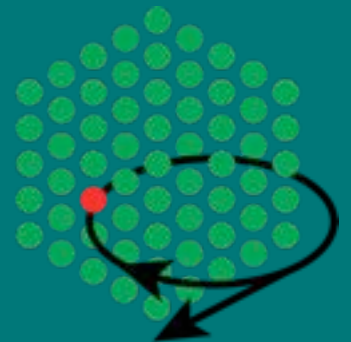


Every thing you always “wanted” to know about  
**ARP/wARP** (and were probably afraid to ask...)

Tim Wiegels  
Lamzin Group (ARP/wARP), EMBL Hamburg

EMBL



# About EMBL

- Established in 1974 to reflect the need for a European centre for research and training in molecular biology
- A basic research institute funded by public research monies from 20 member states
- Over 1300 employees from 60 nations
- Approximately 80 independent research groups



## EMBL's mission

- Basic Research in Molecular Biology
- Advanced Training at all Levels
- Development of New Technologies and Instruments
- Services and Facilities for the Community
- Technology Transfer Discoveries

**Check [www.embl.de](http://www.embl.de) for PhD and postdoctoral job opportunities**

# The EMBL Units

## EMBL Heidelberg (Main Laboratory)

- Gene Expression
- Cell Biology
- Structural and Computational Biology
- Developmental Biology

## EMBL Monterotondo

- Mouse Genetics

## EMBL Grenoble

- Structural Biology

## EMBL / EBI Hinxton (Cambridge)

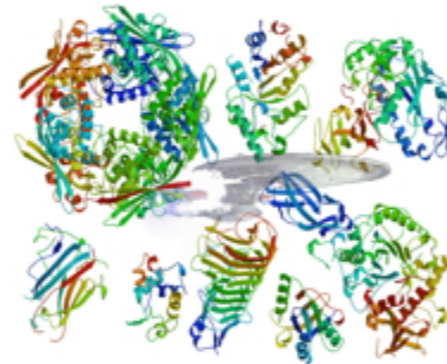
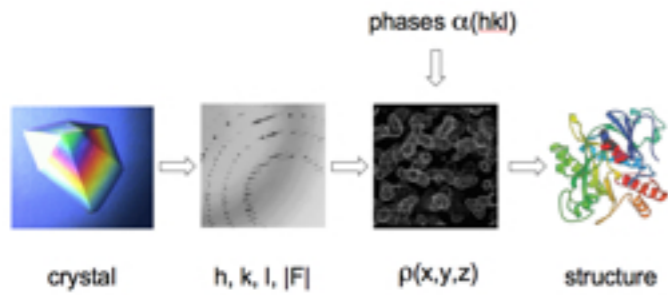
- Bioinformatics

## EMBL Hamburg

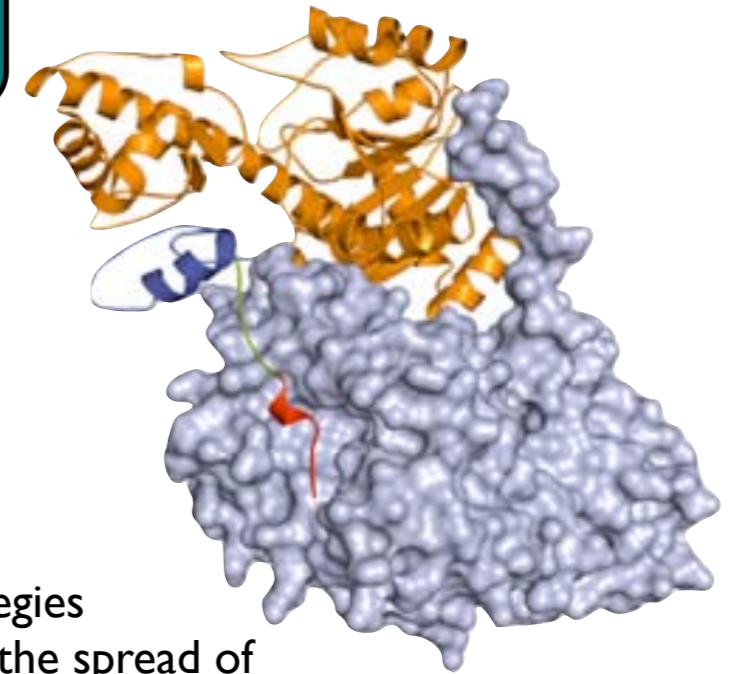
- Structural Biology



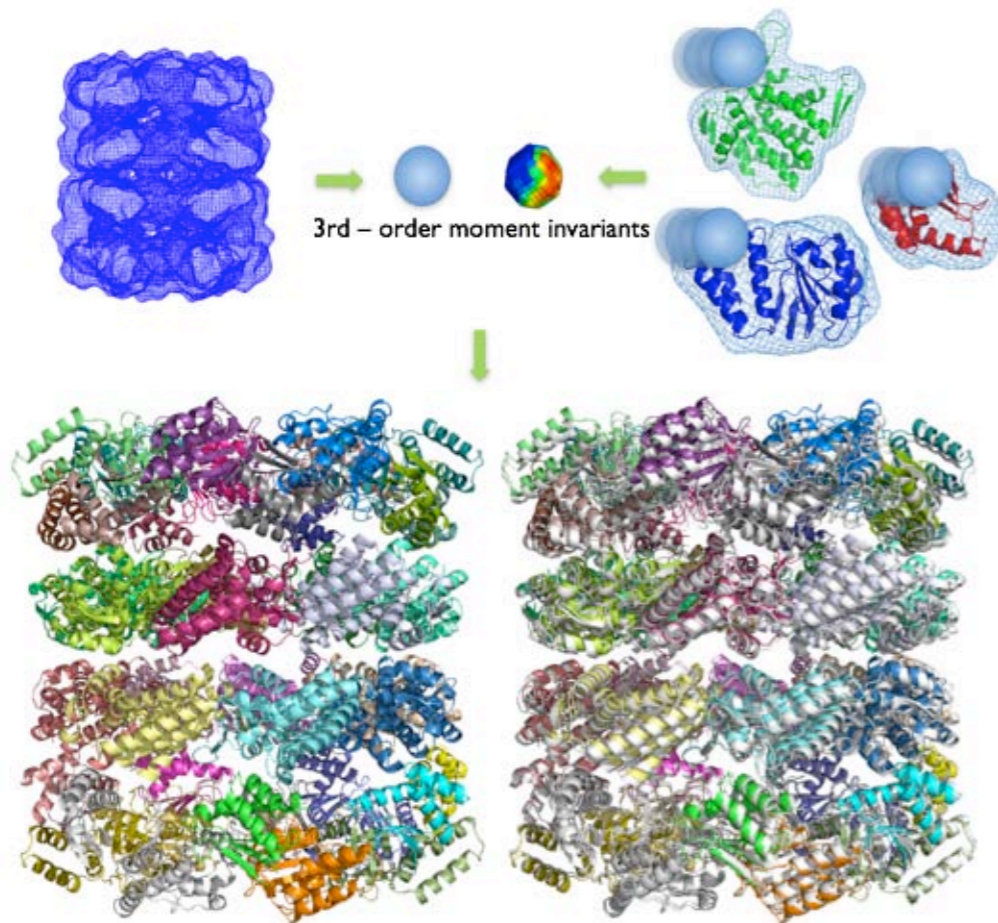
# The Lamzin group:



7.2  
ARP  
wARP

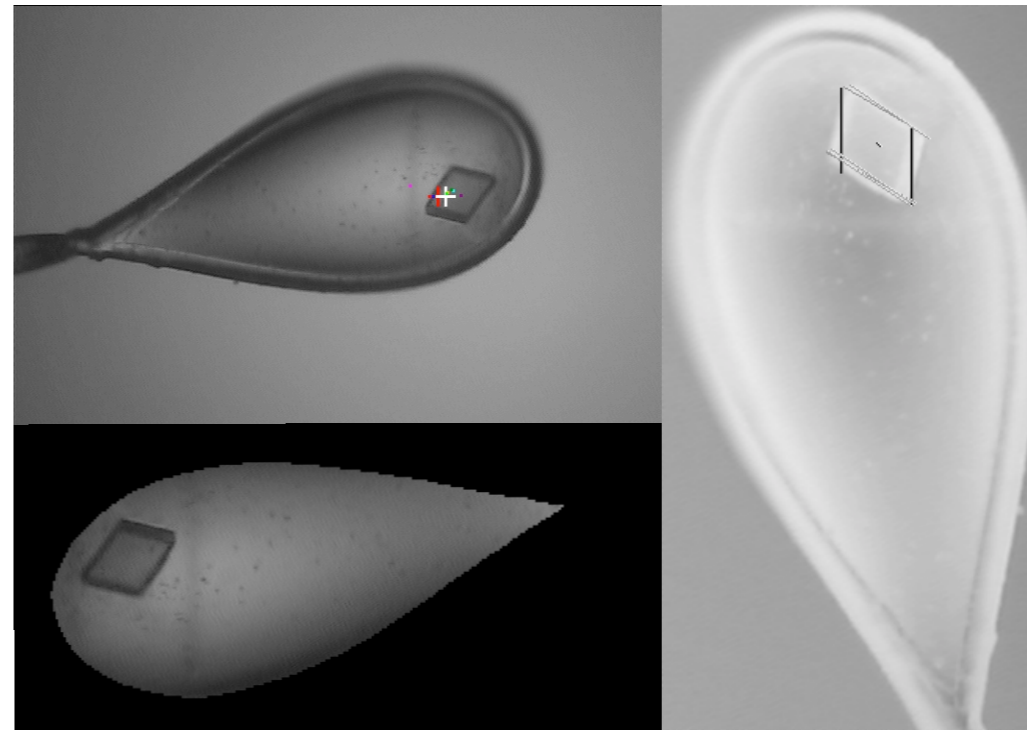


novel strategies  
to combat the spread of  
multi-drug resistant **malaria**

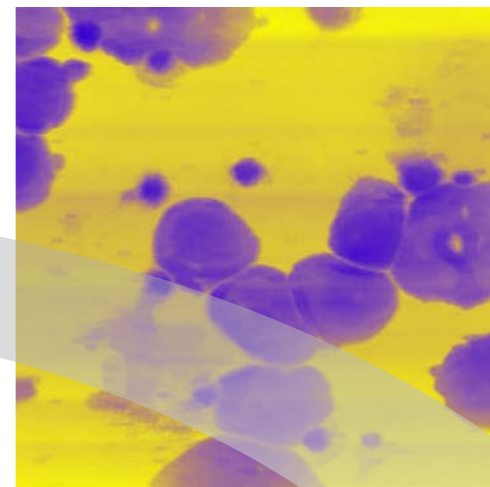
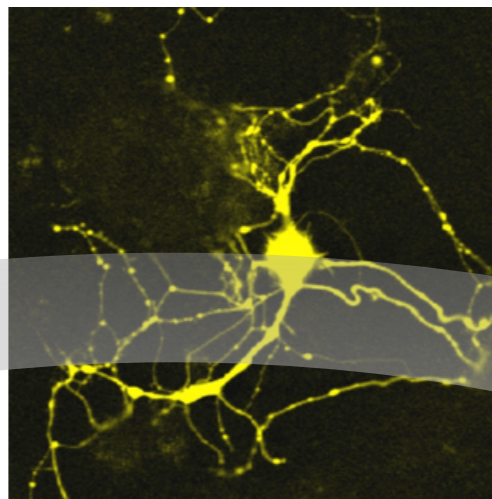
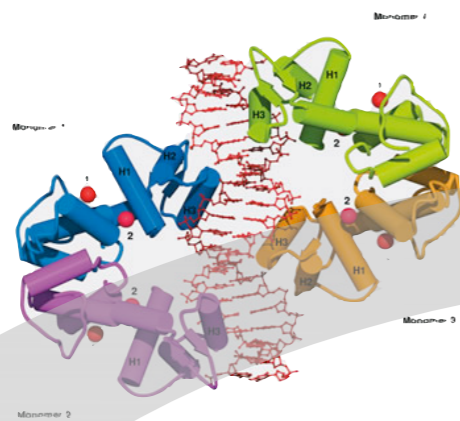


Placing of known 3D domain-structure in very **low resolution** density maps

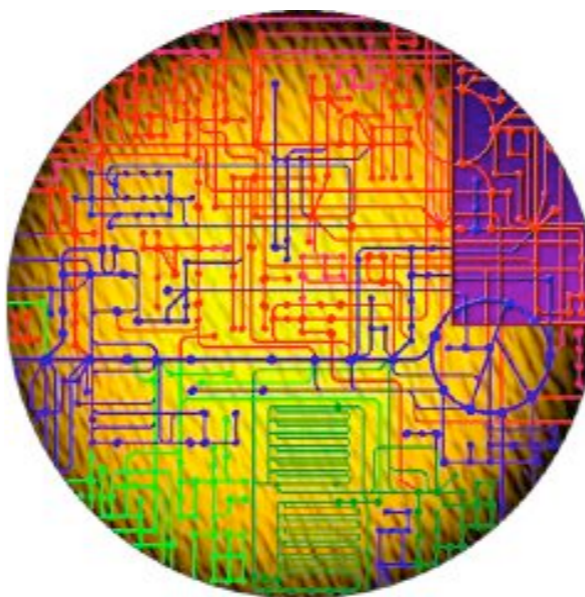
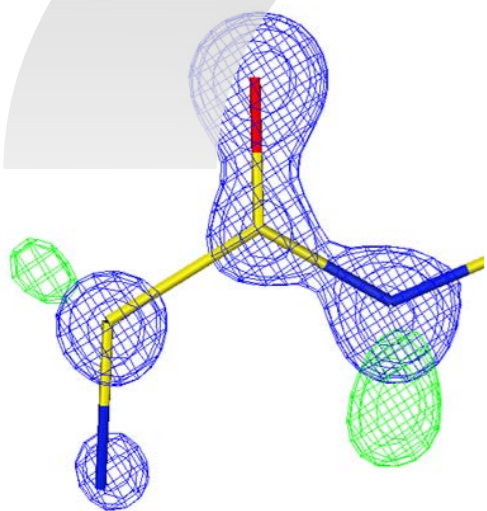
## XREC - Automated crystal recognition



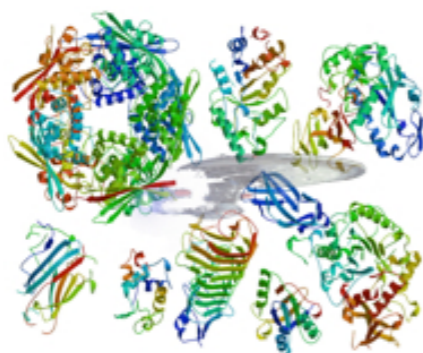
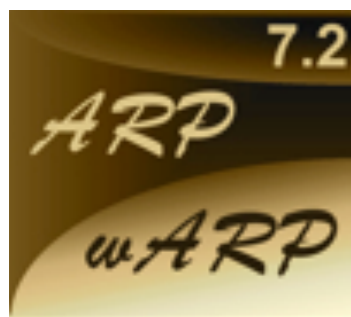
# Towards structural systems biology



$10^{25}$  atoms

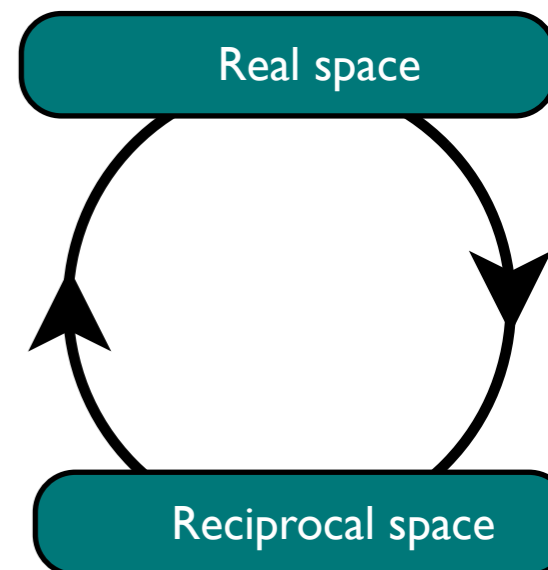


# Fundamental ARP/wARP concepts



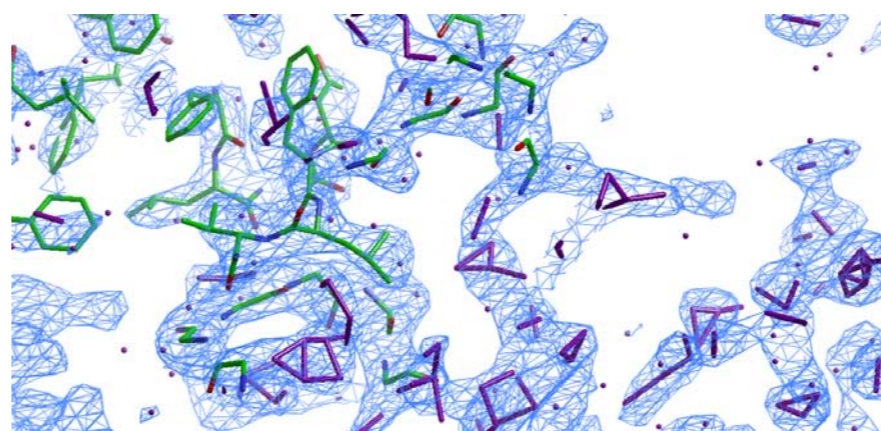
## Combined model building and refinement

