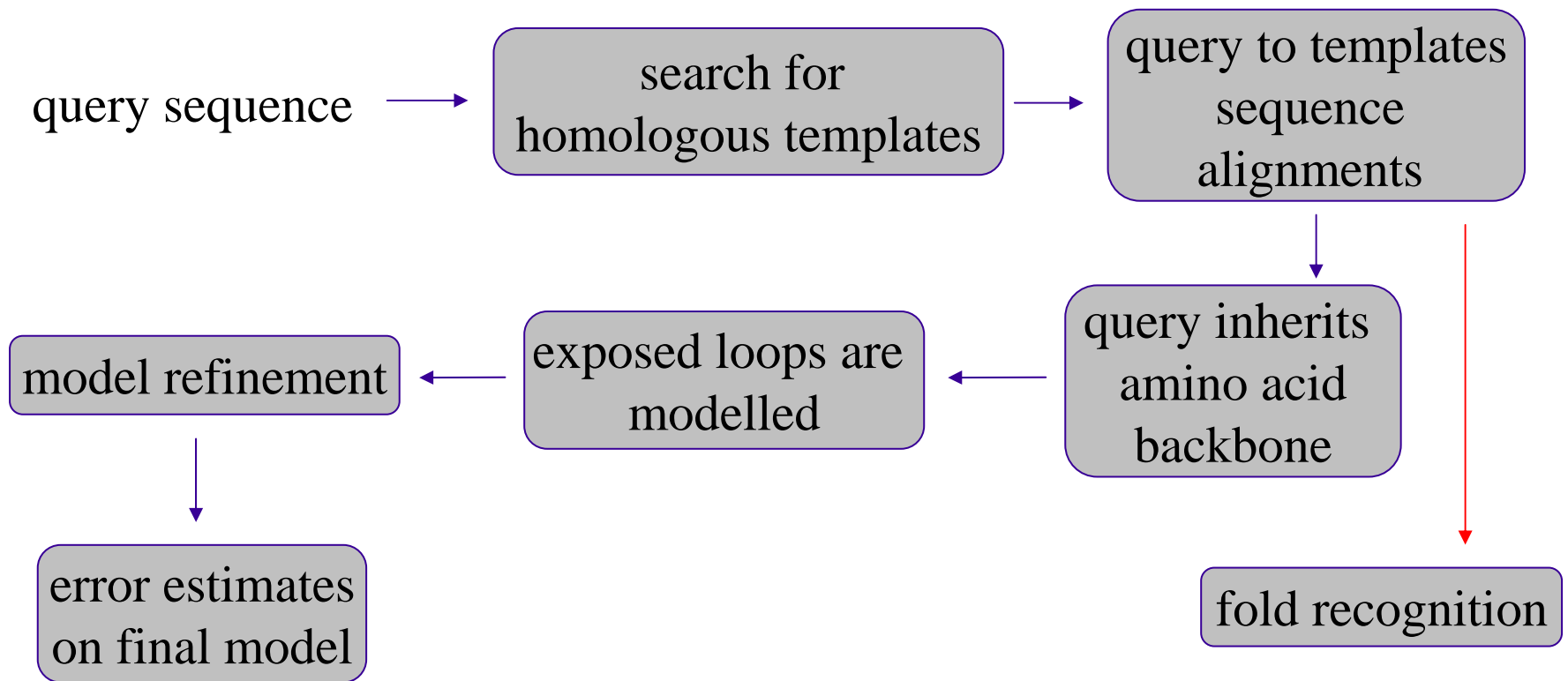


Domain Fishing and 3D-JIGSAW: tools for protein comparative modelling

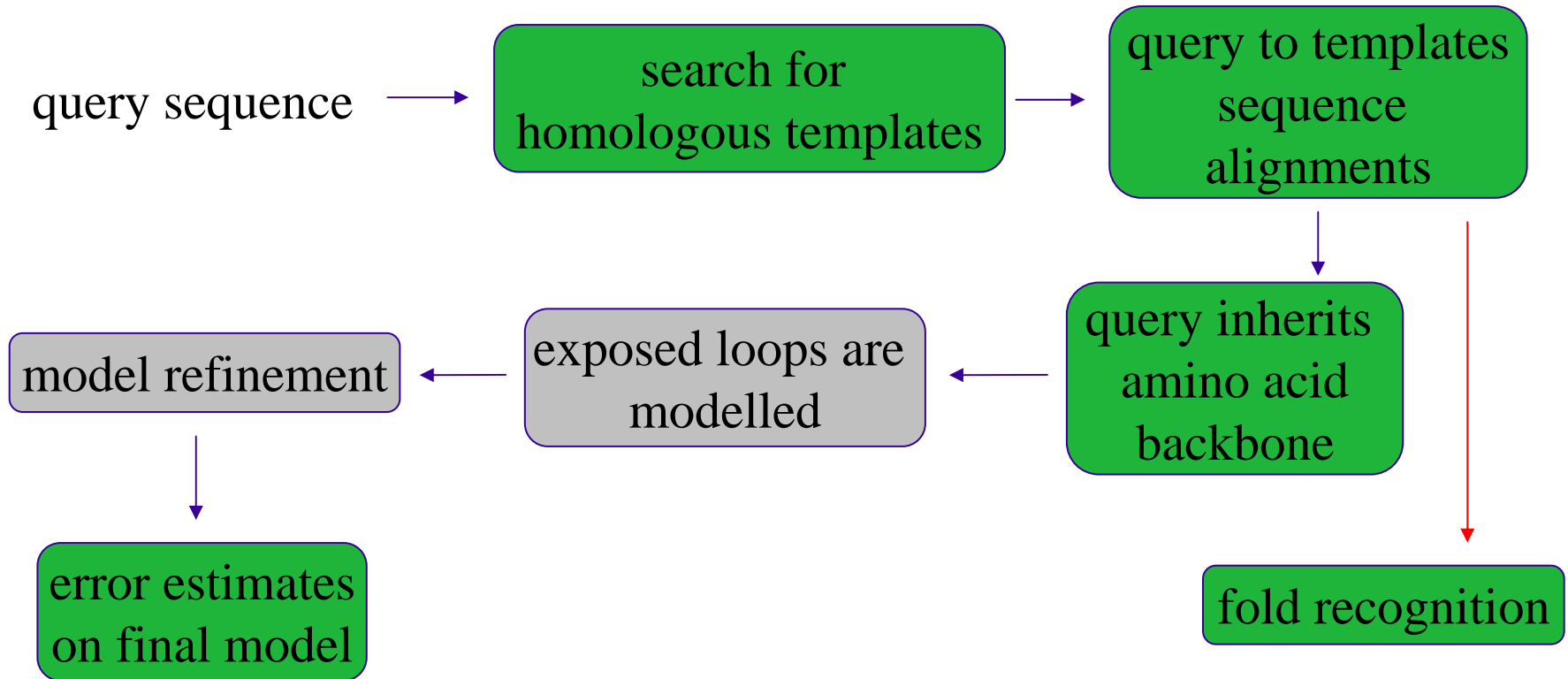
Bruno Contreras-Moreira
and Paul A. Bates
Biomolecular Modelling Laboratory
London Research Institute
Salamanca, September 2002

Protein comparative modelling?

Predictive technique to build a molecular model for a sequence based on homologous proteins whose structure is known.



