Simple, Intuitive Calculations of Free Energy of Binding for Protein−Ligand Complexes. 1. Models without Explicit Constrained Water

Pietro Cozzini,[†] Micaela Fornabaio,[‡,§] Anna Marabotti,[‡,||] Donald J. Abraham,[⊥] Glen E. Kellogg,*[,⊥] and Andrea Mozzarelli*[,‡,§]

Department of General and Inorganic Chemistry, Department of Biochemistry and Molecular Biology, National Institute for the Physics of Matter, University of Parma, 43100 Parma, Italy, and Department of Medicinal Chemistry & Institute for Structural Biology and Drug Discovery, School of Pharmacy, Virginia Commonwealth University, Richmond, Virginia 23298-0133

Received January 22, 2002

The prediction of the binding affinity between a protein and ligands is one of the most challenging issues for computational biochemistry and drug discovery. While the enthalpic contribution to binding is routinely available with molecular mechanics methods, the entropic contribution is more difficult to estimate. We describe and apply a relatively simple and intuitive calculation procedure for estimating the free energy of binding for 53 protein−ligand complexes formed by 17 proteins of known three-dimensional structure and characterized by different active site polarity. HINT, a software model based on experimental Log$P_{o/w}$ values for small organic molecules, was used to evaluate and score all atom−atom hydropathic interactions between the protein and the ligands. These total scores (H_{TOTAL}), which have been previously shown to correlate with $\Delta G_{interaction}$ for protein−protein interactions, correlate with $\Delta G_{binding}$ for protein−ligand complexes in the present study with a standard error of ±2.6 kcal mol^{-1} from the equation $\Delta G_{binding} = -0.001\ 95\ H_{TOTAL} - 5.543$. A more sophisticated model, utilizing categorized (by interaction class) HINT scores, produces a superior standard error of ±1.8 kcal mol^{-1}. It is shown that within families of ligands for the same protein binding site, better models can be obtained with standard errors approaching ±1.0 kcal mol^{-1}. Standardized methods for preparing crystallographic models for hydropathic analysis are also described. Particular attention is paid to the relationship between the ionization state of the ligands and the pH conditions under which the binding measurements are made. Sources and potential remedies of experimental and modeling errors affecting prediction of $\Delta G_{binding}$ are discussed.

Introduction

The exponential increase of three-dimensional structures of proteins determined by X-ray crystallography has been the most important driving force for the structure-based drug design paradigm. The search for new drugs by targeting pharmaceutically relevant proteins requires both the identification of lead compounds and an estimation of their binding affinities. To this end, a collection of potential ligands is subjected to some form of "virtual screening". This is achieved by the use of one of several commonly available docking algorithms based on approaches that, in effect, systematically move the putative ligands into proposed positions and then "score" each position.[1,2] Ultimately, the best "scoring" ligand position/orientation is chosen as the final "docked" structure for the ligand. The best ligands from the collection may then be subjected to further experimental evaluation and, if promising, to drug design and development cycles. The state of the art in docking and scoring has recently been reviewed.[3−7] Typically, these scoring methodologies are reporting enthalpy-related parameters. However, the evaluation of the binding free energy is a much more difficult problem due to a variety of issues. Probably foremost among these is accounting for entropic effects arising from water molecules that are initially bound at both the isolated ligand and the protein binding site but released upon the formation of the complex. This phenomenon is most of the basis for the hydrophobic interaction.

Several computational strategies have been proposed for quantitatively describing the formation of complexes between a protein and ligand(s).[8−12] Most programs of this type utilize Newtonian physics and molecular mechanics force field parametrization to calculate atom−atom interaction forces. Hydrophobicity, when included at all, is generally evaluated by considering hydrophobic surface contact area as a term in the scoring function. In contrast, the program HINT (Hydropathic INTeractions), is based on parameters derived from experimentally determined solvent partition Log*P* measurements between water and 1-octanol. This program was developed to highlight hydrophobic interactions,[13,14] but it can be applied to a wide variety of biological problems.[15] For example, it has been used as a docking score calculator to locate ligands within a protein active site,[16] to investigate sequence specificity for doxorubicin antibi-

* To whom correspondence should be addressed. A.M.: Tel.: +39-0521-905138. Fax: +39-0521-905151. E-mail: biochim@unipr.it. G.E.K.: Tel.: +1-804-828-6452. Fax: +1-804-827-3664. E-mail: glen.kellogg@vcu.edu.

[†] Department of General and Inorganic Chemistry, University of Parma.

[‡] Department of Biochemistry and Molecular Biology, University of Parma.

[§] National Institute for the Physics of Matter, University of Parma.

[||] Present address: CRIBI Biotechnology Centre, University of Padua, 35121 Padua, Italy.

[⊥] Virginia Commonwealth University.

10.1021/jm0200299 CCC: $22.00
Published on Web 05/04/2002

otic intercalation in DNA,[17,18] to aid in the design of potent cyclin-dependent kinase inhibitors,[19] and to predict the free energy changes associated with native and mutant hemoglobin dimer−tetramer assembly.[20−22] HINT has been shown in these and other studies to give intuitively reasonable interaction models with quantitatively useful scores.

In the present study, we evaluate whether HINT is able to effectively predict the binding free energy of a wide variety of proteins that interact with either hydrophobic or polar ligands. In a previous work, we analyzed the binding free energy for three ligands of retinol binding protein and found an excellent correlation with the experimental dissociation constants.[23] In this paper, we optimize molecular models without explicit and constrained water molecules in the binding region (vide infra); a later paper will describe the modeling of bridging water molecules for these and other complexes and their effect on the free energy of binding.

Results

Hydropathic Analysis. The program HINT calculates a "score" for each atom−atom interaction in a biomolecular association. The basis of the HINT model is that quantitatively significant data of biomolecular association are encoded in the experimental determination of hydrophobicity, particularly from the water−octanol system. The constant, $\mathrm{Log}P_{o/w}$, is a thermodynamic quantity representing the free energy of solvent transfer for partitioning between the two solvents. As such, it includes the effects of entropy,[15] solvation,[24] and enthalpic terms such as hydrogen bonding, Coulombic attractions, and hydrophobic attractions. Each atom−atom score is a partial δg that has a character representing the type of interaction (hydrogen bond, hydrophobic, acid−base, base−base, acid−acid, or hydrophobic−polar) and a magnitude representing the potency of the interaction. The sum of atom−atom scores for an association represents the total strength of the interaction. We have shown that these total scores can be correlated with $\Delta G_{\mathrm{interaction}}$ or with $\Delta\Delta G$ for a series of related biomolecules.[22,23,25]

Molecular Models. A database was created of 210 ligand−protein complexes for which the three-dimensional structure and binding affinity were determined. (Protein data bank (PDB) codes of structures in initial database were as follows: 1a50, 1aaq, 1abe, 1abf, 1adb, 1adf, 1adl, 1aid, 1apb, 1apv, 1apw, 1b0h, 1b1h, 1b2h, 1b3h, 1b40, 1b4h, 1b5h, 1b60, 1b61, 1b6h, 1b6j, 1b6k, 1b6m, 1b6n, 1b6p, 1b7h, 1ba8, 1bap, 1bb0, 1bmm, 1bmn, 1bra, 1bxo, 1bxq, 1c29, 1c8v, 1c9d, 1ca8, 1cw2, 1cx9, 1d3d, 1d3p, 1d3q, 1d3t, 1d4p, 1dbb, 1dbj, 1dbk, 1dbm, 1dih, 1dog, 1dwb, 1dwc, 1dwd, 1dwe, 1ebg, 1eed, 1ent, 1epl, 1epm, 1epn, 1epo, 1epp, 1epq, 1epr, 1er8, 1erb, 1etr, 1ets, 1ett, 1fel, 1fen, 1fq4, 1fq5, 1fq6, 1fq7, 1hbv, 1hpb, 1hpv, 1hsl, 1htf, 1htg, 1hvi, 1hvj, 1hvk, 1hvl, 1hvr, 1jet, 1kce, 1lca, 1lcb, 1lce, 1lgr, 1lhc, 1lhd, 1lhe, 1lhf, 1lhg, 1lid, 1lie, 1lif, 1lyb, 1mbi, 1mcb, 1mcf, 1mch, 1mcj, 1mcs, 1mfe, 1nrn, 1nro, 1nrp, 1nrq, 1nrr, 1nrs, 1nsc, 1nsd, 1pgp, 1phf, 1phg, 1ppc, 1pph, 1ppk, 1ppl, 1ppm, 1qb0, 1qb1, 1qb6, 1qb9, 1qbn, 1qop, 1qrp, 1sre, 1srf, 1srg, 1srh, 1sri, 1srj, 1tdu, 1thl, 1tlc, 1tlp, 1tls, 1tmn, 1tmt, 1tng, 1tnh, 1tni, 1tnj, 1tnk, 1tnl, 1trg, 1tsd, 1tsn, 1ulb, 1uvs, 1uvt, 2cgr, 2cpp, 2dbl, 2er0, 2er6, 2er7, 2er9, 2gbp, 2ifb, 2olb, 2pph, 2tmn, 2trs, 2tsc, 2tsc, 2tsy, 2xis, 2yas, 2ypi, 3er1, 3er3, 3er3, 3er5, 3ptb, 3yas, 4apr, 4dfr, 4er1, 4er2, 4er4, 4hmg, 4hvp, 4tln, 4tln, 4tmn, 5abp, 5apr, 5cna, 5cpp, 5er2, 5hvp, 5tmn, 5yas, 6abp, 6apr, 7abp, 7dfr, 7hvp, 8abp, 8tln, 9abp, and 9hvp.) Proteins represented in the database are characterized by a wide difference of active site polarity and ligand polarity and chemical structures. A grid of selection conditions was imposed to create a "homogeneous" subset. These conditions are (i) the absence of metal ions in the binding pocket; (ii) the absence of a third ligand that may cooperate in ligand−protein interaction, considering water molecules bridging ligand and protein a special case, and (iii) crystallographic resolution lower than 3.2 Å. Only 53 complexes fulfill these multiple conditions and include 17 proteins forming either a single ligand−protein complex or a system in which the same protein binds three or more ligands (Table 1). The corresponding experimentally determined binding affinities vary over about 9 orders of magnitude (Table 1).