Refinement engine, REFMAC5 – Murshudov et al., (1997)  
unified process



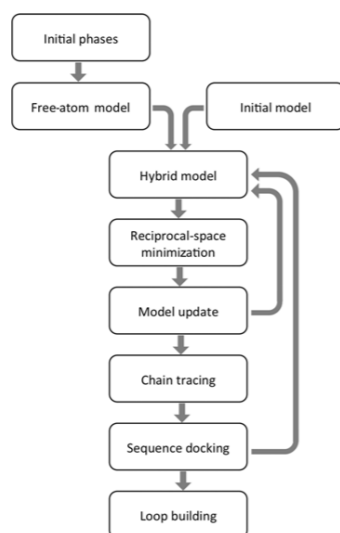
## Hybrid models

'Free atoms' – Agarwal and Isaacs (1985)  
combined with built fragments



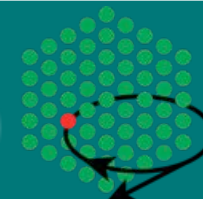
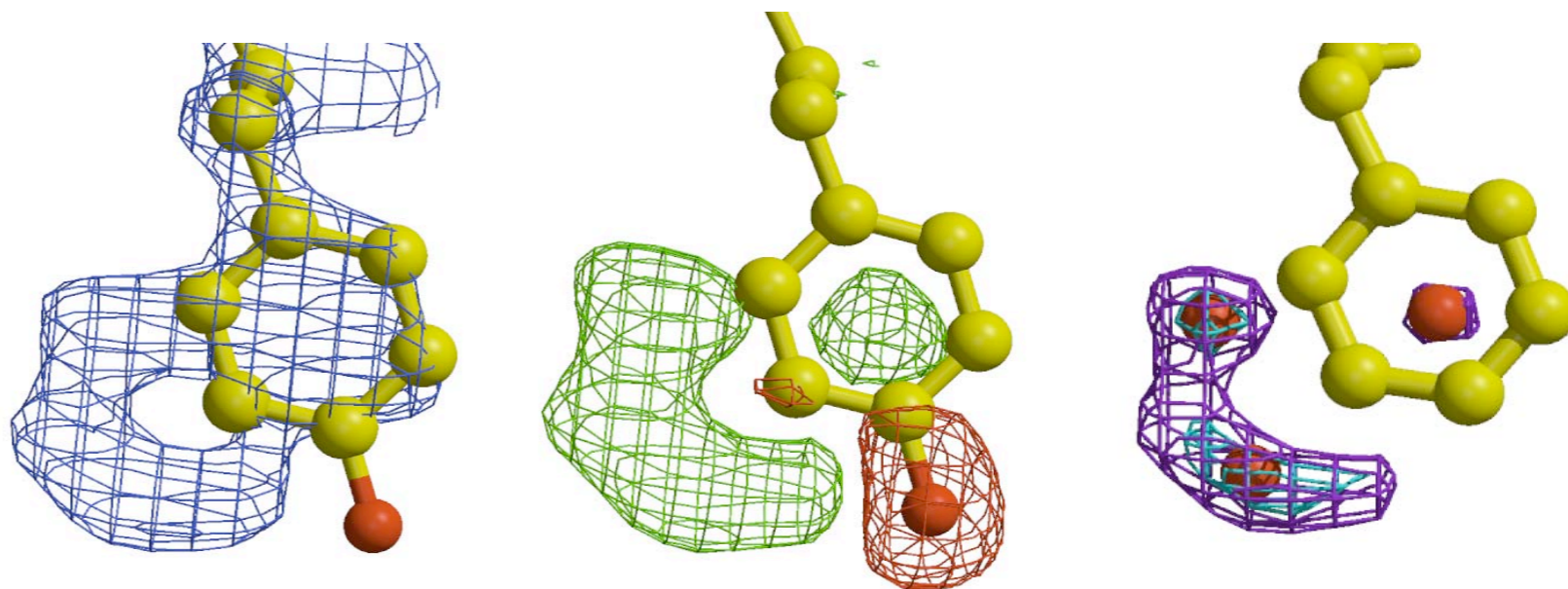
## Feed-back loops:

Iterative process

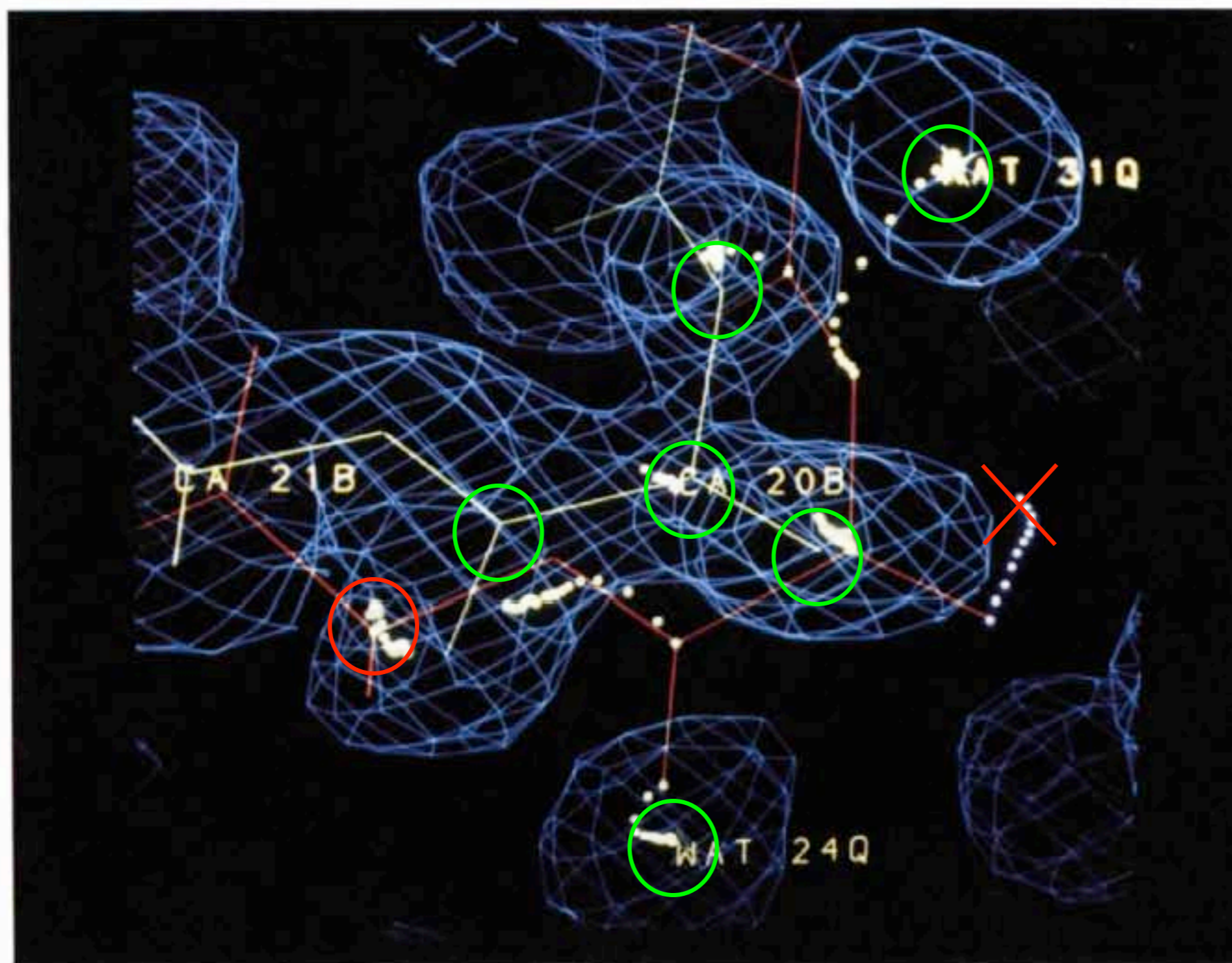


## Model update

Removal and addition of atoms based on density at atomic centres



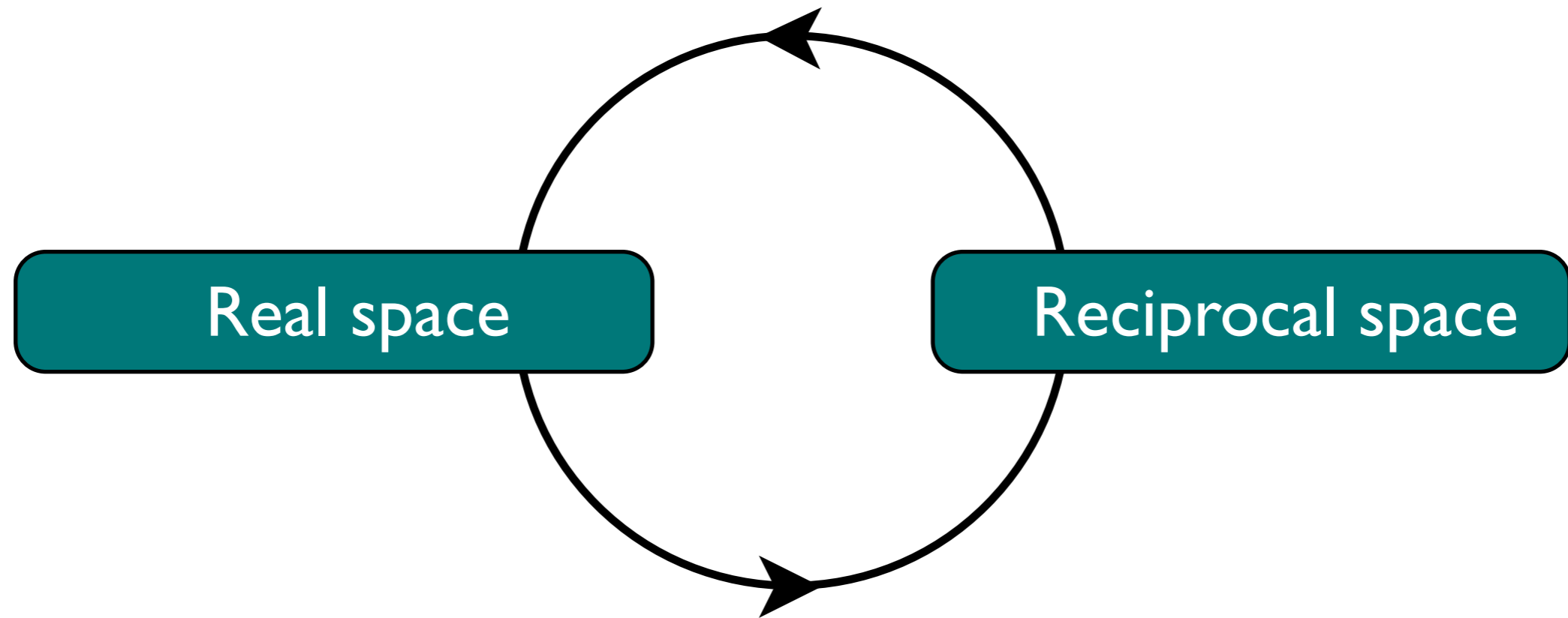
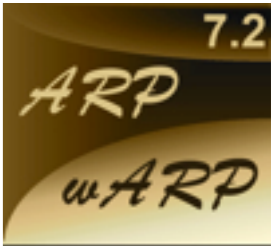
# Fundamental ARP/wARP concepts



Lamzin, V.S. & Wilson, K.S. (1993)  
Automated refinement of protein  
models. *Acta Cryst.* **D49**, 129-147.

- one of the leading **MX** software projects
- aimed at building structures of **proteins**, nucleotides, ligands and their complexes
- pattern recognition to build **models** from MX **electron density maps**

# Fundamental ARP/wARP concepts



- **Real space**

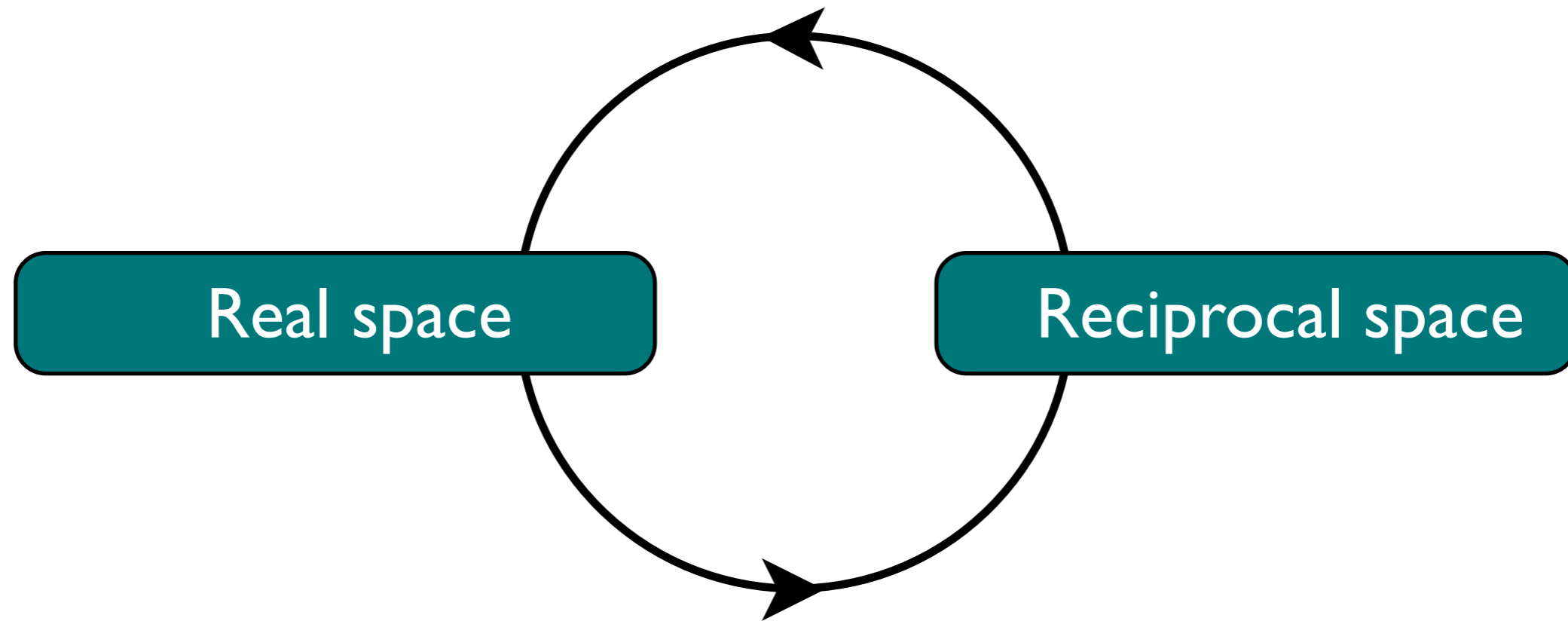
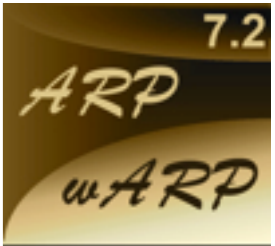
- **atom update** by removing and adding free atoms based on current density
- **hybrid models** made up of atoms having chemical identity and free atoms

- **Reciprocal space**

- **unrestrained refinement** of free atoms
- **restrained refinement** of chemically assigned atoms

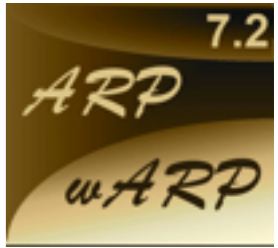


# Fundamental ARP/wARP concepts

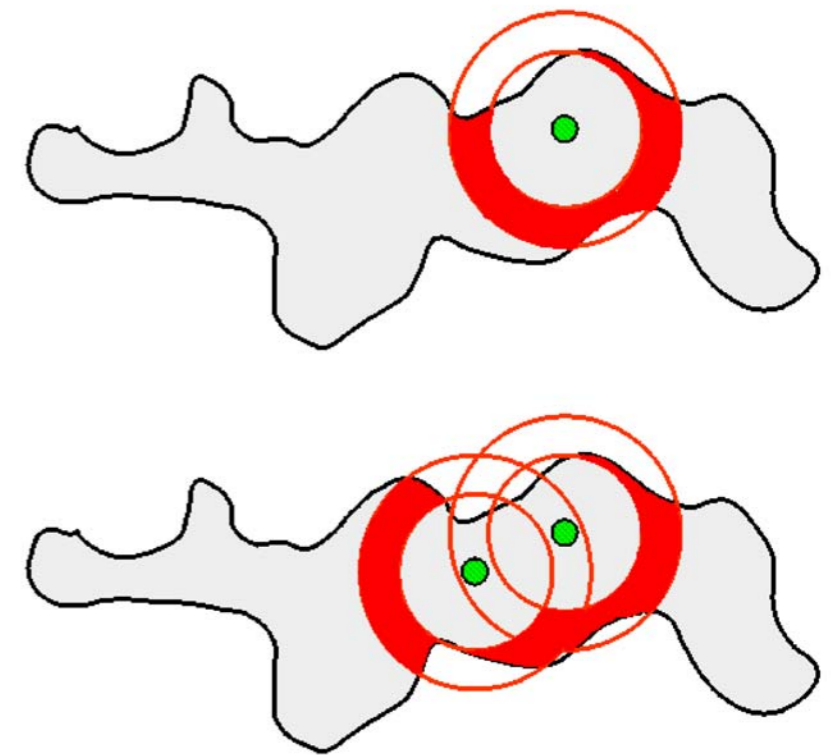
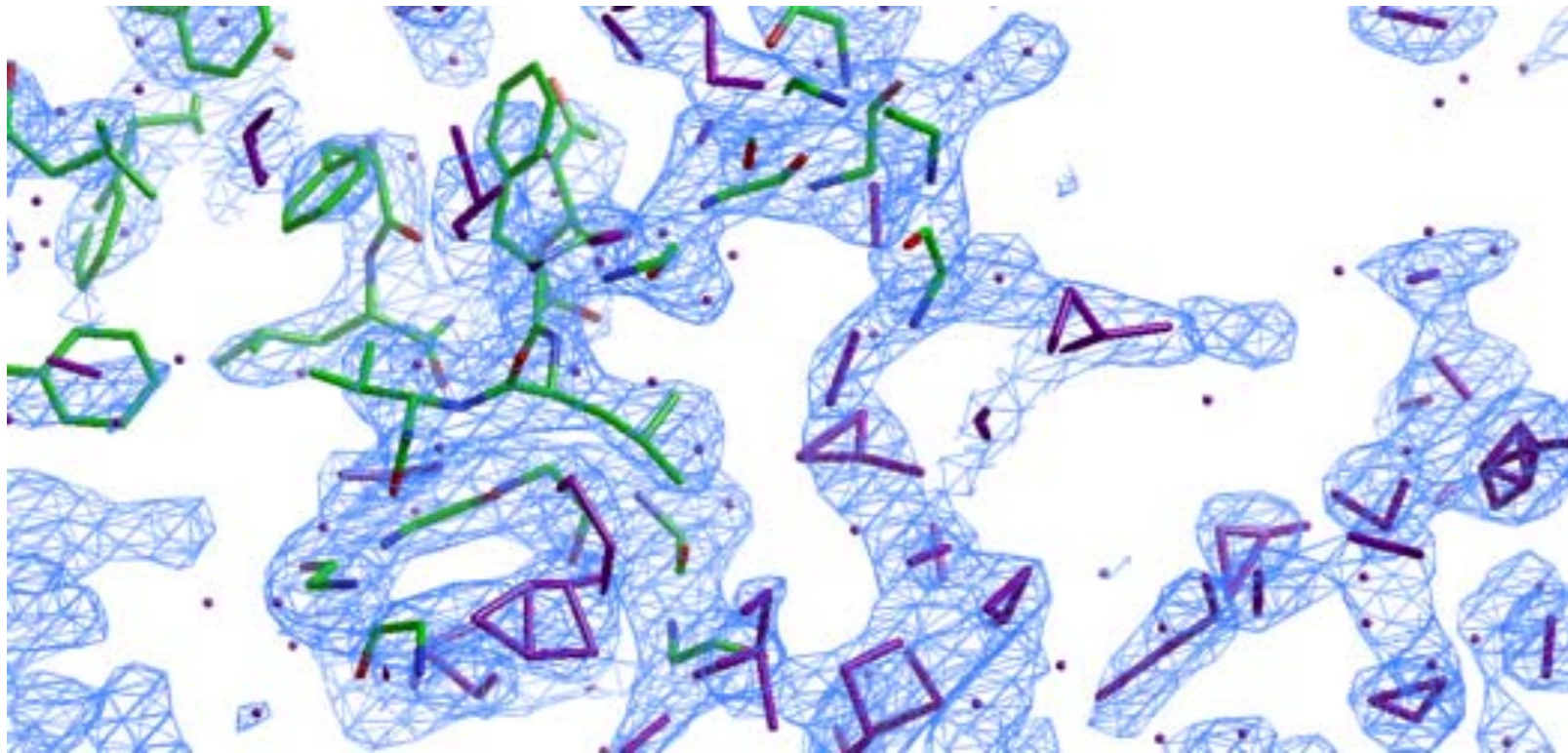


