

A unifying motif of intermolecular cooperativity in protein associations

S.R. Accordino¹, J.A. Rodriguez Fris¹, G.A. Appignanesi^{1,a}, and A. Fernández²

¹ Sección Fisicoquímica, INQUISUR-UNS-CONICET-Departamento de Química, Universidad Nacional del Sur, Avda. Alem 1253, 8000 Bahía Blanca, Argentina

² Instituto Argentino de Matemática “Alberto P. Calderón”, CONICET, Buenos Aires 1083, Argentina

Received 5 April 2012 and Received in final form 25 May 2012

Published online: 16 July 2012 – © EDP Sciences / Società Italiana di Fisica / Springer-Verlag 2012

Abstract. At the molecular level, most biological processes entail protein associations which in turn rely on a small fraction of interfacial residues called hot spots. Our theoretical analysis shows that hot spots share a unifying molecular attribute: they provide a third-body contribution to intermolecular cooperativity. Such motif, based on the wrapping of interfacial electrostatic interactions, is essential to maintain the integrity of the interface. Thus, our main result is to unravel the molecular nature of the protein association problem by revealing its underlying physics and thus by casting it in simple physical grounds. Such knowledge could then be exploited in rational drug design since the regions here indicated may serve as blueprints to engineer small molecules disruptive of protein-protein interfaces.

1 Introduction

Protein associations are basic molecular processes in biology [1–13]. In spite of their importance, their biophysical underpinnings remain a subject of debate [1–13]. A challenging standing problem involves the characterization of hot spots [1–12]. These are few in number and provide the most significant contribution to the stability of the protein-protein interface. Knowledge-based and first-principle docking potentials have been relatively successful at predicting these singular sites [1–12], fitting the outcome of probes for experimental identification such as site-directed mutation or alanine-scanning [1]. These techniques assess the impact on binding free energy of the truncation of an individual residue side chain at the β -carbon. Notwithstanding these predictive successes, the physical nature of hot spots remains elusive. Even the establishment of general rules for hot-spot characterization has proven unfeasible so far, as has been explicitly recognized [1, 4, 5] and constitutes the scope of this work. Since attempts at rationalizing the stability of protein-protein interfaces based on pairwise interactions between the two chains is inconclusive at best, as also demonstrated in this work, we focus our attention on higher-order energetic contributions as a theoretical framework to explain and predict binding hot spots.

To prevail in water environments, soluble proteins protect their backbone hydrogen bonds (BHBs) from the disruptive effect of water attack by clustering nonpolar

residues around them [14, 15]. This exclusion of surrounding water, or wrapping effect, also enhances the electrostatic contribution by modulating the local dielectric (de-screening the partial charges) and thus stabilizes the HB [14–21]. In turn, as demonstrated previously [14, 15], underwrapped interactions are adhesive, hence promoters of protein associations because their inherent stability increases upon approach of additional nonpolar residues. Thus, the integrity of the protein-protein interface in protein complexes becomes extremely reliant on intermolecular cooperativity [14, 15]. We make this concept precise by invoking three-body correlations, whereupon a third nonpolar body protects an electrostatic interaction pairing the other two by contributing to their dehydration. Since these three-body correlations must engage the two protein molecules, the correlations must be subject to an additional constraint: One body belongs to a protein chain and the other two to its binding partner. To complete this description it is necessary to classify pairwise electrostatic interactions and detect underprotected interactions (UPIs). In a similar way as done for single free proteins [14–17], we shall make use of an abundance distribution $P(\rho)$, where ρ is the number of three-body correlations associated with an interaction within the protein-protein complex. The UPIs are crucial in defining protein associations due to their sensitivity to critical changes in intermolecular cooperativity brought about by site-directed aminoacid substitution. And since UPIs are adhesive [14–20], this physical picture leads us to characterize hot spots as the residues whose alanine substitution most drastically affects intermolecular cooperativity. This

^a e-mail: appignan@criba.edu.ar

conjecture is validated computationally in this work by combinatorially dissecting the protein-protein interfaces of structurally reported complexes that have been independently studied by alanine scanning through experimental means. The analysis boils down to a decomposition of the interface into a web of three-body cooperative interactions, easily identified from structural coordinates. Besides its scientific interest, the knowledge gained from our approach may significantly impact drug discovery endeavors [22], especially since hot spots are expected to constitute the blueprint for the design of small molecule drugs disruptive of protein-protein associations.

