PROTEINS: Structure, Function, and Genetics 37:565–575 (1999)

Hydrophobicity at the Surface of Proteins

Marco Scarsi, Nicolas Majeux, and Amedeo Caflisch*
Department of Biochemistry, University of Zürich, Zürich, Switzerland

ABSTRACT **A new method is presented to quantitatively estimate and graphically display the propensity of nonpolar groups to bind at the surface of proteins. It is based on the calculation of the binding energy, i.e., van der Waals interaction plus protein electrostatic desolvation, of a nonpolar probe sphere rolled over the protein surface, and on the color coding of this quantity on a smooth molecular surface (hydrophobicity map). The method is validated on ten protein–ligand complexes and is shown to distinguish precisely where polar and nonpolar groups preferentially bind. Comparisons with existing approaches, like the display of the electrostatic potential or the curvature, illustrate the advantages and the better predictive power of the present method. Hydrophobicity maps will play an important role in the characterization of binding sites for the large number of proteins emerging from the genome projects and structure modeling approaches. Proteins 1999;37:565–575.**

Key words: hydrophobicity map; continuum electrostatics; molecular surface; solvent accessible surface; Poisson equation; thrombin-NAPAP; farnesyltransferase-farnesylpyrophosphate; MDM2-p53; HoxB1-Pbx1; p11-annexin II

INTRODUCTION

The accurate prediction of the binding modes of nonpolar molecules to proteins in a polar solvent is useful for ligand design and docking,[1] and to understand molecular recognition.[2] Hydrophobicity literally indicates the reluctance of nonpolar compounds to dissolve in water. Consequently, the hydrophobicity of (part of) the surface of a molecule is considered as the propensity to bind nonpolar groups. Accordingly, "hydrophobic association" indicates the binding of nonpolar groups to an hydrophobic part of the surface. Inversely, the regions where the association of nonpolar compounds is less favorable are termed hydrophilic. Previous studies[2–5] have suggested that hydrophobicity is a determinant factor in the association of ligands to proteins (here referred to as receptors). The electrostatic desolvation of the receptor as well as the nonelectrostatic terms (i.e., dispersion forces and perturbation of water structure at the binding interface) contribute to direct the association of a nonpolar compound at the surface of a receptor. Ligand–receptor electrostatic intermolecular interactions do not play a significant role in the case of a nonpolar ligand.

Different quantities are often used to predict the hydrophobicity at the surface of a receptor. One approach is based on the evaluation of the local receptor surface curvature.[6] The curvature models the nonelectrostatic contributions to association. The more concave is a part of the receptor surface, the more favorable are these contributions to ligand binding. Receptor electrostatic desolvation is not accounted for. As a consequence, the curvature does not distinguish between pockets lined by charged and uncharged groups, the former ones being less prone to bind nonpolar groups. A second approach makes use of the electrostatic potential around the receptor in solution calculated via a finite-difference solution of the Poisson equation.[7–10] The hydrophobic zones can then be identified as those parts of the surface where the potential is close to zero (provided that the potential has been set to zero at infinite distance). Alternatively, the electrostatic potential can be used in combination with the surface curvature to distinguish between polar and nonpolar concave zones. This distinction is often only qualitative because in general it is not clear how to weight the contribution of the potential with the one of the curvature. Moreover, the electrostatic potential describes the magnitude of the electrostatic intermolecular energy between the receptor and a hypothetic ligand, but not the receptor electrostatic desolvation. The latter acts on a much shorter range, as shown below at the end of Binding Energy of the Nonpolar Probe Sphere. Hence, while the electrostatic potential is very valuable in distinguishing between binding modes of different charged and polar compounds to the receptor, it is not always useful in discriminating between zones which favor the binding of nonpolar or polar compounds.

A new method (here termed hydrophobicity map) to calculate and visualize the hydrophobicity at the surface of a protein is presented in this paper. The first step consists of the evaluation of the binding energy of a nonpolar probe sphere to the receptor. The binding energy includes both the electrostatic and the nonelectrostatic contributions to the association of a nonpolar compound at the surface of a receptor. A continuum approach is used for the electrostatic component, whereas the van der Waals interaction describes the nonpolar contribution. In a second step, the

Grant sponsor: Swiss National Science Foundation; Grant number: 31-5360498; Grant sponsor: Olga Mayenfisch Foundation; Grant sponsor: Helmut Horten Foundation.
The program for the calculation of the hydrophobicity maps is available from the corresponding author.
*Correspondence to: Amedeo Caflisch, Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. E-mail: caflisch@bioc.unizh.ch
Received 29 April 1999; Accepted 2 July 1999

binding energy is displayed by color-rendering on the surface of the receptor. This yields a precise visualization of the surface hydrophobicity as well as a clear distinction between hydrophobic and hydrophilic zones in close proximity. At present, the binding energy of the nonpolar probe sphere is used also in the docking program SEED[11] to direct the docking of nonpolar molecular fragments (e.g., propane, benzene, cyclohexane).

THEORY AND METHODS

Binding Energy of the Nonpolar Probe Sphere

The center of a sphere rolling over the van der Waals (vdW) surface of a molecule spans the solvent accessible surface (SAS).[12] In the present approach, a number of points are distributed uniformly on the SAS of the receptor with a given surface density (input value) to describe in a discrete manner the different positions of the center of the probe sphere. The binding energy of the probe sphere to the receptor (ΔE) is assumed to consist of two terms, the vdW interaction energy (E^{vdW}) and the electrostatic desolvation of the receptor (ΔE^{desolv}):

$$\Delta E = E^{vdW} + \Delta E^{desolv} \tag{1}$$

vdW parameters and partial charges from the all-hydrogen MSI CHARMm22 parameter set[13,14] were used in all test cases.

The evaluation of ΔE for about 55,000 positions of the probe sphere on the thrombin surface (see Results) requires about 35 s on an SGI workstation (processor R10000, clock frequency 195 MHz). No particular attention was given to the efficiency of the program up to date, and a reduction of the computation time will be addressed in the future without major losses in accuracy.

Van der Waals interaction energy

The nonelectrostatic contributions to binding are a fine balance between the solute–solute vdW interactions (favorable to binding), the solute–solvent vdW interactions (unfavorable) and the disruption of water structure that is an entropic effect at room temperature[6] (favorable). A number of approaches have been proposed to describe these contributions.[6,15–19]

Interestingly, models resulting from different assumptions (e.g., nonelectrostatic contributions represented by solute–solute vdW interactions or by a buried surface curvature-dependent model) can lead to the same predictions,[18] partly depending on the model parameterization. In the case of thrombin, the values of the molecular surface curvature calculated according to Nicholls et al.[6] correlate with the vdW interaction energy of a probe sphere rolled over the same surface (correlation coefficient of 0.85, data not shown). Here it is assumed that solute–solvent vdW interactions and disruption of water structure compensate each other, (see Caflisch et al.[19] and Fig. 6 of Vorobyev et al.[20]), and that solute–solute vdW interactions can account for the nonelectrostatic part of the binding energy. Therefore, the vdW energy between the probe sphere and the receptor atoms (E^{vdW}) is assumed to account for all the nonelectrostatic contributions to the association of the probe sphere to the receptor. It is calculated as:

$$E^{vdW} = \sum_{i \in \text{receptor}} \sqrt{\varepsilon_i \varepsilon_{probe}} \cdot \left\{ \left(\frac{R_i + R_{probe}}{r_i} \right)^{12} - 2 \left(\frac{R_i + R_{probe}}{r_i} \right)^{6} \right\} \tag{2}$$

where r_i is the distance between the receptor atom i and the probe sphere. ε_i and R_i are the vdW energy minimum and radius of atom i. The probe sphere vdW radius (R_{probe}) and energy minimum (ϵ_{probe}) are input values. In the calculations presented in Results, they have been set to 1.4 Å and 0.008 kcal/mol, respectively. Since the probe sphere is rolled over the receptor vdW surface, it does not clash with it and is always at optimal vdW distance from at least one receptor atom.

