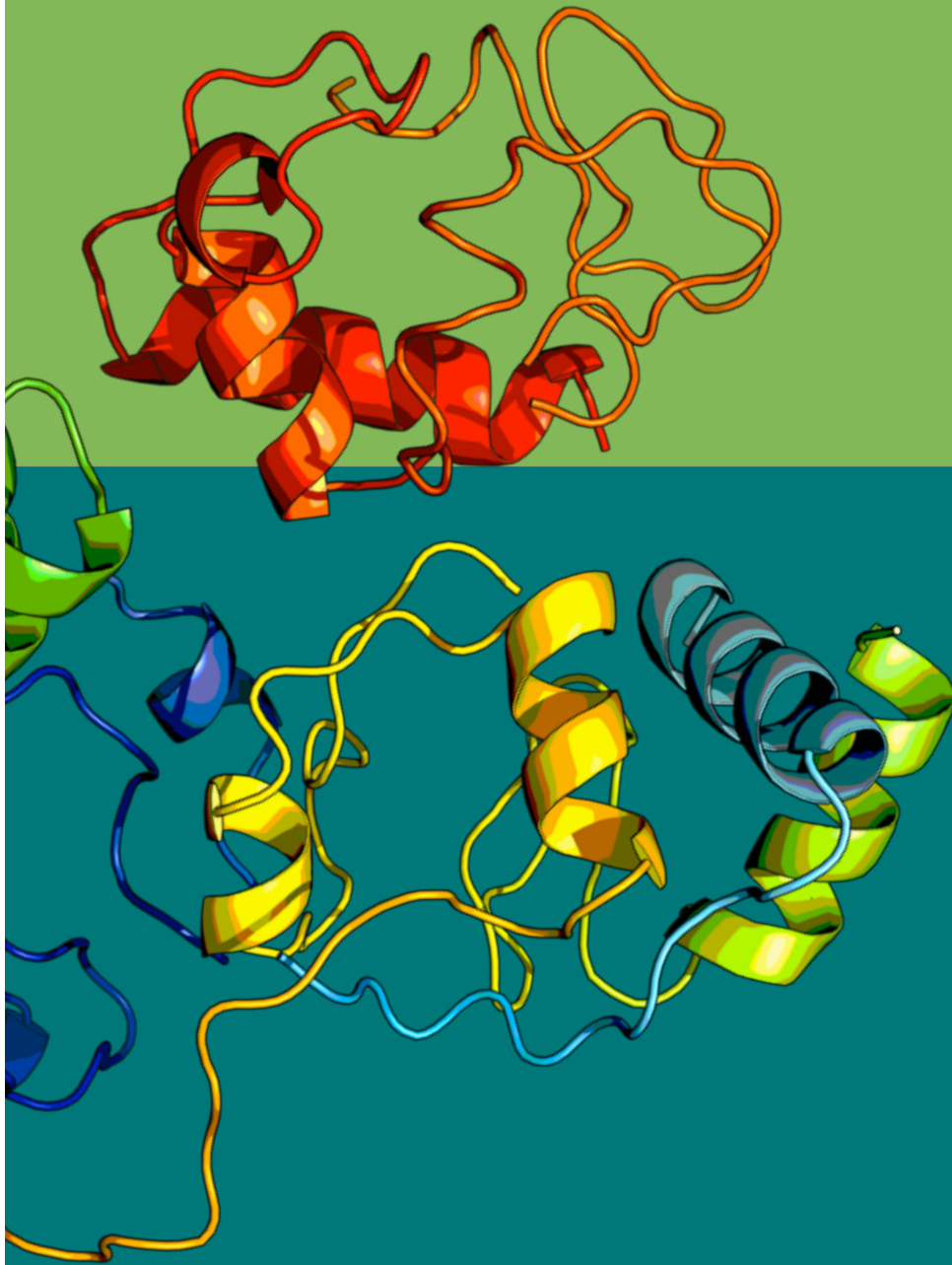
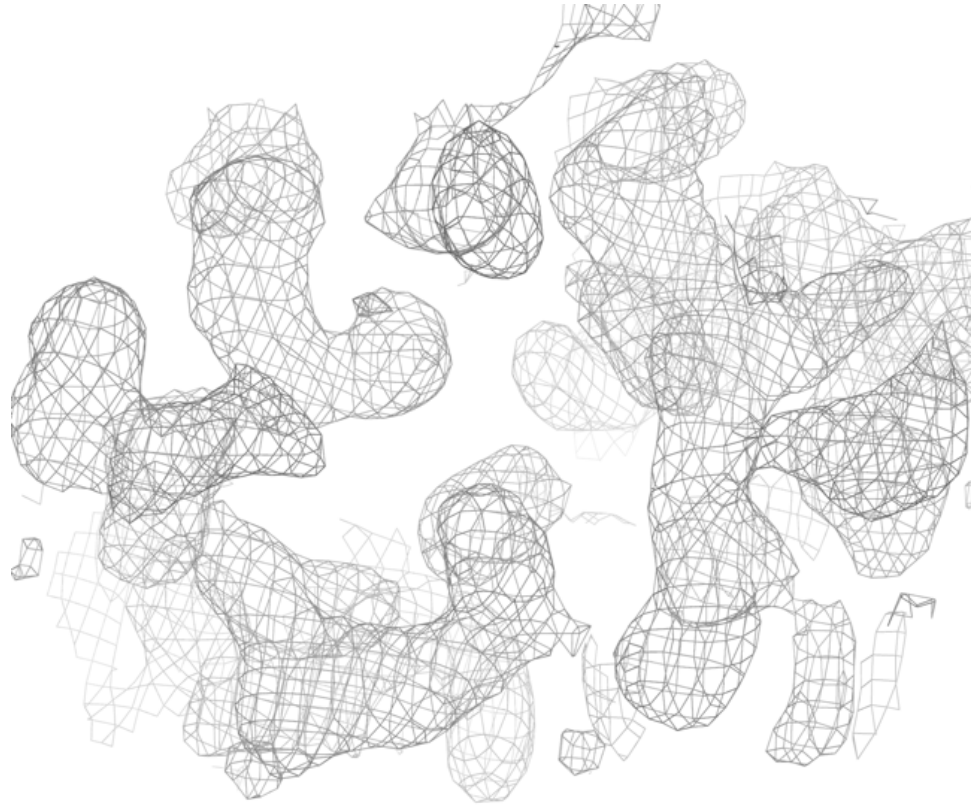


Grzegorz Chojnowski
Lamzin Group
EMBL Hamburg, Germany

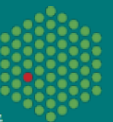


Automated Model Building with ARP/wARP

Electron density map interpretation

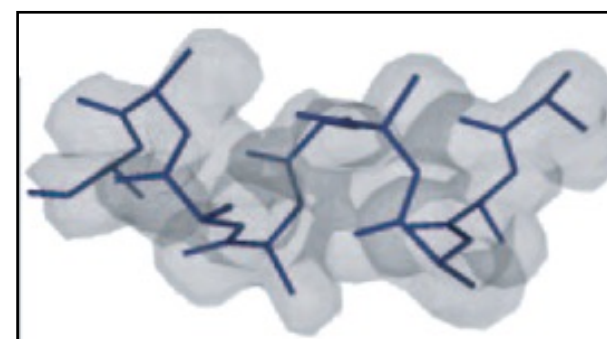
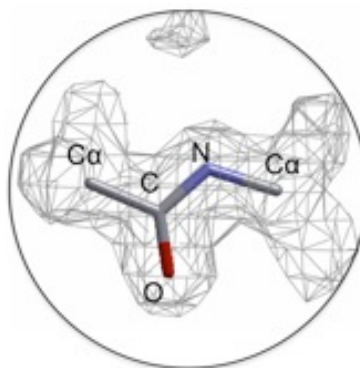
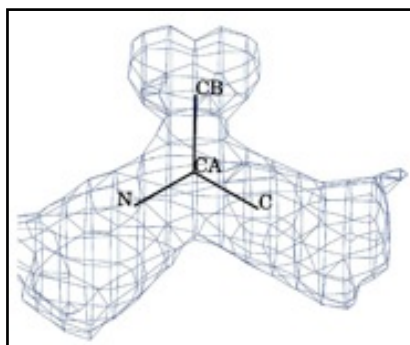
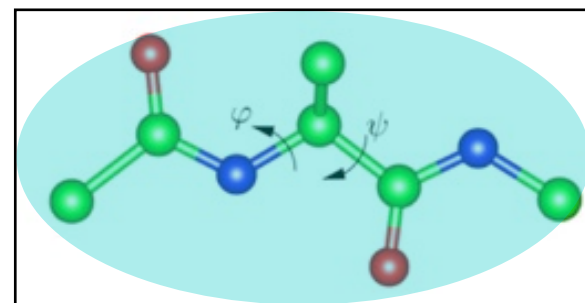
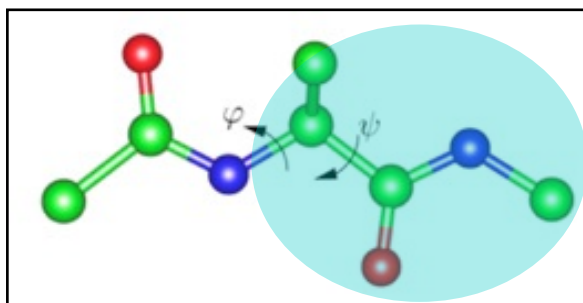
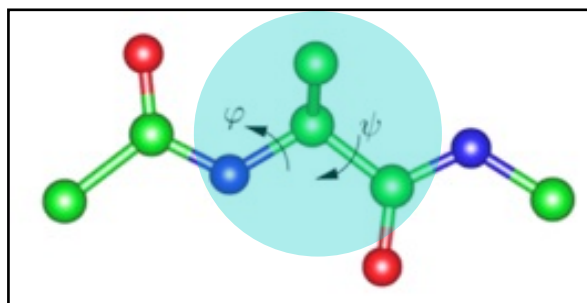


Electron density map interpretation



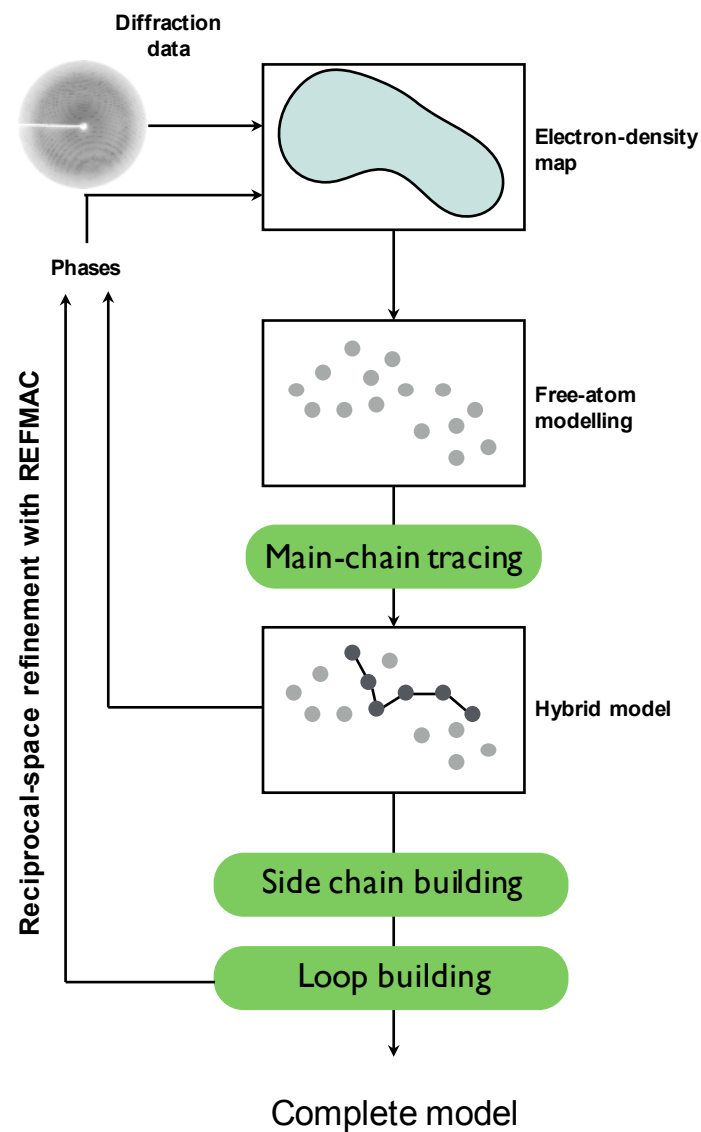
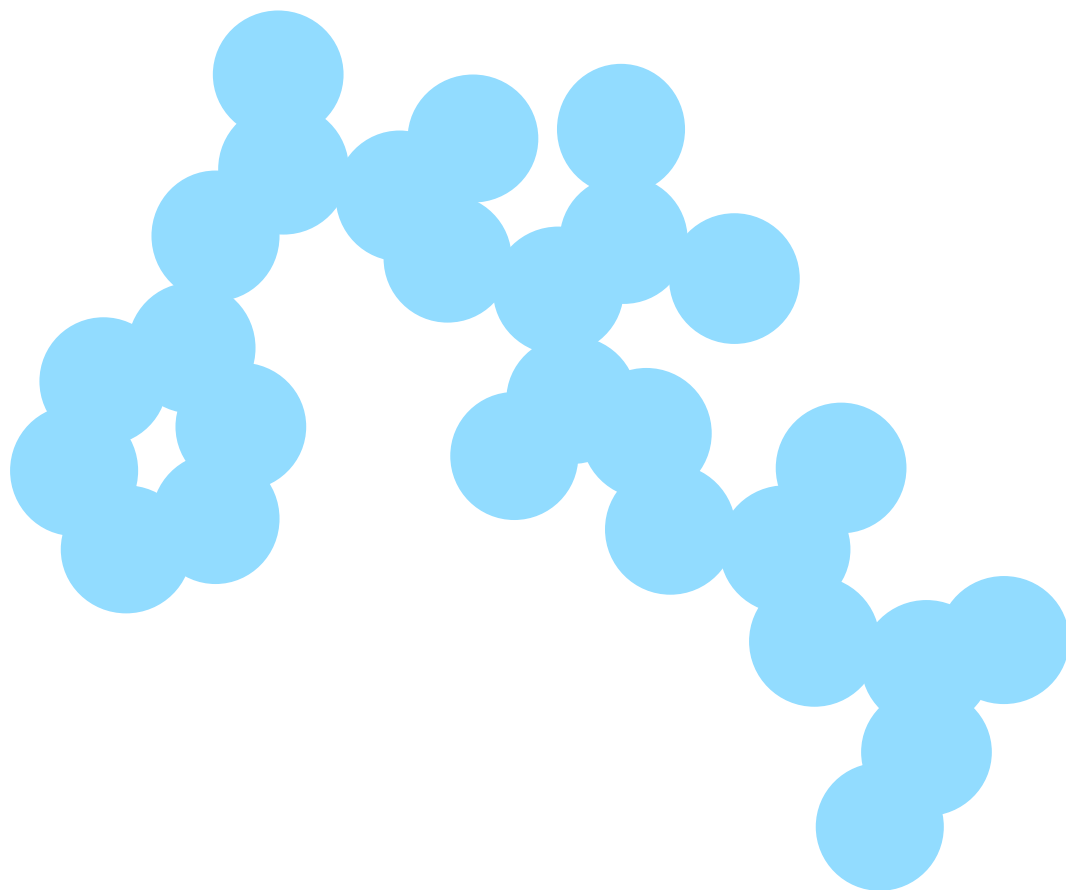
Methods for automated model building

Most current methods search for different patterns that are then used to build and extend a main chain:

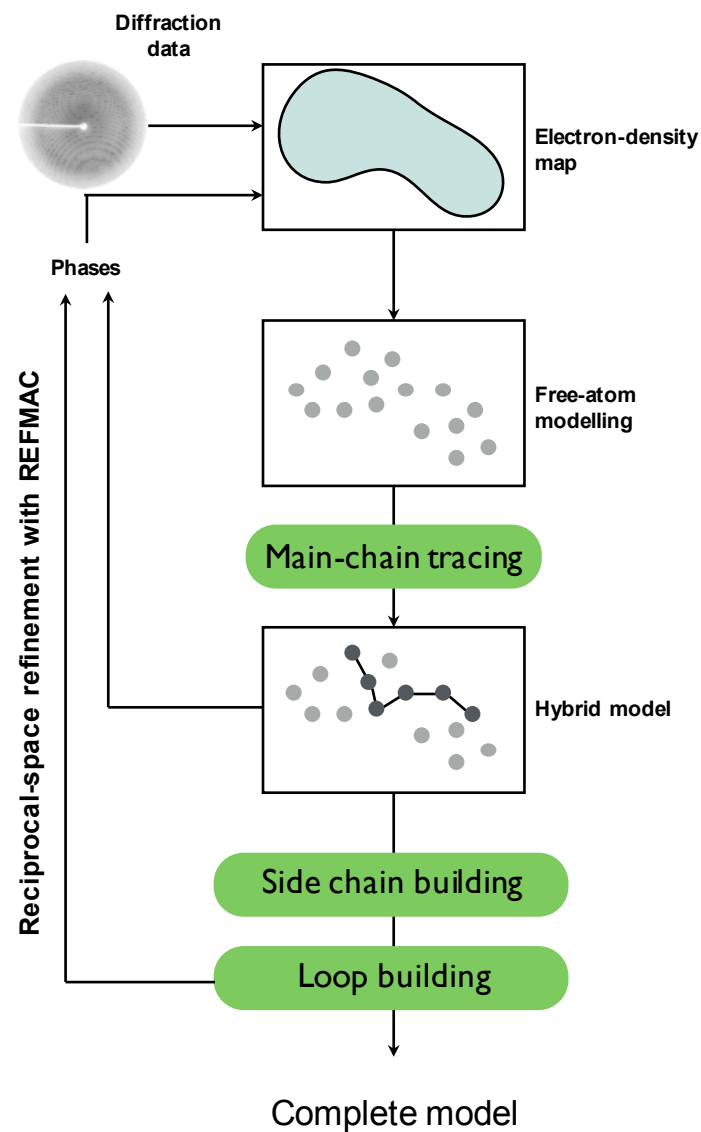
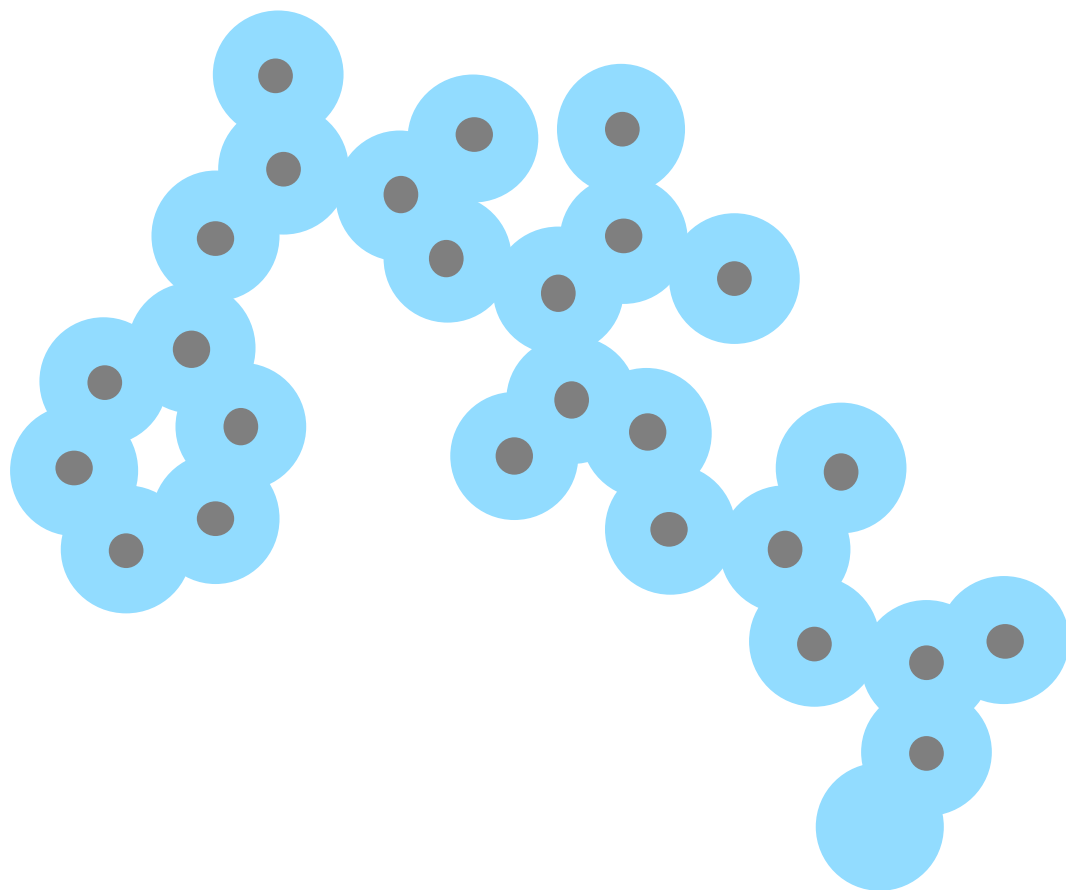


Buccaneer	ARP/wARP 	RESOLVE
$\text{C}\alpha$	(Di-) Peptides	Fragments

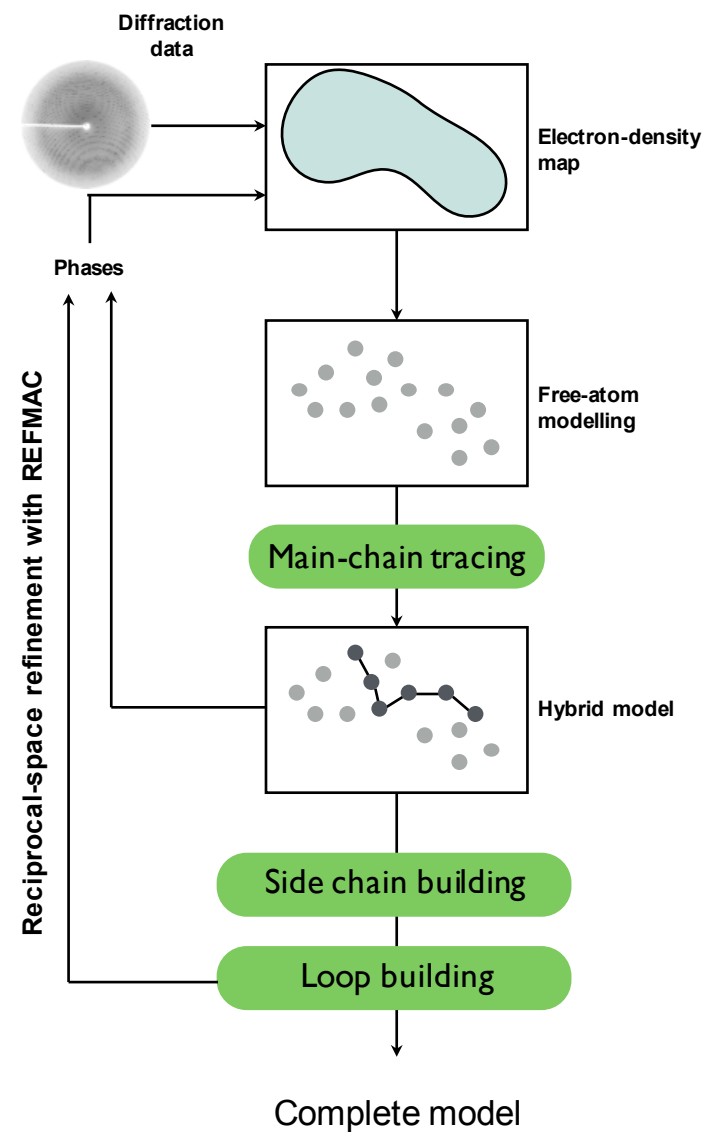
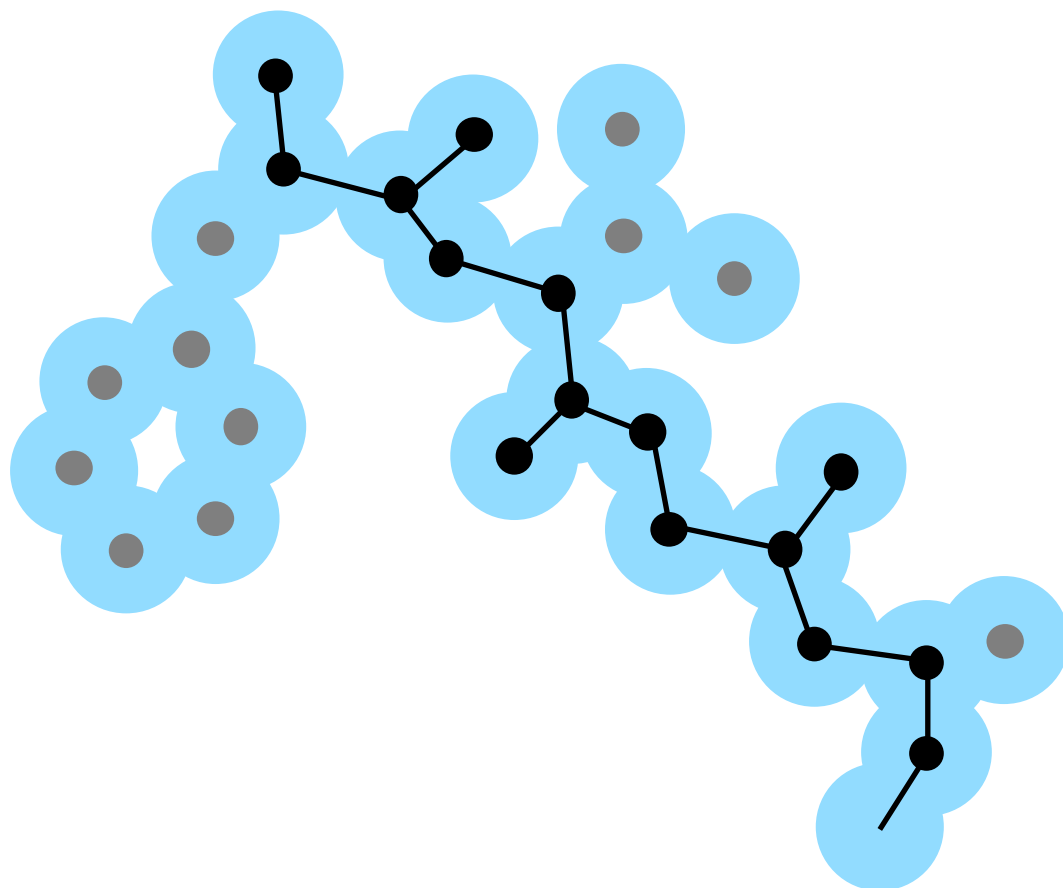
ARP/wARP workflow



