



β α λ β ϵ σ

A Molecular Replacement Pipeline

Fei Long
MRC-LMC Cambridge
UK



MRC

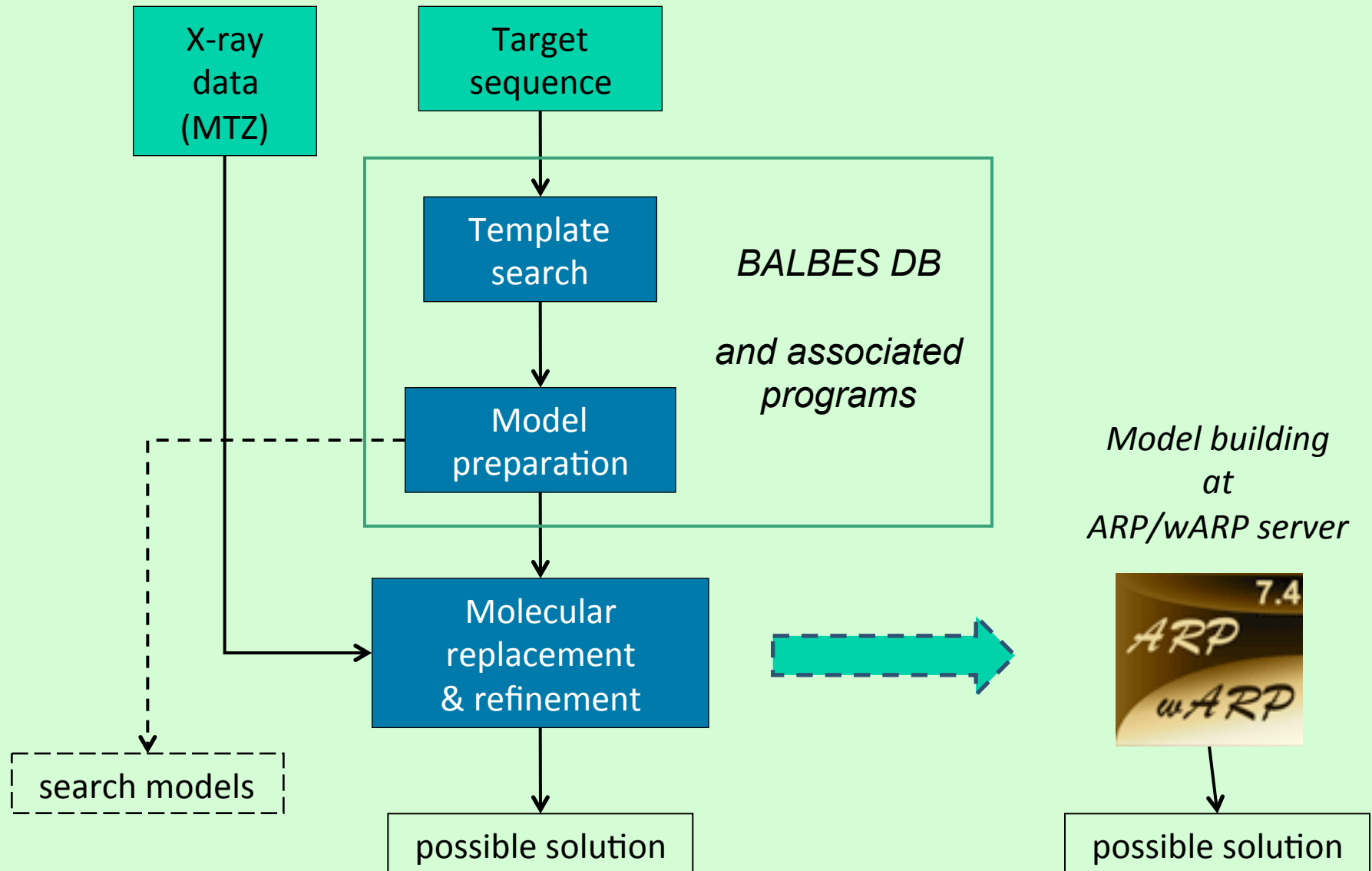
Laboratory of
Molecular Biology

BALBES: An automated system integrating all components required to solve protein structures via molecular replacement methods(MR).

- An internal database specifically designed for MR.
- Decision making protocols and algorithms for tasks at different stage during the structure-solving processes.
- The state of art engine programs such as MOLREP, REFMAC and others.

