Unraveling 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

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

Received: August 23, 2005; In Final Form: March 12, 2006

Alanine-scanning mutagenesis of protein−protein interfacial residues is a very important process for rational drug design. In this study, we have used the improved MM-PBSA approach that combining molecular mechanics and continuum solvent permits one to calculate the free energy differences through alanine mutation. To identify the binding determinants of the complex formed between the IgG1 (immunoglobulin-binding protein G) and protein G, we have extended the experimental alanine scanning mutagenesis study to both proteins of this complex and, therefore, to all interfacial residues of this binding complex. As a result, we present new residues that can be characterized as warm spots and, therefore, are important for complex formation. We have further increased the understanding of the functionality of this improved computational alanine-scanning mutagenesis approach testing its sensitivity to a protein−protein complex with an interface made up of residues mainly polar. In this study, we also have improved the method for the detection of an important amino acid residue that frequently constitutes a hot spot— tryptophan.

Introduction

Protein−protein interactions form the basis for most biological processes including events such as intercellular communication and programmed cell death.[1] A diverse study of protein−protein complexes have shown that shape complementarity is important for complex formation and that it is usually accompanied by a high degree of chemical complementarity.[1,2] The factors affecting protein shape complementarity include the size of the buried interface, alignment of polar and nonpolar residues, and number of buried waters.[3]

Crystallographic structures and alanine scanning mutagenesis of protein−protein interfacial residues have generated a large amount of information that allowed the discovery of energetically important determinants of specificity at intermolecular protein interfaces. They are compact, centralized regions of residues crucial for protein association and have been named hot spots.[1,4]

Therefore, a hot spot has been defined as an amino acid residue where alanine mutations cause an increase in the binding free energy larger than 4.0 kcal/mol, even though lower values are used for statistical analyses.[5,6] The warm spots are those with binding free energy differences between 2.0 and 4.0 kcal/mol, and the null spots are the residues with binding free energy differences lower than 2.0 kcal/mol.[5] There is some tendency for these residues (the hot spots or the warm spots) to be close together and organized in clusters.[5]

Hot spots are generally located near the center of protein−protein interfaces, away from the solvent, and the residues that most frequently form a hot spot are tryptophan (21%), arginine (13%), and tyrosine (12%). Since most protein−protein interaction surfaces are flat, it has been proposed that the exclusion of the hot spot from water requires that these residues are surrounded by a set of contacts that are energetically unimportant forming an O-ring structure.[7]

One of the most interesting features of hot spots is that they are complementary to each other, with buried charged residues forming salt bridges and hydrophobic residues from one surface fitting into depressions of the binding partner.[1] Hot spots present high functional and structural adaptability. Different proteins partners tend to bind to the same hot spot, which adapts to present the same residues in different structural contexts.[1,8]

Alanine-scanning mutagenesis of protein−protein interfacial residues continues to induce interest since further understanding of the nature of binding in complexes, in terms of the diverse biophysical features of the process, is essential to a phenomenological interpretation of the results. The reliable prediction of key residues in the interface has immediate applications in protein engineering, and it is an attractive alternative therapy for many diseases (structure-based drug design).[9]

Even though binding free energies can be measured experimentally, the quantification of the free energy components, essential to a phenomenological interpretation of the results, is hard to obtain.[10] Theoretical and computational methods represent a technique capable of resolving this type of problem.[11] Therefore, great effort has been invested in attaining a computational method to predict $\Delta\Delta G_{binding}$ upon alanine mutation that is simultaneously detailed and fully atomistic and with a high rate of success. In a recent paper we have published a method capable of achieving an overall success rate of 80% and a 100% success rate in residues for which alanine mutation causes an increase in the binding free energy higher than 4.0 kcal/mol (hot spots).[11]

In this paper, we stress the advantage of a computational approach to a comprehensive molecular thermodynamic view of affinity and, consequently, to detect drug resistant mutations. Our study defines the energetic contributions of all residues of

* Corresponding author. E-mail: mjramos@fc.up.pt.

10.1021/jp054760d CCC: $33.50
Published on Web 05/12/2006

this complex interface to achieve a deeper knowledge of the key determinants leading to new therapeutics strategies for streptococcal infections and novel immunochemical reagents.

Immunoglobulin (Ig)-binding bacterial proteins are very important for a diversity of biochemical and immunochemical functions.[12] Numerous infectious microorganisms exhibit proteins on their cell surfaces which interact with mammalian immunoglobulins (Igs) allowing the pathogenic bacterium to evade the host immune response by coating the invading bacteria with host antibodies.[13] Binding to Igs is expected to weaken the host's immune response and to make infection possible.[13]

The X-ray structure of a complex between the IgG1 and the 56 amino acid Fc-binding domain (C2 fragment) of streptococcal protein G shows that about 13 residues on its surface are directly involved in the binding site.[14] Compared to the common protein−protein interactions, the Fc binding site on the C2 fragment of streptococcal protein G is rather uncharacteristic because it is predominantly polar rather than hydrophobic.[15] The interfacial area is lower than the medium value observed (700 Å^2 rather than the average 1200 Å^2),[16] and the interface is not planar as is typically observed in heterodimers. Instead, it is formed by a double "knobs-into-holes" interaction in which a knob from the protein G protrudes into a hole in the IgG1, and vice versa.[17,18]

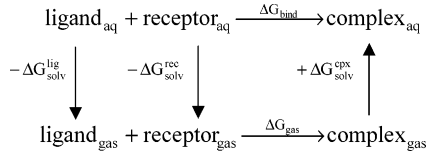
In this paper, we show the advantage of a computational approach to systematically screen a receptor−ligand interface and to provide a molecular view of the structural and energetic consequences of mutations. This computational approach leads to the interpretation of the mutations in terms of the contributions of the electrostatic and van der Waals interactions and those of solvation.

