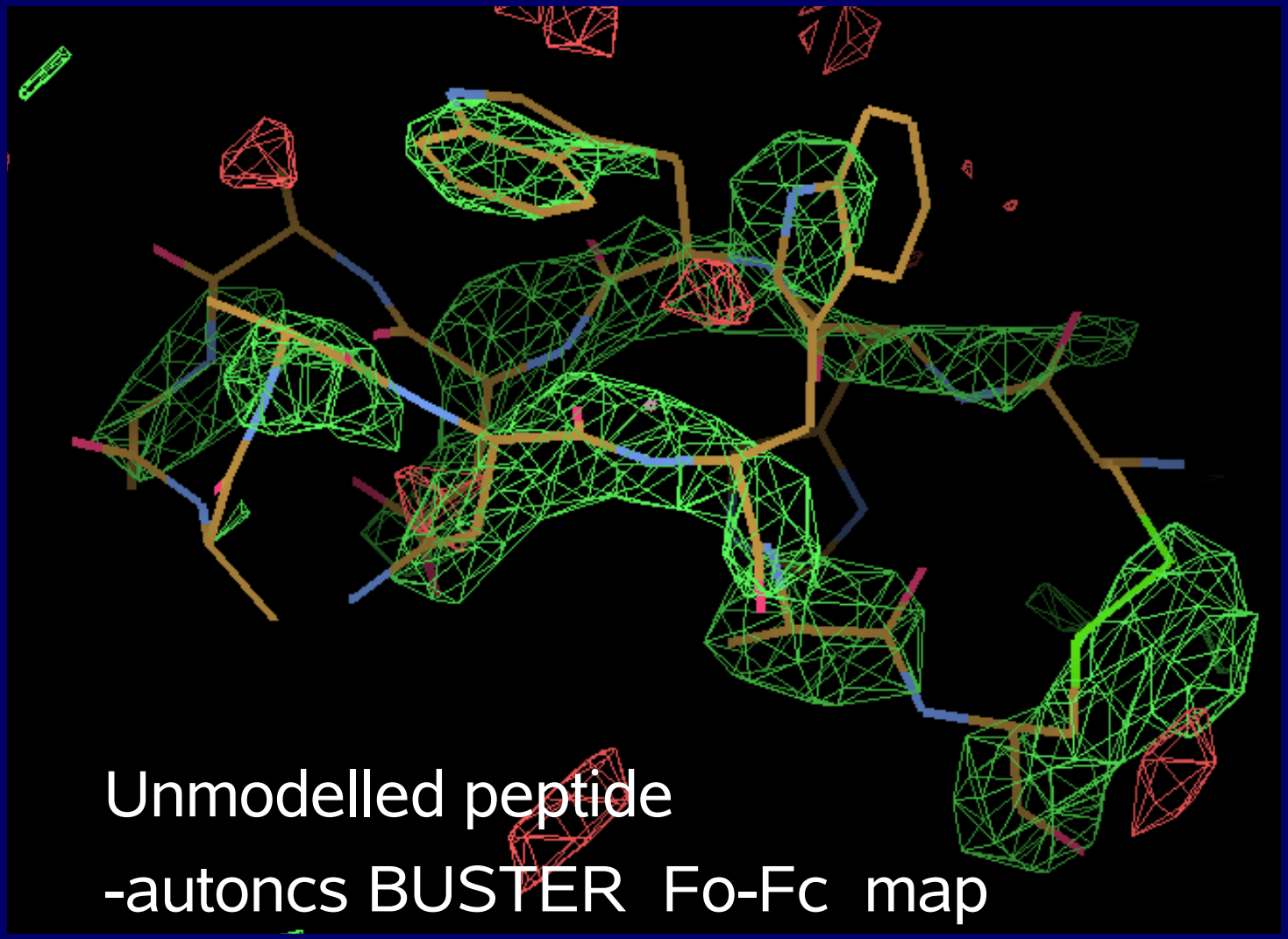
There is an extra peptide at a crystal contact: not clear from EDS



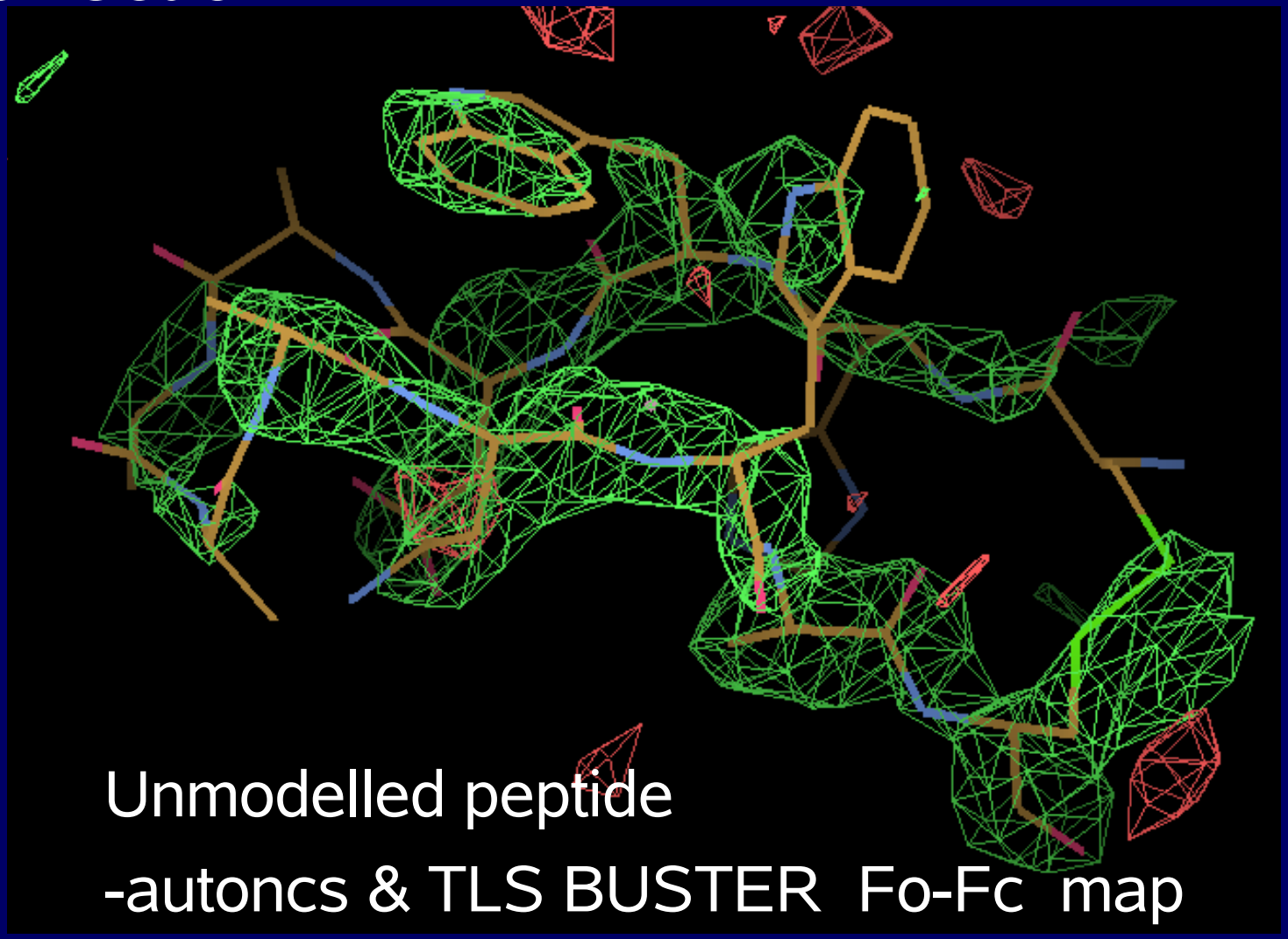
Extra peptide: initial BUSTER map



Extra peptide: -autoncs could just interpret



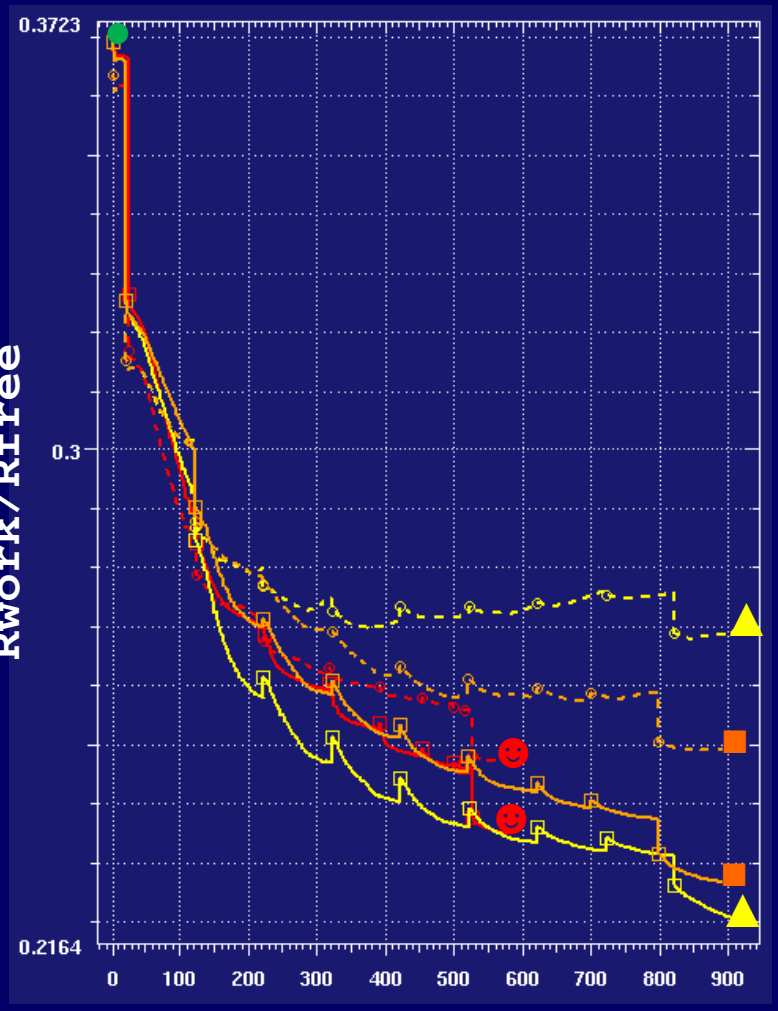
Extra peptide: clear from difference map with -autoncs & TLS despite bulk solvent correction