Electrostatic desolvation energy

The electrostatic desolvation of the receptor accounts for the loss of receptor–solvent favorable electrostatic interactions due to the removal of part of the highly polarizable solvent to accommodate a nonpolarizable probe sphere. This contribution always disfavors association and can be calculated within the assumption of continuum electrostatics.[7,21–27] The system is partitioned into solvent and solute regions, and two different dielectric constants are assigned to each region. The electrostatic energy E of the receptor in solution can be expressed in terms of the electric displacement vector $\vec{D}(\vec{x})$ and of a location dependent dielectric constant $\epsilon(\vec{x})$ as an integral over the three-dimensional space R^3:[28]

$$E = \frac{1}{8\pi} \int_{R^3} \frac{\vec{D}^2(\vec{x})}{\epsilon(\vec{x})} \, d^3x \tag{3}$$

Since $\vec{D}(\vec{x})$ is additive, for point charges it can be rewritten as a sum over all charges i of the receptor:

$$\vec{D}(\vec{x}) = \sum_i \vec{D}_i(\vec{x}) \tag{4}$$

For what concerns the electrostatics, docking an uncharged sphere at the surface of the receptor has the only effect of modifying the dielectric properties in the space occupied by the sphere. Over this volume, the dielectric constant changes from the solvent value (ϵ_w) to the interior value (ϵ_p). In the limit in which $\vec{D}(\vec{x})$ does not change significantly upon docking of the sphere, the variation of the electrostatic energy of the receptor (i.e., the desolvation) can be written according to Eq. (3) as an integral over the volume occupied by the probe sphere (V_{probe}):

$$\Delta E^{desolv} = \frac{\tau}{8\pi} \int_{V_{probe}} \vec{D}^2(\vec{x}) \, d^3x \tag{5}$$

where $\tau = 1/\epsilon_p - 1/\epsilon_w$. In the Results the radius of the probe sphere was assumed to be equal to its vdW radius (1.4 Å) augmented by 0.3 Å to include small voids between the probe and receptor surfaces. A three-dimensional (3D) grid is built around the receptor, and Eq. (5) becomes

$$\Delta E^{desolv} = \frac{\tau}{8\pi} \sum_{k \in V_{probe}} \vec{D}^2(\vec{x}_k) \Delta V_k \tag{6}$$

where the index k runs over the grid points occupied by the probe sphere. The electric displacement of every charge of the receptor can be approximated by the Coulomb field[24,26,29]:

$$\vec{D}(\vec{x}) = \sum_i q_i \frac{(\vec{x} - \vec{x}_i)}{|\vec{x} - \vec{x}_i|^3} \tag{7}$$

where $\vec{x}_i$ is the position of atom i and q_i its partial charge. Equation (7) is an analytical approximation of the total electric displacement and fulfills the condition of validity of Eq. (5) and (6), i.e., $\vec{D}(\vec{x})$ is independent of the dielectric environment. The receptor desolvation in the Coulomb field approximation results from Eq. (7) together with Eq. (6):

$$\Delta E^{desolv} = \frac{\tau}{8\pi} \sum_{k \in V_{probe}} \left(\sum_i q_i \frac{(\vec{x}_k - \vec{x}_i)}{|\vec{x}_k - \vec{x}_i|^3} \right)^2 \Delta V_k \tag{8}$$

The accuracy of this approximation was tested by comparison with finite difference solutions of the Poisson equation for a set of small molecules (including acetate ion, benzoate ion, methylsulfonate ion, methylammonium ion, methylguanidinium ion, 2,5-diketopiperazine, and benzene) distributed over the binding site of thrombin for a total of 1,025 receptor–ligand complexes (Fig. 1).[11] The agreement between Eq. (8) and the solution of the Poisson equation is remarkable (correlation coefficient 0.93, slope of the fitting line 0.78). All these calculations, as well as the ones of the Results were performed with an interior dielectric of 4, solvent dielectric of 78.5, and grid spacing of 0.5 Å.

The desolvation of a charged ion by a small nonpolar sphere at a distance r from the ion varies approximately as $1/r^4$ [Eq. (8)]. This is a very short range effect if compared with the electrostatic potential generated by the same ion that varies as $1/r$. For this reason the potential alone cannot properly describe electrostatic desolvation.

Displaying the Hydrophobicity on the Receptor Surface

The hydrophobicity is color-displayed over the molecular surface (MS),[30] which is the contact + reentrant surface spanned by the probe sphere rolling over the receptor. The MS is preferred to the SAS because it gives a more precise description of the small details at the surface of the receptor. A smooth MS covering the receptor is generated via the molecular graphics package GRASP[6] as an ensemble of triangles. The hydrophobicity at each vertex can be calculated as the value of the binding energy of the probe sphere in the closest position to the vertex. This value is then visually displayed with the help of colors ranging from green (hydrophobic) through white (intermediate) to blue (hydrophilic). The displayed binding energies are restricted in the range from −0.20 to 0.36 kcal/mol in order to provide a sufficient and clear color contrast. This implies that all the values lower than −0.20 are colored dark green, and all the values higher than 0.36 dark blue. In the case of thrombin the lowest and the highest binding energies are −0.55 and 3.44 kcal/mol, respectively.

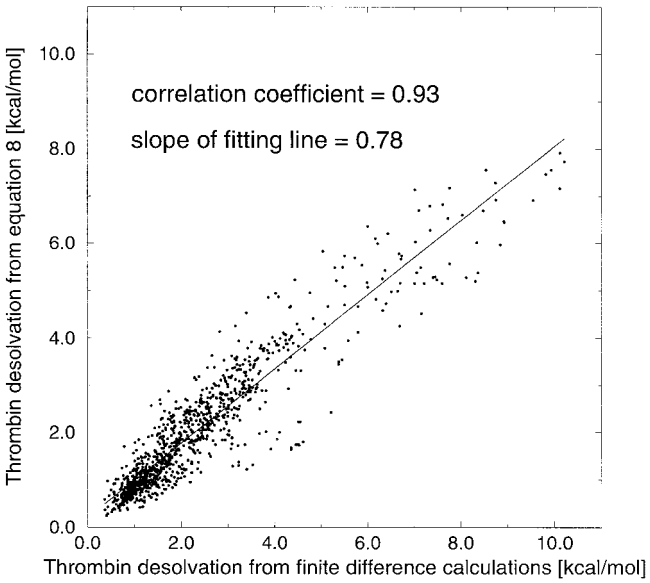


Fig. 1. Correlations in the thrombin desolvation energies calculated by finite difference solution of the Poisson equation (*x*-axis) and Eq. (8) (*y*-axis). Values are plotted for 1,025 complexes of thrombin with small molecules. An interior dielectric of 4, solvent dielectric of 78.5, and grid size of 0.5 Å were used. The finite difference calculations were performed with the program UHBD.[59–61]