What I have been doing



Possible applications of comparative modelling*

Depending on the sequence identity between query and template:

- >90% virtual ligand screening
- >40% defining antibody epitopes
- >40% molecular replacement in X-ray crystallography
- >20% support site directed mutagenesis
- >20% fitting into low resolution electron density maps

* Baker & Sali (2001) Science 294: 93-96

Alignment accuracy (I)

We tested three alignment methods on a set of 428 pairs of homologous proteins:

pairwise **Clustal**
sequence to sequence,
Gonnet,BLOSUM matrices:
avlwqstw
aviwqs-y
232843-2

profile-1
profile+pss to sequence+kss:
avlwqstw
aviwqs-y
aviwqs--
2328321

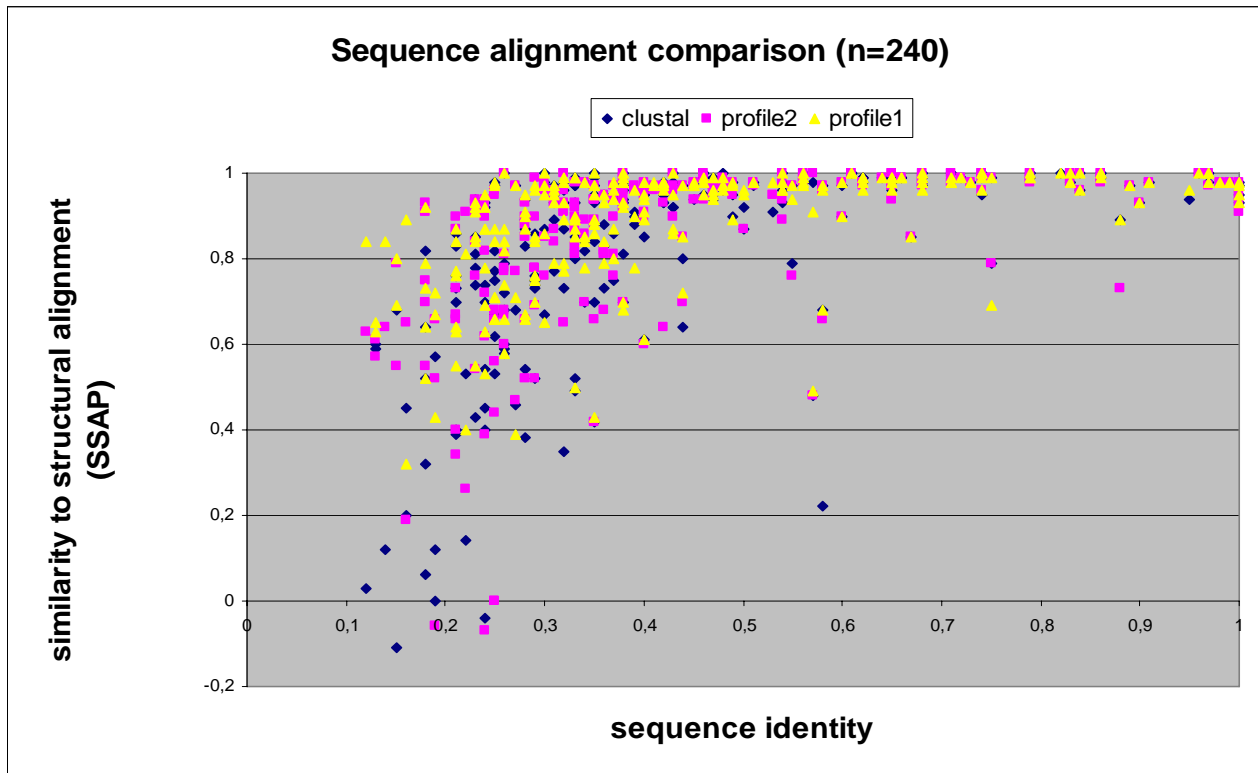
bit-score:
 $\Sigma s_i/n$

profile-2
profile+pss to profile+kss:
avlwnstw
avlwqstw
aviwqs-y
aviwqs--
2328230

psipred

pss: predicted secondary structure
kss: known secondary structure

Alignment accuracy (II)



In addition, a natural cutoff for the bit-score was found to identify accurate alignments (240/428)

