

ARP-Ligand – Locating, identifying and building ligands in electron density

Ciarán Carolan

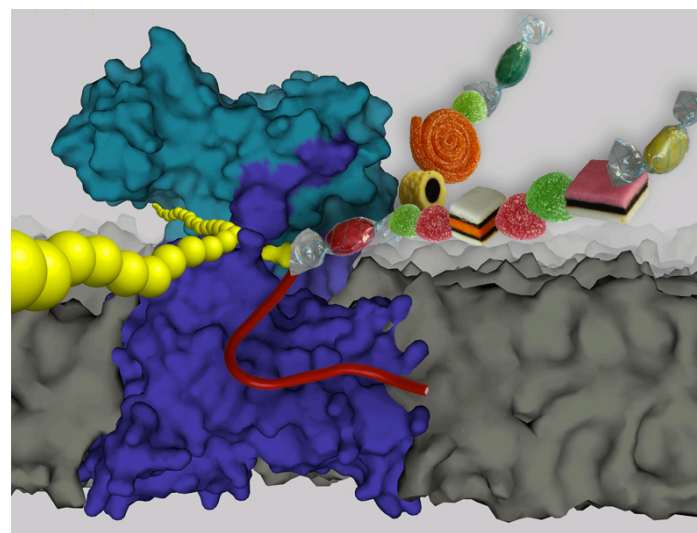
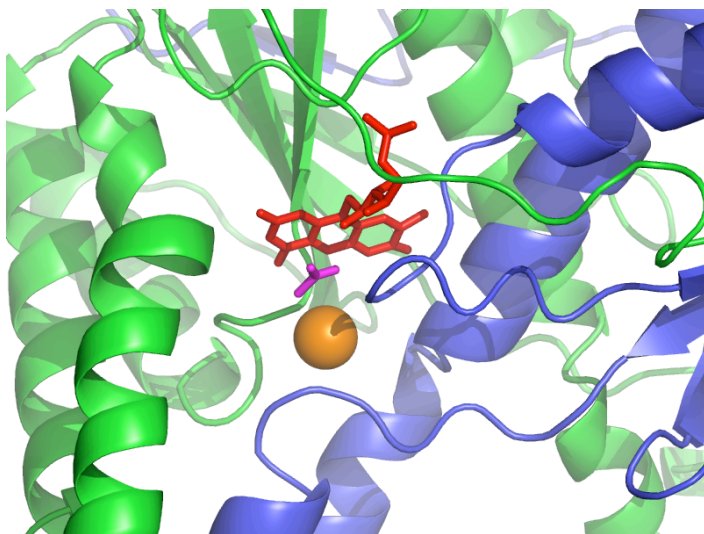
EMBL-Hamburg

ciaran.carolan@embl-hamburg.de

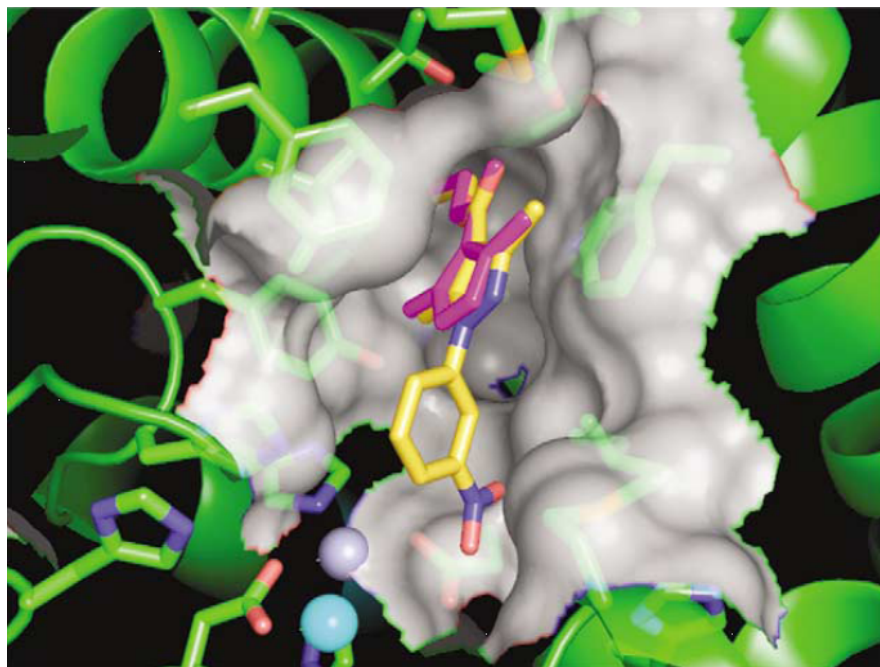


Why Ligands?

- Endogenous substrates and analogues thereof
- Known drugs and biological modulators
- Probe molecules
- Post-translational modifications
- Buffer constituents



Structure based drug design



Purple: A lead for the inhibition of PDE4D;

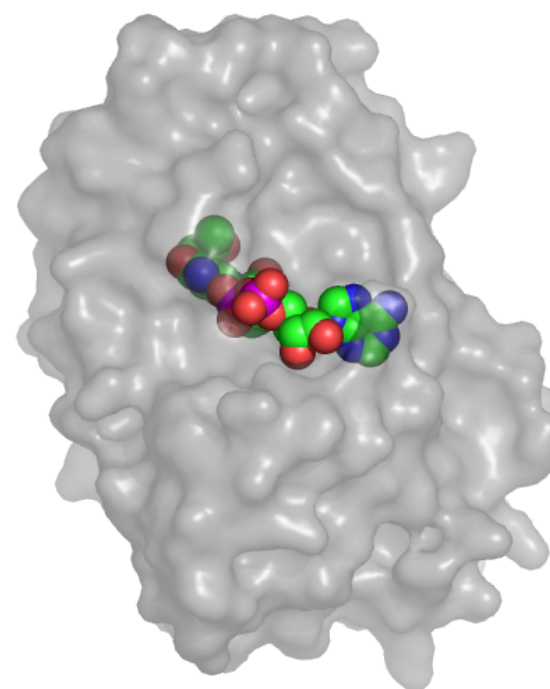
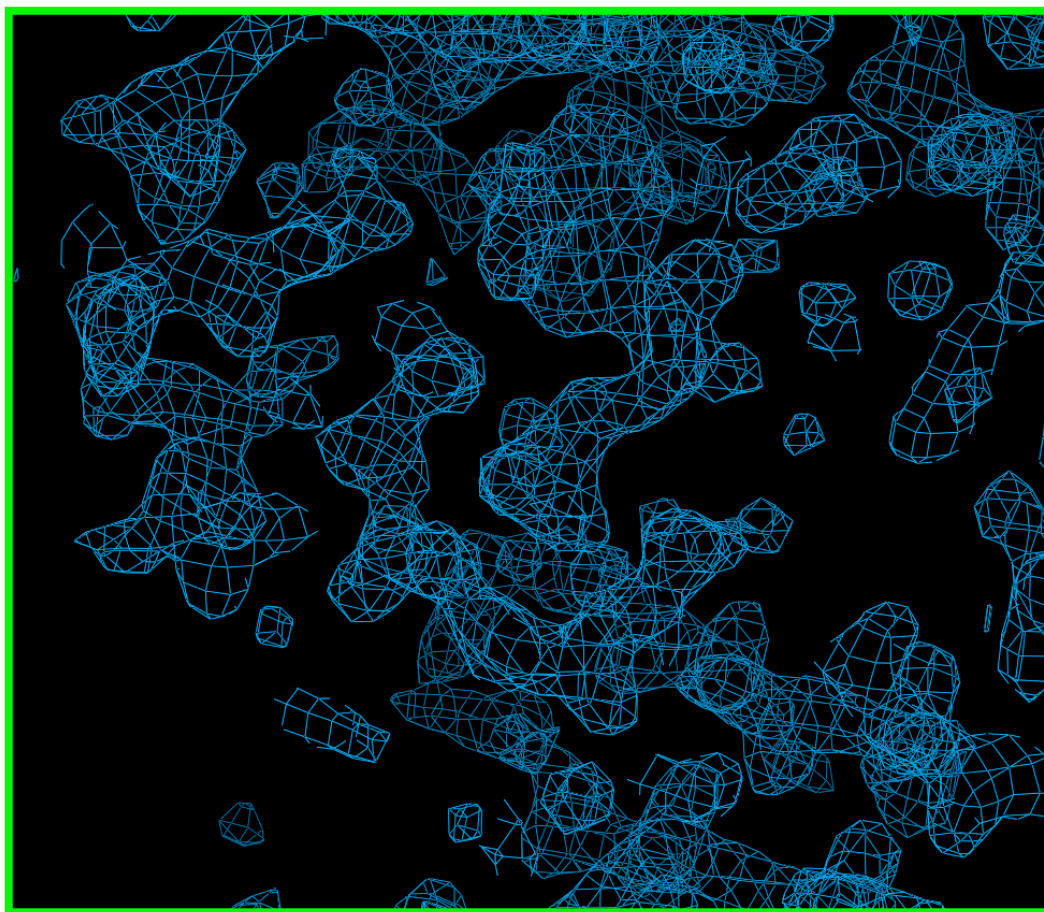
Yellow: the end result from SBDD.

Card et al. (2005) *Nat. Biotechnol.*
23: 201-207

- Structural work, principally involving MX and NMR, allows the elaboration of a lead to a candidate drug.
- High-throughput crystallography is an up-and-coming technique in lead generation (e.g. fragment based drug design). It requires rapid and automated ligand identification and construction, however.
- Elaboration of a lead to produce a drug depends on accurate knowledge of ligand binding modes, and can also be aided by automated ligand building procedures.

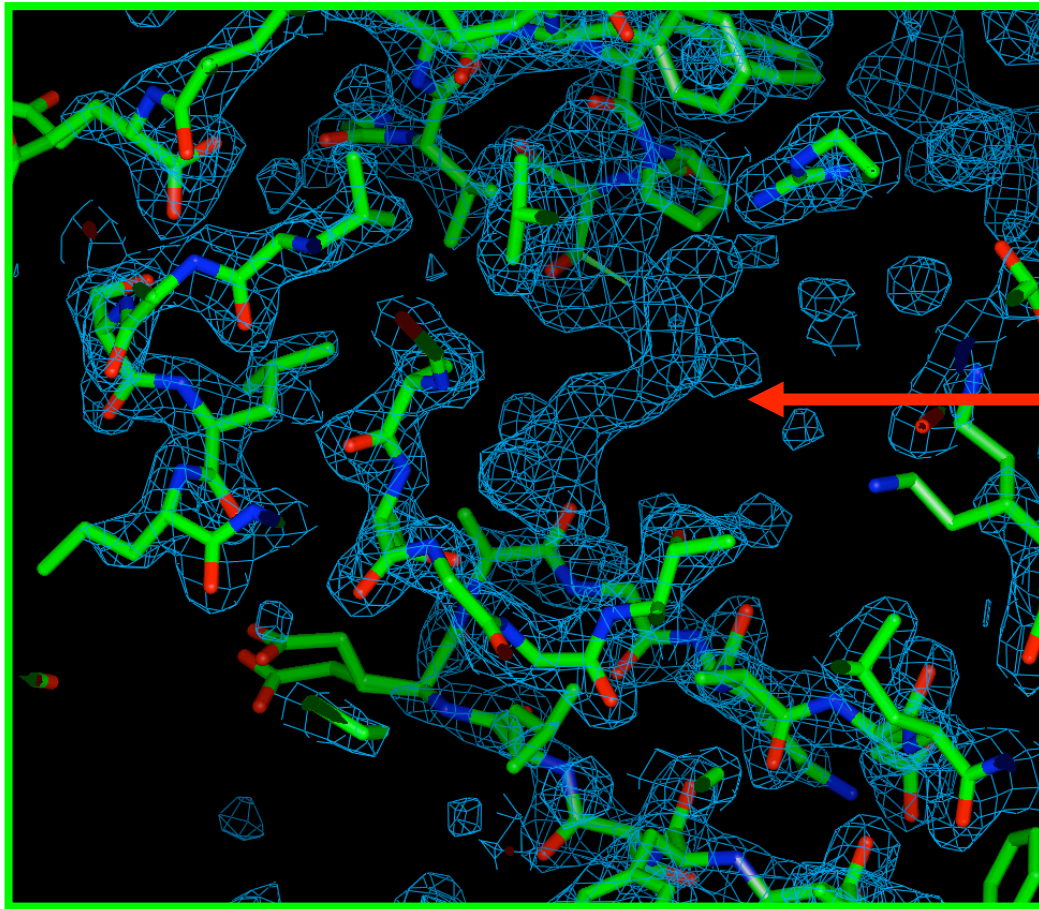
ARP/wARP and Ligand Building

electron density map (good resolution)