As an automated pipeline, BALBES tries to minimise users' intervention. The only thing a user needs to do is to provide two input files (a structure factor file and a sequence file)

What is BALBES



1. Run BALBES from CCP4 on-line web server

**Collaborative Computational Project No. 4**
Software for Macromolecular X-Ray Crystallography

Home (Logout) > Login > Programs > Balbes > New Balbes RunUsername: flong123

New Balbes Run

The file formats accepted for input are **mtz** and **cif** (structure factors) and **FASTA** (sequence target). **Note:** checking the ARP/wARP checkbox will send Balbes's results to the **ARP/wARP** server (it is assumed that you agree to the ARP/wARP academic license conditions)

Structure Factors:
Sequence Target:

No file selected.
 No file selected.

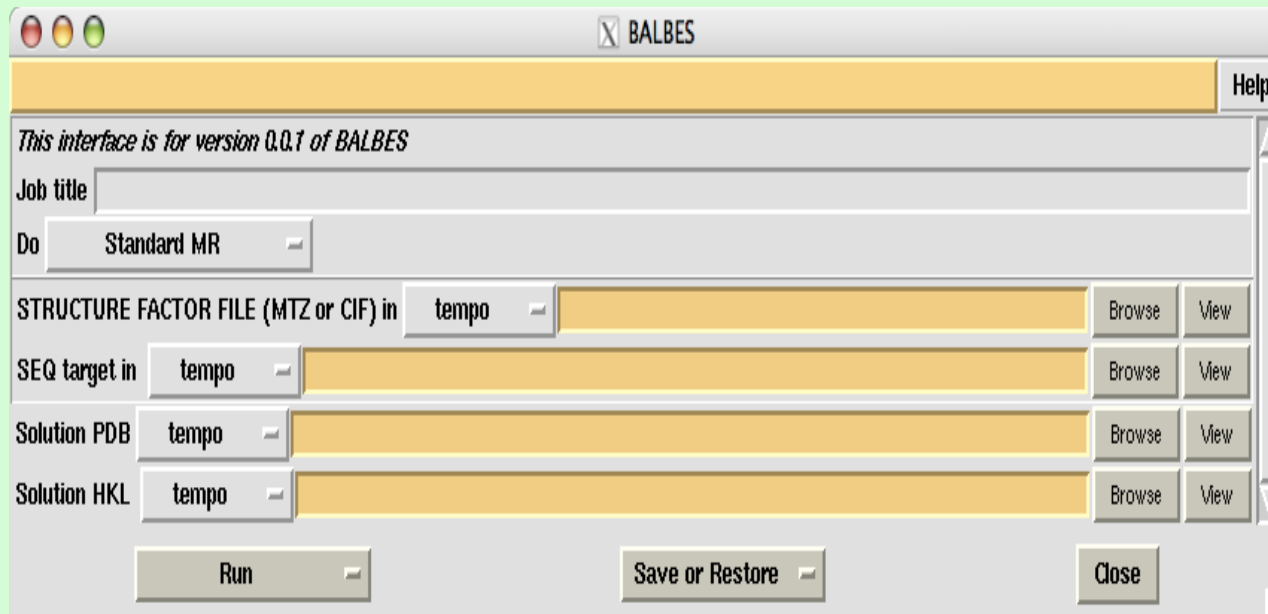
Instead of entering a Sequence Target file you can paste your **FASTA sequence below:**
(Note that a comment line beginning with a '>' character must precede each sequence)

☒ Check Full Spacegroup:
☒ Run ARP/wARP (on the Balbes solution):

Dissemination Level: Confidential

(after clicking submit, **PLEASE WAIT** for your files to upload - this may take some time)

