Structural Determinants of a Typical Leucine-Rich Repeat Protein

Joao M. Martins, Rui M. Ramos and Irina S. Moreira*

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

Abstract. The structural and functional description of protein-protein complexes and their comprehension is a key concept, not only to increase the scientific knowledge in basic terms but also for the application to the biomedical and pharmaceutical industry. The binding association between proteins is nowadays attribute to a few key residues at the interface – the hot-spots. The complex between the RNase inhibitor (RI) and RNaseA protein provides an excellent system to study the role of the functional epitope as it is essential in various molecular recognition processes and constitute one of the tightest complexes known. An energetic pattern of the interface is accomplished by computational alanine scanning mutagenesis and a dynamical characterization is accomplished by a detailed study of the molecular dynamical simulations. A special emphasis is given to the role of solvation across the interface and the shielding of warm- and hot-spots from water.

1 Introduction

The challenging process of drug development is time-consuming, labor intensive, and expensive, but has as a final goal finding, developing and marketing new useful chemical entities. These new drugs can be used against currently untreatable diseases, or as replacements to available but less effective compounds. In the last few years, we observed a change in the tendency for drug design, which not only focus on the binding of a small molecule to a biomolecular target but also to a crucial and popular target class: protein-protein interactions (PPIs). Its exploitation is still taking the first steps due to the

---

*Corresponding author. *Email address:* `irina.moreira@fc.up.pt` (I. S. Moreira)

exceptional complexity of these systems that makes them resistant as pharmaceutical targets. Therefore, it is essential to explore PPIs at an atomic level in order to understand the forces that drive their interaction. Since its initial application to human growth hormone and the growth hormone binding protein, alanine scanning mutagenesis continues to be a valuable procedure for both hot-spot detection and analysis of a wide range of protein-protein interfaces [1]. Although slow and labour-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 [2]. 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. Clarkson and Wells demonstrate, via alanine scanning mutagenesis on the human growth hormone, that there is a highly uneven distribution of energetic contributions of individual residues across each protein [3]. Only a few key residues do contribute significantly to the binding free energy of protein-protein complexes: the hot-spots. Hot-spots have been defined as those sites where alanine mutation cause a significant increase in the binding free energy of at least 4.0 kcal/mol [3,4]. Warm-spots were defined as residues that upon alanine mutation generate a binding free energy difference between 2.0 and 4.0 kcal/mol and null-spots lower than 2.0 kcal/mol [5]. Other values can be used for statistical purposes. This way, it is possible to differentiate the structural epitope, the amino-acids that interact at the tridimensional proteic complexes, and the functional epitope composed of the amino-acid important for protein-protein association [5]. The structural epitopes are normally large and composed of 10-40 residues from multiple discontinuous segments on each protein. In contrast, the functional epitopes are assigned by mutagenesis studies and composed by only a small number of residues, typically two to five on each protein. As said, detailed knowledge of the hot- and warm-spots on an interface and their importance and function has important implications for the design of small molecules that disrupt protein-protein interactions or to substitute one of the protein components. The complex between the RNase inhibitor (RI) and RNase A provides an excellent system to study the role of the functional epitope.

The RNase inhibitor (RI) is a leucine-rich repeat (LRR) protein that binds diverse proteins in the pancreatic RNase superfamily [6,7]. The LRR motif is essential in various molecular recognition processes such as signal transduction, cell adhesion, cell development, DNA repair and RNA processing [8]. LRRs are present in over 2000 proteins and have been identified in viruses, bacteria, archaea and eukaryotes [9]. RNase A complexes comprises important and intriguing systems to study protein-protein association process [10]. RI complexes are some of the tightest complexes known with dissociation constants between 10^{-13} to 10^{-15} M [10]. RI adopts a "horseshoe" fold, formed by symmetrical arrangement of the 15 homologous tandem LRR units (alternately 28 and 29 residues long, that comprise nearly the entire molecule [6,10]. LLRs are $\beta-\alpha$ hairpin units in which the -strand and the -helix are approximately parallel in individual $\beta-\alpha$ units and the units are all aligned roughly parallel to a common axis [8]. Only one of the proteic members of this complex was studied by experimental alanine scanning muta-

genesis. Three warm- and one hot-spot were detected at this interface: Trp384, Asp556, Tyr558 and Tyr555. A curious and important aspect can also be studied at this interface, the effect of alanine-shaving. Alanine shaving is the concerted mutation of two or more residues at a time to evaluate their cooperative effect at the protein-protein interface [7]. By calculating the $\Delta\Delta\Delta G$ values described by Eq. (1.1) it is possible to compare the effect of replacing two or more residues both separately and together.

$$\Delta\Delta\Delta G = \Delta\Delta G_{\text{multiple mutations}} - \sum \Delta\Delta\Delta G_{\text{single mutation}} \cdot \qquad (1.1)$$

If $\Delta\Delta\Delta G = 0$ suggests that the amino-acid residues are functionally independent; if $\Delta\Delta\Delta G > 0$ it is verified a superadditivity effect; if $\Delta\Delta\Delta G < 0$ there is a subadditivity effect. Both these effects can be caused by changes in local or global protein conformation, solvent structure, electrostatic fields or dielectric constants, and protein dynamic properties [7]. Subadditivity reflects the fragility of the specific intermolecular interactions (i.e., how easily the loss of one interaction can impact on others), and superadditivity reflects the plasticity and adaptability of the interface (i.e., how readily the interactions lost can be compensated for) [7]. The structural comprehension of PPIs can benefit a lot of the use of capable computational tools that will generate fundamental knowledge based, not on a static structure, but on an ensemble generated by Molecular Dynamic (MD) simulations. We performed an exhaustive study of the RNaseA:RI interface in order to increase our knowledge about the structural basis of RI action. These results can help the design of RI derivatives or mimics that do not suffer some of the most common limitations of this kind of system, particularly its broad specificity and large size.

2   Methodology

2.1   System setup

The tridimensional structure from the complex between the ribonuclease inhibitor and ribonuclease A was taken from the Protein Data Bank with the PDB ID: 1DFJ [8]. The protonation state of the different residues were determined using the PDB2PQR server at `http://kryptonite.nbcr.net/pdb2pqr/` [11].

