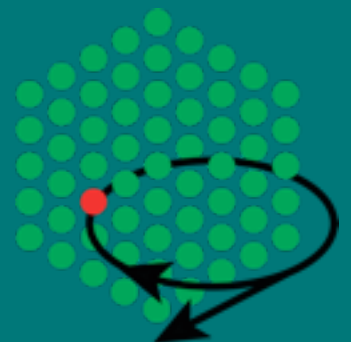


Automated Protein Model Building with ARP/wARP version 7.3

Tim Wiegels
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 for PhD and postdoctoral job opportunities

The EMBL Units

EBML 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”)

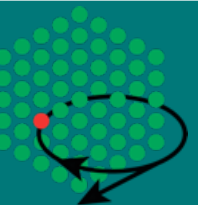
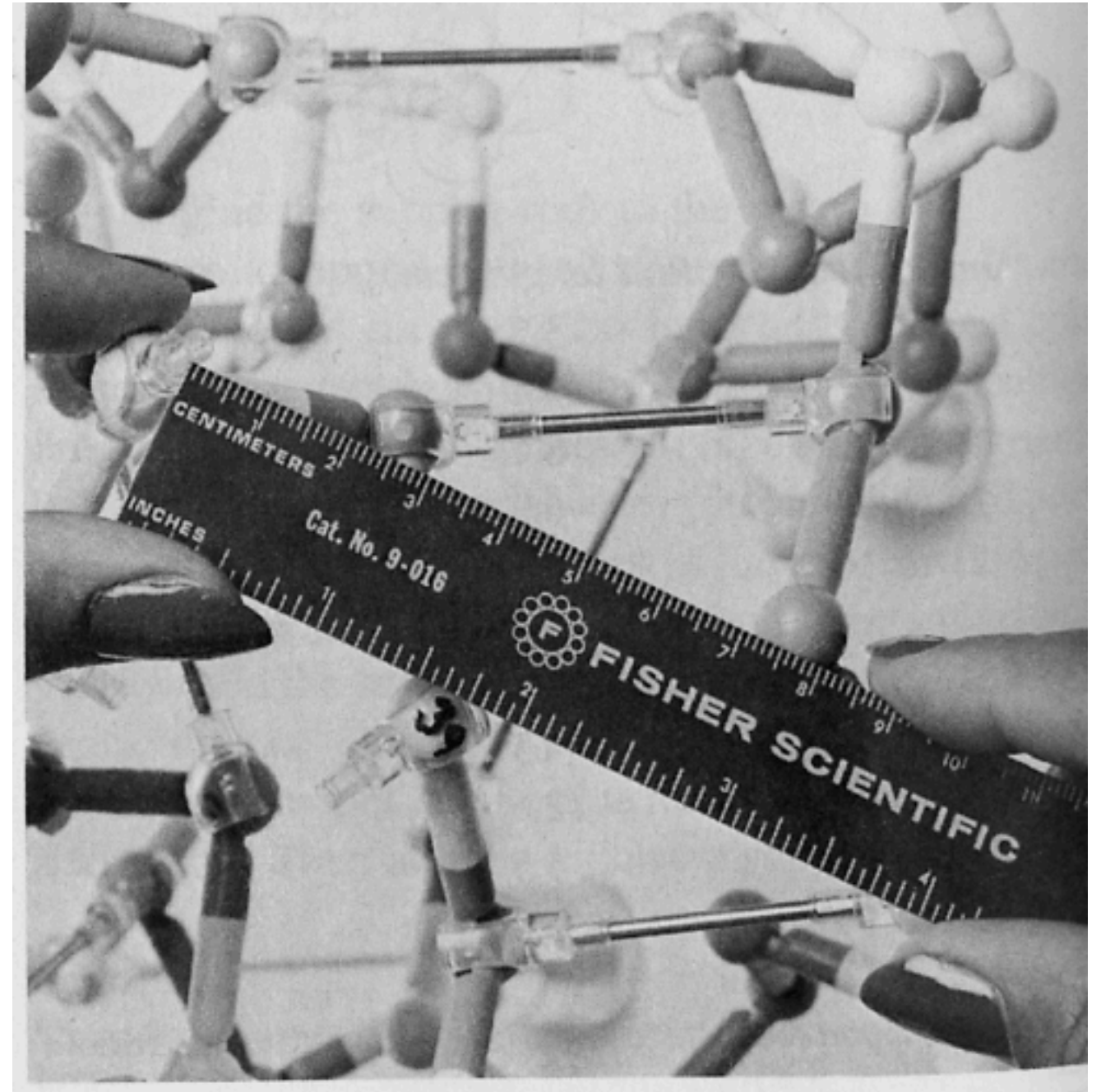
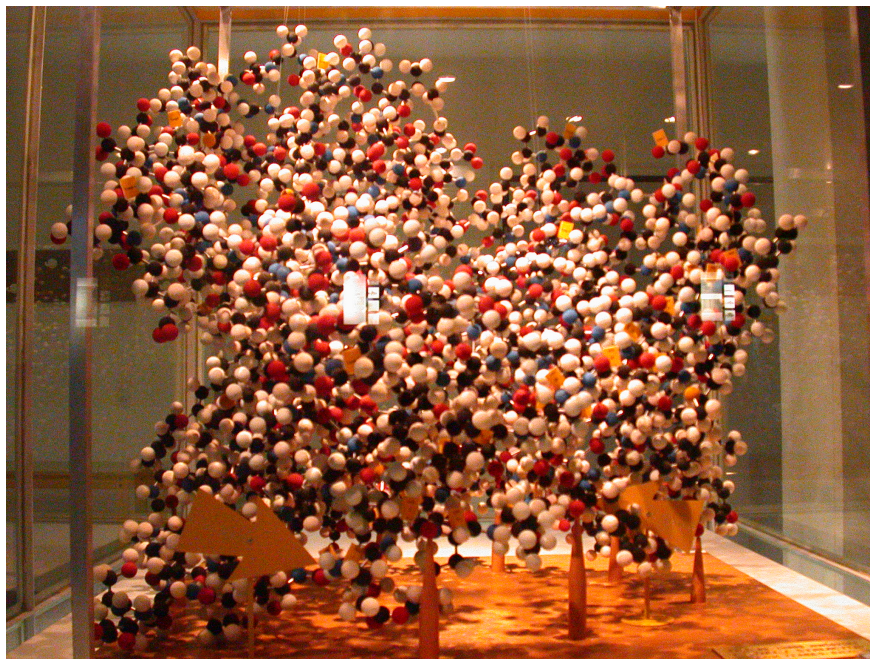
- Bioinformatics

EMBL Hamburg

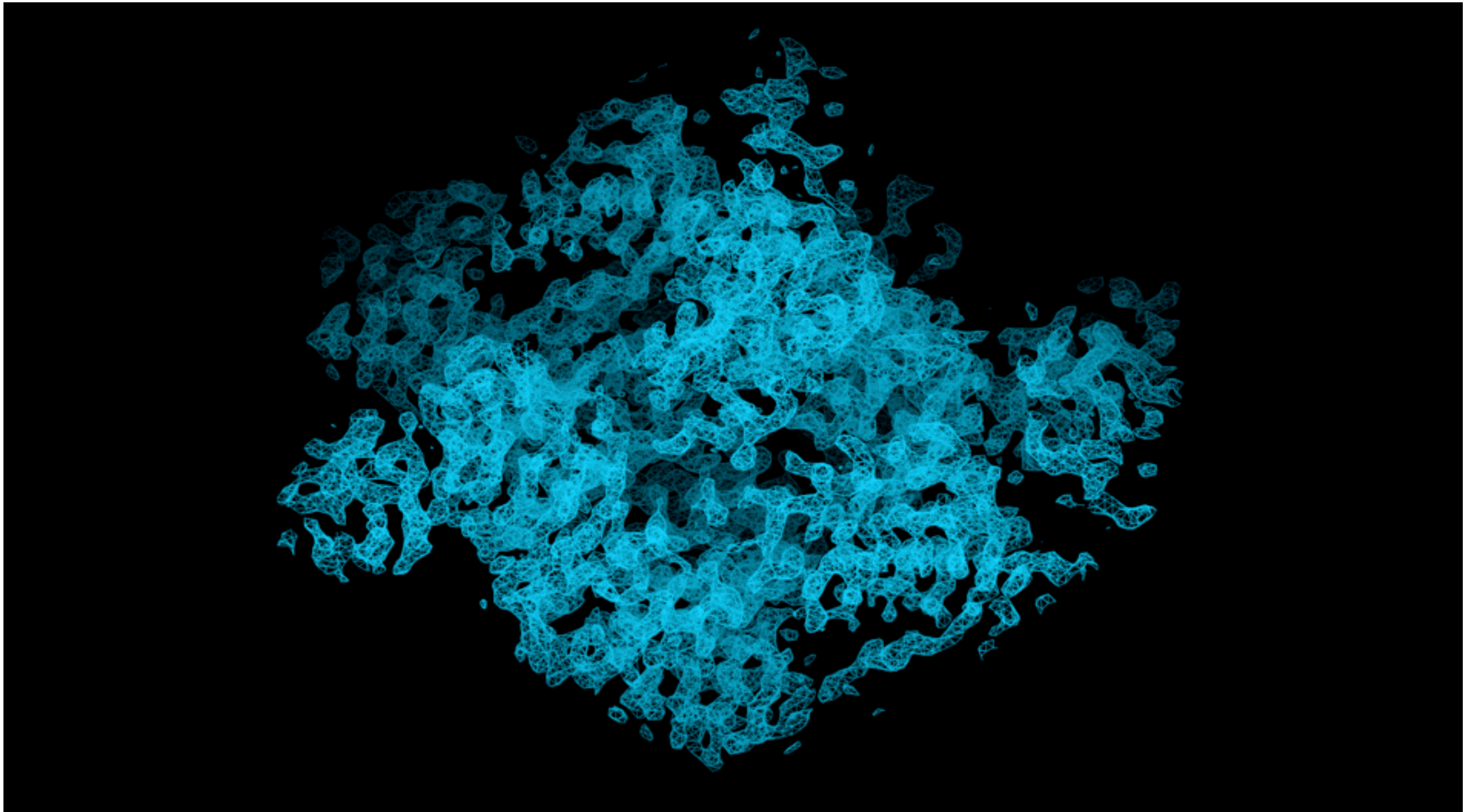
- Structural Biology



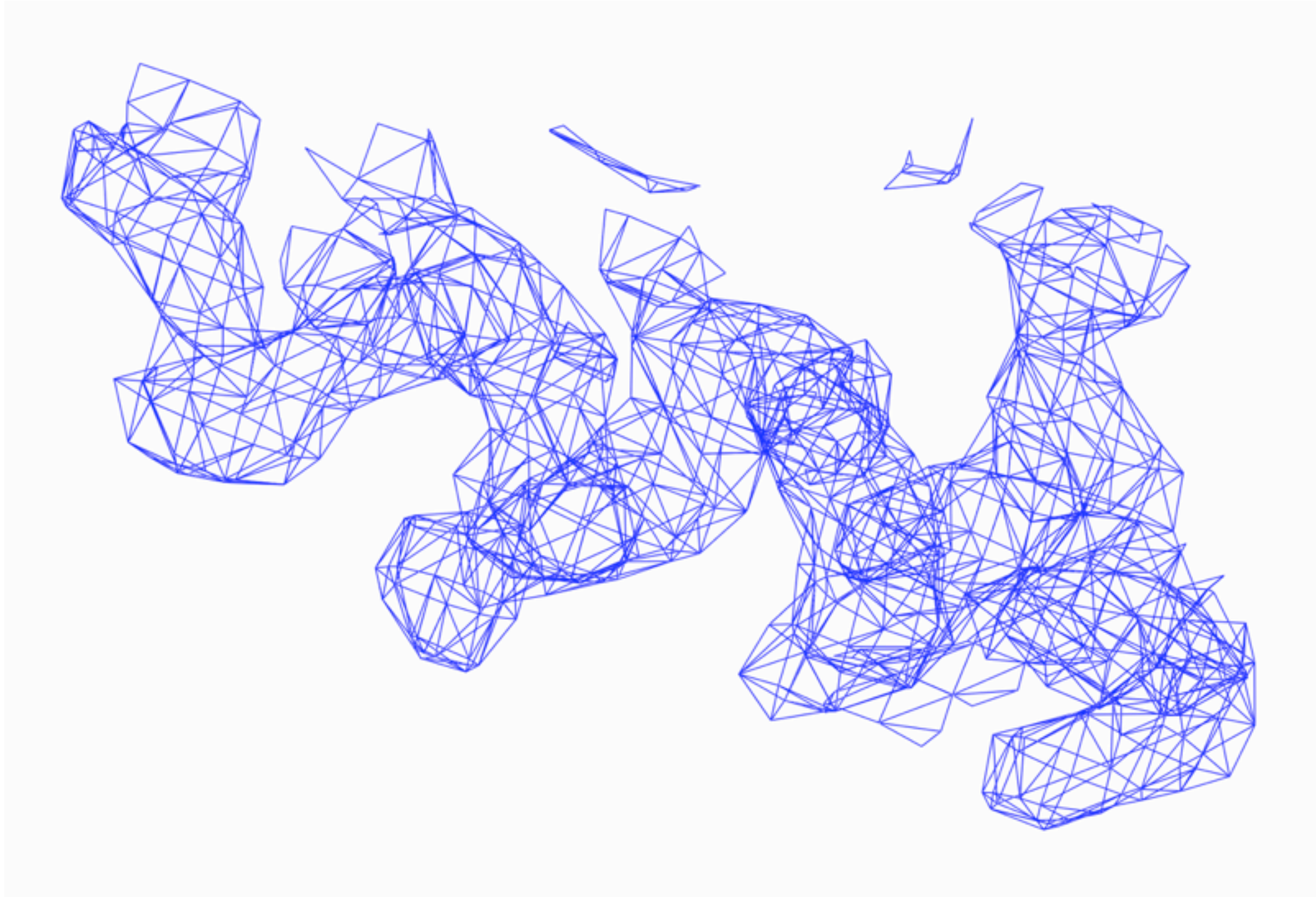
First map interpretations



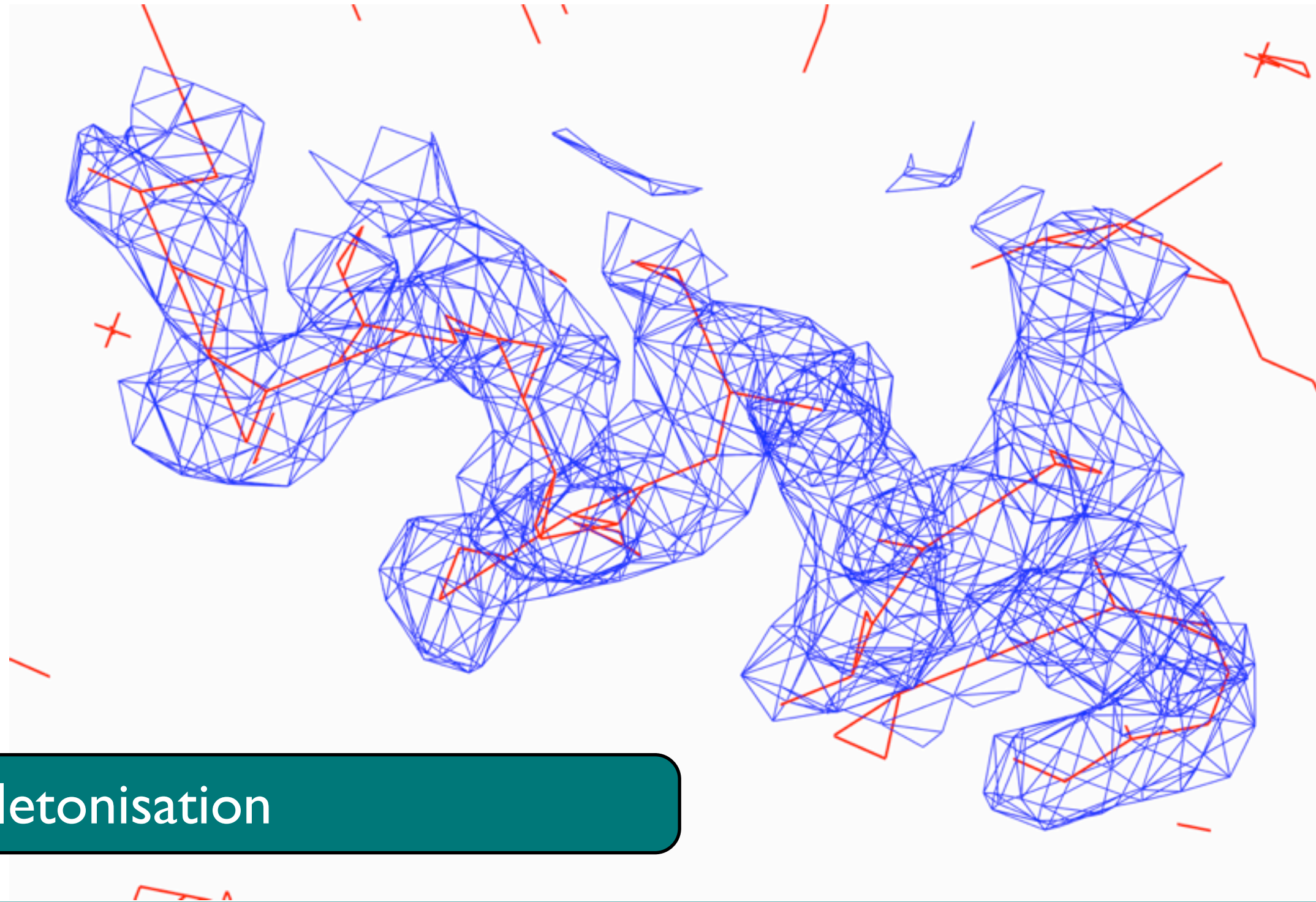
Interprete This Map!



