

Interpretation of
3D electron density
and model building
with ARP/wARP

What is an Electron Density Map?



- 1. A result of an X-ray diffraction experiment**
- 2. A result of an Electron Microscopy experiment**
- 3. A molecular energy landscape**
- 4. A distribution of electrons**
- 5. A smeared representation of my protein structure**
- 6. A probability density function**
- 7. A nice picture that my boss told me to get**

Fundamental Problems in Map Interpretation

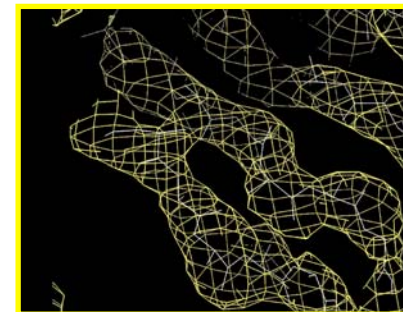
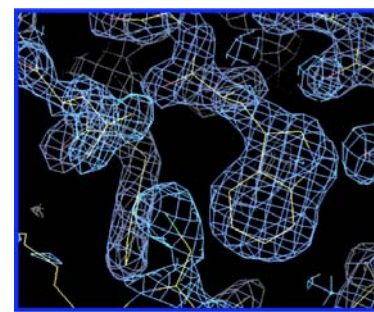
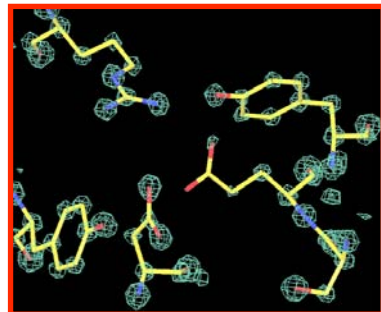


1.0 Å

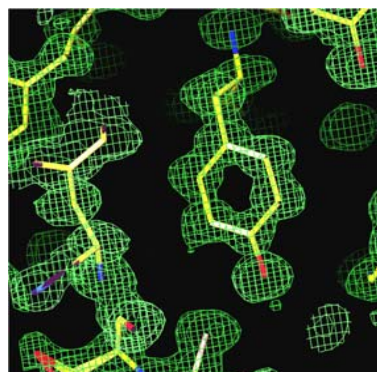
2.0 Å

3.0 Å

Limited resolution



Errors in phases
(in general the amount
and type of noise
present)

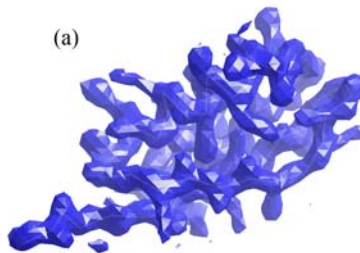


(Limited) Resolution of the X-ray Data

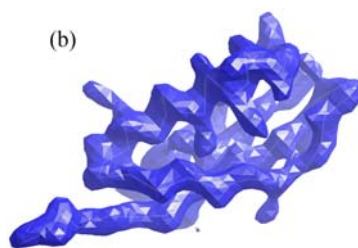


Calculated map of protein G

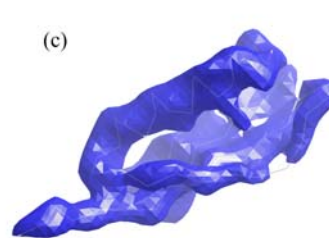
3 Å



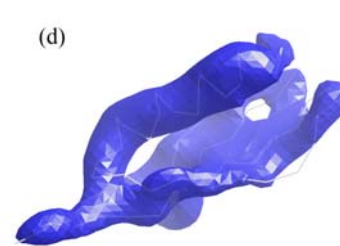
4 Å



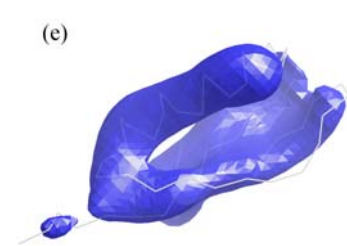
5 Å



6 Å

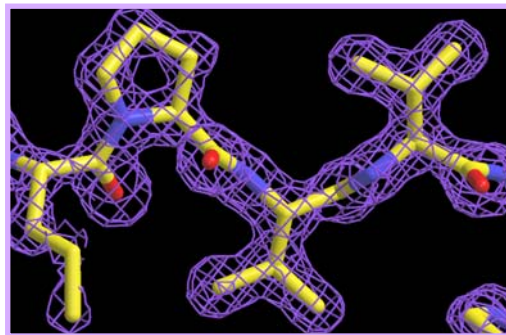
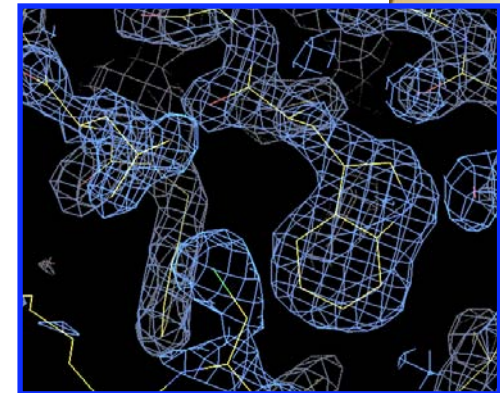
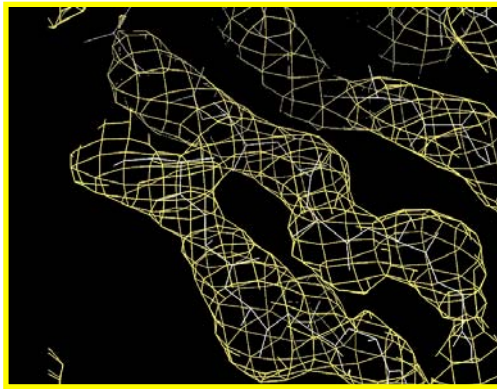
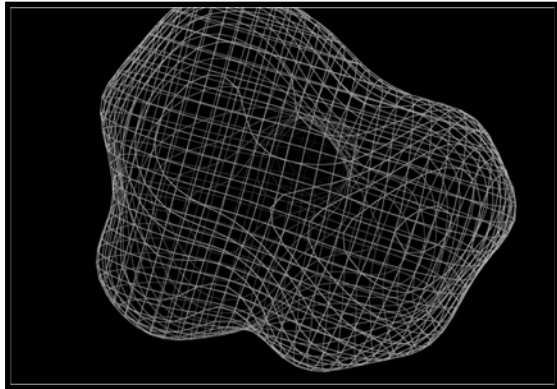


8 Å

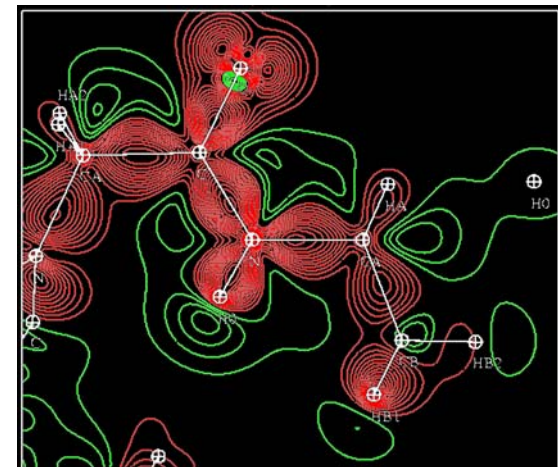
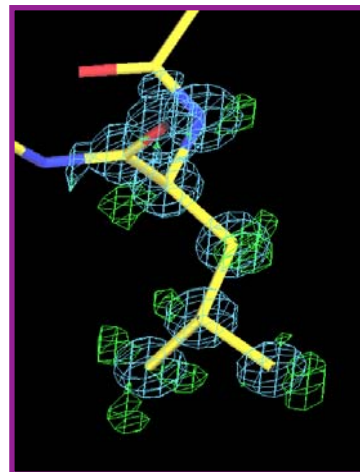
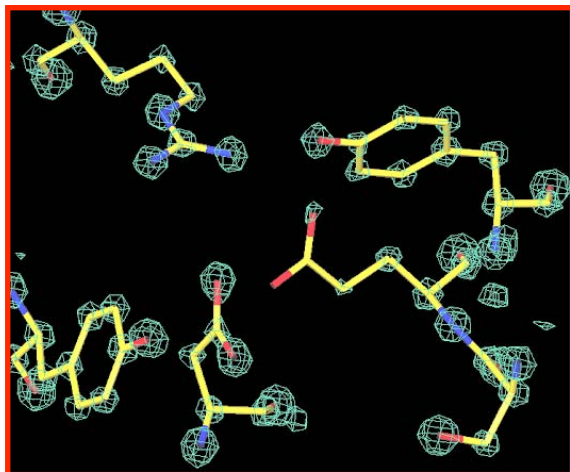
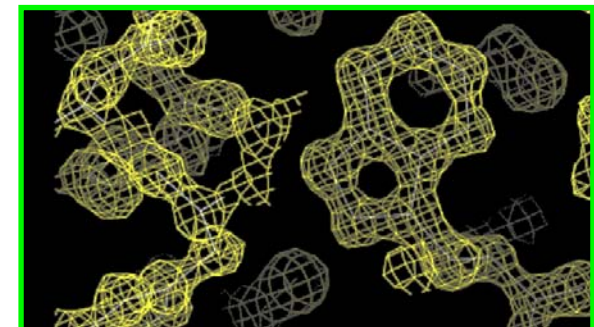