Thus, the key aim of this study is to obtain a deeper knowledge of this protein−protein interface, subsequently allowing control over complex formation and guiding new experimental investigations. This study contributes also to increase the understanding of the functionality of this computational alanine-scanning mutagenesis approach. We have also further studied ways to improve the success rate for an important amino acid residue that commonly constitutes a hot spot—tryptophan. This method is simple and fast, has a low computational cost, and can be applied to a wide range of proteins providing a correct anatomic image of an interface. It is important to highlight that the complex study in this paper is not common, having a highly polar protein−protein interface. Therefore, we are going to show the competence of the method, which has already been proven with more nonpolar interfaces, by using it to perform an alanine-scanning mutagenesis study in an atypical polar interface. Hence, this method can be used for any type of protein−protein interface prior to an experimental mutagenesis investigation helping in the hot spot detection and the choice of the amino acids to mutate.

Material And Methods

1. Model Setup. The starting crystallographic structure for the simulation, the complex formed between a human immunoglobulin IgG complexed with the C2 fragment of streptococcal protein G, was taken from the RCSB Protein Data Bank with PDB entry 1fcc[14] and with a resolution of 3.50 Å. For the alanine-scanning mutagenesis, we have followed a protocol already established to give a coherent agreement with the experimental results.[11] In the molecular simulations the solvent was modeled through a modified generalized Born solvation model[19] being the structure first minimized with 1000 steps of steepest decent followed by 1000 steps of conjugated gradient to release the bad contacts in the crystallographic structure. Subsequently a 10000 ps molecular dynamics (MD) simulation was performed starting from the minimized structure. All molecular mechanics simulations presented in this work were carried out using the sander module, implemented in the Amber8[20] simulations package, with the Cornell force field.[21] Bond lengths involving hydrogens were constrained using the SHAKE algorithm,[22] and the equations of motion were integrated with a 2 fs time-step being the nonbonded interactions truncated with a 16 Å cutoff. The temperature of the system was regulated by the Langevin thermostat.[23−25]

SCHEME 1: Thermodynamic Cycle Used To Calculate the Complexation Free Energy



The MM-PBSA script implemented in Amber8[20] was used to calculate the binding free energies for the complex and for the alanine mutants. The MM-PBSA script[26] was used to perform a postprocessing treatment of the complex by using the structure of the complex and calculating the respective energies for the complex and all interacting monomers. To generate the structure of the mutant complex a simple truncation of the mutated side chain was made, replacing Cγ with a hydrogen atom and setting the Cβ-H bond direction to that of the former Cβ-Cγ. For the binding free energy calculations, 375 snapshots of the complexes were extracted every 20 ps for the last 7500 ps of the run.

2. Alanine-Scanning Mutagenesis. The complexation free energy can be calculated using the thermodynamic cycle in Scheme 1, where ΔG_{gas} is the interaction free energy between the ligand and the receptor in the gas phase and ΔG^{lig}_{solv}, ΔG^{rec}_{solv}, and ΔG^{cpx}_{solv} are the solvation free energies of the ligand, the receptor, and the complex, respectively. The binding free energy difference between the mutant and wild-type complexes is defined as

$$\Delta\Delta G_{binding} = \Delta G_{binding-mutant} - \Delta G_{binding-wild\ type} \quad (1)$$

The binding free energy of two molecules is the difference between the free energy of the complex and the respective monomers (the receptor and the ligand):

$$\Delta G_{binding-molecule} = G_{complex} - (G_{receptor} + G_{ligand}) \quad (2)$$

The free energy of the complex and respective monomers can be calculated by summing the internal energy (bond, angle. and dihedral), the electrostatic and the van der Waals interactions, the free energy of polar solvation, the free energy of nonpolar solvation, and the entropic contribution for the molecule free energy:

$$G_{molecule} = E_{internal} + E_{electrostatic} + E_{vdw} + G_{polar\ solvation} + G_{nonpolar\ solvation} - TS \quad (3)$$

The first three terms were calculated using the Cornell force field[22] with no cutoff. The electrostatic solvation free energy was calculated by solving the Poisson−Boltzmann equation with the software Delphi v.4,[27,28] using the same methodology of previous works which has been shown in an earlier work to constitute a good compromise between accuracy and computing time.[29] For the energy calculations three internal dielectric

TABLE 1: Results of the Computational Alanine Screening Mutagenesis[a]

mutation	$\Delta\Delta E_{electrostatic}$	$\Delta\Delta E_{vdw}$	$\Delta\Delta G_{nonpolar\ solvation}$	$\Delta\Delta G_{polar\ solvation}$	$\Delta\Delta G_{binding}$	$\Delta\Delta G_{exp}$
			Protein G (C2 Fragment)			
Thr25Ala	−0.25	0.30	0.05	−0.01	0.09	0.24
Glu27Ala	14.78	−0.18	0.02	−4.38	10.23	>4.90
Lys28Ala	23.2	2.72	0.64	−23.54	3.01	1.30
Lys31Ala	11.09	5.85	0.24	−12.55	4.64	3.50
Asn35Ala	1.16	2.27	0.33	−2.52	1.25	2.40
Asp40Ala	−2.40	1.85	0.10	0.42	−0.04	0.30
Glu42Ala	−2.21	0.31	0.05	1.78	−0.08	0.40
Trp43Ala ($\epsilon = 2$)	1.78	3.18	−0.02	−4.44	0.47	3.80
Trp43Ala ($\epsilon = 8$)	0.44	3.18	−0.02	−0.94	2.66	3.80
Thr44Ala/Tyr45Ala	0.43	2.67	0.11	−0.99	2.24	2.00
			IgG1			
Ile253Ala	0.68	4.13	0.23	−1.82	3.22	NA
Ser254Ala	6.94	−1.20	0.00	−2.61	3.13	NA
Glu380Ala	4.86	−0.25	0.01	−3.69	0.92	NA
Glu382Ala	8.62	−0.29	0.05	−7.67	0.70	NA
Ser426Ala	−1.77	0.16	0.00	1.82	0.21	NA
Met428Ala	0.21	0.92	0.02	−1.10	0.05	NA
His433Ala	1.27	5.30	0.34	−3.05	3.86	NA
Asn434Ala	1.18	4.90	0.13	−4.56	1.65	NA
His435Ala	1.65	2.12	0.12	−3.64	0.24	NA
Tyr436Ala	1.77	5.77	0.33	−4.17	3.71	NA
Thr437Ala	−0.07	0.14	0.02	0.05	0.13	NA
Gln438Ala9	−1.08	0.75	0.06	0.60	0.32	NA