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



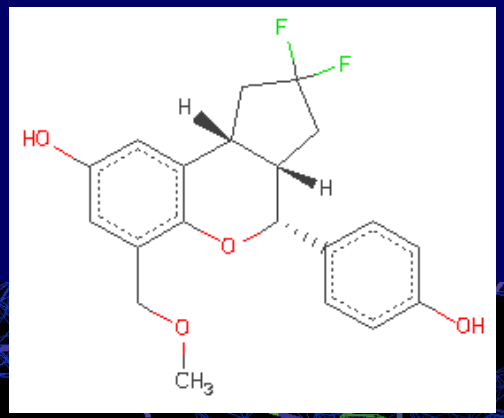
structure	Rwork/ Rfree	Ramach % core	molprob score
MR solution	*0.354* *0.351*	91.7%	3.03
AB control no NCS no target	0.220 0.269	87.4%	3.22
AB -autoncs	0.227 0.249	91.0%	3.06
AB -target 2e4u -autoncs	0.235 0.247	92.2%	2.95

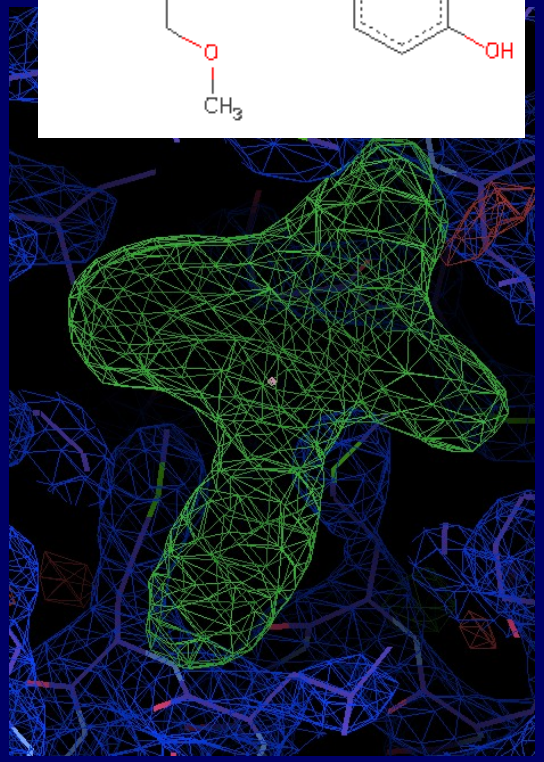
www.globalphasing.com/buster/wiki

Better ligand restraints

Fitting & refining a ligand in protein-complex

- Task is to fit ligand into F_o-F_c and then refine complex
- need prior knowledge of ligand chemistry to interpret density
- Use this in fitting to assess accessible low strain conformations that ligand can adopt
- And then in refinement to keep ligand conformation realistic





We want ligand restraints that are compatible with protein restraints

- For refinement & fitting use restraints like

$$V_{angle} = \Sigma[(\theta_{actual} - \theta_{ideal})/\sigma]^2$$

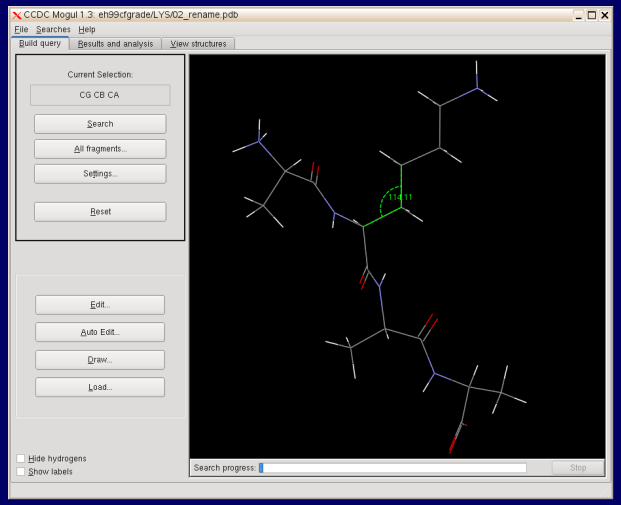
- For the protein we use restraints:

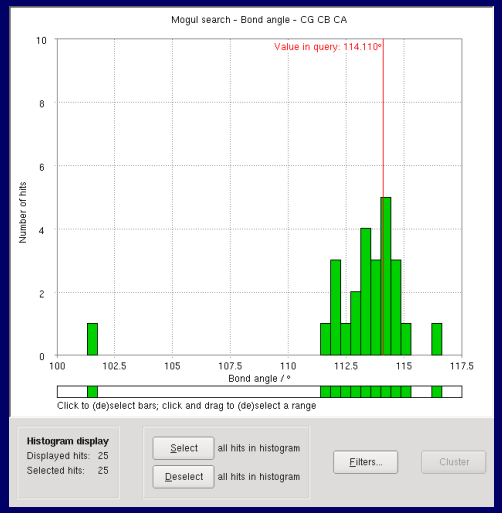
term	source	from
BOND	Engh & Huber EH99	CSD small molecule crystal structure
ANGLE	Engh & Huber EH99	CSD
χ TORSION	CCP4: Priestle, Richardsons	Protein structures
PLANE	TNT	

- We want EH99-like restraints for ligands

grade: ligand dictionaries based on CSD information where possible

- Use CCDC mogul program to survey CSD

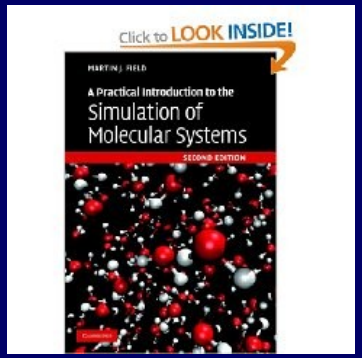




- grade uses mogul in batch mode!
- Use mogul as source of information for restraints not in validation
- Including for torsion restraints

grade: ligand dictionaries use QM if no CSD information

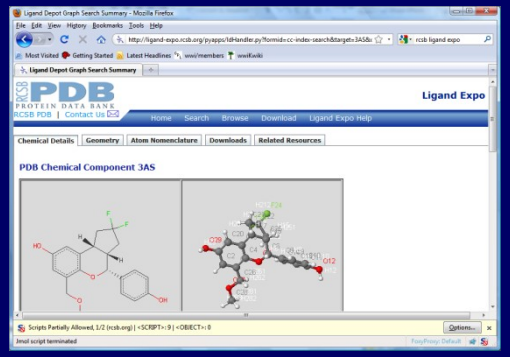
- What about cases where too little information from CSD?
- Use quantum chemistry
  - as pioneered by eLBOW
  - normally use RM1 semi-empirical method as implemented in dynamo (Martin Field, IBS Grenoble)
- Only used where mogul does not provide CSD data.



pDynamo

grade tool

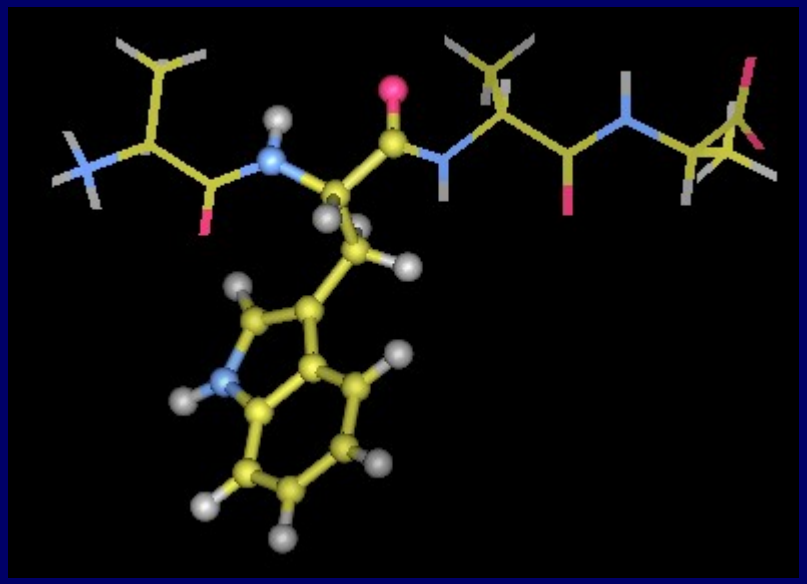
- Needs CCDC mogul licence
- Input:
  - mol2 coordinates
  - cif dictionary from other dictionary generators
  - smi smiles string *uses libcheck*
- *grade_PDB_ligand*
  - tool for existing PDB ligand
  - fetches from **rcsb ligand_expo site**
  - produces grade dictionary



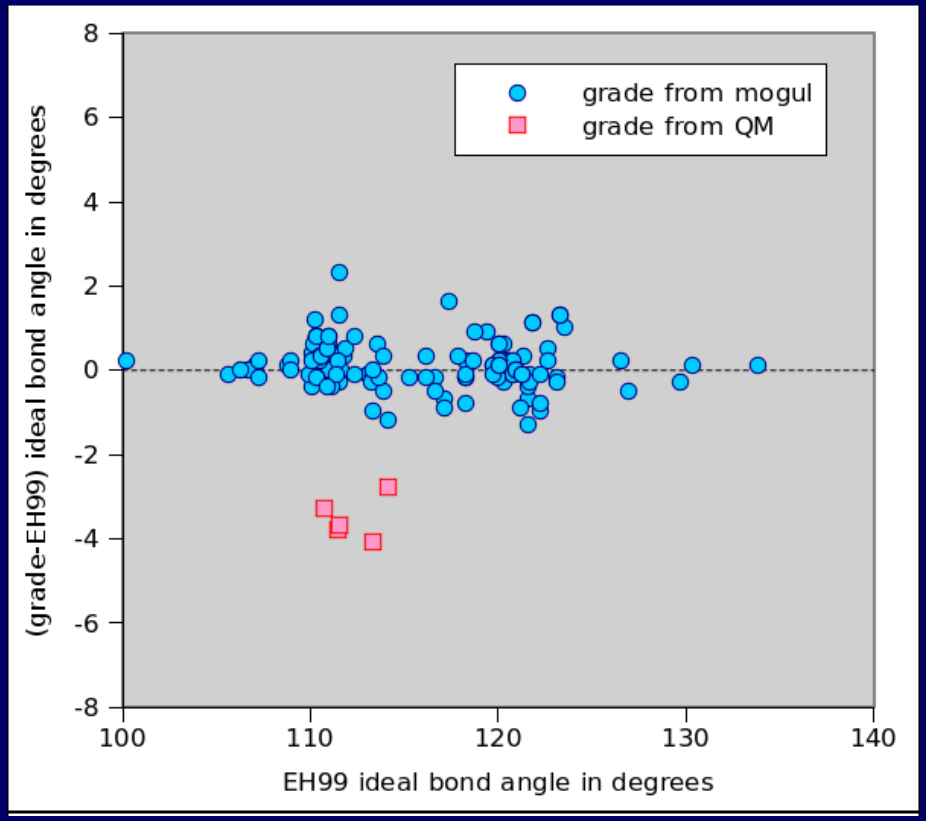
- Produces cif restraint dictionary for use in
  - BUSTER and rhofit
  - coot
  - refmac
  - ....
- Will be included in upcoming BUSTER Academic Release January 2011