- hybrid model undergoes **optimization** with the REFMAC engine to better fit **experimental data** and **stereochemical expectations**
- if quality is high enough **phases** should **improve** overall and a more complete model can be constructed from the **enhanced density**
- forming an **unified process** of model building and refinement

# Basics: The Hybrid Model



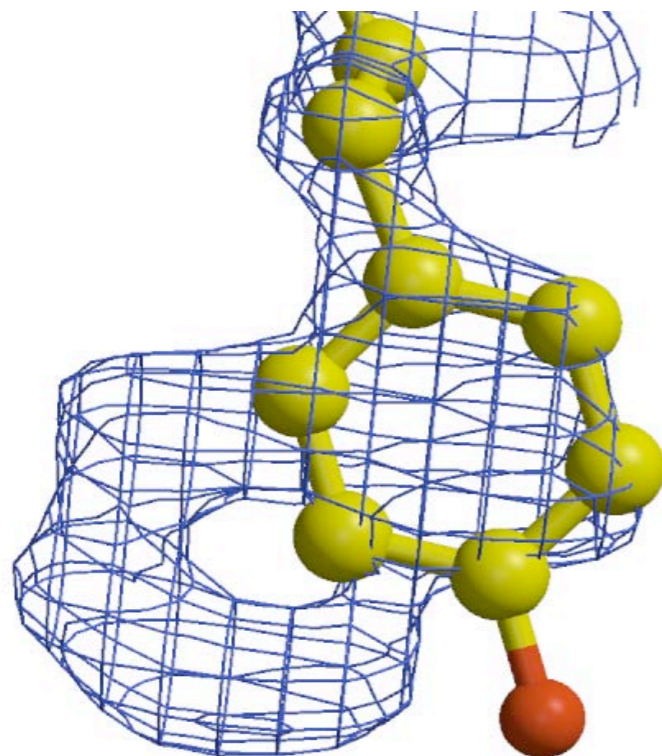
- A **partial** model is used together with a **free-atom** model
- Chemically assigned parts of the model provide **restraints** for refinement
- The scheme of restraints and the free atoms are **iteratively updated**
- The hybrid model is **converging** to the final model
- finding a **parametrisation** of the spatial density distribution



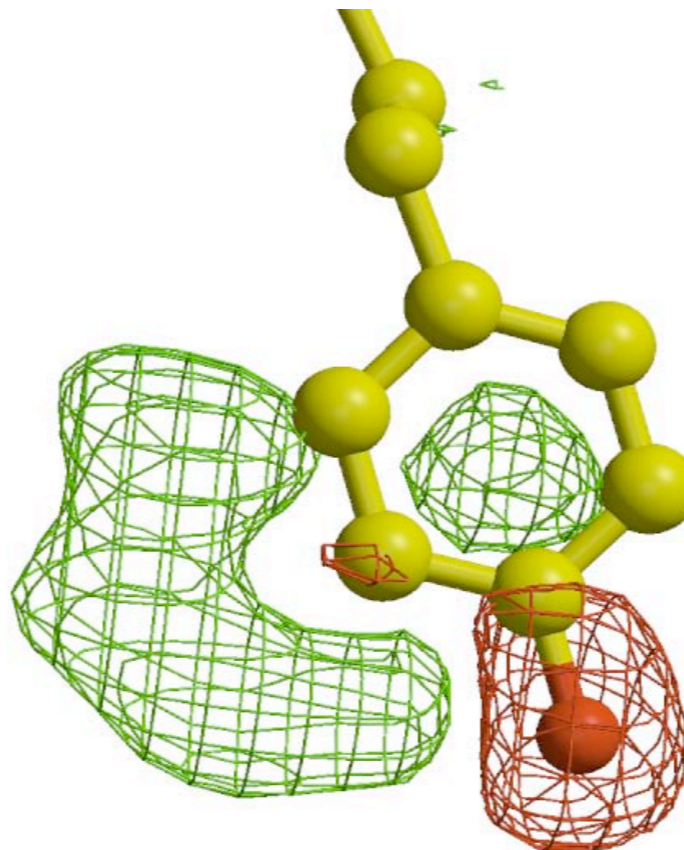
# Basics: The Atom Update



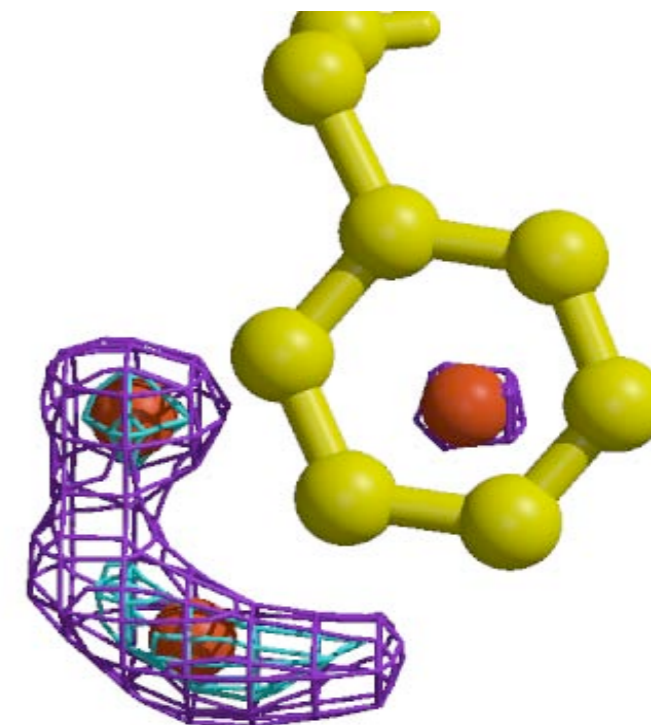
- Removal and addition of atoms based on density at atomic centers



2Fo-Fc map



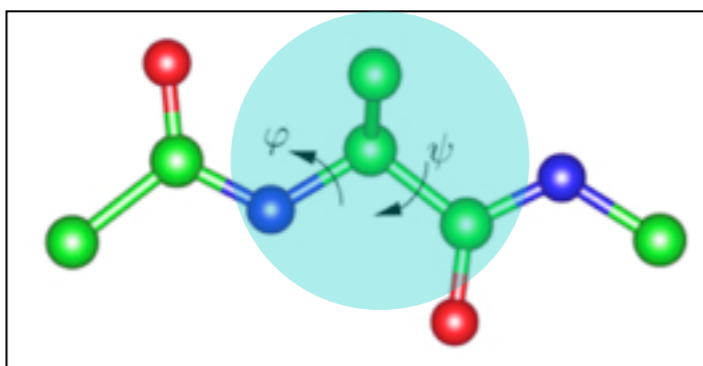
Fo-Fc map



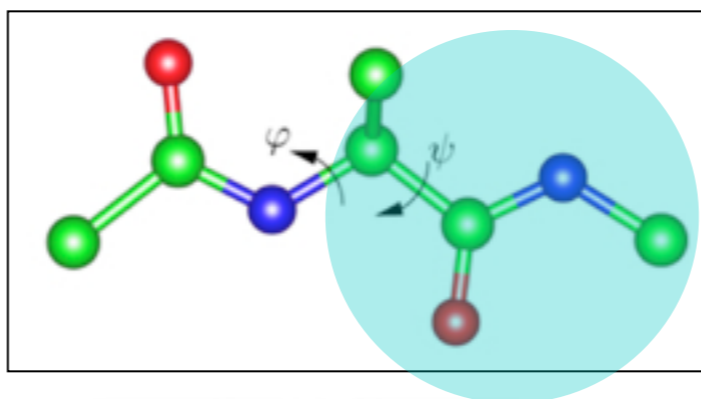
```
ARP/wARP will be iterated with REFMAC5  
50 refinement / model update cycles will be run in total.  
Atoms will be removed below 1.0 sigma in 2mFoDFc map and added above 3.2 sigma in mFoDFc  
map.
```

# Methods for Building Protein Structure

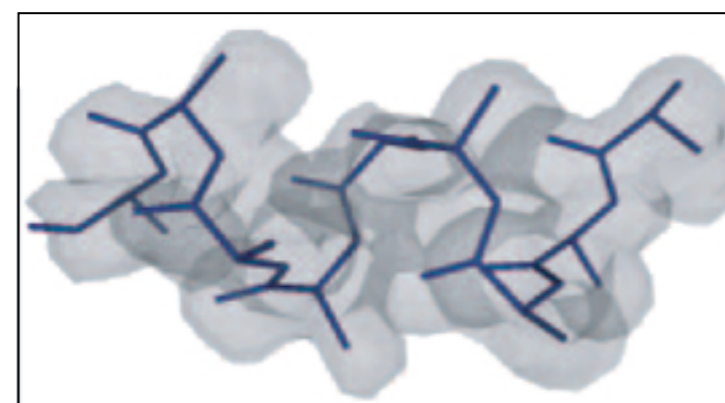
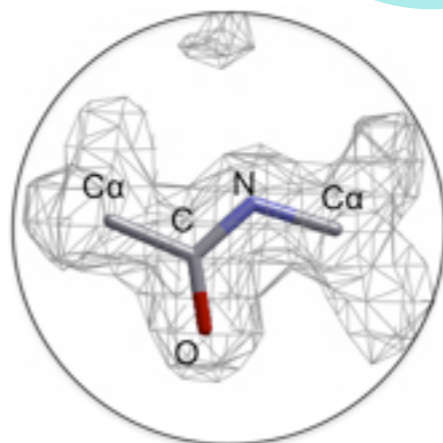
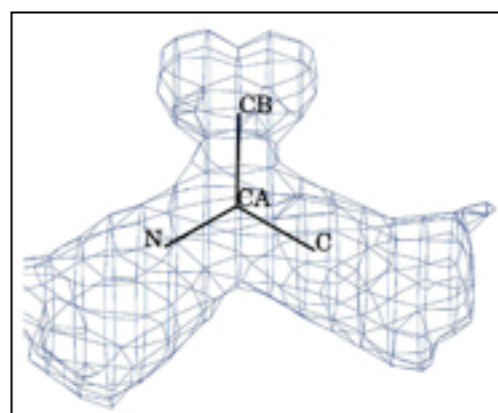
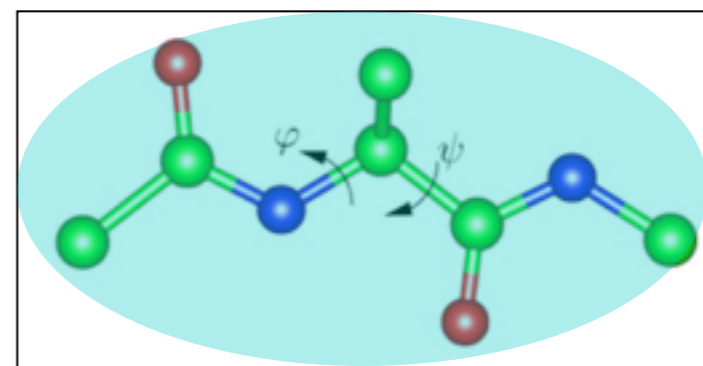
TEXTAL, Buccaneer



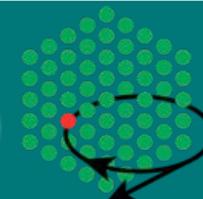
ARP/wARP



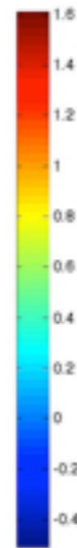
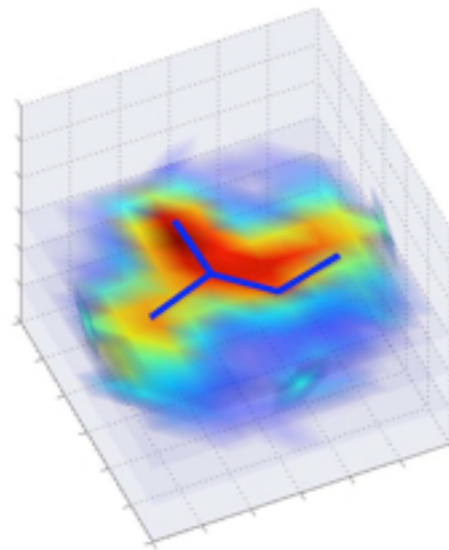
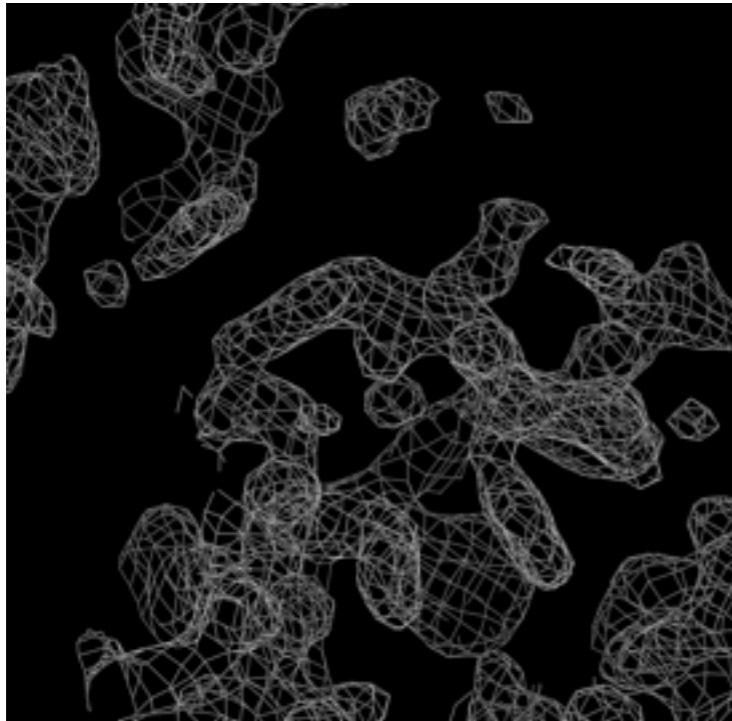
Resolve ACMI



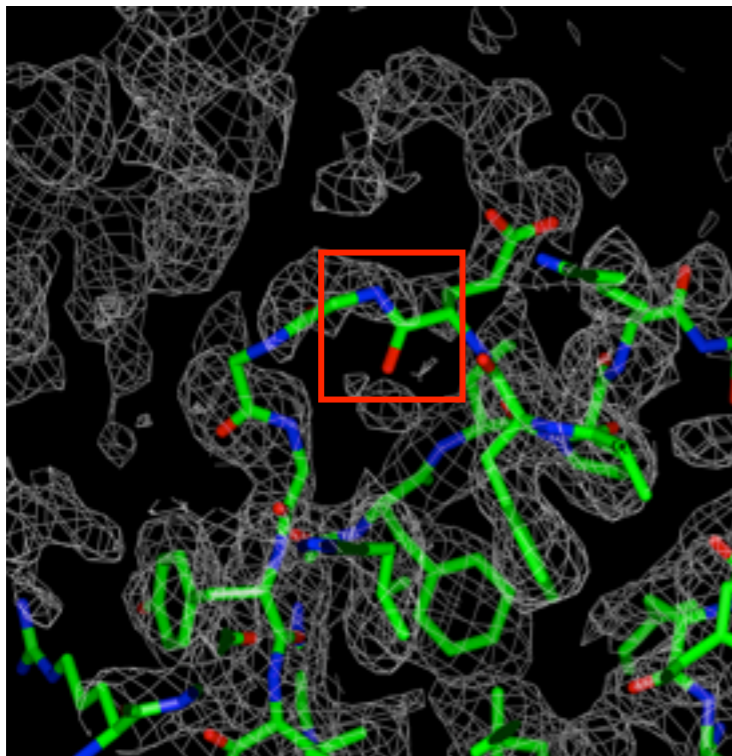
Peptides	-	✓	✓
Di-peptides	-	-	✓
Short fragments	-	-	✓
Long chains	-	-	-



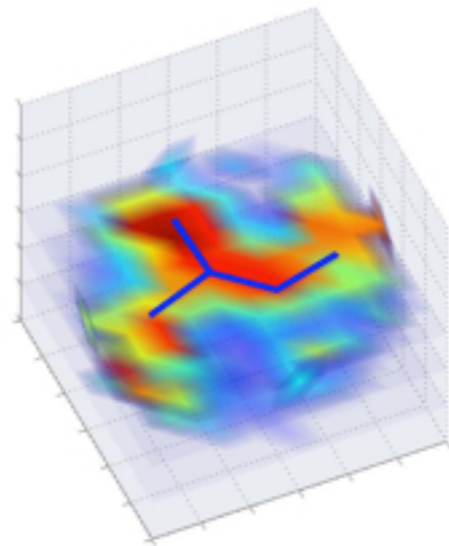
# Autotracing: Recognising peptide units



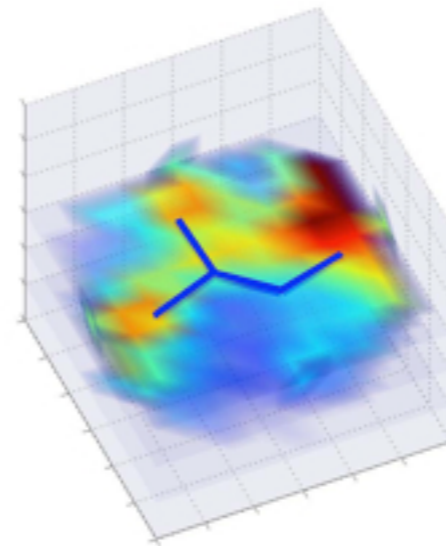
The task:  
Separating **true** peptides  
from **false** ones.