(Limited) Resolution of the X-ray Data

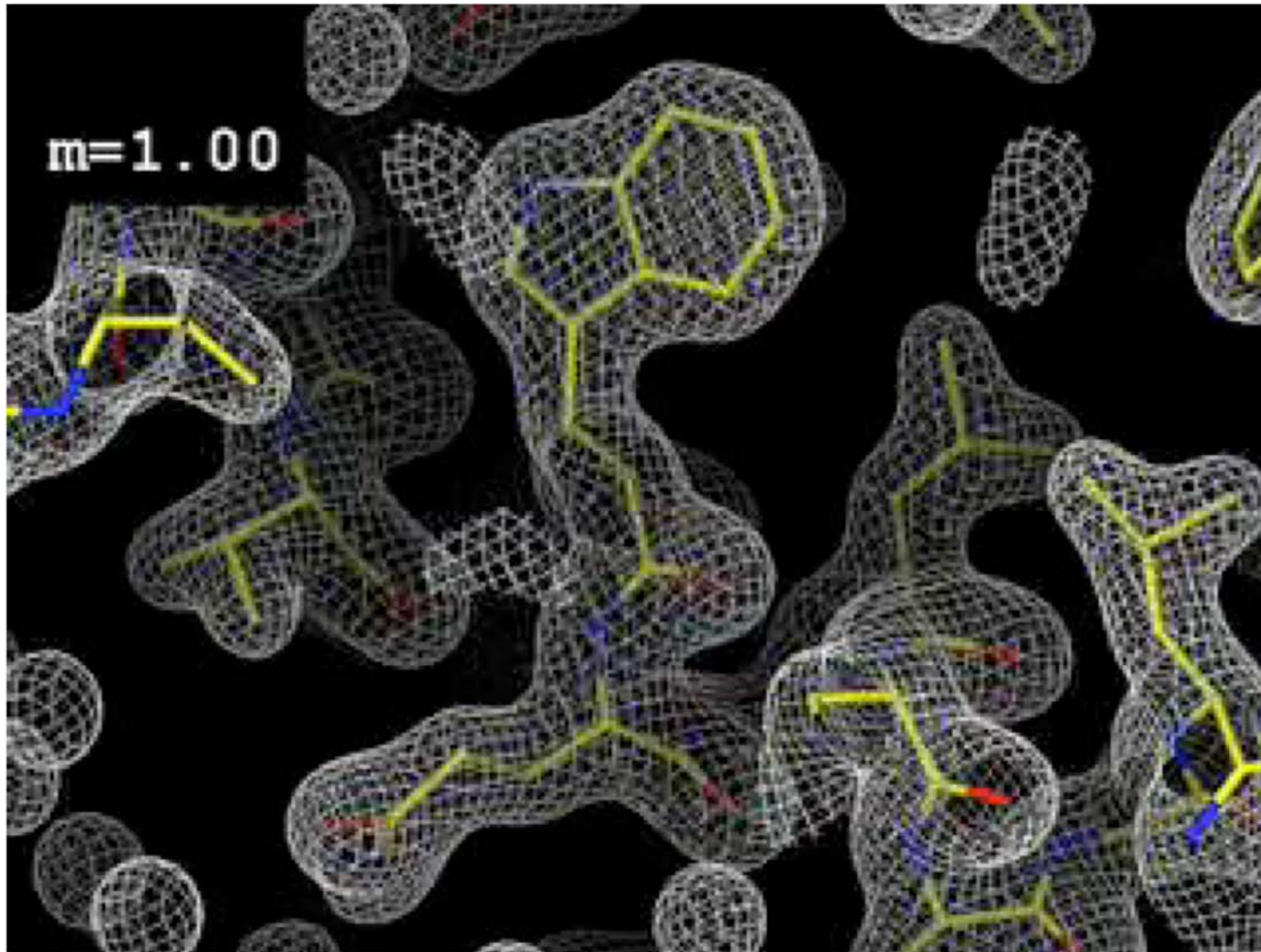
7.2
ARP
wARP



20 Å	3.0 Å	1.9 Å
1.6 Å		1.2 Å
1.0 Å	0.8 Å	ultra

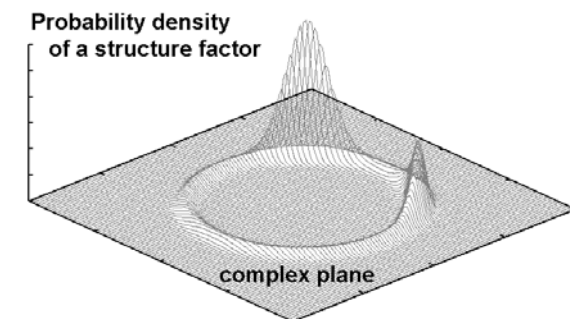
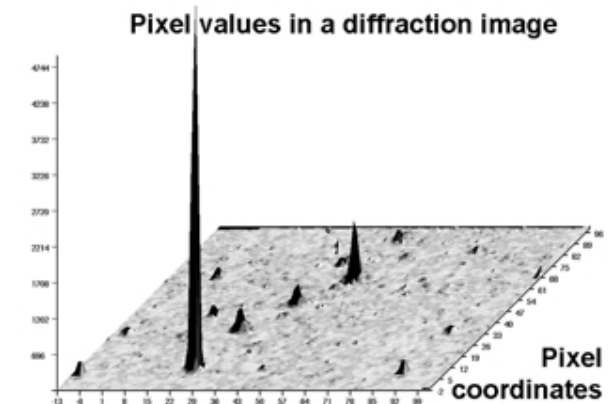
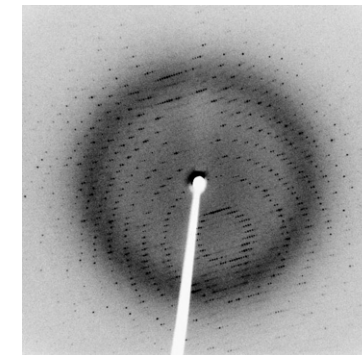
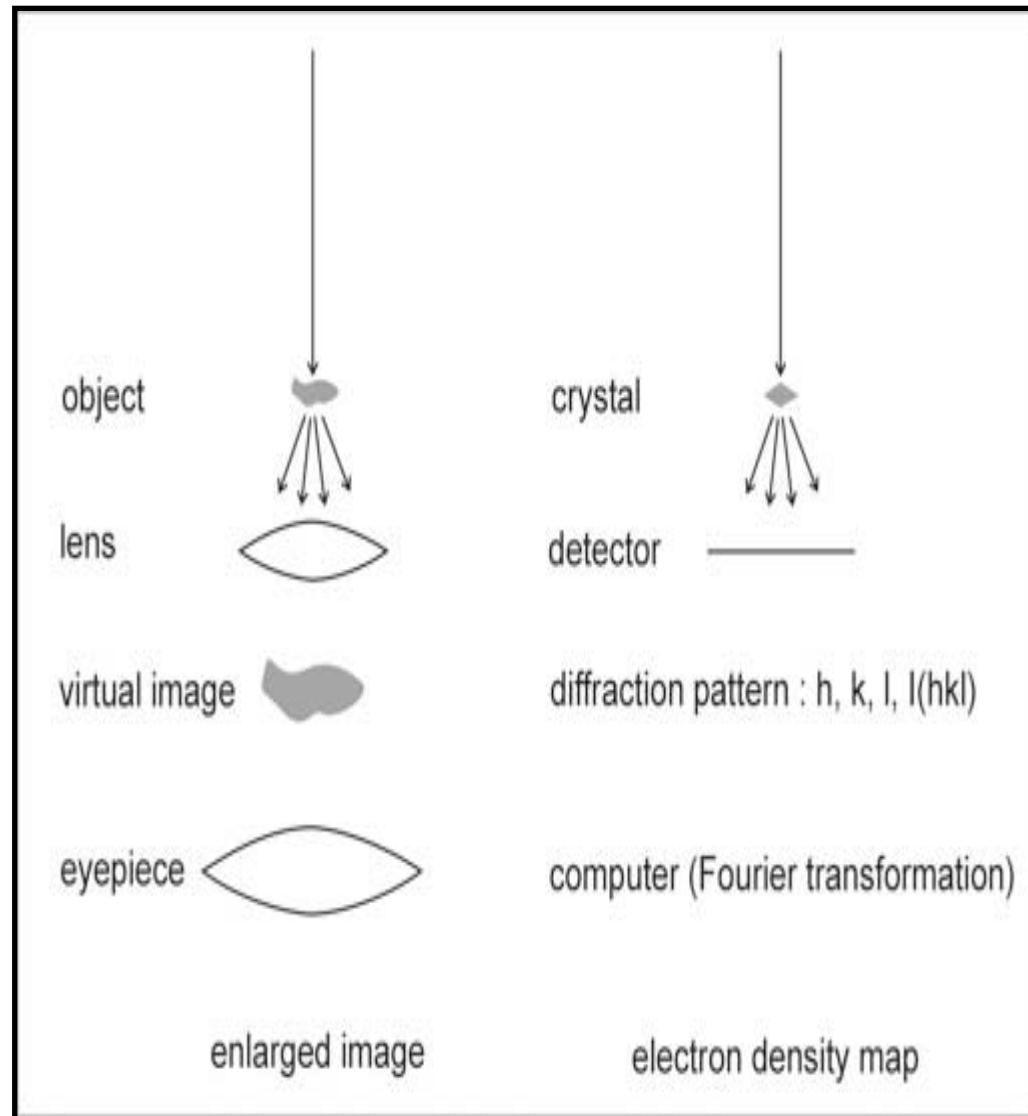


Importance of the Phases

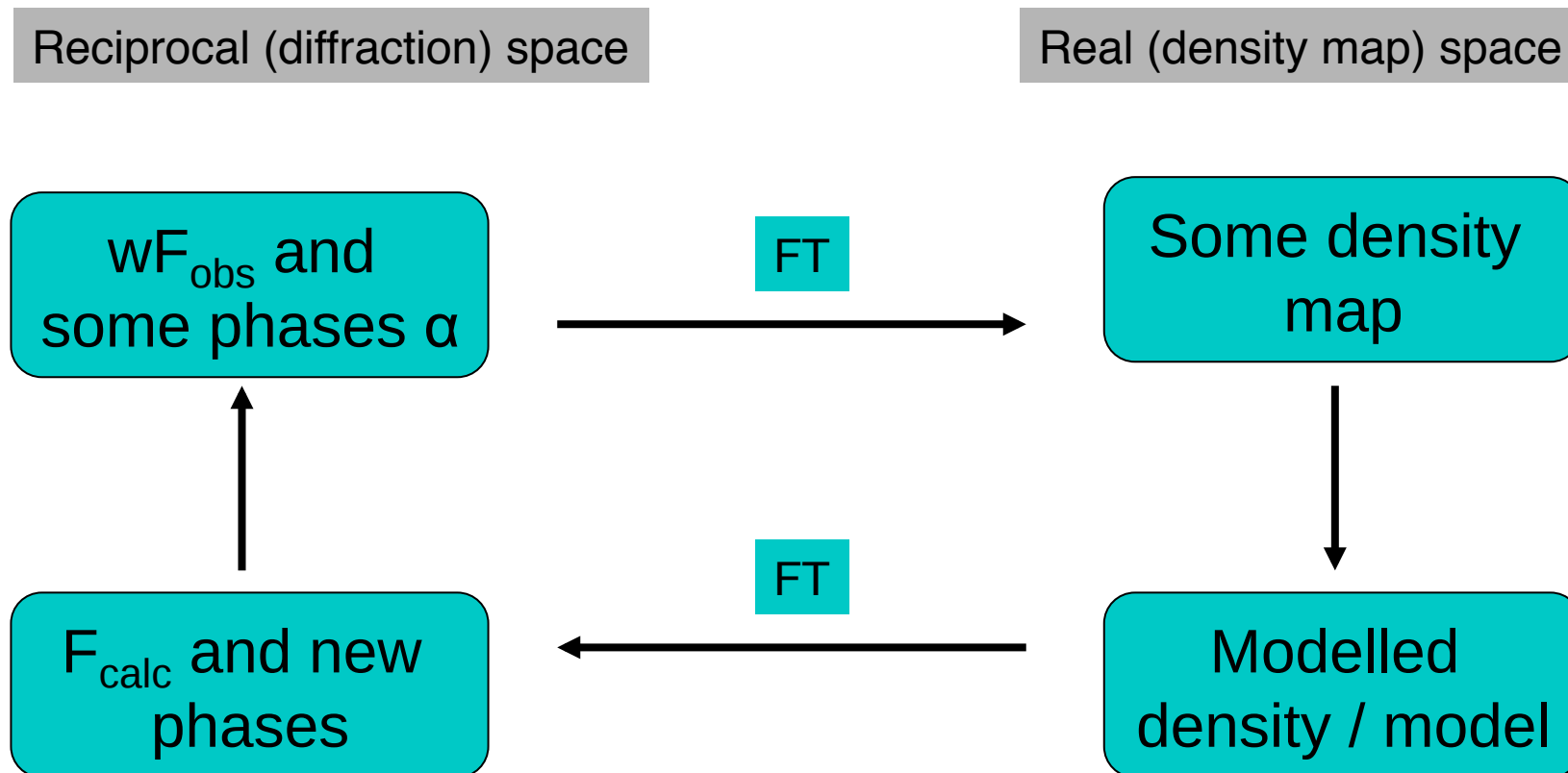


Why Do We Worry So Much About Phases?

7.2
ARP
wARP



Iterative Solution to the Crystallographic Phase Problem



Interpretability and Information Content



Interpretability of a density map is defined by its information content

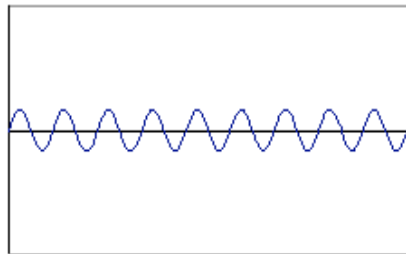
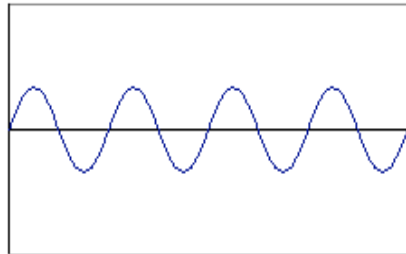
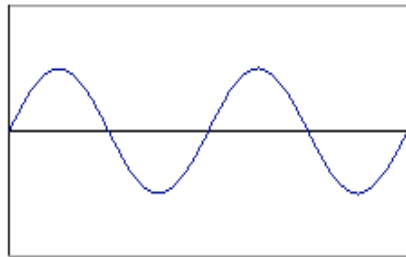


The information content of our modelling should match the information content of the map

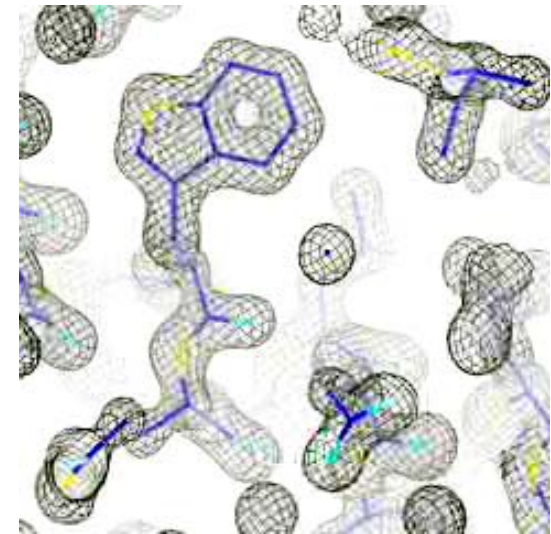
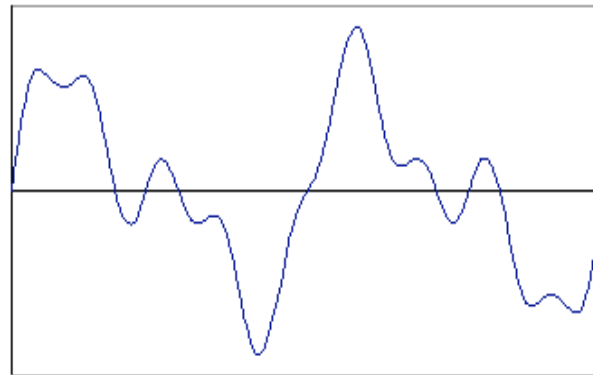
Modelling the Density: Fourier Transform



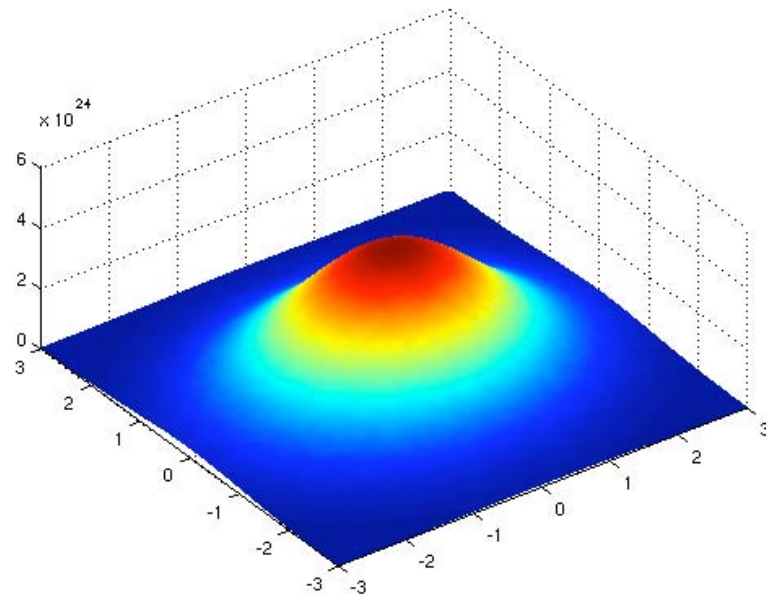
$$p(\mathbf{x}) = \sum_{\mathbf{h}} |F_{\mathbf{h}}| \cos(2\pi \mathbf{h} \mathbf{x} - \alpha_{\mathbf{h}})$$



=



Modelling the Density: Radial Basis Functions

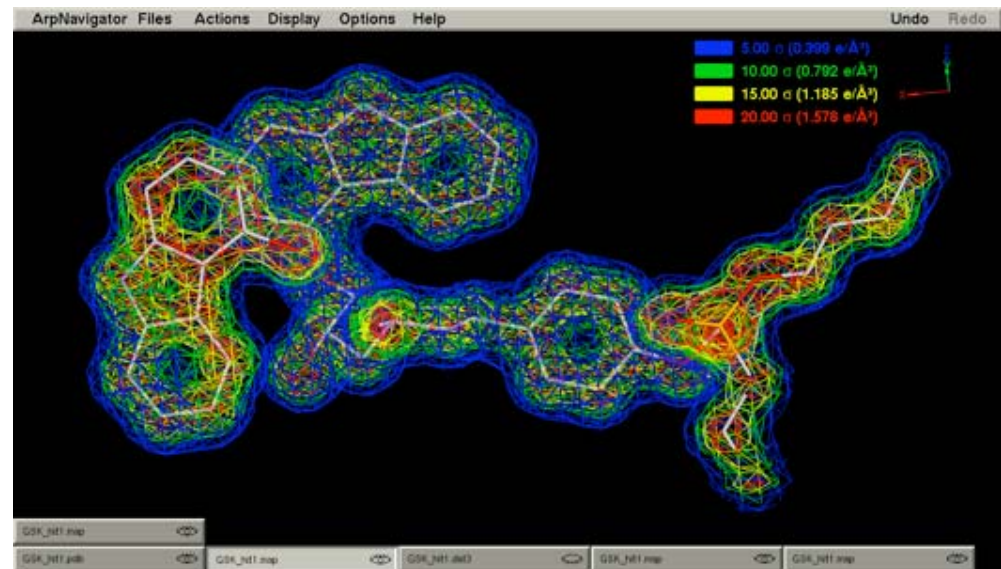


$$f(x) = \sum_i w_i \exp\left(-\frac{(x - x_i)^2}{2\sigma^2}\right)$$

One per atom

Advantage:

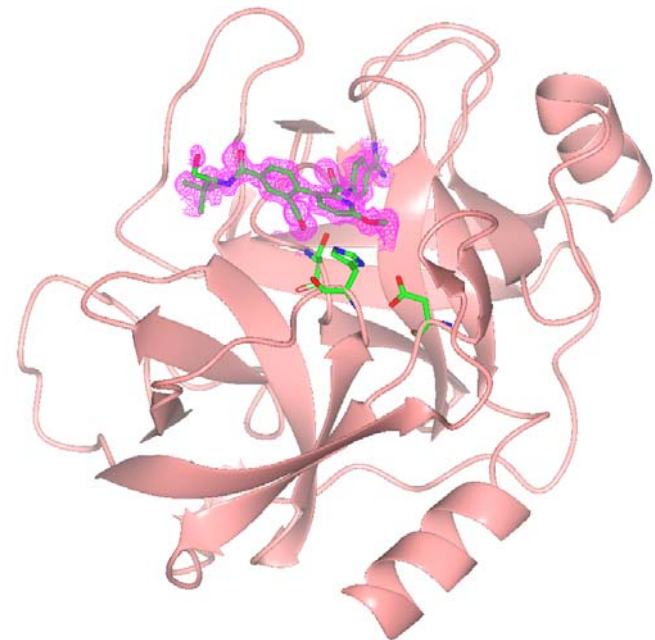
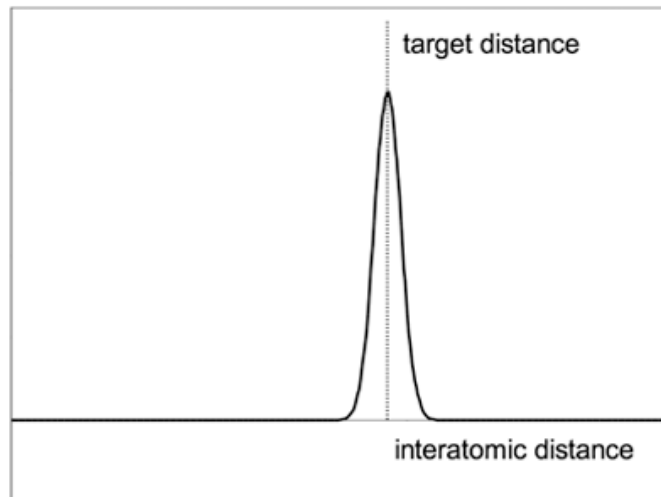
Gaussians in real space are Gaussians in reciprocal space, too!



Modelling the Density: Ball & Stick



xyz1 xyz2



Number of X-ray Observations

The number of diffracted X-ray reflections for a complete data in P1 lattice

$$N_{refl} = \frac{2\pi}{3d^3} V$$

Assuming 50% solvent content and an MW for an average residue of 110 Da, we arrive at:

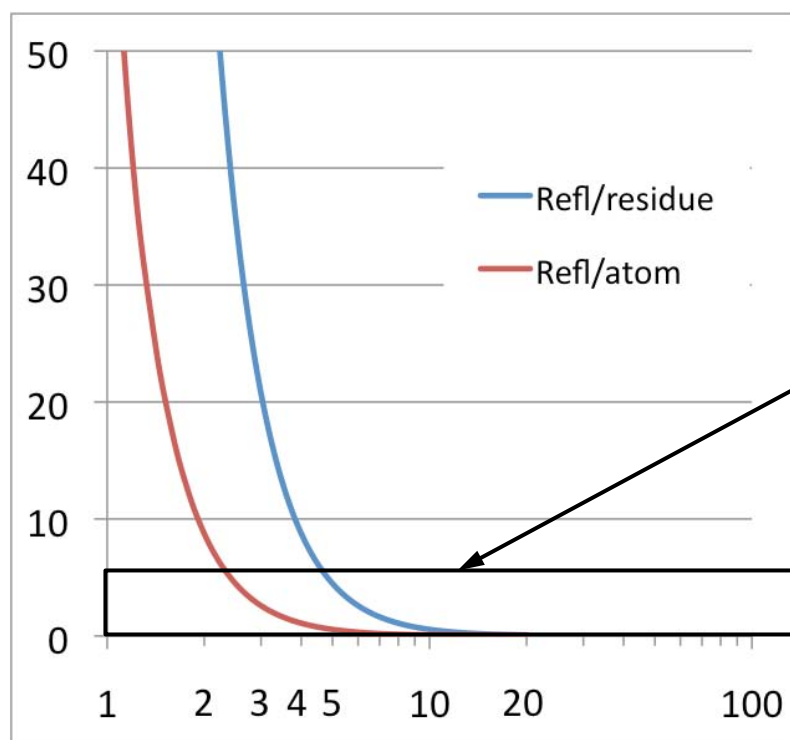
$$N_{refl} = \frac{560}{d^3} N_{res}$$

Or:

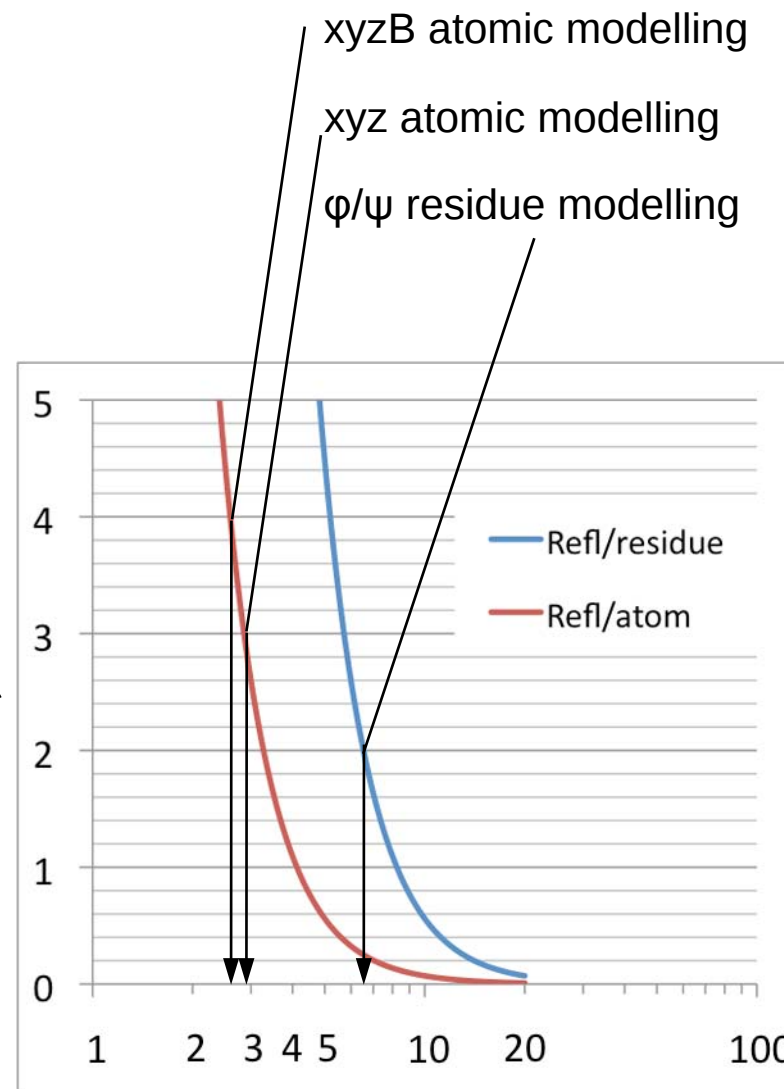
$$\frac{\text{reflections}}{\text{residue}} = \frac{560}{d^3}$$

$$\frac{\text{reflections}}{\text{atom}} = \frac{70}{d^3}$$

Number of X-ray Observations



$d (\text{\AA}^3)$



$d (\text{\AA}^3)$

Information Content



In the absence of additional data the information content of the model **cannot exceed the information content of the map**

One may try to increase the information content of the map by complementing it with additional data based on statistical grounds

Exercise: Children

John has got 2 children
What is the probability that both children are boys?

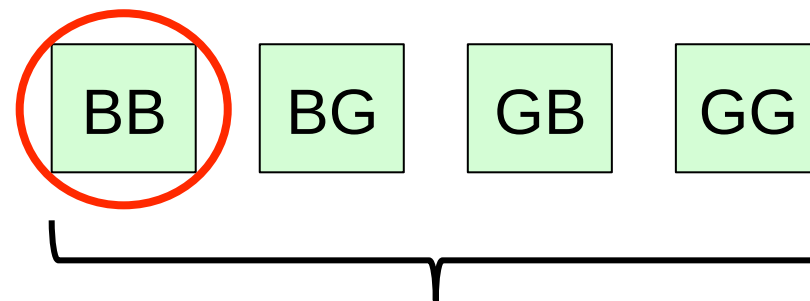
$$f(x = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$p = 0.5; n = 2; k = 2$$

$$f(x = 2) = \frac{2!}{2!0!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^0 = \frac{1}{4}$$



1/4



$$f(x) = \frac{P_success}{P_all_events}$$

Exercise: Children

John has got 2 children
At least one is definitely a boy
What is the probability that both children are boys?

$$f(x = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$p = 0.5; n = 2; k = 2$$

$$f(x = 2) = \frac{2!}{2!0!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^0 = \frac{1}{4}$$

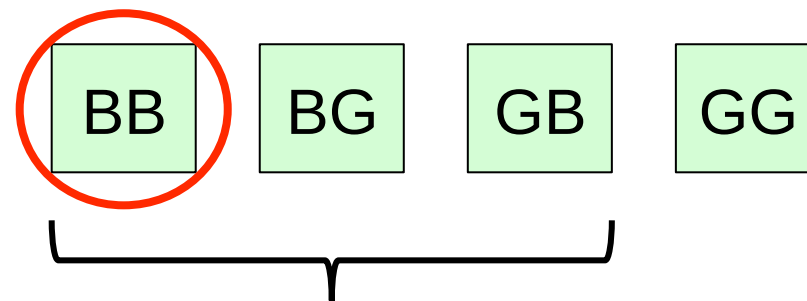
$$p = 0.5; n = 2; k = 1$$

$$f(x = 1) = \frac{2!}{1!1!} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^1 = \frac{1}{2}$$

$$f(x = k | k \geq 1) = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{2}} = \frac{1}{3}$$



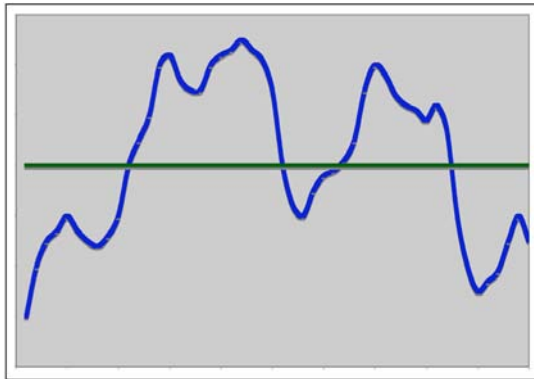
1/3



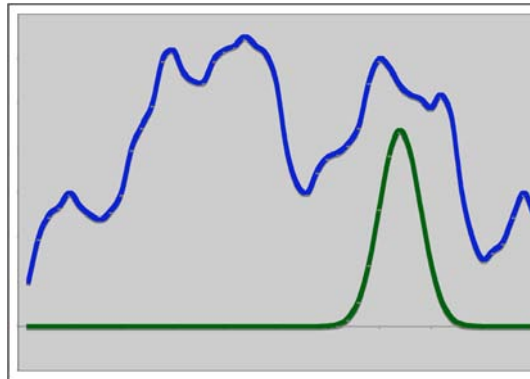
$$f(x) = \frac{P_success}{P_all_events}$$

Knowledge/Model-Based Map Improvement

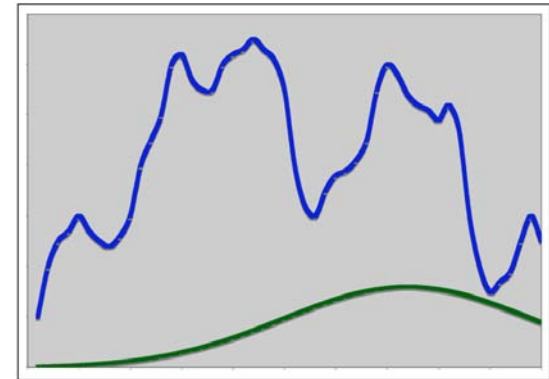
Uniform prior



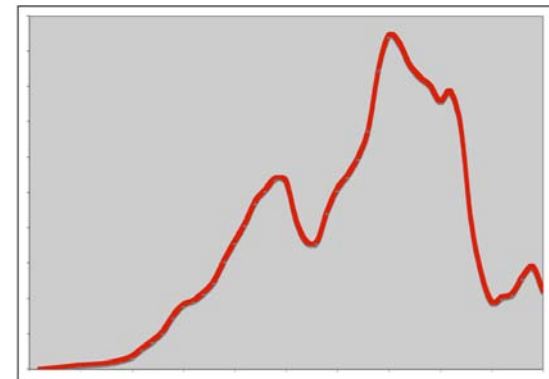
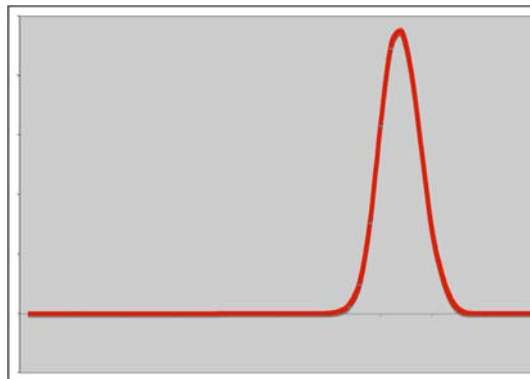
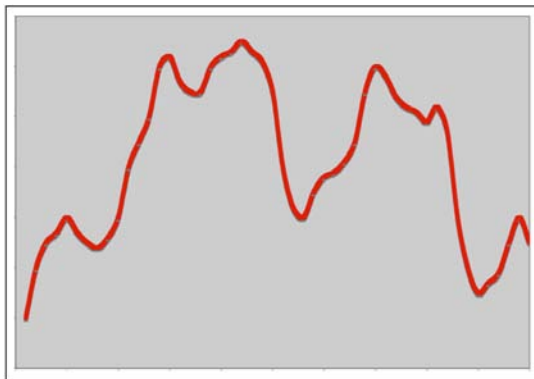
Tight prior (~constraint)



