WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Hot spots—A review of the protein–protein interface determinant amino-acid residues

Irina S. Moreira, Pedro A. Fernandes, and Maria J. Ramos*

REQUIMTE/Departamento de Química, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal

ABSTRACT

Proteins tendency to bind to one another in a highly specific manner forming stable complexes is fundamental to all biological processes. A better understanding of complex formation has many practical applications, which include the rational design of new therapeutic agents, and the analysis of metabolic and signal transduction networks. Alanine-scanning mutagenesis made possible the detection of the functional epitopes, and demonstrated that most of the protein–protein binding energy is related only to a group of few amino acids at intermolecular protein interfaces: the hot spots. The scope of this review is to summarize all the available information regarding hot spots for a better atomic understanding of their structure and function. The ultimate objective is to improve the rational design of complexes of high affinity and specificity as well as that of small molecules, which can mimic the functional epitopes of the proteic complexes.

Proteins 2007; 68:803–812.

Key words: hot spots; protein–protein interactions; alanine scanning mutagenesis; O-ring; alanine-shaving; functional epitopes.

INTRODUCTION

One fundamental aim of molecular biology is the discovery of all protein–protein interactions in an organism as well as their biochemical and biological functions.[1] Protein–protein interactions are central to most biological processes, and detection of specific amino acid residues that contribute to the specificity and strength of protein interactions is a problem of the utmost importance. Therefore, an atomic exploration of the proteic systems will allow a better understanding of the forces that drive protein–protein interactions and will elucidate how molecular recognition processes govern complex biological systems.[2–15]

Crystallographic structures of proteins cocrystallized with various ligands,[16] structural and thermodynamic studies[17,18] that identify structural epitopes (residues in contact with a ligand), and alanine-scanning mutagenesis of protein–protein interfacial residues (specially the computational approach) allow a more detailed comprehensive knowledge of the principles that govern complex formation.[19] The understanding of protein–protein associations is a useful link between structure and function of biomolecular systems, and allows the characterization of the energetics of molecular complexes.[20,21] A number of studies have focused on the physical and chemical properties of protein–protein interfaces of complexes to determine their unique features.[22–24]

PROTEIN–PROTEIN INTERFACIAL CHARACTERISTICS

Size

Protein–protein interactions are very complex and can be characterized by their size, shape, and surface complementarity.[23] The hydrophobic[8] and electrostatic interactions[4] they establish, as well as the flexibility of the molecules involved, are very significant.[2–4]

Protein–protein recognition sites are formed by protein surfaces with good shape and electrostatic complementarity.[2–4,25] It has been described that the standard-size interfaces[26] have 1200–2000 Å^2. A few smaller interfaces with 1150–1200 Å^2 normally constitute short-lived and low-stability complexes.[14] Large interfaces in the range 2000–4660 Å^2 occur mostly between proteases and a particular class of inhibitors and between G-proteins and other components of the signal transduction system.[25,26]

The vast majority of protein heterodimer interfaces[16] are larger than 600 A^2 and it is often assumed that the energy of protein–protein binding is directly related to the buried hydrophobic surface area.[4,5,27–29] This 600 Å^2 cutoff should correspond to a minimum

*Correspondence to: Maria J. Ramos, REQUIMTE/Departamento de Química, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal. E-mail: mjramos@fc.up.pt
Received 20 June 2006; Revised 4 October 2006; Accepted 11 December 2006
Published online 1 June 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21396

area required to make a water-tight seal around a critical set of energetically favorable interactions.[5]

Interface character

Protein–protein interfaces are frequently hydrophobic and bury a large extent of nonpolar surface area.[30] Hence, hydrophobicity is a leading force in protein–protein interactions.[30,31] Hydrophobic interactions in proteins occur between nonpolar regions of their amino acid residues through van der Waals contacts and are driven by the gain in free energy that results from their movement from polar (aqueous) to nonpolar environment.[32] These interactions lead to tight packing of residues that are organized as patches that tend to protrude from the surface. The number of patches may vary from 1 to 15 and their sizes[33] are within 200–400 Å resulting in the expulsion of water molecules in the interface, and causing an increase in entropy that favors complex formation.[34] Summing the free energy gain produced by all the individual van der Waals interactions and adding the energy gain associated with desolvation, the total gain in free energy is substantial and may produce a higher stabilization of the protein–protein complex.[23]

Electrostatic forces are the other noteworthy force involved in protein–protein interactions[35–39] because electrostatic complementarity of interacting protein surfaces[24] promotes complex formation[40,41] and defines the lifetime of complexes.[42] The primary determinants[43] are conserved potentials that increase the association rate.[44] Thus, electrostatics may provide a steering force for the diffusion process or transportation across the enzyme surface.[45] It has also been found that the average number of hydrogen bonds is proportional to the area of subunit surfaces[46,47]: one bond per each 100–200 Å. On average, side chains of amino acids form 76% of the hydrogen bonds in protein interfaces. Other hydrogen bonds are formed between protein contact surfaces and the surrounding water molecules.[48–50]

HOT SPOTS

Since its initial application to human growth hormone and the growth hormone binding protein[51] (shown in Fig. 1), alanine scanning mutagenesis continues to be a valuable procedure for both hot spot detection and analysis of a wide range of protein–protein interfaces.[5,14,45,53–55] Although slow and labor-intensive, alanine-scanning mutagenesis is the most trendy method for mapping functional epitopes, as alanine substitutions remove side-chain atoms past the β-carbon without introducing additional conformational freedom.[56–59] Thus, the role of side-chain functional groups at specific positions and the energetic contributions of individual side-chains to protein binding can be inferred from alanine mutations. Glycine would also nullify the side chain but could introduce conformational flexibility into the protein backbone, and therefore is not commonly used.[60] With the application of this methodological approach, it has been discovered a highly uneven distribution of energetic contributions of individual residues across each interface, and that only a few key residues do contribute significantly to the binding free energy of protein–protein complexes: the hot spots.[4,14,61–67]