2.2   Molecular dynamic simulations

The MD simulations were performed using the AMBER10 package [12] with the Cornell force field [13]. Two different simulations were made, one in an implicit solvent using the Generalized Born (GB) solvent [14], and other using TIP3P explicit water molecules. The system subjected to the GB simulation was constituted by 581 residues that correspond to 8738 atoms. In the explicit solvent simulation, besides the 581 amino acid residues, there were also 15 sodium ions and 18904 water molecules to a total of 65455 atoms. The GB simulation (GB$^{\text{OBC}}$) [15] was used in the alanine scanning mutagenesis protocol because it was shown before to reproduce more accurately the experimental binding free energy

values [16, 17]. We used the GB^{OBC} modified model with α, β and γ values of 1.0, 0.8, and 4.85 as GB authors have shown that it allows a better agreement with the Poisson-Boltzmann treatment in calculating the electrostatic part of the solvation free energy [15]. The explicit simulation was used to study the role of water molecules in the function of warm- and hot-spot residues. The complex was solvated by explicit waters that extended 10 Å from any edge of the box to the protein atoms. Counter ions were added to the box to neutralize the system. In the GB simulation the ionic strength was set to 0. In each of the simulations, the system was initially minimized to remove bad contacts by steepest descent followed by conjugated gradient. The systems were then subjected to 2 ns of heating procedure in which the temperature was gradually raised to 300 k followed by 6 ns runs. The Langevin [18, 19] thermostat was used and the electrostatics interactions were calculated by using the particle mesh Ewald (PME) method [20]. Both lengths involving hydrogens were constrained using the SHAKE algorithm [21]. The equations of motion were integrated with a 2 fs time-step and the non-bonded interactions were truncated with a 16 Å and a 10 Å cutoff, in the GB and in explicit solvent simulation respectively.

2.3 Mutagenesis protocol

The MM-PBSA (Molecular Mechanics Poisson Boltzmann Surface Area) script [22] integrated into the AMBER10 package [12] was used to calculate the binding free energy difference ($\Delta\Delta G$) upon alanine mutation. It combines a continuum approach to model solvent interactions with a MM-based approach to atomistically model protein-protein interactions. This provides speed and accuracy and has been quite used in the last years [16, 17, 22–31]. The MM-PBSA approach first developed by Massova *et al* [22]. was improved by Moreira *et al*. [17] and can now be applied with an accuracy of 1 kcal/mol. The mutant complexes are generated by a single truncation of the mutated side chain, replacing $C\gamma$ with a hydrogen atom and setting the $C\beta$-H direction to that of the former $C\beta$-$C\gamma$. For the binding energy calculations, a total of 26 snapshots of the complexes were extracted in the last 1 ns of the run. The $\Delta\Delta G$ is defined as the difference between the mutant and wild type complexes defined as:

$$\Delta\Delta G = \Delta G_{RI:RNaseA\text{-mutant}} - \Delta G_{RI:RNaseA\text{-wild type}}. \tag{2.1}$$

Typical contributions to the free energy include the internal energy (bond, dihedral, and angle), the electrostatic and the van der Waals interactions, the free energy of polar solvation, the free energy of non-polar solvation, and the entropic contribution:

$$G_{molecule} = E_{internal} + E_{electrostatic} + E_{vdW} + G_{polar\ solvation} + G_{non\text{-}polar\ solvation} - TS. \tag{2.2}$$

For the calculations of relative free energies between closely related complexes it is assumed that the entropic contribution is negligible as it essentially cancel each other on Eq. (2.1) [30]. The first three terms of Eq. (2.2) were calculated with no cutoff. The

$G_{\text{polar solvation}}$ was calculated by solving the Poisson-Boltzmann equation with the software DELPHI [32, 33]. In this continuum method, the protein is modeled as a dielectric continuum of low polarizability embedded in a dielectric medium of high polarizability. We use a set of values for the DELPHI parameters that proven in a previous study to constitute a good compromise between accuracy and computing speed [34]. So, we used a value of 2.5 grids/Å for scale (the reciprocal of the grid spacing); a value of 0.001 kT/c for the convergence criterion; a 90% for the fill of the grid box; and the coulombic method to set the potentials at the boundaries of the finite-difference grid. The dielectric boundary was taken as the molecular surface defined by a 1.4 Å probe sphere and by spheres centered on each atom with radii taken from the Parse [35] vdW radii parameter set. The key aspect of the new improved approach is the use of a three dielectric constant set of values (2 for non-polar residues, 3 for polar residues and 4 for charged residues plus histidine) to mimic the expected rearrangement upon alanine mutation. It is important to highlight that we used only one trajectory for the computational energy analysis as it has proven to give the best results [17]. Side-chain reorientation was implicitly included in the formalism by raising the internal dielectric constant. The non-polar contribution to the solvation free energy due to van der Waals interactions between the solute and the solvent was modeled as a term dependent of the solvent accessible surface area (SASA) of the molecule. It was estimated by $0.00542 \times \text{SASA} + 0.92$ using the molsurf program developed by Mike Connolly [36].

2.4   Analysis

VMD [37] and PTRAJ module from AMBER10 package [12] were used in the different analysis carried on the two MD simulations. RMSDs (root mean square deviation) were calculated for each simulation to assure their equilibration. For the GB$^{\text{OBC}}$ simulation we reached a plateau of 3 Å and for the explicit simulation a plateau of 1.8 Å. All the chemical and physical characteristics described next were evaluated at the production phase. Different solvent accessible surface area (SASA) calculations were made to evaluate the importance of water molecules in the warm- and hot-spots microenvironment. The SASA values were calculated with a probe sphere of radius 1.4 Å following Lee and Richards algorithm. The SASA of each interfacial residue within the complex (SASA_{cpx}) and within the monomers (SASA_{mon}) was calculated at the production phase of the explicit MD simulation. ΔSASA and relSASA defined by Eqs. (2.3) and (2.4), respectively, were also calculated. relSASA allows the differentiation of residues with equal ΔSASA but different solvent exposure such as, i.e, a residue with a 50 Å^2 solvent accessibility in the monomer and 0 Å^2 in the complex from a residue that has a value of 150 Å^2 solvent accessibility in the monomer and 100 Å^2 in the complex. In both cases ΔSASA is 50 Å^2 but solvent accessibility importance is strikingly different between the two. relSASA was already shown to be important in previous work [38].

$$\Delta\text{SASA} = \text{SASA}_{\text{cpx}} - \text{SASA}_{\text{mon}}, \tag{2.3}$$

$$\text{relSASA} = \frac{\text{SASA}_{\text{cpx}} - \text{SASA}_{\text{mon}}}{\text{SASA}_{\text{mon}}} = \frac{\Delta\text{SASA}}{\text{SASA}_{\text{mon}}}. \tag{2.4}$$

The singular SASA behavior of the different amino-acid residues was also taken into account by considering their average SASA($<\textbf{SASA}_{\textbf{res}}>$). $<\textbf{SASA}_{\textbf{res}}>$ of all the various amino-acid residues types are (in Å^2): Gly$=85$, Ala$=113$, Cys$=140$, Asp$=151$, Glu$=183$, Phe$=218$, His$=194$, Ile$=182$, Lys$=211$, Leu$=180$, Met$=204$, Asn$=158$, Pro$=143$, Gln$=189$, Arg$=241$, Ser$=122$, Thr$=146$, Val$=160$, Trp$=259$, Tyr$=229$. These values were taken out form the relationship found out by Miller and colleagues [39, 40] that protein SASA values are approximately given by A$=6.3$M$_i^{0.73}$, where M$_i$ is the residue molecular weight. Therefore, we calculated SASA$_{\text{cpx/res}}$, ΔSASA$_{\text{res}}$ and relSASA$_{\text{res}}$ defined by Eqs. (2.5), (2.6) and (2.7) respectively.