RESULTS

Ten complexes are analyzed with the help of the hydrophobicity maps to test their ability in predicting hydrophobic association. Comparisons are made with existing methods based on the electrostatic potential and on the surface curvature. The fraction of the most hydrophobic points on the surface of the receptor that is buried is shown in Table I. All the pictures were generated with GRASP.[6]

Thrombin-NAPAP

Thrombin is one of the best characterized enzymes from a structural point of view. It binds a series of diverse inhibitors without major rearrangements of its conformation.[31–35] The active site presents hydrophobic and hydrophilic parts relatively close in space. Its S3 and S2 precleavage subpockets have hydrophobic character, whereas at the bottom of the S1 or recognition pocket the carboxyl group of Asp189 is a salt bridge partner for basic side chains. *N*α-((2-naphthylsulfinyl)glycyl)-DL-*p*-amidinophenylalanylpiperidine (NAPAP) is an archetypal active

TABLE I. Hydrophobic Regions Buried at the Binding Interface

Complex (PDB code)	N_{100}[a]	N_{1000}[b]	N_{SAS}[c]
Thrombin ÷ NAPAP (1ets)	13	118	54938
FTase ÷ FPP (1ft2)	0	51	105357
MDM2 ÷ p53 (1ycq)	69	353	21129
Pbx1 ÷ HoxB1 (1b72)	30	219	20475
Pbx1 ÷ DNA (1b72)	33	171	20475
p11 ÷ annexin II N-terminal peptide (1bt6)	7	277	41688
Leucine zipper (2zta)	74	536	13179
DB3 antibody Fv ÷ progesterone (1dbb)	34	282	41313
M02/05/01 antibody Fab ÷ traseolide	23	216	74502
Kallikrein ÷ hirustasin (1hia)	25	202	39251
HIV1 protease ÷ acetyl-pepstatin (5hvp)	40	279	36993

[a] N_{100} indicates how many of the 100 most hydrophobic points on the SAS of the receptor are buried at the binding interface.
[b] N_{1000} indicates how many of the 1,000 most hydrophobic points on the SAS of the receptor are buried at the binding interface.
[c] Total number of points on the SAS of the receptor.

site inhibitor of thrombin. It fills the S3 and S2 pockets with its naphthalene and piperidine groups, respectively (Fig. 2a). Moreover it is anchored by its basic group (benzamidine) into S1 to form a salt bridge with Asp189.[36]

Figure 2a shows the hydrophobicity map of the nonprime region of the thrombin active site (PDB code 1ets). NAPAP and Asp189 are also shown. S3 and part of S2 are identified as hydrophobic, while S1 shows an hydrophilic character. The naphthalene and piperidine groups of NAPAP are in contact with hydrophobic zones and bury 13 of the 100 most hydrophobic points (over a total of about 55,000) on the SAS of thrombin (Table I). The polar groups of NAPAP bind to hydrophilic zones. The energy loss of removing water from the hydrophilic zones is compensated upon binding by favorable electrostatic ligand–receptor interactions. S1, despite being a narrow concave cavity, is identified as hydrophilic, since electrostatic desolvation of Asp189 dominates over the favorable vdW interactions between the probe sphere and the surrounding thrombin atoms. The curvature mapped on the MS (Fig. 2b) does not take into account thrombin electrostatic desolvation and suggests that the bottom of the S1 pocket is the most hydrophobic region, in contrast with the actual binding mode of NAPAP.

Farnesyltransferase-farnesylpyrophosphate

Farnesyltransferase (FTase) catalyzes the transfer of the hydrophobic farnesyl group from farnesylpyrophosphate (FPP) to protein substrates with a carboxy-terminal CaaX sequence motif, where C is the cysteine at which the hydrophobic farnesyl group is transferred, "a" is an aliphatic amino acid, and X is generally a methionine. FTase is a target for the development of new anticancer therapies. It was shown that inhibition of FTase causes tumor regression in animal models.[37–39]

The complex FTase-FPP (PDB code 1ft2)[40] is shown in Figure 3A, B, and C with color coding of FTase by the hydrophobicity map, the curvature, and the electrostatic potential, respectively. Only the hydrophobicity map suggests clearly that the hydrophobic moiety of FPP binds in the lower part of the pocket, and the diphosphate moiety of FPP in the upper part. The hydrophobic moiety of FPP buries 51 of the 1,000 most hydrophobic points (over a total of about 105,000) on the SAS of FTase, while it does not bury any of the 100 most hydrophobic points (Table I). These numbers, relatively low if compared with the ones originating from the other complexes, are probably due to the fact that FPP binds FTase together with the hydrophobic CaaX sequence of the protein substrate which may occupy the hydrophobic regions at the bottom of Figure 3A. In Figure 3B the highly positive potential resulting from a zinc ion and from charged residues surrounding the diphosphate moiety of FPP tends to identify the whole pocket as favorable to bind negatively charged groups. From a comparison between Figure 3A and C, it is evident that the electrostatic potential has a longer range than the electrostatic desolvation, as mentioned at the end of Binding Energy of the Nonpolar Probe Sphere. The curvature does not distinguish the lower part of the binding pocket from the upper one (Fig. 3B), where electrostatic desolvation is dominant. Also, a combined analysis of the potential and the curvature would not easily allow a distinction between the hydrophobic and the hydrophilic parts of the FTase binding site.

MDM2-p53

The MDM2 oncoprotein is a cellular inhibitor of the p53 tumor suppressor. It can bind the transactivation domain of p53 and downregulate its ability to activate transcription.[41] Figure 4 shows the 109-residue amino-terminal domain of MDM2 bound to a 15-residue transactivation domain peptide of p53 (PDB code 1ycq).[42] The p53 peptide binds as an amphipathic α-helix by burying its nonpolar side chains in a deep hydrophobic cleft of MDM2.

The MS of MDM2 colored according to the hydrophobicity allows to easily identify the actual binding site as the most probable region where the hydrophobic face of the amphipathic α-helix can bind. It is remarkable that 69 of the 100 most hydrophobic points (over a total of about 21,000) on the SAS of MDM2 are buried by the Phe19, Trp23, and Leu26 side chains of p53 (Table I).