Hot spots have been defined as those sites where alanine mutations cause a significant increase in the binding free energy of at least 2.0 kcal/mol.[55] To have a strong impact in protein building the binding free energy should be higher than 4 kcal/mol (three orders of magnitude in the binding affinity constant). However, residues whose mutation results in such a large difference are quite unusual, and the threshold for the hot spots had to be lowered to 2 kcal/mol to get enough data for statistical analysis. Therefore, in a protein–protein interface, a small subset of the buried amino acids typically contribute to the majority of binding affinity as determined by the change in the free energy of binding ($\Delta\Delta G_{binding}$) upon mutation of the residue to an alanine. Of the analysis of the database of Thorn and Bogan[55] we have determined that on an average of 9.5% of the interfacial residues are hot spots (although this value could be overestimated because of the reduced information available for some of the complexes). The same hot spot adapts to the same residues in different structural contexts, and therefore is used by proteins that function by binding to multiple partners, showing a high functional and structural adaptivity.[65,68–71] The high propensity shown for interaction with a diversity of partners suggests that the understanding of hot spots may be helpful not only for the study of a single protein–protein dimer, but also for determining likely sites of interaction for other binding partners.[65]

Although there is no purely geometric reason, these energetic determinants are compact, centralized regions of residues crucial for protein association.[72,73] Thus, very few hot spot residues are at the edge of an interface.[5] Hot spots have been shown to overlap with structurally conserved residues,[27,74] and their number tend to increase with size of the analyzed system.[45]

Amino acid composition

Systematic analysis of hot spots has shown a nonrandom composition. Instead they have a distinctive amino acid composition.[75] The fundamental ones are tryptophan (21%), arginine (13.3%), and tyrosine (12.3%). As an example, the importance of the tryptophan residue can be clearly seen in Figure 2, which illustrates the complex formed between the human growth hormone and the growth hormone binding protein. In 29 interfacial residues only four are hot spots, and two of them are tryptophan (with $\Delta\Delta G_{binding}$ higher than 4.5 kcal/mol).

DOI 10.1002/prot

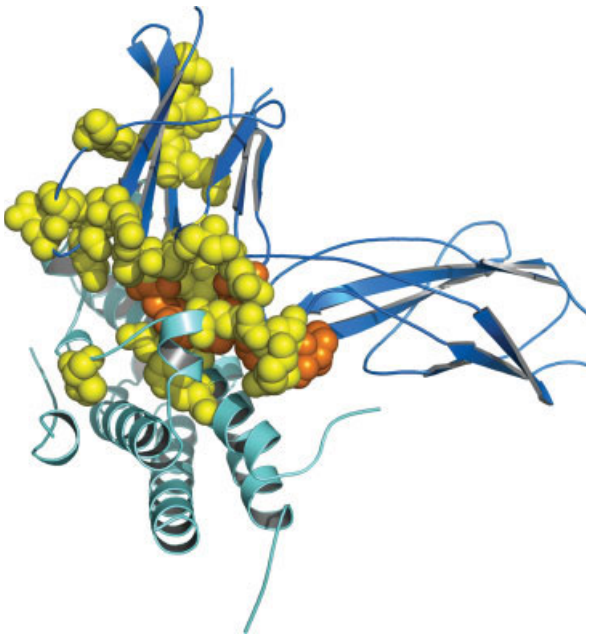

Figure 1

The human growth hormone (in cyan) complexed with its receptor (in blue). In yellow are represented residues with a relative binding energy lower than 2.0 kcal/mol and in orange the residues with a relative binding energy higher than 2.0 kcal/mol. The PDB ID is 1A22. The Pymol software[52] was used to produce this picture.

On the other hand, leucine (not isoleucine), serine, threonine, and valine residues are disfavored and essentially absent as hot spots albeit their importance for distinct proteic structures.[5] Tryptophan appears to play a unique function, probably owing to its large size and aromatic nature.[76] It can contribute with aromatic π-interactions, it is a hydrogen bonding donor, it has a large hydrophobic surface, and can protect fragile hydrogen bonds from water.[77] Finally, tryptophan mutation to an alanine generates a large cavity, due to the significant difference in sizes,[5] which can create a highly complex destabilization. Tyrosine has the third highest conservation propensity on binding sites offering a hydrophobic surface, and both aromatic π-interactions and the hydrogen bonding ability of its 4-hydroxyl group.[5] It has a three times higher probability of being a hot spot than phenylalanine probably because of the ability of tyrosine to participate in hydrogen bonds.[5] The average percentage of aromatic residues as hot spots[5] clearly demonstrate their importance to protein interactions. Arginine is capable of multiple types of favorable interactions, such as forming a hydrogen bond arrangement with up to five hydrogen-bonds and a salt-bridge with its positive charge on its guanidinium motif.[5] Analysis of various complexes have also shown that aspartate and aspargine are favored over glutamate and glutamine, which can be explained presumably due to differences in side-chain conformational entropy.[78] Despite the fact that they are isomers with

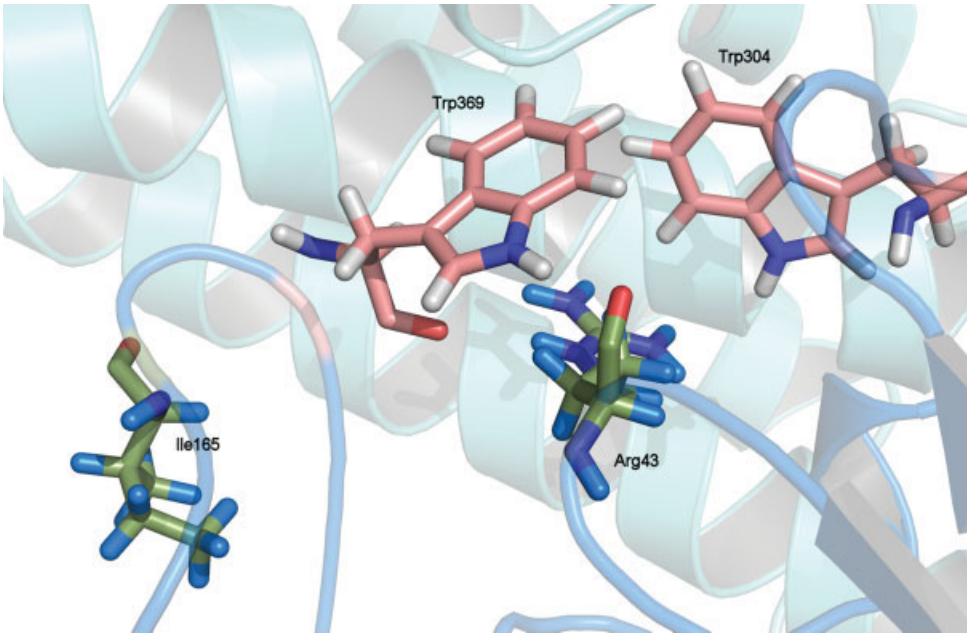

Figure 2

The human growth hormone (in cyan) complexed with its receptor (in blue). The four hot spot residues are highlighted by a stick representation and the two tryptophan residues are colored in pink. The PDB ID is 1a22. The Pymol software[52] was used to produce this figure.