Searching for templates

sequence against
profile library

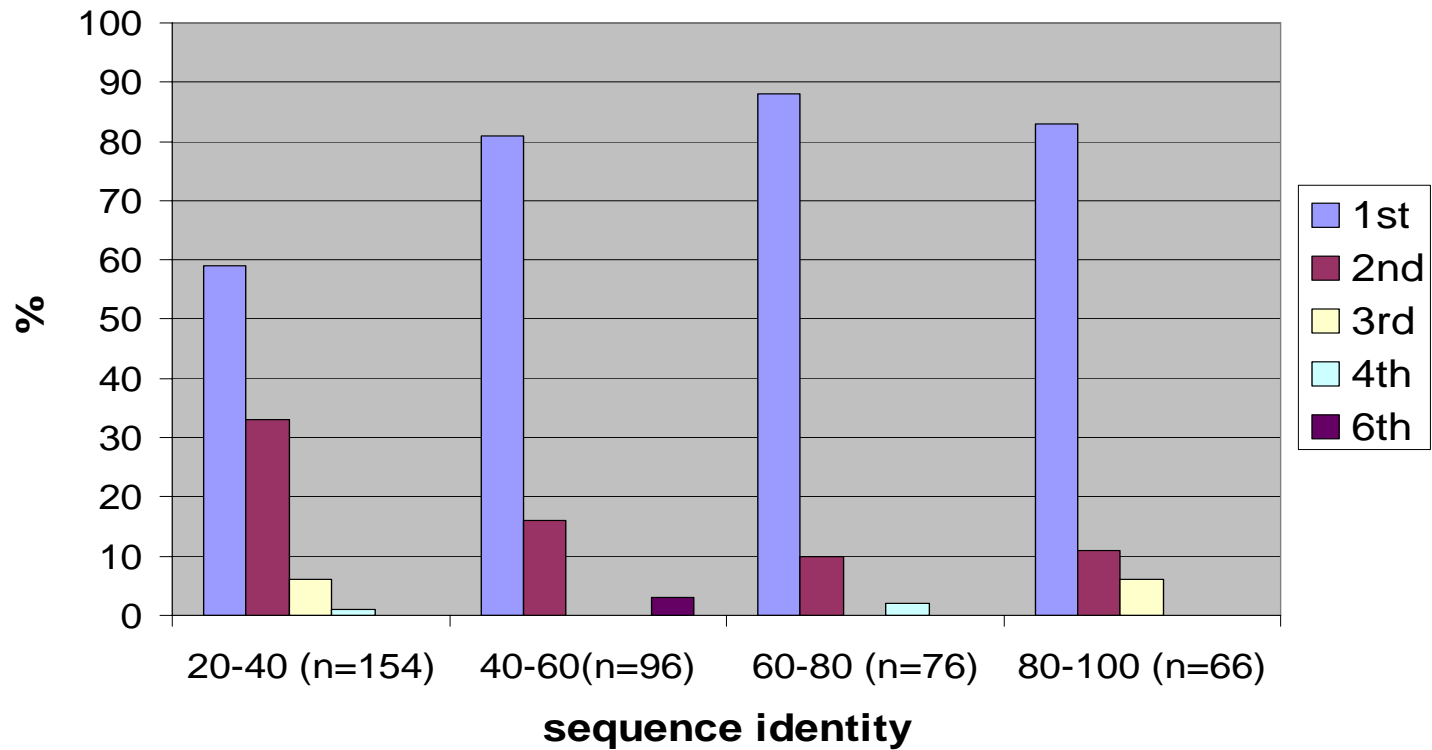
- PFAM profiles: 290/300
- in-house SCOP profiles