2. Run BALBES using CCP4i if you have installed CCP4 suite at your local computers

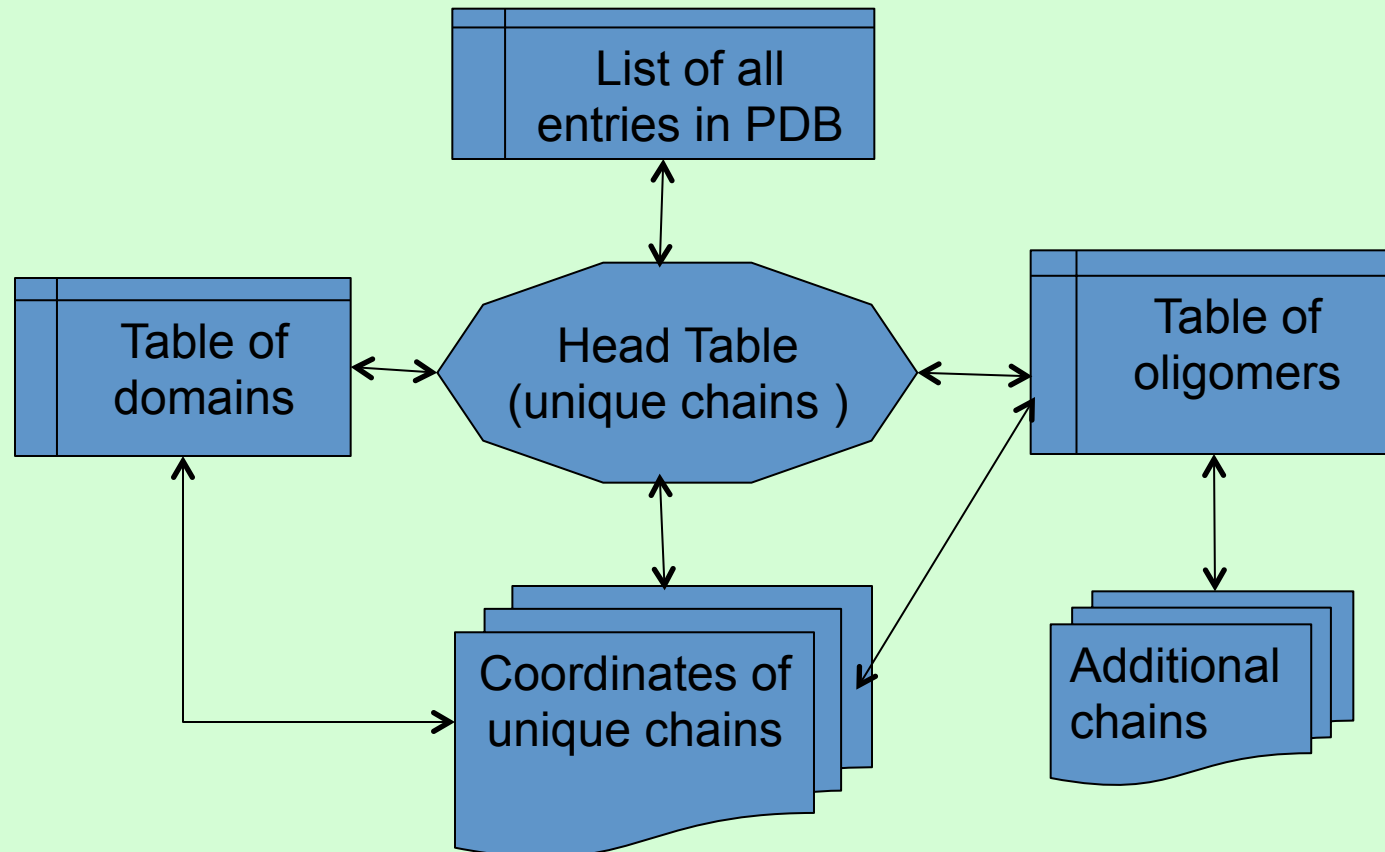


3. Run BALBES from command line if you have installed CCP4 suite at your local computers (for advanced users)

- More options
- Embedding BALBES as a component in other programs and pipelines.

```
balbes -f structure_factors_file  
      -s sequence_file  
      -o output_directory  
  
-f required  
-s required  
-o optional
```

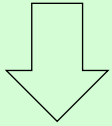
Organization of BALBES database



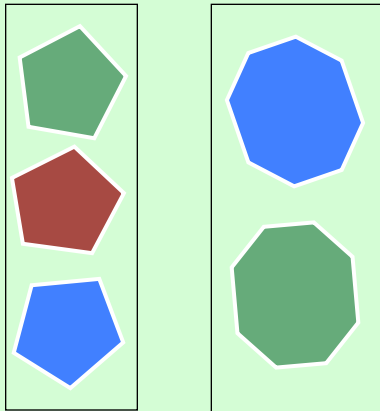
Some details of BALBES database

- About 30000 Non-redundant chains selected from protein structures the of high resolutions in PDB.
- More than 50000 domain definitions
 - ✓ defined by their three-dimensional compactness and separability from other parts
 - ✓ flexible parts remove.
 - ✓ hierarchically organized according to 3D similarity
- Multimers
 - generated using PISA definition

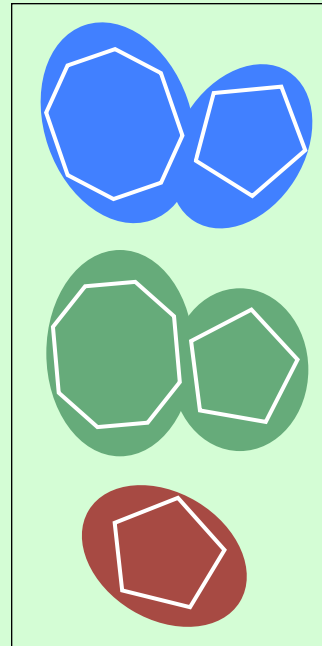
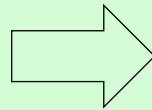
MARPVPEETVATERVKE