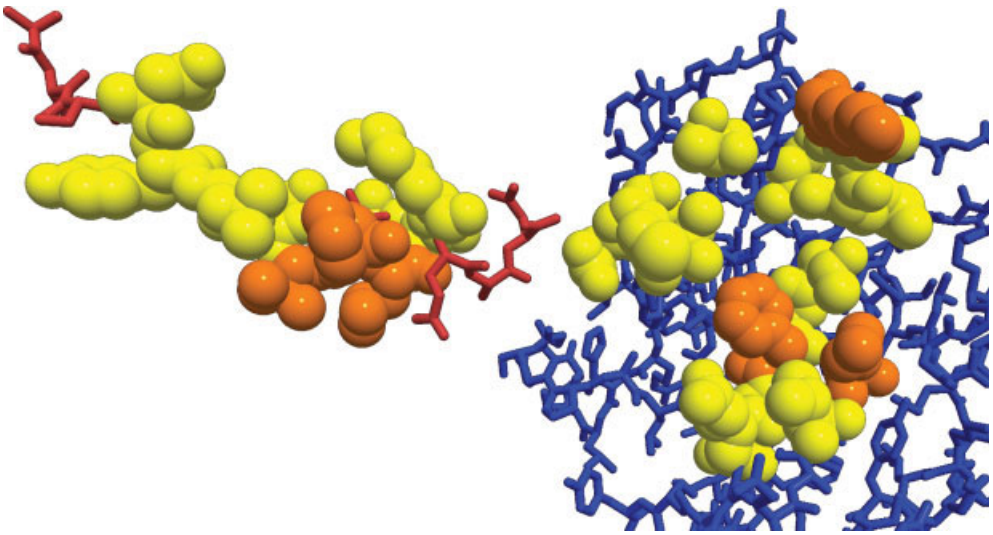

Figure 3

Complex formed between the bacterial cell-division protein ZipA (in blue) and the FtsZ (in red) fragment highlighting the mutated residues upon alanine scanning mutagenesis of the complete protein–protein interface by means of a van der Waals representation. In yellow are represented the residues with a relative binding free energy lower than 2.0 kcal/mol and in orange the residues with a relative binding energy higher than 2.0 kcal/mol. The PDB ID is 1F47. The VMD software[83] was used to produce this figure. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

essentially identical chemistry, isoleucine, with a frequency of 9.62% as a hot spot, is more than 10 times as frequent as leucine (0.83%).[5]

Complementarity—clusters of hot spots

As already mentioned most interfaces are optimal tight fitting[15] regions characterized by complementary pockets scattered through the central region of the interface, and enriched in structurally conserved residues.[79] These pockets are classified as "complementary"[80] because there is a large complementarity both in shape and in the juxtaposition of hydrophobic and hydrophilic hot spots, with buried charged residues forming salt bridges and hydrophobic residues from one surface fitting into small nooks on the opposite face.[12] Usually, the hot spot of one face packs against the hot spot of the other face establishing a region determinant for complex binding,[81,82] which may provide sites for drug discovery.[15] This can be seen in Figure 3, which represents the interface between FtsZ and ZipA. The complex is shown open to exhibit the notorious shape complementarity between the hot spots through their close contact across the interface. Usually the hot spot residues present in these clusters or pockets are involved in hydrogen bonds. This point is illustrated in Figure 4, which represents the complex formed between the C2 fragment of streptococcal protein G and the human immunoglobulin IgG. We can observe the hydrogen bonds (2.43 and 1.69 Å) formed between the hot spot Glu27 of the C2 fragment and the hot spot Ser254 of IgG as well as two hydrogen bonds between the NH atom of the main chain of the hot spot Ile253 and the Oϵ1 and Oϵ2 atom of Glu27 with distances of 2.95 and 1.65 Å, respectively.

The number of these hot spots within densely packed regions is correlated essentially with the interface size,[15] and local organization of these hot spots is a critical factor in stabilizing protein–protein interactions.[27] Complementarity is basically affected by the size of the buried surface, alignment of polar and nonpolar residues, number of buried waters, and the packing densities of atoms involved in the protein–protein interface.[84] Residues across the protein–protein interface often coevolve,[85,86] forming complemented pockets abundant with enriched conserved residues, and the corresponding protruding residues are also frequently conserved.[15]

It has been determined that on average 79% of the hot spot residues are located on complemented pockets,[27] and 93% of residues with a free energy difference of binding ($\Delta\Delta G_{binding}$) higher than 4 kcal/mol upon alanine mutagenesis are found as protruding or complemented pocket residues.[27] Complemented pockets contain smaller fractions of polar or ionizable residues, such as arginine, lysine, glutamate, and aspartate, than the other surface pockets.[15] This suggests that the desolvation barrier for protruding residues to anchor into the complemented pocket is not high, since there are few

DOI 10.1002/prot

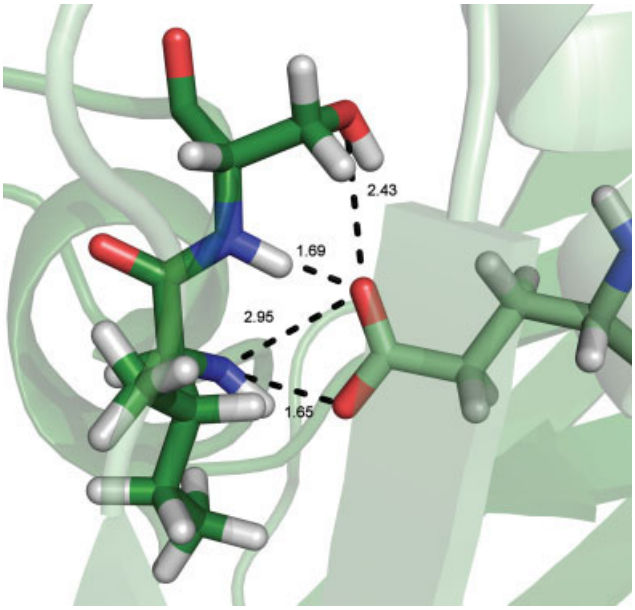

Figure 4

Detail of the molecular interactions involving the hot spot Glu27 of the C2 fragment of streptococcal protein G and the warm spots Ser254 and Ile253 of the human immunoglobulin IgG. The PDB ID is 1FCC. The Pymol software[52] was used to produce this figure. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

polar residues, especially ionizable residues, in complemented pockets. However, there is a high presence of polar and ionizable residues located at the bottom of the complemented pockets to increase binding stability, by enhancing polar–polar interactions in a hydrophobic environment.[15]