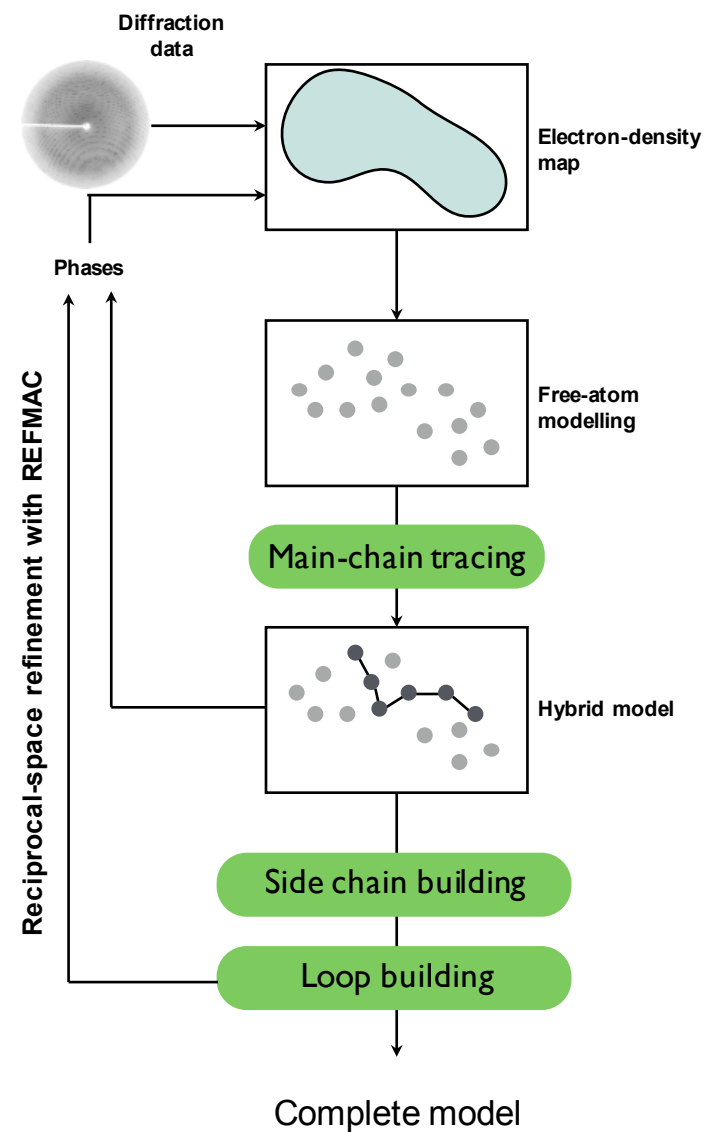
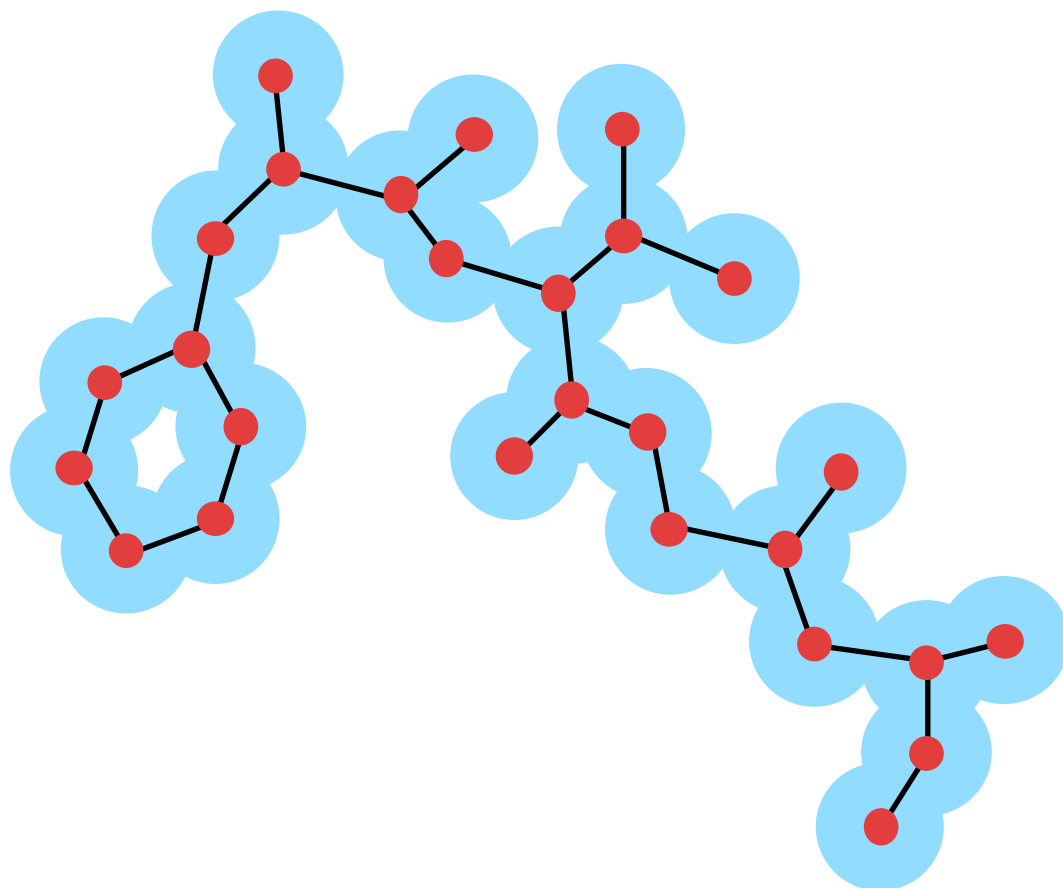
ARP/wARP workflow



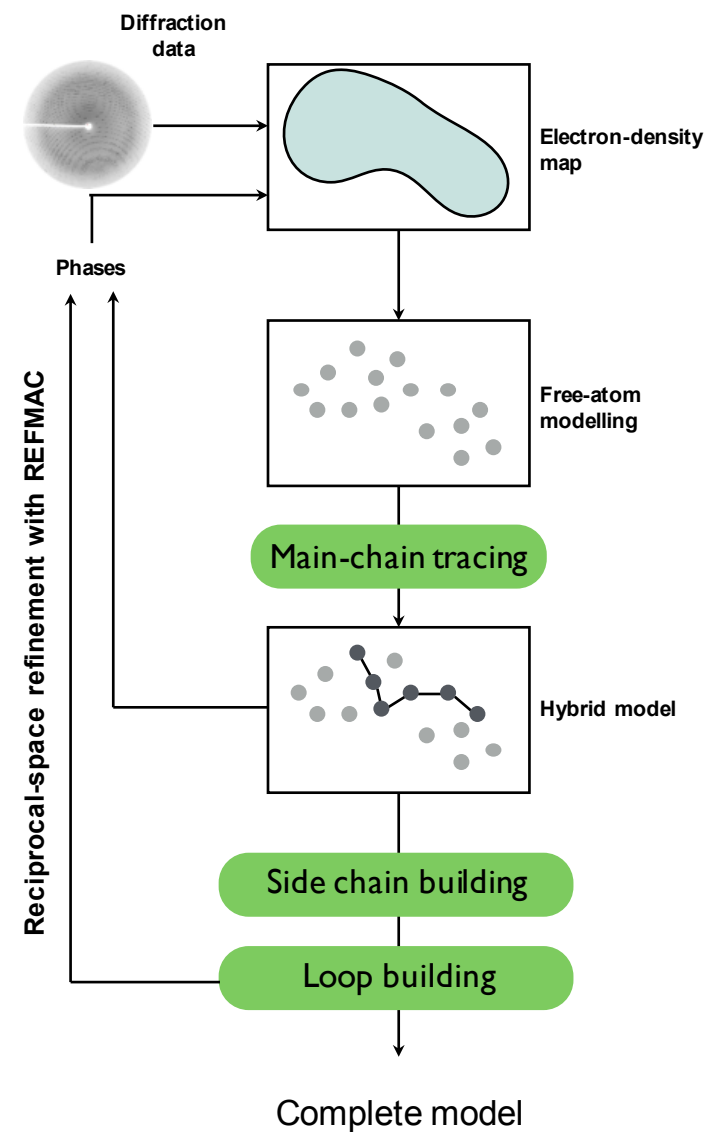
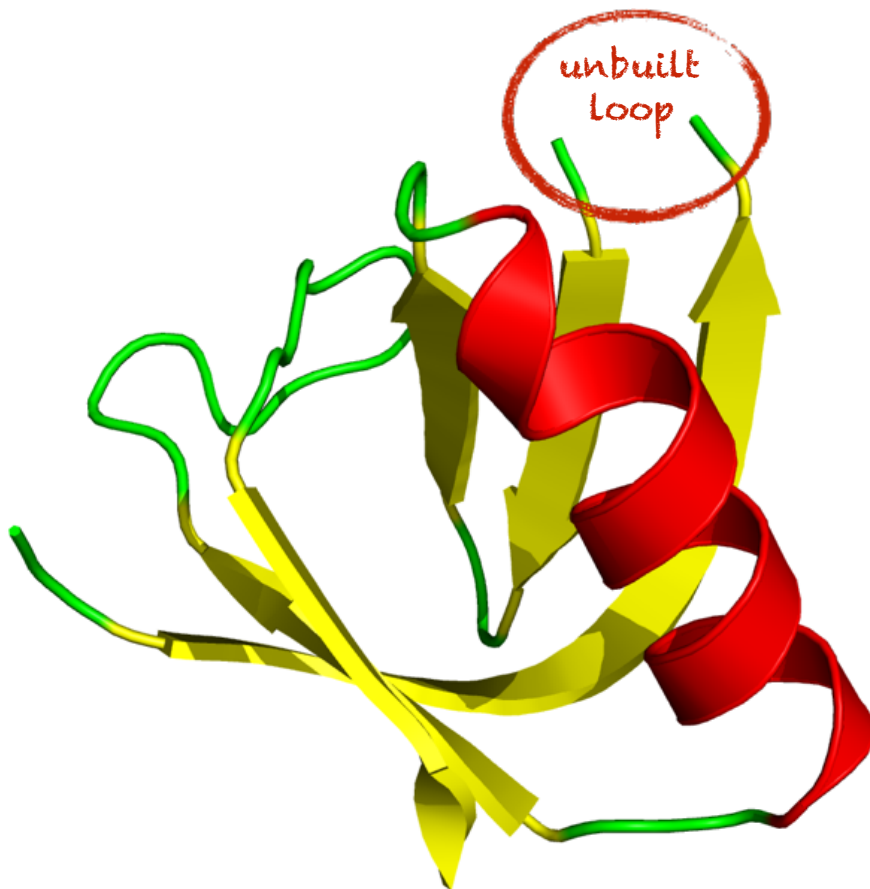
ARP/wARP workflow



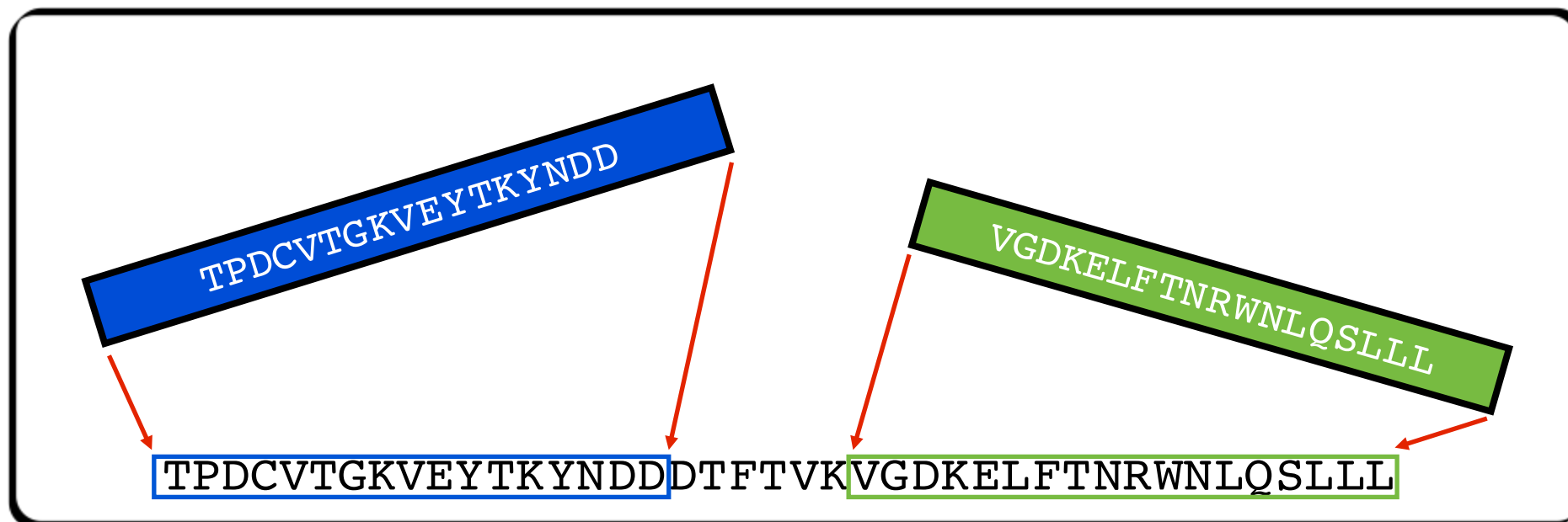
ARP/wARP workflow



ARP/wARP workflow



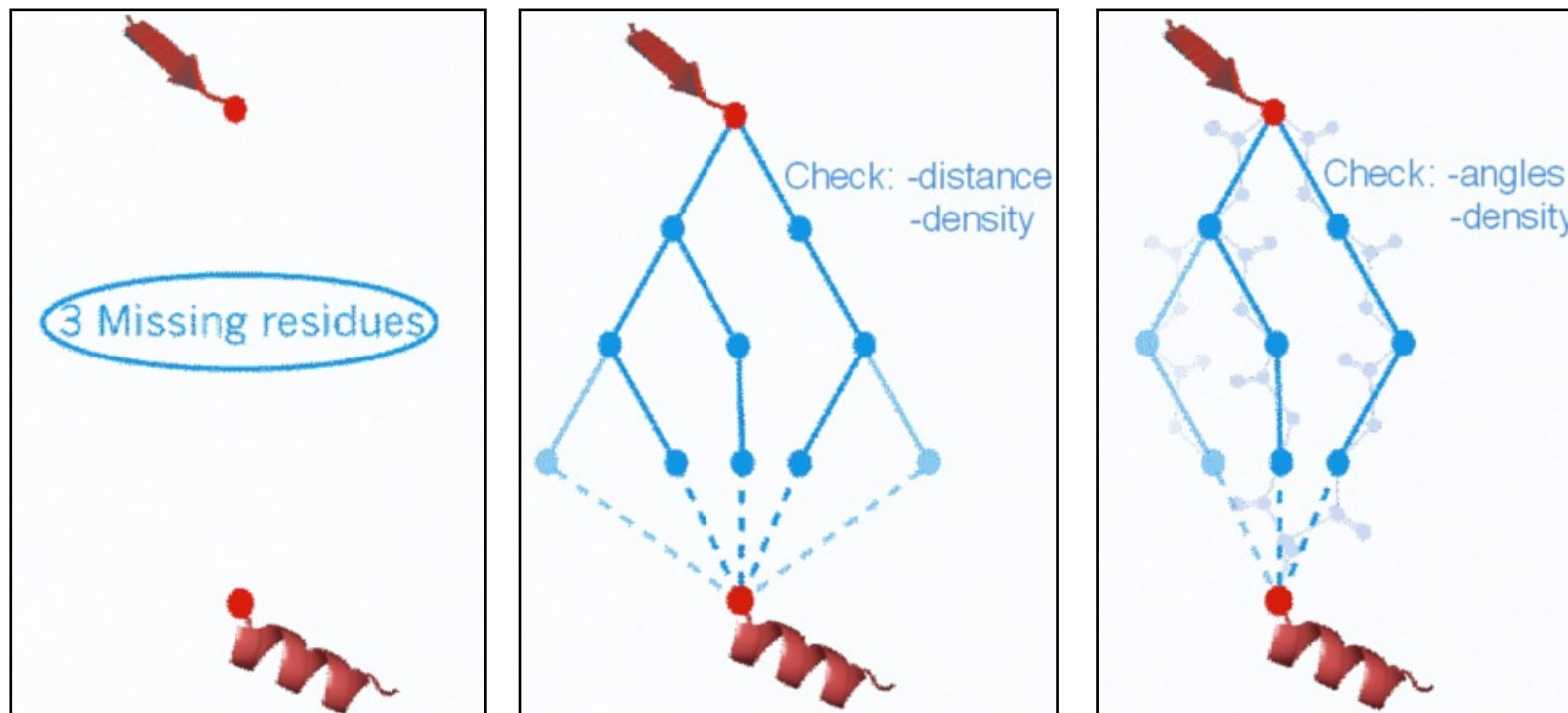
ARP/wARP workflow



ARP/wARP workflow

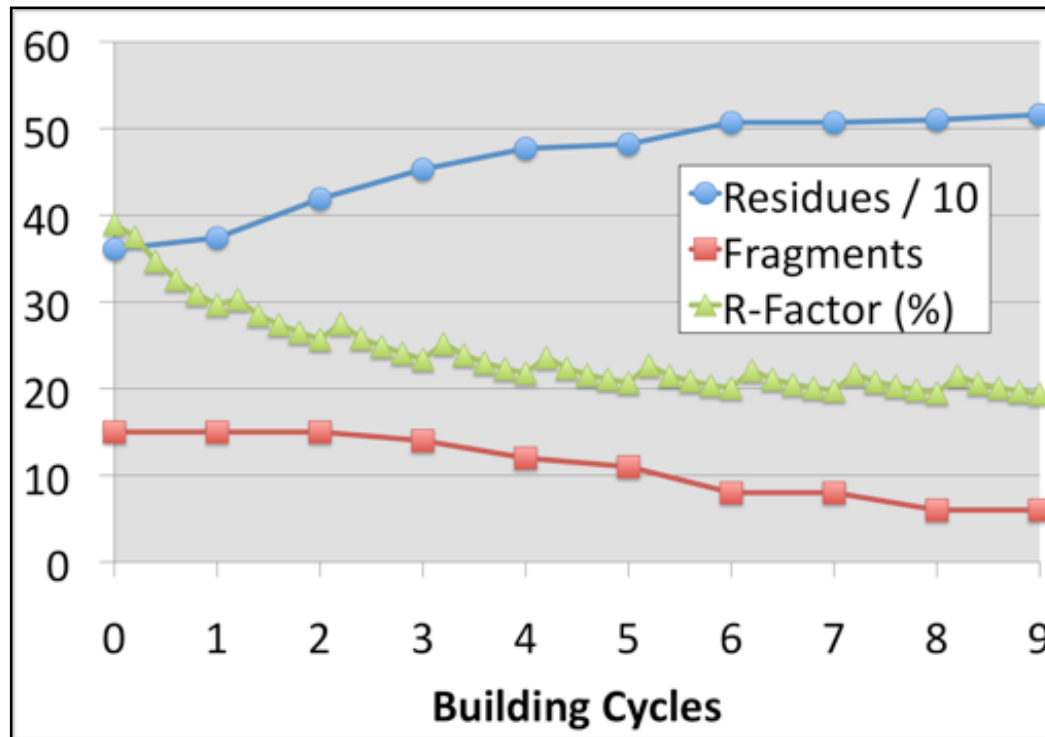


Even after a successful automated model-building run, some regions of known start, end, length, and amino-acid sequence remain to be built, and ARP/wARP builds these **loops** by:

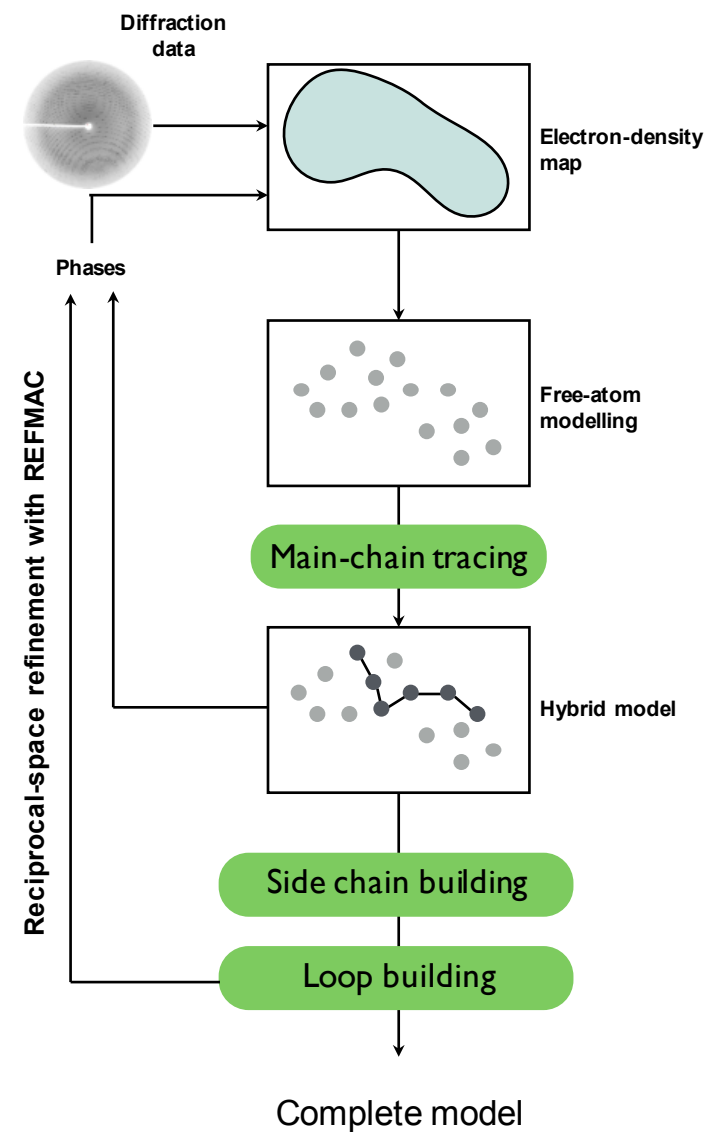
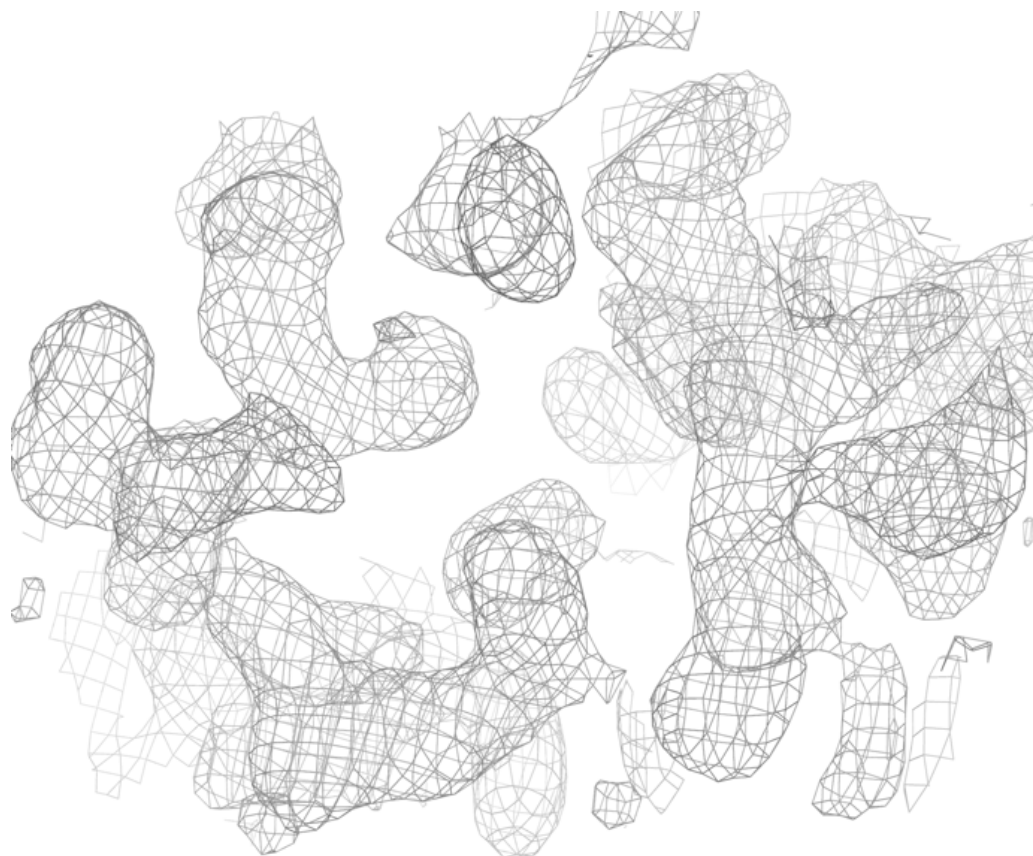


ARP/wARP workflow

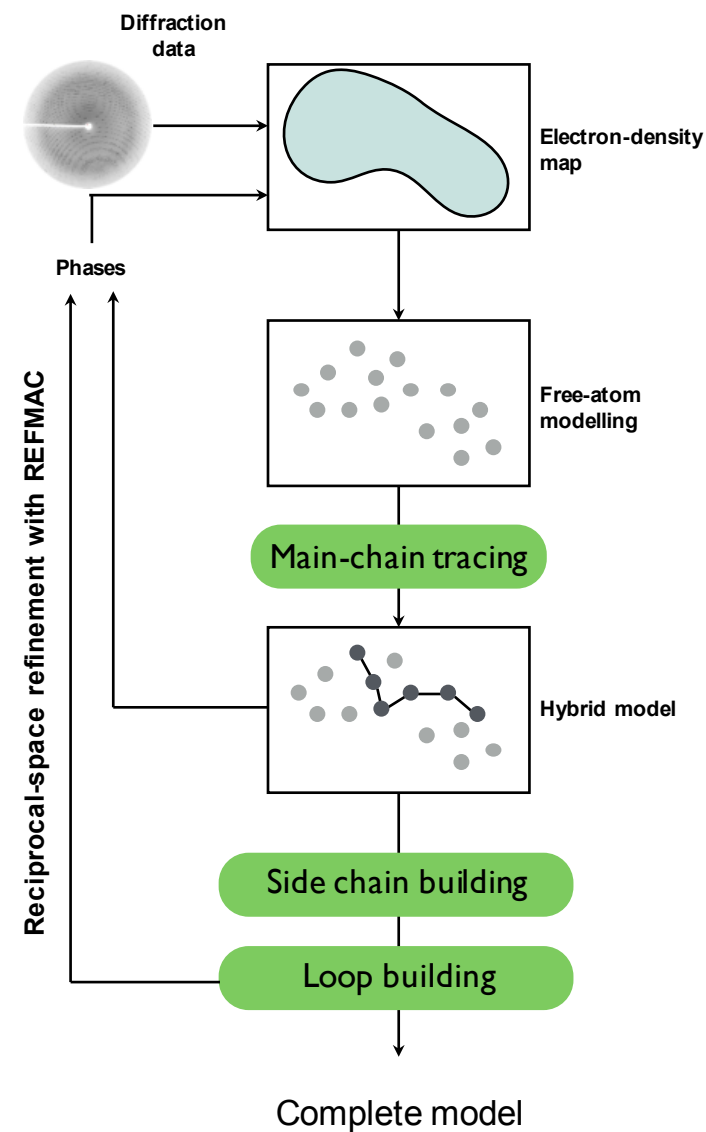
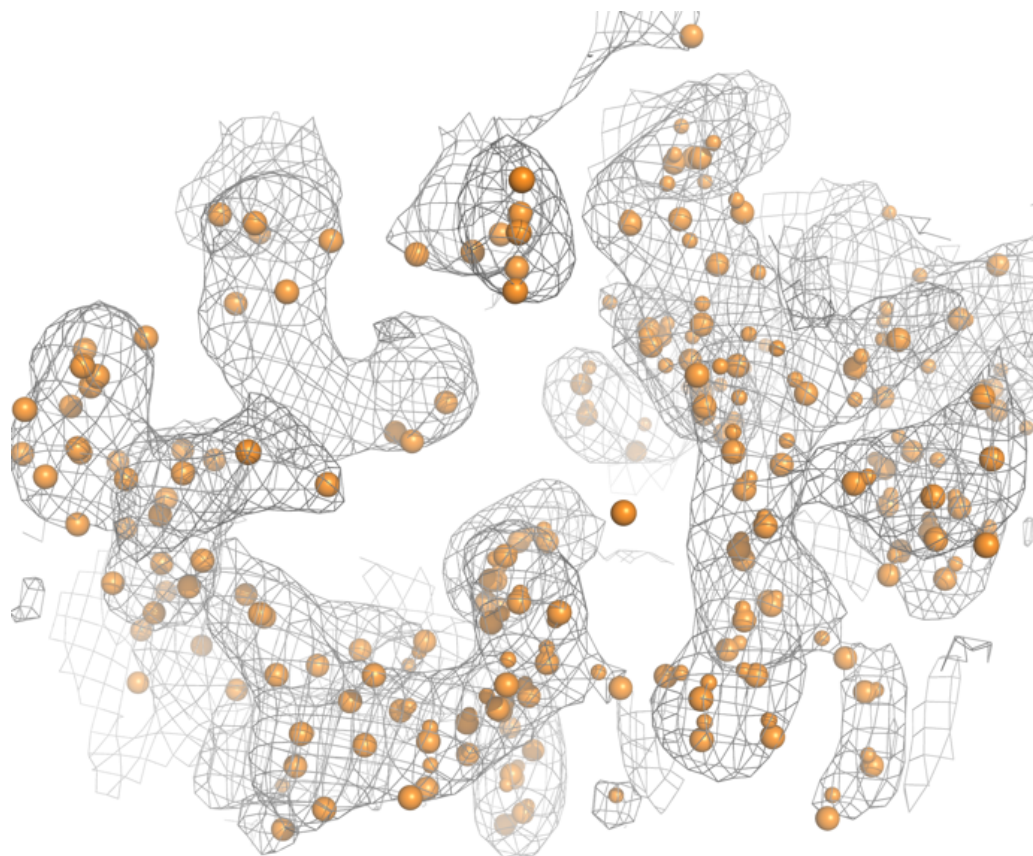
ARP/wARP iterates between main chain tracing, refinement and model update to build models that converge to the final model:



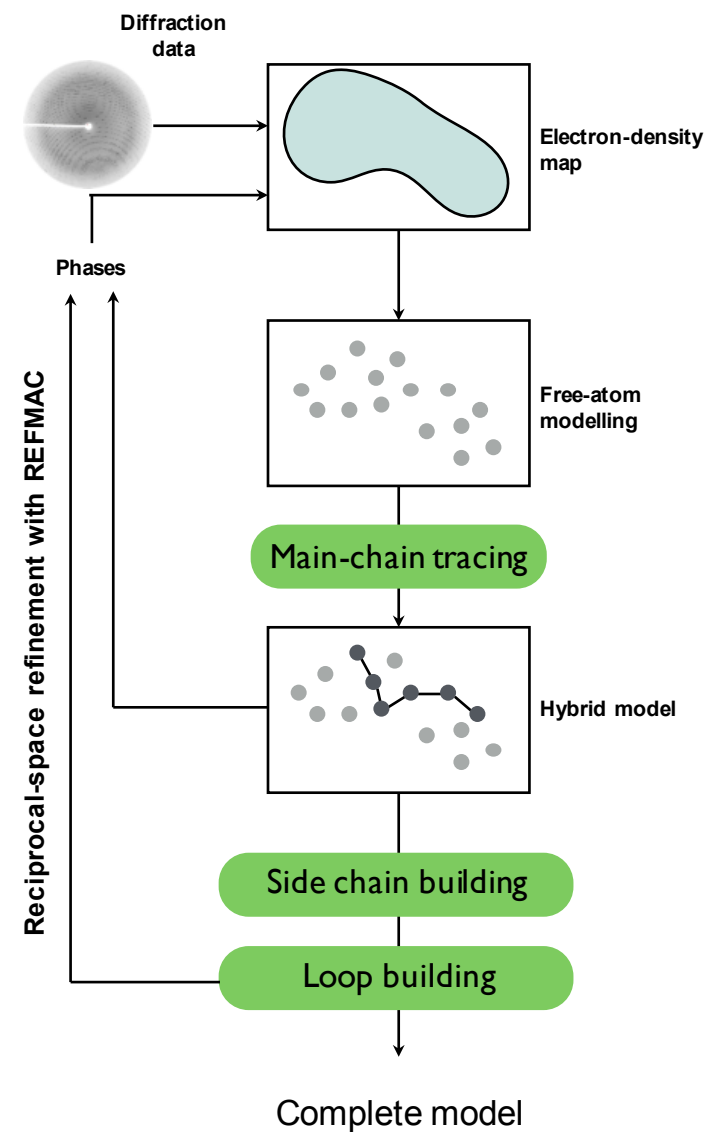
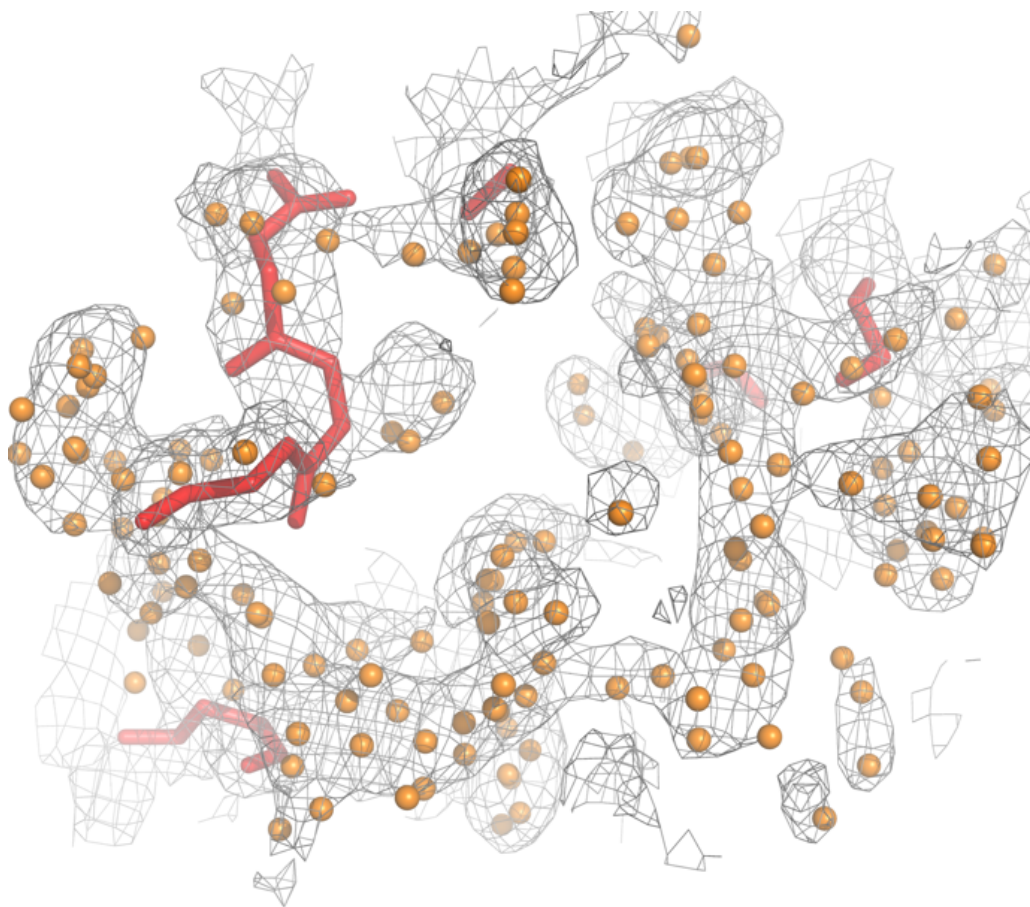
ARP/wARP workflow



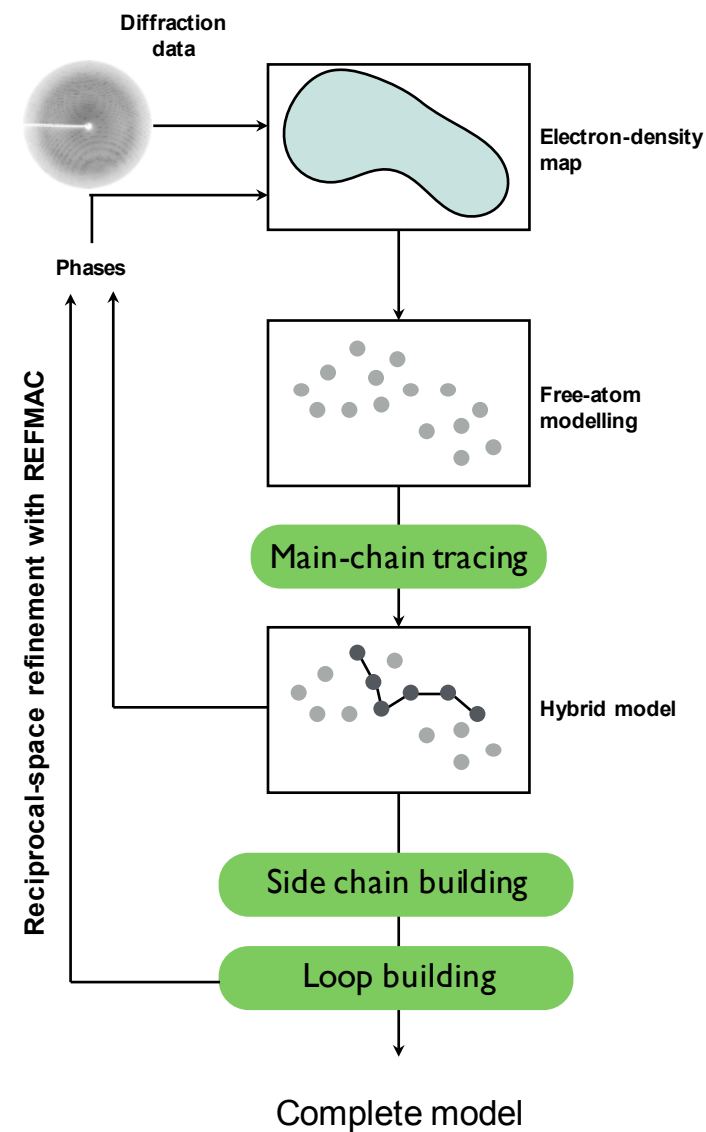
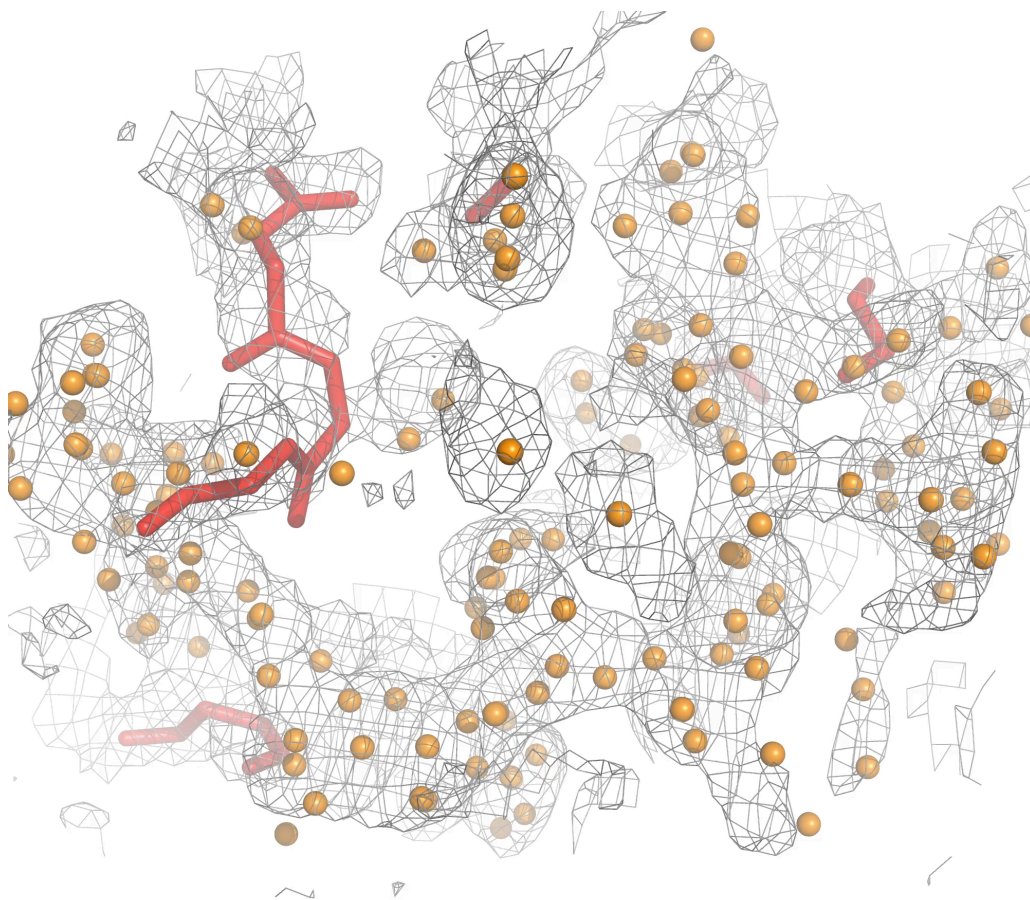
ARP/wARP workflow



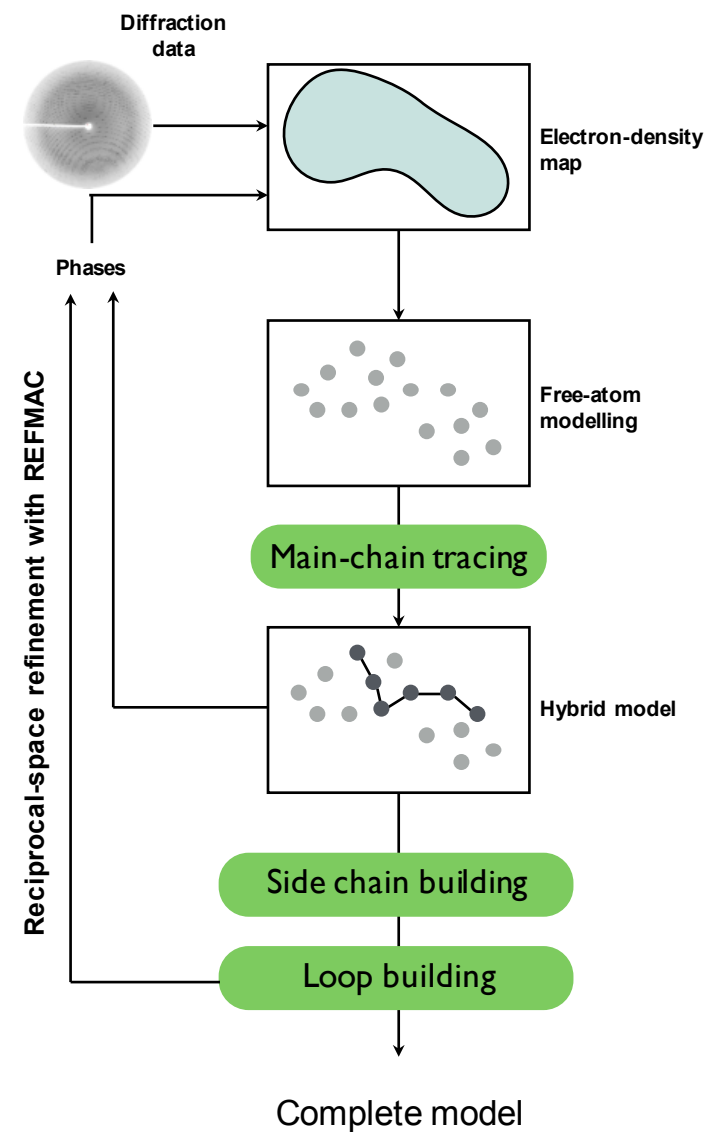
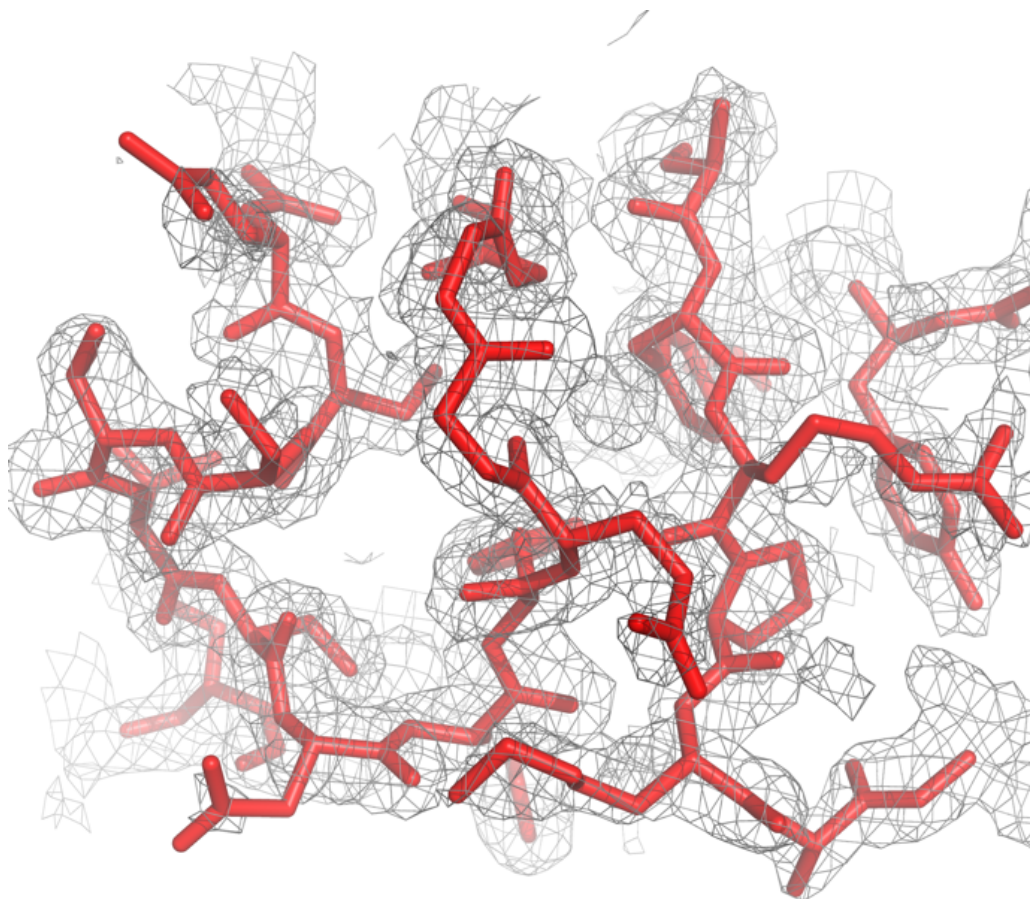
ARP/wARP workflow



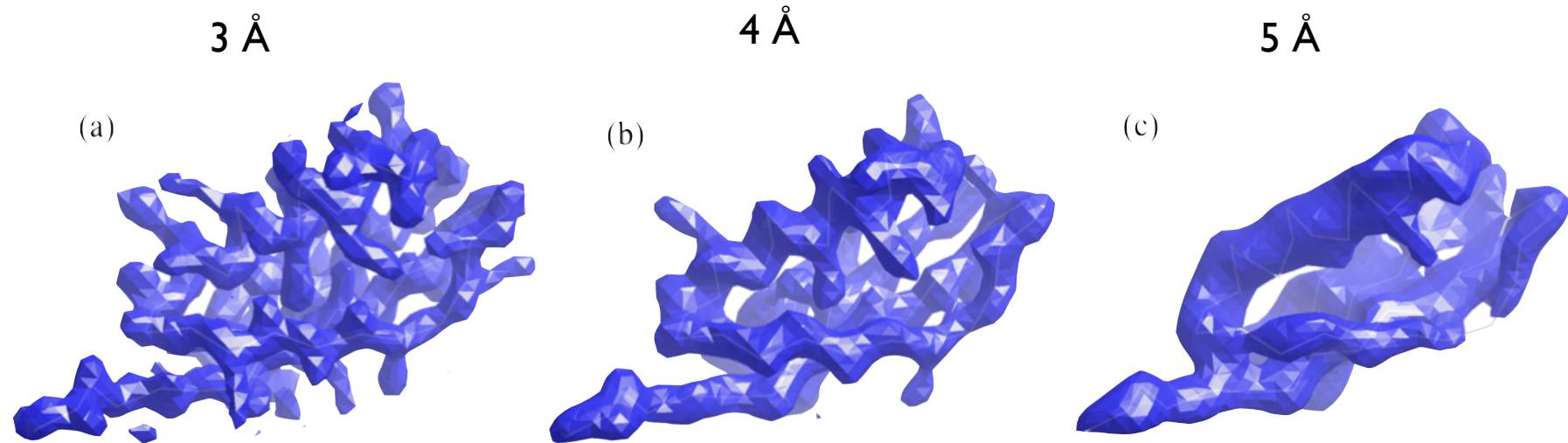
ARP/wARP workflow



ARP/wARP workflow

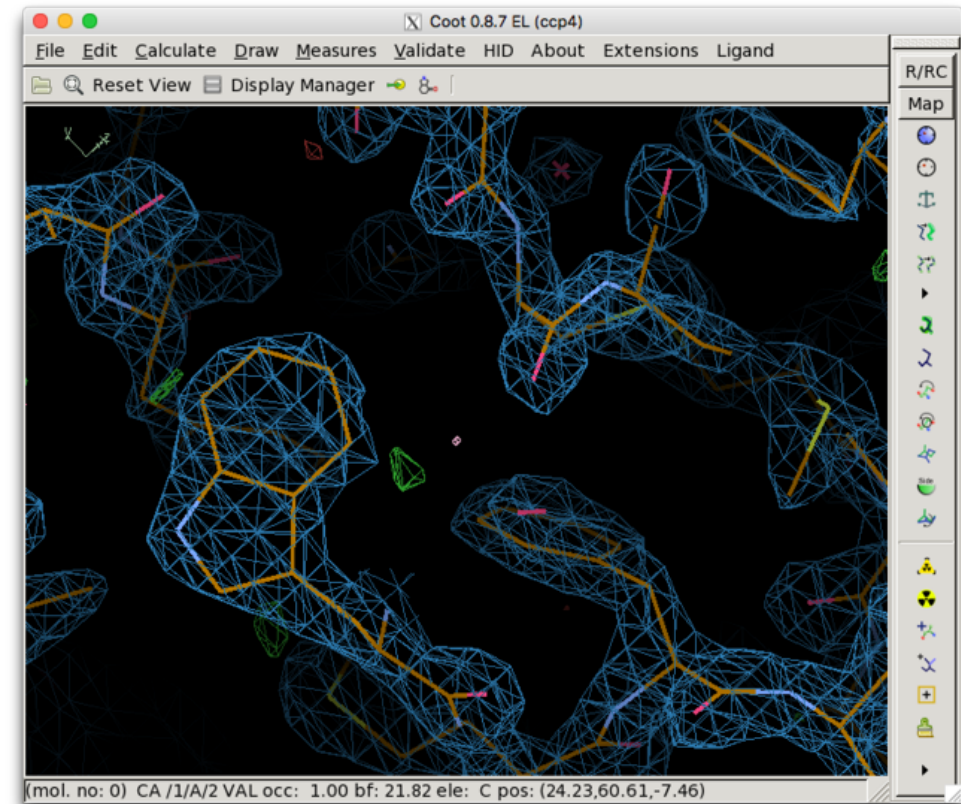
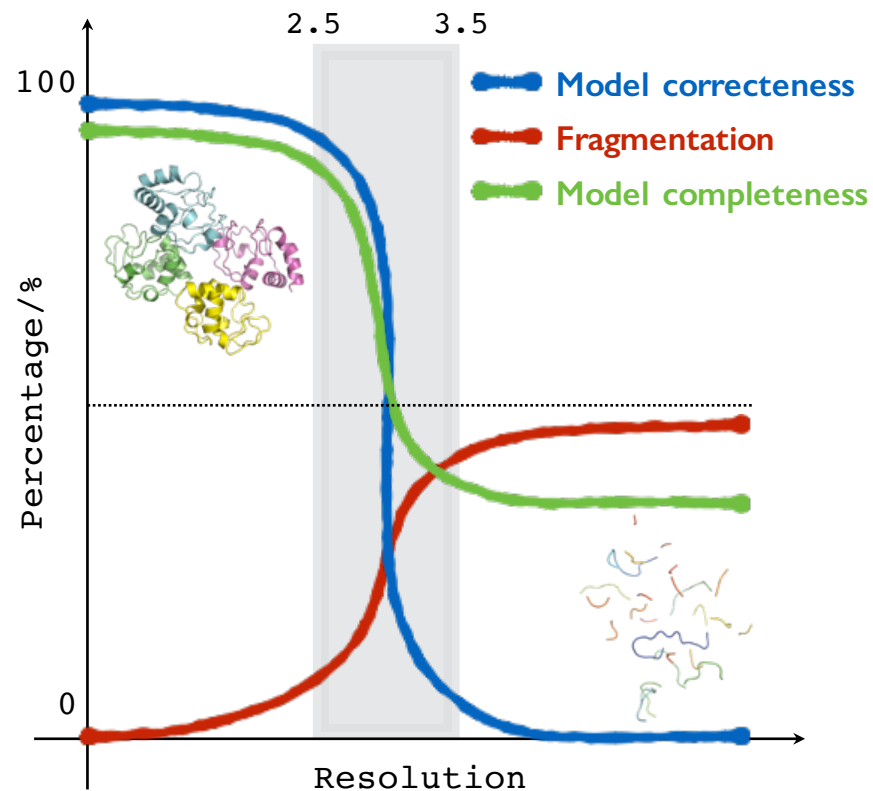


Low resolution limit



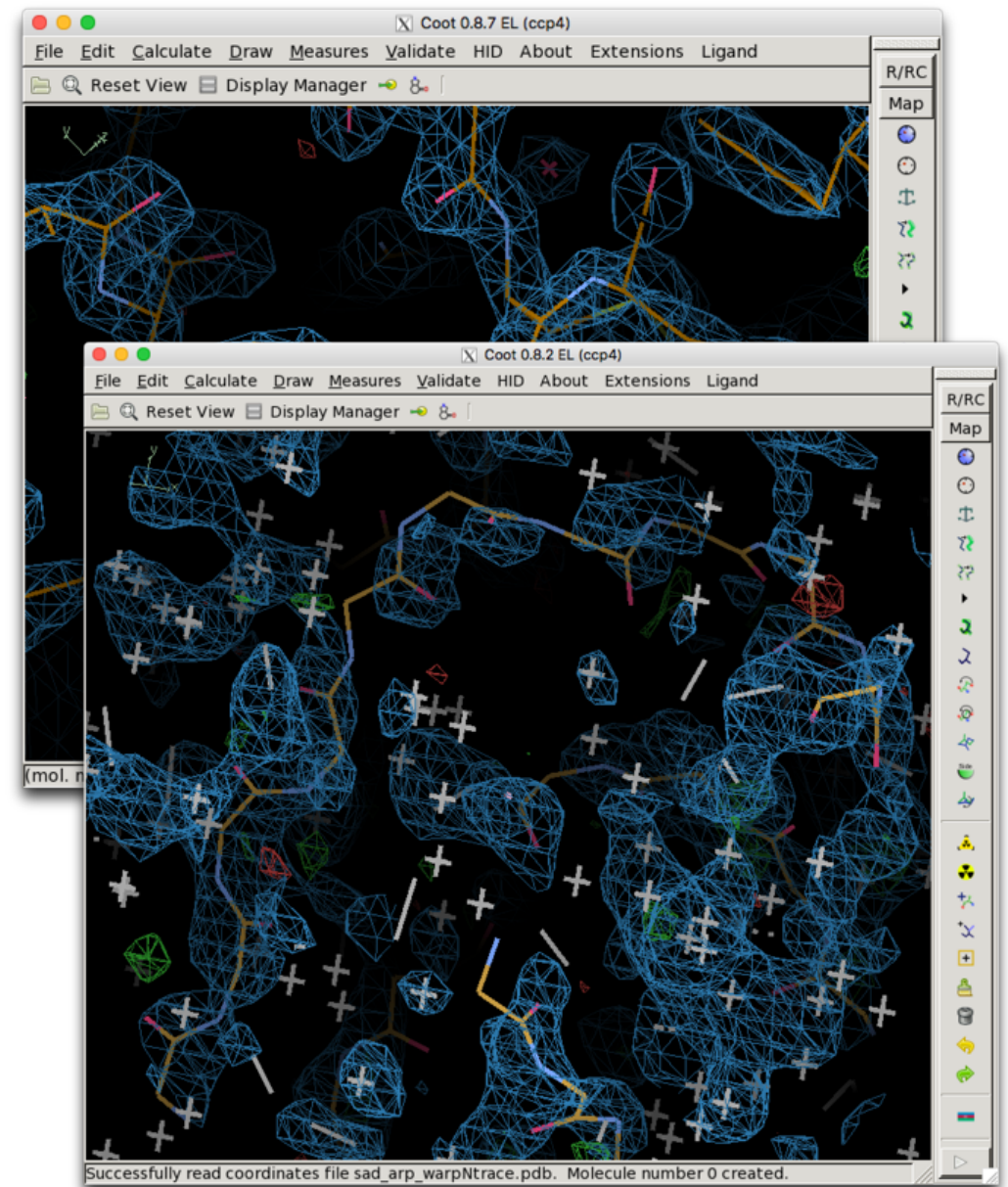
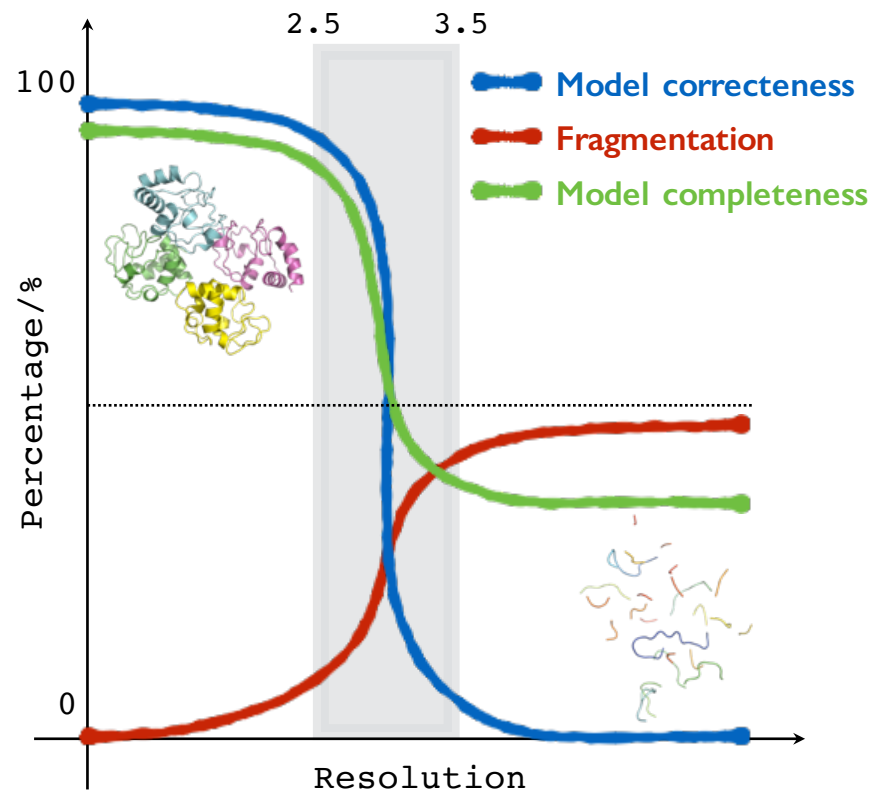
Low resolution limit