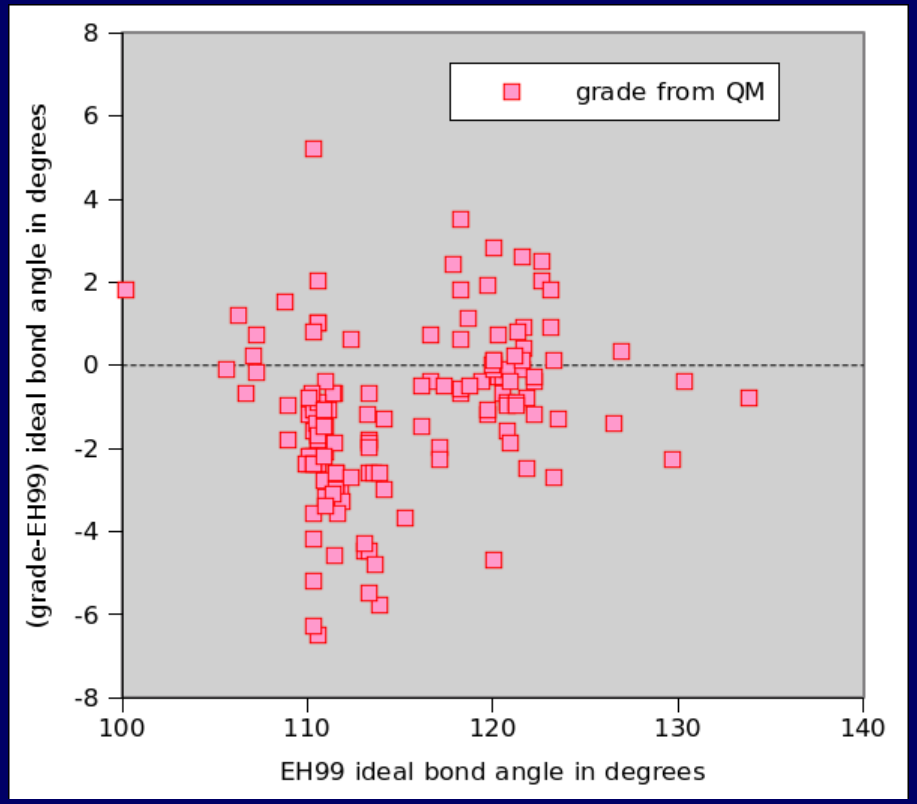
Test grade on standard amino acids

- Test grade on “ligands” like
  ALA-**TRP**-*ALA-ALA*
- grade produce bonds and angles like EH99 including sigmas
- Sets sensible torsion restraints
- Gets planes correct
- By-product is new EH-like protein dictionary: `protgeo_grade2011.dat`
- Based on current much larger CSD 2011 database cf 1999



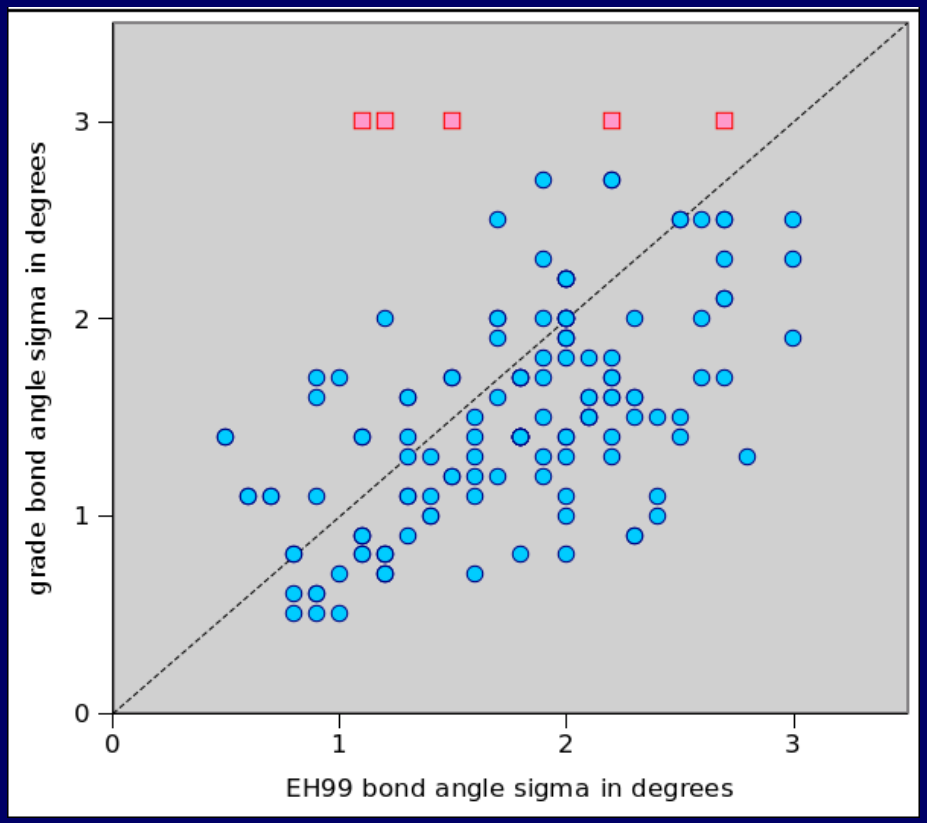
standard amino acids: ideal bond angles

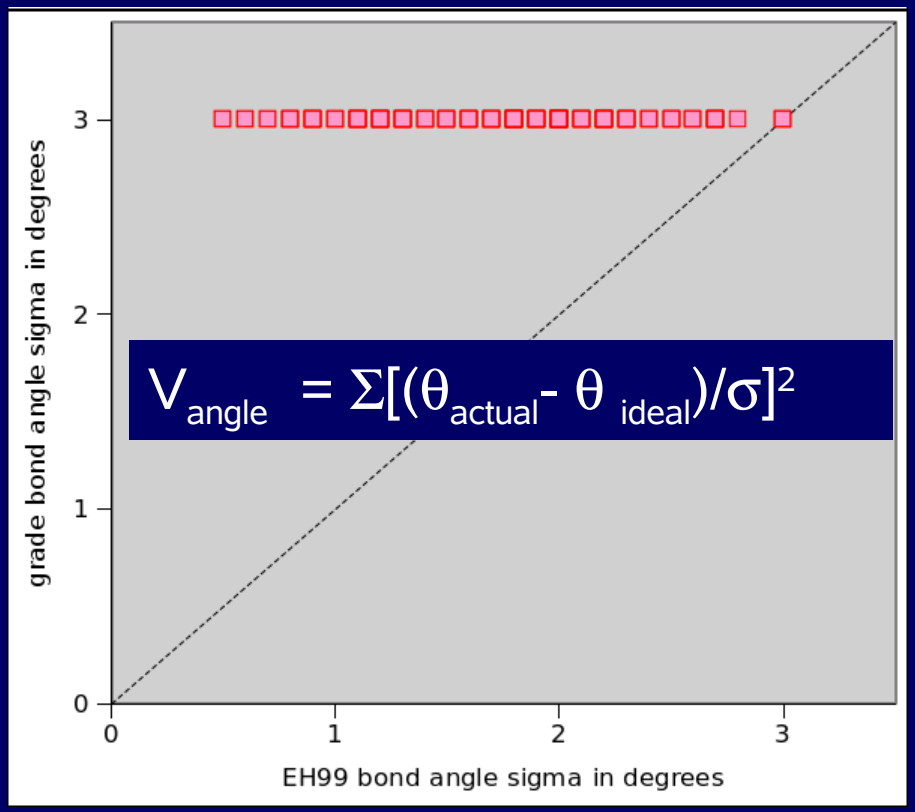




- Using mogul grade can quickly reproduce EH99 ideal bond angles
- Just using QM (RM1) results in much greater deviations

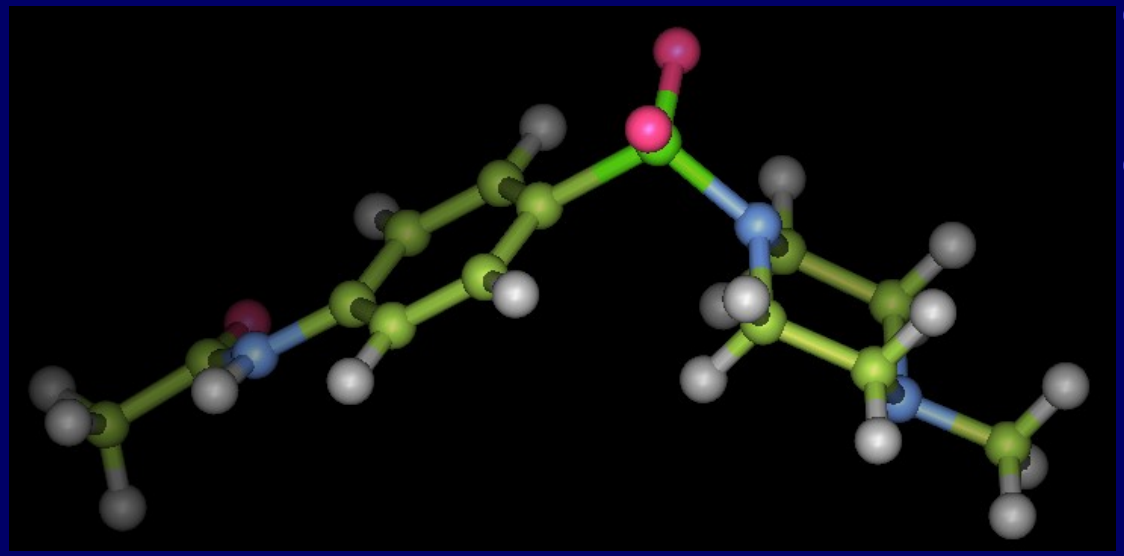
standard amino acids: bond angle σ





- σ value important
- Using mogul grade yield EH99-like angle σ's
- No method (yet) to get σ information from QM: set to 3.0