[a] The units of free energies and potential energies are kcal/mol. NA = not available.

constant values, exclusively characteristic of the mutated amino acids, were used: 2 for the nonpolar amino acids; 3 for the polar residues; 4 for the charged amino acids and histidine.[11] Recalling that we used only one trajectory for the computational energy analyses, it is important to highlight that side chain reorientation is not included explicitly in the formalism. As amino acid polarity increases, the structural effect beyond the neighbor residues also increases, and the conformational reorganization after alanine mutagenesis should be more extensive. This effect can be mimicked with the use of a set of three different internal dielectric values.

The nonpolar contribution to solvation free energy due to van der Waals interactions between the solute and the solvent and cavity formation was modeled as a term that is dependent on the solvent-accessible surface area of the molecule. It was estimated using an empirical relation, $\Delta G_{nonpolar} = \sigma A + \beta$, where A is the solvent-accessible surface area that was estimated using the molsurf program, which is based on the idea primarily developed by Connolly.[30] σ and β are empirical constants, and the values used were 0.005 42 kcal Å^{-2} mol^{-1} and 0.92 kcal mol^{-1}, respectively. The entropy term, obtained as the sum of translational, rotational, and vibrational components, was not calculated because it was assumed, on the basis of a previous work, that its contribution to $\Delta\Delta G_{binding}$ is negligible.[26]

The IC_{50} values of the experimental alanine mutagenesis data[31] were used to calculate the experimental binding free energies using the relationship $\Delta G_{binding} = RT \ln IC_{50}$, where R is the ideal gas constant and T is the temperature in K.

Results

The accurate theoretical prediction of binding affinities is one of the most significant problems in computational chemistry. The improved[11] MM-PBSA is a fully atomistic method that merges molecular mechanics and continuum solvent as the foundation for the methodology to explore protein−protein interactions by calculating differences in binding free energy upon alanine mutation.[26,32−39] This fully atomistic computational approach consists of performing a molecular dynamics simulation (MD) using a single trajectory of the wild-type structure to calculate $\Delta\Delta G_{binding}$, and it complements experimental analysis adding molecular insight into the macroscopic properties measured therein and decomposing $\Delta\Delta G_{binding}$ into its internal, interaction, and solvation components.

To obtain trustworthy estimates of the free binding energy differences of the 1fcc complex we had to make sure that the proteins were fully equilibrated. Hence, to assess the quality of the simulations, we monitored the root-mean-square deviations (rmsd) of the trajectory structures compared to the starting structures. In Figure 1 we have plotted the rmsd from the X-ray crystal structure of the backbone atoms of the complex and the respective separate proteins for the dynamics simulation. It can be observed that, after an initial period of about 2500 ps of equilibration, the system has achieved equilibrium remaining stable. It is important to highlight that the structure of the interface, which is going to be subjected to the alanine-scanning mutagenesis study, remains very stable (rmsd < 2 Å).

Table 1 resumes the results of the computational alanine-scanning mutagenesis study of the protein−protein complex formed between IG γ-1 chain c region with the streptococcal protein G (C2 fragment).

To determine the contribution of each residue in the protein−protein interface and to identify key residues, computational alanine-scanning mutagenesis was extended to the whole interface. The first nine interaction free energies in this table are energies for residues for which experimental values exist, assessing the worth of the computational alanine-scanning mutagenesis approach.

If we consider a deviation of ±1.3 kcal/mol from the experimental value as an accurate result (to predict binding affinities within an accuracy of 1 order of magnitude) and omit Glu27 (which has not a defined experimental value), we obtain an overall success rate of either 75% or 88%, depending of the ϵ value (2 or 8, respectively) used to treat the tryptophan residue (whose appropriate dielectric constant has not been parametrized before, as explained below). This corresponds to experimental/theoretical correlation coefficients of 0.29 and 0.65, respectively. The standard deviation of the mean for the $\Delta\Delta G_{binding}$, defined as $\sigma/\sqrt{n}$, where n is the number of snapshots, ranges from 0.01