ARP/wARP methods work very well
at high-resolution



Low resolution limit

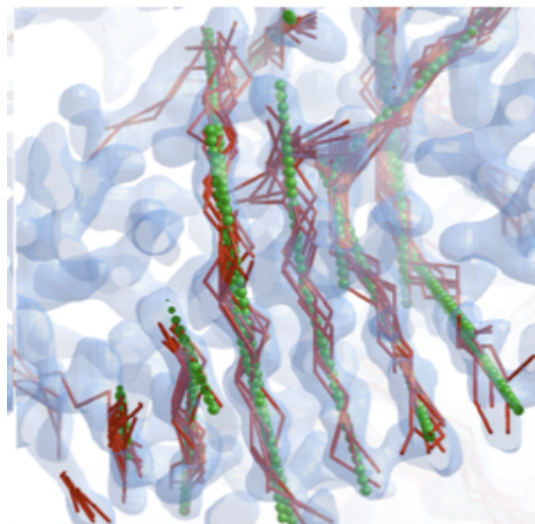
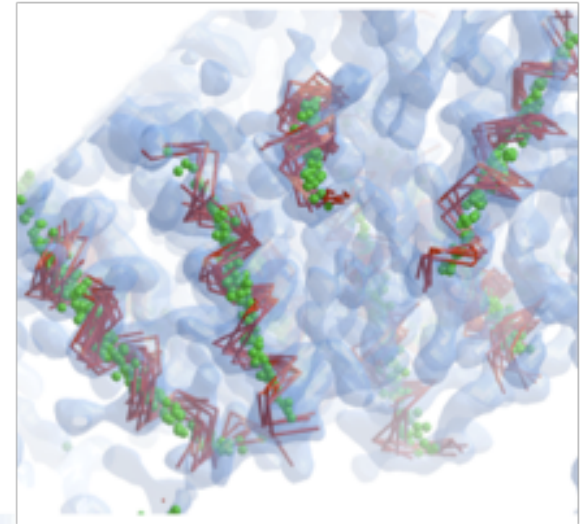
ARP/wARP methods work very well
at high-resolution



Secondary structure modelling

At low resolution, one can try to find **secondary structure elements** on the map:

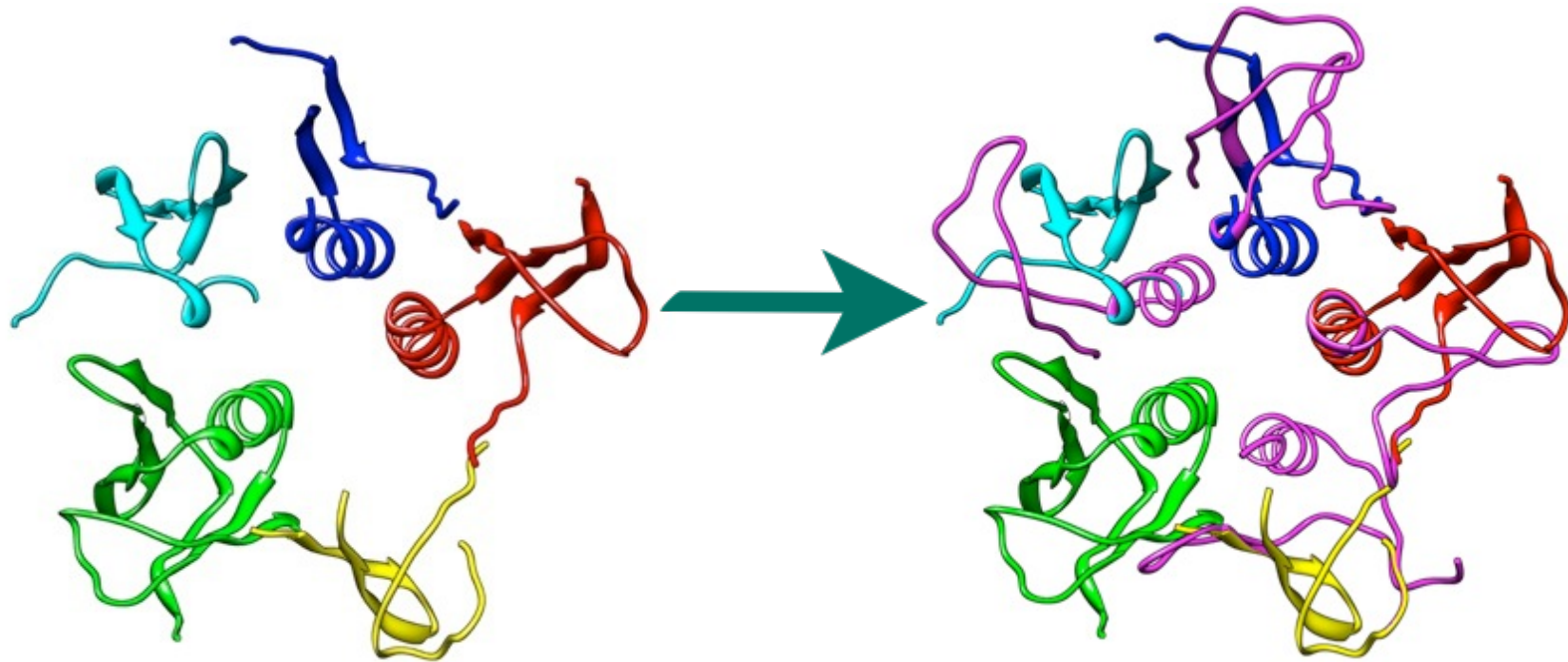
- Short helix/strand fragments (3 to 5 C α candidates) are built
- Traces are clustered, averaged, and extended
- Default below 2.7Å resolution



NCS Detection and Extension

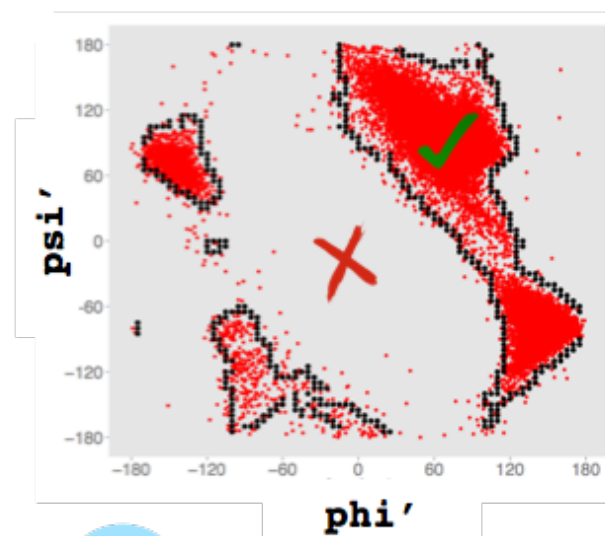
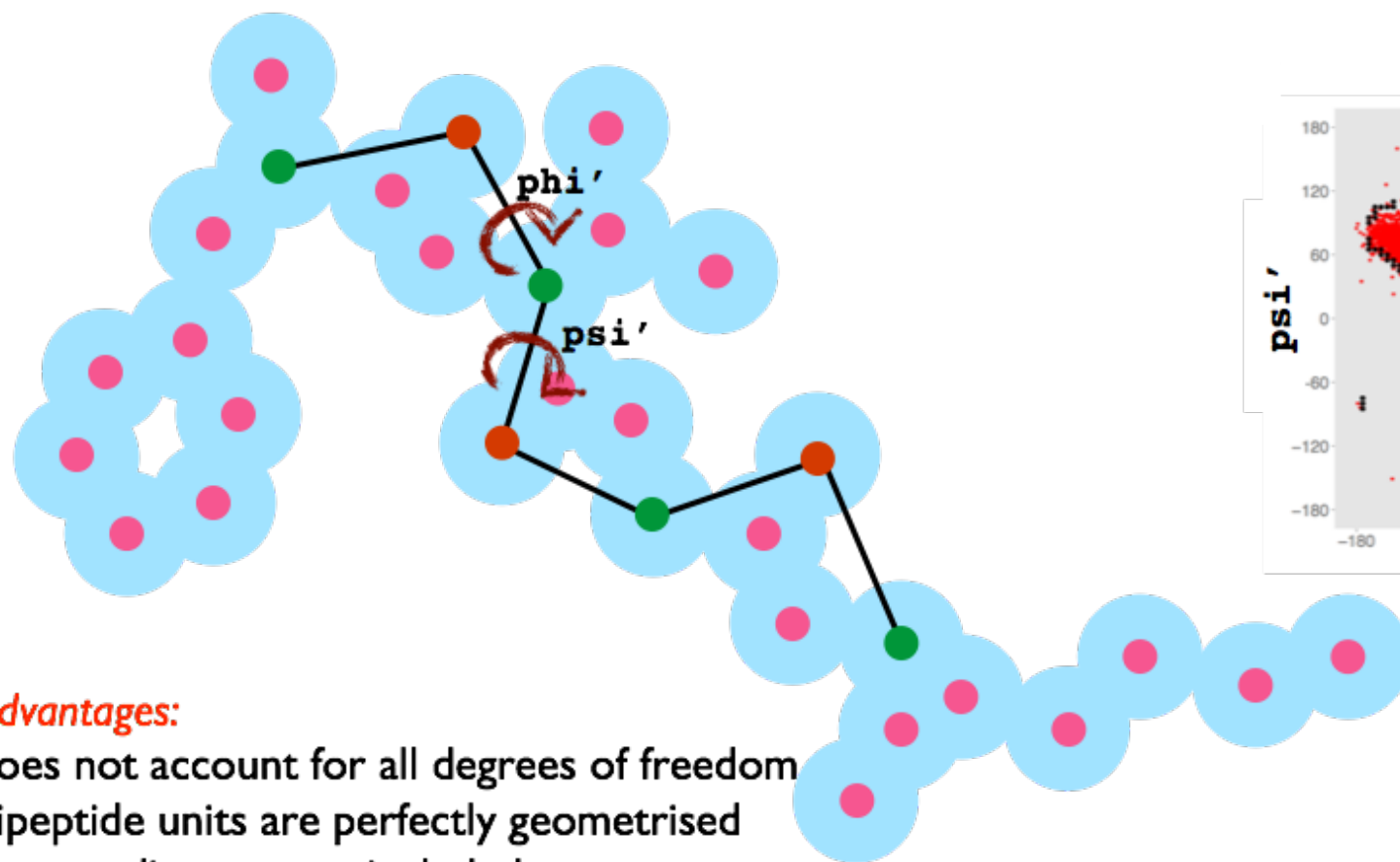
Non-Crystallographic Symmetry occurs if there are multiple subunits in the asymmetric unit of a crystal and some of them adopt almost the same tertiary structure:

We can use this information to extend fragmented models!



DipCheck: main-chain geometry validation

Pereira, J., & Lamzin, V.
under review



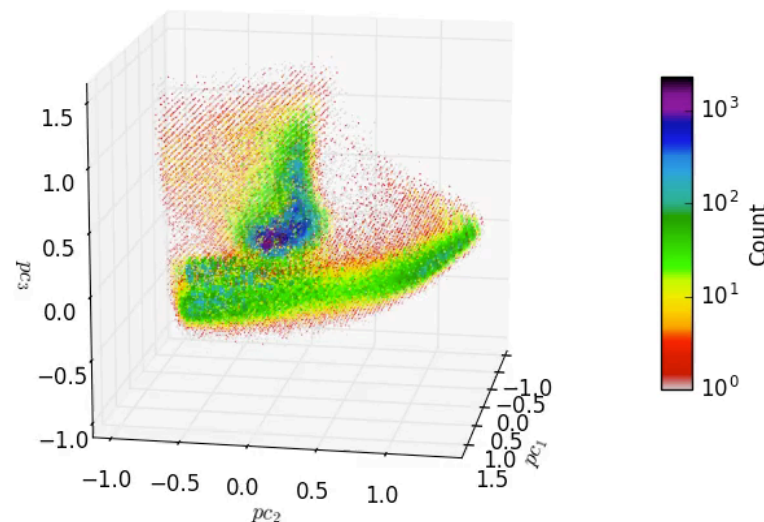
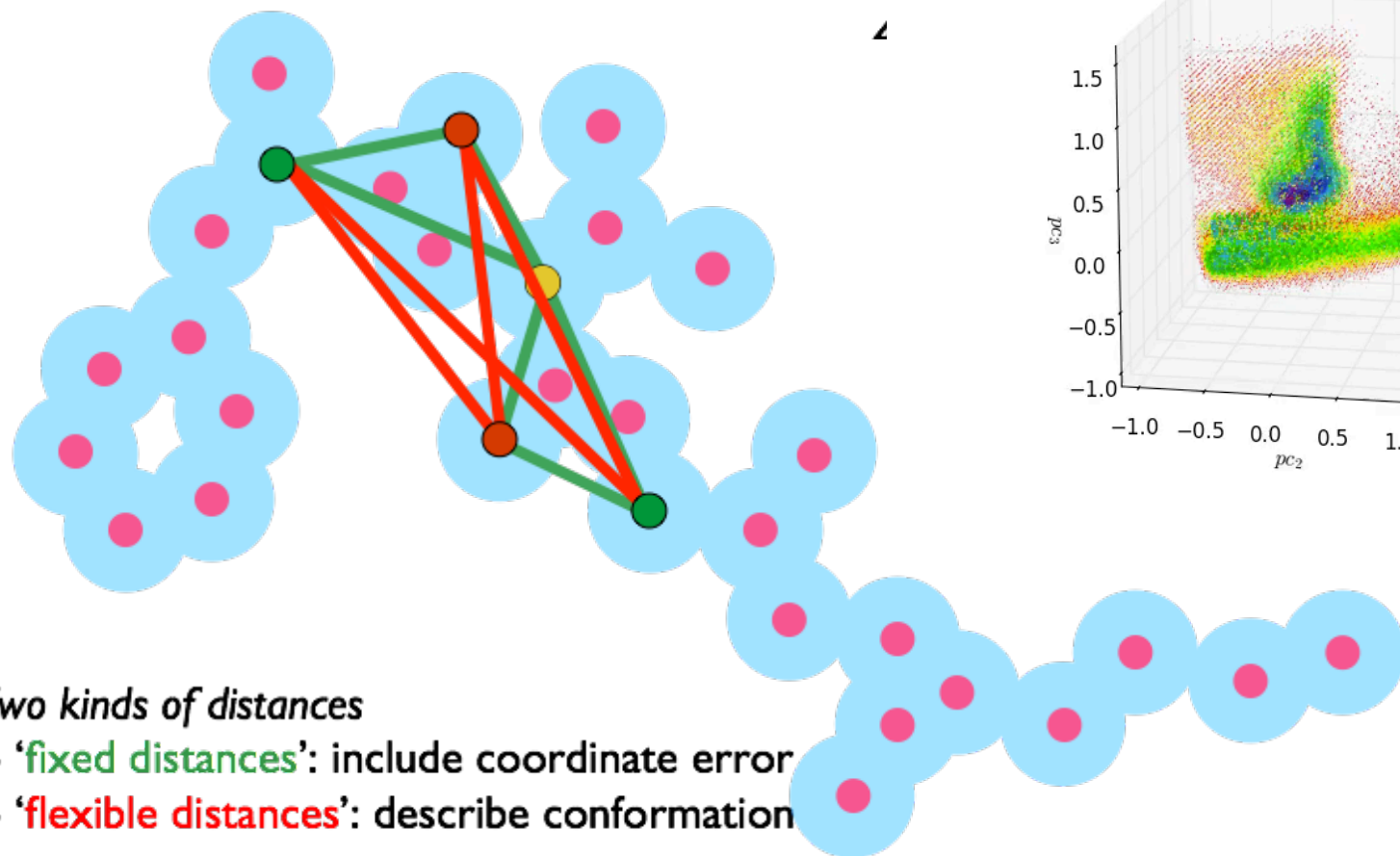
Disadvantages:

- Does not account for all degrees of freedom
- Dipeptide units are perfectly geometrised
- No coordinate error included

... as it is not suitable for lower resolution.

DipCheck: main-chain geometry validation

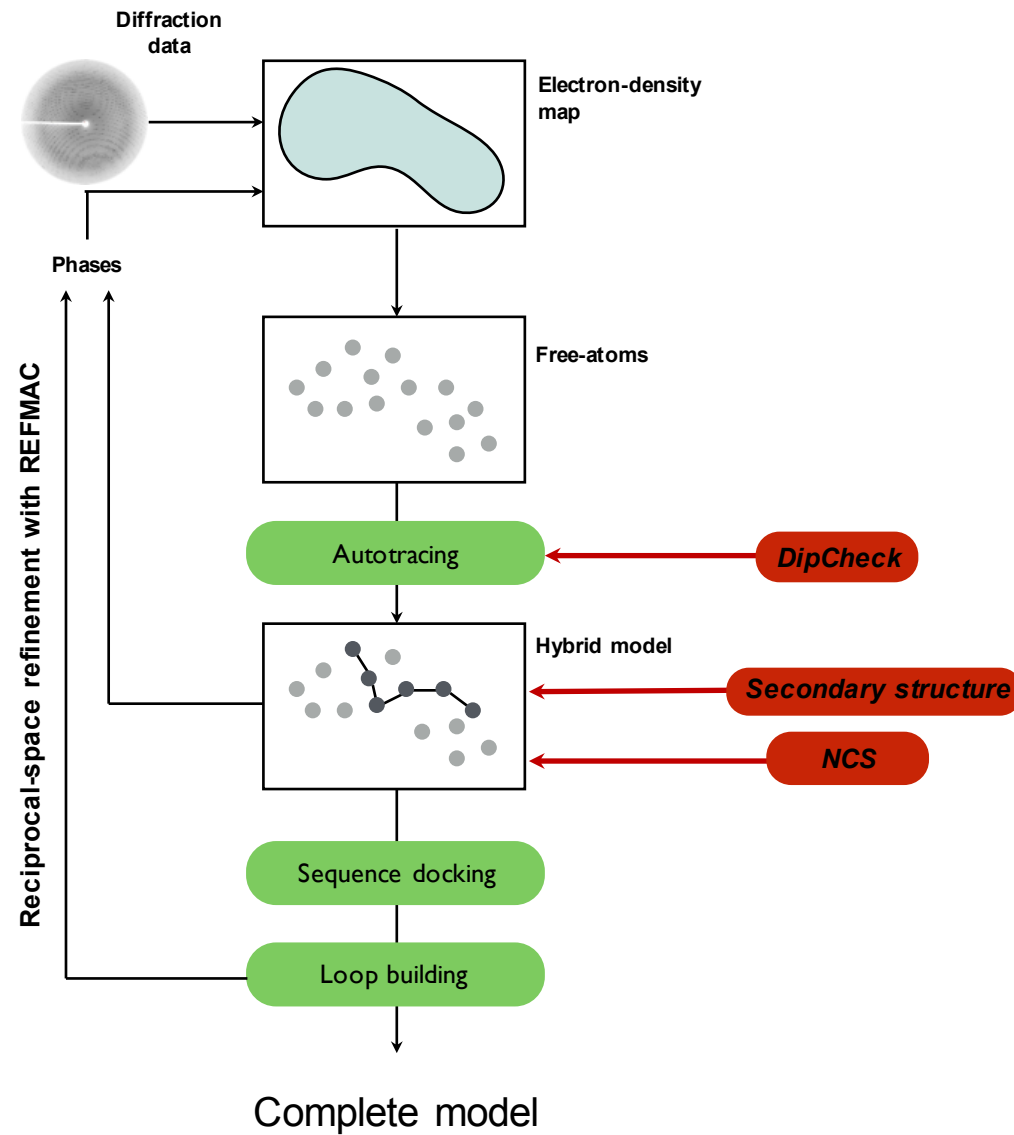
Pereira, J., & Lamzin, V.
under review



Two kinds of distances

- 'fixed distances': include coordinate error
- 'flexible distances': describe conformation

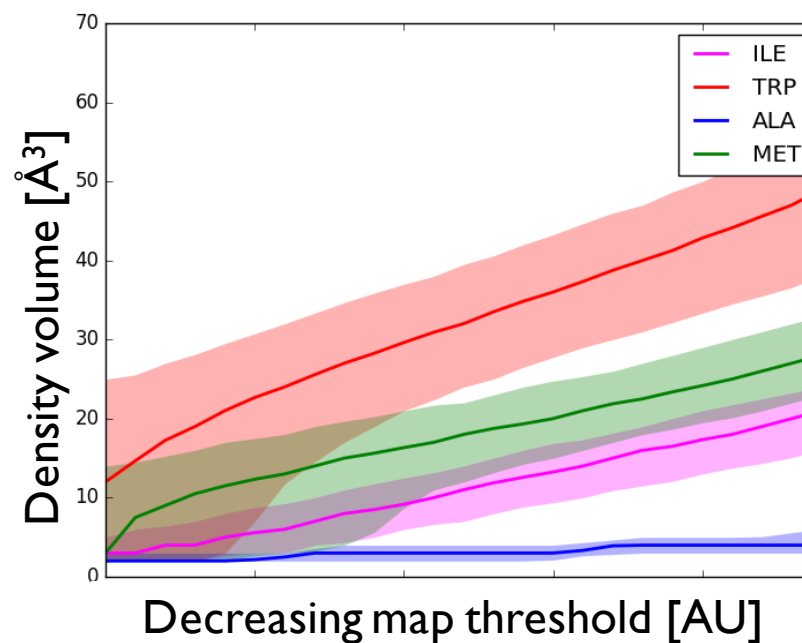
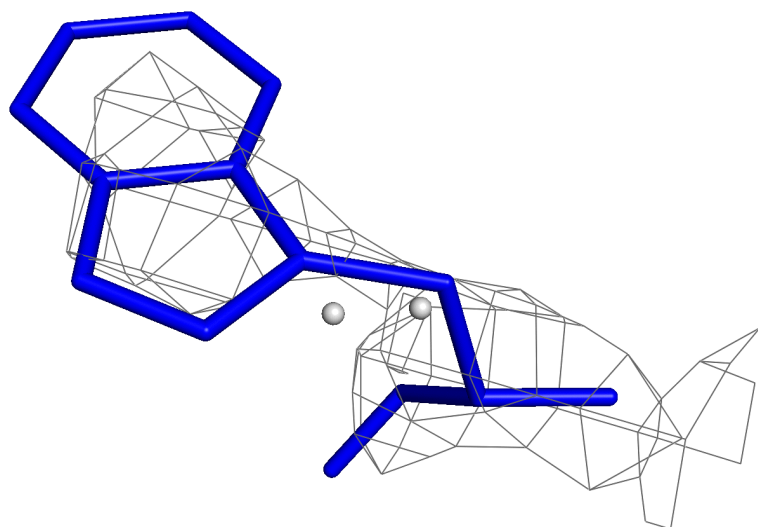
How to get better models at lower resolution?



New side-chain building module

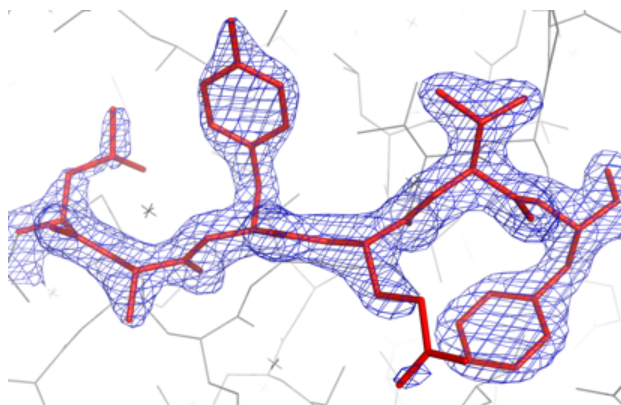
coming
soon!

Side-chain electron-density volume changes correlate with residue type.



New side-chain building module

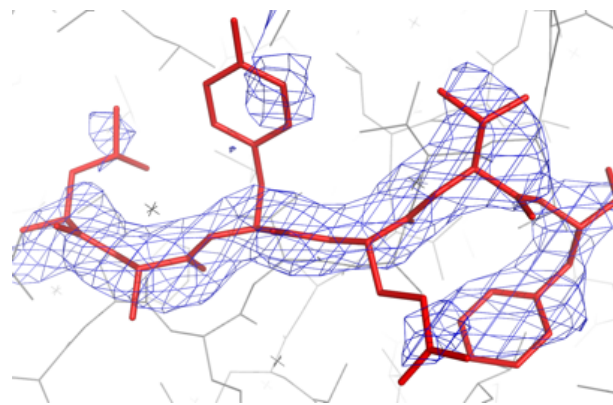
coming
soon!



