CMLS, Cell. Mol. Life Sci. 57 (2000) 1970–1977
1420-682X/00/141970-08 $ 1.50 + 0.20/0
© Birkhäuser Verlag, Basel, 2000

CMLS Cellular and Molecular Life Sciences

Review

Protein surface similarities: a survey of methods to describe and compare protein surfaces

A. Via[a], F. Ferrè[a], B. Brannetti[a] and M. Helmer-Citterich[a,b,*]

[a]Centro di Bioinformatica Molecolare, Department of Biology, University of Rome 'Tor Vergata', via della Ricerca Scientifica, 00133 Rome (Italy), Fax + 39 06 2023500, e-mail: citterich@uniroma2.it
[b]I.N.F.M., University of Rome 'Tor Vergata', via della Ricerca Scientifica, 00133 Rome (Italy), Fax + 39 06 2023500, e-mail: citterich@uniroma2.it

Received 23 May 2000; accepted 15 June 2000

Abstract. Many methods have been developed to analyse protein sequences and structures, although less work has been undertaken describing and comparing protein surfaces. Evolution can lead sequences to diverge or structures to change topology; nevertheless, surface determinants that are essential to protein function itself may be mantained. Moreover, different molecules could converge to similar functions by gaining specific surface determinants. In such cases, sequence or structure comparisons are likely to be inadequate in describing or identifying protein functions and evolutionary relationships among proteins. Surface analysis can identify function determinants that are independent of sequence or secondary structure and can therefore be a powerful tool to highlight cases of possible convergent or divergent evolution. This kind of approach can be useful for a better understanding of protein molecular and biochemical mechanisms of catalysis or interaction with a ligand, which are usually surface dependent. Protein surface comparison, when compared to sequence or structure comparison methods, is a hard computational challenge and evaluated methods allowing the comparison of protein surfaces are difficult to find. In this review, we will survey the current knowledge about protein surface similarity and the techniques to detect it.

Key words. Protein surface; surface comparison; patterns; molecular evolution.

Introduction

If a protein is found to be evolutionarily related to another protein, then information about the function or structure of that protein can be inferred from the other, with varying degrees of reliability, depending on the evolutionary distance between them.

At short evolutionary distances, string comparisons between two protein sequences are usually sufficient to extend to one the information derived from the other. At larger evolutionary distances (sequence similarity below 20–25%), more sensitive methods, such as the sequence profile methods [1–3], must be used. These methods allow one to detect subtle sequence similarities between proteins. However, three-dimensional (3D) structure comparison methods can eventually detect evolutionary links which are no longer identifiable by the most sophisticated sequence analysis methods, since structure similarity is more conserved through evolution and therefore remains detectable even when the sequence may have changed beyond recognition in the course of evolution.

Sequence comparison methods can take advantage of the enormous amount of data coming from the various

* Corresponding author.

genome sequencing projects and are generally simpler than structure comparison methods. Moreover, the latter can only be applied when protein structures are available, and this is often not the case. The Protein Data Bank (PDB) now contains about 12,000 protein structures, while the protein sequence database contains about 150,000 sequences.

Nevertheless, even if with various ranges of sensitivity and degrees of accuracy, sequence and structure comparison methods are generally able to identify links between proteins depending on their molecular evolution.

The problem of comparing protein surfaces seems to be intrinsically more complicated than sequence or structure comparison and, to date, has attracted fewer scientists for a variety of reasons: (i) the algorithms to compare protein surfaces are generally less simple and slower than their sequence or structure counterparts; (ii) possible findings of protein surface similarity searches risk to be obvious consequences of their sequence and/or structure similarity and, most probably, the great majority of protein surface similar patches will involve cases which could be much more easily detected by less expensive computational methods; (iii) protein surface analysis can be used only when 3D structures are available.

Protein surface comparison is a very challenging problem and findings of such an analysis may highlight proteins that do not share a common ancestor but show the convergence of different sequences or structures into a unique pattern of exposed residues univocally identifying function, independently of the polymer backbone.

Cases have been described where active or binding sites can be encoded in proteins displaying non-homologous structures [46], and cases where proteins with a similar structure display differences in their active or binding sites and therefore do not share the same function. In such cases, sequence or structure comparisons might be misleading and miss a correct analysis of the region involved in the interaction, whereas a protein surface comparison would be more effective (fig. 1).

