THE PROTEIN SOCIETY

FOR THE RECORD

Protein–protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area

Jieming Chen,[1,2] Nicholas Sawyer,[1,3] and Lynne Regan[1,2,3,4]*

[1]Integrated Graduate Program in Physical and Engineering Biology, Yale University, New Haven, Connecticut 06520
[2]Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut 06520
[3]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520
[4]Department of Chemistry, Yale University, New Haven, Connecticut 06520

Received 21 December 2012; Revised 27 January 2013; Accepted 28 January 2013
DOI: 10.1002/pro.2230
Published online 7 February 2013 proteinscience.org

Abstract: Protein–protein interactions play key roles in many cellular processes and their affinities and specificities are finely tuned to the functions they perform. Here, we present a study on the relationship between binding affinity and the size and chemical nature of protein–protein interfaces. Our analysis focuses on heterodimers and includes curated structural and thermodynamic data for 113 complexes. We observe a direct correlation between binding affinity and the amount of surface area buried at the interface. For a given amount of surface area buried, the binding affinity spans four orders of magnitude in terms of the dissociation constant (K_d). Across the entire dataset, we observe no obvious relationship between binding affinity and the chemical composition of the interface. We also calculate the free energy per unit surface area buried, or "surface energy density," of each heterodimer. For interfacial surface areas between 500 and 2000 Å^2, the surface energy density decreases as the buried surface area increases. As the buried surface area increases beyond about 2000 Å^2, the surface energy density levels off to a constant value. We believe that these analyses and data will be useful for researchers with an interest in understanding, designing or inhibiting protein–protein interfaces.

Keywords: protein–protein interaction; hydrophobicity; surface energy density; hot spot; dissociation constant; buried surface area

Introduction

Protein–protein interactions (PPIs) are vital for cellular function. Moreover, the unique function of each interaction determines its affinity and specificity. Perturbation of such interactions, including loss of vital PPIs or gain of inappropriate PPIs, is often associated with disease states.[1,2] Recently, small molecule inhibitors of specific PPIs have been identified as potential therapeutic agents.[3–8] It is therefore of significant interest and importance to understand protein–protein interfaces, particularly the relationship between the area and chemical nature of the interaction interface and the affinity of the interaction.

Early investigations on this topic date back more than a decade,[9,10] when the majority of the

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

Grant sponsor: Raymond and Beverly Sackler Institute for Biological, Physical, and Engineering Sciences.

*Correspondence to: Lynne Regan, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520. E-mail: lynne.regan@yale.edu

cocrystal structures available were enzyme-inhibitor or antibody-antigen complexes. The topic was revisited in 2003 using a set of manually curated protein complexes for which both structural and thermodynamic data were available. There was a particular emphasis on "transient" complexes, which were defined as those in which the component proteins can exist as either monomers or in the complex.[11] Although this dataset contained only 10 homodimers and 9 heterodimers, the data nevertheless hinted at a direct relationship between the surface area buried upon complex formation and the strength of the interaction. Since these pioneering studies, the number of structures in the Protein Data Bank (PDB)[12] has increased to almost 70,000 in 2011.[13] Therefore, we now revisit this topic with more available structures, thermodynamic data, and a focus on heterodimers, which tend to be involved in essential cellular processes such as signal transduction and histone modification.

Results

Our extensive and carefully curated dataset includes 113 heterodimeric complexes. Since all the structures in our dataset are obtained from the PDB, hereafter, we use its PDB ID when we refer to a specific structure. We plotted the logarithm (base 10) of the measured dissociation constant (K_d) versus the calculated surface area buried upon complex formation [Fig. 1(A)]. There is a large difference in surface area buried between the smallest and largest interfaces. The smallest complex buries 381 Å^2 (2R0Y) whereas the largest buries nearly 3393 Å^2 (2WWX). The binding affinity of different complexes ranges from the weakest with a K_d of 1 mM (2R0Y) to the strongest with a K_d of 3 pM (3JZA). For a given buried surface area, the dissociation constants can vary over four orders of magnitude. A least squares fit of the data gives a slope that corresponds to a free energy change per unit area of 1.6 cal mol^{-1} Å^{-2}, where free energy change is $\Delta G = -\mathrm{RT}\ln K_d$ (R = 2.0 cal K^{-1} mol^{-1} and T = 298 K). This trend indicates that as the buried interfacial surface area increases, the binding energy increases (i.e., decrease in K_d).

