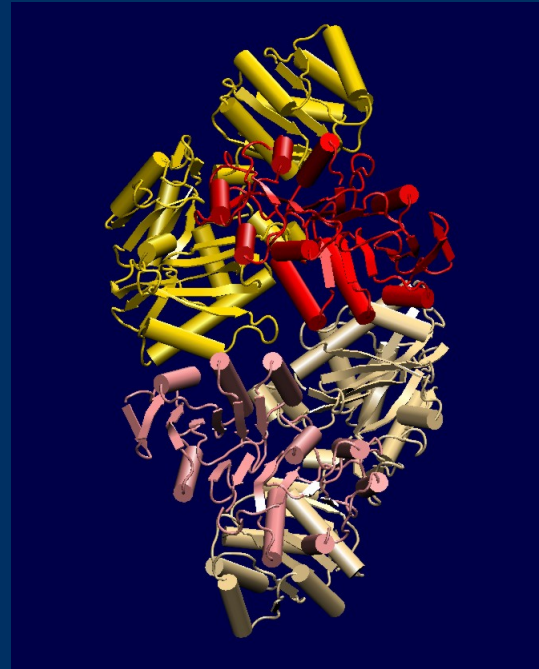
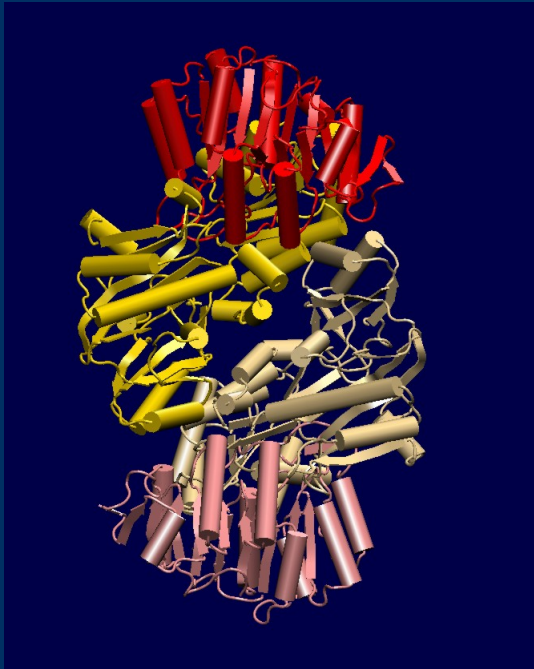


Inference of Protein Quaternary Structure from Crystals

Hannes Ponstingl

Thornton Group, EMBL-EBI
hpo@ebi.ac.uk

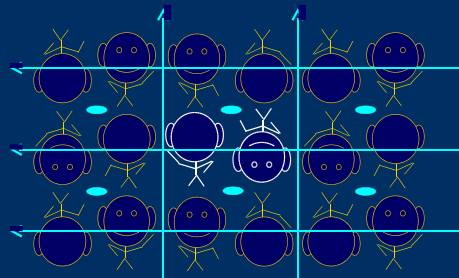
PDB file may not show the correct assembly



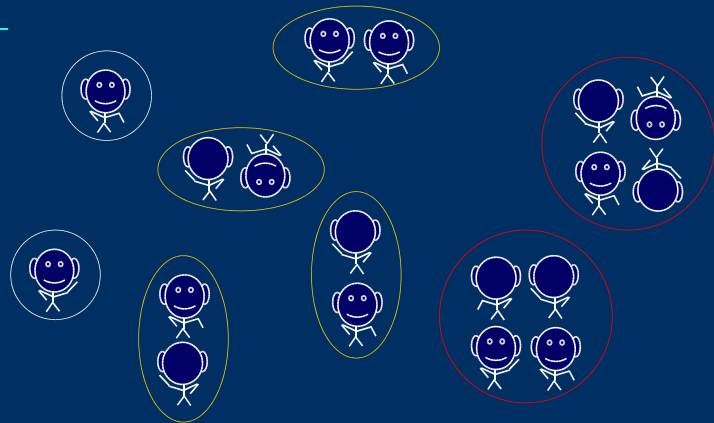
Succinyl-CoA Synthetase from *E.coli* (PDB code: 2SCU)

Wolodko *et al.* (1994), Fraser *et al.* (1999)