ARP/wARP and Ligand Building

electron density map (good resolution) + protein
model

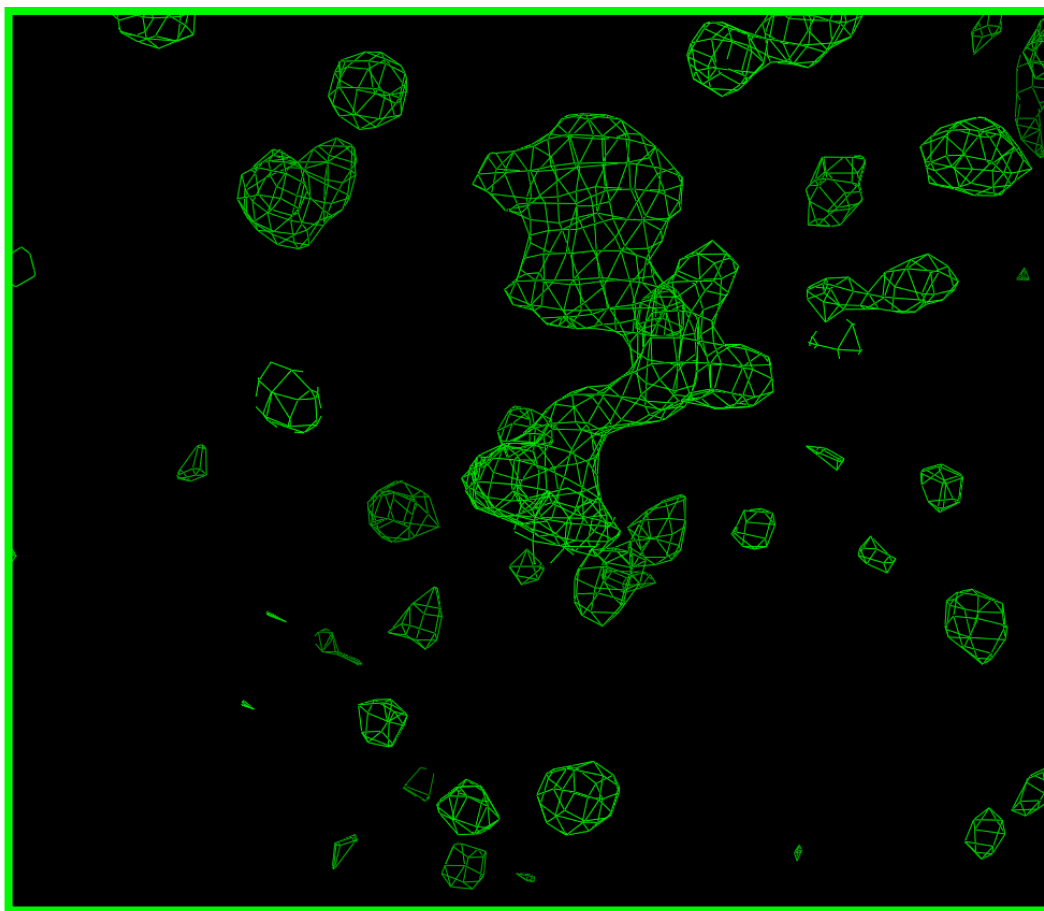


excess density

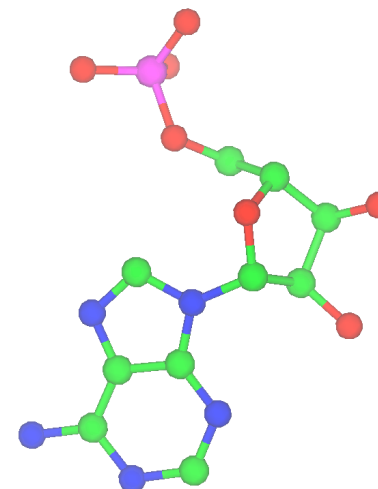
not
accounted for
by the protein

ARP/wARP and Ligand Building

difference density map...



... into which we have
to
FIT
the
the
APPROPRIATE
ligand



The protocol of a modeling task

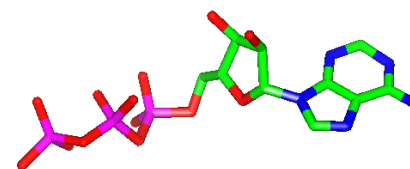
Fundamental initial knowledge determines the protocol.

		Binding site	
		known	unknown
Ligand identity	known	X	
	unknown		

1 density cluster

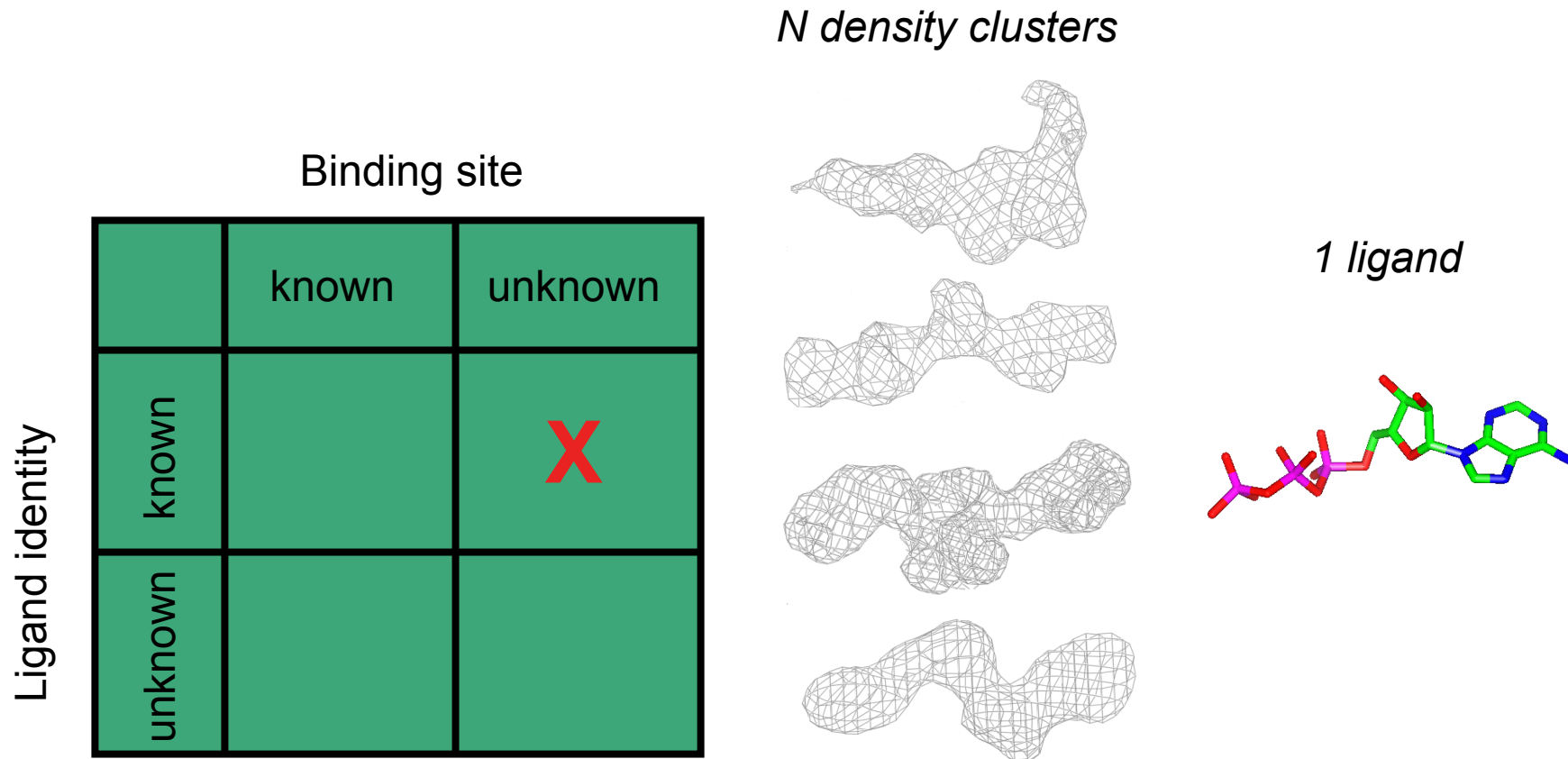


1 ligand



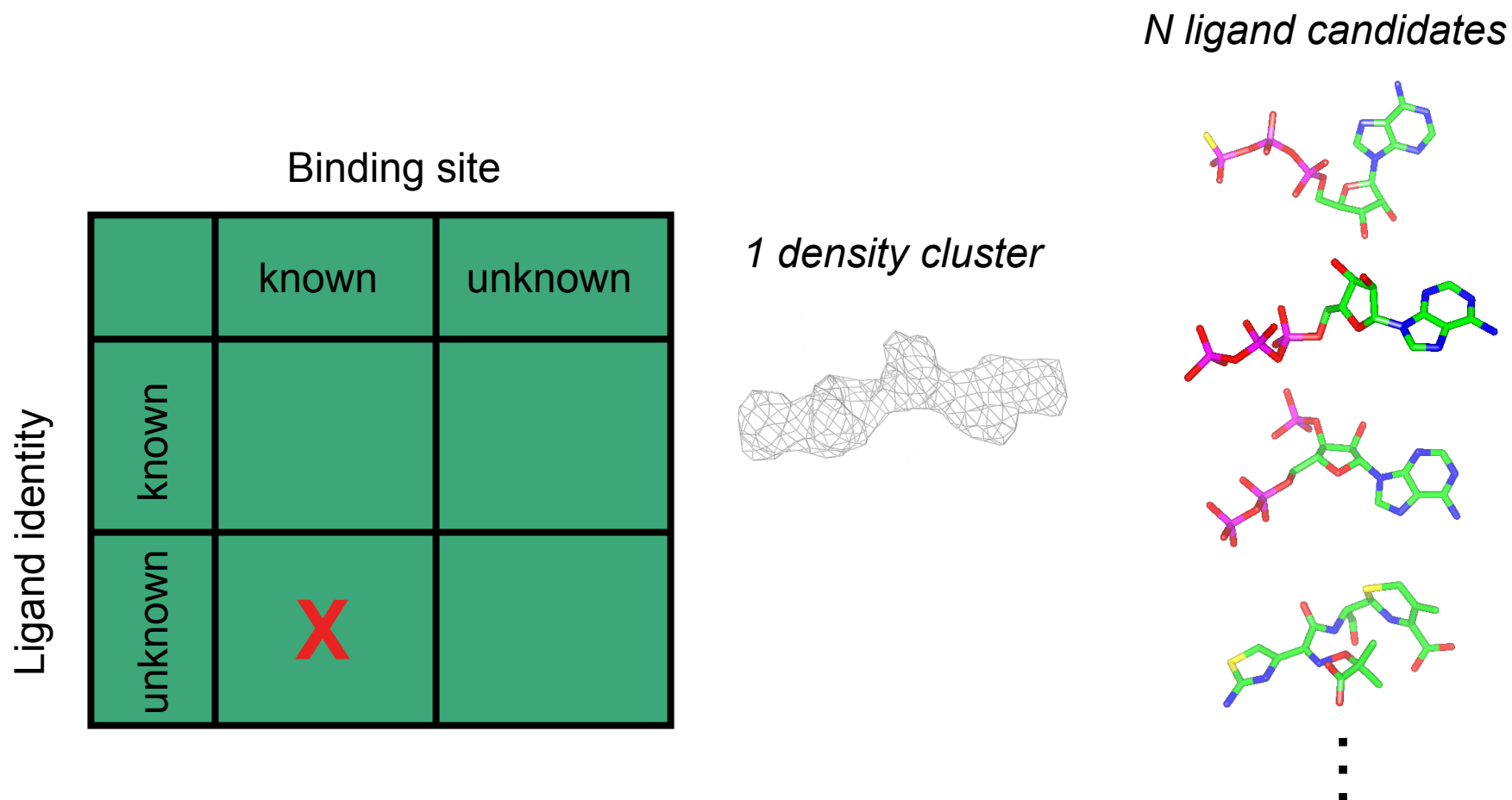
The protocol of a modeling task

Fundamental initial knowledge determines the protocol.



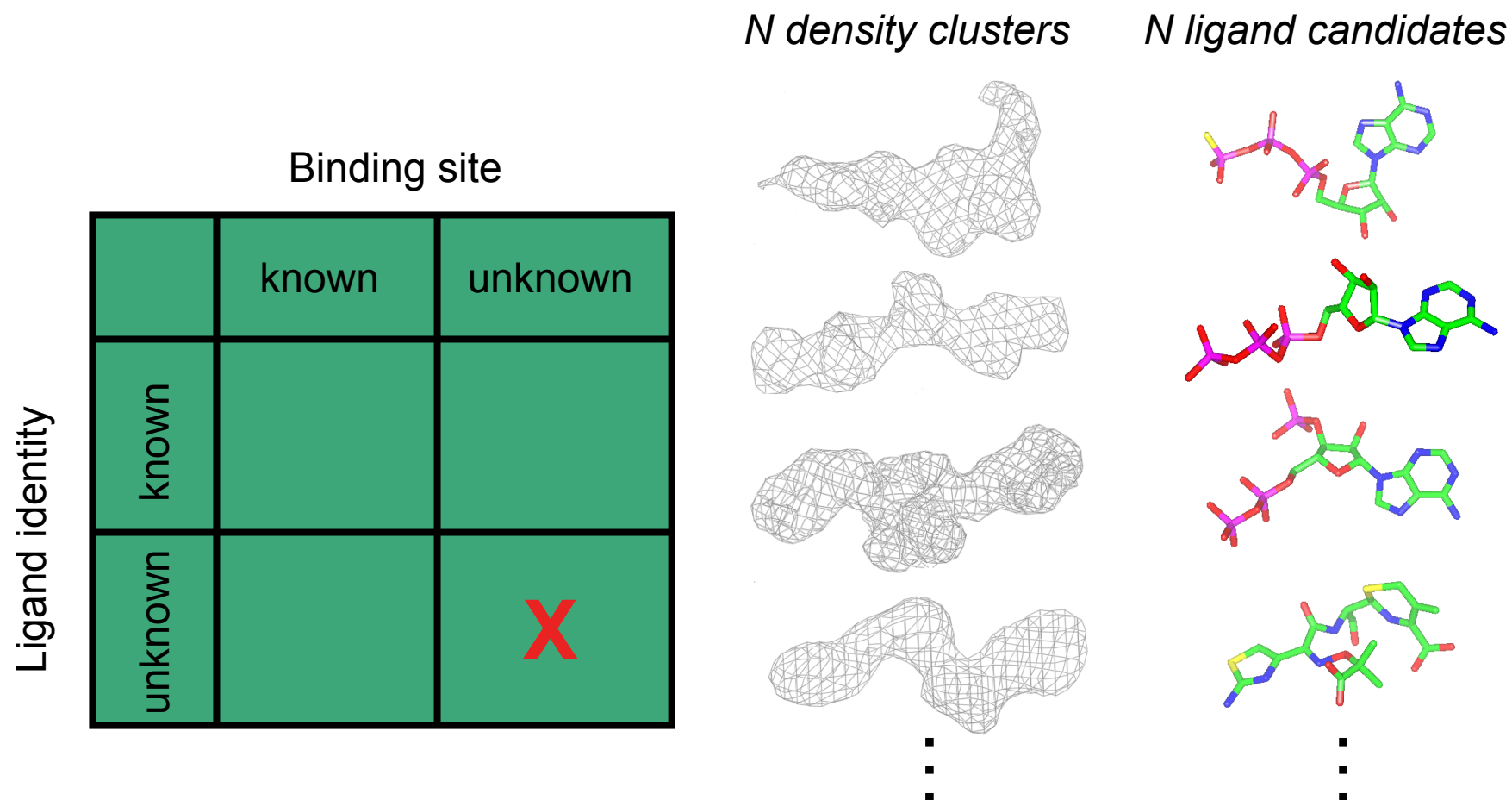
The protocol of a modeling task

Fundamental initial knowledge determines the protocol.



