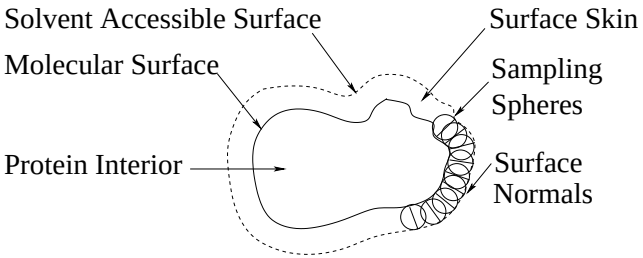
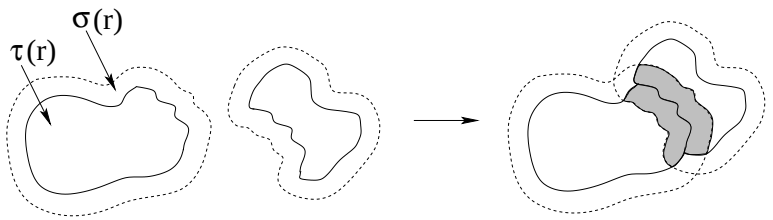
3D protein shape density representation in Hex



$$\sigma(\underline{r}) = \begin{cases} 1; & \underline{r} \in \text{surface skin} \\ 0; & \text{otherwise} \end{cases} \qquad \tau(\underline{r}) = \begin{cases} 1; & \underline{r} \in \text{protein atom} \\ 0; & \text{otherwise} \end{cases}$$

[Ritchie & Kemp (2000) Proteins **39:178–194**]

Protein shape complementarity



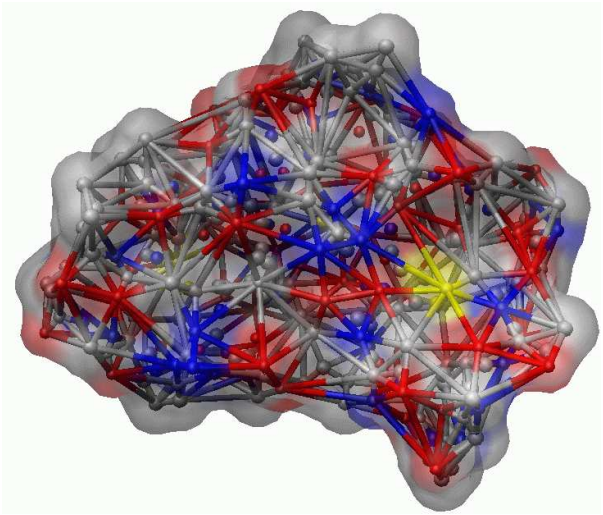
Favourable: $\int(\sigma_A(\underline{r}_A)\tau_B(\underline{r}_B) + \tau_A(\underline{r}_A)\sigma_B(\underline{r}_B))\mathrm{d}V$

Unfavourable: $\int \tau_A(\underline{r}_A)\tau_B(\underline{r}_B)\mathrm{d}V$

Score: $S_{AB} = \int(\sigma_A\tau_B + \tau_A\sigma_B - Q\tau_A\tau_B)\mathrm{d}V$

Penalty Factor: $Q = 11$

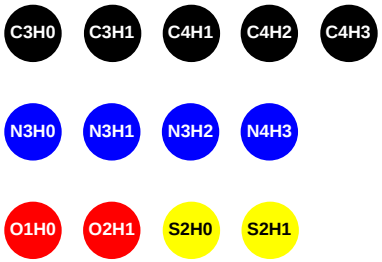
Surface representation



Atomic group in proteins

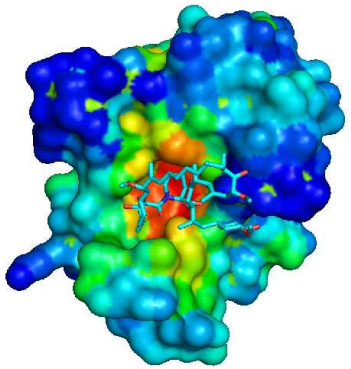
Classifcation proposed by Tsai et al. (J. Mol. Biol., 1999, 290:253-266), based on:

- heavy-atom types,
- the number of covalently attached hydrogen atoms, and
- the number of all covalently attached atoms.