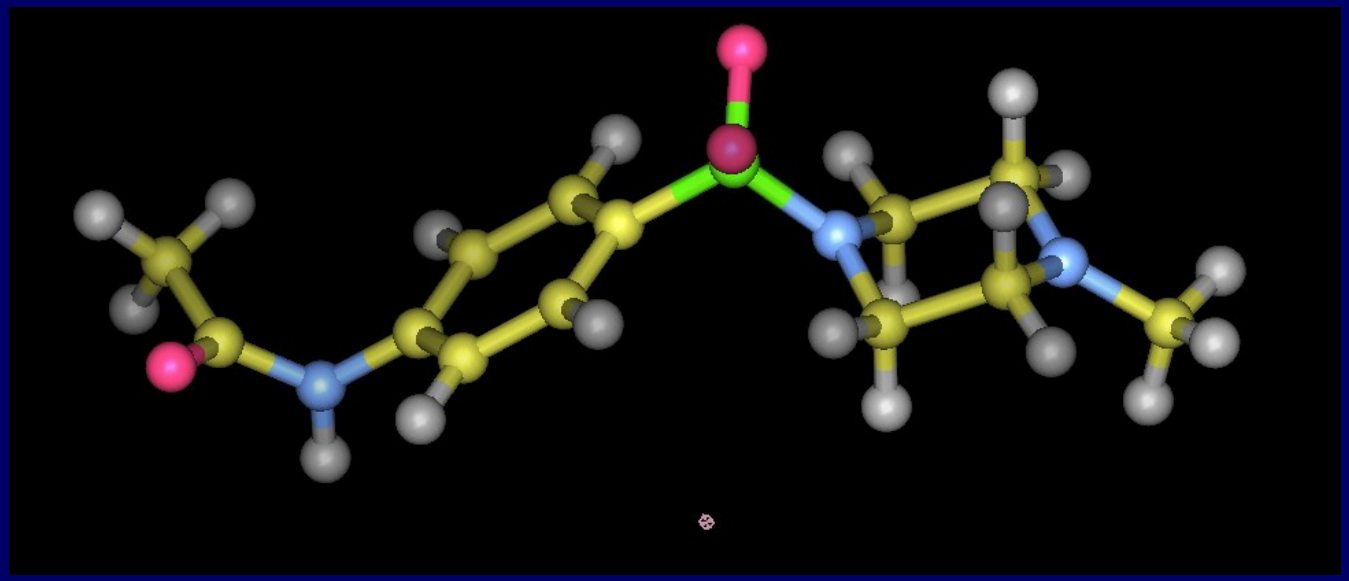
Lets look at one ligand-like example from csd: EVIDUI



`S(=O)(=O)(N1CCN(CC1)C)c1ccc(NC(=O)C)cc1`

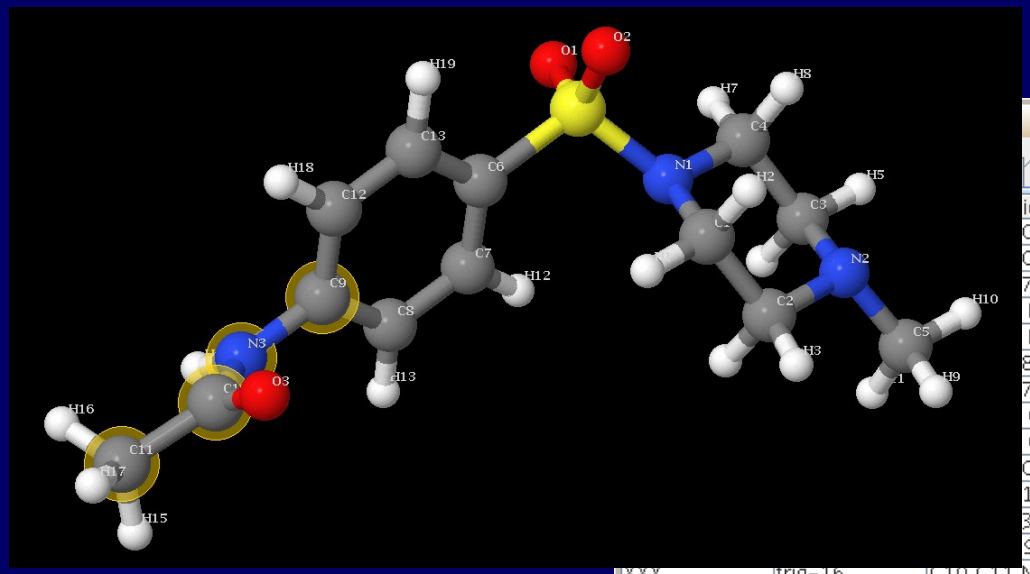
- Like ½ viagra with a peptide attached
- csd structure:
  - peptide planar but not coplanar with phenyl ring
  - Both nitrogen atoms pyramidal
- Create dictionaries from smiles
- Score dictionary against csd structure

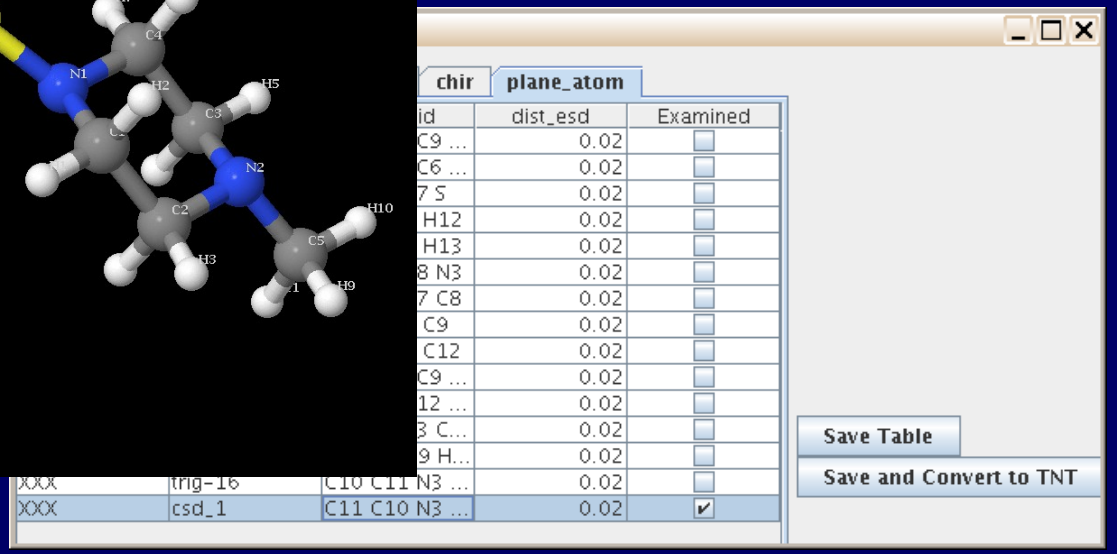
EVIDUI – 3rd party dictionary



- dictionary from *X*, ideal coords
- Test dictionary by “scoring” csd structure
- piperazine nitrogens exact planar but should be ~tetrahedral
- peptide plane missed.
- rms bond deviation 0.048Å rms angle 5.0°

EVIDUI grade with mogul





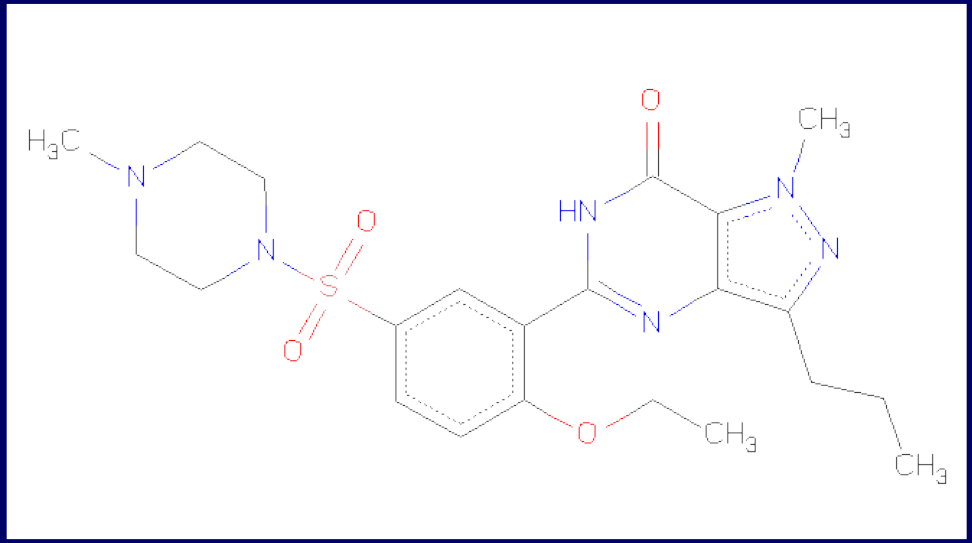

```
grade 'S(=O)(=O)(N1CCN(CC1)C)c1ccc(NC(=O)C)cc1'
```

- Piperazine nitrogen atoms correct pyramidal
- Peptide bond: torsions restraints hold trans.
- rms bond 0.006Å rms angle 0.6°

Example: grade dictionary for Viagra

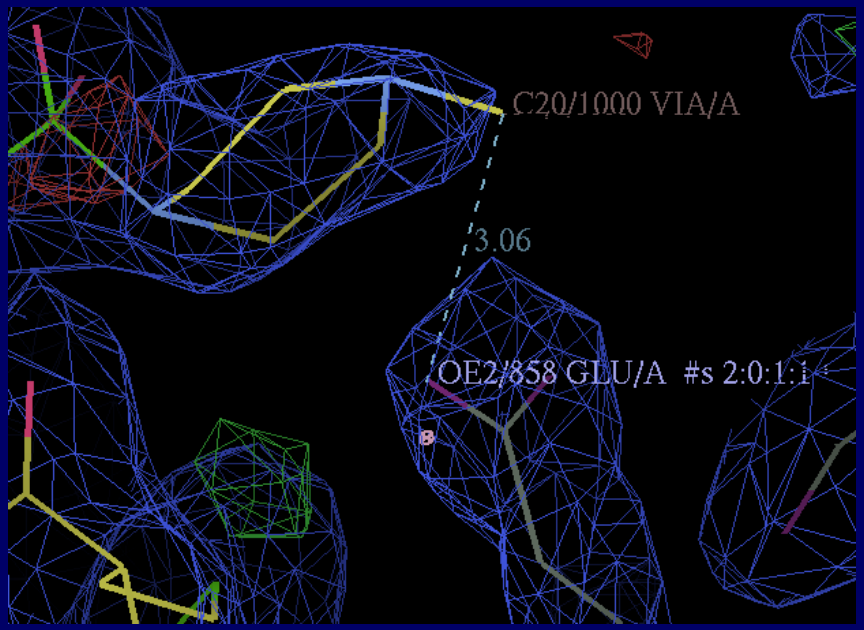
BUSTER refinement of

- 1udt : Human Phosphodiesterase 5 complexed with Sildenafil (Viagra)
- 2.3Å resolution