Model Building: Approaches



Model Building: Approaches

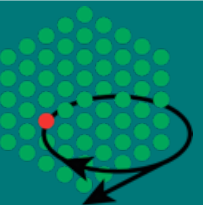


Skeletonisation

Model Building: Approaches



Skeletonisation

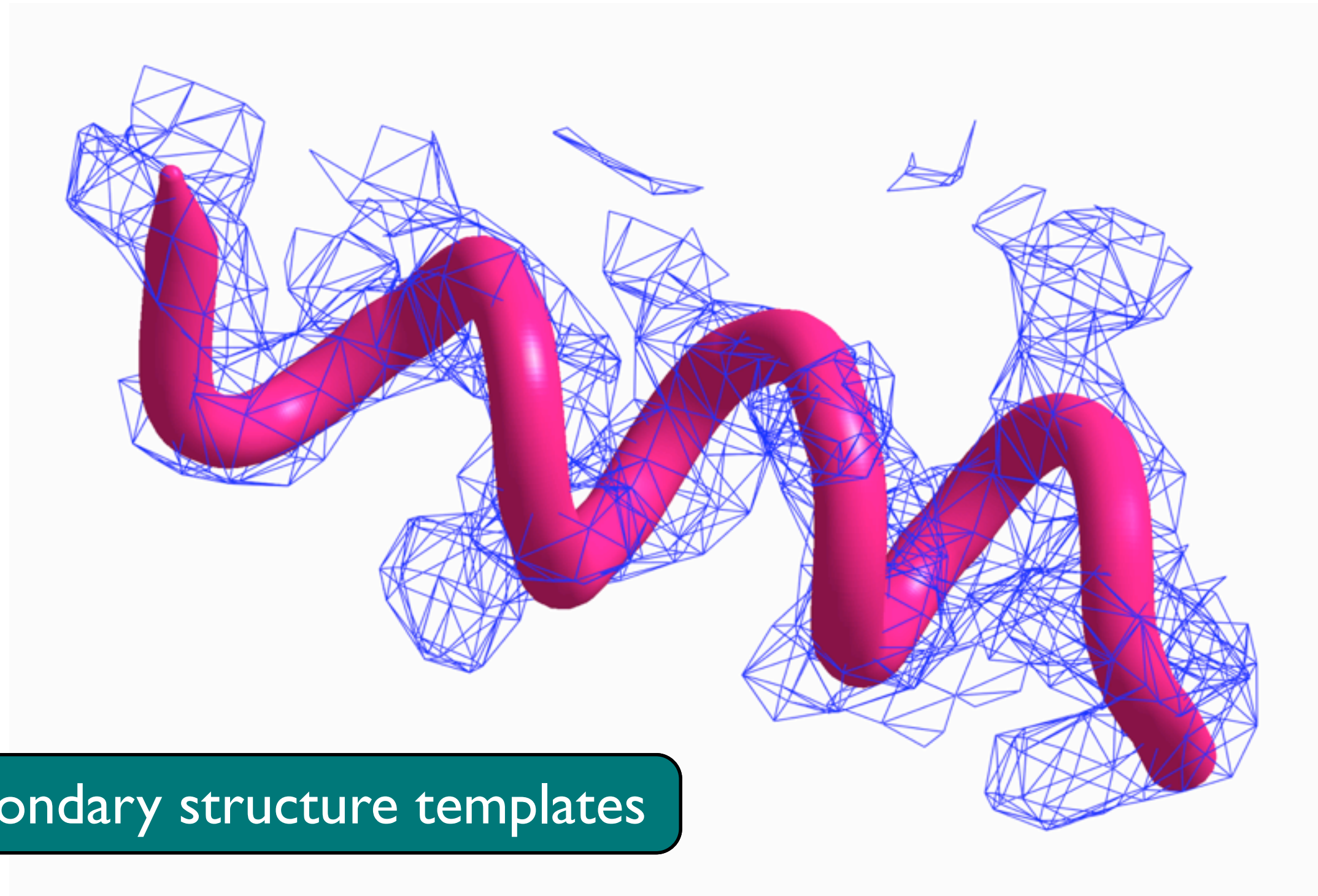


Model Building: Approaches



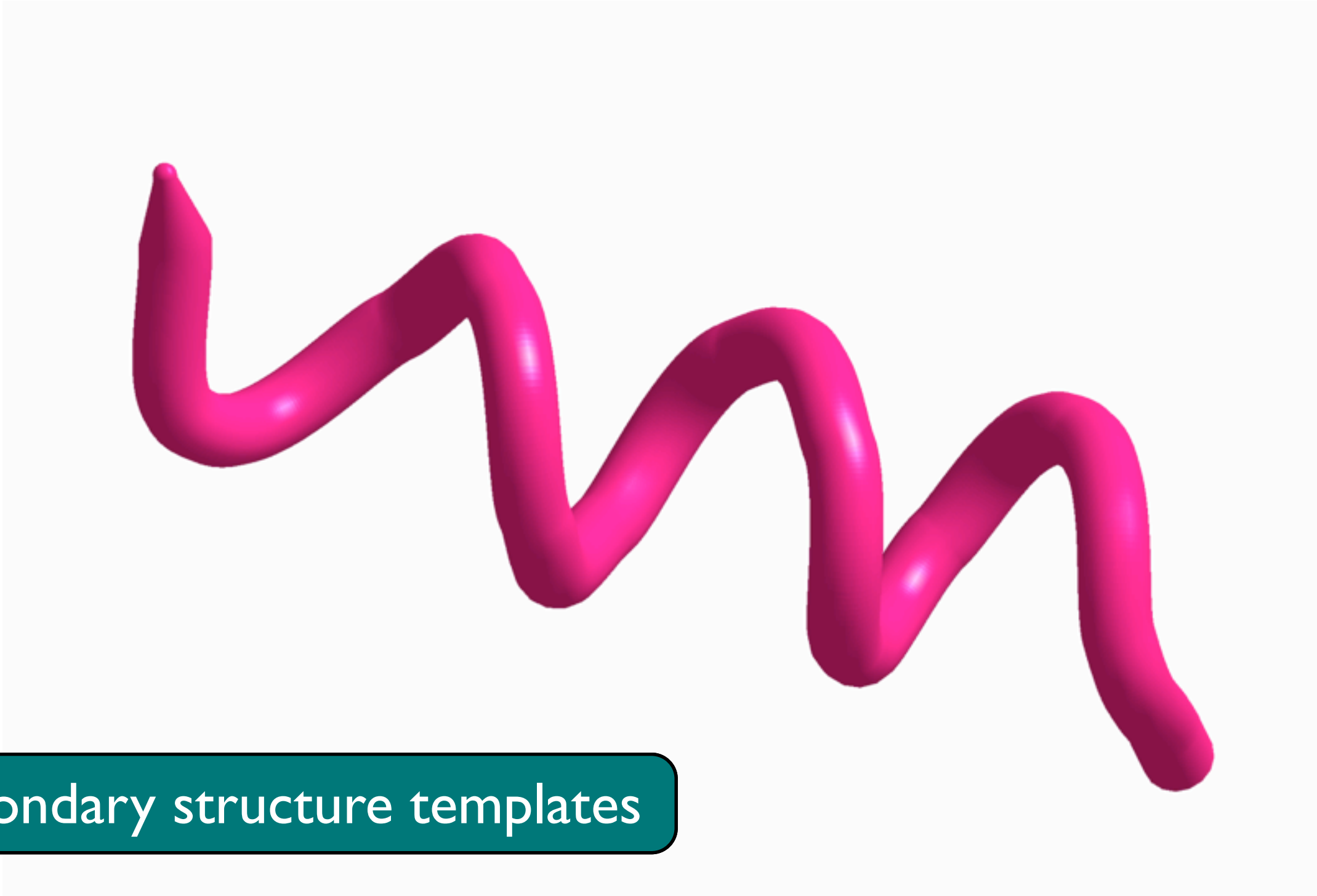
Skeletonisation

Model Building: Approaches

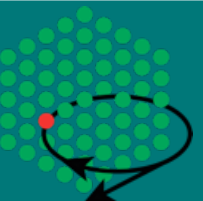


Secondary structure templates

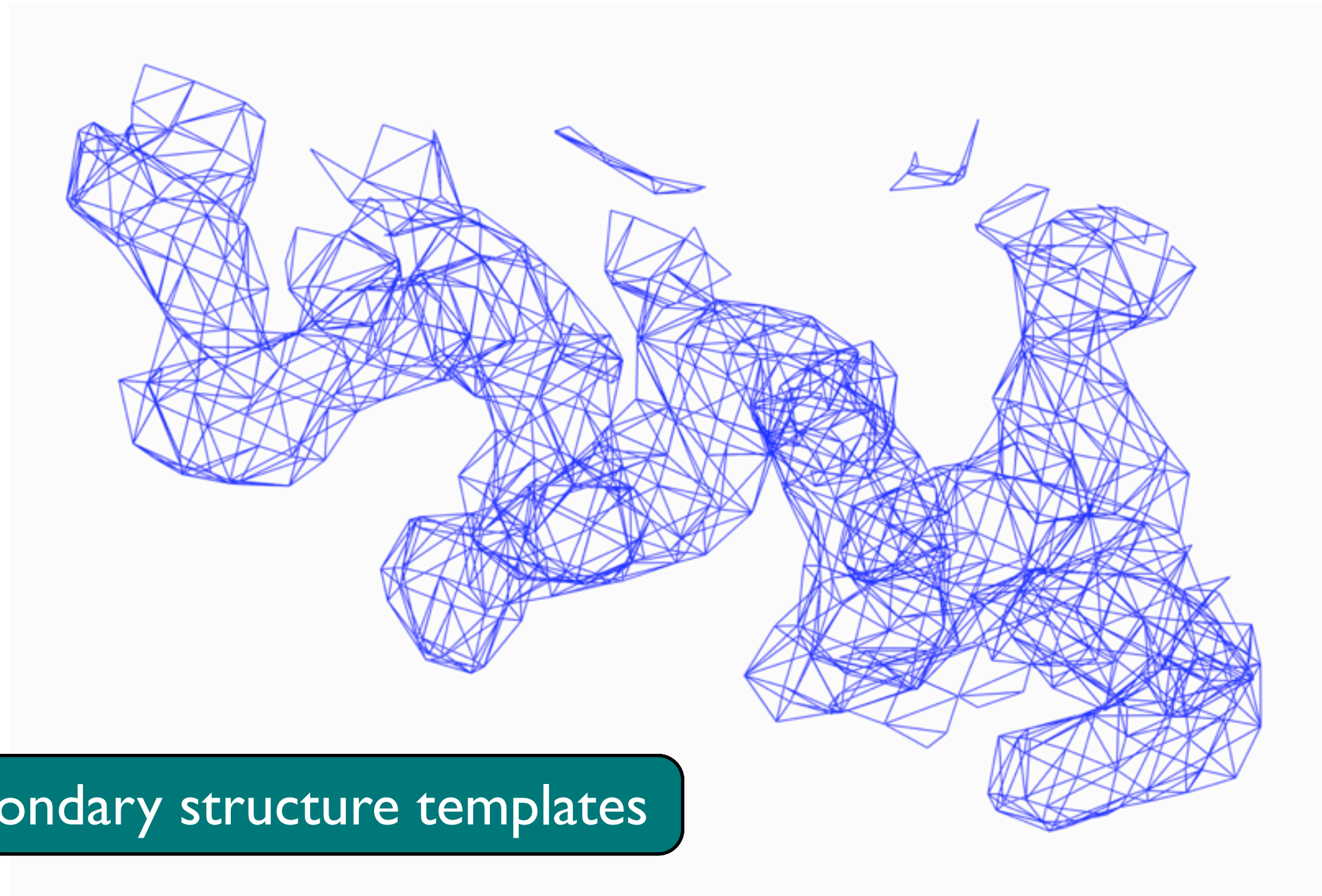
Model Building: Approaches



Secondary structure templates

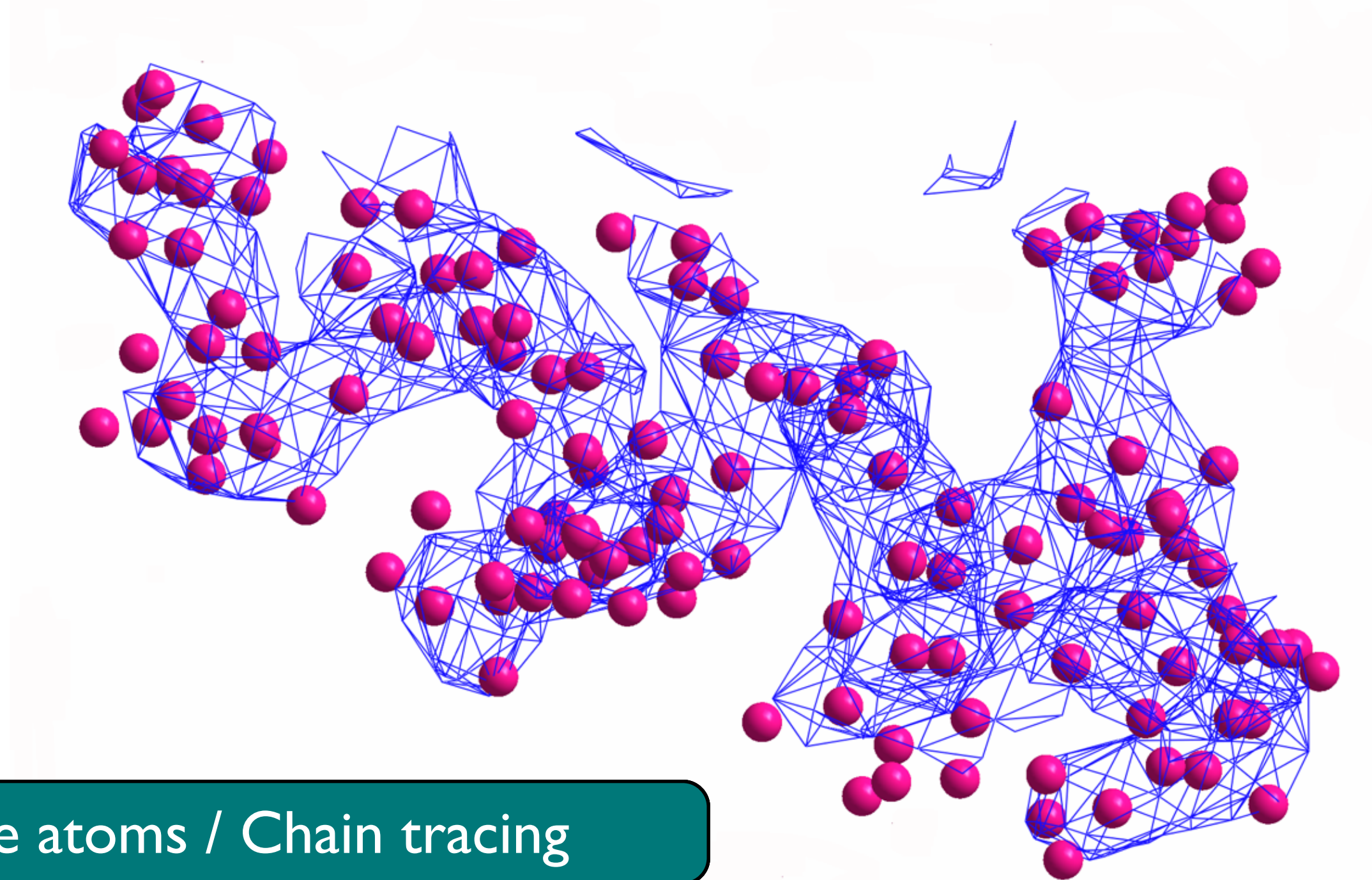


Model Building: Approaches



Secondary structure templates

Model Building: Approaches

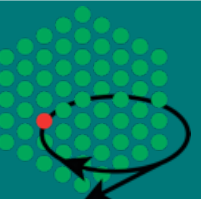


Free atoms / Chain tracing

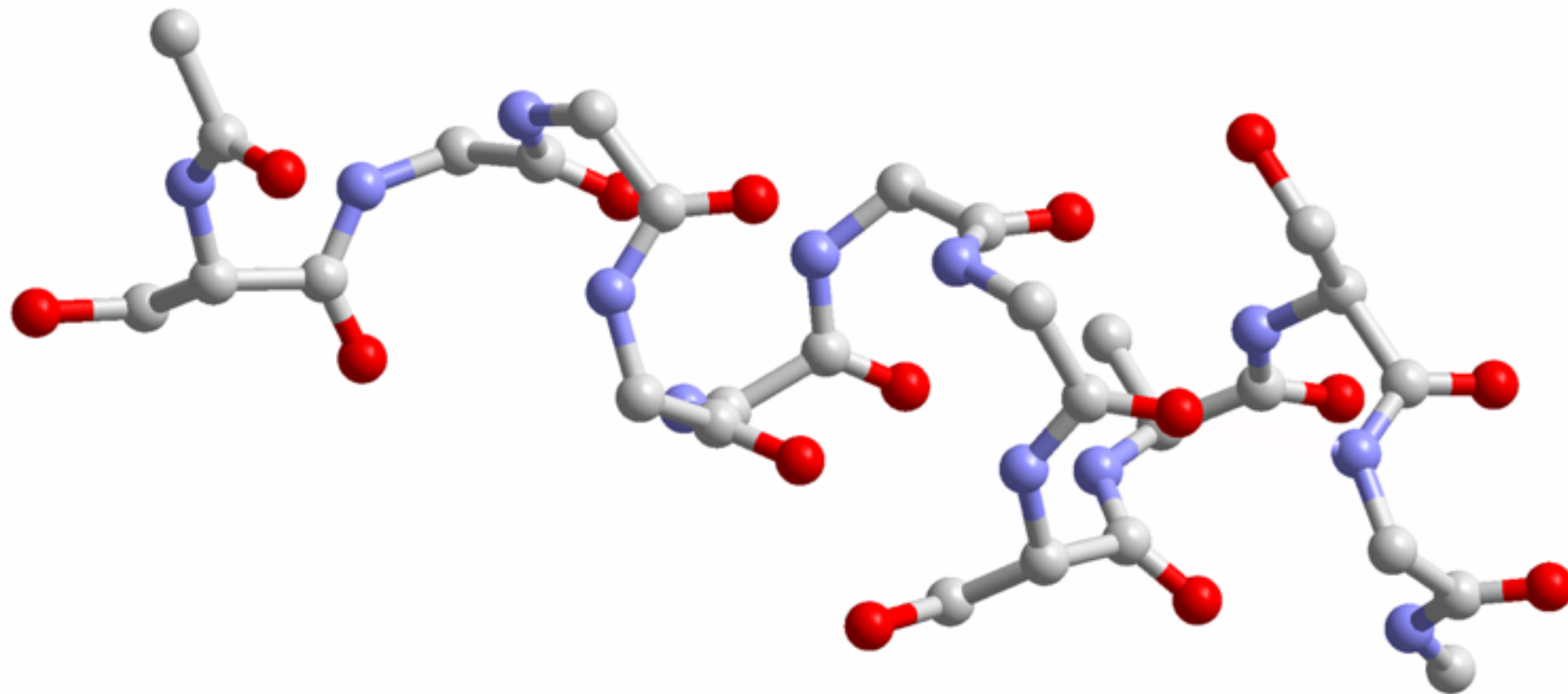
Model Building: Approaches



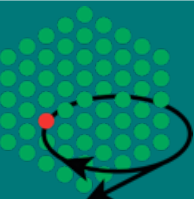
Free atoms / Chain tracing



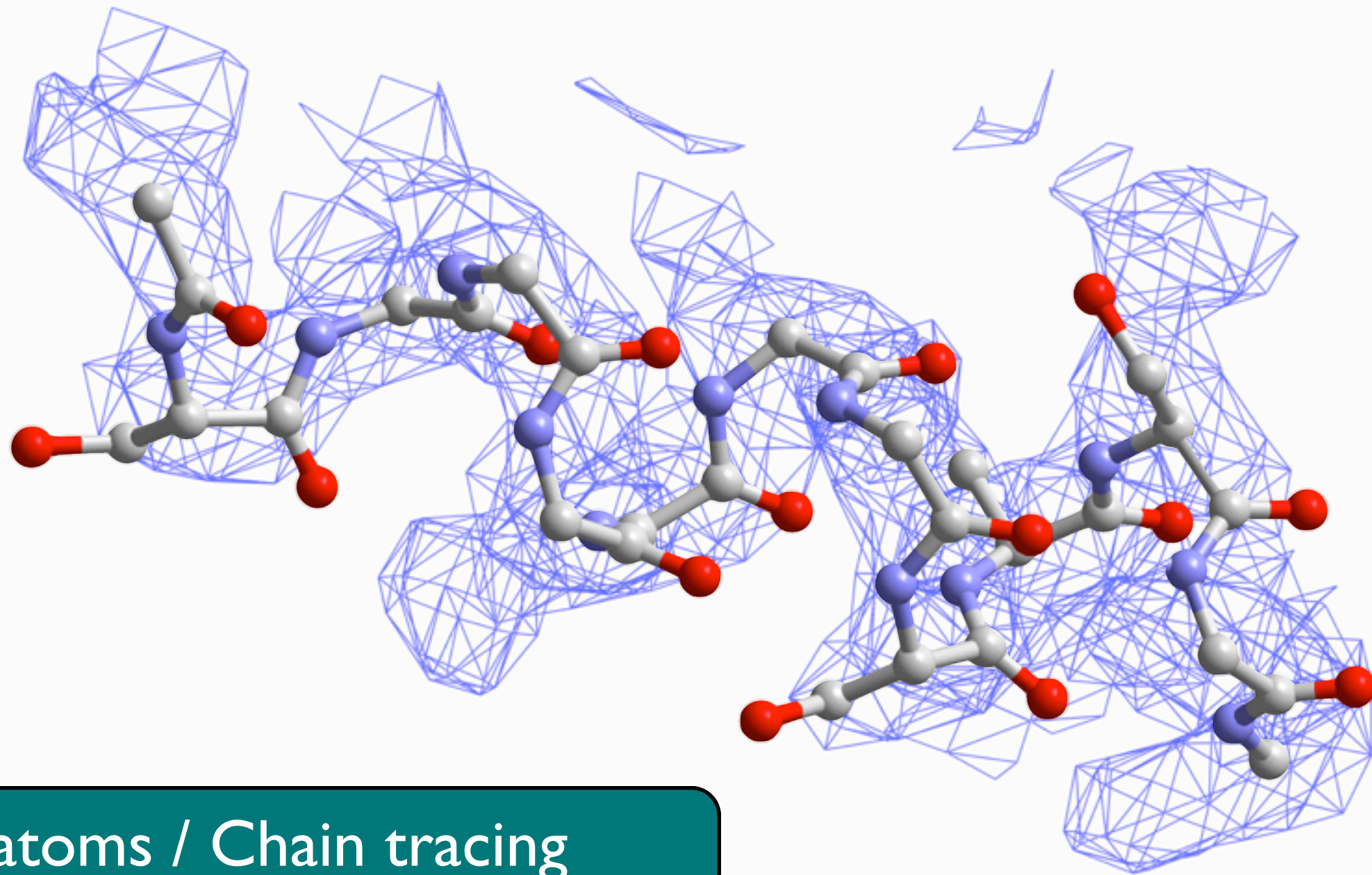
Model Building: Approaches



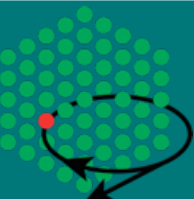
Free atoms / Chain tracing



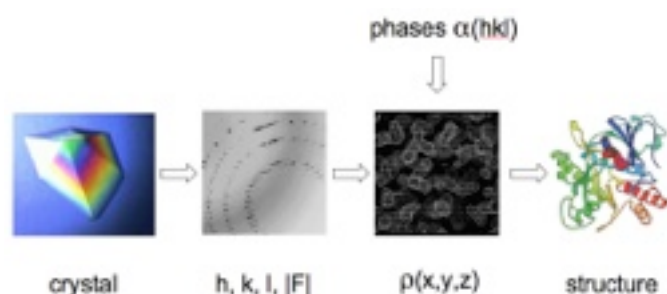
Model Building: Approaches



Free atoms / Chain tracing



Methods for Building Protein Structure



Skeletonisation:

Greer (1974)

Core-tracing - Swanson (1994)

Mol. scene analysis - Fortier *et al.*, (1997)



Molecular-graphics packages:

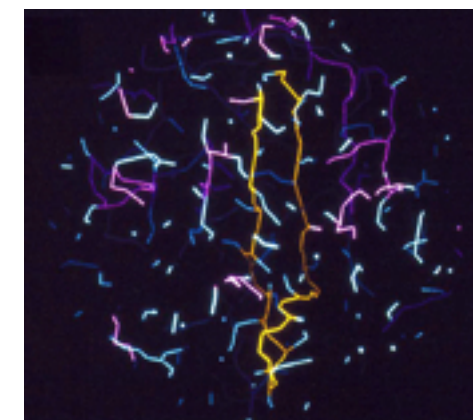
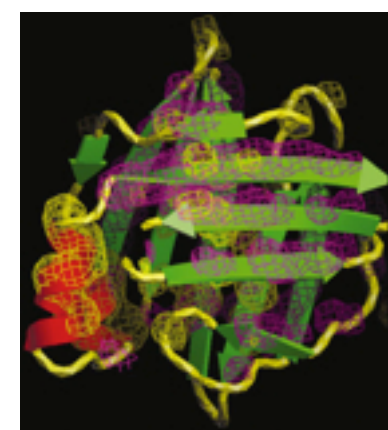
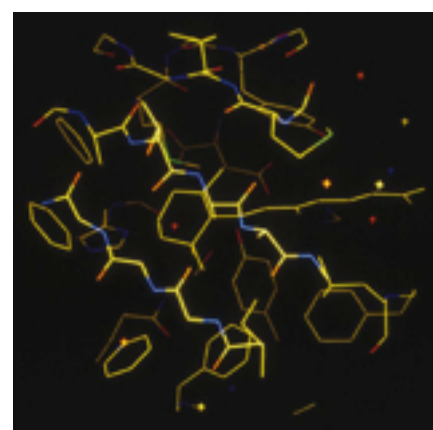
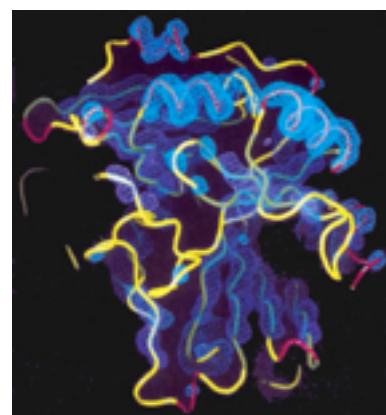
FRODO - Jones (1978)

O - Jones *et al.*, (1991)

Secondary structure:

an exhaustive search
by **ESSENS** -

Kleywegt and Jones,
(1997)

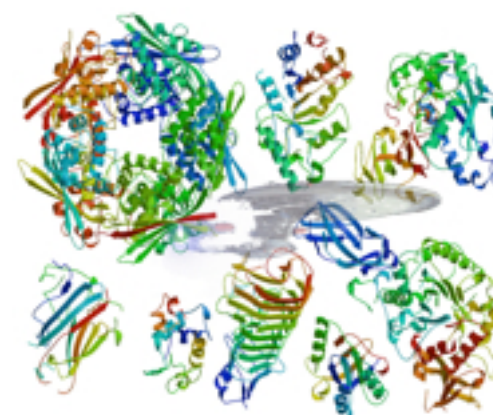


Chain tracing:

ARP/wARP – Perrakis *et al.*, (1999)

Concept of '**free atoms**' – Agarwal and Isaacs (1985)

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



TEXTAL/CAPRA - Ioerger *et al.*, (1999)

CONFMATCH - Wang (2000)

RESOLVE – Terwilliger (2000)

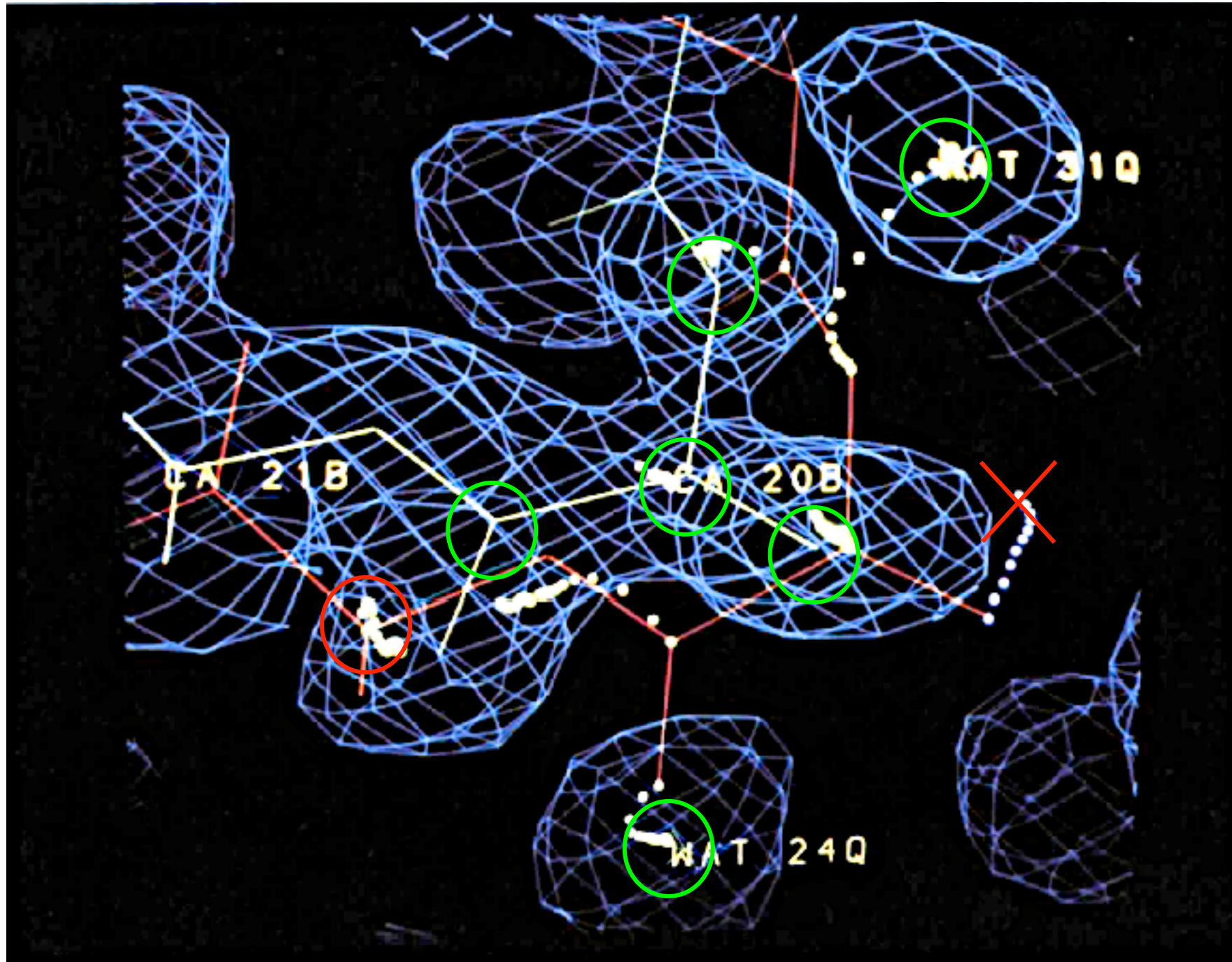
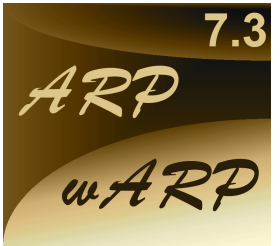
MAID - Levitt, (2001)

FSM - Pavelcik (2003)

LAFIRE - Yao *et al.*, (2006)

BUCCANEER - Cowtan (2006)

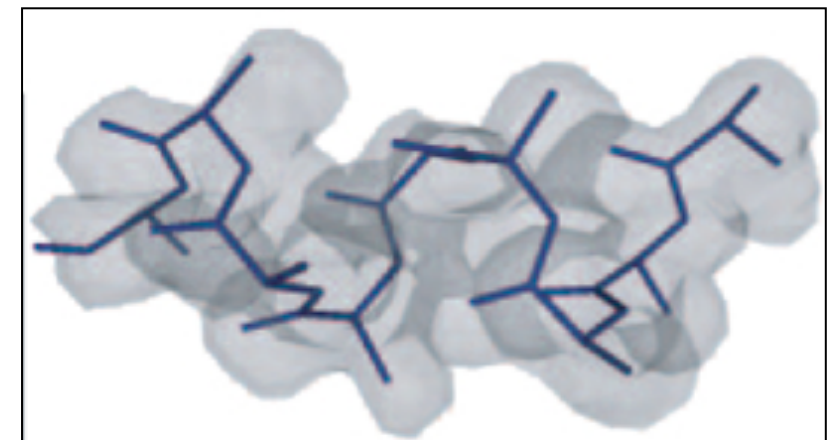
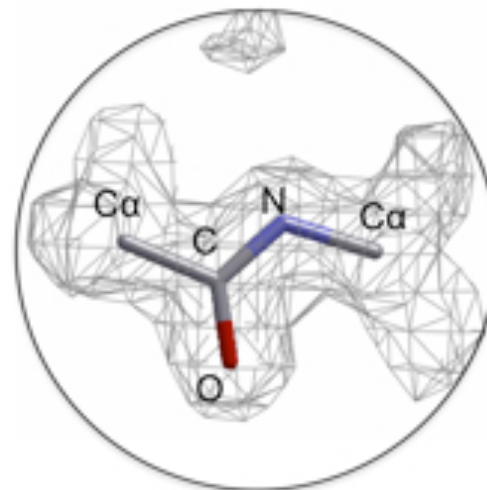
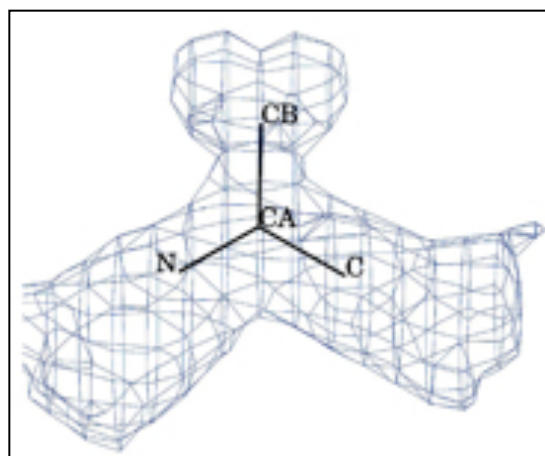
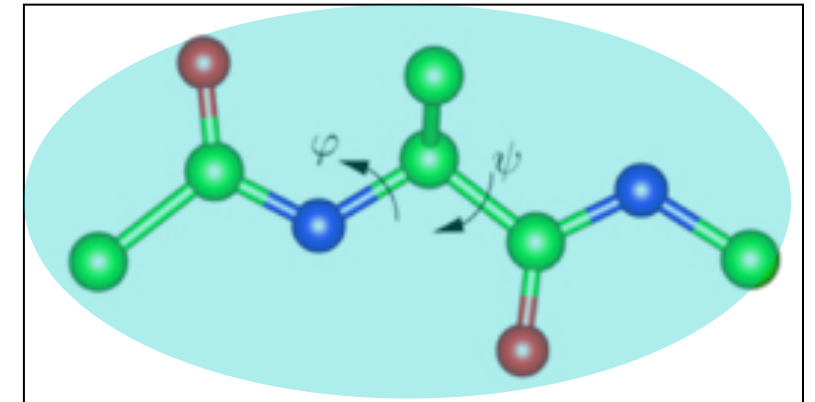
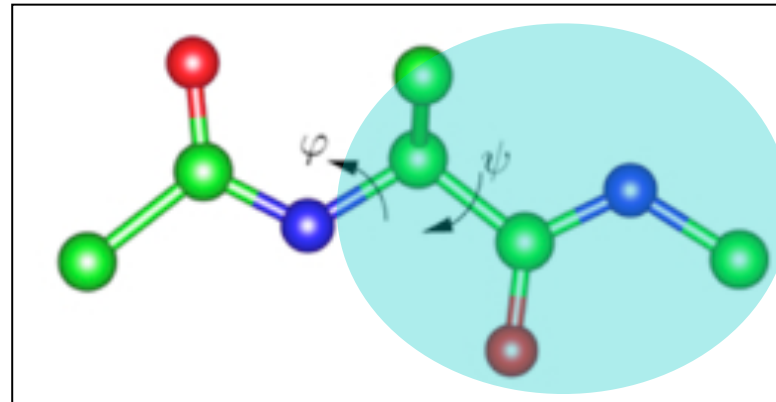
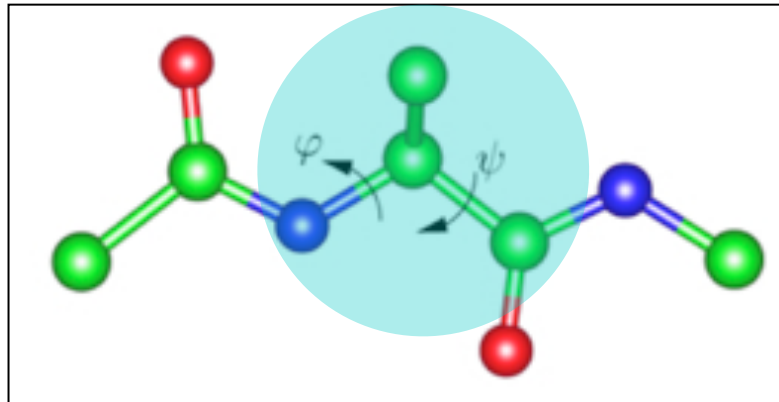
What is this “ARP/wARP”?



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

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

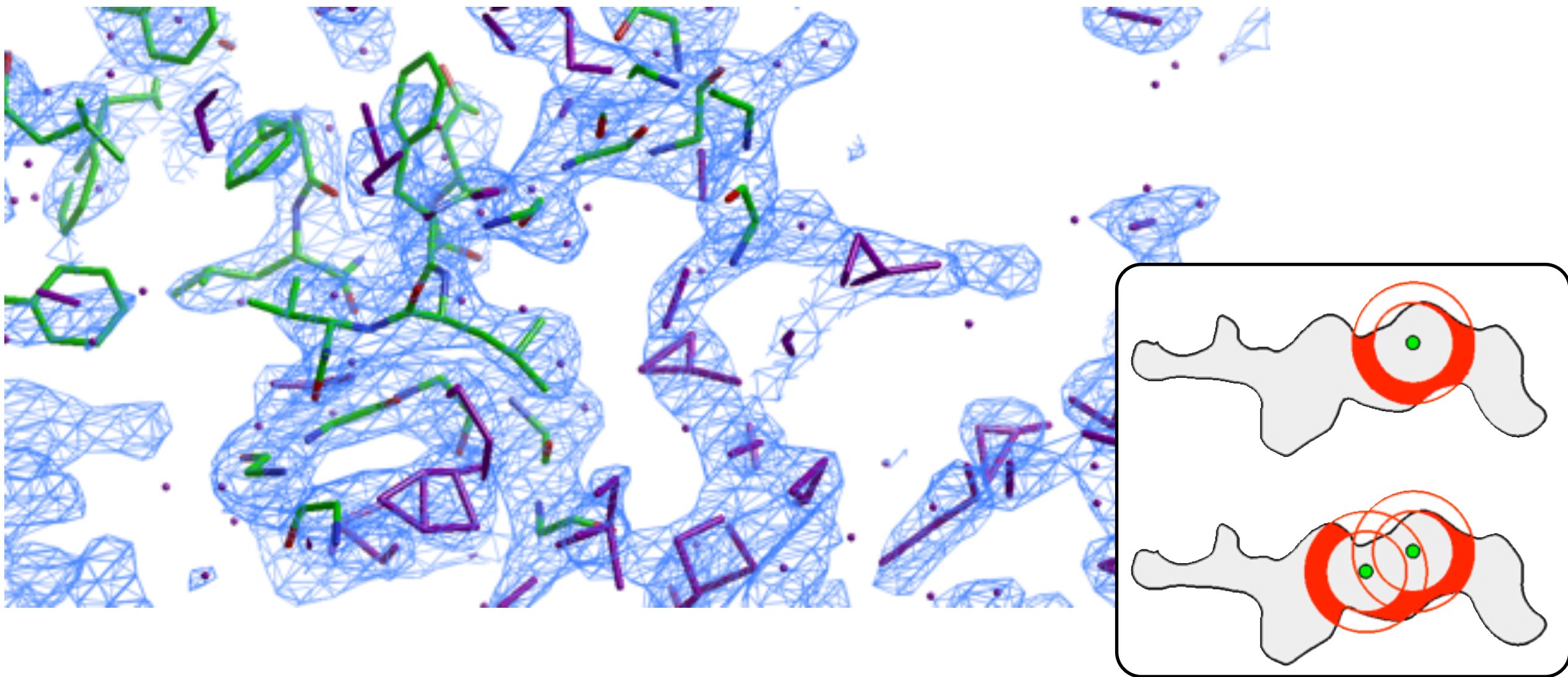
Methods for Building Protein Structure



TEXTAL / Buccaneer	ARP/wARP	RESOLVE / ACMI
C α	(Di-) Peptides	Fragments

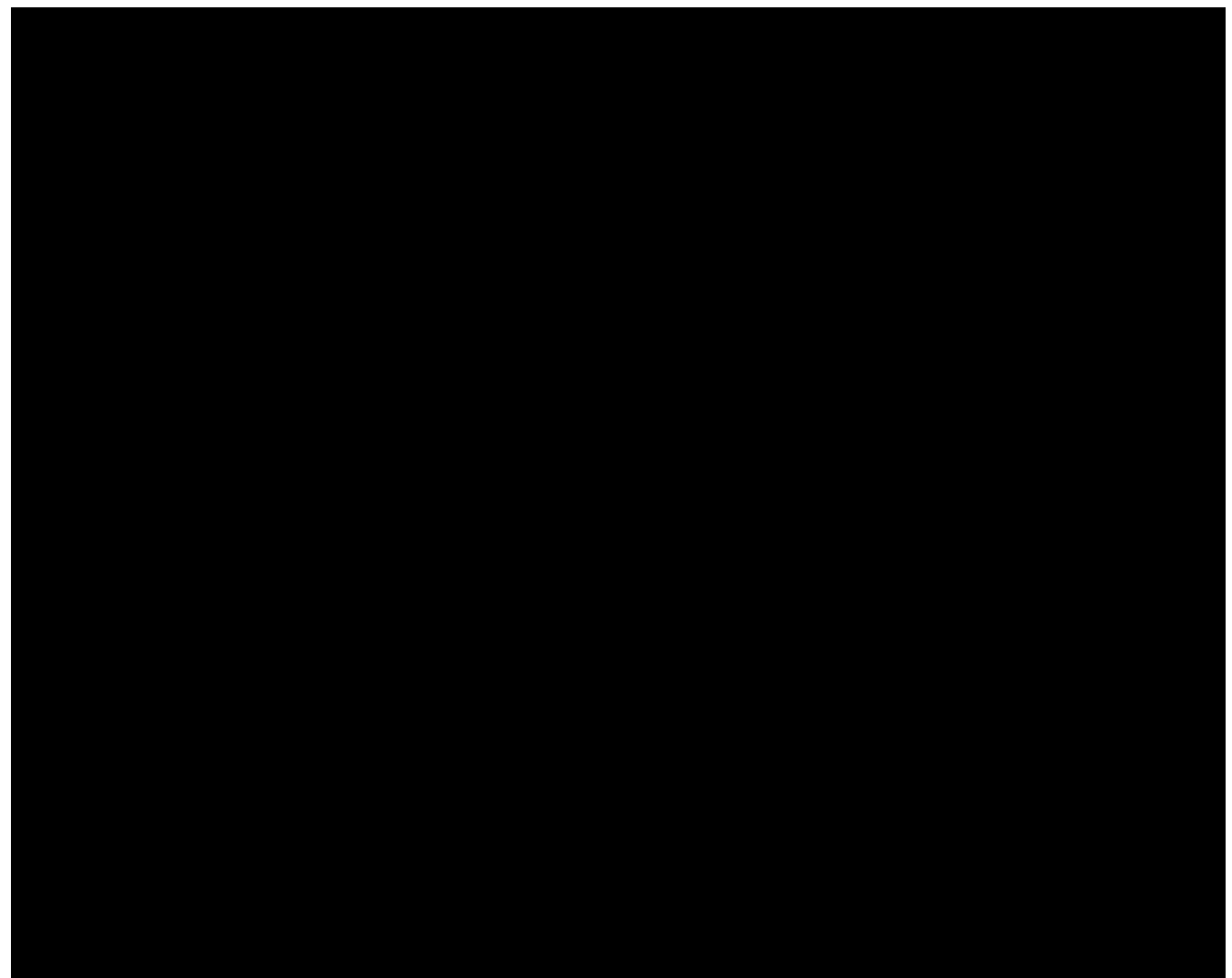
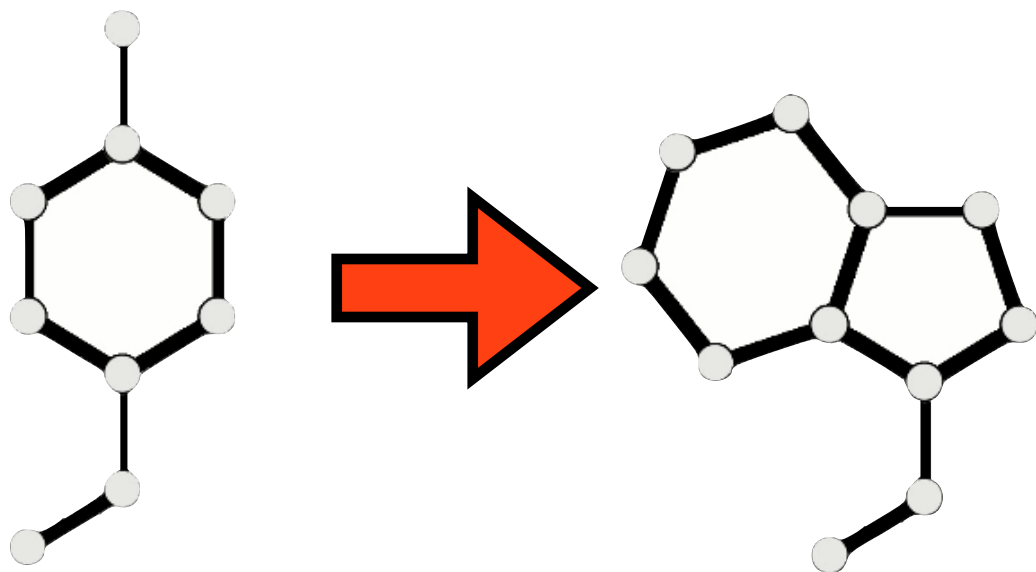
The Hybrid Model

- **Partial model** is used together with a **free-atom** model
- Chemically assigned parts provide **restraints** for refinement

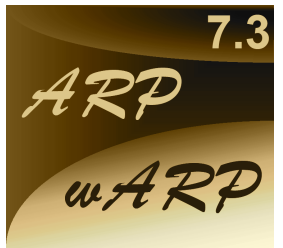


The Atom Update

- **Removal** and **addition** of atoms based on **density at atomics centers**



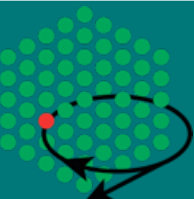
The fundamental ARP/wARP concept



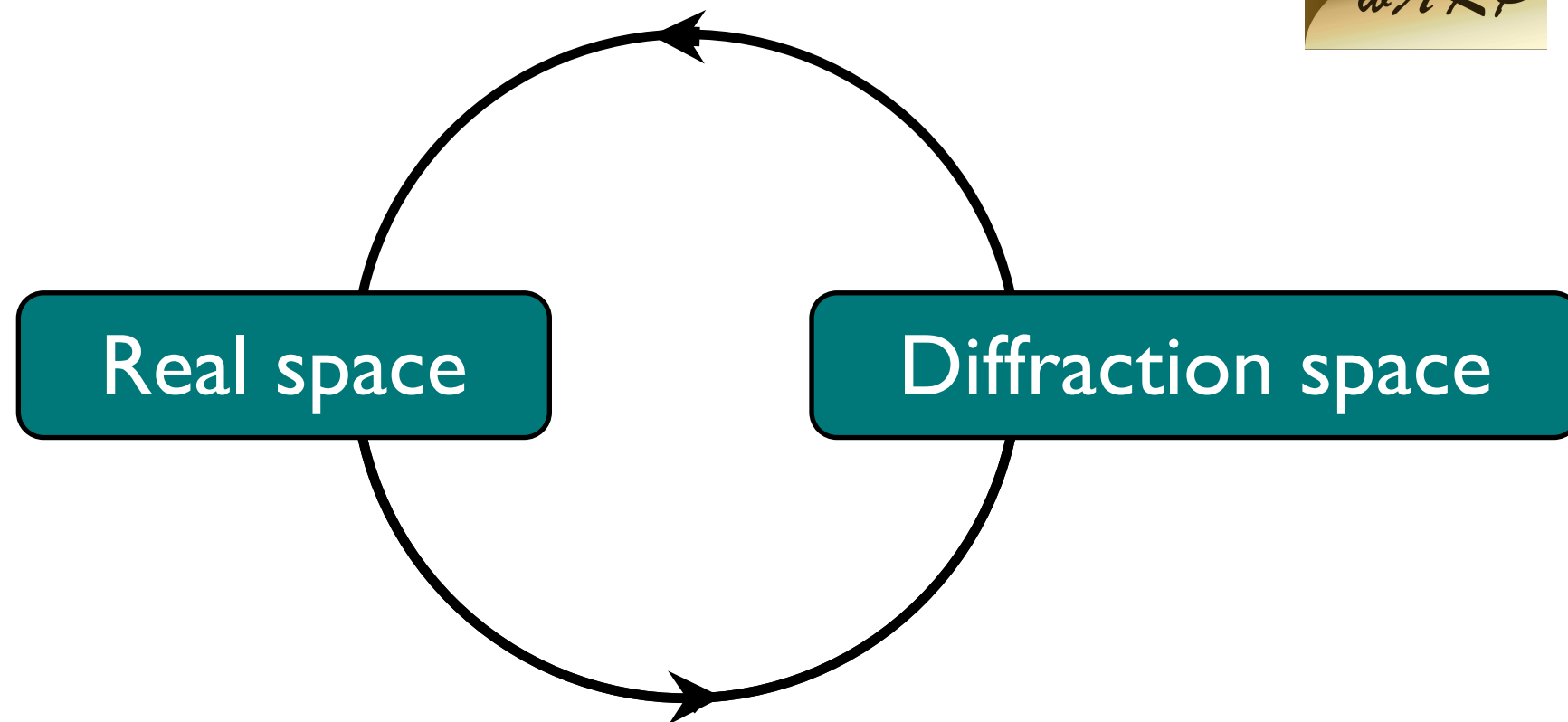
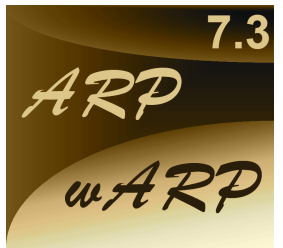
Real space

Real space

- **Hybrid model**: atoms having chemical identity and free atoms
- **Model update**: removing and adding parts of the model



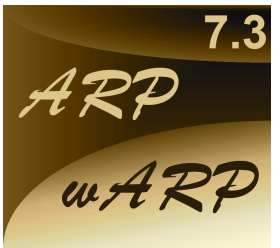
The fundamental ARP/wARP concept



Diffraction space

- Unrestrained refinement of free parts of the model
- Restrained refinement of chemically assigned atoms

The fundamental ARP/wARP concept



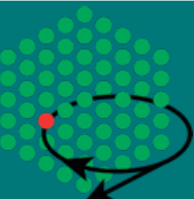
Pattern space

Real space

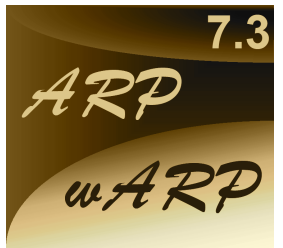
Diffraction space

Pattern space

- **Electron density**: local object interpretation
- **Hybrid Model**: local motif interpretation



The fundamental ARP/wARP concept



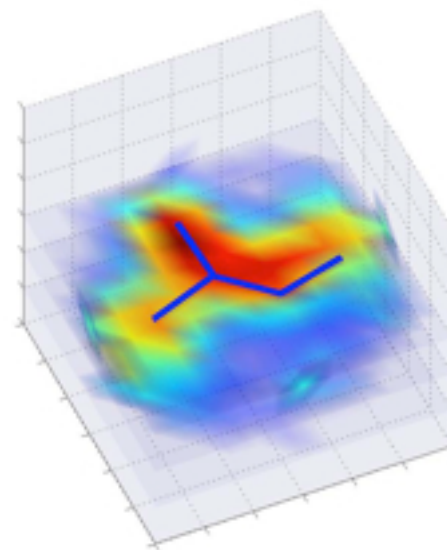
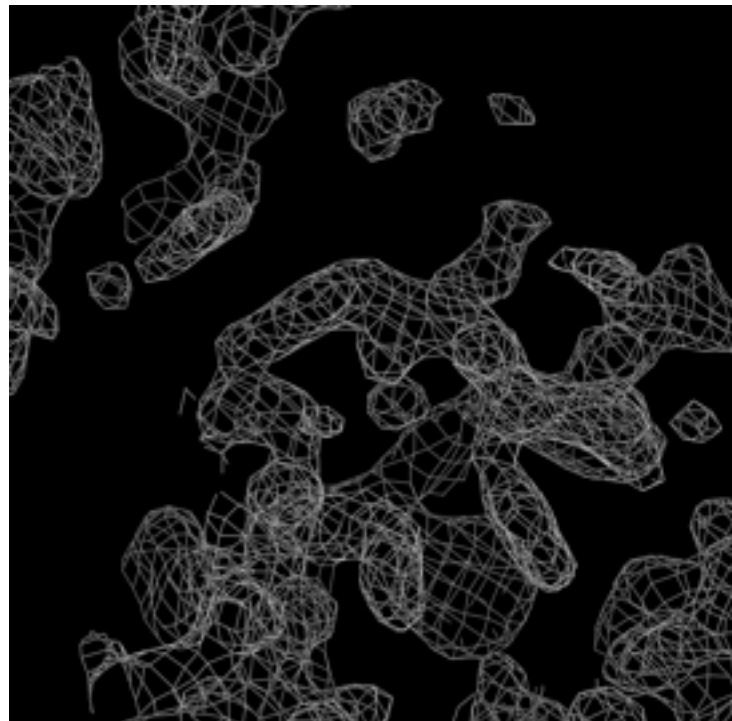
Pattern space

Real space

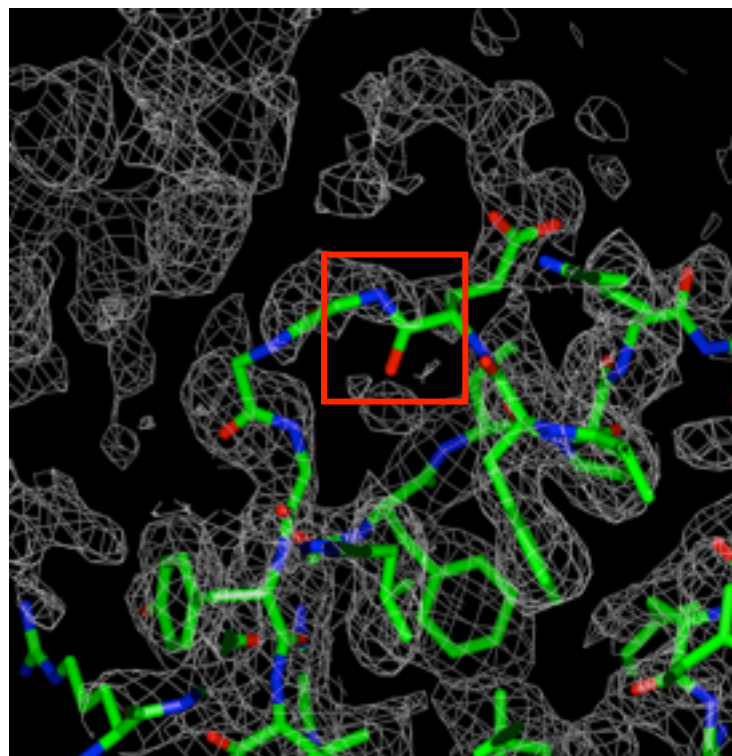
Diffraction space

- forming **unified process of model building and refinement**
- Scheme of restraints and free atoms are **iteratively updated**
- **Hybrid model** is **converging** to the final model

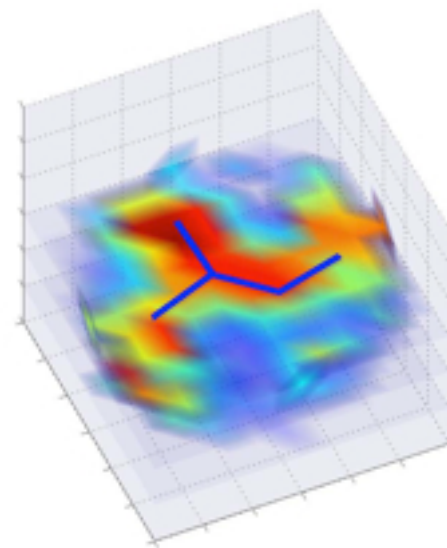
Autotracing: Recognising peptide units



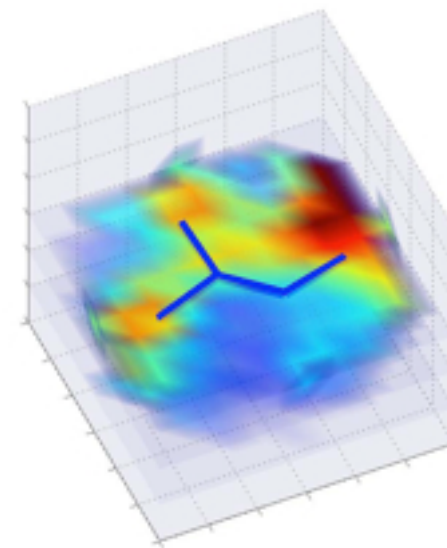
The task:
Separating **true** peptides
from **false** ones.