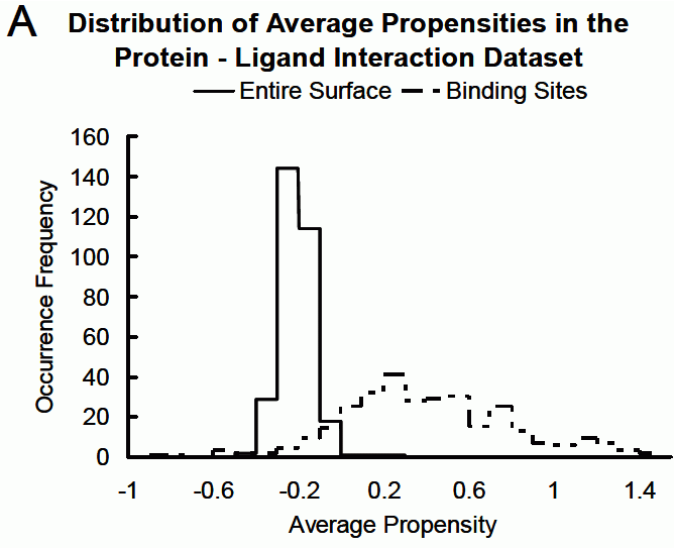


Surface Triplet Propensities

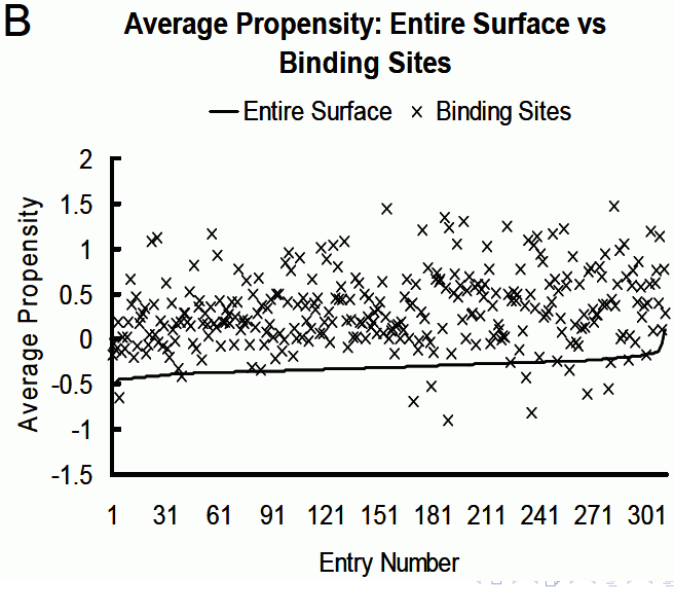
Mehio, W., Kemp, G.J.L., Taylor P. and Walkinshaw, M.D. (2010) Identification of Protein Binding Surfaces using Surface Triplet Propensities. *Bioinformatics*