Moreover, only a protein surface analysis can, in principle, detect convergent evolution involving functional sites which arose independently in evolution. In general, detection of surface similarity between protein molecules may provide us with a better understanding of biomolecular recognition.

We will describe a number of tools which have been developed for protein surface description. Some graphic tools are able to produce meaningful images for visual inspection while others provide numerical descriptions for computational analysis. Many ways of describing protein surfaces have been developed for docking studies (i.e. to study protein-protein interaction) and only occasionally have these been applied to the analysis of protein surface similarity. We will review a few 3D-pattern-matching methods which have been applied to the study of conserved residues in protein structures. Some of these methods can search for sequence-independent patterns and, if the residues are exposed on the protein surface, they may help to identify similar protein surfaces. Finally, we will describe some computational methods specifically developed for the analysis and comparison of protein surfaces and discuss their application to a few biological cases.

3D-pattern matching

An interesting consequence of the increasing availability of 3D structures is the development of 3D-pattern-matching algorithms. Methods allowing the detection of a 3D pattern in a set of proteins may spot active or binding sites common to a set of diverse protein structures, thus identifying similar functions on different folds as well as cases of convergent or divergent evolution. A brief description of some of these methods may contribute considerably to a comprehension of protein surface analysis and comparison techniques.

A 3D-pattern-matching algorithm can be either sensitive or insensitive to the amino acid order in the protein sequence. In general, patterns sharing the same amino acid sequence order are much more likely to have been derived from a common ancestor; therefore, when searching very distant relationships among proteins, a method which only detects patterns that have the same ordering of residues along the sequence is to be preferred. However, a sequence-independent 3D-pattern search is desirable when one is interested in the conservation of active or binding site residues, to find examples of possible convergent evolution.

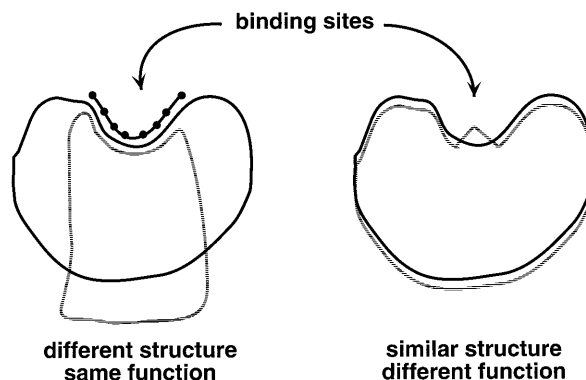

Figure 1. Schematic representation of two proteins with similar function and different structure and of two proteins with similar structure and a different binding site.

We review here a few methods which analyse the conservation of 3D patterns independently of amino acid sequence order.

The graph-theoretic method proposed by Mitchell et al. [4] allows recognition of user-defined spatial motifs in protein structures. It is based on a subgraph-matching algorithm [5, 6] and on a simplified representation of the amino acid side chain positions, in which each side chain is characterized by two pseudo-atoms (considered as the nodes of a graph) while the relative orientations of side chains in space are described by means of distances (the edges of the graph) between the pseudo-atoms. A graph consists of a set of nodes and a set of edges and is used to represent each PDB protein structure, while the query (user-defined) pattern is a subgraph. The Ullman subgraph-isomorphism algorithm [7] is used to determine whether one graph (the query 3D pattern) is actually a subgraph of another, larger graph (a protein structure).

Russell [8] described a method that detects 3D pattern similarities between two protein structures. A protein is represented by a string of residues sorted in alphabetical order by residue type. In a protein, each residue of a particular type is compared to those in a second protein thus generating an exhaustive list of matches. This procedure is iterated for each amino acid type. The aim of this method is to identify groups of residues involved in a common active or binding site. Therefore, there are distance constraints that the amino acids of a matched pair must satisfy: the distances between a residue added to a pattern in a protein and the other residues of the pattern must be similar to the corresponding distances in the second protein. Moreover, to save computing time, amino acids which are not supposed to be directly involved in the function are eliminated: positions in the structure that are not well-conserved across known homologues, amino acids with side chains containing only hydrogen and carbon atoms and residues involved in disulphide bonds are ignored. A final weighted root mean square deviation (RMSD) is calculated for each pair of common side chain patterns identified and its statistical significance is assigned from the analysis of randomly generated side chain patterns.