Quaternary Structure Inference from Crystals



Crystal



Solution

Overview

Morphology of Oligomers

- 'bona fide' data set
 - ▷ crystals of soluble oligomers
 - ▷ low redundancy (sequence & structure)
 - ▷ 141 homomers, 22 heteromers, 54 monomers (pseudo-dimers)
- contact size, hydrophobicity

heuristic inference of physiologically relevant assembly

- neglect (local) conc. of constituent subunits, pH, salinity, . . .
- crystal contacts \longleftrightarrow physiological interfaces by contact size, chemical complementarity
- contact score \longrightarrow assembly by successive bi-partitioning of contacts, symmetry constraints

performance assessment

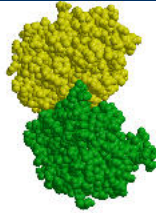
- overall error rate 16%

limitations

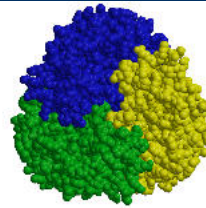
- dependence on thresholds, overfitting, . . .

Examples of Multimers

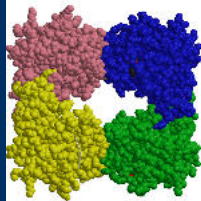
Triose Phosphate
Isomerase (1AMK)



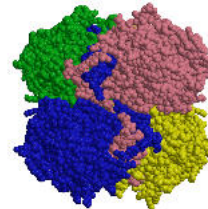
Chloramphenicol
Acetyltransferase
(3CLA)



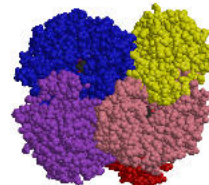
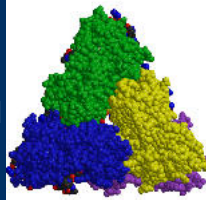
Human β -Tryptase
(1A0L)



Catalase (1A4E)



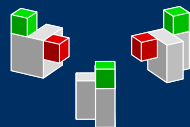
Branched-Chain Amino Acid
Aminotransferase (1A3G)



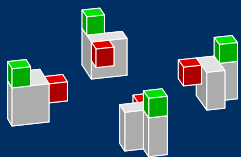
Symmetries in Oligomer Data Set



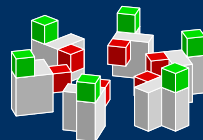
C_2



C_3



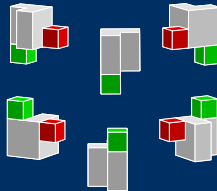
C_4



C_6



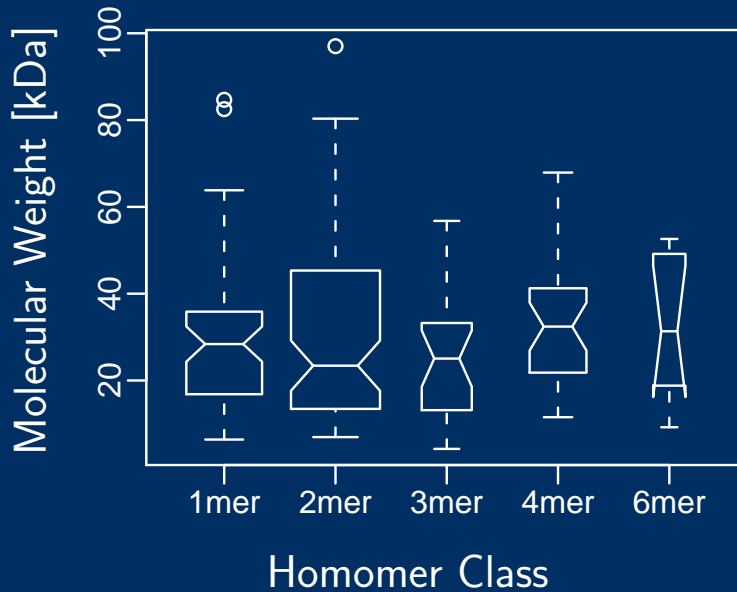
D_2



D_3

(courtesy of J. Barker, EBI)

