PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Non-interacting surface solvation and dynamics in protein–protein interactions

Koen M. Visscher, Panagiotis L. Kastritis, and Alexandre M. J. J. Bonvin*

Bijvoet Center for Biomolecular Research, Faculty of Science–Chemistry, Utrecht University, 3584CH Utrecht, The Netherlands

ABSTRACT

Protein–protein interactions control a plethora of cellular processes, including cell proliferation, differentiation, apoptosis, and signal transduction. Understanding how and why proteins interact will inevitably lead to novel structure-based drug design methods, as well as design of *de novo* binders with preferred interaction properties. At a structural and molecular level, interface and rim regions are not enough to fully account for the energetics of protein–protein binding, even for simple lock-and-key rigid binders. As we have recently shown, properties of the global surface might also play a role in protein–protein interactions. Here, we report on molecular dynamics simulations performed to understand solvent effects on protein–protein surfaces. We compare properties of the interface, rim, and non-interacting surface regions for five different complexes and their free components. Interface and rim residues become, as expected, less mobile upon complexation. However, non-interacting surface appears more flexible in the complex. Fluctuations of polar residues are always lower compared with charged ones, independent of the protein state. Further, stable water molecules are often observed around polar residues, in contrast to charged ones. Our analysis reveals that (a) upon complexation, the non-interacting surface can have a direct entropic compensation for the lower interface and rim entropy and (b) the mobility of the first hydration layer, which is linked to the stability of the protein–protein complex, is influenced by the local chemical properties of the surface. These findings corroborate previous hypotheses on the role of the hydration layer in shielding protein–protein complexes from unintended protein–protein interactions.

Proteins 2015; 83:445–458.

Key words: molecular dynamics; hydration; surface properties; binding affinity; residence time; complexes.

INTRODUCTION

With the rapid advances in structural techniques, like nuclear magnetic resonance (NMR)[1,2] and X-ray diffraction,[3] a growing number of structures of protein–protein complexes are being deposited[4] into the Protein Data Bank (www.pdb.org).[5] Large-scale annotation of biomolecular complexes is coming within reach, resulting in a deeper understanding of the nature of biomolecular interactions. Despite current progress in dissecting the structural dimension of protein–protein interactions, it is not only important to understand how protein–protein interactions can occur, but also if an interaction will take place. The latter is defined by the binding affinity of the interaction. In physicochemical terms, the binding affinity is described by the equilibrium dissociation constant, K_d. For a simple binary interaction, K_d measures the concentration of Protein A at which half of the binding sites of Protein B are occupied. Mutagenesis studies have shown that interfacial and rim mutations significantly impact binding affinity,[6,7] the interface being defined as the main binding surface of a protein–protein complex and the rim being formed by residues adjacent to interfacial residues.[8] Although there have been multiple attempts to describe binding affinity with a variety of models,[9,10] no theoretical model so far has been able to accurately reproduce dissociation constants.[11] Most classical models aim to describe binding affinity using simple descriptors that are derived exclusively from interface properties.[7,11,12] Simple mathematical models account for interfacial contributions using one or few descriptors,[13] like the buried surface area or associated

Additional Supporting Information may be found in the online version of this article.

Koen M. Visscher and Panagiotis L. Kastritis contributed equally to this work.

Panagiotis L. Kastritis's current address is EMBL Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany

*Correspondence to: Alexandre M.J.J. Bonvin, Bijvoet Center for Biomolecular Research, Faculty of Science–Chemistry, Utrecht University, Padualaan 8, 3584CH Utrecht, The Netherlands. E-mail: a.m.j.j.bonvin@uu.nl

Received 18 July 2014; Revised 10 November 2014; Accepted 26 November 2014
Published online 18 December 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24741

structural/chemical properties. These have been able to correlate at most modestly their predicted affinities with experimental dissociation constants but are not yet ready for accurately predicting the binding affinity of protein–protein complexes.[7,11,12,14] Recently, we have hypothesized that the non-interacting protein surface (NIS) has a modulating role in binding affinity,[7,15] where not only interfacial interaction descriptors play a major part in binding equilibrium, but also non-interacting surface residues have an intrinsic effect on the binding strength.

A reasonable explanation for the role of the non-interfacial surface contributions could be a modulating effect of the hydration layer around the protein complex. It is known from NMR experiments[16–18] and high-resolution X-ray diffraction data[19] that distinct solvent molecules are stabilized at the protein surface. These stabilized water molecules can form a hydration layer around the protein surface,[20,21] visible in crystallographic density,[22] and showing longer residence times than the bulk solvent as demonstrated by NMR.[16,17,23] Stabilized water molecules can have residence times up to hundreds of picoseconds or even nanoseconds before exchanging with the bulk solvent.[16–18,24–29]

Water is the active matrix for biomolecular folding, binding, and recognition events.[30] Because of water-mediated contacts and water-mediated electrostatic interactions, tertiary and quaternary protein structure can be defined in a precise manner.[31–34] Inclusion of water molecules during docking exercises of biomolecules can increase the success rate of correctly predicted interactions.[35,36] Despite the appreciation of the fundamental role of the solvent during the last 20 years,[37] its contribution to the thermodynamics of interactions is not particularly well understood. The classical model to describe the contribution of water to binding is described by the hydrophobic effect. This effect is stemming from the severe loss of translational and rotational entropy of the solvent near apolar surfaces due to the reorientation of the dipole moment to minimize unfavorable interactions. Upon binding, these water molecules are released into the bulk solvent with which they interact more favorably. Water molecules also preferably interact with each other as compared with their interaction with hydrophobic surfaces; therefore, the hydrophobic effect is also of enthalpic origin. The resulting enthalpy and entropy gains, however, describe only a fraction of the total thermodynamic contribution of the entire hydration layer.

In this work, we probe by molecular dynamics (MD) simulations the relationship between the mobility of the hydration layer and the local chemical properties of the surface of the free proteins and their products, the corresponding protein–protein complexes. We do not consider hydrophobic patches, but focus on understanding the effects of polar and charged (highly polarized) surface patches that interact favorably with the surrounding solvent layer, potentially stabilizing higher-order water shells. We have recently shown that their percentages on the non-interacting surface show significant correlations with binding affinity.[15] Our detailed analysis provides a first milestone toward a better understanding of the role of the hydration layer in biomolecular interactions and of its associated changes upon complex formation.

MATERIALS AND METHODS

Systems studied

A total of five protein complexes, including three protein–inhibitor complexes (trypsin–inhibitor,[38] E7–Im7,[39] and barnase–barstar[40]) and two enzyme–substrate complexes (ubiquitin ligase–ubiquitin[41] and adenovirus FK protein–receptor[42]) were simulated. They cover very high (6E-05 M) to very low (2E-13 M) dissociation constants[12] (Fig. 1).

The binding mechanism of the chosen complexes follows a simple lock-and-key recognition mechanism, meaning that there are almost no significant conformational changes taking place upon binding [<1 Å root mean square deviation (RMSD)] and that the unbound proteins have shape complementarity. In this way, the analysis is not hampered by additional energetic factors stemming from conformational changes upon binding. In order to assess the surface properties and possible changes between the free proteins and their complexes, both the protein–protein complexes and their free constituents were simulated, leading to a total of 15 MD simulations, amounting to ~0.9 μs total simulation time. The simulations were run on a local AMD Opteron Processor 6172 cluster, using the multithreading capability of GROMACS on eight threads, resulting on a production rate of ~8 and ~3 ns per day on average for the free proteins and their complex, respectively.