In general, recognition of spatial motifs can be performed by comparing a motif with a database of structures or, alternatively, a protein structure can be compared with a database of predefined motifs [9]. The latter approach, based on the comparison of a database of motifs to a single structure, depends essentially on the quality of such databases. Residues are represented by the co-ordinates of their alpha carbons and by pseudo-atoms located at the centre of gravity of their side chain atoms. The algorithm calculates the distances between all the pseudo-atoms in a user-defined motif and tries to identify groups of amino acids of the same type and with a similar spatial arrangement in a protein structure. Distances between pseudo-atoms in the query pattern are compared with distances between pseudo-atoms in a database protein and a combination of residues that potentially matches the 3D motif is rejected if distance pairs differ by more than a preset cut-off value. Finally, the query pattern and the corresponding matched set of residues are superimposed and their RMSD calculated.

Protein surface description

Over the years, different methods have been developed to describe a protein surface visually or analytically. We will briefly overview a few methods for protein surface description and, for a more complete and thorough description of the argument, we suggest to refer to works [10–14] or to two web pages dedicated to molecular surface representation [15, 16].

The earliest definition of protein solvent-accessible surface was published by Lee and Richards in 1971 [17]. They described a method which allows one to distinguish surface atoms from buried atoms and to describe the total surface of the molecule as the sum of atomic surfaces. Based on this idea, an algorithm was proposed to compute and display a protein surface [18]. Connolly [19] then proposed an improved algorithm where a protein surface is traced by a water-sized probe sphere rolled over the atoms of the molecule. This representation of protein surface was developed for graphical display and proved to be particularly efficient in visualizing shape complementarity at protein-protein interfaces (fig. 2).

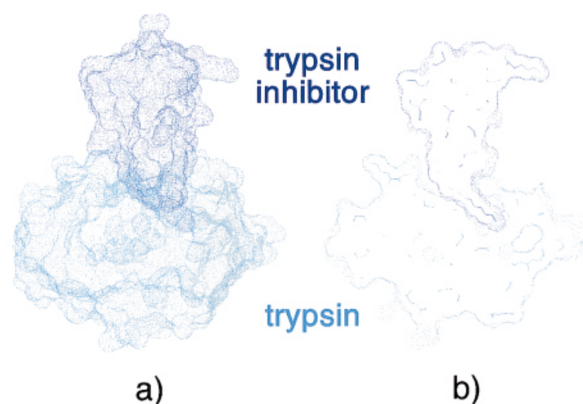

Figure 2. Connolly surface representation of the interface between trypsin and the trypsin inhibitor (PDB code: 2tpi). (*a*) The Connolly surface of the trypsin-trypsin inhibitor complex is shown in cyan and blue, respectively. (*b*) A 2 Å-thick slice of the trypsin-trypsin inhibitor complex highlights the almost perfect complementarity between the surfaces of the two proteins.

Other methods were developed ad hoc to represent a protein surface for protein docking which were used as a starting point to score the solution complexes for protein shape complementarity. A cube representation of molecular surface was also developed [20]. In the first step, the protein surface is represented by surface dots with the attached surface normals. The protein is then immersed in a grid space and the surface dots are converted into surface cubes. Shoichet and Kuntz [21] described protein surfaces as sets of spheres of varying sizes, while Walls and Sternberg [22] used a series of slices cut from the van der Waals surface of a protein. A molecular surface representation was proposed by Lin et al. [23] which consisted of a limited number of critical points disposed at key locations over the surface. Helmer Citterich and Tramontano [24] and Ausiello et al. [25] started from a dot representation of the Connolly surface to describe a protein surface as a set of parallel polygons.

These different surface descriptions are generally able to give a simplified and 'soft' representation of the protein surface shape. A simplified representation proved to be more efficient in protein docking than a more detailed one, because it implicitly accommodates the effects of minor conformational changes upon docking.