2 Methods

2.1 Quantifying wrapping and identifying UPIs

UPIs that involve hydrogen bonds (HBs) are named dehydrons. This structural motif has been extensively discussed in the literature and identified in soluble proteins with PDB-reported structure [13–17]. Thus, the extent of hydrogen-bond protection can be determined directly from atomic coordinates. This parameter indicates the number of three-body correlations engaging the HB and is also known as the wrapping of the bond and denoted ρ . It is given by the number of side-chain carbonaceous nonpolar groups (CH_n , $n = 0, 1, 2, 3$, where the carbon atom of these groups is not bonded to an electrophilic atom or polarized group) contained within a desolvation domain around the HB. Each wrapping nonpolar group represents the third body within a three-body correlation involving the HB. This domain is typically defined as the reunion of two intersecting spheres of fixed radius (\sim thickness of three water layers) centered at the α -carbons of the residues paired by the hydrogen bond. In structures of PDB-reported soluble proteins, BHBs are protected on average by $\rho = 26.6 \pm 7.5$ side-chain nonpolar groups for a desolvation sphere of radius 6 Å [16]. Thus, structural deficiencies lie in the tail of the ρ -distribution, *i.e.* their microenvironment contains 19 or fewer nonpolar groups, so their ρ -value is below the mean (26.6) minus one standard deviation (7.5). While the statistics on ρ -values for BHBs vary with the radius, the tails of the distribution remain invariant, thus enabling a robust identification of structural deficiencies [14–17]. In the present work we are dealing with protein complexes and accordingly we compute the ρ -values arising from intra and inter-molecular correlations. Additionally, we consider both intramolecular and (less frequent) intermolecular BHBs. Thus, for each BHB we consider the domain formed by the reunion of the two intersecting spheres of radius 6 Å centered at the α -carbons of the corresponding HB-paired residues. Then, we compute the wrapping extent or ρ -value as the number of carbonaceous nonpolar groups (CH_n , $n = 0, 1, 2, 3$, where the carbon atom of these groups is not bonded to an electrophilic atom or polarized group) that reside within such domain and which come from the side chains of any of the residues of the two proteins engaged in the protein complex. Thus, a BHB can be wrapped by apolar groups within side chains of different

residues from one or from both proteins, all of which contribute to the overall ρ -value of such BHB. The algorithm to identify dehydrons, named “Dehydron Calculator”, is freely accessible from the Web at the following location: <http://www.owl.net.rice.edu/~arifer/courses/DehydronCalculator.exe>. The wrapping concept may be spatially represented as shown in fig. 1, where two different types of three-body correlations are illustrated. Figure 1a shows an instance of intermolecular wrapping of an intramolecular BHB, while fig. 1b shows the wrapping of an intermolecular BHB. We wish to note that for the sake of clarity each of these two examples illustrates the wrapping provided by only one residue to one or a few given BHBs. Usually, each BHB is simultaneously wrapped by many different residues.

2.2 Cooperativity-based computational alanine scanning

Our virtual alanine-scanning procedure is performed by computationally replacing each residue of a protein chain (one at a time) with alanine within the 3D structure of the complex and assessing the impact of the substitution on intermolecular cooperativity (we make no simulation but work directly with the atomic coordinates of the PDB reported structure). For most residues (those with a side chain larger than that of alanine) this means truncating the residue side chain at the β -carbon so that the whole side chain is replaced by a methyl group, thus significantly reducing the extent of wrapping involving the residue. In the special case of glycine (which lacks a β -carbon) we include a methyl at the corresponding position, increasing the extent of wrapping enabled by the residue. The *in silico* scanning process entails computing the change in ρ -value generated by each Ala-substitution on each intra and intermolecular BHBs of the complex. In a first stage, we calculate the ρ -value for all BHBs from the complex structure, producing a set of wild-type ρ -values. For each mutated residue we perform the corresponding ala-substitution leaving all other coordinates unchanged and we recalculate the full set of ρ -values (mutated ρ -values). Then, in accord with our premise of intermolecular cooperativity, hot spots are predicted taking into account their role as intermolecular wrappers according to the following classes:

a) The Ala-substitution of a residue on one chain lowers the ρ -value of a BHB (an intramolecular BHB in the partner protein or an intermolecular BHB) and the mutated ρ -value of this BHB falls below $\langle \rho \rangle$. These predicted hot spots will be labeled class-A hot spots. In the cases where the final ρ -value falls below the dehydron threshold, $\rho = 19$, (dehydron creation) these class-A hot spots will be labeled A*. *The mutation of these protective (nonpolar) residues lowers significantly the wrapping they provided to BHBs at the interface, even in some cases generating dehydrons within the structure of the protein-protein complex and hence destabilizing its structure.*

b) The Ala-substitution procedure replaces a nonwrapper residue (glycine, serine, cysteine, aspartic acid or asparagine) located within the desolvation environment of

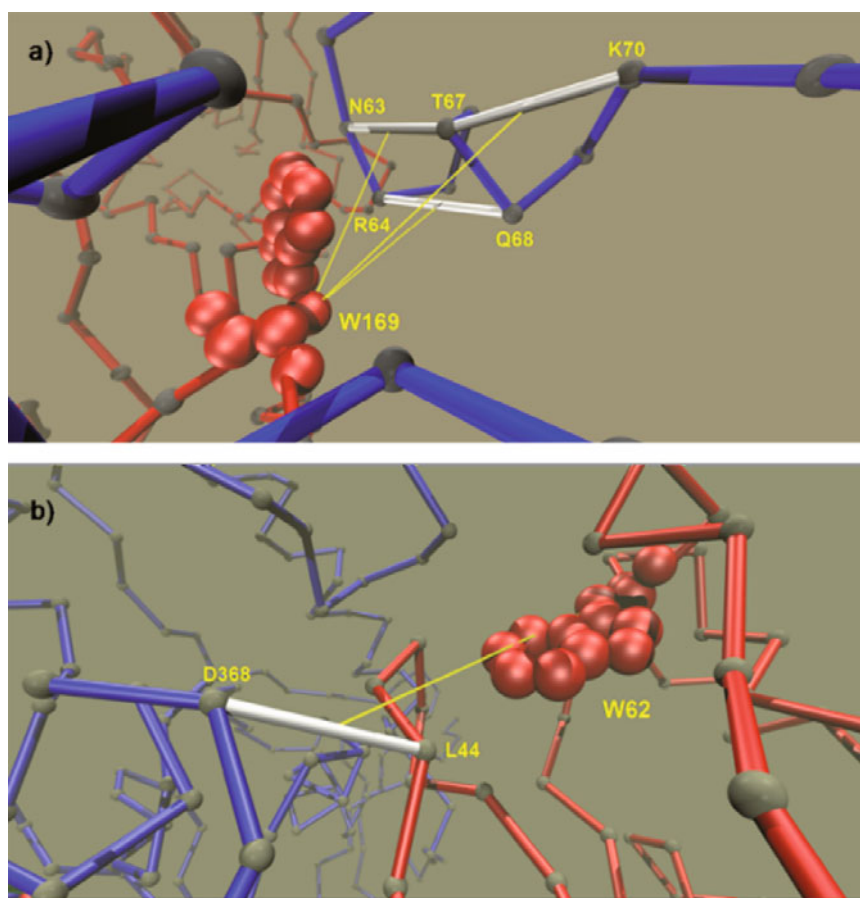


Fig. 1. (Colour on-line) Illustration of intermolecular cooperativity represented by three-body correlations: a) Trp 169 (full atomic detail) of hGHbp (red chain —light gray) wrapping three intramolecular BHBs of the hGH chain (blue chain —dark gray). The BHBs of the hGH chain are indicated by white sticks between the corresponding α -carbons. b) Similar to a) but for the complex between the HIV glycoprotein gp120 and the CD4 receptor. Here a Trp residue of the CD4 chain wraps an intermolecular BHB.