Our dataset contains both protein–peptide and PPIs, which are distinguished by the definition of a peptide as a chain of 20 or fewer amino acids. We investigated the dataset for any fundamental differences between the protein–peptide and protein-protein complexes. The protein–peptide complexes have binding affinities in the low milli- to nanomolar range [denoted by magenta crosses in Fig. 1(A)]. They also tend to have interface areas of 2000 Å^2 or less, with one exception (3L6X). The protein–protein complexes have binding affinities in the micro- to sub-nanomolar range. Although most of the protein–protein complexes have interfaces areas more than 2000 Å^2, there are also many of them with interface areas less than 2000 Å^2. Hence, the two groups of complexes overlap in terms of their range of affinities and buried surface areas, with no clear boundary that distinguishes them.

We also investigated the dataset to see if there is any relationship between the buried surface area and the chemical nature of the interface. To address this point, we calculated the percent of hydrophobic surface area buried in each complex. For this analysis, we divided the data into five bins according to their buried surface area: <1000, 1000–1500, 1500–2000, 2000–2500, and >2500 Å^2. For each of these bins, the mean percent of buried hydrophobic surface area is ~60%, with the largest range (36–82%) observed in the 1000–1500 Å^2 bin and the smallest range (51–62%) observed in the >2500 Å^2 bin. We observe no direct relationship between hydrophobicity and the buried surface area for any of the bins [Fig. 2(A)] or between hydrophobicity and binding affinity. We also performed similar analyses on aliphatic, aromatic, polar charged and polar uncharged surface areas separately and found no correlation between buried surface area and binding affinity (data not shown).

Finally, we calculated the free energy change per unit surface area, or "surface energy density", for each complex, that is, ΔG/(buried surface area). The results are quite striking [Fig. 1(B)]. Below 2000 Å^2 of buried surface area, the surface energy density decreases linearly as the area of the interface increases, ranging from energy densities of about 13 cal mol^{-1} Å^{-2} in smaller interfaces to energy densities of about 4 cal mol^{-1} Å^{-2} in larger interfaces. In other words, the burial of 1 Å^2 of surface area is more energetically favorable for smaller interfaces than for larger interfaces. Beyond 2000 Å^2, the surface energy density levels off to a fairly constant value of ~3–4 cal mol^{-1} Å^{-2}.

Discussion

With a larger set of protein heterodimer structures, more thermodynamic data and careful manual curation, we are able to examine the relationships between the affinity, buried surface area and chemical nature of the interface in greater detail than was possible in earlier studies. We observe a direct relationship between buried interfacial surface area and affinity [Fig. 1(A)], that is, as buried surface area increases, binding affinity increases.

In addition, we observe no direct relationship when we correlate the hydrophobicity and binding affinity of the complexes within each bin of buried surface area (data not shown). This suggests that hydrophobic interactions might not be the dominant driving force in many of these PPIs. The average percent of hydrophobic buried surface is 60%, but the range of percent hydrophobic surface area varies

