PROTEINS: Structure, Function, and Genetics 48:645–653 (2002)

Stability of Macromolecular Complexes

Natasja Brooijmans,[1] Kim A. Sharp,[2] and Irwin D. Kuntz[3]*
[1]*Graduate Program in Chemistry and Chemical Biology, University of California-San Francisco, San Francisco, California*
[2]*Department of Biochemistry and Molecular Biophysics, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania*
[3]*Department of Pharmaceutical Chemistry, University of California-San Francisco, San Francisco, California*

ABSTRACT **Macromolecular interactions are crucial in numerous biologic processes, yet few general principles are available that establish firm expectations for the strength of these interactions or the expected contribution of specific forces. The simplest principle would be a monotonic increase in interactions as the size of the interface grows. The exact relationship might be linear or nonlinear depending on the nature of the forces involved. Simple "linear–free energy" relationships based on atomic properties have been well documented, for example, additivity for the interaction of small molecules with solvent, and, recently, have been explored for ligand–receptor interactions. Horton and Lewis propose such additivity based on buried surface area for protein–protein complexes. We investigated macromolecular interactions and found that the highest-affinity complexes do not fulfill this simple expectation. Instead, binding free energies of the tightest macromolecular complexes are roughly constant, independent of interface size, with the notable exception of DNA duplexes. By comparing these results to an earlier study of protein–ligand interactions we find that: (1) The maximum affinity is approximately 1.5 kcal/mol per nonhydrogen atom or 120 cal/mol Å^2 of buried surface area, comparable to results of our earlier work; (2) the lack of an increase in affinity with interface size is likely due to nonthermodynamic factors, such as functional and evolutionary constraints rather than some fundamental physical limitation. The implication of these results have some importance for molecular design because they suggest that: (1) The stability of any given complex can be increased significantly if desired; (2) small molecule inhibitors of macromolecular interactions are feasible; and (3) different functional classes of protein–protein complexes exhibit differences in maximal stability, perhaps in response to differing evolutionary pressures. These results are consistent with the widespread observation that proteins have not evolved to maximize thermodynamic stability, but are only marginally stable. Proteins 2002;48:645–653.**

Key words: macromolecular complexes; thermodynamics of binding; interface size; protein evolution; protein folding

INTRODUCTION

We recently analyzed the interaction of ligands with macromolecular receptors focusing on the highest-affinity complexes that have been reported.[1] These ligands span a considerable range in molecular weights from monatomic ions to drug molecules of 50–200 atoms. We concluded that the binding data for the strongest binding ligands reflected two tendencies: a linear increase in the free energy of binding with increasing molecular size and a plateau regime in which the best binders (e.g., subpicomolar dissociation constants) were insensitive to molecular size. We suggested that the first trend reflected the thermodynamic principle of additive free energy, while the plateau was likely to arise from kinetic considerations rather than thermodynamic or physical constraints on binding energies. While ligand–receptor interactions are of great interest, they constitute only a fraction of the molecular interaction universe. In this article, we explore another large group of molecular interactions: those involving macromolecular partners. Protein–protein, protein–nucleic acid, and nucleic acid–nucleic acid interactions are crucial in numerous biologic processes including enzymatic control, signaling, and information storage and retrieval. We ask if the same general principles apply to these systems as to those with smaller ligands. Differences might arise because of the (relative) flatness of macromolecular interfaces compared with the deeply buried ligand binding pockets associated with enzyme active sites and small molecule receptor pockets,[2] the atomic composition of macromolecular ligands that differ from low-molecular-weight ligands, or evolutionary pressures operating on activity or regulation.

Protein–protein interactions, in particular, have been the subject of study in recent years, especially since the availability of cocrystal structures has allowed the dissection of structural aspects of macromolecular recognition as well as increased understanding of the forces that drive complexation. Among the interesting findings are the

Abbreviations: ASA, accessible surface area; e.u., entropy units; IA, interface area; PDB, Protein Data Bank; standard one-letter abbreviations for amino acid residues.

*Correspondence to: Irwin D. Kuntz, Department of Pharmaceutical Chemistry, University of California-San Francisco, San Francisco, CA 94143-0446. E-mail: kuntz@cgl.ucsf.edu

Received 20 September 2001; Accepted 8 February 2002

observations that the interfaces often resemble protein surfaces in juxtaposition rather than protein interiors[3–8] and that much of the interaction energy arises from a few critical residues.[9,10] If the formation of protein–protein complexes buries more charged and hydrophilic residues than protein folding itself,[8] the balance of forces might be different, with entropy losses playing a larger role in protein folding than in complex formation.[8,11] The occurrence of "hot spots" that dominate the free energy of binding in protein complexes raises the question of whether there is a special character to these interfaces that differs from the interactions seen in small molecule complexes.[9,10]

Another interesting issue is whether there are different thermodynamic/kinetic compromises made for the different classes of protein–protein complexes. Homodimeric complexes are usually long-lived and some exhibit coordinated folding/dimerization. The components of heterodimeric systems often fold independently and are more transient than homodimeric complexes.[5,6] It has been reported that homodimeric protein–protein interfaces are more closely packed, more hydrophobic, and less planar than heterodimeric complexes.[5,6] It has also been found that the interacting surfaces of protein–inhibitor interfaces are more complementary than the interacting surfaces of antibody–antigen complexes. These differences in complementarity might have arisen from different evolutionary pressures on these complexes.[5]

Nucleic acid structural motifs include the familiar helical structures, which would be expected to exhibit linear increases in free energy as a function of molecular weight, as well as compact structures with tertiary interactions whose behavior is less predictable (e.g., tRNA). Protein–nucleic acid complexes are of great biologic relevance but relatively little thermodynamic data are currently available.

In this article, we extend the analysis of Kuntz et al.[1] to the tightest binding macromolecular complexes. We also make use of structural information, where available, to provide additional molecular interpretations of the binding behavior. In brief, we find that: (1) The maximal interactions, per atom, are similar to those found in our earlier study (ca. 1.5 kcal/mol per nonhydrogen atom), and these maximal values can be converted to 120 cal/mol Å^2 of buried surface area based on analysis of structures of selected complexes; (2) except for nucleic acid helical duplexes, most macromolecular complexes lie in the plateau regime where there is little change in complex affinity as a function of molecular weight; (3) there appear to be differences in maximal affinity based on biological function.

