

Grzegorz Chojnowski
Lamzin Group
EMBL Hamburg, Germany

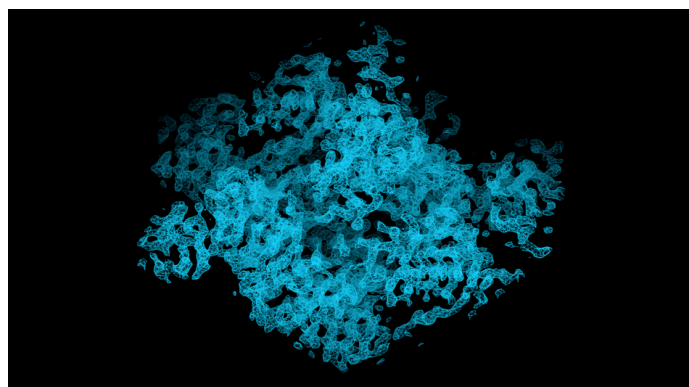
Automated Model Building with ARP/wARP



EMBL 40 YEARS | 1974-2014

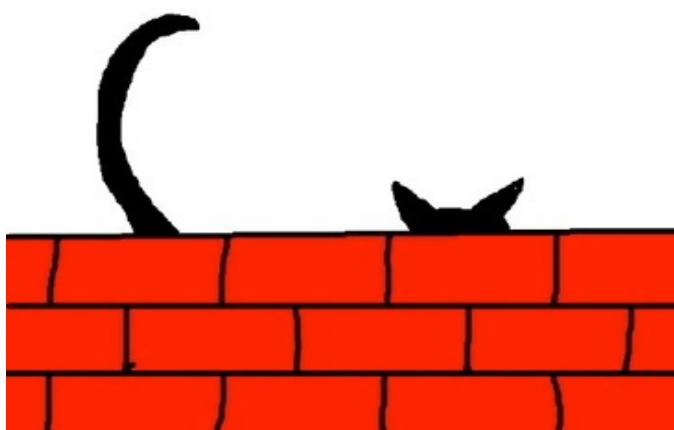


Electron density map



Grzegorz Chojnowski

Automation



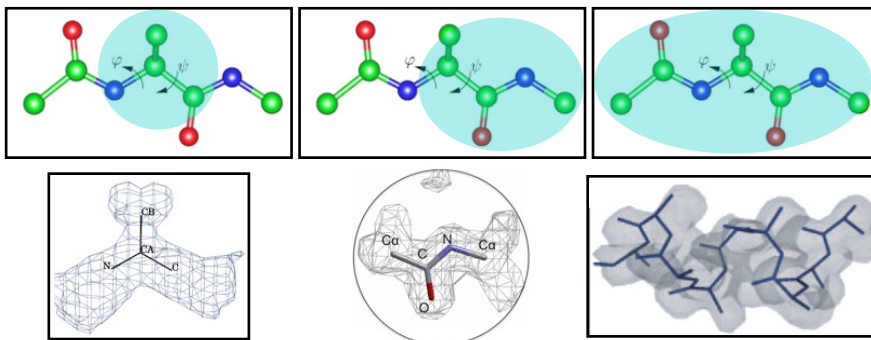
Grzegorz Chojnowski

24 January 2017



Methods for automated model building

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

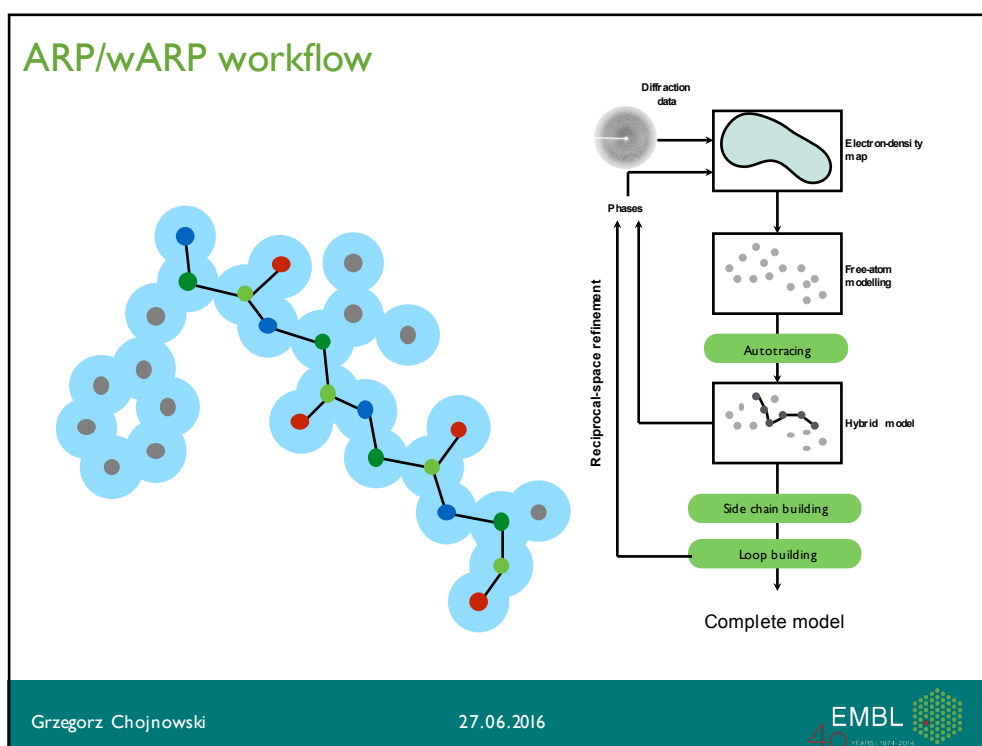
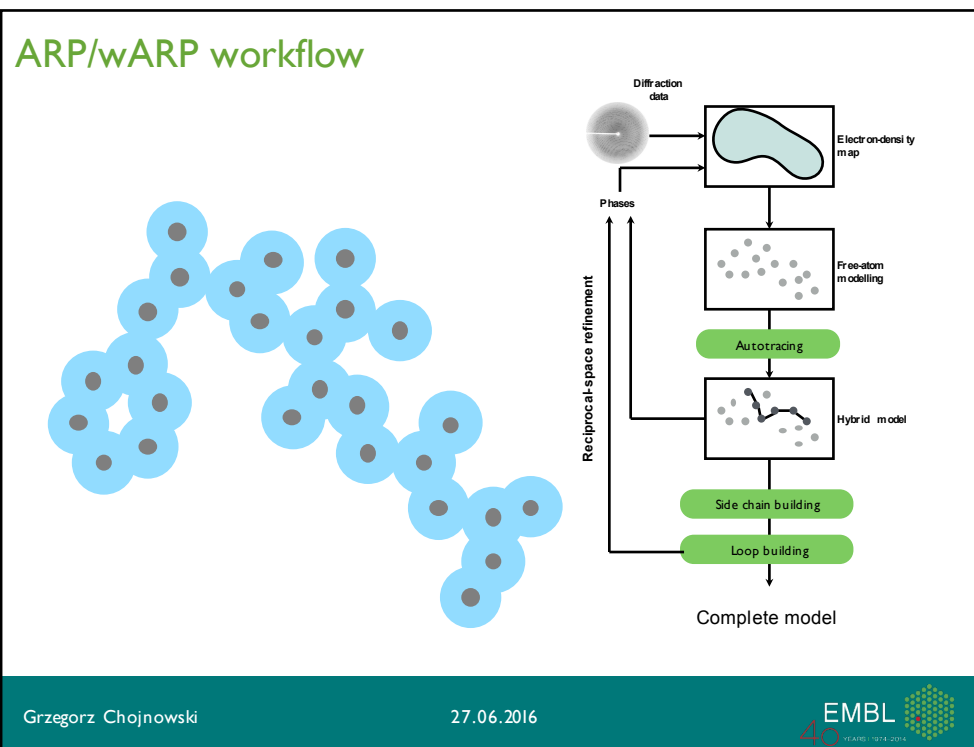


