

FOR THE RECORD

Reassessing buried surface areas in protein–protein complexes

Devlina Chakravarty,¹ Mainak Guharoy,² Charles H. Robert,²
 Pinak Chakrabarti,^{1*} and Joël Janin^{3*}

¹Department of Biochemistry, Bose Institute, P-1/12 CIT Scheme VIIM, Kolkata 700 054, India

²CNRS Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique (IBPC), Université Paris Diderot, Sorbonne Paris Cité, Paris, France

³IBBMC, CNRS UMR 8619, Université Paris-Sud 11, Orsay, France

Received 26 June 2013; Accepted 5 August 2013

DOI: 10.1002/pro.2330

Published online 12 August 2013 proteinscience.org

Abstract: The buried surface area (BSA), which measures the size of the interface in a protein–protein complex may differ from the accessible surface area (ASA) lost upon association (which we call DSA), if conformation changes take place. To evaluate the DSA, we measure the ASA of the interface atoms in the bound and unbound states of the components of 144 protein–protein complexes taken from the Protein–Protein Interaction Affinity Database of Kastiris *et al.* (2011). We observe differences exceeding 20%, and a systematic bias in the distribution. On average, the ASA calculated in the bound state of the components is 3.3% greater than in their unbound state, and the BSA, 7% greater than the DSA. The bias is observed even in complexes where the conformation changes are small. An examination of the bound and unbound structures points to a possible origin: local movements optimize contacts with the other component at the cost of internal contacts, and presumably also the binding free energy.

Keywords: protein–protein interaction; solvent accessible surface; conformation changes; binding free energy

Introduction

Protein–protein recognition is essential to all aspects of life. Its physical chemical basis has long been known to reside in desolvated atoms that form non-

covalent electrostatic and van der Waals interactions at molecular interfaces,^{1–6} which the many entries of the Protein Data Bank (PDB⁷) reporting structures of protein–protein complexes illustrate. Yet, structure-based models of binding thermodynamics and kinetics are still far from quantitative, and they have little predictive value,^{8,9} in part because they take into consideration the structure of the complex but not that of its components, and thus, they ignore conformation changes which contribute to the mechanism and affect the free energy balance of the reaction.

Here, we examine the effect of conformation changes on a geometric quantity, the area of the protein surface that becomes buried upon association.

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

Grant sponsor: Indo-French Centre for the Promotion of Advanced Research; Grant number: CEFIPRA project number 4003-2.

*Correspondence to: Joël Janin, IBBMC, UMR 8619 CNRS, Université Paris-Sud 11, Orsay, France.

E-mail: joel.janin@u-psud.fr or Pinak Chakrabarti, Department of Biochemistry, Bose Institute, P-1/12 CIT Scheme VIIM, Kolkata 700 054, India. E-mail: pinak@jcb Bose.ac.in

The buried surface area (BSA), commonly used to estimate the size of the interface between two macromolecules, is calculated on atomic coordinates of the complex alone.¹ We introduce a novel quantity, the decrease in surface area (DSA), derived from solvent accessible surface areas (ASA) measured on both the bound (in the complex) and the unbound (the free protein) states of the components. In a set of 281 bound/unbound structure pairs taken from the Protein–Protein Interaction Affinity Database (PPIAD),⁹ we observe differences between the two values that exceed 20% as the result of conformation changes between the two states. Even in complexes that show small conformation changes, the ASA of interface atoms tends to be greater in bound than unbound state, and thus, the BSA tends to be larger than the DSA, due to local movements of interface atoms that augment their contacts with the partner molecule. These findings have implications on the mechanism of protein–protein recognition and the thermodynamics of the association reaction.

Results

Dataset and definitions

The PPIAD comprises 144 binary protein–protein complexes and their free components.⁹ As the antibody moiety of seven antibody–antigen complexes are missing, this yields 281 B (bound in the complex) and 281 U (unbound, free) component structures. ASA values are measured separately in U, in B and in the complex C.

Atom i contributes to the BSA:

$$BSA_i = ASA_i(B) - ASA_i(C)$$

Its contribution to the DSA is the ASA lost upon association:

$$DSA_i = ASA_i(U) - ASA_i(C)$$

This differs from the contribution to the BSA by:

$$\Delta ASA_i = BSA_i - DSA_i = ASA_i(B) - ASA_i(U).$$

Summing over all the interface atoms of a component yields $ASA(B)$, $ASA(U)$, BSA , DSA , and ΔASA values for each protein in the dataset.