METHODS

Selection of Complexes

Protein–protein complexes

We choose heterodimeric protein–protein complexes from a database compiled by LoConte et al.[12] containing 75 nonredundant X-ray structures of protein–protein complexes with resolution of 3.1 Å or better. Only the 43 complexes for which equilibrium binding data (IC_{50}, K_i, K_a, or K_d) could be found in the literature were included in this study.

Homodimeric protein–protein complexes were included only if denaturation studies showed three-state dimer denaturation and if resolution was 3.1 Å or better. Dimeric complexes that show three-state denaturation complexes dissociate to their constitutive native monomers before they unfold and the equilibria associated with these separate processes can be described by

$$N_2 \overset{K_1}{\rightleftharpoons} 2N \overset{K_2}{\rightleftharpoons} 2D$$

where N_2 represents the native complex, $2N$ represents two native monomers, and $2D$ represents two denatured monomers. The dimeric dissociation constant K_d can be obtained by dilution and is often assumed to be equivalent to K_1.[13] Homodimeric complexes for which K_d was determined were included in this study. Homodimeric complexes whose denaturation thermodynamics precludes a simple decomposition into dimerization (i.e., complexes with two-state thermodynamics) and folding were excluded.

The data for both the heterodimeric and homodimeric complexes can be found in Table I.

Bogan and Thorn[14] compiled a database of alanine mutants in protein–protein interfaces. Only those mutations were selected that showed $\Delta\Delta G$ values of more than 5 kcal/mol and for which an X-ray structure of the wild-type complex was available, which are shown in Table II. For our work, the mutant structures were modeled on the wild-type structure using Sybyl, without any further minimization of the structure.

The data collected by Kuntz et al.[1] have also been included for comparison.

Nucleic acid complexes

Data for DNA duplex formation were compiled by Allawi and SantaLucia[15] (supplemental information) from their own work as well as others, and only those duplexes that show two-state thermodynamics are included. Duplexes that show two-state thermodynamics can be described by the following equation

$$2A \rightleftharpoons A_2$$

where A represents the random coil and A_2 the duplex. In the case of two-state thermodynamics of duplex formation/unfolding, no alternative secondary structures such as slipped duplexes or hairpins contribute to the equilibrium thermodynamics. We note, however, as can be seen from the equation, that these energies are not true association free energies because they include the folding free energy of each of the strands. Data for DNA duplexes that show two-state thermodynamics are in Table III.

RNA and DNA aptamers, shown in Table IV, were selected from a recent review by Hermann and Patel[16] on nucleic acid aptamers. While these interactions are typically with low-molecular-weight ligands, they serve as a

TABLE 1. Data for Enzyme–Enzyme Inhibitor, Signalling, Homodimeric, and Miscellaneous Complexes

PDB[†]	Receptor	Ligand	Type[‡]	IA (Å^2)	No. of interface atoms ligand	ΔG (kcal/mol)[§]	References[‖]
1tgs	Trypsinogen	PSTI	P–Pi	924.7	70	−4.1	1
3tpi	Trypsinogen	PTI	P–Pi	860.7	59	−7.8	2
1bth	Thrombin	BPTI	P–Pi	1382.2	113	−10.7	3
1hia	Hirustatin	Kallikrein	P–Pi	849.3	61	−10.8	4
4cpa	CPA	PCI	P–Pi	722.2	52	−11.3	5
1avw	STI	PPT	P–Pi	1008	74	−12.3	6
1fle	Elafin	PPE	P–Pi	893.6	66	−12.3	7, 8
2kai	Kallikrein	BPTI	P–Pi	945.2	57	−12.5	9
1ppf	HLE	OMTKY3	P–Pi	773.3	53	−13.4	10
2sic	SSI	BPN	P–Pi	940.9	66	−13.4	11, 12
1stf	Cystatin B	Papain	P–Pi	750.5	67	−13.5	13
3sgb	SGPB	OMTKY3	P–Pi	603.3	51	−14.7	14
4htc	Thrombin	Hirudin	P–Pi	1421.5	116	−14.8	15
1cho	Chymotrypsin	OMTKY3	P–Pi	574.4	77	−15.4	16
1acb	Chymotrypsin	Eglin C	P–Pi	748.4	62	−16.1	17
1tbq	Thrombin	Rodniin	P–Pi	1529.7	146	−17.3	18
2ptc	Trypsin	BPTI	P–Pi	855.5	60	−18.0	19
1gla	GK	III^{Gic}	E–Ei	419.5	55	−7.1	20
2pcc	CCP	Cytochrome C	E–Ei	181	40	−10.0	21
1dhk	PPA	α-AJ	E–Ei	1167.9	150	−14.3	22
1brs	Barnase	Barstar	E–Ei	677.1	76	−17.3	23, 24
1dfj	RnaseA	RI	E–Ei	684.3	114	−18.0	25, 26
1mlc	D 44.1	HEL	Ab–Ag	743.9	58	−9.7	27, 28
1nmb	NC10	Neuraminidase	Ab–Ag	521.4	56	−10.0	29
1mel	cAb-Lys3	HEL	Ab–Ag	934.4	87	−10.5	30
1jhl	D 11.15	PHL	Ab–Ag	536.3	54	−10.7	31
1dvf	FvD1.3	FvE5.2	Ab–Ag	942.8	82	−11.2	32
1vfb	D 1.3	HEL	Ab–Ag	555.8	85	−11.5	33
2jel	Jel42	HPr	Ab–Ag	653.6	53	−11.5	34
1fbi	F9.13.7	GEL	Ab–Ag	751	85	−11.6	28
1nsn	Fab N10	Snase	Ab–Ag	530.6	78	−11.8	35
3hfm	Hy-HEL10	HEL	Ab–Ag	863.9	74	−12.8	36
3hfl	Hy-HEL5	HEL	Ab–Ag	659.5	70	−14.5	37
1a2k	GDP-Ran	NTF2	Misc. sign.	814.9	71	−9.3	38
1gua	Rap	RafRBD	Misc. sign.	543.6	58	−10.1	39
1tx4	RhoA.GDP	P50RhoGAP	Misc. sign.	994.2	91	−7.1	40
1ak4	CypA	CA151	Misc.	452.1	46	−6.5	41
1efn	HIV-1 Nef	Fyn tyrosine kinase	Misc.	672.7	56	−8.8	42
1dkg	GrpE	DnaK	Misc.	656.3	67	−10.3	43
1ycs	P53	53BP2	Misc.	661.3	70	−10.3	44
1ebp	EPO receptor	EPO	Misc.	944.6	87	−11.7	45
1efu	EF-Tu	EF-Ts	Misc.	1791.8	143	−12.2	46
1hwg	HGH receptor	HGH	Misc.	2060.1	197	−13.0	47
1myk	ARC repressor	ARC repressor	Homodimer	2086.2	145	−10.0	48
1hvh	HIV-1 protease	HIV-1 protease	Homodimer	2165.5	162	−12.0	48
1run	cAMP rec. prot.	cAMP rec. prot.	Homodimer	1619.7	136	−12.0	48
1rti	TIM	TIM	Homodimer	2168.0	162	−14.3	49
1www	NGF	NGF	Homodimer	1245.9	133	−16.0	48