Buccaneer	ARP/wARP 	phenix.autobuild
C α	di-peptides (C α)	fragments

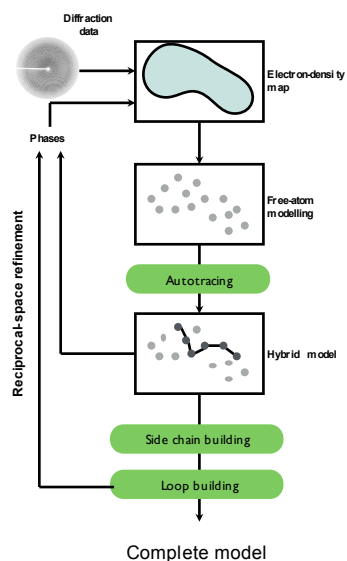
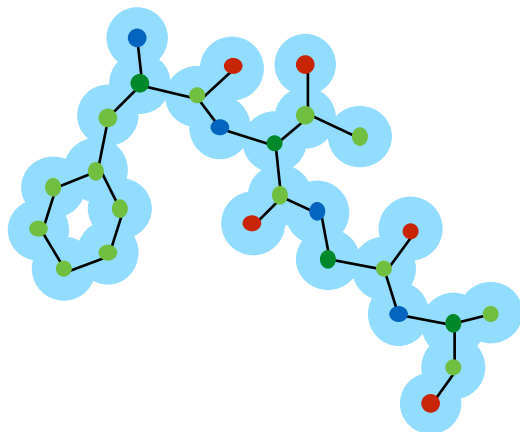
Grzegorz Chojnowski

24 January 2017





ARP/wARP workflow



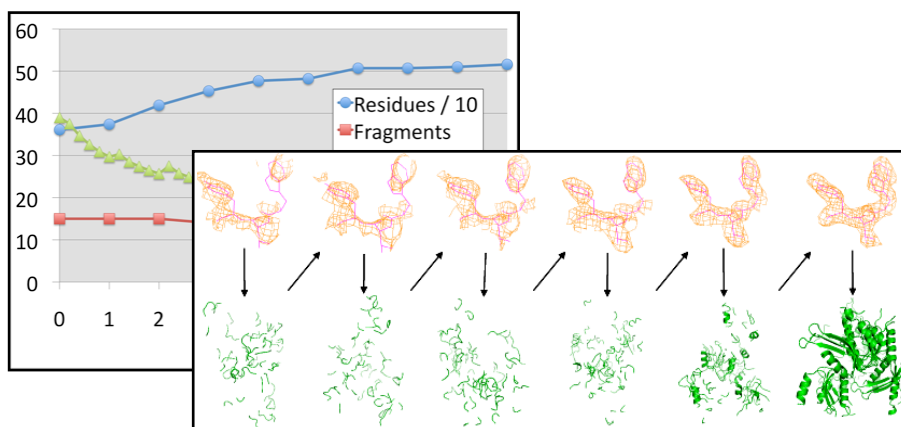
Grzegorz Chojnowski

24 January 2017



ARP/wARP workflow

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



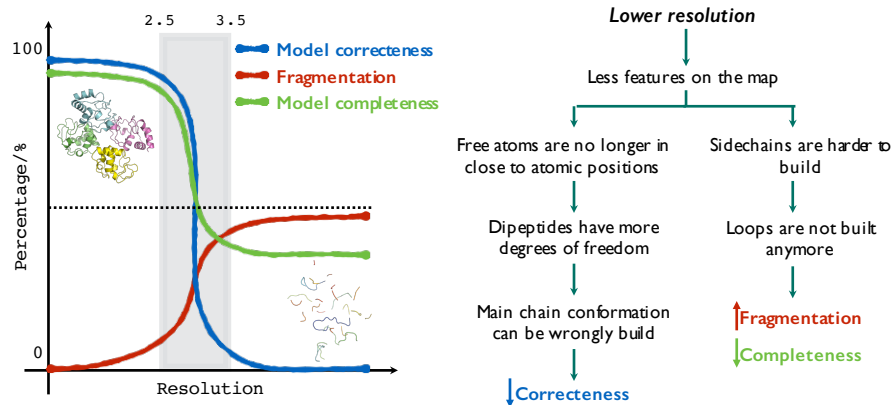
Grzegorz Chojnowski

24 January 2017



Low resolution as a limit

ARP/wARP methods work very well at high-resolution, but



Grzegorz Chojnowski

24 January 2017



ARP/wARP and resolution



Model building methods were historically developed for and work considerably better with high-resolution diffraction data:

Results from the ARP/wARP webservice (tracing performance)

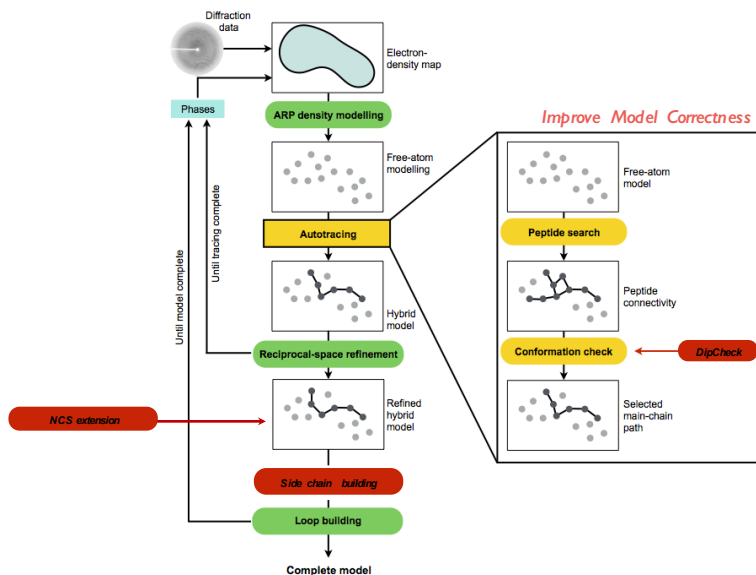
Resolution	Estimated fraction of automatically built protein structure
1.5 Å and higher	98% and higher
1.5-2.0 Å	91 - 98%
2.0-2.5 Å	81 - 91%
2.5-3.0 Å	67 - 81%
3.0-3.5 Å	51 - 67%

Protein model building at medium-to-low resolution in the best case generates incomplete and fragmented macromolecular models!

Grzegorz Chojnowski



Methods to improve model building at lower resolution



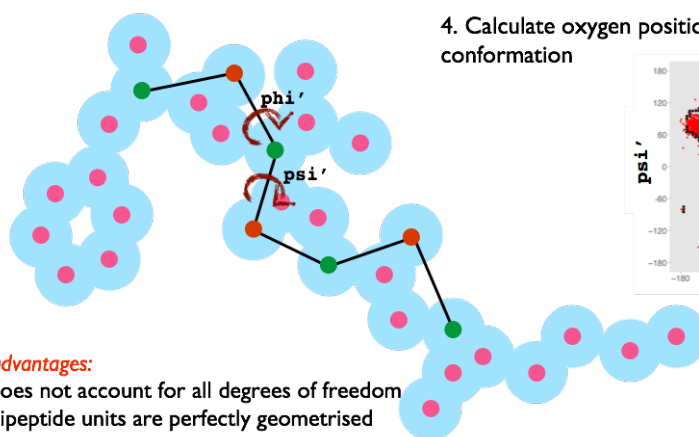
Grzegorz Chojnowski



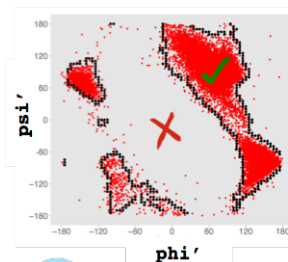
DipCheck

Pereira, J., & Lamzin, V.
in preparation

On the replacement of the Ramachandran-like plot...



4. Calculate oxygen position and check dipeptide conformation



Disadvantages:

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

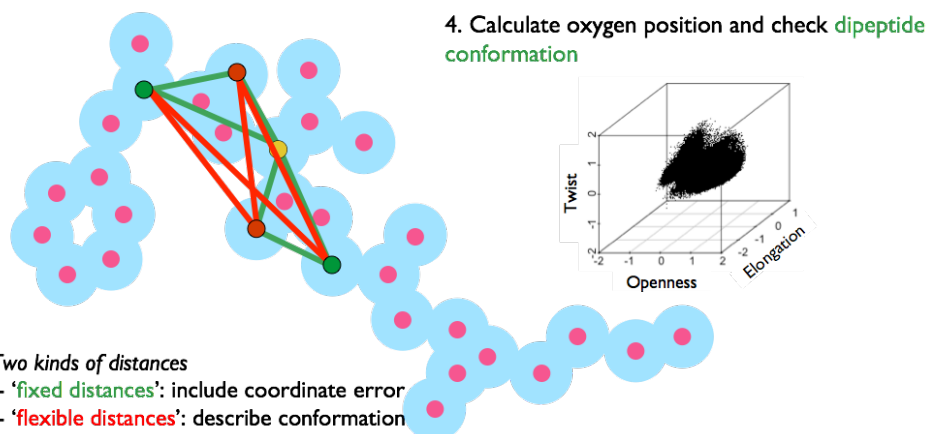
... as it is not suitable for lower resolution.

Grzegorz Chojnowski



DipCheck

... by applying a **distance geometry-based approach** to describe local C α geometry.



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

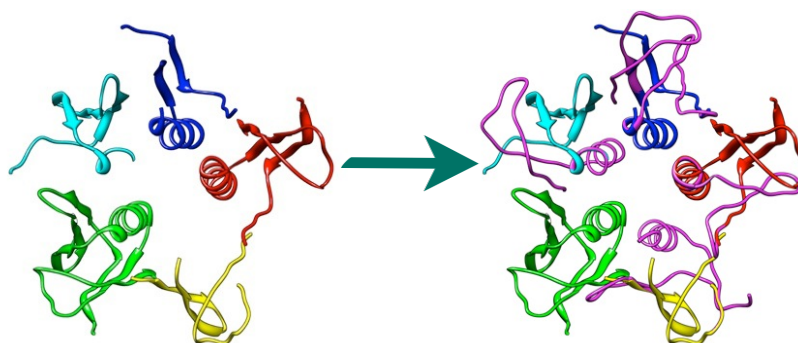
Grzegorz Chojnowski



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:

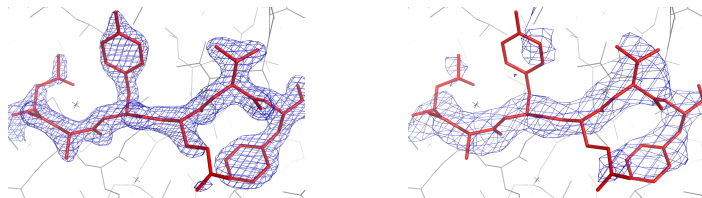
We can use this information to extend fragmented models!