We also consider the fractional excess ASA of the interface atoms in B vs. U state:

$$\delta A = \Delta ASA / ASA(B)$$

Those calculations involve mapping atom i of B to the corresponding atom in U, a non-trivial process

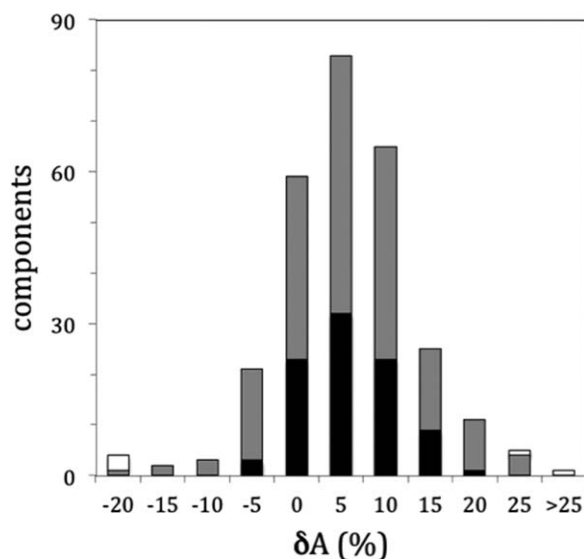


Figure 1. Distribution of δA values. δA is the fractional excess ASA of the interface atoms in B vs. U state. The histogram represents its distribution over the 281 protein components of the Protein–Protein Interaction Affinity Database. Dark columns represent the 91 components of the core set where conformation changes are small. The average δA is the same (3.3%) as in the whole sample, but the standard deviation is less (4.9% vs. 7.2%). Five components (empty columns) are excluded from those statistics for reasons explained in the text.

when B and U are from different PDB entries. Although they represent the same protein, B and U may be different genetic constructs, or have disordered segments with missing atoms. Local alignments reveal that 134 of the 281 U/B pairs have sequence identity >98%, with the rest in the range 90–98% referring to the shorter sequence. In addition, 36 interface residues (0.5% of all interface residues) belonging to 22 components are absent or different in the U and B sequences, and therefore excluded from the analysis.

Ambiguous labels, such as OD1/OD2 in Asp, occur in all PDB entries. If an interface atom is marked OD1 in B and OD2 in U, a spurious ΔASA will result. We circumvent the problem by taking both to be part of the interface. While this increases the number of the interface atoms by 8%, the BSA is unchanged, because the added atoms do not contribute to it. On the other hand, 4% of the interface atoms, belonging to 95 of the 281 components, have no counterpart in U. They are excluded from the summation yielding $ASA(B)$ to make it consistent with $ASA(U)$.

The ASA of interface atoms in the bound and unbound states

Figure 1 shows a histogram of the δA values observed in the database component proteins. Individual values and overall statistics are reported in

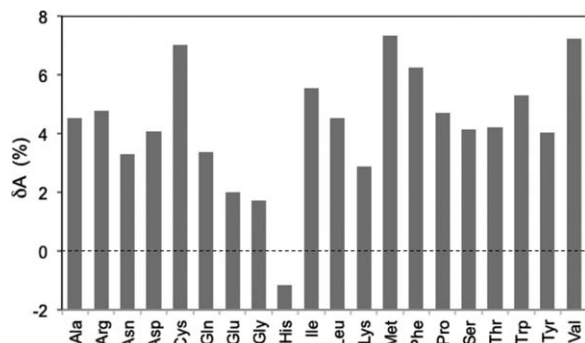


Figure 2. Average δA per residue type. Here, δA is the ratio $[\text{ASA}(\text{B}) - \text{ASA}(\text{U})]/\text{ASA}(\text{B})$, where the ASA of interface atoms in the B or U state is summed over all the residues of a given type present in the data set, which contains 7206 interface residues in total.

Supporting Information. δA ranges from -50% to $+50\%$, ΔASA , from -340 \AA^2 to 940 \AA^2 . However, the proteins yielding such extreme values appear as grossly different constructs in B and U, or contain disordered segments in U, which leads to abnormal values of $\text{ASA}(\text{U})$. After excluding five proteins for those reasons, the range of ΔASA is still -250 \AA^2 to 610 \AA^2 , that of δA -21% to 28% , which proves that the interface solvent accessibility can be very different in the U and B states, and therefore, the BSA very different from the DSA. The difference can be of either sign, but ΔASA is positive, and the BSA greater than the DSA, in 69% of cases. As a result, the average δA is positive: $3.3 \pm 7.2\%$ (mean \pm standard deviation), and the BSA, which represents about 61% of $\text{ASA}(\text{B})$, exceeds the DSA by 7.2% on average.

These differences are only marginally related to the chemical and amino acid composition of the interfaces. Dividing interface groups into polar (N, O, and S containing) and nonpolar (C-containing), δA averages 2.2% for polar, 3.6% for non-polar groups. Polar groups contribute 42%, non-polar groups 58% of the BSA in protein–protein complexes.⁴ Moreover, the average δA calculated for each of the 20 types is positive in the range 1–7% for all types except histidine ($\delta A = -1.1\%$), and it exceeds 4% for all non-polar residue types (Fig. 2).