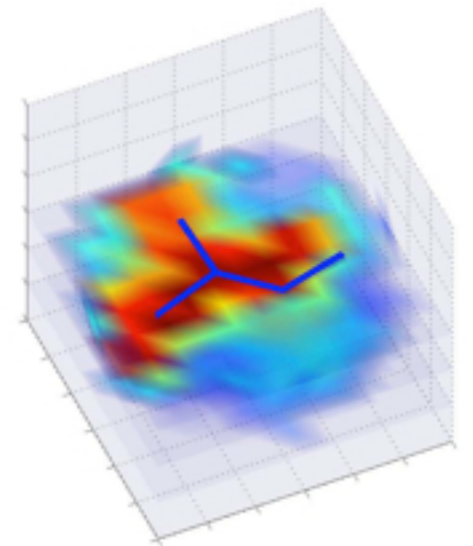
peptide  
good fit



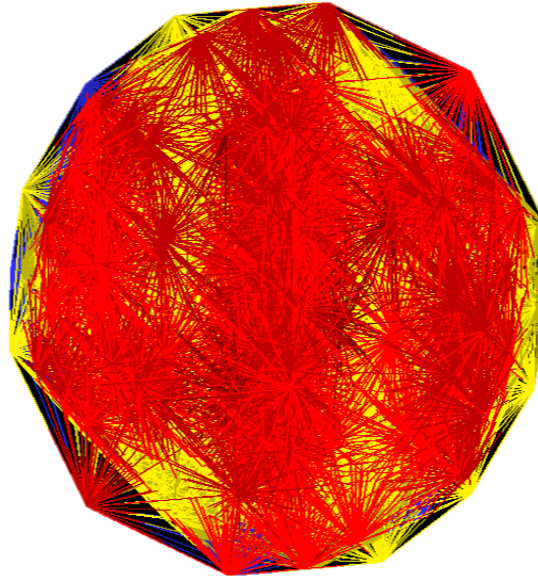
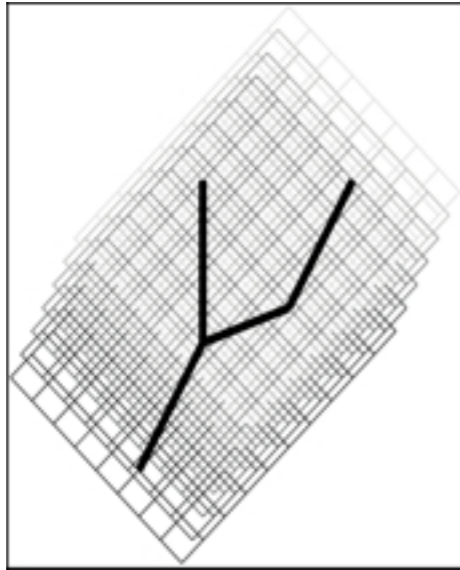
peptide  
but poor fit



non-peptide  
but good fit



# Autotracing: Recognising peptide units

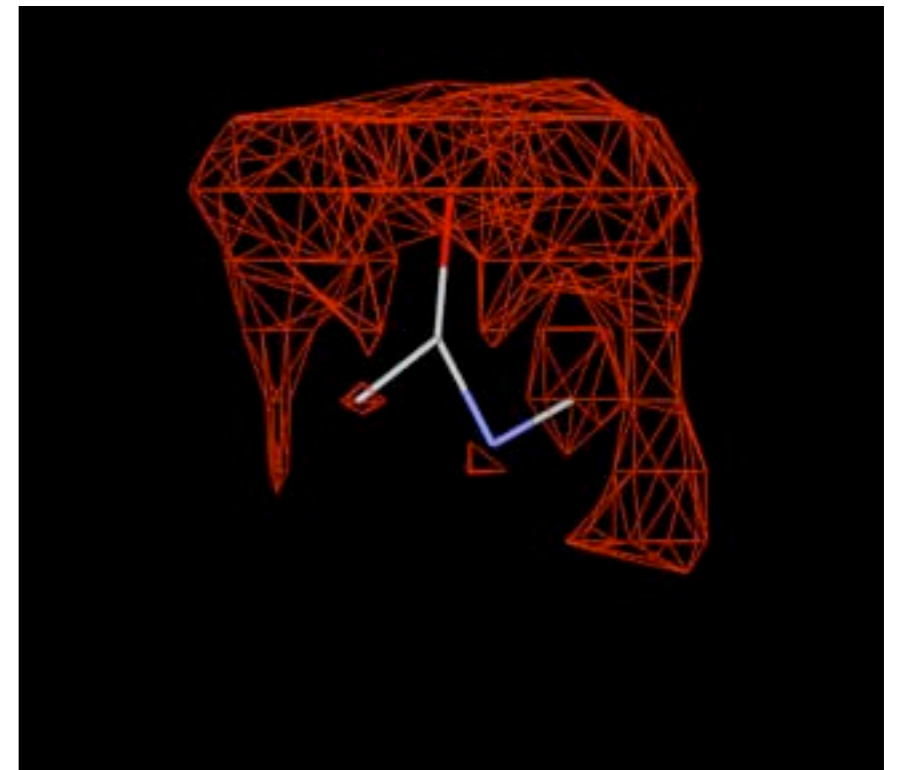
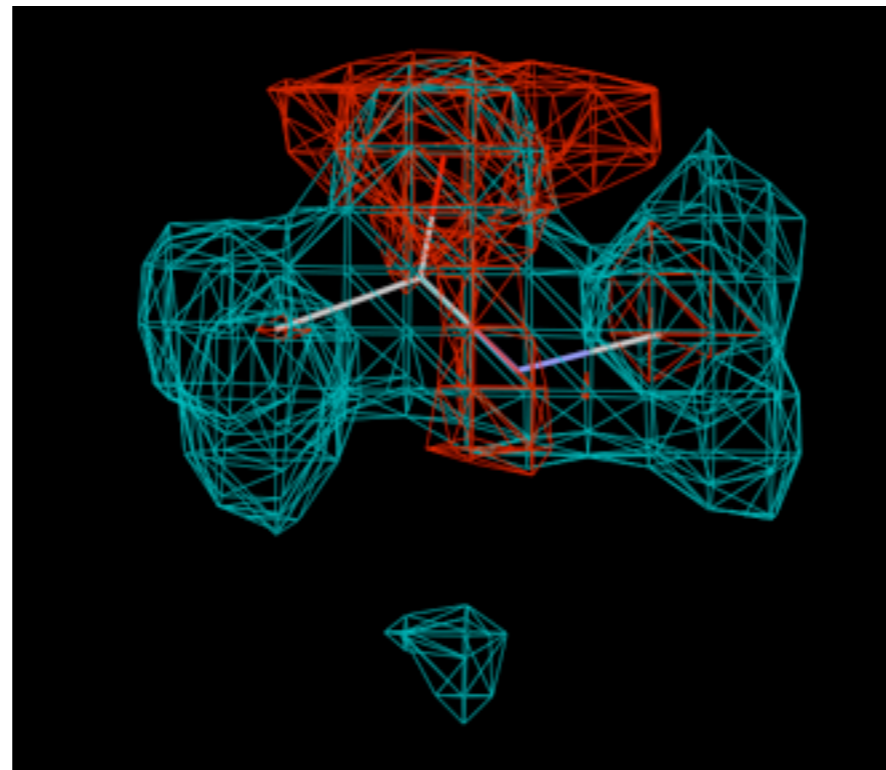


A peptide density shape is described by ~ **1,000** parameters, which are then **reduced to 1 parameter**

$$Y = \sum_{i=1,N} w_i (p_{obs,i} - p_{template,i})^2$$

1.9Å resolution

Target density ■  
Weight function ■  
Similarity searched for



# Dependence in the Data Resolution



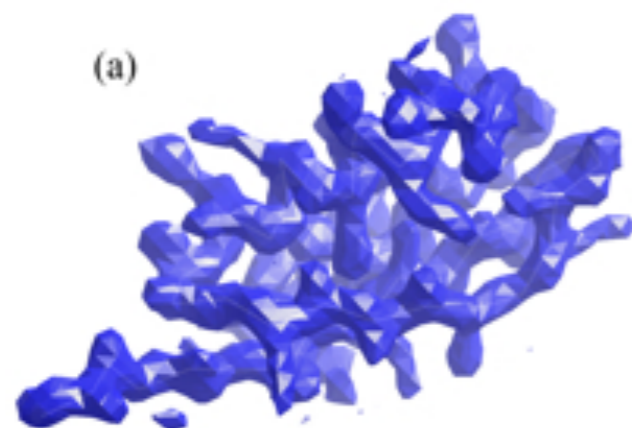
Results from the ARP/wARP 7.1 webservice (tracing performance)

<b>Resolution</b>	<b>Estimated fraction of automatically built protein structure (7/2010)</b>	<b>Average residues per chain of automatically built protein structure (7/2010)</b>
< 2.0Å	Over 90%	70
2.3Å	84%	47
2.6Å	80%	23
3.0Å	74%	13
3.5Å	65%	6

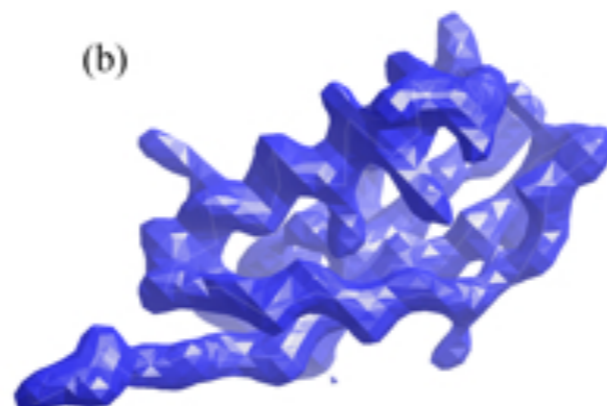
# Dependence in the Data Resolution



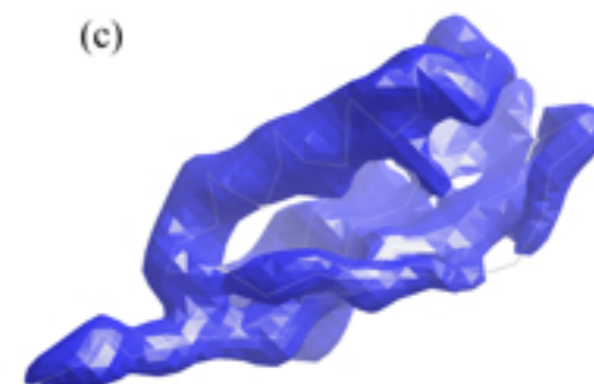
3 Å



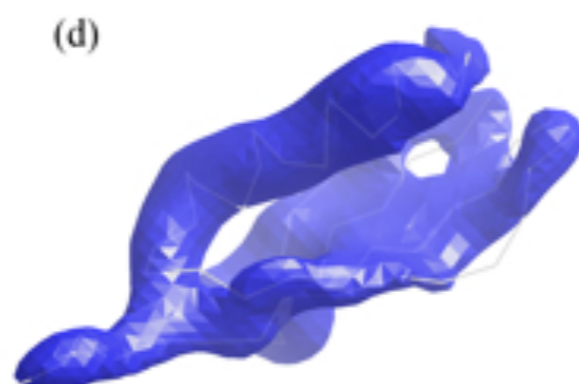
4 Å



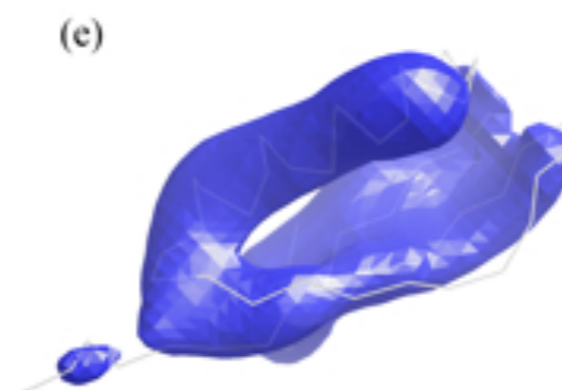
5 Å



6 Å

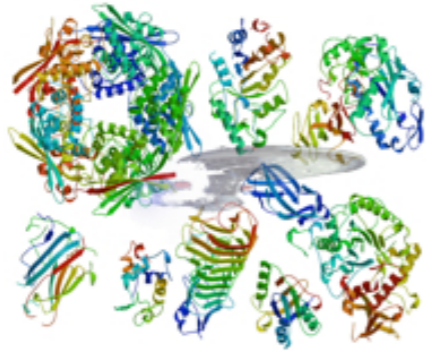
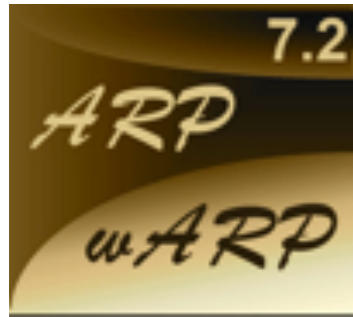


8 Å





# ARP/wARP: Main functions

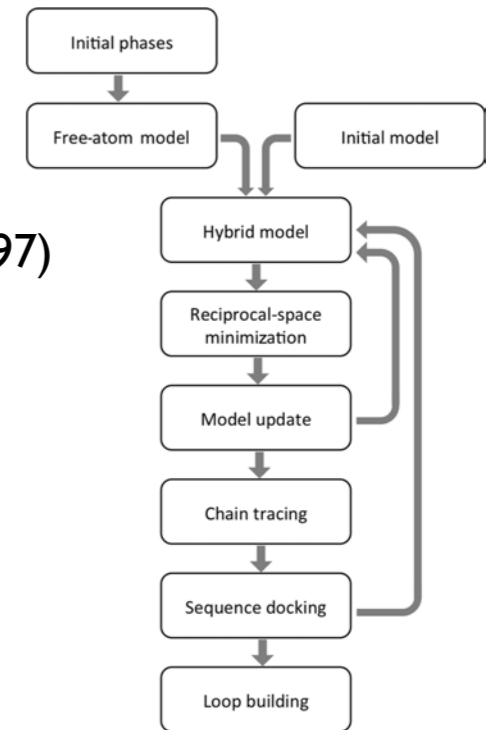


## Iterative protein-model building

ARP/wARP – Perrakis *et al.*, (1999)

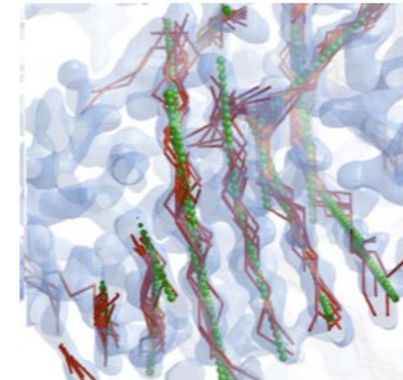
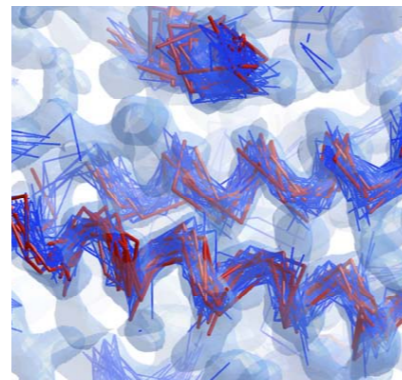
'Free atoms' – Agarwal and Isaacs (1985)

Refinement engine, REFMAC5 – Murshudov *et al.*, (1997)



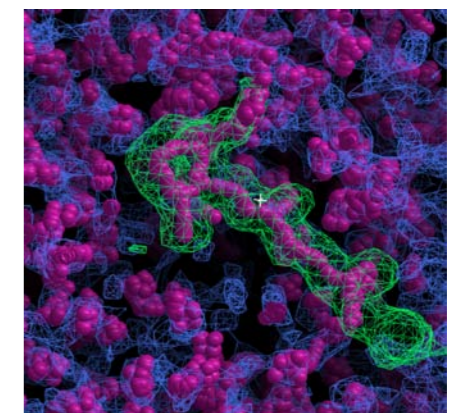
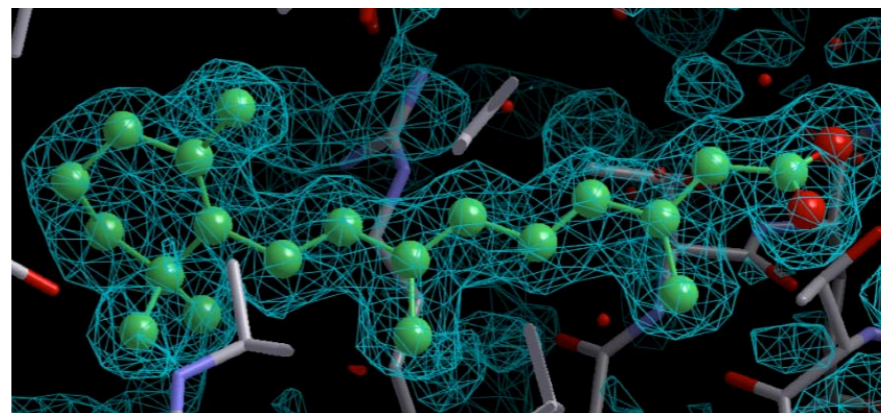
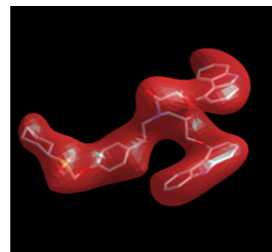
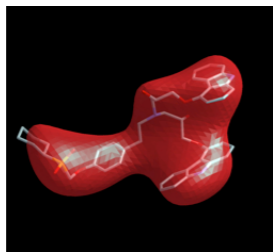
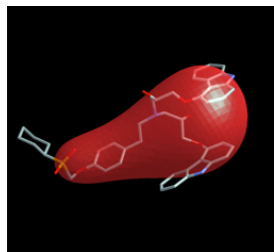
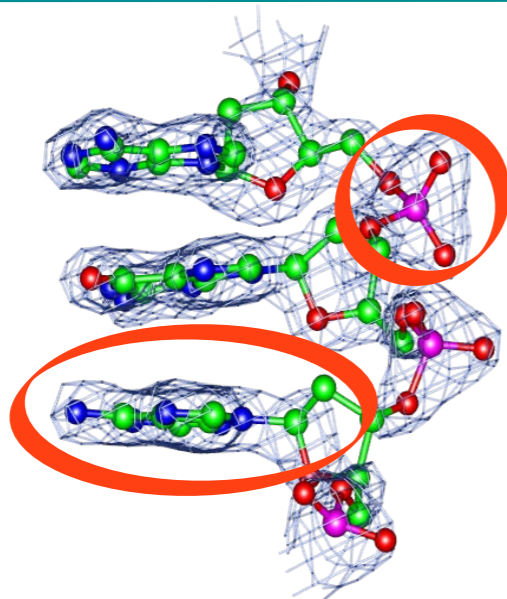
## Recognition of secondary structure:

fast approach for helices and sheets at low resolution ( $< 4.5 \text{ \AA}$ )