Grzegorz Chojnowski



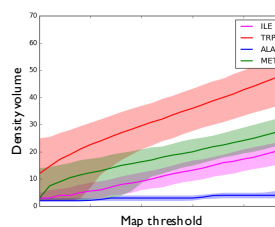
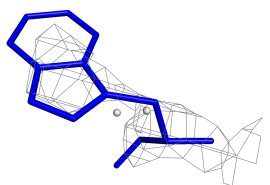
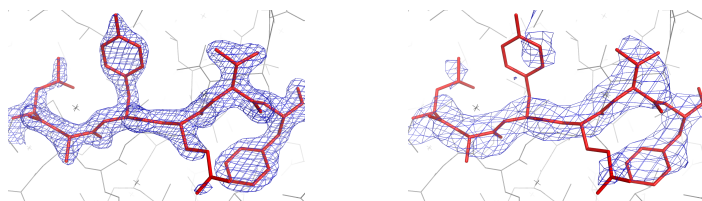
New side-chain building module (available soon)



Grzegorz Chojnowski

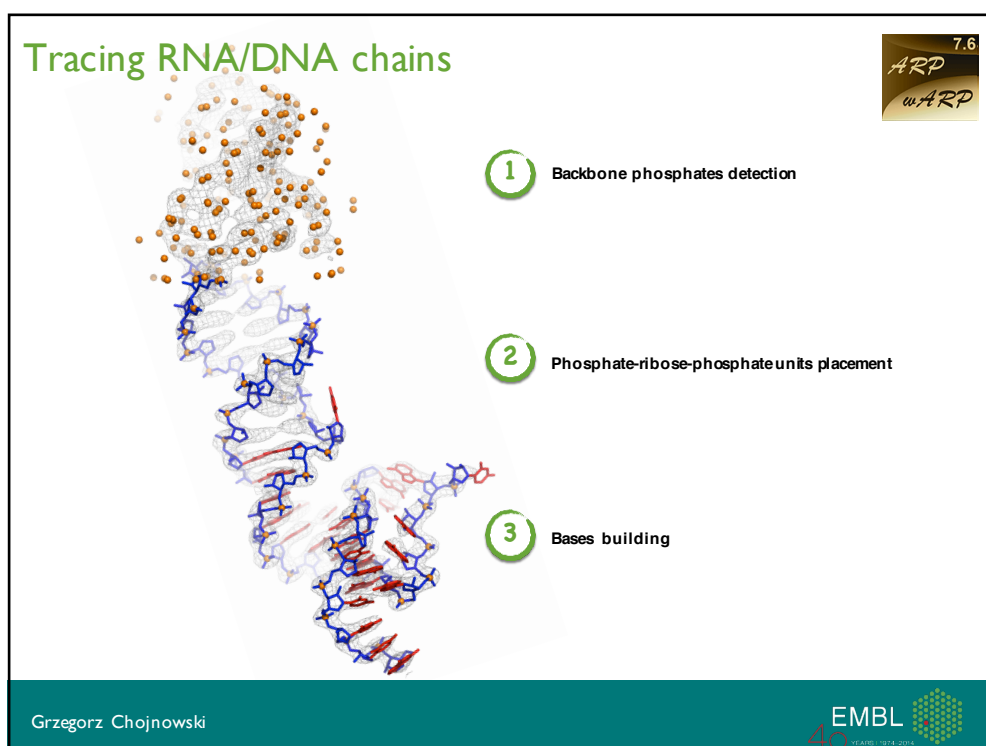
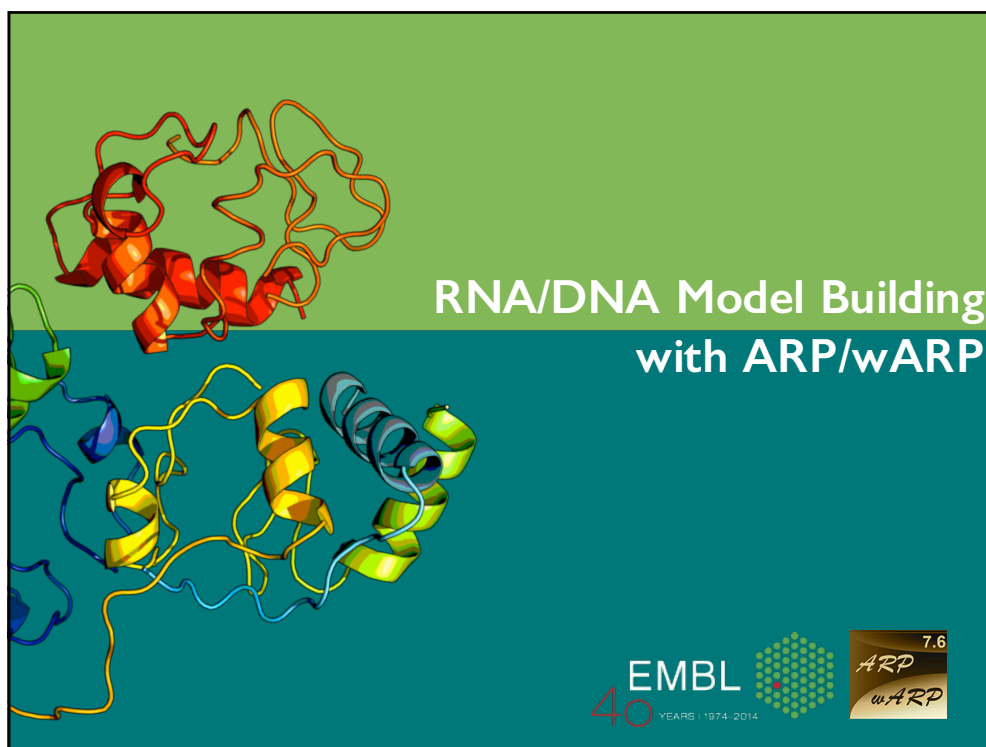