In searching for protein surface similarity, one would ideally like to be able to select similar surfaces even when analysing proteins crystallized at different resolutions, solved with different methods, in different experimental conditions and sometimes also crystallized both in their bound and unbound states. The ability to describe a protein surface without too many details is therefore also precious in protein surface similarity searches.

In any case, the above-mentioned methods were developed to describe protein surface shape. Various molecular properties, other than geometric shape (such as atomic charge, electrostatic potential, hydrophobicity), may turn out to be interesting in analysing a protein surface. These additional properties may be encoded into a deeper description of the molecular surface and used to increase the information contained in the description.

One of the most widely used procedures to visualize protein surfaces associated with specific properties is the Grasp program [26]. Grasp's rendered surface can be colour coded by electrostatic potential derived from its internal Poisson-Boltzmann solver or from DelPhi [26–28]. This representation (fig. 3) has become a standard tool in assessing proteins electrostatic character. Surfaces can also be coloured by other properties, such as those of the underlying atoms (e.g. hydrophobicity) or by its own intrinsic properties, such as local curvature.

Surface comparison methods

Since protein surfaces are critically involved in selective binding, recognition and interaction with molecular partners, methods for surface comparison may give new insights into protein function analysis.

The comparison of two structures (and also of two surfaces) generally requires prior identification of the transformation (rotation and translation) able to achieve an optimal superposition between selected atom pairs of the two structures. As a consequence, methods useful for docking or structure comparison may also be used for comparing protein surfaces. However, other methods have been specifically developed for surface similarity analysis.

The choice of the surface representation method is crucial since the surface description must be sufficiently detailed but must not affect the performance of the surface comparison algorithms; most of the methods use surface representation derived from the above-mentioned Connolly program [19, 29].

A method based on Connolly algorithm was proposed by Brickmann's group [30]. They set up a procedure that calculates the local canonical curvature [31, 32] at each surface-representing point, generated from Connolly's method. Moreover, the authors introduce the concept of global curvature to represent the average curvature of a surface region. Global curvature is calculated considering the direct neighbours of a reference point, determined by a triangulation algorithm [33] at a selected distance. Global curvatures can be used to define a 'curvature profile' of a specific region, which

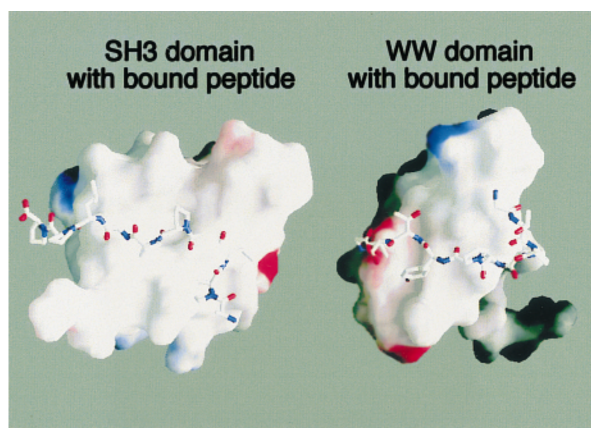

Figure 3. Grasp [26] surface representation of SH3 [51] and WW domains [48] with their bound peptide. The molecular surface is colour coded by electrostatic potential, while the atoms of the bound peptides are displayed as sticks and coloured by atom type. No trivial similarity between the surfaces of the two polyproline-binding domains can be recognized by visual inspection.

can be used to detect surface complementarity and surface similarity.

In their approach, Fischer et al. [34] established a method that can be applied to different issues: (i) backbone comparison; (ii) search for a predefined 3D motif within the full backbone of a domain; (iii) surface comparison; (iv) docking. Indeed, this procedure can be used for searching geometric matches among any given set of points in the space, such as Cα, Cβ or surface atoms; any 3D comparison of structures without any sequence order constraint is therefore allowed. This goal is achieved by the geometric hashing paradigm [35], a technique suitable for detecting matches between a model object and a database of objects. The overall method has three major critical points: (i) finding seed matches: the structures to be compared are covered by a set of balls, then the algorithm based on the geometric hashing paradigm counts the matching pairs of atoms for every rigid transformation; (ii) clustering seed matches: seed matches representing almost identical transformations are clustered; a seed match representing the cluster is obtained by choosing matching pairs that appear at least in a certain percentage of the seed matches that compose the cluster; (iii) extending seed matches: to obtain additional information, one of the structures to be compared is translated and rotated as the seed match representing the cluster, and then other matching atoms are selected.

