Total surface comparison

Glimpse into the morphogenesis of proteins

Andreas Hoppe and Cornelius Frömmel

Institut für Biochemie
Charité, Universitätsmedizin Berlin
Monbijoustr 2A, D-10117 Berlin, Germany

Abstract

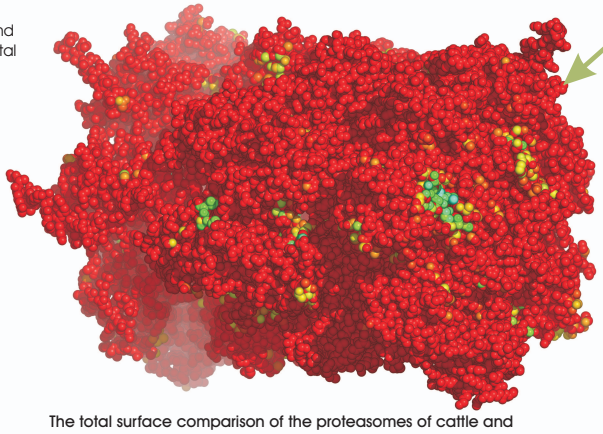
Proteins successful in evolution proved two main aspects: stability and function. The stability relates to the fold and the function relates to the molecular surface. It is a duality of core and periphery. Structure alignment programs can discover deep evolutionary relationships of proteins via their folds, even if there is no sequence similarity. The importance to look at the other side of the duality, the evolution of molecular surfaces, is beginning to be recognized [1-3]. It is generally accepted that the geometry of active sites and other important surface areas is more conserved than the actual amino acids that build them (similar to the fold). The present work aims to get more detailed knowledge.

We present an analogue to structure alignment for protein surfaces. The main difference to structure alignment is that the protein fold is a global property whereas surface 3D-similarity (and hence protein function) is a local property. Thus the result of a total surface comparison cannot be a single alignment but a collection of corresponding surface patches.

The technical challenges are to split the surface in compartments such that similarity scores of the respective surface patches are comparable and to interpret a multi-dimensional result vector.

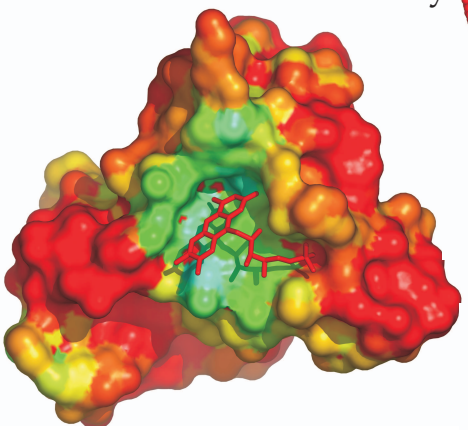
Discover peripheral conservation in evolution.

Color scheme: rainbow
Bad scores = high varibility = red
Good scores = low varibility = blue

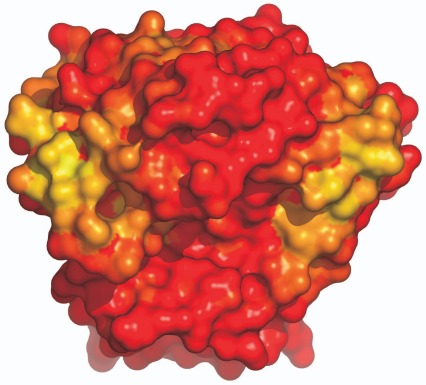


The total surface comparison of the proteasomes of cattle and yeast. The well-known key sites (active sites, passage between the chambers, docking site of the activator) are clearly recognizable as structurally conserved which underlines the validity of the method. But there are other structurally conserved areas whose function has not been known beforehand. Most interestingly, our results indicate substrate tracks inside the proteasome chambers.

Discover surface variability

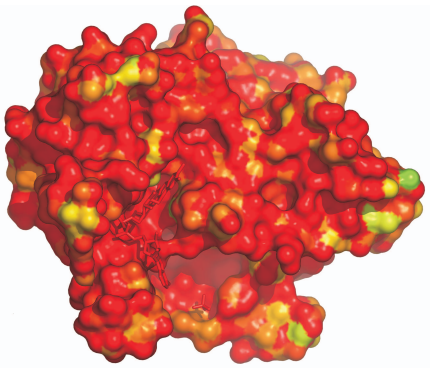


Surface variability. The surface of a given molecule is not rigid. Single atoms, atom groups and whole molecule parts move stochastically. External molecules and other inuences can change the geometry of the molecule dramatically. It is clear that some areas of the molecular surface are more variable than others, relating to the function and stability of the molecule. No other method catches more details of this variability than the different models of an NMR analysis. We compared the surfaces of each possible pair of the 20 models for the FMN-binding protein from desulfovibrio vulgaris (PDB: 1AXJ). For each atom we calculated the average score for all comparisons this atoms appears at the surface which translates to the color.