Molecular dynamics simulations

MD simulations were performed with the $AMBER_{99}$-ILDN corrected force field[43] using GROMACS 4.5.4.[44,45] Because water–protein interactions are central in this study, the TIP3P water model[46] was chosen. TIP3P is commonly accepted as the best-benchmarked water model for the $AMBER_{99}$-ILDN force field.[43] All the simulations were performed under constant pressure conditions at 0.1 MPa using the Berendsen pressure coupling[47] and at constant temperature using the v-rescale thermostat.[48] Temperature and salt concentration (NaCl) were set at 297 K and 150 m*M*, respectively, consistent with the experimental conditions used for determining the binding affinity of the complexes under study.[12] Bond lengths were constrained with the linear constraint solver algorithm,[49] and integration step of 2 fs was used. Electrostatic interactions were calculated using Particle Mesh Ewald[50] with a cubic interpolation mode and a Fourier grid spacing of 1.6 Å. Non-bonded interactions

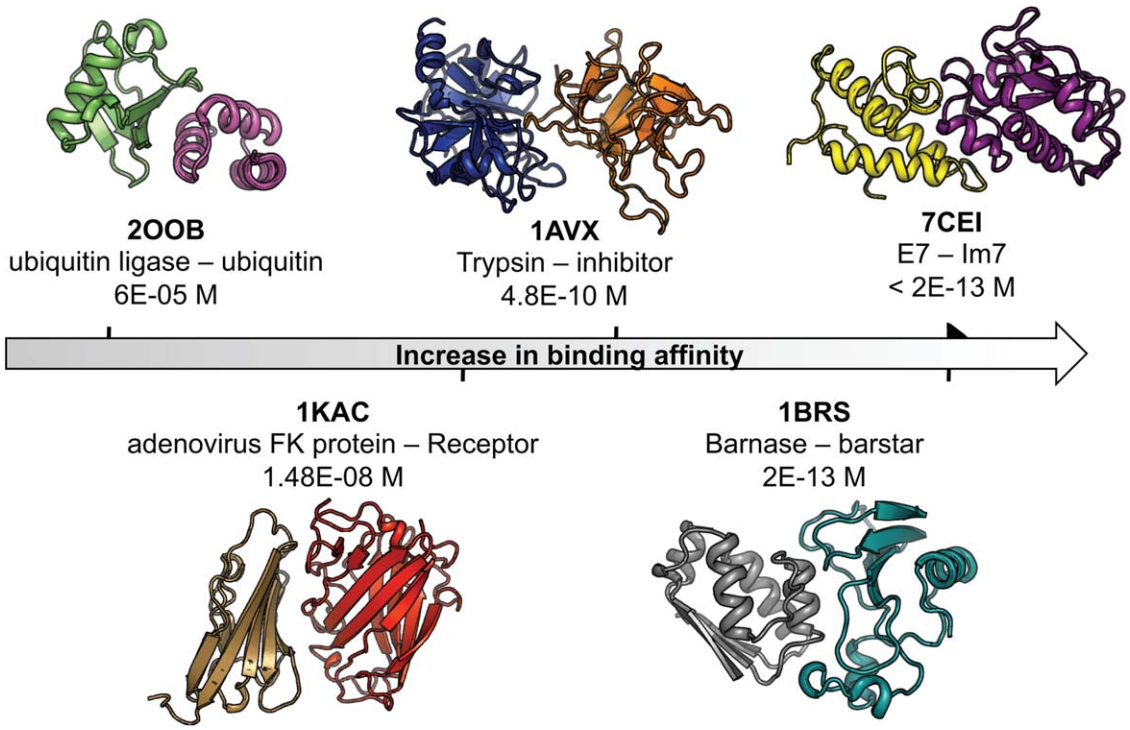

Figure 1

Protein–protein complexes used in this study. The corresponding PDB of the bound complex and affinity data are shown (Reproduced from Ref. 12, with permission from Wiley-Blackwell).

were updated every 10 fs with a 10 Å cut-off distance for the short-range neighbor list. Random starting velocities were assigned. An optimal dodecahedron solvent box was defined with an initial protein-boundary minimum distance of 16 Å. The system was slowly equilibrated, which included a 3 ns simulation under production conditions. This was followed by a 50 ns production run, of which all were used for all subsequent analysis.

Because water molecule diffusion is occurring at time scales in the order of picoseconds, coordinate frames were saved every picosecond for analysis. Each 50 ns simulation thus contained 50,000 frames including the full protein and water coordinates.

Surface residues assignment

The surface of the complexes was divided into three parts, which were subsequently mapped onto the corresponding surface of the free proteins. These are (Fig. 2) as follows:

- the interface where residues that bury surface upon complexation are located,
- the rim region that surrounds the actual interface, and
- the non-interacting surface.

Surface residues are those that have an absolute solvent accessibility $\geq$20 Å^2 or a relative solvent accessibility $\geq$2%. They were identified from the separated components of the complex. Solvent accessibilities were calculated using NACCESS V2.1.1 (www.bioinf.manchester.ac.uk/naccess/),

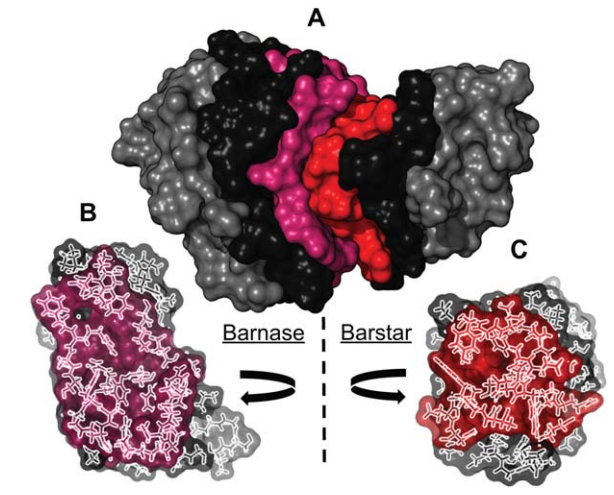

Figure 2

Illustration of the interface (red/magenta), rim (black), and non-interacting surface (gray) regions for the barnase–barstar complex (PDB ID: 1BRS[40]). (**A**) Surface representation of the complex (**B**, **C**) rotation by 90° and −90° for barnase and barstar, respectively; combined surface–stick representation is shown.

by probing the van der Waals surface with a sphere of 1.4 Å radius, approximating that of a water molecule.

The interface residues are defined as the subset of surface residues that are within 5 Å distance from the partner molecule based on intermolecular atomic contacts. Deeply buried interfacial residues falling outside the defined cut-offs (rare occurrence) were not considered in the interface analysis.

Rim surface residues are defined as those within 5 Å intramolecular atomic contact distance from assigned interfacial residues. These do not make direct contacts (within 5 Å) with the partner protein, but are within potential intermolecular interaction distance through, for example, long-range ionic interactions or water-mediated intermolecular contacts.

Finally, the non-interacting surface residues are defined as those fulfilling the surface residue criteria and not located on either the interface or the rim surface.

Assessment of the B-factors

The Debye–Waller factor, also called temperature- or B-factor provides a measure for the vibrational motion of a specific atom, when calculated from simulations. When taken from experimental crystal structures, these factors also account for other sources like crystal packing defects and heterogeneity, for example. They can be calculated directly from the atom coordinates derived from an MD trajectory. When analyzing the motion of protein residues, the entire trajectory was used. In case of water molecules, however, the analysis was limited to the time window that a given water molecule spends in the first hydration layer at proximity of a given residue. The corresponding frames were extracted from the trajectory and aligned using a standard least squares fit protocol considering all atoms of the specific residue's corresponding chain to remove contributions from the internal protein dynamics. The atomic or residue root mean square fluctuation (RMSF_a), which defines how much does an atom a or group of atoms move in space through time, was calculated as:

$$\mathrm{RMSF}_a = \sqrt{\frac{1}{N}\sum_{i=1}^{N} ||\vec{r}_i - \langle\vec{r}\rangle||^2} \qquad (1)$$

where $<r>$ corresponds to the time averaged position, and N is the number of frames considered. This RMSF_a value can be easily translated into an actual B-factor with:

$$\text{B-factor}_a = \frac{\mathrm{RMSF}_a^2 8\pi^2}{3} \qquad (2)$$

Assessment of water molecule residence times

Each water molecule in the trajectory has a unique identification number allowing tracking of its position through time. Its average residence time near a given residue was calculated as:

$$\overline{T_{\mathrm{res}}} = \frac{\sum_{i=1}^{N_{\mathrm{uniq}}^{\mathrm{water}}} A_i \mathrm{d}t}{N_{\mathrm{uniq}}^{\mathrm{water}}} \qquad (3)$$

where T_{res} denotes the average residence time of a water molecule close to a given residue, A_i denotes the amount of frames spent by water molecule i close to the specific residue surface and $\mathrm{d}t$ the time period between two consecutive frames. $N_{\mathrm{uniq}}^{\mathrm{water}}$ denotes the total number of unique water molecules, in sense of individually tracked water molecules that come and stay close to the residue for a minimum period of time.