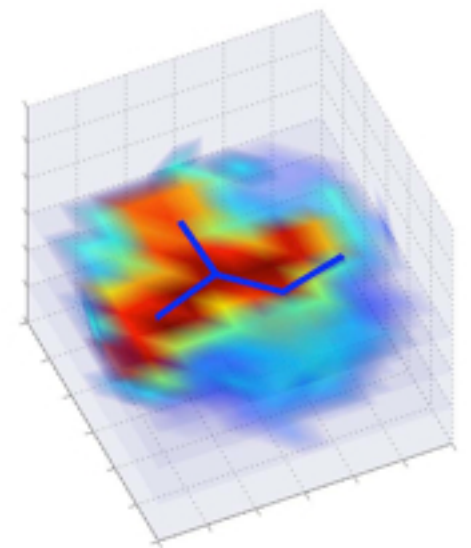
peptide
good fit



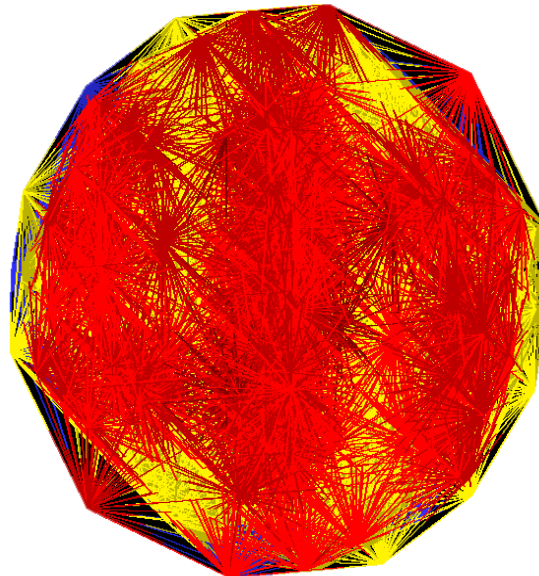
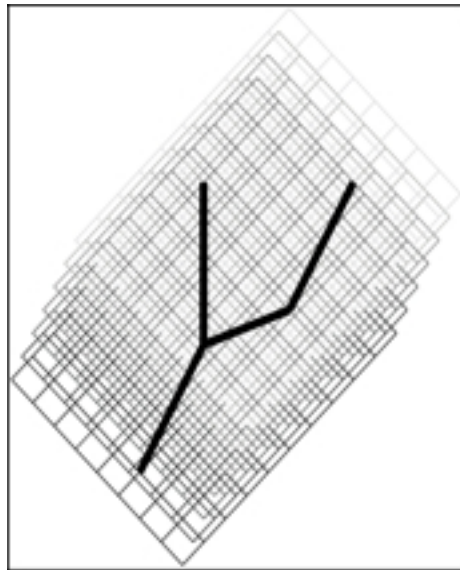
peptide
but poor fit



non-peptide
but good fit



Autotracing: Recognising peptide units

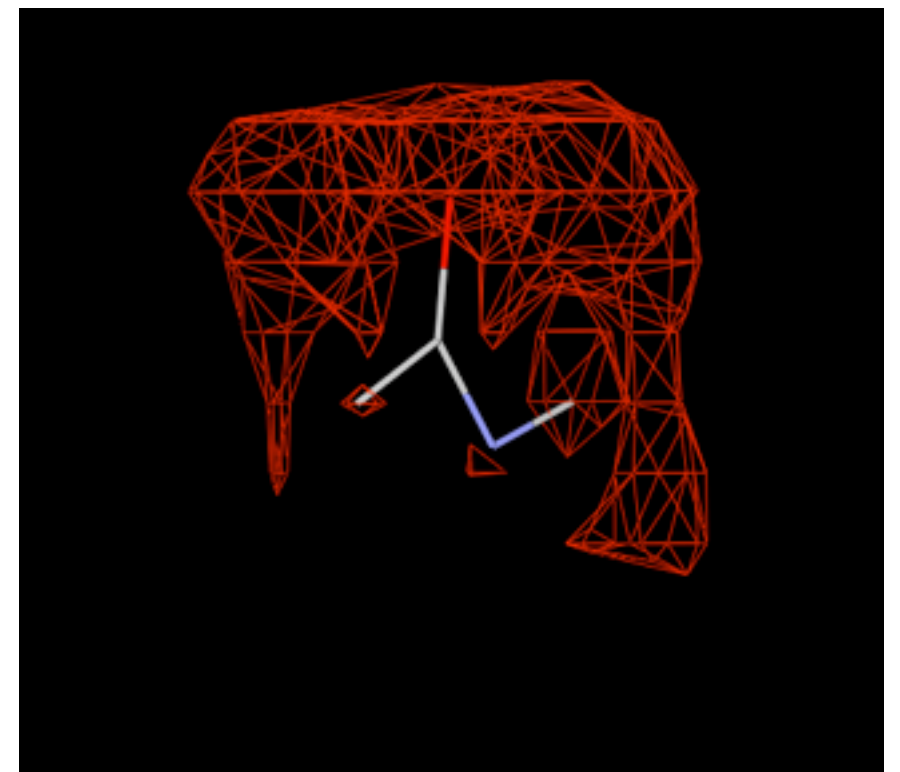
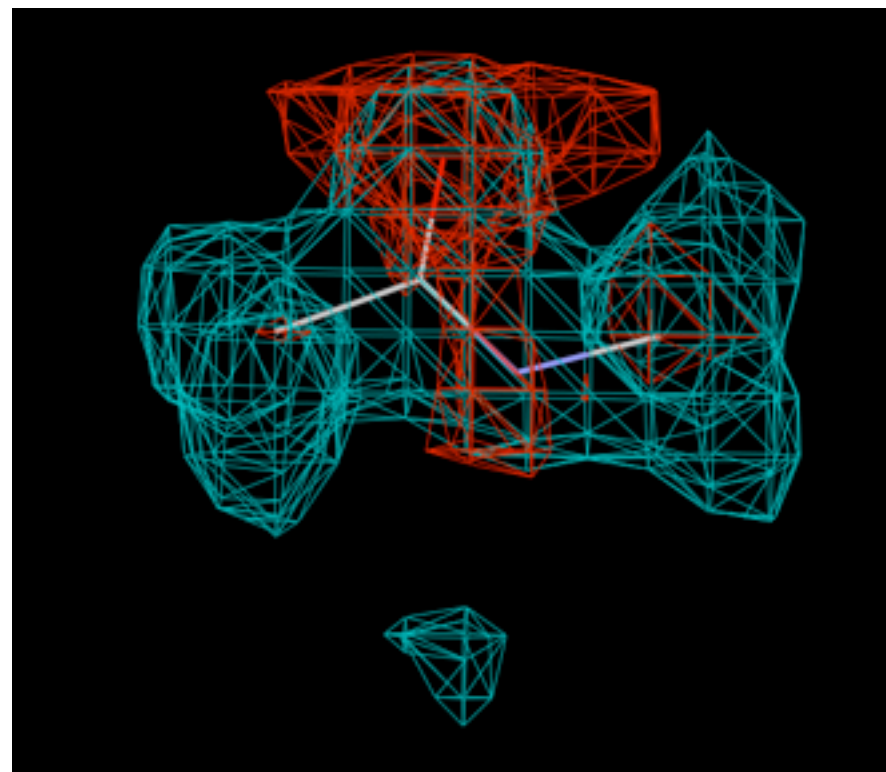


A peptide density shape is described by ~ **1,000** parameters, which are then **reduced to 1 parameter**

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

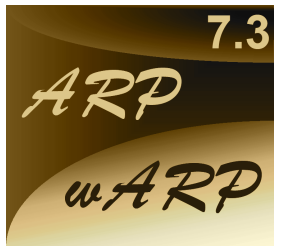
1.9Å resolution

Target density ■
Weight function ■
Similarity searched for



ARP/wARP Protein Model Building

“Modern Classic Version”

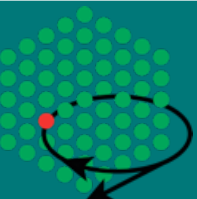
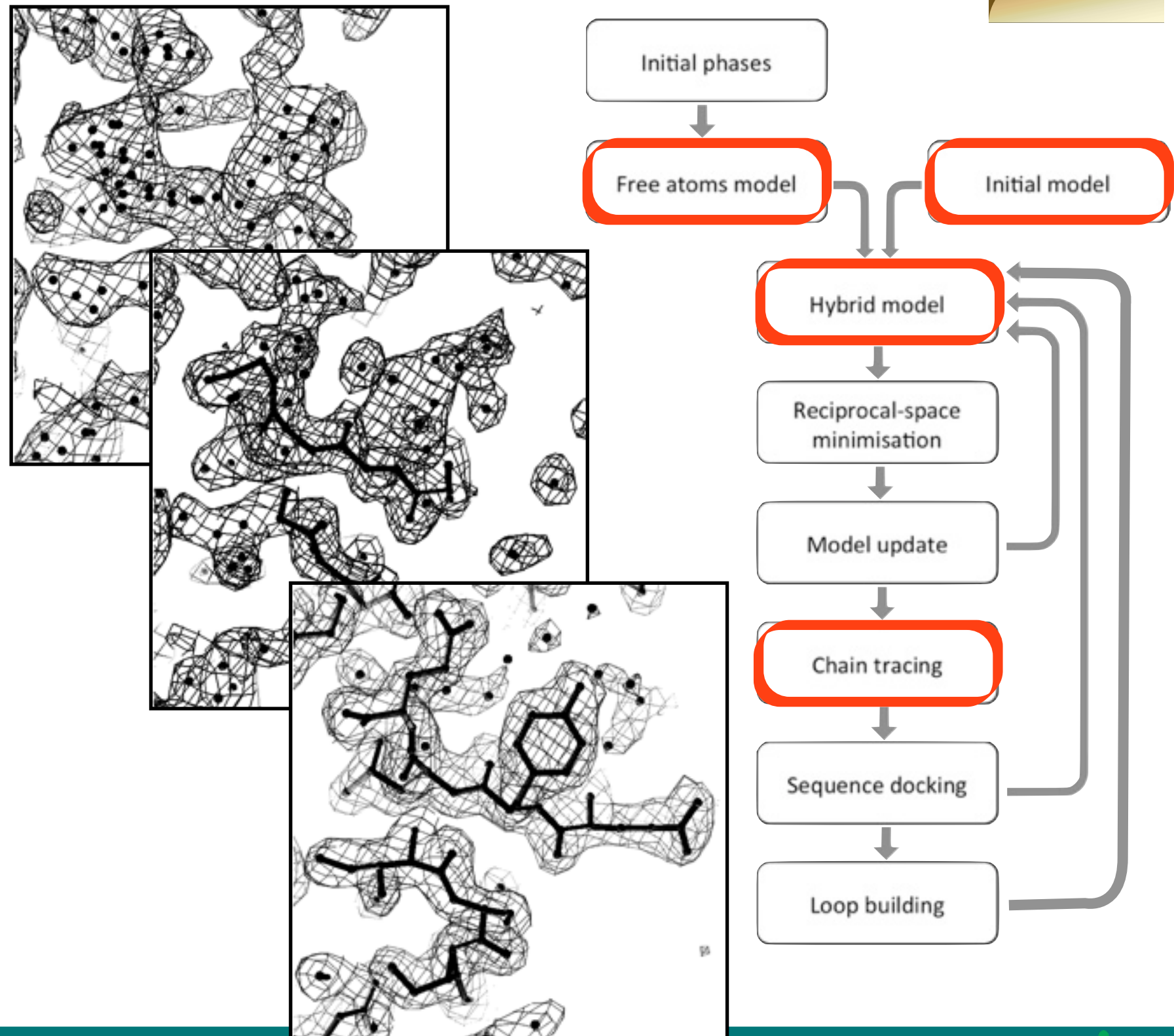


Free Atoms

Pattern Recognition

Hybrid Models

Iterations



ARP/wARP Model Building

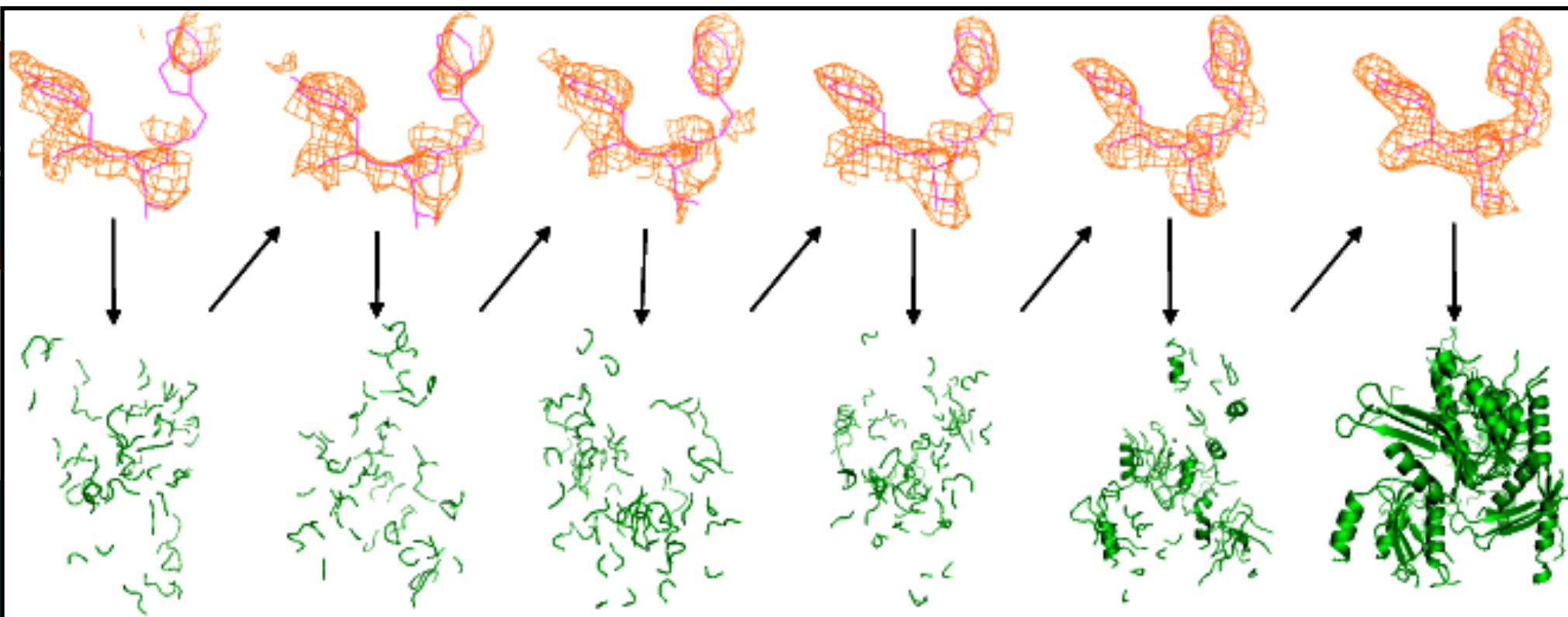
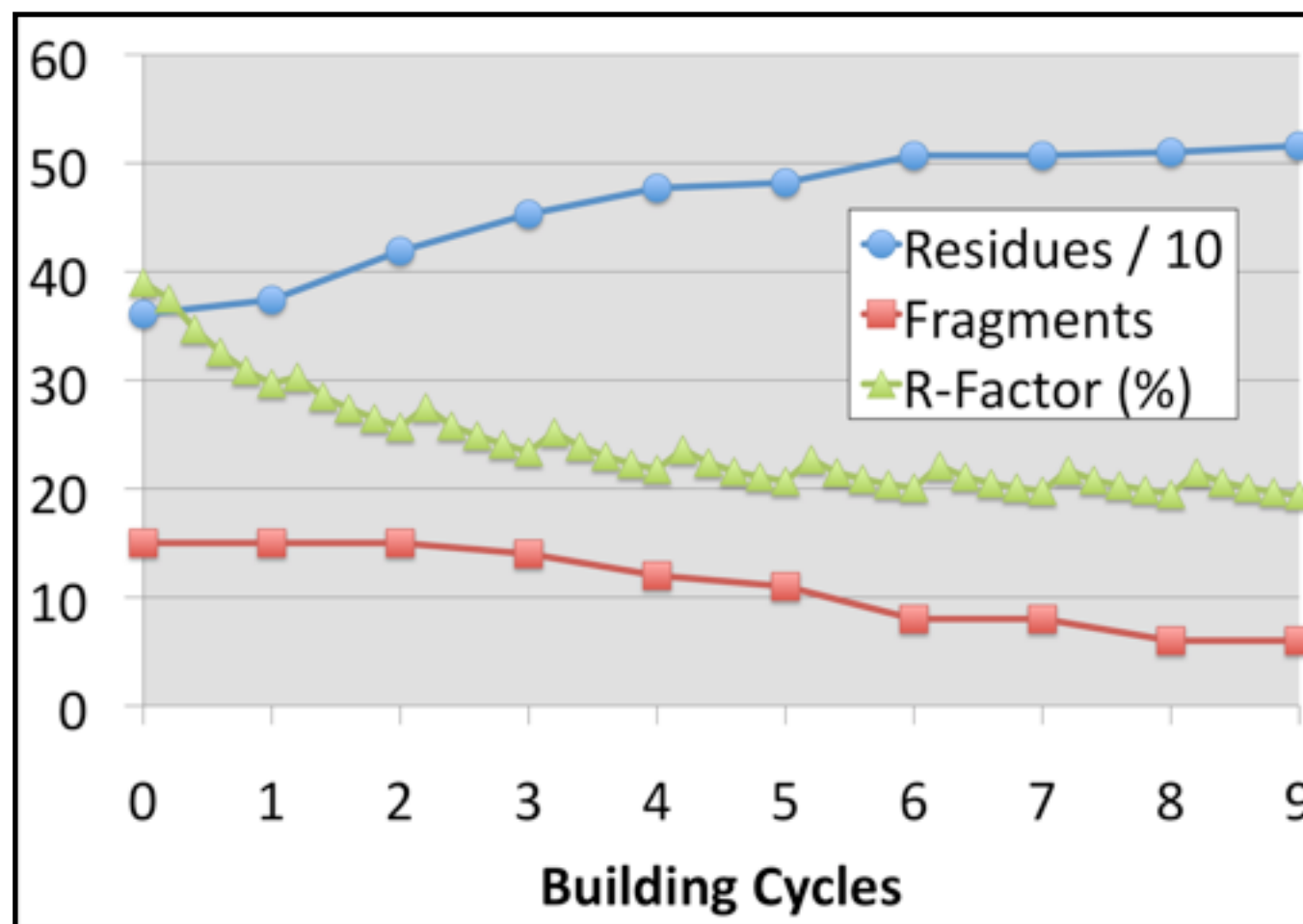
“Modern Classic Version”



Free Atoms

Pattern Recognition

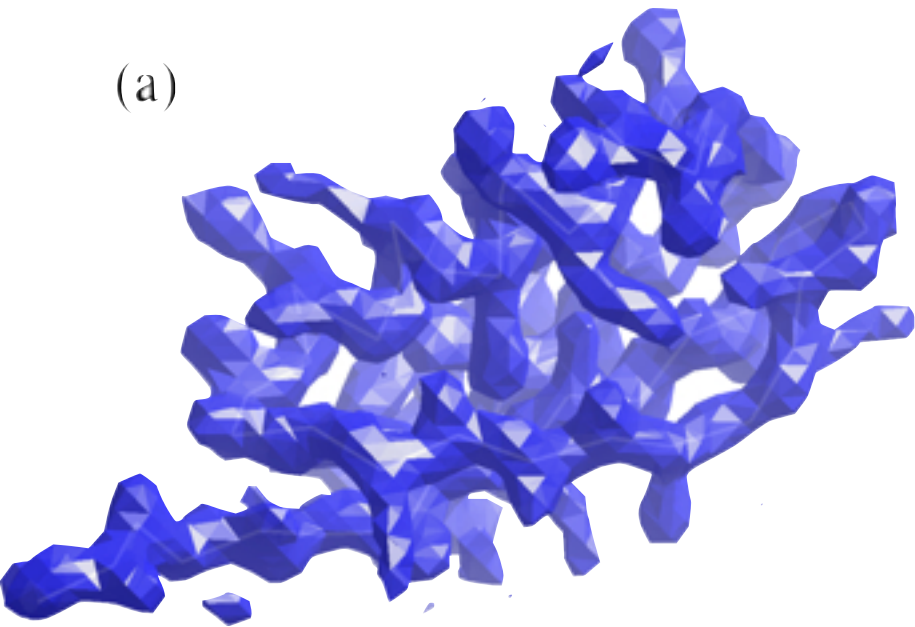
Hybrid Models



Dependence on Data Resolution

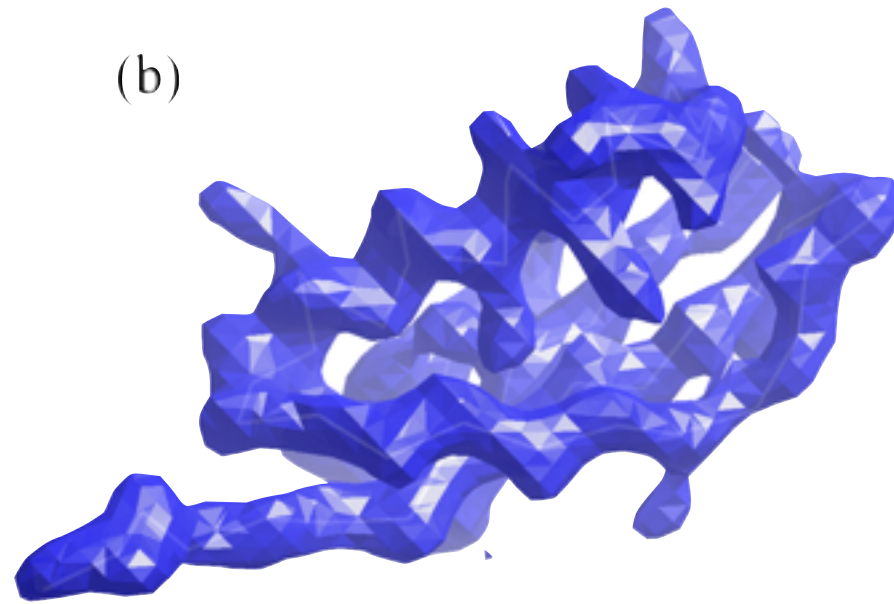
3 Å

(a)



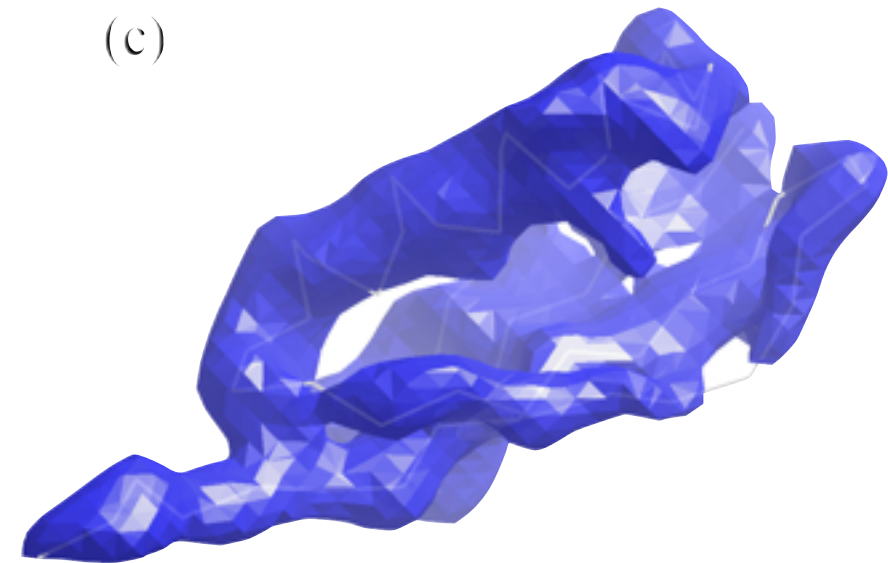
4 Å

(b)



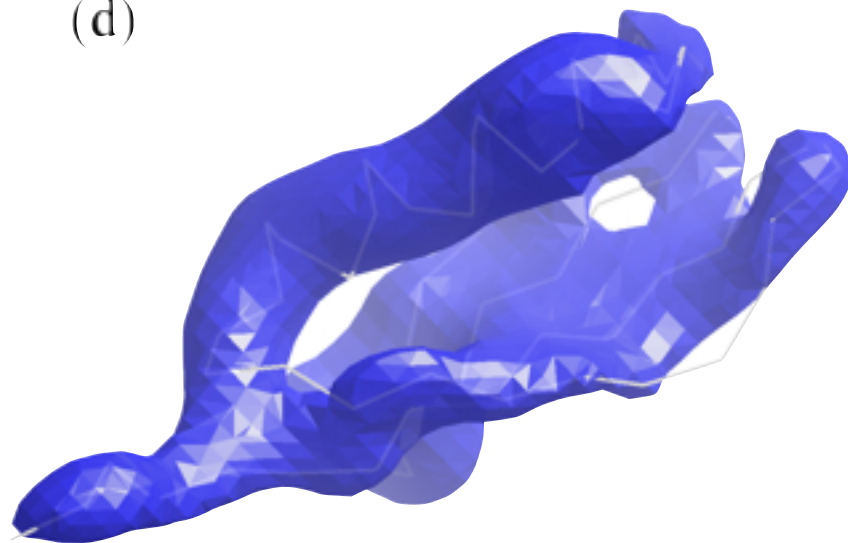
5 Å

(c)



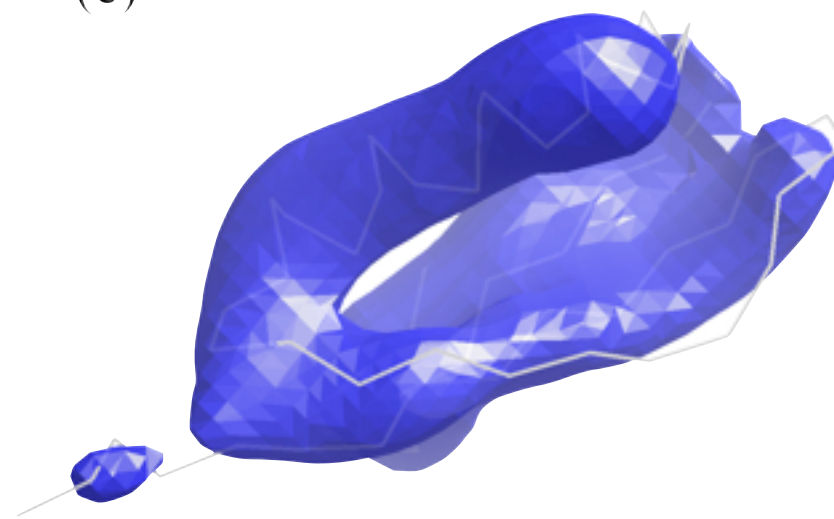
6 Å

(d)



8 Å

(e)



Dependence on Data Resolution



Results from the ARP/wARP 7.2 webservice (tracing performance)

Resolution	Estimated fraction of automatically built protein structure (October 2011 – March 2012)
1.5 Å and higher	98% and higher
1.5 - 2.0 Å	91 – 98%
2.0 - 2.5 Å	81 - 91%
2.5 - 3.0 Å	67 - 81%
3.0 – 3.5 Å	51 - 67%

Ease of Access - ARPnavigator



Ease of Access - ARPnavigator



ArpNavigator Files Actions Display Options Help

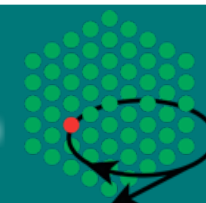
ArpNavigator

Use the Files menu to open files.

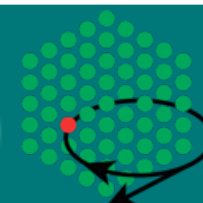
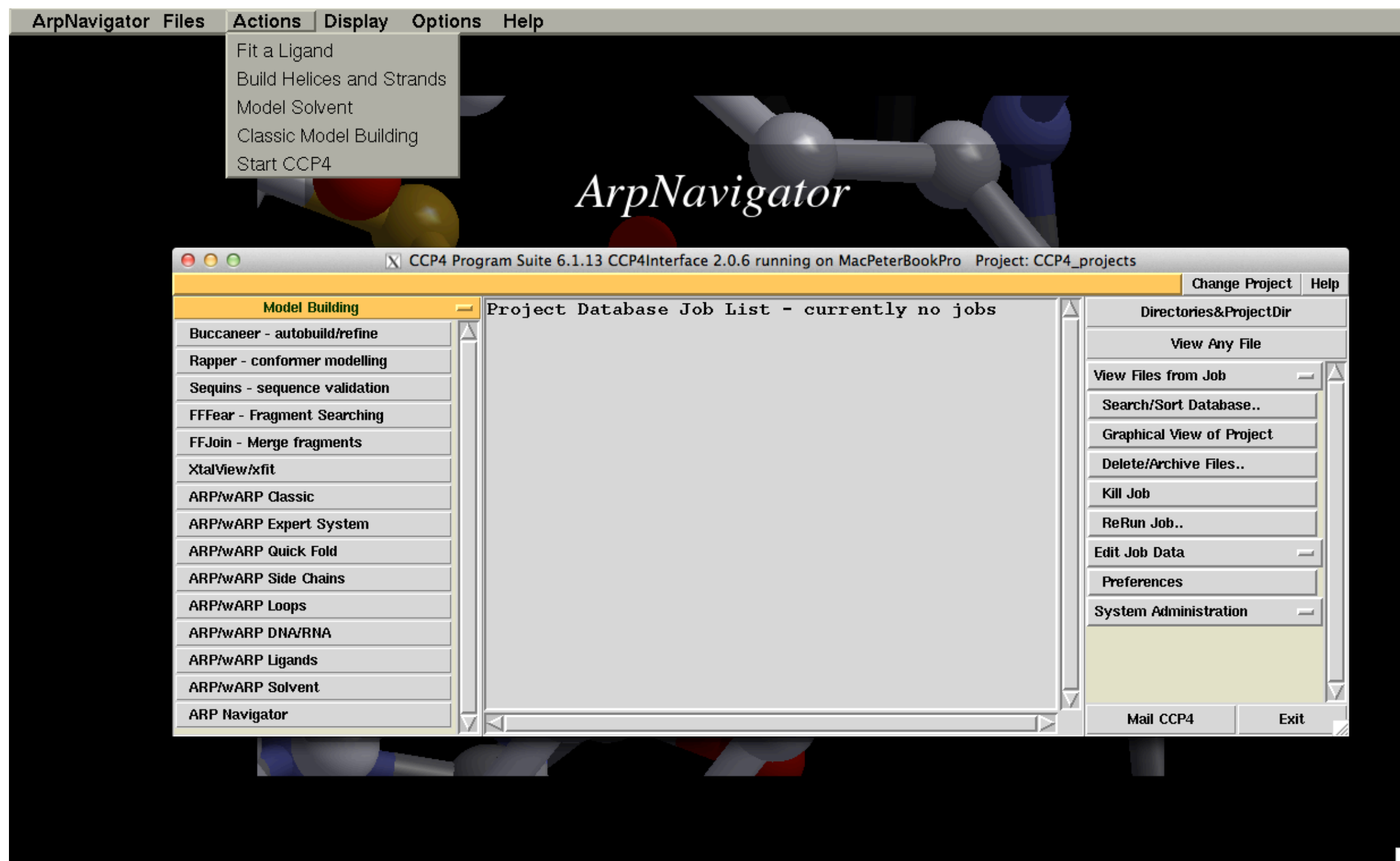
Use the Actions menu to build a protein model, fit a ligand, build helices and strands, model solvent or start CCP4.