Resolution: 2.5Å
Mean phase error: 20deg

Docked residues: 99%

old algorithm: 70%

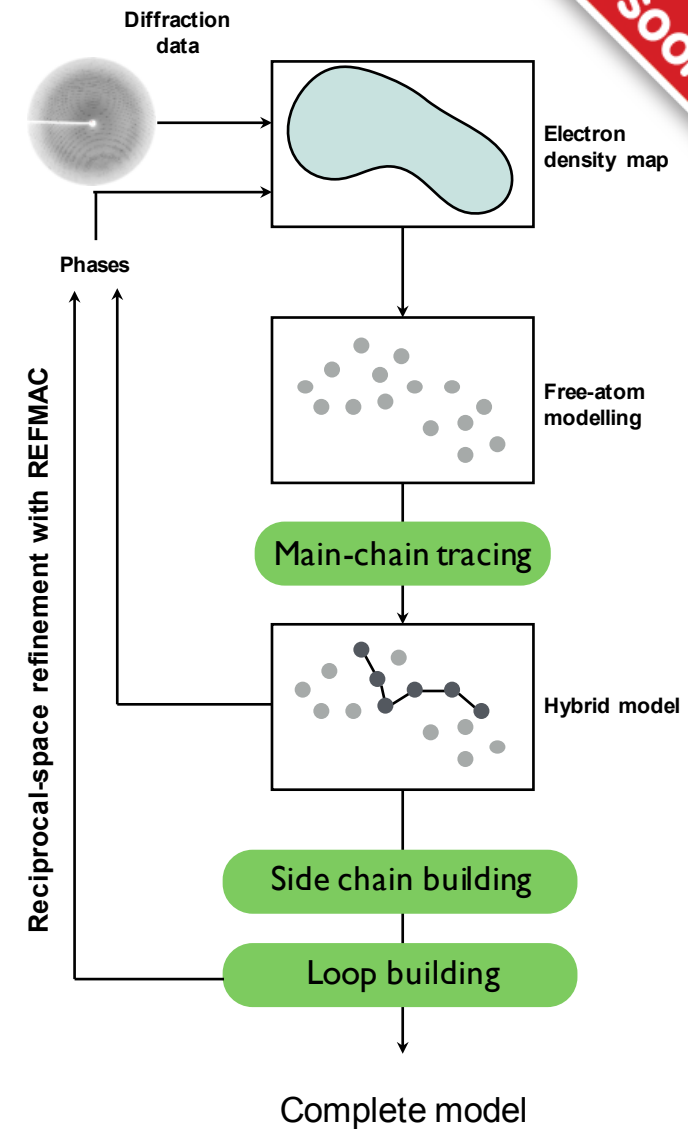
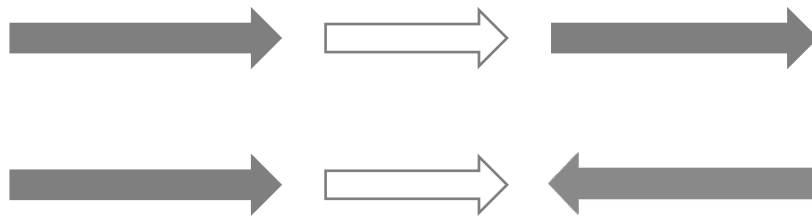
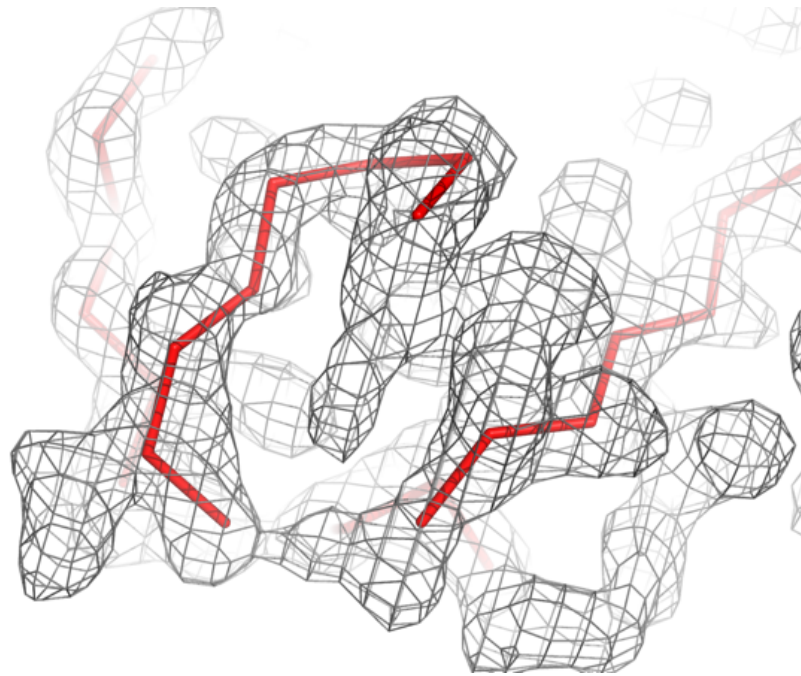


Resolution: 3.2Å
Mean phase error: 40 deg

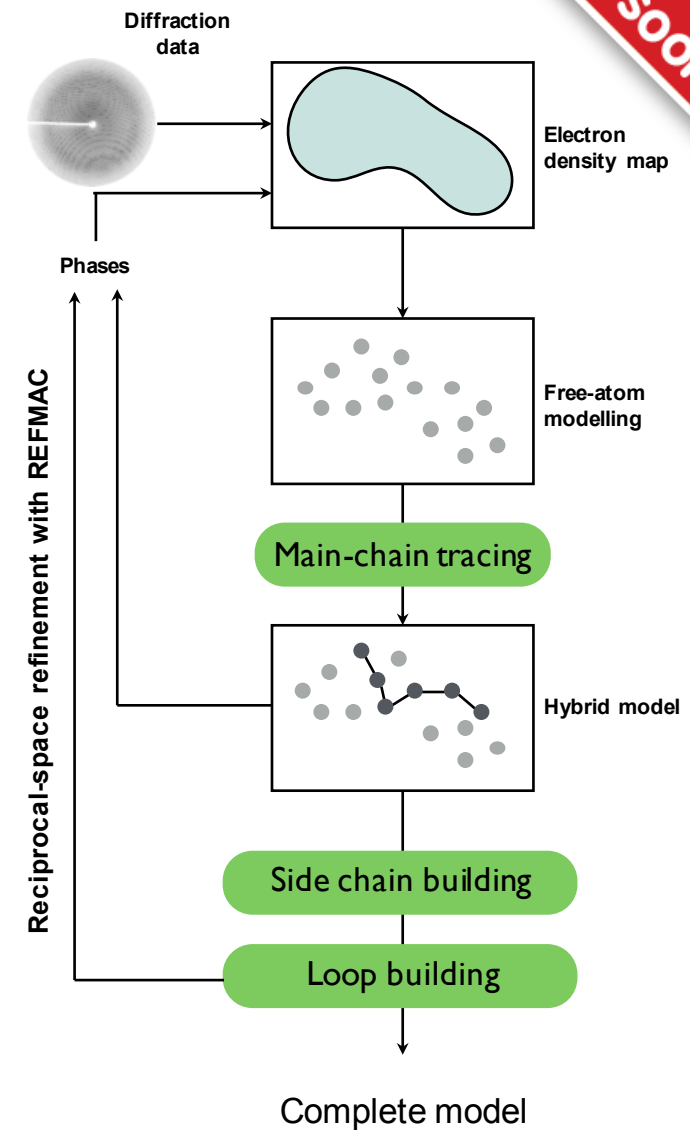
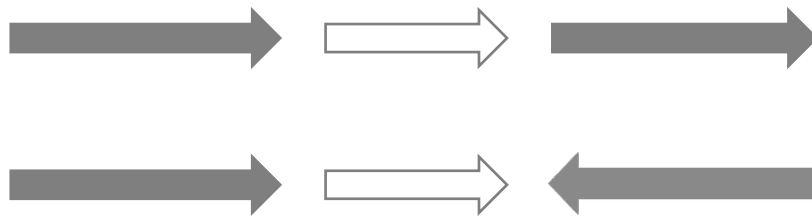
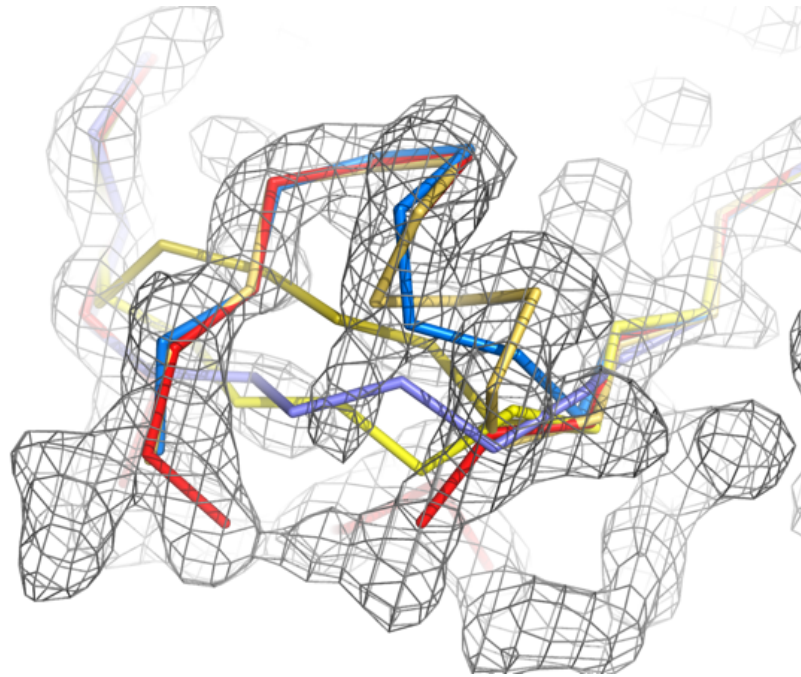
Docked residues: 69%

old algorithm: 40%

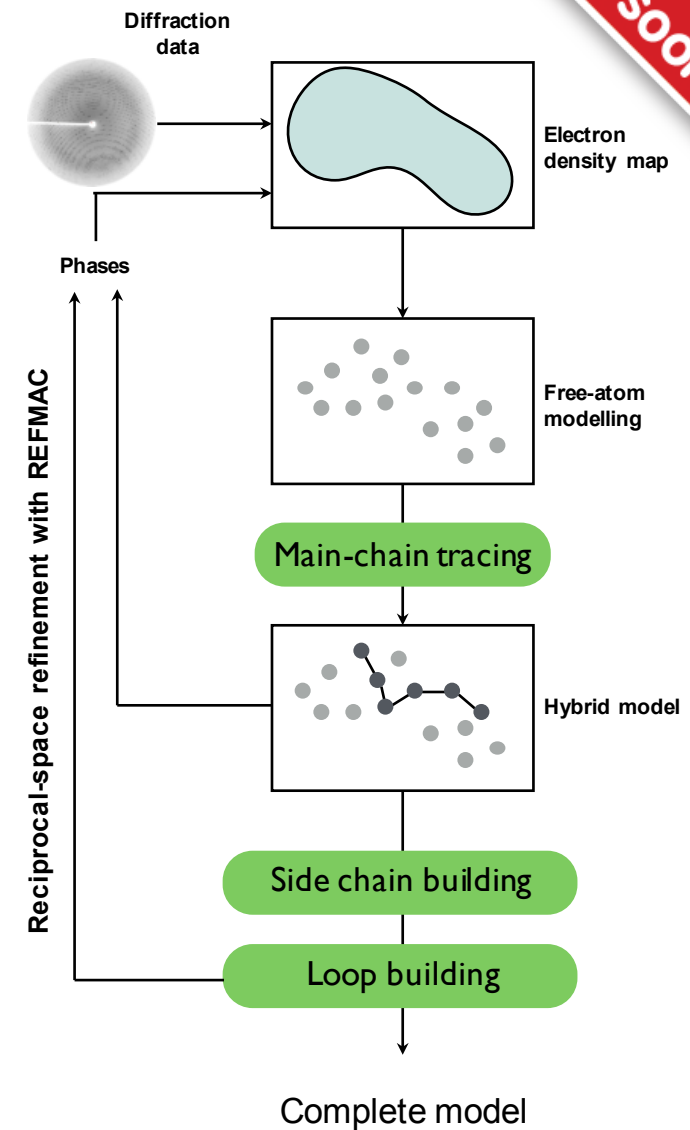
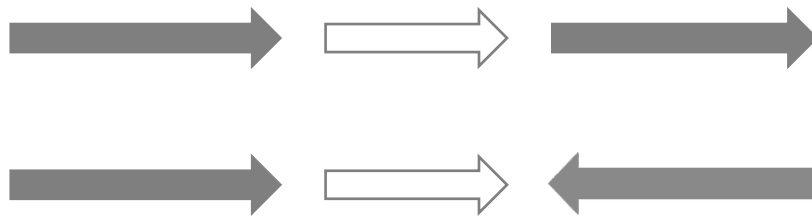
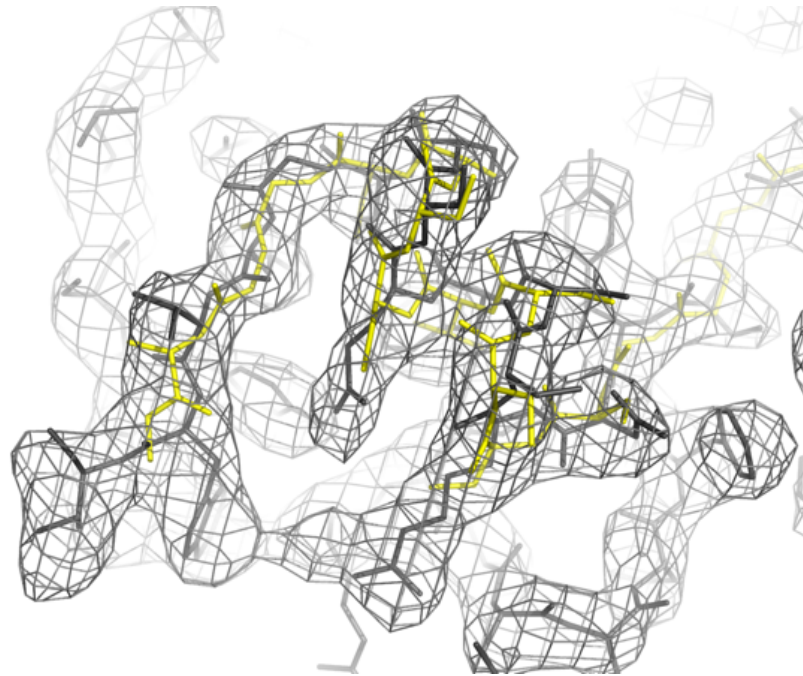
New loop building algorithm

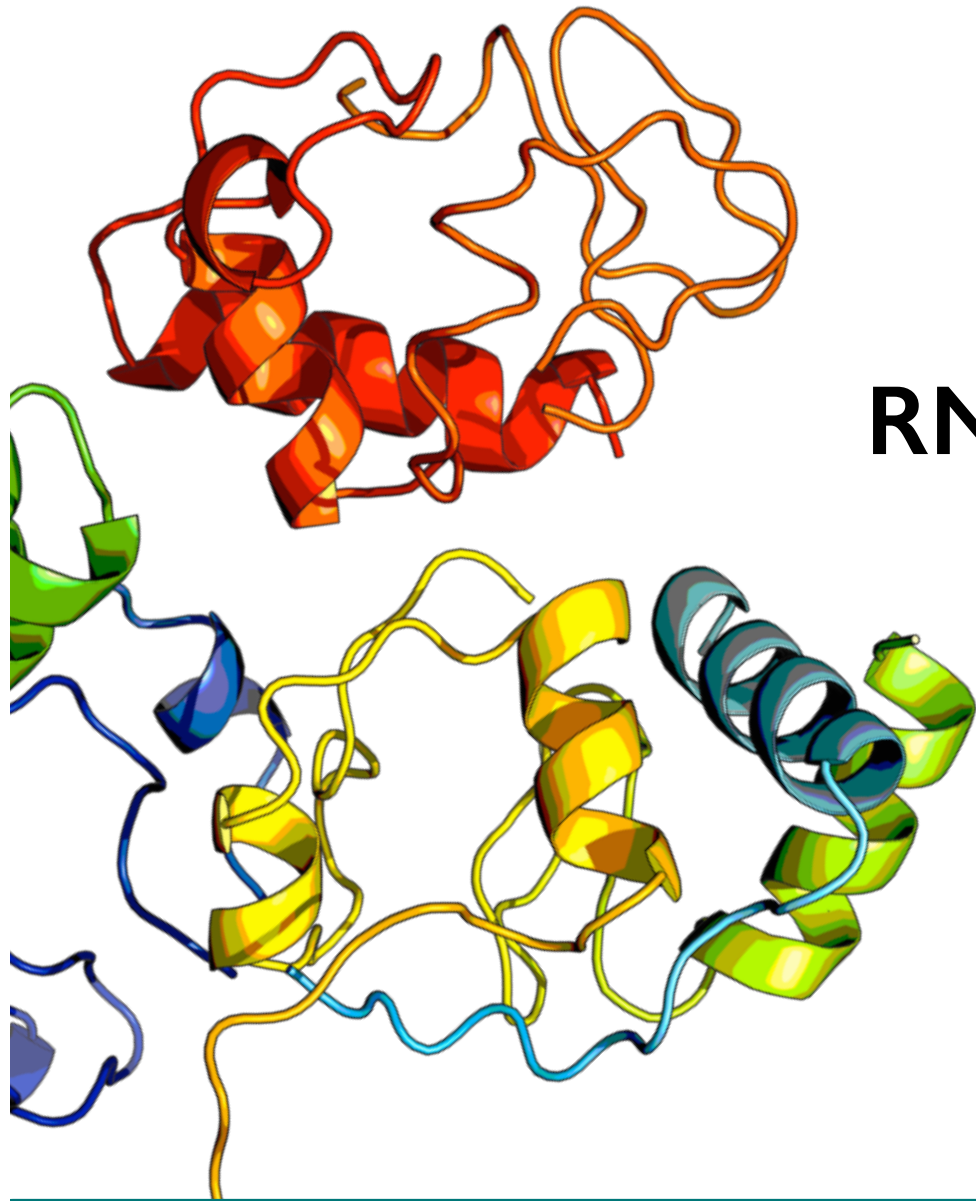


New loop building algorithm



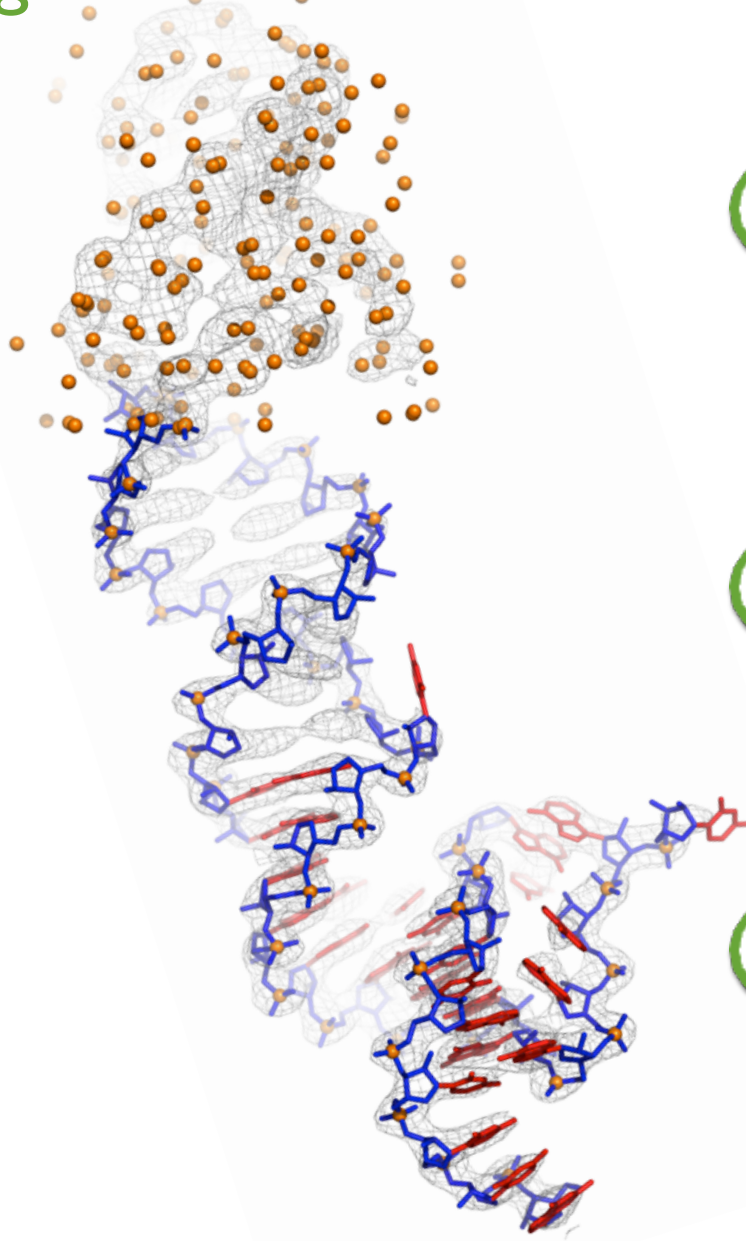
New loop building algorithm





RNA/DNA Model Building with ARP/wARP

Tracing RNA/DNA chains with ARP/wARP



1

Backbone phosphates detection

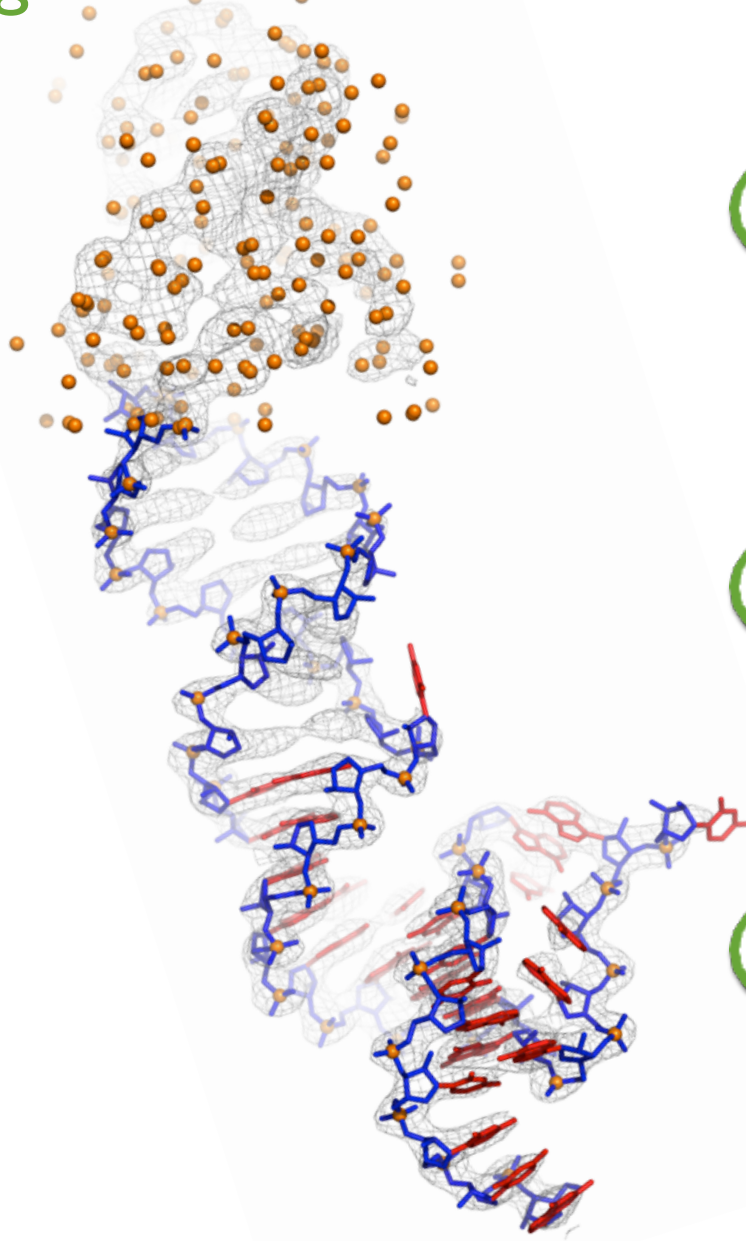
2

Phosphate-ribose-phosphate units placement

3

Bases building

Tracing RNA/DNA chains with ARP/wARP



1

Backbone phosphates detection

NEW

Machine learning algorithm

2

Phosphate-ribose-phosphate units placement

NEW

Updated geometry restraints

NEW

Iterative chain tracing (like for proteins)

3

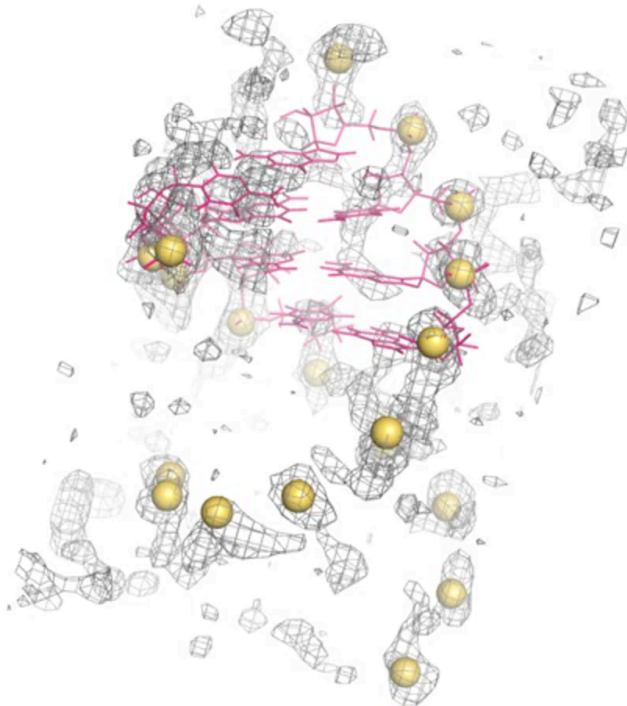
Bases building

Fitting recurrent RNA/DNA fragments with Brickworx

Brickworx builds recurrent RNA and DNA structural motifs into medium-and **low-resolution** electron-density maps.

RNA motifs taken from  **RNA Bricks**
<http://iimcb.genesilico.pl/rnabricks>

A web-server and standalone version available at
<http://iimcb.genesilico.pl/brickworx>



BrickworX Home Help Downloads

BrickworX web server builds RNA and DNA models into electron density maps at resolutions as low as 4 Å. On input the program requires a MTZ file with structure factor amplitudes and phases. Optionally, the user may provide phosphate group phosphorus atom positions (predicted internally by default). On output Brickworx provides two files in PDB format; the predicted phosphorus atom pattern (if applicable) and a model in a full-atom representation.

Input data

Input MTZ file (maximum size 50MB): 4jrt_besttls.mtz

Nucleic acid type:

Phosphorus atoms (optional PDB file): No file chosen

Job finished.

Job 5e192a62ce7caa54c1e03a674102f3609747f8f1

[Link to this job](#) [Phosphates .pdb file](#) [Model .pdb file](#)

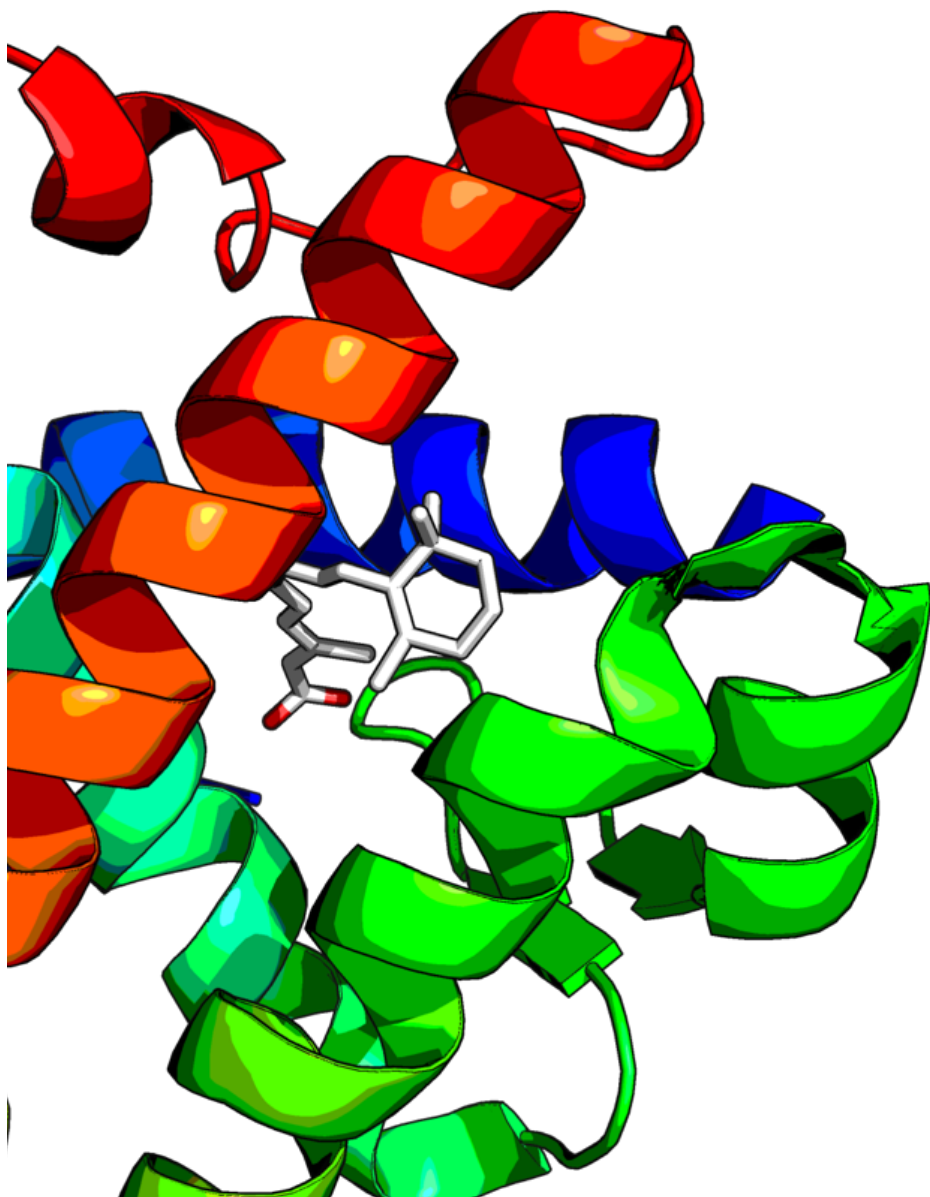
MAP labels

MAP labels:

/10 (3 points/0.56A rmsd/cc 0.1434)
0/10 (3 points/0.68A rmsd/cc 0.1113)
ched motifs/residues: 2/20

^^^ (1/3)
--> Found a loop! (1vx6_loop_363_mcanotate.pdb.gz, score = 6.66)
^^^ (2/3)
--> Found a loop! (1vy1_loop_348_mcanotate.pdb.gz, score = 8.81)
^^^ (3/3)
=> Found 18 residues and 4 chain segments
Finished.