a BHB of its own protein chain whose *intramolecular* wrapping value is $\rho \leq 19$ and that is *intermolecularly* wrapped within the complex. These alanine substitutions raise the intramolecular ρ value by $\Delta\rho = +1$. The resulting predictions will be labeled class-B hot spots. In particular, when the intramolecular wrapping value of the BHB amounts exactly to $\rho = 19$, we will denote such cases as class-B* hot spots. *These mutations increase the intramolecular wrapping of a dehydron formerly existing in the same chain, thus lowering the need for intermolecular wrapping upon protein association. In the case of B* hot spots, the mutation implies a net intramolecular removal of a dehydron, thus eliminating the adhesive nature of such site.*

We decided to leave aside side-chain–side-chain HBs from the cooperativity analysis based on the following grounds: The fluctuational nature of surface side chains imposes an entropic cost associated with HB formation which makes the latter marginally stable at best [13]. Also, the wrapping statistics for side chain HBs are essentially flat with no clear distinction of the tails of the distribution do to the conformational richness of the side chains. An

a posteriori justification for the exclusion arises from the very artifactual nature of surface side-chain HBs. Particularly misleading are the large B-factors of solvent-exposed side chains and the large hydration demands of exposed polar groups, which hinder HB formation. These artifacts would yield an overwhelming number of false positives in the cooperativity analysis of the protein-protein interface (most interfacial residues would be hot spots). In turn, we shall not take into account salt bridges in our analysis, since they are not expected to significantly stabilize protein structure. These bridges are destabilizing with respect to hydrophobic replacement of both charged partners and charge burial has been shown to be usually destabilizing (see [23] and references therein). However, it is also known that for a pair of complimentary buried charges it is preferable for them to be paired by a salt bridge than to be buried isolated from each other [23]. In fact, the burial process strenghtens the ionic interaction by providing it with a desolvated environment. Thus, an Ala-mutation of a residue engaged in an intermolecular salt bridge with its complex partner protein would be destabilizing since a net buried ionic interaction would be lost. This trivial