To be considered in the residence time analysis, a water molecule must fulfill the following conditions:

- A water molecule is considered part of the local hydration layer of a residue if the Euclidean distance between a heavy atom of the side chain of the residue and the oxygen atom of the water molecule is smaller than 3.9 Å (using an alternative cutoff of 4 Å between the center of mass of the residue and the oxygen of the water molecule did not alter the results).
- A water molecule leaving the local hydration layer but returning within 10 ps is counted as continuously present during that period, although the time absent is not included as time spent in the hydration layer. If the same water molecule returns after 10 ps, it will be counted as a new water molecule.

For the analysis of residue type characteristics, only water molecules with a residence time longer than 100 ps were considered. These molecules show a clearly more stable behavior than bulk water and are expected to contribute to the stabilization of the first hydration layer, being more probable starting points for the formation and propagation of higher-order solvation shells around surfaces.[37]

The protein surface residues were classified according to their hydrophobicity using the standard Kyte–Doolittle hydrophobicity scale.[51] This classification led to two groups of different polarization states, namely the polar and charged surface residues.

RESULTS

In order to study the effect of solvent on the surface of protein–protein complexes and in particular in the rim and non-interface regions, we performed MD simulations of five different complexes and their free components ranging in binding affinity from very high (6E-05 M) to very low (2E-13 M) dissociation constants (see Material and Methods section). Each system was simulated for 50 ns using GROMACS,[44,45] with the AMBER_{99}-ILDN force field[43] and the TIP3P water model.[46]

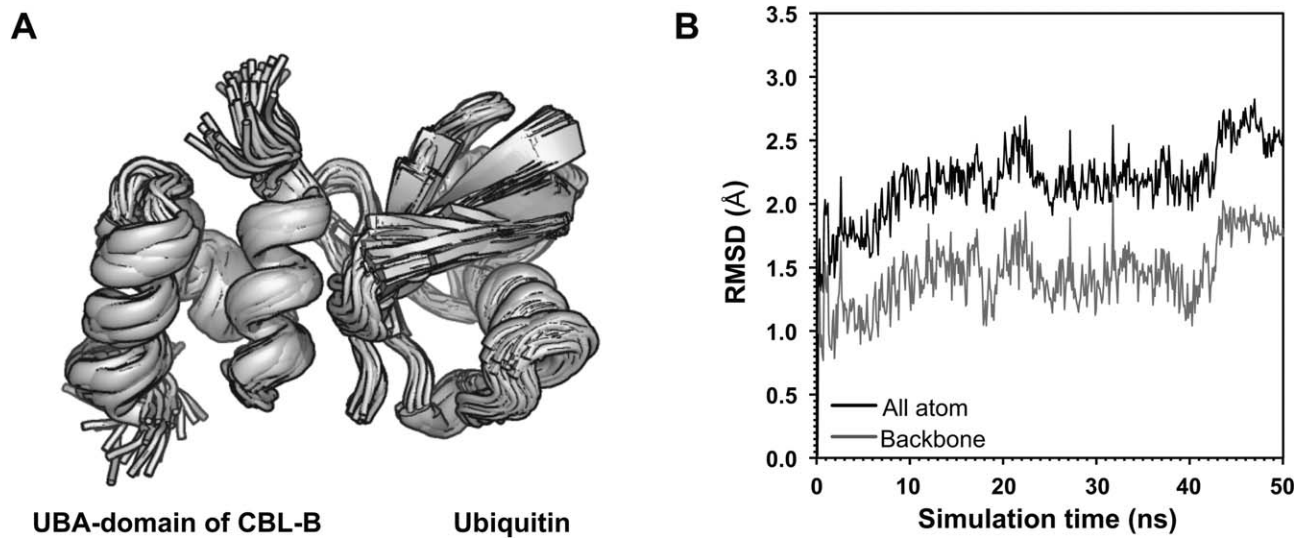

Figure 3

(**A**) Snapshots taken every 1 ns from the 50-ns MD trajectory of an ubiquitin ligase–ubiquitin complex (PDB ID 2OOB:A_B[41]). (**B**) All-atom/backbone atom positional RMSD deviations as a function of the simulation time. Structures were superimposed on the backbone using the g_rmsdist function of GROMACS.[44,45]

Quality assurance

To make sure that all systems were stable under the chosen simulation conditions (see Materials and Methods section), positional RMSDs from the respective starting structures and secondary structure content were monitored as a function of time. In all cases, the overall secondary and tertiary structure was well preserved throughout the entire simulation.

As an example, a superposition of 50 snapshots of the UBA domain of ubiquitin ligase–ubiquitin complex[41] taken at 1-ns intervals is shown in Figure 3(A). Secondary structure and global fold are well maintained. Both all-atom and backbone RMSDs from the starting crystal structure of the bound complex do not exceed 3.0 Å and 2.0 Å, respectively [Fig. 3(B)]. No sign of protein unfolding is observed, and deviations in terms of RMSDs from the reference structure are mainly due to flanking N- and C-terminal loops. Overall, the mobility of the other systems is comparable with that of the UBA domain of ubiquitin ligase–ubiquitin complex [see the average RMSDs (all-atom and Ca) reported in Table I]. Backbone RMSDs as a function of simulation time and superimposed snapshots are shown for all 15 systems in Supporting Information Figures S1 and S2, respectively.

Changes in surface dynamics upon complexation

To assess the flexibility of the interface upon complexation, B-factors for both polar and charged residues located at the interface were calculated. For this, we binned the B-factor values and defined a residue fraction ($Q_{residues}$) as the proportion of residues that fall within a certain B-factor bin, with values between 0 and 1. The B-factors for both polar and charged interfacial residues for all five complexes exhibit, as expected, a significant decrease in mobility upon complexation (P-value$^{t\text{-test(paired)}} = 3.5E\text{-}03$

Table I

Time-Averaged all-Atom and Ca Positional Root Mean Square Deviations from the Reference Crystal Structures Calculated from the Last 55 ns of the Simulations Performed in This Study

PDB (bound) RMSD (Å)	1AVX (A:B)	1BRS (A:D)	1KAC (A:B)	2OOB (A:B)	7CEI (A:B)
All atom-complex	2.14 ± 0.30	1.60 ± 0.21	1.56 ± 0.18	2.21 ± 0.29	2.09 ± 0.21
Backbone-complex	1.61 ± 0.31	1.08 ± 0.18	1.18 ± 0.21	1.47 ± 0.27	1.61 ± 0.22
PDB (Unbound 1) RMSD (Å)	1QQU (A)	1A2P (A)	1NOB (F)	2OOA (A)	1UNK (D)
All atom-Unbound 1	1.28 ± 0.16	1.50 ± 0.17	1.55 ± 0.16	2.29 ± 0.24	2.03 ± 0.23
Backbone-Unbound 1	0.83 ± 0.14	1.08 ± 0.15	1.11 ± 0.14	1.66 ± 0.20	1.05 ± 0.16
PDB (Unbound 2) RMSD (Å)	1BA7 (B)	1A19 (B)	1F5W (B)	1YJ1 (A)	1M08 (B)
All atom-Unbound 2	2.41 ± 0.29	1.47 ± 0.12	1.57 ± 0.20	3.42 ± 0.58	2.78 ± 0.46
Backbone-Unbound 2	1.71 ± 0.27	0.76 ± 0.09	0.98 ± 0.17	2.62 ± 0.67	2.29 ± 0.45

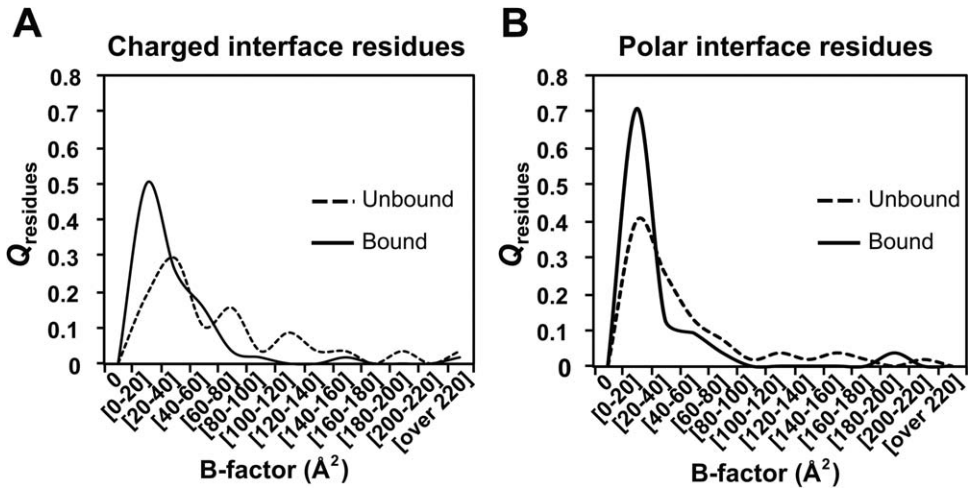

Figure 4

Shift in B-factor (Å^2) distributions upon complexation for (**A**) charged and (**B**) polar residues present at the interface, for all the five protein–protein complexes studied.

for polar and 5.9E-06 for charged; Fig. 4 and Supporting Information Table S1). This effect is preserved when the simulations are split into halves, indicating good convergence of those properties [Supporting Information Fig. S3(A,B)]. Overall, the distributions of B-factors for both charged and polar residues in the interface show a clear decrease in mobility in the transition from an unbound to a bound state (Fig. 4); this is less apparent for the rim [P-value$^{t\text{-test(paired)}} = 3.0\text{E-}03$ for polar and 8.3E-02 for charged; Supporting Information Fig. S4(A,B)] and almost no clear trend is observed for the NIS residues [P-value$^{t\text{-test(paired)}} = 2.5\text{E-}02$ for polar and 3.8E-01 for charged; Supporting Information Fig. S4(C,D)].