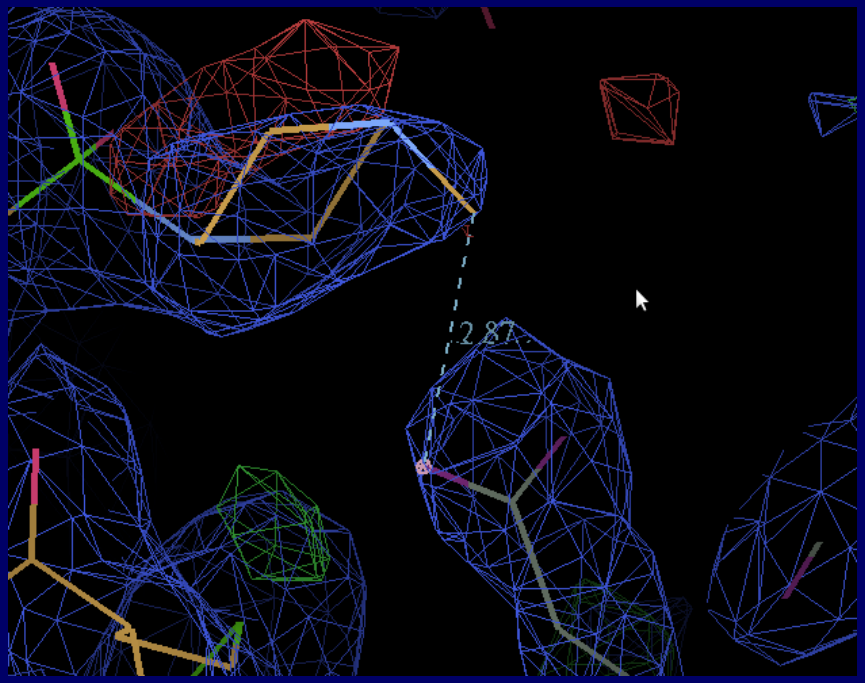
CCCc1nn(C)c2C(=O)NC(=Nc12)c3cc(ccc3OCC)[S](=O)(=O)N4CCN(C)CC4

1udt refinement VIA frozen



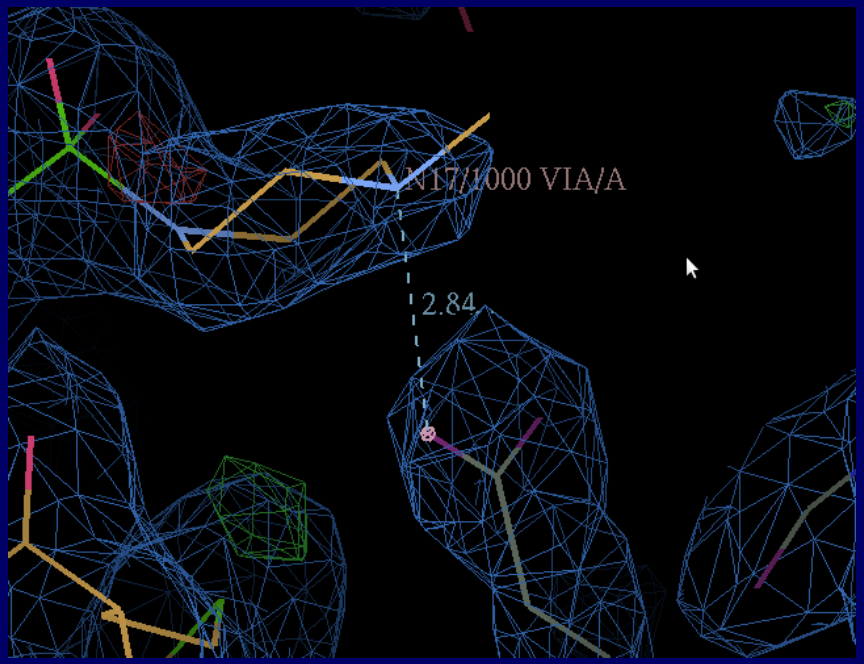
- Protein refined with BUSTER
- But ligand frozen at 1udt.pdb position
- Piperazine ring has poor stereochemistry

1udt refinement grade dictionary



- VIA grade dictionary
- piperazine ring forced to boat form
- Rfree improved by 0.3%
- Methyl group in conflict with GLU
- Creates difference density
- Conclusion: methyl should be equatorial not axial

1udt refinement VIA rebuilt



- methyl tweaked in coot
- Re-refined with grade
- N17 known to be protonated in Viagra
- Forms salt bridge to 858* GLU
- A good dictionary prevents ligand from adopting unrealistic conformation and so reveals mistakes

If QM is good for dictionaries: why not use it directly?

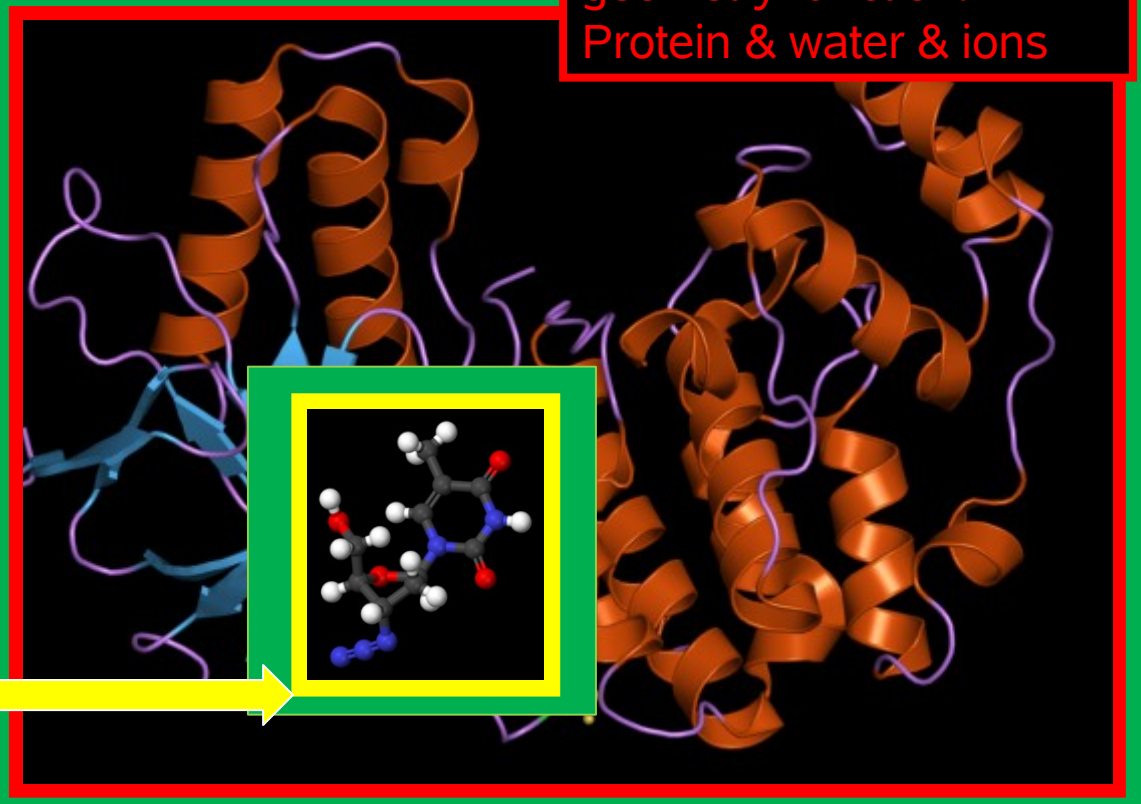
- Why approximate RM1 into a model function?
- RM1 Energy/gradient calculation is faster than BUSTER ML X-ray
- So adapt modelling MM/QM method to refinement

Direct use of quantum chemical method as part of geometry function

- Simply use a weighted QM energy as part of the geometry function

X-ray BUSTER ML for everything!

conventional TNT+ geometry function: Protein & water & ions



Weighted QM energy for ligand

Direct use of quantum chemical method as part of geometry function

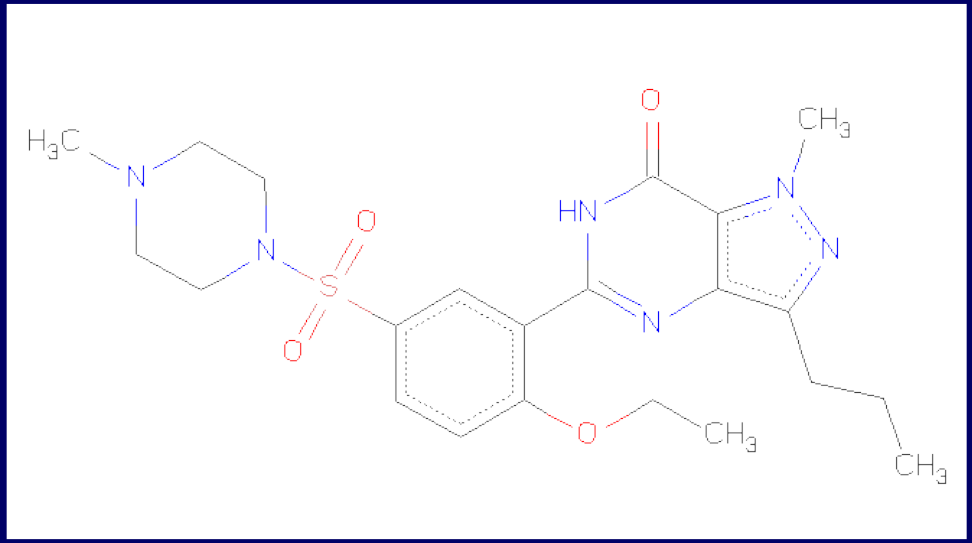
- Refine objective function:

$$V_{tot}(xyz,B,occ) = V_{xrayML}(xyz,B,occ) + V_{bond}(xyz) + V_{angle}(xyz) + V_{torsion}(xyz) + V_{plane}(xyz) + V_{contact}(xyz) \ldots + V_{bcorrel}(B) + w_{QM} E_{QM}(xyz_{ligand})$$