## Building polynucleotides:

planar bases and phosphate groups for tracing



## Ligand building methods:

Cocktail screening, ligand identification using shape descriptors, conformational fit, sparse grids, fine skeletons

# What ARP/wARP 7.2 can do for you



Protein  
model building

Model completion

Classical protein chain  
tracing

Loop building

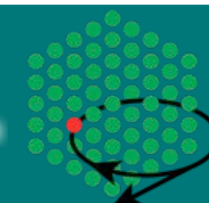
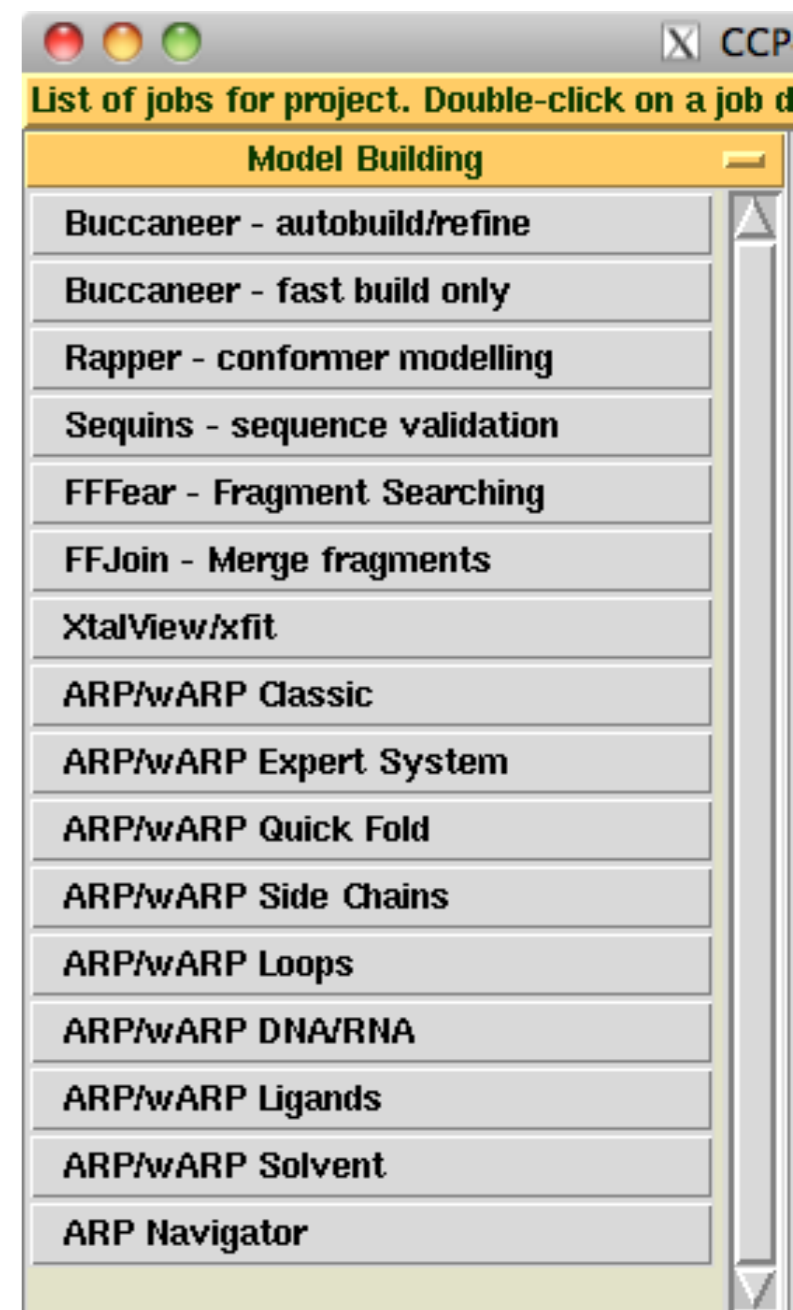
Helices & Strands

Nucleotides

Ligands

Graphics Front  
End!!!

Model improvement



# ARP/wARP Model Building

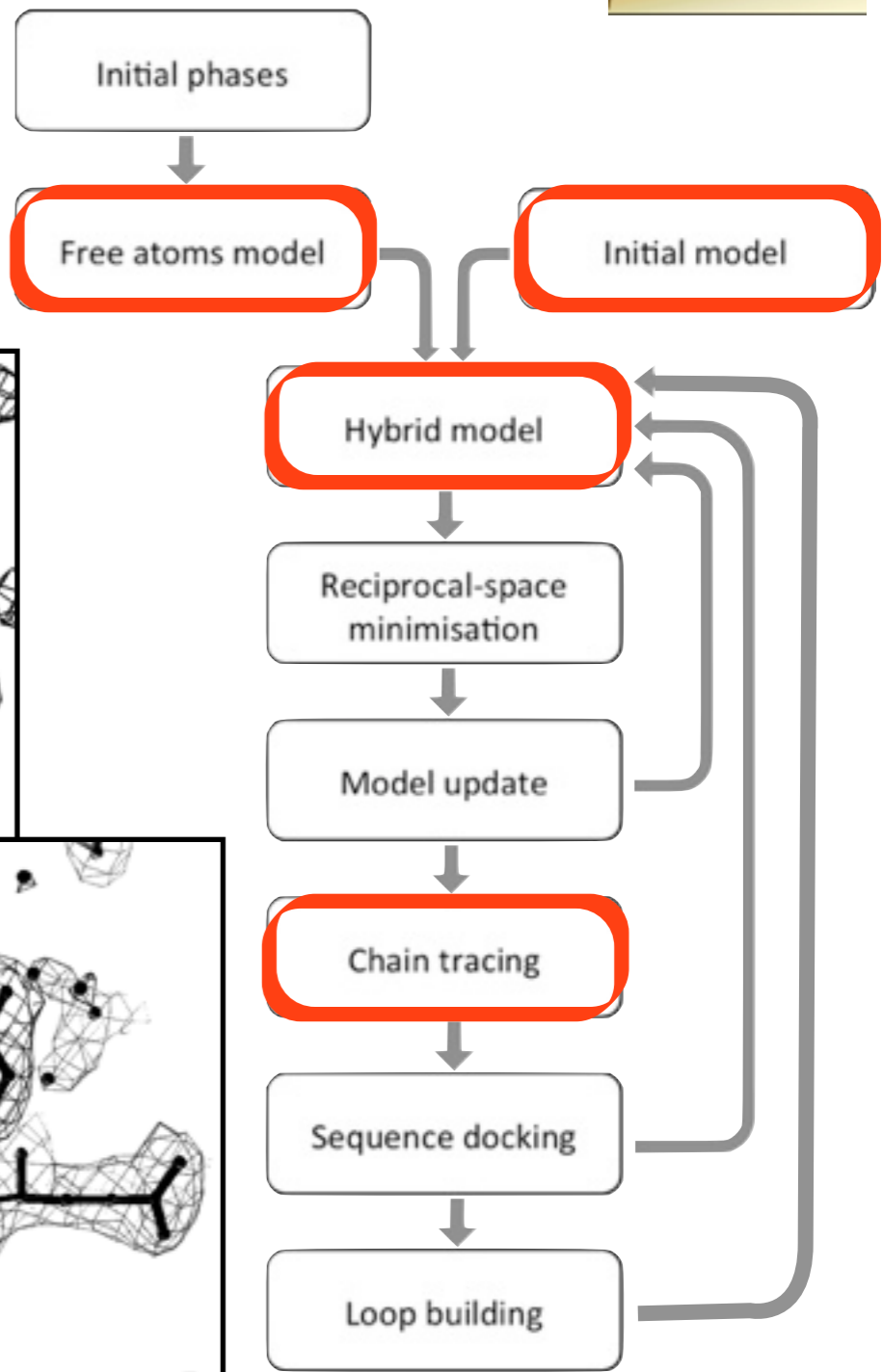
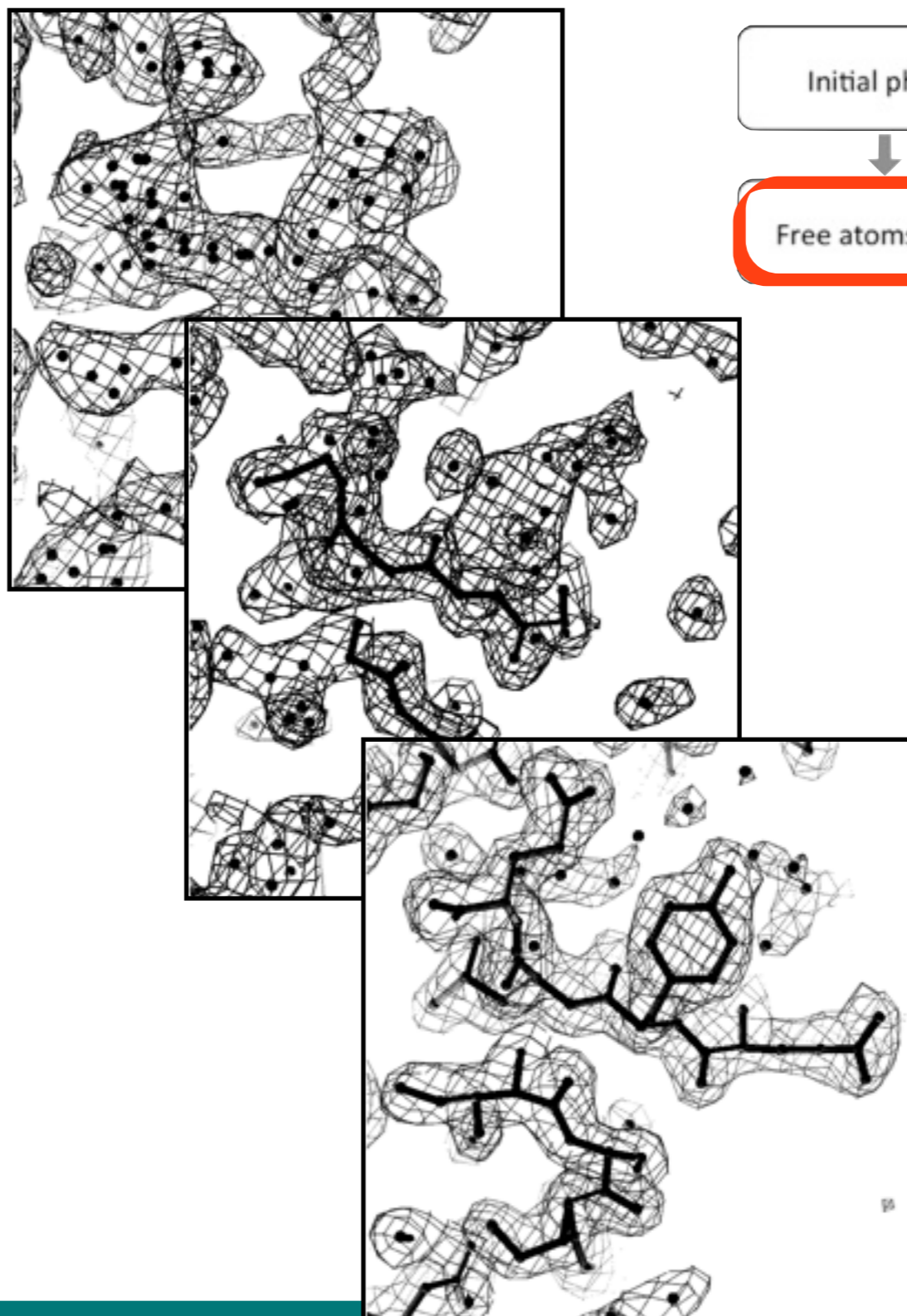


Free Atoms

Pattern Recognition

Hybrid Models

Iterations



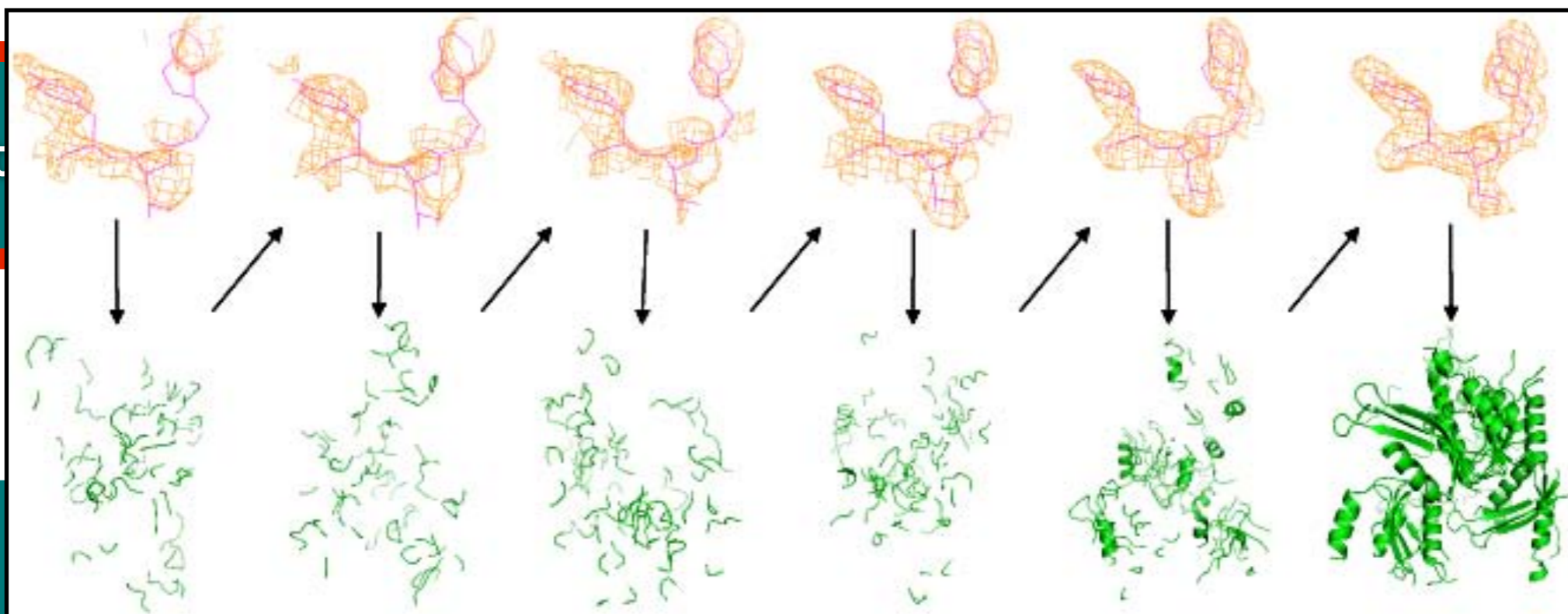
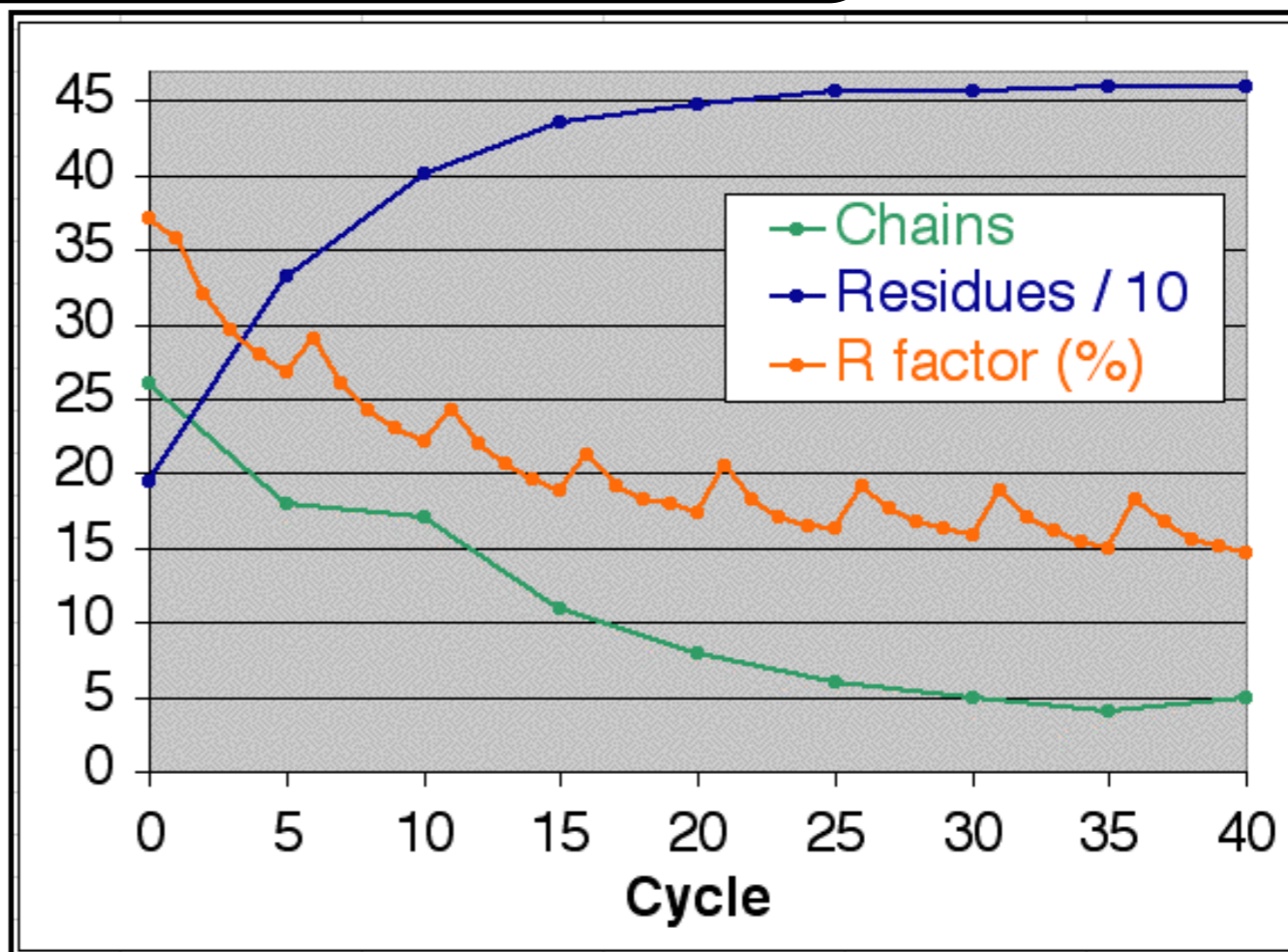
# ARP/wARP Model Building