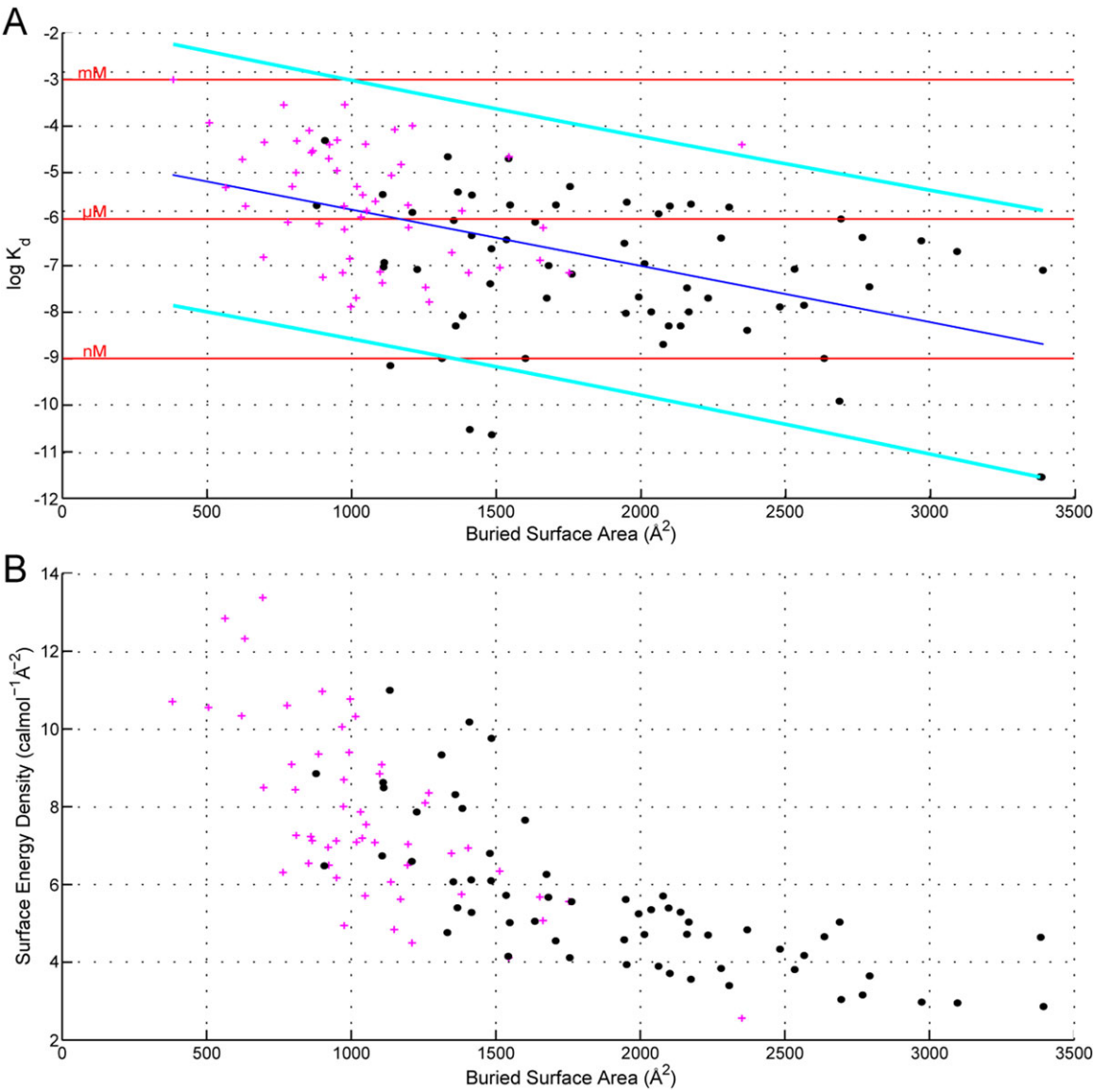

Figure 1. (A) Scatter plot showing the relationship between the logarithm (base 10) of dissociation constant (log K_d) and the buried surface area in Å^2 of 113 heterodimers. The blue line is the linear least-squares regression fit of the data with slope corresponds to a free energy change per unit area of 1.6 cal mol^{-1} Å^{-2} (Spearman's correlation $\rho = -0.53$, P value = 2.5E-9; $R^2 = 0.25$), indicating that as buried surface area increases, the affinity increases (i.e. K_d decreases). Each point represents a co-crystal structure of a heterodimer. Protein-peptide complexes, which contain a component polypeptide chain with less than 20 amino acids, are denoted by magenta crosses while all other protein-protein complexes are black circles. Horizontal red lines indicate 1 mM, 1 μM, and 1 nM dissociation constants for calibration of the log K_d values. The cyan lines are the 90% predicted interval for each new predicted value (or 'predicted interval'). For comparison, we also computed the slopes for the protein-peptide and protein-protein complexes. They correspond to a free energy change per unit area of 1.6 cal mol-1 Å^{-2} (Spearman s correlation ρ = -0.39, P value = 0.0039; $R^2 = 0.15$) for the protein-peptide complexes and 0.91 cal mol-1 é-2 ($\rho = -0.24$, P value = 0.062; $R^2 = 0.058$) for the protein-protein complexes. (B) Scatter plot showing the relationship between the free energy change per Å^{-2}, or 'surface energy density', and the buried surface area in Å^2 of 113 heterodimers. It shows two regimes: below 2000 Å^{-2}, energy density decreases with increasing area; above 2000 Å^2, a plateau is observed. As in 1A, each point represents a co-crystal structure of a heterodimer, protein-peptide complexes (as defined previously) are denoted by magenta crosses and all other protein-protein complexes are black circles.