$$\text{SASA}_{\text{cpx/res}} = \frac{\text{SASA}_{\text{cpx}}}{<\textbf{SASA}_{\textbf{res}}>}, \tag{2.5}$$

$$\Delta\text{SASA}_{\text{res}} = \frac{\Delta\text{SASA}}{<\textbf{SASA}_{\textbf{res}}>}, \tag{2.6}$$

$$\text{relSASA}_{\text{res}} = \frac{\text{relSASA}}{<\textbf{SASA}_{\textbf{res}}>} \times 1000. \tag{2.7}$$

These three equations allow the normalization of the three SASA descriptors listed before: SASA$_{\text{cpx}}$, ΔSASA and relSASA. We have also analyzed the radial distribution function, $g(r)$ and the average number of waters within a given distance, of all interfacial residues. G(r) gives the probability of finding an atom a distance r from another atom, in relation to the probability expect for a bulk solvent distribution at the same density. It was calculated by compiling a histogram with a spacing of 0.02 and a range of 8 Å. Residence water times in the first coordination shell were calculated by in house scripts. In house scripts were also used to evaluate the microenvironment surrounding each interfacial amino-acid residue. The distances between the interacting heavy atoms of each pair of residues were calculated using the VMD package using a threshold of 3.5 Å. Hydrogen bonds between proteic residues, hydrogen bonds between protein and solvent and salt-bridges were analyzed.

3   Results

The correct detection of the residues that are energetically essential in protein-protein interfaces is a key issue with huge practical application such as rational drug design and protein engineering. Alanine scanning mutagenesis (ASM) has been applied successfully to the characterization of these interfaces. However, experimental ASM is a costly and time consuming task, which urged the need for fast and accurate theoretical methods. A huge amount of algorithms of increasing complexity have been employed to address the binding energy between biological molecules. These can be divided essentially in

three types: empirical functions or simple physical methods that use knowledge-based simplified models to evaluate complex association; fully atomistic methods that estimate the binding free energy as a result of mutating the residues of the interacting molecules; or, more recently, feature-based approaches. The feature-based approaches tend to be more qualitative than quantitative [38, 41–47]. Although, no without of merit, an atomistic and accurate quantitative ASM method is crucial to detect warm- and hot-spots as it adds also an energetic characterization of the terms responsible for it warm- or hot-spots character. Only a few amino-acid residues of the RNaseA:RI interface possess an experimental $\Delta\Delta G_{binding}$ value. This way it was important to apply our ASM computational methodological approach and get a broad picture of all interface. Our method opened the possibility of decomposing the binding free energy ($\Delta\Delta G_{binding}$) into the various energetic factors, such as the electrostatic energy, $\Delta\Delta E_{electrostatic}$; the van der Waals energy, $\Delta\Delta E_{vdW}$; the free energy of polar solvation, $\Delta\Delta G_{polar\ solvation}$; and the free energy of non-polar solvation, $\Delta\Delta G_{non\text{-}polar\ solvation}$. The results are listed in Table 1.

Alanines, glycines or prolines could not be tested by computational ASM. For example, glycines could not be tested as the ASM method only works when an amino-acid is mutated by a smaller one, in this case alanine. As proline mutations are disruptive and are associated with conformational changes, they could not be considered in our study. To ensure that our predictions are valid it is first necessarily to evaluate the computational ASM behavior and accuracy. Previous studies demonstrated that the method possess an overall success rate (theoretical result within 1.0 kcal/mol of the experimental result) higher than 80% with an unsigned error of 0.80 kcal/mol [17]. Fig. 1 demonstrates that the correlation between the $\Delta\Delta G_{experimental}$ and the $\Delta\Delta G_{theoretical}$ for this particular system is of 89% with an average error of 0.65 kcal/mol, and therefore we have the necessary confidence to retrieve important structural and energetic information from our results.
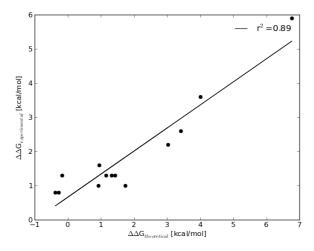

Figure 1: $\Delta\Delta G_{experimental}$ versus $\Delta\Delta G_{theoretical}$.

Table 1: Results of the Computational Alanine Scanning Mutagenesis Method. All values are in kcal/mol. Highlighted in light shading are the hot- and warm-spots detected at this interface. Each monomer results were sorted by $\Delta\Delta G_{binding}$ for an easier understanding.

Protein	#	AA	$\Delta\Delta E_{electrostatic}$	$\Delta\Delta E_{vdW}$	$\Delta\Delta G_{non\text{-}polar\ solvation}$	$\Delta\Delta G_{polar\ solvation}$	$\Delta\Delta G_{binding}$	Score	$\Delta\Delta G_{exp}$	Ref
RNaseA	43	VAL	8.33	1.21	0.03	-0.74	8.84	Hot		
	41	LYS	100.34	2.16	0.16	-94.19	8.47	Hot		
	98	LYS	69.22	0.07	0.00	-66.06	3.23	Warm		
	119	HIS	1.41	2.02	0.26	-2.49	1.20	Null		
	89	SER	3.57	-1.07	0.02	-1.84	0.68	Null		
	120	PHE	0.21	0.06	-0.02	0.41	0.67	Null		
	90	SER	0.02	0.12	0	0.12	0.25	Null		
	7	LYS	0.80	0.10	0.03	-0.68	0.24	Null		
	11	GLN	-1.24	0.26	0.05	1.16	0.22	Null		
	87	THR	0.18	0.1	-0.02	-0.36	-0.09	Null		
	8	PHE	-0.13	0.05	-0.03	-0.39	-0.50	Null		
	12	HIS	0.82	0.20	0.00	-1.74	-0.71	Null		
	118	VAL	-0.68	0.46	0.03	-0.56	-0.75	Null		
	35	LEU	-0.19	0.64	0.02	-1.27	-0.81	Null		
	38	ASP	-78.01	2.34	0.07	72.17	-3.44	Null		
	86	GLU	-70.81	0.67	0.2	66.32	-3.62	Null		
RI	555	TYR	2.39	10.58	0.44	-6.64	6.77	Hot	5.9	[10]
	556	ASP	37.56	1.4	0.11	-35.06	4.01	Hot	3.6	[10]
	558	TYR	0.78	4.88	0.5	-2.76	3.41	Warm	2.6	[10]
	384	TRP	-0.95	7.73	0.29	-4.05	3.02	Warm	2.2	[6]
	351	GLU	20.45	-0.06	-0.01	-18.38	1.99	Null		
	353	ASP	27.4	-0.58	-0.03	-24.98	1.81	Null		
	439	TRP	0.55	3.53	0.27	-2.61	1.73	Null	1	[6]
	524	ASP	18.85	0.51	0.02	-17.78	1.6	Null		
	385	GLU	19.52	0.55	0.16	-18.76	1.47	Null		
	522	GLU	24.92	-1.05	0.02	-22.46	1.43	Null	1.3	[6]
	408	GLU	15.53	0.09	0	-14.3	1.32	Null	1.3	[6]
	553	VAL	1.08	1.7	-0.04	-1.55	1.19	Null		
	441	LYS	-2.85	2.11	0.46	1.43	1.15	Null	1.3	[6]
	465	GLU	14.96	0.06	0	-14.07	0.95	Null	1.6	[6]
	327	GLU	20.37	0.41	0.02	-19.88	0.92	Null	1	[6]
	551	GLN	0.25	1.57	0.01	-1.04	0.79	Null		
	494	VAL	0.81	0.19	0.02	-0.34	0.67	Null		
	499	ASP	11.47	0.26	-0.01	-11.36	0.36	Null		
	470	SER	-0.3	0.09	-0.01	0.38	0.15	Null		
	356	SER	-0.37	0.13	0	0.38	0.14	Null		
	442	SER	-0.17	0.12	0	0.17	0.12	Null		
	526	SER	0.47	0.32	0	-0.75	0.03	Null		
	554	LEU	0.10	0.08	0.00	-0.20	-0.03	Null		
	382	TRP	2.62	3.89	0.21	-6.9	-0.18	Null	1.3	[6]
	467	GLN	-0.86	0.65	0.01	-0.03	-0.23	Null		
	527	ASN	0.08	0	0	-0.33	-0.24	Null		
	578	ARG	-11.96	1.47	0.16	10.05	-0.28	Null	0.8	[6]
	410	SER	-0.77	0.25	-0.02	0.16	-0.38	Null	0.8	[6]
	325	ARG	-19.96	0.64	0.01	18.36	-0.95	Null		