The protocol of a modeling task

Fundamental initial knowledge determines the protocol.



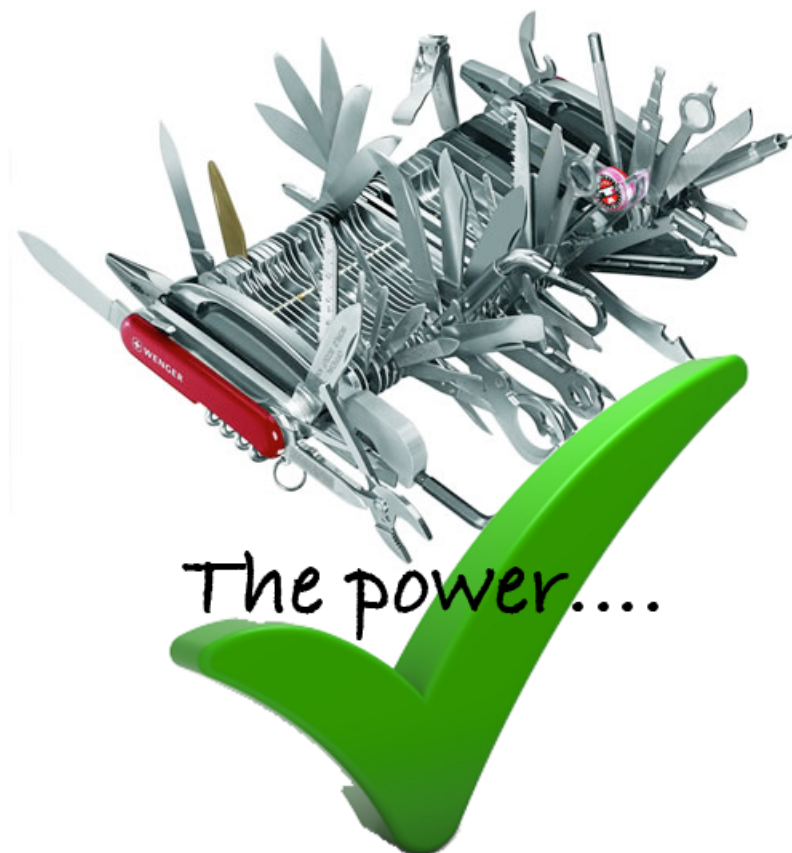
The task of ligand fitting

- Locate

ARP-Ligand

- Identify

- Fit



Researcher's Needs #1 - Identification

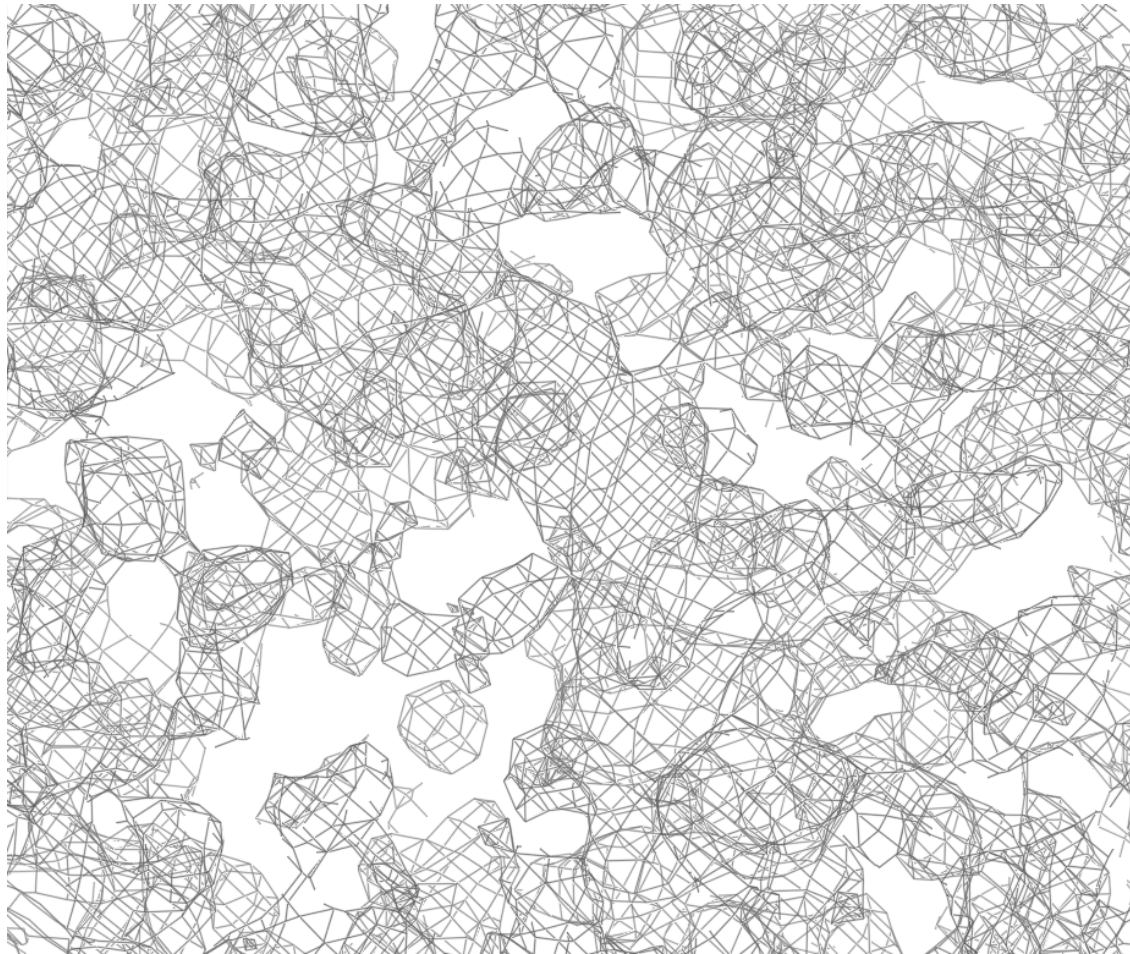
CCP4BB:

Is there a simple way to determine whether ligand is bound or not by comparing the diffraction patterns between ligand-free (structure known) and ligand-soaked protein crystals? I would like to solve the ligand bound protein structure, but before I do so, I have to find out if the ligand is actually bound and if so, where. Thank you very much!

*Best,
Joe*

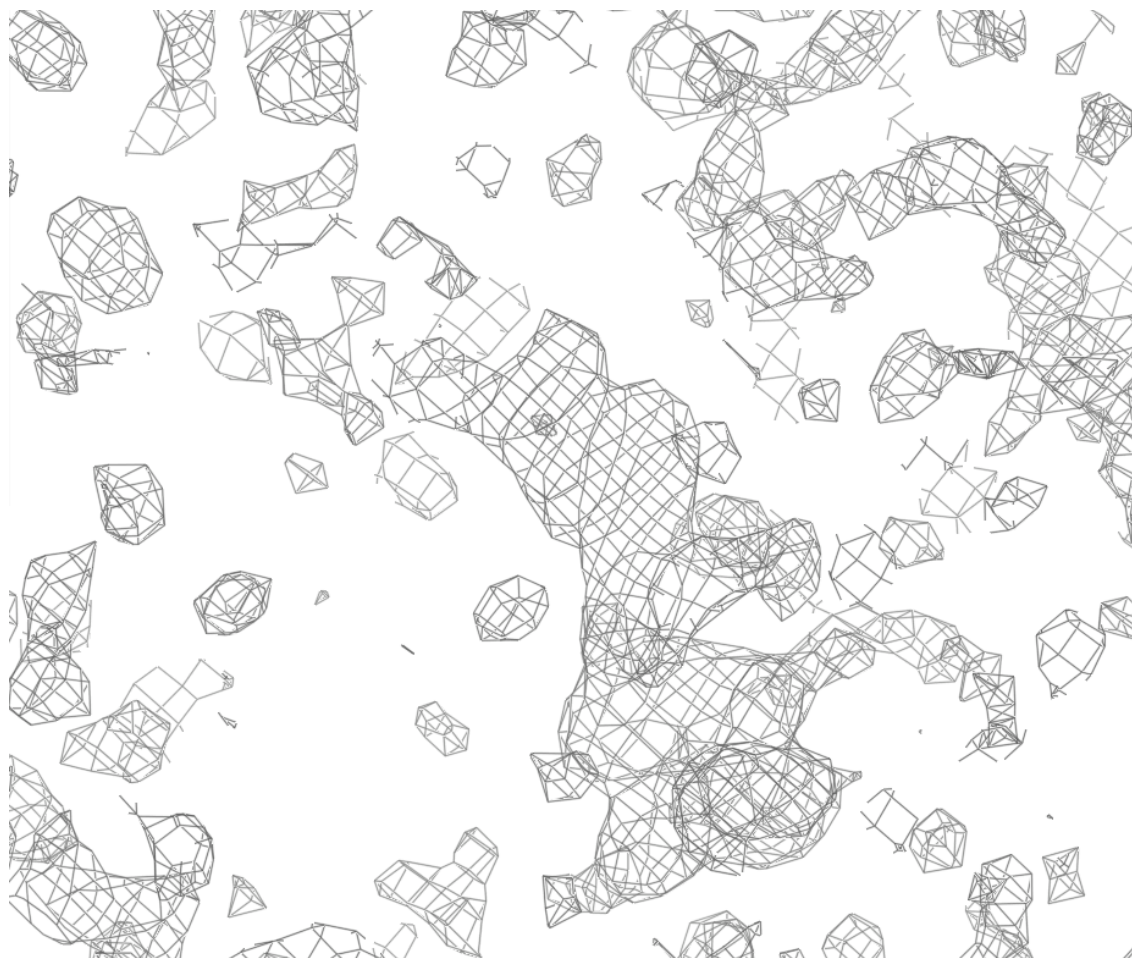
Automatic binding site identification

The difference density map at a low contour threshold.



Automatic binding site identification

The difference density map at a medium contour threshold.



Automatic binding site identification

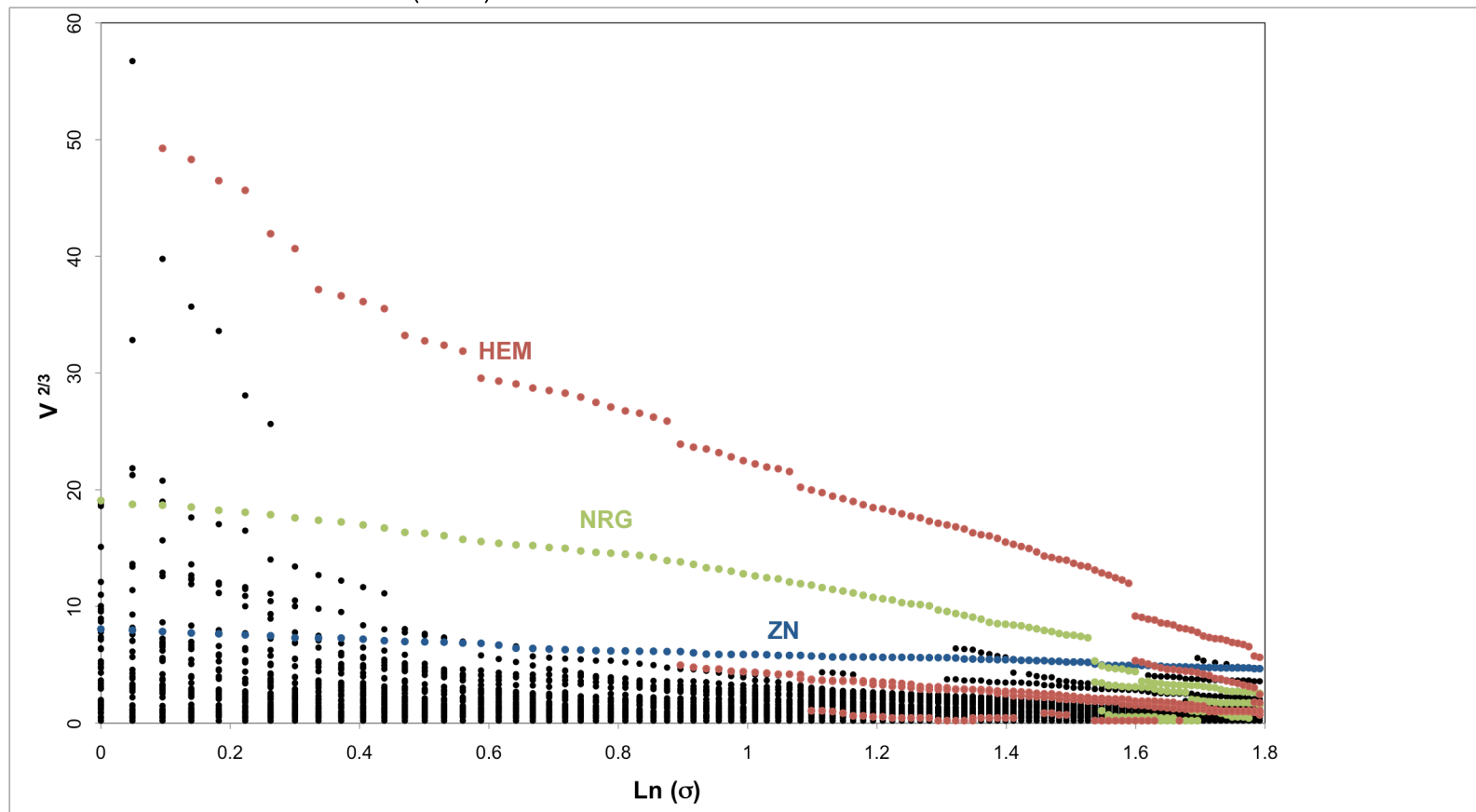
The difference density map at a high contour threshold.