An analysis of the dynamics of polar and charged residues on the non-interfacial surface reveals global differences. On both rim and non-interacting surfaces, charged residues are more mobile than polar ones. This behavior is characterized by B-factor distributions of polar residues shifted toward lower calculated B-factors both for the rim [P-value$^{t\text{-test(unpaired)}} = 1.2\text{E-}01$ for bound and 3.6E-03 for unbound; Fig. 5(A,B) and Supporting Information Table S1] and, more pronounced, for the NIS residues [P-value$^{t\text{-test(unpaired)}} = 2.0\text{E-}05$ for bound and 7.1E-04 for unbound; Fig. 5(C,D) and Supporting Information Table S1] compared with charged residues. These effects are robust and still observed even when the simulations are split in two halves (Supporting Information Fig. S3).

We also analyzed the probability of a residue losing mobility upon complex assembly as function of its distance (measured on the Ca atom) from the center of mass of the interface. A lower B-factor in the bound structure suggests that the residue becomes less mobile (more stable) upon assembly, whereas a higher B-factor points to the opposite. We find that residues close to the interface periphery have a much higher chance of becoming less mobile than residues further away on the non-interacting surface. A clear trend between location with respect to the interface and probability for loss of mobility can be observed [Fig. 6(A)]. In contrast, on average, residues further than 23 Å from the center of mass of the interface exhibit a higher probability for increased flexibility upon complex formation contrary to those close to the interface, for which a decrease in mobility is observed. Residues further than 23 Å from the center of mass of the interface are mostly located on the non-interacting surface, with very few located at the interface periphery [Fig. 6(B)].

Using the same classification as mentioned earlier, different effects of similar magnitude can be observed in the interface and the noninterfacial regions (rim and non-interacting surface) of the individual complexes [Fig. 6(C)]. Most residues in the interface (60%–100%) are becoming less mobile, but this effect decays when residues on the non-interface surface are considered. The non-interacting surface shows either the same increase in mobility as the rim or becomes significantly more mobile. This phenomenon is observed for all protein–protein complexes studied [Fig. 6(C)].

Hydration layer dynamics

To characterize the dynamics of the hydration layer, B-factors of water molecules and their corresponding residence times around surface residues were calculated as described in Materials and Methods section and shown in Figures 7 and 8. These figures show violin plots, which represent a combination of a box plot and a kernel density representation of the distributions of these two water molecule properties.

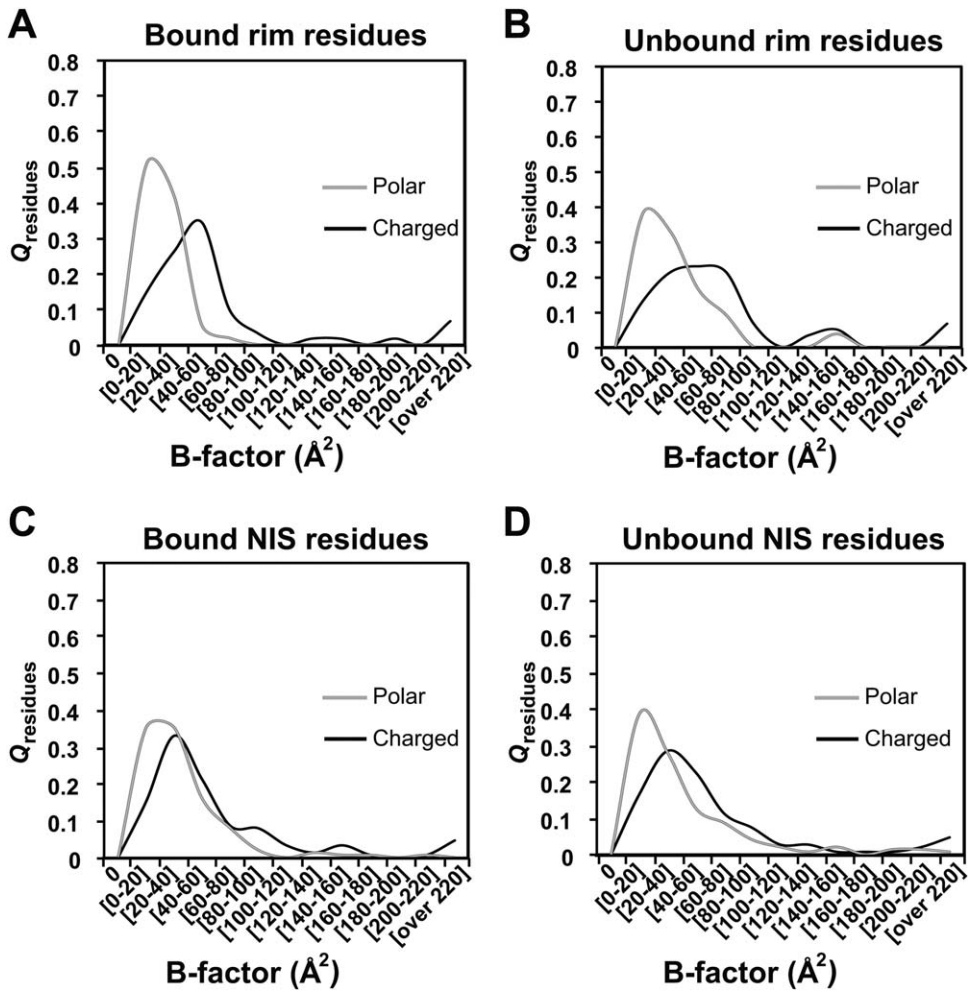

Figure 5

Shift in B-factor (Å^2) distributions of polar residues compared with charged residues, present in the rim (**A**, **B**) and non-interacting surface (NIS) (**C**, **D**) regions, in the complexed (A, C) and uncomplexed (B, D) states of the proteins. The shift to lower B-factor values indicates a stabilization of the rim and NIS side chains upon complexation.

As far as their B-factors are concerned (Fig. 7), water molecules at the NIS show broader distributions compared with the rim independently of the nature of the residue with which they interact. We can clearly observe that water molecules close to charged residues have higher B-factors compared with those interacting with polar residues on the non-interface surface [both rim (P-value$^{t\text{-test(unpaired)}} = 2.5\text{E-}08$ for unbound and 1.9E-08 for bound) and NIS (P-value$^{t\text{-test(unpaired)}} = 9.7\text{E-}06$ for unbound and 3.2E-09 for bound)]. This shows that in both the unbound and the bound conformation of the protein–protein complexes, the solvation layer has preferential mobility depending on the amino acid charge state.

Interacting water molecules located at the rim are more stable close to polar residues even when considered in a per-complex basis [Fig. 7(B)]. The same trend is identified for NIS residues [Fig. 7(C)] corroborating the robustness of the observed trend.

The majority of water molecules present in the first hydration layer on the protein surface have residence times within the 100–200 ps timescale (Fig. 8). We observe no significant change in residence times for the free and complexed states. A majority of the water molecules have a fast exchange with the bulk solvent, with only 25% of the bound water molecules showing residence times above 200 ps with maximum values reaching 30 ns. These molecules are often located in surface pockets and form optimal hydrogen bonds with protein atoms (see for example, Supporting Information Fig. S5).