[†]Structures can be retrieved from the Protein Data Bank (www.rcsb.org).

[‡]P–Pi, protease–protease inhibitor; E–Ei, enzyme–enzyme inhibitor (other than P–Pi); Ab–Ag; antibody–antigen; misc. sign., miscellaneous signalling (other than Ab–Ag); misc., miscellaneous complexes.

[§]$\Delta G = -RT \ln k_{eq}$.

[‖]References refer to the article containing the experimental binding data: 1, Bolognesi et al. (1982); 2, Vincent and Lazdunski (1976); 3, Guinto et al. (1994); 4, Soliner et al. (1994); 5, Rees and Lipscomb (1982); 6, Baugh and Trowbridge (1972); 7, Tsunemi et al. (1996); 8, Wiedow et al. (1990); 9, Krystek et al. (1993); 10, Bode et al. (1989); 11, Takeuchi et al. (1991); 12, Keitaro (1985); 13, Green et al. (1984); 14, Read et al. (1983); 15, Degryse et al. (1989); 16, Lu et al. (1997); 17, Qasim et al. (1997); 18, Van de Locht et al. (1995); 19, Vincent and Lazdunski (1972); 20, Novotny et al. (1985); 21, Corin et al. (1991); 22, Bompard-Gilles et al. (1996); 23, Buckle et al. (1994); 24, Hartley (1993); 25, Kobe and Deisenhofer (1995); 26, Vicentini et al. (1990); 27, Schwarz et al. (1995); 28, Tello et al. (1993); 29, Malby et al. (1994); 30, Desmyter et al. (1996); 31, Chitarra et al. (1993); 32, Dall'Acqua et al. (1996); 33, Bhat et al. (1994); 34, Smallshaw et al. (1998); 35, Smith et al. (1991); 36, Padlan et al. (1989); 37, Hibbits et al. (1994); 38, Chaillan-Huntington et al. (2000); 39, Nassar et al. (1996); 40, Zhang and Zheng (1998); 41, Gamble et al. (1996); 42, Lee et al. (1996); 43, Harrison et al. (1997); 44, Gorina and Pavletich (1996); 45, Philo et al. (1996); 46, Zhang et al. (1998); 47, Clackson et al. (1998); 48, Neet and Timm (1994); 49, Rietveld and Ferreira (1998).

TABLE II. Hot Spot Data

PDB	Receptor	Ligand	Mutation	$\Delta\Delta G$(kcal/mol)[†]	ΔNo. of ligand atoms	$\Delta\Delta G$/ΔNo. of ligand atoms
1brs	Bamase	Barstar	D39A	−7.7	7	−1.1
1brs	Barnase	Barstar	K27A	−5.4	6	−0.9
1brs	Barnase	Barstar	R59A	−5.2	11	−0.5
1dfj	RnaseA	RI	Y430A	−5.9	6	−1.0
2ptc	Trypsin	BPTI	K15A	−10.0	3	−3.3

[†]Bogan and Thom[14] and references therein.

preliminary source of data for nucleic acid–ligand interactions.

Calculations

The equilibrium binding data collected from the literature can be related to the binding free energy ΔG_{bind} with the following equation:

$$\Delta G_{\text{bind}} = -RT \ln K_{\text{eq}}$$

Because the binding data used in this article come from different experiments in different research groups, the reference states of all the experiments might not be the same. For simplicity, we assume all experiments used a standard biochemical reference state.[1] We do not correct for any experimental differences. This assumption is not meant to imply that reference states are unimportant or that they can be in general ignored. They are unlikely, however, to impact the broad trends we are exploring in this article, where only the highest-affinity complexes are examined.

We can present the binding data in several ways. First, similarly to Kuntz et al.,[1] we can attribute the total binding free energy to the ligand atoms alone. We do this because it avoids calculating the contribution that each atom makes individually to the total binding free energy and allows comparisons where the 3D structures of the complexes are not available.

Second, we can make explicit use of the structural information to calculate the interface areas and the number of interface atoms. Especially in the case of protein ligands, most atoms do not contribute to the interface area, and these are less likely to contribute significantly to the binding free energy. Although a few noninterface hot spots have been found, these are exceptions and their contribution in general does not exceed 1.1 kcal/mol.[14] Thus, it is reasonable to focus on the atoms that actually form the interface, thereby testing the hypothesis that noninterface residues do not contribute significantly to the binding free energy.