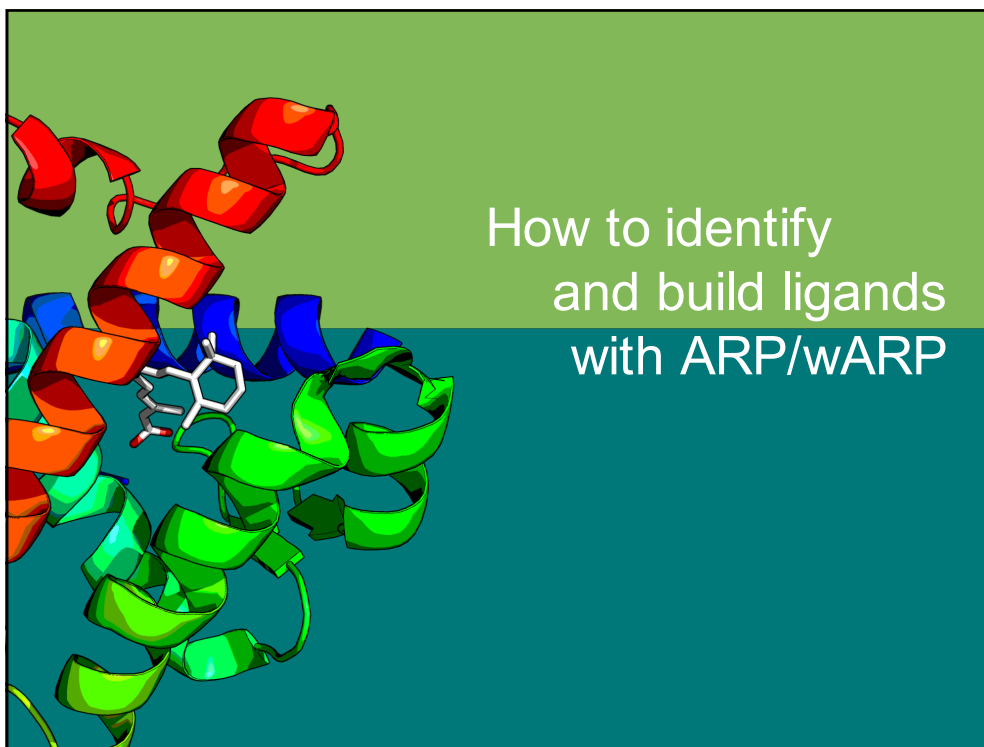
New side-chain building module (available soon)



Grzegorz Chojnowski



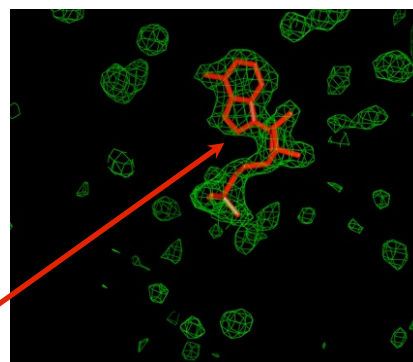
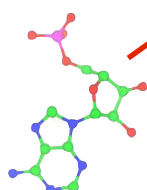




Fitting a ligand

We have:

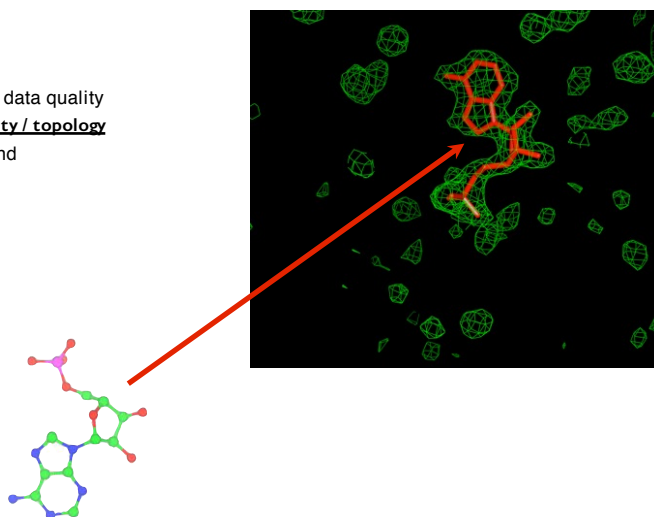
- electron density map (good resolution)
- protein model
- difference density map
 - shows excess density not explained by the protein model
 - to fit the appropriate ligand into



Fitting a ligand

Challenges:

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



Grzegorz Chojnowski



The protocol of a modelling task

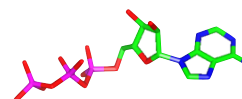
What we know at the beginning will determine the protocol:

		Binding site	
		Known	Unknown
Ligand identity	Known	X	
	Unknown		

1 ligand density



1 ligand



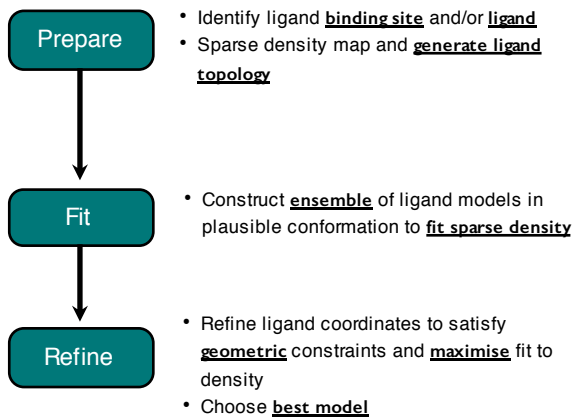
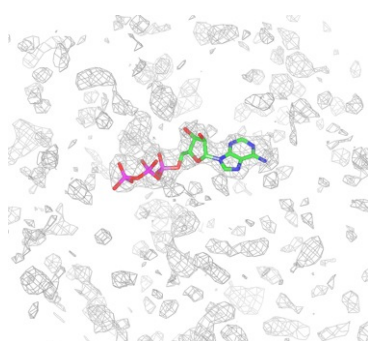
Grzegorz Chojnowski



