Automated Protein Model Building with ARP/wARP version 7.3

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

EMBL’s mission
• **Basic Research** in Molecular Biology
• **Advanced Training** at all Levels
• **Development** of New Technologies and Instruments
• **Services** and **Facilities** for the Community
• **Technology Transfer** Discoveries

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

**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
Interprete This Map!
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

- Skeletonisation
- Secondary structure templates
- Free atoms / Chain tracing

Free atoms / Chain tracing
Model Building: Approaches

Free atoms / Chain tracing
**Methods for Building Protein Structure**

**Molecular-graphics packages:**
- FRODO - Jones (1978)
- O - Jones et al., (1991)

**Secondary structure:**
- an exhaustive search by ESSENS - Kleywegt and Jones, (1997)

**Skeletonisation:**
- Greer (1974)

**Chain tracing:**
- ARP/wARP – Perrakis et al., (1999)
- Concept of ‘free atoms’ – Agarwal and Isaacs (1985)

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

**Textual content list:**
- TEXTAL/CAPRA - Ioerger et al., (1999)
- MAID - Levitt, (2001)
- FSM - Pavelcik (2003)
- BUCCANEER - Cowtan (2006)
What is this “ARP/wARP”?

- 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**


Tim Wiegels 01.11.12
Methods for Building Protein Structure

<table>
<thead>
<tr>
<th>TEXTAL / Buccaneer</th>
<th>ARP/wARP</th>
<th>RESOLVE / ACMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cα</td>
<td>(Di-) Peptides</td>
<td>Fragments</td>
</tr>
</tbody>
</table>
• **Partial model** is used together with a **free-atom** model
• Chemically assigned parts provide **restraints** for refinement
• **Removal** and **addition** of atoms based on **density at atomics centers**
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

Real space

Diffraction space

Diffraction space

- Unrestrained refinement of free parts of the model
- Restrained refinement of chemically assigned atoms
The fundamental ARP/wARP concept

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

Real space

Diffraction space
The fundamental ARP/wARP concept

- **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

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

- 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
The task: Separating true peptides from false ones.

peptide good fit
peptide but poor fit
non-peptide but good fit
A peptide density shape is described by ~ **1,000** parameters, which are then **reduced to 1 parameter**.

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

**1.9Å resolution**

- Target density
- Weight function
- Similarity searched for
ARP/wARP Protein Model Building

Free Atoms

Pattern Recognition

Hybrid Models

Iterations

"Modern Classic Version"

Initial phases

Free atoms model

Initial model

Hybrid model

Reciprocal-space minimisation

Model update

Chain tracing

Sequence docking

Loop building
ARP/wARP Model Building

Free Atoms

Pattern Recognition

Hybrid Models

“Modern Classic Version”

Graph:
- Residues / 10
- Fragments
- R-Factor (%)

Building Cycles:
- 0 to 9

Images:
- Iterative building process
- Protein model progression

Tim Wiegels
Dependence on Data Resolution

3 Å (a) 4 Å (b) 5 Å (c)

6 Å (d) 8 Å (c)
## Dependence on Data Resolution

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

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Estimated fraction of automatically built protein structure (October 2011 – March 2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5 Å and higher</td>
<td>98% and higher</td>
</tr>
<tr>
<td>1.5 - 2.0 Å</td>
<td>91 – 98%</td>
</tr>
<tr>
<td>2.0 - 2.5 Å</td>
<td>81 - 91%</td>
</tr>
<tr>
<td>2.5 - 3.0 Å</td>
<td>67 - 81%</td>
</tr>
<tr>
<td>3.0 – 3.5 Å</td>
<td>51 - 67%</td>
</tr>
</tbody>
</table>
Ease of Access - ARPnavigator

This is the ArpNavigator

version 1.0
written by Gerrit G. Langer @ EMBL Hamburg, 2011
Ease of Access - 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
Ease of Access - ARPnavigator

Build helices and strands

- Helices and strands
- MTZ file input
- F: None
- PHI + FOM: None
- Limit the resolution
- Number of residues in the asymmetric unit: 0
- Keep all settings of this interface
- Show results in real time

Run now

r - Rotate  s - Scale  d - Depth  f - Front  b - Back
Ease of Access - ARPnavigator
Ease of Access - ARPNavigator
Ease of Access - ARPnavigator
Ease of Access - ARPnavigator
ARPnavigator - improved visualisation!
ARPnavigator - improved visualisation!
Modelling Secondary Structure

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

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

- Traces are **clustered**

- Assemblies are **averaged**
Building Helices with ARPNavigator

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

Tim Wiegels

01.11.12
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

335res 5chn

2.0Å avg: 70
• 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**

275res
19chn

3.0Å avg: 14

2.0Å avg: 70
335res
5chn
Automatic NCS Detection / Extension
Automatic NCS Detection / Extension