It is important to highlight that one of the major reasons behind the success of the method is the use of three different internal dielectric constants that mimic the rearrangement expected upon alanine mutation. Conformational sampling, the relaxation and reorganization due to the mutation for an alanine, and maybe even specific water binding as well as the electronic polarization that affects the charge-charge are not ex-
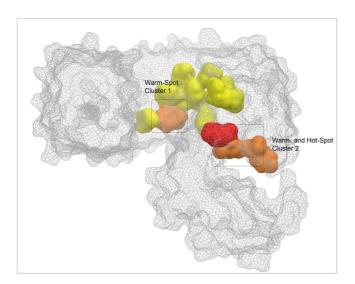
Figure 2: Warm- and hot-spots clusters at the RNaseA:RI interface. Hot-, warm, and null-spots are in red, orange and yellow color respectively.

plicitly included in the formalism. Therefore, the scaling of the macroscopic parameter (internal dielectric constant) to larger values when larger re-organizations are expected mimics these effects. By using only a three-internal dielectric constant set exclusively characteristic of the mutated amino acid (2 for the non-polar amino acids, 3 for the polar residues, and 4 for the charged amino acids), it was possible to obtain an excellent agreement with the experimental results for the $\Delta\Delta G_{\text{binding}}$ values. Charged residues were one of the major problems of the fully atomistic ASM approaches [22]. The use of a set of 3 internal dielectric constants is essential for the calculation of the electrostatic components as $\Delta\Delta G_{\text{polar solvation}}$ obtained using the DELPHI software, which is bigger for charged residues [33]. This fact explains our high success rate concerning this kind of residues. Besides the correct detection of residues Tyr555, Asp556, Tyr558 and Trp384 as energetically key residues at the RNaseA:RIinterface, new residues were also detected such as Lys41, Val 43 and Lys98. The first two have a $\Delta\Delta G_{\text{binding}}$ higher than 4.0 kcal/mol and therefore are hot-spots and the last one has a warm-spot character. This way, the RNaseA:RIinterface presents two different clusters of warm- and hot-spots residues: cluster one is formed by RNaseA Lys98 and RI Trp384, cluster two is formed by RI Tyr555, Asp556 and Tyr558 and RNaseA Lys41 and Val43. Their spatial distribution can be seen in Fig. 2.

From experiments some key questions were posed. What is the relative importance of H-bonds versus van der Waals contacts and the hydrophobic effect in the interactions of the various hot- and warm-spots at the RNaseA:RI interface? Do these residues form contacts with either protein that are not apparent from the crystal structures [7]? At the time, the authors performed mutations for other residues besides alanine trying to achieve the answers to some of these questions. For example, the role of Tyr555 was analyzed not

Table 2: Interactions at the RNaseA:RI interface. The values presented are the average and the standard deviation error for the distances between the heavy atoms of each pair of residues.

Type of interaction	Residue pair	Distance [Å]
Salt-bridge	Asp556-Lys 41	2.82±0.12
Hydrogen bond	Asp556-Lys 41	3.84±0.17
Hydrogen bond	Lys41-Water	4.99±0.35
Hydrogen bond	Val43-Water	2.91±0.13
Hydrogen bond	Val43-Water	4.16±0.36
Hydrogen bond	Tyr555-Water	2.75±0.06
Hydrogen bond	Tyr558-Water	4.69±0.85

only by alanine mutation but also upon phenylalanine mutation. The authors conclude that the attenuated affinity of Tyr434Ala was due only to the loss of the phenyl group. In contrast, replacement of Tyr558 by Phe weakened the RI:RNase A interaction almost to the same extent as substitution with Ala, and therefore the OH group rather than the phenyl ring of this Tyr appears to provide the energetically important interactions [7]. By inspection of Table 1 it is easily perceived that for Tyr555Ala, $\Delta\Delta E_{vdW}$ has a value of 10.58 kcal/mol being the major energetic component, which contrasts with the value of 4.88 kcal/mol for Tyr558Ala. These facts gave us, once more, confidence in our results. For the new detected warm- and hot-spots at the RI protein, the major energetic component seems to be the electrostatic contribution ($\Delta\Delta E_{electrostatics} + \Delta\Delta G_{polar\ solvation}$) that ranges from 3.16 to 7.59 kcal/mol. Electrostatics is one of the most noteworthy interactions at a protein-protein interface, and it seems to play a major role at the RNaseA:RI complex. The analyses of the PPIs requires dynamical information as the data extracted only from crystal structures is limited to that particular, and most likely, not representative conformation. Proteins in solution are mobile molecules that do not exist in a single conformation, but instead in a set of different conformational states. MD simulations are undoubtedly the cornerstone when it comes to the study of the dynamics of a system and have proved to be a reliable tool. By applying an explicit MD approach it was possible to generate an ensemble of complex conformations, in which different physical and chemical characteristics could be studied. From the MD simulation it was possible to retrieve the average distances of the most stable hydrogen bonds and salt-bridges across the warm- and hot-spots at the RNaseA:RI interface, which are listed in Table 2.