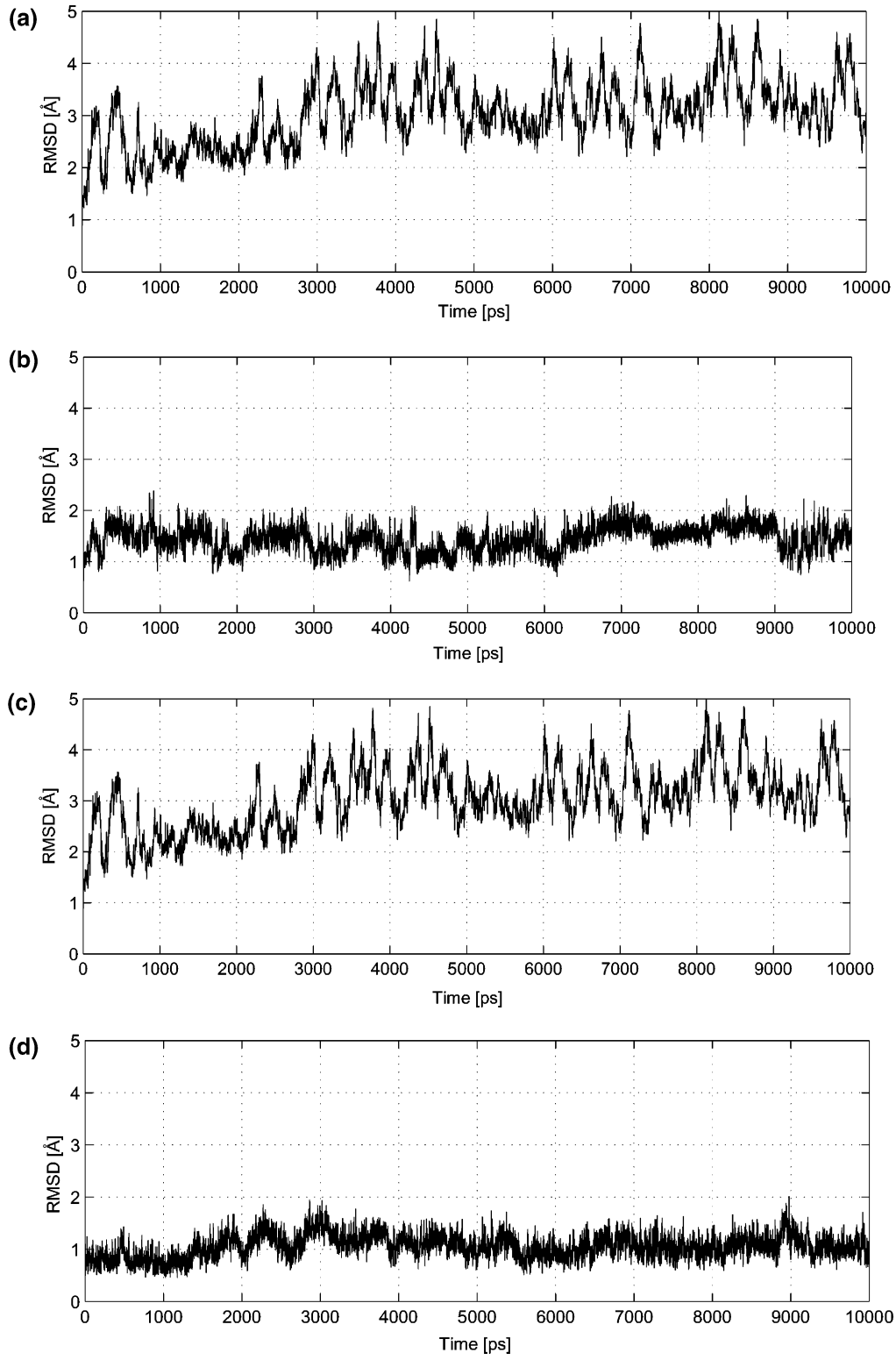


Figure 1. Rmsd plots for the protein backbone of the complex formed between the human immunoglobulin IgG and the C2 fragment of streptococcal protein G relative to its initial structure: (a) the complex; (b) the ligand; (c) the receptor; (d) the interface.

to 0.09 kcal/mol for all the mutations analyzed. Depending on the ϵ value (2 or 8), the method gives a mean and maximum unsigned error of 1.07 and 3.33 kcal/mol or 0.79 and 1.71, respectively. As result, the predictions of this work are validated by the correct reproduction of the experimental values.

It is important to recall that one of the objectives of this work is to expand the knowledge of the way the improved method holds. Therefore, we tested the precision of this theoretical approach by calculating the $\Delta\Delta G_{\text{binding}}$ every 500 ps after equilibration for the nine mutations for which exist experimental values. This way we have tested the sensitivity of the $\Delta\Delta G_{\text{binding}}$ to the use of different sets of macromolecular structures. In Figure 2, we have plotted the $\Delta\Delta G_{\text{binding}}$ as a function of simulation time.

We have also further studied an important amino acid residue that commonly constitutes a hot spot, tryptophan. Tryptophan is a large aromatic residue, and it is the most important residue in terms of both conservation and free energy binding difference