HoxB1-Pbx1 Heterodimer Bound to DNA

Hox homeodomain proteins regulate the development in a variety of organisms in that they specify patterning of the embryo.[43,44] Most of the vertebrate Hox proteins bind DNA as heterodimers with the Pbx1 homeodomain protein. The crystal structure of a human HoxB1-Pbx1 heterodimer bound to DNA (PDB code 1b72)[45] is analyzed. HoxB1 and Pbx1 bind to overlapping binding sites located on opposite faces of the DNA. An hexapeptide of HoxB1 (Thr-Phe-Asp-Trp-Met-Lys) mediates the heterodimeric

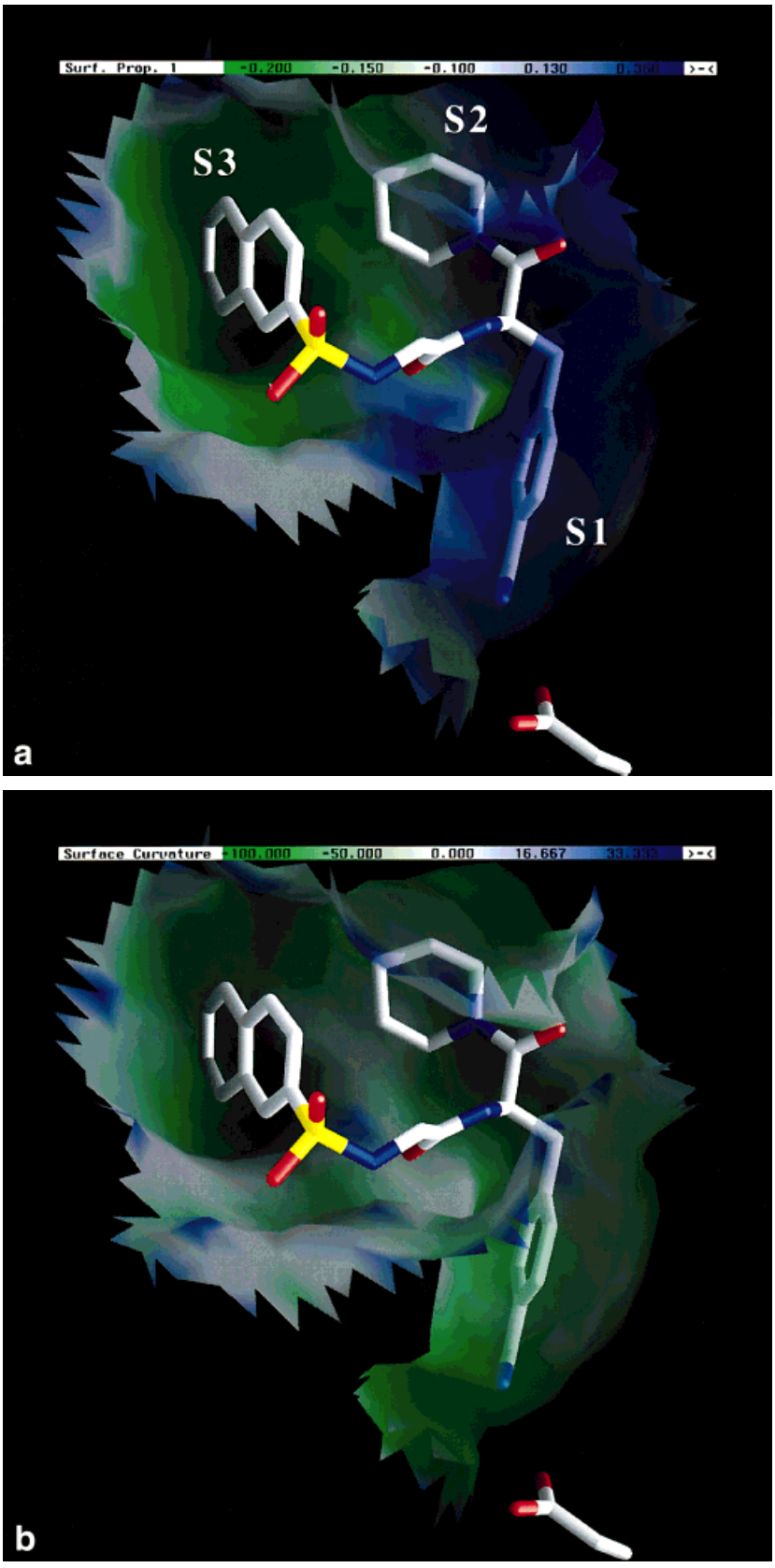
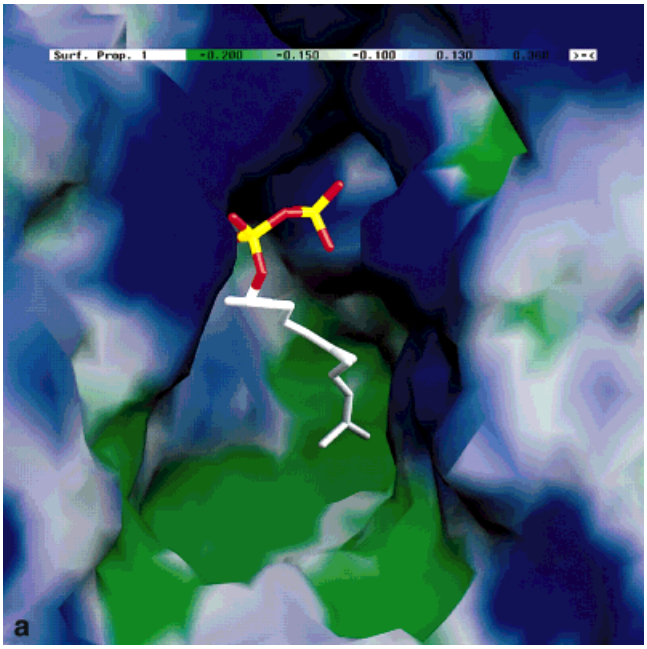
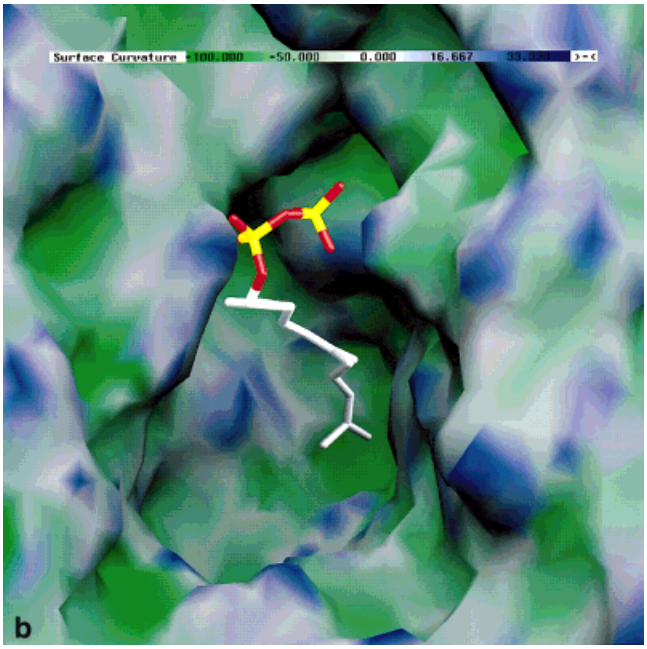


Fig. 2. Thrombin–NAPAP complex. The transparent MS of the thrombin active site is displayed together with the side chain of Asp189 and NAPAP in a cylinder model. **a:** The MS is colored according to the hydrophobicity map. The colored bar on the top indicates the binding energy of the probe sphere in kcal/mol. **b:** The MS is colored according to the curvature as described by Nicholls et al.[6] The default range of colors gray–white–green has been exchanged with green–white–blue. The colored bar on the top indicates the value of the curvature in arbitrary units. A value of 100 would imply that the surface point is completely accessible (limiting case), while a value of −100 would imply that the probe sphere used to evaluate the curvature is completely buried (e.g., in a cavity).