The correlation coefficients are 0.06, 0.58, 0.21, 0.06, 0.63, 0.10 for $SASA_{cpx}$, $\Delta SASA$, $relSASA$, $SASA_{cpx/res}$, $\Delta SASA_{res}$, and $relSASA_{res}$, respectively. SASA descriptors are some of the most commonly used features at the machine learning algorithms developed recently. [38, 41–47] At this interface $\Delta SASA$ and $\Delta SASA_{res}$ descriptors already show a significant correlation with $\Delta\Delta G_{binding}$. However, it seems that the best SASA descriptor that shows a higher correlation is $\Delta SASA_{res}$. The main explanation can be the fact that $\Delta SASA_{res}$ includes the importance of SASA upon complex formation as well as the normalization by the different types of amino-acid residues. If we only considered the four warm- and hot-spots the correlation for the six SASA descriptors increase. Although we

Table 3: Values for the six SASA descriptors ($SASA_{cpx}$, $\Delta SASA$, $relSASA$, $SASA_{cpx/res}$, $\Delta SASA_{res}$ and $relSASA_{res}$) for all interfacial residues. The units are Å^2 and the average and standard deviation error for each SASA descriptor is presented.

Protein	#	AA	$SASA_{cpx}$	$\Delta SASA$	$relSASA$	$SASA_{cpx/res}$	$\Delta SASA_{res}$	$relSASA_{res}$	$\Delta\Delta G_{exp}$
RNaseA	7	LYS	23.74 ± 10.43	-73.74 ± 12.83	-0.76 ± 0.1	0.11 ± 0.05	-0.35 ± 0.06	-3.58 ± 0.49	
	8	PHE	2.80 ± 1.47	-2.02 ± 1.22	-0.42 ± 0.21	0.01 ± 0.01	-0.01 ± 0.01	-1.94 ± 0.97	
	11	GLN	5.20 ± 2.94	-14.41 ± 5.17	-0.73 ± 0.14	0.03 ± 0.02	-0.08 ± 0.03	-3.85 ± 0.74	
	12	HIS	1.37 ± 0.99	-12.98 ± 2.11	-0.91 ± 0.05	0.01 ± 0.01	-0.07 ± 0.01	-4.67 ± 0.26	
	35	LEU	8.08 ± 7.61	-7.56 ± 4.86	-0.54 ± 0.35	0.04 ± 0.04	-0.04 ± 0.03	-2.98 ± 1.94	
	38	ASP	84.01 ± 16.62	-30.02 ± 19.72	-0.26 ± 0.16	0.56 ± 0.11	-0.2 ± 0.13	-1.7 ± 1.04	
	41	LYS	2.00 ± 2.71	-56.03 ± 4.31	-0.97 ± 0.04	0.01 ± 0.01	-0.27 ± 0.02	-4.58 ± 0.21	
	43	VAL	18.11 ± 6.00	-45.83 ± 5.69	-0.72 ± 0.08	0.11 ± 0.04	-0.29 ± 0.03	-4.49 ± 0.48	
	86	GLU	30.65 ± 5.59	-15.42 ± 4.62	-0.34 ± 0.1	0.17 ± 0.03	-0.08 ± 0.03	-1.84 ± 0.56	
	87	THR	16.19 ± 3.19	-34.85 ± 5.88	-0.68 ± 0.06	0.11 ± 0.02	-0.24 ± 0.04	-4.66 ± 0.42	
	89	SER	22.35 ± 4.95	-80.84 ± 5.07	-0.78 ± 0.04	0.18 ± 0.04	-0.66 ± 0.04	-6.43 ± 0.33	
	90	SER	5.89 ± 1.82	-14.10 ± 1.71	-0.71±0.08	0.05 ± 0.01	-0.12±0.01	-5.80±0.62	
	98	LYS	100.08 ± 12.92	-46.11 ± 15.96	-0.31±0.1	0.47 ± 0.06	-0.22±0.07	-1.48±0.47	
	118	VAL	16.00 ± 5.11	-12.08 ± 6.05	-0.42 ± 0.15	0.10 ± 0.03	-0.08 ± 0.04	-2.63 ± 0.96	
	119	HIS	16.86 ± 7.34	-89.78 ± 7.65	-0.84 ± 0.07	0.09 ± 0.04	-0.46 ± 0.04	-4.32 ± 0.34	
	120	PHE	15.91 ± 4.24	-10.84 ± 3.70	-0.41 ± 0.12	0.07 ± 0.02	-0.05 ± 0.02	-1.87 ± 0.56	
RI	325	ARG	53.25 ± 5.21	-0.02 ± 0.12	0 ± 0	0.22 ± 0.02	-9.43E-005 ± 0	0 ± 0.01	
	327	GLU	20.37 ± 4.99	-19.41 ± 2.91	-0.49 ± 0.08	0.11 ± 0.03	-0.11 ± 0.02	-2.69 ± 0.44	1.00
	351	GLU	37.94 ± 7.84	-0.23 ± 0.75	-0.01 ± 0.02	0.21 ± 0.04	0 ± 0	-0.04 ± 0.13	
	353	ASP	13.61 ± 2.78	-6.38 ± 1.76	-0.32 ± 0.07	0.09 ± 0.02	-0.04 ± 0.01	-2.12 ± 0.44	
	356	SER	32.93 ± 7.21	-2.27 ± 3.37	-0.07 ± 0.09	0.27 ± 0.06	-0.02 ± 0.03	-0.56 ± 0.78	
	382	TRP	29.97 ± 9.24	-40.00 ± 12.09	-0.57 ± 0.15	0.12 ± 0.04	-0.15 ± 0.05	-2.18 ± 0.57	1.30
	384	TRP	15.70 ± 4.14	-72.55 ± 4.27	-0.82 ± 0.04	0.06 ± 0.02	-0.28 ± 0.02	-3.18 ± 0.15	2.20
	385	GLU	70.66 ± 11.43	-22.27 ± 7.81	-0.24 ± 0.08	0.39 ± 0.06	-0.12 ± 0.04	-1.32 ± 0.46	
	408	GLU	55.70 ± 10.71	-9.62 ± 6.89	-0.15 ± 0.1	0.30 ± 0.06	-0.05 ± 0.04	-0.8 ± 0.55	1.30
	410	SER	1.40 ± 1.23	-4.31 ± 0.86	-0.78 ± 0.16	0.01 ± 0.01	-0.04 ± 0.01	-6.39 ± 1.28	0.80
	439	TRP	15.66 ± 5.86	-64.94 ± 6.38	-0.81 ± 0.07	0.06 ± 0.02	-0.25 ± 0.02	-3.11 ± 0.27	1.00
	441	LYS	36.09 ± 6.43	-47.98 ± 19.97	-0.55 ± 0.13	0.17 ± 0.03	-0.23 ± 0.09	-2.6 ± 0.63	1.30
	442	SER	57.54 ± 8.55	-2.15 ± 4.42	-0.04 ± 0.08	0.47 ± 0.07	-0.02 ± 0.04	-0.32 ± 0.68	
	465	GLU	31.80 ± 6.32	-6.45 ± 7.08	-0.16 ± 0.17	0.17 ± 0.04	-0.04 ± 0.04	-0.88 ± 0.9	1.60
	467	GLN	24.14 ± 6.57	-2.05 ± 2.76	-0.08 ± 0.1	0.13 ± 0.03	-0.01 ± 0.01	-0.44 ± 0.54	
	470	SER	37.44 ± 6.66	-17.15 ± 3.28	-0.32 ± 0.06	0.31 ± 0.05	-0.14 ± 0.03	-2.6 ± 0.53	
	494	VAL	27.72 ± 7.81	-0.01 ± 0.04	0 ± 0	0.17 ± 0.05	-5.08E-005 ± 0	0 ± 0.01	
	499	ASP	31.71 ± 8.47	-32.14 ± 6.93	-0.51 ± 0.11	0.21 ± 0.06	-0.21 ± 0.04	-3.35 ± 0.73	
	522	GLU	36.19 ± 6.04	-20.52 ± 3.76	-0.36 ± 0.04	0.20 ± 0.03	-0.11 ± 0.02	-1.98 ± 0.19	1.30
	524	ASP	11.01 ± 3.66	-16.21 ± 4.34	-0.59 ± 0.12	0.07 ± 0.02	-0.11 ± 0.03	-3.93 ± 0.81	
	526	SER	2.54 ± 1.89	-18.25 ± 2.85	-0.88 ± 0.09	0.02 ± 0.02	-0.15 ± 0.02	-7.19 ± 0.78	
	527	ASN	21.39 ± 7.10	-52.78 ± 7.97	-0.71 ± 0.09	0.14 ± 0.04	-0.33 ± 0.05	-4.5 ± 0.6	
	551	GLN	35.74 ± 12.12	-14.77 ± 3.46	-0.31 ± 0.12	0.19 ± 0.06	-0.08 ± 0.02	-1.66 ± 0.66	
	553	VAL	0.87 ± 0.97	-21.60 ± 5.95	-0.96 ± 0.04	0.01 ± 0.01	-0.13 ± 0.04	-6.01 ± 0.25	
	554	LEU	8.35 ± 3.09	0.00 ± 0.00	0 ± 0	0.05 ± 0.02	0 ± 0	0 ± 0	
	555	TYR	25.09 ± 8.63	-111.62 ± 7.61	-0.82 ± 0.06	0.11 ± 0.04	-0.49 ± 0.03	-3.57 ± 0.26	5.90
	556	ASP	13.31 ± 3.98	-79.65 ± 7.46	-0.86 ± 0.04	0.09 ± 0.03	-0.53 ± 0.05	-5.67 ± 0.28	3.60
	558	TYR	44.55 ± 12.96	-78.14 ± 34.07	-0.61 ± 0.15	0.19 ± 0.06	-0.34 ± 0.15	-2.68 ± 0.65	2.60
	578	ARG	130.78 ± 20.43	-18.31 ± 14.97	-0.12 ± 0.1	0.54 ± 0.08	-0.08 ± 0.06	-0.51 ± 0.41	0.80