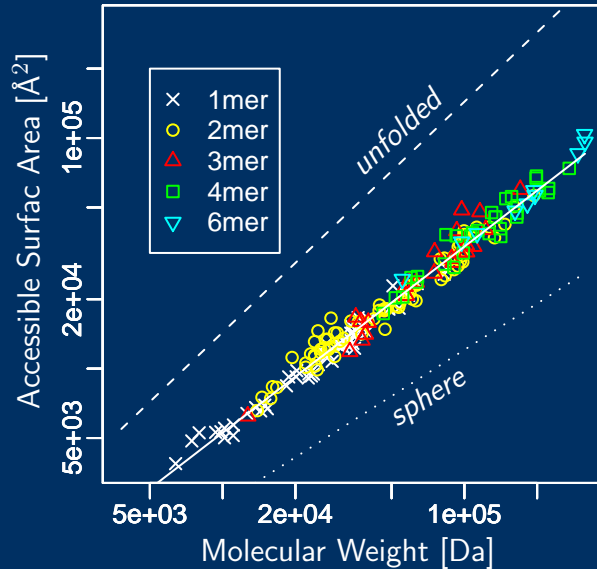
Molecular Weight of Homomer Subunits



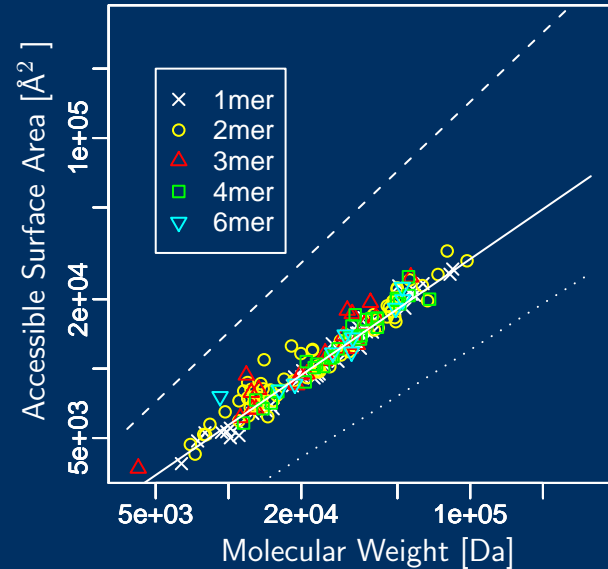
Homomeric Proteins

Surface Area vs Molecular Weight

Entire Multimer



Isolated Subunit



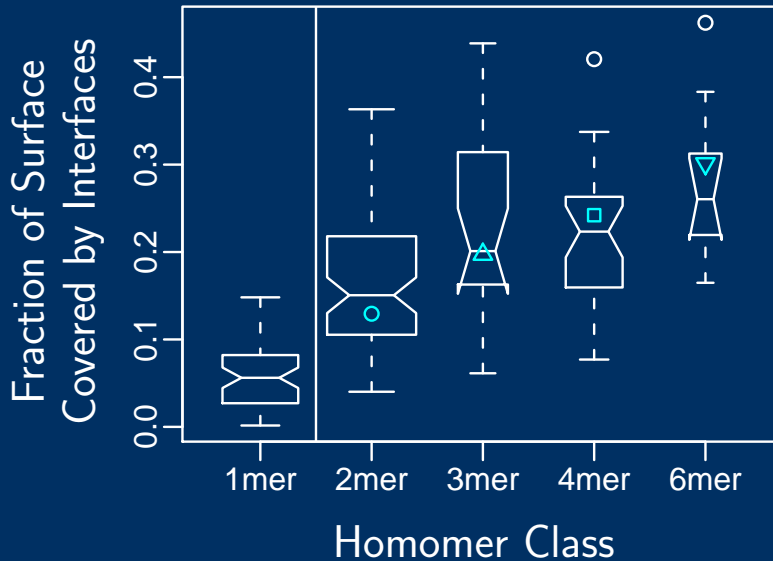
$$A \approx \alpha M^\beta$$

$$\beta = 0.81 \text{ (SE 0.01)}$$
$$\alpha = 3.1 \text{ (SE 0.3)}$$

$$\beta = 0.71 \text{ (SE 0.02)}$$
