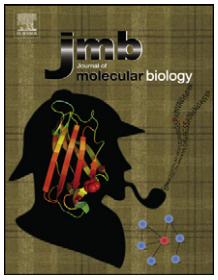

Proteins Feel More Than They See: Fine-Tuning of Binding Affinity by Properties of the Non-Interacting Surface

Panagiotis L. Kastritis, João P.G.L.M. Rodrigues, Gert E. Folkers, Rolf Boelens and Alexandre M.J.J. Bonvin

Bijvoet Center for Biomolecular Research, *Faculty of Science, Department of Chemistry, Utrecht University, Padualaan 8, 3584CH Utrecht, The Netherlands*

Correspondence to Alexandre M.J.J. Bonvin: *a.m.j.j.bonvin@uu.nl.*
http://dx.doi.org/10.1016/j.jmb.2014.04.017
Edited by A. Panchenko

Abstract

Protein–protein complexes orchestrate most cellular processes such as transcription, signal transduction and apoptosis. The factors governing their affinity remain elusive however, especially when it comes to describing dissociation rates (k_{off}). Here we demonstrate that, next to direct contributions from the interface, the non-interacting surface (NIS) also plays an important role in binding affinity, especially polar and charged residues. Their percentage on the NIS is conserved over orthologous complexes indicating an evolutionary selection pressure. Their effect on binding affinity can be explained by long-range electrostatic contributions and surface–solvent interactions that are known to determine the local frustration of the protein complex surface. Including these in a simple model significantly improves the affinity prediction of protein complexes from structural models. The impact of mutations outside the interacting surface on binding affinity is supported by experimental alanine scanning mutagenesis data. These results enable the development of more sophisticated and integrated biophysical models of binding affinity and open new directions in experimental control and modulation of biomolecular interactions.

Introduction

In biology, description of any process occurring in the cell leads inevitably to a direct listing of protein complexes that are implicated as its essential participants [1,2]. Cellular events encompass non-obligate (transient) protein–protein interactions that accurately define the interactome (the complete set of macromolecular interactions occurring within a cell) in a time- and location-dependent manner [3]. Knowledge on the three-dimensional structure of a complex allows describing its interactions at atomic detail. However, it is the binding affinity, in relation to the protein concentrations present in a cell, that defines whether or not complex formation occurs, while the underlying k_{on} and k_{off} rates determine the timescale of association and dissociation, respectively. The corresponding equilibrium dissociation constant ($K_d = k_{off}/k_{on}$), which can be empirically translated into the Gibbs free energy of binding ΔG ($\Delta G = -RT\ln K_d$), is commonly used to describe the affinity of an interaction. Unsurprisingly, since the timescales of cellular processes are extremely diverse, K_d values span more than 11 orders of magnitude, from high millimolar (mM) to low femtomolar (fM) concentrations [4]. Accordingly, k_{on} and k_{off} rates of protein–protein complexes underline this broad spectrum of affinities, covering 8 (10^2–10^9 M^{-1} s^{-1}) and 11 (10^2–10^{-8} s^{-1}) orders of magnitude, respectively [4,5].

Weak transient protein–protein interactions are responsible for various sequences of rapid chemical conversion events that occur in the cell (e.g., signaling cascades and electron transfer) [6]. These interactions fall on the high end of the K_d spectrum ($K_d < 10^{-6}$ M) and have a half-life of less than 1 s, comparable to that

of an enzymatic reaction such as phosphotransfer [7]. Recently, with the emergence of highly sensitive experimental techniques, transient interaction analysis has become tractable [8]. As a consequence, in a recently compiled benchmark of protein–protein complexes with known structure and experimental affinity, 26% of the entries correspond to weak transient cellular interactions of different nature, showing various amounts of conformational changes upon binding [4].

Modulating the binding affinity of transient protein–protein complexes offers new opportunities to control interaction profiles and design innovative therapeutics [9], as illustrated by structure-based design of protein–protein interaction inhibitors targeting complexes involved in cancer [8] and neurodegenerative processes [10]. Although deciphering the underlying details has proved to be a daunting task [7], a systematic and meticulous approach can set the basis for the discovery of new interaction profiles or of inhibitors that can serve as therapeutic agents [1,7]. Such an approach must integrate biophysical models that aim at calculating binding affinities of transient complexes and, ideally, predicting changes in the binding affinity of the partners as a consequence of point mutations [11,12].

Basic models for binding affinity estimation were already proposed more than 20 years ago [13,14]. Horton and Lewis considered both polar and apolar fractions of the surface that are buried upon complexation [buried surface area (BSA): BSA_{pol} and BSA_{apol}]. Their model (will be referred to from now on as the *classical interface model*) was in fact based on the initial observation back in the 1970s from Chothia and Janin [15] that the BSA is directly related to binding affinity. A basic model for predicting changes in heat capacity (ΔCp) of protein–protein interactions was developed independently by Freire at the same time, albeit initially introduced for protein folding [14]. It accounts for several thermodynamic parameters [ΔCp, the enthalpy change ΔH and (partially) the entropy change ΔS] that are explained using BSA_{pol} *and* BSA_{apol}. Since then, even the most recent models [16–18] estimate the energetics of a protein–protein complex by considering only features of the interface. Their performance is limited however when tested against larger sets [11,12,16–19]. The models developed to date still fail to accurately predict binding affinity [20] or discriminate binders from non-binders [21,22]. These weaknesses can be attributed to several factors, including the quality of the experimental data used to parameterize the models, conformational changes of the proteins occurring upon binding, (the absence of) co-factors required for binding and possible solvent and allosteric effects [23]. While some models can reasonably well describe the energy of a (near) rigid binding complex [16,17], affinity prediction is not accurate for complexes of different functions that undergo substantial conformational changes upon binding, regardless of the model used [20].

All present models are based on the hypothesis that properties of the BSA and the area in its close vicinity (rim) are sufficient to provide a complete description of the binding affinity of protein–protein interactions [11,12]. Two arguments are usually given to support this conjecture. First, the BSA has been proven to correlate with binding affinity [15]. When tested against a large dataset, the relation was partially holding (r = 0.5) [4]. Still, for most complexes tested, the predicted ΔG values were lower than the experimentally determined ones indicating that other physicochemical descriptors must play a role in determining the binding affinity. Second, hot spot residues, occasionally found in protein–protein interfaces via alanine scanning mutagenesis [24], clearly illustrate that very few interfacial residues may account for a large fraction of the interaction energy. Evidence for such hot spots comes primarily from rigid and tight protein–protein interactions. This remains to be experimentally explored for transient complexes in particular [7]. Consequently, until now, research in modeling protein–protein interaction affinity has solely focused on the interface, neglecting the potential role of the remainder of the surface, the non-interacting surface (NIS) [11].

It is long known that protein interactions may involve long-range effects from non-interacting regions (both surface and protein interior), including classical examples of allostery [25,26] but also through long-range inter-residue communication pathways [25]. In enzymatic catalysis, mutations in residues distant in space from the binding site can change the enzymatic rate up to a factor of ~200 [27]. Such effects have been rationalized by the energy landscape theory of protein binding [28], being attributed to deviations [29] from the "principle of minimal frustration" [30], stemming from the unbound state, the latter signifying that a strong energetic bias toward the bound state exists in the case of protein interactions [31]. Local deviations from minimal frustration present on the protein structure (usually stemming from polar and charged surface residues [32]) contribute to the functional characteristics of proteins and their complexes and correlate with long-range effect on protein dynamics [33]. Structural dynamics and charge distribution not only are correlated to function but also reflect evolutionary pathways of macromolecular assemblies [34,35]. It is also known that hydration, even of residues that are not directly present at the interface, may drastically alter protein dynamics [36] and assembly conformation [37,38] and may improve docking prediction of interacting biomolecules [39–42].

So far, a quantitative relationship between NIS characteristics and binding affinity of protein–protein interactions has never been derived mainly due to the absence of a proper structure-based dataset. In this study, we probe the relationship between protein structure and binding affinity. We demonstrate that binding affinity is also affected by properties of the NIS, something that all biophysical models aiming to

explain the interaction affinity have been neglecting so far and propose a novel structure-based global surface model for binding affinity prediction.

Our results collectively favor protein–protein interaction regulation through long-range electrostatics and preferential solvation of non-interacting protein complex surface patches. The physical chemistry of the NIS fractions responsible for affinity regulation is very similar to those reported by the Wolynes group [32] regarding residues with increased frustration that causally contribute to functional characteristics of proteins and their complexes.

Results

Properties of the NIS are related to binding affinity

In light of the limitation of current models [4,19–22], we sought to quantitatively assess if and which physicochemical properties of protein–protein complexes are capable of describing binding affinity. The protein–protein interactions that are being studied here occur at distinct timescales and in several cellular compartments, with known binding affinity. For 51 complexes, protein–protein association (k_{on}) and dissociation (k_{off}) rates could be compiled (see Table S1). This means that structure–affinity relationships derived for these complexes, if causal, should represent fundamental properties of the interaction process. We evaluated 39 different properties of these complexes for correlation with binding affinity (see Tables S2 and S3). A surprising and unanticipated outcome of this analysis is that, next to the expected properties of the interface, a significant correlation of few NIS characteristics with binding affinity (Fig. 1a and b) is observed. In particular, the percentage of polar residues on the NIS ($^{\%}AA^{NIS}_{pol}$) is found to relate to binding affinity (K_d) ($N = 144$, $r = 0.35$, $p < 0.0001$). This effect is even stronger when considering only complexes with limited conformational change (≤ 1 Å RMSD) ($N = 73$, $r = 0.42$, $p < 0.0001$). We next sought to determine if the $^{\%}AA^{NIS}_{pol}$ is related to the k_{on} and/or k_{off} of protein–protein interactions. A significant correlation of $^{\%}AA^{NIS}_{pol}$ with the k_{off} rates ($-\log k_{off}$) was observed ($N = 51$, $r = 0.30$, $p < 0.05$) (Fig. S1 and Table S4), whereas the relation to k_{on} was insignificant (Table

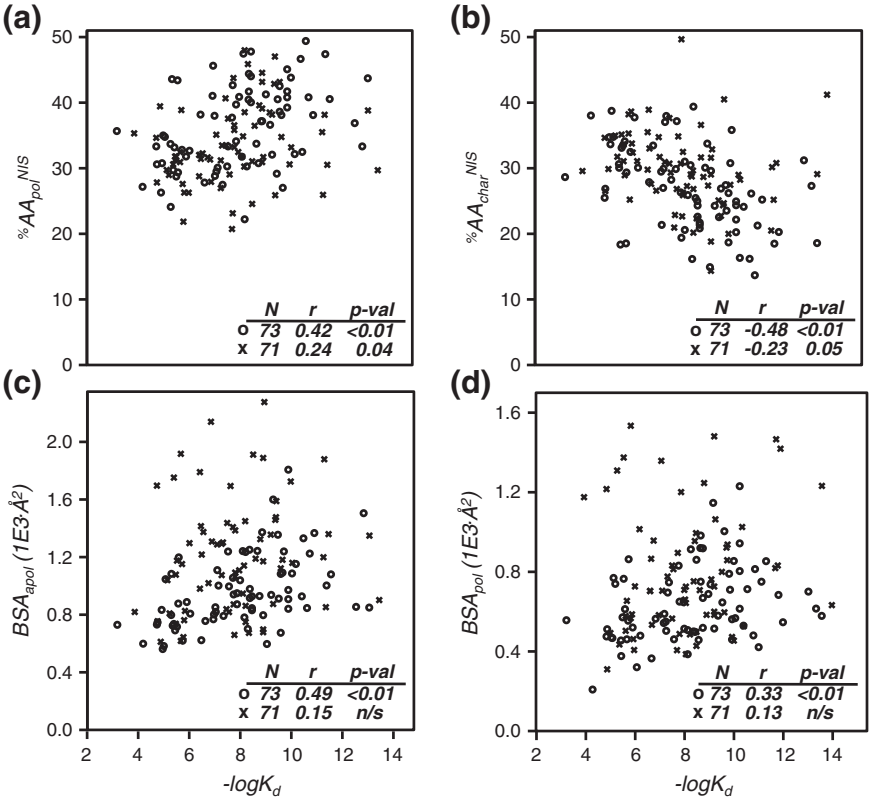

Fig. 1. Structural properties of protein–protein complexes showing correlation with binding affinity. Correlations of the percentage of polar ($^{\%}AA^{NIS}_{pol}$) and charged residues ($^{\%}AA^{NIS}_{char}$) on the NIS of protein–protein complexes with binding affinity are shown in plots a and b, respectively. Plots c and d illustrate correlations of the apolar (BSA_{apol}) and polar (BSA_{pol}) BSA with binding affinity.