Smooth prior (restraint)



↓ $p = p_1 p_2$

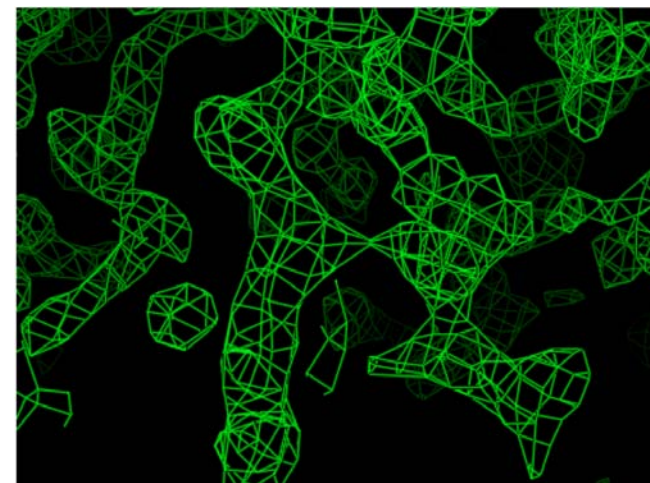
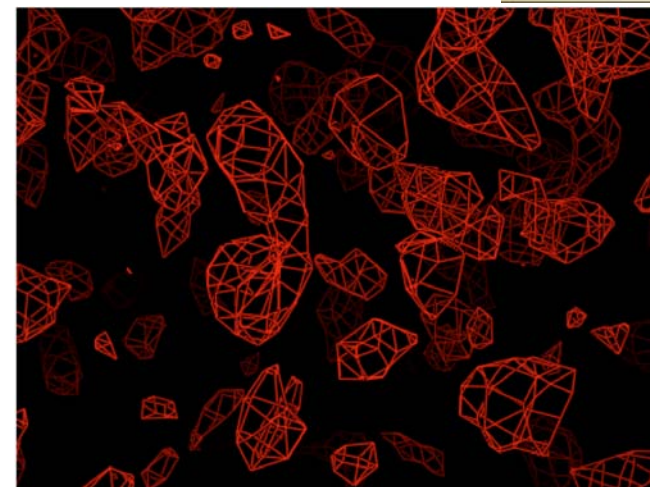


Improvement of a Density Map Prior to its Interpretation



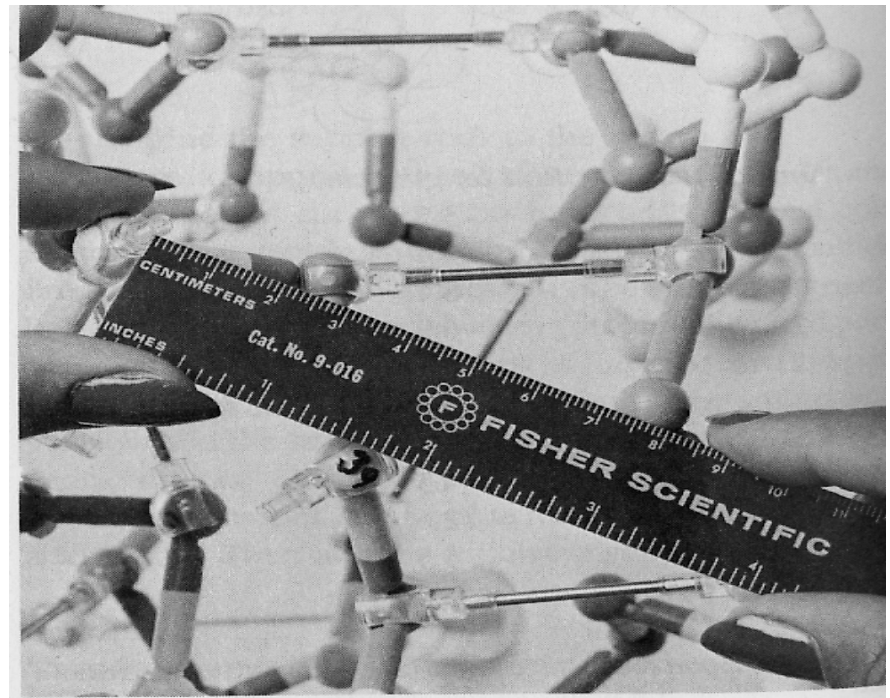
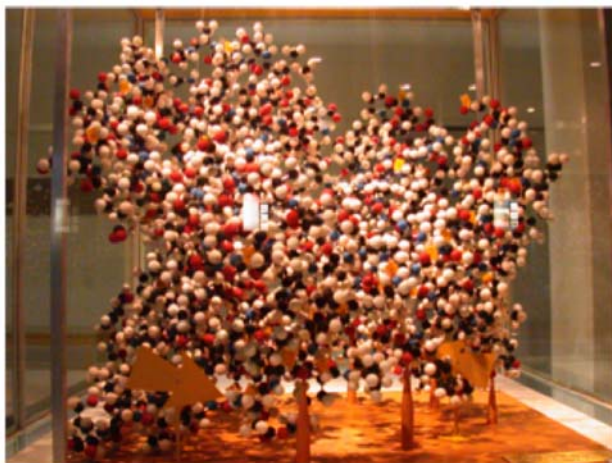
Successful key-concepts

- solvent flattening/flipping
- histogram matching
- non-crystallographic averaging/symmetry



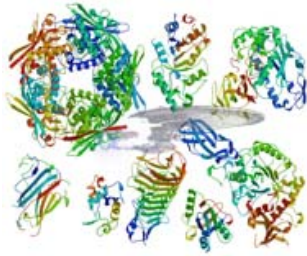
First Map Interpretations

7.2
ARP
wARP

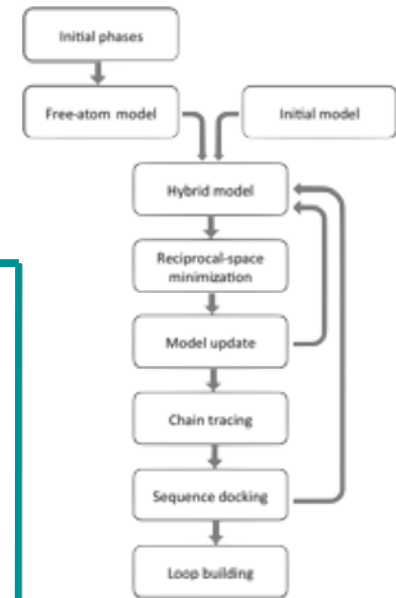


Victor Lamzin, CCP4 workshop, Okinawa, December 2011

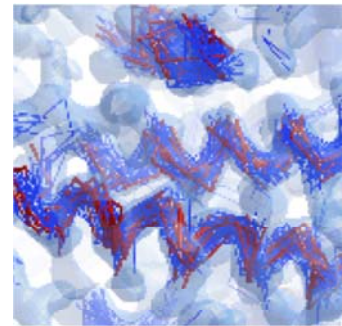
The ARP/wARP Project



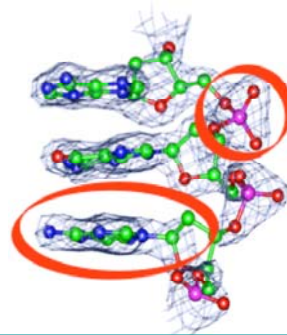
Iterative protein-model building



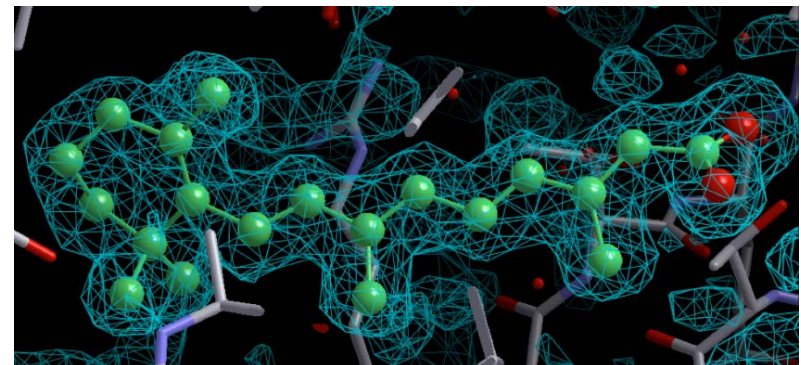
Recognition of secondary structure



Building polynucleotides



Ligand building and screening



Major Software Releases



1998

1999

2002

2004

2007

2009

2011

Protein chain tracing

Main chain



Auto-NCS



Side chains

Helices / strands

Loop completion

Non-protein parts

Ligands



Nucleotides

Solvent

Refmac: twin,
SAD, bulk
solvent, jelly, etc

Front End

Terminal / shell scr



CCP4 GUI

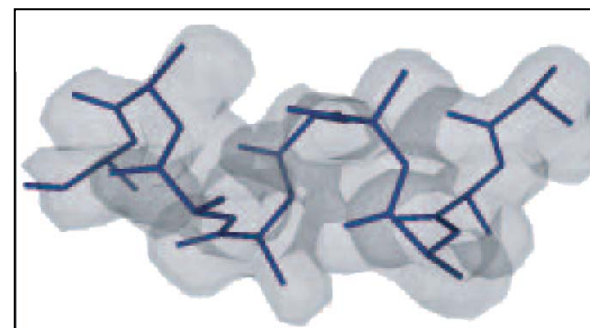
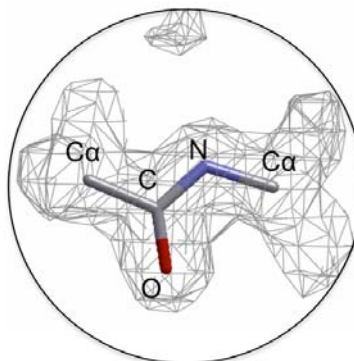
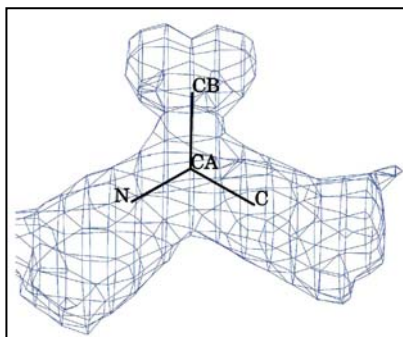
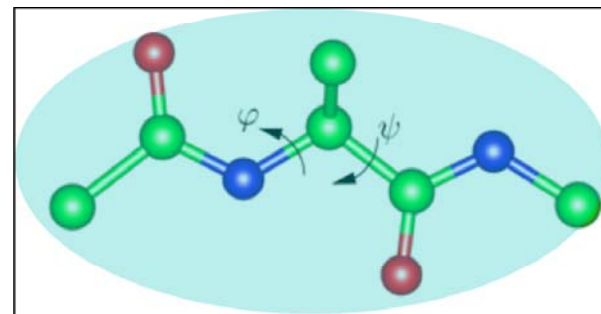
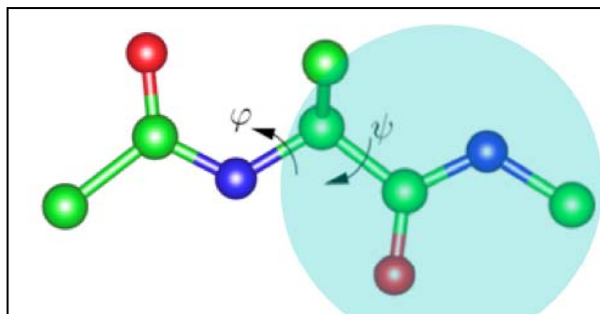
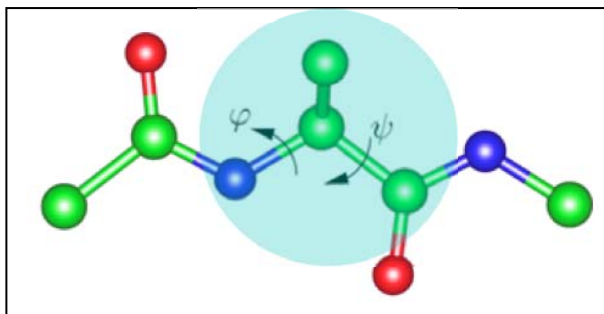
Web service