depending on the total surface area buried. We speculate that evolutionary limits are being imposed on the total hydrophobic surface area of an interface to prevent non-specific interactions and inappropriate aggregation when the interfaces of the uncomplexed proteins are exposed.

When we analyze the data in terms of free energy per unit surface area buried, an interesting trend emerges—we find two regimes of buried surface areas [Fig. 1(B)]. For complexes that bury less than 2000 Å^2 surface area, the free energy per unit surface area buried is inversely

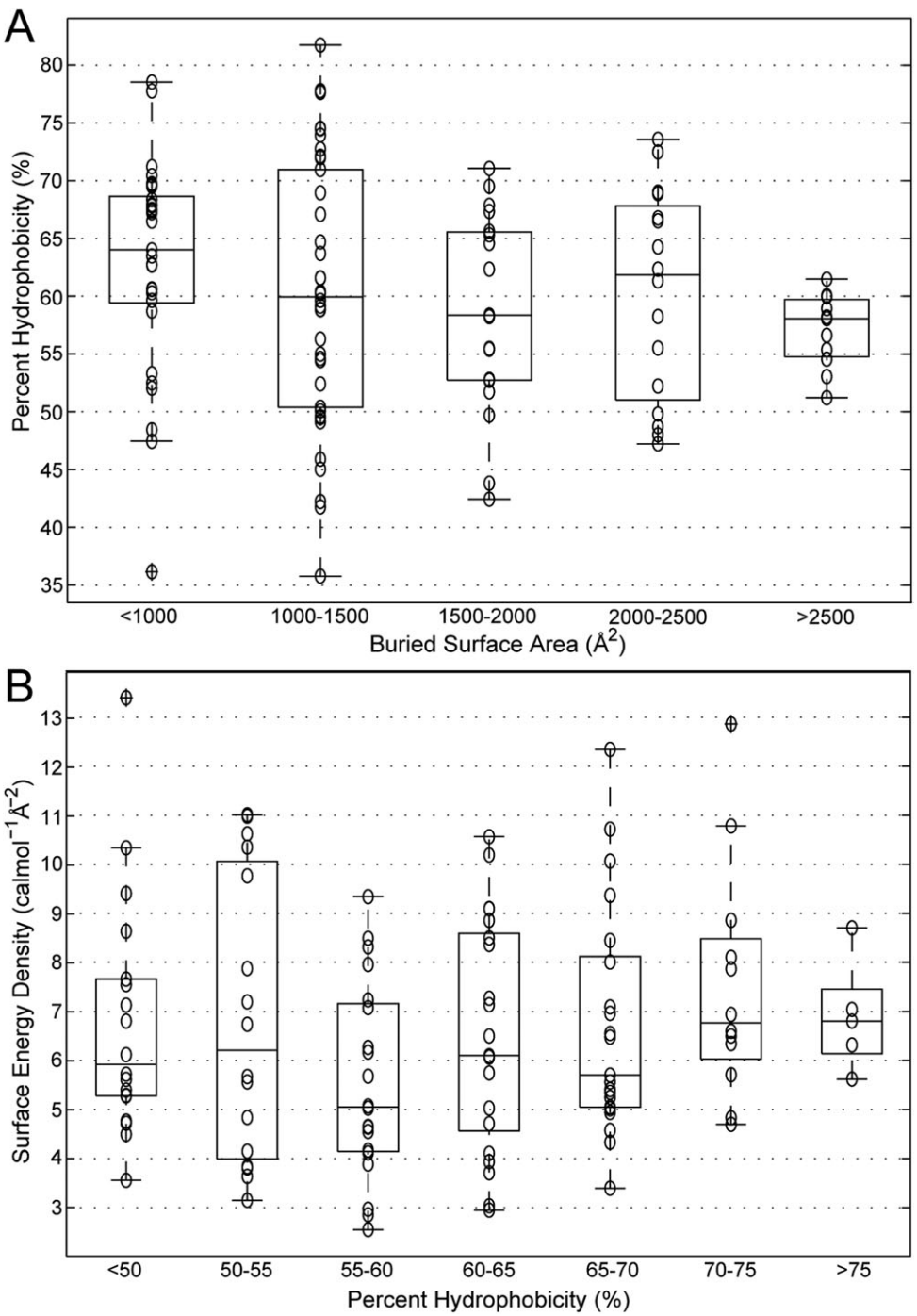

Figure 2. (A) Box plots showing the lack of correlation between the buried surface area in é2 and percent hydrophobicity. The buried surface area is divided into five bins: <1000, 1000-1500, 1500-2000, 2000-2500 and > 2500 Å^2. Each box plot indicates the quartiles of the distributions of the points within each bin of buried surface area. The lower edge of the box represents the 25th percentile (first quartile), while the upper edge indicates the 75th percentile (third quartile) of hydrophobicity. The line within the box denotes the median (second quartile) hydrophobicity. The uppermost and lowermost lines, or more commonly known as whiskers, indicate the observations that are not within the interquartile range but within 1.5 times of the interquartile range above the 75th percentile and below the 25th percentile. Points that lie beyond the whiskers are considered outliers and marked with a cross. Each open circle overlaid on the box plot represents a heterodimer in the corresponding bin of buried surface area. (B) Box plots showing the relationship between percent hydrophobicity and surface energy density. The percent hydrophobicity is divided into seven bins: < 50, 50-55, 55-60, 60-65, 65-70, 70-75 and > 75%. As in 2A, each box plot indicates the quartile distribution of each bin and each open circle represents a heterodimer.