Missing atoms and different constructs in U and B state entries certainly contribute artifacts to ΔASA in some of the proteins of the dataset. To test the possibility that they introduce a bias toward positive values, we check the ASA of surface residues, which shows no significant excess in the bound state ($\delta A = 0.9 \pm 6.0\%$). We also consider a core set of 91 component proteins that have sequences at least 99% identical in B and U, no more than two missing interface atoms in U, and a RMSD less than 1.5 \AA (see Supporting Information). The histogram of their δA values is shown alongside that of the larger set

in Figure 1. It is narrower, but the average is the same ($3.3 \pm 4.9\%$) and only 25 out of 91 values are negative, a proportion that would have $P < 10^{-5}$ if the two signs were equally probable.

Local conformation changes make interface atoms more accessible in the bound state

In the core set, B and U contain the same atoms, and the different accessibilities of the interface atoms in the two states must result from small conformation changes. PPIAD derives from a docking benchmark in which the complexes that display a RMSD less than 1.5 \AA are described as undergoing rigid-body association.¹⁰ No disorder–order transition, secondary structure change, or domain rotation occurs in the core set, but side chain rotations and local movements of the polypeptide chain do, and Figure 3 shows how they can affect the accessibility of interface atoms. When ferredoxin binds to ferredoxin-NADP reductase,¹¹ its conformation remains unchanged (the RMSD is only 0.8 \AA), but the movement of a tyrosine side chain allows its phenol group to make an interface H-bond and

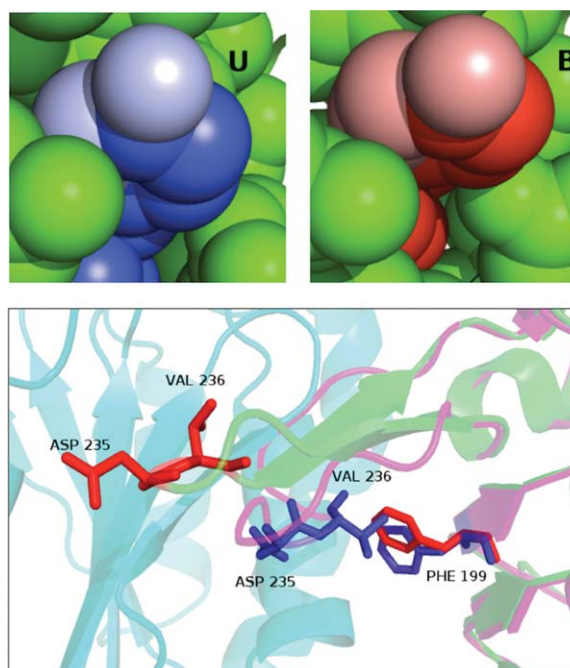


Figure 3. Local movements that make interface residues more accessible in bound state. A: Tyr 25 of ferredoxin in U state (1CZP) and B state (the complex with ferredoxin-NADP reductase,¹¹ 1EWY); the interface atoms (lightly shaded) see their ASA increase by 17 \AA^2 in B state. B: A surface loop of glycoprotein IB- α shifts position when it binds to von Willebrand factor domain A1 (cyan). Interface residues Asp 235, Val 236, and Phe 199 are colored red and the main chain green in B state (the complex,¹² 1M10), blue with the main chain pink in the free glycoprotein (1M0Z). Their ASA increases by 45 to 75 \AA^2 from U to B, which accounts for most of the ΔASA seen in the complex.

optimize contacts with atoms of the reductase [Fig. 3(A)]. In doing so, the group loses some contacts with neighboring ferredoxin atoms, and this causes ASA(B) to be larger than ASA(U). When glycoprotein IB- α binds to the von Willebrand factor,¹² a loop bearing interface residues 235–236 shifts by about 5 Å toward the partner protein [Fig. 3(B)], which accounts for most of the 1.7 Å RMSD. The shift increases the accessibility of the loop residues, and also that of Phe 199, part of the interface but not involved in the main chain movement.

Discussion

Whereas the BSA calculation uses only the coordinates of the complex, computing the DSA also requires the coordinates of the free components, which come from separate X-ray or NMR studies. Then, experimental errors and random changes in atomic positions lead to different ASA values in the B and U states, although the conformation changes may be insignificant. The difference can be of either sign, and it could be expected to cancel after summing over all the atoms of an interface (about 100 per component⁴). The present study shows that this is not the case. Even when conformation changes are small, the interface atoms tend to have a larger ASA in the B than the U state. The excess is only a few per cents, but the bias toward positive values is highly significant. Greater differences of either sign are observed when the conformation changes are large, and their average is also positive.