Interface areas and the number of interface atoms were calculated with the program dms, a variant of the Connolly molecular surface algorithm[17] built into MidasPlus.[18] The interface area (IA) can be obtained by determining the accessible surface area (ASA) for the complex and its constitutive components followed by taking the following difference:

$$\text{IA} = (\text{ASA}_{\text{receptor}} + \text{ASA}_{\text{ligand}}) - \text{ASA}_{\text{complex}}$$

Dms was used with a probe radius of 1.4 Å, a density of 0.5, and all nonhydrogen atoms were included in the calculations. Interface atoms were defined as atoms that lose *any* ASA upon complex formation based on the dms output.

For the nucleic acid aptamers, the number of ligand atoms is defined as the number of heavy atoms in the small molecule ligand. For peptide ligands, both the backbone and side-chain heavy atoms were counted.

The number of "ligand" atoms for the DNA duplexes was taken as the number of heavy atoms *in each base for one strand* (cytosine = 8, thymine = 9, adenine = 10, guanine = 11). Sugars and phosphates were ignored in this calculation. No structural information about the DNA duplexes was included in this study.

RESULTS

In Figure 1, we plot the total binding free energy versus the number of ligand heavy atoms from Tables I–IV, where the number of ligand heavy atoms is either the total number of ligand heavy atoms or the number of interface ligand heavy atoms as defined in the Methods section. The "small protein inhibitor" data illustrate the initial linear increase of ΔG_{bind} with the number of ligand atoms and the formation of a maximal affinity plateau region beyond 15–20 ligand atoms as reported earlier.

The protein ligand data (first three entries in Fig. 1) indicate that the tightest binding macromolecular ligands clearly lie in the plateau region observed earlier. This is also illustrated in Figure 2, where ΔG_{bind}/ligand atoms decreases smoothly with increasing number of ligand heavy atoms for protein–ligand complexes.

The DNA duplexes, however, show additive behavior for the size range available for this study. The duplex interaction, per atom, is in general less than seen for the other systems (Fig. 2), even though only a fraction of the atoms are used for DNA duplexes. Most likely, this is a result of the energetic cost of restructuring the monomer upon duplex formation. While our simple accounting scheme for DNA duplexes certainly impacts the binding free energy per atom, it should not obscure the linear relationship between ΔG_{bind} and the number of DNA "interface" atoms. This behavior contrasts nicely with that of all the other macromolecular complexes.

RNA and DNA aptamers are nucleic acid sequences that bind small molecules or peptide ligands highly selectively (see Hermann and Patel[16] for a review). The interactions between the nucleic acid and the ligand are thus specific

TABLE III. DNA Duplex Unfolding Data

DNA sequence	Atoms in one strand	ΔG_{37}†	ΔG_{37}/atoms
Two-state thermodynamics			
CCGG	38	−3.52	−0.09
CGCG	38	−4.05	−0.11
CACAG	47.5	−3.62	−0.08
AGCCG	47.5	−5.50	−0.12
GTGAAC	57	−5.14	−0.09
CGTTGA	57	−5.20	−0.09
CGTCAG	57	−5.35	−0.09
GCATGC	57	−5.60	−0.10
CGTTGT	57	−5.65	−0.10
CCAACG	57	−5.68	−0.10
AGTTGC	57	−5.69	−0.10
TGTTGC	57	−5.79	−0.10
GCCTGC	57	−6.45	−0.11
GGGACC	57	−6.49	−0.11
CGGACG	57	−6.62	−0.12
ACCGCA	57	−6.70	−0.12
CGTGCC	57	−6.90	−0.12
TGCGCA	57	−6.90	−0.12
CGGTGC	57	−7.20	−0.13
ATGCGC	57	−7.30	−0.13
GCGAGC	57	−7.67	−0.13
GGCGCC	57	−7.93	−0.14
CCGCGG	57	−7.97	−0.14
CGCGCG	57	−8.31	−0.15
CGGCCG	57	−8.31	−0.15
GCCGGC	57	−8.51	−0.15
GCGCGC	57	−9.14	−0.16
CAAAAAG	66.5	−4.82	−0.07
GGACTTA	66.5	−5.60	−0.08
AATACCG	66.5	−5.90	−0.09
AGCTTCA	66.5	−6.10	−0.09
AGCCGTG	66.5	−8.50	−0.13
GAATATTC	76	−4.30	−0.06
GATTAATC	76	−4.30	−0.06
TCTATAGA	76	−4.33	−0.06
AAAAAAAA	76	−5.10	−0.07
TAGATCTA	76	−5.10	−0.07
GGTATACC	76	−5.50	−0.07
CAAAAAAG	76	−5.65	−0.07
ACATATGT	76	−5.70	−0.08
GAGTACTC	76	−5.80	−0.08
TATGCATA	76	−5.90	−0.08
ACCTAGTC	76	−6.50	−0.09
GTCTAGAC	76	−6.50	−0.09
GACTAGTC	76	−6.60	−0.09
GAAGCTTC	76	−6.70	−0.09
AATCCAGT	76	−6.80	−0.09
CCATATGG	76	−6.80	−0.09
CGATATCG	76	−6.88	−0.09
GTAGCTAC	76	−6.99	−0.09
GTACGTAC	76	−7.05	−0.09
CGATATCG	76	−7.30	−0.10
GATGCATC	76	−7.34	−0.10
AGAGAGAG	76	−7.40	−0.10
GGAATTCC	76	−7.40	−0.10
AGTCCTGA	76	−7.50	−0.10
CTAGTGGA	76	−7.50	−0.10
CATCGATG	76	−7.59	−0.10
TGAGCTCA	76	−7.73	−0.10
AGAGCTCT	76	−7.76	−0.10
AGCGTAAG	76	−7.80	−0.10
GCATATGC	76	−7.80	−0.10
CGCTGTAA	76	−7.90	−0.10
AAGCGTAG	76	−8.00	−0.11
CTGAGTCC	76	−8.00	−0.11
GCCAGTTA	76	−8.20	−0.11
TAGGCCTA	76	−8.20	−0.11
GTCGAACA	76	−8.30	−0.11
CGCGTATA	76	−8.40	−0.11
ACGACCTC	76	−8.70	−0.11
GGAGCTCC	76	−8.70	−0.11
GGACGTCC	76	−8.99	−0.12
ATGCGCAT	76	−9.00	−0.12
GGTCTCCA	76	−9.00	−0.12
GGACCTCG	76	−9.03	−0.12
GGAGCACG	76	−9.66	−0.13
CGTCGTCC	76	−9.67	−0.13
CGTCGACG	76	−9.79	−0.13
CTCACGGC	76	−9.80	−0.13
CACGGCTC	76	−10.00	−0.13
AAAAAAAAA	85.5	−6.20	−0.07
CAAATAAAG	85.5	−6.49	−0.08
CAAAAAAAG	85.5	−7.21	−0.08
CAAAGAAAG	85.5	−7.28	−0.09
CAAACAAAG	85.5	−7.74	−0.09
ATCTATCCG	85.5	−8.70	−0.10
GCCAGTTAA	85.5	−8.80	−0.10
ATAACTGGC	85.5	−9.00	−0.11
CGCTGTTAC	85.5	−9.90	−0.12
AAAAAAAAAA	95	−6.70	−0.07
TAGGTTATAA	95	−7.00	−0.07
ATGAGCTCAT	95	−10.00	−0.11
GCGAAAAGCG	95	−11.86	−0.12
GCGAATTCGC	95	−12.91	−0.14
CGGCAAGCGC	95	−13.30	−0.14
Marginally two-state thermodynamics			
CCATTGCTACC	104.5	−12.17	−0.12
CCATCGCTACC	104.5	−13.30	−0.13
ACGTATTATGC	104.5	−10.40	−0.10
ATTGGATACAAA	114	−10.30	−0.09
GGAACTTGATGC	114	−12.85	−0.11
GGAACAAGATGC	114	−12.85	−0.11
CTTCCTTCCTTC	114	−13.20	−0.12
CAACCAACCAAC	114	−13.75	−0.12
CGCGAATTCGCG	114	−16.52	−0.14
ACATTATTATTACA	133	−11.30	−0.08
CAACTTGATATTAATA	152	−12.37	−0.08