Right click and drag to move objects.
Right click on a loaded file to change its parameters.
Right click on the background for quick actions.

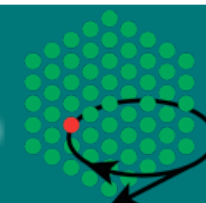
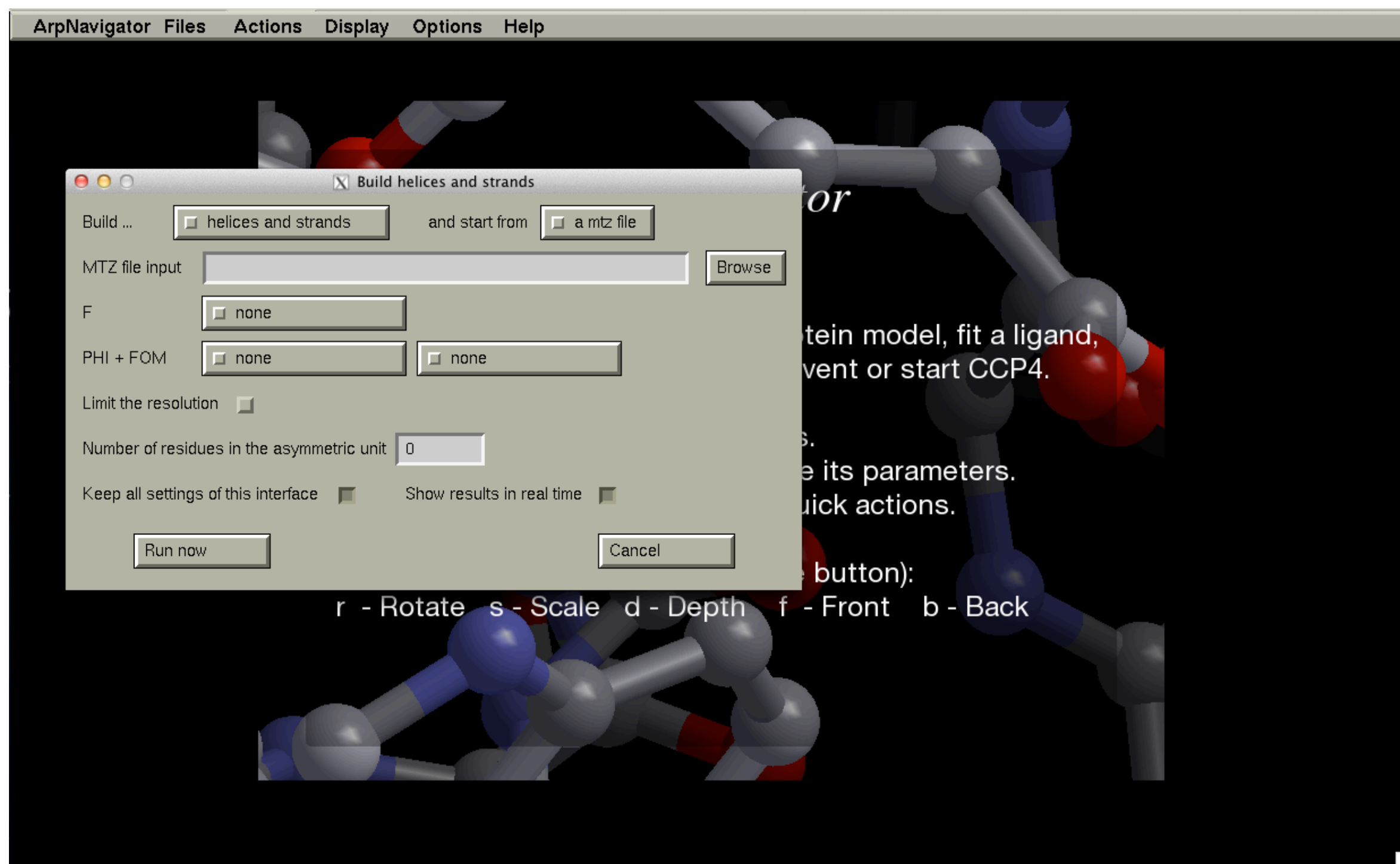
Keyboard shortcuts (with left mouse button):
r - Rotate s - Scale d - Depth f - Front b - Back

A 3D molecular model showing a protein backbone in grey, with various atoms represented by colored spheres (red for oxygen, blue for nitrogen, yellow for sulfur, and grey for carbon). The model is set against a dark background within the ArpNavigator software window.

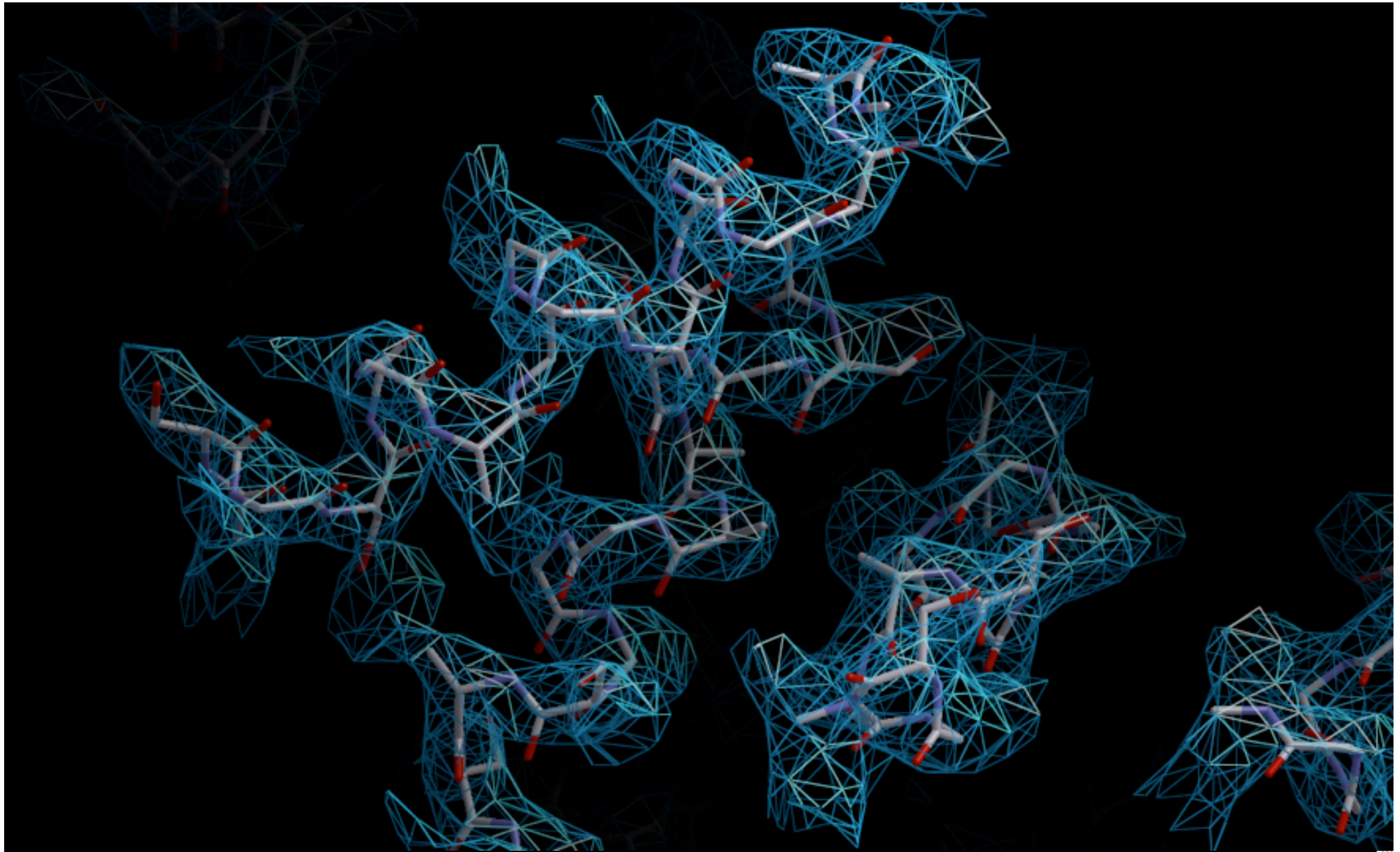
Ease of Access - ARPnavigator



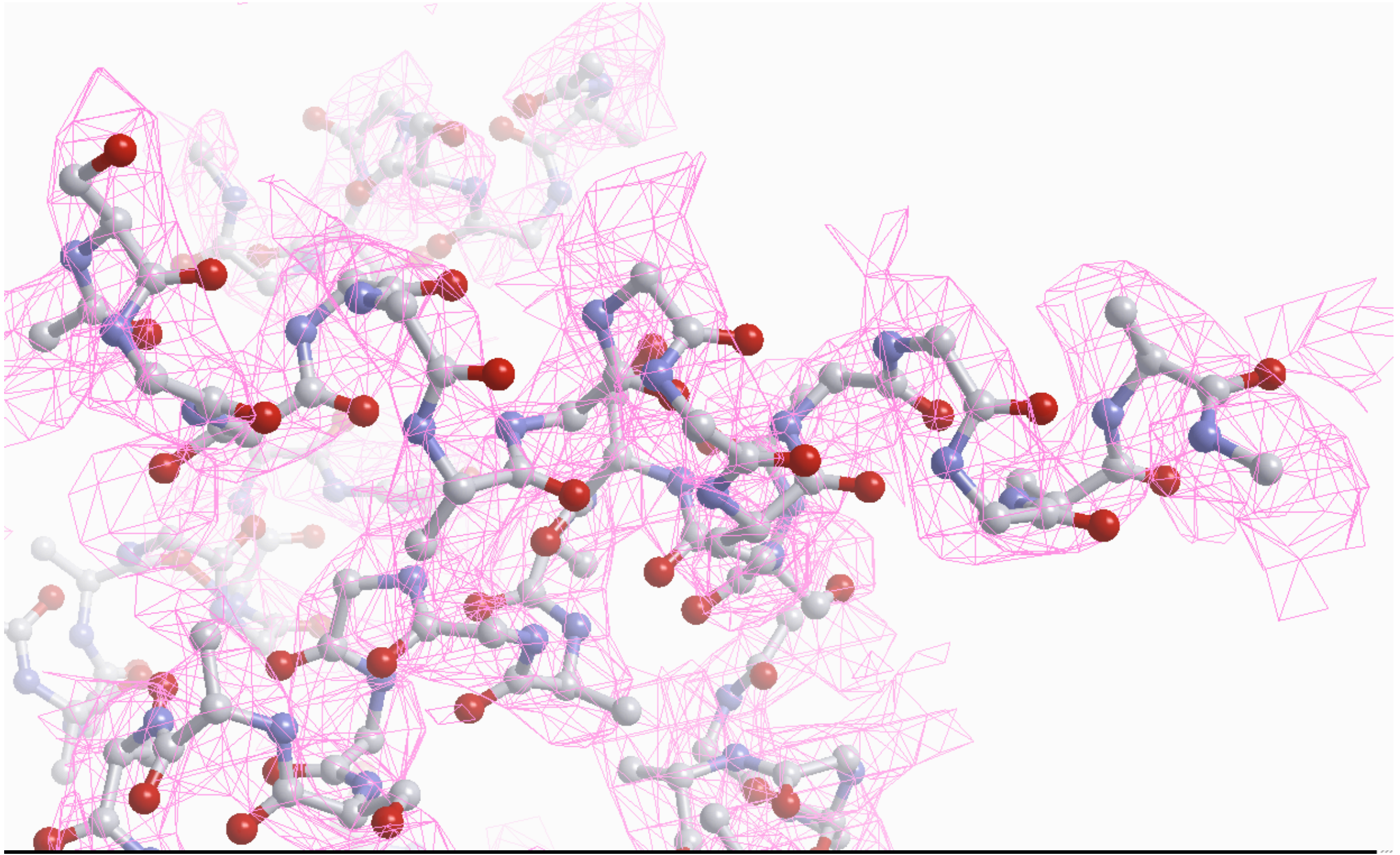
Ease of Access - ARPnavigator



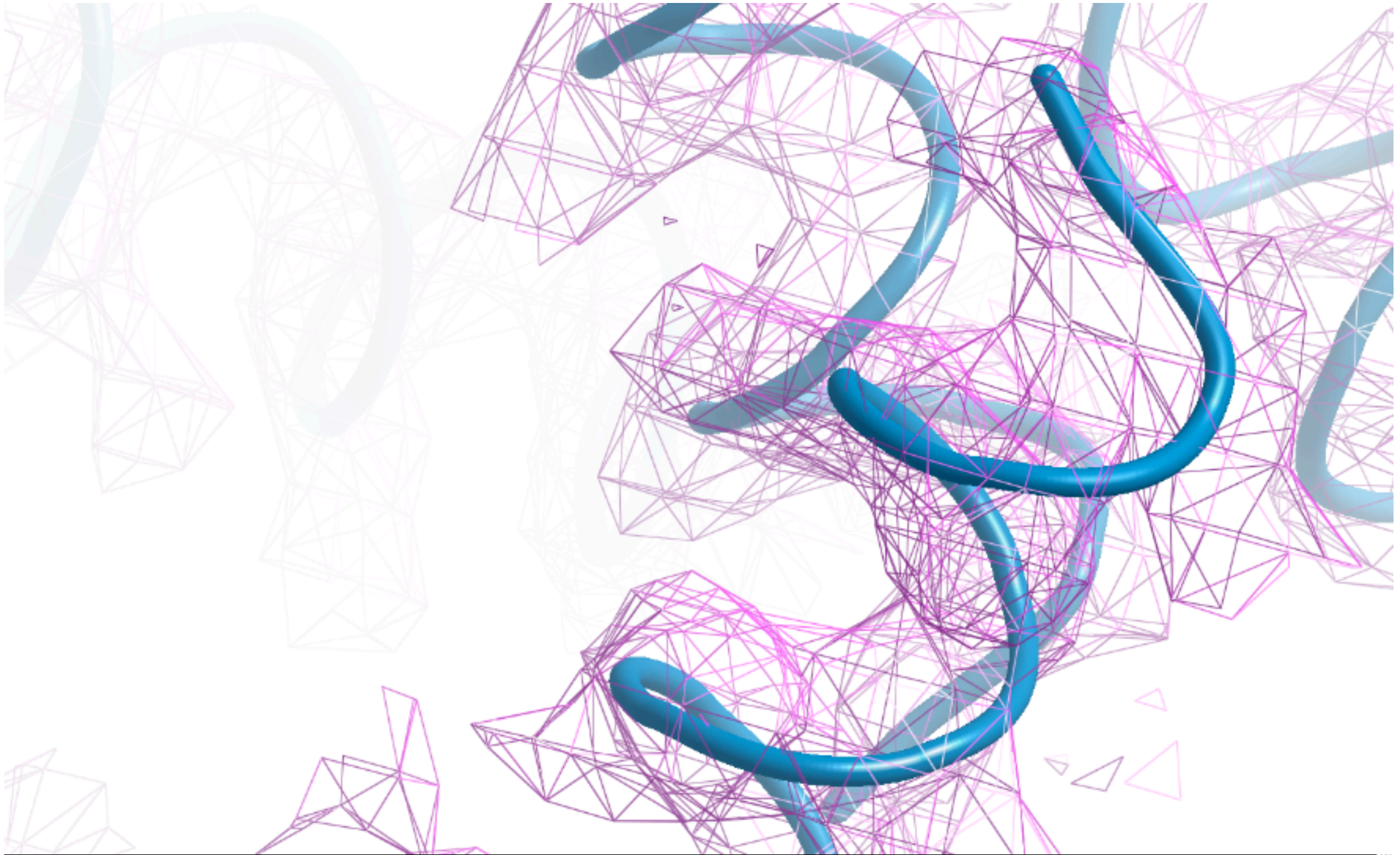
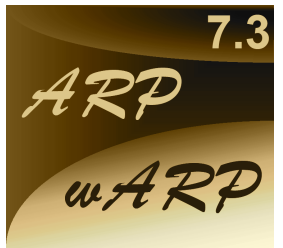
Ease of Access - ARPnavigators



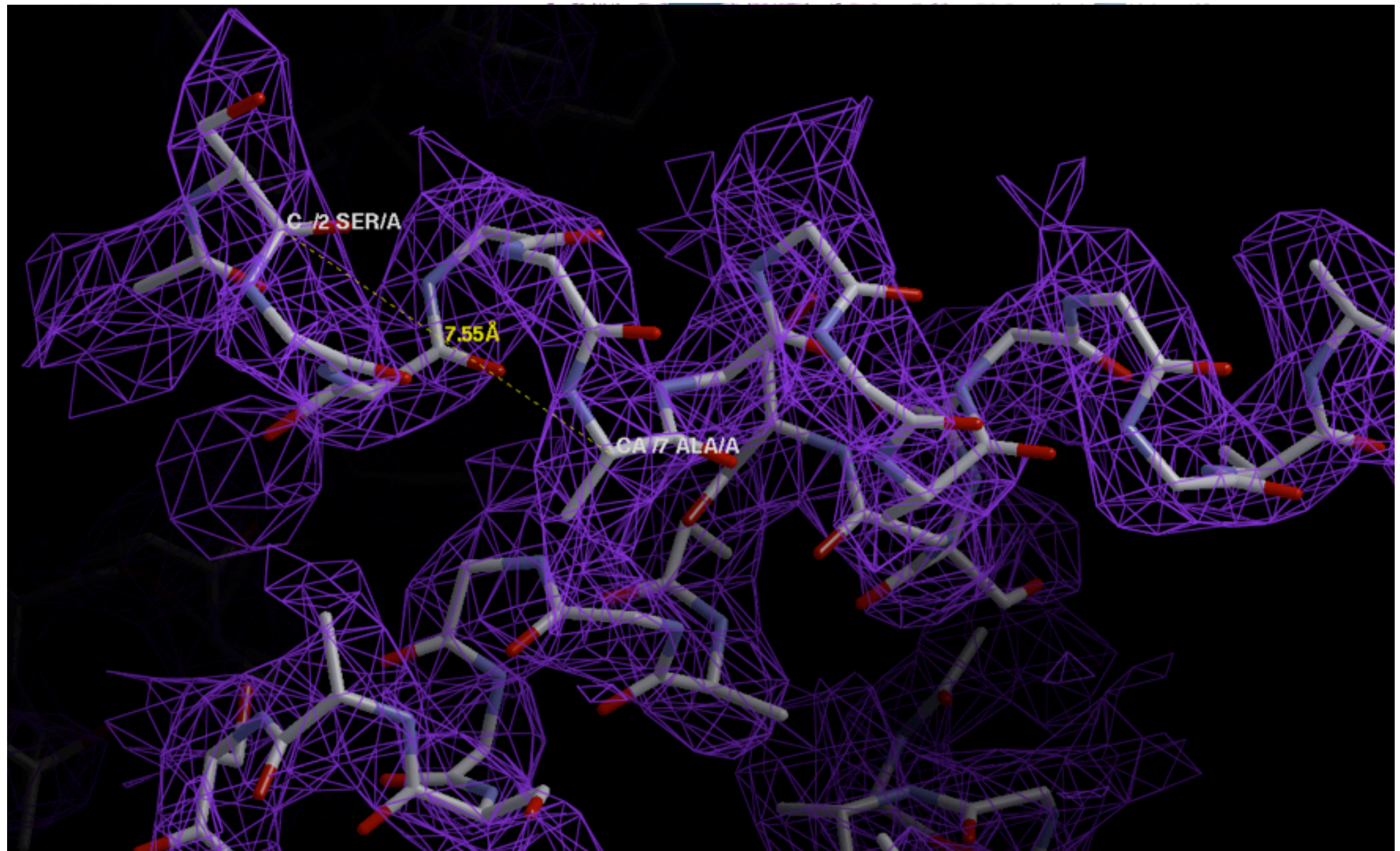
Ease of Access - ARPnavigators



Ease of Access - ARPnavigator

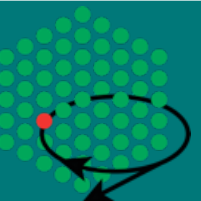
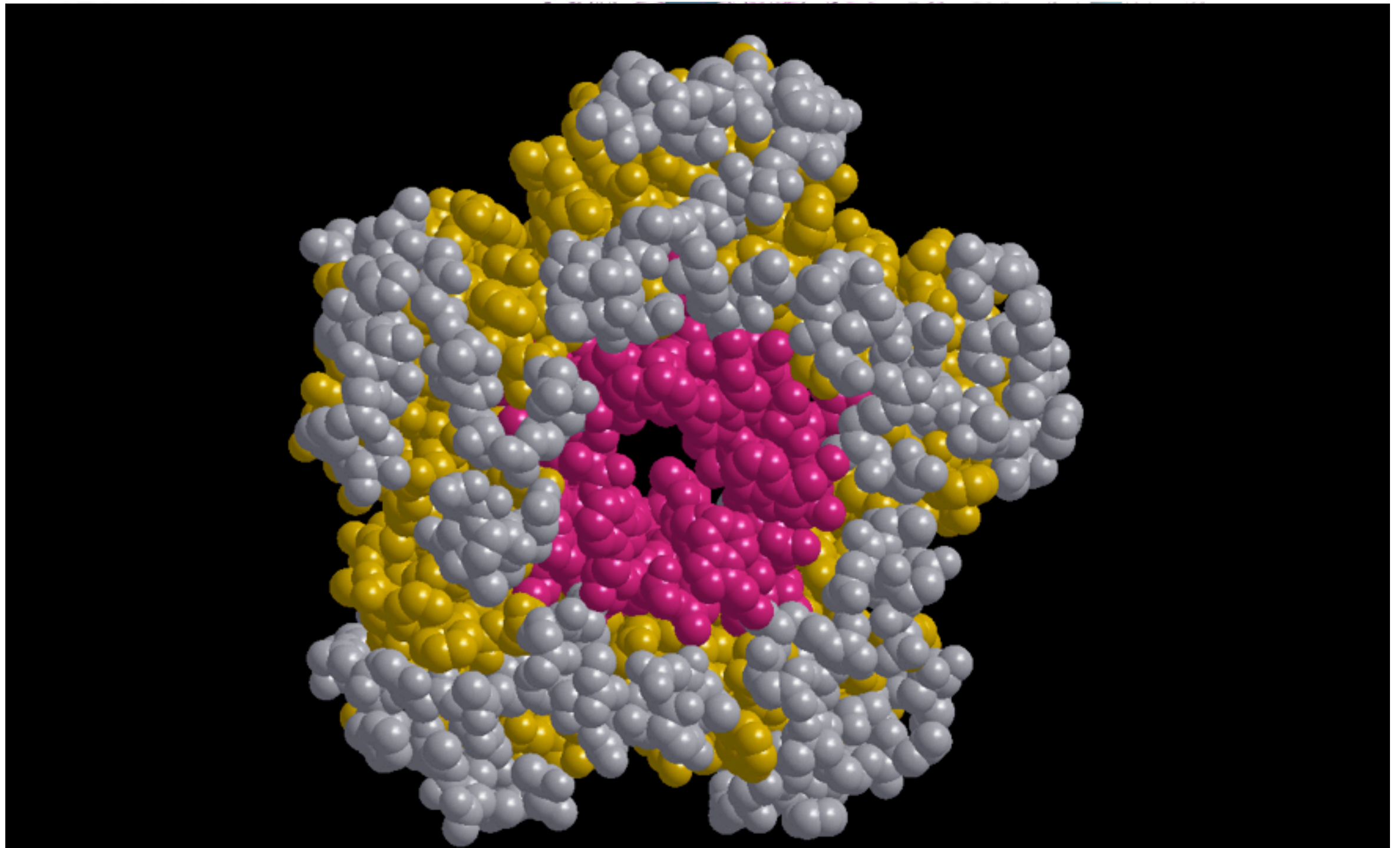


Ease of Access - ARPnavigator

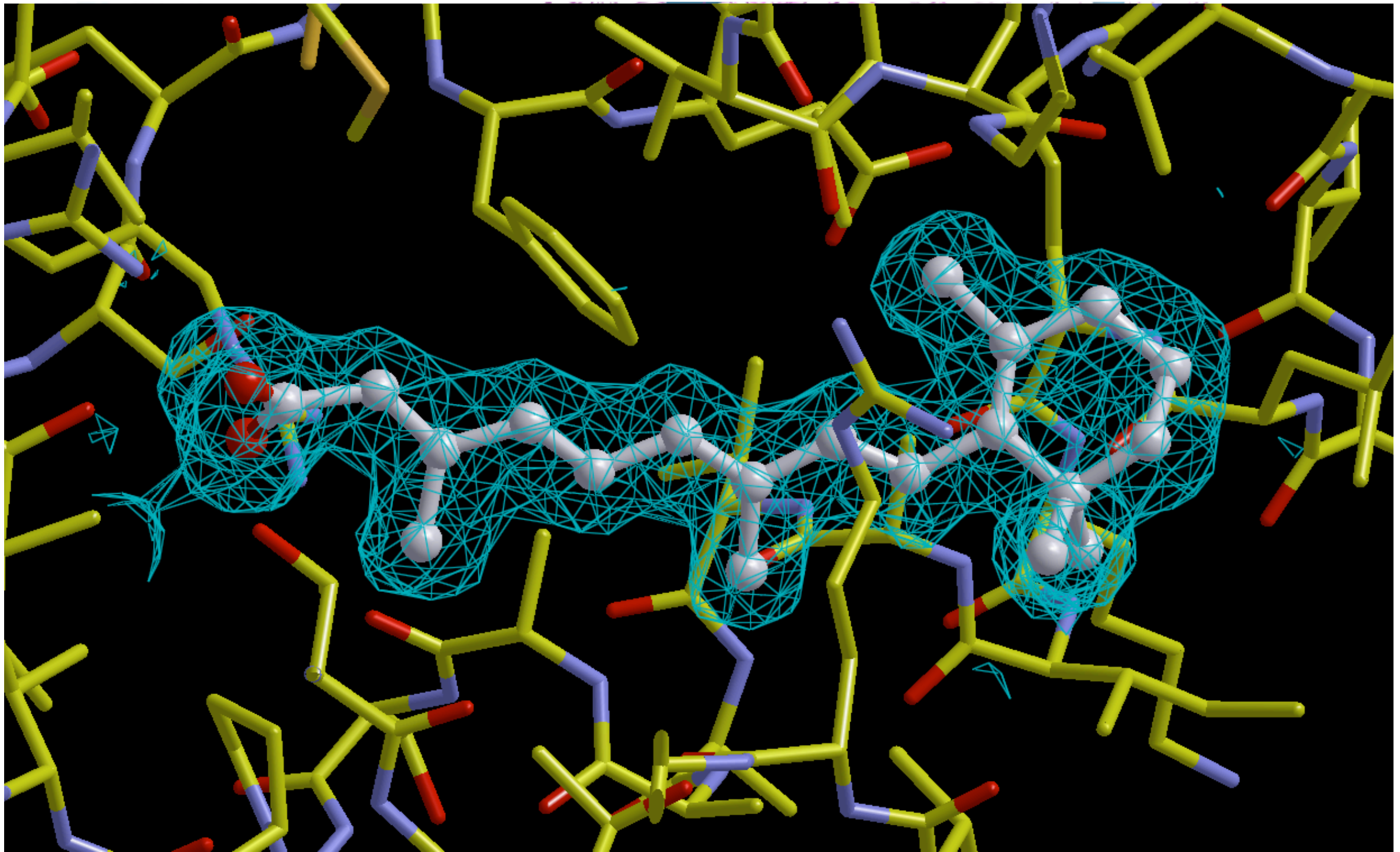




Ease of Access - ARPnavigator



Ease of Access - ARPnavigators

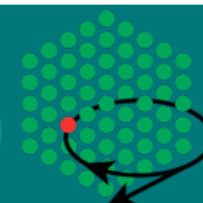
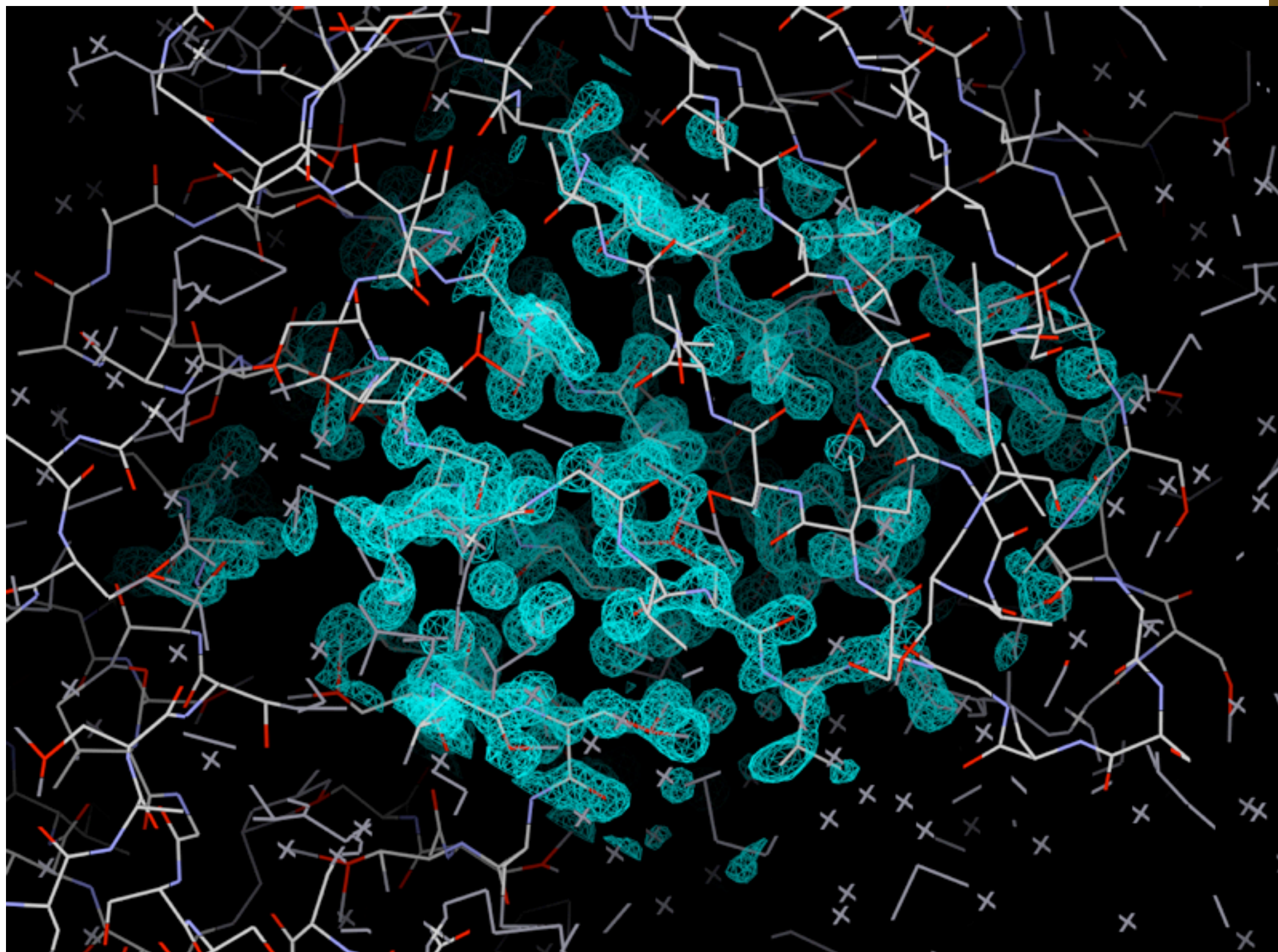


ARPnavigator - improved visualisation!

7.3

ARP

wARP

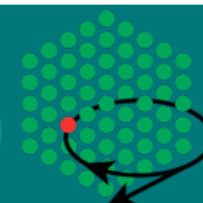
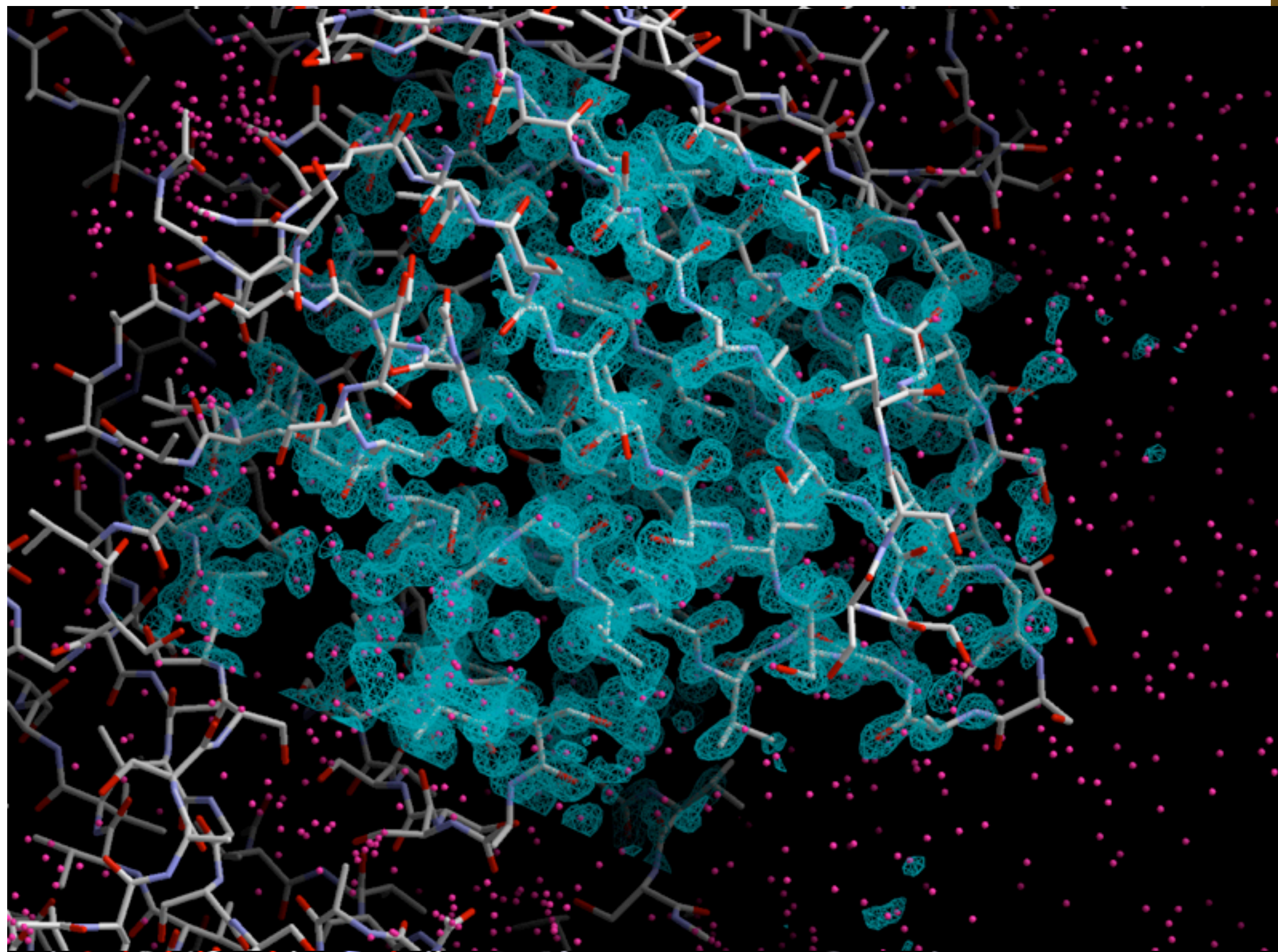


ARPnavigator - improved visualisation!