Phase 2 (3/3)



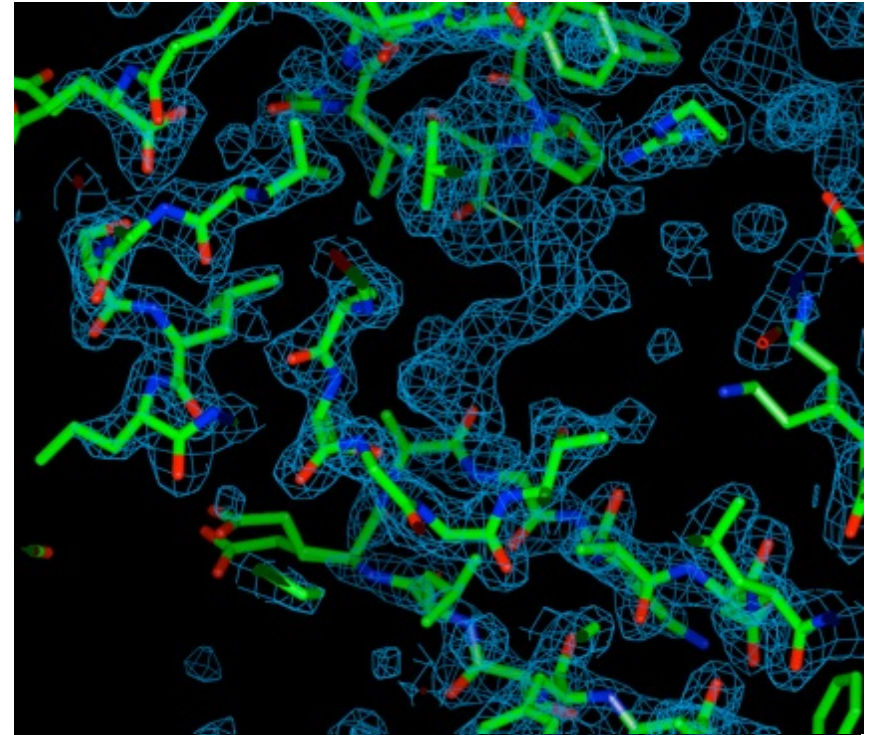
Building ligands with ARP/wARP

Ligand fitting with ARP/wARP



We have:

- electron density map (good resolution)
- protein model (refined)
- difference density map
 - shows excess density not explained by the protein model
 - to fit the appropriate ligand into

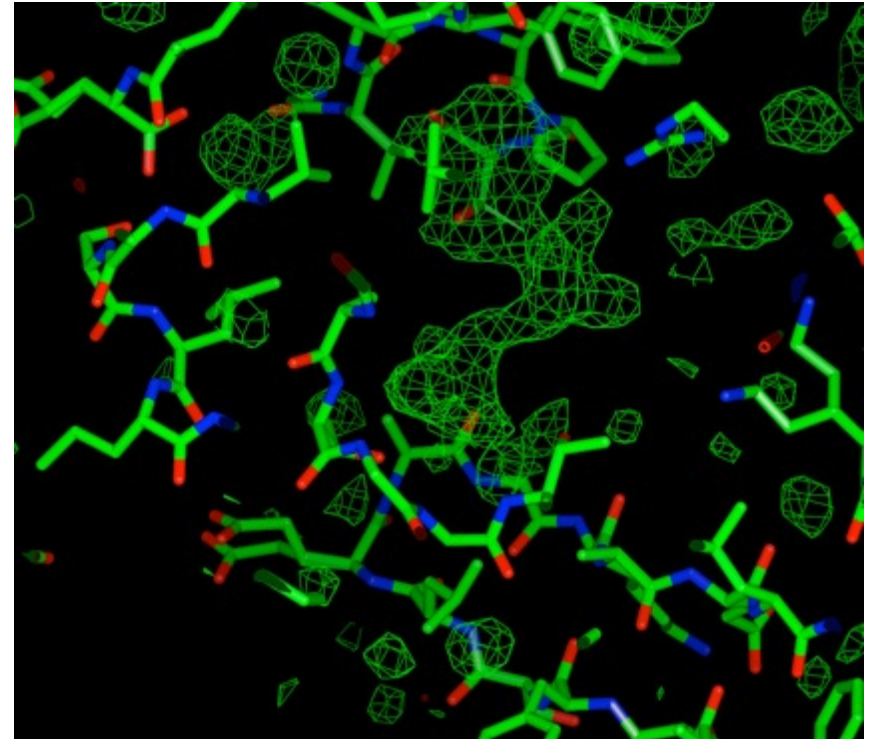


Ligand fitting with ARP/wARP



We have:

- electron density map (good resolution)
- protein model (refined)
- difference density map
 - shows excess density not explained by the protein model
 - to fit the appropriate ligand into

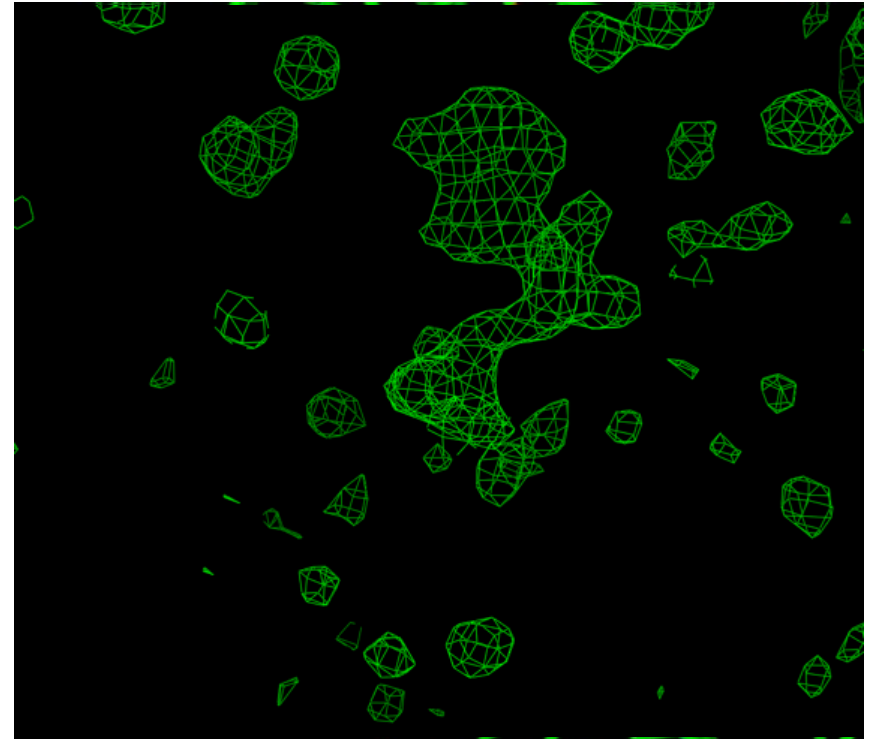


Ligand fitting with ARP/wARP



We have:

- electron density map (good resolution)
- protein model (refined)
- difference density map
 - shows excess density not explained by the protein model
 - to fit the appropriate ligand into

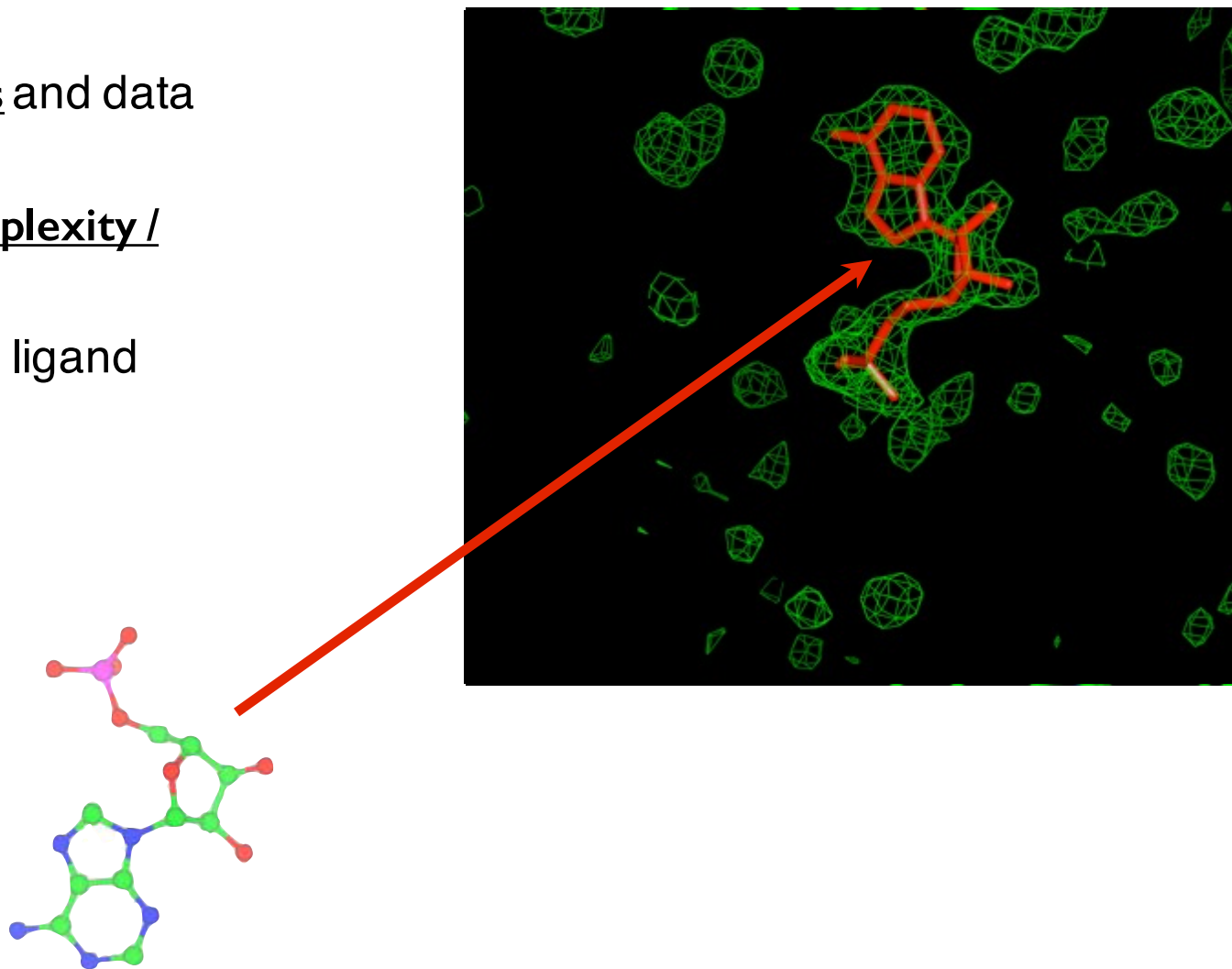


Ligand fitting with ARP/wARP



Challenges:

- different resolutions and data quality
- different ligand complexity / topology
- partial disorder of a ligand



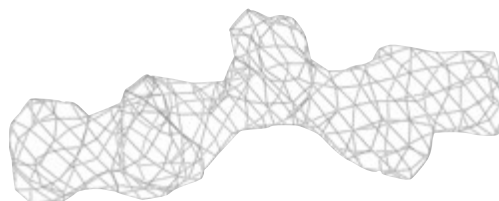
Ligand fitting with ARP/wARP



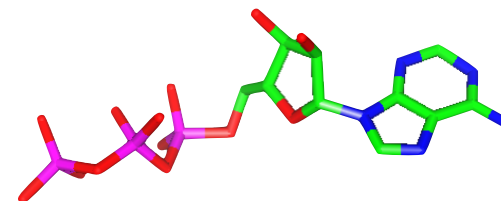
What we know at the beginning will determine the protocol:

		Binding site	
		Known	Unknown
Ligand identity	Known	X	
	Unknown		

1 ligand density



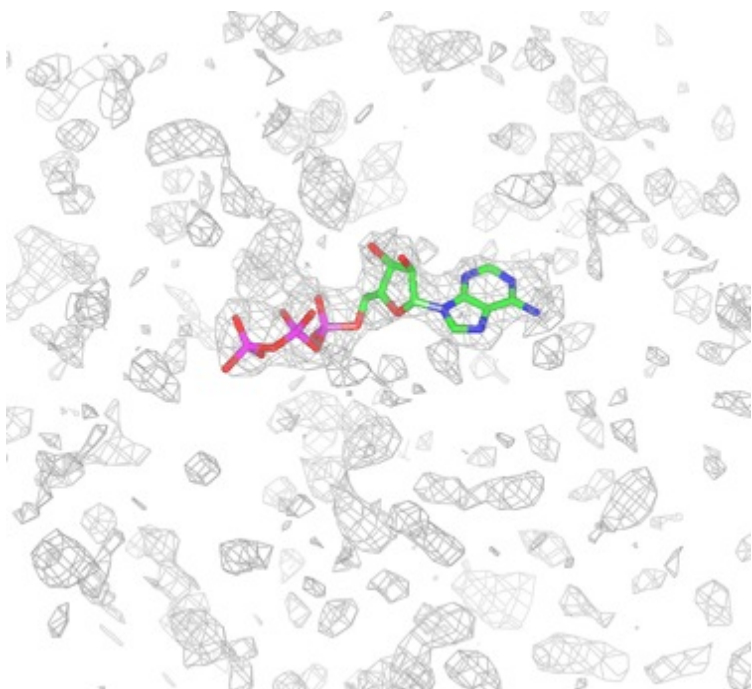
1 ligand



Ligand fitting with ARP/wARP



Organised as a pipeline of core modules for specific sub-tasks following an intuitive approach



Prepare

- Prepare Sparse density map
- Generate ligand topology

Fit

- Construct ensemble of ligand models
- fit sparse density

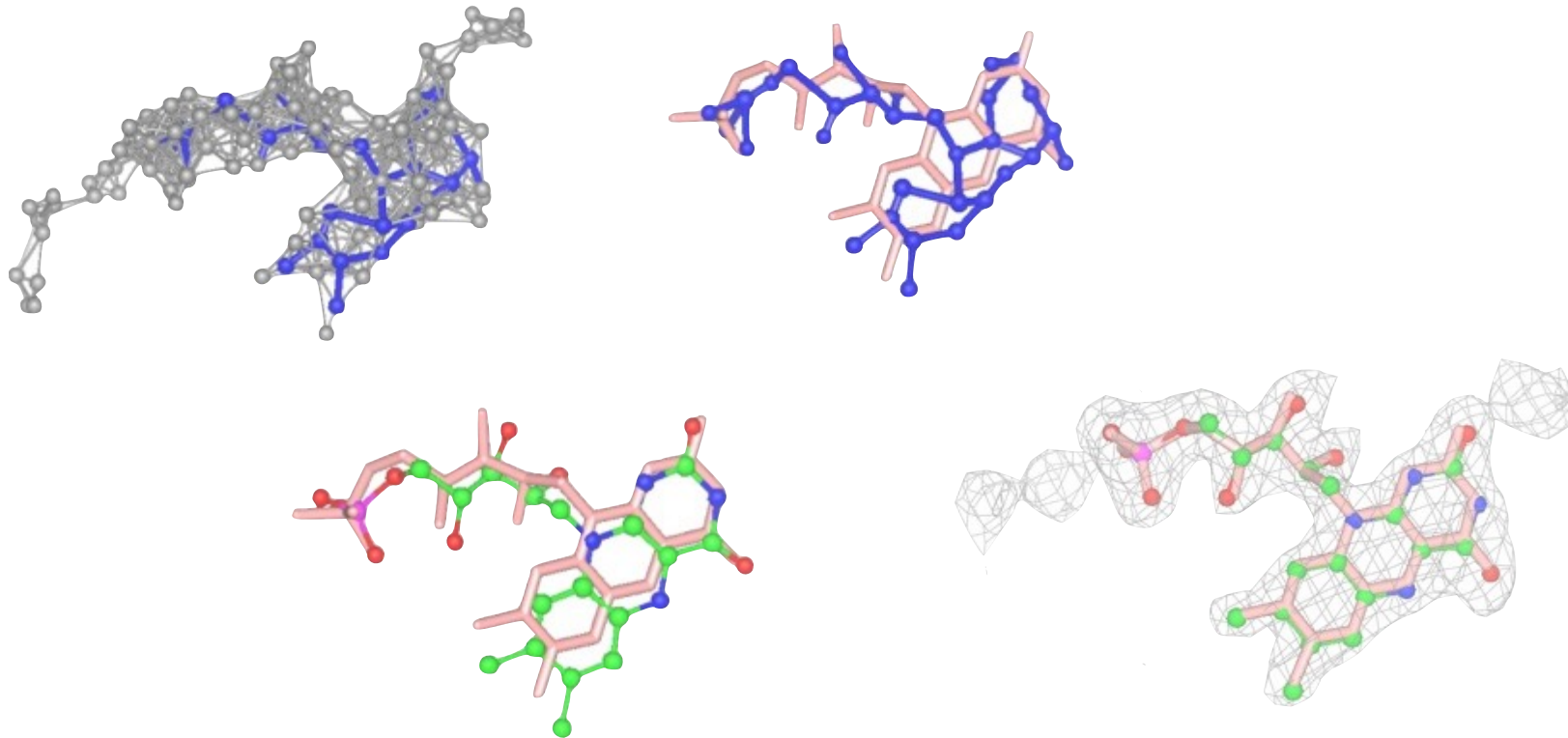
Refine

- Refine ligand coordinates
- Choose best model

Ligand fitting with ARP/wARP



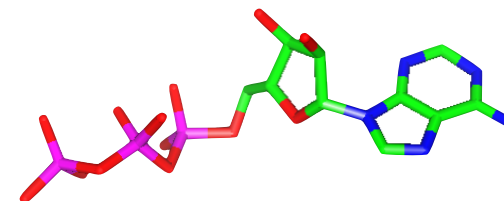
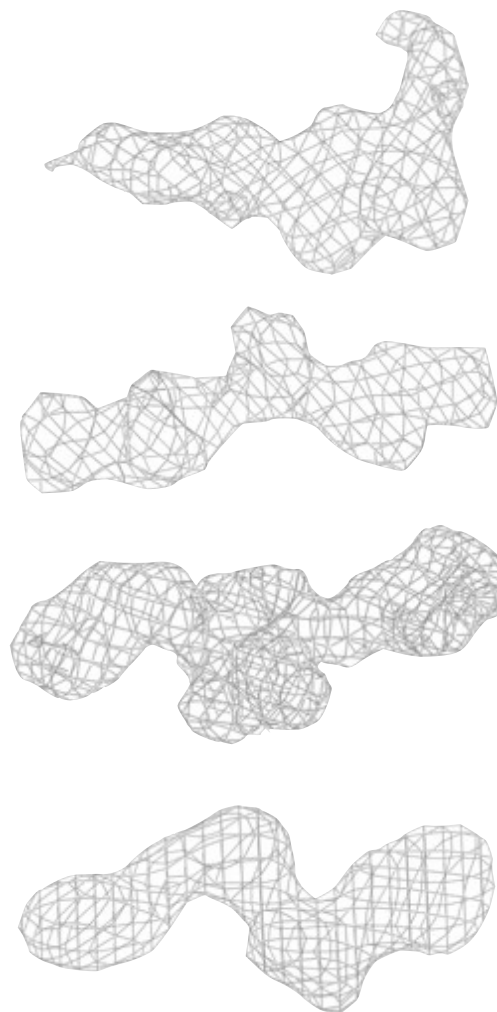
- Uses the full density and topology information
- Employs a gradient method to optimise fit to density and ligand geometry simultaneously



Ligand fitting with ARP/wARP

What we know at the beginning will determine the protocol:

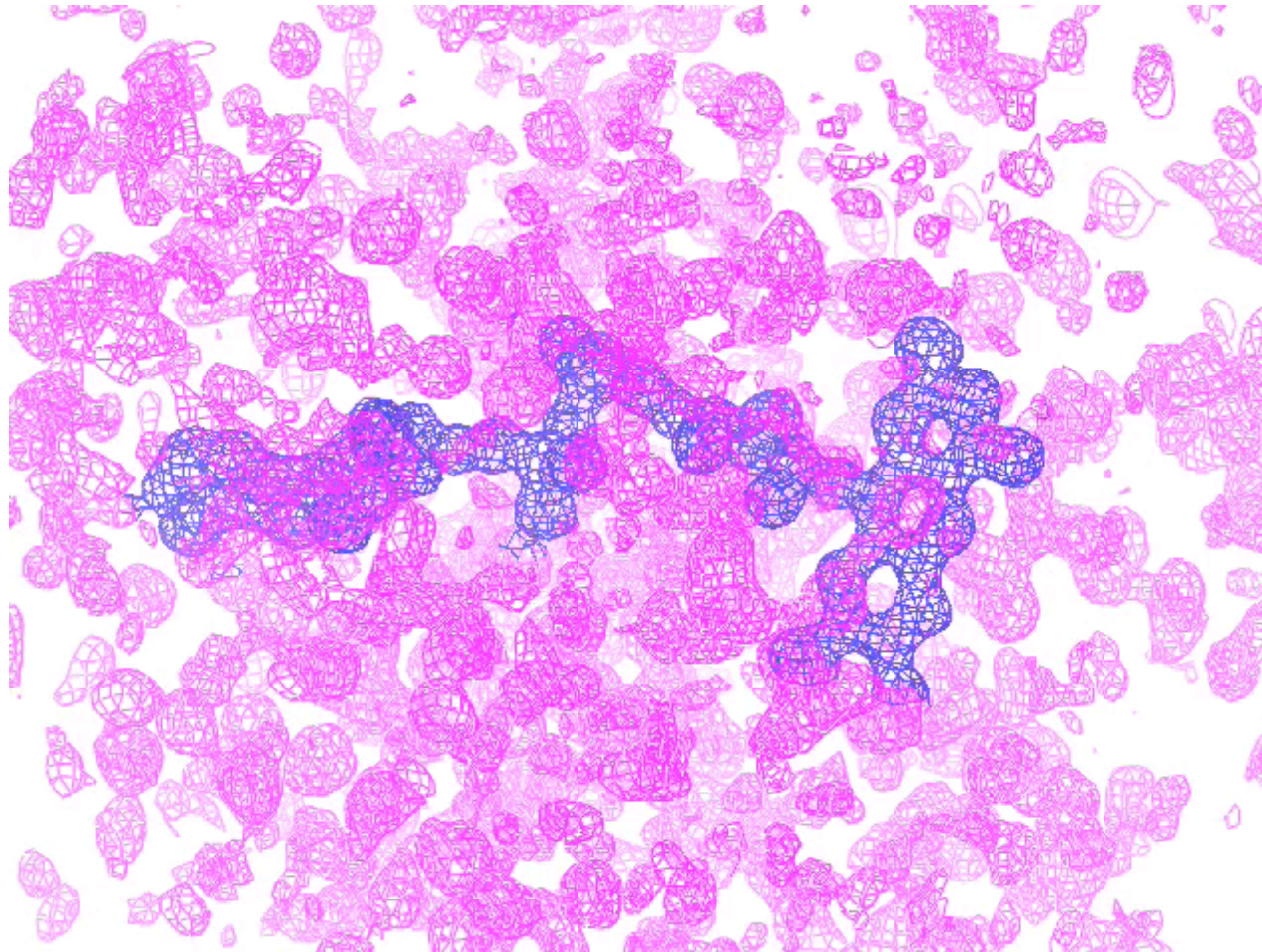
Ligand identity	Binding site	
	Known	Unknown
	Known	X
Unknown		



Ligand fitting with ARP/wARP



- The difference density map from low to high contour threshold

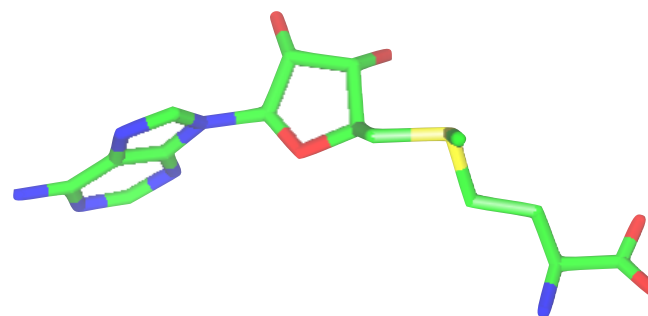
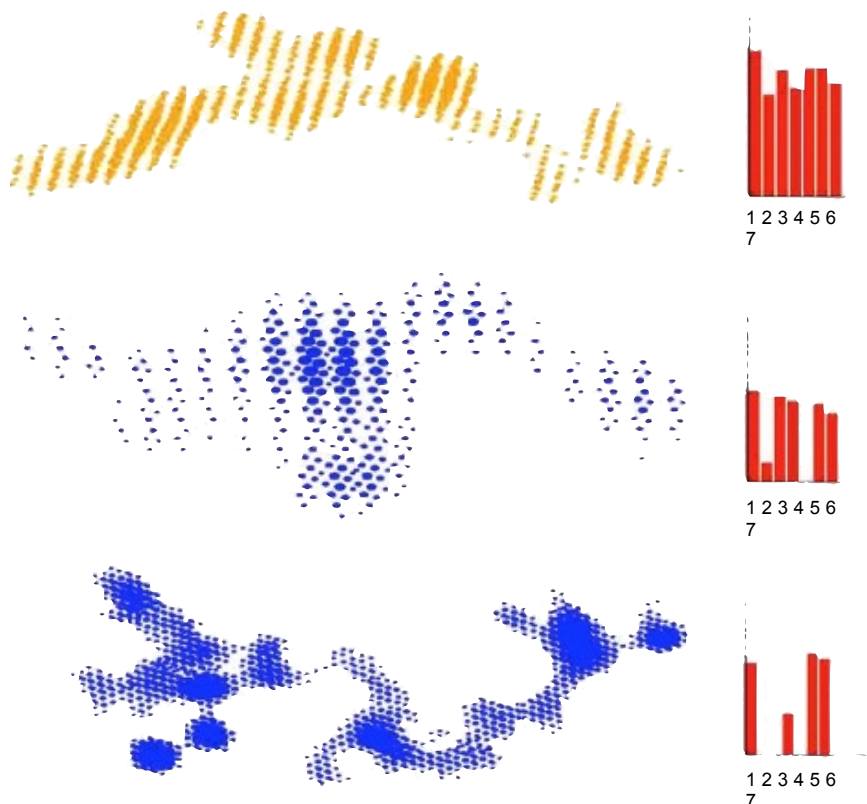


Ligand fitting with ARP/wARP



An electron density map calculated from the ligand to be fit is compared to each potential density cluster:

Find the matching cluster for this ligand!