†Allawi and SantaLucia.[15]

and we might expect the binding free energy per ligand heavy atom to be quite high. Figures 1 and 2 show that that does not seem to be the case, perhaps because these nucleic acid sequences are unstructured without their ligands, and ligand binding thus involves also refolding of the nucleic acid,[16] similar to DNA duplex formation.

TABLE IV. Nucleic Acid Aptamer Binding Data

Ligand	No. of ligand heavy atoms	ΔG (kcal/mol)[†]	ΔG/No. of ligand atoms
RNA aptamers			
Theophylline	13	−8.90	−0.68
FMN	31	−8.60	−0.28
AMP	23	−6.82	−0.30
Arginine	12	−5.76	−0.48
Citrulline	12	−5.71	−0.48
Tobramycin	31	−10.98	−0.35
Neomycin B	42	−9.47	−0.23
HIV-1 Rev peptide	172	−11.46	−0.07
HTLV-1 Rex peptide	147	−10.37	−0.07
DNA aptamers			
AMP	23	−7.12	−0.31
Arginine	12	−5.32	−0.44

[†]Hermann and Patel[16] and references therein.

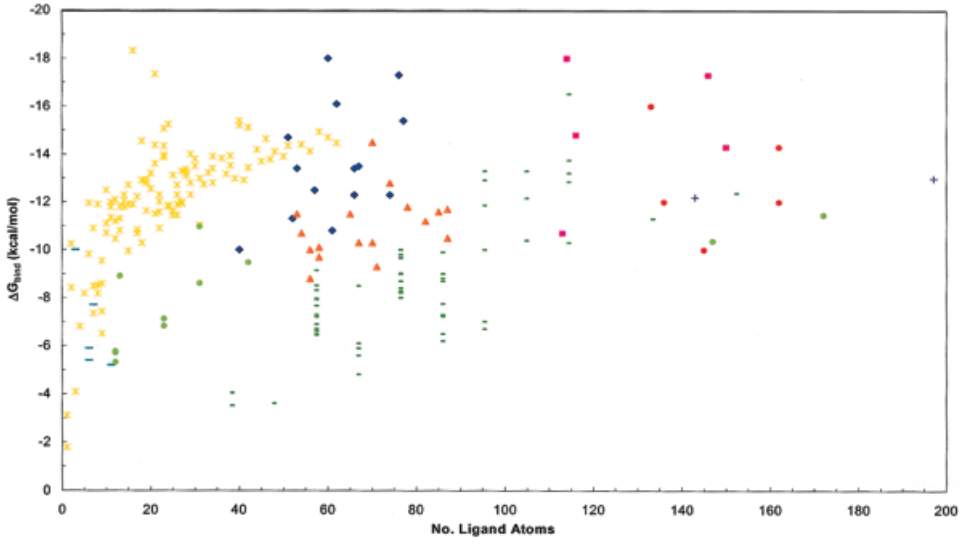

Fig. 1. (**A**) Relationship between the binding free energy and the number of ligand atoms in the interface. Blue diamonds, protein–inhibitor complexes (E–Ei & P–Pi in Table I), average size interface; pink squares, large protein–inhibitor complexes; orange triangles, signaling complexes (Ab–Ag & misc. sign. in table I), average size interface; yellow stars, protein–low-molecular-weight inhibitors (from Kuntz et al.[1]); red circles, homodimeric complexes; blue plus signs, large signaling complexes; small green dashes, DNA duplexes; large blue dashes, hot spots; green circles, nucleic acid aptamers. For clarity, some of the lower-lying points have been omitted.

Another expectation might be that hot spots, small regions in protein–protein interfaces that contribute a disproportionate amount of energy to the total binding free energy, have unusual free energy/atom values. However, they do not seem to exceed the 1.5 kcal/mol per heavy atom (Table II), which was previously found as a limit for tight ligand interactions.[1] One exception is the K15A mutation in the BPTI–trypsin complex, which has a contribution of over 3 kcal/mol per ligand heavy atom. The K15A mutation in the BPTI–trypsin complex is among the most detrimental mutations observed so far.[14]

We can also compare the binding free energy and ΔG/ligand atoms of the different classes of protein–protein complexes. The maximal affinities of protein–inhibitor complexes and signaling complexes, within the same interface size range, indicate that both the total binding free energy (Fig. 1) and the free energy per ligand heavy atom (Fig. 2) are higher for protein–inhibitor complexes, offering support to the hypothesis that the evolutionary pressures are different for these classes.[5] Averages and the variance for both the total binding free energy and the contribution of each ligand atom to complex stability are reported in Table V.