$$\alpha = 9.1 \text{ (SE 1.5)}$$

Homomeric Proteins

Fraction of Surface Covered by Interfaces

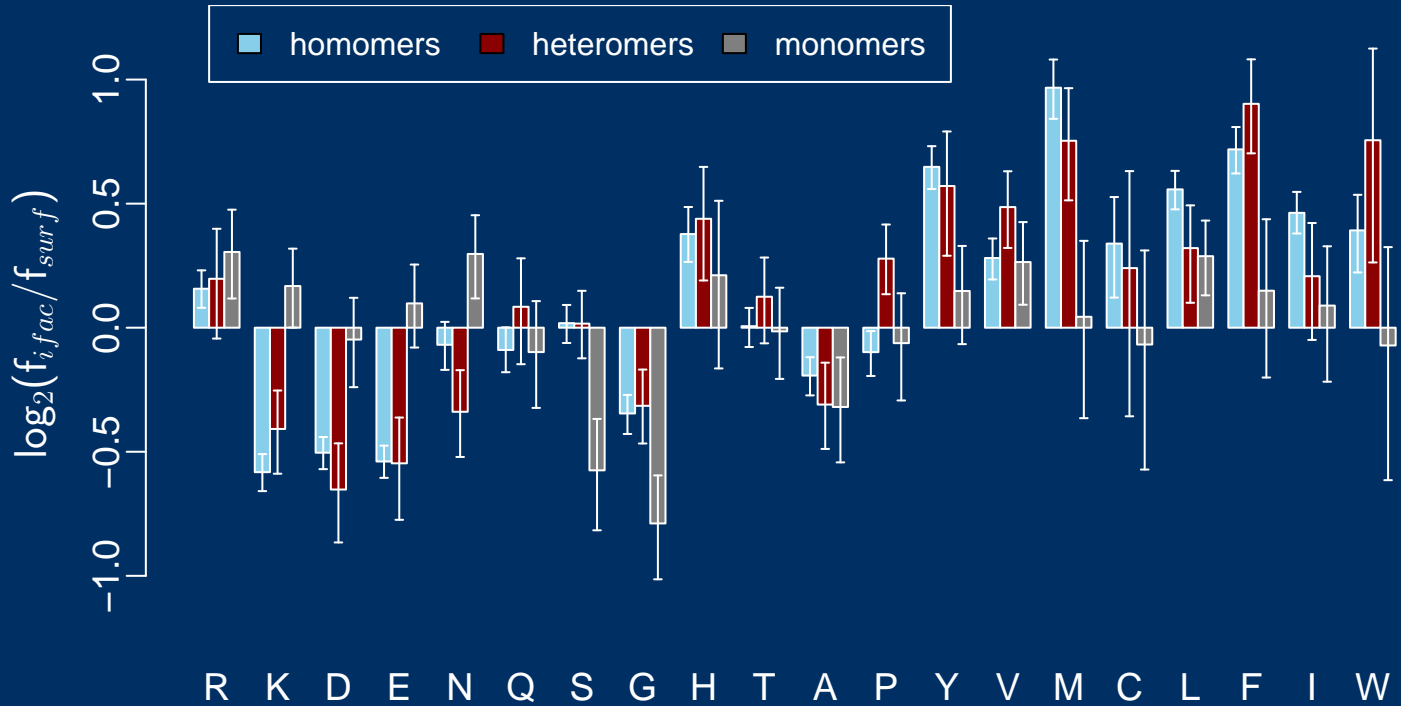


$$\frac{\Delta ASA}{ASA} \approx 1 - n^{\beta-1} \approx 1 - n^{-0.2}$$

subunit multiplicity $n > 1$

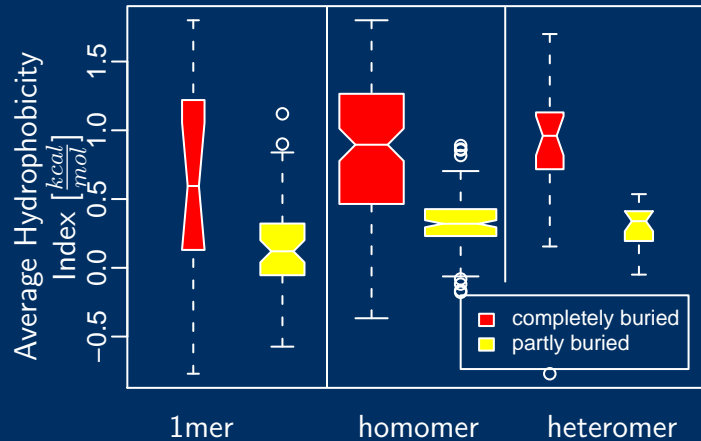
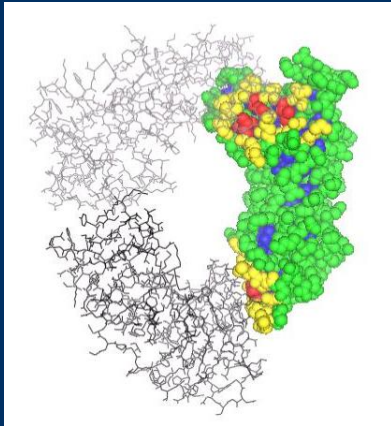
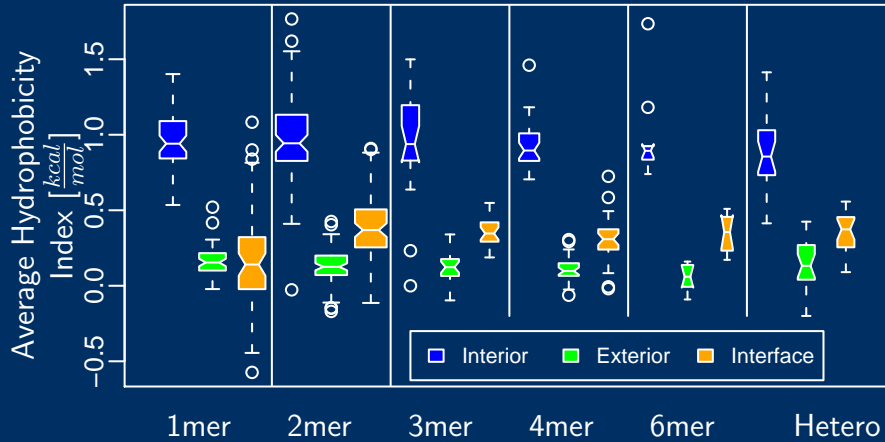
Amino Acid Propensities