S5). For the percentage of charged residues located on the NIS ($^{\%}AA^{NIS}_{char}$), inverse correlations are found against k_{off} rates ($-\log k_{off}$) ($N = 51$, $r = -0.46$, $p = 0.0005$). $^{\%}AA^{NIS}_{char}$ is also significantly related to K_d for all complexes of the benchmark ($N = 144$, $r = -0.37$, $p < 0.0001$). Again this correlation improves when only protein–protein complexes with limited conformational changes are considered ($N = 73$, $r = -0.48$, $p < 0.0001$) (Fig. 1).

Notably, only few other properties (both interfacial and non-interfacial; see Table S2) share significant relations with binding affinity, k_{on} or k_{off} (Tables S3–S5). In particular, the BSA-related properties (Fig. 1c and d), for example, BSA_{pol}, BSA_{apol}, BSA_{total} [15], the number of atoms/residues in the interface ($N_{atomsINT}$ and $N_{residINT}$) show comparable correlation coefficients to the ones shared by properties of the NIS (Fig. 1 and Tables S3–S5). However, even for these, the correlation vanishes for complexes that undergo substantial conformational change upon binding (for details, see Table S3).

In order to ensure that NIS parameters derived in this study are not the negative image of their respective interface accessible surface areas or their percentages, we correlated $^{\%}AA^{NIS}_{pol}$ with BSA_{pol} and fraction of BSA_{pol} of the interface. No correlation was observed ($N = 144$, $|r| < 0.10$, $p = 0.2279$) indicating that the BSA effect is not being double-counted (Fig. S2a and b). $^{\%}AA^{NIS}_{pol}$ and $^{\%}AA^{NIS}_{char}$ are strongly anticorrelated ($N = 144$, $r = -0.78$, $p < 0.0001$; Fig. S2c) and their percentages on the NIS varies by more than 35% among the complexes (Fig. S2d). In addition, complexes with the same percentage of charged residues on the NIS can have a different net charge hence different overall electrostatic contribution to the interaction. Electrostatics is a well-known contributor in association of near-rigid, diffusion-limited complexes and is fundamental in transient-complex theory [5,43–46]. Complexes in this study are mainly k_{off} limited (Fig. 2) and, therefore, significant, but predictive correlations between charge calculated with HyPare [47] and k_{on} are derived for all but very slow binders ($N = 48$, $r = 0.59$, $p < 0.0001$), consistent with the transient-complex theory (Fig. S3a).

We observe a strong correlation between the k_{off} rates ($-\log k_{off}$) and the K_d of the complexes, reaching $r = 0.88$ and $r = 0.72$ for 26 rigid and 25 flexible protein–protein complexes, respectively (Fig. 2a). Therefore, compared to k_{on}, the physical principles that govern the k_{off} rates can explain a larger fraction of the free energy of a complex (Fig. 2a and b).

This extensive analysis confirmed that (a) the extent of conformational changes is a limiting factor for correlation with binding affinity for all studied properties and (b) NIS properties, $^{\%}AA^{NIS}_{char}$ and $^{\%}AA^{NIS}_{pol}$, are related to the binding affinity and in particular to the dissociation rate of protein–protein complexes. The correlation coefficients calculated for these NIS properties are significant and, among the highest observed, next to the well-known relations of BSA and its apolar fraction with binding affinity [4,13]. The lower correlations observed for flexible complexes indicate that there must be contributions to binding affinity that are currently beyond our understanding and predictive capability (e.g., conformational entropy changes).

Chemical properties of the NIS are conserved within orthologous complexes

The observation that the percentage of polar and charged residues on the NIS influences binding affinity needs further rationalization since correlation does not imply causality *per se*. We therefore postulated a simple hypothesis: If these fractions indeed modulate binding affinity, one would expect some degree of

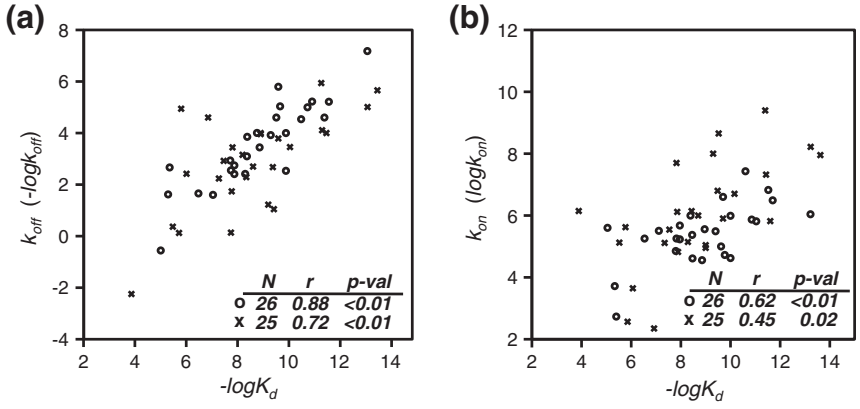

Fig. 2. Kinetic properties of protein–protein complexes showing correlation with binding affinity. Correlations between k_{on} and k_{off} rates with binding affinity of protein–protein complexes are shown in plots a and b, respectively. Complexes that bind with limited conformational changes (i-RMSD < 1.0 Å) are indicated by empty circles whereas complexes showing larger conformational changes are indicated with an "*x*" in the plots.

conservation among orthologues (homologous complexes across species sharing the same function). To answer this question, we assembled a dataset of 47 binary complexes, constituting a subset of the full set being investigated [4]. For these, both interacting molecules are composed of a single polypeptide chain and both chains correspond to the same species. We then searched for sequence homologues (orthologues) using sensitive HMMER profiles [48], finding 2896 homologous interactions in total for all 47 complexes (Table S6). The Markov search did not have a threshold for sequence identity but a bias toward the protein size; the latter should be ±30% when compared to the template sequence. This bias was set in order to be able to generate good-quality homology models, avoiding large insertions/deletions that would hamper a reliable analysis of the surface properties. The resulting dataset encompasses various orthologues for each complex, ranging in number from 2 to 226. The 47 complexes represent various types of interactions, of different nature, affinity and degree of conformational changes. The corresponding $^{\%}AA^{NIS}_{char}$ and $^{\%}AA^{NIS}_{pol}$ cover a wide range (from 18% to 50% and from 22% to 47% for $^{\%}AA^{NIS}_{char}$ and $^{\%}AA^{NIS}_{pol}$, respectively). The differences in $^{\%}AA^{NIS}_{char}$ and $^{\%}AA^{NIS}_{pol}$ observed between the 47 complexes are much larger than those observed among orthologues for a given complex (compare Fig. 3a and b and c). For all complexes, $^{\%}AA^{NIS}_{char}$ and $^{\%}AA^{NIS}_{pol}$ are very well preserved across orthologues, down to approximately

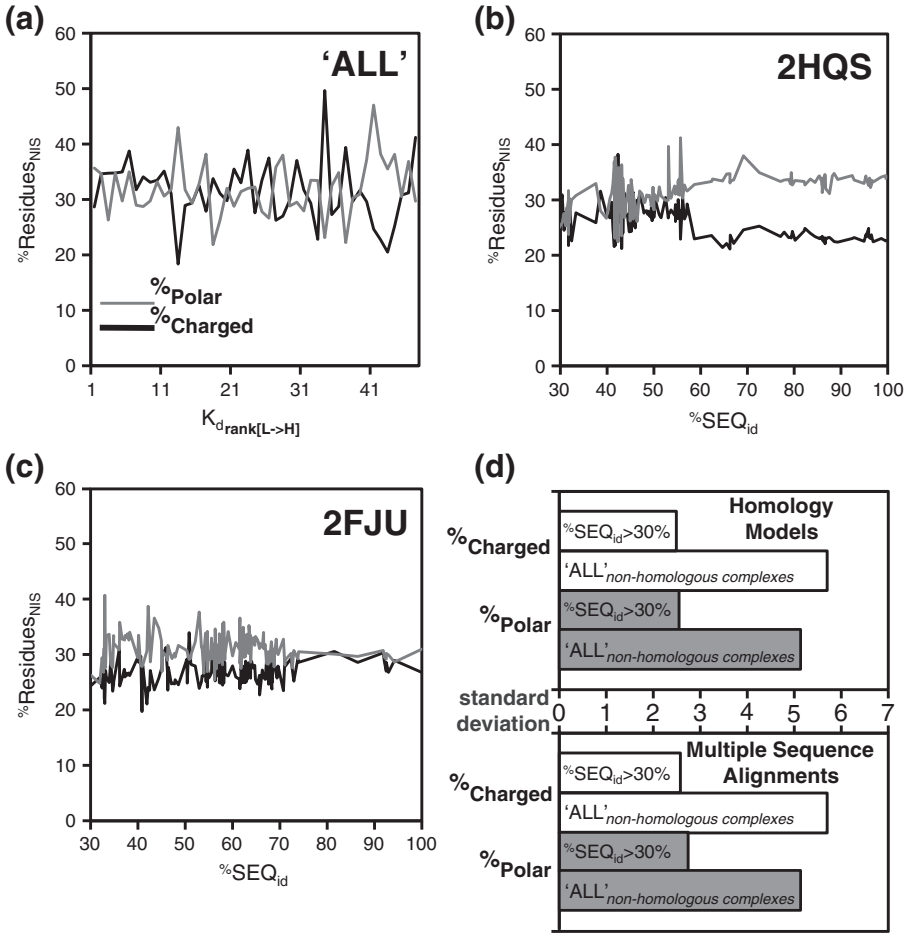

Fig. 3. Evolutionary conservation of properties of the NIS among orthologues. (a) Percentage of polar and charged residues on the NIS ($^{\%}AA^{NIS}_{pol}$ and $^{\%}AA^{NIS}_{char}$) plotted for all complexes sorted according to their binding affinity (from low to high) ($K_{drank(L\rightarrow H)}$). This plot highlights the large variation in $^{\%}AA^{NIS}_{pol}$ and $^{\%}AA^{NIS}_{char}$, which translates into large standard deviations [see (d)]. (b and c) Percentage of polar and charged residues on the NIS as a function of sequence identity for selected complexes (2HQS and 2FJU). These clearly show conservation of the surface properties among orthologues. (d) Bar charts of the pooled standard deviations of the percentage of charged and polar residues on the NIS calculated for all orthologues found for each protein–protein complex compared to the standard deviation for all non-homologous 47 protein–protein complexes. Top chart concerns data derived from constructed homology models; bottom chart concerns data derived from multiple sequence alignments.

30% sequence identify. This is further highlighted by the corresponding standard deviations: the average standard deviation of $^{\%}AA^{NIS}_{char}$ and $^{\%}AA^{NIS}_{pol}$ over the sets of orthologues is ~2.6 times smaller than those across the set of 47 unrelated complexes (Fig. 3d). This difference is equally pronounced when $^{\%}AA^{NIS}_{char}$ and $^{\%}AA^{NIS}_{pol}$ are calculated either from the generated molecular models (Fig. 3d, upper plot) or directly from multiple sequence alignments (Fig. 3d, bottom plot).

Overall, this analysis reveals that properties of the NIS ($^{\%}AA^{NIS}_{char}$ and $^{\%}AA^{NIS}_{pol}$) for all binary protein–protein interactions studied remain fairly constant when orthologues are considered, even for complexes that share very low sequence identity (down to 30%). For non-homologous complexes, these attributes vary much more.

Distant electrostatic effects are important in protein–protein interactions

Having identified discriminating relations of chemical properties of the NIS with binding affinity and highlighted that they are conserved across orthologues, we next investigate their origin. A logical cause to explain the effect of the percentage of charged residues of the NIS on affinity could be electrostatics. In order to test this hypothesis, we created a simple biophysical model to estimate the electrostatic contribution to the binding energy as a function of the distance from the center of the interface: for all protein–protein complexes, $E_{elec,binding}$ defined as $E^{AB}_{elec} - (E^{A}_{elec} + E^{B}_{elec})$ was calculated for increasing concentric shells around the center of the interface (Fig. 4a). Our simple model considers the electrostatic contribution of each titratable group within a defined sphere and assigns the pK_a and charge state according to the Henderson–Hasselbalch equation; then the electrostatic energy is calculated using a simple coulomb function. The average electrostatic contribution over all 144 complexes increases (becomes more negative) as a function of the sphere radius up to approximately 40 Å after which a plateau value is reached (Fig. 4b). This simple analysis reveals a profound effect of electrostatics on protein–protein interactions even when distances up to 40 Å from the interface are considered.