The water molecules in the first hydration layer of the NIS surface were sorted based on their residence time around specific surface residues, charged or polar. The enrichment ratio (or fold-change) in the residence time of these water molecules is calculated by dividing the number of water molecules in the complex by the number of water molecules in the free protein for a given

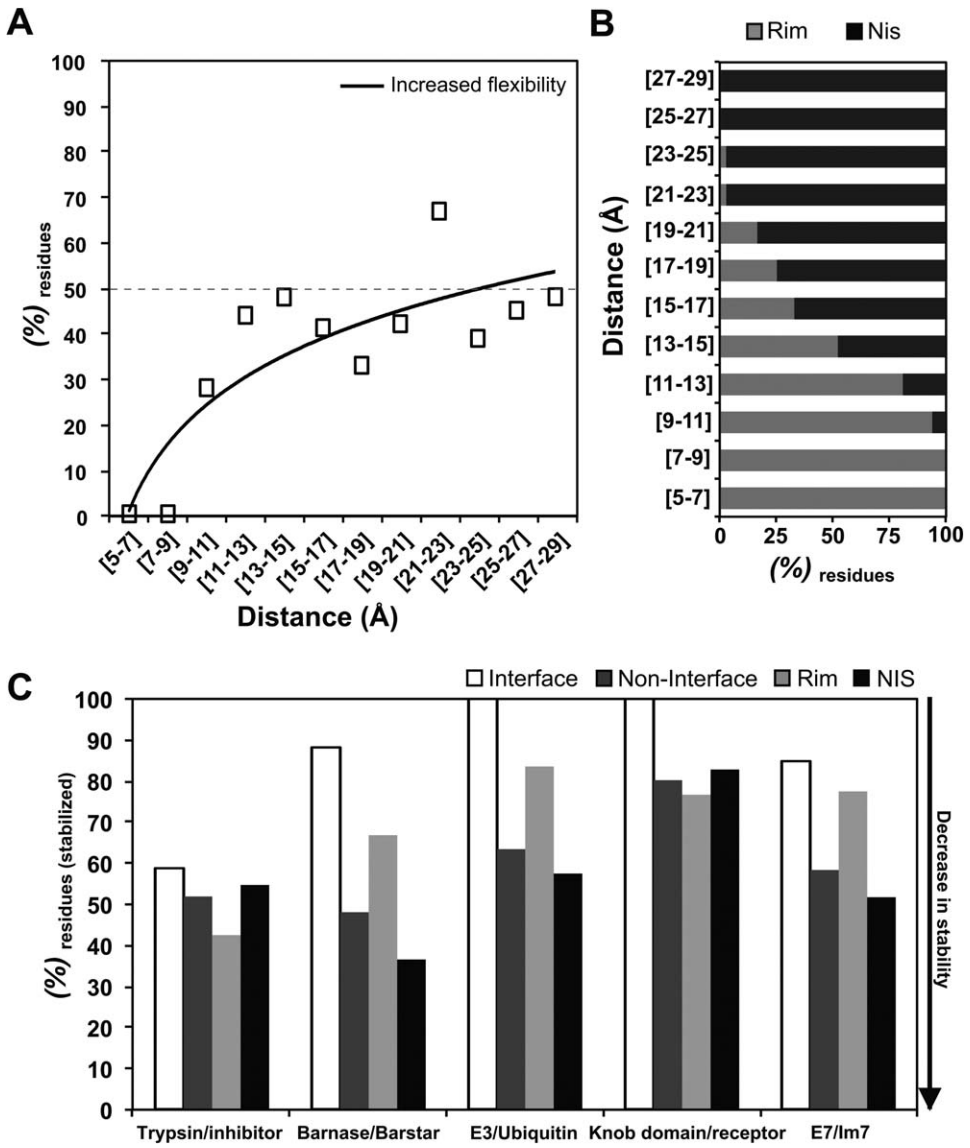

Figure 6

Percentage of residues becoming more mobile (B-factor shift to higher values) (**A**) and portion of surface to be rim or non-interacting (**B**) as a function of distance from the center of mass of the interface (see Results section). Line at 50% denotes that the majority of residues above or below the line are more flexible or less flexible, respectively. (**C**) Per-complex quantification of percentage of residues becoming more stable on interface, rim, and non-interacting surface regions.

residence time bin. We only observe an enrichment of very stable water molecules (corresponding to Quartile 4 of the distribution of residence times) around polar surfaces. In contrast to polar surfaces, we observe the opposite effect close to charged surfaces (Fig. 9). This effect is observed both in the full-length simulation analysis and in the split halves, corroborating the robustness of the observed trend. Interestingly, in the second halves of the simulations, a decrease in stable water molecules is observed for both water molecules close to charged and polar residues. The decrease close to charged residues is more pronounced. Note that water molecules with residence times above 6 ns are not analyzed because of their small number in the bins considered (see Supporting Information Table S2 for details in number of water molecules in each bin).

DISCUSSION

Entropic contribution of the non-interacting surface residues

The well-known and expected loss of mobility and stabilization of interface residues upon complexation was

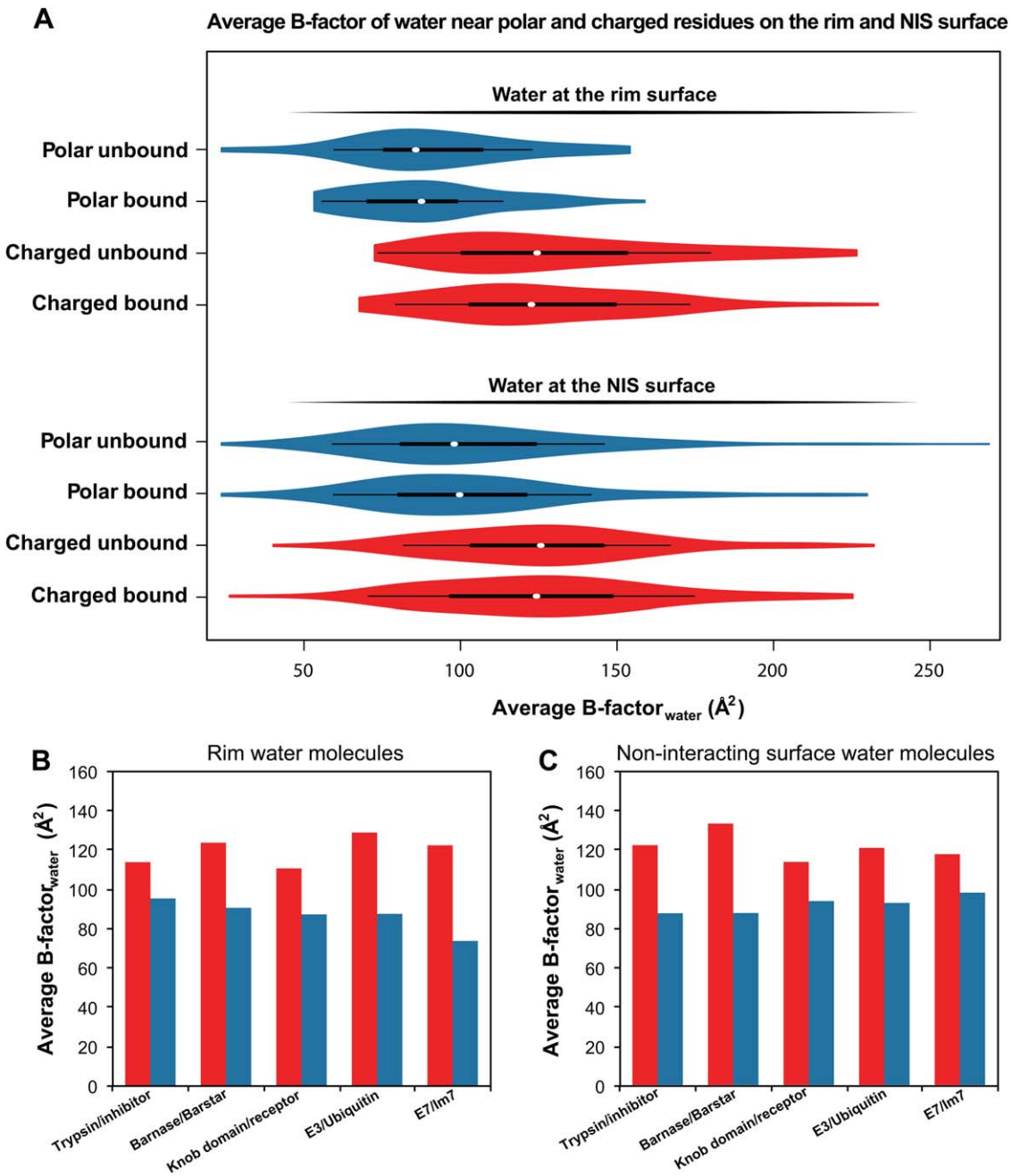

Figure 7

(**A**) Distributions of average B-factors (Å^2) of water molecules near polar and charged residues on the rim region and on the non-interacting surface represented with violin plots. Averages over the polar and charged residues are shown as a white point; one and two standard deviation units are represented with black lines of different thickness. Red and blue density represents the probability density of the data at different values in the distribution for water molecules around charged and polar residues, respectively. (**B**, **C**) Average B-factors (Å^2) of water molecules close to polar (indicated with blue bars) and charged (indicated with red bars) residues located on the rim (B) and non-interacting surface (C). These plots (B, C) concern all simulated complexes in their bound states.

observed in the MD simulation trajectories performed in this study. A clear shift in the B-factor distribution of interfacial residues toward lower values was observed. As expected, intermolecular interactions in the complex do stabilize the interface in all cases studied.