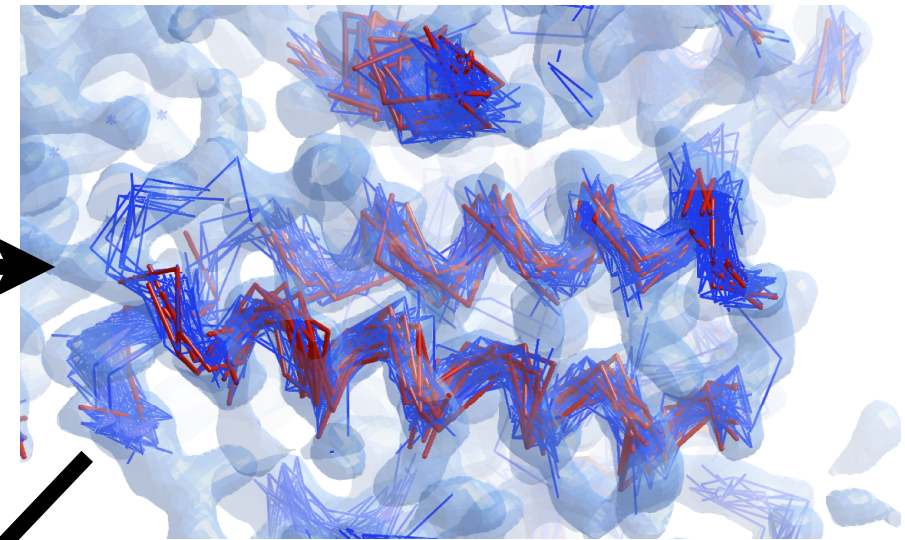
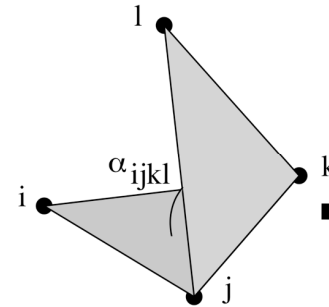
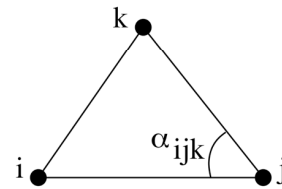
7.3

ARP

wARP



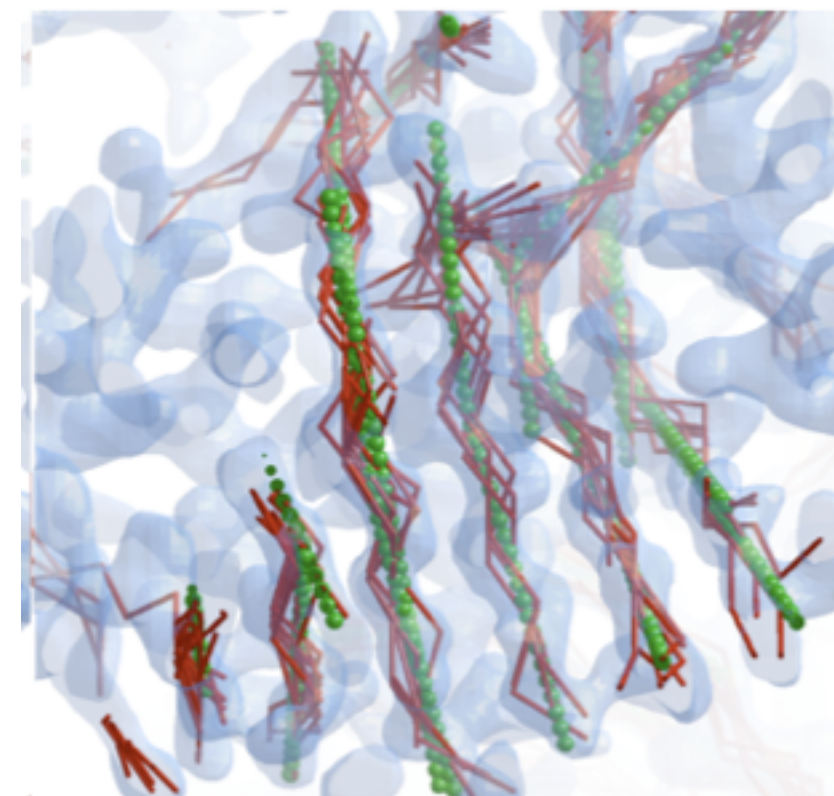
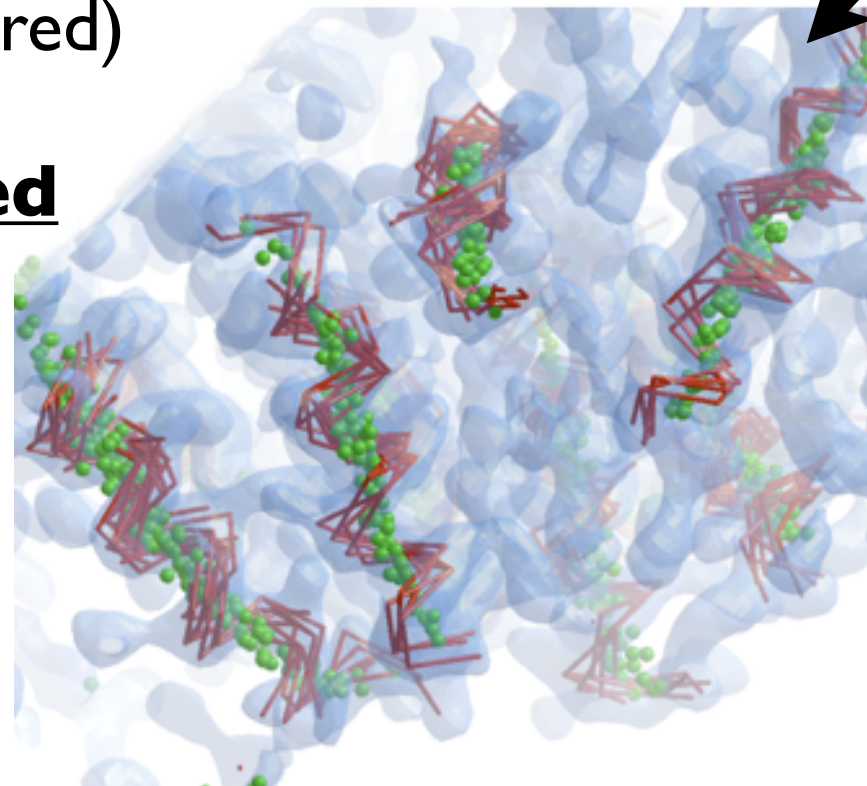
- **Short helix/strand fragments** (3 to 5 C α candidates) are built.

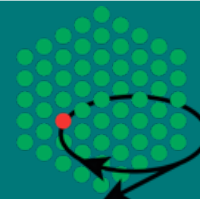
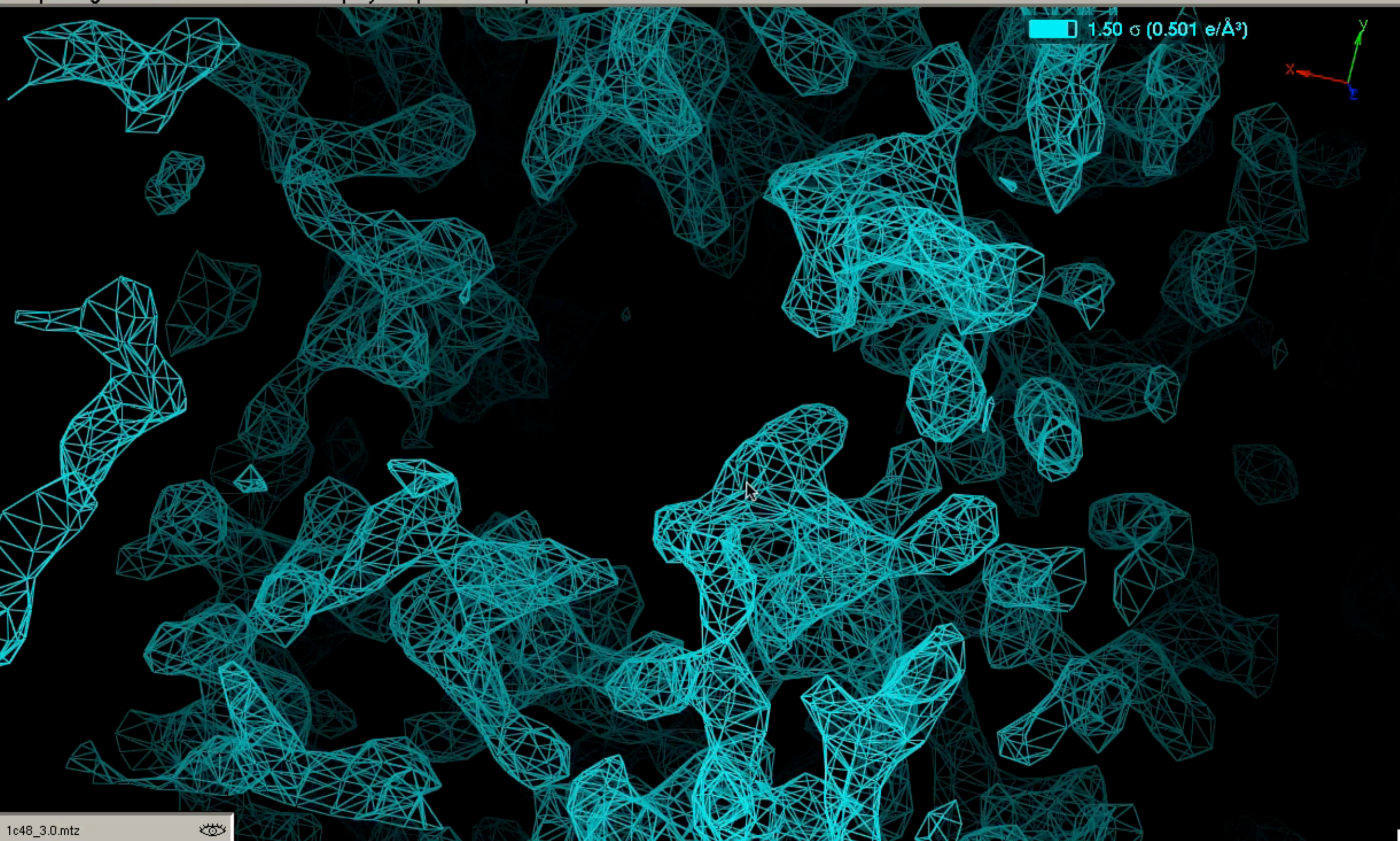


- Longer traces are formed or which the **best are kept** (in red)

- Traces are **clustered**

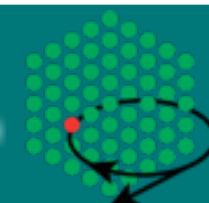
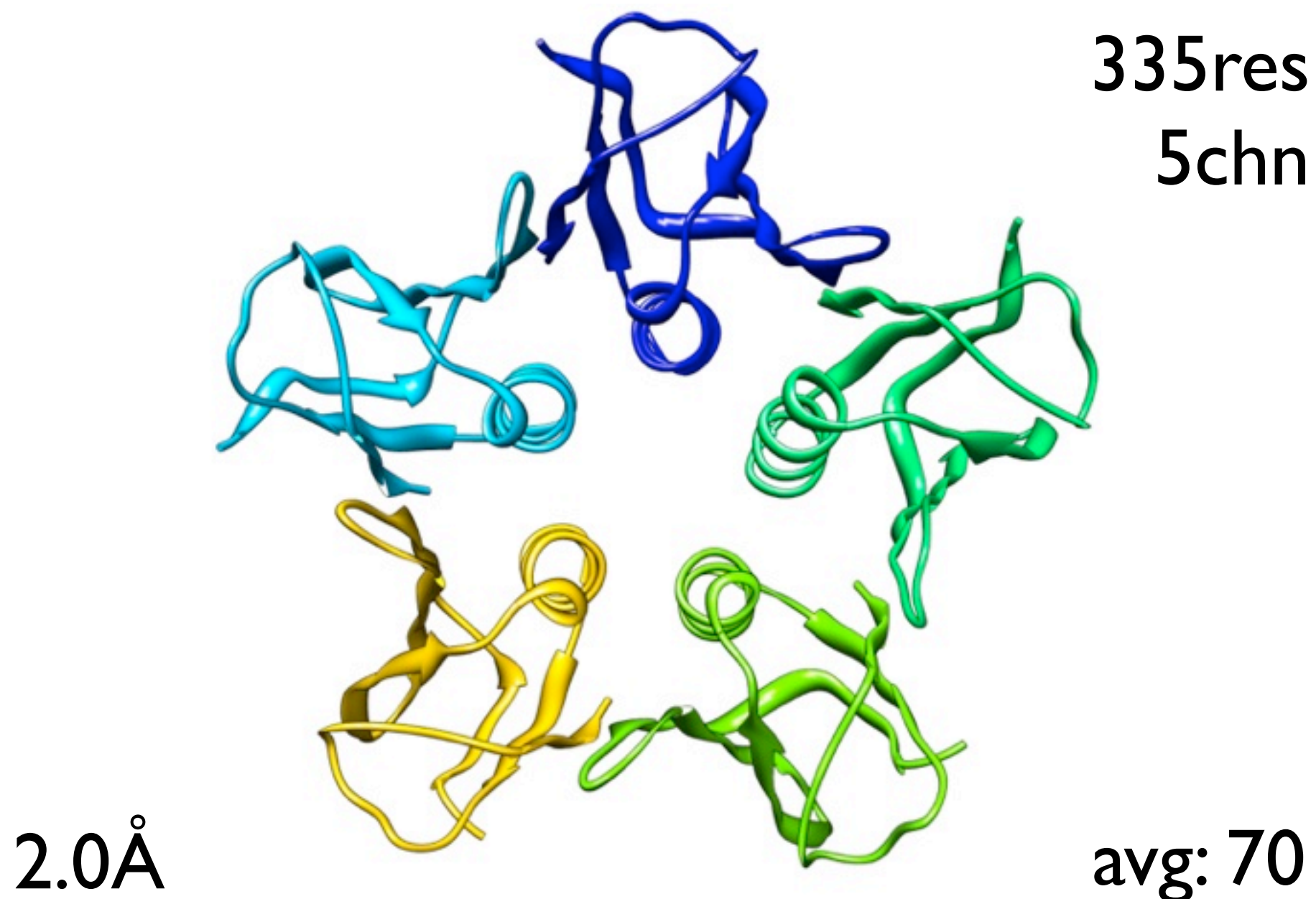
- Assemblies are **averaged**





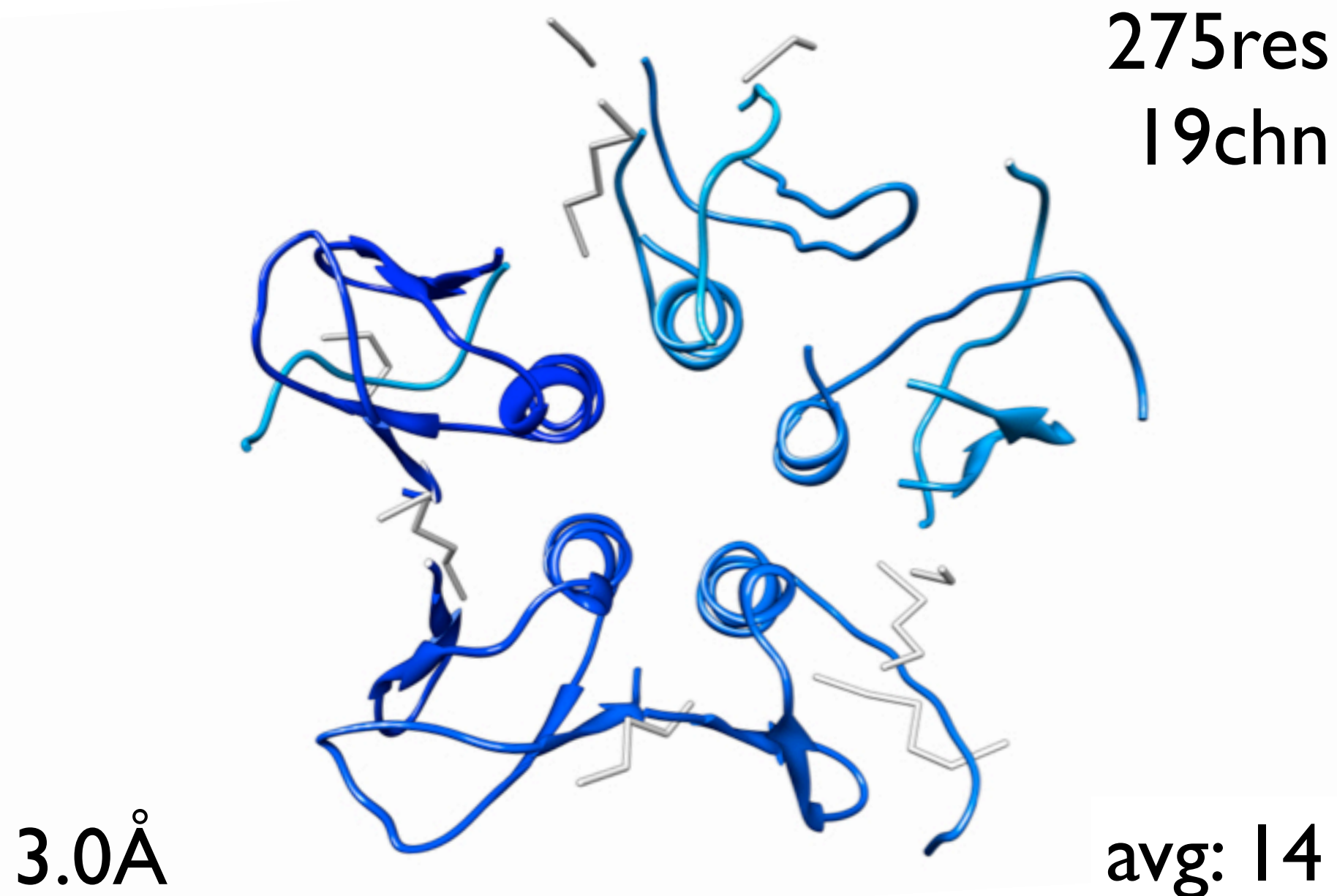
Automatic NCS Detection / Extension

- historically, the development of automatic modeling software has been **focused** on solving **high** resolution structures
- applied at **low** resolution, the solution is often **highly fragmented**

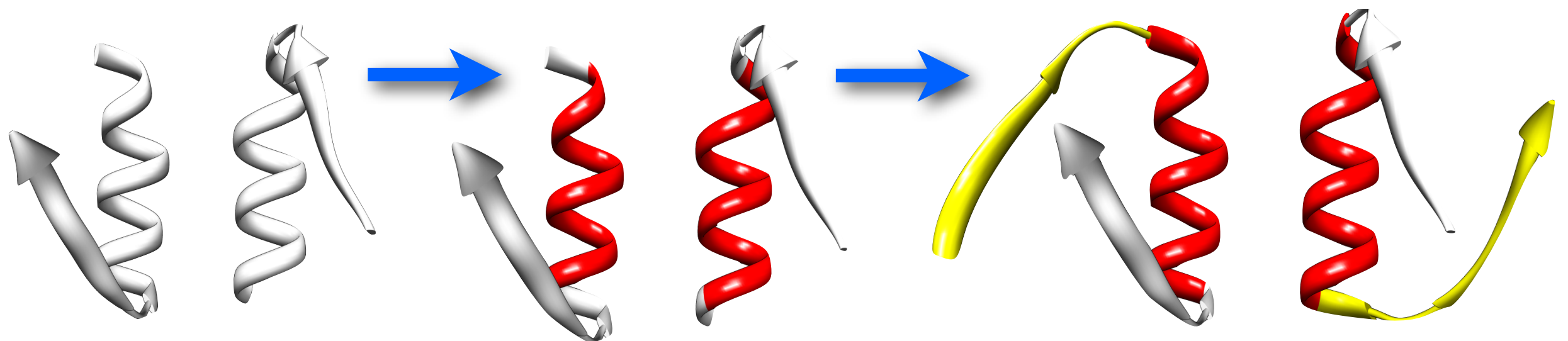
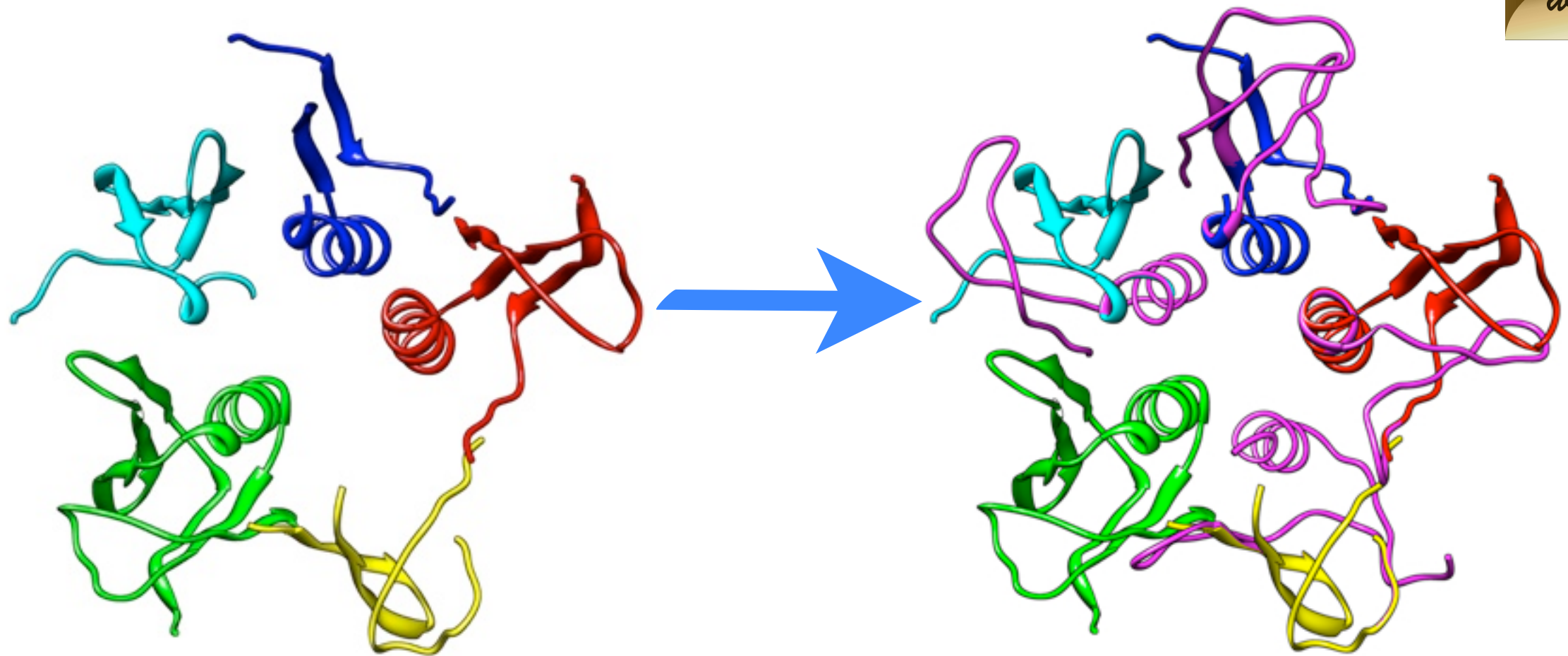


Automatic NCS Detection / Extension

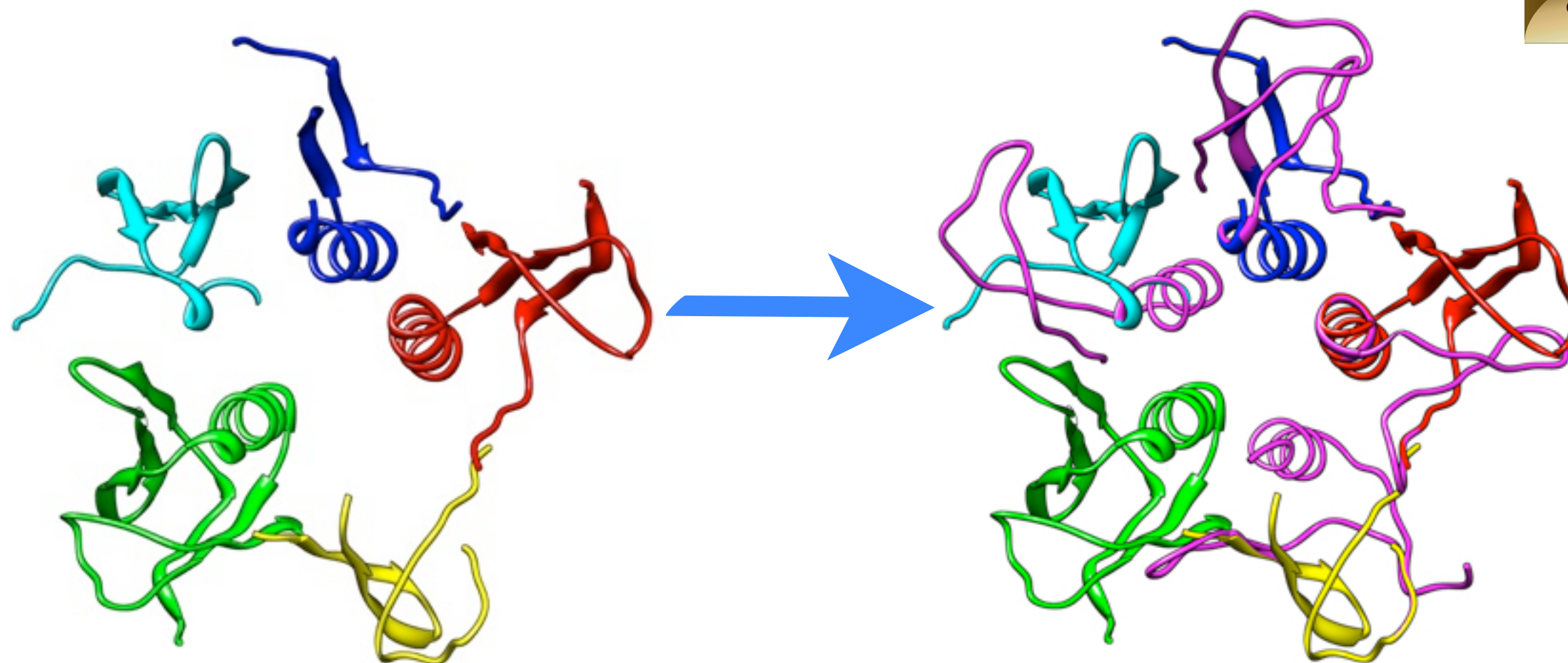
- historically, the development of automatic modeling software has been **focused** on solving **high** resolution structures
- applied at **low** resolution, the solution is often **highly fragmented**



Automatic NCS Detection / Extension

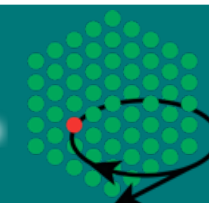


Automatic NCS Detection / Extension



14 protein test structures with resolution 1.9 to 3.2 Å, NCS order 2 - 10

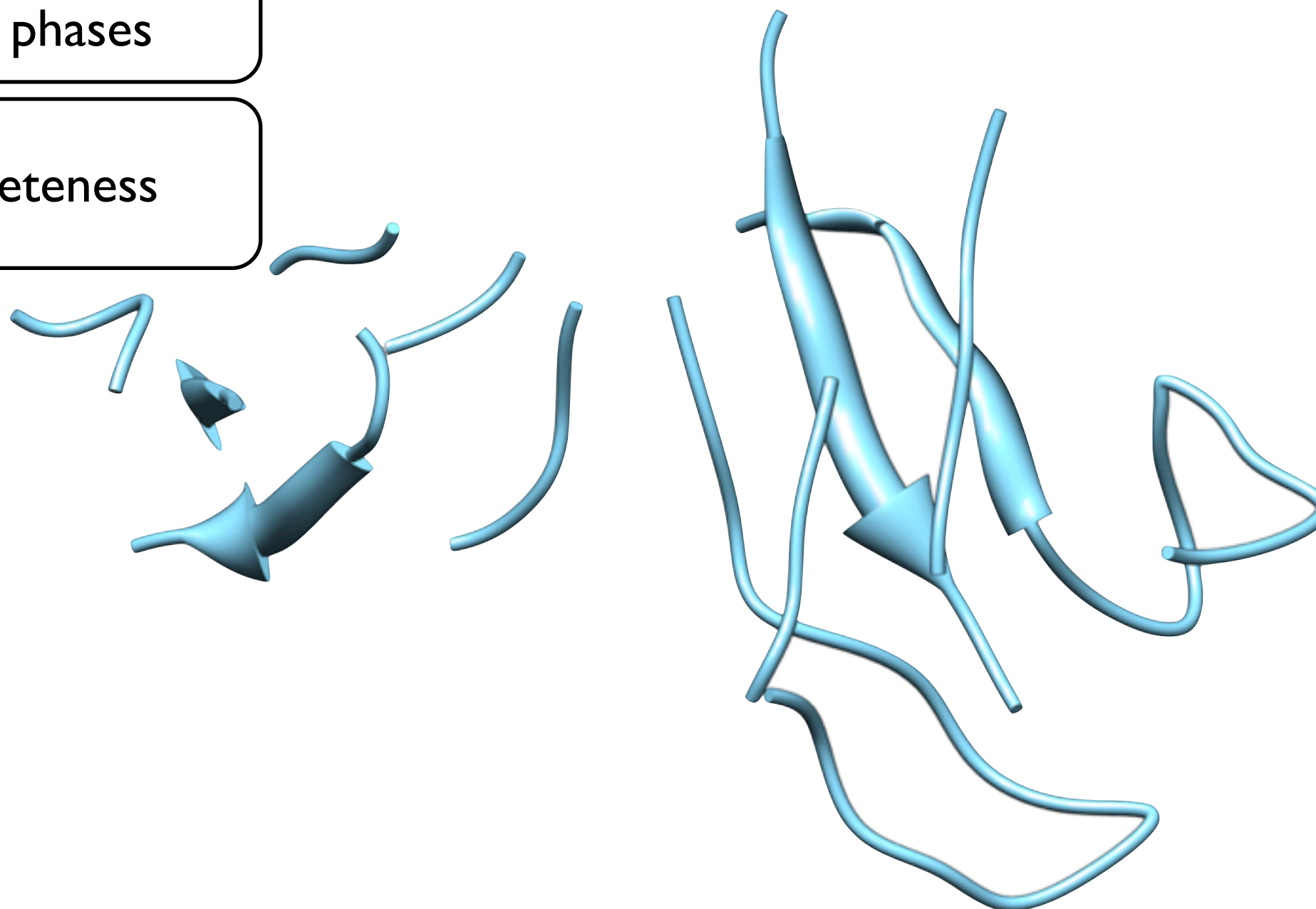
	Model Completeness	Residues / Fragment	Sequence Coverage
Top Results	+19.5%	+ 220%	+ 56%
Average in 7.3	+ 7.0%	+ 40%	+ 8%