Each complex has, however, its own electrostatic profile, meaning that the distance at which the

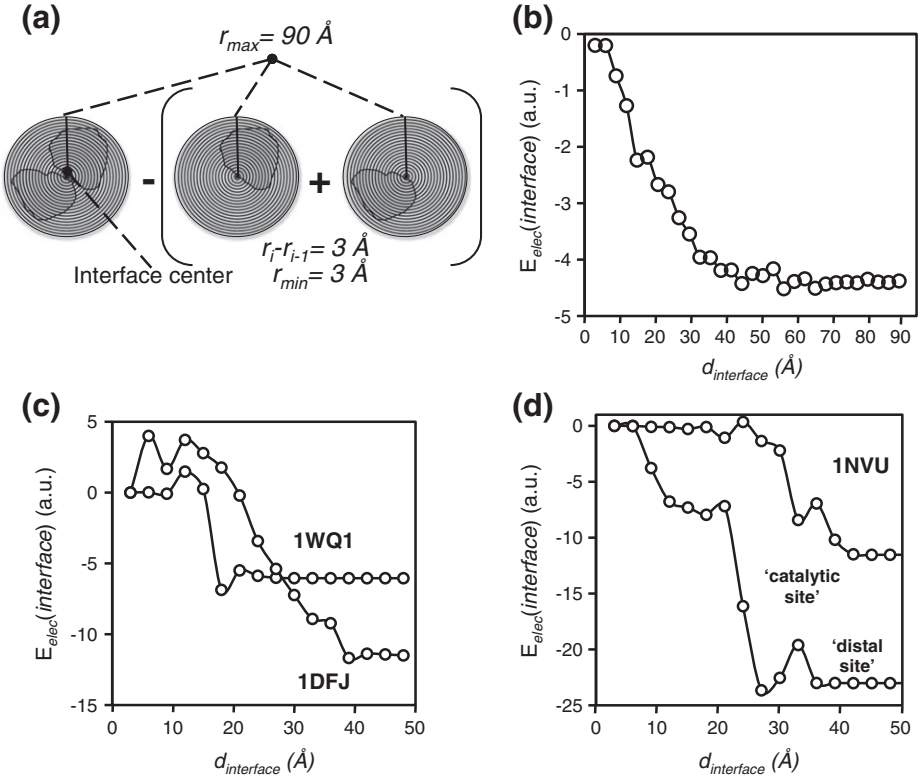

Fig. 4. (a–d) Electrostatic effects on the affinity of protein–protein complexes stemming from the NIS as a function of distance from the center of the interface. (a) Schema of the methodology used to calculate the coulomb potential as a function of distance from the center of the interface. (b) Average electrostatic energy as a function of the distance from the interface over the 144 complexes. (c and d) Examples of electrostatic profiles for three complexes as a function of distance from the center of the interface (see Results for details).

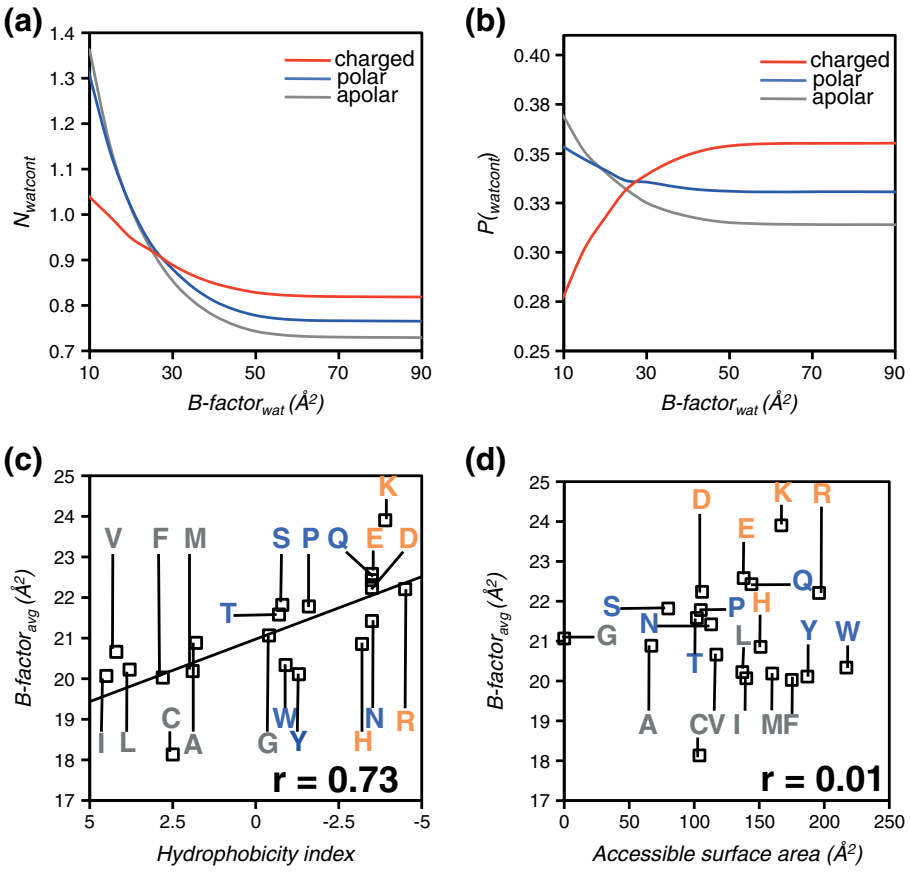

Fig. 5. (a–d) Analysis of water–surface residue contacts as a function of the residue type: charged (red), polar (blue) and apolar (gray). (a) Average number of contacts of water molecules with the protein surface as a function of the waters corresponding Debye–Waller factor (*B*-factor). (b) Contact probabilities of water molecules with fractions of the protein surface as a function of its corresponding Debye–Waller factor. (c) Average *B*-factor of water molecules contacting the various amino acids classified according to the Kyte–Doolittle hydrophobicity scale. (d) Average *B*-factor of water molecules contacting surface residues as a function of the residue-type accessible surface area demonstrating that there is no correlation with the relative accessible surface area of these residues.

electrostatic effect does no longer change is unique for each complex in the benchmark. For example, the interaction between human H-Ras bound to guanosine diphosphate and the guanosine triphosphatase (GTPase)-activating domain of the human GTPase-activating protein p120GAP (GAP-334) (PDB ID: 1WQ1) exhibits electrostatic effects that are significant up to approximately 20 Å after which a plateau is observed (Fig. 4c). The extremely affine complex of ribonuclease A with the RNase inhibitor (PDB ID: 1DFJ) shows a similar decay but only reaches a plateau at 40 Å (Fig. 4c). We do not observe a dependency of the electrostatic contribution profile on the size of the complex. A very interesting case from the benchmark is the complex formed between Ras GTPase•GTP and son of sevenless protein. This binary complex is striking in the sense that Ras has two binding sites for SoS: the "catalytic" site and the "distal" site, both having different affinities toward SoS (PDB ID: 1NVU). In line with this, the electrostatic contribution of the NIS on the two sites differs: the electrostatic contribution levels off at 25 Å for the distal site (K_d = 3.6E-06 M), whereas for the catalytic site (K_d = 1.9E-06 M), it increases up to about 40 Å (Fig. 4d). Our analysis was performed considering 100 mM salt concentration. When considering increasing salt concentrations (150 mM, 200 mM and 250 mM), the effect significantly becomes less pronounced. At physiological salt concentrations (100–150 mM), we estimate that such an effect would level off beyond 30 Å from the center of the interface (Fig. S3b). We should note, however, that salt dependence calculations are an overestimate because salt effects will not affect the interactions mediated through the low dielectric interior of the proteins.

This analysis shows that, although the electrostatic signature of the non-interaction surface is unique for each complex, distant effects are observed for a

Table 1. Correlation coefficients of various structural parameters and regression models for binding affinity predictors

(a) Performance of the components of the prediction models against $-\log K_d$

Descriptors	Class of complexes[a] (number of complexes in that class)				Original finding
	All (143)	**≤1** (72)	**≤1.5** (110)	**>1.5** (33)	
Number of atoms in the interface	**0.28** (<0.001)	**0.48** (<0.001)	**0.36** (<0.001)	0.11 (n/s)	Chothia and Janin as BSA [15]
Buried polar surface area	**0.17** (0.027)	**0.33** (<0.005)	**0.23** (0.012)	0.18 (n/s)	Horton and Lewis [13]
Buried apolar surface area	**0.26** (<0.002)	**0.51** (<0.001)	**0.34** (<0.001)	0.10 (n/s)	Horton and Lewis [13]
%Polar residues on the surface	**0.34** (<0.001)	**0.42** (<0.001)	**0.37** (<0.001)	0.22 (n/s)	This work
%Charged residues on the surface	**−0.37** (<0.001)	**−0.46** (<0.001)	**−0.35** (<0.001)	**−0.56** (<0.001)	This work

(b) Performance of the binding affinity prediction models

Models	**All** 143	**≤1** 72	**≤1.5** 110	**>1.5** 33	Proposed contributors to affinity
"Classical interface model" [13]	**0.26** (<0.002)	**0.49** (<0.001)	**0.34** (<0.001)	**0.15** (n/s)	Polar and apolar BSA
"Global surface model" (this work)	**0.48** (<0.001)	**0.64** (<0.001)	**0.54** (<0.001)	**0.36** (0.040)	Polar and Charged surface, number of atoms in the interface

[a] **All** denotes the entire dataset; the **≤1.0, ≤1.5** and **>1.5** classes denote complexes with i-RMSD ≤1.0 Å, ≤1.5 Å and >1.5 Å, respectively. The first number corresponds to the Pearson's product-momentum correlation coefficient (r) and the number in parenthesis indicates the p-value ($p < 0.05$ is considered significant).

plethora of these complexes with, on average, a significant contribution up to 40 Å from the interface and are related to the k_{off}.

More stable water molecules are observed near polar residues in high-resolution crystal structures

The observed relation between binding affinity and the percentage of polar residues on the NIS is remarkable. How can we rationalize their effect on binding affinity? Physicochemical properties of protein surfaces define a unique hydration layer around the protein. Due to their nature, polar and charged residues on the protein surface are long known to be "hot" hydration spots (also referred to as "wet spots") [49], directly influencing the formation of this hydration layer. Although all surface residues can contribute to the stabilization of the water layer, water near polar uncharged residues should only exchange with other water molecules, whereas water near charged groups might exchange with both water and soluble ions. We hypothesize that the polar residue contribution to the stabilization of a complex could stem from a more "stable" hydration layer that might "protect" the complex from other interactions, explaining the observed effect on the dissociation constant. We tested this hypothesis by analyzing water properties on the surface of ultra-high-resolution (<1 Å) crystal structures of proteins (see Table S7). Their surfaces were divided into polar, apolar and charged parts according to the Kyte–Doolittle hydrophobicity scale [50]. Since the oxygen atoms of water molecules can be safely recognized in such high-resolution electron density maps, contact frequencies of these with surface residues were calculated (N_{cont}^{wat}) and classified according to the type of residue (apolar, polar, charged). We analyzed first the number of contacts formed as a function of the *B*-factor of a water molecule. A contact was counted if the distance between the water oxygen and a protein heavy atom was ≤3.9 Å. Water molecules with high *B*-factors (approximately ≥30) make on average less than 2.5 contacts with the protein surface (Fig. 5a) while water molecules with *B*-factor values of 30 or lower form more interactions with the protein surface, reaching an average value of 3.6 contacts for waters with *B*-factor ≤ 10. Water molecules with *B*-factor ≤ 10 tend to form more contacts with polar ($N\text{polar}_{cont}^{wat} = 1.25$) and apolar surface fragments ($N\text{apolar}_{cont}^{wat} = 1.30$), compared to contacts with charged surface fragments ($N\text{charged}_{cont}^{wat} < 1.00$). When contact frequencies are transformed into contact propensities of water with the three types of amino acids, polar, apolar and charged, «stable» waters (*B*-factor < 30) are less probable to contact charged amino acids (Fig. 5b). Further analysis indicated that the average *B*-factor of a water close to charged amino acids is significantly higher than that for polar and apolar amino acids independently of the *B*-factor threshold used. This can clearly be seen when plotting the average *B*-factor against the Kyte–Doolittle hydrophobicity scale of amino acids (Fig. 5c). This relation is not affected by the