ArpNavigator

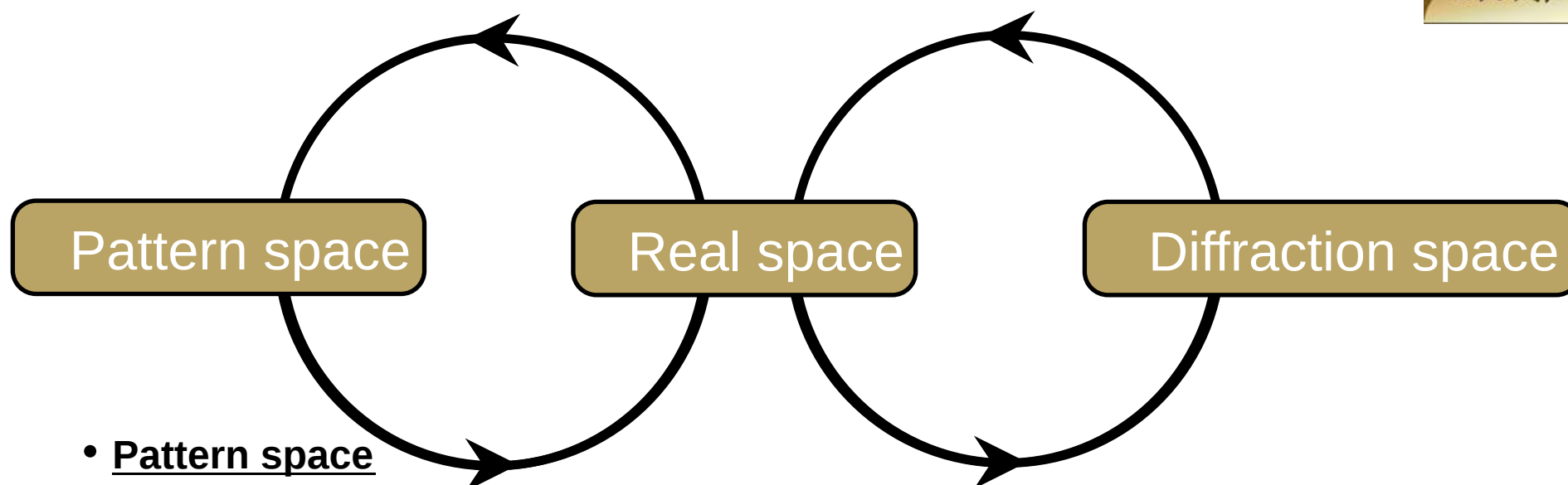


Methods for Protein Model Building



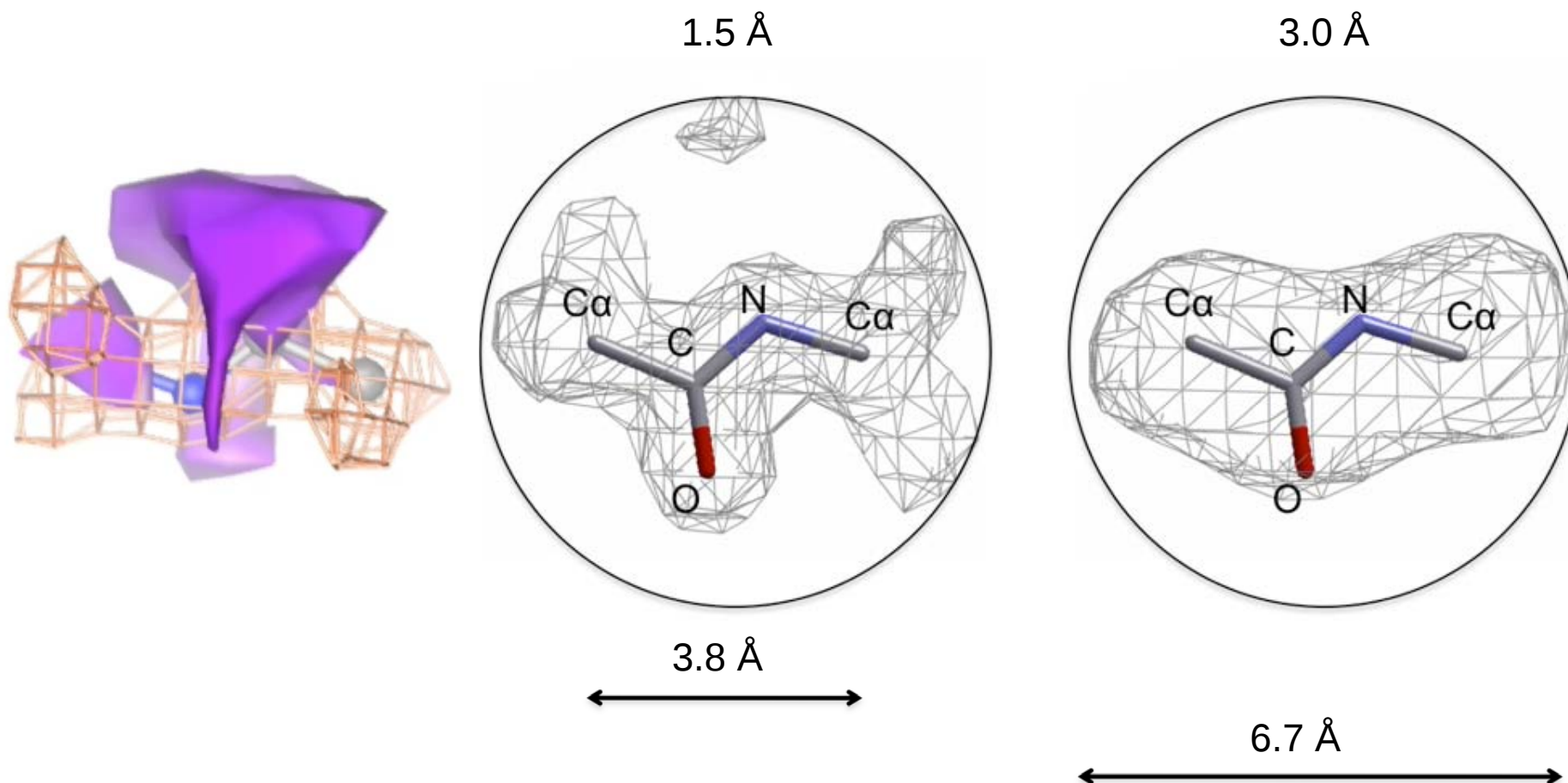
TEXTAL/Buccaneer	ARP/wARP	Resolve/ACMI
Ca atoms	Peptides/Dipeptides	Fragments

Fundamental ARP/wARP Concepts

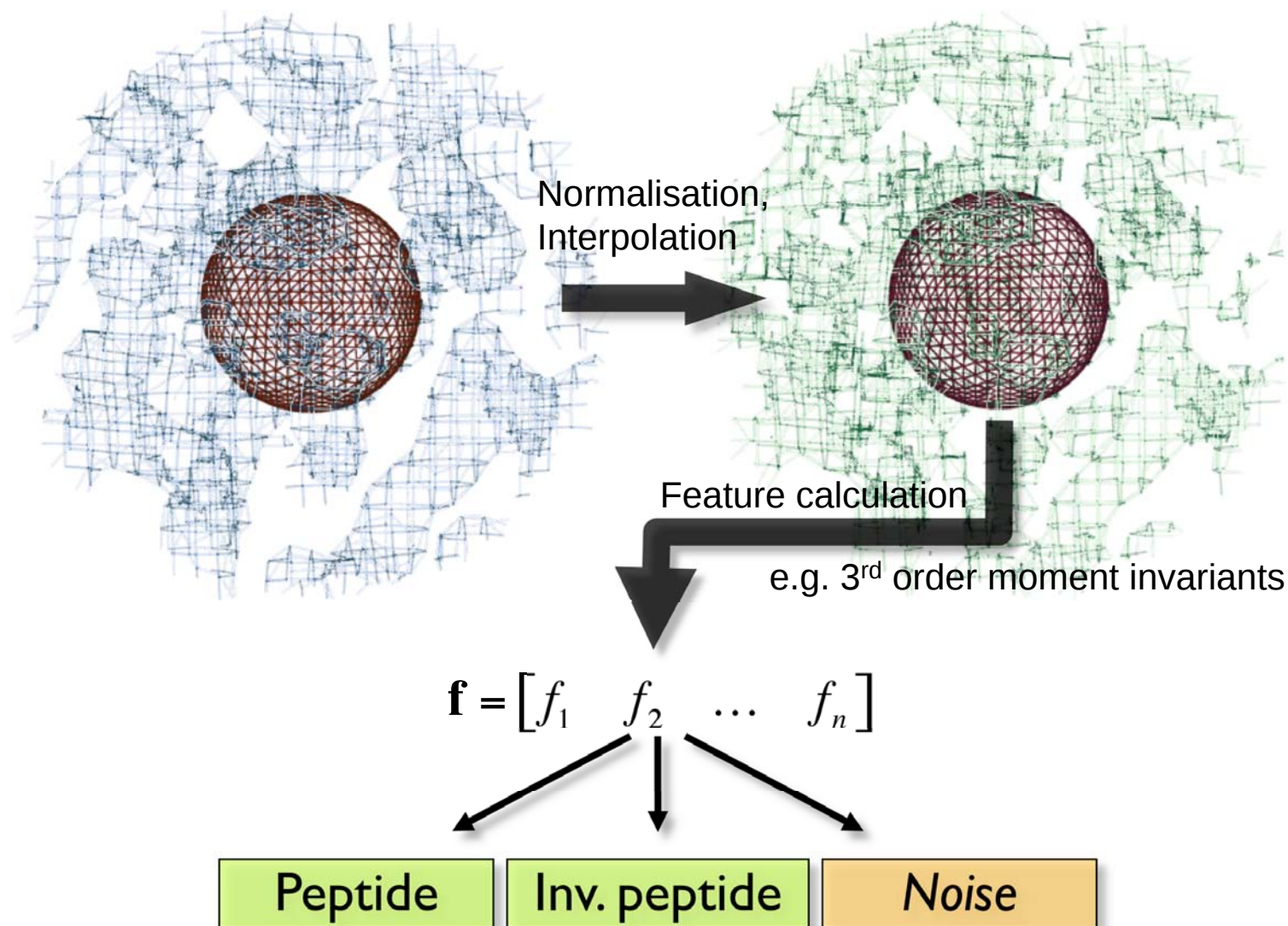


- Pattern space
 - Electron density: local object interpretation
 - Hybrid Model: local motif interpretation
- Real space
 - Hybrid model: atoms having chemical identity and free atoms
 - Model update: removing and adding parts of the model
- Diffraction space
 - Unrestrained refinement of free parts of the model
 - Restrained refinement of chemically assigned atoms

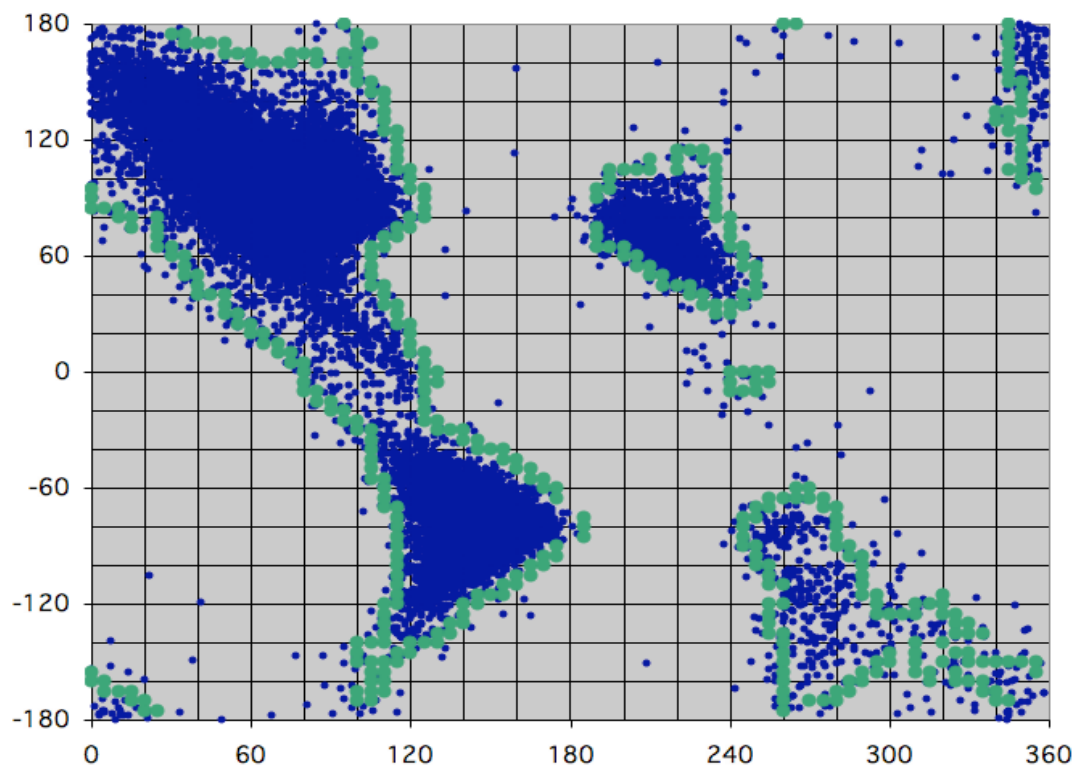
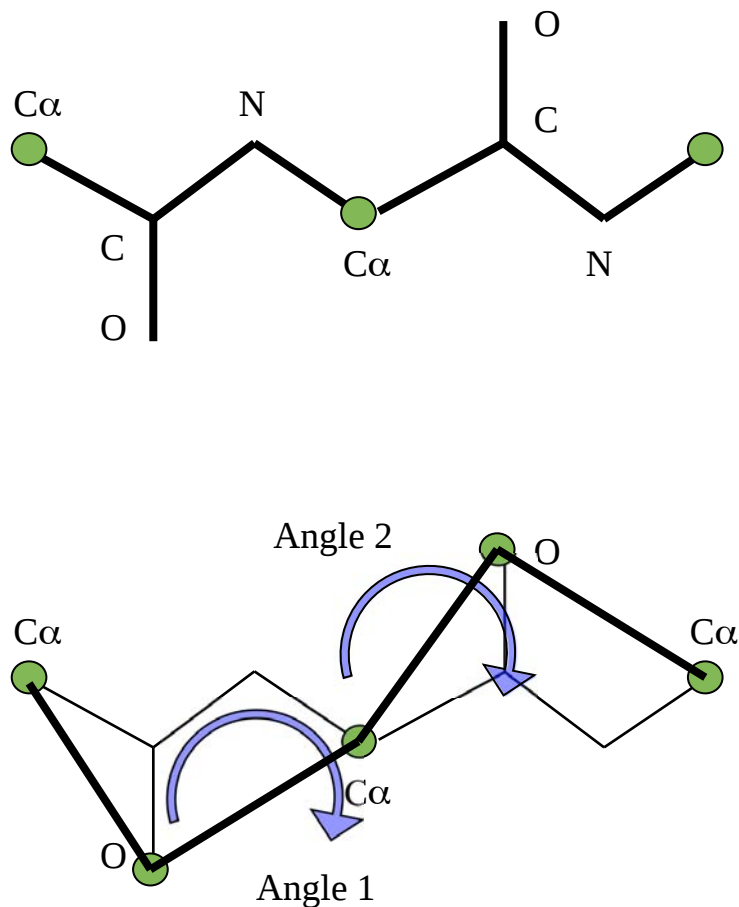
Concept #1: Pattern Recognition