proportional to the total surface area buried. For the highest free energy per unit area (~13 cal mol^{-1} Å^{-2}), the surface area buried is only about 694 Å^2. Beyond 2000 Å^2, the free energy per unit surface area buried for each complex is about 4 cal mol^{-1} Å^{-2}. In other words, the surface energy

density is greater for smaller complexes than for larger ones. If we consider our observed plateau at 4 cal mol^{-1} $Å^{-2}$ as the "basal" average contribution of free energy change per $Å^2$, this suggests that in complexes with buried surface areas $\leq$2000 $Å^2$, some fraction of each of these the interfaces is making greater energetic contributions than the basal level. This conclusion is consistent with the idea of "hot spots."[14,15] A hot spot residue has been previously defined as one whose mutation to alanine results in a decrease in binding energy of at least 2 kcal/mol.[16] Our results imply that as the buried surface area increases, the fraction of the buried surface area that is occupied by the hot spots decreases until $\sim$2000 $Å^2$. Beyond this point, the hot spots become a relatively small fraction of the total buried surface area. Conversely, for smaller interfaces, hot spots represent a larger fraction of the buried surface area.

Are our observations then quantitatively consistent with 2 kcal/mol? Two amino acid residues that are often identified as hotspots are tryptophan and isoleucine.[16] When each of these residues is mutated to alanine, the change in surface area is $\sim$140 and 60 $Å^2$, respectively.[17] If we assume them to be hot spot residues by the definition above, we can estimate the associated surface energy density to be 14 and 33 cal mol^{-1} $Å^2$ for tryptophan and isoleucine, respectively. In our dataset, we observe that the largest values of surface energy density are $\sim$12–14 cal mol^{-1} $Å^{-2}$. Hence, the estimated values above are consistent with the observation that smaller interaction surfaces have a larger fraction of hot spot residues.