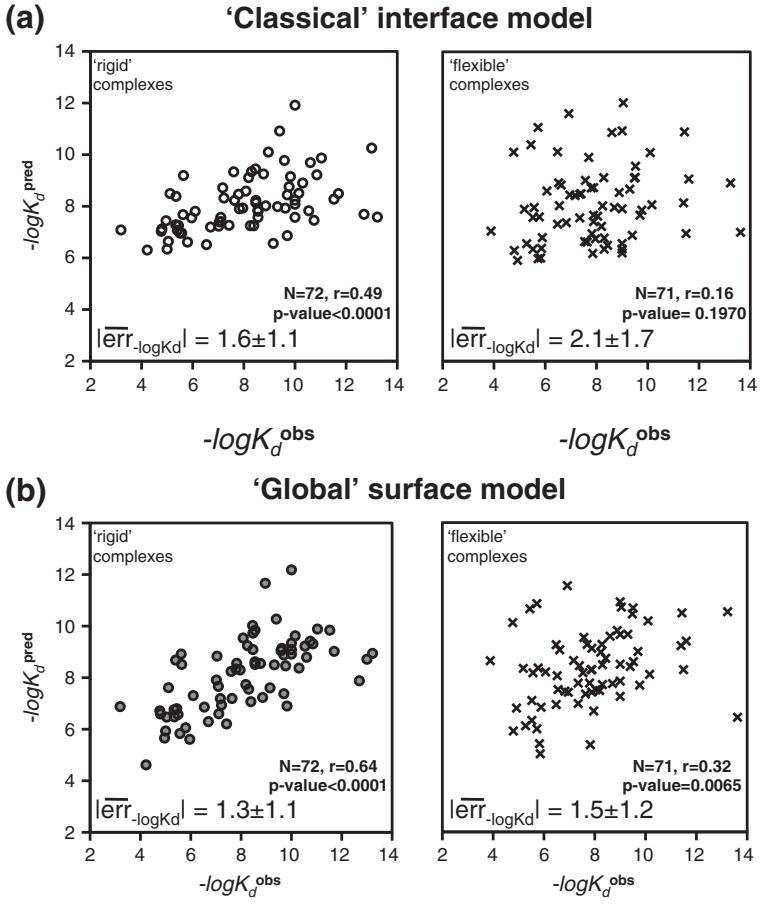

Fig. 6. (a) Classical interface model [13] [Eq. (1)] and (b) global surface model introduced in this work [Eq. (2)] for the prediction of protein–protein binding affinities. The global surface model accounts for both interface and non-interface surface contributions (see the main text). Both models were optimized on rigid complexes using 4-fold cross-validation (left panels) and blindly tested on the flexible complexes (right panels). The left panels show the cross-validated predictions.

corresponding accessible surface area of the residues (Fig. 5d).

The main result from this analysis is that the surface of proteins shows different contact propensities with water molecules depending on the chemical nature of the amino acids. Proteins with highly polar but uncharged surfaces tend to have more stable waters in their hydration layer and the underlying *B*-factors of those water molecules close to polar residues is on average lower compared to the corresponding ones close to charged amino acids.

In combination with the observation that polar non-interface surface regions do experience selection pressure [51,52], the abovementioned results suggest a mechanism in which highly polar but uncharged surfaces result in a stable water layer that might protect the generated complex from ion penetration and unintended protein–protein interactions. This would thus rationalize the observation that both $^{\%}AA^{\mathrm{NIS}}_{\mathrm{char}}$ and $^{\%}AA^{\mathrm{NIS}}_{\mathrm{pol}}$ are contributors to the binding affinity of protein–protein interactions.

The global surface model: development and validation

The prevailing model of Horton and Lewis based on a dataset of 15 protein–protein complexes decomposes the BSA into polar and apolar contributions to predict the affinity of protein–protein interactions [13]. This is consistent with the results presented here, particularly for the apolar BSA (BSA_{apol}) and the overall contribution of the BSA (Fig. 1). In our extended dataset, however, only complexes undergoing minor conformational changes [interface-RMSD (i-RMSD) $\leq$ 1.0 Å] still retain some correlation between buried apolar and polar surface areas and affinity (Fig. 1). We reparametrized the classical interface model [13] [Eq. (1)] based on our set of rigid 72 complexes:

$$-\log K_{\mathrm{d}} = \alpha \cdot BSA_{\mathrm{apol}} + \beta \cdot BSA_{\mathrm{pol}} + \gamma \qquad (1)$$

Table 2. Benchmark of structure-based binding affinity data of dimeric complexes with calculated conformational change from their known unbound structures

Binary Complex	PDB1	PDB_ID[a]	PDB2	PDB_ID[b]	BSA (Å^2)[c]	i–rmsd (Å)[d]	K_d (M)[e]
1CLV_A:I	α–amylase	1JAE_A	α–amylase inhibitor	1HTX_A	2085	0.896	1.00E–09
1LW6_E:I	subtilisin BPN'	1SUP_A	subtilisin–chymotrypsin inhibitor–2A	1YPC_I	1706	0.430	2.00E–12
1P6A_A:B	fiber protein KNOB domain	1NOB_A	coxsackie/adenovirus receptor	1EAJ_A	1547	0.851	3.50E–08
1PK1_A:B	Ph SAM domain	1KW4_A	Scm SAM domain	1PK3_A	933	0.379	5.40E–08
2XQR_A:B	β–fructofuranosidase	2AC1_A	cell–wall inhibitor of β–fructosidase	2CJ4_A	1988	0.459	3.10E–08
3D7T_A:B	Tyrosine–protein kinase CSK	1BYG_A	proto–oncogene tyrosine–protein kinase Src	1YOJ_A	1179	0.806	7.00E–05
3G3A_A:B	variable lymphocyte receptor VLRB.2D	3G39_A	lysozyme C	2VB1_A	1657	0.845	4.30E–07
3LB8_A:C	putidaredoxin reductase	1Q1R_A	putidaredoxin	1XLQ_A	1587	0.902	6.60E–05
3M18_A:B	variable lymphocyte receptor A diversity region	3M19_A	lysozyme C	2VB1_A	1760	0.467	1.80E–10
3OXU_B:F	complement c3d	1C3D_A	HF protein	3R62_A	1498	0.839	3.40E–05
3QQ8_A:B	transitional endoplasmic reticulum ATPase	3QQ7_A	FAS–associated factor 1	3QX1_A	1653	0.581	1.50E–06
4ETW_A:B	Pimelyl–[acyl–carrier protein] methyl ester esterase	1M33_A	acyl carrier protein	1T8K_A	1131	0.889	3.10E–06
4HCP_A:B	ATP/GTP binding protein	3GQM_A	NEDD8	1NDD_A	2426	0.957	9.40E–06
1F3V_A:B	TRADD (N–ter domain)	1F2H_A	TRAF domain	1CA4_A	1484	2.020	7.80E–06
1K93_A:D	calmodulin–sensitive adenylate cyclase	1K8T_A	calmodulin	3IF7_A	5468	13.056	2.00E–08
1L0Y_A:B	TCR Vβ8.2	1BEC_A	exotoxin type A	1FNU_A	1133	1.236	6.00E–06
1SQ0_A:B	von Willebrand factor	1IJB_A	Platelet glycoprotein Ib α–chain (L–domain)	1P9A_G	2108	2.231	3.00E–08
1UAD_A:C	Ras–related protein Ral–A	1U8Z_A	exocyst complex component Sec5 (N–ter)	1HK6_A	1014	1.322	1.37E–07
1XT9_A:B	sentrin–specific protease 8	2BKQ_A	Neddylin	1NDD_A	3017	2.473	2.00E–07
2AQ1_A:B	TCR Vβ H72Q	2APB_A	Enterotoxin type C–3	1UNS_A	1126	1.451	5.50E–09
2FU5_A:D	Guanine nucleotide exchange factor MSS4	1FWQ_A	Ras–related protein Rab–8A	4LHW_A	2196	3.364	7.00E–10
2G45_A:B	Ubiquitin carboxyl–terminal hydrolase 5	2G43_A	Ubiquitin	1YJ1_A	1017	7.259	2.82E–06
2J7P_A:D	signal recognition particle protein	1LS1_A	FTSY cell division protein	2IYL_D	3008	2.640	1.00E–08
2JGZ_A:B	phospho–CDK2	4FKL_A	G2/mitotic–specific cyclin–B1	2B9R_A	2977	5.350	1.00E–03
2JJS_A:C	tyr–protein phosphatase substrate 1 (N–ter)	2UV3_A	leucocyte surface antigen CD47	2VSC_A	1833	6.828	1.20E–06
2OT3_A:B	Rab5 GDP/GTP exchange factor	1TXU_A	Ras–related protein Rab–21	1Z08_A	2306	2.594	1.80E–06
2PTT_A:B	CD48 antigen	2PTV_A	Natural killer cell receptor 2B4	2PTU_A	1455	1.051	4.00E–06
2V8S_E:V	clathrin interactor 1 (ENTH)	2QY7_A	HABC domain of VTI–1B	2QYW_A	1333	1.718	2.20E–05
2VSM_A:B	hemagglutinin–neuraminidase	2VWD_A	ephrin–B2	2I85_A	2787	2.770	3.50E–08
2W2X_A:D	ras–related C3 botulinum toxin substrate 2	2W2T_A	1–PtdIns–4,5–bisphosphate phosphodiesterase γ–2	2W2W_A	949	2.980	2.19E–05
2WEL_A:D	CA/calmodulin–dependent protein kinase type II δ–chain	2VN9_A	calmodulin	3IF7_A	3035	23.452	6.00E–07
2WG3_A:C	Desert Hedgehog protein (N–ter)	2WFQ_A	Hedgehog–interacting protein (C–ter)	2WFT_A	1871	1.042	7.36E–08
2XB6_A:C	Neuroligin–4, x–linked	3BE8_A	Neurexin–1–β	3BOD_A	1193	1.372	1.15E–07
2XBB_A:C	E3 ubiquitin–protein ligase NEDD4	2XBF_A	Ubiquitin	1YJ1_A	1799	2.933	1.10E–05

(*continued on next page*)

Table 2 (*continued*)

Binary Complex	PDB1	PDB_ID[a]	PDB2	PDB_ID[b]	BSA (Å²)[c]	i–rmsd (Å)[d]	K_d (M)[e]
2XGY_A:B	Relik capsid N–ter domain	2XGU_A	Peptidyl–prolyl cis–trans isomerase A	2X25_B	1534	2.293	3.00E–05
2Z8V_A:D	Apical membrane antigen 1	1Z40_A	New antigen receptor variable domain	1VES_A	2086	1.303	4.80E–09
3BEG_A:B	Serine/threonine–protein kinase SRPK1	1WAK_A	Splicing factor, arginine/serine–rich 1	2O3D_A	1938	4.703	5.00E–08
3C9A_B:D	Protein giant–lens	3CGU_A	Protein spitz	3CA7_A	2720	5.660	7.70E–09
3FPU_A:B	Evasin–1	3FPR_A	C–C motif chemokine 3	2X6G_A	2689	2.310	1.20E–10
3GC3_A:B	β–arrestin–1	1G4M_A	Clathrin heavy chain 1	2XZG_A	2161	2.916	2.10E–06
3GQI_A:B	Basic fibroblast growth factor receptor 1	3RHX_A	Phospholipase C–γ–1	4FBN_A	1867	3.494	3.30E–08
3H2U_A:B	Vinculin (C–ter)	1QKR_A	Raver–1 (RRM 1–3 domains)	3SMZ_A	1360	3.213	1.26E–05
3JZA_A:B	Ras–related protein Rab–1B	3NKV_A	Uncharacterized protein DrrA	3JZ9_A	3384	4.659	3.00E–12
3K75_B:D	DNA repair protein XRCC1	1XNA_A	DNA polymerase β	2VAN_A	1195	1.314	1.10E–07
3KUD_A:B	Ras–GDP	2CE2_X	RAF proto–oncogeneser/thr–protein kinase (A85K)	1RFA_A	1046	1.706	1.70E–06
3KW5_A:B	Ubiquitin carboxyl–terminal hydrolase isozyme L1	2ETL_A	Ubiquitin	1YJ1_A	2350	1.502	3.85E–07
3MC0_A:B	variable beta 8.2 mouse T cell receptor	2APB_A	Enterotoxin SEG	1XXG_A	1205	1.024	1.25E–07
3MJ7_A:B	Junctional adhesion molecule–like	3MJ6_A	Coxsackievirus and adenovirus receptor homolog	3JZ7_A	1728	2.220	5.00E–06
3ONA_A:B	Tumour necrosis factor receptor, SECRET domain	3ON9_A	CX3CL1 protein, chemokine domain	1F2L_A	1111	1.158	6.80E–07
3VYR_A:B	Hydrogenase expression/formation protein HypC	2Z1C_A	Hydrogenase expression/formation protein HypD	2Z1D_A	2184	3.954	1.40E–07
4DGE_A:C	TRIMCyp (cyclophilin domain)	2X25_B	capsid protein (cyclophilin–binding domain)	2PWO_A	1037	1.606	3.85E–05

[a] Crystal structures and corresponding chains of the unbound partner 1.
[b] Crystal structures and corresponding chains of the unbound partner 2.
[c] Calculated using NACCESS (www.bioinf.manchester.ac.uk/naccess/) using standard van der Waals radii and probe radius 1.4 Å.
[d] i-RMSD (measured in Å), concerning Ca atoms. Interface residues were assigned using an interatomic contact distance cutoff of 10 Å.
[e] Equilibrium constants are collected directly from PDBbind v2013 (www.pdbbind-cn.org). For a full list of references, see Table S12.