Improvements by PNsxtender

- 1.95 Å data
- poor initial phases

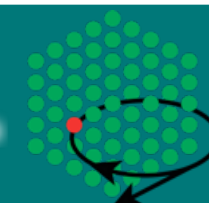
- 42% completeness



ARP/wARP



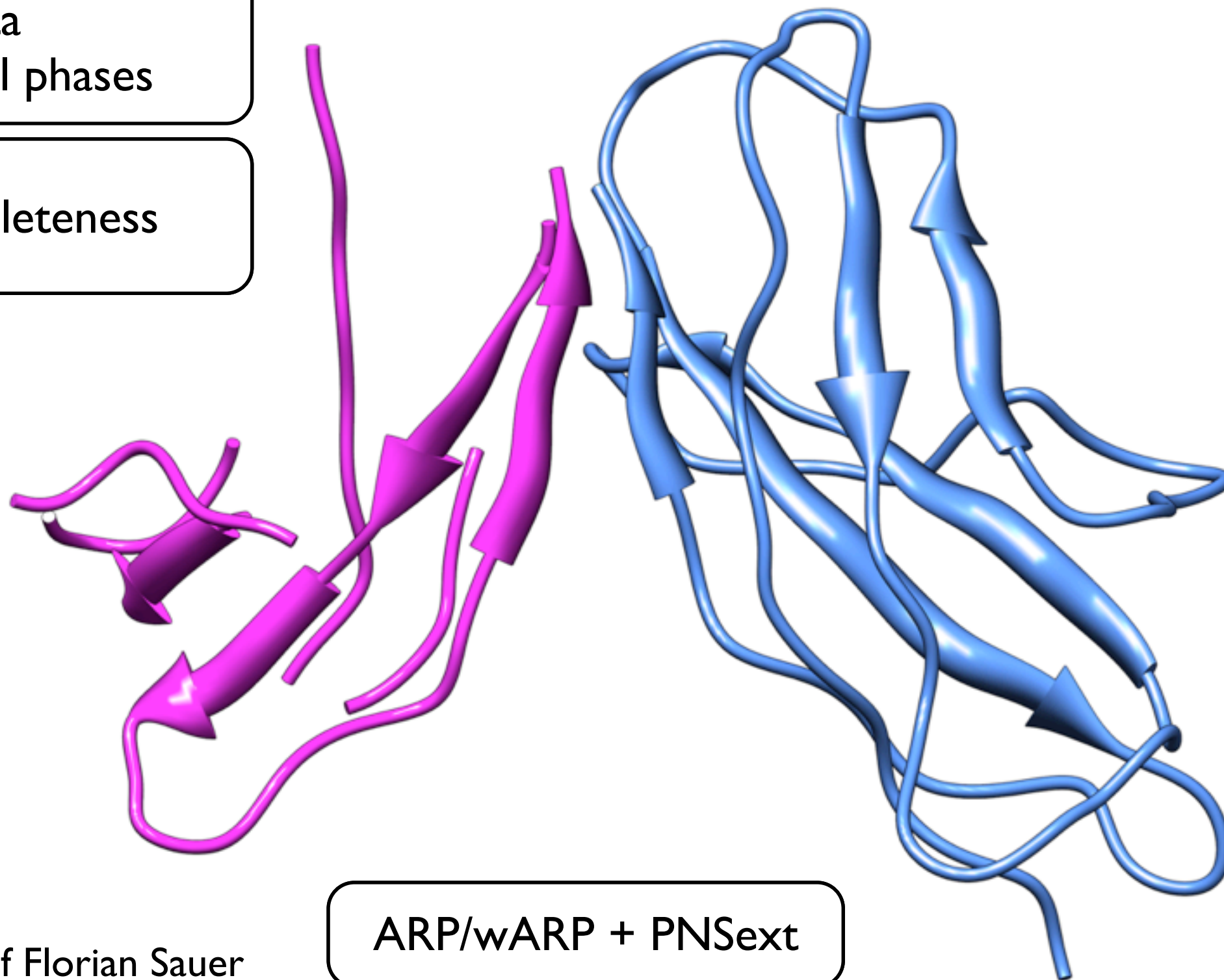
Data courtesy of Florian Sauer



Improvements by PNsxtender

- 1.95 Å data
- poor initial phases

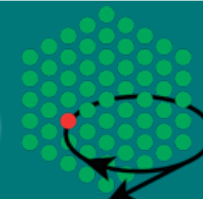
- 75% completeness



ARP/wARP + PNsxt



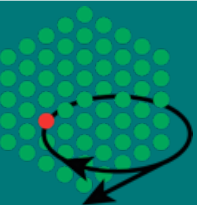
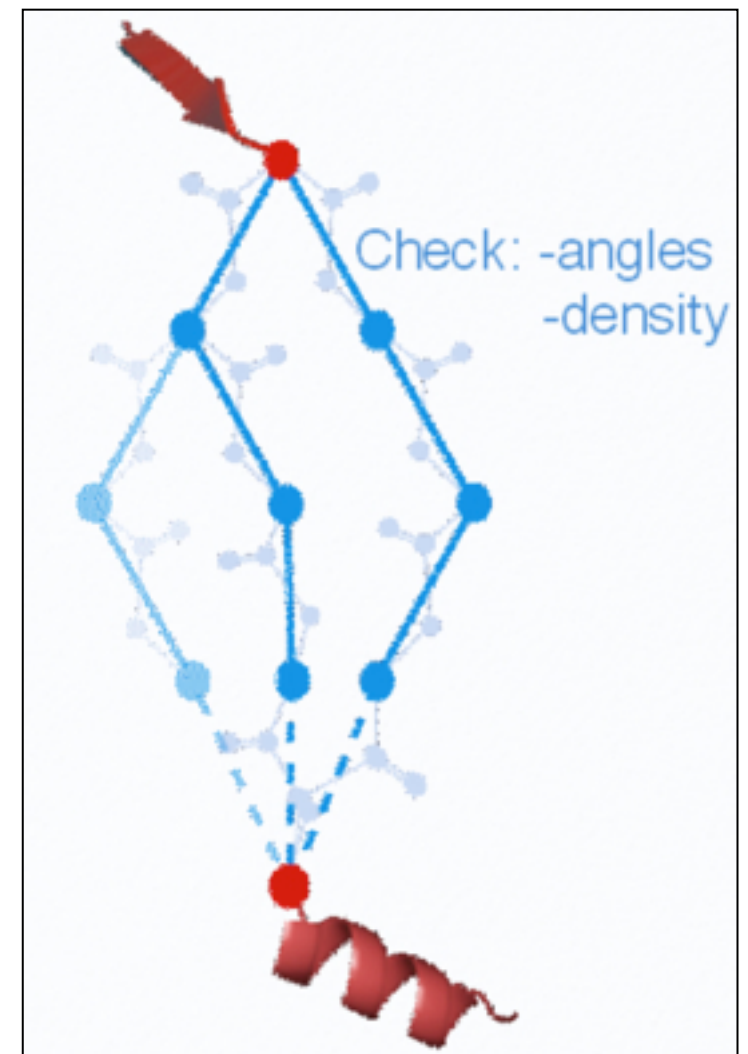
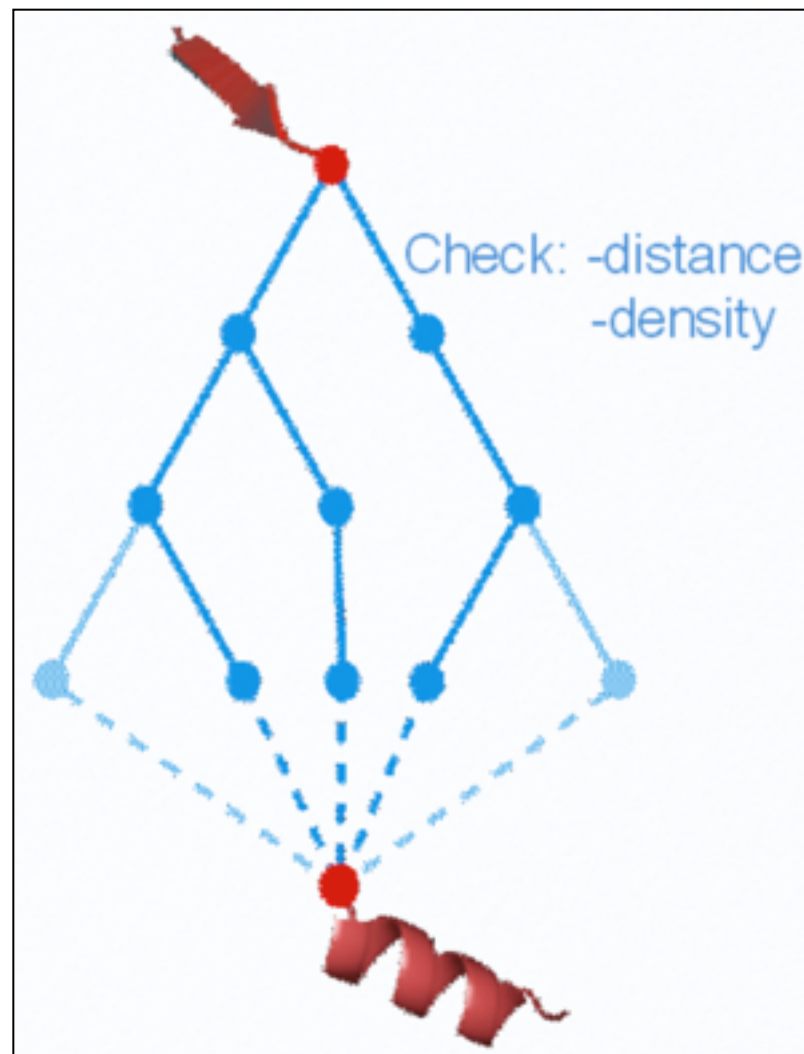
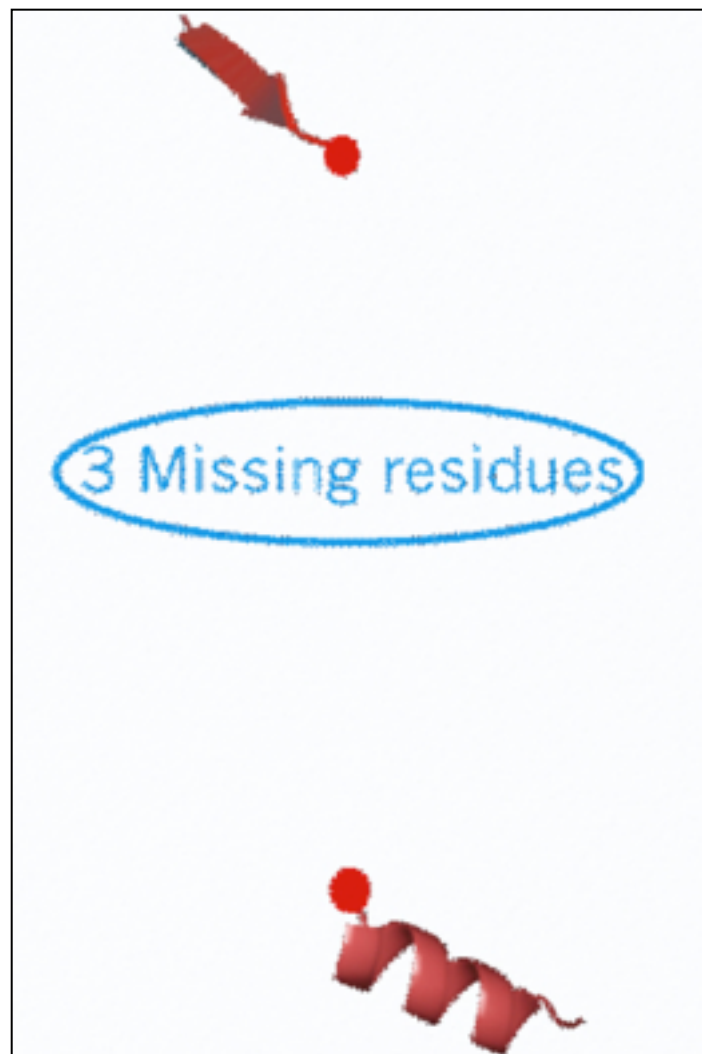
Data courtesy of Florian Sauer



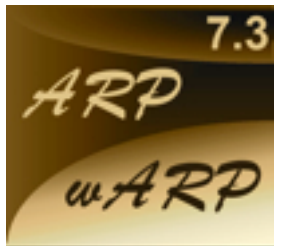
Modelling loops with *a priori* knowledge

Chain fragments are built and **docked to sequence**

- using **knowledge of angles** between anchors
- **check with density** to see if modeled loops are valid

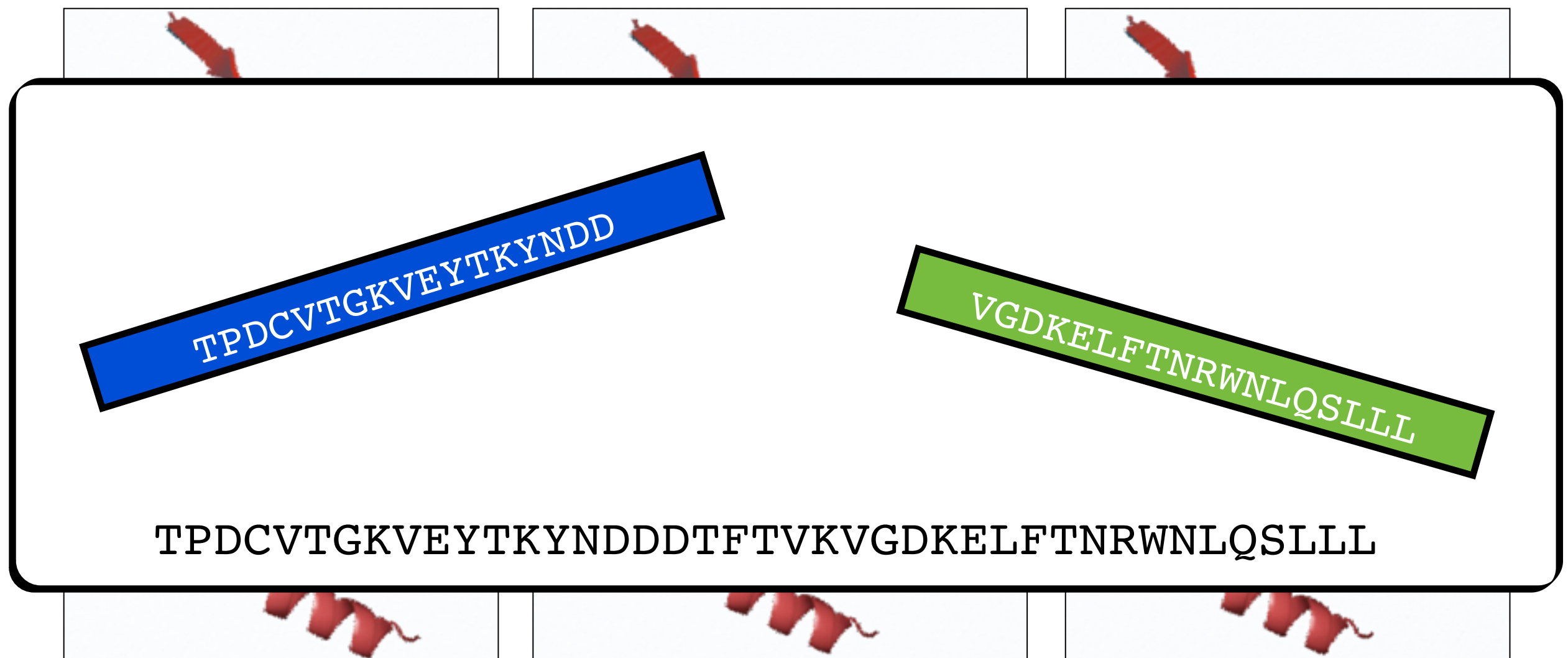


Modelling loops with *a priori* knowledge



Chain fragments are built and **docked to sequence**

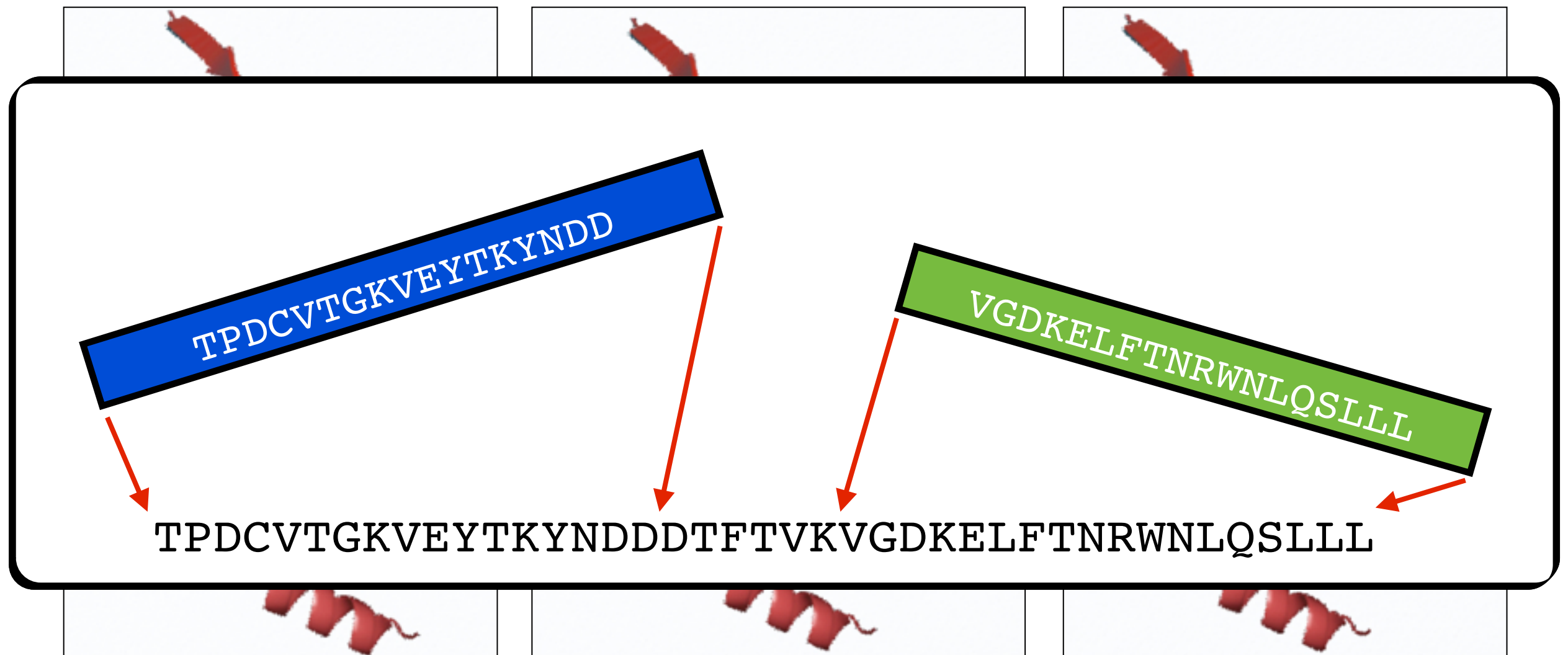
- using **knowledge of angles** between anchors
- **check with density** to see if modeled loops are valid



Modelling loops with *a priori* knowledge

Chain fragments are built and **docked to sequence**

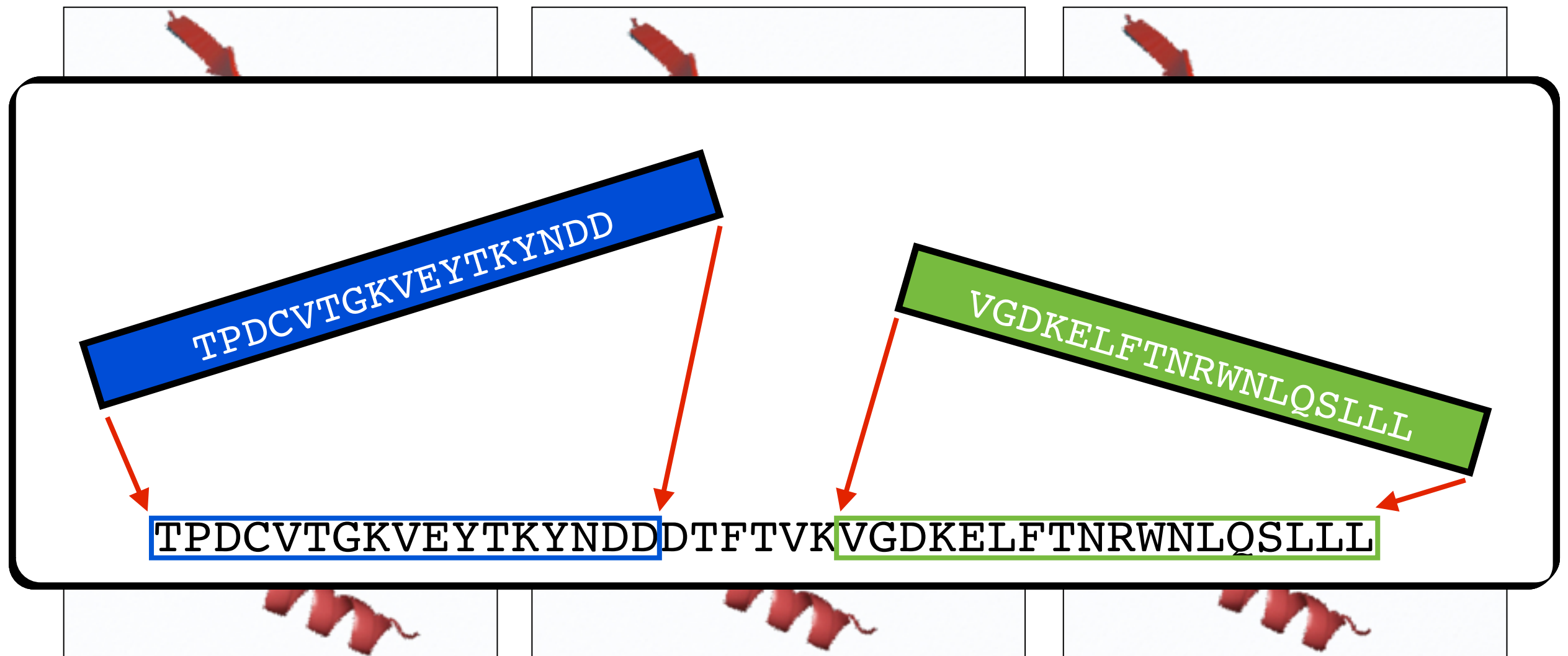
- using **knowledge of angles** between anchors
- **check with density** to see if modeled loops are valid



Modelling loops with *a priori* knowledge

Chain fragments are built and **docked to sequence**

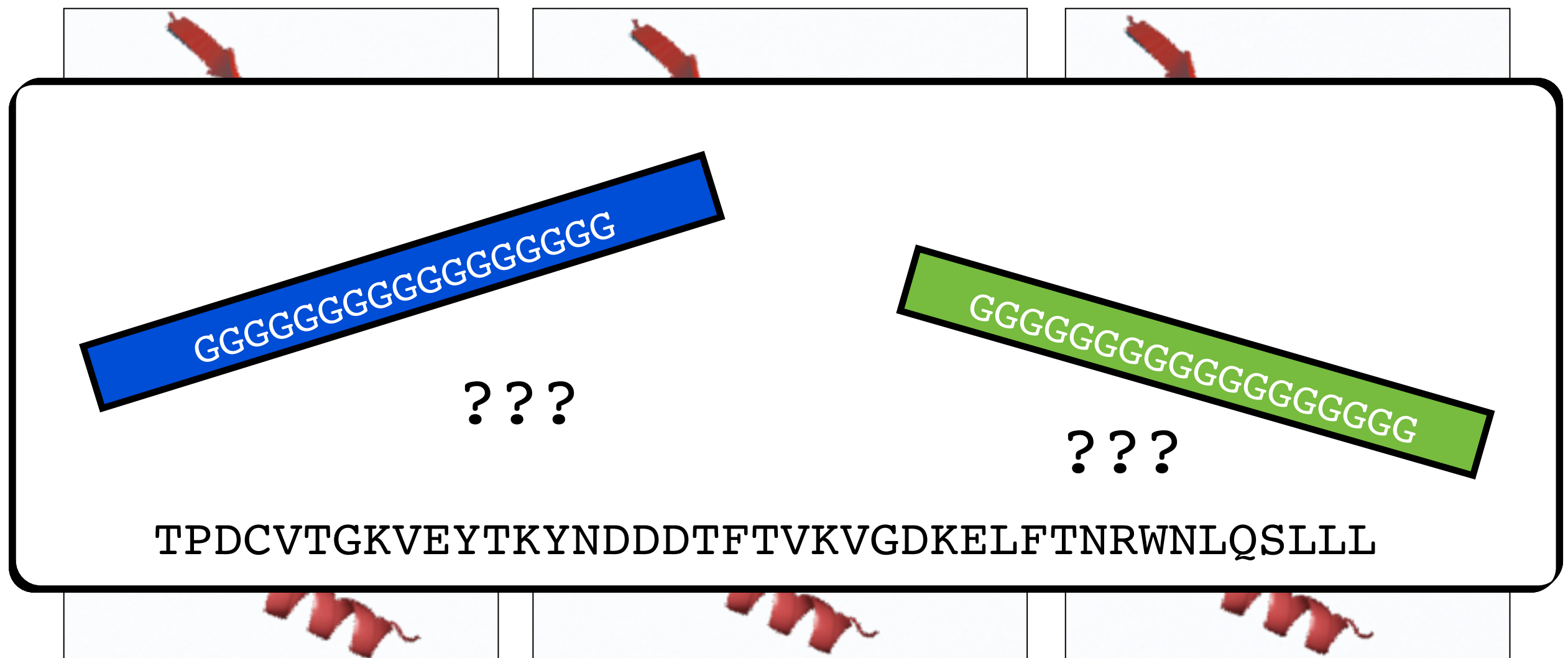
- using **knowledge of angles** between anchors
- **check with density** to see if modeled loops are valid



Modelling loops with *a priori* knowledge

Chain fragments are built and **docked to sequence**

- using **knowledge of angles** between anchors
- **check with density** to see if modeled loops are valid



Use some Bioinformatics?

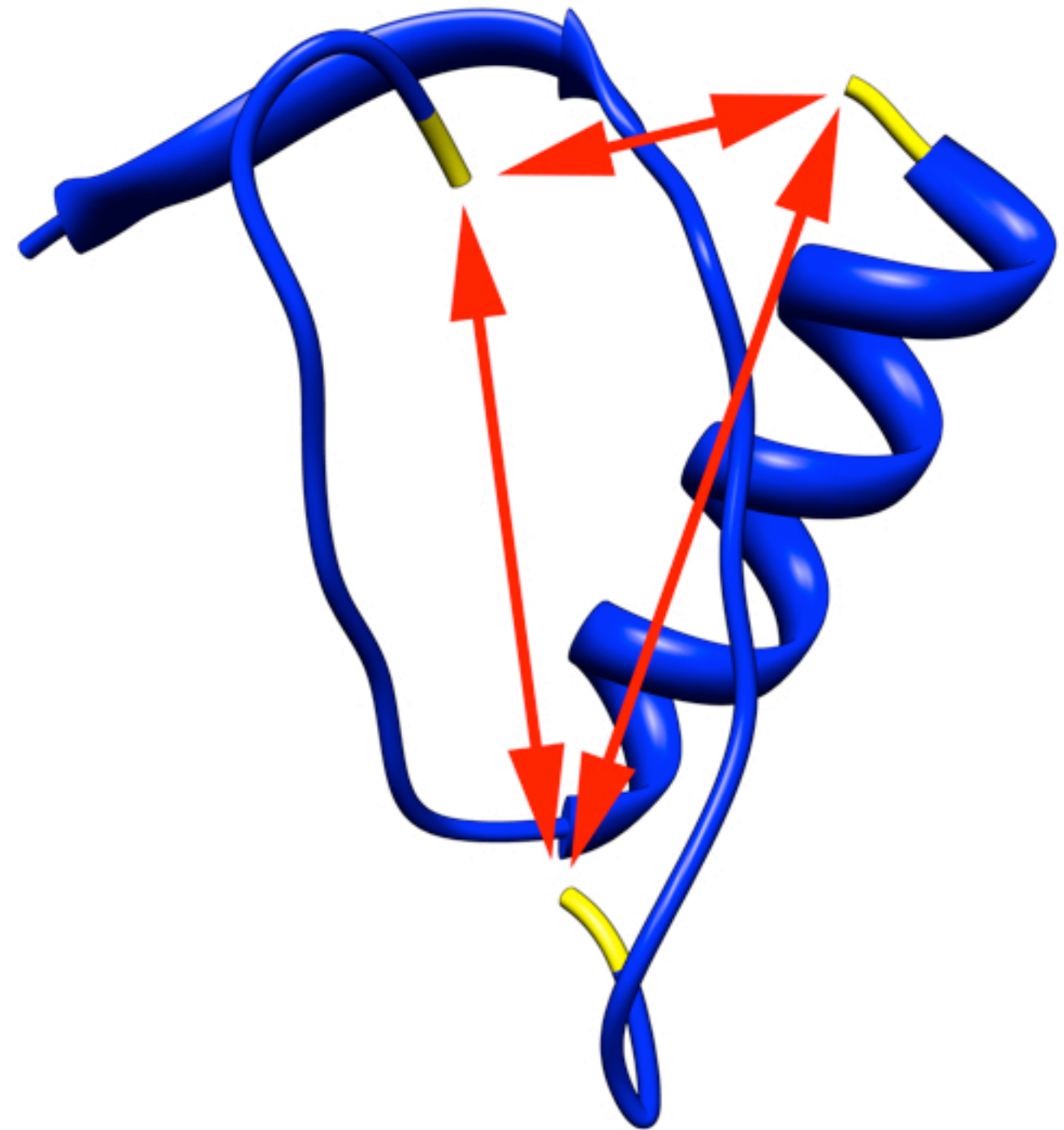


Filling gaps at medium resolution



When **no sequence** has been **docked** - **no loop building!**

- Identify loop regions
 - start / end

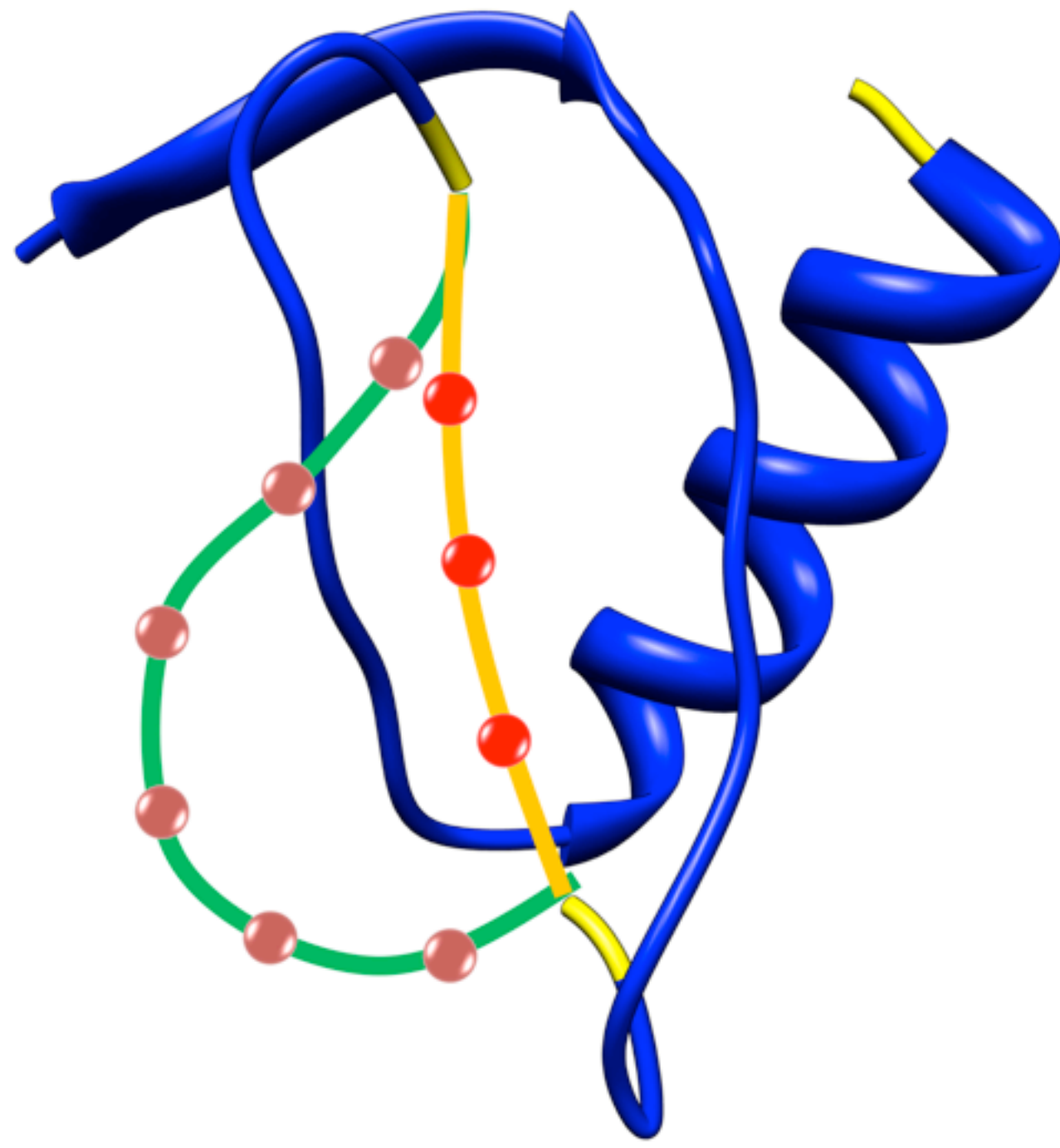


Filling gaps at medium resolution

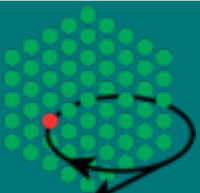
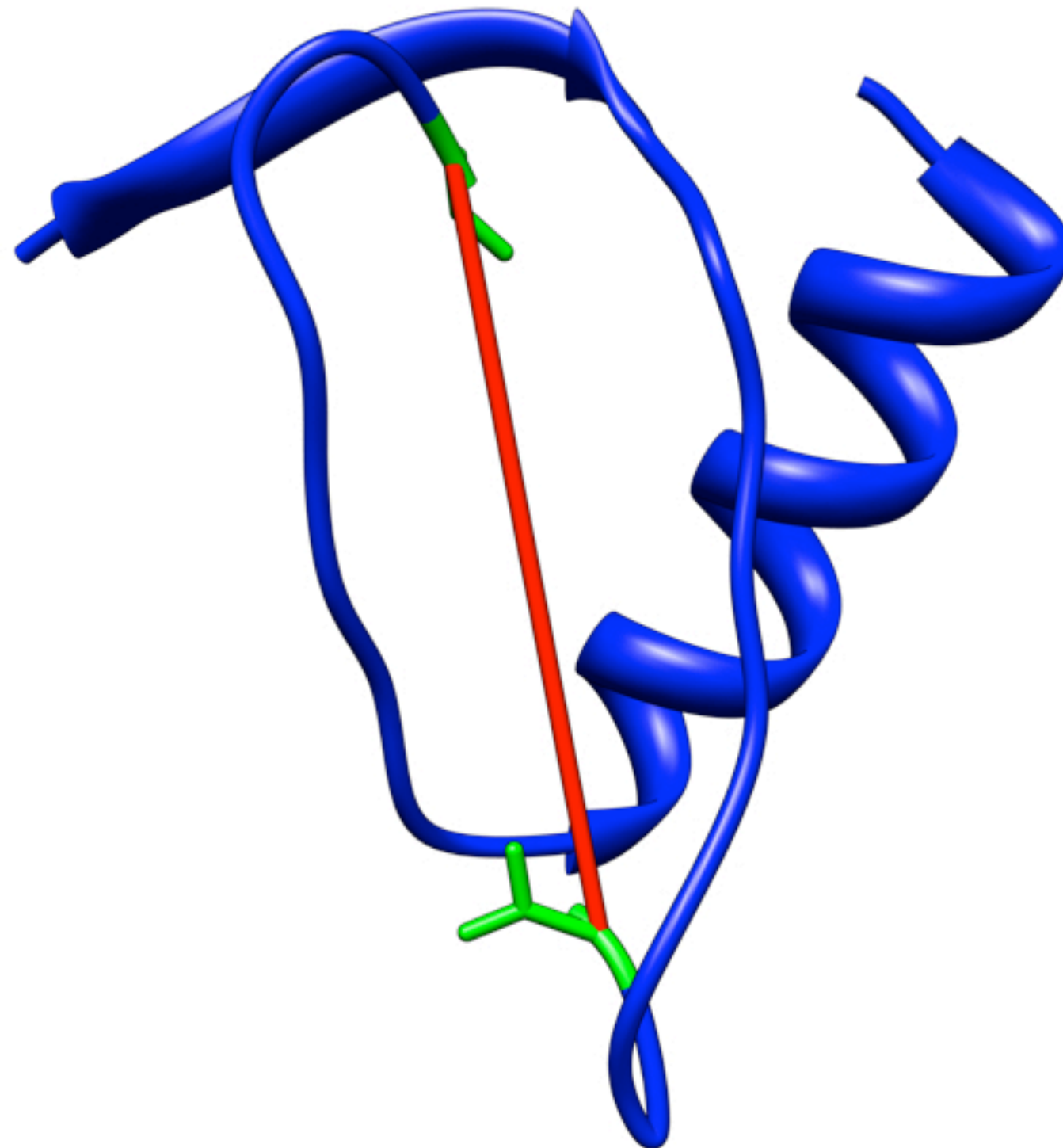


When **no sequence** has been **docked** - **no loop building!**

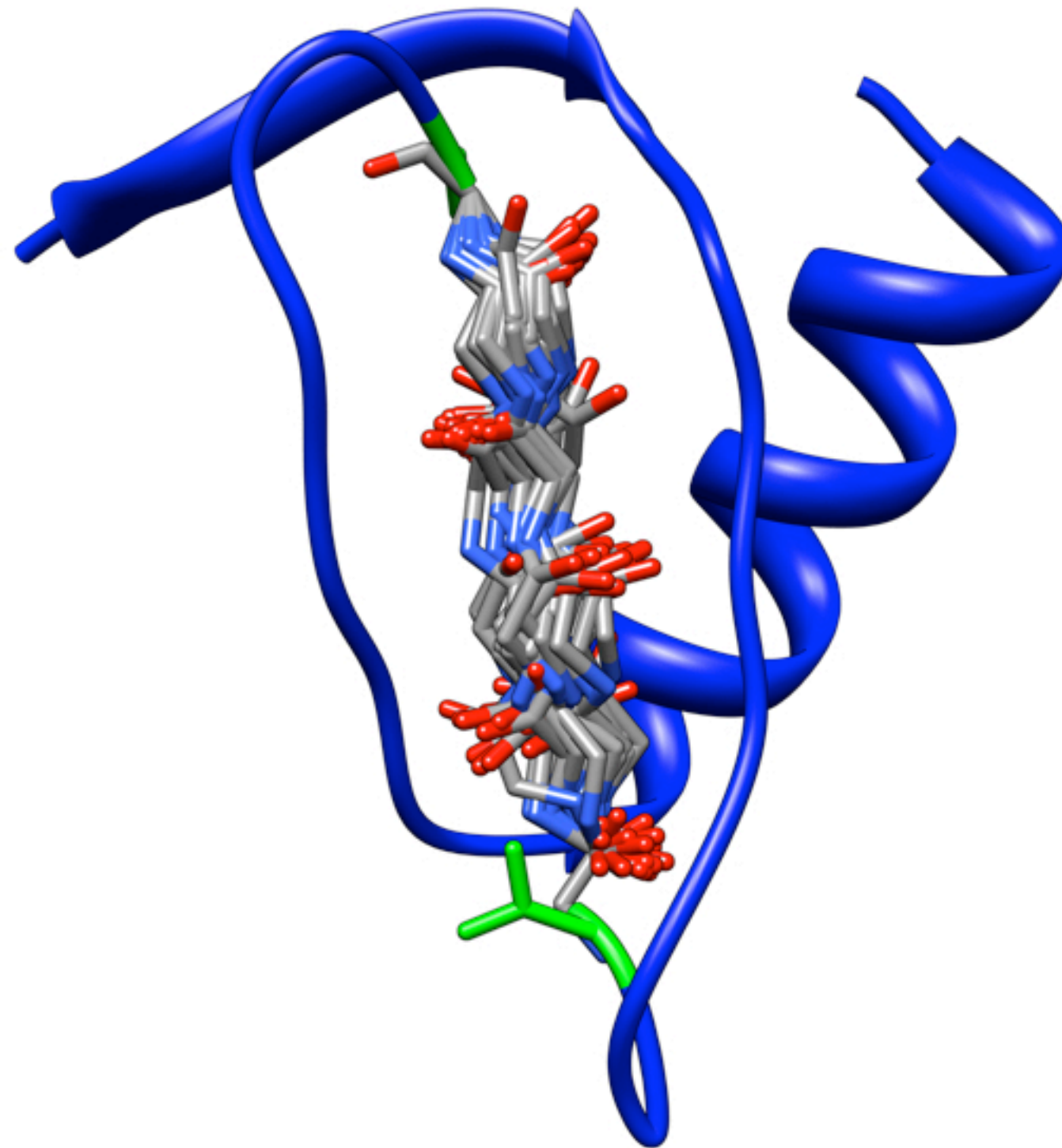
- Identify loop regions
 - start / end
 - length
- Search for **missing fragments** in a **structural database**
- Developed together with Marco Biasini (Schwede group, SWISSmodel)
 - Using the given **density information**