- 1) Surface to volume ratio
- 2) Bounding box limits
- 3) Moments of inertia
- 4) Rotation match score
- 5) Eigenvalues
- 6) Distance histogram
- 7) Geodesic distance histogram

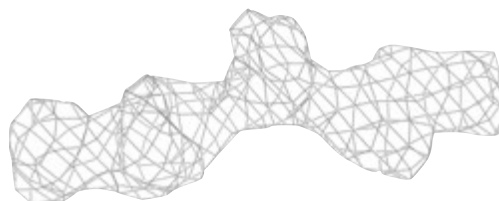
Decision is based on feature vectors

Ligand fitting with ARP/wARP

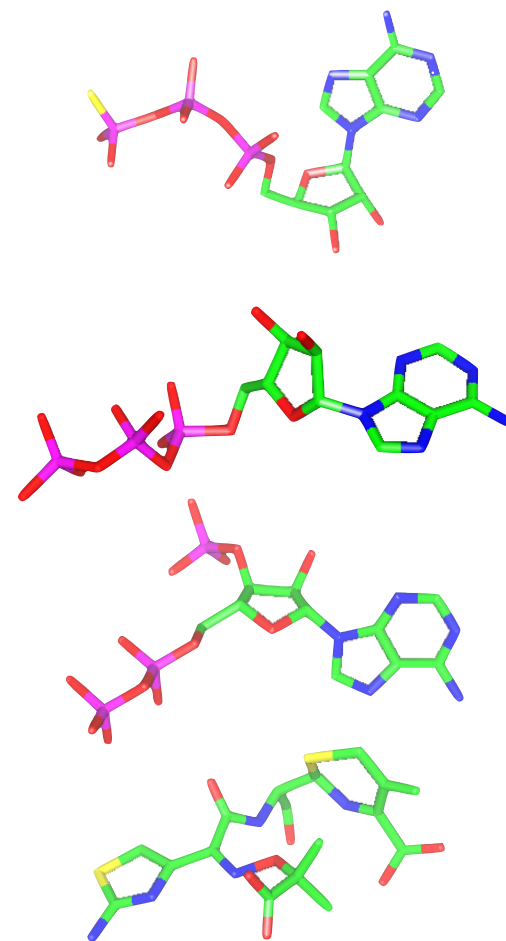
What we know at the beginning will determine the protocol:

		Binding site	
		Known	Unknown
Ligand identity	Known		
	Unknown	X	

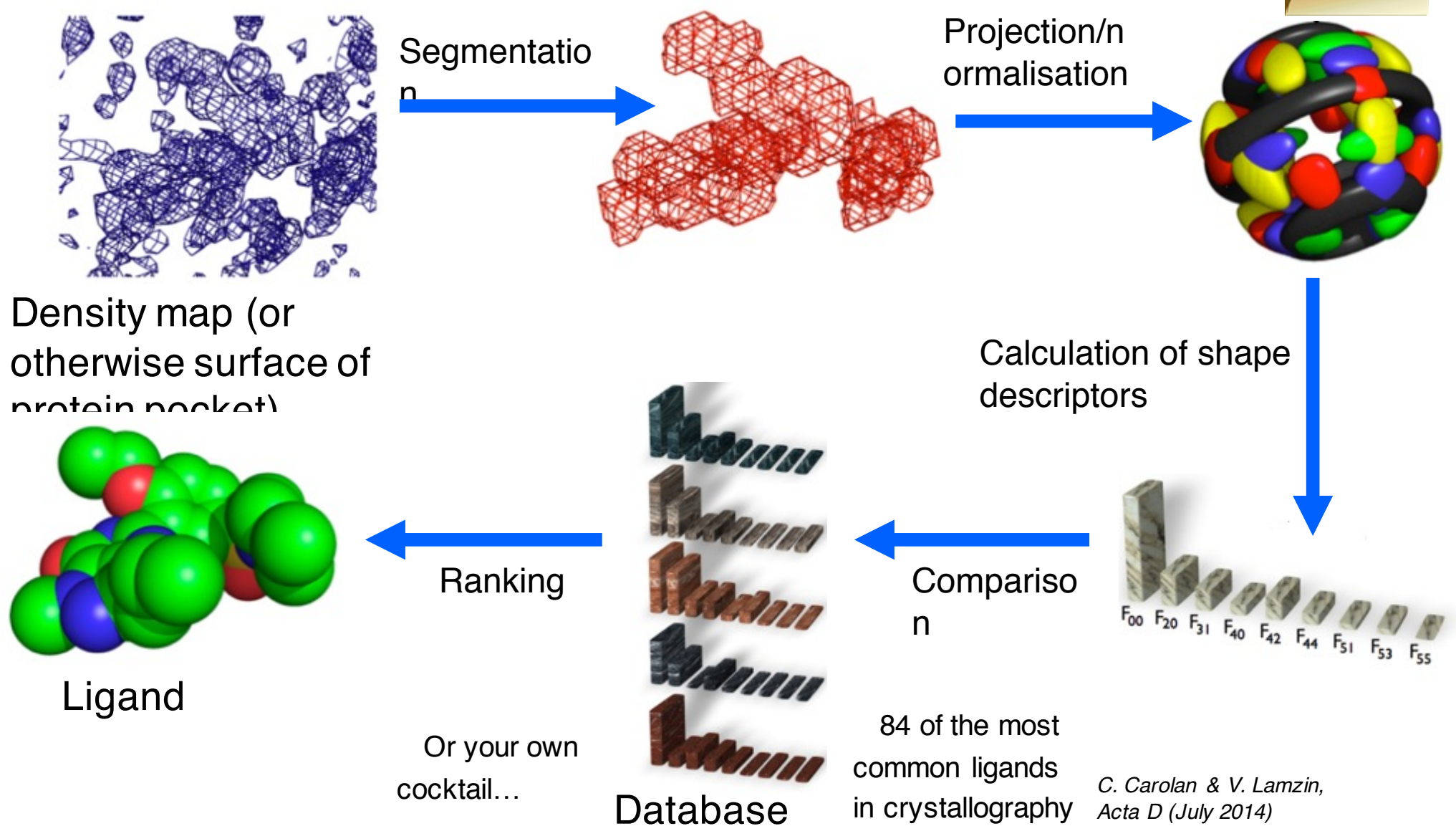
1 ligand density



N ligand candidates



Ligand fitting with ARP/wARP



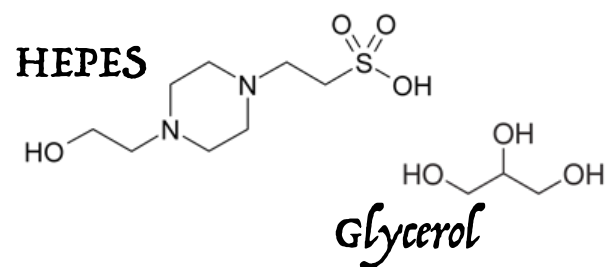
Ligand fitting with ARP/wARP

C. Carolan & V. Lamzin,
Acta D (July 2014)

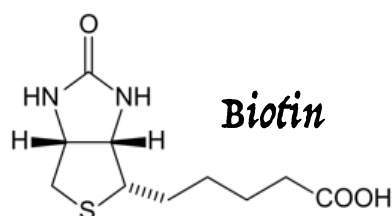


- 84 of the most common ligands in crystallography

Ligand three-letter code	Ligand common name
017	Darunavir
1PE	Pentaethylene glycol
2GP	Guanosine 2-monophosphate
2PE	Nonaethylene glycol
5GP*	Guanosine 5-monophosphate
A3P*	Adenosine 3',5'-diphosphate
ACO*	Acetyl coenzyme A
ADE	Adenine
ADN	Adenosine
ADP	Adenosine 5'-diphosphate
AKG	2-Oxoglutaric acid
AMP	Adenosine monophosphate
ATP*	Adenosine 5'-triphosphate
B3P	2-[3-(2-Hydroxy-1,1-dihydroxymethyl-ethylamino)-propylamino]-2-hydroxymethyl-propane-1,3-diol
BCL	Bacteriochlorophyll A
BTB	Bis-tris buffer
BTN	Biotin
C2E*	Cyclic diguanosine monophosphate
CAM	Camphor
CDL	Cardiolipin
CHD	Cholic acid
CIT	Citric acid
CLA	Chlorophyll A
CMP	Adenosine 3',5'-cyclic monophosphate
COA	Coenzyme A
CXS	3-Cyclohexyl-1-propylsulfonic acid
CYC	Phycocyanobilin
DIO	1,4-Diethylene dioxide
DTT	1,4-Dithiothreitol
EPE	HEPES

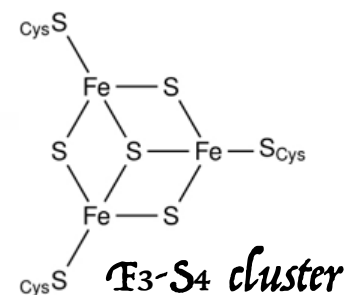
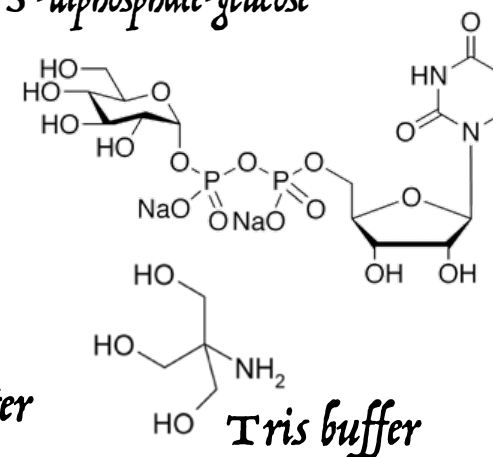


Ligand three-letter code	Ligand common name
F3S	Fe ₃ -S ₄ cluster
FAD*	Flavin-adenine dinucleotide
FMN*	Flavin mononucleotide
FPP	Farnesyl diphosphate
GOL	Glycerol
GSH	Glutathione
H4B	5,6,7,8-Tetrahydrobiopterin
HC4	<i>para</i> -Coumaric acid
HEA*	Haem A
HED	2-Hydroxyethyl disulfide
HEM	Haem
IMD	Imidazole
IPH	Phenol
LDA	Lauryl dimethylamine- <i>N</i> -oxide
MES	2-(<i>N</i> -Morpholino)ethanesulfonic acid
MLI	Malonate ion
MLT	D-Malate
MPD	(4 <i>S</i>)-2-Methyl-2,4-pentanediol
MTE	Phosphonic acid mono-(2-amino-5,6-dimercapto-4-oxo-3,7,8A,9,10,10A-hexahydro-4H-8-oxa-1,3,9,10-tetraaza-anthracen-7-ylmethyl)ester
MYR	Myristic acid
NAD*	Nicotinamide adenine dinucleotide
NAP*	Nicotinamide adenine dinucleotide phosphate
NCO	Cobalt hexammine(III)
NHE	2-(<i>N</i> -Cyclohexylamino)ethanesulfonic acid
OLA	Oleic acid
ORO	Orotic acid
P6G	Hexaethylene glycol
PEG	Di(hydroxyethyl)ether
PEP	Phosphoenolpyruvate
PG4	Tetraethylene glycol
PGA	2-Phosphoglycolic acid
PGO	<i>S</i> -1,2-Propanediol
PHQ	Benzyl chlorocarbonate



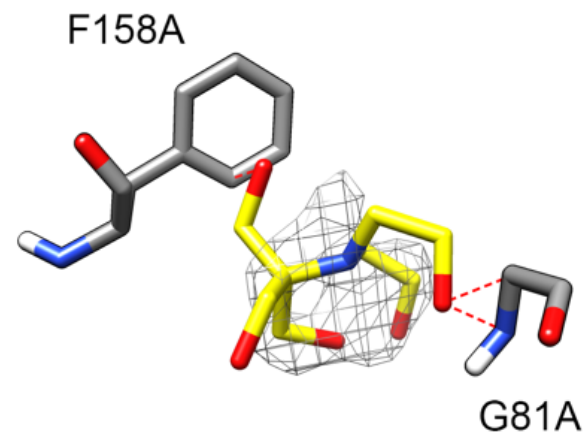
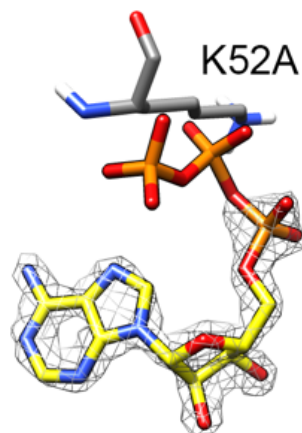
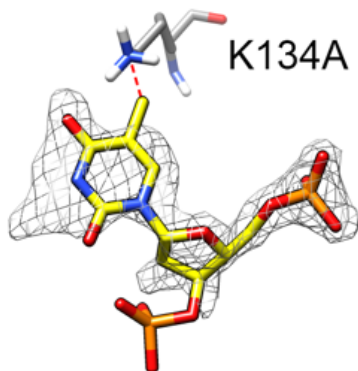
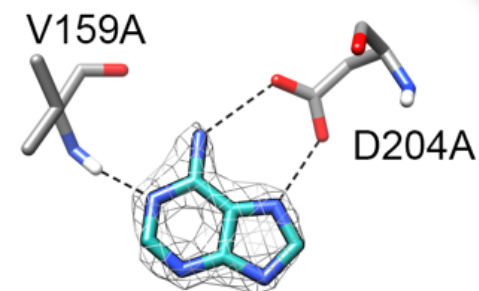
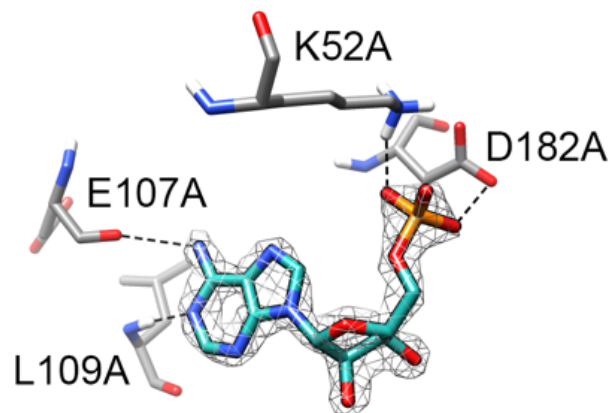
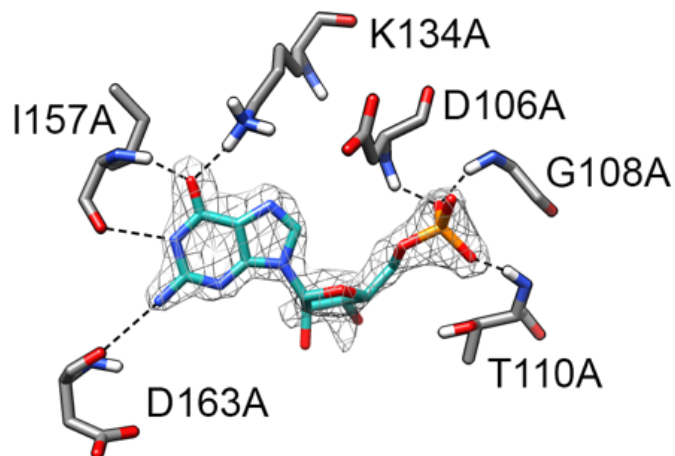
Ligand three-letter code	Ligand common name
PLM	Palmitic acid
PLP	Pyridoxal-5'-phosphate
POP	Pyrophosphate ²⁻
PYR	Pyruvic acid
RET	Retinal
SAM*	<i>S</i> -Adenosylmethionine
SF4	Iron-sulfur cluster
SIA	<i>O</i> -Sialic acid
SO4	Sulfate ion
SPO	Spheroindene
STU*	Staurosporine
TAM	Tris(hydroxyethyl)aminomethane
THP	Thymidine 3',5'-diphosphate
TLA	L-(+)-Tartaric acid
TPP	Thiamine diphosphate
TRS	Tris buffer
TYD	Thymidine 5'-diphosphate
U10	Coenzyme Q10
UPG	Uridine 5'-diphosphate-glucose

Uridine 5'-diphosphate-glucose



Ligand fitting with ARP/wARP

$$V_{bound}^{P-L} = W_{VDW} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + W_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + W_{elec} \sum_{i,j} \left(\frac{q_i q_j}{\epsilon(r_{ij}) r_{ij}} \right) + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i) \exp \left(\frac{-r_{ij}^2}{2\sigma^2} \right)$$



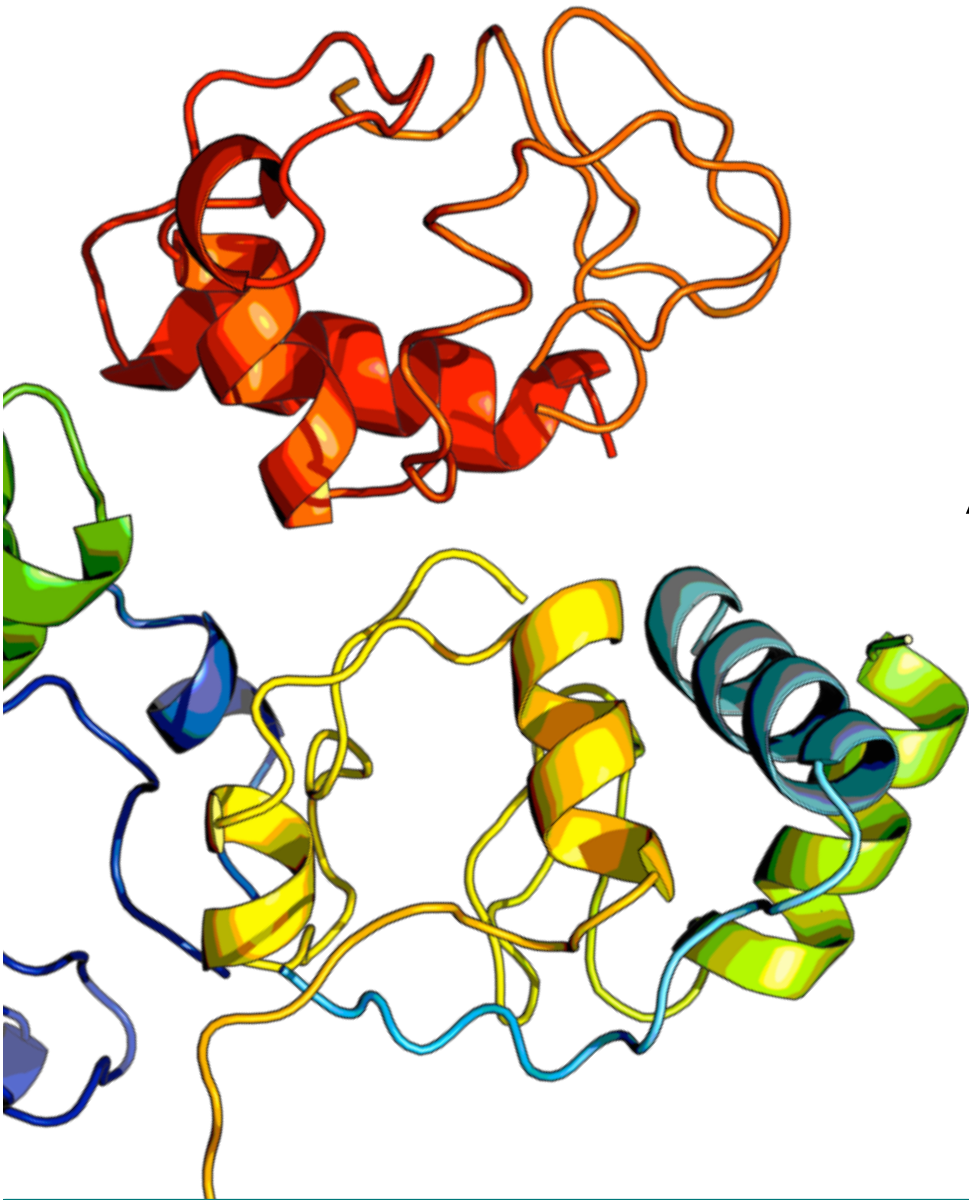
Ligand guessing without (yellow skeleton) and with (blue skeleton) the use of the estimated energy as an additional parameter

D. Beshnova, J. Pereira, V. Lamzin, Acta D (2017)

coming soon!

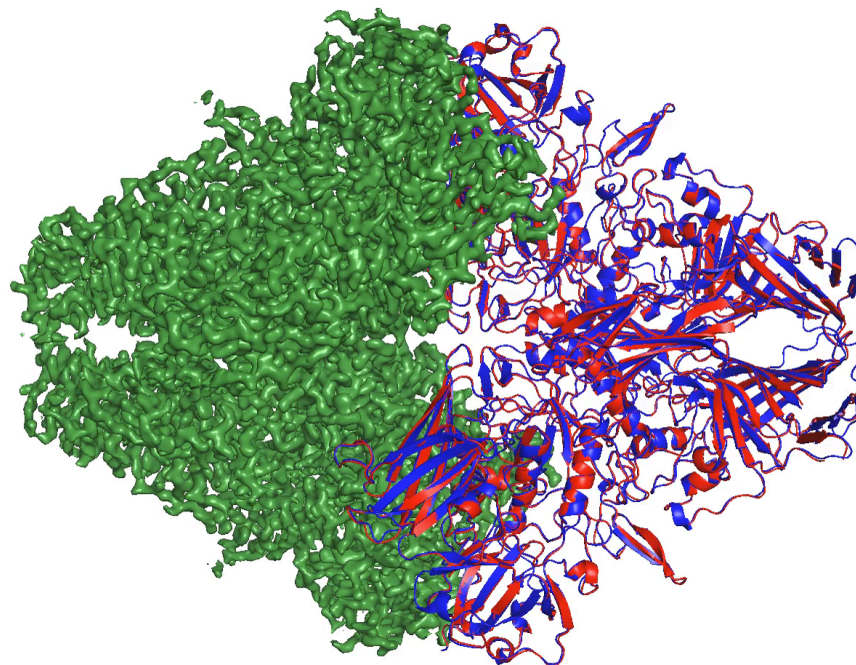
coming
soon!

ARP/wARP for cryo-EM



ARP/wARP for cryo-EM

After ARP/wARP (4 hours of CPU time)
~ **93%** (3800 out of 4088) of the
residues traced and docked

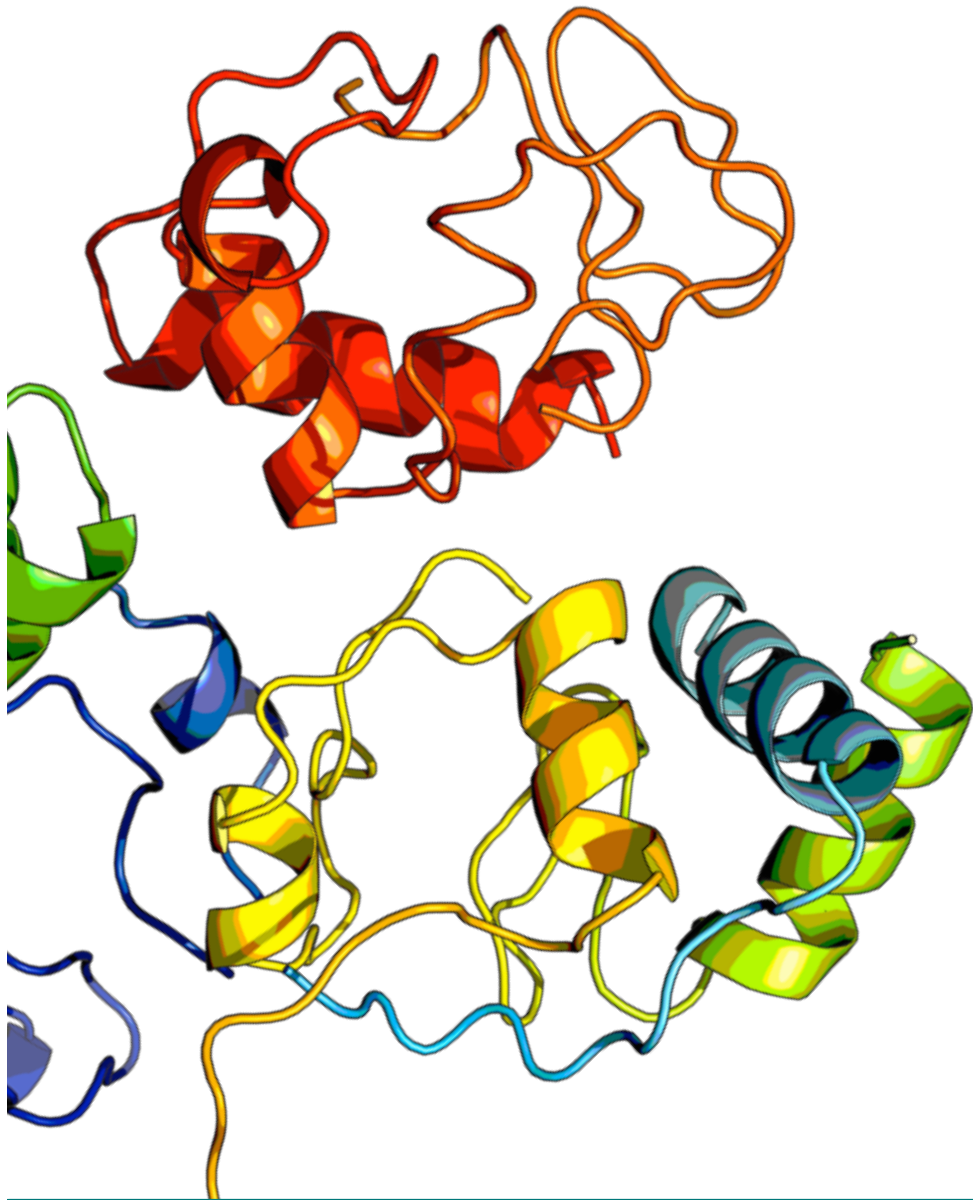


β -galactosidase @ 2.2 Å resolution (EMD-2984, 5a1a)

ARP/wARP model

deposited coordinates

coming
soon!



ARP/wARP tips and tricks

ARP/wARP tips and tricks

Resolution

best results above 2.6Å,
but phases can be poor

Space group

double check if model
building fails

Data completeness

as high as possible.

Data quality (outliers, ice rings)

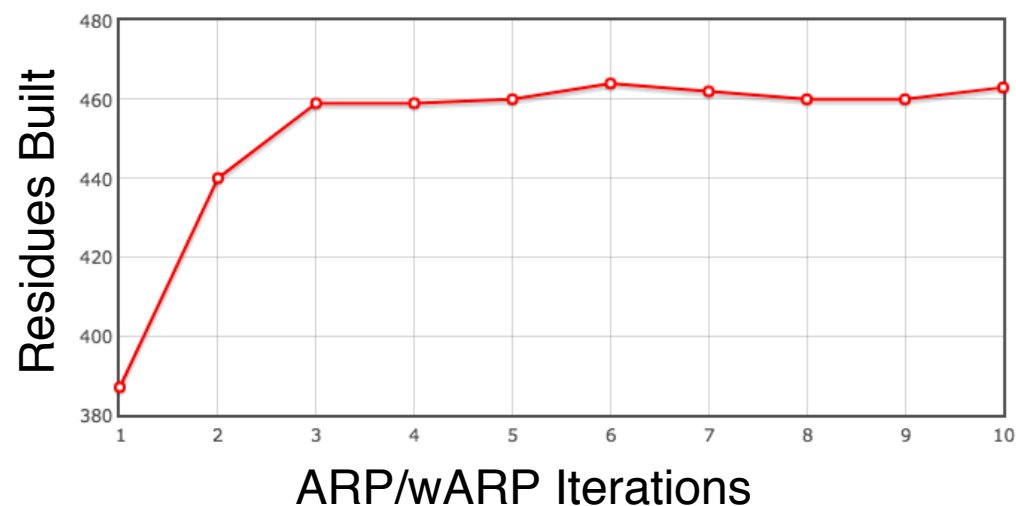
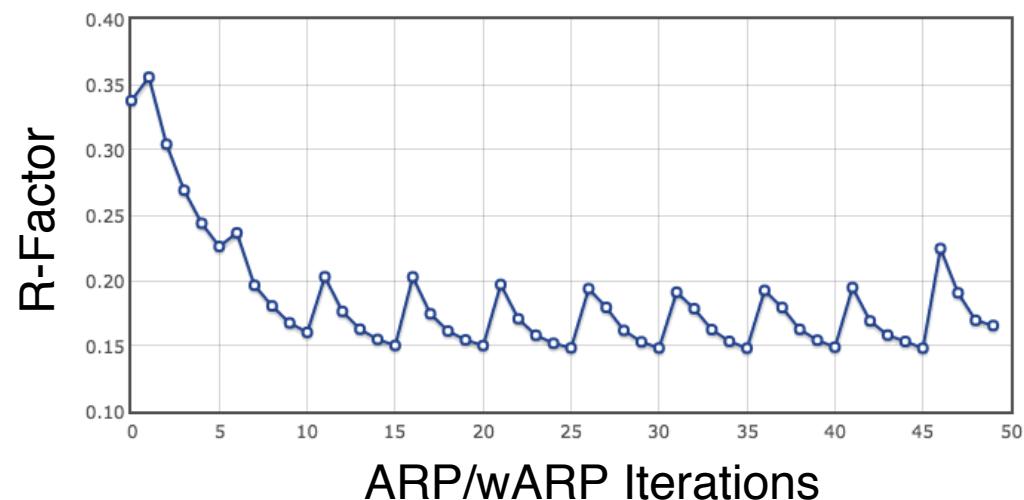
always check Wilson plot

Check intermediate models

use keepjunk option

Use secondary structure modeling

auto_albe.sh or quickfold in CCP4i



ARP/wARP tips and tricks

Resolution

best results above 2.6Å,
but phases can be poor

Space group

double check if model
building fails

Data completeness

as high as possible.