Free Atoms

Pattern Recognition

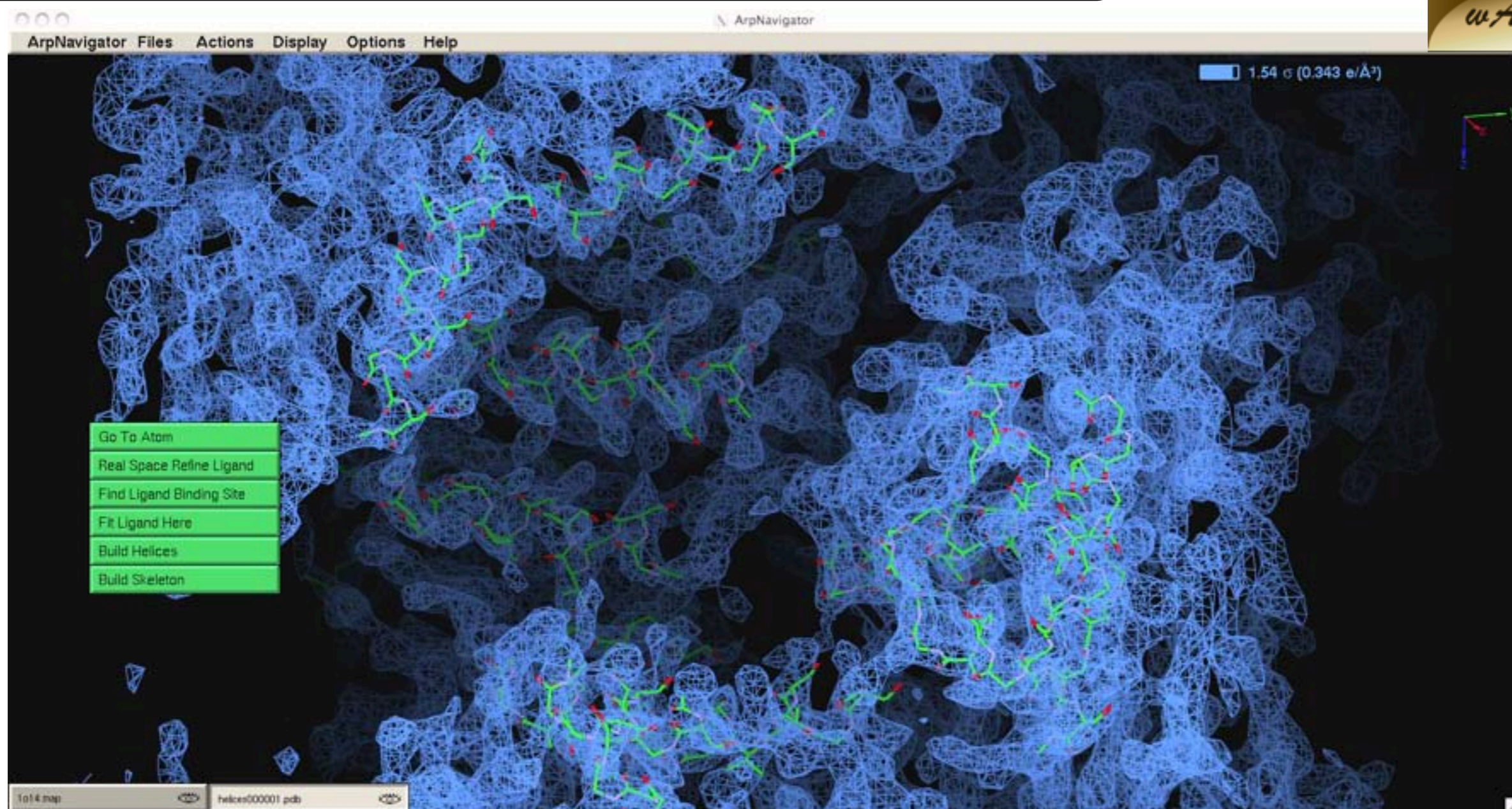
Hybrid Models

It



# Modelling Secondary Structure

7.2  
ARP  
wARP

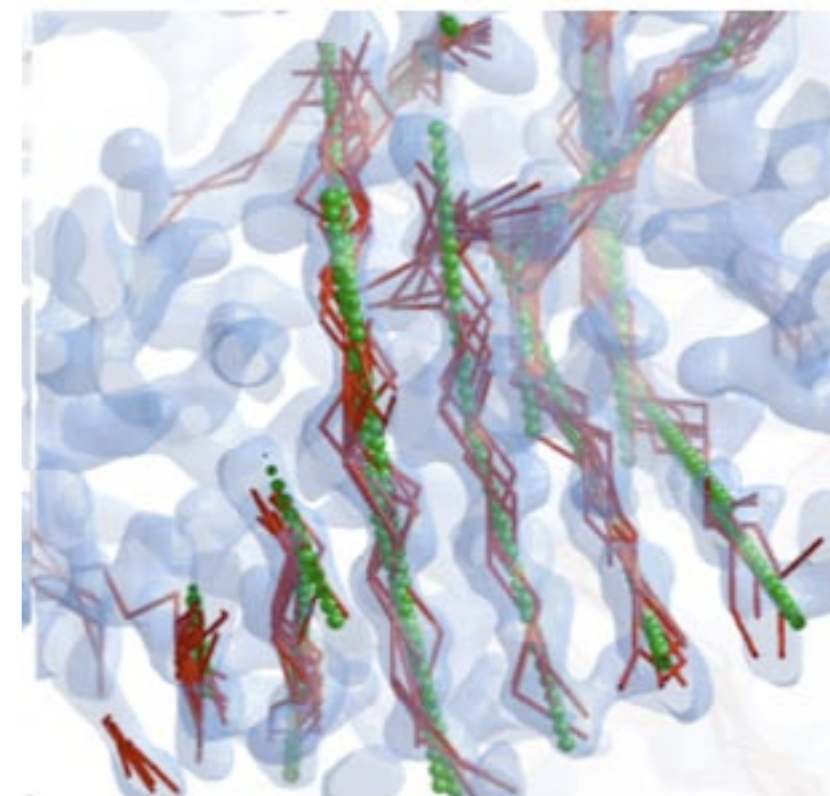
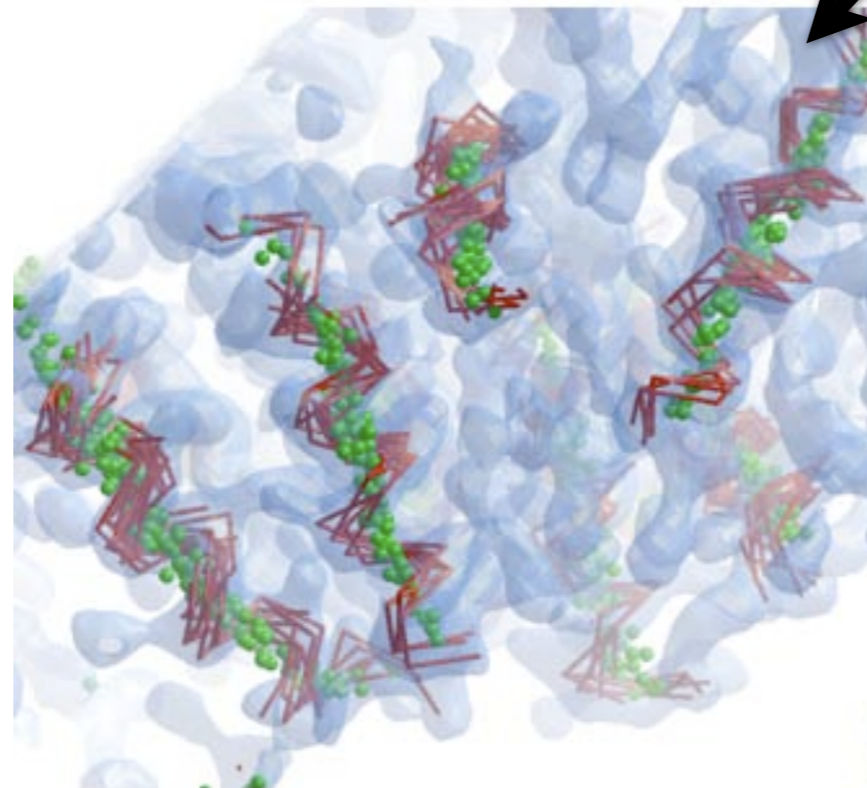
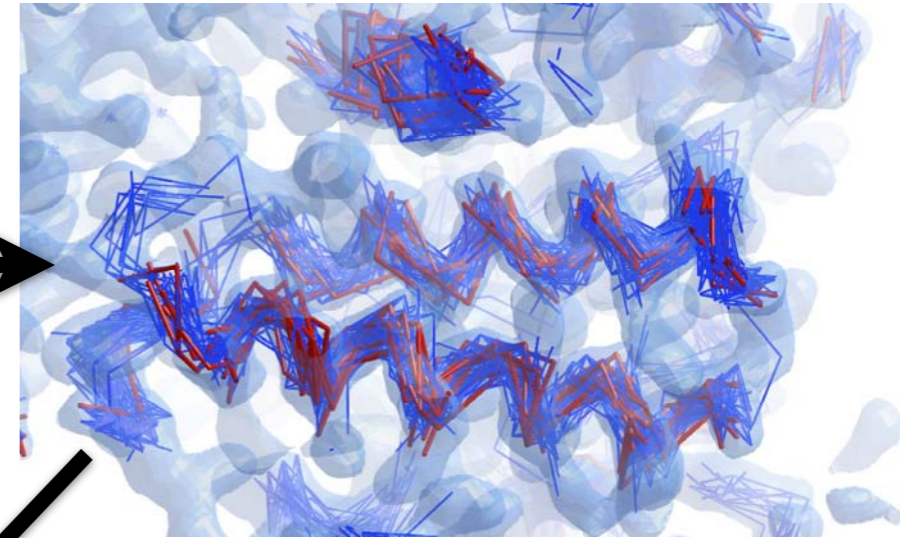
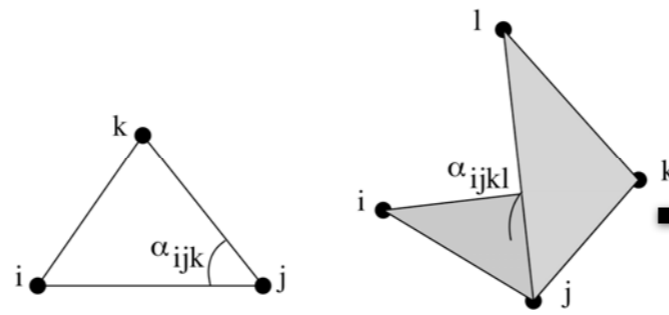


## Helices & Strands

Helices for a 600-residue (3.2 Å) protein can be built in a few seconds on a modern imac

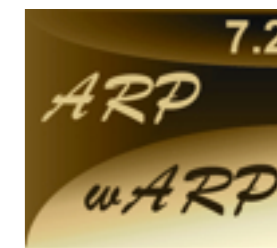
# Modelling Secondary Structure

- Short helix/strand fragments (3 to 5 C $\alpha$  candidates) are built.
- Longer traces are formed or which the best are kept (in red)
- Traces are clustered
- Assemblies are averaged