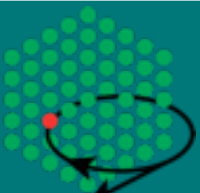
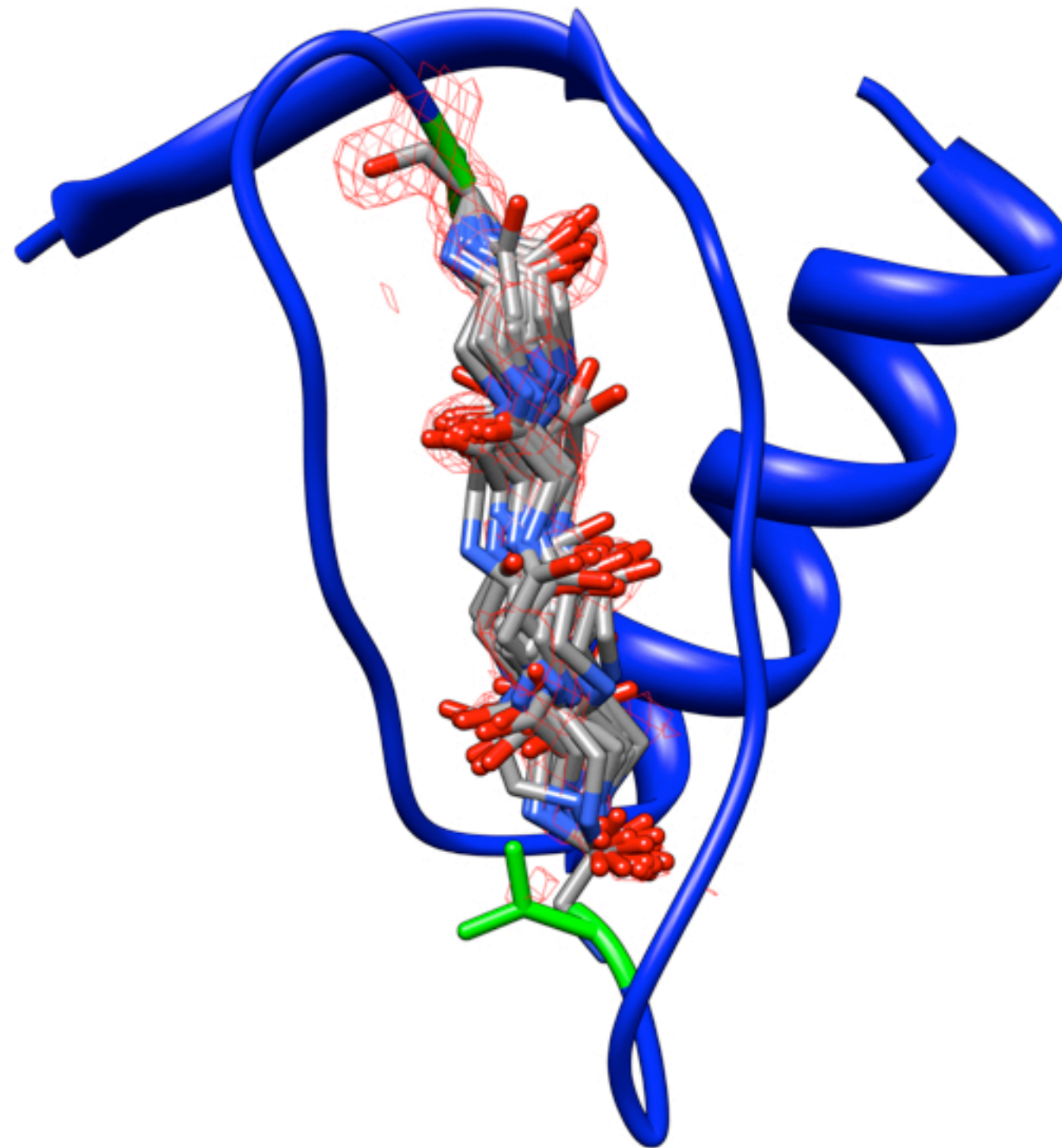
Filling gaps at medium resolution



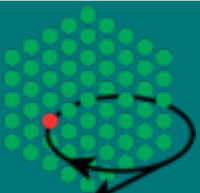
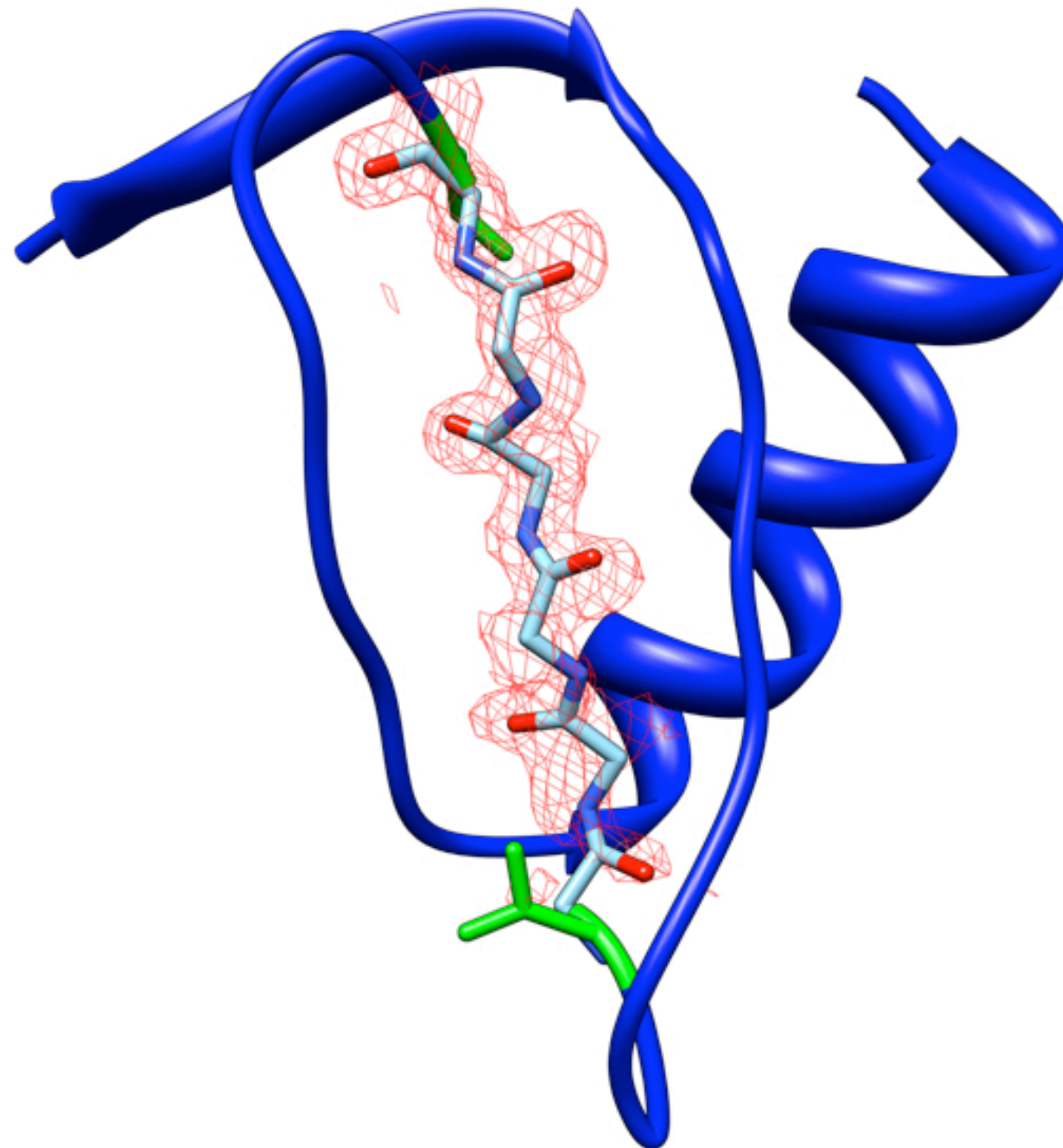
Filling gaps at medium resolution



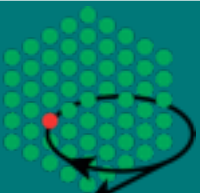
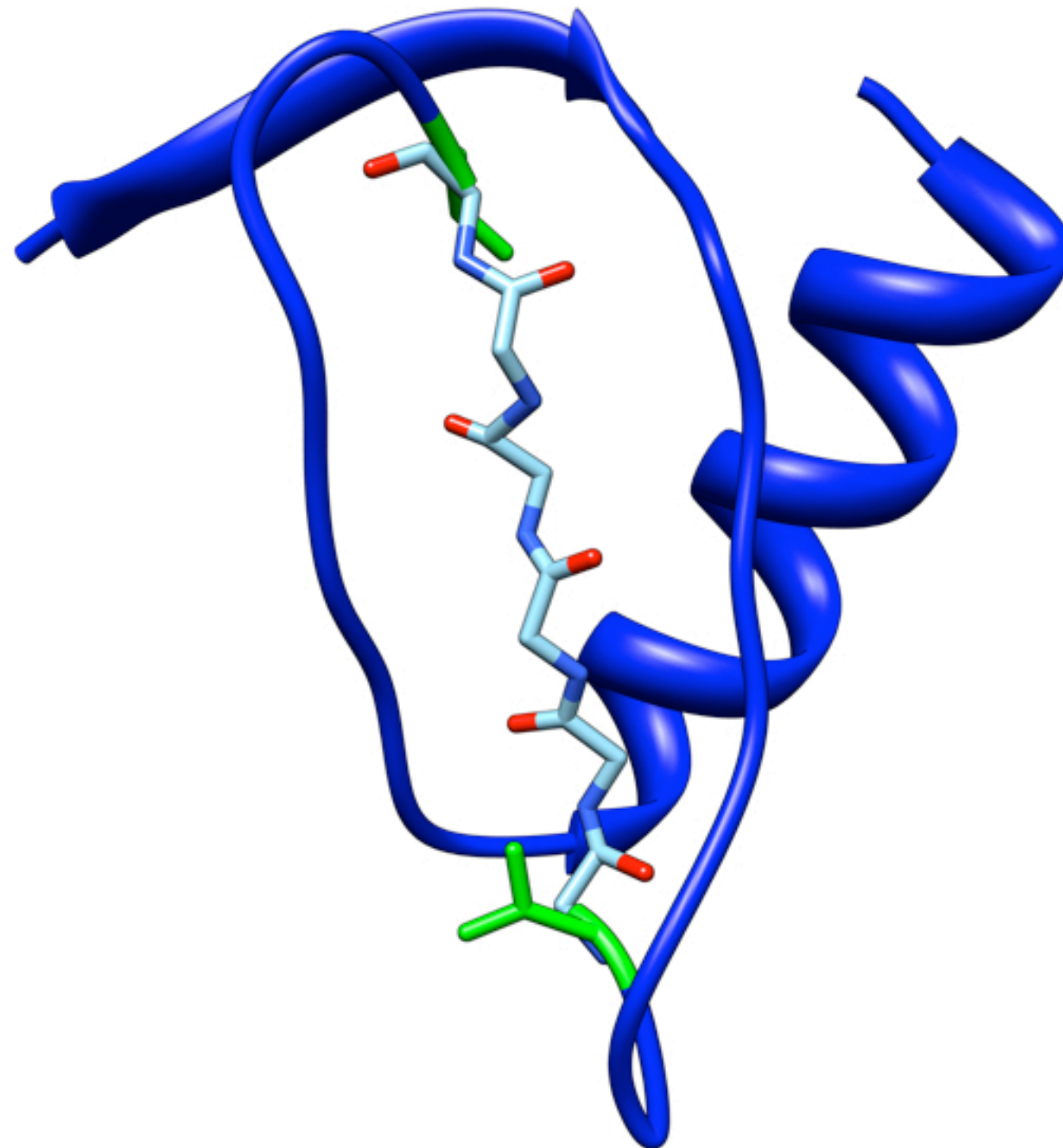
Filling gaps at medium resolution



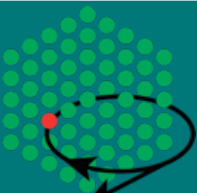
Filling gaps at medium resolution



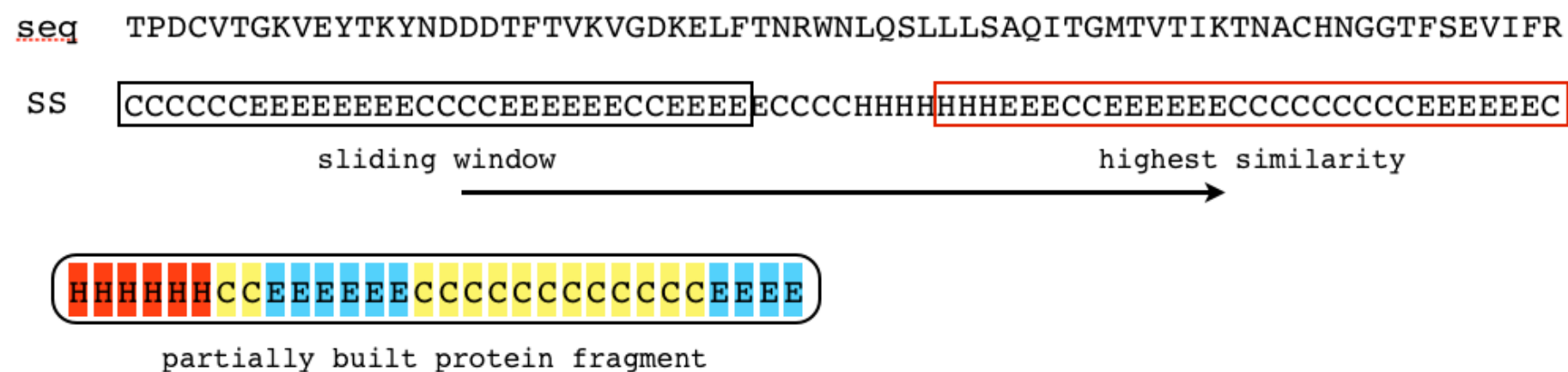
Filling gaps at medium resolution



FittOFF Workflow



Docking results

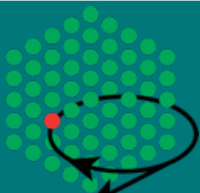
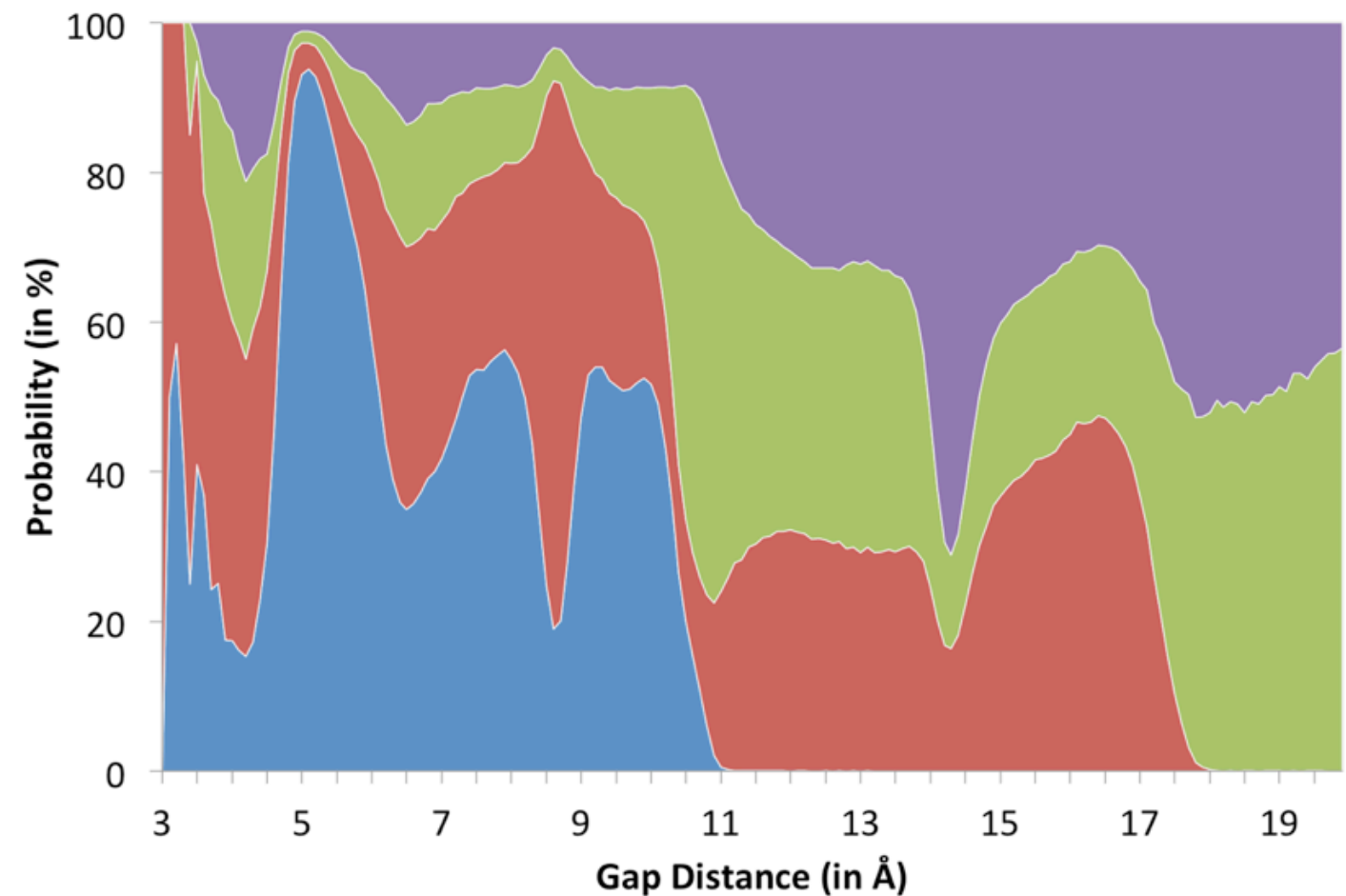


FittOFF Workflow

8.0
ARP
wARP

Docking results

Distance statistics



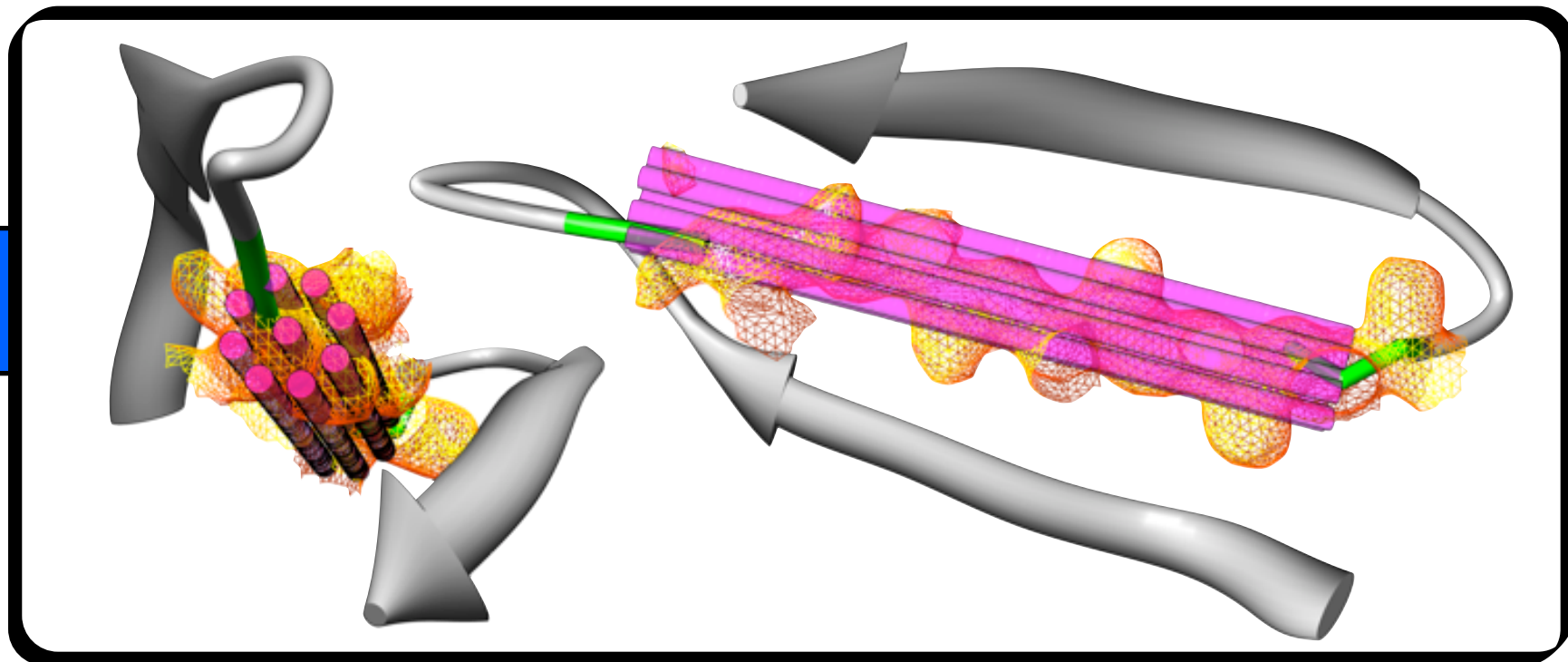
FittOFF Workflow



Docking results

Distance statistics

Density check



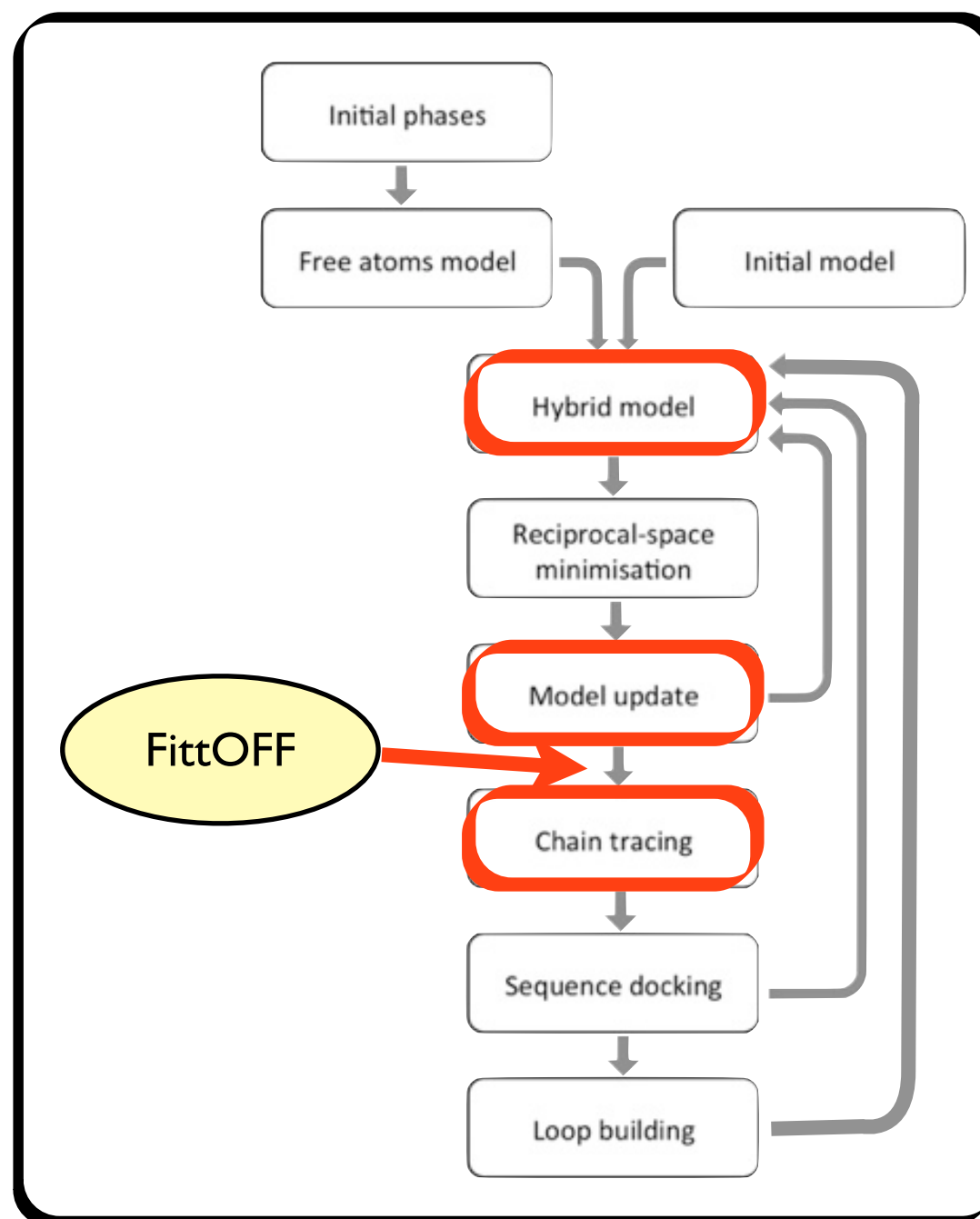
FittOFF Workflow



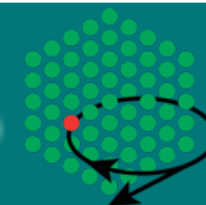
Docking results

Distance statistics

Density check



FRAGRA



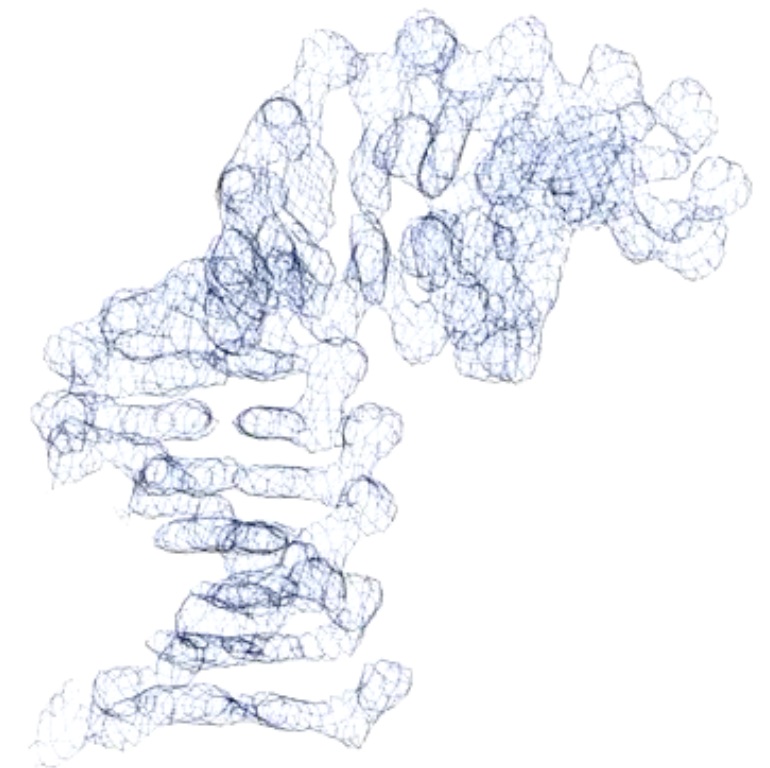
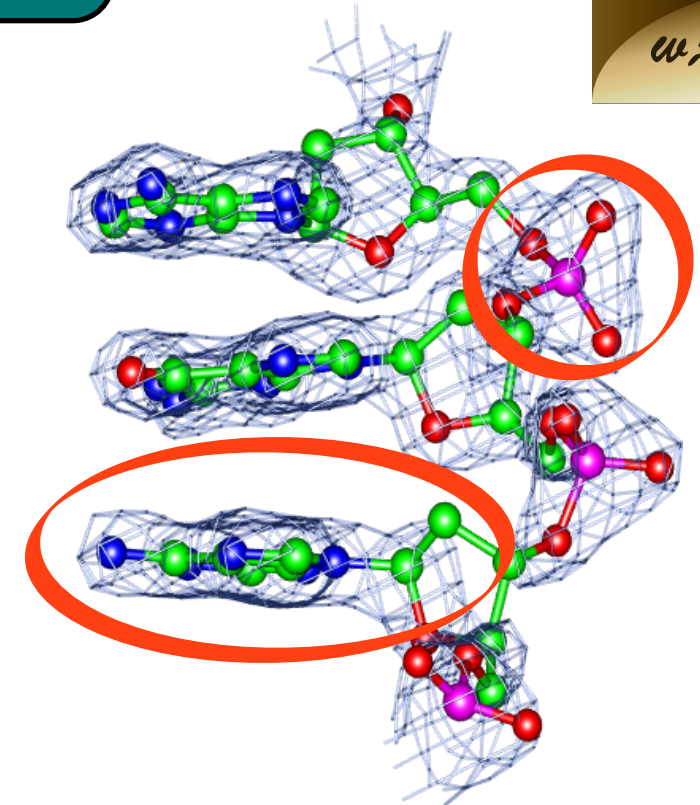
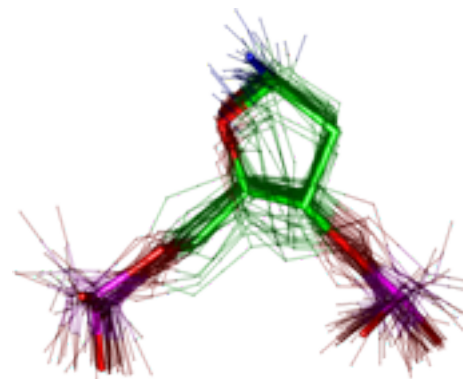
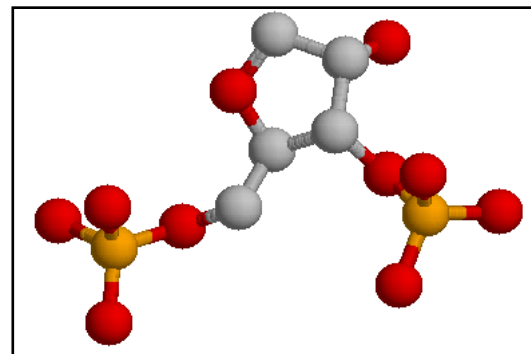
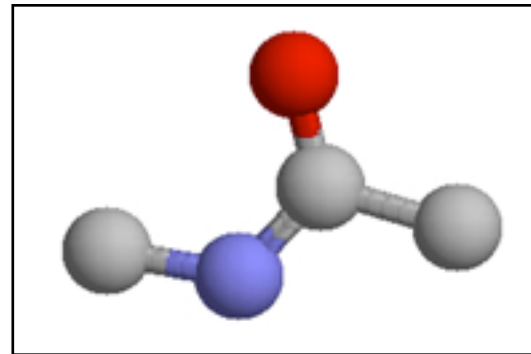
Docking results

10 protein test structures with resolution 3.0 to 3.8 Å, up to 300 residues

	Model Completeness	Residues / Fragment	Sequence Coverage
Top Results	+ 12%	+ 98%	+ 59%
Average	+ 4%	+ 25%	+ 7%

Tracing RNA/DNA Chains

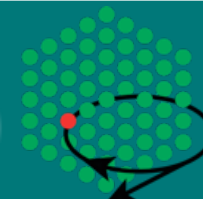
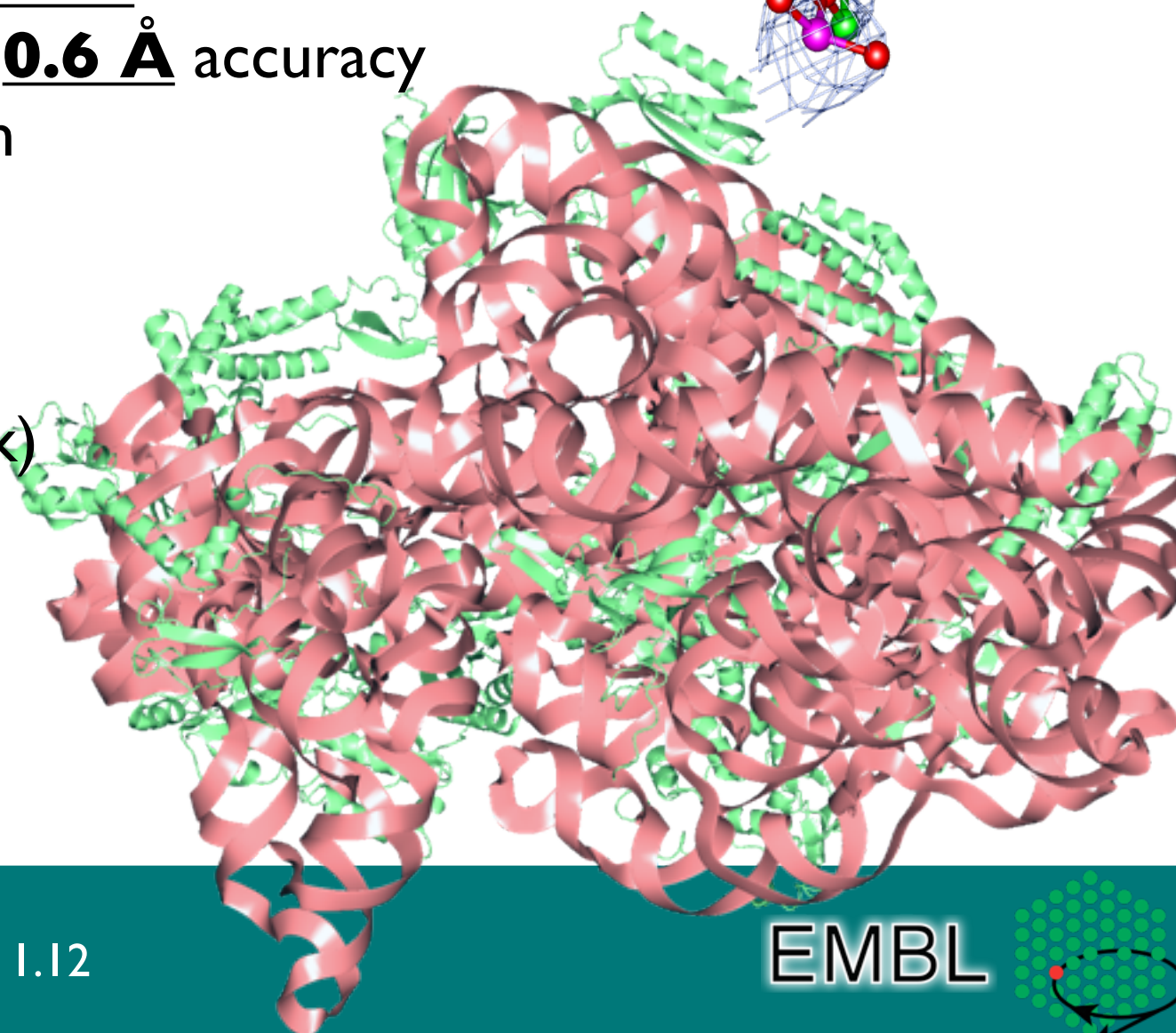
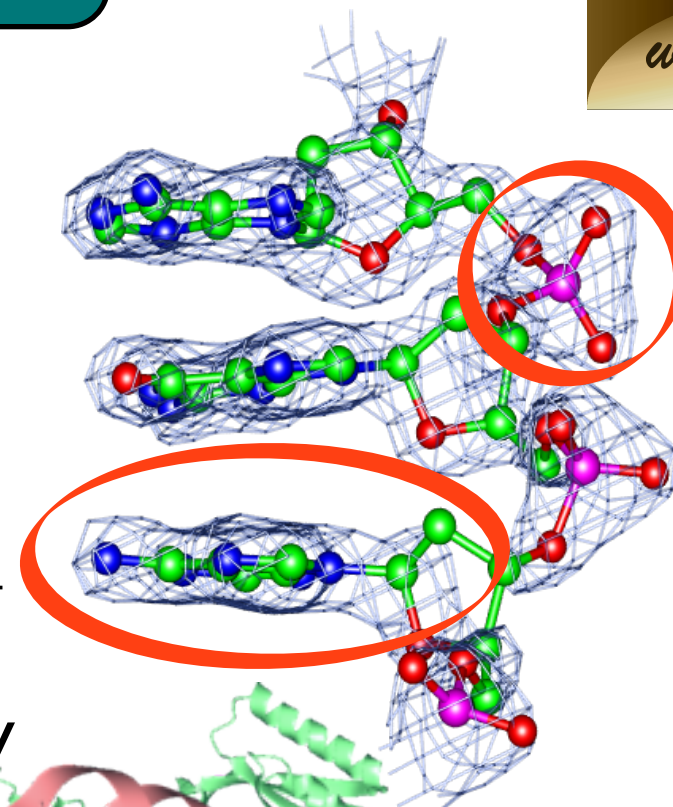
7.3
ARP
wARP