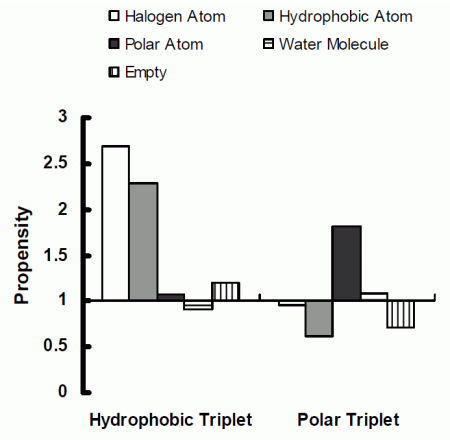
Surface Triplet Propensities



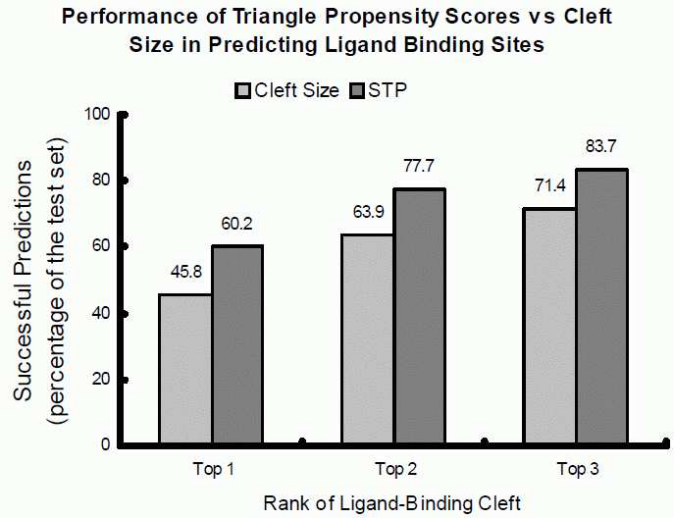
Surface Triplet Propensities



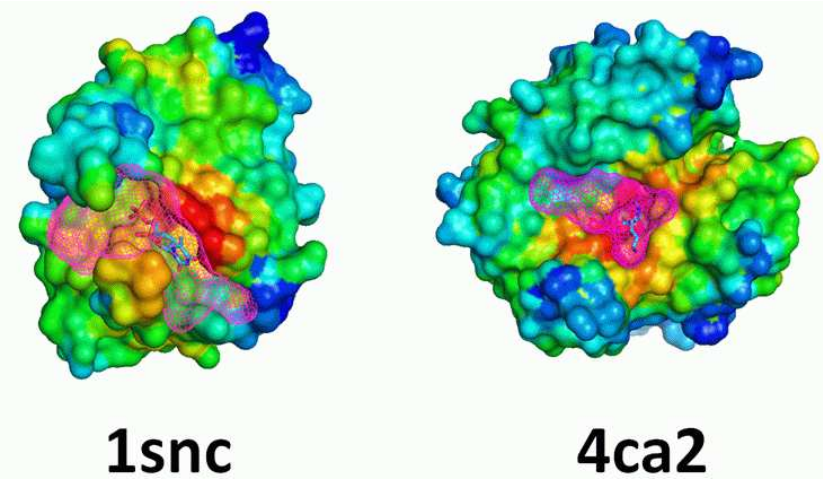
Surface Triplet Propensities



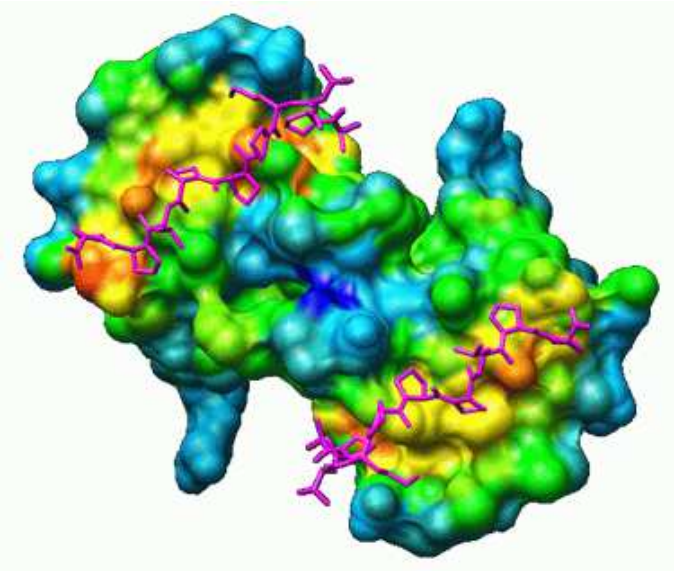
Surface Triplet Propensities



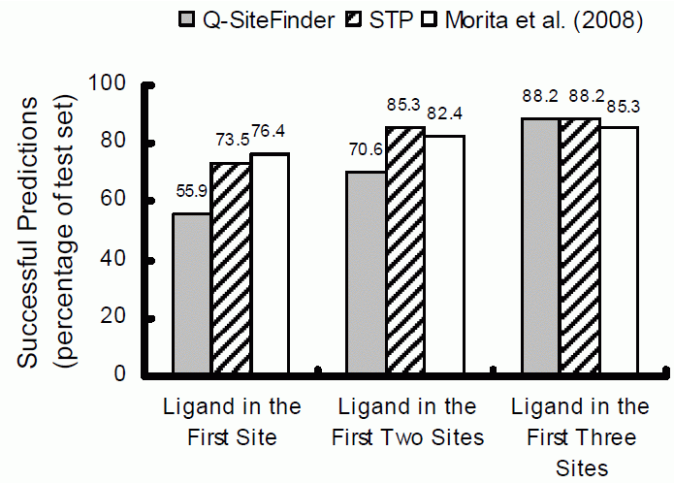
Surface Triplet Propensities



Surface Triplet Propensities



Surface Triplet Propensities



CombDock

[Inbar, Y., Benyamini H., Nussinov R. and Wolfson H.J. (2005) "Prediction of multimolecular assemblies by multiple docking". *J. Mol. Biol.*, **349**, 435-447]

- All pairs docking
  - $N(N-1)/2$ pairs
  - keep best K transformations for each pair
- Combinatorial assembly
  - find best spanning tree representing a valid complex
  - keep best D trees of size s starting at i
- Rescoring
  - cluster (to avoid redundancy in solution set)
  - geometric component
    - large interface area and small steric overlap
  - physico-chemical component
    - count number of buried non-polar atoms

How many spanning trees?

(i) If we have N vertices and 1 edge between each pair there are N^{N-2} spanning trees.

(ii) If we have K edges between each pair of vertices, then there are K^{N-1} graphs of type (i).

So there are $N^{N-2}K^{N-1}$ spanning trees.

Can't search the whole space!
So use a heuristic solution.

CombDock results (1)

Yeast RNA polymerase II elongation complex

10 protein chains

$K = 100$, 15 pairwise interactions predicted

10^{26} possible complexes

50188 complexes generated by combinatorial assembly

1113 complexes left after clustering

2nd ranked complex had RMSD of 1.37.sp 2 Human subunits modelled by homology
6th ranked complex had RMSD of 1.9.

CombDock results (2)