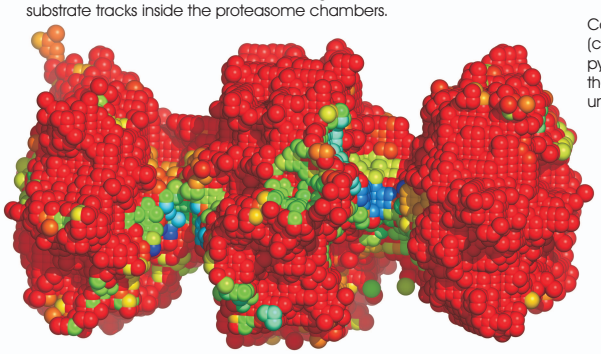


The same analysis for bacillus alcalophilus serine protease PB92 (PDB: 1AH2).

Discover convergent evolution

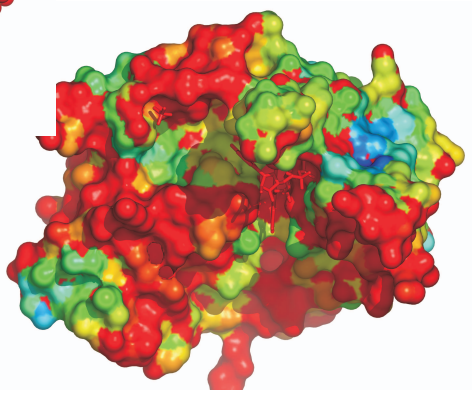


Comparison of ferredoxin-NADP+ reductase (cyanobacterium anabaena, PDB: 1BJK) and human pyridoxine-5'-phophate oxidase (PDB: 1NRG). Both bind the isoalloxasine ring but are phylogenetically completely unrelated.



Visualisation of the interior. The balls are spheres (radius 1.4A) placed on a grid with mesh size 1.4A which do not clash any of the proteasome atoms. In other words, the balls show hypothetical water atoms which might fill the interior of the proteasome (but not all at the same time). The color is the average of colors of surrounding atoms weighted by the distance.

Discover diversity



Ferredoxin reductase from spinach (1BX0) and pea (PDB 1QG0). Related protein in distant plant species, belonging to subclasses Caryophyllidae versus Rosidae.

Method

Central aspects:

1. the asymmetric design: one (focus) molecule is colored based on comparisons with the other (reference) molecule
2. multi-level, vectorial scoring function
3. overlapping compartments of the model: equally sized surface patches with a standard shape,
4. coding the superposition RMSD scores with color,
5. NeedleHaystack [4] superpositions: model surface patches are superposed with the complete reference molecule surface

Algorithm outline:

1. Compute the surfaces of both focus and reference molecule.
2. Compute the surface connection graph of the focus molecule (two atoms count as connected if a 1.4Å can be placed that it touches both atoms).
3. For every atom A of the focus surface repeat
4. Build a normal surface patch upon A using the surface connection graph (minimal number of steps in the connection graph to A, and minimal distance for the same number of steps).
5. Superpose the surface patch with the reference molecule surface, record the score and the atom of the reference molecule which corresponds to A (atom correspondances).
6. goto step 3
7. Collect superpositions if the

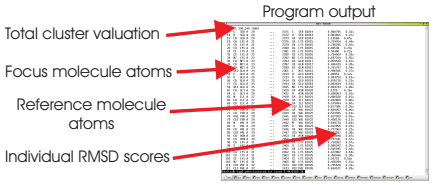
- score is better than a certain bound and
- the atom correspondances refer to neighboring atoms in both focus and reference molecule

Clustering the match pairs

Beyond the above graphical analysis we are able to identify regions of significant similarity.
Patch matches are bundled if

- The score is better than a certain bound
- Neighbouring patches of the focus molecule also match neigbouring patches of the reference molecule.

Based on the size and the individual scores a cluster recieves a score for comparison.



[1] Via A, Ferre F, Brannetti B, Helmer-Citterich M. Protein surface similarities: a survey of methods to describe and compare protein surfaces. Cell Mol Life Sci. 2000;57:1970-7.
[2] Pawlowski K, Godzik A. Surface map comparison: studying function diversity of homologous proteins. J Mol Biol 2001;309:793-806.
[3] Pickering SJ, Bulpitt AJ, Efford N, Gold ND, Westhead DR. AI-based algorithms for protein surface comparisons. Comput Chem 2001;26:79-84.
[4] Hoppe A, Frömmel C. NeedleHaystack: a program for the rapid recognition of local structures in large sets of atomic coordinates. J Appl Cryst 2003;36:1090-7.