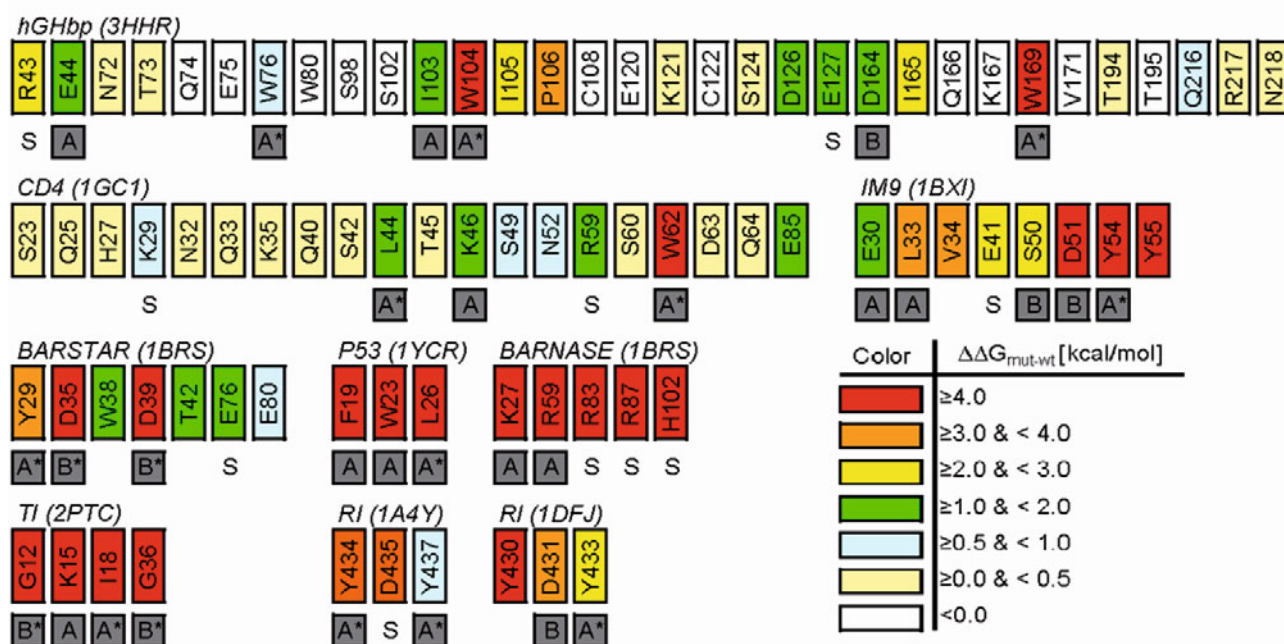


Fig. 2. (Colour on-line) Experimental alanine scanning probes contrasted against cooperativity-based *in silico* scanning for the complexes indicated. For each case we display the portion of the protein chain or the set of residues with experimental data. The colors (different gray shading) indicate the experimentally determined $\Delta\Delta G$ values for the corresponding hot spots, as shown in the scale at the right. The gray squares indicate our computational predictions, and the letter code is explained in the text. Considering the protein complexes where the whole interface has been mutated experimentally (3HHR and 1GC1) our predictions yield a p -value of $p = 0.0048$ for the most energetic hot spots ($\Delta\Delta G \geq 4$ kcal/mol, red residues). The p -value is even lower if we include the whole set of complexes.

Table 1. Detailed predictions of class-A and class-A* hot spots for one example: The human growth hormone receptor/human growth hormone interface. The first column indicates the residue mutated to Ala in the human growth hormone receptor (hGHbp), the second column indicates the BHB in the human growth hormone (hGH) whose ρ -value is affected by the mutation so as to produce a hot-spot prediction. ρ_i and ρ_f are the ρ -values of the corresponding BHB in the protein-protein complex before and after the mutation, respectively.

Residue mutated in hGHbp	BHB in hGH	ρ_i	ρ_f
Glu 44	Gln 68 - Arg 64	20	19
Trp 76	Gln 46 - Tyr 42	21	18
Trp 76	Asn 47 - Ser 43	10	9
Ile 103	Thr 67 - Asn 63	27	25
Trp 104	Thr 67 - Asn 63	27	26
Trp 104	Thr 175 - Asp 171	19	17
Trp 104	Ile 179 - Thr 175	21	19
Trp 169	Thr 67 - Asn 63	27	23
Trp 169	Gln 68 - Arg 64	20	14
Trp 169	Lys 70 - Thr 67	20	18

type of hot spots accounts for approximately 15% of all the hot spots in the complexes considered and obviously lies outside the scope of our cooperativity-based analysis.

Table 2. Predictions obtained for the different protein complexes studied. Column 2 shows the global performance of our method since it includes both class-A and class-B hot spots (that is, A, A*, B and B* hot spots, where the case A*/B* indicates dehydron creation/removal upon mutation). Column 3 corresponds to the contribution of only class-A hot spots (A and A*). Finally, column 4 summarizes the contribution of solely the dehydron creation/removal hot spots (classes A* and B*).

Experimental hot spots ($\Delta\Delta G$ value)	Prediction success (percentages)		
	A+A*+B+B*	A+A*	A*+B*
≥ 4 kcal/mol	89	61	56
≥ 3 kcal/mol	83	58	50
≥ 2 kcal/mol	79	54	46
≥ 1 kcal/mol	74	53	37

3 Results and discussion

We performed a cooperativity-based alanine scanning analysis on several protein-protein interfaces from complexes with PDB reported structure for which experimental alanine scanning results are available [2, 3] (in each case, the first protein of the complex indicated is the one mutated and we provide the PDB entry of the complex and reference of the experimen-