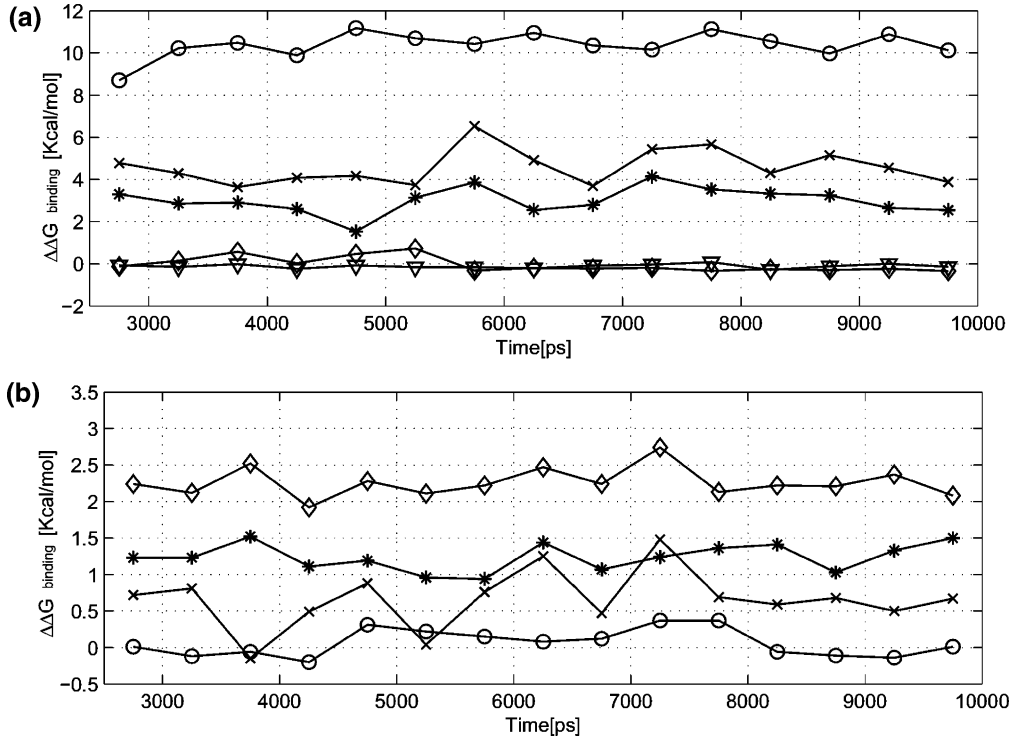


Figure 2. 500 ps averages of $\Delta\Delta G_{binding}$ after 2500 ps of equilibration for the nine mutations for which an experimental value exist as a function of simulation time: (a) (○) $\Delta\Delta G_{binding}$ of Glu27Ala, (∗) $\Delta\Delta G_{binding}$ of Lys28Ala, (×) $\Delta\Delta G_{binding}$ of Lys31Ala, (◇) $\Delta\Delta G_{binding}$ of Asp40Ala, (▽) $\Delta\Delta G_{binding}$ of Glu42Ala; (b) (○) $\Delta\Delta G_{binding}$ of Thr25Ala, (∗) $\Delta\Delta G_{binding}$ of Asn35Ala, (×) $\Delta\Delta G_{binding}$ of Trp43Ala, (◇) $\Delta\Delta G_{binding}$ of Thr44Ala/Tyr45Ala.

upon alanine mutation.[40] This fully atomistic computational method consists of using computational MD simulations with a single trajectory protocol. A set of different internal dielectric constants have to be used for the proteins, depending on the kind of amino acid that is mutated to account for the different degree of relaxation of the interface when different types of amino acids are mutated for alanine.[12] Therefore, for the nonpolar amino acids we established a dielectric constant of 2. However, despite the fact that this value proved to work for the majority of these kinds of residues, for the tryptophan amino acid residue it is predictable that a larger value of ϵ should be used to mimic the structural relaxation upon alanine mutation. Therefore as expected, with an internal dielectric constant of 2 the $\Delta\Delta G_{binding}$ value obtained was of 0.47 kcal/mol presenting an error of 3.33 kcal/mol. After trying a different internal dielectric constant (ϵ), we have achieved a $\Delta\Delta G_{binding}$ value of 2.66 kcal/mol with a ϵ value of 8. Tryptophan can contribute with aromatic π-interactions, it is a hydrogen-bonding donor, and has a large hydrophobic surface, and upon alanine mutation a large cavity is produced due to significant differences in size.[2,7] Consequently, the structural effect beyond the neighbor residues also increases, and the conformational reorganization after alanine mutagenesis should be more extensive assessing the necessity of raising the internal dielectric constant to mimic these effects. However, we should note that the exact value of 8 has been obtained through calibration with a single experimental value and further examples of such mutations can eventually lead in the future to slight adjustments of the value of the internal dielectric constant for tryptophan.

To fully understand the nature of binding between IG γ-1 chain c regions and the streptococcal protein G (C2 fragment), Table 1 present all the individual energy contributions to the free binding energy differences. According, to the use of a single trajectory protocol the internal energies cancel being the $\Delta\Delta E_{internal}$ value is 0. Analyzing Table 1, we observed that the nonpolar solvation contribution to the effective binding free energy difference is relatively small compared with the other contributing factors. The intermolecular electrostatic and the polar solvation free energy are important for protein−protein binding because they cancel each other's effect.

In the majority of the complexes, an analysis of the dominant interactions implies that van der Waals interactions and hydrophobic effects provide a reasonable basis for understanding binding affinities.[41] From the analysis of Table 1 it is perceived that the $\Delta\Delta E_{vdw}$ values are almost all positive indicating that the van der Waals interaction is favorable to the complex binding and that alanine mutation of the residues diminishes the vdW contacts at the interface.

It is important to highlight the fact that the binding site on the C2 fragment of streptococcal protein G is slightly atypical because it is predominantly a polar environment rather than hydrophobic.[15] Therefore, the specificity of association is deeply influenced by the presence of polar residues because electrostatic complementarity between the individual molecules further optimizes binding.[42,43] By examination of Table 1, it can be observed that the $\Delta\Delta G_{electrostatic}$ and the $\Delta\Delta G_{polar\ solvation}$ values for some of the polar residues are 10−15 kcal/mol higher than for the nonpolar amino acid. The high energy values are in agreement with the use of a dielectric constant of 4 to mimic the protein relaxation and reorganization as well as the electronic polarization that affects the charge−charge interactions.[11] During protein binding, the charged groups in the proteins desolvate as their environments change from aqueous to largely nonpolar solvent, causing huge electrostatic and polar solvation energy differences.[42] However, if we only consider their sum, we find that its value ranges between −3.0 and 3.0 kcal/mol. Generally, the contributions of the electrostatic energy together with polar solvation are usually negative showing that the alanine mutants bind better in term of this kind of interaction. Again, we can

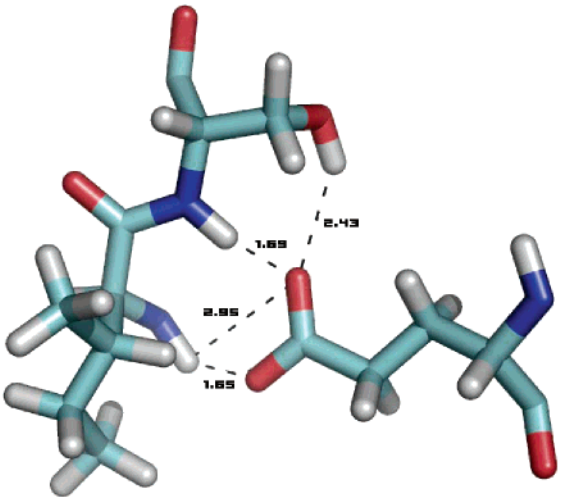


Figure 3. Detail of the molecular interactions involving the hot spot Glu27 of the C2 fragment of streptococcal protein G and the warm spots Ser254 and Ile253 of the human immunoglobulin IgG. These residues form one of the two knob-into-hole packings and are colored by atom type. These residues form one of the two knob-into-hole structures of this interface with Glu27 protruding into the cavity formed by Ser254 and Ile253 (X-ray crystallographic structure).