- QM provides a different sort of ligand restraint function

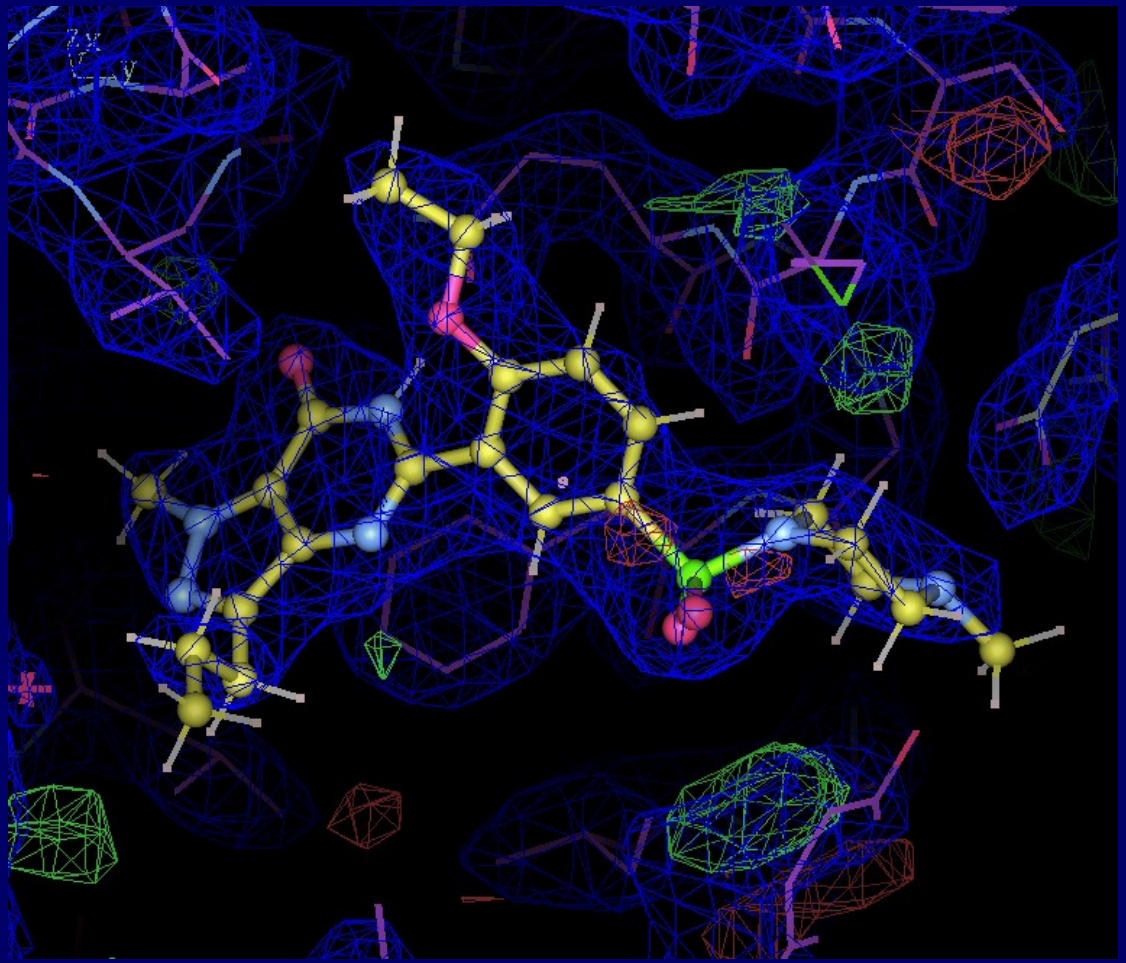
Application to PDE5 1udt

- 1udt “Crystal structure of Human Phosphodiesterase 5 complexed with Sildenafil(Viagra)”
  - 2.3Å resolution