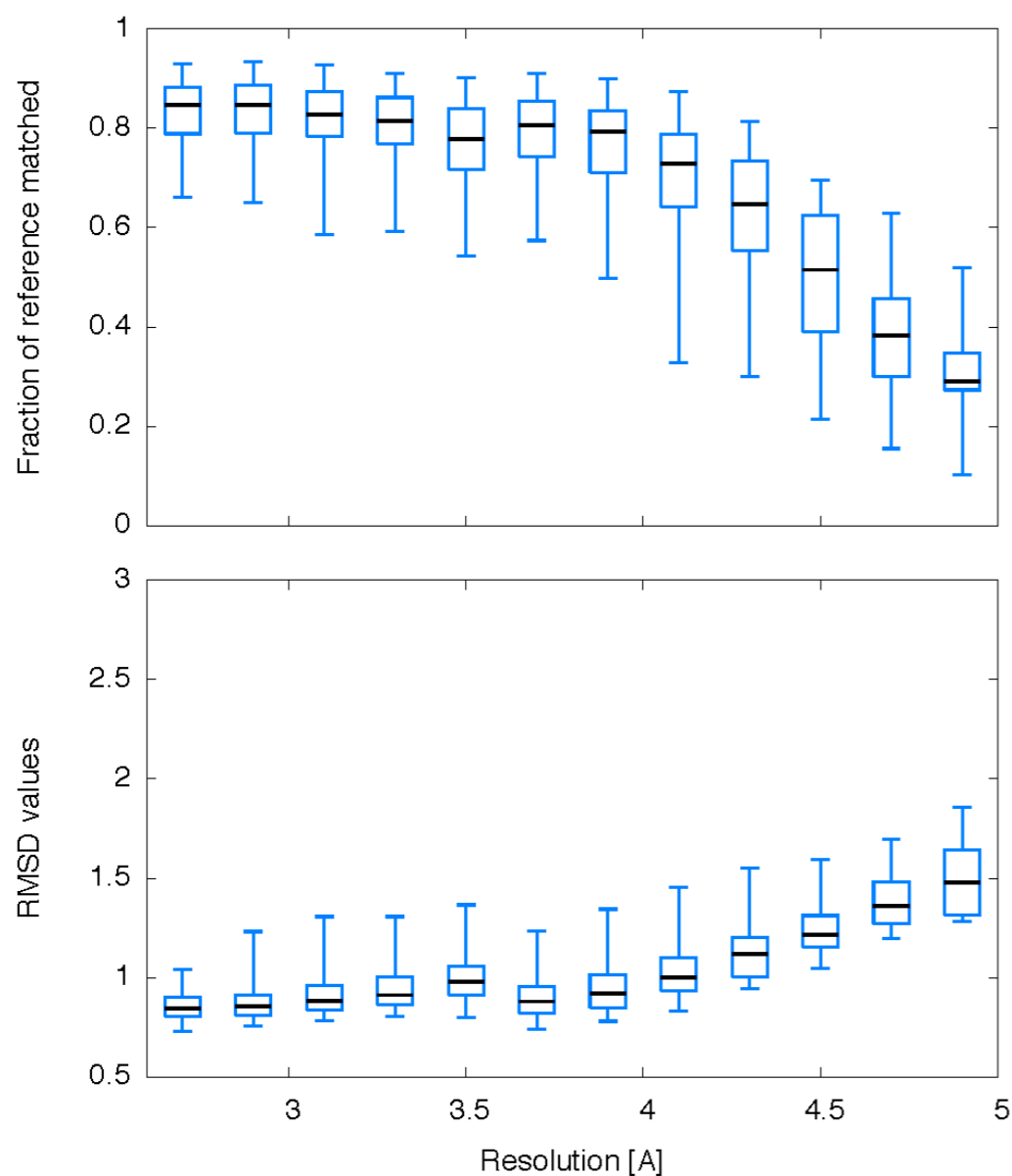


Helices & Strands

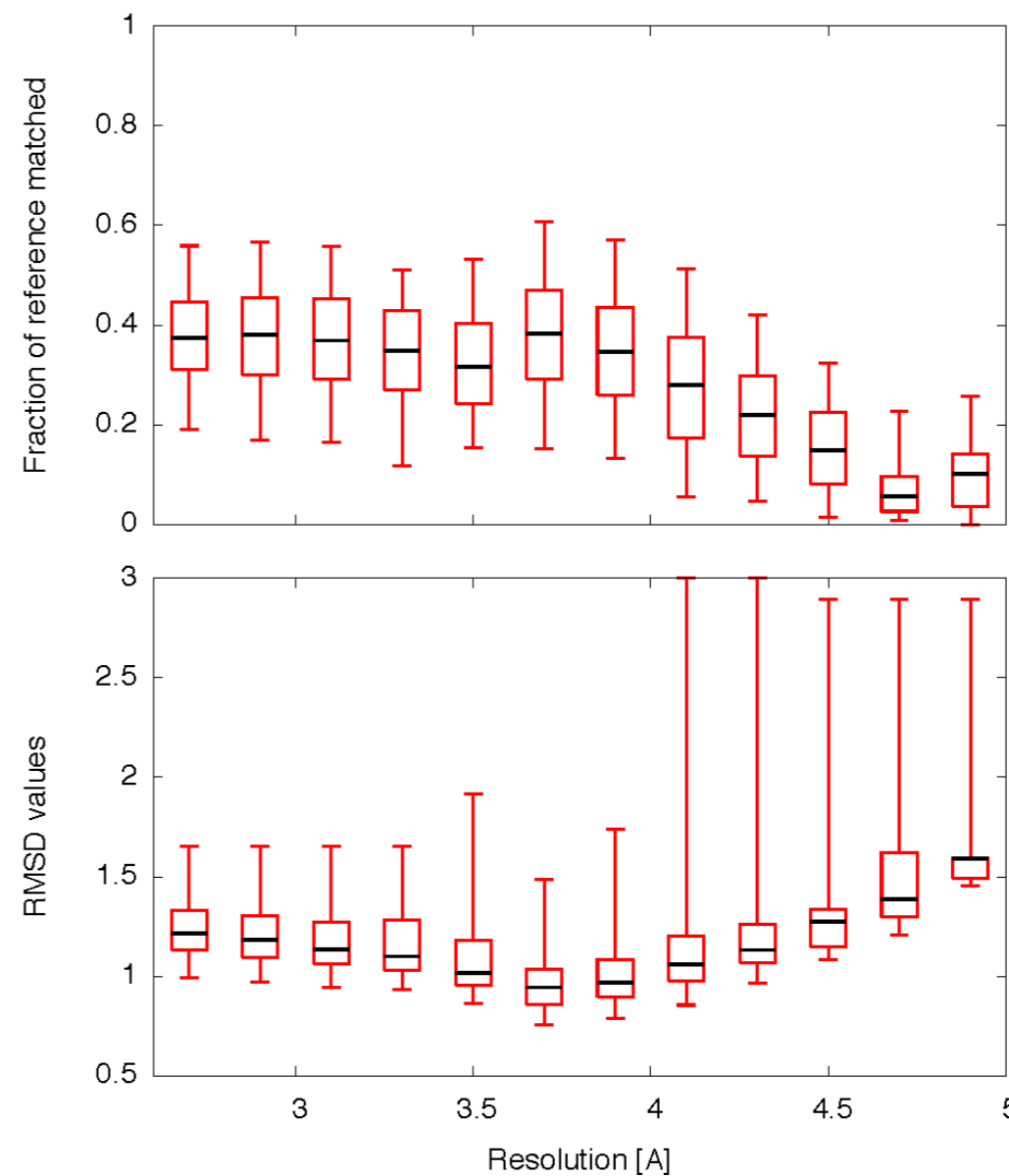
# Modelling Secondary Structure



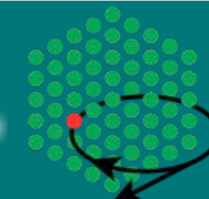
## Helices

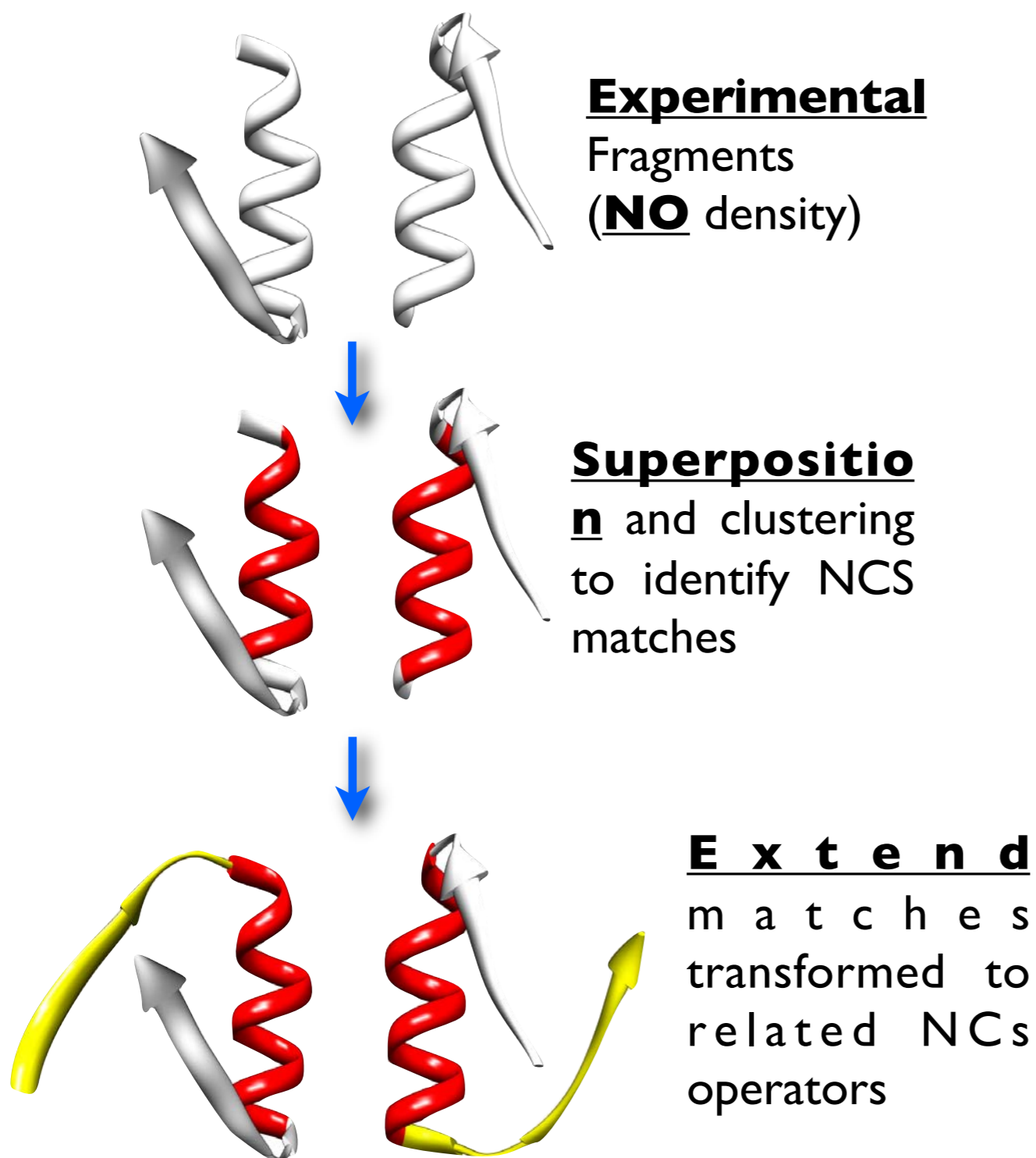


## Strands



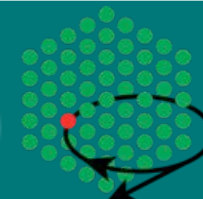
## Helices & Strands





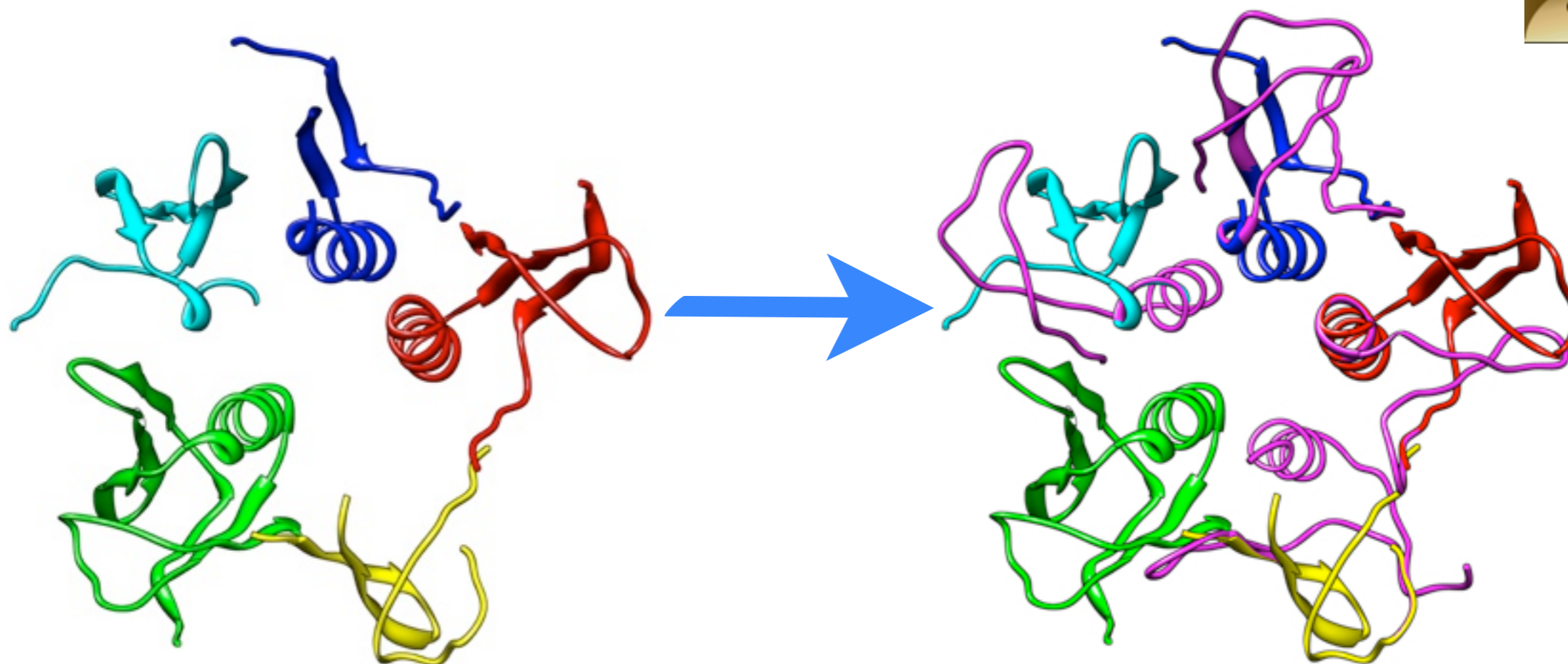
- uses chemically assigned fragments from hybrid model for extensions
- extensions are fed back as candidates for protein chains
- applied between model update and chain tracing
- model building protocol decides on validity of extensions
  - stereo-chemical
  - with respect to density

Model improvement



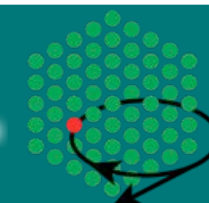


# PNS\_extender: Automatic NCS detection



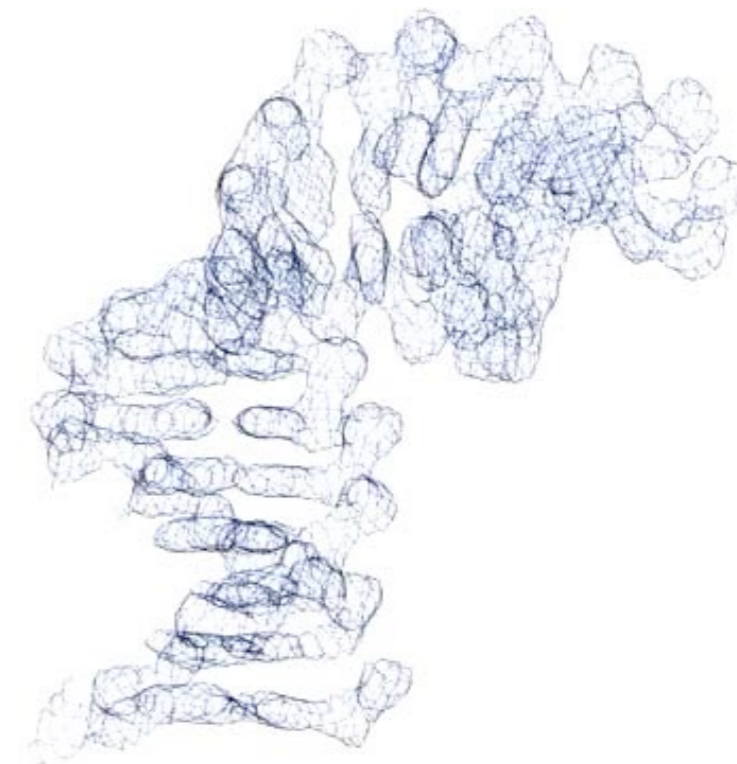
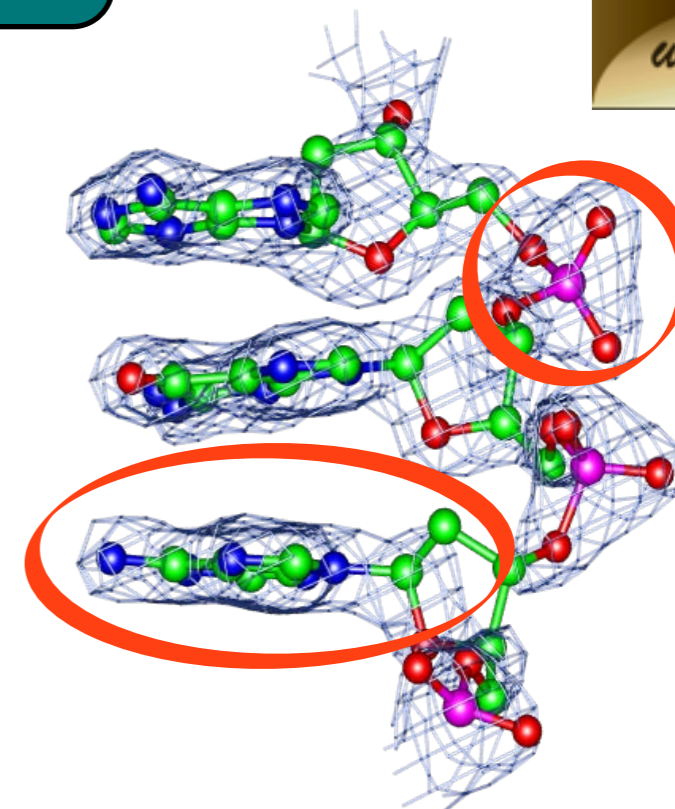
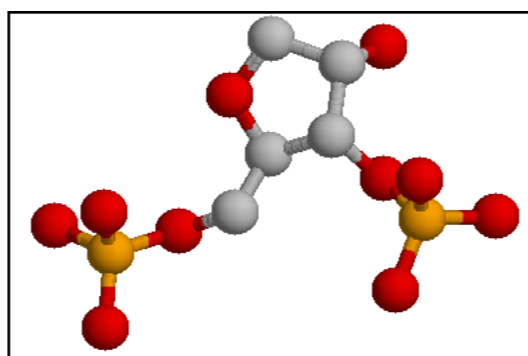
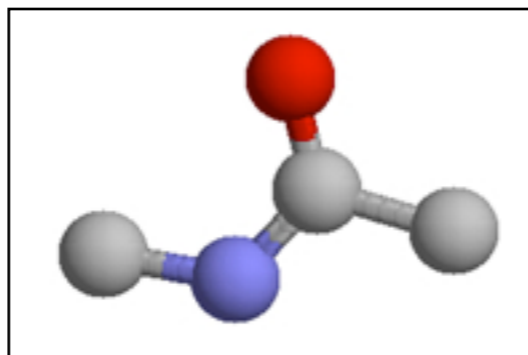
- 12 protein test structures, resolution 2.3 to 3.3 Å, NCS order 2 - 5
  - observed improvement in all cases
  - average model completeness increased by 3%
  - average number of residues per chain increased by up to 25%

Model improvement

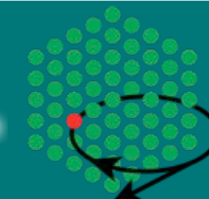
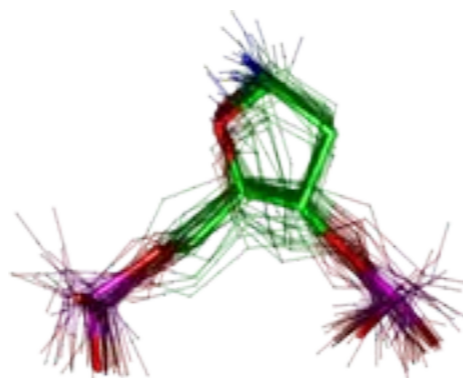


# Tracing RNA/DNA Chains

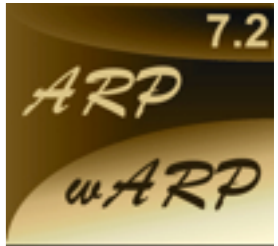
7.2  
ARP  
wARP



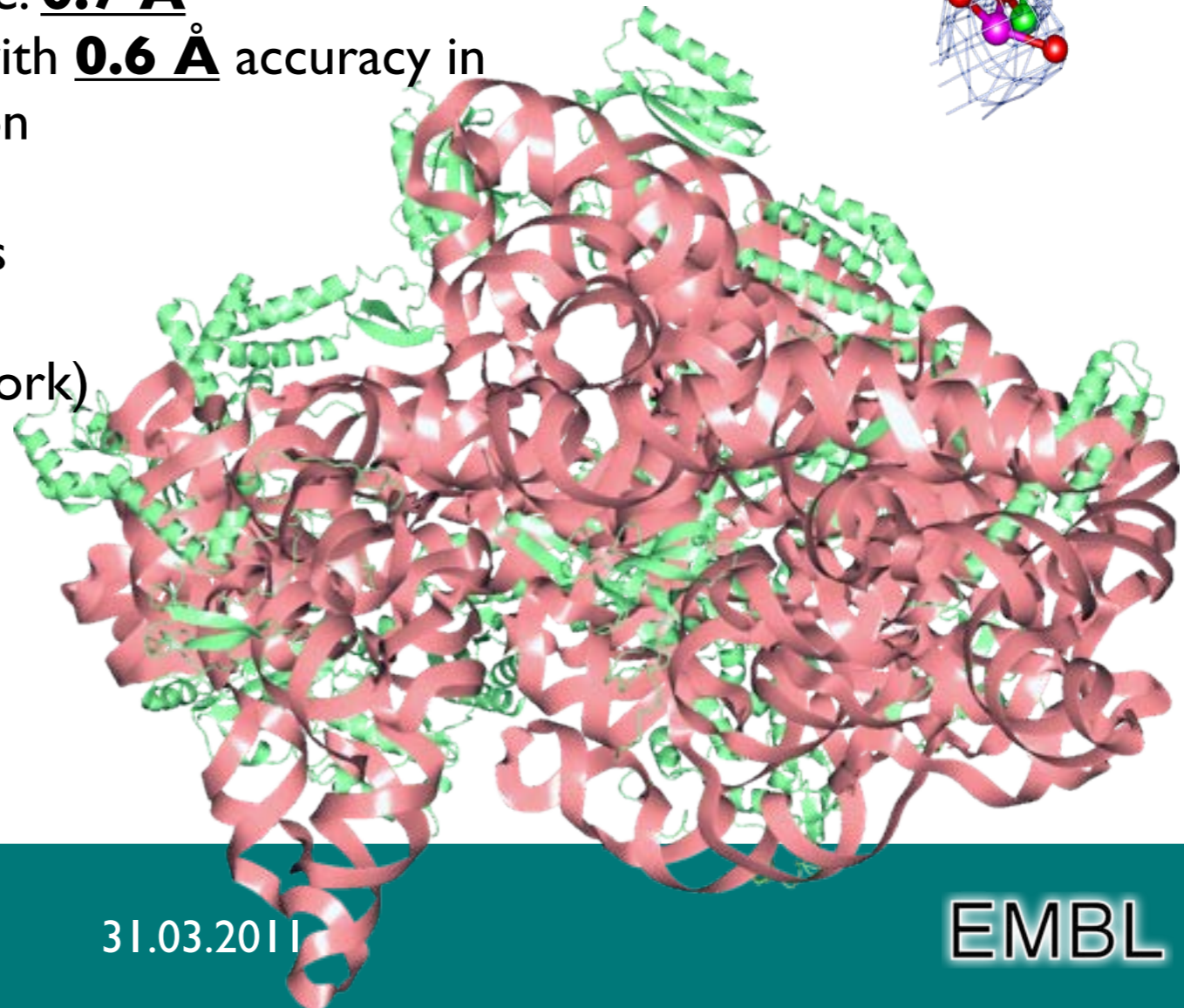
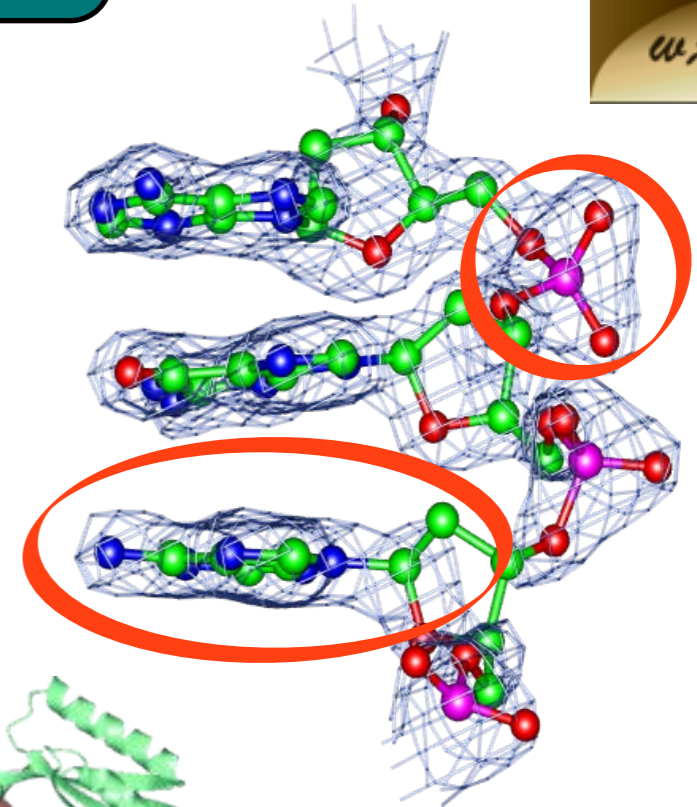
Nucleotides