Although alanine scanning mutagenesis studies have shown that the effect of mutations on the non-interacting surface is not as important as those stemming from mutations in the interfacial or rim region, these can still affect the dissociation constant of protein–protein

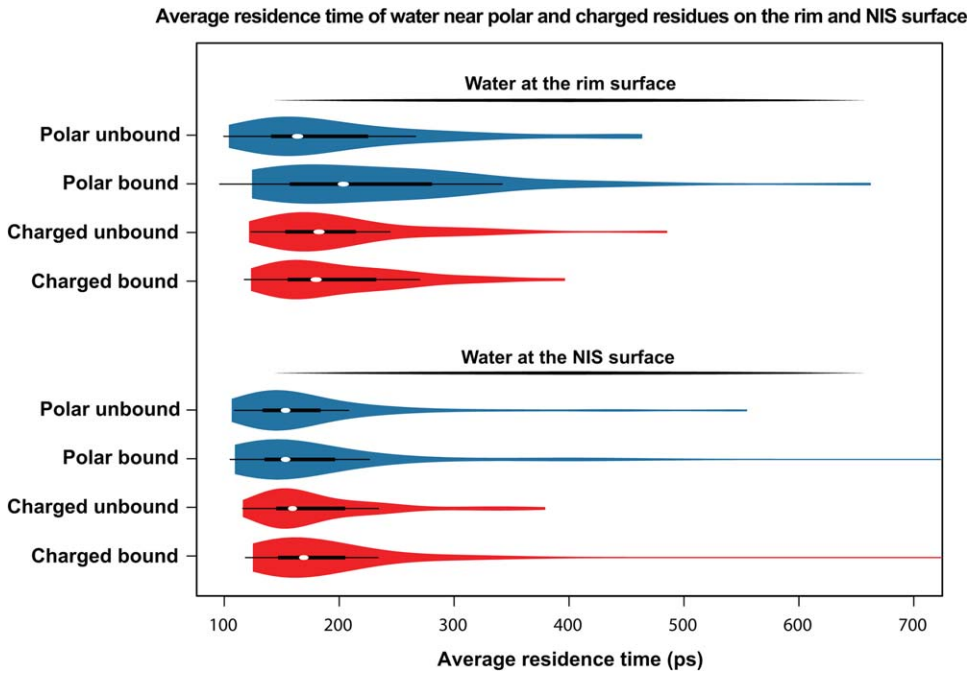

Figure 8

Distributions of average residence time of water molecules near polar and charged residues on the rim region and on the non-interacting surface. Averages over the polar and charged residues are shown as a white point, and first and second standard deviations are shown as black lines. Red and blue density represents the probability density of the data at different values in the distribution for water molecules around charged and polar residues, respectively.

complexes.[7,15] This suggests that the non-interacting surface has a direct contribution to binding affinity. Our analysis revealed that the probability of a residue to be stabilized (as measured by its mobility) upon complex formation is related to its distance from the interface. Residues closer to the interface are losing mobility upon

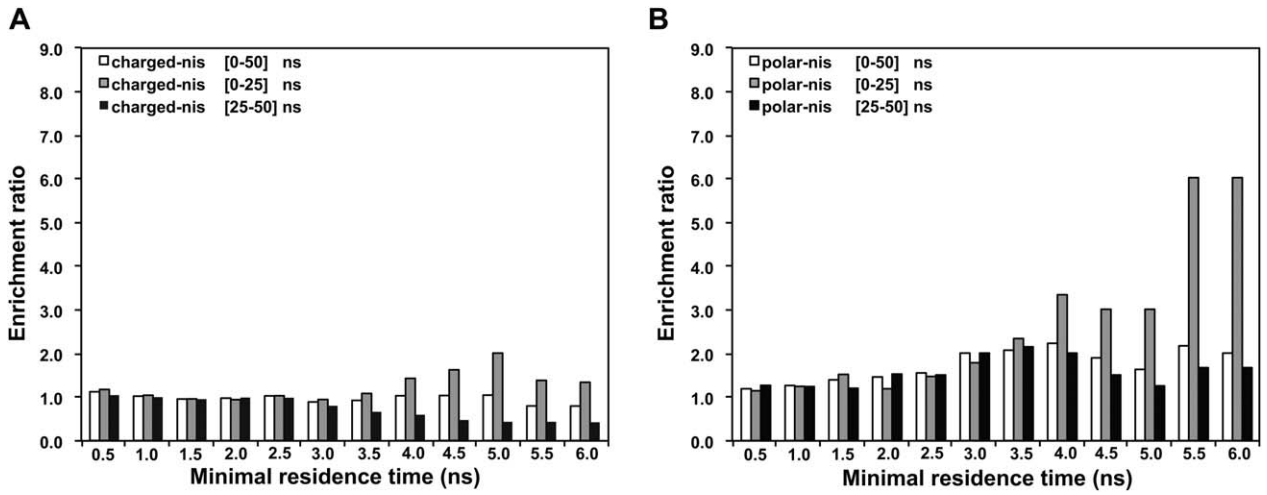

Figure 9

Enrichment ratios, calculated by dividing the number of water molecules in the complex for a given residence time bin by the number of water molecules in the free protein for the same residence time bin, for (**A**) charged and (**B**) polar residues present on the non-interacting surface in waters as a function of the residence time. The analysis was performed for the full trajectory (white bars) and the first (grey bars) and second halves (black bars). All three show similar trends. There are, however, too few water molecules with residence times beyond 6 ns to allow any proper statistical analysis (see Supporting Information Table S2).

complexation. This effect is observed up to ~23 Å away from the center of mass of the interface [Fig. 6(A)], where most residues are classified as being in the interface periphery [Fig. 6(B)]. Effects beyond 23 Å are more probably destabilizing, and these residues become more mobile in the bound state [Fig. 6(A)]. These residues are mostly located at the non-interacting surface [Fig. 6(B)]. This would mean that, upon complexation, the non-interacting surface becomes more mobile and consequently leads to a direct entropic compensation for the overall entropy loss that is observed in the interface and rim regions. To show that there is a net effect, we calculated the percentage of NIS within and outside the 23 Å cutoff. NIS percentages outside the 23 Å cutoff vary for the various complexes, covering 5%–58% of the total NIS in the complexes studied. We thus detect a phenomenon that resembles the long-debatable entropy–enthalpy compensation effect.[52–57]

Note that the complexes selected for this study all follow a lock-and-key recognition mechanism, with only very limited conformational changes upon binding (<1.0 Å RMSD of their interface residues). Many complexes undergo significant conformational changes or even folding transitions upon binding; some even increase flexibility upon complexation. For these, the non-interacting surface properties must also contribute to the binding affinity, but in a much more complex manner than for the complexes studied in this work because additional entropic factors than just side-chain dynamics and solvation will influence the binding affinity. In addition, difference in dissociation constants could be associated to differences in B-factor distributions upon protein binding in a qualitative manner; however, given the few interactions studied herein, conclusions for such relations are very premature to be drawn. Instead, more systems must be analyzed with the methods established in this work for proper quantitative B-factor change/binding affinity relationships.