only have 4 residues at this condition, which is not enough for a statistical analysis, it seems that the warm- and hot-spot character is highly associated with the solvent accessibility of the interfacial residues. Due to the uncertainty (average error of 0.65 kcal/mol) of the computational ASM, these correlations are a bit worse when we considered all amino-acids tested. However, it is clear from Table 3 that solvent accessibility is ex-
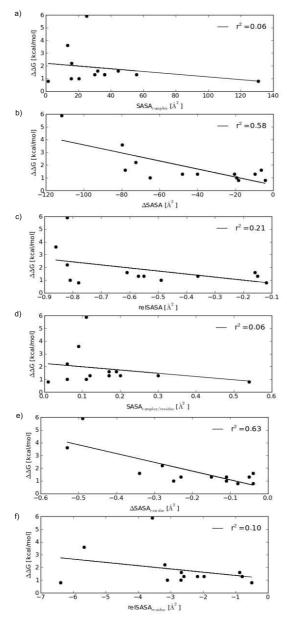
Figure 3: Plots of the six SASA descriptors versus $\Delta\Delta G_{experimental}$. a) $SASA_{cpx}$, b) $\Delta SASA$, c) relSASA, d) $SASA_{cpx/res}$, e) $\Delta SASA_{res}$, and f) relSASA$_{res}$.

tremely correlated with the warm- and hot-spot character of key residues at the PPI. So, it is important to further investigate the fine details of hydration. To that purpose we calculated the radial distribution functions of the interfacial residues in the MD simulation as well as the average number of waters inside a given radius. Table 4 shows the average

Table 4: Average number of waters at a certain radius of the hot-, warm- and null-spots.

r [Å]	Hot-Spots	Warm-Spots	Null-Spots
0	0	0	0
1	0	0	0
2	0	0	0
3	0	0.68	1.49
4	0.13	1.83	3.36
5	0.97	4.28	6.86
6	4.84	9.48	11.70

number of waters at a given distance from the null-, the warm- and the hot-spots. It is clearly shown that warm- and hot-spots tend to be protected from solvent presenting a much lower average number of waters around them. For example, at a maximum of 5 Å, after which pair interactions becomes less meaningful, the hot-spots have on average one water molecule; the warm-spots have four; and the null-spots around seven. Once more, it is stressed that the energetically important residues are shielded from water and that trapped water molecules have crucial roles by establishing important interactions between the warm- and hot-spots.

The overall picture of the RNaseA:RI interface can be seen in Fig. 4. This figure shows the most stable interactions that involve the warm- and hot-spots detected computationally. Panel a) shows the interaction that have a minimum residence time of 80% and panel b) of 90%. This figure allows visualization of the dynamical behavior of the RNaseA:RI interface. It is easily perceived that the two clusters of warm- and hot-spots are very stable as their interactions are not lost even at a 90% threshold. It is also indisputable that although the energetically important residues tend to be occluded from solvent, water still has a major function as a bridge between the amino-acid residues at the cluster 2 of warm- and hot-spots. This fact was not seen in the crystallographic structures, and therefore previous work done on this system was not able to explain how Tyr555 and Tyr558, although at 13 Å apart, could still be energetically coupled. From inspection of Fig. 4 and, for a more detailed comprehension, Fig. 5, it is possible to observe the importance of water connection to the coupling of these two residues. The water molecules present at this network vary in their residence time. The two closest ones to Tyr555 are constant at a residence time of 80% (as seen in Fig. 4) and the others have a more dynamical character.

The flexibility of the residues at the protein surface together with the participation of water molecules help to define the shape and physicochemical complementarities of the binding partners, which determines the specificity and stability of the association. In this interface, as noted for others PPIs [5, 48], the hot-spots are not homogeneously distributed across the interface: rather they are distributed in clusters in which water molecules plays a major role. To have a complete knowledge of this interface we performed alanine shaving in order to access the cooperatively within the most important cluster of warm- and hot-spots. For the simultaneous mutation of the energetically interacting important residues of both interfaces (Lys41/His119/Tyr555/Tyr558;
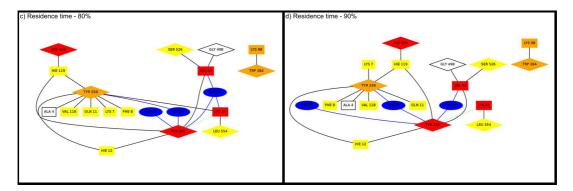
Figure 4: Map of the key interactions between the hot- and warm-spot cluster 1 (Lys98 and Trp384) and cluster 2 (Tyr555, Asp556, Tyr558, Lys41, Val43). RNase A and RI residues are in a rectangle and diamond representations, respectively. Hot-, warm-, and null-spots are in red, orange and yellow color respectively. Protein-protein interactions are in black and water interactions in blue. Green dots indicate that a salt-bridge and a hydrogen bond are made simultaneously between the pair of residues. a) Interactions with a residence time of 80%, b) Interactions with a residence time of 90%. Figure produced by an in-house script which takes advantage of a python library designed to build graphs, python-graph (`http://code.google.com/p/python-graph/`). It allows the 2D understanding of the micro-environment surrounding each interfacial residue.
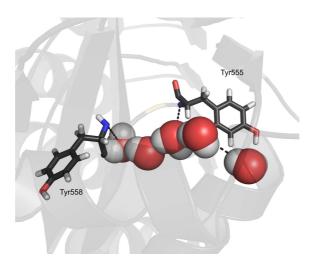

Figure 5: Water network between the two energetically important residues Tyr555 and Tyr558 with a hot and warm-spot character, respectively.