Ligand fitting with ARP/wARP



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

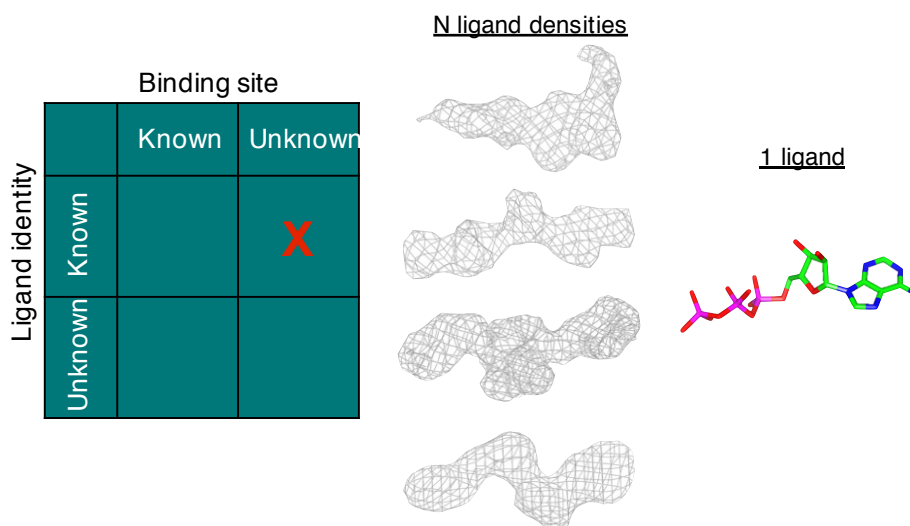


Grzegorz Chojnowski



The protocol of a modelling task

What we know at the beginning will determine the protocol:



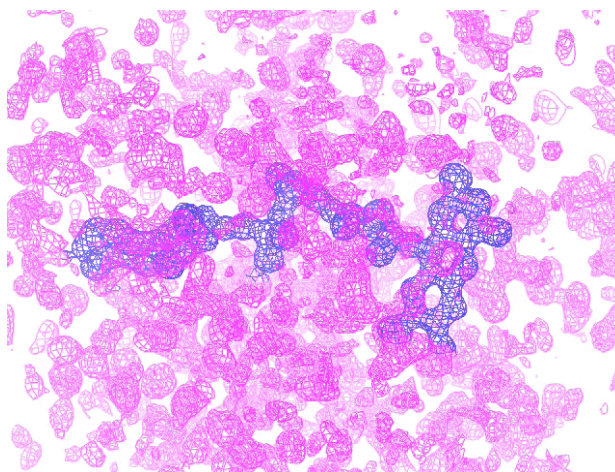
Grzegorz Chojnowski



Automatic binding site identification



- The difference density map from low to high contour threshold



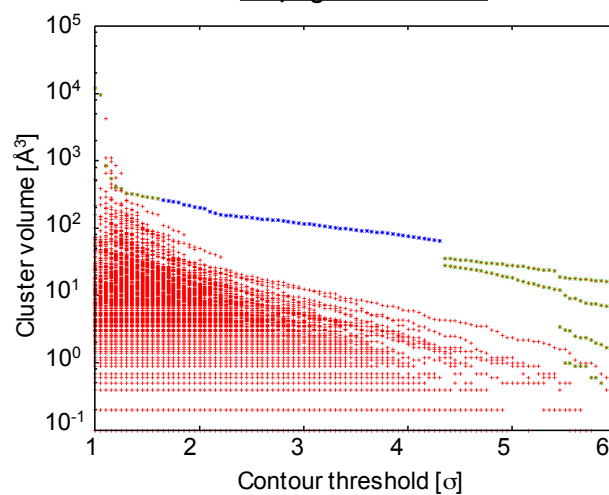
Grzegorz Chojnowski



Automatic binding site identification



- Capturing the dependence of the difference density map on changing the contour thresholds: the fragmentation tree



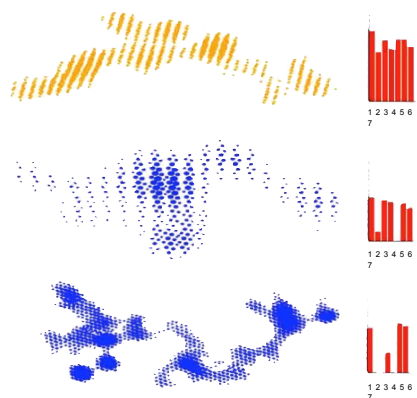
Grzegorz Chojnowski



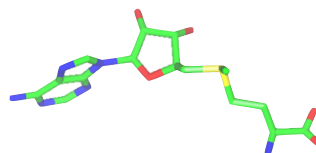
Shape features to measure similarity



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

Grzegorz Chojnowski



The protocol of a modelling task

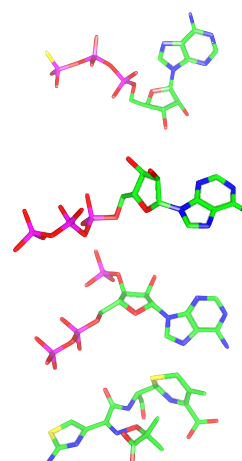
What we know at the beginning will determine the protocol:

		Binding site	
		Known	Unknown
Ligand identity	Known		
	Unknown	X	

1 ligand density



N ligand candidates

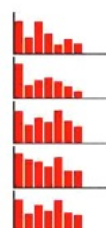
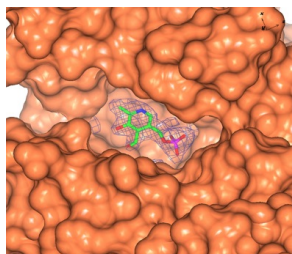


Grzegorz Chojnowski

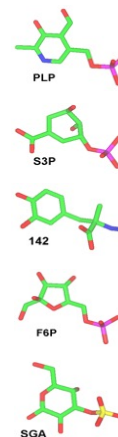


Shape features to measure similarity

- To assign a likelihood to a ligand knowing the binding site



Final ligand ranking



Grzegorz Chojnowski

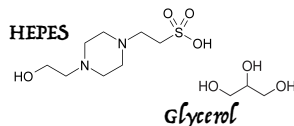


Which ligands can ARP/wARP identify?