The mobility of the hydration layer is affected by surface properties

The mobility of the hydration layer is determined by the permanence, both in terms of localization as measured by the B-factors and residence times of water molecules at charged or polar surfaces. We observe very low B-factors for several water molecules (Fig. 7) both in the rim region and the NIS. A trend related to the residues' polarizability in B-factors of water molecules is evident despite their broad distribution (being in the unbound or bound protein form): Indeed, water molecules close to charged residues have higher B-factors compared with those present close to polar residues on the non-interface surface. This relation is not unexpected because charged residues have longer side chains than polar ones. We have, however, previously shown from an analysis of high-resolution crystal structures that the absolute accessible surface area of the interacting residues does not seem to affect the observed water B-factors.[15] The residence time of water molecules will, however, also depend on a residue's location on the protein surface and the local physicochemical environment of the interacting molecules (residue and water molecule). We have found a simple, yet descriptive proxy for water molecule B-factors, being residue polarizability (Fig. 7).

Considering the residence times of water molecules, the majority of bound water molecules (75%) are exchanging relatively fast with the bulk solvent, with residence times below a few hundred picoseconds. Only a small quartile show residence times from a few hundred picoseconds to a few nanoseconds. These water molecules have plausibly a better chance to be initiators for the formation of an oriented water–water hydrogen-bonding network that is necessary for higher-order hydration shell formation around protein surfaces, in line with previous studies.[17,37] *In silico* formation of higher-order hydration shells is, however, hindered by the diffusion properties of the water model used in this study, TIP3P, whose diffusion coefficient is two-and-a-half times higher than that of the experimental value.[58]

Surprisingly, the stable water molecules close to the rim do not show a significant change in mobility upon complexation because their B-factor and residence time distributions do not shift when comparing the free and bound state simulations. This would mean that water-mediated salt bridges at the rim regions do not stabilize the associated water molecules. Water molecules around polar rim residues show, however, a distribution indicative of higher residence times and lower B-factors compared with water molecules around charged residues (Fig. 7). This could be explained by the lower mobility of polar residues compared with charged residues. Rationally, a more rigid residue provides a better platform for the formation of favorable hydrogen bonds with the first hydration layer, consequently stabilizing the nearby water molecules longer. Especially, when water molecules are located in the rim region, they can be involved in solvent-mediated intermolecular interactions, the latter already shown to significantly contribute to protein–protein recognition.[33,59]

We performed a simple analysis for assessing the probability of water molecules to interact with the charged or the polar surface of protein–protein complexes and their free components as a function of their B-factor (Fig. 7). We clearly observe that stable water molecules are more prone to interact with polar residues on the surface than with charged ones; an opposite effect is observed for waters with B-factors > 30. We have previously performed the same analysis on 186 crystallographic structures of resolution <1.0 Å, revealing a similar trend.[15] Our simulation results are thus in good agreement with crystallographic data (Fig. 10).

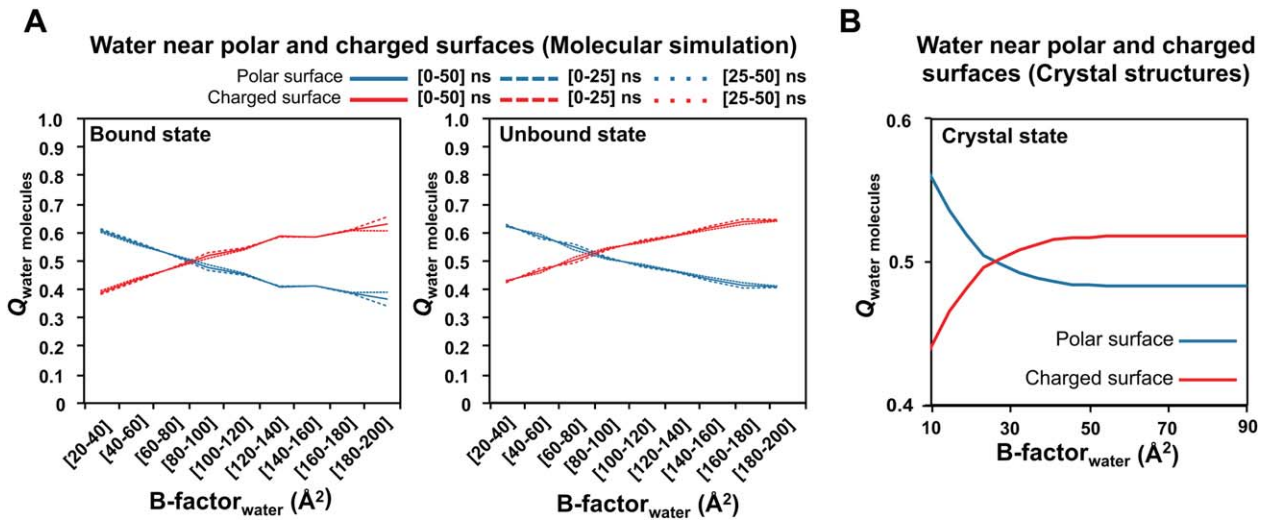

Figure 10

(**A**) Probability of water molecules to interact with charged or polar patches on the protein complex surface and the free components as a function of B-factor range. (**B**) Probabilities of water molecules as a function of their respective B-factors near polar and charged residues in crystallographic structures.[15]

Moreover, we observe a significant enrichment (fold-increase) of very stable water molecules around polar regions on the non-interacting surface upon complex formation. Around charged surface patches, we do not observe any change in the amount of water molecules with relatively high residence times [Fig. 9(A,B)]. This would indeed signify that polar residues may stabilize the hydration layer. This stable hydration layer was proposed to shield the complex from undesirable interactions and thus contribute to its stability.[15] However, the entire phenomenon must be much more complex because, in principle, entropy is gained by more mobile water structure, which should also contribute to a stabilization of the complex.

Overall, we can clearly see that the mobility of charged residues is higher than the mobility of polar residues on the entire protein surface for both unbound and bound proteins, in agreement with previous observations on static structures.[15] The differences in mobility between surface residues on the rim and the non-interacting surface corroborate the Global Surface Model which we previously proposed,[7,15] stating that properties of the non-interacting surface, and in particular interactions with water molecules, do modulate protein–protein dissociation rates, besides the dominant direct contributions from the interface. Polar residues on the non-interacting surface were indeed shown to contribute positively to affinity, whereas corresponding charged residues contribute in a negative manner. The main hypothesis is thus that an increased polarity and decreased charge on the non-interacting surface stabilizes the hydration layer and, consequently, both reactants (the free proteins) and their products, the complexes, are less likely to participate in nonspecific interactions. Regulation of hydration layer stability by properties of the non-interacting surface should affect binding properties of the complexes, such as binding affinity, promiscuity, and specificity. In the crowded cellular environment, these solvent effects must, however, be rather complicated, but may still play a simple, yet dominant role in finely adjusting promiscuity and specificity in protein–protein interactions.

ACKNOWLEDGMENTS

This work was supported by the Dutch Foundation for Scientific Research (NWO) through a VICI grant (n° 700.56.442) to A.M.J.J.B.

REFERENCES

1. Guerry P, Herrmann T. Advances in automated NMR protein structure determination. Q Rev Biophys 2011;44:257–309.
2. Sakakibara D, Sasaki A, Ikeya T, Hamatsu J, Hanashima T, Mishima M, Yoshimasu M, Hayashi N, Mikawa T, Walchli M, Smith BO, Shirakawa M, Guntert P, Ito Y. Protein structure determination in living cells by in-cell NMR spectroscopy. Nature 2009;458:102–105.
3. Joachimiak A. High-throughput crystallography for structural genomics. Curr Opin Struct Biol 2009;19:573–584.
4. Berman HM. The protein data bank: a historical perspective. Acta Crystallogr A 2008;64:88–95.
5. Berman HM, Westbrook J, Feng Z, Gillil G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res 2000;28:235–242.
6. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol 1998;280:1–9.
7. Kastritis PL, Bonvin AM. Molecular origins of binding affinity: seeking the Archimedean point. Curr Opin Struct Biol 2013;23:868–877.