Our protocol for preparing protein−ligand complex structures for the analyses is described, in detail, in the Methods and Materials section of this paper. Before analysis, each PDB file was read into Sybyl. The atom potential types bond orders were carefully checked to evaluate their correctness with respect to the intended structure. The lack of bond order records in the PDB format necessitates that this step be diligently performed before adding hydrogen atoms to the ligand structure. Hydrogen atoms were also added to the protein structure.

We developed a set of rules that were applied throughout the remainder of the model-building process. First, our procedures rarely affected heavy atom (i.e., nonhydrogen) positions as described by crystallographic results. In rare circumstances, described in the following subsections, steric problems or unusual ligand geometries required some energy minimization/structure optimization of all atoms. Generally, however, optimizations are performed on only the hydrogens of the complexes. Second, we manually rotate as necessary $-NH_2$ and $-OH$ groups, e.g., in threonine, serine, and tyrosine residues, as well as on the ligands, to maximize the ligand-to-protein hydrogen bonding. Because hydrogen atoms are only (very) rarely located in biomacromolecular crystallographic studies, there is little experimental guidance as to their actual positions. Also worth noting is that the automated procedures to add hydrogens to biopolymers and small molecules, as in Sybyl or other software programs, often (and randomly) orient the hydrogen-bonding hydrogen atoms away from their intermolecular acceptor atoms. Energy minimization procedures are usually not a remedy for these cases because of local minima in the energy profiles that "pin" the hydrogen atom in the wrong conformation. Third, we carefully examine the ionization state of acid and base moieties to determine the most likely model. While assuming standard ionization states is a good starting point, that may not always be the actual situation in local regions of the complexes. There are limited experimental data on this topic. In particular, the penicillopepsin system that was characterized by Paul

Table 1. Ligand Binding Data and HINT Score Results for 53 Protein−Ligand Complexes

pdb	protein	lig	K_i (μM)	ref	ΔG (kcal/mol)	HS (ess)	H_{HB+AB}[d]	H_{HH}[d]	H_{AA}[d]	H_{BB}[d]	H_{HP}[d]	HS (all)
1ETS	bovine thrombin	TRB1	6.00×10^{-3}	32	−11.17	2692	5670	1222	−55	−2736	−1410	3623
1ETT	bovine thrombin	TRB2	1.3×10^{0}	32	−8.00	679	3689	992	−97	−1949	−1955	2131
1ETR	bovine thrombin	TRB3	19×10^{-3}	33	−10.49	1051	4842	1348	−85	−3017	−2038	2848
1UVT	bovine thrombin	TRB4	23×10^{-3}	29	−10.38	2170	2723	1294	−50	−1101	−702	1834
1D3T	human thrombin	TRH1	0.374×10^{0}	34	−8.73	1730	1305	1227	0	−208	−595	1660
1D3Q	human thrombin	TRH2	3.43×10^{0} [a]	30	−8.88	1714	1106	1414	0	−92	−714	1734
1D3P	human thrombin	TRH3	9.00×10^{-3}	34	−10.87	1864	2280	1388	−18	−839	−948	2094
1D3D	human thrombin	TRH4	1.24×10^{3} [a]	30	−12.35	2101	2549	1520	−19	−917	−1033	2339
1DWB	human thrombin	TRH5	1.24×10^{3}	32	−3.95	1606	4370	141	−61	−2081	−762	2323
1DWB	human thrombin*	TRH5	1.24×10^{3}	32	−3.95	743	3532	333	−57	−2444	−622	909
2YAS	hydroxynitrile lyase	HNL1	5.5×10^{0}	53	−7.14	1735	2115	670	0	−453	−615	1988
5YAS	hydroxynitrile lyase	HNL2	0.55×10^{3}	53	−4.43	27	1707	589	−624	−316	−1329	1012
3YAS	hydroxynitrile lyase	HNL3	30×10^{3}	53	−2.07	1163	1154	851	0	−145	−697	1206
1ADL	adipocyte lipid-binding protein	ALB1	0.18×10^{0} [b]	58	−9.16	2859	3084	1339	0	−324	−1240	3107
1LIE	adipocyte lipid-binding protein	ALB2	83×10^{-3} [b]	58	−9.62	1840	2503	1591	0	−296	−1959	3001
1LID	adipocyte lipid-binding protein	ALB3	58×10^{-3} [b]	58	−9.83	2615	3026	1511	0	−295	−1626	3486
1LIF	adipocyte lipid-binding protein	ALB4	80×10^{-3} [b]	58	−9.64	2191	2750	1648	0	−279	−1928	3445
1HBP	retinol binding protein	75	70×10^{-3} [b]	23	−9.72	285	224	1987	−2	−88	−1837	932
1ERB	retinol binding protein	76	90×10^{-3} [b]	23	−9.57	701	290	2221	−1	−216	−1593	816
1FEL	retinol binding protein	77	0.17×10^{0} [b]	23	−9.19	311	771	2326	−385	−112	−2289	488
1TNJ	bovine trypsin	TPB1	11×10^{3}	35	−2.66	−133	1839	359	−92	−1356	−883	677
1TNK	bovine trypsin	TPB2	32.5×10^{3}	35	−2.02	−350	1599	413	−89	−1230	−1044	720
1TNI	bovine trypsin	TPB3	20×10^{3}	35	−2.30	−364	1698	443	−89	−1311	−1105	834
1TNL	bovine trypsin	TPB4	13.3×10^{3}	35	−2.54	527	2333	400	−50	−1349	−806	1360
1TNG	bovine trypsin	TPB5	1.17×10^{3}	35	−3.98	−78	1840	410	−34	−1393	−901	923
1TNH	bovine trypsin	TPB6	0.43×10^{3}	35	−4.57	755	2275	295	−37	−1320	−458	972
3PTB	bovine trypsin	TPB7	18.4×10^{0}	38	−6.43	2025	4611	152	−205	−1955	−579	2597
3PTB	bovine trypsin*	TPB7	18.4×10^{0}	38	−6.43	1413	4091	327	−164	−2368	−472	1634
1PPH	bovine trypsin	TPB8	1.2×10^{0}	37	−8.04	1652	4976	652	−221	−2576	−1179	2663
1CX9	tryptophan synthase	TS1	0.18×10^{0} [c]	43	−9.58	1129	3850	1347	−205	−1512	−2351	2595
1C29	tryptophan synthase	TS2	0.47×10^{0} [c]	43	−9.00	1529	4366	1130	−214	−1745	−2008	2793
1C9D	tryptophan synthase	TS3	0.50×10^{0} [c]	43	−8.97	1548	4290	1359	−214	−1586	−2301	3094
1CW2	tryptophan synthase	TS4	0.72×10^{0} [c]	43	−8.76	1374	4563	1389	−193	−1910	−2474	3094
1C8V	tryptophan synthase	TS5	0.54×10^{0} [c]	43	−8.92	1160	3733	1221	−168	−1565	−2061	2571
2TRS	tryptophan synthase (K87T)	TS6	5×10^{0}	47	−7.20	1624	3814	1320	−59	−1456	−1996	2646
1QOP	tryptophan synthase	TS6	5×10^{0}	47	−7.20	1538	3645	1524	−68	−1399	−2163	2721
1A50	tryptophan synthase	TS7	0.50×10^{0} [b]	41	−8.56	1874	3916	1457	−60	−1425	−2013	2914
2TSY	tryptophan synthase (K87T)	TS8	0.38×10^{3} [b]	48	−4.65	−645	2865	765	−172	−1821	−2281	905
1BXQ	penicillopepsin	PPA1	42×10^{-3}	49	−10.02	872	7392	2178	−423	−4378	−3897	4294
1BXO	penicillopepsin	PPA2	0.10×10^{-3}	49	−13.59	1326	7005	2151	−375	−4043	−3413	4435
1PPL	penicillopepsin	PPA3	2.8×10^{-3}	26	−11.62	1022	6763	2032	−322	−4216	−3235	3995
1PPM	penicillopepsin	PPA4	0.19×10^{0}	26	−9.13	1688	7120	1978	−290	−4038	−3081	3908
1PPK	penicillopepsin	PPA5	20×10^{-3}	26	−10.40	809	6283	1693	−304	−3988	−2874	3859
1APV	penicillopepsin	PPB1	1.0×10^{-3}	51	−12.23	1822	6512	1469	−760	−3081	−2318	4629
1APW	penicillopepsin	PPC1	10×10^{-3}	51	−10.87	920	5216	1458	−574	−2833	−2348	4206
1FQ4	saccharopepsin	SAC1	0.40×10^{0}	52	−8.70	2272	5857	2241	−427	−2889	−2511	4214
1FQ6	saccharopepsin	SAC2	14×10^{-3}	52	−10.70	1409	4501	2446	−767	−1849	−2921	3323
1FQ7	saccharopepsin	SAC3	3.74×10^{0}	52	−7.37	1576	4749	2511	−487	−2270	−2927	3556
1LGR	glutamine synthetase		0.85×10^{3}	80	−4.17	1491	4624	495	−169	−2341	−1118	2268
1ADF	alcohol dehydrogenase			66	−6.24	1220	6339	1458	−770	−2795	−3012	3467
2YPI	triosephosphate isomerase		15×10^{0}	81	−6.55	−257	3992	891	0	−2683	−2458	1978
1ULB	purine nucleoside phosphorylase			66	−7.23	1444	3889	469	−145	−1511	−1259	2391
1DIH	dihydrodipicolinate reductase		1.8×10^{0} [b]	82	−7.83	3000	8156	1225	−773	−2701	−2906	5517
1LYB	cathepsin		3.8×10^{-6}	83	−15.5	2462	6314	2051	−667	−2621	−2615	5498
4HMG	hemagglutinin			66	−3.48	1666	6101	642	−865	−2230	−1982	3459

[a] K_{ass} app (L/mol × 10^6). [b] K_d. [c] IC_{50}; *, neutral benzamidine model. [d] HINT scores for "essential" hydrogens case, by interaction type. H_{HB+AB} = hydrogen bond + acid−base, H_{HH} = hydrophobic−hydrophobic, H_{AA} = acid−acid, H_{BB} = base−base, and H_{HP} = hydrophobic−polar.