For residues located in complemented pockets, tryptophan, glycine, proline, cysteine, tyrosine, and glutamate are likely to be more conserved. Tryptophan, being a large residue with many neighbors, is often found on the wall of the complemented pocket and may function to occlude interactions inside the pocket from the solvent.[87] Glycine is far more conserved if located at a complemented pocket than if located within the rest of the interface. Glycine lacks a side-chain allowing a tight packing, and consequently is coupled with aromatic, polar, and small hydrophobic residues in the interacting chain, with likely backbone H-bonding across the interface.[88] Although not very common as a hot spot glycine is important in some structural motifs, such as the arginine–glycine–aspartate–serine (RGDS) motif. The RGDS polypeptides derived from cell adhesion molecules are very important, for example, for the control of cancer metastasis.[88]

Hot spots are preferably found in pre-existing pockets in the unbound state, which will be occupied by their binding partners in the bound state. The mechanism of protein–protein interactions often does not involve structural changes that abolish pockets pre-existent in unbound structures.[15,66,68] Packing defects at the protein–protein interface result in these gaps or pockets, and it is unclear whether unfilled pockets contain water molecules or how the dynamics of water molecules entering and escaping these pockets may affect binding stability.[15]

O-ring structure

Through careful analysis of the binding free energy upon alanine mutation and the X-ray structures of 23 protein–protein complexes Bogan and Thorn[5] suggested that the hot spots are usually surrounded by residues not important for binding, whose role would be to shelter the hot spots from the solvent. These structures resembled on O-ring and the idea becomes known as the O-ring theory. The theory was based in three important observations. First, residues that have a large impact on the free energy of binding are largely protected from contact with bulk solvent because in the X-ray structures they have low or zero accessible solvent area (ASA). Second, many of the residues that are occluded from solvent do not make large contributions to the binding energy. Third, there are no residues with high solvent accessibility that make a large contribution to the $\Delta\Delta G_{binding}$.[5] Thus, it was proposed that inaccessibility to the solvent is a necessary albeit insufficient condition to define a residues as a binding hot spot.[5,54] Therefore, the residues surrounding a hot-spot, which generally establish unimportant interactions, were supposed to be related to the shielding of the solvent.

Occlusion of bulk solvent by the O-ring should favor residues (tryptophan, tyrosine, aspargine) that are capable of establishing both hydrogen bonding and hydrophobic interactions. As the primary role of the O-ring is to occlude bulk solvent from the hot spot, an alanine mutation in it might have little or no effect on binding affinity as long as bulk solvent is still stereochemically blocked.[5] On the other hand, the hydrophobic rings around the polar residues that usually constitute hot spots may contribute to more mobile, easily displaced water molecules.[89,90] Nevertheless, the hydrophobic rings around these polar residues may compensate the entropic terms by increasing both side-chain and backbone motions.[66,68]

An interface may contain a single, or a few hot clusters, and the highly packed nature of these hot spots facilitates removal of water molecules upon binding, strengthening the contributions of charge–charge interactions.[27] In protein–protein recognition locations a core and a rim was identified, which was made of a standard size recognition patch, often increased by adding smaller surface patches, and in few cases by duplication. The core regions contain atoms that are buried upon complex formation and are surrounded by a rim of atoms that

remains partly accessible. The two regions differ in their amino acid composition. The rim that form the O-ring structures is similar to the surface, whereas the core has a distinctive composition.[91]

Although the O-ring theory is a hypothesis commonly accepted,[5,27,45,66,91] it did not had a conclusive evidence. The assumption that non-hot-spot residues do not participate in important interactions is only valid if it can be demonstrated that water or nearby side chains are not able to effectively substitute the eliminated atoms.[92] However, the replacement of the eradicated atoms could lead to interactions that provide the same driving force for complex formation as the original side-chain and, thus, O-rings may not directly be related to the binding energy.[54] In a recent work we have shown that within a flexible, dynamic protein framework the hot spot residues are indeed kept sheltered from the bulk solvent during the whole molecular mechanics simulation and all the results obtained in our study supported the O-ring theory.[93] However it must be noticed that this last result has been obtained for one complex only, and therefore insufficient proof is available to strongly favor the O-ring theory. Absence of the solvent is not sufficient evidence to assume that the O-ring theory is its cause, even though this theory is very attractive and intuitive. In fact, a computational analysis would be of the utmost importance to clarify between the role played by the intermolecular interactions and that of the solvent.

Cooperativity—alanine-shaving

Alanine-shaving, which is the process of making multiple simultaneous alanine mutations, has been used to test experimentally the cooperativity between inert side-chains.[5] Cooperativity can be detected by multiple mutation cycles, in which the free energy change caused by the simultaneous mutations at residue positions in a protein is compared with the sum of the free energy changes associated with single mutations at each of the residues positions.[94,95] Deviations from additivity are indicative of cooperative interactions, and in this way, energetic coupling between protein residues can be quantified.[96] Additivity of mutational effects is very common for protein–protein interactions.[55,97,98] The tight-networked hot spot organization in clusters indicates that the contributions of the hot spots within a hot region should be cooperative.[27] Nonetheless, many exceptions to the non-additivity rule have been observed.[96,99,100]

According to a study developed by Jin and Wells, alanine-shaving of 16 amino acid residues that surround five functionally critical residues for binding can be tolerated with minimal change in affinity.[101] The easiest explanation is that several side chains actually hinder binding, and simultaneously mutating them to alanines can improve affinity.[102] However, such kind of incompatibilities should have been eliminated during evolution, and thus there may be other roles for the peripherical contact residues, which limit their ability to change.[102,103] Most of these side-chains are polar or charged and may be important for solubility of the protein even in the unbound state and for conferring specificity to the functional epitope.[4]