Val43/His119/Tyr555/Tyr558; Lys41/Val43/His119/Tyr555/Tyr558) we verified a sub-additivity in which the simulations mutation of the amino-acid residues generate a $\Delta\Delta G_{binding}$ lower than the sum of their single contributions. The losses in binding energy for these multiple mutations are strongly subadditivity implying that the residues are functionally coupled. The simultaneous replacement of multiple residues can weaken multiple interactions and have especially negative effects on solvent structure. Therefore, the three multiple mutations subadditivity effect can be explained by the major role water plays bridging the two binding partners.

Table 5: Results of the Computational Alanine Shaving Mutagenesis Method. All values are in kcal/mol.

#	AA	$\Delta\Delta E_{electrostatic}$	$\Delta\Delta E_{VDW}$	$\Delta\Delta G_{non-polar\ solvation}$	$\Delta\Delta G_{polar\ solvation}$
41/119/555/558	LYS/HIS/TYR/TYR	139.69	18.33	1.38	-140.65
43/119/555/558	VAL/HIS/TYR/TYR	10.58	18.27	1.22	-12.66
41/43/119/555/558	LYS/VAL/HIS/TYR/TYR	139.69	18.33	1.38	-140.65

#	AA	$\Delta\Delta G_{binding}$	$\sum\Delta\Delta G$	$\Delta\Delta\Delta G_{exp}$
41/119/555/558	LYS/HIS/TYR/TYR	17.74	19.85	<0
43/119/555/558	VAL/HIS/TYR/TYR	17.43	20.22	<0
41/43/119/555/558	LYS/VAL/HIS/TYR/TYR	17.74	28.6 9	<0

4   Conclusion

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. The RNase inhibitor is a LRR protein that binds diverse proteins in the pancreatic RNase superfamily and is essential in various molecular recognition processes forming some of the tightest complexes known with dissociation constants between 10^{-13} to 10^{-15} M. These characteristics make RNaseA:RI interface an imperative subject in the field of protein-protein interactions. In this work we have applied a computational alanine scanning mutagenesis protocol that presents a high accuracy (correlation between $\Delta\Delta G_{theoretical}$ and $\Delta\Delta G_{experimental}$ of 89% and an average error of 0.65 kcal/mol) to this particular system. This protocol distinguish itself from others by the use of a set of three internal dielectric constants (2 for non-polar residues, 3 to polar and 4 for charged ones) in order to mimic the expected rearrangement upon alanine mutation that is not explicitly included in the formalism. Besides correctly detecting the warm- and hot-spots in the RI protein, we have also detected 3 more crucial residues for binding (Lys41, Val43, Lys98). The first two have a hot-spot character and the third a warm-spot character. These three residues seem to play a key role in protein-protein binding. We have also analyzed the role of solvation across the interface. Water plays a major role at every biological process and can make important interactions at an interface. A few years ago an O-ring theory was proposed. It stated that warm- and hot-spots should be occluded from water in order to allow an efficient coupling between proteins. We saw by a careful analyze of the radial distribution functions of all interfacial residues, as well as the water micro-environment around each interfacial residue, that warm- and hot-spots seem, indeed, to be shielded from water. Nevertheless, some water molecules can still be trapped at the interface producing crucial interactions. This feature was particularly important to explain the cooperativity between Tyr555 and Tyr558, two energetically important residues that are 13 Å apart. These two residues possess a water network that allows their interaction. We have also calculated six different SASA descriptors to access the importance of the loss of solvent accessibility upon complex formation. Three SASA descriptors ΔSASA, ΔSASA$_{res}$ and relSASA appear to present a significant correla-

tion with $\Delta\Delta G_{binding}$ and can be used in future ASM features-based work. These deeper knowledge about the RNaseA:RI interface can potentially be used to design better and more specific inhibitors.

References

[1] Lee, N. H., N. S. M. Geoghagen, E. Cheng, R. T. Cline, and C. M. Fraser. 1996. Alanine scanning mutagenesis of conserved arginine/lysine-arginine/lysine-X-X-arginine/lysine G protein-activating motifs on m1 muscarinic acetylcholine receptors. Molecular Pharmacology 50:140-148.

[2] Moreira, I., P. Fernandes, and M. Ramos. 2007. Computational Determination of the Relative Free Energy of Binding: Application to Alanine Scanning Mutagenesis. In Molecular Materials with Specific Interactions Modeling and Design. 305-339.

[3] Clackson, T., and J. A. Wells. 1995. A hot-spot of binding energy in a hormone receptor interface. Science 267:383-386.

[4] Jin, L., F. E. Cohen, and J. A. Wells. 1994. Structure from function – screening structural models with functional data. Proceedings of the National Academy of Sciences of the United States of America 91:113-117.

[5] Moreira, I. S., P. A. Fernandes, and M. J. Ramos. 2007. Hot spots-A review of the protein-protein interface determinant amino-acid residues. Proteins-Structure Function and Bioinformatics 68:803-812.

[6] Shapiro, R., M. Ruiz-Gutierrez, and C. Z. Chen. 2000. Analysis of the interactions of human ribonuclease inhibitor with angiogenin and ribonuclease A by mutagenesis: Importance of inhibitor residues inside versus outside the C-terminal "hot spot". Journal of Molecular Biology 302:497-519.

[7] Chen, C. Z., and R. Shapiro. 1999. Superadditive and subadditive effects of "hot spot" mutations within the interfaces of placental ribonuclease inhibitor with angiogenin and ribonuclease A. Biochemistry 38:9273-9285.

[8] Kobe, B., and J. Deisenhofer. 1995. A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature 374:183-186.

[9] Kobe, B., and J. Deisenhofer. 1996. Mechanism of ribonuclease inhibition by ribonuclease inhibitor protein based on the crystal structure of its complex with ribonuclease A. Journal of Molecular Biology 264:1028-1043.

[10] Chen, C. Z., and R. Shapiro. 1997. Site-specific mutagenesis reveals differences in the structural bases for tight binding of RNase inhibitor to angiogenin and RNase A. Proceedings of the National Academy of Sciences of the United States of America 94:1761-1766.