Bartlett and co-workers,[26] including measuring K_i at different pH levels for some of the ligands, is particularly interesting and will be a topic of the Discussion section. In general, our model-building rules are designed to produce the highest (most favorable binding) HINT score. There are certainly a variety of reasons that models thus constructed may not be the most biologically relevant species, but (i) the protons we optimize were never observed experimentally and are, in reality, highly fluxional and not static; (ii) these are

definable endpoints in model building rather than supposedly "ideal" intermediate or average structures, and any errors introduced would be, in principle, systematically applied to all molecular models.

To critically evaluate the HINT results, a few of the calculation options of the program were varied. Most importantly for this study, we investigated the effects of two approaches to "partitioning" the protein and ligand molecular models. In HINT, partitioning refers to calculating the Log$P_{o/w}$ for each molecule and assigning the empirical HINT constants a_i (partial Log$P$) and S_i (partial solvent accessible surface area) to each atom of each species.[13,14] We varied the treatment of hydrogens with two approaches. The first, called "essential", only treats as explicit atoms hydrogens attached to polar heavy atoms, i.e., O, N, S, P, etc., while hydrogens attached to carbons are implicitly evaluated as part of united atom definitions. The second approach, called "all", explicitly evaluates all hydrogen atoms. The following subsections describe the results for several representative proteins with either polar or apolar active sites. Problems that were encountered in the analysis and the applied solutions are described.

Human and Bovine Thrombin. Thrombin is a serine protease involved in blood coagulation, catalyzing the conversion of fibrinogen in fibrin. Clinical studies have shown that the selective inhibition of thrombin allows the control of thrombosis and atherosclerosis pathologies.[27] The human and bovine thrombin proteins are virtually homologous in the binding pocket region. A number of thrombin inhibitors have been designed (see Scheme 1), and the ligand structures bound to the human and/or bovine proteins determined by X-ray crystallography ({1ets, 1ett, 1etr};[28] {1uvt};[29] {1d3t, 1d3q, 1d3p, 1d3d};[30] {1dwb}[31]) show little difference between the bovine and human cases. Measured inhibition constants are also very much the same for human and bovine complexes of the same ligand. Thus, we are treating the two species interchangeably in this work.

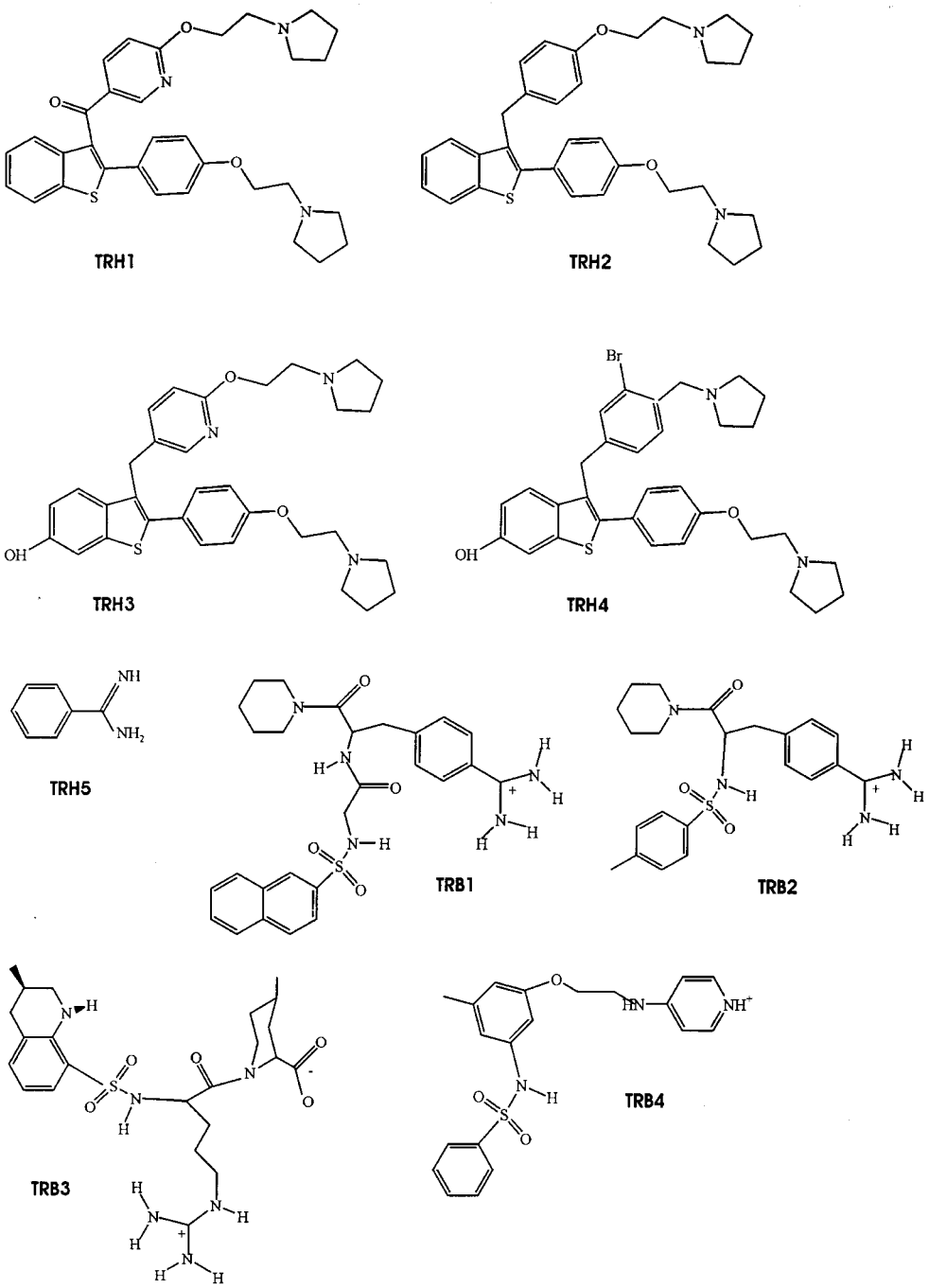
Inhibition constant data for the thrombin inhibitors have generally been recorded at around pH 8.0.[32−34] For some of the ligands, TRH1−4, where the ligands themselves cannot be ionized, the pH is of little consequence. However, for ligands TRH5 and TRB1−4, the actual protonation state of the ligand significantly affects the modeling results. The parent compound of ligands TRB1−2, benzamidine (TRH5), has an apparent pK_a of greater than 10. The modified benzamidines (TRB1−2) will have somewhat higher pK_a values because of electronic effects of the para substitutions. We modeled TRH5 in both the protonated (a) and neutral (b) forms (Scheme 2). TRB1−2, as well as the arginine derivative TRB3, were modeled only in the protonated form. The pyridine derivative, TRB4, was modeled as a pyridinium ion.

While the protonated benzamidine (TRH5a) produces a more favorable HINT score (1606) than that of TRH5b (743), the poor K_i and ΔG for this compound with respect to the other analogues (K_i for TRH5 is 10^3-10^6 less than that of the other ligands) suggest taking a closer look at this complex. Two possible explanations for this discrepancy are (i) some fraction of the ligand may be binding in the neutral form, with inherently weaker interactions with the protein; and (ii) TRH5 is a significantly smaller ligand than the others and the entropy cost of additional water molecules bound in the cavity may noticeably impact K_i and $\Delta G_{binding}$. We are exploring this possibility for a subsequent paper. A confounding factor, however, is that the benzamidine−thrombin complex crystal structure[31] has a relatively poor resolution (3.16 Å) and only located five water molecules in this region. The number and quality of crystallographically identified water molecules in a structure are directly dependent on the structure resolution. HINT scores and $\Delta G_{binding}$ data for the thrombin complexes are listed in Table 1. In some cases (TRH2 and TRH4), ΔG values were calculated from apparent $K_{association}$ rather than from K_i. In cases where both were available, they are similar—within a factor of 2.[34] This should have only a small impact on calculated $\Delta G_{binding}$.

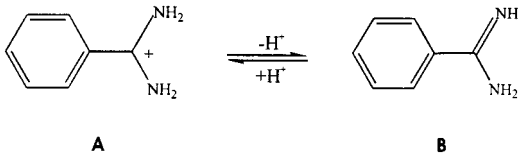
Bovine Trypsin. Trypsins are serine proteases that play important roles in the regulation of biological processes. They are site specific for catalyzing the cleavage of peptide bonds on the carboxyl end of only lysine and arginine residues. Crystallographic structures for these complexes have been reported ({1tnj, 1tnk, 1tni, 1tnl, 1tng, 1tnh};[35] {3ptb};[36] {1pph}[37]). The eight trypsin−ligand complexes examined in this work (Scheme 3, TPB1−8) are of two classes. The first class, TPB1−6, is simple primary amines. The other two ligands are related to benzamidine. The K_i measurements for TPB1−8[35,37−39] were carried out at around pH 8.0. Benzamidine (TPB7) was modeled as before, for thrombin, in both the neutral and the protonated states (see Scheme 2). And, again as before, the HINT score for the neutral benzamidine model (1413) is more consistent with its relatively poor K_i than the HINT score for the protonated model (2025) (see Table 1). The para-substituted benzamidine (TPB8) was modeled in the protonated form.