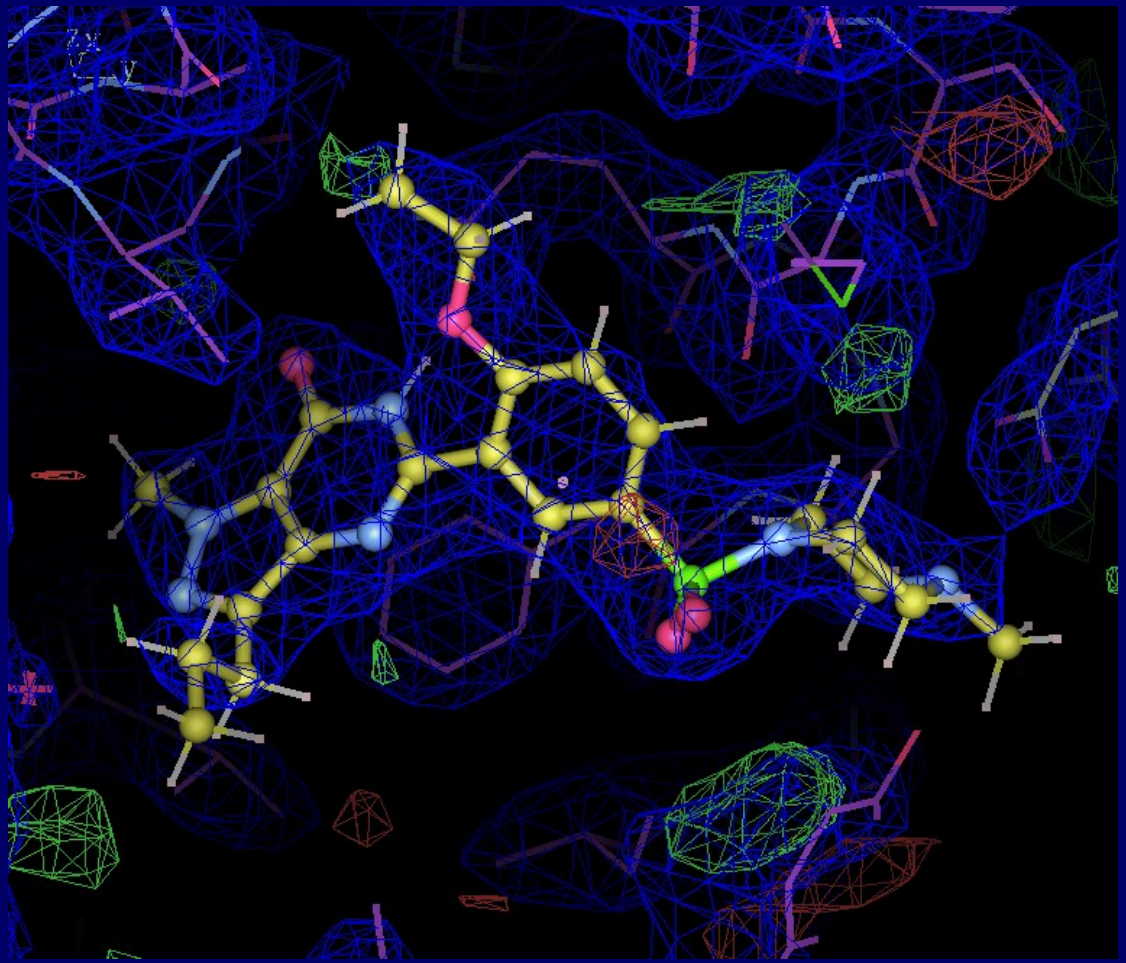


CCCc1nn(C)c2C(=O)NC(=Nc12)c3cc(ccc3OCC)[S](=O)(=O)N4CCN(C)CC4

1udt corrected VIA BUSTER refinement grade

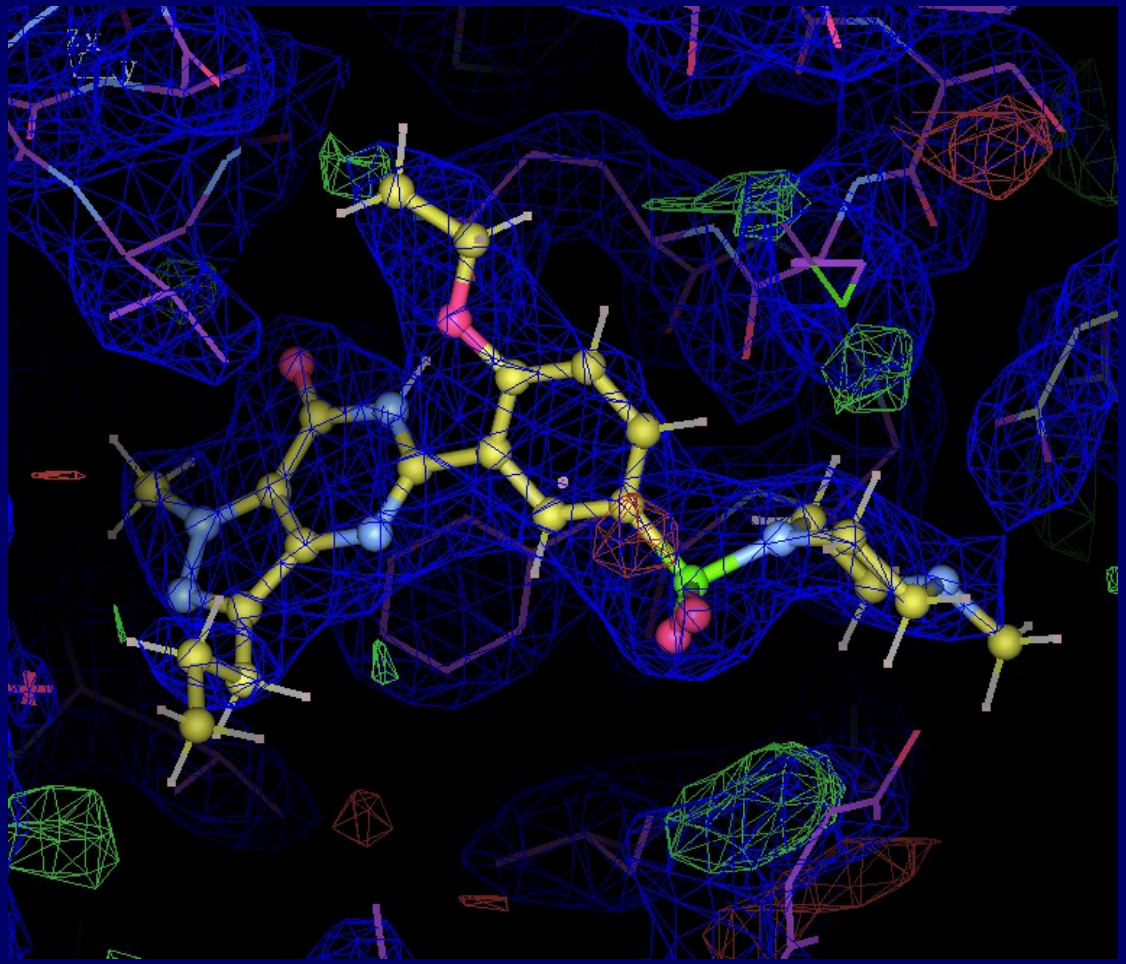


1udt corrected VIA BUSTER refinement RM1



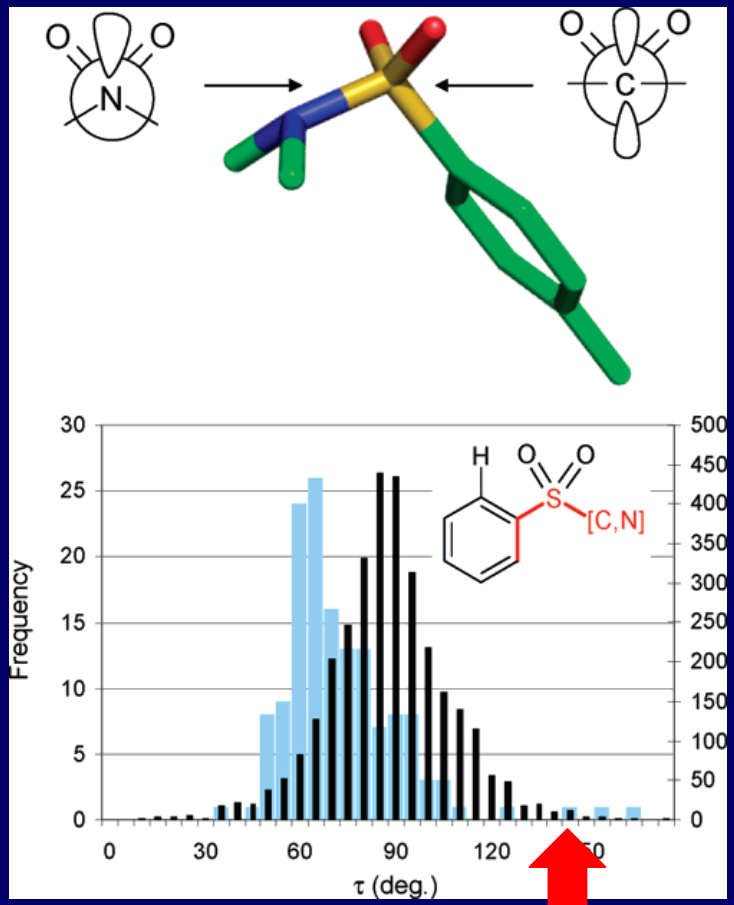
R_{work} Rf_{ree} ligand CC identical

1udt corrected VIA BUSTER refinement RM1



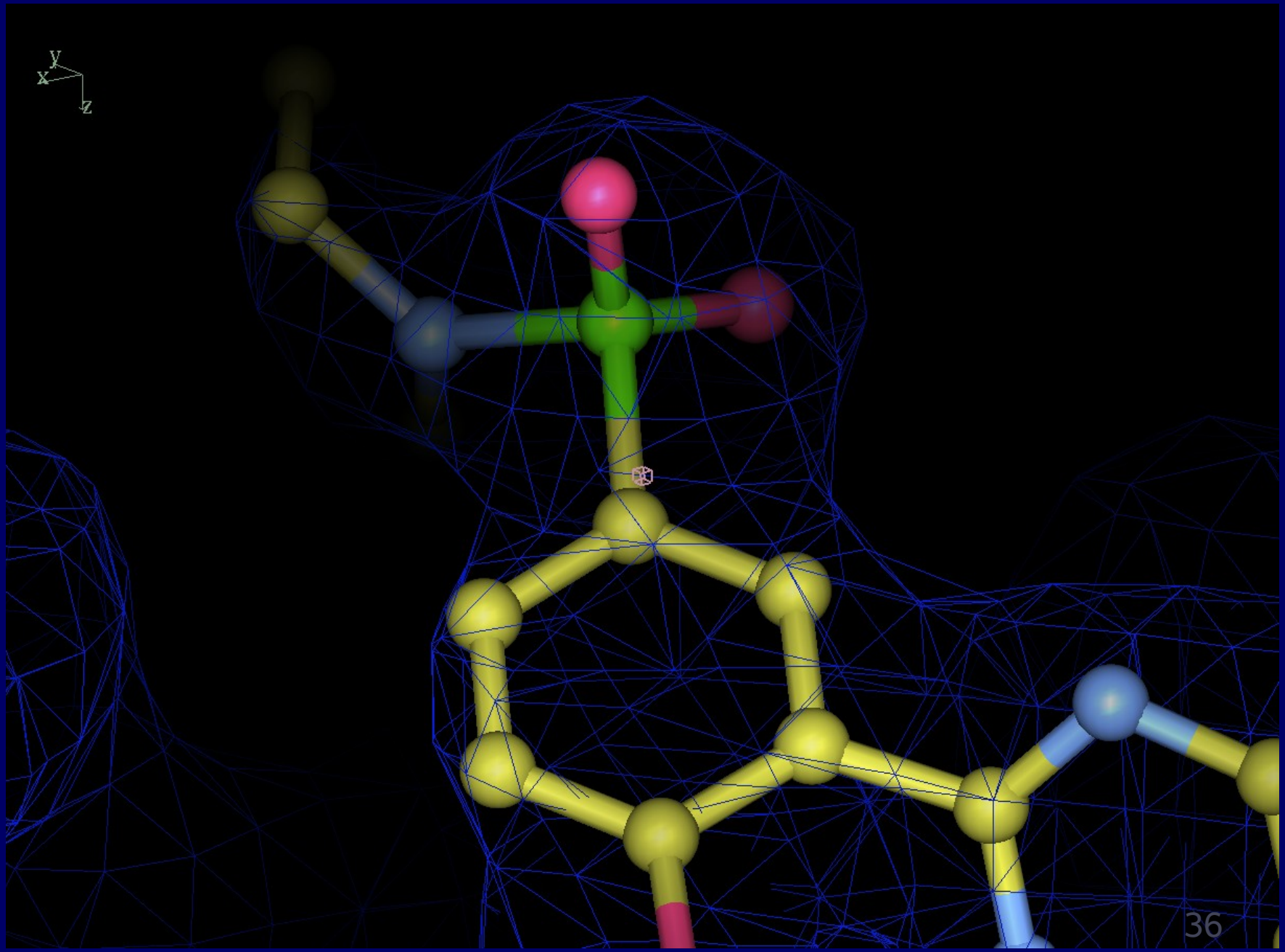
R_{work} Rf_{ree} ligand CC identical

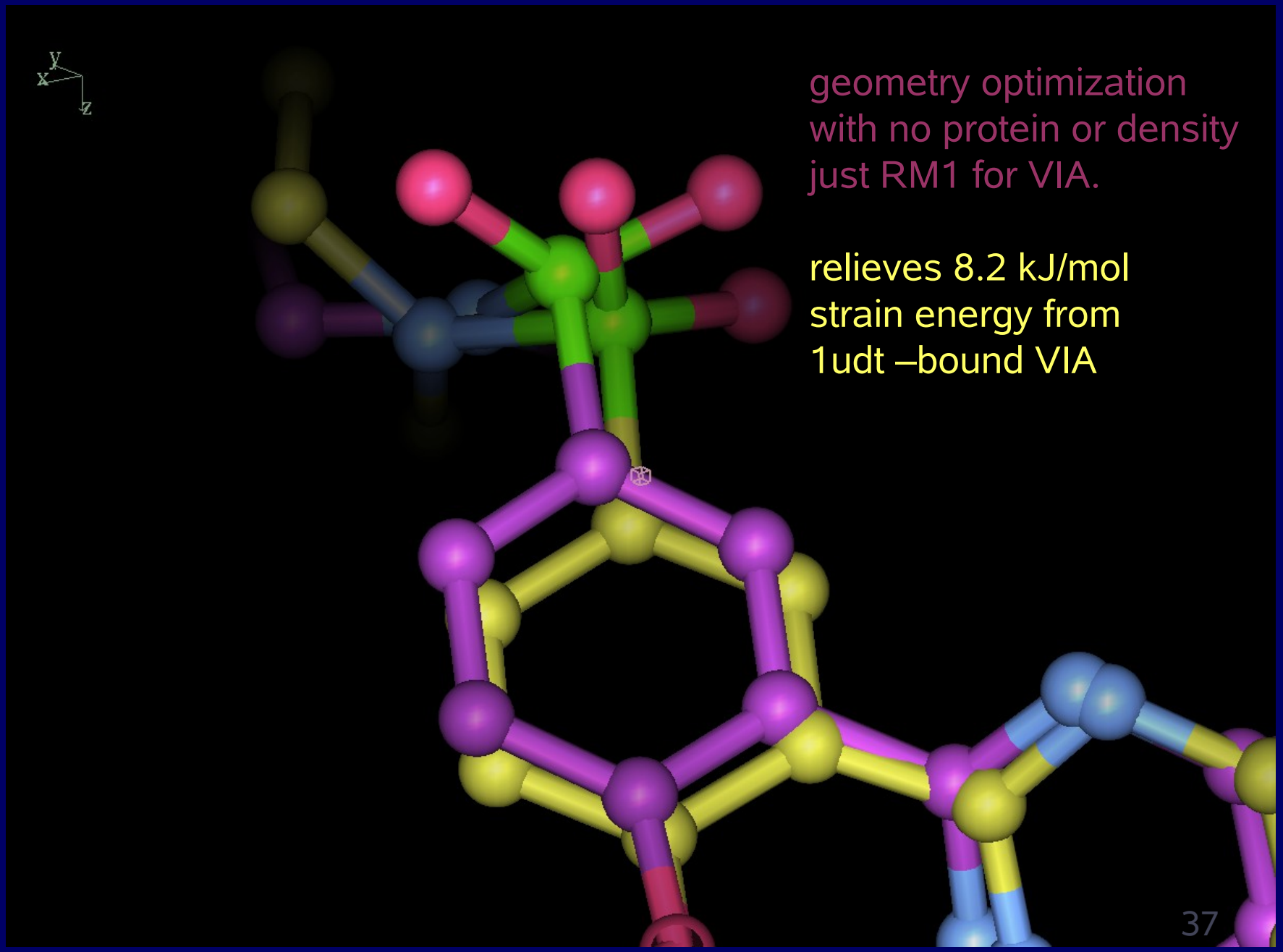
Is the 1udt viagra structure wrong or strained?