However, alanine shaved mutants often have binding affinities similar to those of wild-type proteins because alanine side-chains and the protein backbone appear to be largely sufficient to exclude solvent from a neighboring hot spot. Obviously, the hydrophobic nature of the alanine residue can also contribute to solvent shielding. Therefore, alanine-shaving often fails to significantly increase the solvent accessibility of the hot spot in a complex.[5]

Experimental detection methods

Systematic mutagenesis is very laborious and time-consuming to perform, as individual mutant proteins must be purified and analyzed separately.[54] Each alanine-substituted protein must be separately constructed, expressed and sometimes refolded, and the loss of the side chain functionality is then assessed in an *in vitro* assay of the protein activity.[60] Combinatorial libraries of alanine substitutions are an alternative to the arduous process of scanning individual positions in a protein. Through a single round of site-specific oligonucleotide-directed mutagenesis, "binomial substitutions" of either alanine or a wild-type amino acid residue are readily accessible by conventional oligonucleotide synthesis for seven amino acids (aspartate, glutamate, glycine, proline, serine, threonine, and valine). For these seven amino acids, altering a single encoding nucleotide can result in a codon for alanine. Other method is "shotgun scanning," which implements a simplified format for combinational alanine scanning and uses phage-displayed libraries of alanine-substituted proteins for high-throughput analysis.[60] The rapidity and general applicability of the shotgun-scanning scheme should accelerate the investigation of many other protein–protein interactions.[96]

Erlanson et al.[104] et al. have created a powerful new technology to directly study the binding ability of hot spots with a library of potential organic binding partners. This approach, termed "covalent tethering," utilizes equilibrium disulfide exchange to target potential binding partners at specific regions on a protein surface and to calculate relative binding affinities.[54] It has been automated and is currently being used in a commercial setting to discover new ligands for protein interfaces and enzymes.[105] Tethering provides a context for interpreting results through experimental structure determination or computational modeling.[54] Piehler and Schreiber[106] describe another approach for measuring kinetics and affinities based on reflectrometric interference spectros-

DOI 10.1002/prot

copy, that could enable efficient large-scale alanine scans strengthened with kinetic data.[54]

Besides experimental alanine scanning mutagenesis there are other experimental methods to study protein–protein interactions. They can be essentially divided in three types: molecular biology based methods, such as yeast two-hybrid system, ubiquitin-based split-protein sensor, and Fluorescence Resonance Energy transfer; mass spectrometry based methods; and protein microarrays.[106]

Theoretical detection methods

As experimental hot spots determination is time consuming and involves a high cost, an effort has been made in achieving accurate, predictive computational methodologies for alanine scanning mutagenesis, capable of reproducing the experimental mutagenesis values. For that purpose it is important to accurately calculate the binding free energies of known three-dimensional structures and the effect of mutations on these affinities.

The theoretical prediction of the free binding energy differences and the understanding of the physical foundations of affinity and specificity of the complex interaction prior to the experimental design are crucial in computational biochemistry.[107] To apply a quantitative model for the binding affinity determination of a broad variety of protein–protein interfaces complements experimental analysis and adds molecular insight to the macroscopic properties measured therein.[108–112]

A huge amount of algorithms of increasing complexity has been employed to address the binding energy between biological molecules, and can be divided essentially in two types. First, empirical functions or simple physical methods that use knowledge-based simplified models are used to evaluate binding. Second, there are fully atomistic methods that estimate the free energy of association directly or changes in the binding free energies as a result of mutating the residues of the interacting molecules.

The most rapid methods for estimation of binding free energies are the empirical or knowledge-based (statistical) scoring approaches in conjunction with simple physical models.[113] More time consuming methods involve fully atomistic simulations and include both the rigorous free energy perturbation[114] and thermodynamic integration,[115] and more approximate methods such as MM-PBSA.[116] λ-dynamics,[117–120] chemical Monte–Carlo/molecular mechanics[118] or ligand interaction scanning[119] are also methodological approaches proposed to identify the interfacial hot spots.

Therapeutic implications of studying hot spots

Reliable prediction of functional epitopes has immediate implications for drug design and for protein engineering by specifically targeting them with virtual ligand screening and template-directed combinatorial chemistry.[54] At present, protein–protein contact areas are considered to be new prospective drug targets because numerous physiological and pathological cell processes depend on them, and thus can be influenced by external compounds. Most drugs produce their effect by interacting with a biological macromolecule by entirely nonbonded forces or, in some cases, by a covalent interaction. Drugs that interact with proteins present a tight-binding, and often have a high degree of complementarity with the target. The drug often forms hydrogen bonds with the receptor. However, some targets have hydrophobic pockets into which the drug can put perhaps a hydrophobic group of an appropriate size. The affinity of a protein interface depends on both an energetically critical hot spot located near the center of the interface, and a surrounding seal of contacting residues that may establish the correct solvation environment by occluding bulk solvent from the hot spot.[5] Small design molecules with a built-in O-ring may be able to adjust or even mimic large molecular interfaces[102] much more rapidly in a functionally important protein–protein interaction.[5] Furthermore, the rapid determination of hot spot residues may speed development of small-molecule competitive inhibitors of protein–protein interactions.[5]

Therefore, the identification of critical binding residues on proteins may provide useful insights for medicinal chemistry by optimization of the corresponding ligands.

CONCLUSIONS

The atomic resolution exploration of protein–protein interactions is essential for all biological functions because it allows the comprehensive knowledge of the physical basis of affinity, as well as the understanding of molecular recognition, having broad applications ranging from rational drug design to the analysis of metabolic and signal transduction networks.

Alanine-scanning mutagenesis made possible the detection of the functional epitopes, and demonstrated that most of the protein–protein binding energy is related only to a group of a few amino acids at intermolecular protein interfaces: the hot spots. Hot spots can be clustered in complemented pockets scattered on the interfaces and enriched in structurally conserved residues. These pockets show large complementarity both in shape and in the juxtaposition of the amino acid residues for the hot spots, with the hot spot of one face packed against the hot spot of the other monomer.

An O-ring structure, that leads to the exclusion of bulk solvent from the interacting residues by surrounding the hot spot with a set of contacts that are energetically unimportant, has been proposed to be fundamental to achieve high binding affinity.

The identification of these critical binding residues on proteins permits a rational design of complexes of high

affinity and specificity as well as that of small molecules that can mimic the large interface, which is typical of protein–protein complexes. Therefore, it is fundamental to the development of small-molecule competitive inhibitors of protein–protein interactions, which is crucial in structure based drug design.