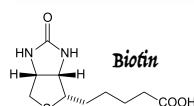
Ć. Carolan & V. Lamzin,
Acta D (July 2014)

- 84 of the most common ligands in crystallography

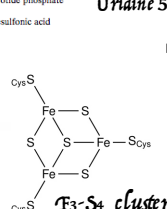
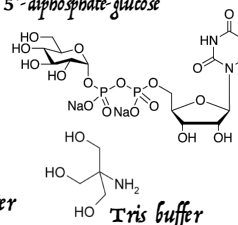
Ligand three-letter code	Ligand common name	Ligand three-letter code	Ligand common name	Ligand three-letter code	Ligand common name
01P	Darunavir	F3S	Fe ₃ -S ₄ cluster	PLM	Palmitic acid
1PE	Pentaerythritol glycol	FAD*	Flavin-adenine dinucleotide	PLP	Pyridoxal-5'-phosphate
2GP	Guanosine 2'-monophosphate	FMN*	Flavin mononucleotide	POP	Pyrophosphate
2PE	Nonathylene glycol	PPP	Fumaryl pyrophosphate	PYR	Pyruvic acid
SGP*	Guanosine 5'-monophosphate	GOL	Glycerol	RET	Retinal
A3P*	Adenosine 3',5'-diphosphate	GSH	Glutathione	SAM*	S-Adenosylmethionine
ACO*	Acetyl coenzyme A	H4B	5,6,7,8-Tetrahydrobiopterin	SF4	Iron-sulfur cluster
ADE	Adenine	HC4	para-Coumaric acid	SIA	O-Sialic acid
ADN	Adenosine	HEA*	Haem A	SO4	Sulfate ion
ADP	Adenosine 5'-diphosphate	HED	2-Hydroxyethyl disulfide	SPO	Spheroide
AKG	2-Oxoglutaric acid	HEM	Haem	STU*	Staurosporine
AMP	Adenosine monophosphate	IMD	Imidazole	TAM	Tris(hydroxyethyl)aminomethane
ATP*	Adenosine 5'-triphosphate	IPH	Phenol	THP	Thymidine 3',5'-diphosphate
B3P	2-[3-(2-Hydroxy-1,1-dihydroxymethyl-ethylamino)-propylamino]-2-hydroxymethyl-propane-1,3-diol	LDA	Lauryl dimethylamine-N-oxide	TLA	L-(+)-Tartaric acid
BCL	Bacteriochlorophyll A	MES	2-(N-Morpholino)ethanesulfonic acid	TPP	Thiamine diphosphate
BTB	Bis-tris buffer	MLI	Malonate ion	TRS	Tris buffer
BTN	Biotin	MLT	m-Maleine	TYD	Thymidine 5'-diphosphate
C2P*	Cyclic diguanosine monophosphate	MFD	(4S)-2-Methyl-2,4-pentanediol	U10	Coenzyme Q10
CAM	Camphor	MTE	Phosphonic acid mono-(2-amino-5,6-dimercapto-4-oxo-3,7,8,9,10,10A-hexahydro-4H-6-oxa-1,3,9,10-tetraaza-anthracen-7-yl)methyl ester	UPG	Uridine 5'-diphosphate-glucose
CDL	Cardiolipin	MYR	Myristic acid		
CHD	Cholic acid	NAD*	Nicotinamide		
CIT	Citric acid	NAP*	Nicotinamide adenine dinucleotide		
CLA	Chlorophyll A	NCO	Nicotinamide adenine dinucleotide phosphate		
CMP	Adenosine 3',5'-cyclic monophosphate	NHE	Cobalt hexamine(III)		
COA	Coenzyme A	OLA	2-(N-Cyclohexylamino)ethanesulfonic acid		
CSS	3-Cyclohexyl-1-propylsulfonic acid	ORO	Oleic acid		
CYC	Phycocyanobilin	PGG	Hexaethylene glycol		
DIO	1,4-Dithiolene dioxide	PEG	Di(hydroxyethyl)ether		
DTT	1,4-Dithiothreitol	PEP	Phosphoenolpyruvate		
EPE	HEPES	PG4	Tetraethylene glycol		
		PGA	2-Phosphoglycolic acid		
		PGO	3-1,2-Propanediol		
		PHQ	Benzyl chlorocarbonate		