The coefficients obtained after 4-fold cross-validation are α = 0.0040, β = 0.0007 and γ = 3.7342. BSA_{apol} and BSA_{pol} denote the apolar and polar BSAs (Å^2), respectively.

The statistics of the individual correlations of the contributors used to build each model with binding affinity are shown in Table 1a. Overall statistics for both models developed are shown in Table 1b. The classical interface model accounts modestly well for the affinity of the complexes of the benchmark when considering only "rigid" protein–protein interactions, with a correlation coefficient of $r = 0.49$ ($p < 0.0001$) (4-fold cross-validation, $N = 72$). However, it fails to describe the affinity of "flexible" protein–protein interactions (those showing conformational changes >1.0 Å) ($r = 0.16$, $p = 0.1970$).

All the data presented in this work collectively argue that the NIS properties have a striking effect on binding affinity. We therefore built a simple "global surface" model that encompasses both the interface and the NIS. It combines the $^{\%}AA_{pol}^{NIS}$ and $^{\%}AA_{char}^{NIS}$ on the NIS with the number of atoms in the interface ($N_{atomsINT}$) and can explain reasonably well the experimental binding affinity data in the dataset for all complexes ($r = 0.50$, $N = 143$) (Table 1b):

$$-\log K_d = \alpha \cdot {}^{\%}AA_{pol}{}^{NIS} + \beta \cdot {}^{\%}AA_{char}{}^{NIS} + \gamma \cdot N_{AtomsINT} + \delta \quad (2)$$

$N_{atomsINT}$ denotes the number of atoms in the interface of the complex and $^{\%}AA_{char}^{NIS}$ and $^{\%}AA_{pol}^{NIS}$ denote the percentages of charged and polar residues on the NIS, respectively, and α = 0.0857, β = −0.0685, γ = 0.0262 and δ = 3.0125 (obtained after 4-fold cross-validation based on the rigid complexes only).

Our global surface model is successful for the "rigid" protein–protein interactions and the corresponding predictions for the "flexible" binders are significantly improved: predictions for flexible complexes become statistically significant ($r = 0.32$, $N =$

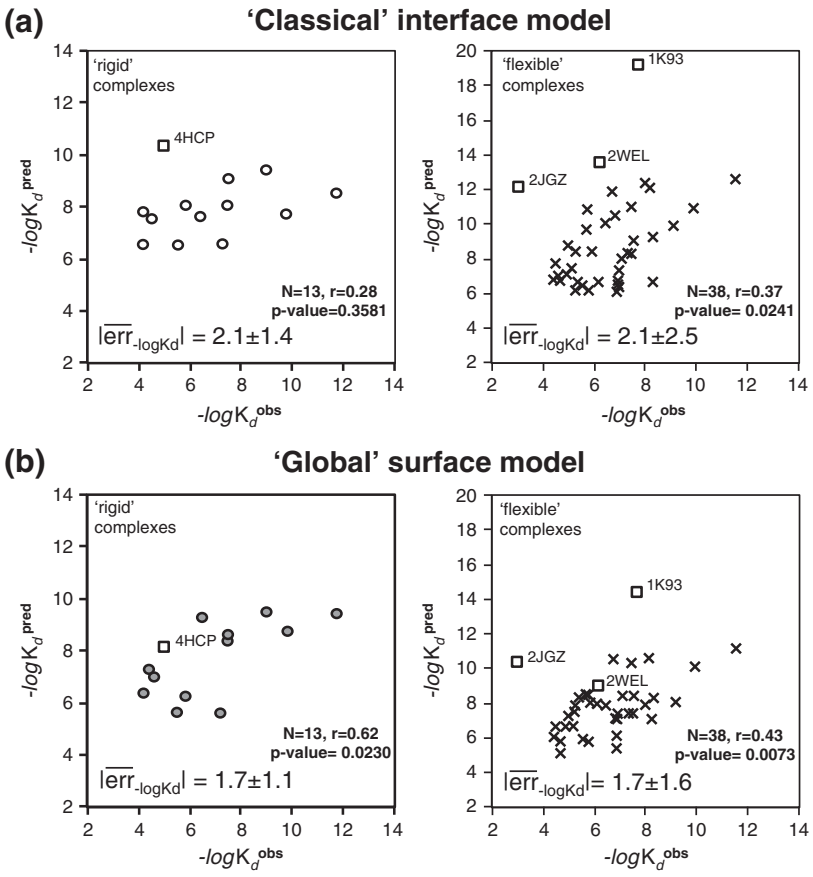

Fig. 7. Independent test of the (a) classical interface model [13] [Eq. (1)] and (b) global surface model introduced in this work [Eq. (2)] for the prediction of protein–protein binding affinities against a compiled test set of protein complexes with known experimental affinities and conformational change. Rigid are shown in the left panels whereas flexible complexes are shown in the right panels; complexes discussed in-text are shown as squares within the correlation plots along with their respective PDB IDs.

71, $p = 0.0065$) when compared to those calculated with the interface model, the latter not being statistically significant ($r = 0.16$, $N = 71$, $p = 0.1970$) (Fig. 6a and b). Note that we chose to parameterize the models based on the rigid complexes only because flexible complexes add an additional level of complexity in affinity prediction, that of energetics stemming from conformational change, mostly linked with entropy changes.

External test of the global surface model: improvement of binding affinity prediction by incorporating NIS properties

Making predictions on the same set that was used for feature selection could lead to possible overfitting despite the fact that our model only has four variables and that 4-fold cross-validation was performed. To test for possible overfitting, we have compiled an external dataset of protein–protein complexes, for which binding affinity data are available in PDBbind [53]. We have thoroughly searched for unbound structures for all dimeric complexes provided in PDBbind to assess the effect of conformational changes on binding affinity prediction (these have already been shown to be critical for accurate affinity estimation). Our external dataset (Table 2) comprises 51 protein–protein complexes with known crystal structures (resolution < 3.0 Å), significantly different in nature from the ones used for training and cross-validation of the models [4]. Thirteen complexes bind in a near-native manner (i-RMSD ≤ 1.0 Å), whereas 38 complexes undergo substantial conformational changes. Extreme examples of conformational change are those of calmodulin in complex with (a) calmodulin-sensitive adenylate cyclase [54] (i-RMSD = 13 Å) and (b) CA/calmodulin-dependent protein kinase type II delta chain [55] (i-RMSD = 23 Å). The prediction results on this independent test set for both the classical interface model and the global surface model are shown in Fig. 7a and b.

Prediction of affinity for "near-rigid binders"

The classical interface model fails to relate to the affinity of near-rigid binders if all complexes are

considered (r = 0.28, N = 13, p = 0.3581), but affinity is overestimated for a complex with large BSA (~2426 Å^2); such interface size is unusual for a "near-rigid" binder (ATP/GTP binding protein in complex with NEDD8 [56], indicated as a square in Fig. 7a and b, "rigid complexes" plots, PDB ID: 4HCP). If this protein–protein complex is excluded from the calculations using the interface model, significant relation between calculated and experimental affinities emerges (r = 0.53, N = 12, p = 0.0625), albeit weaker than the global surface model predicted affinities (r = 0.65, N = 12, p = 0.0231). Overall, the mean absolute prediction error [mean absolute error (MAE)] for the classical interface model for all rigid complexes is 2.1 ± 1.4 in $-\log K_d$ units. On the contrary, a lower MAE in the rigid dataset (MAE = 1.7 ± 1.1) is observed when global surface model is used, and significant correlations with the experimental affinities are derived for all complexes (r = 0.62, N = 13, p = 0.0230); in the case of ATP/GTP binding protein in complex with NEDD8 (for which both models have the highest prediction error), prediction error is 3.1 $-\log K_d$ units with the global surface model instead of 5.3 $-\log K_d$ units with the classical interface model.

Prediction of affinity for flexible complexes

Flexible complexes undergo substantial conformational changes upon binding (i-RMSD > 1.0 Å). Interestingly, the classical interface model still significantly correlates with experimentally measured binding affinities (r = 0.37, N = 38, p = 0.0241; Fig. 7a), which is higher than what was obtained on the training set (r = 0.16, N = 71, p = 0.1970). The MAE is, however, pretty large, reaching 2.7 ± 2.5 in $-\log K_d$ units and predictions for the calmodulin complexes (indicated as squares in the plots) are way off, but this is reasonable: both complexes (PDB IDs: 1K93 and 2WEL) undergo conformational changes reaching 13 Å and 23 Å and bury very large surface areas (5468 Å^2 and 3035 Å^2, respectively). Another interesting complex with an overestimated predicted affinity is for a pCDK2/cyclin, a complex with unusually low affinity (K_d > 1E-03 M measured using surface plasmon resonance [57]), at the boundary of a detectable interaction. The global surface model decreased the MAE (1.7 ± 1.6 $-\log K_d$ units) for these flexible complexes and also yields a better correlation with experimental K_d values with a Pearson's correlation coefficient r = 0.43 (N = 38, p = 0.0073) compared to the standard model.

This independent validation confirms our previously described results: the interface size is a significant contributor to binding affinity, but NIS properties, when properly accounted for, can lead to increase in prediction performance.

Discussion

We have identified and quantified a fundamental principle that, next to interface properties, contributes to the binding affinity of protein–protein interactions, namely the effect of the NIS. Similar effects have been shown to alter through long-range communication the catalytic activity of enzymes [27]. Here they are demonstrated for the first time on experimentally measured dissociation constants of nearly 200 transient protein assemblies, all with known conformational change, binding affinity and high-resolution molecular structures of both their unbound and bound states.

The various biophysical descriptors related to interface properties account but for a fraction of the binding affinity of a complex. Conformational changes are one of the limiting factors for accurate prediction. Nevertheless, even for complexes whose association induces very few or no conformational changes at all, binding affinity calculations using such interface-only descriptors are qualitative, corroborating previous findings [4,19–22]. Even sophisticated energetic calculations using HADDOCK score [58] and applied on the binding affinity benchmark reinforce the abovementioned view: van der Waals energy (a single descriptor) significantly correlates with affinities for rigid binders (N = 73, r = 0.60, p < 0.0001), but then again, the correlation is very low for complexes with conformational changes (N = 71, r = 0.22, p = 0.0665). Similar conclusions hold for the compiled test set in this work (Fig. S4a–d).

We propose that one of the reasons for such a limited performance, especially for flexible complexes, is that NIS effects on binding affinity have so far been neglected. We have now included these in a simple prediction model, expanding the classical interface model (Fig. 8a) into "global surface model" that accounts for both interface and non-interface surface parameters (Fig. 8b). Although it can relate predicted affinities to experimental ones for the largest dataset of protein–protein interactions with known conformational change and binding affinity assembled to date, the correlation coefficients calculated are not particularly impressive, albeit significant. Still, these correlations are the highest observed among all current biophysical models in binding affinity prediction [11,12].