sequence against
database of sequences

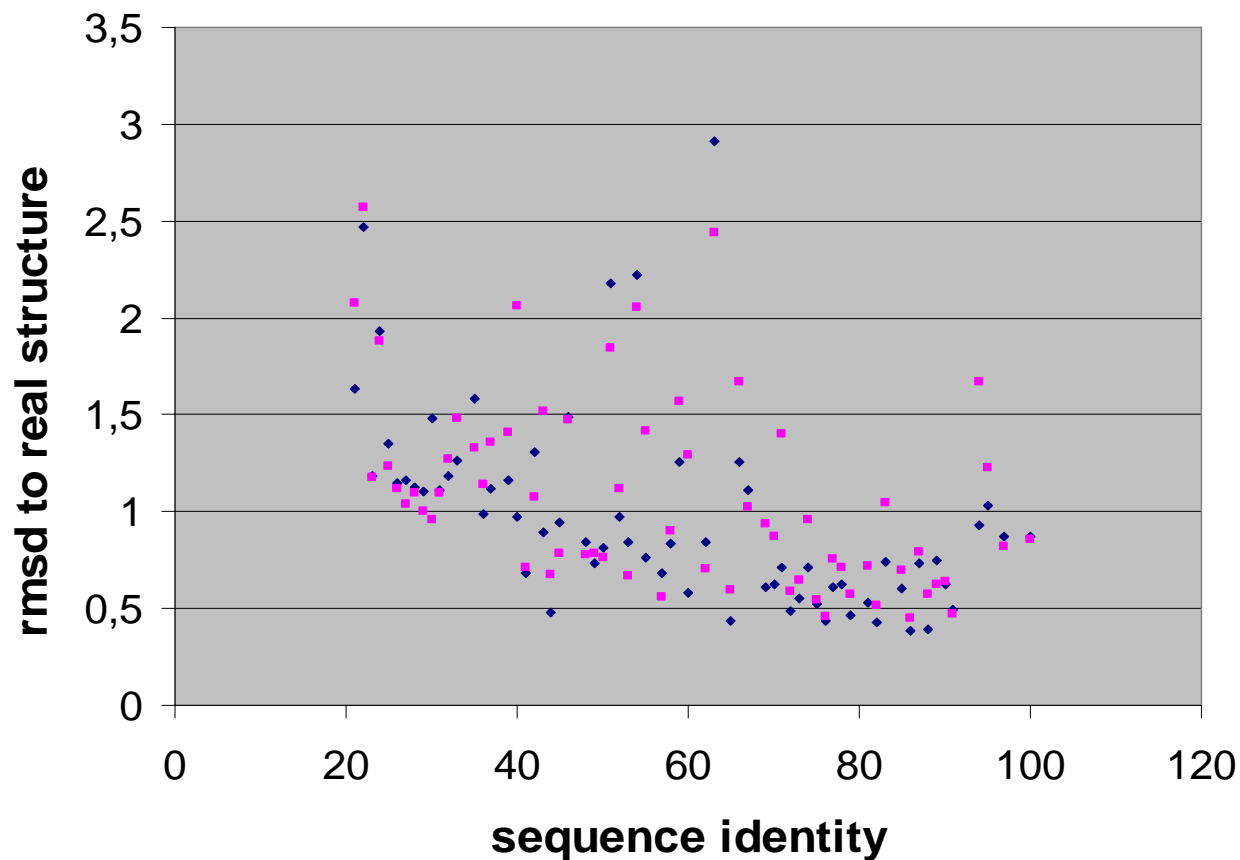
- PFAM + PDB sequences: 300/300
plus: domain splitting