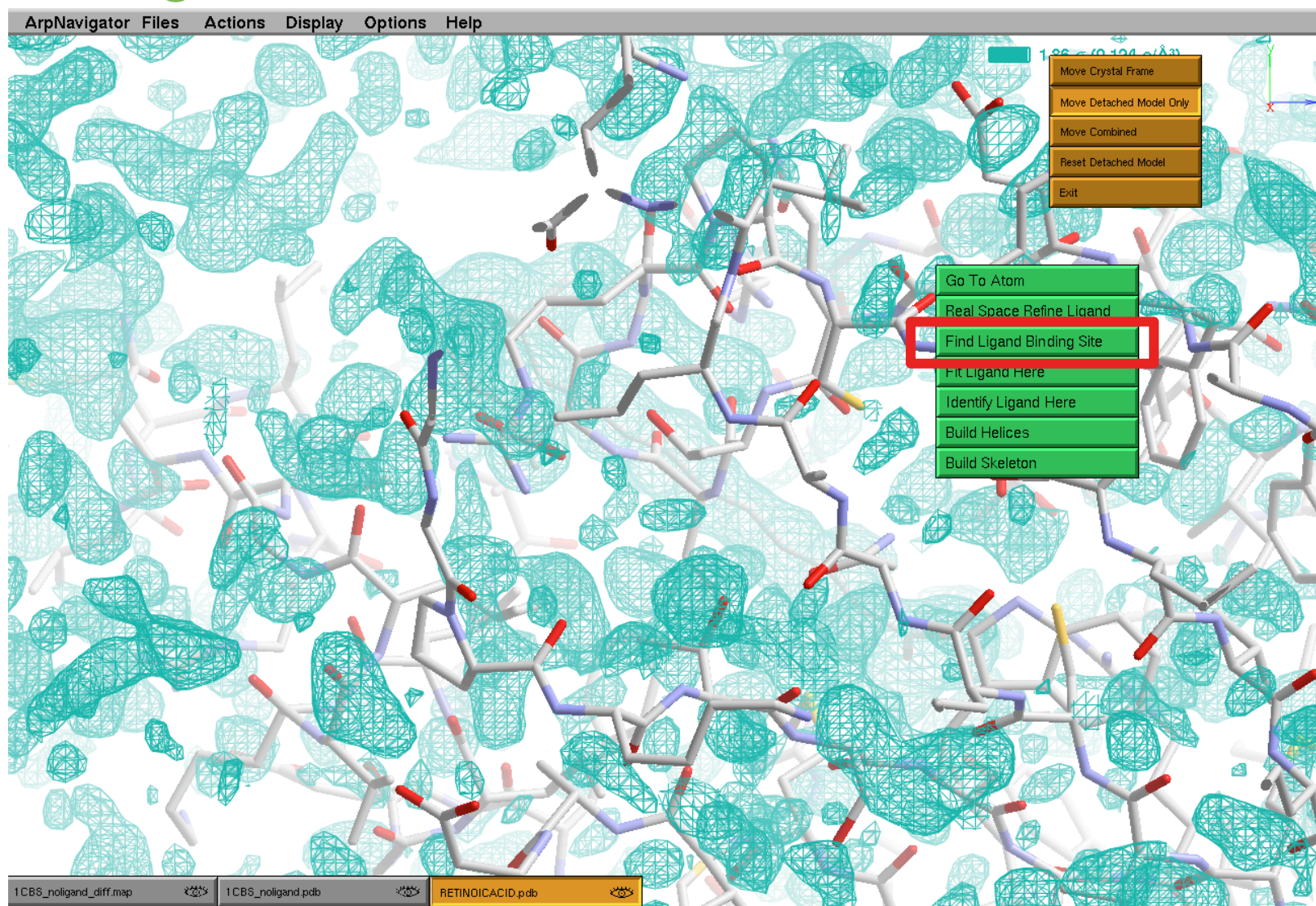
Automatic binding site identification

Capturing the dependence of the difference density map on changing the contour thresholds: *Fragmentation tree*.

$$V^{2/3} = \left(\frac{4\pi}{3}\right)^{2/3} \sigma^2 \left[\ln\left((2\pi)^{-3} \sigma^{-6} Z^2\right) - 2\ln(t) \right]$$



Binding Site Identification in ARP/wARP



Researchers needs #2 - Identification

CCP4BB:

We are working on protein/inhibitor complex structure. However, we did find a strange density at the active site, it looks really like GSH, the natural co-enzyme of this protein. We tried to use very simple solution to get crystal then exclude the possibility of buffer molecules, but that density is always there.

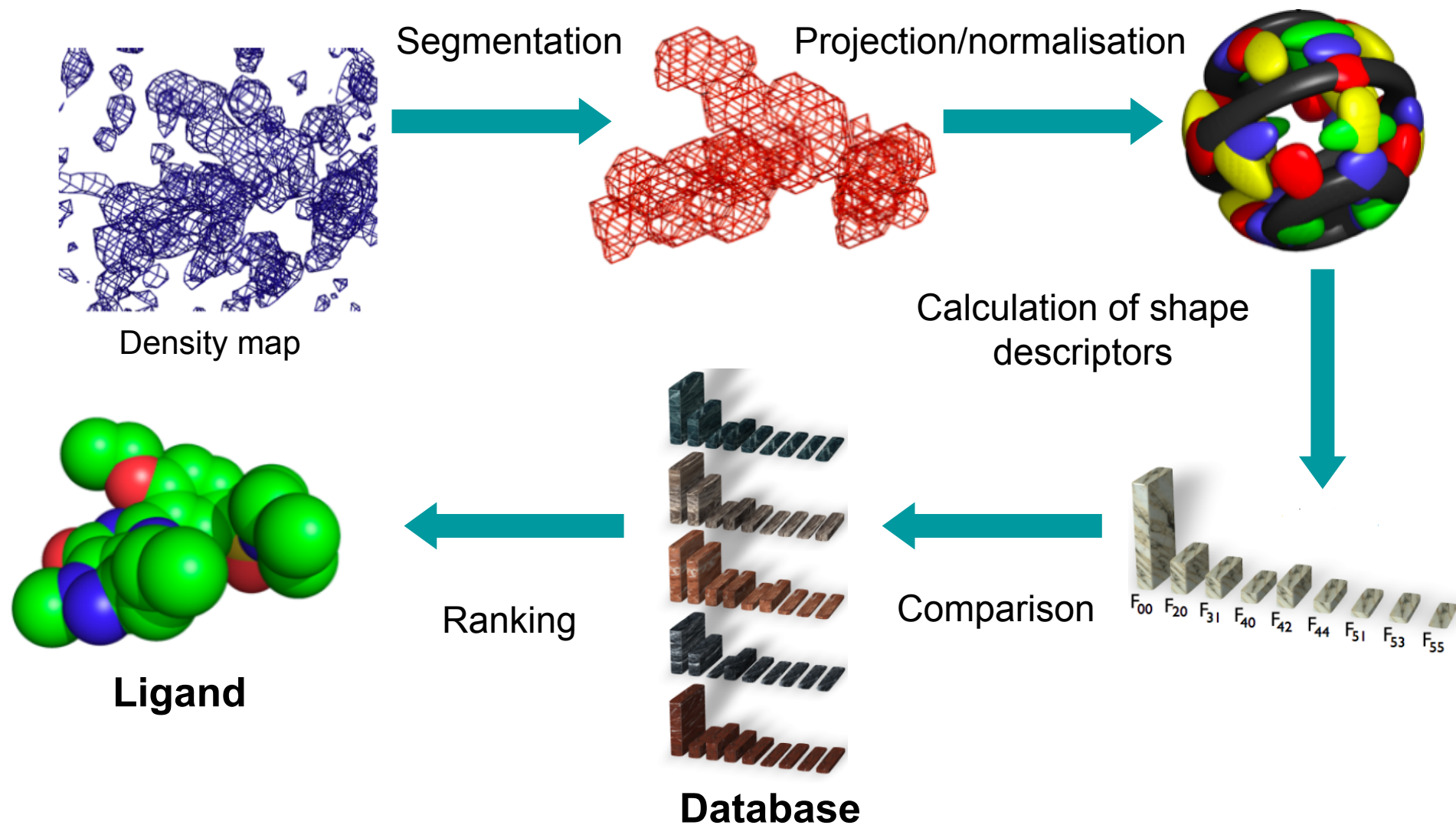
I want to identify this ligand. Are there any methods to do this work?

*Many many thanks!!!
Xiaopeng*

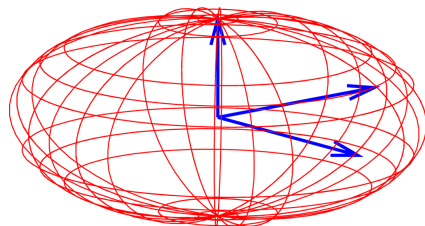
It's all in the shape baby!!!



High throughput ligand identification

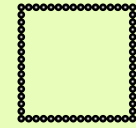
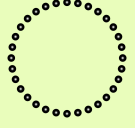
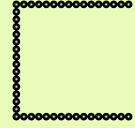
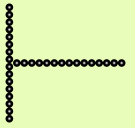
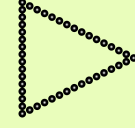


Third Order Moment Invariants



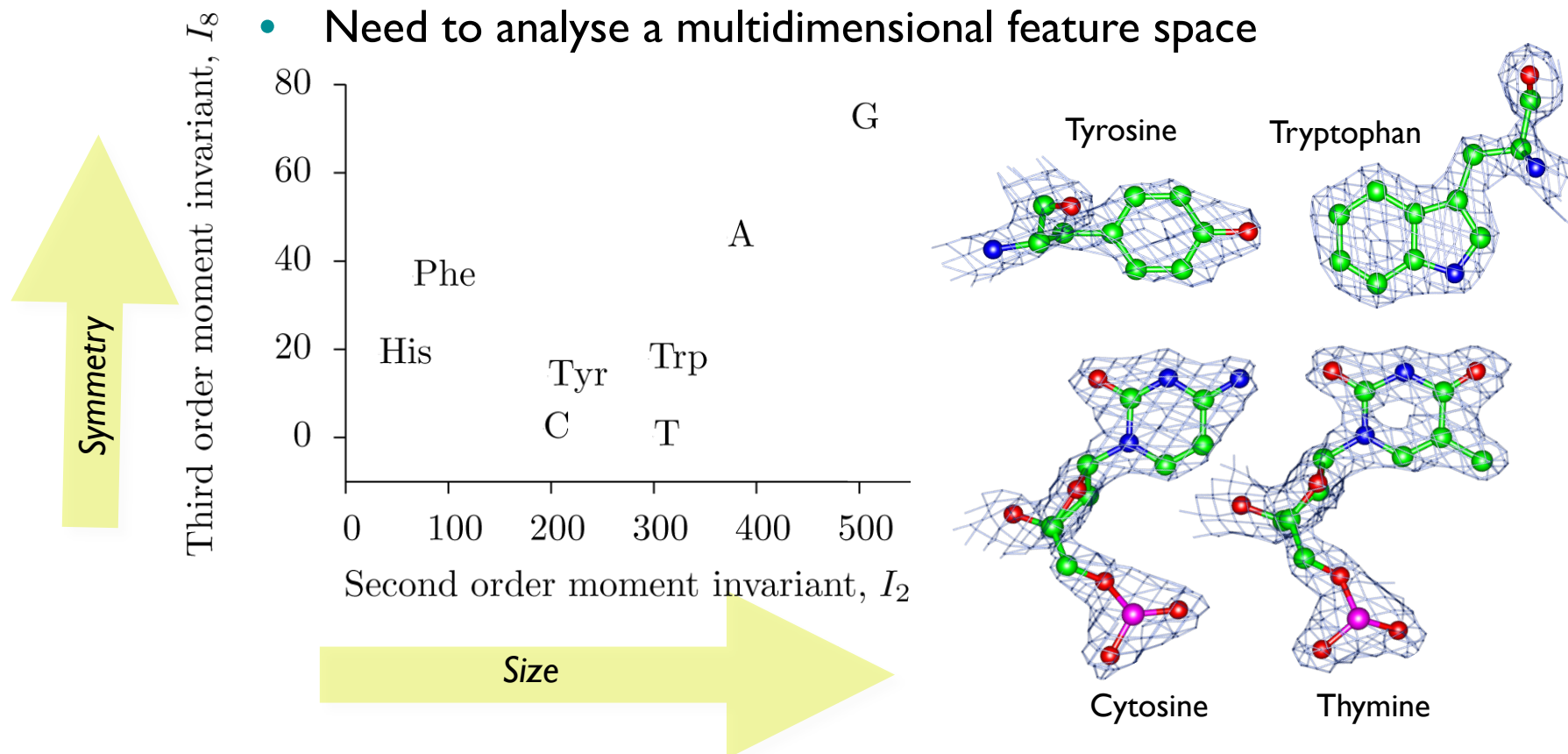
PCA – 2nd order, symmetric shapes

3rd order: Lo & Don (1989) *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 1053–1064

								
μ_{00}	16	32	32	60	32	46	31	45
μ_{20}	170	200	340	2255	900	1228	775	1130
μ_{30}	0	0	0	0	0	3049	3596	2850

Use a Combination of Various Features

- Any one feature is not discriminative enough
- Need to analyse a multidimensional feature space



Important things to consider!