8. Janin J, Bahadur RP, Chakrabarti P. Protein–protein interaction and quaternary structure. Q Rev Biophys 2008;41:133–180.
9. Chothia C, Janin J. Principles of protein–protein recognition. Nature 1975;256:705–708.
10. Horton N, Lewis M. Calculation of the free energy of association for protein complexes. Protein Sci 1992;1:169–181.
11. 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;9:2216–2225.
12. Kastritis PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AM, Janin J. A structure–based benchmark for protein–protein binding affinity. Protein Sci 2011;20:482–491.
13. Kastritis PL, Bonvin AM. On the binding affinity of macromolecular interactions: daring to ask why proteins interact. J R Soc Interface 2013;10:20120835.
14. Fleishman SJ, Whitehead TA, Strauch EM, Corn JE, Qin S, Zhou HX, Mitchell JC, Demerdash ON, Takeda-Shitaka M, Terashi G, Moal IH, Li X, Bates PA, Zacharias M, Park H, Ko JS, Lee H, Seok C, Bourquard T, Bernauer J, Poupon A, Aze J, Soner S, Ovali SK, Ozbek P, Tal NB, Haliloglu T, Hwang H, Vreven T, Pierce BG, Weng Z, Perez-Cano L, Pons C, Fernandez-Recio J, Jiang F, Yang F, Gong X, Cao L, Xu X, Liu B, Wang P, Li C, Wang C, Robert CH, Guharoy M, Liu S, Huang Y, Li L, Guo D, Chen Y, Xiao Y, London N, Itzhaki Z, Schueler-Furman O, Inbar Y, Potapov V, Cohen M, Schreiber G, Tsuchiya Y, Kanamori E, Standley DM, Nakamura H, Kinoshita K, Driggers CM, Hall RG, Morgan JL, Hsu VL, Zhan J, Yang Y, Zhou Y, Kastritis PL, Bonvin AM, Zhang W, Camacho CJ, Kilambi KP, Sircar A, Gray JJ, Ohue M, Uchikoga N, Matsuzaki Y, Ishida T, Akiyama Y, Khashan R, Bush S, Fouches D, Tropsha A, Esquivel-Rodriguez J, Kihara D, Stranges PB, Jacak R, Kuhlman B, Huang SY, Zou X, Wodak SJ, Janin J, Baker D. Community-wide assessment of protein-interface modeling suggests improvements to design methodology. J Mol Biol 2011;414:289–302.
15. Kastritis PL, Rodrigues JP, Folkers GE, Boelens R, Bonvin AM. Proteins feel more than they see: fine-tuning of binding affinity by properties of the non-interacting surface. J Mol Biol 2014;426:2632–2652.
16. Otting G, Wuthrich K. Studies of protein hydration in aqueous-solution by direct NMR observation of individual protein-bound water molecules. J Am Chem Soc 1989;111:1871–1875.
17. Otting G, Liepinsh E, Wuthrich K. Protein hydration in aqueous solution. Science 1991;254:974–980.
18. Zhang L, Wang L, Kao YT, Qiu W, Yang Y, Okobiah O, Zhong D. Mapping hydration dynamics around a protein surface. Proc Natl Acad Sci U S A 2007;104:18461–18466.
19. Nakasako M. Water–protein interactions from high-resolution protein crystallography. Philos Trans R Soc Lond B Biol Sci 2004;359:1191–1204; discussion 1204–1206.
20. Chen XF, Weber I, Harrison RW. Hydration water and bulk water in proteins have distinct properties in radial distributions calculated from 105 atomic resolution crystal structures. J Phys Chem B 2008;112:12073–12080.
21. Raschke TM. Water structure and interactions with protein surfaces. Curr Opin Struct Biol 2006;16:152–159.
22. Rowe AJ. Probing hydration and the stability of protein solutions—a colloid science approach. Biophys Chem 2001;93:93–101.
23. Chakraborty S, Sinha SK, Bandyopadhyay S. Low-frequency vibrational spectrum of water in the hydration layer of a protein: a molecular dynamics simulation study. J Phys Chem B 2007;111:13626–13631.
24. Billeter M, Guntert P, Luginbuhl P, Wuthrich K. Hydration and DNA recognition by homeodomains. Cell 1996;85:1057–1065.
25. Brunne RM, Liepinsh E, Otting G, Wuthrich K, van Gunsteren WF. Hydration of proteins. A comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. J Mol Biol 1993;231:1040–1048.
26. Denisov VP, Carlstrom G, Venu K, Halle B. Kinetics of DNA hydration. J Mol Biol 1997;268:118–136.
27. Liepinsh E, Otting G, Wuthrich K. NMR observation of individual molecules of hydration water bound to DNA duplexes: direct evidence for a spine of hydration water present in aqueous solution. Nucleic Acids Res 1992;20:6549–6553.
28. Sunnerhagen M, Denisov VP, Venu K, Bonvin AM, Carey J, Halle B, Otting G. Water molecules in DNA recognition I: hydration lifetimes of trp operator DNA in solution measured by NMR spectroscopy. J Mol Biol 1998;282:847–858.
29. Bonvin AM, Sunnerhagen M, Otting G, van Gunsteren WF. Water molecules in DNA recognition II: a molecular dynamics view of the structure and hydration of the trp operator. J Mol Biol 1998;282:859–873.
30. Levy Y, Onuchic JN. Water mediation in protein folding and molecular recognition. Annu Rev Biophys Biomol Struct 2006;35:389–415.
31. Jayaram B, Jain T. The role of water in protein–DNA recognition. Annu Rev Biophys Biomol Struct 2004;33:343–361.
32. 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;819:355–374.
33. Papoian GA, Ulander J, Wolynes PG. Role of water mediated interactions in protein–protein recognition landscapes. J Am Chem Soc 2003;125:9170–9178.
34. Ricchiuto P, Rocco AG, Gianazza E, Corrada D, Beringhelli T, Eberini I. Structural and dynamic roles of permanent water molecules in ligand molecular recognition by chicken liver bile acid binding protein. J Mol Recognit 2008;21:348–354.
35. van Dijk M, Visscher KM, Kastritis PL, Bonvin AM. Solvated protein–DNA docking using HADDOCK. J Biomol NMR 2013;56:51–63.
36. Kastritis PL, Visscher KM, van Dijk AD, Bonvin AM. Solvated protein–protein docking using Kyte-Doolittle-based water preferences. Proteins 2013;81:510–518.
37. Rupley JA, Careri G. Protein hydration and function. Adv Protein Chem 1991;41:37–172.
38. Song HK, Suh SW. Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. J Mol Biol 1998;275:347–363.
39. Ko TP, Liao CC, Ku WY, Chak KF, Yuan HS. The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. Structure Fold Des 1999;7:91–102.
40. Buckle AM, Schreiber G, Fersht AR. Protein–protein recognition: crystal structural analysis of a barnase barstar complex at 2.0-angstrom resolution. Biochemistry 1994;33:8878–8889.
41. Peschard P, Kozlov G, Lin T, Mirza A, Berghuis AM, Lipkowitz S, Park M, Gehring K. Structural basis for ubiquitin-mediated dimerization and activation of the ubiquitin protein ligase cbl-b. Mol Cell 2007;27:474–485.
42. Bewley MC, Springer K, Zhang YB, Freimuth P, Flanagan JM. Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. Science 1999;286:1579–1583.
43. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 2010;78:1950–1958.
44. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 2013;29:845–854.
45. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 2008;4:4:435–447.
46. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
47. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular-dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.

48. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys 2007;126:014101.
49. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. J Comput Chem 1997; 18:1463–1472.
50. Darden T, York D, Pedersen L. Particle mesh Ewald—an N.log(N) method for Ewald sums in large systems. J Chem Phys 1993;98: 10089–10092.
51. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982;157:105–132.
52. Starikov EB, Norden B. Enthalpy–entropy compensation: a phantom or something useful? J Phys Chem B 2007;111:14431–14435.
53. Cornish-Bowden A. Enthalpy–entropy compensation: a phantom phenomenon. J Biosci 2002;27:121–126.
54. Leung DH, Bergman RG, Raymond KN. Enthalpy–entropy compensation reveals solvent reorganization as a driving force for supramolecular encapsulation in water. J Am Chem Soc 2008;130:2798–2805.
55. Sharp K. Entropy–enthalpy compensation: fact or artifact? Protein Sci 2001;10:661–667.
56. Hunter CA, Tomas S. Cooperativity, partially bound states, and enthalpy–entropy compensation. Chem Biol 2003;10:1023–1032.
57. Lumry R. Uses of enthalpy–entropy compensation in protein research. Biophys Chem 2003;105:545–557.
58. Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. J Phys Chem A 2001;105:9954–9960.
59. Papoian GA, Ulander J, Eastwood MP, Luthey-Schulten Z, Wolynes PG. Water in protein structure prediction. Proc Natl Acad Sci U S A 2004;101:3352–3357.