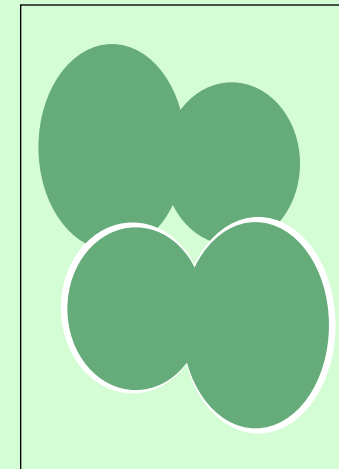
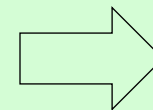
MARPVPEETVATERVKE



Domain models



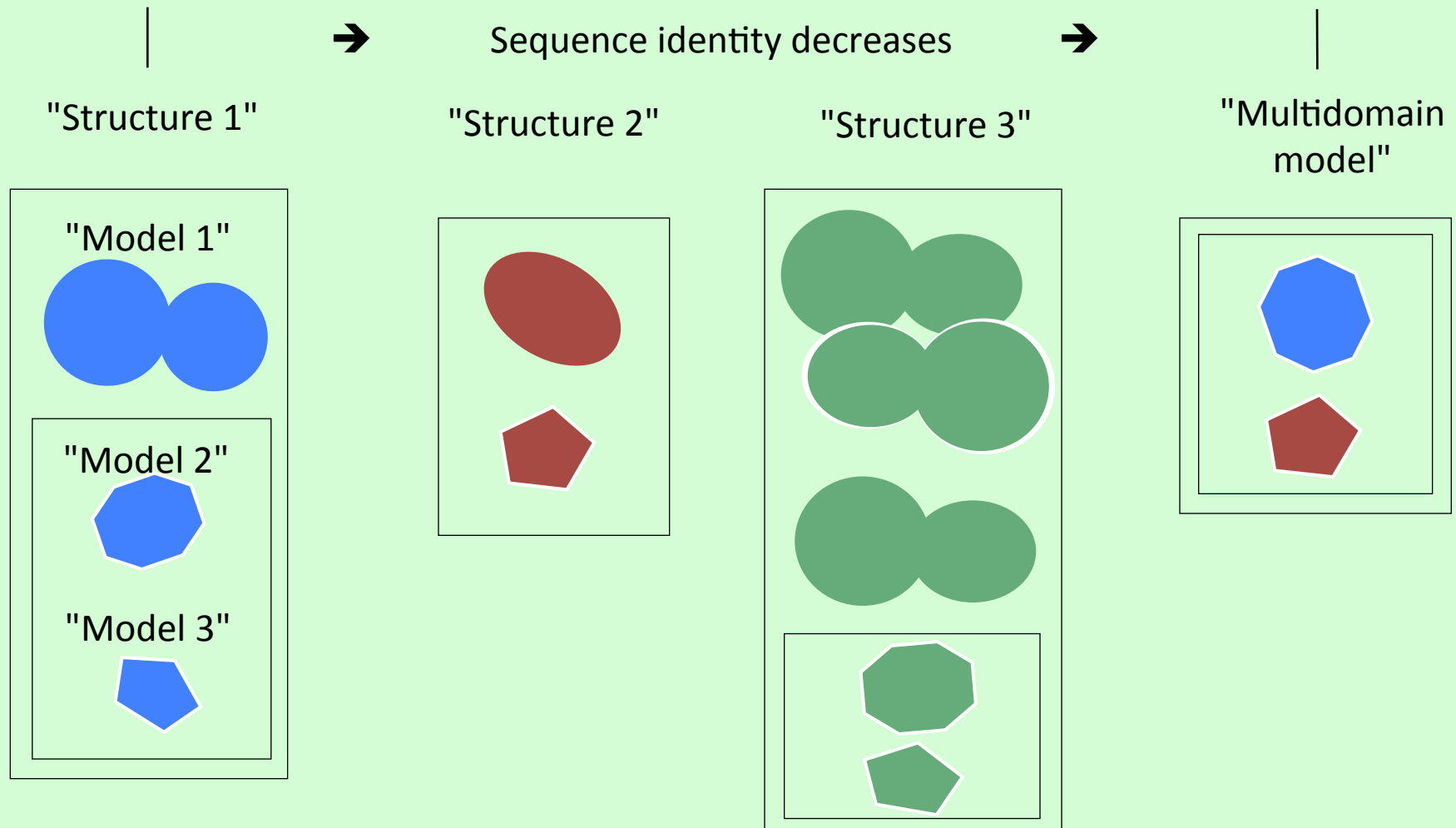
Chain models



Multimeric models

- All models are corrected
 - by sequence alignment
 - by accessible surface area
- Several input sequences mean protein-protein complex
 - hetero-multimeric models will be generated if possible

Homologue Models from one input sequence

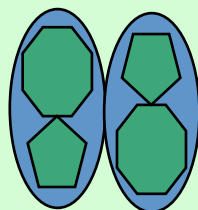


Homologue Models from one input sequence

Homologues

1p7q: homo-dimer;
each monomers is
formed by two domains.