The correlations that we observe and have included in our global surface model are those of the relative abundance of charged and polar residues on the NIS with the binding affinity and the dissociation constant k_{off}. Note that the observed long-range electrostatic effect is a function of the nature of the NISs, the charge model and the dielectric constant used. Charged residues on the NIS affect the electrostatics of the interaction at distances of up to 40 Å away from the center of the

interface (assuming a 100 mM salt concentration), acting on k_{off}, whereas polar residues influence the formation of a stable hydration shell that can shield the protein from unintended interactions. We have also shown that salt concentration, as expected, can influence the electrostatic energy and therefore the energy of the interaction; however, this effect might not be as strong as predicted if the interaction is through the low dielectric interior of the proteins. Modeling ionic content of the solution in which each experiment was performed is difficult but may lead to a decrease in prediction error by more than 2-fold [4].

Interestingly, polar and charged residues are well-known frustration-prone amino acids [32], especially when located on the protein surface [32]. Indeed, deviations from the "principle of minimal frustration" are functionally relevant [59,60] and have already been shown to actually relate to protein binding [61], dynamics [62], allostery [29] and communication pathways by triggering frustration along the protein structure [61]. The rationalization of their effects on binding affinity presented here (long-range electrostatics and preferential solvation) is in line with frustration theory: the long-range electrostatic contribution of charged residues could be propagated within the structure to the interface via a front of highly frustrated residues [61,63].

The observation that properties of the NIS are conserved among orthologues is in line with the hypotheses set by the Drummond and Kortemme groups [64,65] that NIS regions must experience selection pressure to avoid unintended protein–protein interactions. Our results should however be interpreted with caution as orthologous proteins might not bind the same partner, an issue that should be thoroughly investigated in future studies. Note also that, orthologues, by definition, have some sequence similarity and, as such, are expected to have conserved properties of their NIS. We have proposed here an explanation for this conservation as a regulator of protein–protein binding affinity.

Our global surface model predicts that the effect of single mutations on the NIS will be very small, especially for large surfaces, as the residue percentages will not substantially change in such cases. This is also in line with the observation made by Franzosa and Xia [66] that protein surfaces are prone to high mutation rates. A single mutation, even a non-conserved one, on the protein surface will not affect binding affinity in a significant manner. However, the effect can be detected in available experimental data. We performed for this a simple analysis of alanine scanning studies collected in ASEdb [67] for 26 complexes, having 446 single-point mutations in total (Tables S8–S10). We first classified mutations as being in the interface, interface periphery (rim) or on the NIS (details in Table S8), then calculated how far these residues are from the interface periphery. The experimental mutation data show, as expected, that the largest effects on binding affinity occur when the mutations

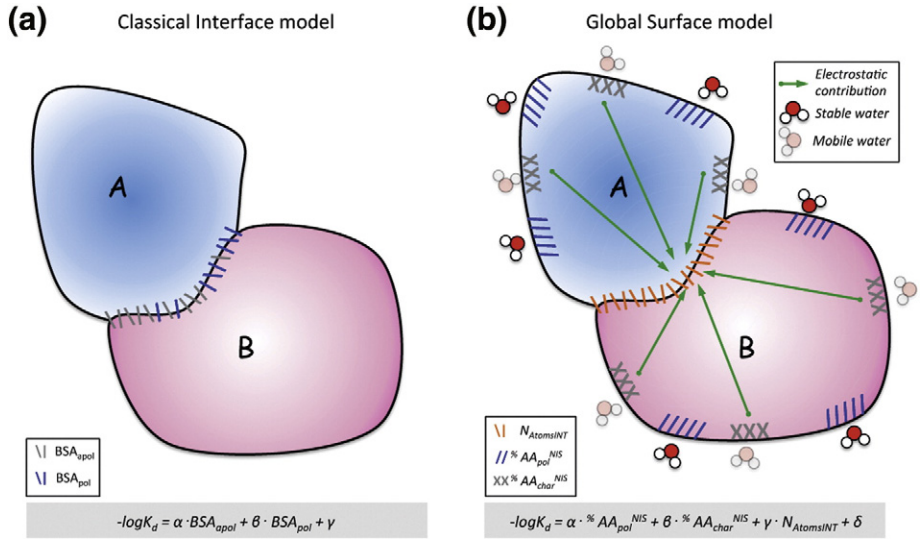

Fig. 8. Models of binding affinity contributions for protein–protein interactions: (a) classical interface model [13]. This classical model accounts for interfacial parameters, namely the apolar (BSA_{apol}) and polar (BSA_{pol}) BSA. (b) Global surface model; this new model accounts for both interfacial (the number of atoms in the interface $N_{atomsINT}$, directly related to the BSA) and non-interfacial properties (the percentages of polar, $^{\%}AA_{pol}^{NIS}$, and charged, $^{\%}AA_{char}^{NIS}$, residues on the NIS); their contribution to the binding affinity can be explained by distant electrostatics and solvent effects (see the main text).

are within the interface. This effect decays as a function of distance from the interface, but detectable changes in binding affinity are still observed far away from the interface periphery (Fig. 9).

However, although surface mutation rates are high, we can directly observe that the underlying properties are preserved. Our conservation analysis results are consistent with the sequence-specific analysis set by the Blundell group, demonstrating that distribution of amino acids in protein–protein complexes is highly environment dependent [51] and that solvent-accessible regions exhibit a certain degree of conservation [52]. Indeed, as demonstrated in this work, polar and charged distributions on the NIS of protein–protein complexes, which are directly related to affinity, remain constant within homologous protein–protein complexes (orthologues) but vary a lot for non-homologous complexes.

From our regression model, we have identified general rules for stabilizing and destabilizing mutations on the NIS. An increase of the fraction of polar residues on the NIS of the protein–protein complexes is predicted to lead to an increase in the stability of the complex through favorable surface–water interactions and therefore of the binding affinity. Although we would expect that whatever conformational change a complex may have undergone during its formation would have no effect on its post-complex stabilization, we do observe a dependence of the contribution of the NIS residues to K_d and k_{off}, but it diminishes with increasing conformational change. This would mean that conformational entropy is correlated with the overall entropy of the interaction (complexes with higher conformational entropy changes are in general more entropic). This is not a rule *per se* but has been shown by NMR for several biomolecular interactions [68–70]. Increase of the concentration of charged residues on the NIS is predicted to have destabilizing effects, but this effect is complex in practice as it may depend both on the residue mutated and on the distance from the interface. Although counterions on protein surfaces should also have an effect on protein–protein interactions, their role is far from being understood [71]. Still, electrostatic pairing on the surface between negatively and positively charged residues should be promoted, even when mutating residues at distances of up to 40 Å from the interface center. Electrostatic pairing on the rim region has already been successfully applied to enhance association rates primarily by the Schreiber group [5], but this

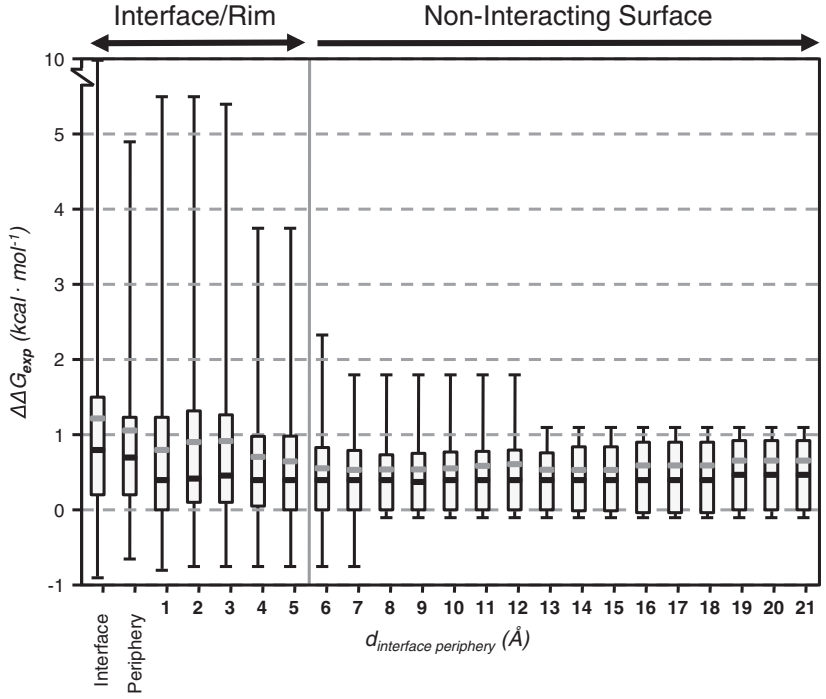

Fig. 9. Analysis of 446 experimental alanine scanning mutations from ASEdb [67] as a function of distance from the interface and its periphery for 26 complexes. Most of the effects cluster around the interface or its rim (up to 5 Å). Significant effects are however still observable at large distances (up to 21 Å away from the interfacial periphery). The $\Delta\Delta G$ values are illustrated using a box-and-whisker diagram. The box parameters are as follows: the box range goes from the first to the third quartile of the distribution; box whiskers identify the minimum and maximum values; the gray and black thick lines in the box identify the median and the average, respectively.

may have a mixed effect on dissociation as also highlighted in this work.

Our global surface model provides thus a test bed to experimentalists who can put our hypothesis to the test and work in that direction is ongoing. Finally, it is clear that more sophisticated models for binding affinity calculation will have to be developed if one wants to bridge the accuracy gap that should bring us within experimental error for any protein–protein interaction. Such models will have to consider properties of both the interface and NIS, account for the effect of conformational changes and properties of the free components as well. Reaching this goal would have a dramatic impact on current and future understanding of protein–protein interactions and open the route to the design of "materia medica" for protein–protein interactions.

Materials and Methods

Dataset

We have analyzed 144 complexes with known binding affinities [4]. For 51 of them, association rate and dissociation rate constants (k_{on} and k_{off}) were manually procured from the literature (Table S1). The structures used for the calculations of the parameters were downloaded from the Protein Data Bank† [72]. Since both bound and unbound experimental structures are available, direct observations can be made that are relevant to the conformational changes that occur upon binding.

Calculation of physicochemical and biophysical parameters

Physicochemical parameters of the complexes were calculated using the PROTORP server [73]. Shortly, PROTORP provides the following:

(a) Interfacial parameters:
   (1) Interface *biochemical parameters*, such as percentage of polar (=O, CH_2P, –OH, =CP–, O^-), non-polar (>S, $–CH_3$, $–CH_2A$, >CH–, =CHA) and neutral (=CA–, CH_2N, >NH, $–NH_2$ and NH_3) atoms (where A denotes aromatic rings), percentage of polar, non-polar and charged residues, as well as the absolute number and type of atoms and residues present in the interface.
   (2) *Shape parameters*, including planarity, eccentricity and gap volume of the interface.
   (3) Percentage and categorization of *secondary structure elements* (alpha, beta, alpha/beta, coil) in the interface.
   (4) *Structural parameters*, including number of hydrogen bonds and number of salt bridges, the BSA of the interface in $Å^2$ and the percentage of the surface area that corresponds to the interface.

(b) Non-interfacial parameters:
   (1) NIS *biochemical parameters*, such as percentage of polar, non-polar and charged residues on the surface, as well as the absolute number and type of residues on the complex's surface. $^{\%}AA^{NIS}_{char}$ and $^{\%}AA^{NIS}_{pol}$ are defined as the number of charged or polar residues, multiplied by 100, present on the NIS divided by the total number of residues on the NIS.
   (2) *Structural parameters* as in (3) and (4), but for the NIS of the complex.

The Protein Interaction Calculator Web server [74] was also used to gain insight into non-covalent interactions. We calculated the number of the following:

(a) Hydrophobic interactions (according to the Kyte–Doolittle hydrophobicity index [50]),
(b) Intermolecular disulfide bonds (at a distance cutoff of 2.2 Å),
(c) Hydrogen bonds [75] [main chain–main chain (MC-MC), main chain–side chain (MC-SC) and side chain–side chain (SC-SC)],
(d) Ionic interactions [76],
(e) Aromatic–aromatic [77] and aromatic–sulfur interactions [78] and
(f) Cation–π interactions [79].