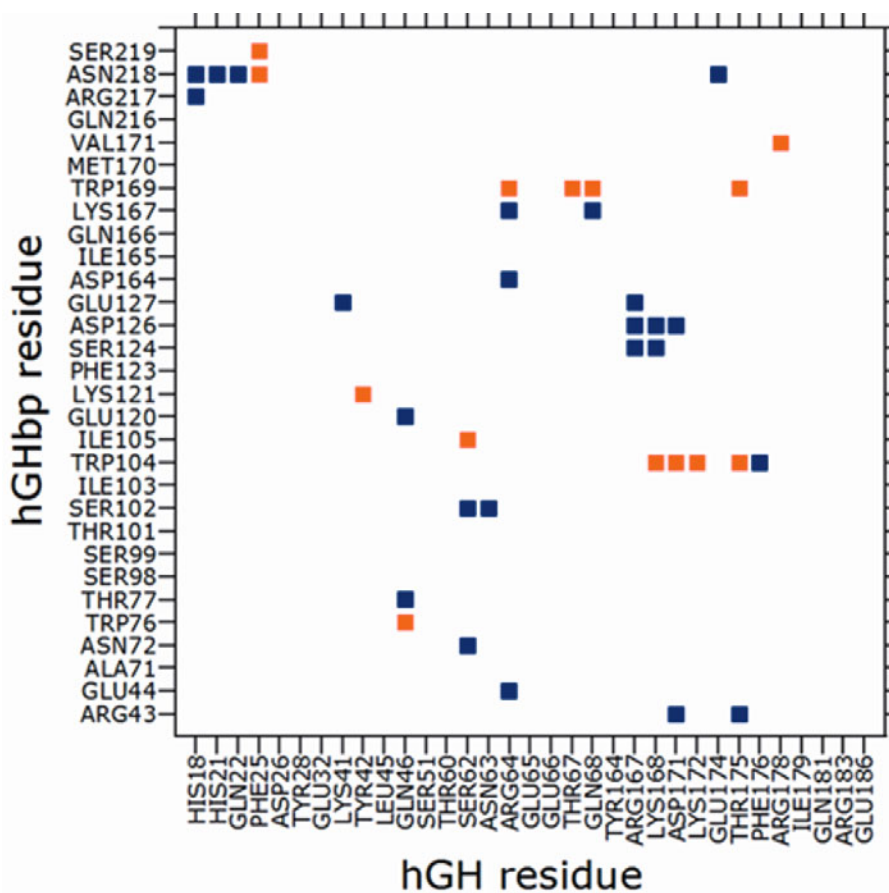


Fig. 3. (Colour on-line) Contact matrix for the hGH/hGHbp complex interface. The distance cutoff for an intermolecular contact was defined as 6 Å (see text). A red (light gray) square indicates a polar-hydrophobic mismatch between the corresponding residues, while a blue (dark gray) square implies a complementary matching contact.

tal alanine scanning results): human growth hormone receptor/human growth hormone [1] (3HHR), Trypsin inhibitor/Beta-Trypsin [24] (2PTC), P53/MDM2 [25] (1YCR), CD4/GP120 [26] (1GC1), Ribonuclease inhibitor/Ribonuclease A [27] (1DFJ), Colicin E9 immunity protein/Colicin E9 DNase domain [28] (1BXI), Barnase/Barstar [29] (1BRS), Barstar/Barnase [29] (1BRS), Ribonuclease inhibitor/Angiogenin [27] (1A4Y).

Figure 2 displays our predictions. The experimental alanine substitution of a native protein subunit yields a change in its binding free energy (ΔG) which is denoted by $\Delta\Delta G = \Delta G_{\text{mut}} - \Delta G_{\text{wt}}$, (mut = mutated, wt = wild type) and is indicated with a color scale. The cooperativity-based hot-spot predictions of our method are indicated with gray squares below the corresponding residues and are denoted by A, A*, B and B*. The letter “S” labels trivial salt bridge hot spots which are removed from the list of experimental hot spots used for the comparison with our computational method, as explained in sect. 2.

In order to display one case in full detail, in table 1 we show the hot-spots class A and class A* obtained by Ala-mutation of the human growth hormone receptor protein mutated within its complex with the human growth hormone protein.