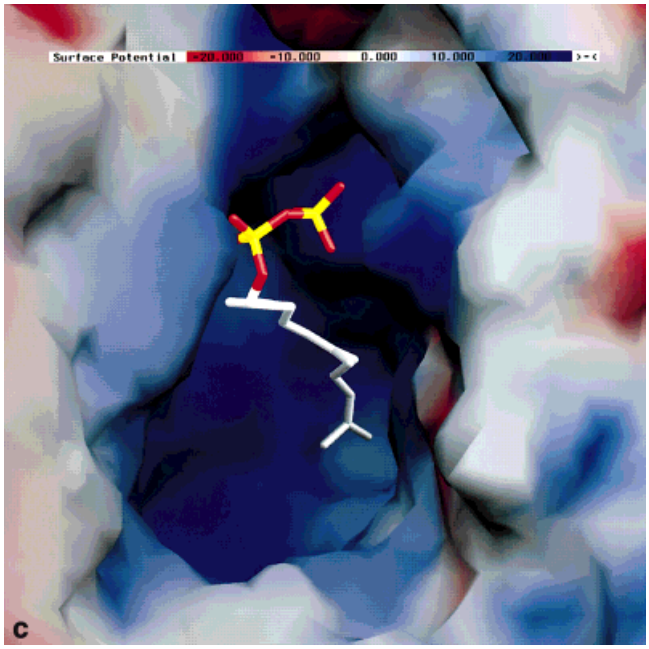


Fig. 3. FTase–FPP complex. **A:** Same as in Figure 2A. **B:** Same as in Figure 2B. **C:** The MS of FTase is colored according to the electrostatic potential obtained by finite-difference solution of the Poisson equation (calculation made with GRASP[6] with an interior dielectric of 4 and a solvent dielectric of 78.5). The colored bar on the top indicates the value of the electrostatic potential generated by the partial charges of the FTase atoms in units of k_BT where k_B is the Boltzmann constant and T is the temperature. Potential values are restricted in the range from -20 k_BT to 20 k_BT.

contacts. It binds in a pocket of Pbx1 via hydrophobic contacts involving the side chains of Phe, Trp, and Met.

Figure 5a shows the contact between Pbx1 and the HoxB1 hexapeptide. The surface of Pbx1 is colored according to the hydrophobicity map. The coloring allows to identify correctly the actual binding site as a markedly hydrophobic pocket. The hexapeptide side chains Phe, Trp, and Met bury 30 of the 100 most hydrophobic points (over a total of about 20,000) on the SAS of Pbx1 (Table I). An additional hydrophobic region on the surface of Pbx1 is visible on the bottom right of Figure 5A. It is probably involved in the association of a Pro and Ala side chains that precede the hexapeptide sequence.[45]

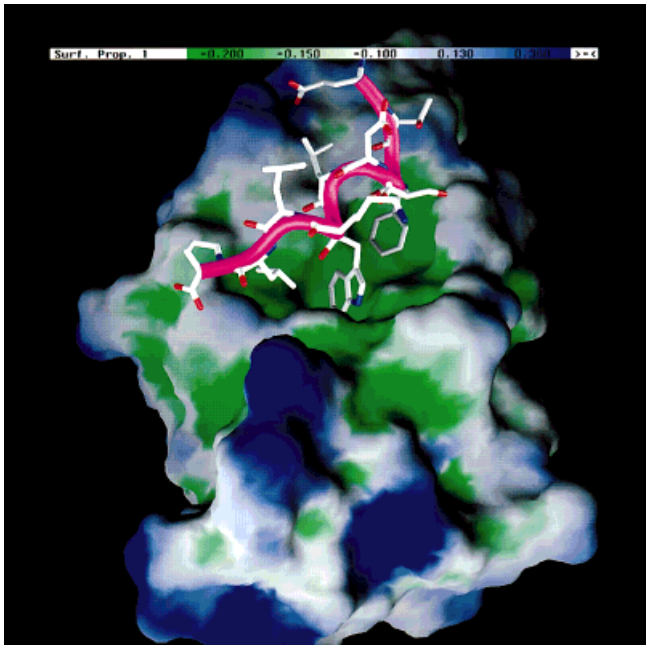


Fig. 4. Same as in Figure 2A for the MDM2–p53 complex. The backbone of the p53 peptide is shown as a magenta ribbon.

Figure 5B shows the contact surface of Pbx1 with the DNA. Most of the contacts are made by hydrophilic regions. The only exception is a small hydrophobic region of the Pbx1 surface which is involved in a contact with a deoxyribose belonging to the DNA. This ring buries 33 of the 100 most hydrophobic points on the SAS of Pbx1 (Table I).

p11-Annexin II N-Terminal Peptide

The p11 protein, known also as calpactin I light chain, is a member of the S100 family. The S100 proteins, apart from p11, bind calcium with different binding affinities.[46] Specific mutations prevent p11 from calcium binding.[47] In vivo p11 forms a complex with annexin II,[48] which can bind negatively charged phospholipids in a calcium-dependent manner.[49] The functions of annexins have been proposed to consist of membrane traffic, membrane-cytoskeleton contacts and ion current across membranes.[49] p11 binding to annexin II results in the formation of an $(\text{annexin II})_2(\text{p11})_2$ heterotetramer, and annexin phospholipid binding is increased by p11 association.[50] The structural reason for this behavior has not been clarified. Both the structure of p11 uncomplexed (PDB code 1a4p) and in complex with annexin II (tetrameric form) (1bt6) have been solved.[51] The N-terminal decapeptide of each annexin II unit binds to the p11 dimer in an amphipathic α-helical conformation, which occupies an hydrophobic cleft formed by two antiparallel helices from two different p11 units.

Figure 6A shows one of the two (symmetric) contacts between the p11 homodimer and the annexin II N-terminal decapeptide. The binding site has an hydrophobic character that favors the burying of the Val, Ile, and Leu side chains on the hydrophobic face of the annexin II helical peptide. As shown in Table I, the fraction of the 100 most hydrophobic points on the SAS of p11 that is buried by annexin II is relatively low, while the fraction of the 1,000 most hydrophobic points is comparable to similar complexes. This indicates that the association is hydrophobic but, probably for structural reasons, the intermolecular contacts do not involve the most hydrophobic region of the p11 dimer. A large part of the 100 most hydrophobic points is grouped in two clusters far away from the annexin II binding site. One of the two clusters is located between the two C-terminal anti-parallel α-helices of p11 (Fig. 6b) and contains 31 hydrophobic points, while the second cluster is between the two N-terminal anti-parallel α-helices and contains 38 hydrophobic points. The cleft between the C-terminal α-helices is formed by the side chains of Phe72, Ala76, and Ile80 from both p11 chains and can accommodate a large hydrophobic compound (Fig. 6b). This could explain the observed increase in phospholipid binding affinity of the $(\text{annexin II})_2(\text{p11})_2$ tetramer with respect to annexin II alone.[50]

Other Complexes

Hydrophobicity maps were calculated and displayed also for the following complexes: leucine zipper (PDB code 2zta),[52,53] progesterone and DB3 antibody Fv (1dbb),[54] traseolide and M02/05/01 antibody Fab,[55] hirustasin and kallikrein (1hia),[56] acetyl-pepstatin and HIV1 protease (5hvp).[57] In all cases, the regions where the nonpolar parts of the ligands bind are correctly predicted by the hydrophobicity maps. The largest fraction of the most hydrophobic points buried is found for the leucine zipper (Table I). This is due to the relatively large contact surface between the two amphipathic α-helices which interact exclusively by hydrophobic contacts. In this case, the curvature analysis[6] does not discriminate between the hydrophilic and hydrophobic face of the two zipper helices. For the two antibody–antigen complexes, considering the small dimensions of the antigens (23 heavy atoms in both cases), the fraction of buried hydrophobic points is remarkably high. Since the remaining hydrophobic points are not clustered in any particular region of the antibody, the identification of the binding site is immediate, given the hydrophobic character of the antigens. For the hirustasin–kallikrein and acetyl-pepstatin–HIV1 protease complexes, where the ligands interact with the receptors via both polar and nonpolar contacts, the hydrophobicity map, unlike the curvature and the potential, distinguish clearly between regions prone to form polar and nonpolar contacts. The values in Table I suggest that hydrophobic contacts are determinant for all the ten complexes examined.