interface vs free surface

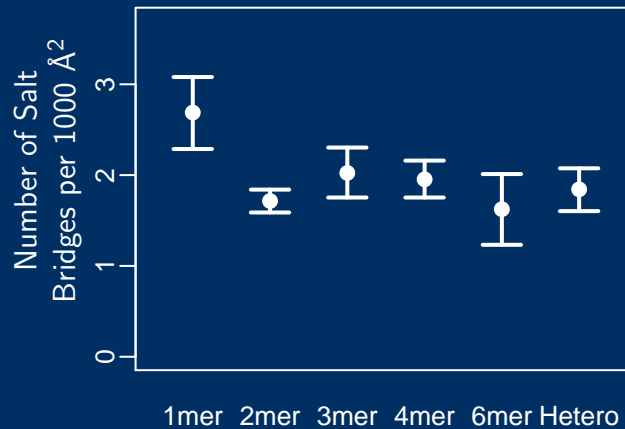
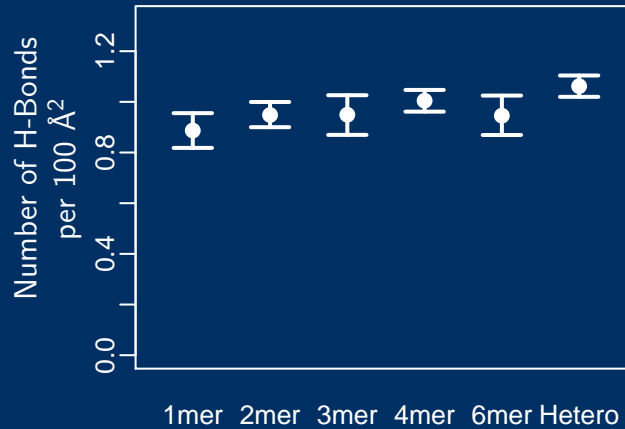


Hydrophobicity Index

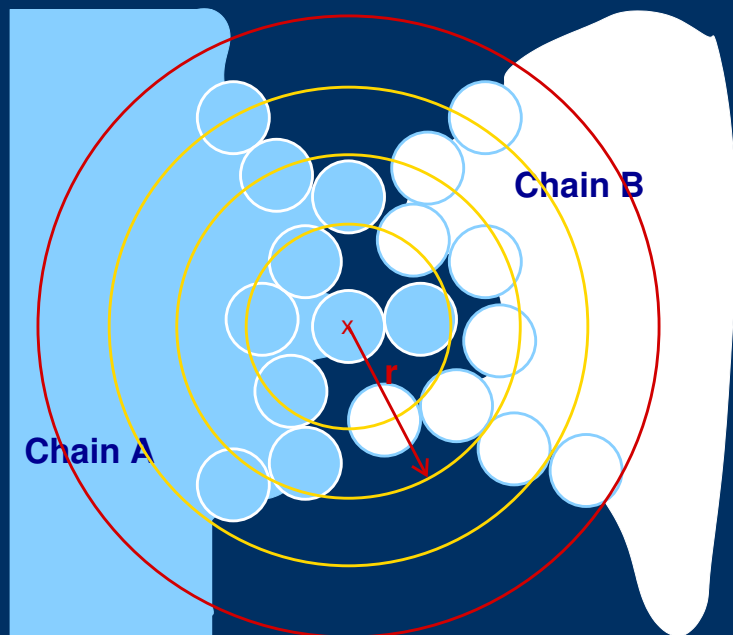
based on Fauchère & Pliška hydrophobicity scale



Polar Interactions across Interfaces



Atom Pair Score



17 atom types

[Mitchell *et al.* (1999) *Chem. Inf. Comp. Sci.*, **39**, 751.]