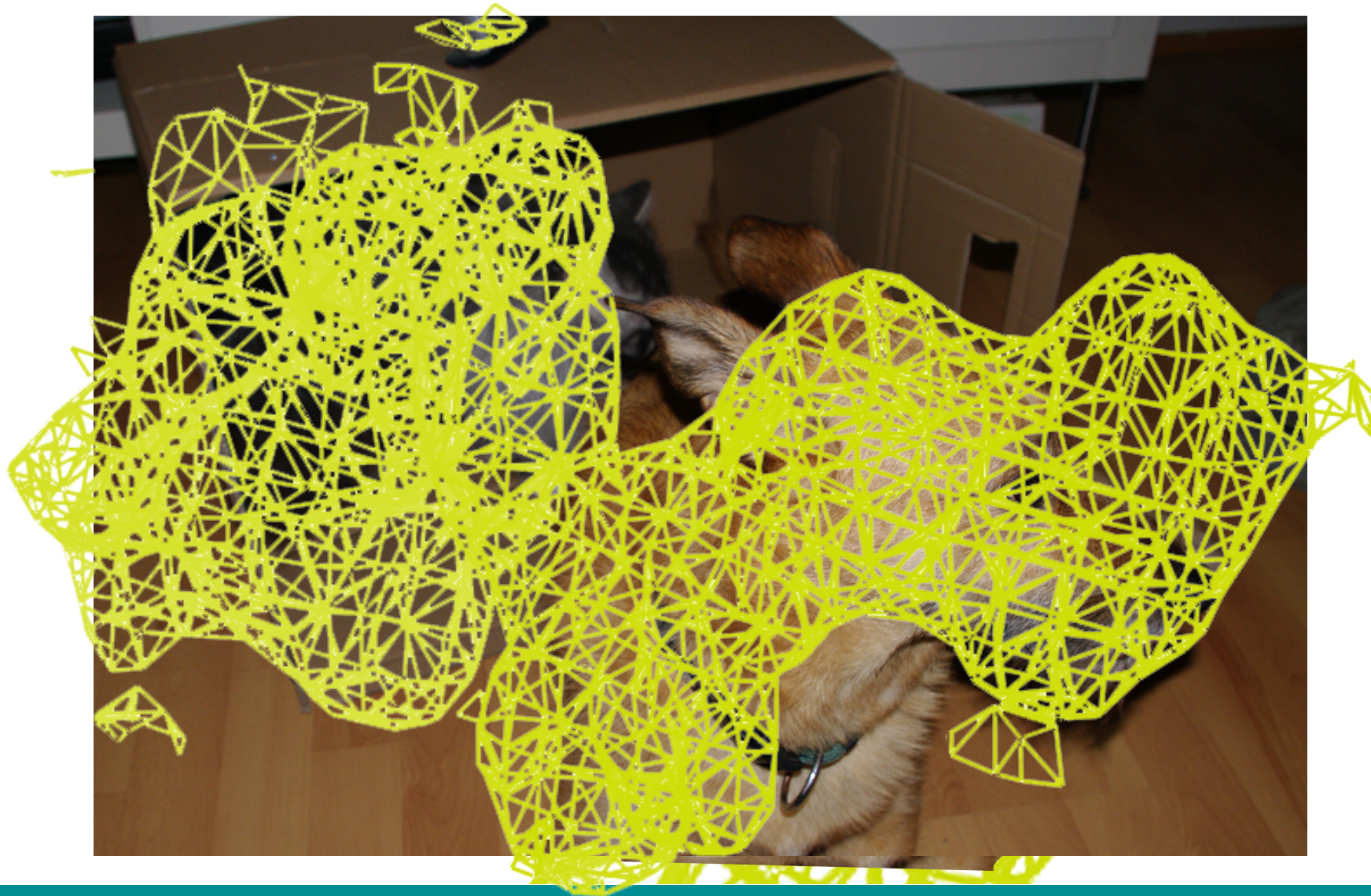
- Flexibility – Ligand Conformations!
 - Full ligand flexibility is considered



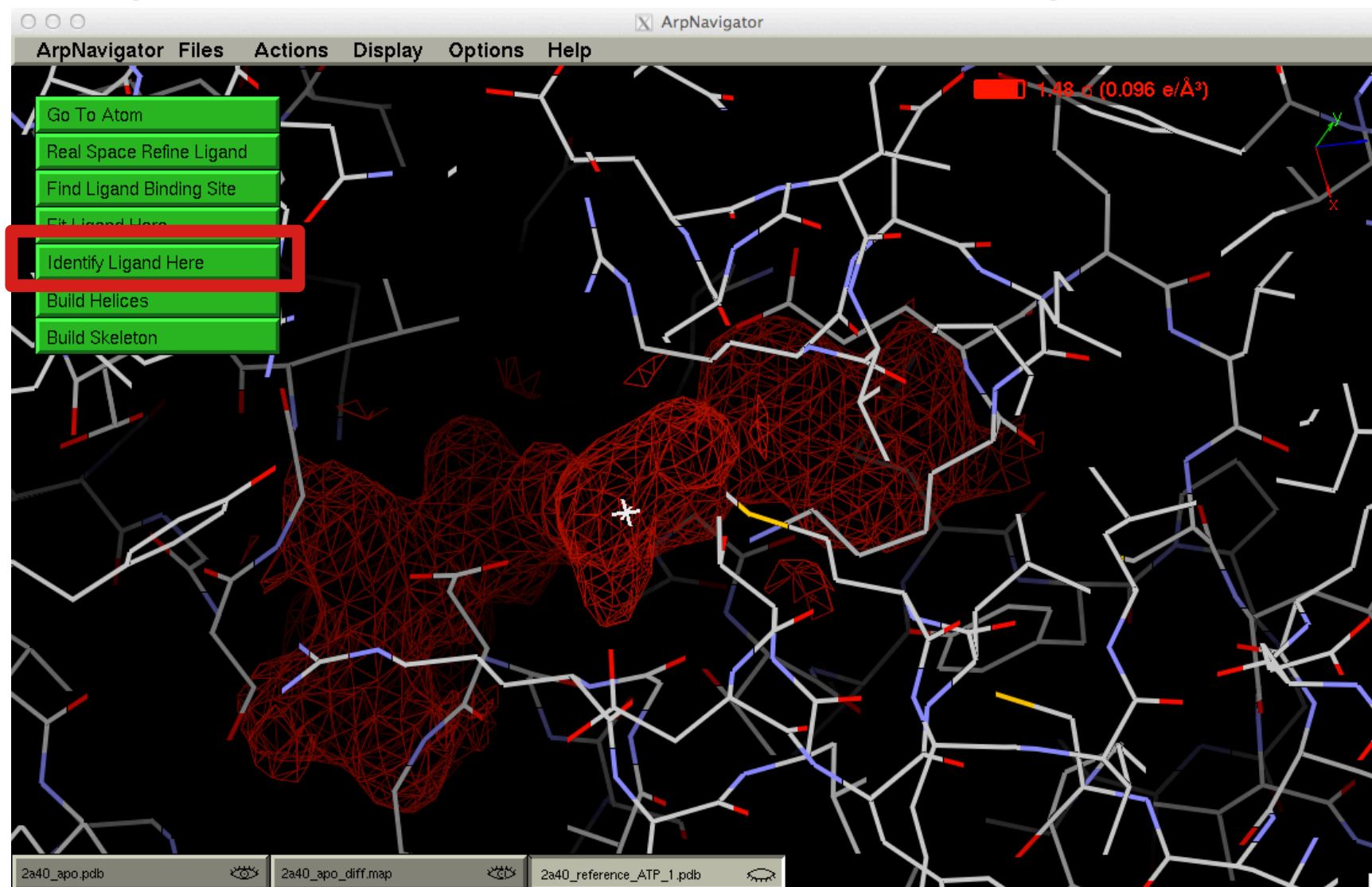
Important things to consider!

- Flexibility – Ligand Conformations!
 - Full ligand flexibility is considered
- Currently, a database of 50 of the most common ligands in crystallography is screened.
- Beware of multiple ligands, possibly covalently bound, and coordinated solvent among other things!

Multiple Ligands



Ligand Identification in ARPNavigator



Researcher's needs - #3 – Model Building

CCP4BB:

I am trying to solve a structure of a protein with some inhibitor. I want to know how I can put in my inhibitor in the density map of the data I got. I can see some density in the active site where the inhibitor should be.....I am not sure of how to do it.

Thanks,

Anshul Awasthi

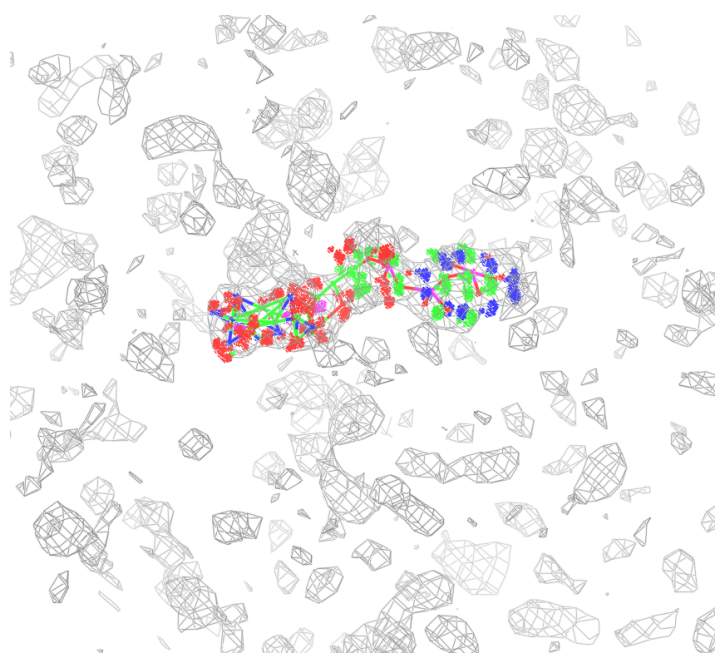
Constructing the ligand

- To address more broadly the different ligand sizes and complexities encountered, to increase the robustness of the software...

2 separate construction methods get used:

- ❑ *Label swapping on a sparsed grid*
- ❑ *Metropolis search in conformation space*

The workflow of ligand fitting in ARP/wARP –



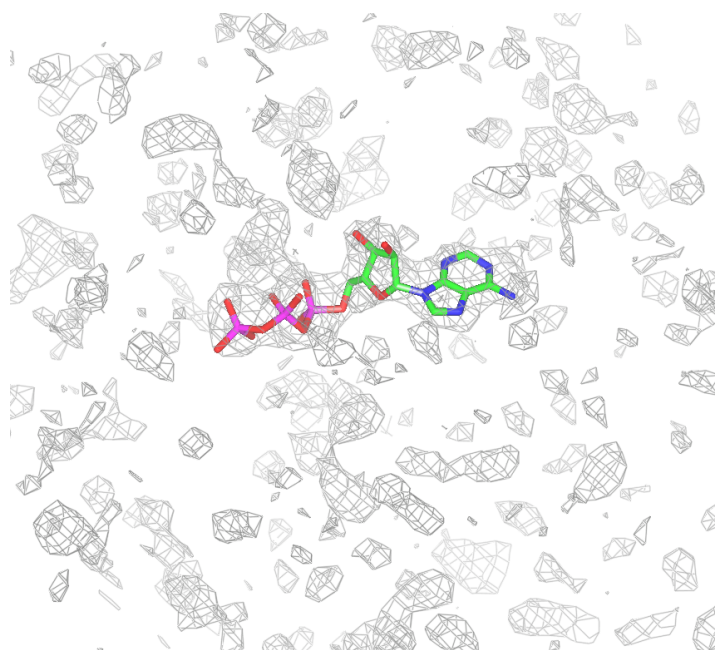
Prepare

- Binding site identified by user.
- Sparse density map and generate ligand topology.

Construct

- Construct ensemble of ligand models in plausible conformations to fit sparsed density.

The workflow of ligand fitting in ARP/wARP – The simplest case



Prepare

- Binding site identified by user.
- Sparse density map and generate ligand topology.

Construct

- Construct ensemble of ligand models in plausible conformations to fit sparsed density.

Refine

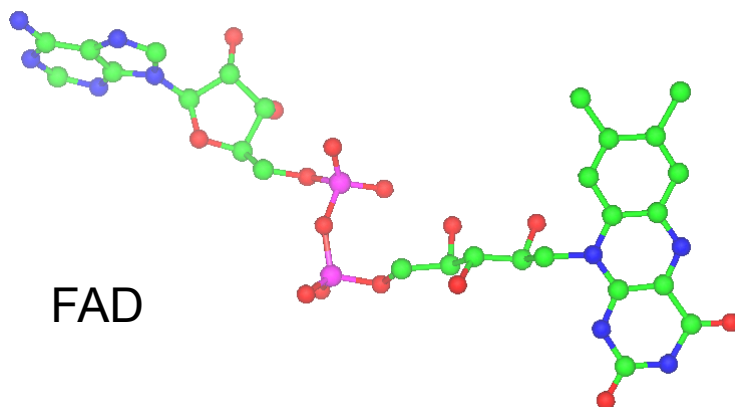
- Refine ligand coordinates to satisfy geometric constraints and to maximise fit to density.
- Choose best model.