Local Pattern Recognition



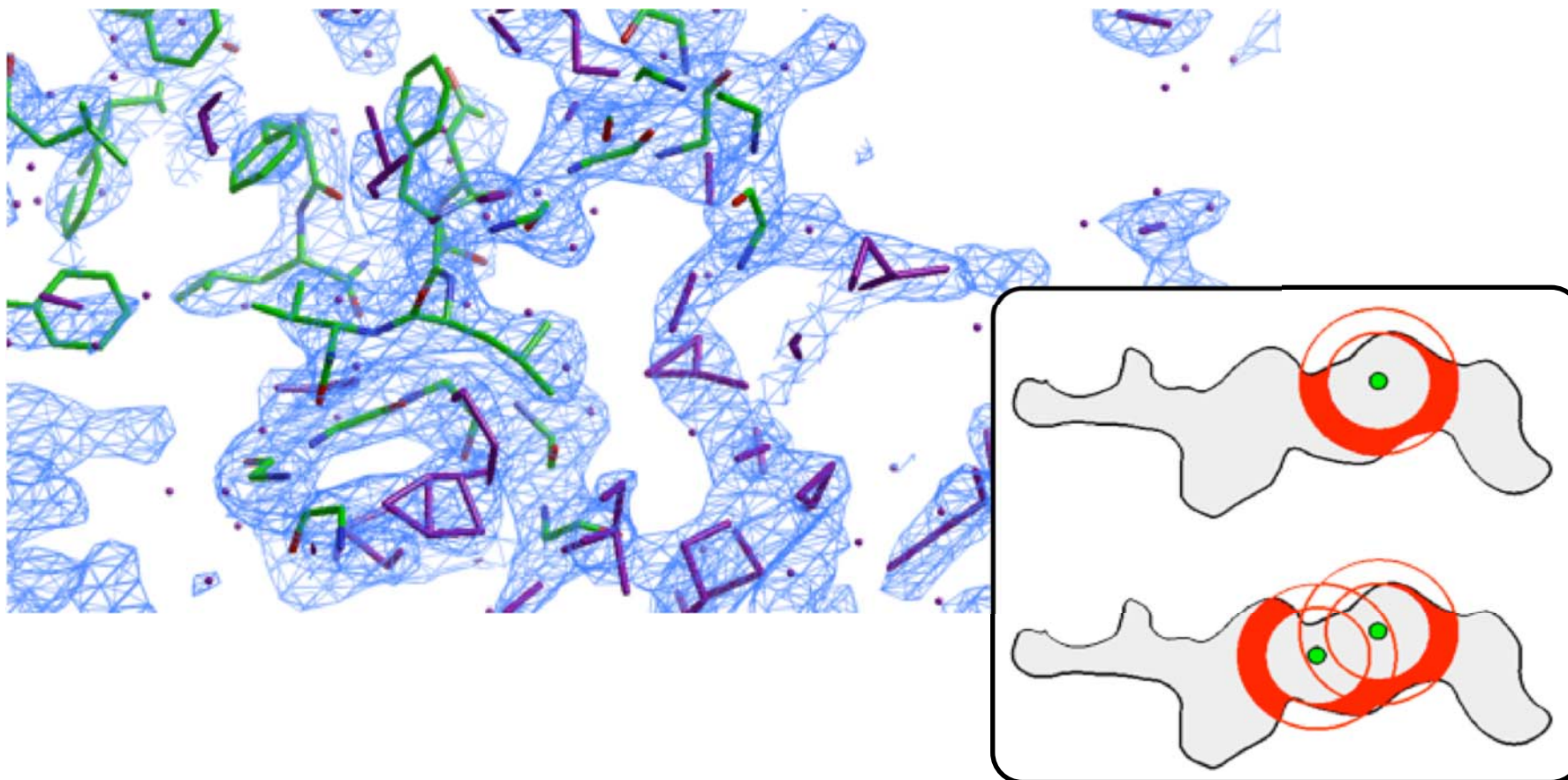
Building Protein Chain: From Peptides to di-Peptides



Restrict possible main-chain conformations

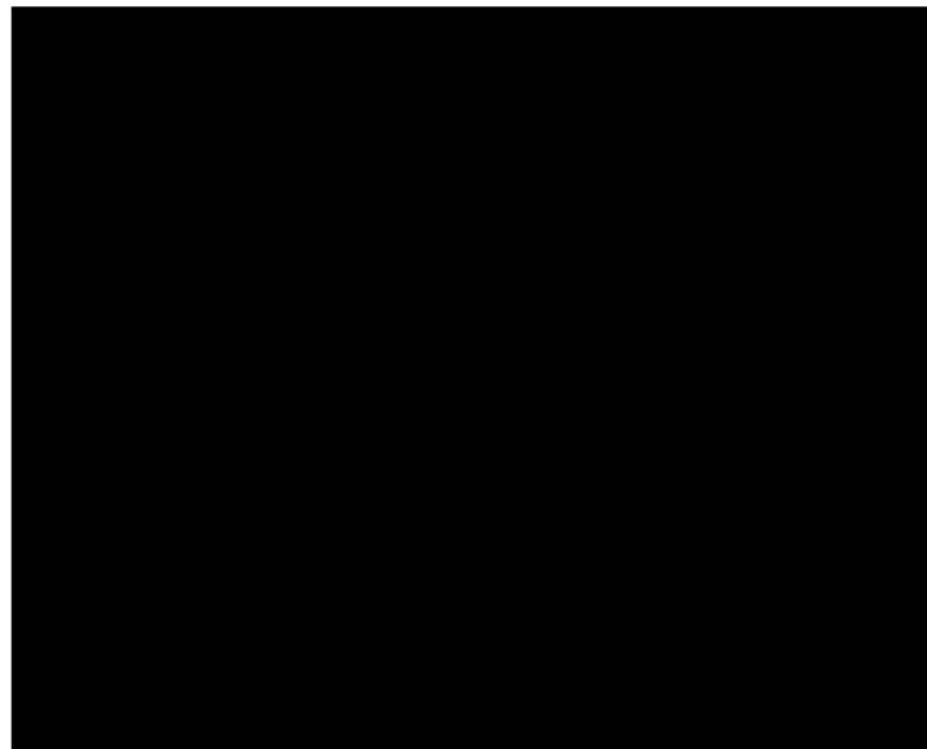
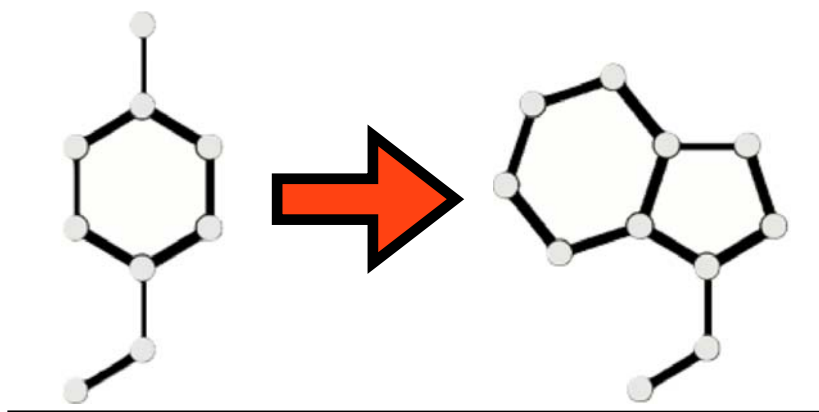
Concept #2: The Hybrid Model

- Partial model is used together with a free-atom model
- Chemically assigned parts provide restraints for refinement
- The hybrid model is converging to the final model



Concept #3: Iterative Update

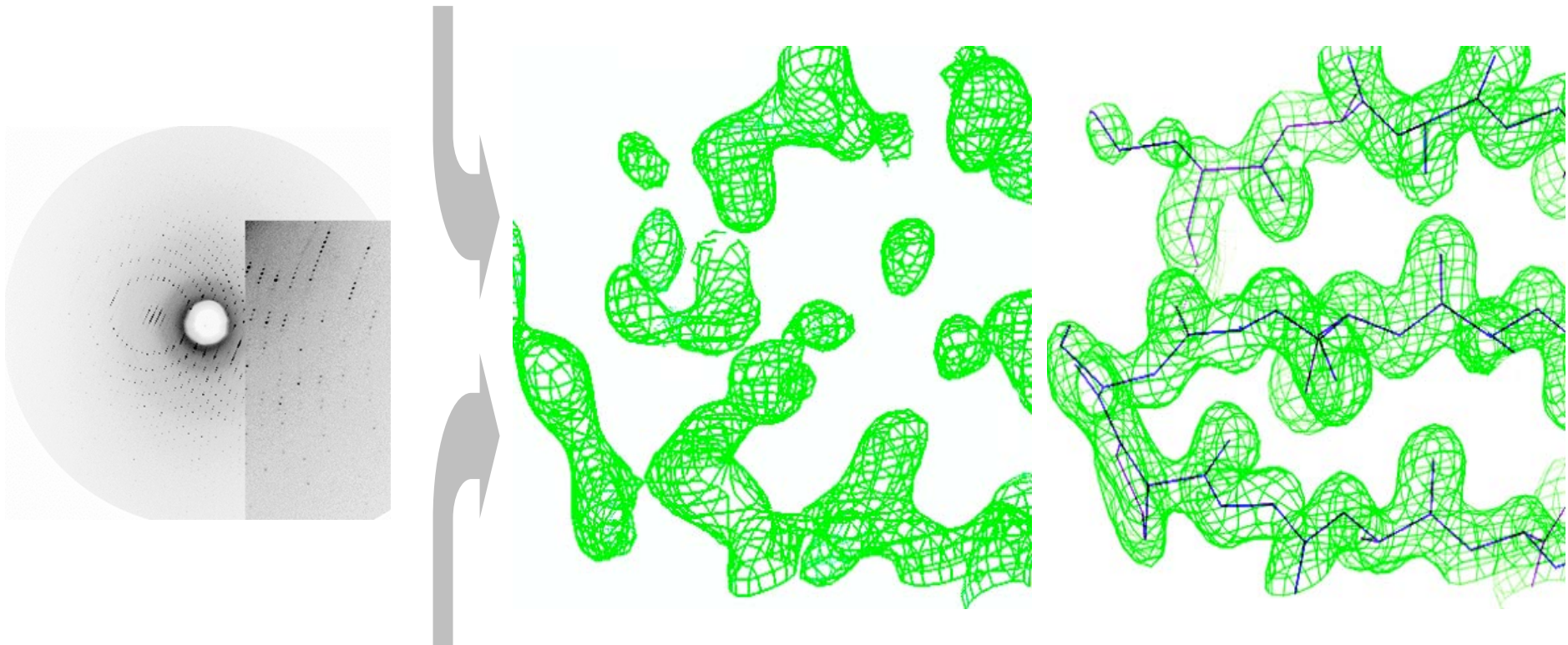
- Iterative update of parts of the model based on density and recognised patterns
- Restraints are re-assigned accordingly



More About the Iterations

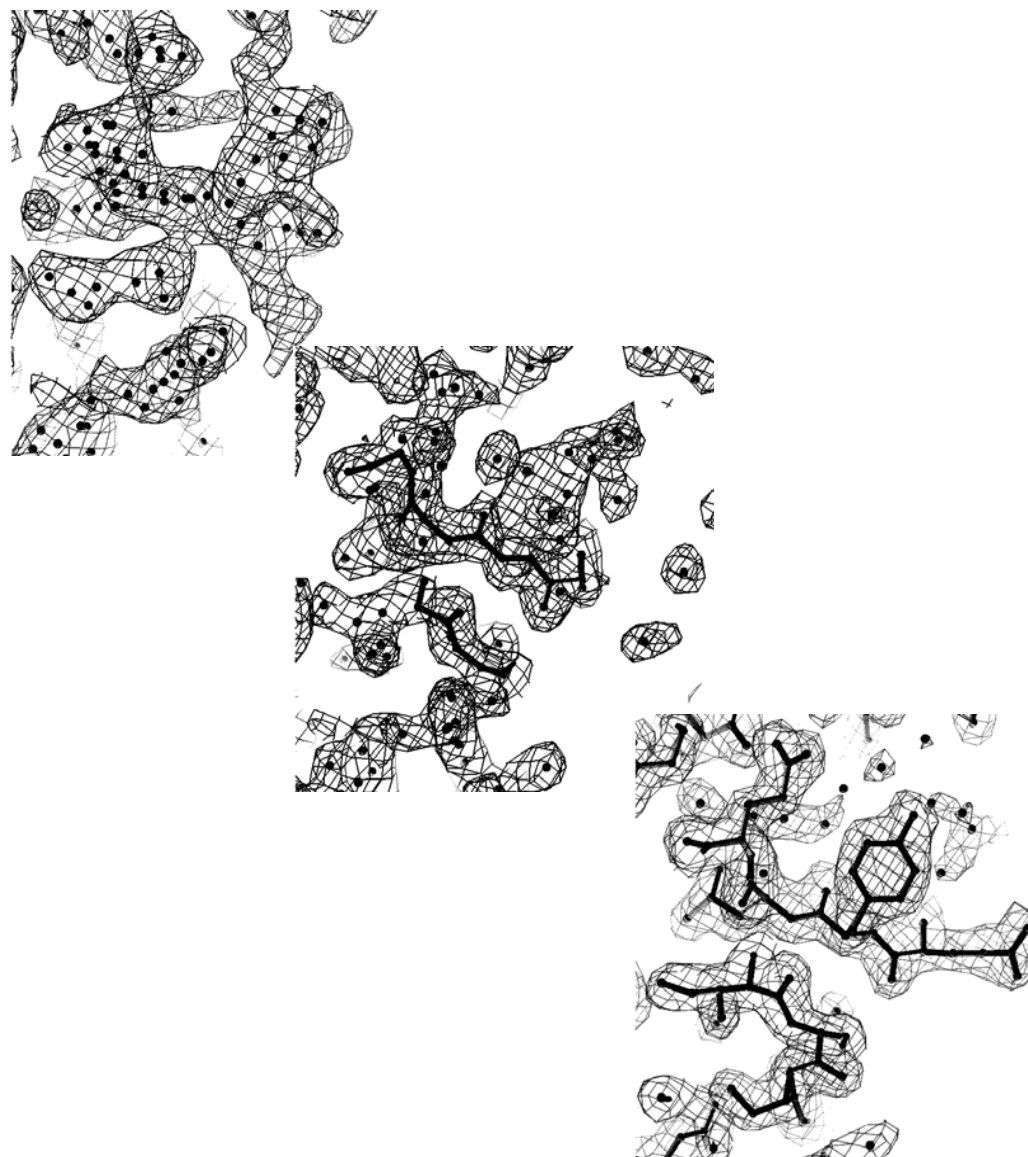
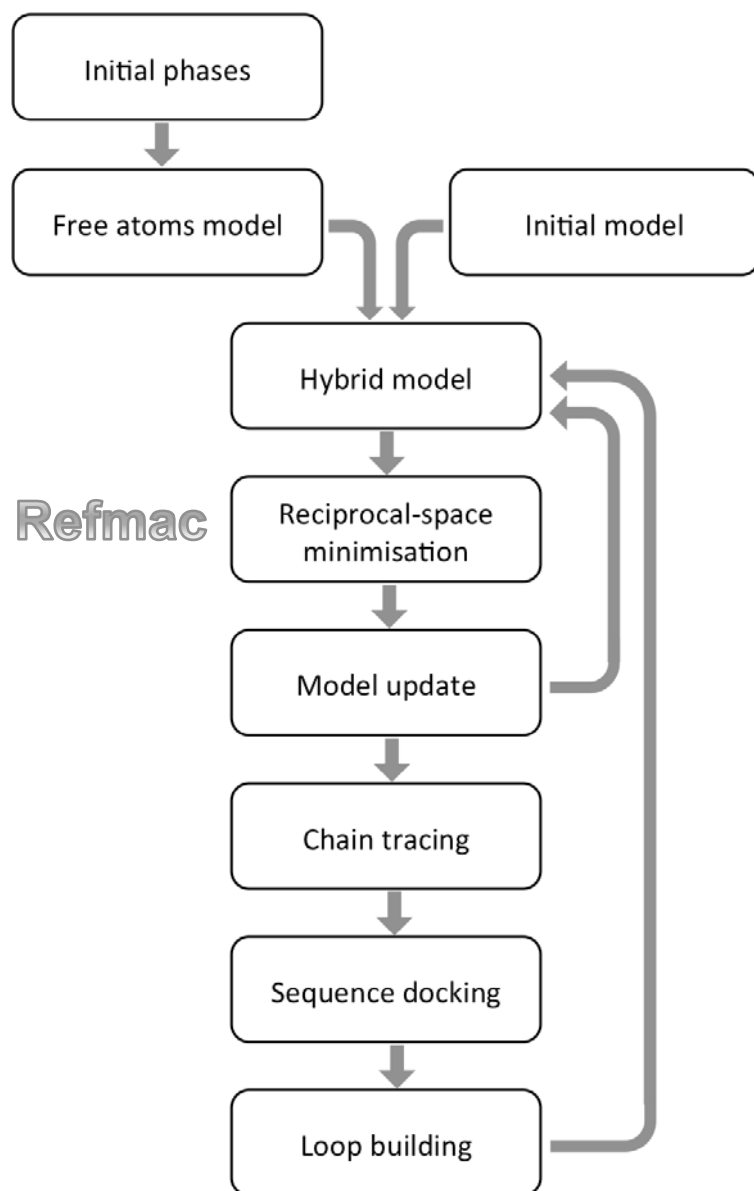