The apolar and polar BSA of the protein–protein complexes were calculated using NACCESS [80].

Correlation studies

We performed correlation studies of 39 structural parameters with binding affinity data previously collected [4]. Correlation of all 39 parameters with the K_d of the complexes, their empirically derived ΔG values and the association and dissociation rates of a subset of the complexes (k_{on} and k_{off}) was calculated using the Pearson product-momentum correlation coefficient (r):

$$r = \frac{1}{N-1}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}\right)\left(\frac{y_i-\bar{y}}{\sigma_y}\right) \quad (1)$$

where r denotes the correlation coefficient, N is the number of complexes and σ_x and σ_y are the standard deviation of the x and y variables.

k_{on} and k_{off} were converted into their logarithmic (log) and (–log) values, respectively. The complexes were classified based on the amount of conformational change occurring upon binding as measured by the interface RMSD (i-RMSD: RMSD calculated on backbone atoms of all residues within 10 Å from the partner molecule). Complexes that undergo minor conformational changes and approximate rigid-body association (i-RMSD ≤ 1.0 Å) constitute the first class, whereas complexes that do not satisfy this criterion (i-RMSD > 1.0 Å) are discussed separately and constitute a second class. The significance of the observed correlations

was estimated using two-tailed *p*-values, both absolute and FDR corrected [81]. FDR correction did not alter significance of the variables discussed in the manuscript (FDR corrected *p*-values are included in Table S11).

Structural definition of NIS

The interface residues of a protein–protein complex were defined as those that show a change between unbound and bound forms of more than 5% of relative accessibility as defined by NACCESS [80]. All other surface residues are defined as being part of the NIS.

Evolutionary conservation of the NIS of protein–protein complexes

Structures of protein–protein complexes in which both partners correspond to the same organism were extracted from our benchmark of 144 protein–protein interactions with known binding affinities [4]. Out of these, only binary complexes were kept; that is, complexes that are formed when single-chain proteins are interacting. We identified 47 protein–protein interactions, corresponding to the biological assemblies of complexes with PDB IDs 1B6C, 1BRS, 1BUH, 1E6E, 1E96, 1EMV, 1EWY, 1F6M, 1FFW, 1FQJ, 1GLA, 1GPW, 1GRN, 1GXD, 1H9D, 1HE8, 1I2M, 1IBR, 1JIW, 1KTZ, 1M10, 1MQ8, 1NVU, 1NW9, 1OC0, 1PVH, 1PXV, 1QA9, 1R6Q, 1S1Q, 1WQ1, 1XD3, 1XQS, 1Z0K, 1ZHI, 2BTF, 2C0L, 2FJU, 2HLE, 2HQS, 2HRK, 2J0T, 2O3B, 2PCC, 2TGP, 2WPT and 3CPH.

Sequences for each protein forming each complex were derived directly from the PDB structures and used as input in HMMER‡ [82]. HMMER is a tool for sequence analysis based on profile hidden Markov models and can be used to identify protein sequence homologues in sequence databases (in this work, UniProt curated sequences were searched). Compared to BLAST [83], FASTA [84] and other sequence alignment and database search tools, HMMER aims to be more accurate and better at detecting remote homologues [48]. All homologous sequences were identified for each protein and only sequences meeting specific criteria for sequence length were kept (see Table S6). This filter was adopted to avoid large insertions/deletions that could hamper the analysis of the NISs.

The homologues for each complex were then analyzed using two different approaches: one, sequence-based, assessing the conservation of polar/charged/apolar character of equivalent residues in a multiple sequence alignment of homologous sequences, and another, structure-based, investigating structural features of the homologues after homology modeling.

For the sequence-based analysis, for each chain of each complex, a multiple sequence alignment of the chain and all its homologues was built using the Clustal Omega [85] Web server available at the EBI Web site [86]. The alignments of the chains for each complex were then concatenated so that each homologue was represented by one sequence only. The concatenated multiple sequence alignments were filtered for NIS residues and these were then quantified based on their character (polar: C,H,N,Q,S, T,Y,W; apolar: A,F,G,I,V,L,M,P; charged: E,D,K,R).

For the structure-based analysis, a pipeline was created, using modules from the Biopython project [87] to align the template sequence and the query (found) sequence of the protein–protein complex using the Needleman–Wunsch algorithm as implemented in NEEDLE [88] and finally construct homology models with MODELLER 9v9 [89]. In this procedure, 10 homology models for each homologue were assembled (>30,000 in total); the best according to the MODELLER score was selected in order to further calculate propensities of polar and charged residues on the NIS, as well as the size of the interface region.

Average and standard deviation of surface properties of the homologous complexes were analyzed as a function of their sequence identity to the original protein–protein complex sequence.

Distant effect of electrostatics in protein–protein interactions

A simple coulomb-based model [90] was used to estimate the electrostatic effect on protein–protein complexes as a function of distance from the center of the interface, considering only charged amino acids. Standard parameters were used (ε = 1.0, ionic strength = 100 mM salt, pH =7.0, T = 298 K). Note that, in our simple model, varying the parameters will only affect the magnitude of the electrostatic effect not the patterns derived for the various complexes. We have also calculated the electrostatic energy in different ionic strengths (100, 150, 200 and 250 mM salt). The dielectric constant was set to 1.0 to estimate the electrostatic contribution "through" the protein–protein complex.

In order to assess the electrostatic effect as a function of distance, we applied the following protocol to each protein–protein complex in the benchmark:

(1) Read structure of complex composed of proteins A and B.
(2) Calculate electrostatic energy (without cutoffs) using standard coulomb potential (only of titratable groups) of native complex AB (${E_{elec}}^{AB}{}_{native}$) and of A (${E_{elec}}^{A}{}_{native}$) and B (${E_{elec}}^{B}{}_{native}$) alone.
(3) Calculate electrostatics of binding using

$$E_{elec}{}^{A-B}{}_{,binding,native} = {E_{elec}}^{AB}{}_{native} - \left({E_{elec}}^{A}{}_{native} + {E_{elec}}^{B}{}_{native}\right)$$

(4) Find the geometrical center of mass of the interface.
(5) Define (4) as the center point and draw a sphere of 3 Å radius.
(6) Mutate all residues outside the sphere to alanine.
(7) Calculate energy (E_{elec}^{AB}) for ${Complex}^{AB}{}_{ala[3Å]}$, for ${Protein}^{A}{}_{ala[3Å]}$ and for ${Protein}^{B}{}_{ala[3Å]}$.
(8) Calculate electrostatics of binding using

$$E_{elec}{}^{A-B}{}_{binding,ala[3Å]} = {E_{elec}}^{AB}{}_{ala[3Å]} - \left({E_{elec}}^{A}{}_{ala[3Å]} + {E_{elec}}^{B}{}_{ala[3Å]}\right)$$

(9) Iterate from step 5 increasing the sphere radius by 3 Å, up to 90 Å.

This model calculates the difference in the coulomb energy (kcal mol^{-1}) of a protein complex from that of its unbound constituents as a function of distance from the interface, considering the electrostatic contribution of each titratable group within a defined sphere and assigning the pK_a and charge state according to the Henderson–Hasselbalch equation.

Analysis of solvent–surface interactions

We collected 186 ultra-high-resolution (≤1 Å) crystal structures of proteins (see Table S7) using the advanced search in PDB [72]. The search options were as follow:

(1) ultra-high-resolution crystal structures (resolution better than 1.0 Å),
(2) protein structures only (no limitation in the number of chains) and
(3) a non-redundancy sequence criterion (30% cutoff) in order to avoid similar structures in our analysis that would influence the subsequent data treatment.

The derived dataset was filtered for structures that miss water, have unusually very low *B*-factors or are designs. The final set is composed of 184 structures (Table S7).

The water contacts were analyzed using a 3.9-Å cutoff (water oxygen–protein heavy atom distance) (also defined in LIGPLOT [91] by default). The surfaces of the proteins were defined using standard NACCESS [80] criteria and classified into polar, charged and apolar fractions using the Eisenberg Hydrophobicity Scale [92] or/and the Kyte–Doolittle scale [50]. The average number of contacts (N_{cont}^{wat}) of water molecules with different fractions of the protein surface was analyzed as a function of the *B*-factor of the water molecules.

Building binding affinity models

Two multiple linear regression models were constructed and validated using a 4-fold cross-validation procedure:

(1) The first follows the idea of Horton and Lewis [13], where apolar and polar BSAs along with a constant term can be used to model the affinity.
(2) The second includes the number of atoms in the interface and the percentages of charged and polar residues on the NIS of the complexes along with a constant term.

The models were derived using the structure-based binding affinity benchmark [4], which includes non-redundant complexes with known unbound structures of the binders and, therefore, known conformational change. The rigid complexes from the dataset having i-RMSD ≤ 1.0 Å (72 complexes in total) were used for training and for 4-fold cross-validation. The reported coefficients were taken as the average of the 4-fold cross-validation optimization runs. The remaining complexes (i-RMSD values >1.0 Å) were blindly predicted (71 complexes in total; complex 2OZA:B_A was removed since its BSA was extraordinary large and detected as an outlier using the standard Grubbs' test).

Testing binding affinity models

Biological assemblies of protein–protein complexes present in PDBbind [53] were downloaded from the Protein Data Bank, with a resolution criterion of <3 Å. Only dimeric complexes having single-chain partners (being in the biological assembly in the form PDBID_*A*:*B*, *A* being protein1, *B* being protein2, respectively) were considered in order to avoid complicated equilibria among three or more chains. Binding affinity data were retrieved directly from PDBbind [53]. For all complexes collected, unbound structures were searched in the PDB, having 95% of sequence identity as a criterion for a successful hit. Unbound components were only used to assess conformational changes and not any binding properties. In total, 51 non-redundant protein–protein complexes with known unbound partners and experimental affinity were found to match the abovementioned conditions. Complexes are shown in Table 2 and associated references can be found in Table S12.

Acknowledgements

The authors thank Prof. Dr. Maarten R. Egmond for valuable discussions and Koen M. Visscher and Manisha Anandbahadoer for their valuable contribution to the processing of alanine scanning mutagenesis data. This work was supported by the Netherlands Organization for Scientific Research (VICI grant 700.56.442 to A.M.J.J.B.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2014.04.017.

Received 6 November 2013;
Received in revised form 11 March 2014;
Accepted 17 April 2014
Available online 25 April 2014

Keywords:
buried surface area;
hydrophobicity;
hydrophilicity;
koff;
protein–protein complexes

Present address: P. L. Kastritis, EMBL Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany.

† www.pdb.org
‡ http://hmmer.janelia.org/

Abbreviations used:
BSA, buried surface area; NIS, non-interacting surface; i-RMSD, interface-RMSD.

References

[1] Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. Transient protein–protein interactions: structural, functional, and network properties. Structure 2010;18:1233–43.

[2] Robinson CV, Sali A, Baumeister W. The molecular sociology of the cell. Nature 2007;450:973–82.

[3] Bonetta L. Protein–protein interactions: interactome under construction. Nature 2011;468:851–4.

[4] Kastritis PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AM, et al. A structure-based benchmark for protein–protein binding affinity. Protein Sci 2011;20:482–91.

[5] Schreiber G, Haran G, Zhou HX. Fundamental aspects of protein–protein association kinetics. Chem Rev 2009;109: 839–60.

[6] Scott JD, Pawson T. Cell signaling in space and time: where proteins come together and when they're apart. Science 2009;326:1220–4.

[7] Rudolph J. Inhibiting transient protein–protein interactions: lessons from the Cdc25 protein tyrosine phosphatases. Nat Rev Cancer 2007;7:202–11.

[8] Charbonnier S, Gallego O, Gavin AC. The social network of a cell: recent advances in interactome mapping. Biotechnol Annu Rev 2008;14:1–28.

[9] Morelli X, Bourgeas R, Roche P. Chemical and structural lessons from recent successes in protein–protein interaction inhibition (2P2I). Curr Opin Chem Biol 2011;15:475–81.

[10] Sievers SA, Karanicolas J, Chang HW, Zhao A, Jiang L, Zirafi O, et al. Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. Nature 2011;475:96–100.

[11] Kastritis PL, Bonvin AM. Molecular origins of binding affinity: seeking the Archimedean point. Curr Opin Struct Biol 2013;23:868–77.