Constructing the ligand: Graph search

Uses a sparse grid representation of the electron density at the chosen binding site & topology of known ligand.

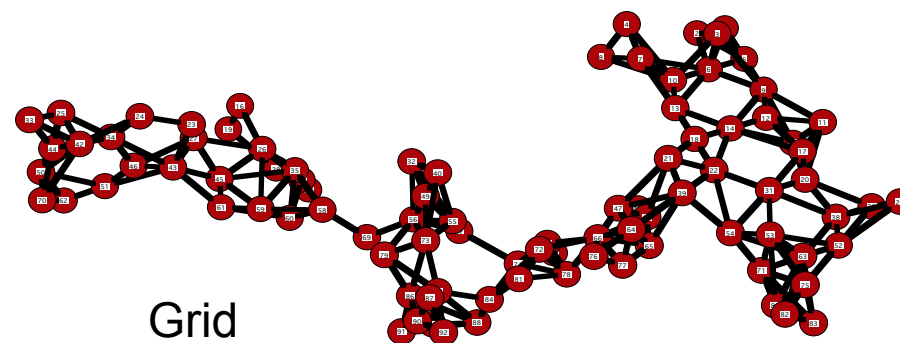
Knowledge about ligand

- A) Connectivity
- B) Distances
- C) Angles
- D) Chirality



Knowledge about grid

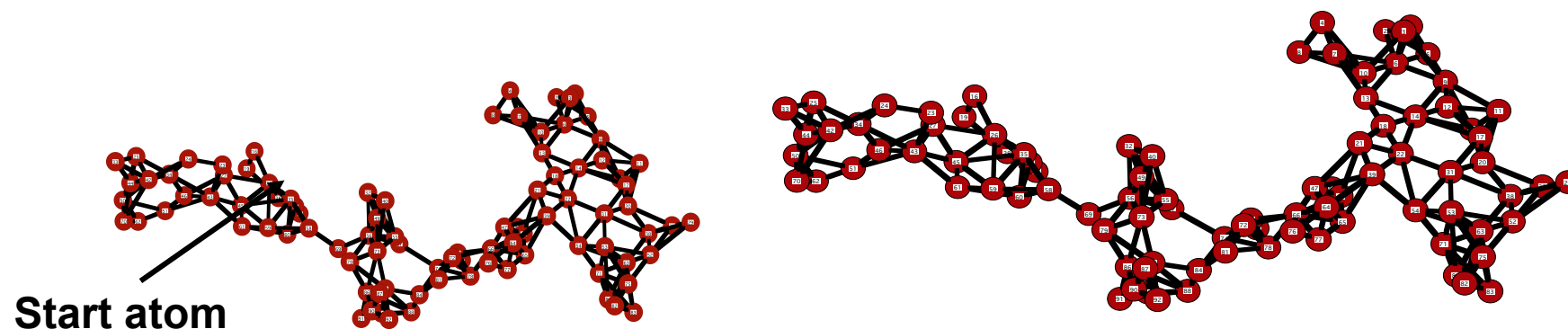
- A) Connectivity
- B) Distances
- C) Density
- D) *NO IDENTITIES!*



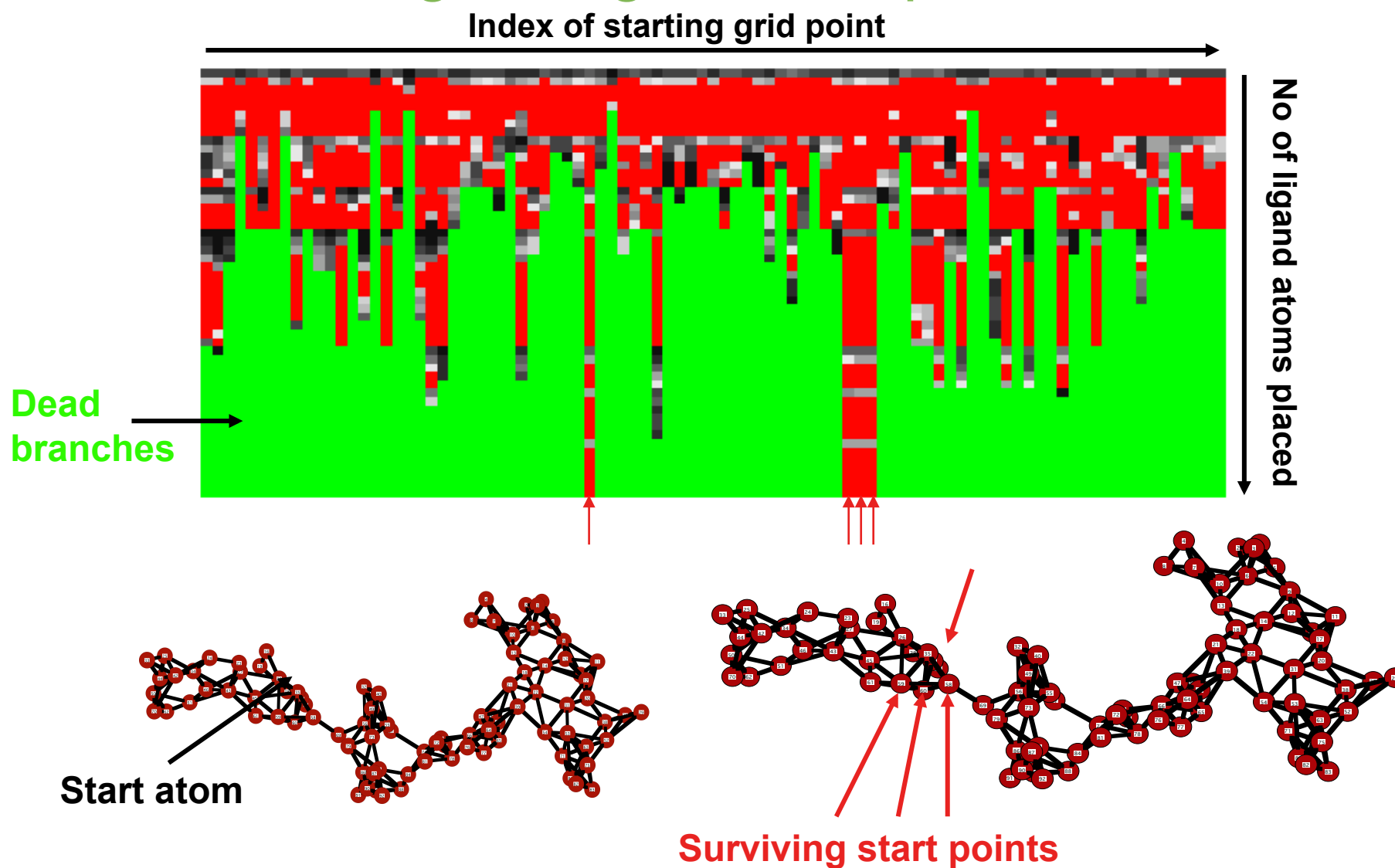
Constructing the ligand: Graph search

‘Label swapping’

- The ligand is expanded on the sparse density preserving connectivities.
- Every dummy ‘atom’ of the sparse grid is tried as a start point.
- An exhaustive graph search is performed.
- Models are scored by their fit to density and expected stereochemical features.



Constructing the ligand: Graph search



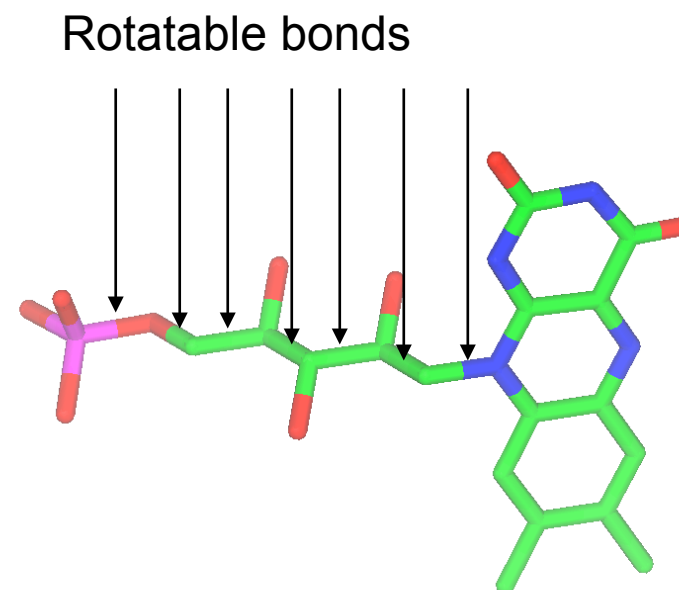
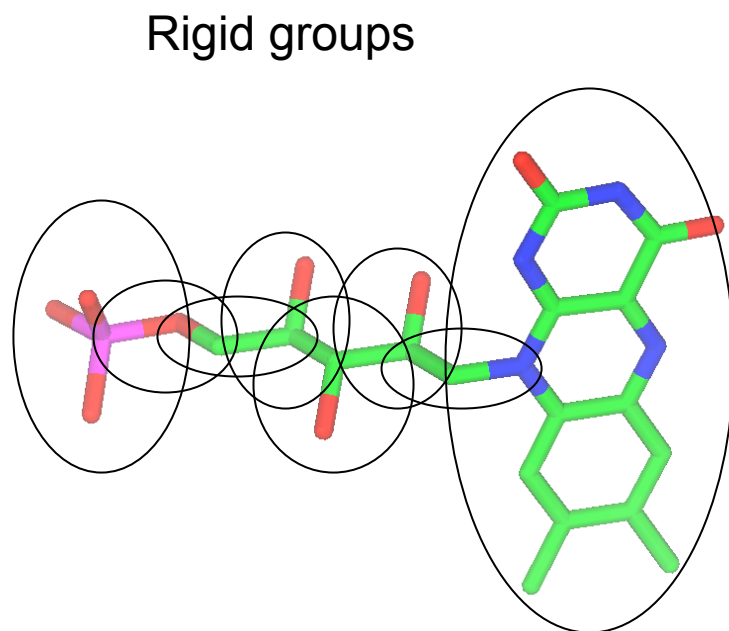
Constructing the ligand: Metropolis search

Perform a random walk in parameter space biased towards the optimum of a score function.

Parameter space: Position, Orientation and Conformation.

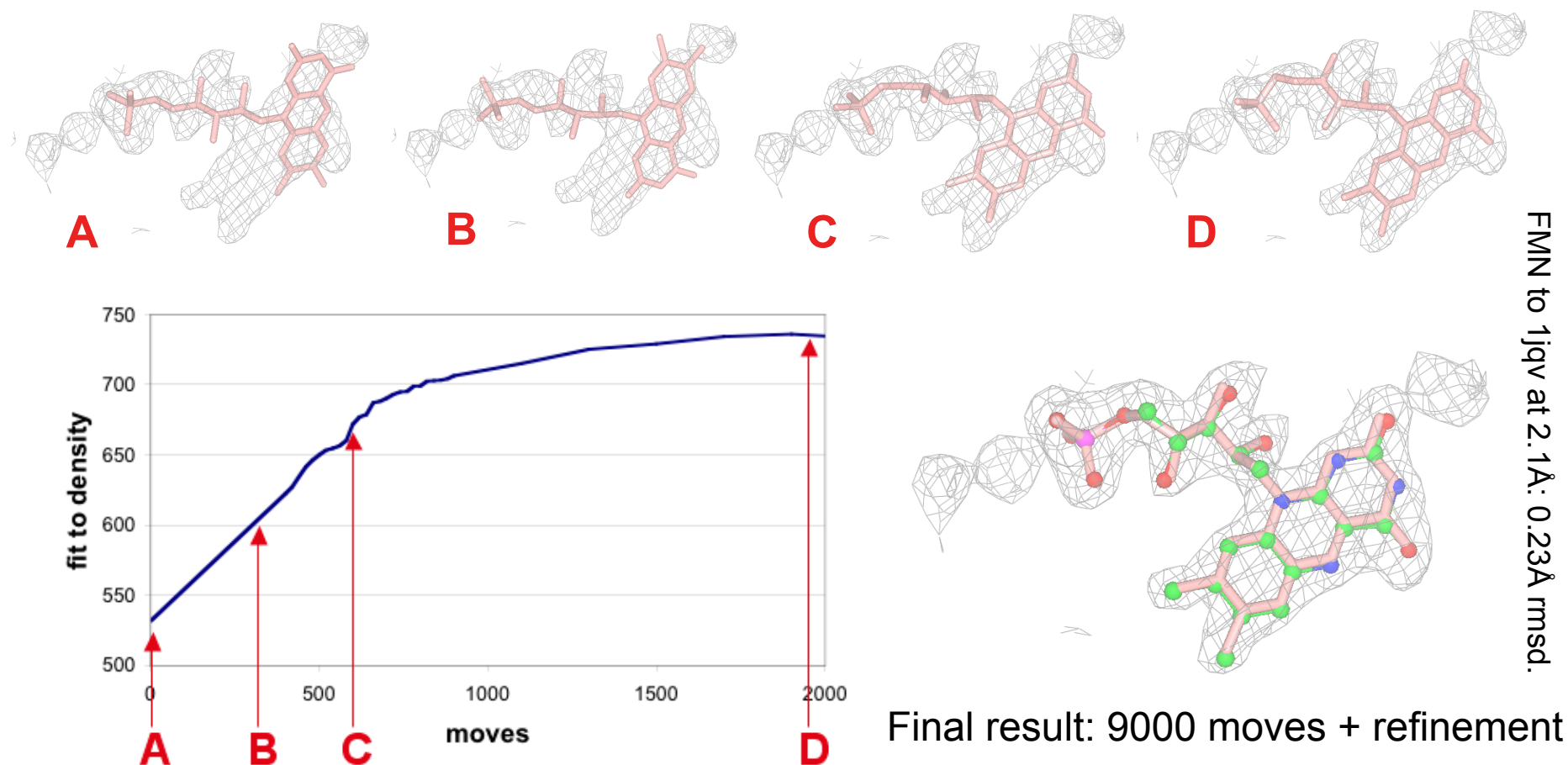
Score function: 'Pseudo' map correlation.