On the other hand, the expectation that homodimeric complexes have a higher maximal affinity than heterodimeric complexes does not seem to hold true (Fig. 1). A possible explanation is that the dissociation constants for more permanent, and thus more stable, homodimeric complexes are more difficult to measure accurately with conventional experimental methods.[19]

If different interface sizes (e.g., enzyme–inhibitor vs. large enzyme–inhibitor complexes) within the same bio-

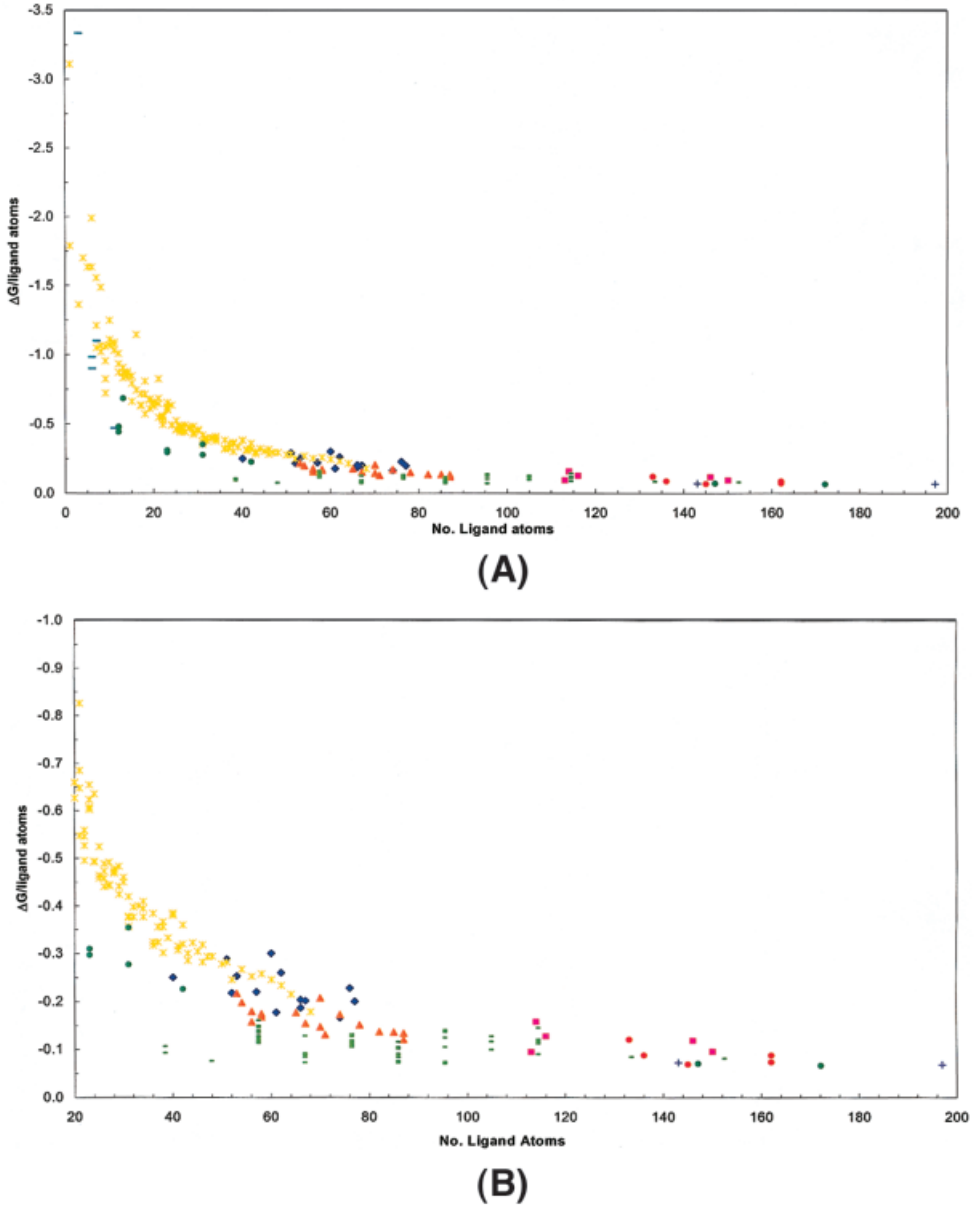

Fig. 2. Relationship between the binding free energy per ligand atom and the number of ligand atoms in the interface. (**A**) See Figure 1(A). (**B**) Same as Figure 2(A), but the points with less than 20 interface ligand atoms have been omitted to illustrate the tail of the curve in Figure 2(A). See Figure 1(A) for symbol/color definitions.

logic class are compared, it can be seen in Figure 2 and Table V that the maximal contribution per ligand interface atom is less for the "large" complexes compared to the average-sized ones, clearly illustrating the "plateau" behavior discussed earlier.

DISCUSSION

In this article, we investigated the thermodynamic properties of macromolecular ligands and compared them to earlier work on low-molecular-weight ligands.[1] What we have shown here is that the maximal stability of macromolecular complexes lies in a plateau region, where there is little, if any, gain in stability as the size of the macromolecular partners increases. Again, DNA duplexes are notable exceptions.

Only plots where the interface size is represented by the number of ligand interface atoms are shown here, but the plots where interface size is measured as ASA are nearly identical (Brooijmans and Kuntz, unpublished results). The conversion factor is approximately 12 Å^2 of molecular surface per atom.