Experimental Phasing (MIR/MAD) Different maps = different models

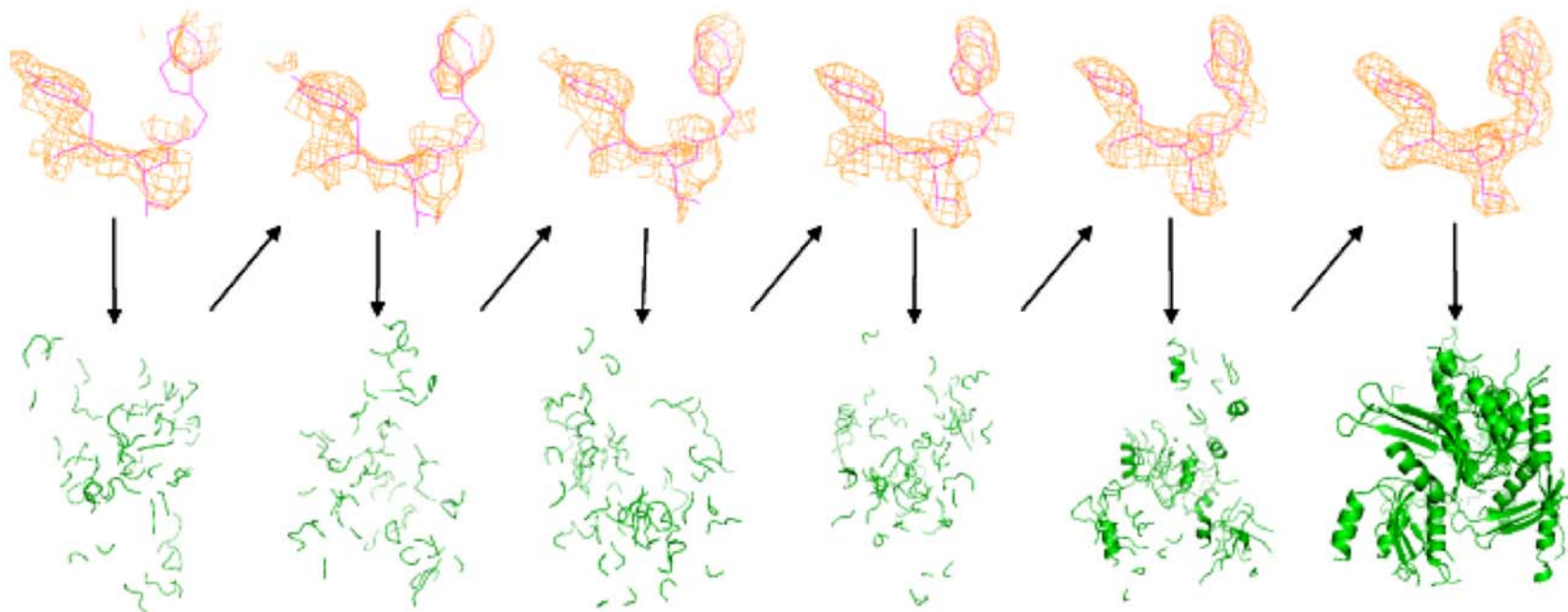
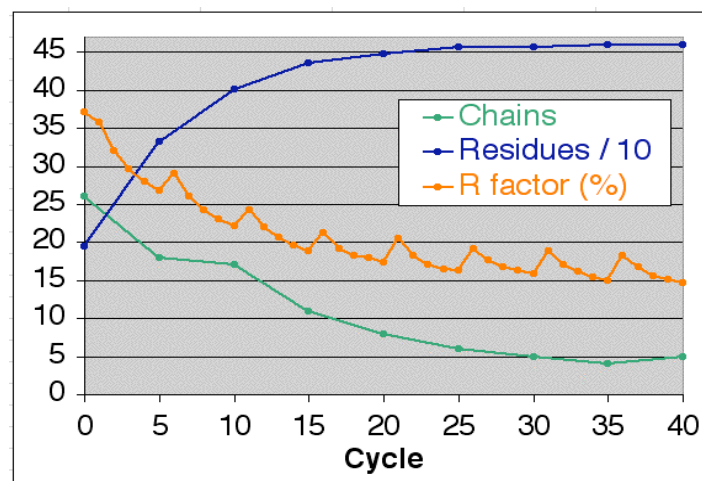


Homologous Model (MR)