α-Subunit of Tryptophan Synthase. The tryptophan synthase $\alpha_2\beta_2$ complex catalyzes the last two steps of L-tryptophan biosynthesis in bacteria and plants. The α-subunit cleaves indole-3-glycerol phosphate in indole and glyceraldehyde-3-phosphate. The three-dimensional structures of the wild-type enzyme or its βK87T mutant in the presence of either phosphate-based inhibitors ({2trs, 2tsy};[40] {1a50};[41] {1qop}[42]) or phosphonate-based inhibitors ({1cx9, 1c29, 1c9d, 1cw2, 1c8v}[43]) have been determined by X-ray crystallography. Two of the crystallographic structures 2trs and 2tsy were collected for the βK87T mutant. This mutation is not near the ligand binding site. The selected complex ligands (TS1−8) are shown in Scheme 4. IC_{50} values[43] for complexes with ligands TS1−5 were converted to K_i assuming $K_i = IC_{50}/2$, i.e., ligand concentration is similar to K_d at assay conditions,[44] and ΔG values were calculated. The inhibition constant of ligand TS6 with wild-type tryptophan synthase has been measured several times,[41,45−47] with values ranging between 1 and 15 μM; we are using an average value of 5 μM. There are two crystal structures for this complex—one for the wild-type tryptophan synthase[42] and another for the βK87T mutant.[40] (No inhibition data have been reported for the βK87T mutant.) HINT scores were calculated for both and are reported in Table 1. The difference in HINT scores is nearly negligible as compared to the differences in K_d measurements for the wild-type com-

Scheme 1. Human (TRH1-5) and Bovine (TRB1-4) Thrombin Ligands



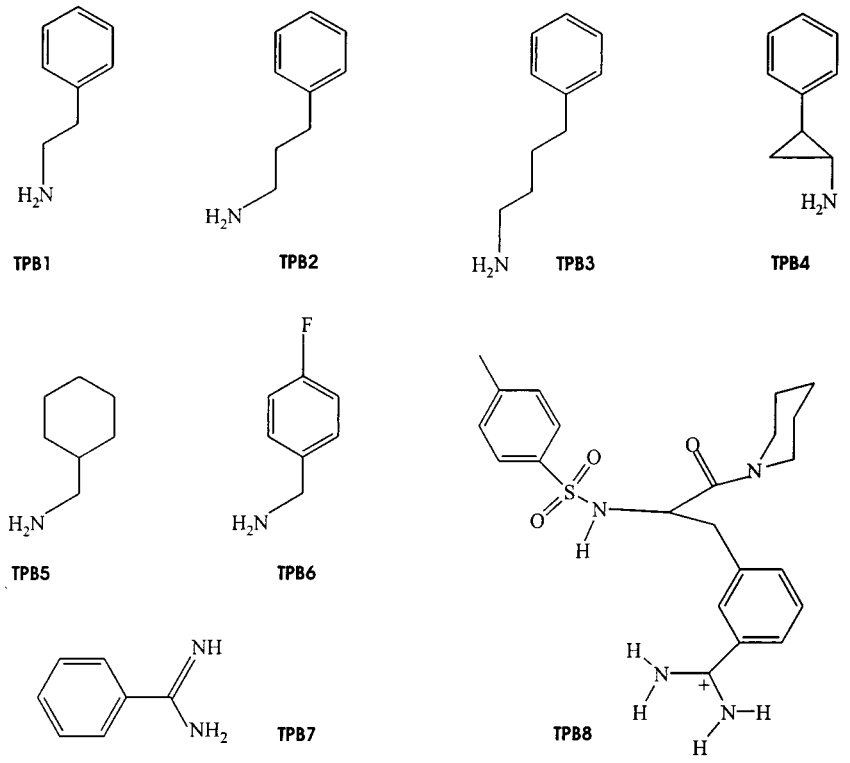
Scheme 2. Protonated and Neutral Forms of Benzamidine



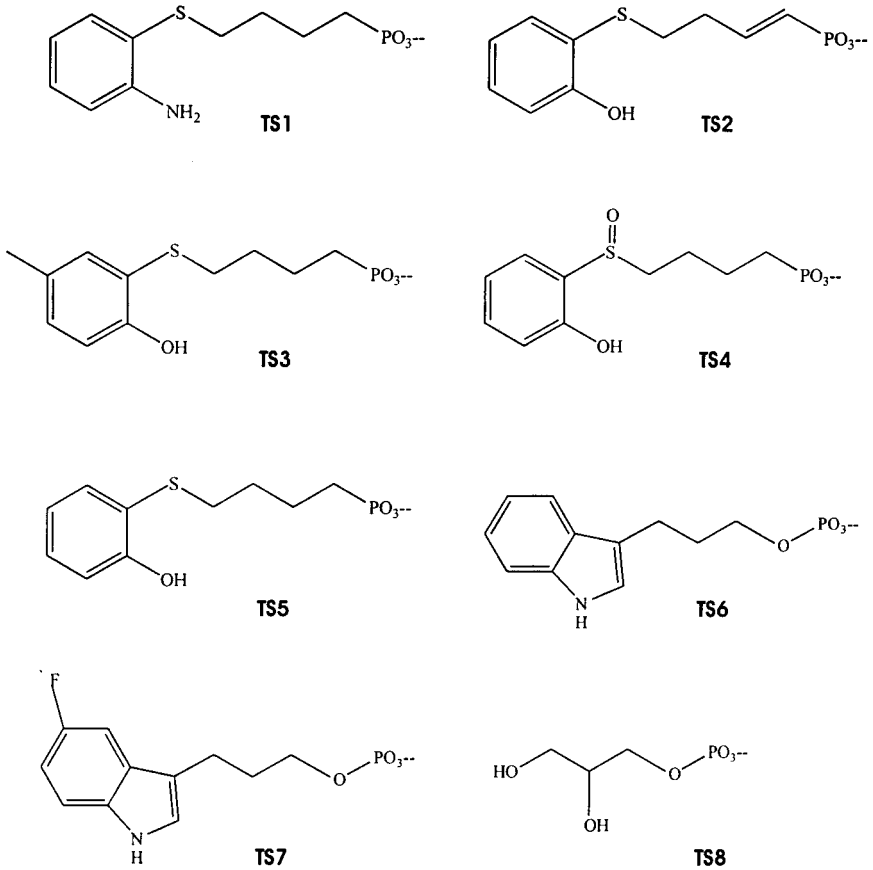
plex. Also, while the three-dimensional structural data for TS8 was available for the βK87T mutant,[40] the inhibition data for this ligand has only been reported for the wild-type TS8 complex.[48] However, within the apparent uncertainty of these measurements, the cross-species comparison is probably valid. Interestingly, the K_d for TS8 is very weak, in the millimolar range. Consequently, the HINT score for this complex is very low, actually, in this single case negative.

Modeling of the ionization state for the phosphates and phosphonates in complexes TS1−8 is complex. The multiplicity of possibilities for ionization state and protonation site is sizable. When that is coupled with the limitations of the phosphorus and oxygen atom types available within the Tripos molecular mechanics force field, we adopted two models for the complexes. The phosphate-containing ligands (TS6−8) were modeled as shown in Scheme 5A, with labeling based on the Tripos force field atom potential definitions. For the phosphonate ligands (TS1−5), one of the phosphonate oxygens was protonated, as shown in Scheme 5B. This proton may interact to form a hydrogen bond with nearby water molecules.

Scheme 3. Bovine Trypsin Ligands

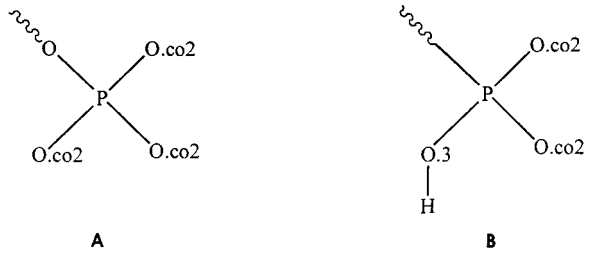


Scheme 4. Tryptophan Synthase Ligands



Penicillopepsin. Penicillopepsin is a monomeric fungal aspartic proteinase, characterized by two predominantly β-sheet domains, each contributing one aspartic acid (Asp33 and Asp213) to the active site. A third aspartic acid (Asp77) also interacts with ligands at the active site. We analyzed penicillopepsin structures complexed with several peptide-based inhibitors of aspartyl proteinases (PPA1−5, PPB1, and PPC1, Scheme 6). The high-resolution X-ray crystallographic structures and inhibition constants of these complexes

Scheme 5. Force Field Model Used for (A) Phosphate-Containing Tryptophan Synthase Ligands and (B) Phosphonate-Containing Tryptophan Synthase Ligands



have been previously reported ({1bxq, 1bxo};[49] {1ppl, 1ppm, 1ppk};[50] {1apv, 1apw}[51]). The inhibition constants[26] for one of the complexes, PPA4 {1ppm}, Cbz-Ala-Ala-Leu^P-(O)Phe-OMe (where Cbz is benzoxycarbonyl, Leu^P is the phosphonic acid analogue of leucine, and (O)Phe is L-β-phenyllactic acid), were recorded as a function of pH, and the associated data can be reconciled in terms of the ionization states of the catalytic aspartates (see Discussion).

Our model building was partially guided by the pH conditions under which the inhibition constants were generally measured (3.5 or 4.5) and under which the crystals were grown. In our models, Asp77 was always protonated, as this improves its interaction with a close amide carbonyl of the ligand backbone. For Asp33 and Asp213, the optimized protonation states varied with ligand structure (see Scheme 7). Note that for the phosphonate cases (PPA1−5), while our models always show Asp213 as being protonated, a choice can be made whether to protonate an oxygen on the phosphonate or an oxygen on Asp33 to create a second hydrogen bond to the phosphonate (Scheme 7A). As this is in many ways arbitrary because it is a shared hydrogen, we chose to protonate the aspartate because HINT parametrization for the highly polar phosphate−phosphonate−etc. system is inferior to that for the more well-understood carboxylic acid−carboxylates system. For inhibitors PPB1 and PPC1, a choice can be made whether to protonate Asp33 or Asp213 (Scheme 7B,C). We protonated Asp 213 in both cases as this produces the highest HINT score. Table 1 sets out the experimental binding energy and HINT score data for the penicillopepsin complexes we examined in this work.