Data quality (outliers, ice rings)

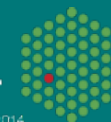
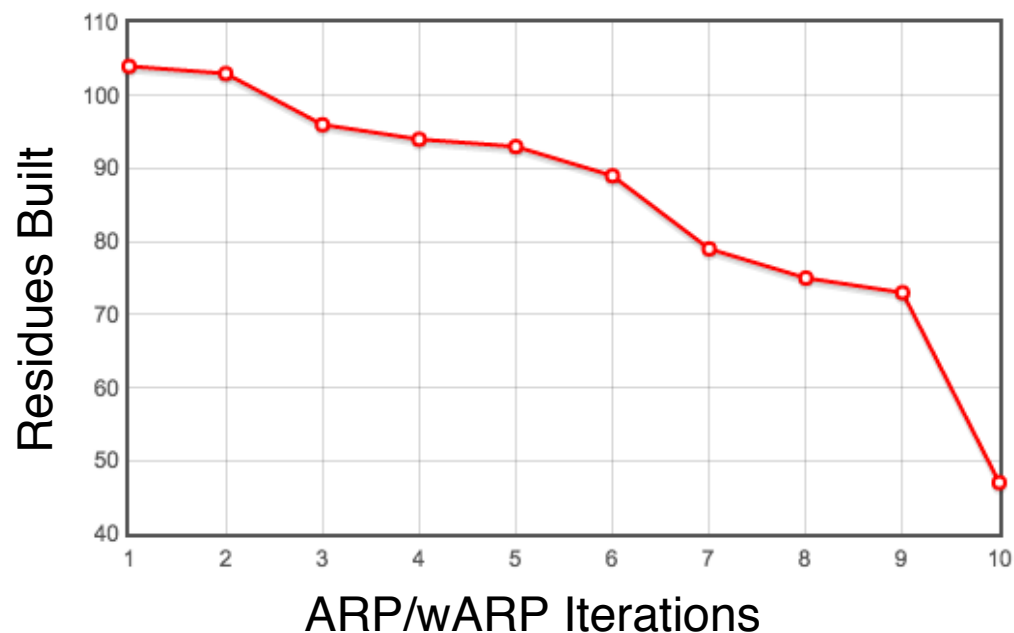
always check Wilson plot

Check intermediate models

use keepjunk option

Use secondary structure modeling

auto_albe.sh or quickfold in CCP4i



ARP/wARP tips and tricks

Resolution

best results above 2.6Å,
but phases can be poor

Space group

double check if model
building fails

Data completeness

as high as possible.

Data quality (outliers, ice rings)

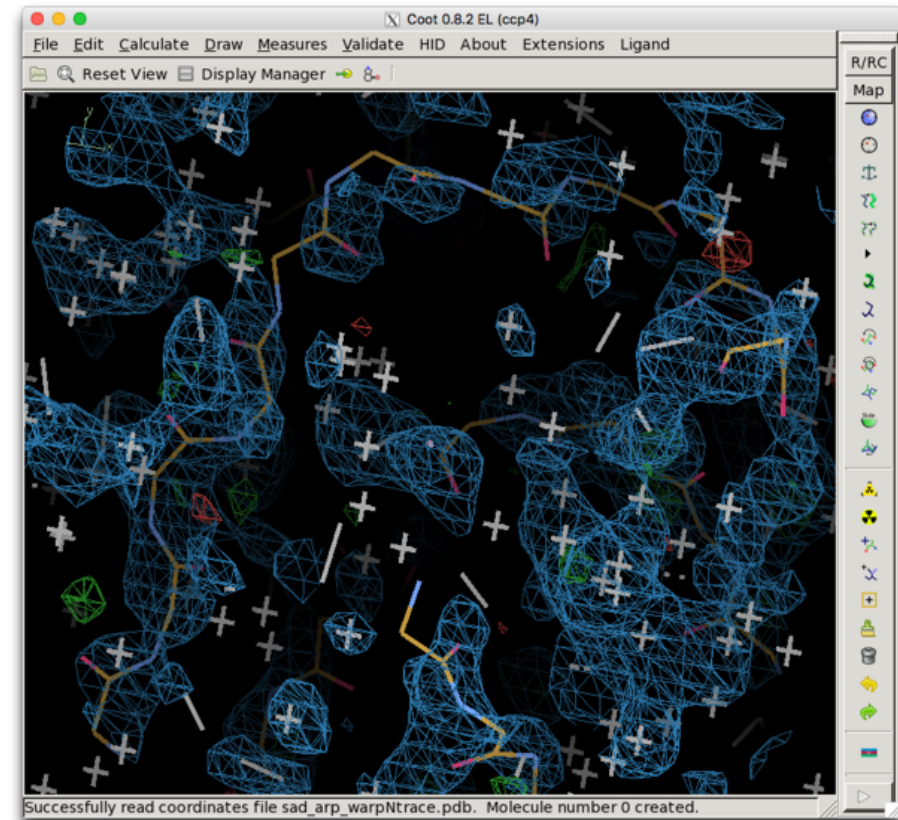
always check Wilson plot

Check intermediate models

use keepjunk option

Use secondary structure modeling

auto_albe.sh or quickfold in CCP4i



ARP/wARP tips and tricks

Resolution

best results above 2.7Å,
but phases can be poor

Data completeness

as high as possible.

Space group

double check if model
building fails

Data quality (outliers, ice rings)

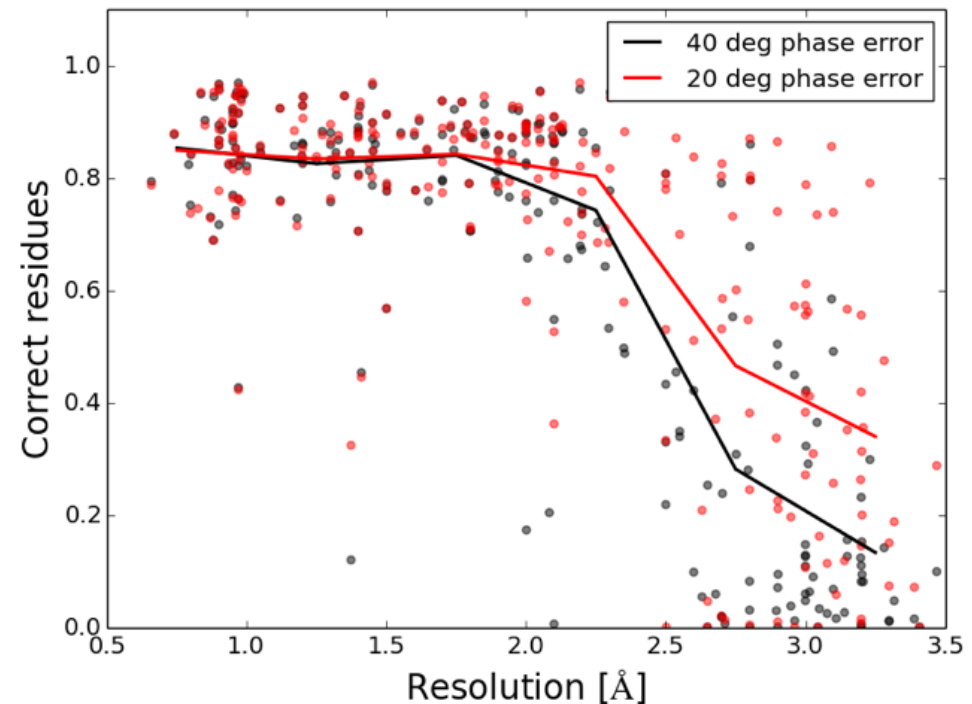
always check Wilson plot

Check intermediate models

use keepjunk option

Use secondary structure modeling

auto_albe.sh or quickfold in CCP4i



ARP/wARP tips and tricks

Resolution

best results above 2.6Å,
but phases can be poor

Data completeness

as high as possible.

Space group

double check if model
building fails

Data quality (outliers, ice rings)

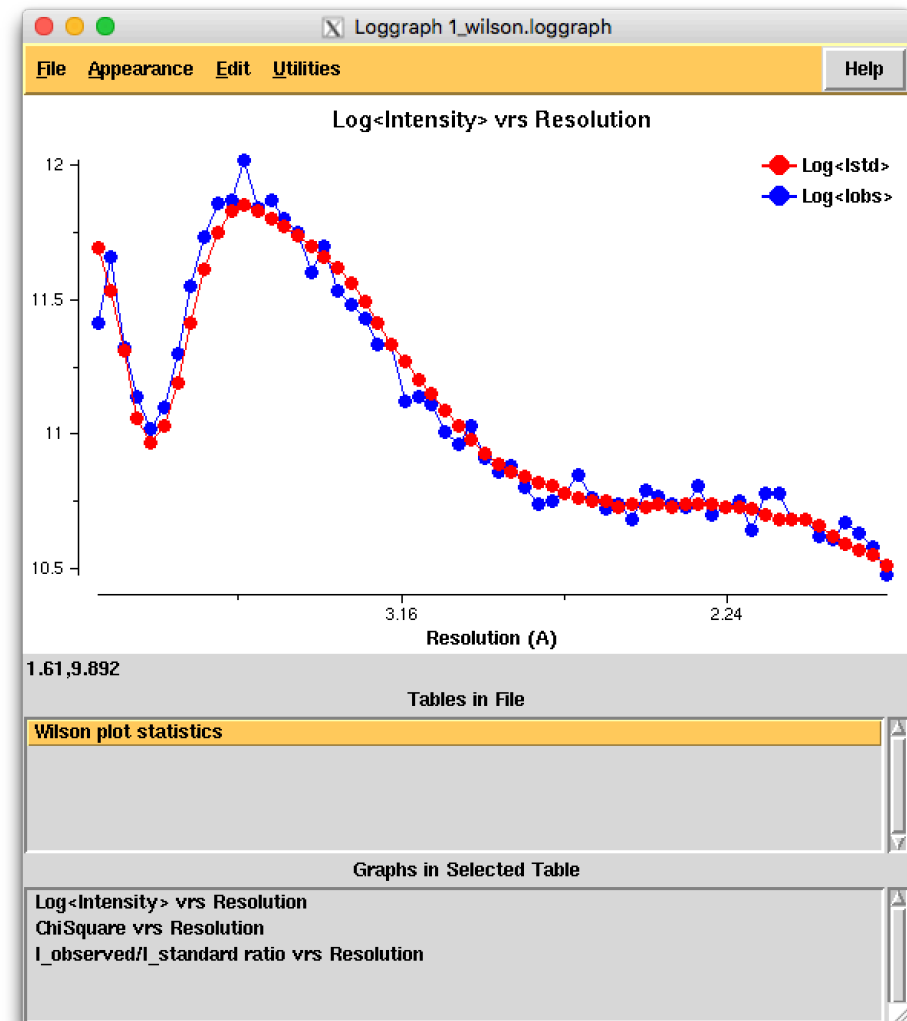
always check Wilson plot

Check intermediate models

use keepjunk option

Use secondary structure modeling

auto_albe.sh or quickfold in CCP4i



ARP/wARP tips and tricks

Resolution

best results above 2.6Å,
but phases can be poor

Data completeness

as high as possible.

Space group

double check if model
building fails

Data quality (outliers, ice rings)

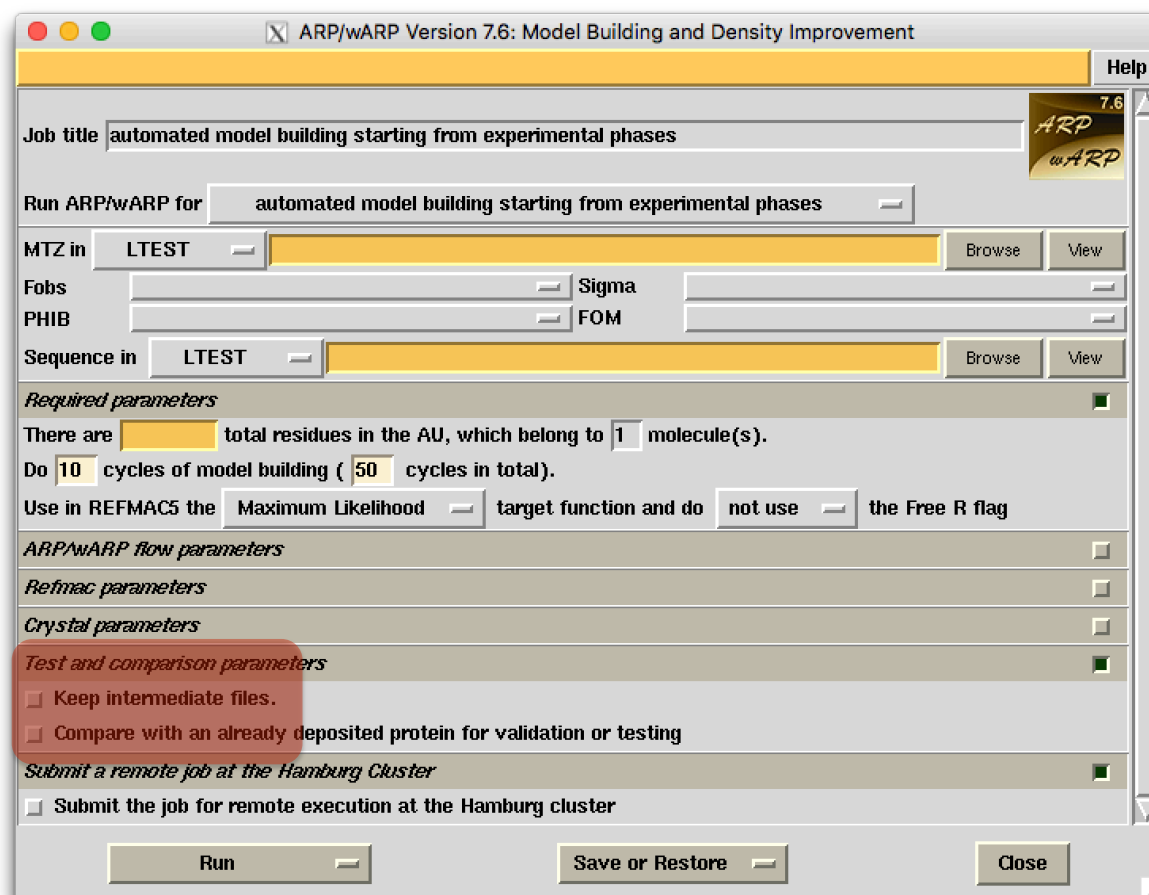
always check Wilson plot

Check intermediate models

use *keepjunk* option

Use secondary structure modeling

auto_albe.sh or quickfold in CCP4i



ARP/wARP tips and tricks

Resolution

best results above 2.6Å,
but phases can be poor

Data completeness

as high as possible.

Space group

double check if model
building fails

Data quality (outliers, ice rings)

always check Wilson plot

Check intermediate models

use keepjunk option

Use secondary structure modeling

auto_albe.sh or quickfold in CCP4i

```
-----
5036 reflections ( 98.84 % complete ) and 831 restraints for refining 813 atoms.
86 conditional restraints added.
Observations/parameters ratio is 1.55
-----
Cycle 1: After refmac, R = 0.3221 (Rfree = 0.387) for 813 atoms.
Found 21 (21 requested) and removed 10 (10 requested) atoms.
Cycle 2: After refmac, R = 0.2895 (Rfree = 0.379) for 824 atoms.
Found 21 (21 requested) and removed 10 (10 requested) atoms.
Cycle 3: After refmac, R = 0.2554 (Rfree = 0.365) for 835 atoms.
Found 13 (21 requested) and removed 10 (10 requested) atoms.
Cycle 4: After refmac, R = 0.2381 (Rfree = 0.344) for 838 atoms.
Found 11 (21 requested) and removed 6 (10 requested) atoms.
Cycle 5: After refmac, R = 0.2265 (Rfree = 0.336) for 843 atoms.
Found 11 (22 requested) and removed 4 (11 requested) atoms.
-----
Building Cycle 2 Atomic shape factors 1.70 1.90
Round 1: 91 peptides, 3 chains. Longest chain 73 peptides. Score 0.936
Round 2: 91 peptides, 2 chains. Longest chain 73 peptides. Score 0.950
Round 3: 88 peptides, 3 chains. Longest chain 69 peptides. Score 0.930
Round 4: 91 peptides, 2 chains. Longest chain 73 peptides. Score 0.950
Round 5: 90 peptides, 3 chains. Longest chain 71 peptides. Score 0.934
Taking the results from Round 4
-----
Chains 2, Residues 89, Estimated correctness of the model 99.6 %
2 chains (89 residues) have been docked in sequence
```

ARP/wARP tips and tricks

Resolution

best results above 2.6Å,
but phases can be poor

Data completeness

as high as possible.

Space group

double check if model
building fails

Data quality (outliers, ice rings)

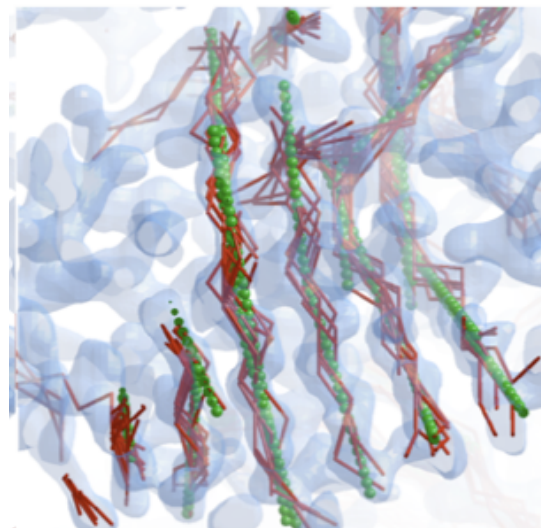
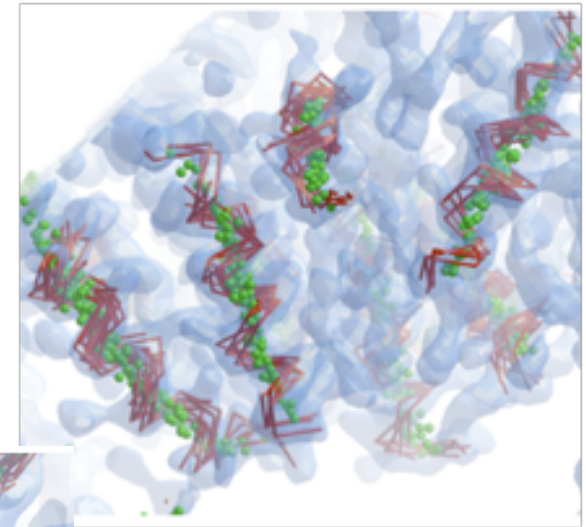
always check Wilson plot

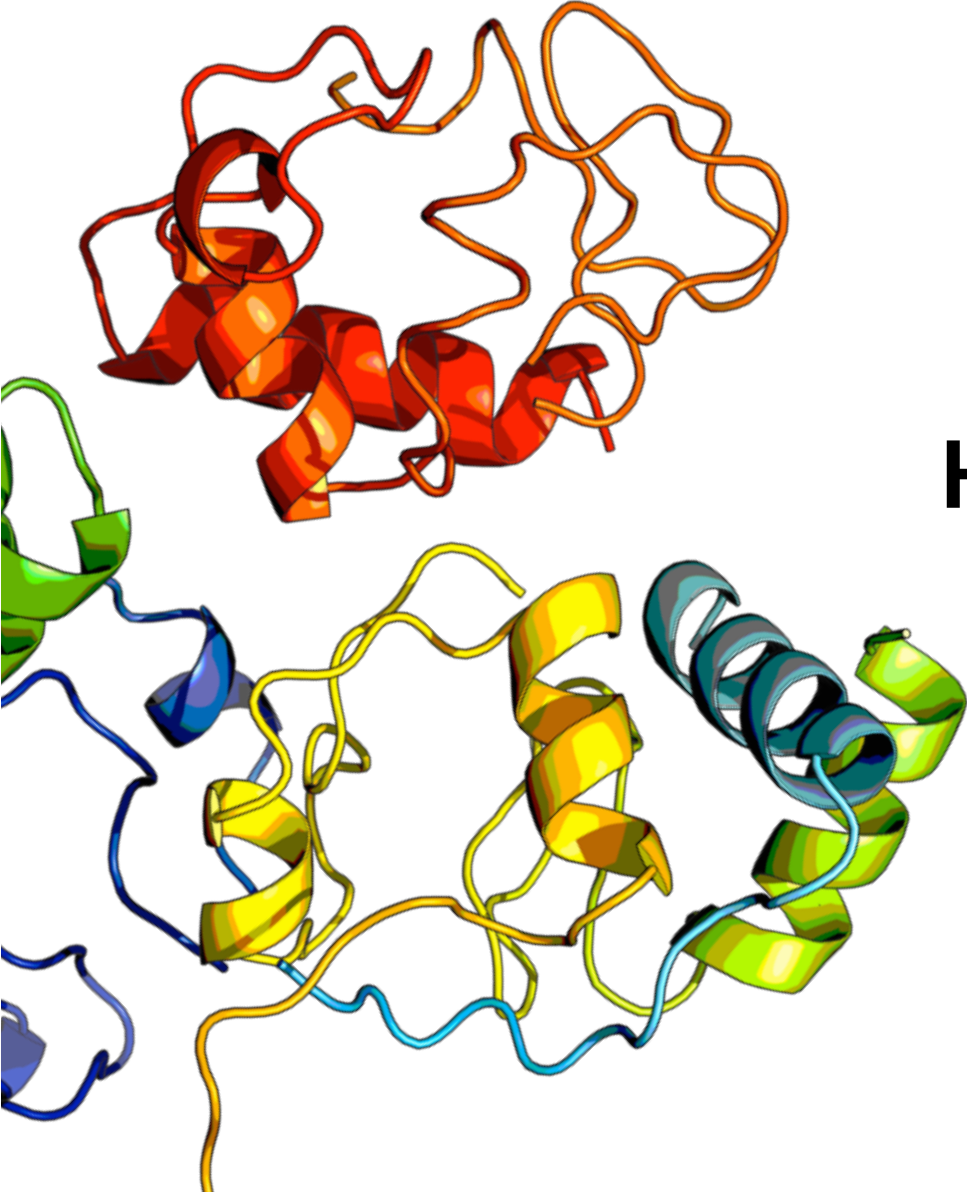
Check intermediate models

use keepjunk option

Use secondary structure modeling

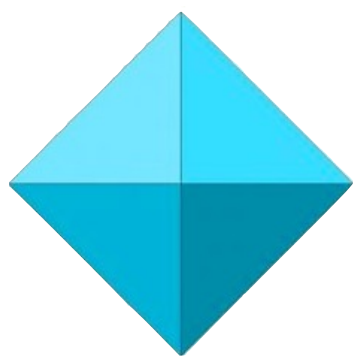
auto_albe.sh or quickfold in CCP4i





How to use ARP/wARP?

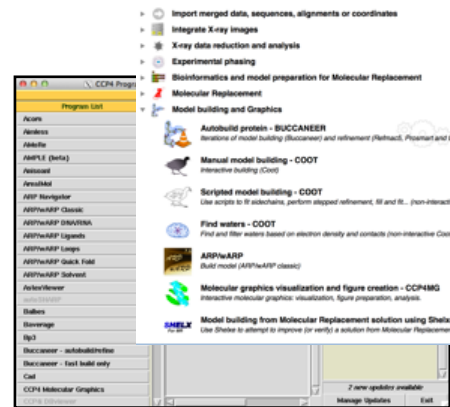
Where can I find ARP/wARP?



CCP4



Joint release - one installer, no worries!



Web Service

Complete
ARP/wARP
interface

Convenient for
parallel computing

Check user guide for the commands

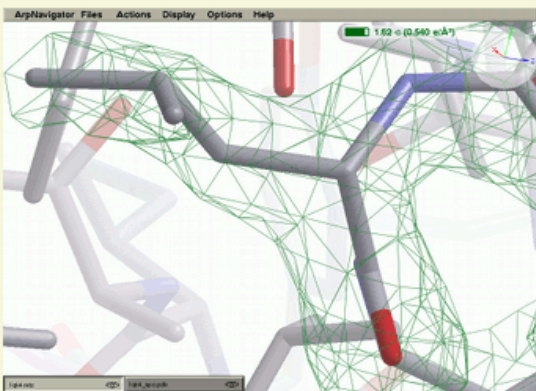
Where can I find ARP/wARP?

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



7.6
ARP/wARP

ARP/wARP Copyright 2016
by European Molecular Biology Laboratory



Click on the images above to zoom.

ARP/wARP: Crystallographic Macromolecular Model Building Version 7.6

Builds proteins, RNA/DNA, secondary structure, side chains, loops, solvent and ligands.

Download available for Mac and Linux.

Model building also available over the web.

[Download/License](#)

[Web Service](#)

DOCUMENTATION

[Download User Guide \(pdf\)](#)
[View User Guide \(html\)](#)
[Download Tutorial \(pdf\)](#)

MAILING LISTS

[Subscribe to ARP/wARP list](#)
[Unsubscribe from ARP/wARP list](#)
[Send message to ARP/wARP list](#)

HELP

[Send email to ARP/wARP developers](#)
[Frequently Asked Questions](#)

Useful Links

[Auto-Rickshaw](#) Automated Crystal Structure Determination Platform
[ViCi](#) In-Silico Ligand-Based Drug Design
[CCP4](#) Software for Macromolecular X-Ray Crystallography
[Refmac](#) Macromolecular Refinement Program

ARP/wARP web service

<https://arpwarp.embl-hamburg.de/>

NEW

ARP/wARP Web Service

Logged in as: gch

ARP/wARP APPS

P_{ph} P_{mr} P_{em}

P_{qf} N_{nt} S_{ol}

L_{ig} P_{dc} L_{vc}

P_{ar} P_{ba}

Dissemination Level: Software Developers ?

ARP/wARP

Welcome to the ARP/wARP webservice for Crystallographic Macromolecular Model Building. Builds proteins, RNA/DNA, secondary structure, side chains, loops, solvent and ligands.

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 publications.

MY RECENT JOBS

P _{qf}	00257_22f949	June 21, 2017, 10 a.m.	?
P _{mr}	00256_2ef3e1	June 21, 2017, 9:58 a.m.	?
N _{nt}	00255_81ffb2	June 21, 2017, 9:58 a.m.	?
S _{ol}	00254_be77a1	June 21, 2017, 9:51 a.m.	✓
S _{ol}	00234_e22118	June 20, 2017, 5:10 p.m.	✓

All my jobs ...

Please cite:
Langer G, Cohen SX, Lamzin VS, Perrakis A. **Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7.** (2008) Nat Protoc. 3, 1171-1179.

ARP/wARP team @ EMBL-Hamburg



Collaborators

MRC Laboratory, Cambridge :

Garib Murshudov's group

Rutherford Appleton

Laboratory: CCP4 team

NKI: Perrakis group

Previous Members

Ciaran Carolan, Tim Wiegels, Ioan Vancea,
Thomas Margraf, Joana Pereira

Funding