Our results disagree with earlier work on protein–protein interactions that reported a positive correlation between change in ASA upon binding and complex stability for proteins that associate as rigid objects.[20] While this relationship is a plausible one, our more extensive data do

TABLE V. Values for the Average Binding Free Energy and Average Ligand Contribution to Complex Stability for the Different Classes of Macromolecular Complexes

Type of complex[†]	Average ΔG (kcal/mol)[‡]	Average ΔG/ (no. of ligand heavy atoms)[‡]
Protein–inhibitor	−12.4 (13.4)	−0.20 (0.004)
Large protein–inhibitor	−15.0 (8.3)	−0.12 (0.0007)
Signaling	−10.5 (3.4)	−0.16 (0.001)
Large signaling	−12.6 (0.3)	−0.08 (0.0002)
Homodimeric	−12.9 (5.4)	−0.09 (0.0004)
DNA duplexes	−7.8 (5.9)	−0.10 (0.0005)

[†]The number of complexes for each class: protein–inhibitor, 17; large protein–inhibitor, 5: signaling, 19: large signaling, 2: homodimer, 5: DNA duplexes, 106.
[‡]Averages are taken over all the data available in the appropriate tables. Variances are in parentheses.

not support this linear relationship. We believe that there are several explanations for this discrepancy. First, as noted, much more data are available. Second, we now expect such linearity to occur at a size (50 ligand interface atoms or below) where little peptide/protein data are available, although we do have information on the interactions of proteins with low-molecular-weight ligands in this regime. Third, we can estimate the per-atom contribution from the plot in Figure 4 in the Horton and Lewis article.[21] We find a contribution of 125 cal/mol per atom (1 atom $\approx$12.5 Å^2 of solvent-accessible surface). This is well below our limiting slope of 1.5 kcal/mol per atom. This underscores our conclusion that the data points of Horton and Lewis[21] lie in the plateau region we observe here. Fourth, from a functional point of view, it is plausible that there is a maximal useful affinity for ligands because a dissociation time of several years for a complex might be undesirable. Thus, although thermodynamically speaking, there is no obvious necessity for the nonadditive behavior observed in this work, we can certainly advance a kinetic rationale.

The plateau in affinity with ligand size necessarily implies a diminishing return from each additional atom (Fig. 1). While it appears that there is no necessary thermodynamic reason for this plateau, the qualitative difference in behavior between nucleic acid duplexes and all the other classes of complexes provides an interesting control. DNA duplexes can be viewed as linear strings of symmetrical elements in distinction to most protein–protein interfaces. As a consequence, for the linear duplex structures beyond a few base pairs in length, addition of bases at the ends has little physical impact on the environment of interface atoms more than a couple of base pairs from the ends. In particular, their interactions with solvent remain strong because DNA is a rather open structure. In contrast, atoms in the center of a large protein–protein interface tend to be further from the solvent than in a smaller interface. Because such interfaces usually contain a significant polar portion,[8] this suggests some kind of general compensation as one increases interface area, between addition of new interactions from peripheral atoms and weakened interactions due to increased desolvation from interior interface atoms.

Although an increase in the desolvation penalty can suggest a physical basis for the lower contribution per atom with increasing interface size, it cannot explain why different biologic classes with similar interface sizes have different maximal affinities (Fig. 1). Instead, this latter observation suggests that different evolutionary pressures may lead to different maximal affinities, although there is overlap between the different biologic classes. Part of this overlap might be due to the inclusion of "nonnatural" complexes that have not evolved together as a complex, but can still bind to each other. Because the interfaces for the different classes are similar in size but show different maximal affinities, this raises the question of how these different maximal affinities are being established. Further investigation of the atomic details of protein–protein interfaces seems to be necessary. It is certainly desirable to be able to design small molecules to specifically inhibit protein–protein interactions. Based on the similar total affinities of the strongest small molecule ligand with high-affinity protein–protein interactions, it is clearly feasible to consider the design of protein–protein interaction inhibitors.

The finding that atoms of low-molecular-weight ligands contribute more than atoms of macromolecular ligands on a per-atom basis has two interesting implications from a design standpoint. First, it implies that more stable protein–protein complexes could be designed. This is supported by the ability to obtain better than wild-type affinity in protein–protein complexes using phage display in vitro evolution.[22] Second, it suggests that the design of small-molecule inhibitors of macromolecular complexes is feasible, as mentioned before.

We realize that counting the number of interface atoms, instead of using the accessible surface area directly, introduces systematic error. After all, in this method every atom that loses any amount of ASA, no matter how small or large, is counted as an interface atom, whether its contribution to the binding free energy is negligible or significant. We mainly adopted this approach so we could directly compare our results on macromolecular ligands to low-molecular-weight ligands, for many of which no structural data are available. Further, this approach leads to an amount of binding free energy per heavy atom, which provides an easy way of thinking about energy contributions. Despite these concerns, the plots for ΔG_{bind} versus ASA are virtually identical to those shown here (Brooijmans and Kuntz, unpublished results), as mentioned above. Thus, the major conclusions we draw would be the same.

It is also interesting to compare the average per-atom contribution to affinity to that seen in protein folding. On a per-residue basis, a typical globular protein has a stability of 0.1 kcal/mol per residue.[23] Averaged over all 20 amino-acid types, the number of side-chain heavy atoms is about 6, and if 50% of them are buried on average the number of atoms buried in protein folding averages about 3 per residue. Clearly, if each atom contributed at the level of maximal affinity attainable by atoms in binding, ca. 1.5 kcal/mol, this would provide about 50 times the observed

stability. Even accounting for the opposing effects of configurational entropy, estimated to be about 4 e.u. per residue,[23–26] this appears to overestimate the contribution. On the other hand, if the median value from Figure 2 of about 0.25 kcal/mol is used it is easier to reconcile with the observed protein folding energies. Again, one concludes that, in principle, globular protein stability could be much larger, but nonthermodynamic reasons (evolutionary, functional, protein lifetime, etc.) limit the average contribution per atom.

The idea that proteins can be readily made more stable has been successfully explored by several research groups using either design methods[27,28] or fast in vitro evolution methods.[29,30]

We would like to emphasize that we do not attempt to predict the binding free energies of specific protein–protein complexes or explain the physical basis for protein–protein complex stability. We only investigated the relationship between the maximal affinity of macromolecular ligands and the interface size, measured by the number of ligand heavy atoms in the interface.