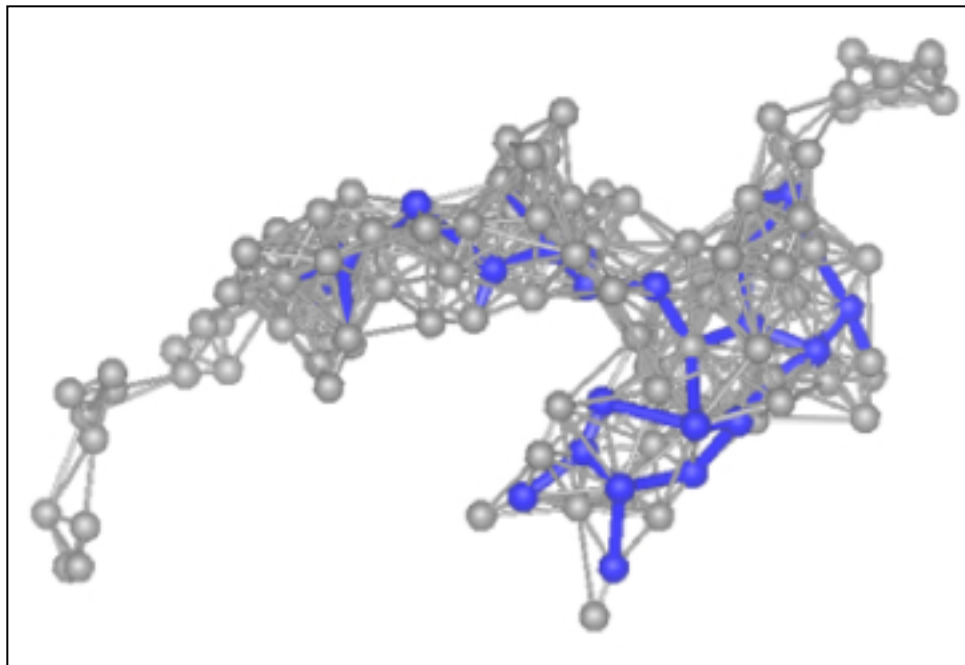
Tracing RNA/DNA Chains

7.3
ARP
wARP

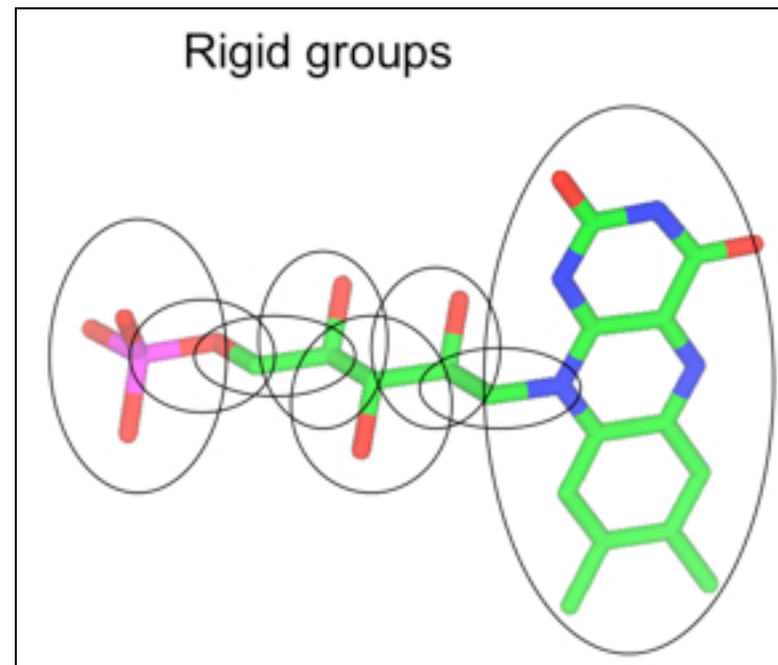
- The **30S ribosomal subunit**
 - Resolution: 3.05 Å
 - Experimental phases (MCC 0.73)
- Automatic model building with ARP/wARP
 - modelled 1,302 out of 1,513 nucleotides (**86%**)
 - backbone r.m.s.d. to reference: **0.7 Å**
 - Located 1,121 nucleobases with **0.6 Å** accuracy in location and 12° in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



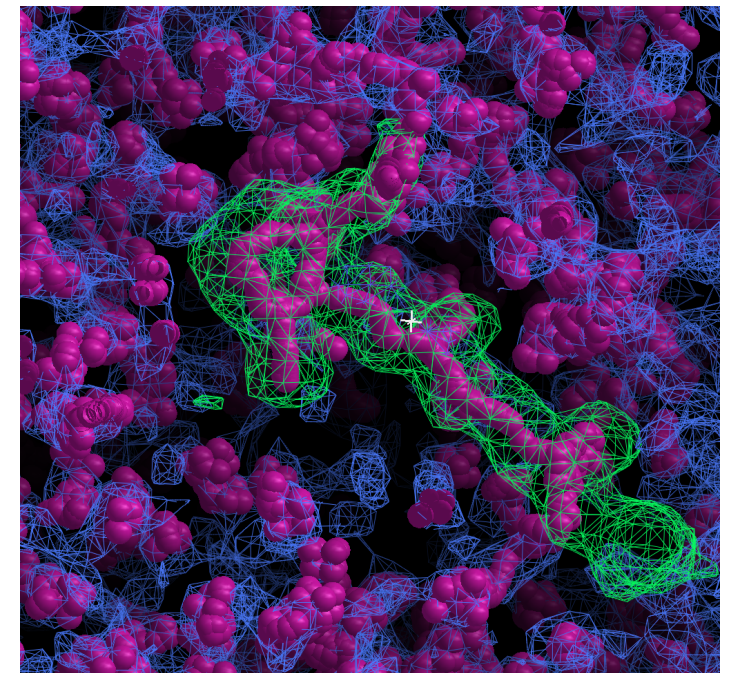
Sparse grids



Conformational fit

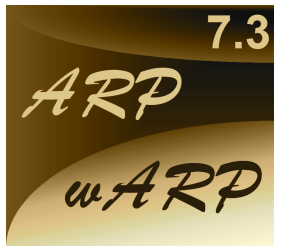


Fine skeletons

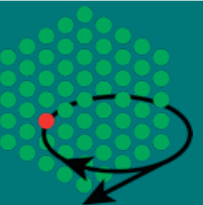
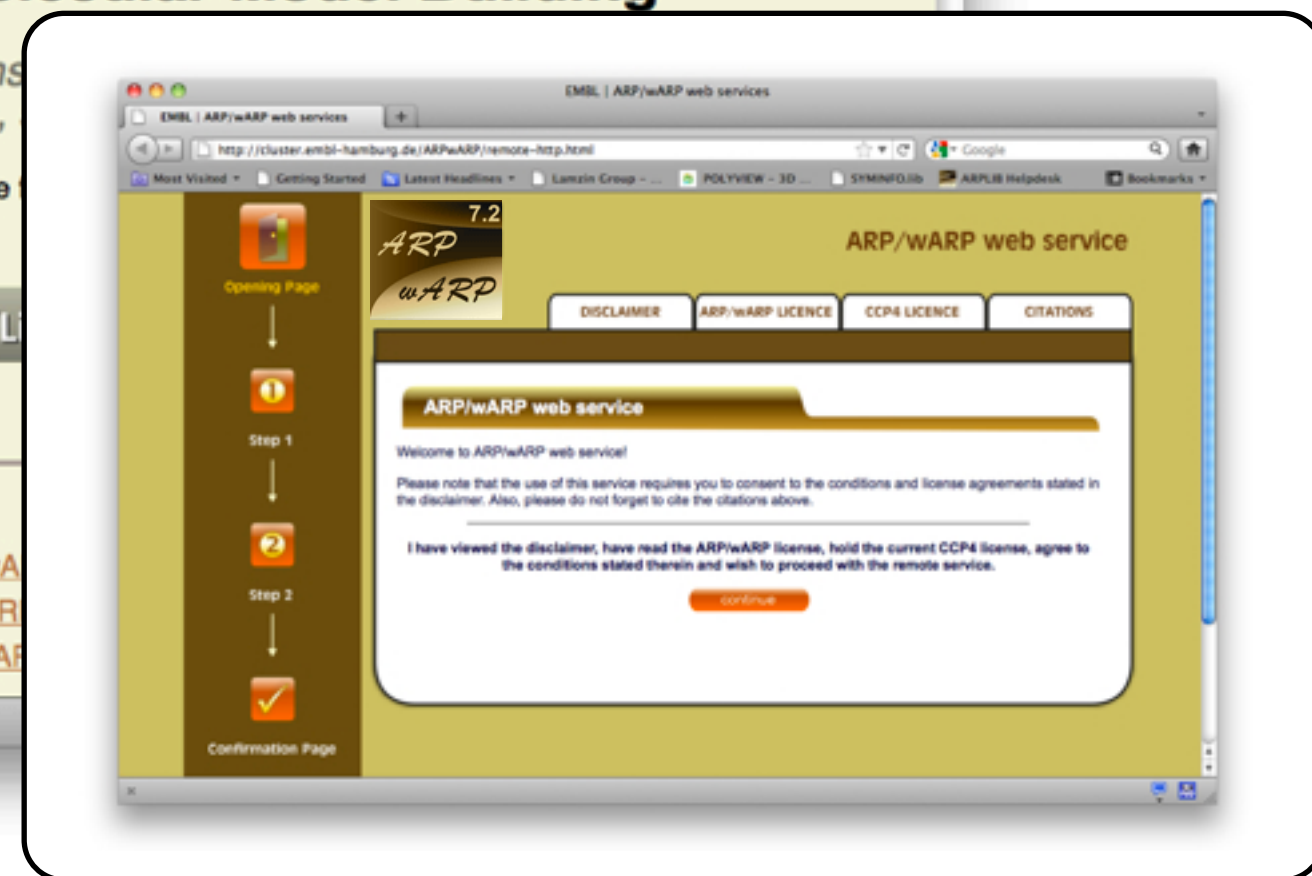
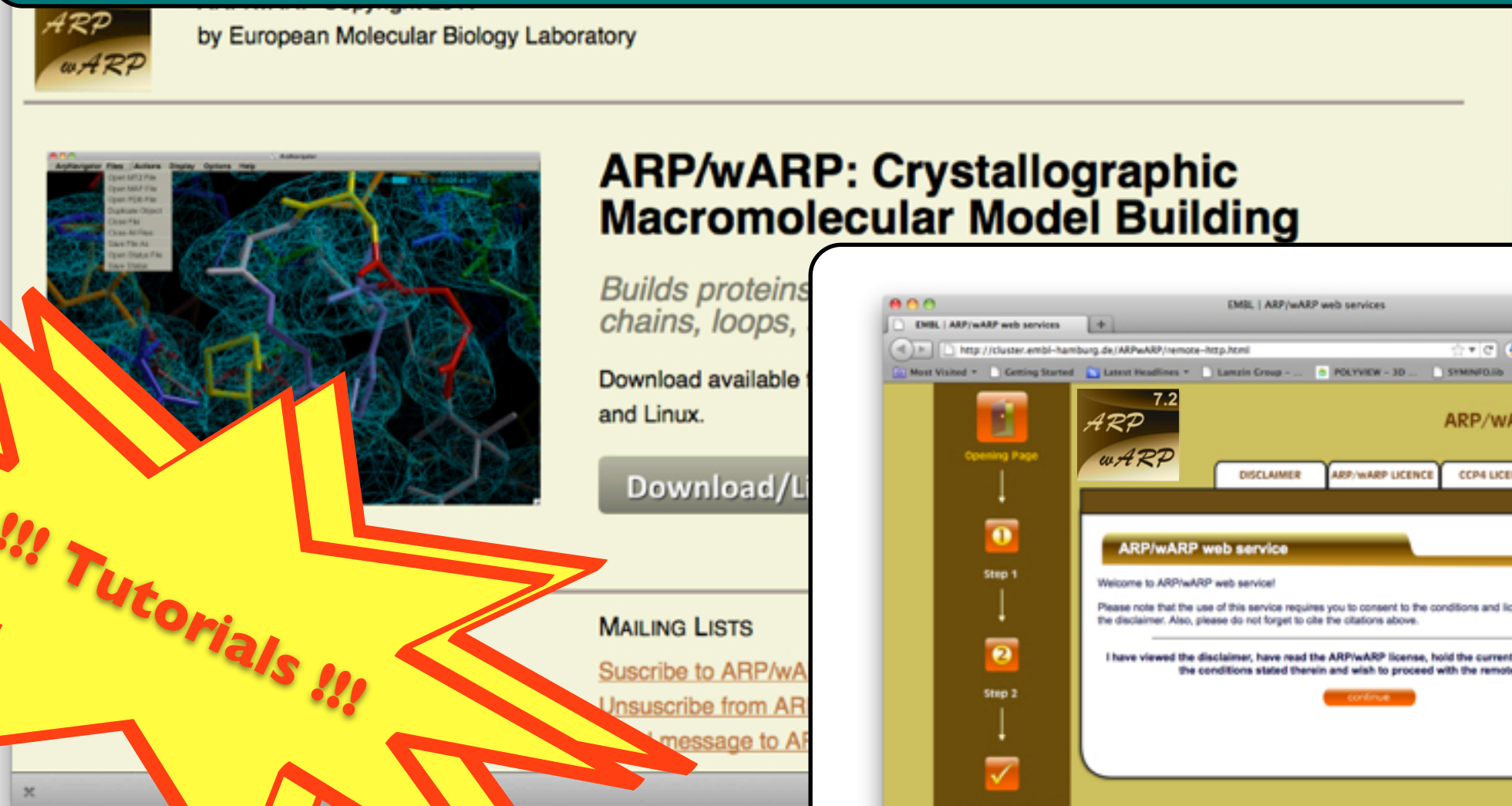


All about ligands tomorrow from Ciaran!

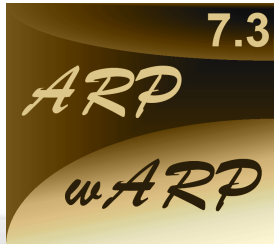
Where Can I Get This???



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



...or...



CCP4 Download

http://www.ccp4.ac.uk/downloads/download-test/#os=mac

Kino-Fahrpl...alfassungen DELUGE (FRE...rian Keene ###RMSD calc### EMBL | ARP/...eb services Zingzing TV ...in TV Guide Macbook - ...Forums.com RCSB Protein Data Bank EMBL Intern...hips - EMBL Everything ...now Leopard

CCP4

CCP4 Download pages

Mac OS X GNU/Linux MS Windows Generic (source code only)

Binary downloads for Apple Mac OS X

Binary packages for Apple Mac OS X are available below. [ARP/wARP](#) requires a separate license from EMBL-Hamburg, so you will first be redirected to their website.

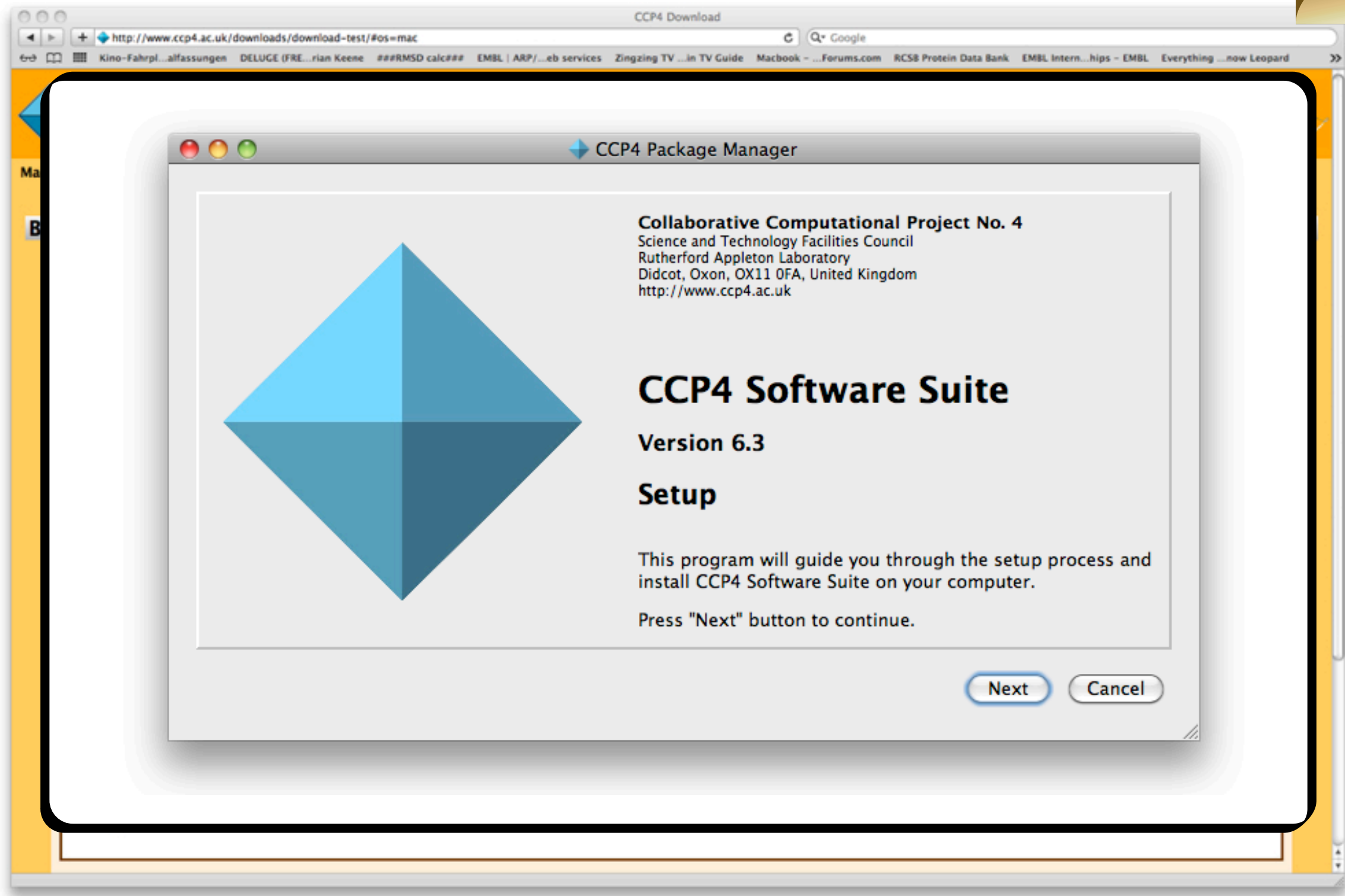
	CCP4 Program Suite v6.3.0 and ARP/wARP v7.3	Agree to License: →	
	CCP4 Program Suite v6.3.0	Agree to License: →	
	ARP/wARP v7.3	Agree to License: →	
	Database for Balbes (Data Files)	Download Now! ↓	Download size: 1172 MB
	COOT v0.6.2	Download Now! ↓	Download size: 292 MB
	CCP4 Molecular Graphics v2.6.0	Download Now! ↓	Download size: 339 MB

Click [here](#) for the MD5 checksums of all files.

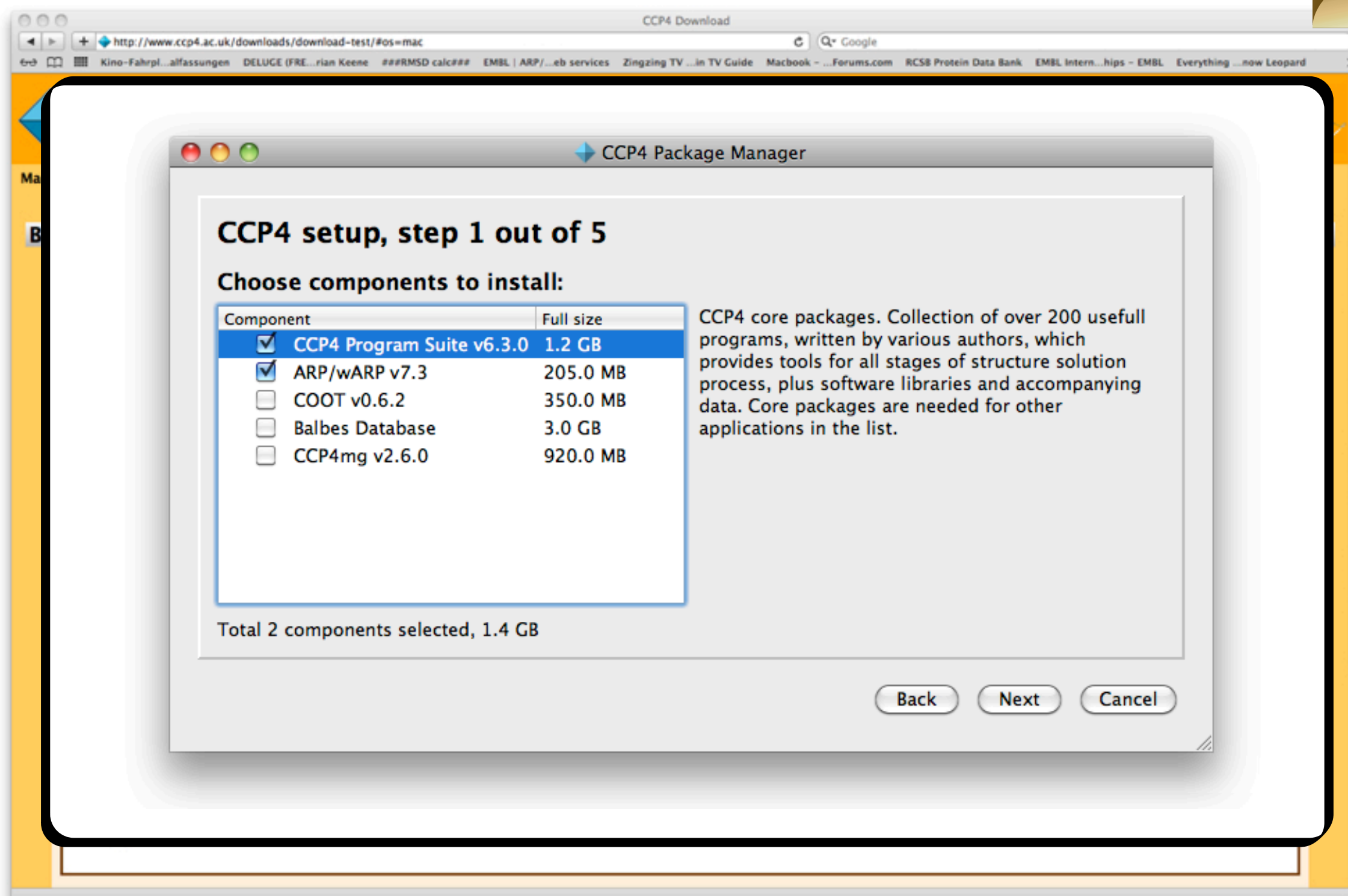
CCP4 Software Suite may also be configured, downloaded and automatically installed with the **Package Manager** :

	Package Manager	Download Now! ↓	Download size: 14 MB
---	-----------------	-----------------	----------------------

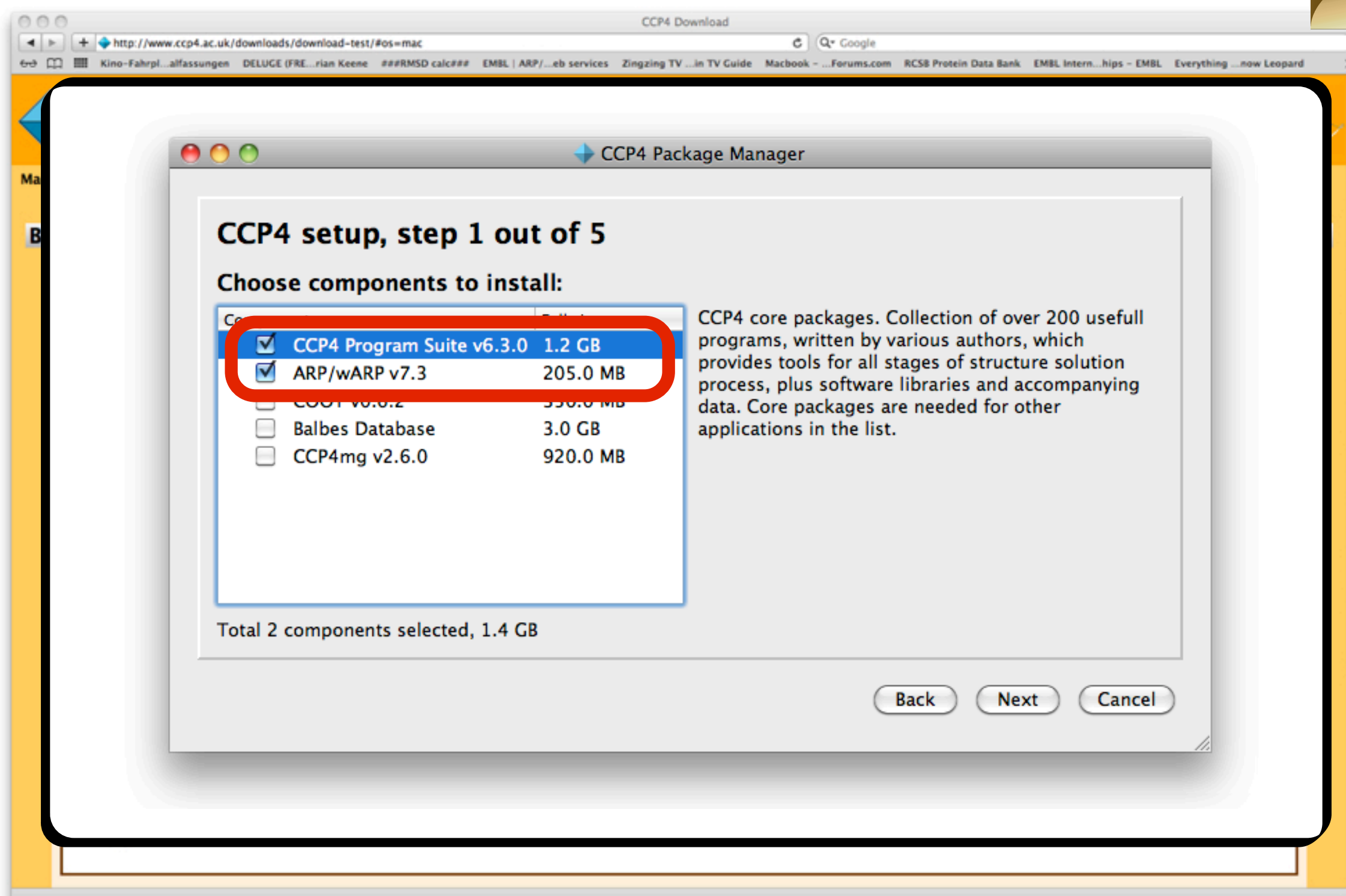
...or...



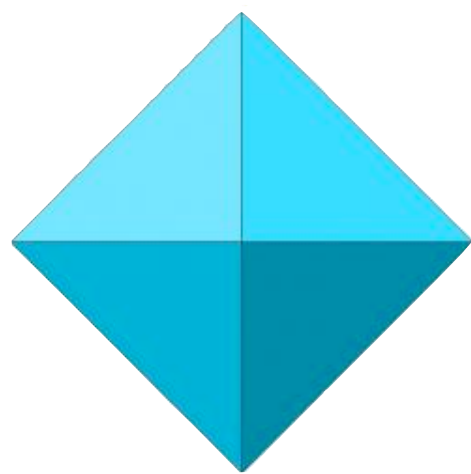
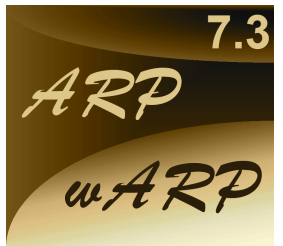
...or...



...or...



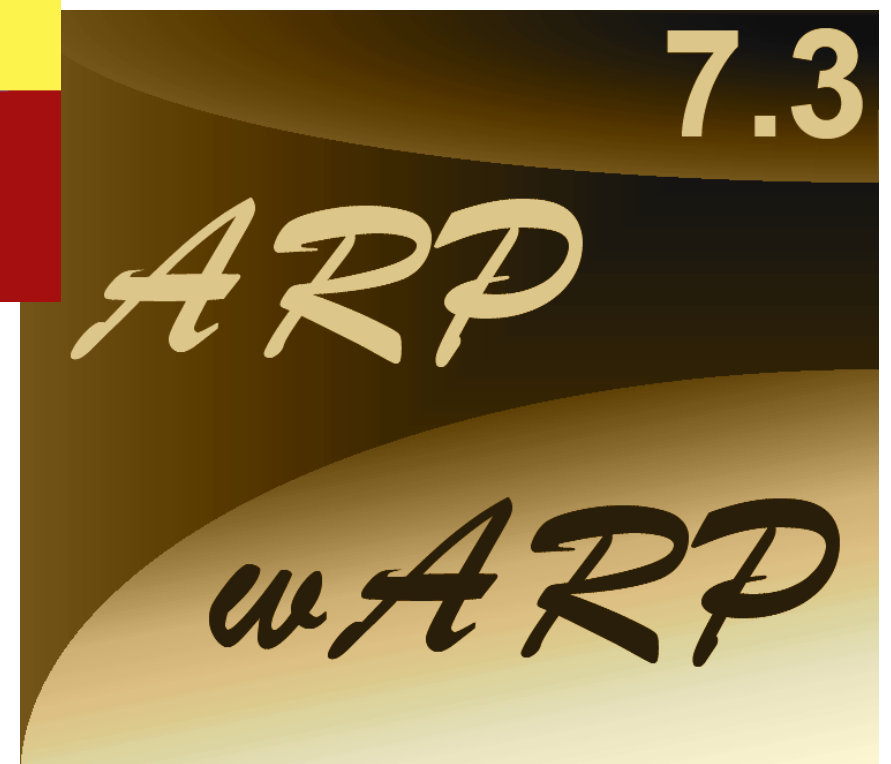
Collaboration with CCP4



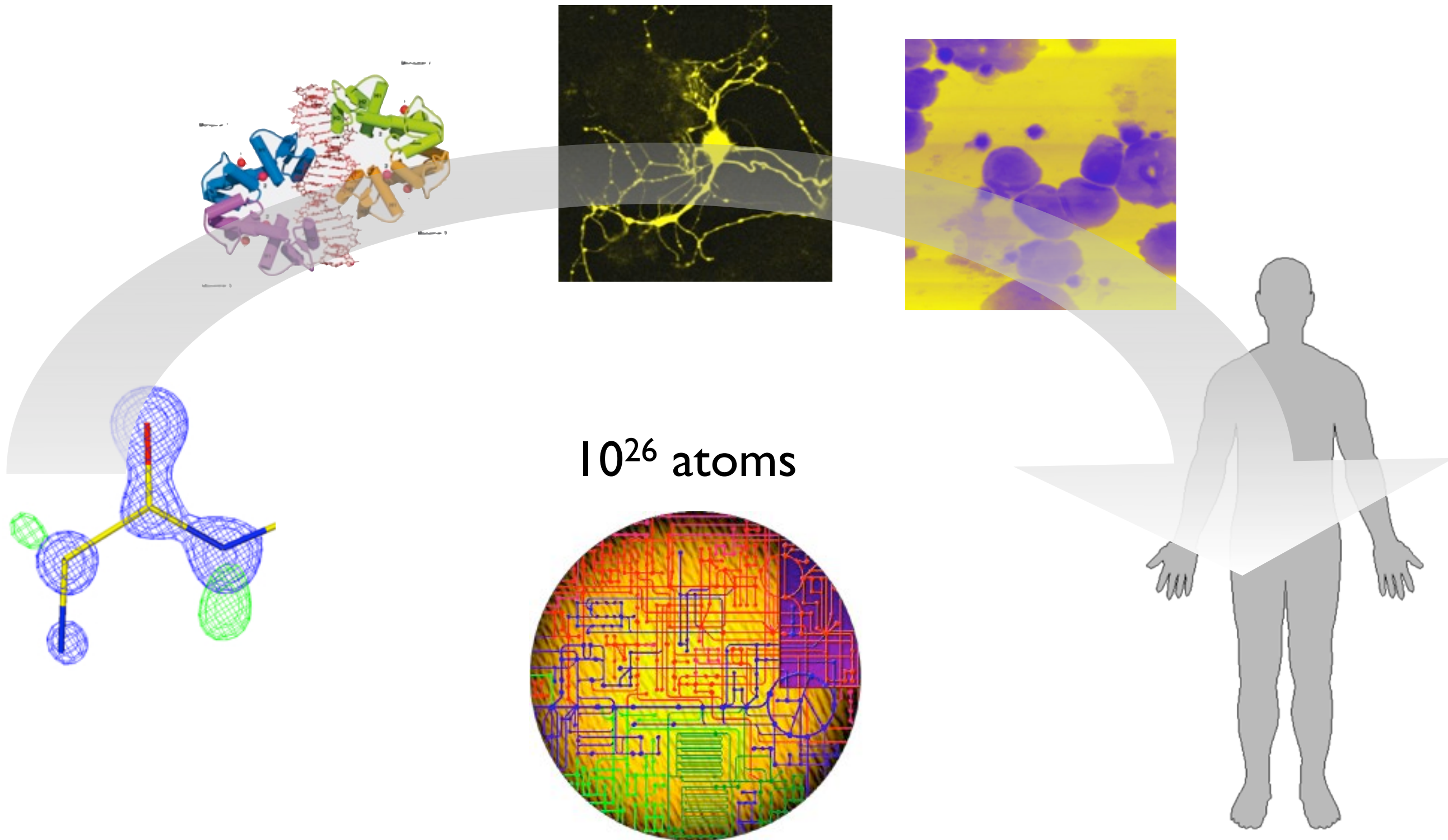
CCP4



- Joint release
 - CCP4 6.3 and ARP/wARP 7.3
 - one installer, no worries ;)



Towards structural systems biology



The people - the power!



Developers

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

NKI Amsterdam: Krista Joosten, Tassos Perrakis

Former ARP/wARP team members

Collaborators

EMBL Hamburg: Gleb Bourenkov, Santosh Panjikar

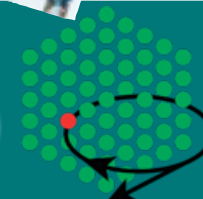
MRC Laboratory, Cambridge : Garib Murshudov's group

Rutherford Appleton Laboratory: CCP4 team

Funding



Bundesministerium
für Bildung
und Forschung



...if you are ever in town

Thank You for Listening!!!

wiegels@embl-hamburg.de
carolan@embl-hamburg.de
victor@embl-hamburg.de