Selecting templates

Best template for comparative modelling

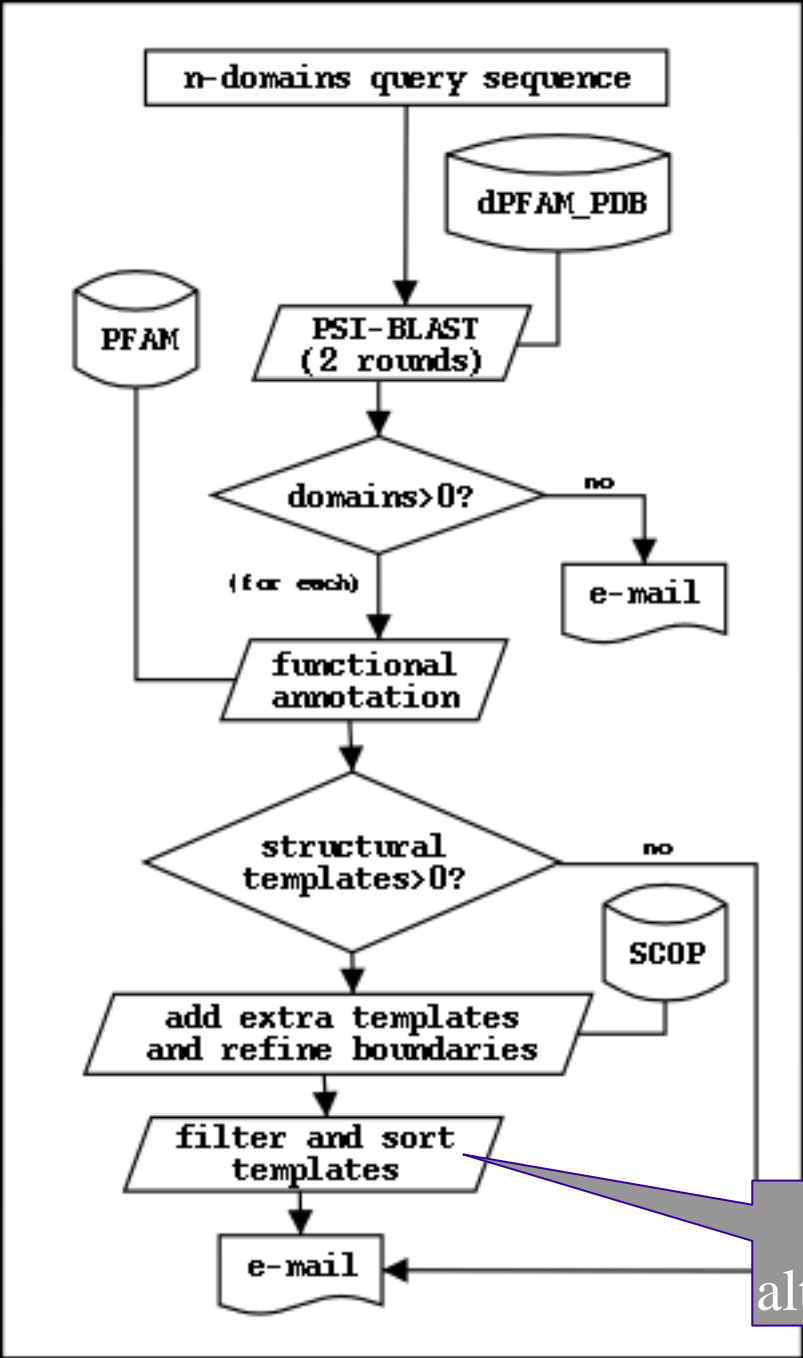


Single vs Multiple template modelling



Overall:
54/97 S
43/97 M

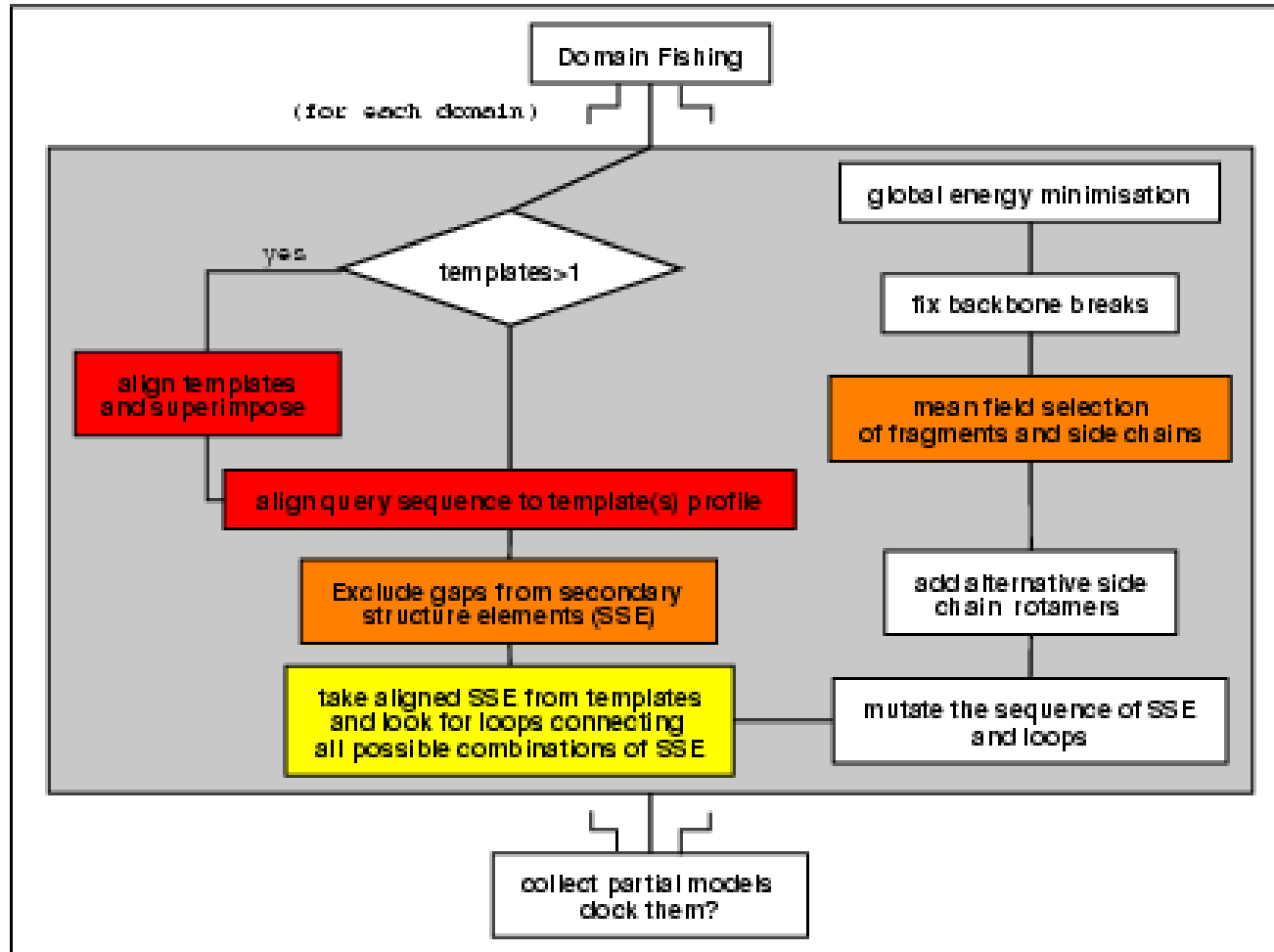
Domain Fishing



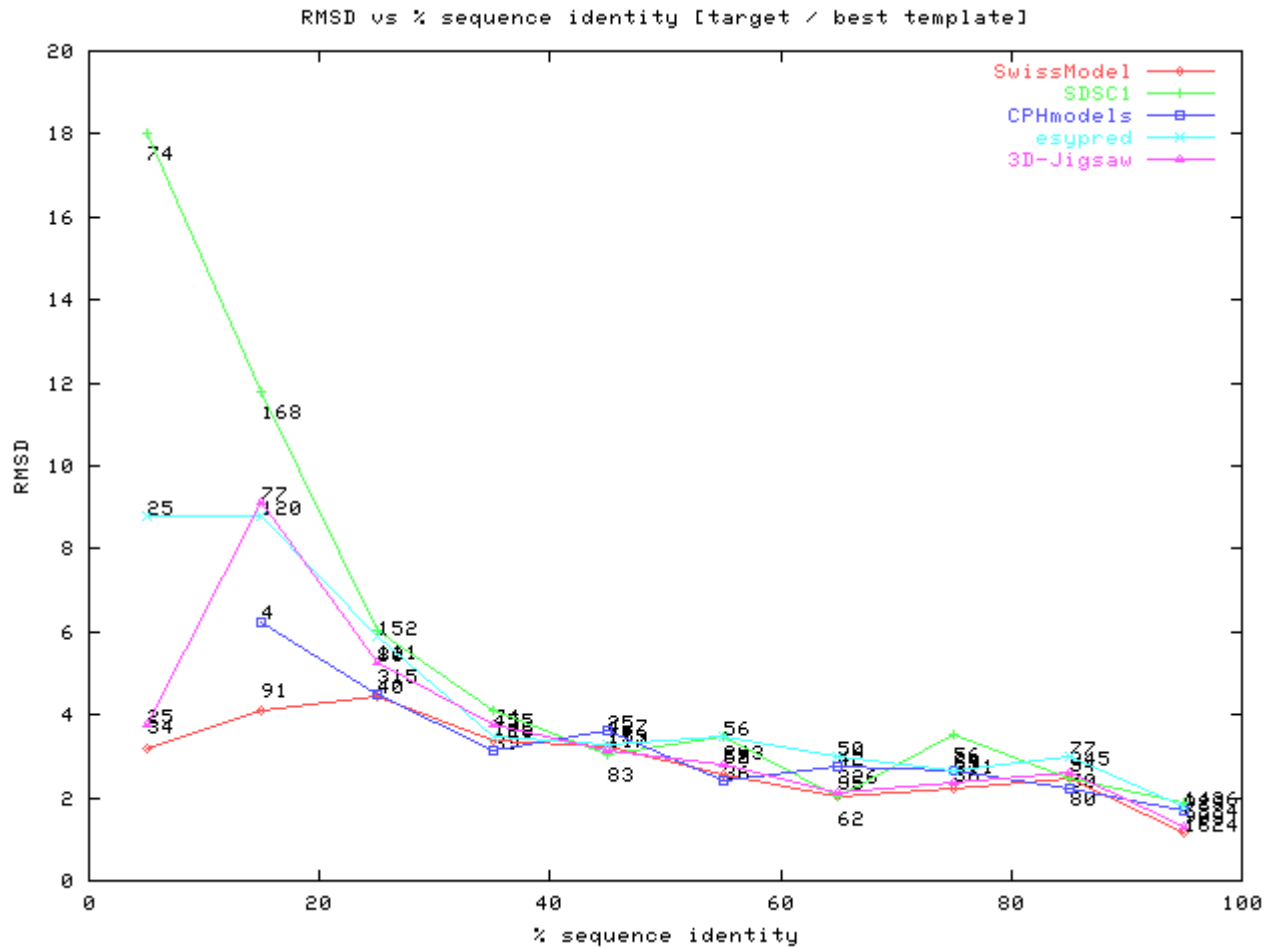
up to 7 alternative alignments

3D-JIGSAW

Example



EVA: continuous evaluation of servers



<http://cubic.bioc.columbia.edu/eva>

Conclusion

We have done:

- automatic domain identification
- improved alignments
- multidomain modelling

We want to do next:

- better template selection (energies)
- connecting domains
- different multi-template strategies

The Biomolecular Modelling Laboratory

Paul Bates

Paul Fitzjohn

Pall Jonsson

Graham Smith

Chris Page

Bruno Contreras Moreira

www.bmm.icnet.uk