Advantage: Less degrees of freedom.



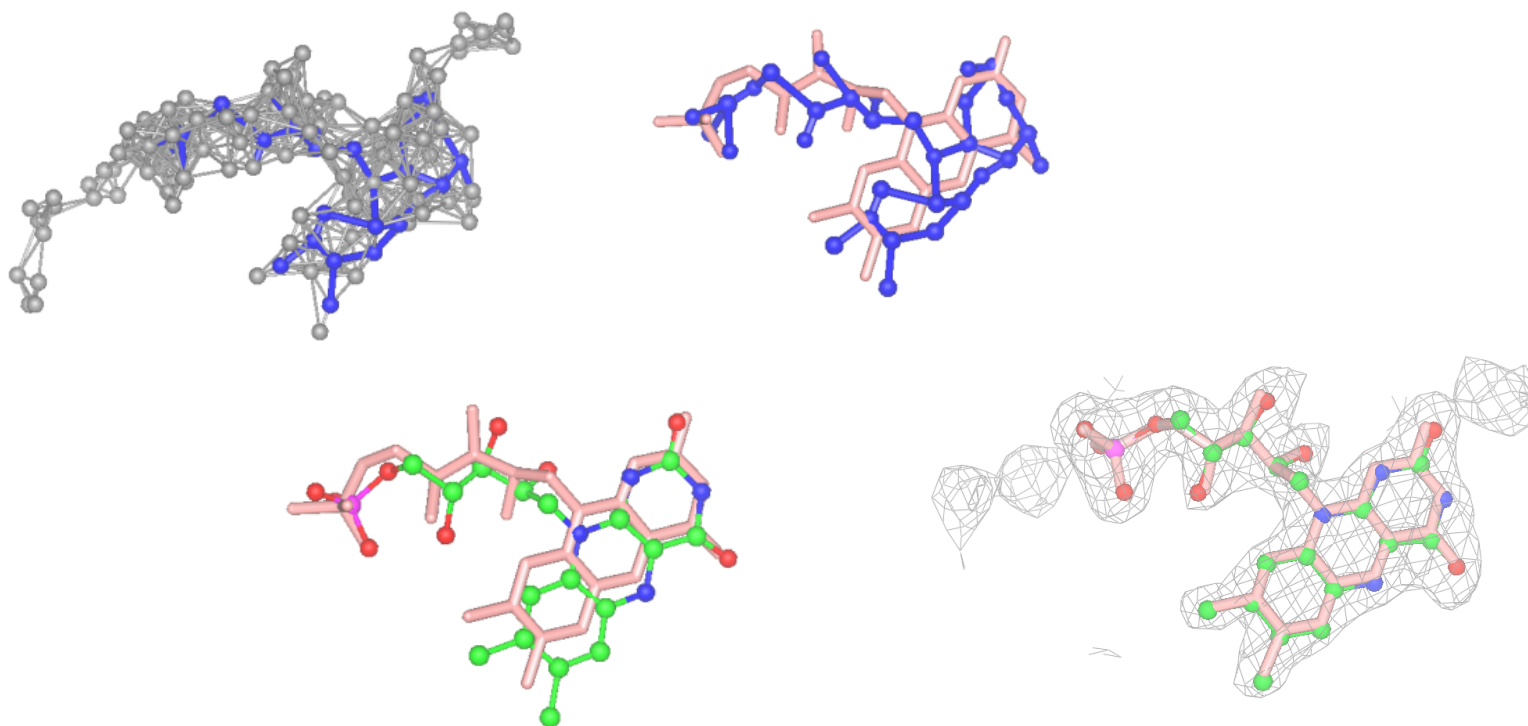
Construction with a Metropolis method

- The evolution of a model during a metropolis optimisation.



Real space refinement

- Uses the full density and topology information
- Employs a gradient method to optimise fit to density and ligand geometry simultaneously



Success rates

Tested on > 20k PDB entries from EDS

X-ray resolution limits: 1.0 to 3.0Å,
Building the largest fully occupied ligand.

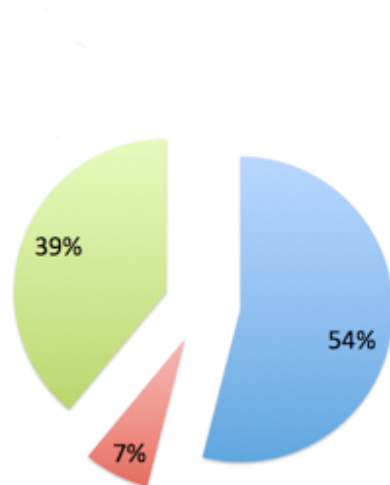
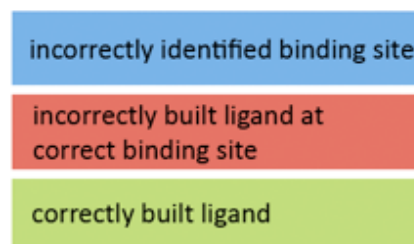
Correctness criterion: rmsd < 1.0Å

(Current Version 7.2)

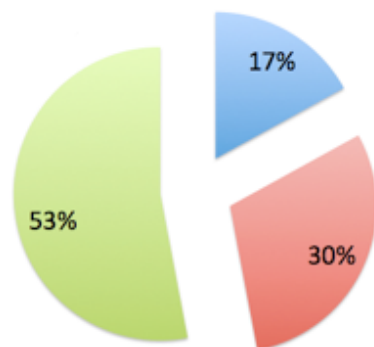
*Fragmentation-tree density representation for
crystallographic modelling of bound ligands.*

Langer GG, Evrard GX, Carolan CG, Lamzin VS.

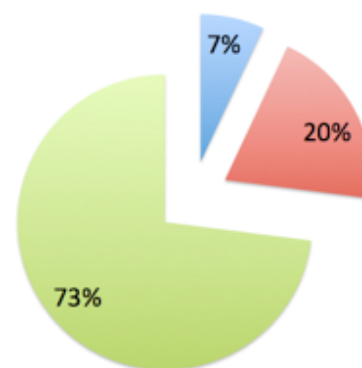
J Mol Biol. **2012** Jun 8; 419(3-4): 211-22.



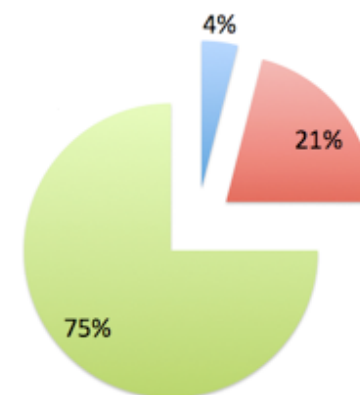
5...6 atoms,
any map corr.



7...100 atoms,
any map corr.

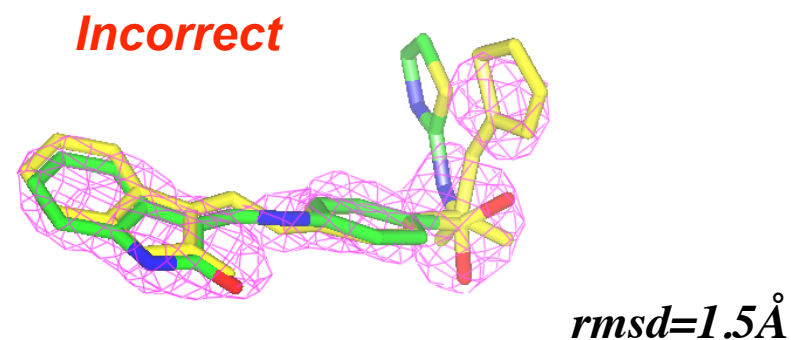
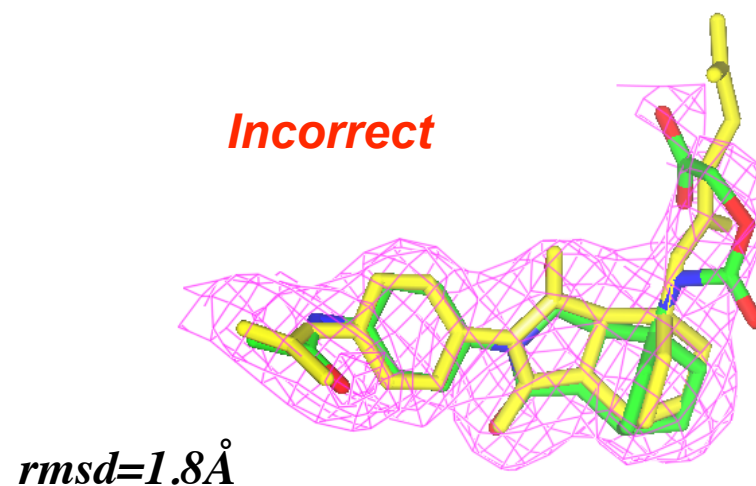
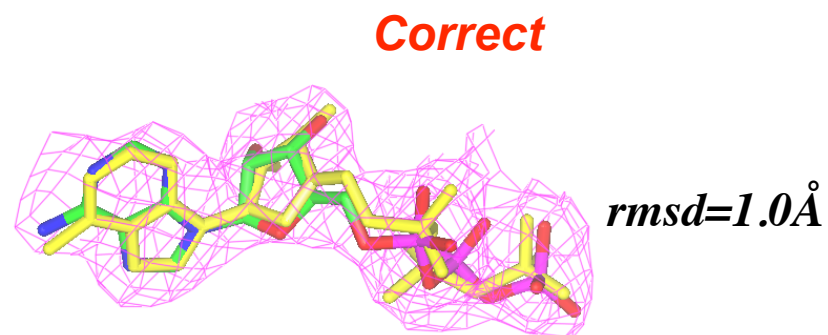
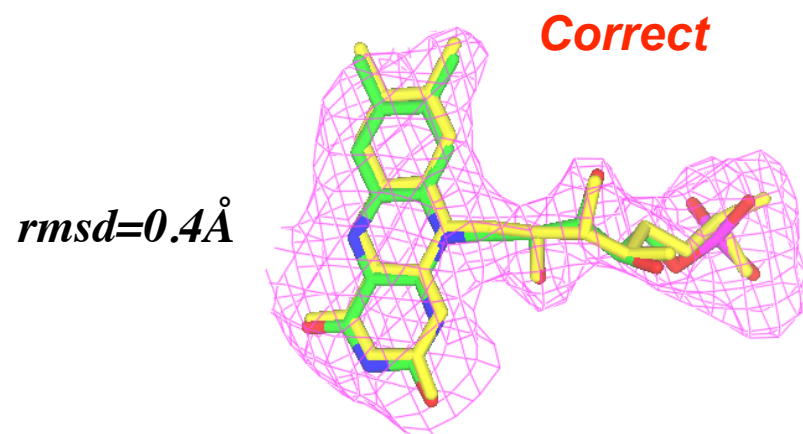


10...40 atoms,
map corr. > 80%



20...40 atoms,
map corr. > 80%

Success criterion: r.m.s.d. < 1.0Å



In yellow: the deposited ligand

Researcher's needs - #4 – partial ligands

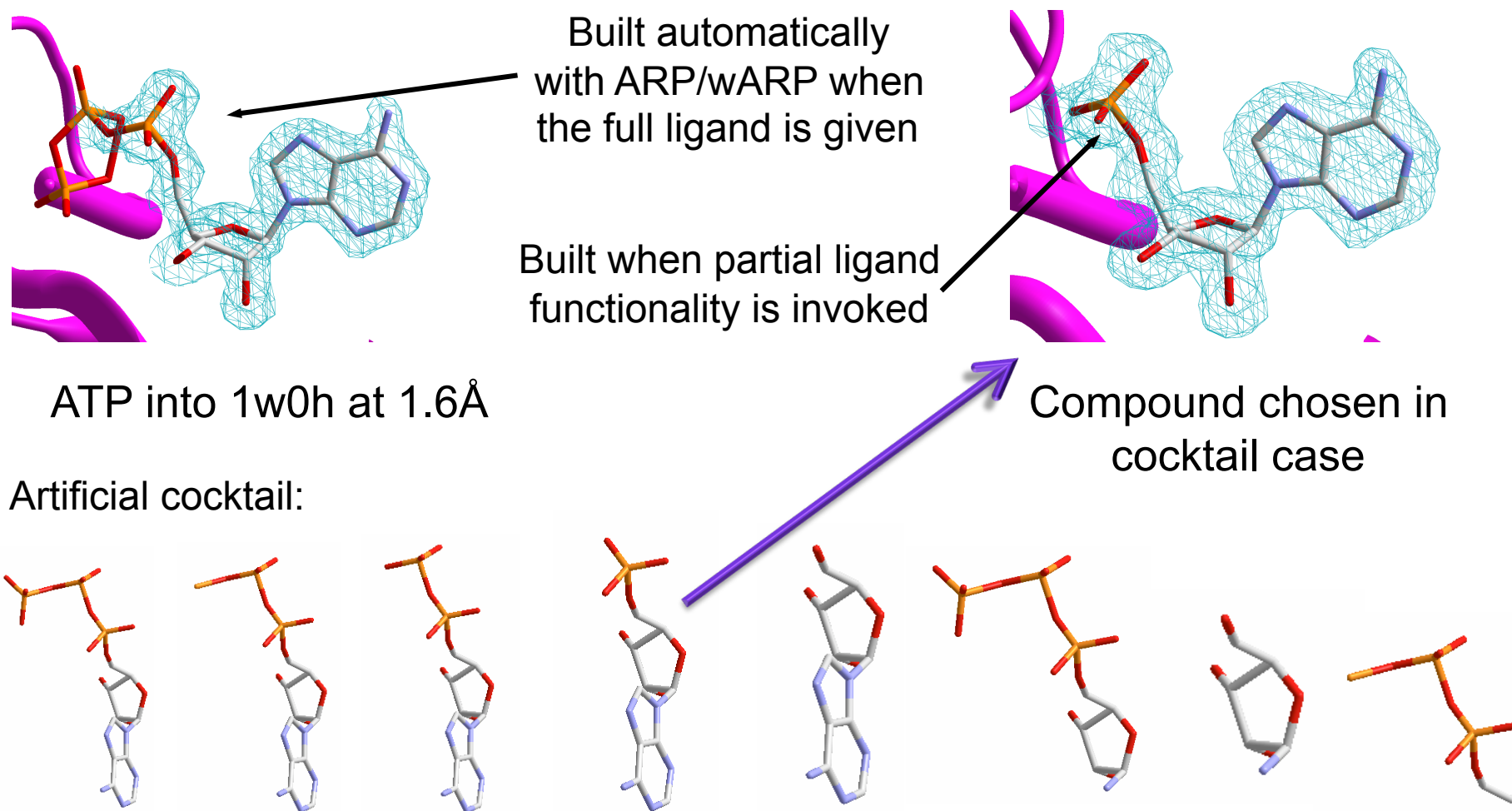
CCP4BB:

I have a ligand in the binding pocket and mediocre data. The core of the ligand is well defined, however there are chains/tails of the ligand which are not....