Saccharopepsin, Hydroxynitrile Lyase, and Other Proteins. Many of the same considerations described above were applied to the remaining families of complexes and a collection of "orphaned" protein−ligand complexes. Some of these data sets are of limited size, only three or four members, while some have a larger membership but have a very small spread in K_i and ΔG. For the three saccharopepsin−ligand complexes, there are two important aspartates in the binding region, Asp32 and Asp215. Crystallographic and kinetic results ({1fq4, 1fq6, 1fq7}[52]) suggest that only one of these aspartate dyad carboxylates is ionized. Furthermore, our HINT score optimization protocol (i.e., modeling and evaluation of all relevant species) confirms this and suggests that the protonated residue is Asp215. The three ligands (SAC1−3, Scheme 8) are structurally very diverse, with little similarity of scaffold, hydropathic, and pharmacophore profile, and even ligand dimensions. They exhibit a concomitant range of inhibition constants and HINT scores (Table 1). The difference of ligand volumes suggests that consideration of the entropic contribution of water may be of value. The three hydroxynitrile lyase complexes were crystallized with very small ligands (HNL1−3, Scheme 8), i.e., from 3 to 11 nonhydrogen atoms ({2yas, 5yas, 3yas}[53]). Hexafluoroacetone (HNL3) was modeled as the gem-diol as suggested by the crystal structure.[53] One particularly close contact between one of the hydroxyl groups and a methyl of Leu148 was remedied by energy minimization in the affected region. Bridging water molecules may also play a role in the binding of this set of complexes. Results are set out in Table 1.

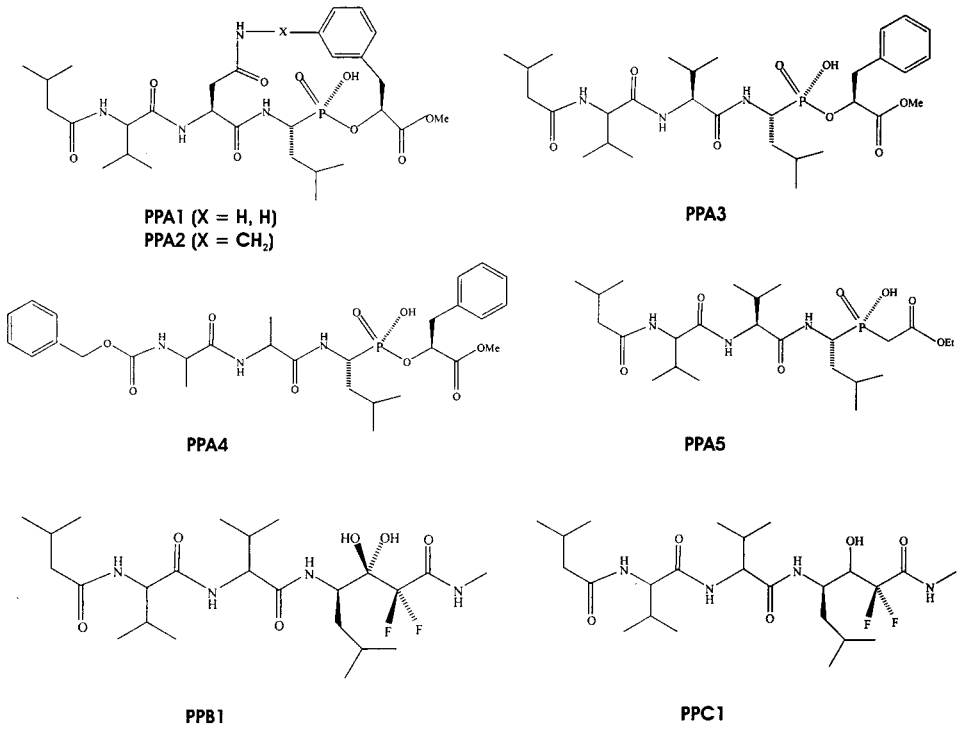
Of the two series of hydrophobic−site complexes, retinol binding protein was described in a previous paper.[23] These data are used in the present work. The very hydrophobic adipocyte lipid binding protein has been extensively studied and crystallized ({1adl},[54] {1lie},[55] {1lid, 1lif}[56]). We report the data from four ligand−protein complexes (ALB1−4, Scheme 8). In two of the PDB files, for ALB1 and ALB2,[54,55] Cys117 was reported as the oxidized $-SO_2$ rather than the normal −SH. In the PDB file for ALB1, Met40 was also oxidized.[54] It is not clear from the observed electron density that these are the predominant forms.[54,55] Presumably, the presence of oxidized sulfurs on cysteine or methionine residues in the PDB files are crystallographic artifacts due to less than full site occupation by the ligand(s). Experimental inhibition measurements for the four ligands bound to the adipocyte lipid binding protein were measured in two laboratories: LaLonde et al.[54] used a cysteine modification assay[57] to determine K_i and ΔG for ALB1 and ALB2 (cysteine is oxidized when a fatty acid is not bound); Richieri et al.[58] used a fluorescence assay to determine K_d for all four (ALB1−4) of the ligands. We have modeled Cys117 and Met40 in the usual forms and report in Table 1 ΔG values derived from the Richieri et al.[58] data for the complexes, because it is a complete and self-consistent set. It should be noted, however, that there are 1−2 orders of magnitude difference in reported K_d for the two measurements of the same complexes.

A number of protein−ligand complexes that met our criteria for analysis were not part of extended series. The modeling and calculation of HINT scores for these complexes proceeded using the set of guidelines described above and with careful attention to protonation state, etc. of the protein and ligand functional groups in response to pH and other experimental conditions. The data for these complexes is provided in Table 1.

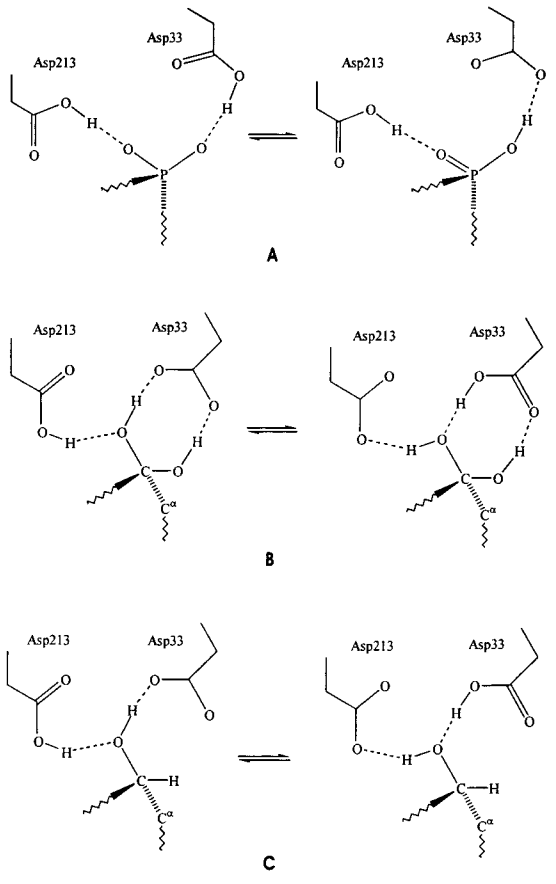
Discussion

HINT is a unique computational model based on experimental Log*P* data that quantitatively evaluates the biological phenomena of ligand binding and macromolecular association. It is an intuitive model because the components of the HINT score are directly related with the type of interactions present between different molecules, and the magnitude of the score is indicative of the strength of the potential interaction. Because Log*P* can be directly correlated with free energy,[15,59] the HINT score reveals information not only about enthalpy but also about entropy. This is because a large propor-

Scheme 6. Penicillopepsin Ligands



Scheme 7. Interaction of Penicillopepsin Asp33 and Asp213 Residues with Ligand Functional Groups