Identity = 45%



1ufu: monomer
formed by two domains.

Identity = 45%



2d3v: monomer
formed by two domains.

Identity = 46%



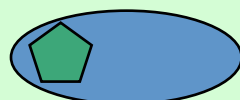
xxxx: contains
domain 1

Identity = 42%

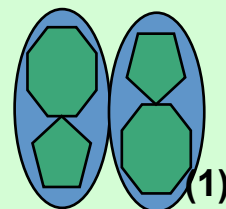


yyyy: contains
domain 2

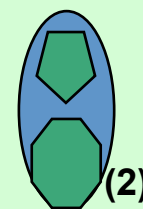
Identity = 56%



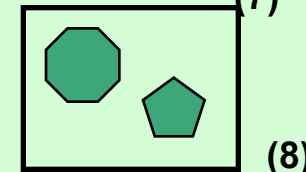
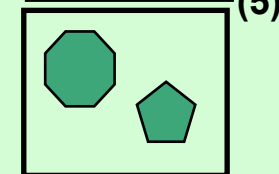
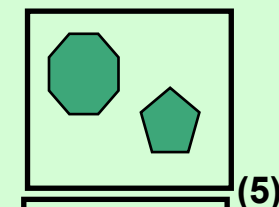
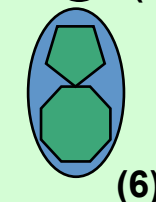
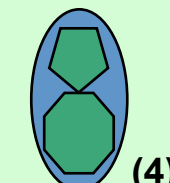
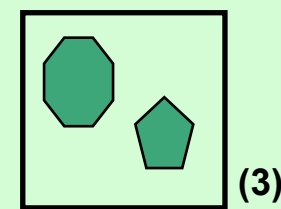
dimeric