REFERENCES

1. Bordner AJ, Abagyan R. Statistical analysis and prediction of protein–protein interfaces. Proteins 2005;60:353–366.
2. Janin J. Elusive affinities. Proteins 1995;21:30–39.
3. Janin J. Protein–protein recognition. Prog Biophys Mol Biol 1995; 64:145–166.
4. Jones S, Thornton JM. Principles of protein–protein interactions. Proc Natl Acad Sci USA 1996;93:13–20.
5. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol 1998;280:1–9.
6. Ma B, Kumar S, Tsai CJ, Nussinov R. Folding funnels and binding mechanisms. Protein Eng 1999;12:713–720.
7. Tsai CJ, Lin SL, Wolfson H, Nussinov R. A dataset of protein–protein interfaces generated with a sequence-order-independent comparison technique. J Mol Biol 1996;260:604–620.
8. Tsai CJ, Nussinov R. Hydrophobic folding units at protein–protein interfaces: implications to protein folding and to protein–protein association. Protein Sci 1997;6:1426–1437.
9. Tsai CJ, Kumar S, Ma B, Nussinov R. Folding funnels, binding funnels and protein function. Protein Sci 1999;8:1181–1190.
10. Kortemme T, Baker D. Computational design of protein–protein interactions. Curr Opin Chem Biol 2004;8:91–97.
11. Russel RB, Alber F, Aloy P, Davis FP, Korkin D, Pichaud M, Topf M, Sali A. A structural perspective on protein–protein interactions. Curr Opin Struct Biol 2004;14:313–324.
12. Arkin MR, Wells JA. Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. Nat Rev Drug Discov 2004;3:301–317.
13. Chothia C, Janin J. Principles of protein–protein recognition. Nature 1975;256:705–708.
14. Conte LL, Chothia C, Janin J. The atomic structure of protein–protein recognition sites. J Mol Biol 1999;285:2177–2198.
15. Li X, Keskin O, Ma B, Nussinov R, Liang L. Protein–protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking. J Mol Biol 2004;344:781–795.
16. Erlandsen H, Abola EE, Stevens RC. Combining structural genomics and enzymology: completing the picture in metabolic pathways and enzyme active sites. Curr Opin Struct Biol 2000;10:719–730.
17. Davies DR, Cohen GH. Interactions of protein antigens with antibodies. Proc Natl Acad Sci USA 1996;93:7–12.
18. Brooijmans N, Sharp KA, Kruntz ID. Stability of macromolecular complexes. Proteins 2002;48:645–653.
19. Pal G, Ultsch MH, Clark KP, Currell B, Kossiakoff AA, Sidhu SS. Intramolecular cooperativity in a protein binding site assessed by combinatorial shotgun scanning mutagenesis. J Mol Biol 2005;347:489–494.
20. Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Freer ST, Rose PM. Computational detection of the binding-site hot spot at the remodelled human growth hormone-receptor interface. Proteins 2003;53:201–219.
21. Aqvist J, Luzhkov B, Brandsdal BO. Ligand binding affinities from MD simulation. Acc Chem Res 2002;35:358–365.
22. Ofran Y, Rost B. Analysing six types of protein–protein interfaces. J Mol Biol 2003;10:325–377.
23. Fernandez A, Scheraga HA. Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. Proc Natl Acad Sci USA 2003;100:113–118.
24. Schmitt S, Kuhn D, Klebe G. A new method to detect related function among proteins independent of sequence and fold homology. J Mol Biol 2002;323:387–406.
25. Janin J, Chothia C. The structure of protein–protein recognition sites. J Biol Chem 1990;265:16027–16030.
26. Horton N, Lewis M. Calculation of the free energy of association for protein complexes. Protein Sci 1992;1:169–181.
27. Keskin O, Ma B, Nussinov R. Hot regions in protein–protein interactions: the organization and contribution of structurally conserved hot spot residues. J Mol Biol 2005;345:1281–1294.
28. Novotny J, Bruccoleri RE, Saul FA. On the attribution of binding energy in antigen–antibody complexes McPC 603, D1.3, and HyHEL-5. Biochemistry 1989;28:4735–4749.
29. Privalov PL. Stability of proteins: small globular proteins. Adv Protein Chem 1979;33:167–241.
30. Young L, Jernigan RL, Covell DG. A role for surface hydrophobicity in protein–protein recognition. Protein Sci 1994;3:717–729.
31. Korn AP, Burnett RM. Distribution and complementarity of hydropathy in multisubunit proteins. Proteins 1991;9:37–55.
32. Kauzmann W. Some factors in the interpretation of protein denaturation. Adv Protein Chem 1959;14:1–63.
33. Lijnzaad P, Argos P. Hydrophobic patches on protein subunit interfaces: characteristics and prediction Proteins 1997;28:333–347.
34. Dill KA. Dominant forces in protein folding. Biochemistry 1990; 29:7133–7155.
35. Stevens JM, Armstrong RN, Dirr HW. Electrostatic interactions affecting the active site of class sigma glutathione S-transferase. Biochem J 2000;347:193–197.
36. Sheinerman FB, Norel R, Honig B. Electrostatic aspects of protein–protein interactions. Curr Opin Struct Biol 2000;10:153–159.
37. Xu D, Lin SL, Nussinov R. Protein binding versus protein folding: the role of hydrophilic bridges in protein associations. J Mol Biol 1997;265:68–84.
38. Xu D, Tsai CJ, Nussinov R. Hydrogen bonds and salt bridges across protein–protein interfaces. Protein Eng 1997;10:999–1012.
39. Ivanov YD, Kanaeva IP, Karuzina IP, Archakov AI. Molecular recognition in the p450cam monooxygenase system: direct monitoring of protein–protein interactions by using optical biosensor. Arch Biochem Biophys 2001;391:255–264.
40. Vijayakumar M, Wong KY, Schreiber G, Fersht AR. Electrostatic enhancement of diffusion-controlled protein–protein association: comparison of theory and experiment on barnase and barstar. J Mol Biol 1999;278:1015–1024.
41. Camacho CJ, Weng ZP, Vajda S. Free energy landscapes of encounter complexes in protein–protein association. Biophys J 1999;76: 1166–1178.
42. Archakov AI, Ivanov YD, Biophysics of electron transfer and molecular bioelectronics. New York: Plenum Publication; 1999. p 173.
43. Wade RC, Gabdoulline RR, Ludemann SK, Lounnas V. Electrostatic steering and ionic tethering in enzyme-ligand binding: insights from simulation. Proc Natl Acad Sci USA 1998;95:5942–5949.
44. Zhang C, Chen J, DeLisi C. Protein–protein recognition: exploring the energy funnels near the binding sites. Proteins 1999;34:255–267.
45. Hu Z, Ma B, Wolfson H, Nussinov R. Conservation of polar residues as hot spots at protein interfaces. Proteins 2000;39:331–342.
46. Jones S, Thornton JM. Analysis of protein–protein interaction sites using surface patches. J Mol Biol 1997;272:121–132.
47. Jones S, Thornton JM. Prediction of protein–protein interaction sites using patch analysis. J Mol Biol 1997;272:133–143.
48. Laskowski RA, Luscombe NM, Swindells MB, Thornton JM. Protein clefts in molecular recognition and function. Protein Sci 1996;5:2438–2452.
49. Vaughan CK, Buckle AM, Fersht AR. Structural response to mutation at a protein–protein interface. J Mol Biol 1999;286: 1487–1506.