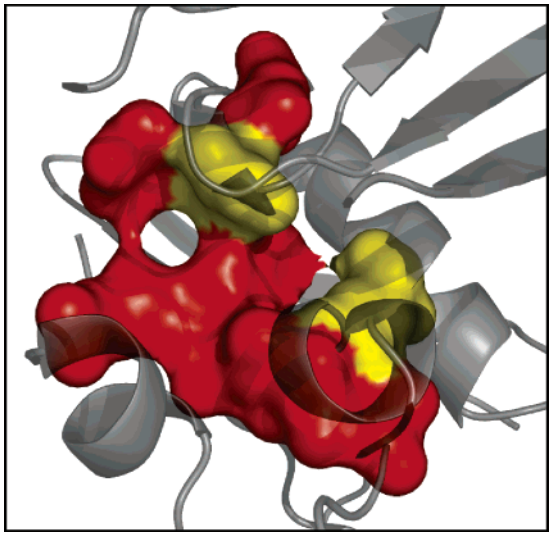


Figure 4. Surface representation of the complex formed between the human immunoglobulin IgG and the C2 fragment of streptococcal protein G highlighting the residues involved in the two knob-into-hole packings. In yellow are presented the two knobs (Glu27 and Asn434), and in red, the two holes (Ile253 and Ser254; Asn35, Asp40, Glu42, and Trp43).

say that changing the amino acid nature from hydrophilic to hydrophobic improves binding.

This protein−protein interface is formed by a double knobs-into-holes interaction in which a knob from the protein G projects into a hole in the IgG1, and vice versa.[17,18] Specificity is provided by the complementary shape of the interacting surfaces, fitting knobs into holes, and by the complementary arrangement of hydrogen-bonding groups and charge−charge pairs.

These knobs-into-holes interactions compose the main structural motif of the binding site of the complex. The hot spots are usually related with these structural motifs, and therefore, they should be complementary to each other, with buried charged residues forming salt bridges and hydrophobic residues from one surface fitting into depressions of the binding partner.[1] Thus, it is expected that the hot spot of a binding partner has a geometrical arrangement, which is complementary to the arrangement in the other, this fit being specific for the complex. On the basis of earlier results from alanine mutagenesis investigation obtained on numerous protein−protein interfacial residues[44] and the existence of this motif formation, it should be expected that functionally important residues on IgG1 are those that are in direct contact with the key binding determinants of protein G: Glu27, Lys31, Asn35, Trp43, Trp44, and Tyr45. The existence of such structural organization at the IgG1:PG(C2) interface could not be confirmed experimentally because the mutagenesis experiments have been only performed at one side of the interface. However, in this work we have extended the mutagenesis to the other side of the interface. The obtained results did reveal an amazing spatial complementarity between binding determinants, confirming the expected structural organization at the protein interface. If we examine the binding free energy values in Table 1, we can observe the existence of five complementary warm spots in protein IgG1: Ile253, Ser254, His433, Asn434, and Tyr436.

Glu27 of the C2 fragment fits into a hole delimited by Ile253 and Ser254 on the surface of the IgG1, where it forms hydrogen bonds with Ser254 side chain. In Figure 3, it is present as an image of the hydrogen bond interaction (2.43 Å) between the Ser254 Oγ atom and the Oϵ1 atom of Glu27 and the NH of the Ser254 backbone and the Oϵ1 of Glu27 with a distance of 1.69 Å.

There are also two hydrogen bonds between the NH atom of the main chain of Ile253 and the Oϵ1 and Oϵ2 atom of Glu27 with distances of 2.95 and 1.65 Å, respectively. These interactions contribute to the warm character of Ile253 and Ser254. Glu27 is the only hot spot present in the protein−protein complex with a experimental $\Delta\Delta G_{binding}$ higher than 4.5 kcal/mol. Given the fact that Glu27 establishes four hydrogen bonds with important residues of the other binding partner, it makes sense that this residue is the only hot spot. All the other important residues have a warm spot character.

The projection of the Asn434 side chain from the surface of IgG1 into a hole formed by Asn35, Asp40, Glu42, and Trp43 constitutes the second knobs-into-holes structural motif. From the IgG1 it is important to emphasize the contribution to this structural motif of residues His433, Asn434, and Tyr436. Therefore, as it can be seen in Figure 4 from the X-ray crystallographic structure, the protein−protein complex has two knobs-into-holes motifs that reduces the distance between residues from both binding partners allowing the formation of important hydrogen and salt bond interactions that provide the warm or/and hot spot character to their amino acid residues.

To evaluate the possibility of cooperativity between the amino acid residues present in the binding interface we have also performed simultaneous alanine mutagenesis (alanine shaving), and the results are presented in Table 2.

For the warm spots cluster of IgG1 (Ser254, His433, Asn434, and Tyr436) we have obtained a $\Delta\Delta G_{binding}$ value of 10.74 kcal/mol, lower than the one predicted by assuming simple additivity of the effects of individual alanine mutation. We also have done alanine shaving of the cluster of the IgG1 protein and the Asn35, Thr44, and Tyr45 of protein G, and once again, we detected a decrease in the binding free energy in comparison with the individual contribution sum. We also tested the effect of alanine shaving of the hole constituted by Ile253 and Ser254. Due to the different polarity of the residues involved we had to carry out two calculations with different internal dielectric constants (ϵ = 2 and 3). The results presented in Table 2 show a