The geometric hashing paradigm procedure itself is divided into two steps. In the preprocessing stage, the features of the model molecule are converted into a hash-table by means of a mathematical procedure which measures the distance between each atom of the molecule and every other pair of atoms. In the recognition stage, a similar representation is calculated for the target molecule and used to access the hash-table, looking for matches between substructures of the two molecules. If the recognition step generates a sufficient score, the target molecule is superposed to the model by a rigid transformation (translation and rotation), and the number of matching points is expanded exploring the 'neighbours' with a heuristic iterative algorithm.

The method is extended to surface comparison representing surfaces by Connolly's Molecular Surface program and Kuntz's SPHGEN program [36], which reduces the number of surface points and selects local surface invaginations. In comparing structures, co-ordinates of the atoms selected in the clustering procedure of the SPHGEN program are used instead of the co-ordinates of the spheres of the SPHGEN output itself.

Another procedure for protein surface comparison [37] takes advantage of the geometric hashing paradigm and uses the face-centre critical point technique [23, 38], also derived from Connolly's method. The points used in the calculation are the Connolly face-centres, chosen at key locations on the surface of the examined structure. Each Connolly face is substituted by a surface point, its surface normal and the face area; these three items represent a 'critical point'. The algorithm used to compare different groups of 'critical points' is essentially similar to the one previously described.

An alternative approach [39] defines molecular 'skins' for protein surface comparisons or docking studies. A molecular skin is defined as the space between an outer and an inner surface. Once the inner surface has been calculated (for example by means of the van der Waals radii or by identifying the solvent-accessible residues), the outer surface is generated by adding the predefined value of 0.4 Å, corresponding to the thickness of the skin, uniformly to the inner surface. The method compares two molecular surfaces intersecting their molecular skins; the volume of intersection provides a measure of their shape similarity. The overlap is maximized by rigid transformations (rotations and translations).

One of the critical points in surface comparison by superposition of two different proteins is the great number of rigid transformations that need to be evaluated. Poirrette et al. [40] used a genetic algorithm (GA) to detect the best transformation superposing two molecular surfaces. The surface is represented through Connolly's program which reports 3D co-ordinates, a normal vector and an indicator of the surface shape, for each surface point; this work uses an additional parameter, namely an indicator of the ability to form hydrogen bonds. The surfaces represented in this way are inserted into a 3D grid, and a GA is used to generate translations and rotations of the query molecule with respect to a target, and to evaluate the superposition of each transformation. The genetic algorithm is a procedure that generates random solutions to a certain problem as vectors (the so-called chromosomes) and picks out the best ones by a score value (the so-called fitness value); the worst are discarded and substituted by new randomly generated ones. New variability is generated by mutations and recombination between chromosomes, and scores are recalculated. This cycle is repeated a prefixed number of times, or until the score reaches a desired value (or a plateau). In this work, each chromosome contains six randomly generated numbers representing the co-ordinates of a rigid transformation, as rotations and translations in the xyz space. The match between the transformed structure and the target molecule is evaluated by a fitness function that controls whether or not two points are in similar space position and have similar characteristics (Connolly's indicators previously mentioned). This approach has been successfully used to compare binding pockets and binding sites of a number of proteins.

A recent work [41] adopts another type of algorithm, the neural network, to generate a map of protein sur-

faces. For automated detection of the protein surface pockets, a grid with 1 Å spacing is generated around the examined protein and the cells corresponding to the protein/solvent interface are evaluated. For each 'solvent' cell, the program scans the grid around it detecting contiguous clusters defining protein surface pockets. Then, solvent-accessible surfaces are calculated by the Connolly algorithm. An automated procedure assigns an interaction type (aliphatic, hydrogen bond donor, hydrogen bond acceptor, aromatic face, aromatic edge) to each pocket cell on the basis of the amino acid type and its orientation on the surface pocket. This analysis was applied to a training set of 175 zinc-ion-containing protein structures. The dataset was submitted to a Kohonen self-organizing neural network [42], which provides a method to project a high-dimensional space onto a low-dimensional display called the self-organizing map (SOM). In this case, the SOM was a two-dimensional map of 20 rows per 20 columns (rows and columns are called 'neurons'). The resulting SOM represents the distribution of protein surface cavities of the training set of proteins, with adjacent neurons containing similar protein cavity structures. Pockets belonging to each class of the enzymes considered form a distinct cluster in the SOM, and active and inactive pockets are then detected. Furthermore, the trained SOM can be used to make predictions for enzymes not present in the training set.