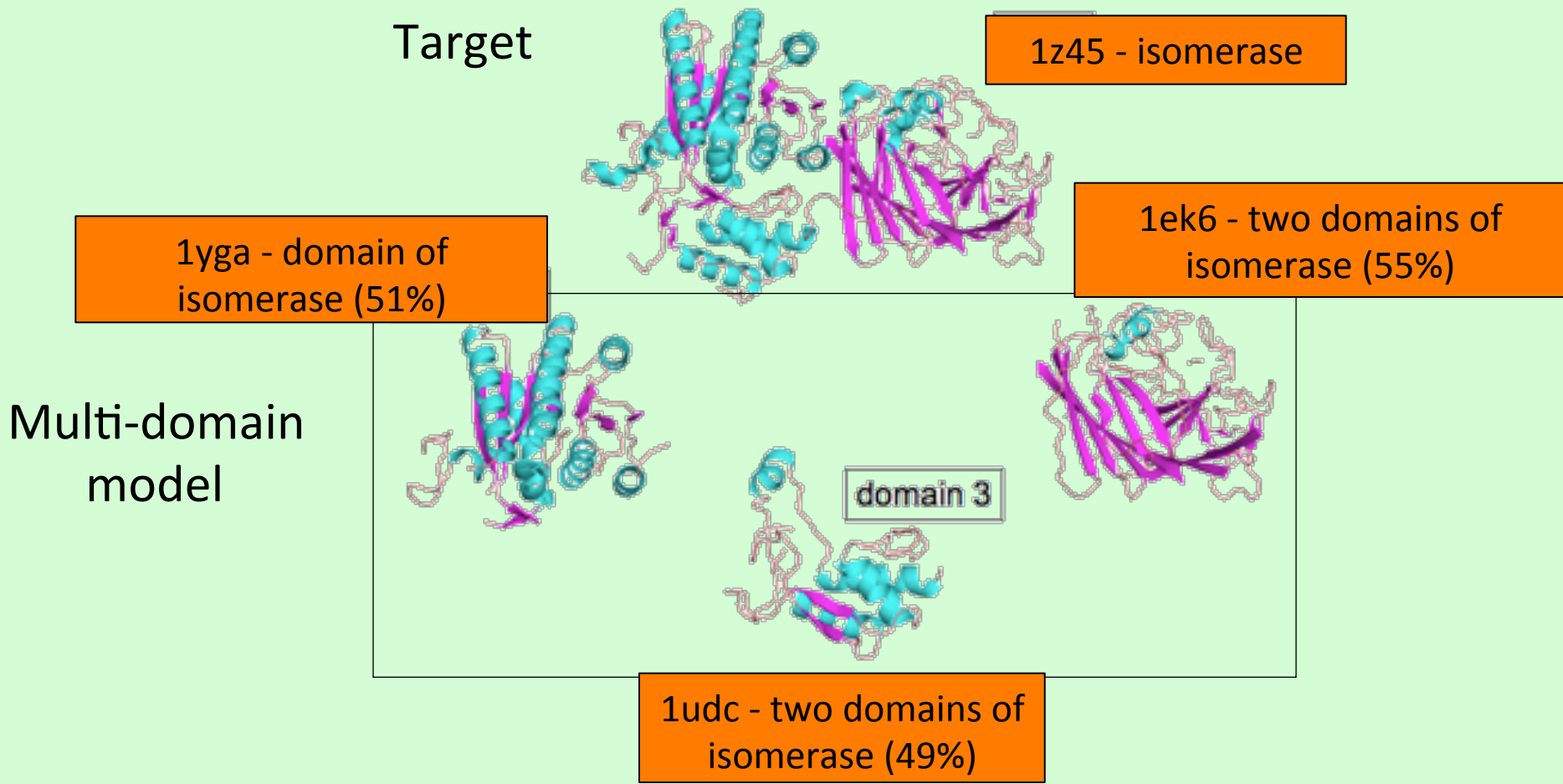
monomeric



“multi-domain”



“Multi-domain” models:
placing domains one by one and
attempting to maintain proper
composition of the asymmetric unit

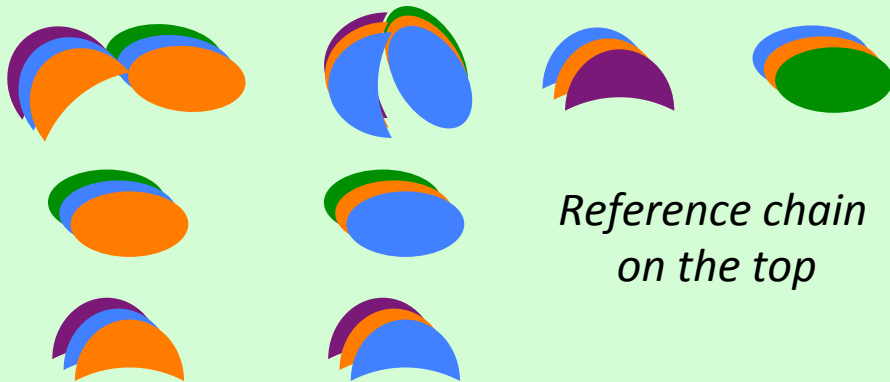


Ensemble models are generated if possible

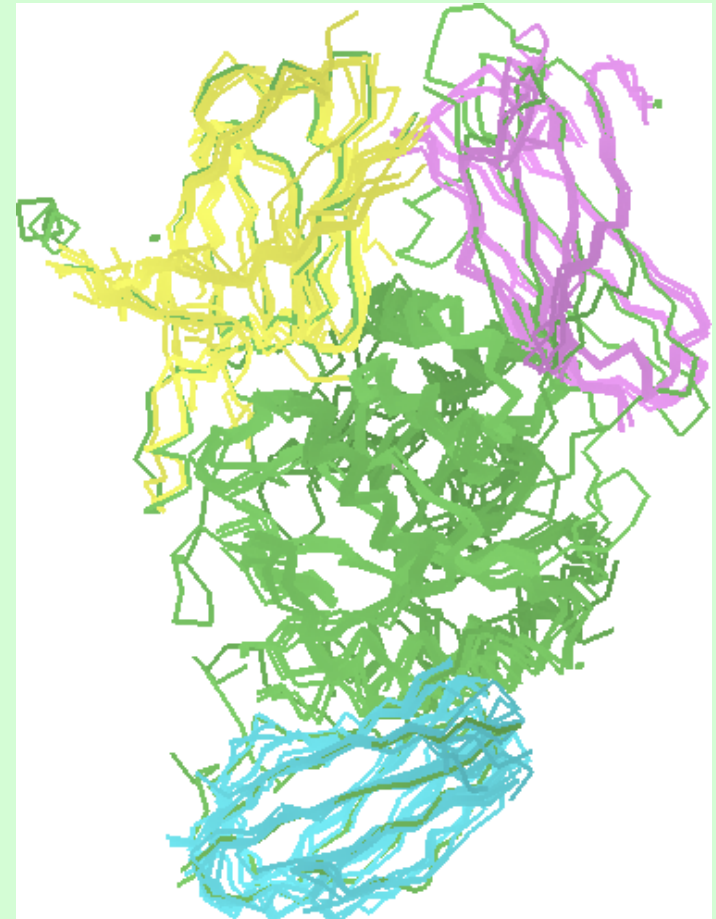
Homologues from the Balbes database:



Ensemble search models:



*Reference chain
on the top*



Ensemble models from Balbes



THE UNIVERSITY of York

York Structural Biology Laboratory

University | Chemistry | YSBL

Home (Logout) > Login > Programs > Balbes > View Results

Username: alezia

View Results

PROCESS 7587375708 IS RUNNING [stop process]

PLEASE CLICK HERE TO REFRESH THIS PAGE (or use the REFRESH PAGE link at the bottom)

FILES

output/

process_details/

template_solutions/

results/

template_models/

CURRENT FILE: summary.log [download this file]

BALBES Version is 1.1.1_DB_Oct_1_2010

process_details/

✓ process_details/
sfcheck_gfobs.log
sfcheck.log
sfcheck.btc
sfcheck.log

mod_2b1qA_mono2.pdb
mod_2b1qA_mono2_ens.pdb
mod_2b1qA_1dom.pdb
mod_2b1qA_tmp.pdb
mod_2b1qA_1dom_ens.pdb
mod_2b1qA_2dom.pdb
mod_2b1qA_2dom_ens.pdb
mod_2b1rA_mono.pdb
mod_2b1rA_mono1.pdb

For the structure 1 in sequence 1

PDB_CODE		2b1qA				
NO. of MODEL		3				
Model	Chain ID	Similarity	Residues	Multimer?	Domain?	Monomers(expected)
1	A	0.857(ENS)	244	Monomer	No	1
2	A	0.846(ENS)	162	Monomer	Yes	1
3	A	0.861(ENS)	72	Monomer	Yes	1

Similarity

0.857(ENS)

0.846(ENS)

0.861(ENS)

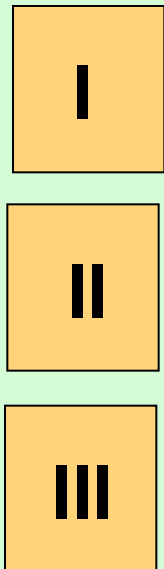
14

14

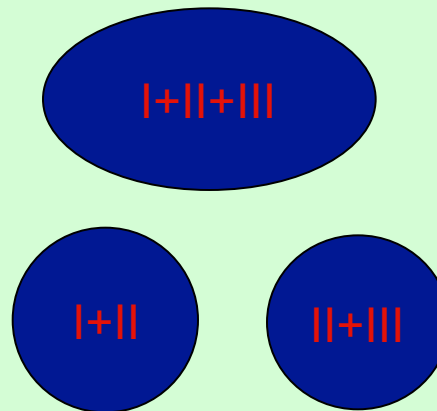
Complex Models from multiple input sequences

BALBES will search the database for complexes of models containing all or some of the sequences if a user provides more than one sequences

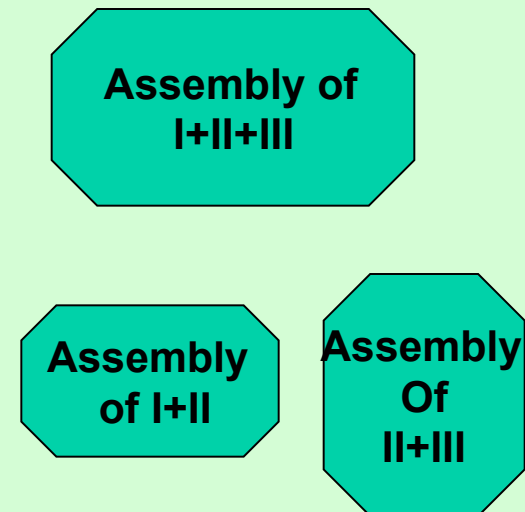
User's sequences



DB



Complex Models

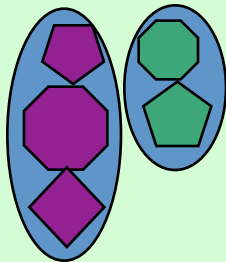


Complex Models from two input sequences

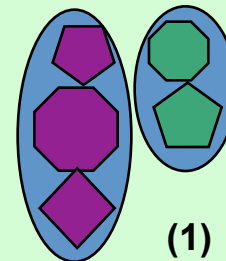
Homologues structure:

Derived search models (and their priority):

2b3t: hetero-dimer;
monomers are formed by
two and three domains.



assembly



(1)

Other homologues (1t43, 1nv8, 1zbt, 1rq0) are matching only one of two sequences. Priority rules applied to them are as in previous examples.

Note: If the system cannot find a good solution from assembly then it tries to solve using individual models and combine them. Individual models may come from different proteins.