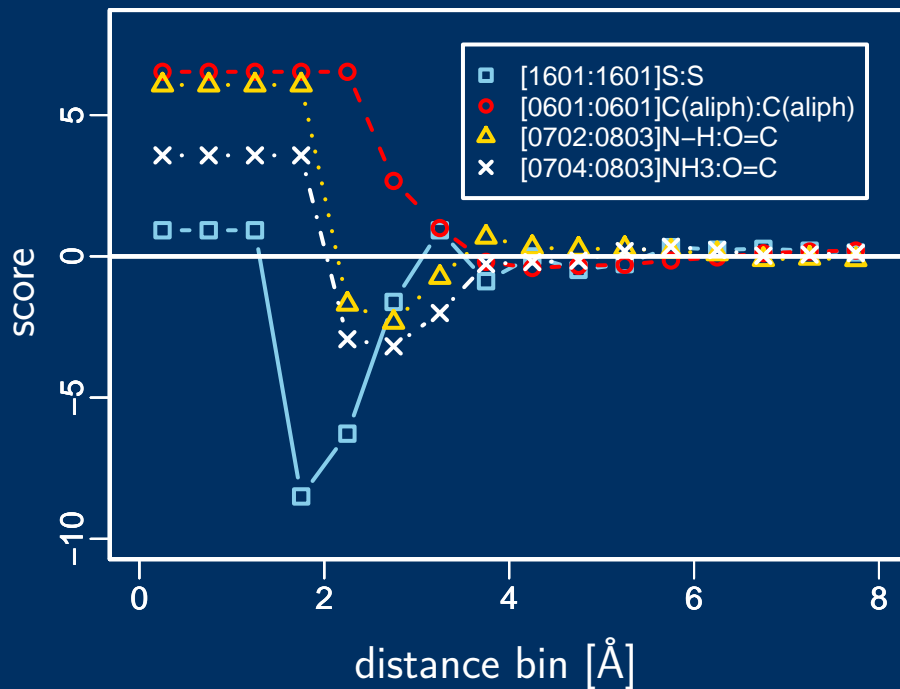
sampling: up to 8 Å, 0.5 Å bin width

$$s_{ijr} = \ln\left(1 + \sigma \sum_r n_{ijr}\right) - \ln\left(1 + \sigma \frac{\sum_r n_{ijr}}{\sum_{ij} n_{ijr}} n_{ijr}\right)$$

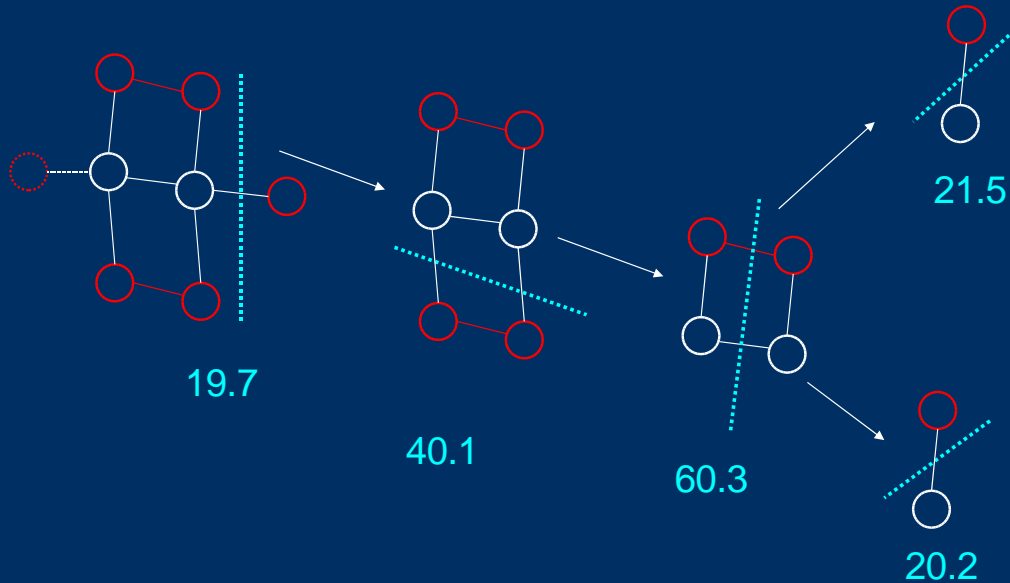
n_{ijr} : number of atom-pairs of types i and j at distance bin r

[Sippl (1990) *J. Mol. Biol.*, **213**, 859.]