DOI 10.1002/prot

50. Davies DR, Cohen GH. Interactions of protein antigens with antibodies. Proc Natl Acad Sci USA 1996;93:7–12.
51. Clackson T, Wells JÁ. A hot spot of binding energy in a hormone–receptor interface. Science 1995;267:383–386.
52. DeLano WL. The PyMOL molecular graphics system. San Carlos, CA: DeLano Scientific; 2002. Available at: http://www.pymol.org
53. DeLano WL, Ultsch MH, de Vos AM, Wells JÁ. Convergent solutions to binding at a protein–protein interface. Science 2000;287: 1279–1283.
54. DeLano WL. Unraveling hot spots in binding interfaces: progress and challenges. Curr Opin Struct Biol 2002;12:14–20.
55. Thorn KS, Bogan AA. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. Bioinformatics 2001;17:284–285. Also available at http://thornlab.cgr.harvard.edu/hotspot/index.php
56. Wells JA. Additivity of mutational effects in proteins. Biochemistry 1990;29:8509–8517.
57. Cunningham BC, Wells JA. Highresolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science 1989;244:1081–1085.
58. Cunningham BC, Wells JA. Comparison of a structural and a functional epitope. J Mol Biol 1993;234:554–563.
59. Skolnick J, Fetrow JS, Kolinski A. Structural genomics and its importance for gene function analysis. Nat Biotechnol 2000;18: 283–287.
60. Morrison KL, Weiss GA. Combinatorial alanine-scanning. Curr Opin Chem Biol 2001;5:302–307.
61. Wells JA. Systematic mutational analyses of protein–protein interfaces. Methods Enzymol 1991;202:390–411.
62. Wells JA, De Vos AM. Structure and function of human growth hormone: implications for the hematopoietins. Annu Rev Biophys Biomol Struct 1993;22:329–351.
63. Wells JA. Structural and functional basis for hormone binding and receptor oligomerization. Curr Opin Cell Biol 1994;6:163–173.
64. Well JA. Binding in the growth hormone receptor complex. Proc Natl Acad Sci 1996;93:1–6.
65. Thornton JM. The Hans Neurath Award lecture of The Protein Society: proteins—a testament to physics, chemistry, and evolution. Protein Sci 2001;10:3–11.
66. Ma B, Wolfson HJ, Nussinov R. Protein functional epitopes: hot spots, dynamics and combinatorial libraries. Curr Opin Struct Biol 2001;11:364–369.
67. Buckle AM, Schreiber G, Fersht AR. Protein–protein recognition: crystal structural analysis of a barnase–barstar complex at 2.0-Å resolution. Biochemistry 1994;33:8878–8889.
68. Ma B, Shatsky M, Wolfson, Nussinov R. Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. Protein Sci 2002;11:185–197.
69. Barnett P, Bottger G, Klein AT, Tabak HF, Distel B. The peroxisomal membrane protein Pex13p shows a novel mode of SH3 interaction. EMBO J 2000;19:6382–6391.
70. Harris BZ, Hillier BJ, Lim WA. Energetic determinants of internal motif recognition by PDZ domains. Biochemistry 2001;40:5921–5930.
71. Luque I, Freire E. Structural stability of binding sites: consequences for binding affinity and allosteric effects. Proteins 2000;54:63–71.
72. Arkin MR, Wells AJ. Small-molecules inhibitors of protein–protein interactions: progressing towards the dream. Drug Discov 2004;3:301–317.
73. Arkin MR, Randal M, DeLano WL, Hyde J, Luong TN, Oslob JD, Raphael DR, Taylor L, Wang J, McDowell RS, Wells JA, Braisted AC. Binding of small molecules to an adaptive protein–protein interface. Proc Natl Acad Sci USA 2003;100:1603–1608.
74. Li W, Hamill SJ, Hemmings AM, Moore GR, James R, Kleanthous C. Dual recognition and the role of specificity-determining residues in colicin E9 DNase–immunity protein interactions. Biochemistry 1998;37:11771–11779.
75. Lichtarge O, Bourne HR, Cohen FE. An evolutionary trace method defines binding surfaces common to protein families. J Mol Biol 1996;257:342–358.
76. Samanta U, Pal D, Chakrabarti P. Environment of tryptophan side chains in proteins. Proteins 2000;38:288–300.
77. Fernandez A. Desolvation shell of hydrogen bonds in folded proteins, protein complexes and folding pathways. FEBS Lett 2002;527:166–170.
78. Lee KH, Xie D, Freire E, Amzel LM. Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. Proteins 1994;20:68–84.
79. Hubbard SJ, Argos P. Cavities and packing at protein interfaces. Protein Sci 1994;3:2194–2206.
80. Binkowski T, Adamian L, Liang J. Inferring functional relationships of proteins from local sequence and spatial surface patterns. J Mol Biol 2003;332:505–526.
81. Moreira IS, Fernandes PA, Ramos MJ. Detailed microscopic study of the full ZipA:FtsZ interface. Proteins 2006; 2006:63:811–821.
82. Moreira IS, Fernandes PA, Ramos MJ. Unravelling the importance of protein–protein interaction: application of a computational alanine scanning mutagenesis to the study of the IgG1: Streptococcal Protein G (C2 Fragment) complex. J Chem Phys B 2006;110: 10962–10969.
83. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:33–38.
84. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. J Mol Biol 1993;234:946–950.
85. Del Sol Mesa A, Pazos F, Valencia A. Automatic methods for predicting functionally important residues. J Mol Biol 2003;326: 1289–1302.
86. Goh CS, Cohen FE. Co-evolutionary analysis reveals insights into protein–protein interactions. J Mol Biol 2002;324:177–192.
87. Glaser F, Steinberg DM, Vakser IA, Tal N. Residue frequencies and pairing preferences at protein–protein interfaces. Proteins 2001;43: 89–102.
88. Halperin I, Wolfson H, Nussinov R. Protein–protein interactions: coupling of structurally conserved residues and of hot spots across protein–protein interfaces. Structure 2004;12:1027–1038.
89. Lee T, Kollman PA. Theoretical studies suggest a new antifolate as a more potent inhibitor of thymidylate synthase. J Am Chem Soc 2000;122:4385–4393.
90. Forman-Kay JD. The 'dynamics' in the thermodynamics of binding. Nat Struct Biol 1999;6:1086–1087.
91. Chakrabarti P, Jani J. Dissecting protein–protein recognition sites. Proteins 2002;47:334–343.
92. Janin J. Wet and dry interfaces: the role of solvent in protein–protein and protein–DNA recognition. Struct Fold Des 1999;7:R277–R279.
93. Moreira IS, Fernandes PA, Ramos MJ. Hot spots occlusion from bulk water—a comprehensive study of the complex between the lyzozyme HEL and the antibody FVD1.3. J Phys Chem B 2007;111: 2697–2706.
94. Carter PJ, Winter G, Wilkinson AJ, Fersht AR. The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). Cell 1984;38: 835–840.
95. Horovitz A. Double-mutant cycles: a powerful tool for analyzing protein structure and function. Fold Des 1986;1:R121–R126.
96. Pal G, Ultsch MH, Clark KP, Kossiakoff AA, Sidhu SS. Intramolecular cooperativity in a protein binding site assessed by combinatorial shotgun scanning mutagenesis. J Mol Biol 2005;347:489–494.
97. Lu SM, Lu W, Qasim MA, Anderson S, Apostol I, Ardelt W. Predicting the reactivity of proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases. Proc Natl Acad Sci USA 2001;98:1410–1415.
98. Gregoret LM, Sauer RT. Additivity of mutant effects assessed by binomial mutagenesis. Proc Natl Acad Sci USA 1993;90:4246–4250.