An example

THREE SEQUENCES ARE SUBMITTED

Sequence 1: 199 residues ,sequence 2: 241 residues, sequence 3: 215 residues

Assembly model 1: 2vlmD + 2vlmE

Assembly model 2: 2eyrB + 2eyrA

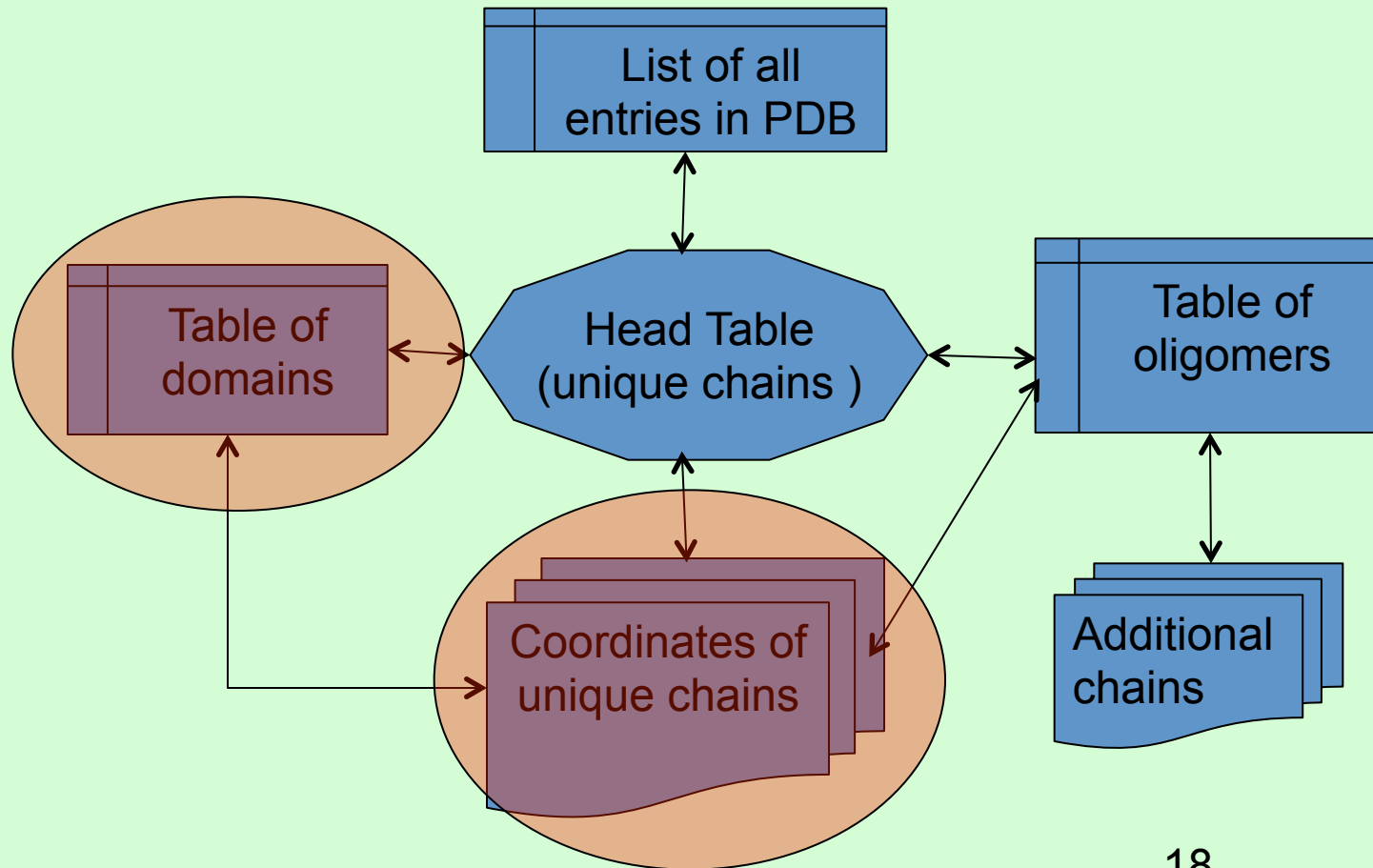
Sequence 1--> models: 2vlmD, **1ogaD**, 2cdfA, 3mv7D, 2wbjC

Sequence 2 --> models: **1ogaE**, 2axhA, 3kxfP, 2po6D, 2eyrB

Sequence 3 --> models: **1enfA**, 1esfA, 1fnuA, 2g9hD, 3ea6A

Final Results	
Models	1ogaE+1enfA+1ogaD
R_ini/R_fin	0.388/0.250
Rfree_ini/Rfree_fin	0.404/0.310

Re-classification of BALBES database



BALBES Development:

Alexey Vagin

University of York

Garib Murshudov

MRC LMB

Web server support:

Ville Uski STFC RAL

Ronan Keegan STFC RAL

Charles Ballard STFC RAL

PISA:

Eugene Krissinel STFC RAL

ARP/wARP development team