CONCLUSIONS

A new method was presented to evaluate the hydrophobicity at the surface of a protein, which is considered as the local propensity to form complexes with nonpolar molecules. It is based on the estimation of the binding energy between a nonpolar probe sphere and the protein itself. This includes the main contributions directing the association of a nonpolar compound to a receptor, i.e., electrostatic desolvation and vdW interactions. The evaluation of the binding energy is accurate (within the continuum approximation of electrostatics) without requiring long computation times. Hydrophobicity maps, i.e., the color rendering of the binding energy on a smooth MS covering the protein, allow a clear and intuitive visualization of the regions where nonpolar moieties of the ligand preferably bind.

The ability of the hydrophobicity maps to help in the prediction of binding modes of nonpolar groups was tested extensively for a set of nonpolar fragments and thrombin[11] and was validated in this paper for ten different protein–ligand complexes. In five of these complexes, the ligand is a natural peptide or protein, while in the other five it is an organic compound with hydrophobic moieties. The thrombin–NAPAP, FTase–FPP, MDM2–p53, HoxB1-Pbx1–DNA, and p11–annexin II complexes were discussed in detail. In all cases, the hydrophobicity maps are in agreement with the association of nonpolar groups to the receptor. The present approach allows easily to discriminate between hydrophobic and hydrophilic surface regions that are close in space. For the thrombin–NAPAP and FTase–FPP complexes, it was shown that existing approaches based on the analysis of the surface curvature and/or the electrostatic potential are not as valuable in distinguishing regions where nonpolar and polar groups can bind.

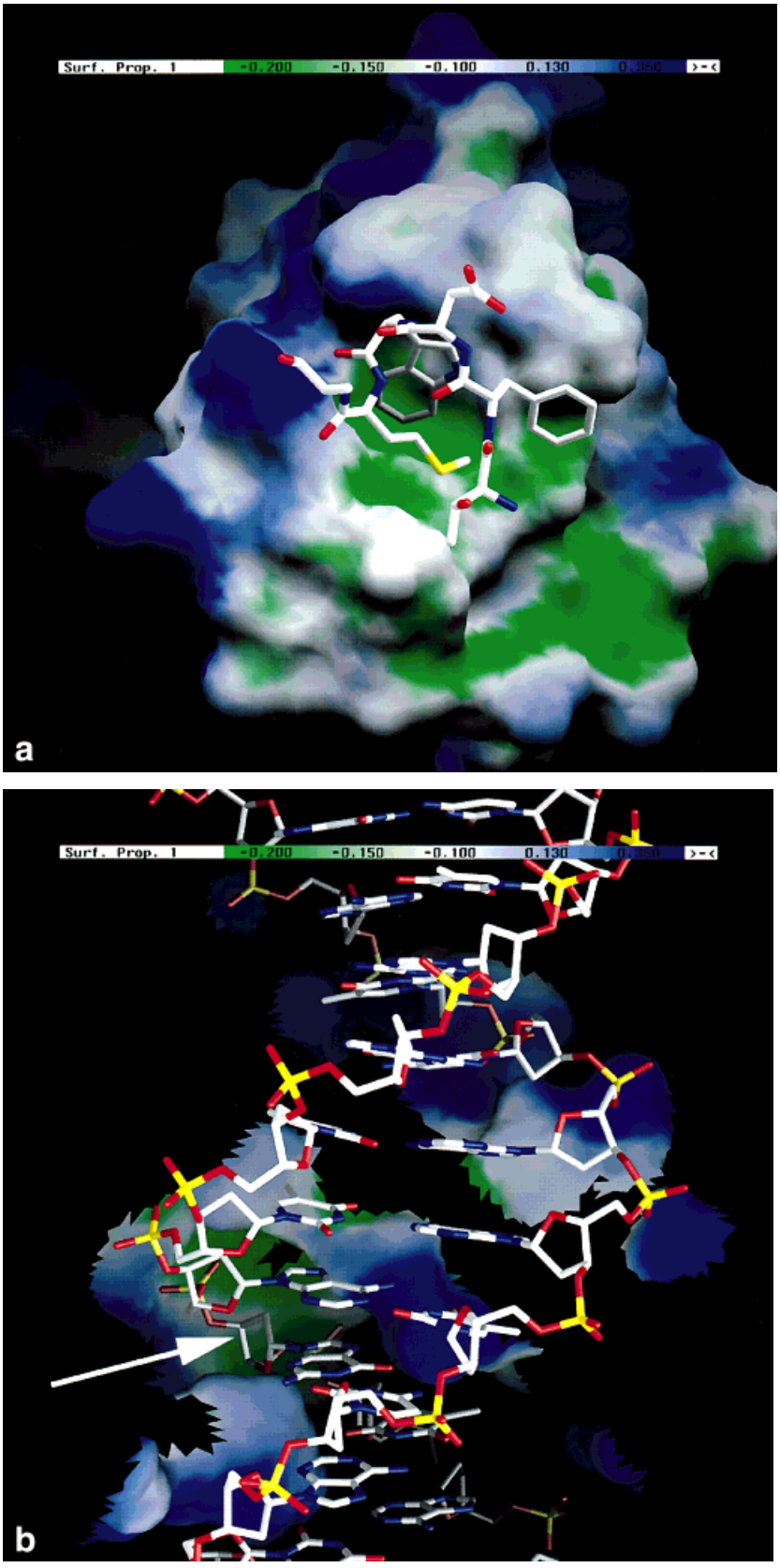


Fig. 5. HoxB1-Pbx1–DNA complex. **a:** Same as in Figure 2A for the HoxB1 hexapeptide–Pbx1 interface. **b:** Same as in Figure 2A for the DNA–Pbx1 interface. Only the contact surface between Pbx1 and the DNA is displayed. The arrow on the bottom left indicates the deoxyribose which buries an hydrophobic region of Pbx1.