Additionally, we investigated the relationship between the fraction of hydrophobic surface area buried and the surface area density and found none [Fig. 2(B)]. The results show that the chemical nature of hot spot residues can be hydrophobic or hydrophilic, which is consistent with previous observations.[16,18,19]

Though we interpret our surface energy density in terms of hot spots, we also recognize that our analysis cannot formally rule out the alternative possibility in which the decrease in binding free energy is uniformly distributed throughout the entire interface as the buried surface area increases. It is hard to picture why this might be, and this explanation is not consistent with the wealth of mutagenesis data available in the literature.[14,15,19,20]

Finally, we find that our analyses have important implications in protein design and inhibitor drug design. The well-defined ranges of hydrophobicity and binding affinity suggest that PPIs can be tuned by modifying the size of the interactive area or the hydrophobic fraction of the buried surface area. Moreover, even though the percent of hydrophobic buried surface area can reach fairly high (>75%), such percentages are only found in smaller interfaces ($\leq$1500 $Å^2$). Hence, the absolute amounts of hydrophobic buried surface area are in fact low—the highest percent of hydrophobicity (81.7%) corresponds to an absolute buried surface area of only about 1200 $Å^2$ (3AJB). Additionally, our hot spot analysis implies that a relatively small fraction of the protein–protein interface contributes a large proportion of the binding energy. Taken together, these conclusions suggest that small inhibitors are still feasible, for the disruption of protein-protein interfaces.

Methods

Structural and thermodynamic data curation

To obtain a nonredundant set of heterodimeric structures from the PDB, we applied a series of filters. From the starting dataset of 73,988 structures in the PDB (July 5, 2011), we included only X-ray crystal structures annotated with the keyword "complex" that were solved at a resolution of better than 3 Å and which contained exactly two different protein entities (heterodimers). We removed complexes that contain DNA or RNA. We then included only protein complexes in which each protein has less than 100% sequence identity to another protein in our dataset. After these filters, 2188 complex structures remained in our dataset. We then searched the PDBbind v2011 database[21] for thermodynamic data on these complexes. This database contained dissociation constants (K_d) for 263 of the complex structures. Finally, we manually curated our list of structures and dissociation constants by scanning the literature to ensure that (1) the K_d listed in PDBbind matches the PPI in the PDB structure, (2) small molecules or ions are not involved in the interaction, and (3) the biological assembly of the complex has not been demonstrated to be an oligomer with order >2. The final dataset consists of 113 distinct heterodimers with validated thermodynamic binding data.

Buried surface area calculation and analyses

The surface area buried at a protein–protein interface was calculated as the sum of the solvent accessible surface area of the monomers minus the solvent accessible surface area of the complex (ignoring any conformational changes to the monomers upon complex formation). The calculations were performed using the program NACCESS v2.1.1 with water represented as a sphere of radius 1.4 Å.[22] Note that some prior calculations have assumed interface symmetry and therefore divided the total surface area buried calculated by two.[11,23] The numbers we report for the total calculated interface area are not divided by two.

Hydrophobicity calculation

The buried surface area of each interface was further analyzed to determine hydrophobic content

using the method described by Kajander *et al.*[24] Each atom at the interface was assigned to one of four categories: polar charged, polar uncharged, aliphatic, or aromatic. The interface hydrophobicity was then calculated as the percentage of buried surface area composed of aliphatic and aromatic atoms.

Electronic Supplementary Material

We provide a supplementary table as a Microsoft Excel file ("suppTable_curated_113_heterodimers.xls"). It contains the data for each of the 113 heterodimers used in this study, including the PDB ID, common name of the complex (PDB title), K_d value, buried surface area, percent of hydrophobic surface area buried, and the method used to measure the dissociation constant. We also include the PubMed IDs (PMIDs), referencing the associated publications for the corresponding structures and K_d. When there are two PMIDs, the first one corresponds to publication where the structure was solved and the second one corresponds to the publication in which the K_d was actually measured. One exception is 1XG2, where the K_d was measured and documented in a chapter of a book, hence the full reference is provided.