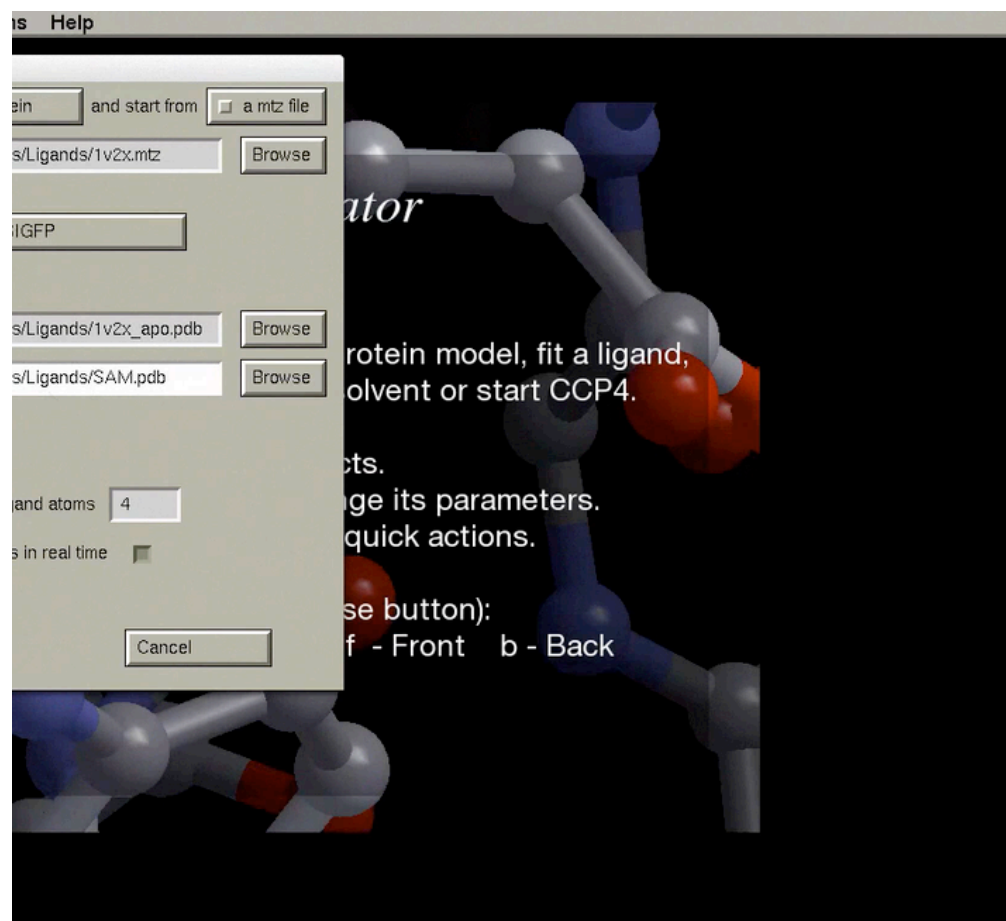
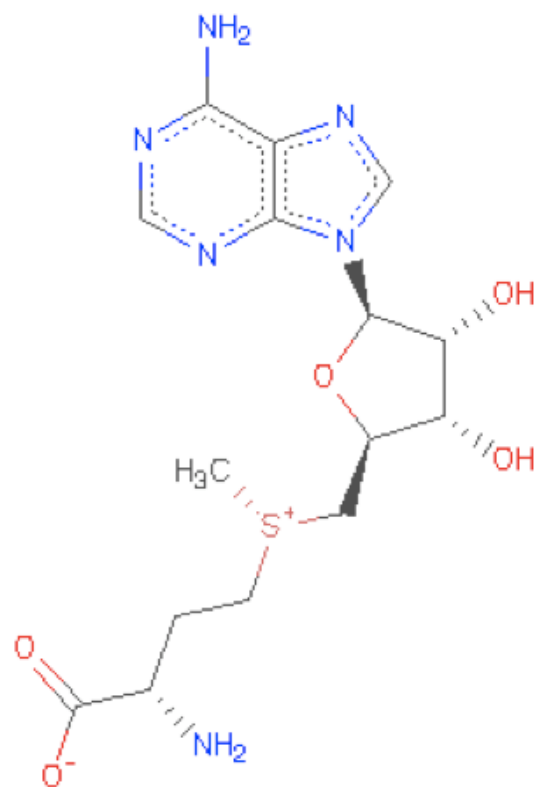
The question here is how the chains/tails should be modelled (if at all)

Francis E. Reyes

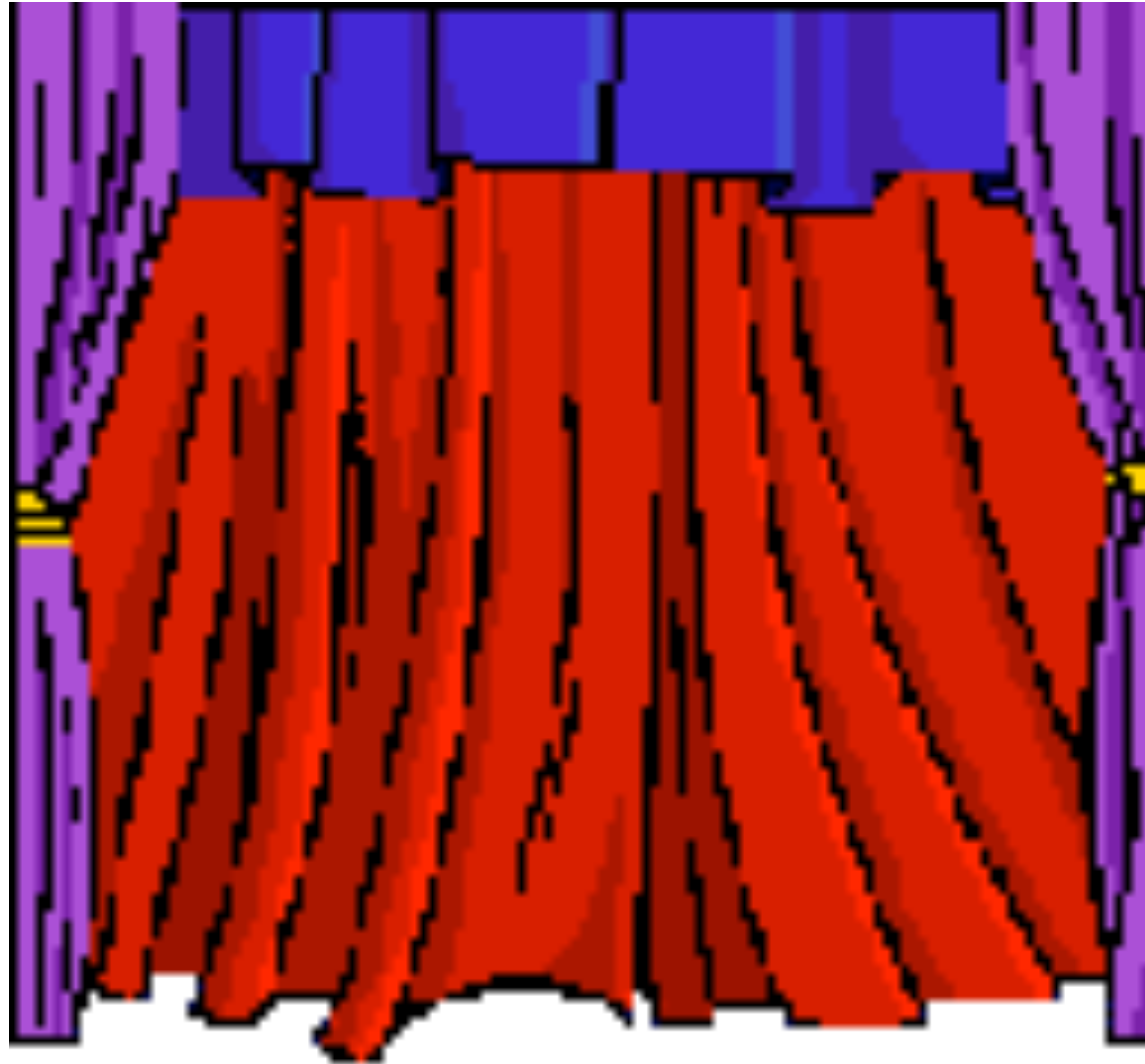
Partial disorder, partially occupied ligands



Partial ligand building in ARPNavigator



A little something new for your perusal



A little something new for your perusal



Building from 3-letter ligand ID

ARP/wARP Version 7.3.0: Ligand Build

Help

Job title

Build the ligand

MTZ in

Fobs

Protein model without ligand

Ligand molecule coordinates

Or ligand molecule three letter code

Ligand output file

Parameters

Use with REFMAC5 the protocol and do the Free R flag

Ligand building cycles

☐ Assume partial occupancy of ligand

☐ Keep waters

Refmac parameters ☐

Crystal parameters ☐

Test and comparison parameters ☐

Specification of restraints with external cif

```
[zsh] MALLORCA::/Users/carolanc
(~)auto_ligand.sh

Mon 22 Oct 2012 18:15:13 CEST

Usage:
auto_ligand.sh
    datafile {either mtzfile or mapfile}
    protein {starting_PDB_file_without_ligand}
    [ligand {PDB_file_with_ligand_to_fit}]
    [ligandcode {3-letter code of a ligand molecule,
    the code must be present in the refmac library}]
    [workdir {FULLPATH_WORKING_DIRECTORY}]
    [ligandfileout {output_PDB_file}]
    [fp {fp_label}] [sigfp {sigfp_label}] [freer {freer_label}]
    [nligandcycles {number_of_ligandbuild_cycles (default is 2)}]
    [search_model {PDB_file_with_model_at_expected_ligand_site}]
    [search_position {X Y Z}]
    [search_radius {radius_in_angstroms}]
    [reflist {textfile with FULLPATHnames of fitted ligands for comparison}]
    [extralibrary {user_defined_library_for_Refmac5}, additionally
    if this library contains data for the ligand to be built, then these
    parameters are used to derive ligand topology to the highest level
    of priority (ahead of REFMAC cif or coordinate-derived topology)]
    [model {modeling the ligand as a whole or as a fragment}]
    [keepwaters {1 for keeping them before computing the difference map}]
    [parfile {parfilename_if_only_parfile_is_to_be_created}]

- Optional command line arguments are given in square parentheses
- All input files are assumed to be located in working directory
  unless they are given with full path
- If workdir is not given, the current directory will be assumed
- All output files will be written into workdir/subdirectory
- If no ligand is specified then auto identification of
  the ligand will be attempted provided that a search position
  is given (experimental)

(carolanc@MALLORCA)-(06:15P--10/22)--
(carolanc@MALLORCA)-(06:16P--10/22)--
(~)█
```

Note that in general, all that is required for ligand fitting, whether started from ARPNavigator, CCP4I or the command line, is a protein PDB file and either a map or MTZ file!

See the tutorial later!

Acknowledgements

Developers

EMBL Hamburg: Ciaran Carolan, Philipp Heuser, Victor Lamzin, Tim Wiegels, Ioan Vancea

NKI Amsterdam: Krista Joosten, Robbie Joosten, Tassos Perrakis

Collaborators

EMBL Hamburg: Gleb Bourenkov

MRC/LMB-Cambridge: Garib Murshudov's group

Rutherford Laboratory: CCP4 team

Former members

Saul Hazledine, Gerrit Langer, Richard Morris, Peter Zwart,
Francisco Fernandez, Matheos Kakaris, Olga Kirillova, Wijnand
Mooij, Diederick De Vries, Marouane Ben Jelloul, Johan Hattne,
Tilo Strutz, Guillaume Evrard, Serge Cohen, Venkat
Parthasarathy, Babu Pothineni, Helene Doerksen

Funding



A Vision for the Future