[12] Kastritis PL, Bonvin AM. On the binding affinity of macromolecular interactions: daring to ask why proteins interact. J R Soc Interface 2013;10:20120835.

[13] Horton N, Lewis M. Calculation of the free energy of association for protein complexes. Protein Sci 1992;1:169–81.

[14] Murphy KP, Freire E. Thermodynamics of structural stability and cooperative folding behavior in proteins. Adv Protein Chem 1992;43:313–61.

[15] Chothia C, Janin J. Principles of protein–protein recognition. Nature 1975;256:705–8.

[16] Audie J, Scarlata S. A novel empirical free energy function that explains and predicts protein–protein binding affinities. Biophys Chem 2007;129:198–211.

[17] Moal IH, Agius R, Bates PA. Protein–protein binding affinity prediction on a diverse set of structures. Bioinformatics 2011;27:3002–9.

[18] Su Y, Zhou A, Xia X, Li W, Sun Z. Quantitative prediction of protein–protein binding affinity with a potential of mean force considering volume correction. Protein Sci 2009;18:2550–8.

[19] Moretti R, Fleishman SJ, Agius R, Torchala M, Bates PA, Kastritis PL, et al. Community-wide evaluation of methods for predicting the effect of mutations on protein–protein interactions. Proteins 2013;81:1980–7.

[20] Kastritis PL, Bonvin AM. Are scoring functions in protein–protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. J Proteome Res 2010;9:2216–25.

[21] Sacquin-Mora S, Carbone A, Lavery R. Identification of protein interaction partners and protein–protein interaction sites. J Mol Biol 2008;382:1276–89.

[22] Fleishman SJ, Whitehead TA, Strauch E-M, Corn JE, Qin S, Zhou H-X, et al. Community-wide assessment of protein-interface modeling suggests improvements to design methodology. J Mol Biol 2011;414:289–302.

[23] Melquiond ASJ, Karaca E, Kastritis PL, Bonvin AMJJ. Next challenges in protein–protein docking: from proteome to interactome and beyond. WIREs Comput Mol Sci 2011;2:642–51.

[24] DeLano WL. Unraveling hot spots in binding interfaces: progress and challenges. Curr Opin Struct Biol 2002;12:14–20.

[25] Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science 1999;286:295–9.

[26] Changeux JP. Allostery and the Monod-Wyman-Changeux model after 50 years. Annu Rev Biophys 2012;41:103–33.

[27] Benkovic SJ, Hammes-Schiffer S. A perspective on enzyme catalysis. Science 2003;301:1196–202.

[28] Zheng W, Schafer NP, Davtyan A, Papoian GA, Wolynes PG. Predictive energy landscapes for protein–protein association. Proc Natl Acad Sci U S A 2012;109:19244–9.

[29] Ferreiro DU, Hegler JA, Komives EA, Wolynes PG. On the role of frustration in the energy landscapes of allosteric proteins. Proc Natl Acad Sci U S A 2011;108:3499–503.

[30] Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 1995;21:167–95.

[31] Levy Y, Wolynes PG, Onuchic JN. Protein topology determines binding mechanism. Proc Natl Acad Sci U S A 2004;101:511–6.

[32] Ferreiro DU, Hegler JA, Komives EA, Wolynes PG. Localizing frustration in native proteins and protein assemblies. Proc Natl Acad Sci U S A 2007;104:19819–24.

[33] Hegler JA, Weinkam P, Wolynes PG. The spectrum of biomolecular states and motions. HFSP J 2008;2:307–13.

[34] Marsh JA, Teichmann SA. Parallel dynamics and evolution: protein conformational fluctuations and assembly reflect evolutionary changes in sequence and structure. BioEssays 2014;36:209–18.

[35] Liberles DA, Teichmann SA, Bahar I, Bastolla U, Bloom J, Bornberg-Bauer E, et al. The interface of protein structure, protein biophysics, and molecular evolution. Protein Sci 2012;21:769–85.

[36] Fischer S, Verma CS. Binding of buried structural water increases the flexibility of proteins. Proc Natl Acad Sci U S A 1999;96:9613–5.

[37] De Simone A, Dodson GG, Verma CS, Zagari A, Fraternali F. Prion and water: tight and dynamical hydration sites have a key role in structural stability. Proc Natl Acad Sci U S A 2005;102:7535–40.

[38] Papoian GA, Ulander J, Eastwood MP, Luthey-Schulten Z, Wolynes PG. Water in protein structure prediction. Proc Natl Acad Sci U S A 2004;101:3352–7.

[39] Kastritis PL, van Dijk AD, Bonvin AM. Explicit treatment of water molecules in data-driven protein–protein docking: the solvated HADDOCKing approach. Methods Mol Biol 2012;819:355–74.

[40] van Dijk M, Visscher KM, Kastritis PL, Bonvin AM. Solvated protein–DNA docking using HADDOCK. J Biomol NMR 2013;56:51–63.

[41] Kastritis PL, Visscher KM, van Dijk AD, Bonvin AM. Solvated protein–protein docking using Kyte-Doolittle-based water preferences. Proteins 2013;81:510–8.
[42] Lensink MF, Moal IH, Bates PA, Kastritis PL, Melquiond AS, Karaca E, et al. Blind prediction of interfacial water positions in CAPRI. Proteins 2013;82:620–32.
[43] Vijayakumar M, Wong KY, Schreiber G, Fersht AR, Szabo A, Zhou HX. Electrostatic enhancement of diffusion-controlled protein–protein association: comparison of theory and experiment on barnase and barstar. J Mol Biol 1998;278: 1015–24.
[44] Selzer T, Schreiber G. Predicting the rate enhancement of protein complex formation from the electrostatic energy of interaction. J Mol Biol 1999;287:409–19.
[45] Alsallaq R, Zhou HX. Electrostatic rate enhancement and transient complex of protein–protein association. Proteins 2008;71:320–35.
[46] Qin S, Pang X, Zhou HX. Automated prediction of protein association rate constants. Structure 2011;19:1744–51.
[47] Shaul Y, Schreiber G. Exploring the charge space of protein–protein association: a proteomic study. Proteins 2005;60: 341–52.
[48] Eddy SR. A probabilistic model of local sequence alignment that simplifies statistical significance estimation. PLoS Comput Biol 2008;4:e1000069.
[49] Saenger W. Structure and dynamics of water surrounding biomolecules. Annu Rev Biophys Biophys Chem 1987;16: 93–114.
[50] Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157: 105–32.
[51] Bickerton GR, Higueruelo AP, Blundell TL. Comprehensive, atomic-level characterization of structurally characterized protein–protein interactions: the PICCOLO database. BMC Bioinform 2011;12:313.
[52] Gong S, Worth CL, Bickerton GR, Lee S, Tanramluk D, Blundell TL. Structural and functional restraints in the evolution of protein families and superfamilies. Biochem Soc Trans 2009;37:727–33.
[53] Wang R, Fang X, Lu Y, Yang CY, Wang S. The PDBbind database: methodologies and updates. J Med Chem 2005;48:4111–9.
[54] Drum CL, Yan SZ, Bard J, Shen YQ, Lu D, Soelaiman S, et al. Structural basis for the activation of anthrax adenylyl cyclase exotoxin by calmodulin. Nature 2002;415:396–402.
[55] Rellos P, Pike AC, Niesen FH, Salah E, Lee WH, von Delft F, et al. Structure of the CaMKIIdelta/calmodulin complex reveals the molecular mechanism of CaMKII kinase activation. PLoS Biol 2010;8:e1000426.
[56] Yao Q, Cui J, Wang J, Li T, Wan X, Luo T, et al. Structural mechanism of ubiquitin and NEDD8 deamidation catalyzed by bacterial effectors that induce macrophage-specific apoptosis. Proc Natl Acad Sci U S A 2012;109:20395–400.
[57] Brown NR, Lowe ED, Petri E, Skamnaki V, Antrobus R, Johnson LN. Cyclin B and cyclin A confer different substrate recognition properties on CDK2. Cell Cycle 2007;6:1350–9.
[58] Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc 2003;125:1731–7.
[59] Frauenfelder H, Sligar SG, Wolynes PG. The energy landscapes and motions of proteins. Science 1991;254:1598–603.
[60] Plotkin SS, Wolynes PG. Buffed energy landscapes: another solution to the kinetic paradoxes of protein folding. Proc Natl Acad Sci U S A 2003;100:4417–22.
[61] Zhuravlev PI, Papoian GA. Protein functional landscapes, dynamics, allostery: a tortuous path towards a universal theoretical framework. Q Rev Biophys 2010;43:295–332.
[62] Fuglestad B, Gasper PM, McCammon JA, Markwick PR, Komives EA. Correlated motions and residual frustration in thrombin. J Phys Chem B 2013;117:12857–63.
[63] Datta D, Scheer JM, Romanowski MJ, Wells JA. An allosteric circuit in caspase-1. J Mol Biol 2008;381:1157–67.
[64] Wilke CO, Drummond DA. Signatures of protein biophysics in coding sequence evolution. Curr Opin Struct Biol 2010;20: 385–9.
[65] Eames M, Kortemme T. Structural mapping of protein interactions reveals differences in evolutionary pressures correlated to mRNA level and protein abundance. Structure 2007;15:1442–51.
[66] Franzosa EA, Xia Y. Structural determinants of protein evolution are context-sensitive at the residue level. Mol Biol Evol 2009;26:2387–95.
[67] Thorn KS, Bogan AA. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. Bioinformatics 2001;17:284–5.
[68] Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ. The role of conformational entropy in molecular recognition by calmodulin. Nat Chem Biol 2010;6:352–8.
[69] Tzeng SR, Kalodimos CG. Protein activity regulation by conformational entropy. Nature 2012;488:236–40.
[70] Akke M. Conformational dynamics and thermodynamics of protein–ligand binding studied by NMR relaxation. Biochem Soc Trans 2012;40:419–23.
[71] Collins KD. Why continuum electrostatics theories cannot explain biological structure, polyelectrolytes or ionic strength effects in ion–protein interactions. Biophys Chem 2012;167: 43–59.
[72] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res 2000;28:235–42.
[73] Reynolds C, Damerell D, Jones S. ProtorP: a protein–protein interaction analysis server. Bioinformatics 2009;25:413–4.
[74] Tina KG, Bhadra R, Srinivasan N. PIC: Protein Interactions Calculator. Nucleic Acids Res 2007;35:W473–6.
[75] Overington J, Johnson MS, Sali A, Blundell TL. Tertiary structural constraints on protein evolutionary diversity: templates, key residues and structure prediction. Proc Biol Sci 1990;241:132–45.
[76] Barlow DJ, Thornton JM. Ion-pairs in proteins. J Mol Biol 1983;168:867–85.
[77] Burley SK, Petsko GA. Aromatic–aromatic interaction: a mechanism of protein structure stabilization. Science 1985;229:23–8.
[78] Reid KSC, Lindley PF, Thornton JM. Sulphur–aromatic interactions in proteins. FEBS Lett 1985;190:209–13.
[79] Sathyapriya R, Vishveshwara S. Interaction of DNA with clusters of amino acids in proteins. Nucleic Acids Res 2004; 32:4109–18.
[80] Jones S, Thornton JM. "NACCESS" Computer Program. London, UK: Department of Biochemistry and Molecular Biology, University College London; 1993.
[81] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc B 1995;57:289–300.
[82] Finn RD, Clements J, Eddy SR. HMMER Web server: interactive sequence similarity searching. Nucleic Acids Res 2011;39:W29–37.

[83] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10.
[84] Lipman DJ, Pearson WR. Rapid and sensitive protein similarity searches. Science 1985;227:1435–41.
[85] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 2011;7:539.
[86] McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis tool Web services from the EMBL-EBI. Nucleic Acids Res 2013;41:W597–600.
[87] Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 2009;25:1422–3.
[88] Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 1970;48:443–53.
[89] Eswar N, Eramian D, Webb B, Shen MY, Sali A. Protein structure modeling with MODELLER. Methods Mol Biol 2008;426:145–59.
[90] Fitzkee NC, Garcia-Moreno EB. Electrostatic effects in unfolded staphylococcal nuclease. Protein Sci 2008;17:216–27.
[91] Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. Protein Eng 1995;8:127–34.
[92] Eisenberg D, Weiss RM, Terwilliger TC. The hydrophobic moment detects periodicity in protein hydrophobicity. Proc Natl Acad Sci U S A 1984;81:140–4.