Figure 3 suggests a possible explanation for the bias toward a positive Δ ASA. In a complex, the interface atoms tend to make fewer contacts within their component as they interact with the other component. As a consequence, some internal energy must be spent going from the U to the B state, and therefore, any model calculation that is done on the complex assuming the unbound components to have the same structure will overestimate the contribution of atom–atom interactions to the binding free energy. In a first approximation, we may take this contribution to scale linearly with the number of atom pairs at the interface, itself is a linear function of the BSA. Experimental values of the Gibbs free energy of dissociation (ΔG_d) are available for all the complexes of the PPIAD.⁹ They display no correlation with the BSA in the whole dataset, but the correlation is significant ($R = 0.54$) for a subset of 70 complexes with a RMSD < 1 Å, and it remains the same after replacing the BSA with the DSA. The two components of 29 of those complexes also belong to the core set of the present study. If the linear correlation reported in Ref. 9 is applied to their Δ ASA, the cost of the U to B transition evaluates to only between -0.2 and 0.7 kcal mol⁻¹, but significant contributions are expected in complexes that undergo larger conformation changes.

Methods

Atomic coordinates were obtained from the PDB for the 144 complexes of the PPIAD (<http://bmm.cancerresearchuk.org/~bmmadmin/Affinity/>), and for 281 of their free components. In 17 entries that report NMR structures, only the first model was retained. Ubiquitin entry 1yj1, which contains a D-glutamine, was replaced by 1ubq with the natural residue. Modified residues were converted to the original (e.g., selenomethionine to methionine) reported in the sequence. All other HETATM lines were omitted.

Sequence alignments used the Smith–Waterman algorithm and the EMBOSS software.¹³ Conformation changes were estimated by the root-mean-square distance (RMSD) between equivalent C α atoms after least-square superposition with PROFIT.¹⁴ Solvent ASA were calculated with NACCESS,¹⁵ which implements the Lee and Richards algorithm.¹⁶ Any atom losing more than 0.1 Å² ASA between states B and C was considered as part of the interface. If this concerned an atom with an ambiguous label, the other atom in the pair was made part of the interface: Asp OD1/OD2, Glu OE1/OE2, Phe and Tyr CD1/CD2 and CE1/CE2, and also Arg NH1/NH2, Asn OD1/ND2, Gln OE1/NE, His CD1/ND2 and CE1/NE2, which are difficult to distinguish in X-ray structures.

References

1. Chothia C, Janin J (1975) Principles of protein–protein recognition. *Nature* 256:705–708.
2. Jones S, Thornton JM (1996) Principles of protein–protein interactions. *Proc Natl Acad Sci USA* 93:13–20.
3. Lo Conte L, Chothia C, Janin J (1999) The atomic structure of protein–protein recognition sites. *J Mol Biol* 285:2177–2198.
4. Chakrabarti P, Janin J (2002) Dissecting protein–protein recognition sites. *Proteins* 47:334–343.
5. Nooren IM, Thornton JM (2003) Diversity of protein–protein interactions. *EMBO J* 22:3486–3492.
6. Janin J, Bahadur RP, Chakrabarti P (2008) Protein–protein interaction and quaternary structure. *Q Rev Biophys* 41:133–180.
7. Berman H, Henrick K, Nakamura H, Markley JL (2007) The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic Acids Res* 35:D301–D303.
8. Kastiris PL, Bonvin AM (2010) Are scoring functions in protein–protein docking ready to predict interactions? Clues from a novel binding affinity benchmark. *J Proteome Res* 9:2216–2225. Erratum in: *J Proteome Res* (2011) 10:921–922.
9. Kastiris PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AM, Janin J (2011) A structure-based benchmark for protein–protein binding affinity. *Protein Sci* 20:482–491.
10. Hwang H, Vreven T, Janin J, Weng Z (2010) Protein–protein docking benchmark version 4.0. *Proteins* 78:3111–3114.
11. Morales R, Kachalova G, Vellieux F, Charon MH, Frey M (2000) Crystallographic studies of the interaction between the ferredoxin–NADP⁺ reductase and ferredoxin from the cyanobacterium *Anabaena*: looking for the elusive ferredoxin molecule. *Acta Crystallogr D* 56:1408–1412.

12. Huizinga EG, Tsuji S, Romijn RA, Schiphorst ME, de Groot PG, Sixma JJ, Gros P (2002) Structures of glycoprotein Ibalpha and its complex with von Willebrand factor A1 domain. *Science* 297: 1176–1179.
13. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* 6:276–277.
14. McLachlan AD (1982) Rapid comparison of protein structures. *Acta Crystallogr A* 38:871–873.
15. Hubbard SJ, Thornton JM (1993) “NACCESS”, computer program. Department of Biochemistry and Molecular Biology, University College London.
16. Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 3:379–400.