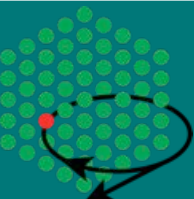
# Tracing RNA/DNA Chains

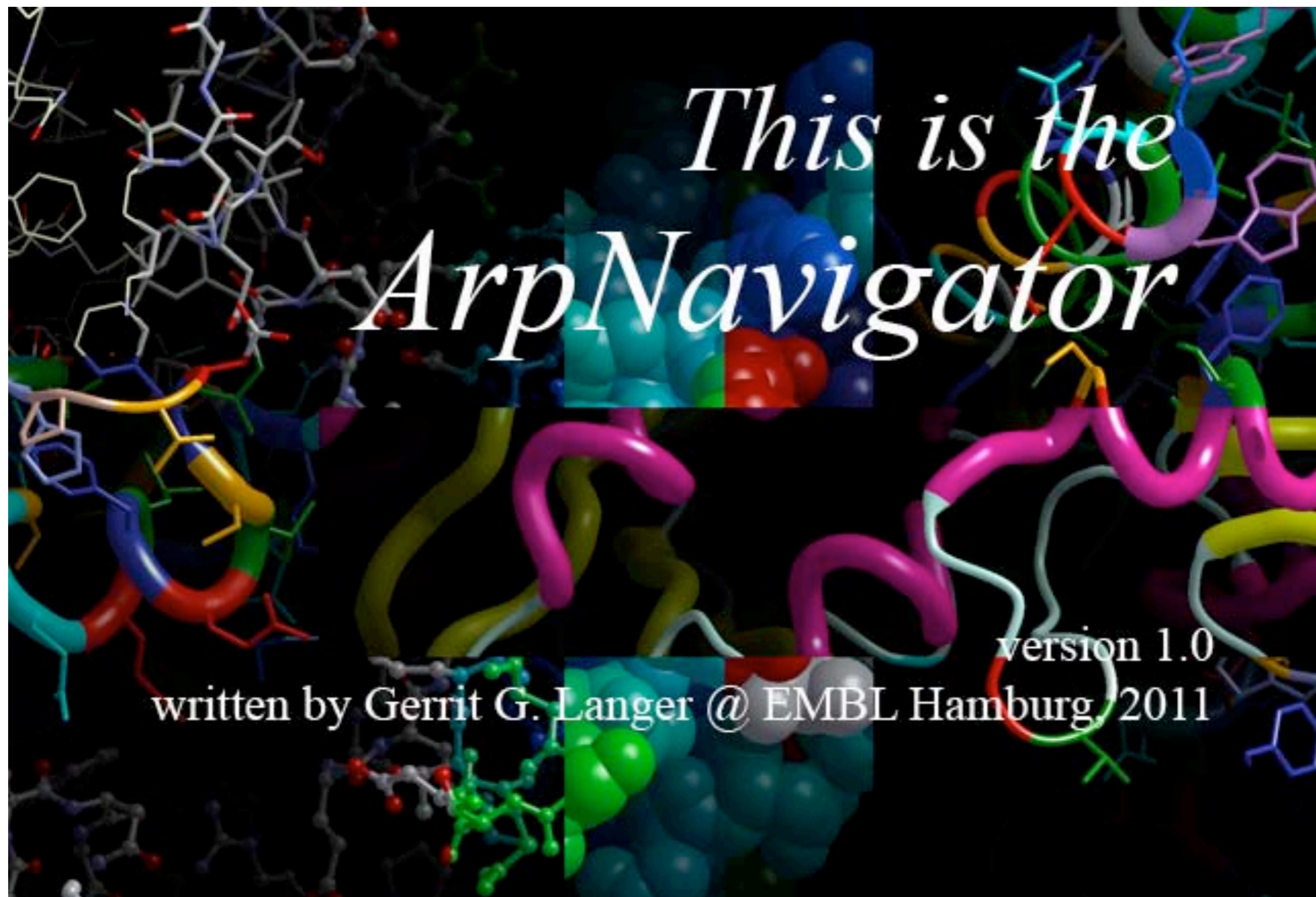


- The 30S ribosomal subunit
  - Resolution: 3.05 Å
  - Experimental phases (MCC 0.73)
- Auto-building with ARP/wARP
  - modelled 1,302 out of 1,513 nucleotides (**86%**)
  - backbone r.m.s.d. to reference: **0.7 Å**
  - Located 1,121 nucleobases with **0.6 Å** accuracy in location and 12° in orientation
- Backbone fragmented in 75 chains
- Built in **around 6h**
- (cf. several **months** of manual work)



Nucleotides

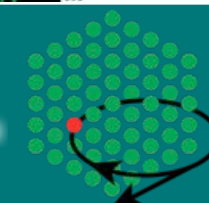
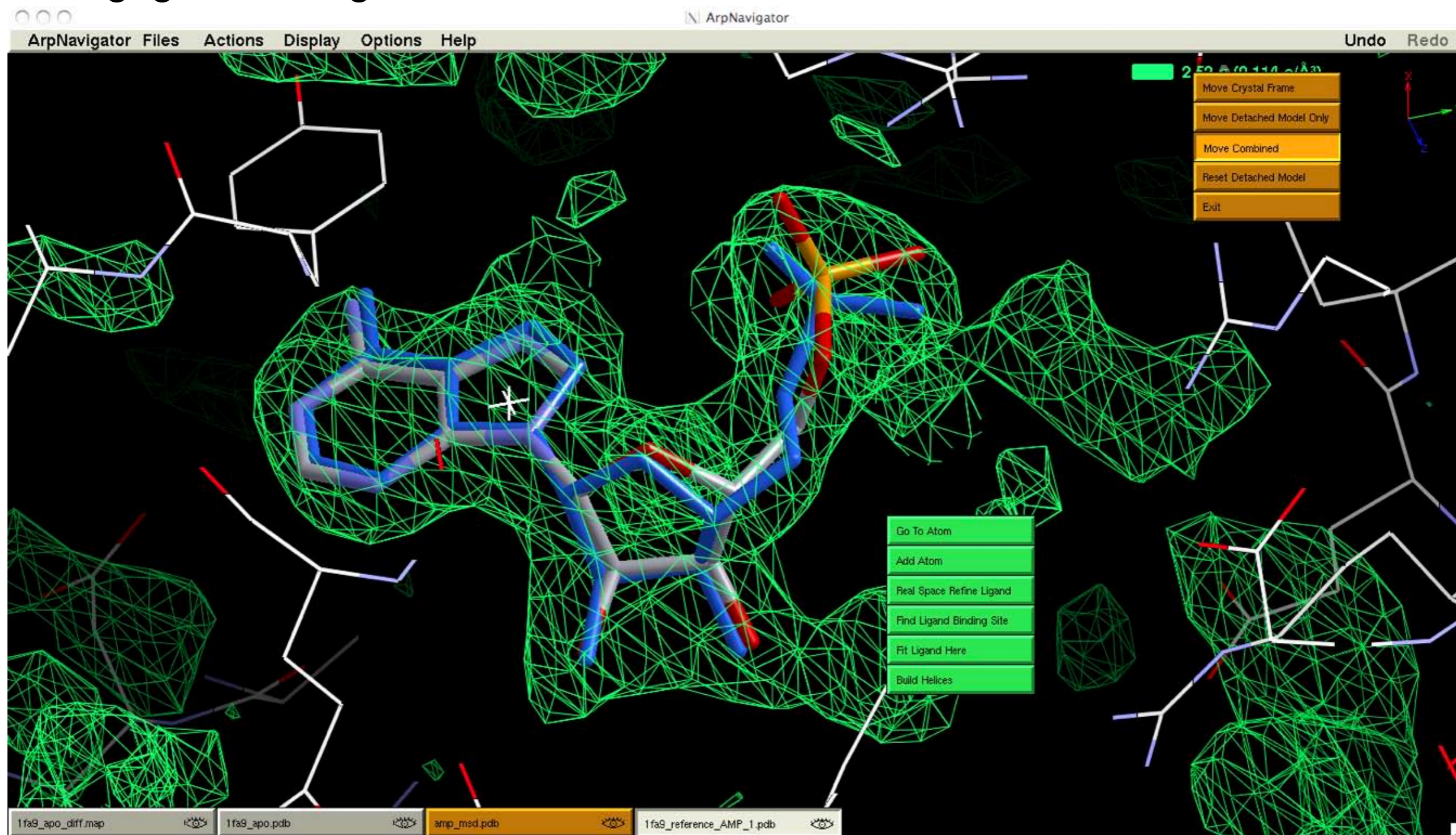




# ARP/wARP Graphics Front End



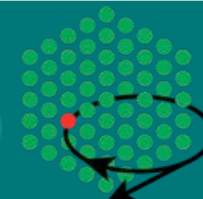
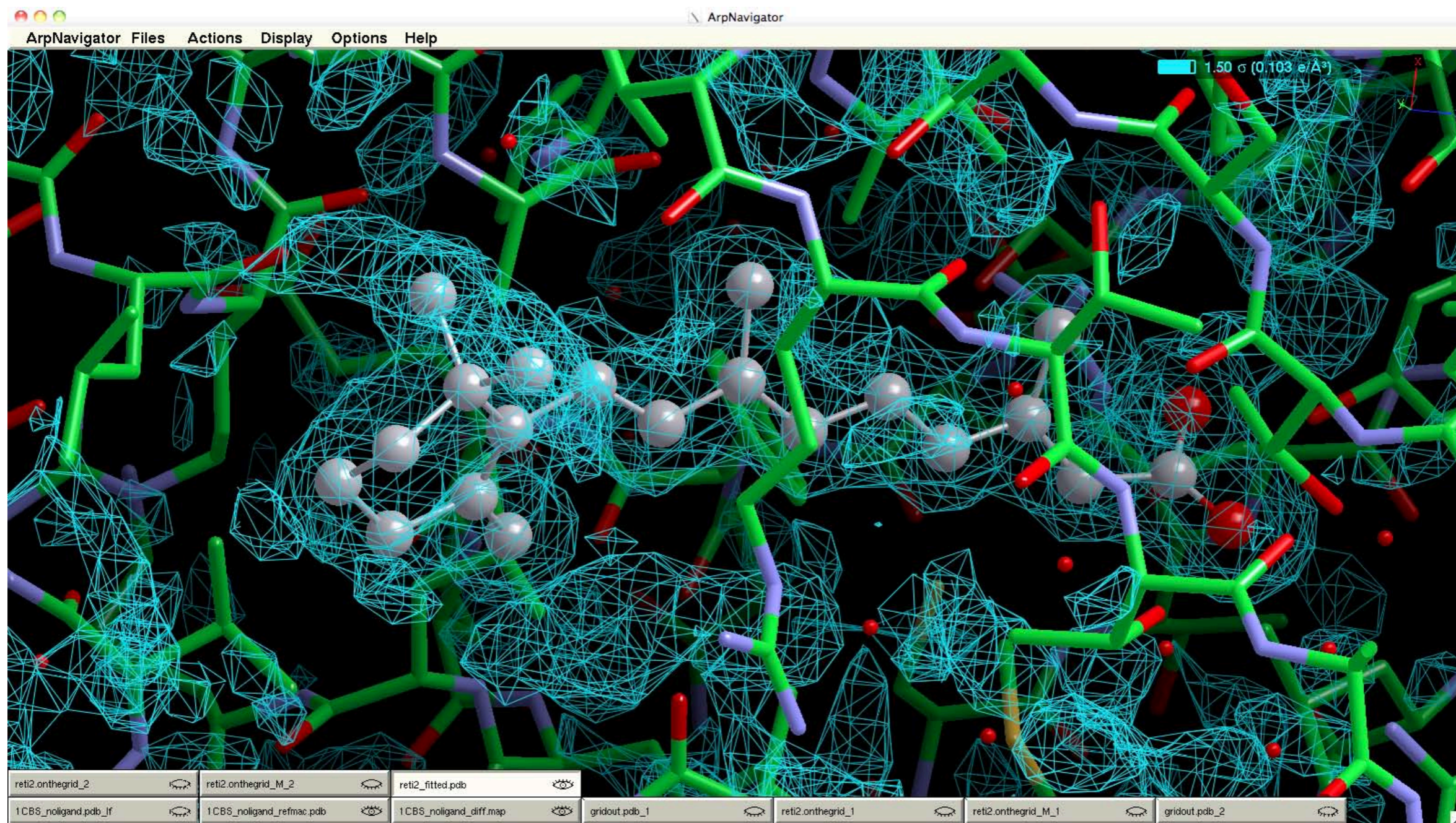
This should give a more intuitive access to functions such as:  
Building ligands, tracing helices/strands and model solvent.



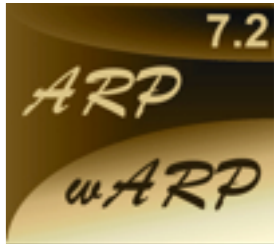
# ARP/wARP Graphics Front End



This should give a more intuitive access to functions such as:  
Building ligands, tracing helices/strands and model solvent.



# Computational Services



Submit a remote job at the Hamburg Cluster

Submit the job for remote execution at the Hamburg cluster

Your Email address

Job data

```
Cycle 17: After reffmac, R = 0.202 (Rfree = 0.000).  
Found 118 (155 requested) and removed 9 (77 requested) atoms.  
Cycle 18: After reffmac, R = 0.188 (Rfree = 0.000).  
Found 68 (133 requested) and removed 10 (66 requested) atoms.  
Cycle 19: After reffmac, R = 0.179 (Rfree = 0.000).  
Found 57 (121 requested) and removed 35 (60 requested) atoms.  
Writing map files covering molecule ...  
Normal termination of warpNtrace
```

[Download Results](#)

After you download all the required files click below to indicate that you are done

[DONE](#)

If the job fails or you have any other comments click below to send comments

[COMMENT](#)


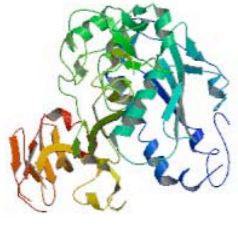


Image created using molscript Image created using PYMOL

ARP/wARP web service

7.1

DISCLAIMER | ARP/wARP LICENCE | CCP4 LICENCE | CITATIONS

**ARP/wARP web service**

Welcome to ARP/wARP web service!

Please note that the use of this service requires you to consent to the conditions and license agreements stated in the disclaimer. Also, please do not forget to cite the citations above.

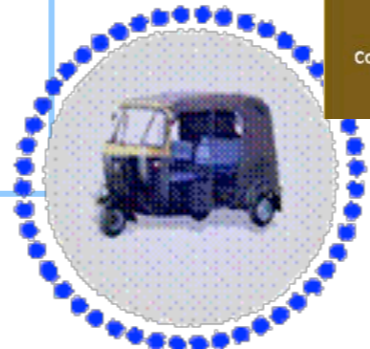
I have viewed the disclaimer, have read the ARP/wARP license, hold the current CCP4 license, agree to the conditions stated therein and wish to proceed with the remote service.

Opening Page

Step 1

Step 2

Confirmation Page



[www.embl-hamburg.de/Auto-Rickshaw/](http://www.embl-hamburg.de/Auto-Rickshaw/)

βαλβεσ



Sequence Target:

Instead of entering a Sequence Target file you can paste your FASTA sequence below:  
(Note that a comment line beginning with a '>' character must precede each sequence)

```
>ADKELFLVV DDFSTMRRIV RNLLKELGFN NVEEAEDGVD ALNKLQAGGY GFVISDWNMP NMDGLELLKT  
IRADGAMSAL PVLMTAEAK KENIIAAAQA  
GASGYVVKPF TAATLEEKLN KIFEKLG  
>MGDSILSQAE IDALLN
```

Check Full Spacegroup:

Run ARP/wARP (on the Balbes solution):  Dissemination Level:

(after clicking submit, PLEASE WAIT for your files to upload - this may take some time)

# Downloading ARP/wARP



Browser address bar: <http://www.embl-hamburg.de/ARP/>

**Software for Macromolecular Crystallography**  
**ARP/wARP Copyright 1998**  
**by European Molecular Biology Laboratory**  
Authors: [Victor S. Lamzin](#) and [Anastassis Perrakis](#)

ARP/wARP is a software suite for improvement and objective interpretation of crystallographic electron density maps and automatic construction and refinement of macromolecular models

**ARP/wARP 7.1 download and license**  
[Download User Guide \(pdf\)](#)  
[View User Guide \(html\)](#)  
[Frequently Asked Questions and Software Patches](#)

**ARP/wARP protein model building via the web**

Resolution	Estimated fraction of automatically built protein structure (last updated on Jun 4, 2010)
< 2.0 Å	Over 90%
2.3 Å	84%
2.6 Å	80%
3.0 Å	74%
3.5 Å	65%

Buttons:  
Send E-mail to the ARP/wARP list (subscribers only)  
Subscribe to ARP/wARP list  
Send E-mail to the authors  
Unsubscribe from ARP/wARP list



# People & Acknowledgements

## Developers

**EMBL Hamburg:** Ciaran Carolan, Saul Hazledine, Philipp Heuser, Victor Lamzin, Tim Wiegels

**NKI Amsterdam:** Krista Joosten, Tassos Perrakis

## Collaborators

**EMBL Hamburg:** Santosh Panjekar

**York University:** Garib Murshudov's group

**Daresbury Laboratory:** CCP4 team

**Former ARP/wARP team members**

## Funding



Bundesministerium  
für Bildung  
und Forschung



...if you are ever in town...