TABLE 2: Results of Alanine Shaving (kcal/mol)

mutation	$\Delta\Delta G_{calculated}$	$\sum\Delta\Delta G_{calculated}$	cooperativity
Ser254Ala/His433Ala/Asn434Ala/Tyr436Ala	10.74	12.35	−1.61
Asn35Ala/Thr44Ala/Tyr45Ala/Ser254Ala/His433Ala/Asn434Ala/Tyr436Ala	13.47	15.84	−2.37
Asn35Ala/Thr44Ala/Tyr45Ala	3.30	3.49	−0.19
Thr25Ala/Ser426Ala	0.29	0.30	−0.01
Asp40Ala/Glu42Ala	0.08	-0.08	0.16
Ile253Ala/Ser254Ala ($\epsilon = 2$)	7.25	6.35	0.90
Ile253Ala/Ser254Ala ($\epsilon = 3$)	6.50	6.35	0.15

cooperative effect confirming the importance of this structure to complex formation. It is difficult to estimate the exact value of this cooperativity using the present protocol, although the calculations show that it lies between 0.15 and 0.90 kcal/mol. The molecular basis for cooperativity relies on the partial unshielding of the knot from the solvent, since the molecular interactions do not change with this protocol. Therefore, the cooperativity in this case can be seen as a measure of the importance of solvent shielding to the formation of the hot spots. To confirm the reliability of the alanine-shaving calculations, we did alanine shaving of null spot clusters of this complex (Thr25 and Ser426; Asp40 and Asp42) and we have confirmed that in this case the effects were additive. Consequently, we can conclude that the cooperative effect of the binding warm or/and hot spots detected in previous reports is not a universal truth, depending on the protein−protein complex studied.

An O-ring structure had been previously reported. This structure is supposed to lead to the exclusion of bulk solvent from the interacting residues by surrounding the hot spots by a set of contacts that are energetically unimportant.[9] In this complex, this kind of formation was not observed. In an earlier work, we had proposed that solvent occlusion could be achieved in two different ways: the formation of an O-ring if the protein has a planar, large interface that allows for such a macromolecular structure or, alternatively, when the interface is too small, with the formation of a hydrophobic pocket. Now we are prepared to expand this perspective; probably for proteins with a polar interface and with an intermediate size, solvent occlusion can be attained by the formation of a correct environment to shield the charge interaction—the knobs-into-holes. Hence, the main purpose of Nature is to protect the warm and hot spots from the aqueous surrounding environment by allowing the formation of important contacts to complex binding. The way this is achieved is not universal and can vary widely depending of the size of the interface and the nature of the residues that constitute it.

Conclusion

Alanine-scanning mutagenesis of protein−protein interfacial residues is a very important process for rational drug design. In this study, we have used the improved MM-PBSA approach that combining molecular mechanics and continuum solvent permits one to calculate the free energy differences upon alanine mutation.

We have extended the experimental alanine scanning mutagenesis study to both proteins of this complex and, therefore, to all interfacial residues of this binding complex to determine the contribution of each residue in the protein−protein interface, to investigate the special distribution and complementarity of the binding hot spots across the interface, and to identify the binding determinants of the complex formed between the IgG1 and protein G. As a result, we present new residues that can be characterized with warm spots and, therefore, are important for complex formation.

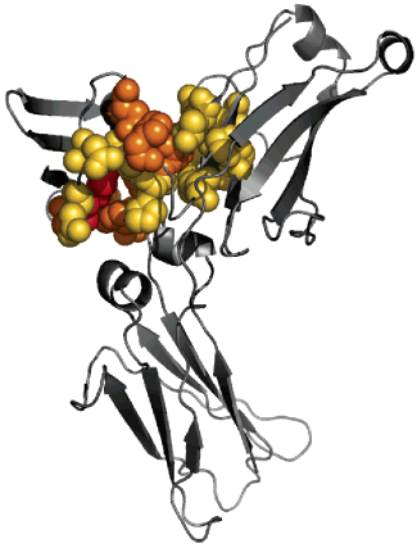


Figure 5. Complex formed between human immunoglobulin IgG and the C2 fragment of streptococcal protein G highlighting the mutated residues by a ball-and-stick representation. In yellow are represented the null spots (relative binding energy < 2.0 kcal/mol), in orange the warm spots (relative binding energy between 2.0 and 4.0 kcal/mol), and in red the hot spots (residues with a relative binding energy higher than 4.0 kcal/mol).

On the basis of earlier results from alanine mutagenesis investigation obtained on numerous protein−protein interfacial residues[44] and on the existence of this motif formation, it should be expected that functionally important residues on IgG1 are those that are in direct contact with the key binding determinants of the protein G: Glu27, Lys31, Asn35, Trp43, Trp44, and Tyr45. We have indeed confirmed such predictions and determined the major protein G-binding determinants, which are represented in Figure 5. Thus, we have obtained five warm spots in protein IgG1: Ile253, Ser254, His433, Asn434, and Tyr436.

The close proximity of residues of either binding partner provided by the two knobs-into-holes structural motifs allows the formation of important hydrogen and salt bond interactions. Consequently, these residues are warm or/and hot spots, being important for the complexation between protein G and IgG1.

This work contributes also to amplify the understanding of the functionality of this computational alanine-scanning mutagenesis approach. Therefore, we tested the precision of this theoretical approach by calculating the $\Delta\Delta G_{binding}$ every 500 ps after equilibration for the nine mutations for which exists an experimental value and we had concluded that as the $\Delta\Delta G_{binding}$ value for a specific mutation increases the fluctuations over time also increase. Therefore, the amino acid residues that present a higher $\Delta\Delta G_{binding}$ value have a higher oscillation over time of the $\Delta\Delta G_{binding}$ value.

It can be perceived by observation of these graphics that as the magnitude of $\Delta\Delta G_{binding}$ for a specific mutation rises, its

fluctuations also increase. Glu27Ala, Lys28Ala, and Lys31Ala are the three residues correctly determined with a higher $\Delta\Delta G_{binding}$ value. Therefore, it is natural that they present a larger amplitude of oscillation over time.