[11] Dolinsky, T. J., J. E. Nielsen, J. A. McCammon, and N. A. Baker. 2004. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Research 32:W665-W667.

[12] D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R.C.Walker, W. Zhang, K.M. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M.G. Seetin, C. Sagui, V. Babin, and P.A. Kollman. 2008. AMBER 10. University of California, San Francisco.

[13] Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A 2nd generation force-field

for the simulation of proteins, nucleic-acids and organic molecules. Journal of the American Chemical Society 117:5179-5197.

[14] Tsui, V., and D. A. Case. 2001. Theory and applications of the Generalized Born solvation model in macromolecular simulations. Biopolymers 56:275-291.

[15] Onufriev, A., D. Bashford, and D. A. Case. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins: Structure, Function, and Bioinformatics 55:383-394.

[16] Moreira, I. S., P. A. Fernandes, and M. J. Ramos. 2007. Unravelling Hot Spots: A comprehensive computational mutagenesis study. Theoretical Chemistry Accounts 117:99-113.

[17] Moreira, I. S., P. A. Fernandes, and M. J. Ramos. 2007. Computational alanine scanning mutagenesis - An improved methodological approach. Journal of Computational Chemistry 28:644-654.

[18] Izaguirre, J. A. 2001. Langevin stabilization of molecular dynamics. Journal of Chemical Physics 114:2090-2098.

[19] Loncharich, R. J., B. R. Brooks, and R. W. Pastor. 1992. Langevin dynamics of peptides- the frictional dependence of isomerization rates of n-acetylalanyl-n-methylamide Biopolymers 32:523-535.

[20] Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald- an n.log(n) method for ewald sums in large systems. Journal of Chemical Physics 98:10089-10092.

[21] Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of cartesian equations of motion of a system with constraints- molecular dynamics of n-alkanes. Journal of Computational Physics 23:327-341.

[22] Huo, S., I. Massova, and P. A. Kollman. 2002. Computational alanine scanning of the 1 : 1 human growth hormone-receptor complex. Journal of Computational Chemistry 23:15-27.

[23] Moreira, I. S., P. A. Fernandes, and M. J. Ramos. 2006. Detailed microscopic study of the full ZipA: FtsZ interface. Proteins-Structure Function and Bioinformatics 63:811-821.

[24] Moreira, I. S., P. A. Fernandes, and M. J. Ramos. 2006. 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. Journal of Physical Chemistry B 110:10962-10969.

[25] Moreira, I. S., P. A. Fernandes, and M. J. Ramos. 2007. Hot spot computational identification: Application to the complex formed between the hen egg white lysozyme (HEL) and the antibody HyHEL-10. International Journal of Quantum Chemistry 107:299-310.

[26] Moreira, I. S., P. A. Fernandes, and M. J. Ramos. 2007. Backbone importance for protein-protein binding. Journal of Chemical Theory and Computation 3:885-893.

[27] Moreira, I. S., P. A. Fernandes, and M. J. Ramos. 2007. Hot spot occlusion from bulk water: A comprehensive study of the complex between the lysozyme HEL and the antibody FVD1.3. Journal of Physical Chemistry B 111:2697-2706.

[28] Moreira, I. S., P. A. Fernandes, and M. J. Ramos. 2008. Protein-protein recognition: A computational mutagenesis study of the MDM2-P53 complex. Theoretical Chemistry Accounts 120:533-542.

[29] Chong, L. T., Y. Duan, L. Wang, I. Massova, and P. A. Kollman. 1999. Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. Proceedings of the National Academy of Sciences of the United States of America 96:14330-14335.

[30] Kollman, P. A., I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham. 2000. Calculating structures and free energies of complex molecules: Combining molecular mechanics

and continuum models. Accounts of Chemical Research 33:889-897.

[31] Bradshaw, R. T., B. H. Patel, E. W. Tate, R. J. Leatherbarrow, and I. R. Gould. 2011. Comparing experimental and computational alanine scanning techniques for probing a prototypical protein-protein interaction. Protein Engineering Design & Selection 24:197-207.

[32] Rocchia, W., E. Alexov, and B. Honig. 2001. Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions. Journal of Physical Chemistry B 105:6507-6514.

[33] Rocchia, W., S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, and B. Honig. 2002. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. Journal of Computational Chemistry 23:128-137.

[34] Moreira, I. S., P. A. Fernandes, and M. J. Ramos. 2005. Accuracy of the numerical solution of the Poisson-Boltzmann equation. Journal of Molecular Structure-Theochem 729:11-18.

[35] Sitkoff, D., K. A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free-energies using macroscopic solvent models Journal of Physical Chemistry 98:1978-1988.

[36] Connolly, M. L. 1983. Analytical molecular surface calculation J. Appl. Crystallogr. 16:548-558.

[37] Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual molecular dynamics. Journal of Molecular Graphics 14:33-8, 27-8.

[38] Cho, K.-i., D. Kim, and D. Lee. 2009. A feature-based approach to modeling proteinprotein interaction hot spots. Nucleic Acids Research 37:2672-2687.

[39] Miller, S., J. Janin, A. M. Lesk, and C. Chothia. 1987. Interior and surface of monomeric proteins. Journal of Molecular Biology 196:641-656.

[40] Miller, S., A. M. Lesk, J. Janin, and C. Chothia. 1987. The accessible surface-area and stability of oligomeric proteins. Nature 328:834-836.

[41] Tuncbag, N., A. Gursoy, and O. Keskin. 2009. Identification of computational hot spots in protein interfaces: Combining solvent accessibility and inter-residue potentials improves the accuracy. Bioinformatics 25:1513-1520.

[42] Xia, J. F., X. M. Zhao, J. N. Song, and D. S. Huang. 2010. APIS: Accurate prediction of hot spots in protein interfaces by combining protrusion index with solvent accessibility. BMC Bioinformatics 11:174.

[43] Darnell, S. J., L. LeGault, and J. C. Mitchell. 2008. KFC Server: Interactive forecasting of protein interaction hot spots. Nucleic Acids Research 36:W265-W269.

[44] Darnell, S. J., D. Page, and J. C. Mitchell. 2007. An automated decision-tree approach to predicting protein interaction hot spots. Proteins-Structure Function and Bioinformatics 68:813-823.

[45] Liu, Q. A., and J. Y. Li. 2010. Protein binding hot spots and the residue-residue pairing preference: A water exclusion perspective. BMC Bioinformatics 11:244.

[46] Cho, K. I., D. Kim, and D. Lee. 2009. A feature-based approach to modeling proteinprotein interaction hot spots. Nucleic Acids Research 37:2672-2687.

[47] Guharoy, M., and P. Chakrabarti. 2009. Empirical estimation of the energetic contribution of individual interface residues in structures of protein-protein complexes. Journal of Computer-Aided Molecular Design 23:645-654.

[48] Keskin, O., B. Y. Ma, and R. Nussinov. 2005. Hot regions in protein-protein interactions: The organization and contribution of structurally conserved hot spot residues. Journal of Molecular Biology 345:1281-1294.