Iterative Protein Building in ARP/wARP

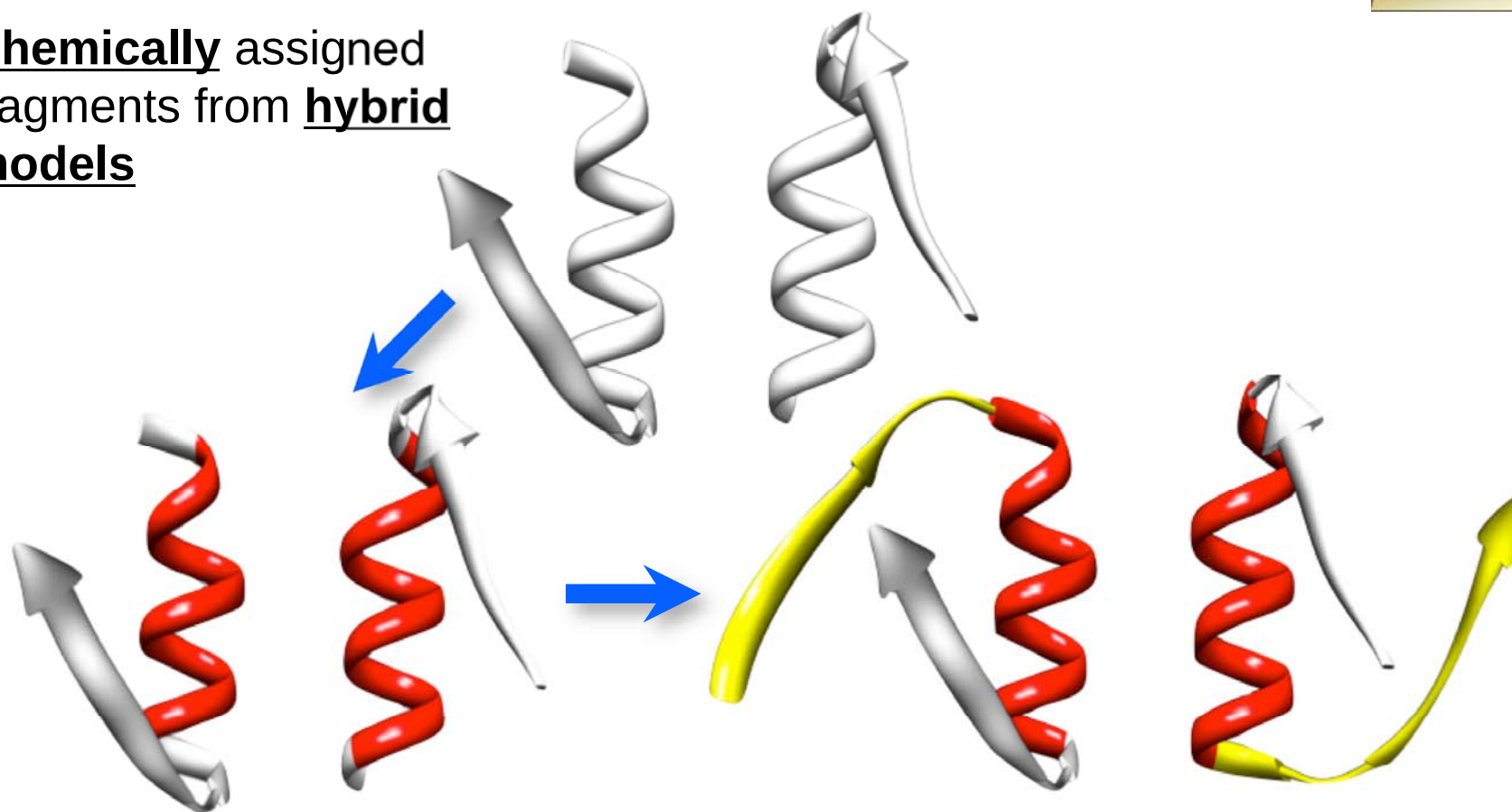


Iterative Protein Building in ARP/wARP



Auto-NCS Detection and Use

Chemically assigned fragments from hybrid models

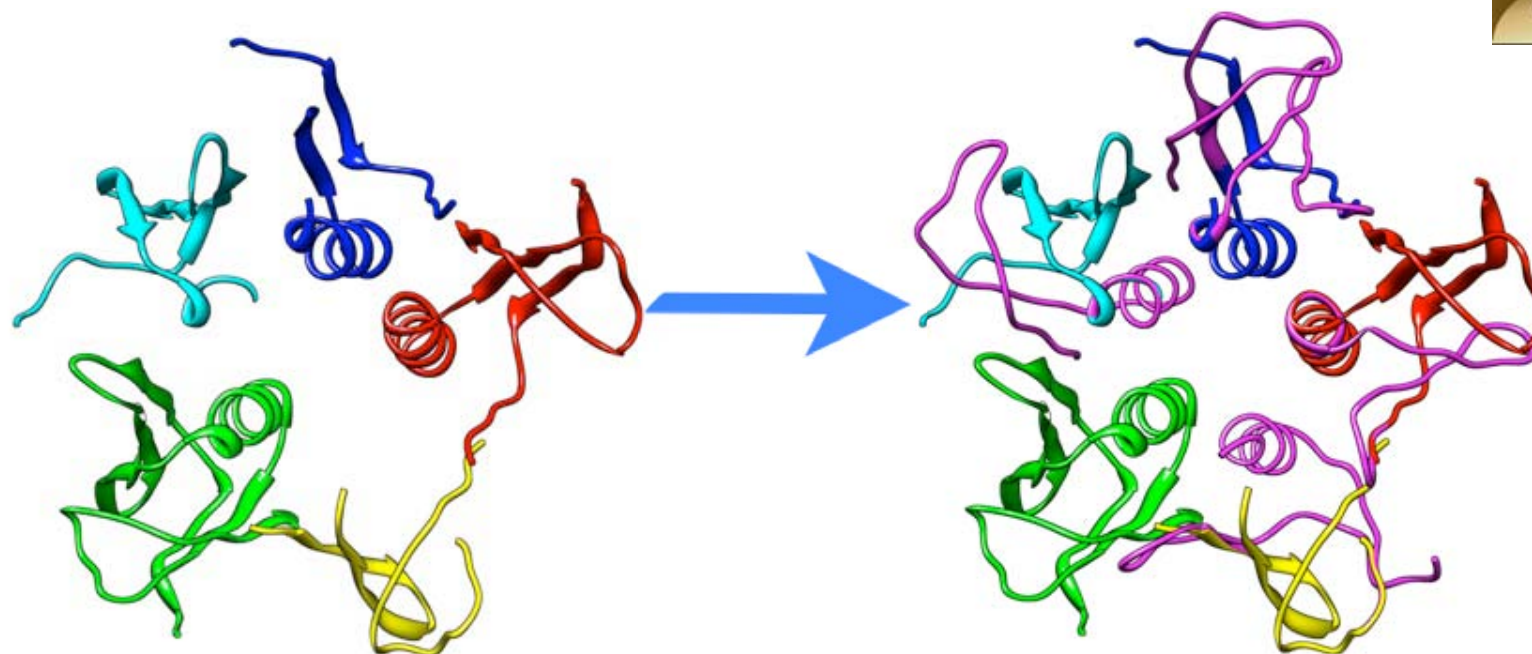


Superposition and clustering to identify NCS matches

Extended matches are transformed to related NCS operators

Seeds for further chain tracing
Restrains for re mac

Auto-NCS Detection and Use



13 protein test structures with resolution 2.1 to 3.2 Å, NCS order 2 - 10

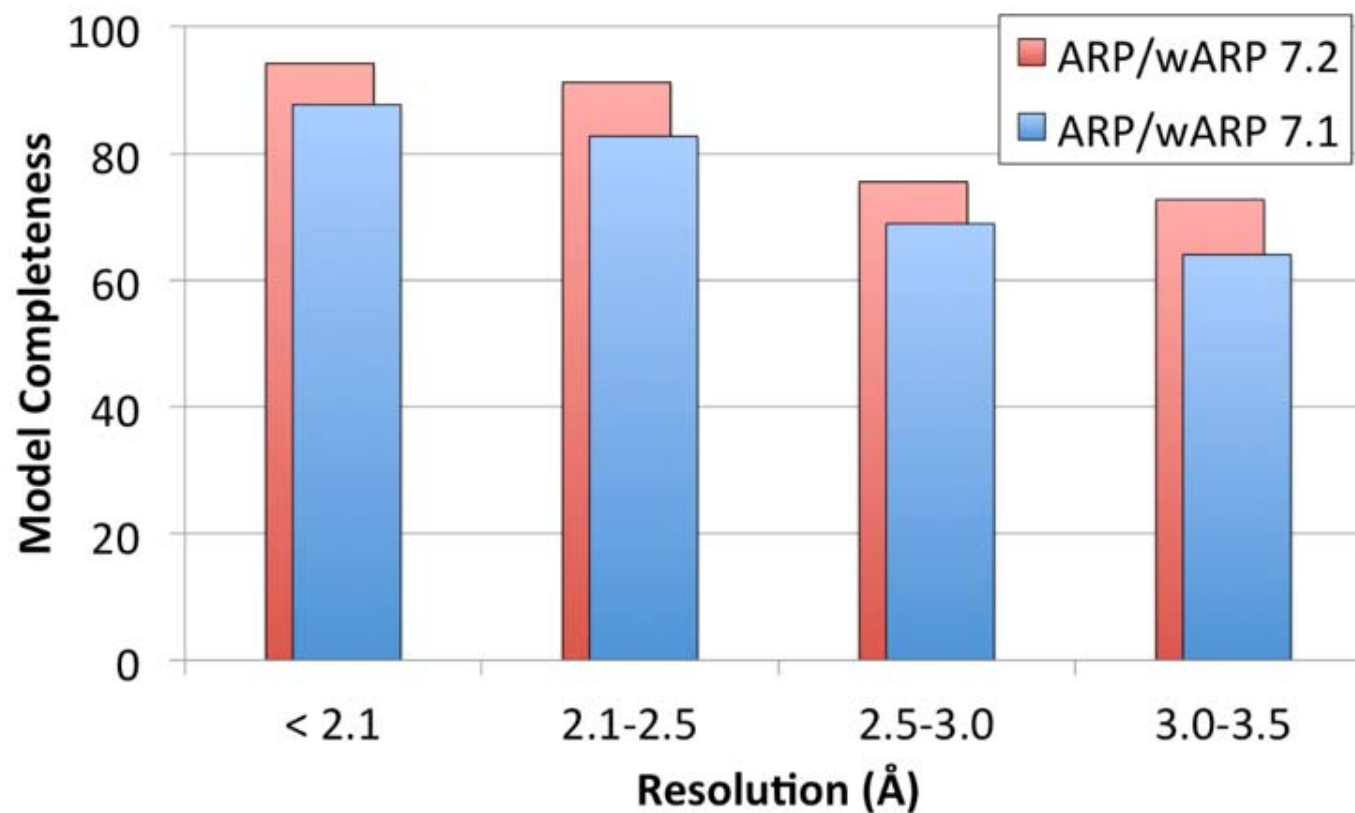
	Built Residues	Residues / Fragment	Sequence Coverage
Top Results	+15.3%	+ 220%	+ 56%
Average in 7.2	+ 4.5%	+ 25%	+ 10%

Dependence on the Resolution of the Data



Resolution	Estimated fraction of automatically built protein structure (last updated on Aug 15, 2011)
2.1 Å and higher	over 90%
2.1 - 2.5 Å	75 - 90%
2.5 - 3.0 Å	60 - 75%
3.0 – 3.5 Å	up to 60%

Dependence on the Resolution of the Data



Remote Computational Services



CCP4 Program Suite 5.0.2 CCP4Interface 1.3.20 running on lamzin-mac.embl-hamburg.de Project: MIRBUILD

Help

Model Building

FFFear - Fragment Searching

FFJoin - Merge fragments

XtalView/xfit

ARP/wARP

ARP/wARP ReBuild

ARP/wARP LigandBuild

ARP/wARP HelixBuild

Job title: automated model building starting from experimental phases

Run ARP/wARP for: automated model building starting from experimental phases

☒ Dock the autotraced chains to sequence after 0 autobuilding cycles.

MTZ in: MIRBUILD psp.mtz Browse View

Fobs: FP Sigma SIGFP

PHIB: PHIDM FOM FOMDM

Sequence in: MIRBUILD psp.pir Browse View

Required parameters

There are 475 total residues in the AU, which belong to 1 molecule(s).

Do 6 cycles of autobuilding (30 total cycles).

Use for REFMAC5 the Fast protocol and do not use the Free R flag

ARP/wARP flow parameters

Refmac parameters

Crystal parameters

Submit a remote job at the Hamburg Cluster

Submit the job for remote execution at the Hamburg cluster

Email address: victor@embl-hamburg.de

Data can be archived and made available to any software developer that requests them

Run

ARP/wARP web service

DISCLAIMER ARP/wARP LICENCE CCP4 LICENCE CITATIONS

Opening Page

Step 1

Step 2

Confirmation Page

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.

continue



Normal termination of wARPtrace

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

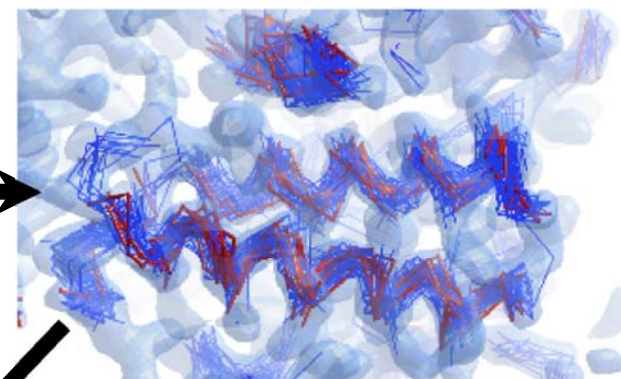
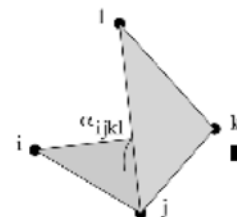
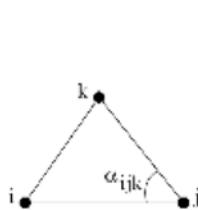
Image created using molecript Image created using P1N04



$\beta\alpha\lambda\beta\epsilon\sigma$

Modelling Secondary Structure

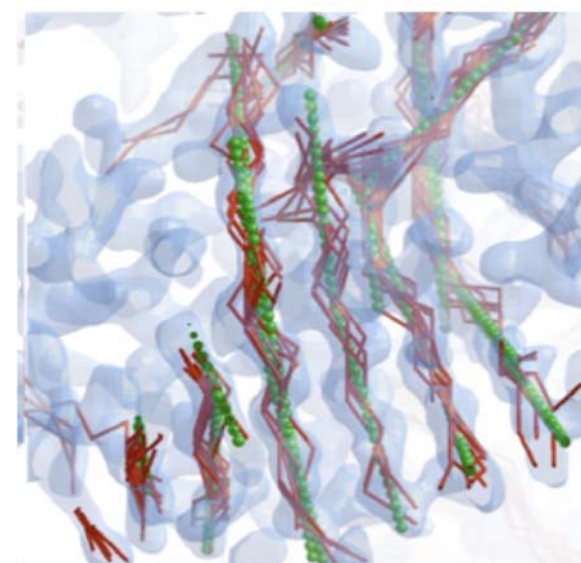
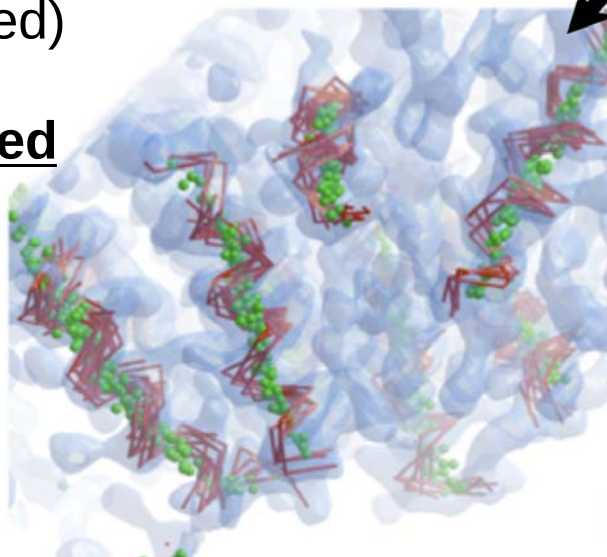
- Short helix/strand fragments (3 to 5 C α candidates) are built.

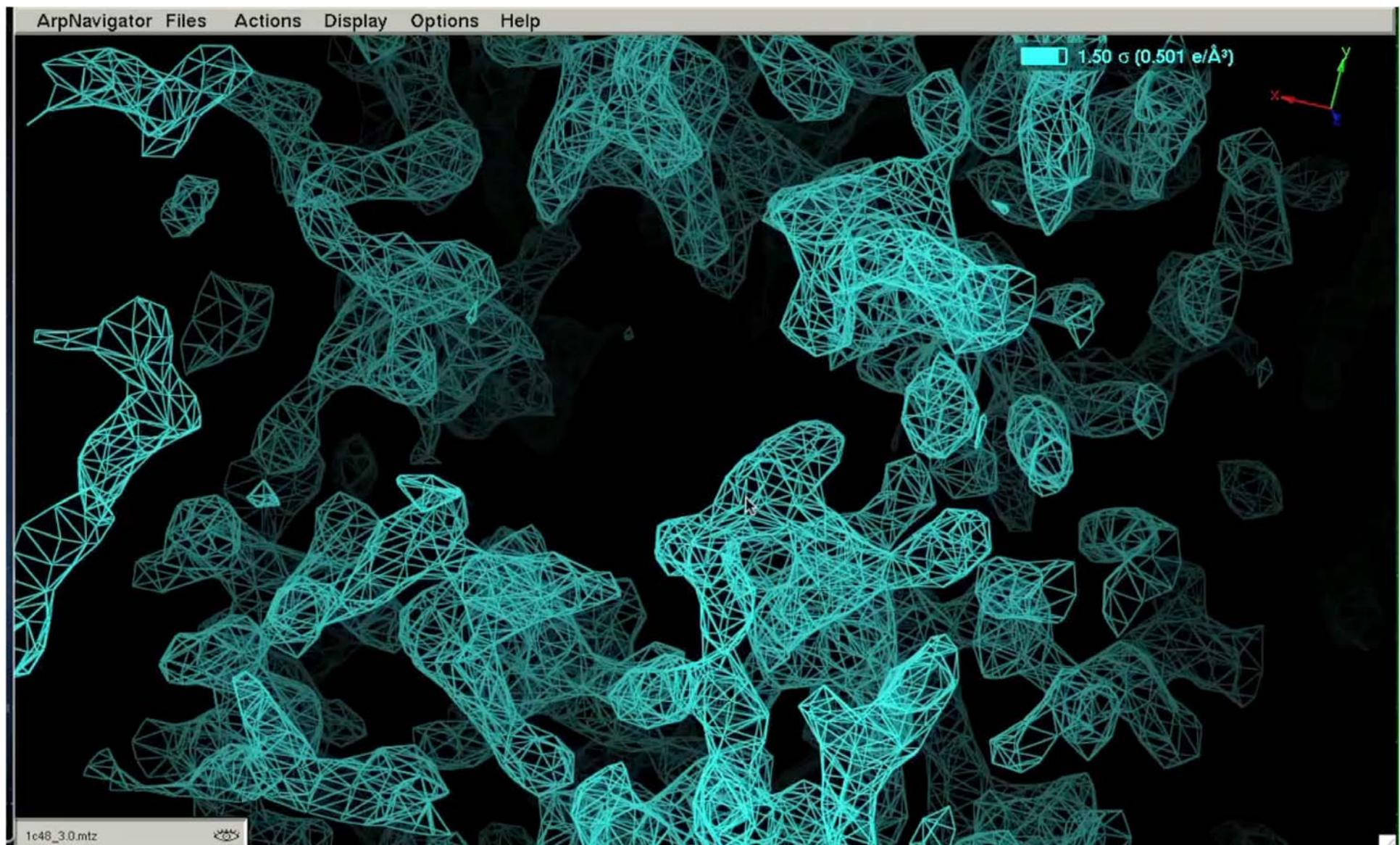


- Longer traces are formed or which the best are kept (in red)

- Traces are clustered

- Assemblies are averaged





Helices for a 350-residue (3.0 Å) protein can be built in under 5 seconds on a modern MacBookPro

The people



Developers

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

NKI Amsterdam: Krista Joosten, Tassos Perrakis

Collaborators

Santosh Panjikar, Garib Murshudov's group, Raj Pannu's group, the CCP4 team

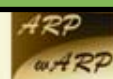
Former members

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

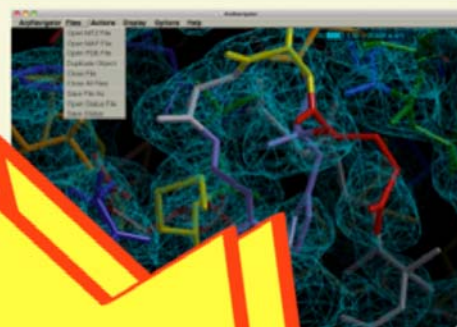
FAQ: Where Can I Get This?



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



by European Molecular Biology Laboratory



ARP/wARP: Crystallographic Macromolecular Model Building

Builds proteins chains, loops,

Download available
and Linux.

Download/L

MAILING LISTS

[Subscribe to ARP/wA](#)

[Unsubscribe from AR](#)

[message to A](#)

!!! Tutorials !!!