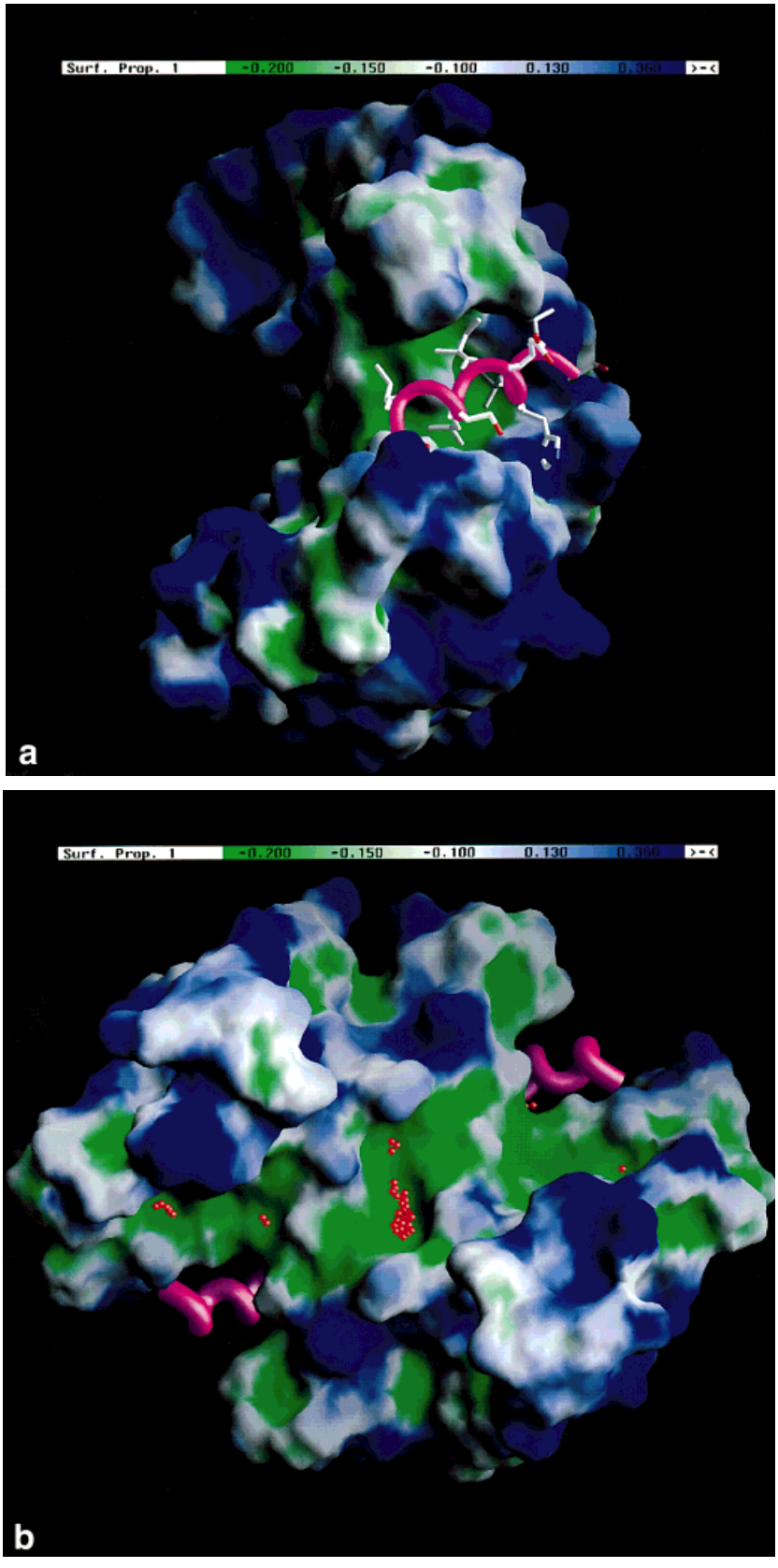


Fig. 6. $(\text{Annexin II})_2(\text{p11})_2$ heterotetramer. **a:** Same as in Figure 2A for the binding site of one of the two annexin II N-terminal peptides. The backbone atoms of annexin II are shown as a magenta ribbon. A lysine side-chain of annexin II pointing toward the solvent has been undisplayed to enhance clarity. **b:** Same as in (A) for the hydrophobic cleft between the two C-terminal anti-parallel α-helices of the p11 dimer. The 100 most hydrophobic points on the SAS of the p11 dimer are shown in red. Annexin II side-chain atoms are not displayed.

The role of hydrophobicity in protein–ligand complexes has been investigated. The fraction of the most hydrophobic receptor regions that are buried at the binding interface is in general particularly high (Table I), suggesting that hydrophobic association is often determinant in protein–ligand binding. This confirms previous findings.[1,5]

The efficient approach illustrated in this article will be useful for the characterization of binding sites for the 3D structures[58] of the large amount of sequences that are emerging from the many genome projects. The location of a binding or association site can be predicted from the clusters of most hydrophobic points on the surface, and the size of the ligand could be estimated from the area and/or the volume of the binding site cleft.

ACKNOWLEDGMENTS

We thank Armin Widmer and Shaheen Ahmed for helpful discussions, and Professor J.A. McCammon for providing the UHBD program, which was used for the finite difference calculations of Figure 1. We also thank Dr. S. Spinelli and Professor C. Cambillau for providing the coordinates of the traseolide-M02/05/01 antibody Fab complex.

NOTE ADDED IN PROOFS

A recent crystal structure of FTase complexed with a CaaX peptide and a farnesyl diphosphate analogue[62] confirms our prediction of the hydrophobicity maps of FTase. In the crystal structure, the hydrophobic regions at the bottom of Figure 3A are occupied by the Ile and Met sidechains of the Cys-Val-Ile-Met substrate.