The Dictionary of Interfaces in Proteins (DIP) is a database of interacting protein surfaces [43]. DIP has a retrieval system to search for complementary molecular surface patches (MSPs) with an automatic superposition procedure; it can also be useful for similar-surface-patch searches. The MSPs are defined as sets of atoms belonging to a given secondary structure in contact with a molecular partner (a ligand, a protein or a solvent molecule); the centres of mass of the MSPs are superimposed to compare two surface patches. One of the patches is then rotated to align the patches along the major directions, and matching atoms are calculated; this step is followed by an expansion of the superposition to the neighboring region [43, 44]. A detailed description of the algorithm has not yet been published.

The 3D profile procedure to describe and compare protein surfaces developed by de Rinaldis et al. [45], is based on a 3D transposition of the profile method for sensitive protein homology sequence searches [1].

In analogy to the profiles derived from the multiple alignment of protein sequences, the 3D profile is derived from a multiple alignment of several protein structures sharing some functional or binding property (e.g. the ability to bind a phosphorylated tyrosine, a polyproline II helix, a nucleotide). Each aligned structure is first transformed into a 'surface' structure, by retaining only the exposed residues, and then immersed in a 3D grid. Each element of the grid is assigned to the residue occupying it. The grids obtained from all the selected superposed proteins are then merged in a unique 3D profile grid, where each element contains information on the superposed residues falling within it and belonging to the different structures.

The elements of the 3D profile grid correspond to the columns of the multiple alignments of protein sequences. It is important to note, however, that the residues contained in the grid elements are aligned and analysed on the protein surface independently from their order in the protein sequence.

The 3D profile obtained from a multiple superposition of protein surfaces sharing a specific function can be considered a 'descriptor' of the specified function and can be used for analysing the pattern of conserved residues associated with a specific function or as a query in searching similar surfaces in the PDB.

Conclusions

One ligand may display affinity for different protein surfaces and similar protein surfaces may bind different ligands [46], thus an ideal one-to-one relationship between surface pockets and ligand cannot always be established. Since the PDB is growing rapidly, many structures are now available crystallized with their ligand and more information can be used to search for similarities among binding sites and to characterize the similarity or diversity of ligand-specific binding sites.

We have analysed a few binding pockets (SH2 and SH3 domains, nucleotide-binding pocket associated with the ploop structure [46, 47] and a few ligands (phosphorylated tyrosines, polyproline II peptides and nucleotides) and tried to tackle the rules of correspondence between the binding and the bound moieties. In each case, we were able to define a unique 3D motif able to select all and only the true positive hits.

We could not define a unique motif able to select all the PDB structures able to bind one specified ligand. The SH3 3D motif does not share similarity with the WW domain or with profilin, even though they also bind proline-rich peptides [45, 48–50] and the phosphotyrosine-binding pocket of SH2 domains is not similar to the PTB (phosphotyrosine-binding) domain structure [45]. These findings seem to indicate that different ways of binding the same ligand may have evolved in different proteins (see also fig. 3).

Nevertheless, analysis of protein surfaces able to bind the same consensus and belonging to the same family may help in understanding the rules governing the specificity of interaction. Comparison of the surfaces of the different SH3 domains may help in constructing a tree where domains with the same or similar specificity of interaction may be clustered together.

On the other hand, analysis of the nucleotide-binding region in the ploop-containing proteins of known structure seems to indicate at least one case of convergent evolution. Functionally important residues, namely positively charged residues aligned to ras K117, are sometimes provided by very different regions of the primary structure of the proteins analysed, though they are perfectly aligned on the protein surfaces [45].

In summary, analysis of protein surfaces may offer new insights into the fascinating world of sequences, structures and functions, and may also offer substantial support in the study of interaction specificity and protein function evolution.