1udt VIA value

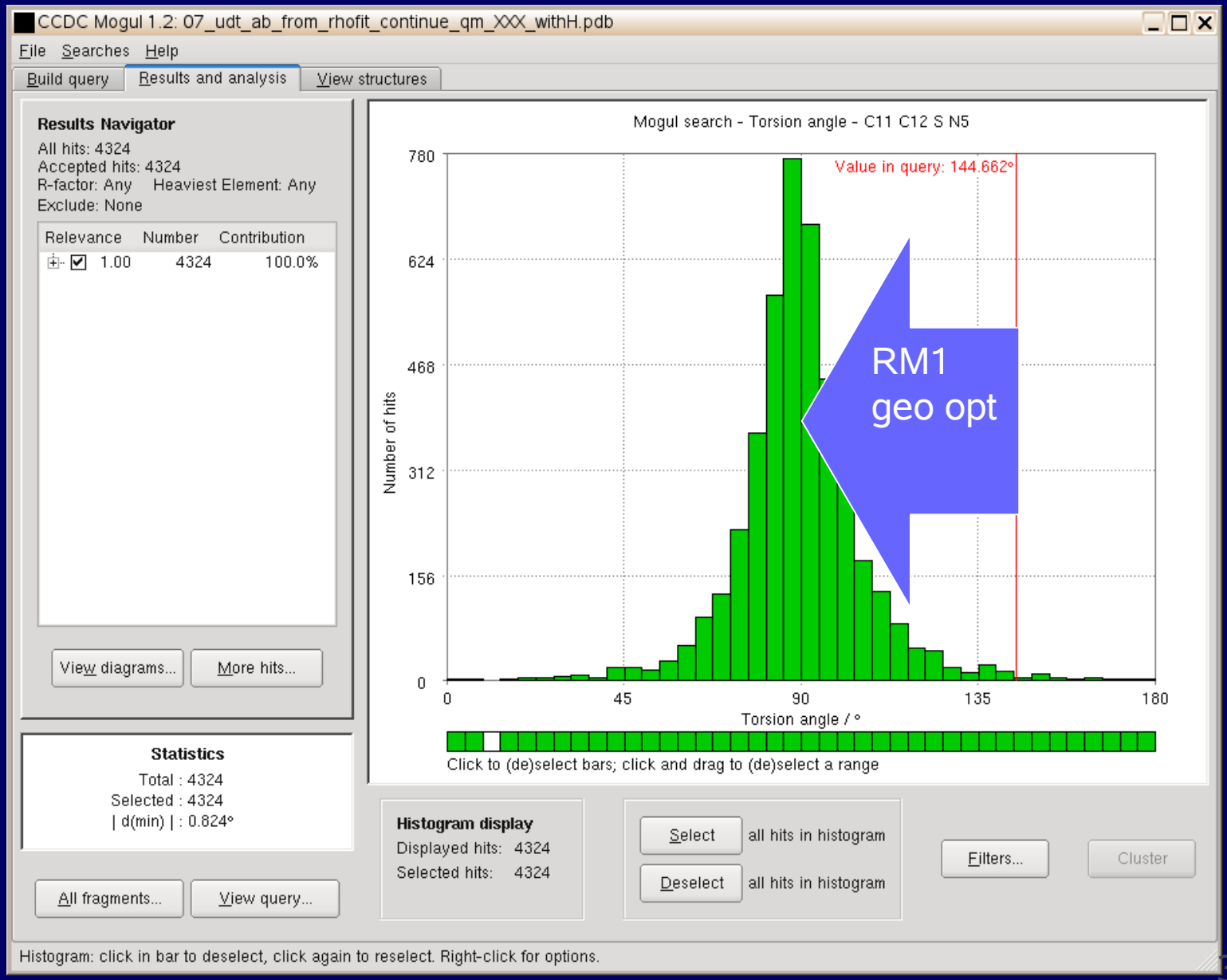
- Figure 9. Ken A. Brameld, Bernd Kuhn, Deborah C. Reuter, and Martin Stahl* J. Chem. Inf. Model. 2008, 48, 1-24
- CSD/mogul indicates that phenyl-SO2-torsion is ‘unlikely’


Y
X
Z

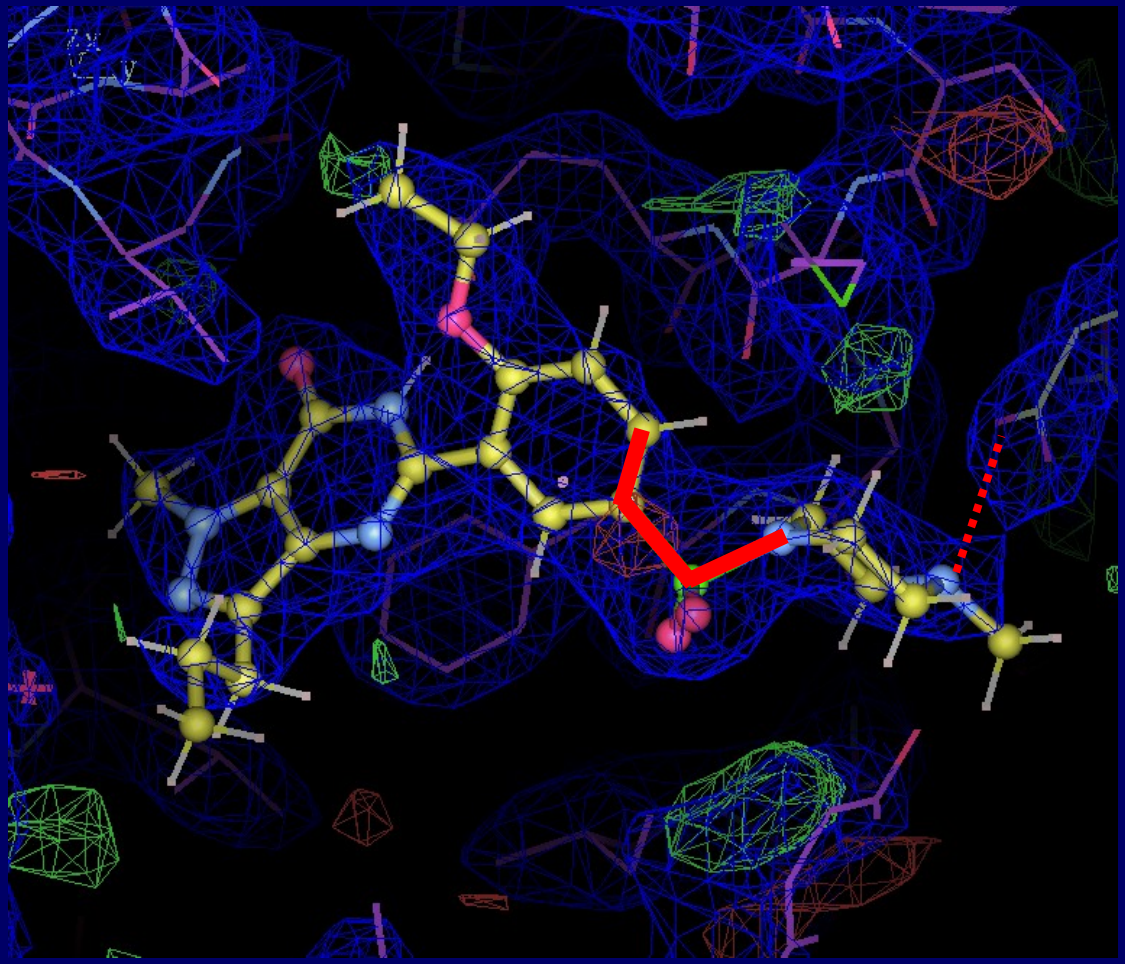


geometry optimization with no protein or density just RM1 for VIA.

relieves 8.2 kJ/mol strain energy from 1udt –bound VIA


CCDC Mogul 1.2: 07_udt_ab_from_rhofit_continue_qm_XXX_withH.pdb
File Searches Help
Build query
Results and analysis
View structures
Results Navigator
All hits: 4324
Accepted hits: 4324
R-factor: Any Heaviest Element: Any
Exclude: None
Relevance Number Contribution
1.00 4324 100.0%
View diagrams...
More hits...
Statistics
Total : 4324
Selected : 4324
| d(min) | : 0.824°
All fragments...
View query...
Mogul search - Torsion angle - C11 C12 S N5
Value in query: 144.662°
RM1
geo opt
Number of hits
780
624
468
312
156
0
0
45
90
135
180
Torsion angle / °
Click to (de)select bars; click and drag to (de)select a range
Histogram display
Displayed hits: 4324
Selected hits: 4324
Select all hits in histogram
Deselect all hits in histogram
Filters...
Cluster
Histogram: click in bar to deselect, click again to reselect. Right-click for options.

Why is this torsion strained?



- salt bridge piperazine to GLU 858*
- caused by Xtal contact

Direct use of weighted QM for ligands in refinement

- Provides alternative to using dictionary
- Comparable to dictionary in applying stereochemical information
- Useful for new chemistry
- Gives direct measure of ligand strain motion and energy at end of refinement
- Transferability and recording
  - RM1 or HF/6-31G* is a known entity
  - Useful for passing results to modellers!

Direct use of quantum chemical method for ligand

- BUSTER distribution provides dynamo helper for quick semi-empirical RM1
- GAMESS can be used for *ab initio*
- Easy to use `-qm VIA`
- To be included in Jan 2011 release of BUSTER

Acknowledgements

- CCDC
- Martin Field, Alexei Vagin, Garib Murshudov, openbabel developers
- BUSTER users
- Support:
  - Members of the Global Phasing Consortium,
  - Vizier Project FP6
- GΦL
  - Tom Womack
  - Maria Brandl
  - Clemens Vonrhein, Claus Flensburg, Wlodek Paciorek,, Peter Keller, Andrew Sharff, Gerard Bricogne