One question that still remains to be answered is whether it is possible to assess the "quality" of a binding site. We and others observed that protein interaction sites are indistinguishable from the rest of the protein surface in terms of fractality[2] or curvature (Brooijmans and Kuntz, unpublished results). Others have tried to predict protein binding sites with a combination of structural and physicochemical properties,[31,32] but with mixed successes for the different biologic classes of protein interaction sites. Investigation of residue conservancy has revealed that, in homodimers at least, interface surface residues are more conserved than noninterface surface residues.[33] If this holds true for heterodimeric interfaces as well, than perhaps a combination of sequence conservation analysis, structural analysis, and physicochemical properties analysis of protein surfaces might lead to more accurate predictions of potential protein interaction sites.

CONCLUSION

In summary, we have shown here that there is a maximal affinity for macromolecular ligands, like there is for low-molecular-weight ligands, resulting in different maximal binding free energy contributions per ligand heavy atom. We also observed that current data support the view that different biologic classes have different maximal affinities, perhaps due to different evolutionary pressures. These findings also have interesting implications for molecular design.

ACKNOWLEDGMENTS

The authors thank Peter Kollman for helpful discussions. N.B. acknowledges UCSF for a Graduate Dean's Health Science Fellowship for support. I.D.K. thanks the National Institute of General Medical Sciences for support of this work. K.A.S. thanks the National Science Foundation for support.

REFERENCES

1. Kuntz ID, Chen K, Sharp KA, Kollman PA. The maximal affinity of ligands. Proc Natl Acad Sci USA 1999;96:9997–10002.
2. Pettit FK, Bowie JU. Protein surface roughness and small molecular binding sites. J Mol Biol 1999;285: 1377–1382.
3. Argos P. An investigation of protein subunit and domain interfaces. Protein Eng 1988;2:101–113.
4. Janin J, Miller S, Chothia C. Surface, subunit interfaces and interior of oligomeric proteins. J Mol Biol 1988;204:155–164.
5. Jones S, Thornton JM. Protein–protein interactions: a review of protein dimer structures. Prog Biophys Mol Biol 1995;63:31–65.
6. Jones S, Thornton JM. Principles of protein–protein interactions. Proc Natl Acad Sci USA 1996;93:13–20.
7. Miller S. The structure of interfaces between subunits of dimeric and tetrameric proteins. Protein Eng 1989;3:77–83.
8. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R. Studies of protein–protein interfaces: a statistical analysis of the hydrophobic effect. Protein Sci 1997;6:53–64.
9. Clackson T, Wells JA. A hot spot of binding energy in a hormone–receptor interface. Science 1995;267:383–386.
10. Cunningham BC, Wells JA. JComparison of a structural and a functional epitope. J Mol Biol 1993;234:554–563.
11. Sheinerman FB, Norel R, Honig B. Electrostatic aspects of protein–protein interactions. Curr Opin Struct Biol 2000;10:153–159.
12. LoConte L, Chothia C, Janin J. The atomic structure of protein–protein recognition sites. J Mol Biol 1999;285:2177–2198.
13. Neet KE, Timm DE. Conformational stability of dimeric proteins: quantitative studies by equilibrium denaturation. Protein Sci 1994;3:2167–2174.
14. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol 1998;280:1–9.
15. Allawi HT, SantaLucia JJ. Thermodynamics and NMR of internal GT mismatches in DNA. Biochemistry 1997;36:10581–10594.
16. Hermann T, Patel DJ. Stitching together RNA tertiary architectures. J Mol Biol 1999;294:829–849.
17. Connolly ML. Analytical molecular surface calculation. J Appl Crystallogr 1983;16:548–558.
18. Ferrin TE, Huang CC, Jarvis LE, Langridge R. The Midas display system. J Mol Graph 1988;6:13–27.
19. Kaplan AP, Bartlett PA. Synthesis and evaluation of an inhibitor of carboxypeptidase A with a K_i value in the femtomolar range. Biochemistry 1991;30:8165–8170.
20. Chothia C, Janin J. Principles of protein–protein recognition. Nature 1975;248:338–339.
21. Horton N, Lewis M. Calculation of the free energy of association for protein complexes. Protein Sci 1992;1:169–181.
22. Dalby PA, Hoess RH, DeGrado WF. Evolution of binding affinity in a WW domain probed by phage display. Protein Sci 2000;9:2366–2376.
23. Privalov PL, Gill SJ. Stability of protein structure and hydrophobic interaction. Adv Protein Chem 1988;39:191–234.
24. Murphy KP, Freire E. Thermodynamics of structural stability and cooperative folding behavior in proteins. Adv Protein Chem 1992; 43:313–361.
25. Privalov PL. Thermodynamics of protein folding. J Chem Thermodyn 1997;29:447–474.
26. Baldwin RL. Temperature dependence of the hydrophobic interaction in protein folding. Proc Natl Acad Sci USA 1986;83:8069–8072.
27. Shoichet BK, Baase WA, Kuroki R, Matthews BW. A relationship between protein stability and protein function. Proc Natl Acad Sci USA 1995;92:452–456.
28. Marshall SA, Mayo SL. Achieving stability and conformational specificity in designed proteins via binary patterning. J Mol Biol 2001;305:619–631.
29. Powell SK, et al. Breeding of retrovirused by DNA shuffling for improved stability and processing yields. Nat Biotech 2000;18: 1279–1282.
30. Sun L, Petrounia IP, Yagasaki M, Bandara G, Arnold FH. Expression and stabilization of galactose oxidase in *Escherichia coli* by directed evolution. Protein Eng 2001;14:699–704.
31. Jones S, Thornton JM. Prediction of protein–protein interaction sites using patch analysis. J Mol Biol 1997;272:133–143.
32. Jones S, Thornton JM. Analysis of protein–protein interaction sites using surface patches. J Mol Biol 1997;272:121–132.
33. Valdar WSJ, Thornton JM. Protein–protein interfaces: analysis of amino acid conservation in homodimers. Proteins 2001;42:108–124.