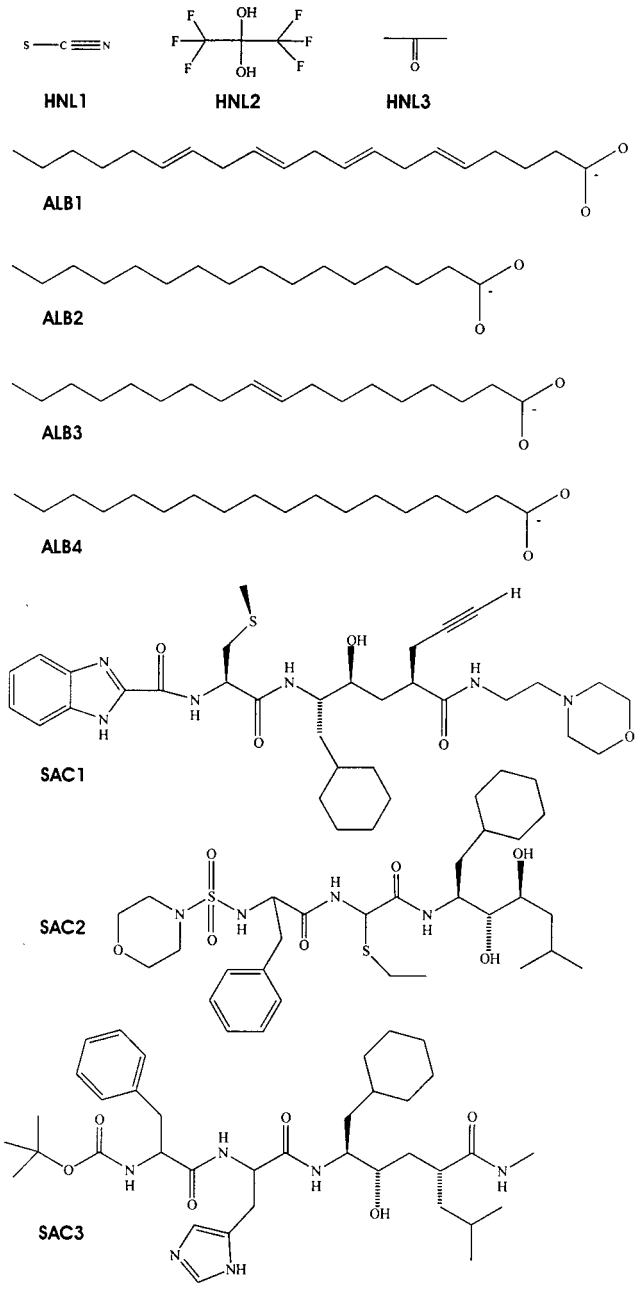


tion of entropy in the biological environment resulting from biomolecular associations arises from the transfer of the solute (ligand) from the solvent (water) to bound position. Concomitant with that process are the transfer of water molecules bound and semibound to the ligand and protein to the (more disordered) solvent, i.e., an increase in entropy. These effects are directly related to the phenomenon that is called hydrophobicity. HINT score components can also be correlated with desolvation energy.[24] The basic premise of HINT is that the energetics of biomolecular association in water are basically the same as those recorded in the experimental measurement of $\mathrm{Log}P_{\mathrm{octanol/water}}$. Importantly, each atom−atom term of the HINT score is a partial δg, which encodes a fraction of the total $\Delta G_{\mathrm{interaction}}$.

We have reported above (Table 1) the $\Delta G_{\mathrm{binding}}$ and HINT scores for the 53 complexes examined in this study. A key goal of this work is the development of a simple and robust method for the prediction of binding affinity, especially in hydrophobic systems where entropy may be the driving force for the molecular association. As the HINT model has evolved, we have been continually interested in understanding the relationship between HINT score and free energy of binding.[15] In this sense, a promising result was the very good correlation found between the HINT score and the K_{diss} of retinol binding proteins and its ligands.[23] Studies on the free energy of association for hemoglobin dimers also showed a good correlation between HINT score and $\Delta G_{\mathrm{association}}$.[21,22] To better understand this relationship, we have extended the analysis to other protein−ligand systems with different polarity of the active sites for which experimental binding data and crystallographic structures are available. For the validation of the method, we have applied quite rigid conditions in order to obtain a subset of protein−ligand complexes of homogeneous quality. The subset is representative of proteins that bind ligands with different structures and polarity.

We also address here a few other topics related to the accurate modeling of protein−ligand complexes and the

Scheme 8. Hydroxynitrile Lyase, Adipocyte Ligand Binding Protein and Saccharopepsin Ligands



estimation of free energy of binding: (i) Does the relationship between HINT binding score and Δ*G* show evidence of a "global HINT constant" relating HINT score units to free energy of binding? (ii) Can other quantitative structure−activity relationship (QSAR)-like equations be constructed utilizing the HINT score and/or its components to estimate free energy of binding? (iii) Is the relationship between HINT score and Δ*G* more accurate for families of ligands bound at the same protein site? (iv) What is the effect of pH on K_i, Δ*G*, and model building? (v) What are the sources (and importance) of errors in model building and empirical free energy calculations from PDB data? How can these errors be ameliorated?

Global Relationship Between HINT Score and Δ*G*. Plots of Δ*G* vs HINT score for the 53 complexes, grouped by protein, are provided in Figures 1 and 2 for HINT scores calculated with only essential hydrogens

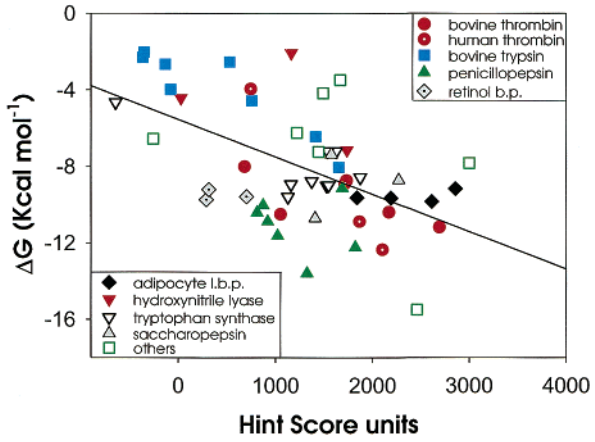


Figure 1. Plot of experimental Δ*G* vs HINT score units for 53 protein−ligand complexes where the HINT score is calculated using only essential (polar) hydrogens. The line is the best least-squares fit, as described in the text (eq 1).

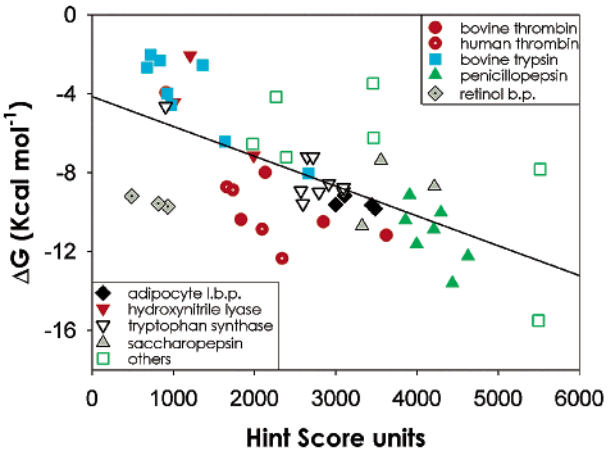


Figure 2. Plot of experimental Δ*G* vs HINT score units for 53 protein−ligand complexes where the HINT score is calculated using all hydrogens. The line is the best least-squares fit, as described in the text (eq 2).

and with all hydrogens, respectively (see Materials and Methods). One of the goals of this study is to determine the optimal protocol for HINT score calculations, and a particular facet of this is to determine what effect, if any, inclusion of nonpolar hydrogens has on the HINT scores. In practice, the all option explicitly catalogs each nonpolar hydrogen and its interactions, while the essential option uses united atoms for $-CH_3$, $-CH_2-$, etc. From the plots of Figures 1 and 2, it is evident that there are only minor differences between the two methods in terms of the overall model. We have earlier asserted[17,22] that there are compelling reasons suggesting that the essential approach is more valid for biomacromolecular structure analysis. First, hydrogen atoms are only very rarely located in crystallographic analyses of biological systems. The apparent presence of hydrogen bonds lends some experimental credence to placing and modeling of polar hydrogen atoms, but there is no experimental basis for location of nonpolar hydrogen atoms. Second, explicit hydrogen atoms add complexity to the calculations. For example, because each methyl is represented by one (united) atom in the essential case but four discrete atoms in the all case, an interaction between two methyl groups would have 16 terms in the all case but only one in the essential case.

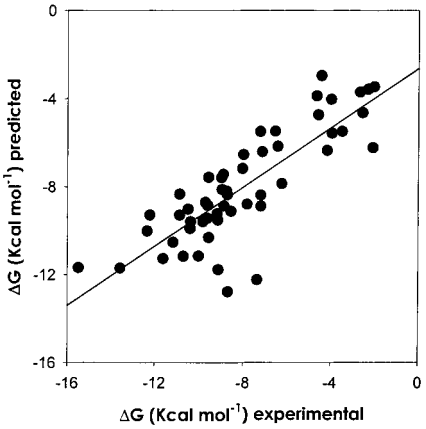


Figure 3. Plot of experimental ΔG vs predicted ΔG (from eq 3) for essential hydrogen HINT models. Correlation (r^2) is 0.67.

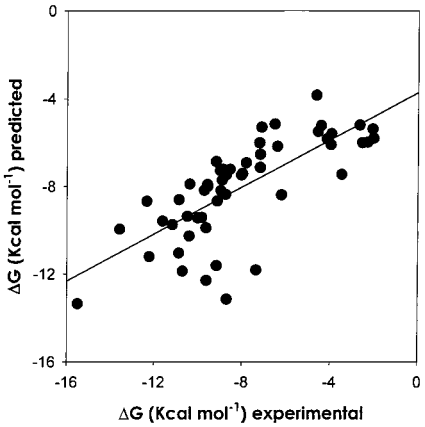


Figure 4. Plot of experimental ΔG vs predicted ΔG (from eq 4) for all hydrogen HINT models. Correlation (r^2) is 0.54.