In turn, to quantify the predicting ability of our method, in table 2 we show our global predictions over the whole set of protein complexes studied. This comparison between theory and experiment reveals that our computational procedure locates most of the experimental alanine-scanning hot spots, with optimal performance (89% prediction success) for the most significant contributors determined experimentally ($\Delta\Delta G \geq 4$ kcal/mol). The greatest contribution to such percentage, 61%, corresponds to class-A mutations (A and A*), while class B (B and B*) provides the remaining 28%. The last column of the table indicates the predictions when considering only dehydron creation, A*, and dehydron removal, B*. In consonance with our cooperativity premises, these cases are expected to constitute very important mutations and this is in fact the case, since such mutations account for 56% of the highly energetic mutations determined experimentally ($\Delta\Delta G \geq 4$ kcal/mol). Additionally, the wild-type ρ -values averaged over the residues wrapped in class-A hot spots yields $\rho = 20.3$, a value higher than the dehydron threshold ($\rho = 19$). However, when we average the mutated ρ -values we get a final $\rho = 18$, that is, below the dehydron threshold. Thus, the dehydron threshold is in fact statistically framed by the averaged wild-type and mutated ρ -values for A-class hot spots, thus revealing the

relevance of the qualitative wrapping differences for protein affinity. At this point it is worth recalling that our method disregards two-body terms unless they are engaged in a three-body correlation. This approach seems natural in view of the fact that no protein-protein interface has proven trivial at the conventional pairwise level analysis [1–12] and given the absence of clear rules for hot-spot prediction [1–12]. This last point also makes it difficult to establish a control for our results, but we have nonetheless defined an elementary one based on polar and hydrophobic complementarities. To this end, we have simply characterized residues as hydrophobic (nonpolar aromatic or aliphatic side chains) or polar (polar or charged side chains) and built a contact matrix for the complex interface. For each residue we calculated the minimum distance between its α -carbon and the α -carbons of the residues of the partner protein and between the centroid of its side chain and those of the partner side chains. When this minimum distance was below 8 Å we regarded such residue as an interfacial one. Thus, having identified the interfaces of both protein chains, a contact between a residue of one chain and another one of the other chain was considered to occur when either the distance between α -carbons of such residues or between the centroids of the side chains was less than 6 Å. The results are robust to moderate changes in the contact parameter and fit a criterium previously adopted for protein-protein interfaces [1].

Figure 3 shows the contact matrix for the hGH/hGHbp complex interface where a red mark indicates a polar-hydrophobic mismatch and a blue one indicates a complementary contact. From this figure we can learn that the interface presents a significant level of mismatches (around 37%) and thus the protein association cannot be simply rationalized as a search for pairwise polar-polar and hydrophobic-hydrophobic complementarity. More interestingly, when we restrict the analysis to the experimentally determined hot spots, the percentage of mismatches is slightly higher (42%). And if we look at the two most important hot spots (Trp 104 and Trp 169, the only residues with $\Delta\Delta G \geq 4$ kcal/mol), these residues are involved in 8 mismatches and only 1 hydrophobic-hydrophobic contact. This level of mismatching seems unavoidable given the high polar content at the protein surface which becomes buried upon creation of the complex. However, when we focus on three-body interactions, we discover that many hydrophobic residues at the complex interface approach polar residues in order to wrap BHBs in which the latter are involved.

4 Conclusions

To summarize, this work has shown that protein-protein interfaces elude standard physico-chemical analysis. Their rationalization in terms of pairwise complementarity along the contact region is unsatisfactory, especially in regards to the role of hot spots as determinants of protein associations. Against this reality, this work unravels a seemingly overlooked simple molecular motif that proves to be ubiquitous in determining protein-protein associations. Such motif is an indicator of three-body intermolecular