We have also further studied an important amino acid residue that commonly constitutes a hot spot—tryptophan. The use of a higher internal dielectric constant improves the results for this amino acid residue raising the $\Delta\Delta G_{binding}$ value from 0.47 to 2.66 kcal/mol because the large and aromatic character of this amino acid leads to an increase of the structural effect beyond the neighboring ones. The definition of a higher dielectric constant was predictable a priori, on the basis of the underlying philosophy of the method. Thus, the conformational reorganization after alanine mutagenesis is more extensive accenting the requirement of raising the internal dielectric constant to mimic these effects. A value of 8 has been shown to lead to good agreement with experimental results. Once that this value has been obtained with a single experiment, it could be questioned whether this result will hold for tryptophan in any other environment. Our experience with the other amino acid is that a single internal dielectric constant value suffices to accurately describe its effect upon alanine mutation. The exact value of 8 here proposed should, however, be further validated once more experimental data and theoretical calculations become accessible.

It is important to highlight that by combining this work with the one previously reported,[2] we have demonstrated the universal character of the method that works with a high success rate either with a nonpolar or a polar protein−protein interface environment. The deeper knowledge of the physical nature of the protein−protein interaction surface and the definition of the binding hot spots is invaluable to drug design. These mutational data can be used to design new drugs against infectious microorganisms.

References and Notes

(1) Arkin, M. R.; Wells, A. J. *Drug Discovery* **2004**, *3*, 301.
(2) DeLano, W. L.; Ultsch, M. H.; de Vos, A. M.; Wells, J. A. *Science* **2000**, *287*, 1279.
(3) Sharma, S. K.; Ramsey, T. M.; Bair, K. W. *Curr. Med. Chem. Anti-Cancer Agents* **2002**, *2*, 311.
(4) Michelle, R. A.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; McDowell, R. S.; Wells, J. A.; Braisted, A. C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 1603.
(5) Pons, J.; Rajpal, A.; Kirsch, J. *Protein Sci.* **1999**, *8*, 958.
(6) Keskin, O.; Ma, B.; Nussinov, R. *J. Mol. Biol.* **2005**, *345*, 1281.
(7) Bogan, A.; Thorn, K. S. *J. Mol. Biol.* **1998**, *280*, 1.
(8) Ma, B.; Shatsky, M.; Wolfson, H. J.; Nussinov, R. *Proteins* **2002**, *54*, 63.
(9) Gao, Y.; Wang, R.; Lia, L. *J. Mol. Model.* **2004**, *10*, 44.
(10) Jayaram, B.; Mcconnel, K.; Dixit, S. B.; Das, A.; Beveridge D. L. *J. Comput. Chem.* **2001**, *23*, 1.
(11) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. Submitted for publication.
(12) Boyle, N. D. P. *Bacterial Immunoglobulin Binding Proteins*; Academic Press: San Diego, CA, 1990.
(13) Goward, C. R.; Scawen, M. D.; Murphy, J. P. *Trends Biochem. Sci.* **1993**, *18*, 136.
(14) Sauer-Eriksson, A. E.; Kleywegt, G. J.; Uhlen, M.; Jones, T. A. *Structure* **1995**, *3*, 265.
(15) Stites, W. E. *Chem. Rev.* **1997**, *97*, 1233.
(16) Jones, S.; Thornton, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13.
(17) Crick, F. H. C. *Nature* **1952**, *170*, 882.
(18) Crick, F. H. C. *Acta Crystallogr.* **1953**, *6*, 689.
(19) Tsui, V.; Case, D. A. *Biopolymers* **2001**, *56*, 275.
(20) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, H. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, CA, 2004.
(21) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(22) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1997**, *23*, 327.
(23) Pastor, I. W. R.; Brooks, B. R.; Szabo, A. *J. Mol. Phys.* **1998**, *65*, 1409.
(24) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. *Biopolymers* **1992**, *32*, 523.
(25) Izaguirre, J. A.; Catarello, D. P.; Wozniak J. M.; Skeel, R. D. *J. Chem. Phys.* **2001**, *114*, 2090.
(26) Huo, S.; Massova, I.; Kollman, P. A. *J. Comput. Chem.* **2002**, *23*, 15.
(27) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. *J. Comput. Chem.* **2002**, *23*, 128.
(28) Rocchia, W.; Alexov, E.; Honig, B. *J. Phys. Chem. B* **2001**, *105*, 6507.
(29) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. *J. Mol. Struct. (THEOCHEM)* **2005**, *729*, 11.
(30) Connolly, M. L. *J. Appl. Crystallogr.* **1983**, *16*, 548.
(31) Sloan, D. J.; Hellinga, H. W. *Protein Sci.* **1999**, *8*, 1643.
(32) Kollman, P. A., 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, D. A.; Cheatham, T. E., III. *Acc. Chem. Res.* **2000**, *33*, 889.
(33) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. *Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211.
(34) Massova, I.; Kollman, P. A. *J. Am. Chem. Soc.* **1999**, *121*, 8133.
(35) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. *J. Am. Chem. Soc.* **2001**, *123*, 5221.
(36) Wang, W.; Kollman, P. A. *J. Mol. Biol.* **2000**, *303*, 567.
(37) Reyes, C. M.; Kollman, P. A. *J. Mol. Biol.* **2002**, *295*, 1.
(38) Cohlke, H.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 238.
(39) Xia, B.; Tsui, V.; Case, D. A.; Dyson, J.; Wright, P. E. *J. Biomol. NMR* **2002**, *22*, 317.
(40) Ma, B.; Elkayam, T.; Wolfson, H.; Nussinov, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5772.
(41) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9997.
(42) Sheinerman, F. B.; Norel, R.; Honig, B. *Curr. Opin. Struct. Biol.* **2002**, *10*, 153.
(43) Kumar, S., Nussinov, R. *Chembiochem* **2002**, *3*, 604.
(44) Clackson, T.; Wells, J. A. *Science* **1995**, *267*, 383.