Linear regression of the data in Figures 1 and 2 produces eqs 1 and 2

$$\Delta G = -0.001\ 95\ H_{TOTAL} - 5.543 \quad (1)$$

$$\Delta G = -0.001\ 51\ H_{TOTAL} - 4.137 \quad (2)$$

where H_{TOTAL} is the total HINT score for the complex. Equation 1, the essential hydrogens only case, has a standard error of 2.6 kcal mol^{-1} and a Pearson r of 0.54; eq 2, the all hydrogens case, has a standard error of 2.4 kcal mol^{-1} and a Pearson r of 0.63. The accuracies of these predictions are comparable to those obtained with much more sophisticated and time-consuming methodology,[60,61] such as free energy perturbation and/or linear response methods,[8,10,11,62−65] or with less intuitive free energy scoring algorithms.[9,66−71]

A more QSAR-like model can be constructed by utilizing the six types of interactions that HINT resolves as terms in a multilinear equation. These terms are H_{HH} (hydrophobic−hydrophobic), H_{HB} (hydrogen bond), H_{AB} (acid−base), H_{AA} (acid−acid), H_{BB} (base−base), and H_{HP} (hydrophobic−polar). In the HINT model, hydrogen bonds are special cases of acid−base interactions, largely filtered by a somewhat arbitrary interaction distance cutoff. We have combined these two terms in eqs 3 and 4 below for the essential and all treatment of hydrogen data, respectively

$$\Delta G = -0.001\ 02\ (H_{HB} + H_{AB}) - 0.004\ 83\ H_{HH} - 0.002\ 22\ H_{AA} - 0.000\ 26\ H_{BB} - 0.001\ 81\ H_{HP} - 2.231 \quad (3)$$

$$\Delta G = -0.000\ 49\ (H_{HB} + H_{AB}) - 0.009\ 86\ H_{HH} - 0.000\ 92\ H_{AA} + 0.000\ 01\ H_{BB} - 0.005\ 69\ H_{HP} - 3.179 \quad (4)$$

The standard errors of these correlations are ±1.8 and ±2.2 kcal mol^{-1} and the r^2 values for predicted vs experimental (Figures 3 and 4) are 0.67 and 0.54. It is important to note that all interactions in the HINT model are calculated with the same functions, the classification of interaction type is based only upon the properties of the interacting atoms and has no actual effect on the score calculations. However, the relative magnitudes of the coefficients reveal insight into the driving forces of the binding process. First, note that the H_{AA} scores are generally small as they arise from the fairly infrequent cases of two polar hydrogen atoms being in close proximity even after structure optimization. Many complexes have very small or zero H_{AA} scores (Table 1). So, discounting the coefficients of H_{AA} in eqs 3 and 4, the most significant terms are H_{HH} and H_{HP}, which are the types of interactions most closely associated with entropy and desolvation. As expected, the H_{HH} and H_{HP} coefficients are significantly larger for the all treatment of hydrogens (eq 4) as compared to the essential case (eq 3). This is because the all treatment of hydrogens distributes the lipophilic potential of hydrophobic (CH_n) groups into smaller, less significant information packets. Then, to properly account for the hydrophobic contribution (in both H_{HH} and H_{HP} terms), the regression calculates compensating larger coefficients.

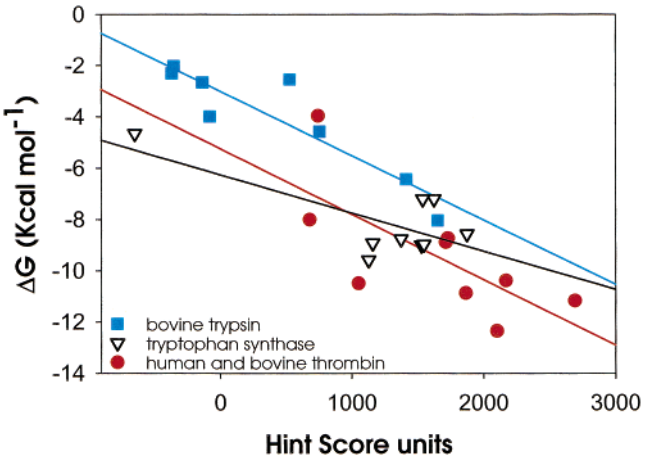
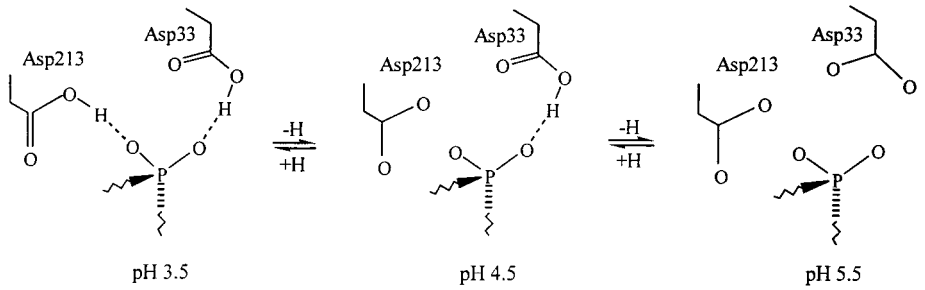


Figure 5. Plots of experimental ΔG vs HINT score units for bovine trypsin (blue squares), tryptophan synthase (black triangles), and human/bovine thrombin (red circles) calculated using only essential (polar) hydrogens. The lines are the best least-squares fits, as described in the text (eq 5−7).

Relationship between HINT Score and ΔG for Specific Systems. Within this data set of 53 complexes, three series of ligands bound to the same protein or enzyme are extensive and robust enough to justify system regression analyses. The trypsin, thrombin, and tryptophan synthase systems have eight or more members and reasonable ranges in ΔG. Figure 5 sets out plots of the regression analyses on these three data series, all calculated using essential hydrogens. The

Scheme 9. Penicillopepsin Asp33 and Asp213 Protonation as a Function of pH When Bound to Inhibitor PPA4



regression equations 5−7

$$\Delta G = -0.002\,51\,H_{\mathrm{TOTAL}} - 2.993 \qquad (5)$$

$$\Delta G = -0.002\,56\,H_{\mathrm{TOTAL}} - 5.234 \qquad (6)$$

$$\Delta G = -0.001\,49\,H_{\mathrm{TOTAL}} - 6.250 \qquad (7)$$

are for trypsin (eight complexes, standard error = ±1.0, $r^2 = 0.83$), thrombin (nine complexes, standard error = ±1.9, $r^2 = 0.50$), and tryptophan synthase (nine complexes, standard error = ±1.1, $r^2 = 0.53$), respectively. Potential sources of modeling errors in the thrombin series (eq 6) have been discussed above, perhaps accounting for the relatively poor correlation statistics. It is interesting that the slopes of the three lines are reasonably consistent and within the regression uncertainty (Figure 5) while there is some variation in the Δ*G* intercept.

By analogy to QSAR analyses, it would appear that the availability of a "learning set" of structurally and thermodynamically characterized complexes for the same protein system should improve the accuracy and precision of Δ*G* predictions for complexes. Even these limited case scenarios, i.e., eqs 5−7 that were assembled from crystallographic and solution binding data collected in multiple laboratories, suggest that accuracies in Δ*G* predictions on the order of ±1 kcal mol^{-1} should be routinely achievable with the HINT free energy scoring methodology applied on a system-by-system basis. In an example of virtual screening,[19] where the model structures of 26 cyclin-dependent kinase inhibitor complexes were generated by docking (and thus not subject to crystallographic errors, etc.), HINT results were shown to highly correlate with inhibition. The regression of Δ*G* vs H_{TOTAL} for these data (not shown) is

$$\Delta G = -0.002\,91\,H_{\mathrm{TOTAL}} - 3.438 \qquad (8)$$

where the standard error is ±0.3 kcal mol^{-1} and r^2 is 0.94.

pH Effects on Model Building and Calculations. As noted above in the results, it is crucial to carefully model the ionization state of the acids and bases in the proteins and on the ligands in order to obtain the most accurate representation of the complex structures. Of course, because the crystallography experiments cannot locate protons, this modeling was done based on first, the pH of the inhibition constant measurements, and second, on the local environment around the ionizable group. (It may occasionally be the case that the crystals were grown under significantly different pH conditions than the solution measurements were made.) In general, our modeling philosophy and procedures sought to maximize the total HINT score for each complex, but divergence between calculated HINT score and measured inhibition occasionally revealed insight into the binding of particular ligands.

Table 2. pH Dependence of Penicillopepsin Inhibition by PPA4

pH	K_i (μM)	Δ*G*	H_{TOTAL} (essential)
3.5	0.19	−9.13	1688
4.5	2.7	−7.56	−1641
5.5	107	−5.39	−5836

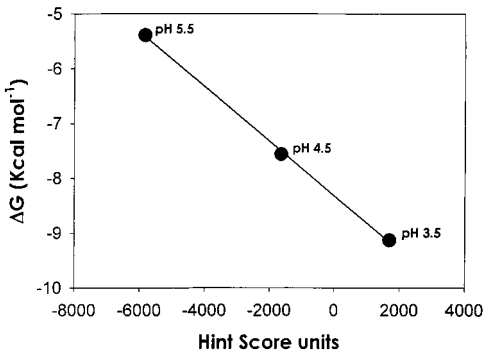


Figure 6. Plot of experimental Δ*G* vs HINT score units for three pH cases of penicillopepsin inhibitor PPA4 (see Scheme 9), where the HINT score is calculated using only essential (polar) hydrogens.

In a particularly lucid series of articles, Paul Bartlett and co-workers have measured the binding of a large variety of ligands to penicillopepsin and reported the three-dimensional structure of several complexes.[26,48−50] For the ligand Cbz-Ala-Ala-Leu^{P}-(O)Phe-OMe (PPA4, Scheme 6), inhibition constants were measured at three pH points, 3.5, 4.5, and 5.5, with significant variance in K_i that was rationalized in terms of the ionization state of the catalytic aspartates (Asp33 and Asp213) at the binding site.[26] Scheme 9 shows the relationship between the three complex structures. We modeled these structures and calculated HINT scores for each (Table 2). Figure 6 demonstrates the nearly perfect linear correlation between HINT score and Δ*G* for the three cases of pH 3.5, 4.5, and 5.5 in the binding of PPA4 to penicillopepsin. Figure 7A−C are HINT interaction maps for the three cases. The progression from two hydrogen bonds between the aspartates and the ligand phosphonate (pH 3.5, Figure 7A), to one hydrogen bond between Asp33 and the ligand phosphonate (pH 4.5, Figure 7B), to no hydrogen bonds (pH 5.5, Figure 7C) is evident. Green contours represent hydrophobic−hydrophobic interactions, blue contours represent favorable polar interactions (i.e., acid−base and hydrogen