cooperativity. In essence, such effects arise as a group in one protein chain stabilizes (wraps) a preformed hydrogen bond in the partner chain or an inter-chain hydrogen bond, so that three bodies intervene in the interaction and not all three belong to the same chain. We have shown that hot-spot predictions based solely on this molecular attribute and defined by two pure combinatorial rules based on structural analysis of protein complexes, account for most (89%) of the hot spots experimentally determined by alanine-scanning in a set of protein complexes. Thus, the simplicity of our method contrasts with the complexity of approaches based on full-fledged potentials with explicit water (where many-body terms are subsumed in all-atom interactions). We do not deny the relevance of these predictive methods, but such avenues have not proven enlightening in terms of identifying clear molecular promoters of protein associations. By contrast, the results presented in this work comply with such imperative and might be instrumental in the rational design of small molecules aimed at disrupting protein-protein interfaces by fulfilling the wrapping capabilities of hot spots. By unraveling the many-body nature of biomolecular recognition, the reported research provides physical underpinnings of protein-protein interfaces, a long-standing problem that cannot be cast in terms of pair-wise additive terms.

Financial support from ANPCyT (PME 2006-1581) and CONICET is gratefully acknowledged. AF, GAA and JARF are research fellows of CONICET. SRA thanks CONICET for a fellowship.

References

1. T. Clackson, J.A. Wells, *Science* **267**, 383 (1995).
2. A.A. Bogan, K.S. Thorn, *J. Mol. Biol.* **208**, 1 (1998).
3. K.S. Thorn, A.A. Bogan, *Bioinformatics* **17**, 284 (2001).
4. Y. Ofran, B. Rost, *PLoS Comp. Biol.* **3**, 1169 (2007).
5. T. Kortemme, D. Baker, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14116 (2002).
6. G.-Y. Chuang *et al.*, *Protein Sci.* **19**, 1662 (2010).
7. P. Chakrabarti, J. Janin, *Proteins* **47**, 334 (2002).
8. P. Privalov *et al.*, *J. Mol. Biol.* **365**, 1 (2007).
9. J. Li, Q. Liu, *Bioinformatics* **25**, 743 (2009).
10. O. Keskin, B. Ma, R. Nussinov, *J. Mol. Biol.* **345**, 1281 (2005).
11. Z. Li, J. Li, *Bioinformatics* **78**, 3304 (2010).
12. T. Geppert, B. Hoy, S. Wessler, G. Schneider, *Chem. Biol.* **18**, 344 (2011).
13. A. Fernández, M. Lynch, *Nature* **474**, 502 (2011).
14. A. Fernández, R. Scott, *Phys. Rev. Lett.* **91**, 018102 (2003).
15. A. Fernández, R. Scott, *Biophys. J.* **85**, 1914 (2003).
16. N. Pietrosevoli, A. Crespo, A. Fernández, *J. Prot. Res.* **6**, 3519 (2007).
17. E. Schulz, M. Frechero, G.A. Appignanesi, A. Fernández, *PLoS ONE* **5**, 12844 (2010).
18. A. Fernández, J. Chen, A. Crespo, *J. Chem. Phys.* **126**, 245103 (2007).
19. A. Fernández, *Phys. Lett. A* **299**, 217 (2002).
20. A. Fernández, R. Scott, H. Scheraga, *J. Phys. Chem. B* **107**, 9929 (2003).

21. E. Schulz, L. Alarcón, G. Appignanesi, Eur. Phys. J. E **34**, 114 (2011).
22. J.A. Wells, C.L. McClendon, Nature **450**, 1001 (2007).
23. C.V. Sindelar, Z.S. Hensch, B. Tidor, Protein Sci. **7**, 1898 (1998).
24. M.J.M. Castro, S. Anderson, Biochemistry **35**, 11435 (1996).
25. A. Böttger *et al.*, J. Mol. Biol. **269**, 744 (1997).
26. A. Ashkenazi *et al.*, Proc. Natl. Acad. Sci. U.S.A. **87**, 7150 (1990).
27. C.-Z. Chen, R. Shapiro, Proc. Natl. Acad. Sci. U.S.A. **94**, 1761 (1997).
28. U.C. Kühlmann *et al.*, J. Mol. Biol. **301**, 1163 (2000).
29. G. Schreiber, A.R. Fersht, J. Mol. Biol. **248**, 478 (1995).