Bovine arp2/3 complex

7 protein chains

$K = 100$

1.68×10^{17} possible complexes

5488 complexes generated by combinatorial assembly

145 complexes left after clustering

3nd ranked complex had RMSD of 1.2.sp 2 *Drosophila melanogaster* subunits modelled by homology
10th ranked complex had RMSD of 1.9.

Structural biology, genomics and proteomics

Structural biology

— gives an understanding of biological function at the molecular level through determination of individual macromolecular structures and complexes.

Structural genomics

— aims to obtain 3D structures for all protein products of the genome.

Structural proteomics

— aims obtain structures of all macromolecular assemblies in the "complexome".

Protein-protein docking

- find sets of candidate docking orientations
- use a scoring function to evaluate these

"Although considerable improvement has been achieved, the scoring functions are still the weakest components of most docking algorithms."

[Inbar Y, Benyamini H, Nussinov R, Wolfson HJ.
Prediction of multimolecular assemblies by multiple docking.
J Mol Biol. 2005 Jun 3;349(2):435-447]

Computational challenges

- to predict atomic models of macromolecules (usually proteins) where experimentally determined structures are not available;
- to predict the structures of pairs of interacting macromolecules;
- to predict assemblies of many macromolecules by multiple docking;
- to fit high-resolution atomic models of macromolecules inside low-resolution envelopes obtained from cryo-EM experiments;
- to build a 3D atlas of a cell at atomic resolution.

Different resolutions

High-resolution

— X-ray, NMR

atomic detail

— atomic coordinates (x,y,z)

— Low-resolution

— (cryo)electron microscopy

— voxels with density values