Acknowledgements. The authors are grateful to Dr. Maria Lo Ponte for carefully revising the manuscript. This review was supported by the CNR Biotechnology programme.

1 Gribskov M., Mclachlan A. D. and Eisenberg D. (1987) Profile analysis: detection of distantly related proteins. Proc. Natl. Acad. Sci. USA **84:** 4355–4358
2 Gribskov M. and Veretnik S. (1996) Identification of sequence pattern with profile analysis. Methods Enzymol. **266:** 198–212
3 Altschul S. F., Madden T. L., Schaffer A. A., Zhang J., Zhang Z., Miller W. et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402
4 Mitchell E. M., Artymiuk P. J., Rice D. W. and Willett P. (1990) Use of techniques derived from graph theory to compare secondary structure motifs in proteins. J. Mol. Biol. **212:** 151–166
5 Artymiuk P. J., Poirrette A. R., Grindley H. M., Rice D. W. and Willett P. (1994) A graph-theoretic approach to the identification of three-dimensional patterns of amino acid side-chains in protein structures. J. Mol. Biol. **243:** 327–344
6 Grindley H. M., Artymiuk P. J., Rice D. W. and Willett P. (1993) Identification of tertiary structure resemblance in proteins using a maximal common subgraph isomorphism algorithm. J. Mol. Biol. **299:** 707–721
7 Ullman J. R. (1976) An algorithm for subgraph isomorphism. J. Assoc. Comput. Machin. **16:** 31–42
8 Russel R. B. (1998) Detection of protein three-dimensional side-chain patterns: new examples of convergent evolution. J. Mol. Biol. **279:** 1211–1227
9 Kleywegt G. J. (1999) Recognition of spatial motifs in protein structures. J. Mol. Biol. **285:** 1887–1897
10 Richards F. M. (1977) Areas, volumes, packing and protein structure. Annu. Rev. Biophys. Bioeng. **6:** 151–176
11 Max N. L. (1983) Computer representation of molecular surfaces. IEEE Comput. Graphics Appl. **3:** 21–29
12 Max N. (1984) Computer representation of molecular surfaces. J. Mol. Graphics **2:** 8–13
13 Richards F. M. (1985) Calculation of molecular volumes and areas for structures of known geometry. Methods Enzymol. **115:** 440–464
14 Mezey P. (1990) Molecular surfaces. In: Reviews in Computational Chemistry, pp. 265–294, Lipkowitz K. and Boyd D. (eds), VCH, Weinheim
15 O'Donnell T. J. http://www.netsci.org/Science/Compchem/feature14.html
16 Connolly M. L. http://www.netsci.org/Science/Compchem/feature15.html
17 Lee B. and Richards F. M. (1971) The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. **55:** 379–400
18 Greer J. and Bush B. L. (1978) Macromolecular shape and surface maps by solvent exclusion. Proc. Natl. Acad. Sci. USA **75:** 303–307
19 Connolly M. L. (1983) Analytical molecular surface calculation. J. Appl. Crystallogr. **16:** 548–558
20 Jiang F. and Kim S. H. (1991) "Soft docking": matching of molecular surface cubes. J. Mol. Biol. **219:** 79–102
21 Shoichet B. and Kuntz I. D. (1991) Protein docking and complementarity. J. Mol. Biol. **221:** 327–346
22 Walls P. H. and Sternberg J. E. (1992) New algorithm to model protein-protein recognition based on surface complementarity. J. Mol. Biol. **228:** 277–297
23 Lin S. L., Nussinov R., Fischer D. and Wolfson H. L. (1994) Molecular surface representations by sparse critical points. Proteins Struct. Funct. Genet. **18:** 94–101
24 Helmer Citterich M. and Tramontano A. (1994) PUZZLE: a new method for automated protein docking based on surface shape complementarity. J. Mol. Biol. **235:** 1021–1031
25 Ausiello G., Cesareni G. and Helmer Citterich M. (1997) ESCHER: a new docking procedure applied to the reconstruction of protein tertiary structure. Proteins **28:** 556–567
26 Nicholls A., Sharp K. A. and Honig B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins **11:** 281–296
27 Gilson M. K., Sharp K. A. and Honig B. (1988) Calculating the electrostatic potential of molecules in solution: method and error assessment. J. Comput. Chem. **9:** 327–335
28 Sharp K. A. and Honig B. (1990) Electrostatic interactions in macromolecules: theory and applications. Annu. Rev. Biophys. Biophys. Chem. **19:** 301–332
29 Connolly M. L. (1983) Solvent-accessible surface of proteins and nucleic acids. Science **221:** 709–713
30 Zachmann C. D., Heiden W., Schlenkric M. and Brickmann J. (1992) Topological analysis of complex molecular surfaces. J. Comp. Chem. **13:** 76–84
31 Spanier E. H. (1966) Algebraic Topology. McGraw-Hill, New York
32 Kay D. C. (1988) Theory and Problems of Tensor Calculus. McGraw-Hill, New York
33 Heiden W., Schlenkrich M. and Brickmann J. (1990) Triangulation algorithms for the representation of molecular surface properties. J. Comp. Aided Mol. Design **4:** 255–269
34 Fischer D., Norel R., Wolfson H. and Nussinov R. (1993) Surface motifs by a computer vision technique: searches, detection, and implications for protein-ligand recognition. Proteins **16:** 278–292
35 Nussinov R. and Wolfson H. J. (1991) Efficient detection of three-dimensional structural motifs in biological macromolecules by computer vision techniques. Proc. Natl. Acad. Sci. USA **88:** 10495–10499
36 Kuntz I. D., Blaney J. M., Oatley S. J., Langridge R. and Ferrin T. E. (1982) A geometric approach to macromolecule-ligand interactions. J. Mol. Biol. **161:** 269–288
37 Rosen M., Lin S. L., Wolfson H. and Nussinov R. (1998) Molecular shape comparisons in searches for active sites and functional similarity. Protein Eng. **11:** 263–277
38 Lin S. I. and Nussinov R. (1996) Molecular recognition via face center representation of a molecular surface. J. Mol. Graphics **14:** 78–90
39 Masek B. B., Merchant A. and Matthew B. (1993) Molecular skins: a new concept for quantitative shape matching of a protein with its small molecule mimics. Proteins Struct. Funct. Genet. **17:** 193–202
40 Poirrette A. R., Artymiuk P. J., Rice D. W. and Willett P. (1997) Comparison of protein surfaces using a genetic algorithm. J. Comp. Aided Mol. Design **11:** 557–569
41 Stahl M., Taroni C. and Schneider G. (2000) Mapping of protein surface and prediction of enzyme class by a self-organizing neural network. Protein Eng. **13:** 83–88
42 Kohonen T. (1982) Self-organized formulation of topologically correct feature maps. Biol. Cybern. **43:** 59–69
43 Preißner R., Goede A. and Frömmel C. (1998) Dictionary of interfaces in proteins (DIP): data bank of complementary molecular surface patches. J. Mol. Biol. **280:** 535–550