99. Bernat B, Sun M, Dwyer M, Feldkamp M, Kossiakoff AA. Dissecting the binding energy epitope of a high-affinity variant of human growth hormone: cooperative and additive effects from combining mutations from independently selected phage display mutagenesis libraries. Biochemistry 2004;43:6076–6084.
100. Greenspan NS, Di Cera E. Defining epitopes: it's not as easy as it seems. Nat Biotechnol 1999;17:936–937.
101. Jin L, Wells JA. Dissecting the energetics of an antibody–antigen interface by alanine shaving and molecular grafting. Protein Sci 1994;3:2351–2357.
102. Clackson T, Ultsch MH, Wells JA, de Vos JA. Structural and functional analysis of the 1:1 growth hormone: receptor complex reveals the molecular basis for receptor affinity. J Mol Biol 1998;277:1111–1128.
103. Tedford HW, Fletcher JI, King GF. Functional significance of the β hairpin in the insecticidal neurotoxin omega-atracotoxin-Hv1a. J Biol Chem 2001;276:26568–26576.
104. Erlanson DA, Braisted AC, Raphael DR, Randal M, Stroud RM, Gordon EM, Wells JA. Site-directed ligand discovery. Proc Natl Acad Sci USA 2000;97:9367–9372.
105. Sunesis Pharmaceuticals, Inc. Available at http://www.sunesis.com
106. Piehler J, Schreiber G. Fast transient cytokine–receptor interactions monitored in real time by reflectometric interference spectroscopy. Anal Biochem 2001;289:173–186.
107. Droit A, Poirier GG, Hunter JM. Experimental and bioinformatic approaches for interrogating protein–protein interactions to determine protein function. J Mol Endocrinol 2005;34:263–280.
108. Zhu H, Snyder M. Protein chip technology. Curr Opin Cell Biol 2003;7:55–63.
109. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein–proteins complexes. Proc Natl Acad Sci USA 2002;99:14116–14121.
110. Gilson MK, Given JA, Bush BL, McCammon JA. The statistical-thermodynamic basis for computation of binding affinities: a critical review. Biophys J 1997;72:1047–1069.
111. Swanson JMJ, Henchman RH, McCammon JA. Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. Biophys J 2004;86:67–74.
112. Simonson T, Archontis G, Karplus M. Free energy simulations come of age: protein–ligand recognition. Act Chem Res 2002;35: 430–437.
113. Moreira IS, Fernandes PA, Ramos MJ. Computational Determination of the relative free energy of binding—application to alanine scanning mutagenesis in molecular materials with specific interactions. In: Andrzej Sokalski W, editor. Modeling and design. Springer; 2007.
114. Kollman PA. Free energy calculations-applications to chemical and biochemical phenomena. Chem Rev 1993;93:2395–2317.
115. Gouda H, Kuntz I, Case D, Kollman PA. Free energy calculations for theophylline binding to an RNA aptamer: comparison of MM-PBSA and thermodynamic integration methods. Biopolymers 2002; 68:16–34.
116. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE III. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 2000;33:889–897.
117. Kong XJ, Brooks CL. Lambda dynamics—a new approach to free energy calculations. J Chem Phys 1996;105:2414–2423.
118. Moreira IS, Fernandes PA, Ramos MJ. Accuracy of the numerical solution of the Poisson-Boltzmann equation. J Mol Struct (THEOCHEM) 2005;729:11–18.
119. Moreira IS, Fernandes PA, Ramos MJ. Hot spots computational identification–application to the complex formed between the hen egg-white lysozyme (HEL) and the antibody HyHEL-10. Int J Quantum Chem 2006;107:299–310.
120. Moreira IS, Fernandes PA, Ramos MJ. Unravelling hot spots—a comprehensive computational mutagenesis study. Theor Chem Acc 2007;117:99–113.

DOI 10.1002/prot