Atom-Pair Scores

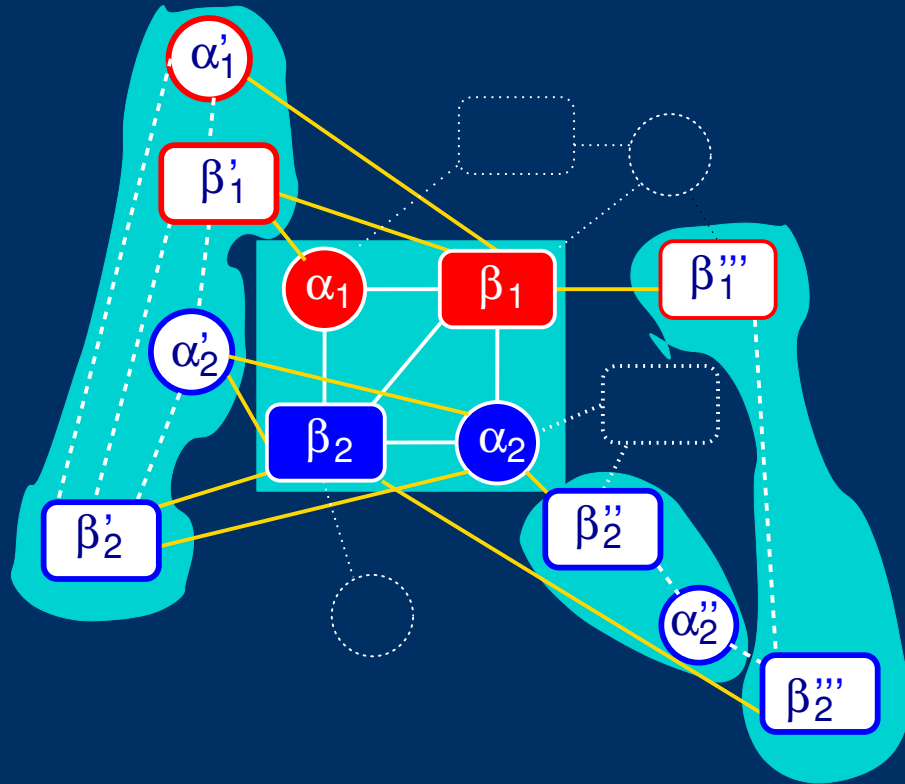


Heuristic Quaternary Structure Inference

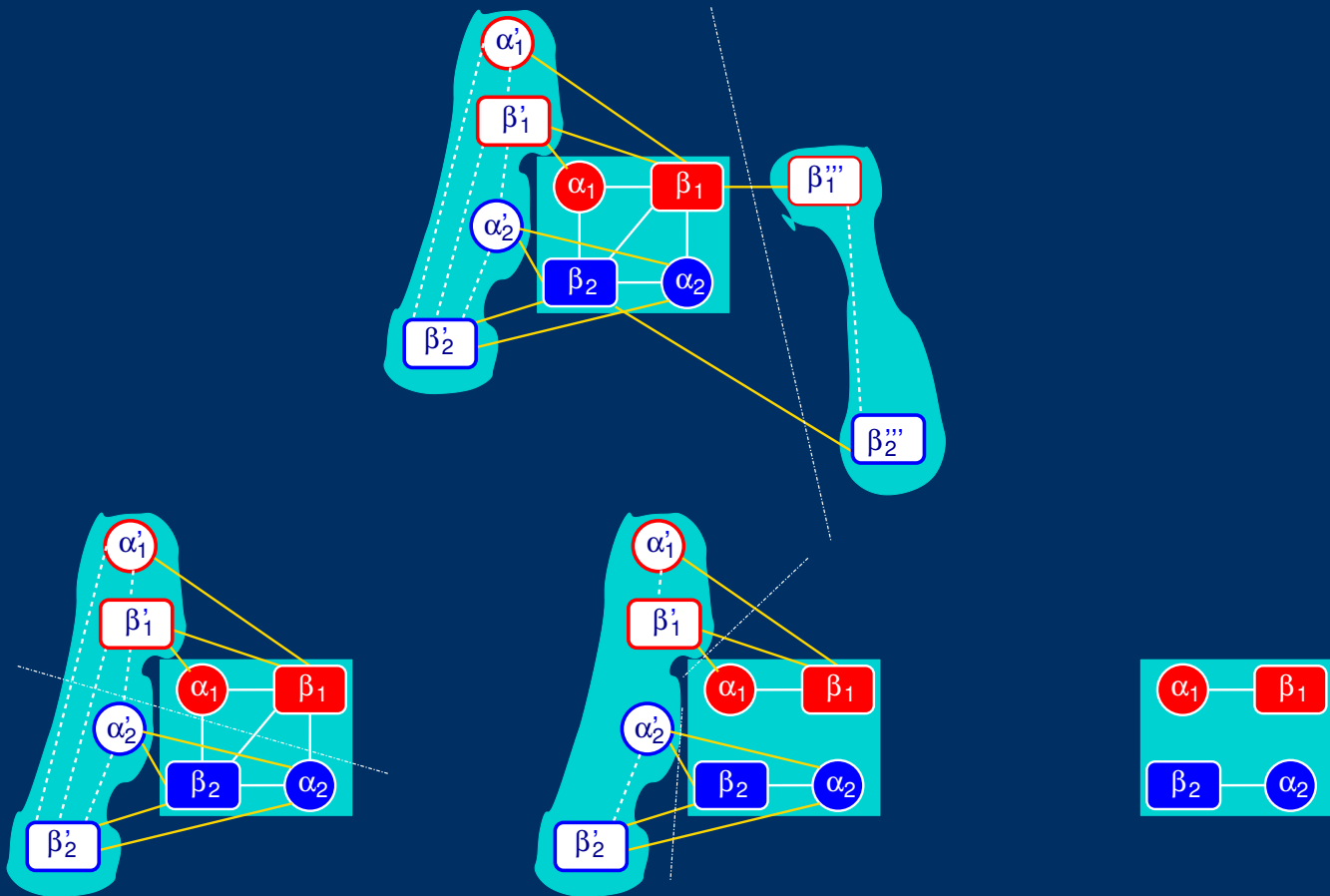


Based on MinCut algorithm by Stoer & Wagner (1994).
Assembly selection by score threshold.

Graph Representation of Subunit Contacts for 2SCU



Successive Partitioning of 2SCU



Performance Assessment

Reference	Predicted						Sum	Correct
	1mer	2mer	3mer	4mer	6mer	other		
1mer	46	6	0	2	0	1	55	84%
2mer	4	79 (11)	1	3	0	1 (1)	88 (12)	88%
3mer	2	0	20	0	2	0	24	83%
4mer	1	5	0	28 (5)	1 (1)	3 (1)	38 (7)	74%
6mer	1	0	1 (1)	0	11 (2)	0	13 (3)	85%

Cramér's Φ for $r \times s$ tables

$$\Phi_C = \sqrt{\frac{\chi^2}{n(k-1)}}$$

where $k = \min(r, s)$, n total number of counts, $0 \leq \Phi_C \leq 1$

Pearson χ^2 statistic

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

where f_{ij} are frequencies in the table cells with row and column indices i, j and $e_{ij} = \frac{f_{i.} f_{.j}}{f_{..}}$ are the expected frequencies.

Performance Assessment dependence on threshold



Generalisation Behaviour

classification problem: interface \longleftrightarrow crystal contact

scoring method	error rate [%]	
	apparent	cross-validated
contact size	13.5	15.4
atom-pair score	7.0	12.5

Summary

performance assessment

- 16% error rate in identifying the subunit multiplicity
 - ▷ relative oligomer frequencies in data set representative of PDB
 - ▷ correct assembly given multiplicity (anecdotal)
- improvement over simple classification rules
- very fast

limitations

- small ligands not considered
- heuristic
 - ▷ does not look at all possible assemblies
 - ▷ may miss correct assembly
- pair-score tends to overfit

<http://www.ebi.ac.uk/Thornton/software.html> (QuasiProX)

Acknowledgements

Janet Thornton (EBI)

Thomas Kabir (UCL)

Thibault Varin (Poitiers)

Kim Henrick (EBI, Macromolecular Structure Database)