URL

PDB: http://www.rcsb.org
PDBbind: http://www.pdbbind-cn.org/
NACCESS: http://www.bioinf.manchester.ac.uk/naccess/
PDB Statistics: http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=explMethod-xray&seqid=100

Acknowledgments

The authors thank the Regan Lab and Dorottya Blaho Noble for critical reading of and suggestions on this manuscript. Thanks to Zhihai Liu of the PDBbind team for the help in consolidating the PDBbind data. Thanks also to Ramza Shahid and Paola Peshkepija for background work in the Regan Lab related to this project. This work was supported, in part, by the Raymond and Beverly Sackler Institute for Biological, Physical and Engineering Sciences.

References

1. Stendel C, Roos A, Kleine H, Arnaud E, Ozçelik M, Sidiropoulos PNM, Zenker J, Schüpfer F, Lehmann U, Sobota RM, Litchfield DW, Lüscher B, Chrast R, Suter U, Senderek J (2010) SH3TC2, a protein mutant in Charcot-Marie-Tooth neuropathy, links peripheral nerve myelination to endosomal recycling. Brain 133: 2462–2474.
2. Jellinger K (2012) Interaction between pathogenic proteins in neurodegenerative disorders. J Cell Mol Med 16:1166–1183.
3. Vassilev LT, Vu BT, Graves B, Carvajal D, Podlaski F, Filipovic Z, Kong N, Kammlott U, Lukacs C, Klein C, Fotouhi N, Liu E (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science 303:844–848.
4. Shangary S, Wang S (2009) Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy. Ann Rev Pharmacol Toxicol 49:223–241.
5. Arkin MR, Wells JA (2004) Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. Nature Rev 3:301–317.
6. Yi F, Zhu P, Southall N, Inglese J, Austin CP, Zheng W, Regan L (2009) An AlphaScreen-based high-throughput screen to identify inhibitors of Hsp90-cochaperone interaction. J Biomol Screen 14:273–281.
7. Yi F, Regan L (2008) A novel class of small molecule inhibitors of Hsp90. ACS Chem Biol 3:645–654.
8. Cortajarena AL, Yi F, Regan L (2008) Designed TPR modules as novel anticancer agents. ACS Chem Biol 3: 161–166.
9. Lo Conte L, Chothia C, Janin J (1999) The atomic structure of protein-protein recognition sites. J Mol Biol 285:2177–2198.
10. Jones S, Thornton JM (1996) Principles of protein-protein interactions. Proc Natl Acad Sci USA 93:13–20.
11. Nooren IMA, Thornton JM (2003) Structural characterisation and functional significance of transient protein-protein interactions. J Mol Biol 325:991–1018.
12. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 112:535–542.
13. Berman H. M., (2012) Creating a community resource for protein science. Protein science 21:1587–1596.
14. Clackson T, Wells JA (1995) A hot spot of binding energy in a hormone-receptor interface. Science 267: 383–386.
15. DeLano WL (2002) Unraveling hot spots in binding interfaces: progress and challenges. Curr Opin Struct Biol 12:14–20.
16. Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. J Mol Biol 280:1–9.
17. Chothia C (1976) The nature of the accessible and buried surfaces in proteins. J Mol Biol 105:1–12.
18. Ofran Y, Rost B (2007) Protein-protein interaction hotspots carved into sequences. PLoS Comp Biol 3:e119.
19. Thorn KS, Bogan AA (2001) ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. Bioinformatics 17: 284–285.
20. Kouadio J-LK, Horn JR, Pal G, Kossiakoff AA (2005) Shotgun alanine scanning shows that growth hormone can bind productively to its receptor through a drastically minimized interface. J Biol Chem 280:25524–25532.
21. Wang R, Fang X, Lu Y, Wang S (2004) The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. J Med Chem 47:2977–2980.
22. Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. J Mol Biol 55:379–400.
23. Ponstingl H, Henrick K, Thornton JM (2000) Discriminating between homodimeric and monomeric proteins in the crystalline state. Proteins 41:47–57.
24. Kajander T, Kahn PC, Passila SH, Cohen DC, Lehtiö L, Adolfsen W, Warwicker J, Schell U, Goldman A (2000) Buried charged surface in proteins. Structure 8: 1203–1214.