44 Preißner R., Goede A. and Frömmel C. (1999) Homonyms and synonyms in the dictionary of interfaces in proteins (DIP). Bioinformatics **15:** 832–836
45 Rinaldis M. de, Ausiello G., Cesareni G and Helmer-Citterich M. (1998) Three-dimensional profiles: a new tool to identify protein surface similarities. J. Mol. Biol. **284:** 1211–1221
46 Kauvar L. M. and Villar H. O. (1998) Deciphering cryptic similarities in protein binding sites. Curr. Opin. Biotechnol. **9:** 390–394
47 Via A., Ferré F., Brannetti B., Valencia A. and Helmer-Citterich M. (2000) Three-dimensional view of the surface motif associated with the P-loop structure: *cis* and *trans* cases of convergent evolution. J. Mol. Biol., in press
48 Pawson T. and Schlessinger J. (1993) SH2 and SH3 domains. Curr. Biol. **3:** 434–442
49 Macias M. J., Hyvonen M., Baraldi E., Schultz J., Sudol M., Saraste M. et al. (1996) Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. Nature **382:** 646–649
50 Sudol M. (1996) The WW module competes with the SH3 domain? Trends Biochem. Sci. **21:** 161–163
51 Musacchio A., Willmanns M. and Saraste M. (1994) High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. Nat. Struct. Biol. **1:** 546–551