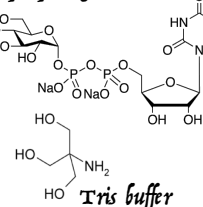
Glycerol



Biotin

Fe₃-S₄ cluster

Uridine 5'-diphosphate-glucose



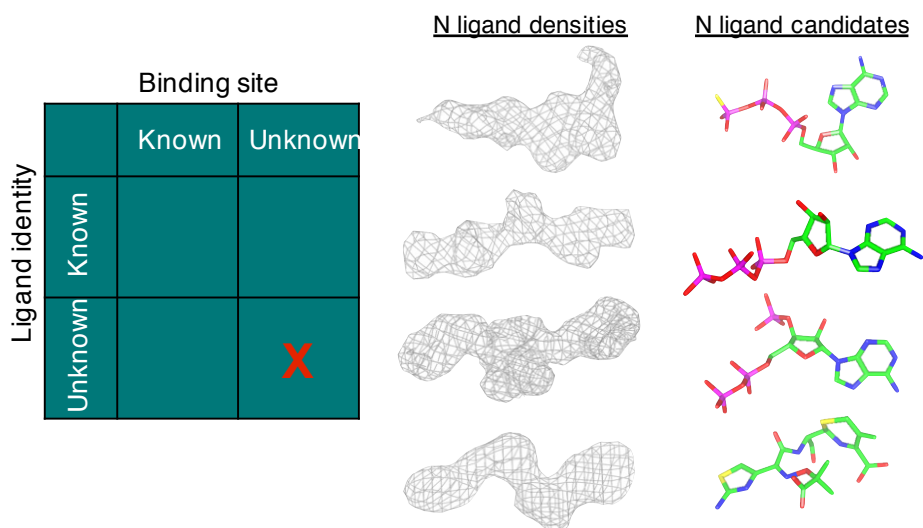
Tris buffer

Grzegorz Chojnowski



The protocol of a modelling task

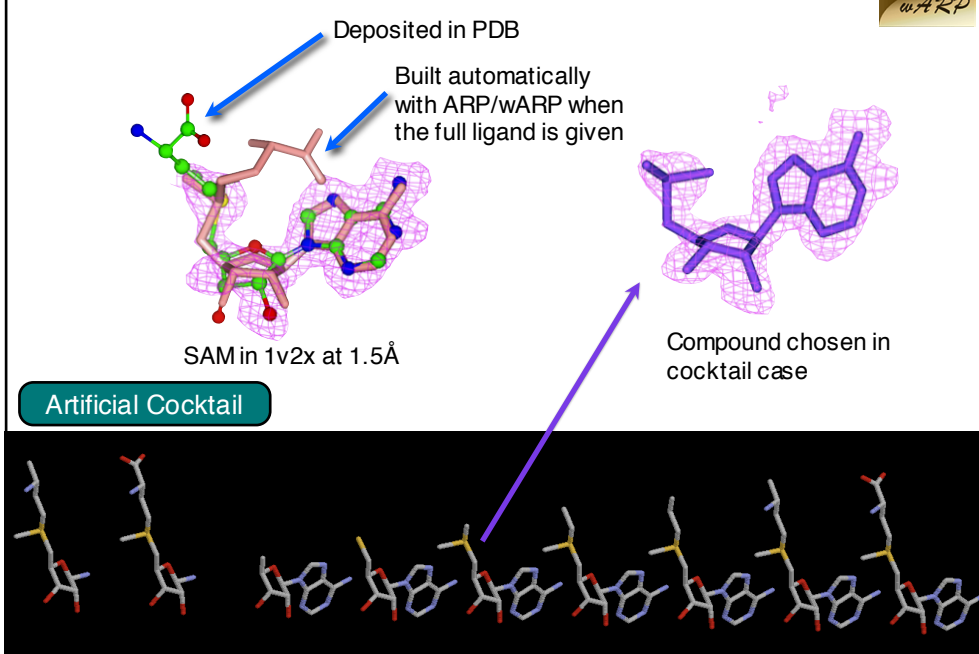
What we know at the beginning will determine the protocol:



Grzegorz Chojnowski



Partial Disorder, Partially Occupied Ligands



What ARP/wARP can do for you?



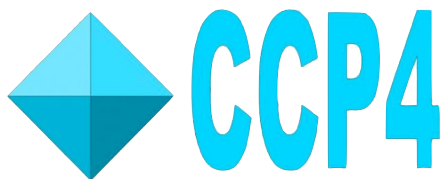
- Automated protein model building pipeline
- Protein main-chain tracing
- Secondary structure modelling
- Solvent modelling
- DNA/RNA model building
- Ligand fitting

Grzegorz Chojnowski

24 January 2017



Collaboration with CCP4



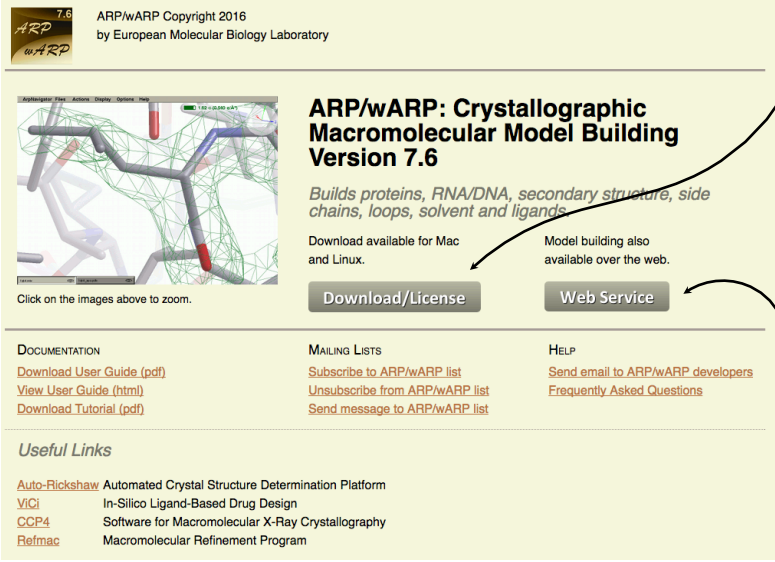
- Joint release
 - CCP4 7.0 and ARP/wARP 7.6
 - one installer, no worries!

Grzegorz Chojnowski



Where can I find it?

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



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

Only for Mac or Linux!!



Good for Windows users!!

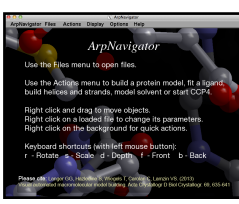


Grzegorz Chojnowski

EMBL 40 YEARS 1974-2014

How do I use ARP/wARP?

There are 4 main ways to use ARP/wARP:



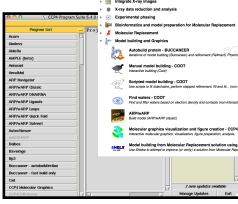
ArpNavigator

Intuitive interaction with automated model building

Real time visualisation of model building

Automated setup of computational tasks

Displays the intermediate steps



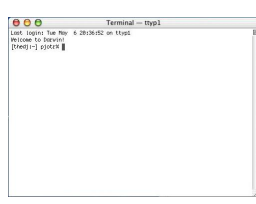
CCP4 Suite

Model building job saved in your project directory

Extra options for automated model building

Job submission to Hamburg cluster

Possibility to share data with developers



Terminal

Good for expert users

Complete control of the parameters to use

Check user guide for the commands



Web Service

Good for Windows users

Only for protein automated model building

Jobs don't run your computer but on our cluster in Hamburg

ARP/wARP team @ EMBL-Hamburg



Collaborators

EMBL Hamburg: Gleb Bourenkov,
Santosh Panjikar

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

West-Life