REFERENCES

1. Davis AM, Teague SJ. Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. Angew Chem Int Ed 1999;38:736–749.
2. Janin J, Chothia C. The structure of protein-protein recognition sites. J Biol Chem 1990;265:16027–16030.
3. Janin J, Chothia C. Role of hydrophobicity in the binding of coenzymes. Biochemistry, 1978;17:2943–2948.
4. Horton N, Lewis M. Calculation of the free energy of association for protein complexes. Prot Sci 1992;1:169–181.
5. Young L, Jernigan RL, Covell DG. A role for surface hydrophobicity in protein-protein recognition. Prot Sci 1994;3:717–729.
6. Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins 1991;11:281–296.
7. Warwicker J, Watson HC. Calculation of the electric potential in the active site cleft due to α-helix dipoles. J Mol Biol 1982;157:671–679.
8. Gilson MK, Honig BH. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. Proteins 1988;4:7–18.
9. Bashford D, Karplus M. pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. Biochemistry 1990;29:10219–10225.
10. Davis ME, Madura JD, Luty BA, McCammon JA. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. Comp Phys Comm 1991;62:187–197.
11. Majeux N, Scarsi M, Apostolakis J, Ehrhard C, Caflisch A. Exhaustive docking of molecular fragments on protein binding sites with electrostatic solvation. Proteins 1999;37:88–105.
12. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379–400.
13. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.
14. Momany FA, Klimkowski VJ, Schäfer L. On the use of conformationally dependent geometry trends from ab initio dipeptide studies to refine potentials for the empirical force field CHARMM. J Comput Chem 1990;11:654–662.
15. Privalov PL, Gill SG. Stability of protein structure and hydrophobic interactions. Adv Prot Chem 1988;39:191–234.
16. Privalov PL, Gill SG. Common features of protein unfolding and dissolution of hydrophobic compounds. Adv Prot Chem 1990;247:559–561.
17. Creighton TE. Stability of folded conformations. Curr Opin Struct Biol 1991;1:5–16.
18. Friedman RA, Honig B. A free energy analysis of nucleic acid base stacking in aqueous solution. Biophys J 1995;69:1528–1535.
19. Caflisch A, Fischer S, Karplus M. Docking by Monte Carlo minimization with a solvation correction: application to an FKBP–substrate complex. J Comput Chem 1997;18:723–743.
20. Vorobjev YN, Almagro JC, Hermans J. Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. Proteins 1998;32:399–413.
21. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127–6129.
22. Hawkins GD, Cramer CJ, Trulhar DG. Pairwise solute descreening of solute charges from a dielectric medium. Chem Phys Lett 1995;246:122–129.
23. Hawkins GD, Cramer CJ, Trulhar DG. Parameterized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. J Phys Chem 1996;100:19824–19839.
24. Schaefer M, Karplus M. A comprehensive analytical treatment of continuum electrostatics. J Phys Chem 1996;100:1578–1599.
25. Qiu DI, Shenkin PS, Hollinger FP, Still WC. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. J Phys Chem A 1997;101:3005–3014.
26. Scarsi M, Apostolakis J, Caflisch A. Continuum electrostatic energies of macromolecules in aqueous solutions. J Phys Chem A 1997;101:8098–8106.
27. Scarsi M, Apostolakis J, Caflisch A. Comparison of a GB solvation model with explicit solvent simulations: potentials of mean force and conformational preferences of alanine dipeptide and 1,2-dichloroethane. J Phys Chem B 1998;102:3637–3641.
28. Jackson JD. Classical electrodynamics. John Wiley & Sons: New York; 1975.
29. Luo R, Moult J, Gilson MK. Dielectric screening treatment of electrostatic solvation. J Phys Chem B 1997;101:11226–11236.
30. Richards FM. Areas, volumes, packing, and protein structure. Ann Rev Biophys Bioeng 1977;6:151–176.
31. Hilpert K, Ackermann J, Banner DW et al. Design and synthesis of potent and highly selective thrombin inhibitors. J Med Chem 1994;37:3889–3901.
32. Bode W, Mayr I, Baumann U, Huber R, Stone SR, Hofsteenge J. The refined 1.9-Å crystal structure of human α-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. EMBO J 1989;8:3467–3475.
33. Banner DW, Hadvary P. Crystallographic analysis at 3.0-Å resolution of the binding to human thrombin of four active site-directed inhibitors. J Biol Chem 1991;266:20085–20093.
34. Obst U, Gramlich V, Diederich F, Weber L, Banner DW. Design neuartiger, nichtpeptidischer Thrombin-Inhibotoren und Struktur eines Thrombin-Inhibitoren-Komplexes. Angew Chem 1995;107:1874–1877.
35. Håkansson K, Tulinsky A, Abelman MM, et al. Crystallographic structure of a peptidyl keto acid inhibitor and human α-thrombin. Bioorgan Med Chem 1995;3:1009–1017.
36. Brandstetter H, Turk D, Hoeffken HW, et al. Refined 2.3 Å x-ray structure of bovine thrombin complexes formed with the benzamidine and arginine-based thrombin inhibitors NAPAP, 4-TAPAP and MQPA. J Mol Biol 1992;226:1085–1099.

37. Buss JE, Marsters JC. Farnesyltransferase inhibitors: The successes and surprises of a new class of potential cancer chemotherapeutics. Chem Biol 1995;2:787–791.
38. Graham SL, Williams TM. Inhibitors of protein farnesylation. Exp Opin Ther Patents 1996;6:1295–1304.
39. Gibbs JB, Oliff A. The potential of farnesyltransferase inhibitors as cancer chemotherapeutics. Ann Rev Pharmacol Toxicol 1997;37: 143–166.
40. Long SB, Casey PJ, Beese LS. Cocrystal structure of protein farnesyltransferase complexed with a diphosphate substrate. Biochemistry 1998;37:9612–9618.
41. Momand J, Zambetti GP, Olson DC, George D, Levine AJ. The MDM-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. Cell 1992;69:1237.
42. Kussie PH, Gorina S, Marechal V, et al. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science 1996;274:948–953.
43. Gehring WJ, Affolter M, Bürglin T. Homeodomain proteins. Ann Rev Biochem 1994;63:487–526.
44. McGinnis W, Krumlauf R. Homeobox genes and axial patterning. Cell 1992;68:283–302.
45. Piper DE, Batchelor AH, Chang CP, Cleary ML, Wolberger C. Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. Cell 1999;96:587–597.
46. Kligman D, Hilt DC. The S100 protein family. Trends Biochem Sci 1988;13:437–443.
47. Gerke V, Weber K. The regulatory chain of the p36-kd substrate complex of viral tyrosine-specific protein kinases is related in sequence to the s-100 protein of glial cells. EMBO J 1985;4:2917–2910.
48. Waisman DM. Annexin II tetramer:structure and function. Mol Cell Biochem 1995;149/150:301–322.
49. Gerke V, Moss S. Annexins and membrane dynamics. Biochim Biophys Acta 1997;1357:129–154.
50. Powell MA, Glenney JR. Regulation of calpacitin I phospholipid binding by calpacitin I light-chain and phosphorylation by $p60^{v\text{-}src}$. Biochem J 1987;247:321–328.
51. Rety S, Sopkova J, Renouard M, et al. Crystal structure of a complex of p11, the ligand of annexin II, with the annexin II N-terminal peptide. Nature Struct Biol 1999;6:89–95.
52. Rasmussen R, Benvegnu D, O'Shea EK, Kim PS, Alber T. X-ray scattering indicates that the leucine zipper is a coiled coil. Proc Natl Acad Sci USA 1991;88:561.
53. O'Shea EK, Rutkowski R, Kim PS. Evidence that the leucine zipper is a coiled coil. Science 1989;243:538–542.
54. Arevalo JH, Stura EA, Taussig MJ, Wilson IA. Three-dimensional structure of an anti-steroid Fab′ and progesterone-Fab′ complex. J Mol Biol 1993;231:103–118.
55. Langedijk AC, Spinelli S, Anguille C, et al. Insight into odorant perception: the crystal structure and binding characteristics of antibody fragments against the musk odorant traseolide. J Mol Biol 1999:in press.
56. Mittl PRE, Di Marco S, Fendrich G, et al. A new structural class of serine protease inhibitors revealed by the structure of the hirustasin-kallikrein complex. Structure 1996;5:253–264.
57. Fitzgerald PMD, McKeever BM, VanMiddlesworth JF, et al. Crystallographic analysis of a complex between human immunodeficiency virus type 1 and acetyl-pepstatin at 2.0 Å resolution. J Biol Chem 1990;265:14209–14219.
58. Sánchez R, Sali A. Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. Proc Natl Acad Sci USA 1998; 95:13597–13602.
59. Davis ME, McCammon JA. Solving the finite difference linearized Poisson-Boltzmann equation: a comparison of relaxation and conjugate gradient methods. J Comput Chem 1989;10:386–391.
60. Davis ME, McCammon JA. Calculating electrostatic forces from grid-calculated potentials. J Comput Chem 1990;11:401–409.
61. Davis ME, McCammon JA. Dielectric boundary smoothing in finite difference solutions of the Poisson equation: An approach to improve accuracy and convergence. J Comput Chem 1991;12:909–912.
62. Strickland CL, Windsor WT, Syto R, Wang L, Bond R, Wu Z, Schwartz J, Le HV, Beese LS, Weber PC. Crystal structure of farnesyl protein transferase complexed with a CaaX peptide and farnesyl diphosphate analogue. Biochemistry 1998;37:16601–16611.