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

<table>
<thead>
<tr>
<th></th>
<th>Model Completeness</th>
<th>Residues / Fragment</th>
<th>Sequence Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top Results</td>
<td>+19.5%</td>
<td>+ 220%</td>
<td>+ 56%</td>
</tr>
<tr>
<td>Average in 7.3</td>
<td>+ 7.0%</td>
<td>+ 40%</td>
<td>+ 8%</td>
</tr>
</tbody>
</table>
Improvements by PNSextender

- 1.95 Å data
- Poor initial phases
- 42% completeness

Data courtesy of Florian Sauer

ARP/wARP
Improvements by PNSextender

- 1.95 Å data
- poor initial phases
- 75% completeness

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

TPDCVTGKVEYTKYND

TPDCVTGKVEYTKYND

VGDKELFTNRWNLQSLLL

TPDCVTGKVEYTKYND

TPDCVTGKVEYTKYND

DTFTVKVGDKELFTNRWNLQSLLL
Modelling loops with \textit{a priori} knowledge

Chain fragments are built and \textbf{docked to sequence}
- using \textbf{knowledge of angles} between anchors
- \textbf{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

```
TPDCVTGKVEYTKYNDD
```

```
VGDKELFTNRWNQLQSLLL
```

```
TPDCVTGKVEYTKYNDD
```

```
VGDKELFTNRWNQLQSLLL
```

```
TPDCVTGKVEYTKYNDD DTFTVK VGDKELFTNRWNQLQSLLL
```
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
When **no sequence** has been **docked** - **no loop building!**

- Identify loop regions
  - start / end
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
Docking results

seq  TPDCVTGKVEYTLYNDTFTVVGKELFTNWRNHLSLQAQITGIVIKTNACHNGTFSEVIFR

ss   CCCCCCCEEEEEEEECCCEEEEEEECCEECEEEEEECCCHHHHHHHHEEECCCCEEEEEE

sliding window  highest similarity

partially built protein fragment
FittOFF Workflow

Docking results

Distance statistics

Probabilities of various gap distances in Ångströms (Å)
FittOFF Workflow

Docking results

Distance statistics

Density check

25.09.12
FittOFF Workflow

Docking results

Distance statistics

Density check

FittOFF

Initial phases

Free atoms model

Initial model

Hybrid model

Reciprocal-space minimisation

Model update

Chain tracing

Sequence docking

Loop building

FRAGRA
FittOFF Workflow

Docking results

Density check

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

<table>
<thead>
<tr>
<th></th>
<th>Model Completeness</th>
<th>Residues / Fragment</th>
<th>Sequence Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top Results</td>
<td>+ 12%</td>
<td>+ 98%</td>
<td>+ 59%</td>
</tr>
<tr>
<td>Average</td>
<td>+ 4%</td>
<td>+ 25%</td>
<td>+ 7%</td>
</tr>
</tbody>
</table>
Tracing RNA/DNA Chains
Tracing RNA/DNA Chains

• 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)
Ligand Building Methods

Sparse grids

Conformational fit

Fine skeletons

All about ligands tomorrow from Ciaran!
Where Can I Get This???

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

!!! Tutorials !!!
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
- CCP4 Program Suite v6.3.0
- ARP/wARP v7.3
- Database for Balbes (Data Files)
- COOT v0.6.2
- CCP4 Molecular Graphics v2.6.0

Download size: 1172 MB
Download size: 292 MB
Download size: 339 MB

Click here for the MDS checksums of all files.

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

Collaborative Computational Project No. 4
Science and Technology Facilities Council
Rutherford Appleton Laboratory
Didcot, Oxon, OX11 0FA, United Kingdom
http://www CCP4.ac.uk

CCP4 Software Suite
Version 6.3
Setup

This program will guide you through the setup process and install CCP4 Software Suite on your computer.
Press "Next" button to continue.
...or...

CCP4 setup, step 1 out of 5

Choose components to install:

<table>
<thead>
<tr>
<th>Component</th>
<th>Full size</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCP4 Program Suite v6.3.0</td>
<td>1.2 GB</td>
</tr>
<tr>
<td>ARP/wARP v7.3</td>
<td>205.0 MB</td>
</tr>
<tr>
<td>COOT v0.6.2</td>
<td>350.0 MB</td>
</tr>
<tr>
<td>Balbes Database</td>
<td>3.0 GB</td>
</tr>
<tr>
<td>CCP4mg v2.6.0</td>
<td>920.0 MB</td>
</tr>
</tbody>
</table>

CCP4 core packages. Collection of over 200 useful programs, written by various authors, which provides tools for all stages of structure solution process, plus software libraries and accompanying data. Core packages are needed for other applications in the list.

Total 2 components selected, 1.4 GB
Tim Wiegels
01.11.12
Collaboration with CCP4

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

01.11.12
Towards structural systems biology

10^{26} atoms
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**
...if you are ever in town

Thank You for Listening!!!

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