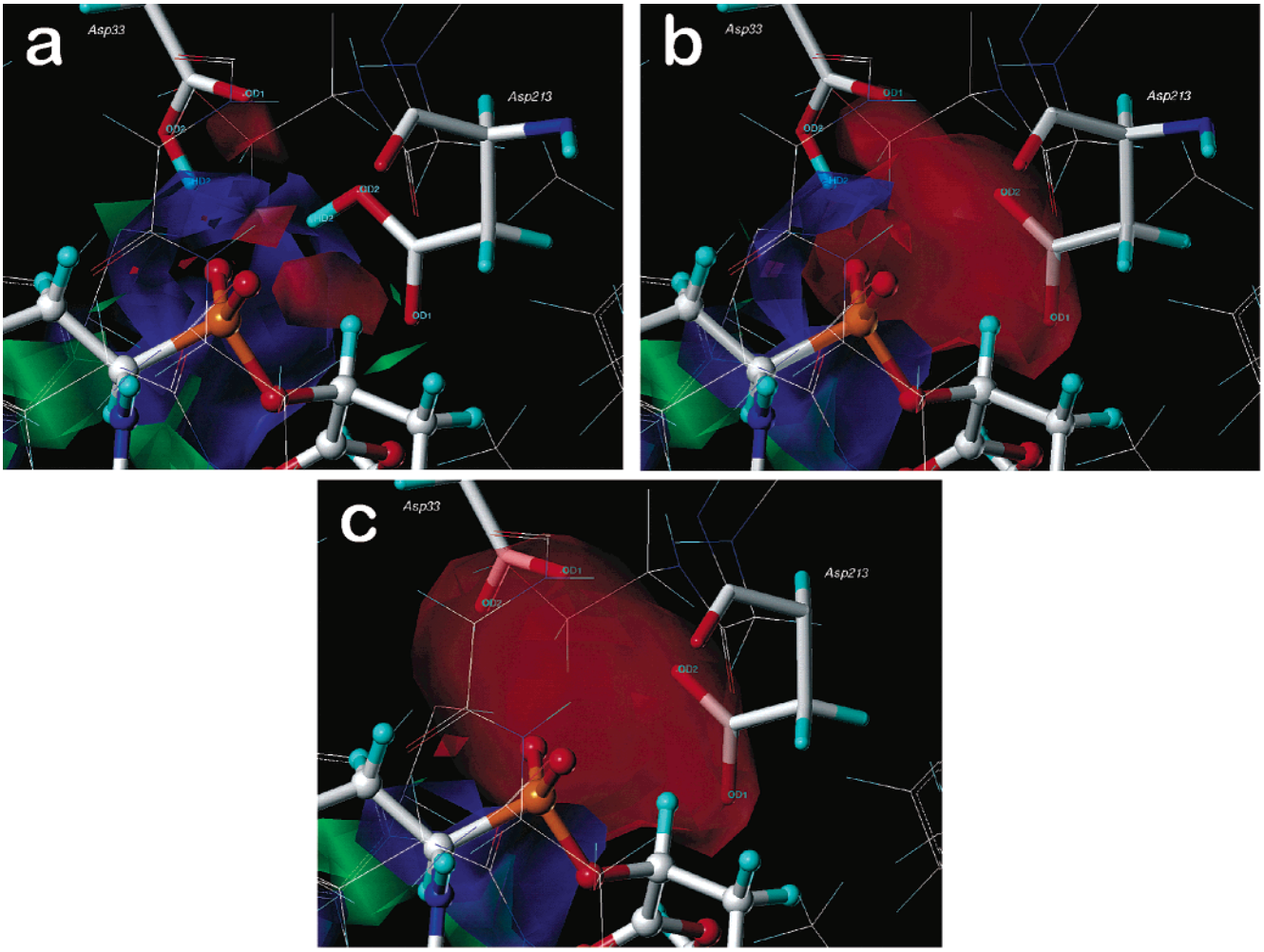


Figure 7. HINT interaction maps showing favorable and unfavorable interactions between penicillopepsin and ligand PPA4 under varying pH conditions; (a) pH 3.5; (b) pH 4.5; and (c) pH 5.5. Red contours indicate regions of unfavorable polar−polar interaction, which in these cases are largely due to "base−base" interactions between the ligand phosphonate and the Asp33 and Asp213 (particularly b and c). Blue contours are representative of hydrogen bonding between the ligand and protein, i.e., between the hydrogen(s) on the protonated Asp33 and Asp213 residues and the phosphonate moiety on the ligand (particularly a and b). Green contours indicate hydrophobic−hydrophobic interactions.

bond), and red contours represent unfavorable polar interactions (i.e., acid−acid and base−base).

Other systems in this study were carefully examined for consistency of binding interactions, as represented by the HINT score, to the measured inhibition constants and calculated ΔG. In particular, the experimentally determined weak binding of benzamidine to both thrombin and trypsin was surprising in light of the excellent interactions the protonated benzamidine ligand could make in the active site of either protein. This would, on the surface, suggest that benzamidine may be (partially) binding in its neutral form. However, as we alluded to above, another possibility is that the loss of entropy associated with trapping water molecules in active sites with smaller ligands such as benzamidine may be a significant factor. We are continuing to explore these phenomena with calculations, but the availability of more extensive binding and inhibition data as a function of pH for structurally characterized systems would be valuable.

Errors and Uncertainties in Estimating ΔG. The first point to be considered is how accurate are the experimental measurements of inhibition from which ΔG is calculated? There is quite often a 1 order of magnitude variation in reported K_i values between laboratories. This suggests that a lower limit on the expected accuracies of ΔG predictions of $\pm 1.0-1.5$ kcal mol^{-1} within the same ligand family for a single protein would not be unreasonable. Application of a free energy scoring method over a large collection of protein systems with, for example, widely varying properties of active site polarity as described in this work or a variety of different assays and solution measurements, etc. should be expected to fare worse.

On the prediction side, other factors contribute significantly to uncertainties in the HINT free energy calculation. First, HINT is very sensitive to structure. It is essential to have an accurately determined three-dimensional structure of the protein−ligand complex for HINT score calculations. Even apparently small differences in structure can significantly influence the score. For example, the chemically identical subunits of streptavidin-HABA (2-[(4′-hydroxyphenyl)azo]benzoic acid) ligand complexes[72] show asymmetric structure when analyzed by HINT (unpublished results). This is not an isolated example of how subtle differences of ligand−protein geometry within the active sites of proteins or at protein−protein interfaces, even within the resolution

and accuracy of crystallographic structures, impact the efficacy and accuracy of computational techniques. Second, the correct atom potential types and bond types must be in place for the ligand. The PDB format does not support bond order, so incorrect assignments are almost invariably made by modeling programs that automatically interpret cofactor or substrate structures from PDB data files. Third, HINT requires that polar hydrogen atoms involved in interactions be accurately positioned with respect to the appropriate hydrogen bond acceptor(s). It is often the case that even extensive molecular mechanics structure optimization fails to properly position these hydrogen atoms. Fourth, cofactors can contribute to the free energy of binding by interacting with both the ligand and the protein. Ternary complexes are formed by the simultaneous presence of analogues of uracyl and methylentetrahydrofolate in the active site of thymidylate synthase.[73] A more common situation is that encountered when water molecules bridge the ligand and active site residues. The HINT score calculations and energetics associated with structurally conserved water molecules have been discussed for protein−protein systems[21,22] and are the focus of our ongoing research.

On the other hand, an approach such as HINT, which estimates free energy as a sum of partial free energies for each atom−atom interaction, is preferable to methods using scoring functions that are sums of a variety of terms. Dill (1997)[74] has commented that simply combining terms from different energy functions, e.g., Coulombic, hydrophobic surface contact area, van der Waals, hydrogen bonding, etc. functions, may not yield meaningful free energy estimates. At the opposite extreme, while difficult and time-consuming to perform, free energy estimates derived from detailed molecular dynamics simulations of the entire biomolecule + ligand + solvent system are reasonably reliable. The ultimate goal of research in this area is development of a robust and rapid method for high-throughput virtual screening. We believe that because of its foundation in the free energy measurement of Log$P_{o/w}$, the further development of the HINT model will contribute to this goal.

Materials and Methods

General. The program Sybyl version 6.7 (Tripos, Inc., St. Louis, MO; www.tripos.com) used for this work was installed on a cluster of Silicon Graphics Irix workstations. The program HINT version 2.35S (eduSoft, LC, Ashland, VA; www.edusoft-lc.com) was used as an add-on module within Sybyl.

Model Building. Three-dimensional coordinates of all protein−ligand complexes were retrieved from the PDB (www.rcsb.org),[75] following these criteria: (i) quality of the crystallographic data, crystallographic resolution lower than 3.2 Å; (ii) homogeneous experimental conditions for each class of complexes—whenever possible we chose data collected with the same technique for all of the ligands by the same group of researchers. All of the structures thus selected were imported as the PDB files into Sybyl and checked for correct atom and/or bond types with respect to the original literature reference to the ligand−protein complexes and in particular to the chemical structure of the ligand. All hydrogens, which are not normally present in the X-ray crystallographic data files in the PDB, were added with tools within the Sybyl Biopolymer and Build/Edit menus. While hydrogens added automatically in this manner are internally correct with respect to their own heavy atom parents, the automatic algorithms do not account for either intermolecular or interresidue steric clashes. To reduce this steric hindrance, hydrogens on the proteins were energy-minimized while keeping the coordinates of all nonhydrogen atoms fixed.

Next, hydrogens on polar heavy atoms, in cases where hydrogen bonds are possible, were manually examined and optimized by torsional rotation for proper orientation. This largely affects hydroxyl groups on the protein (i.e., serine, threonine, and tyrosine residues) and ligands but was occasionally important for amines. In brief, some of these hydrogens were initially placed by the automated procedure in orientations that did not support obvious hydrogen bonding with near neighbors and were trapped in local minima during optimization. Finally, the ionization state of polar residues on the proteins, especially aspartic acid and glutamic acid, and similar functional groups on the ligands were evaluated and modified as necessary. Markers for such cases were excessively large base−base interactions involving atoms in these groups. The final configuration, which in all cases was physically and chemically reasonable, was the one that gave the largest HINT score (vide infra). Except where described in the Results section, no substantial structural changes were made to any heavy atoms of the proteins or ligands.

Hydropathic Analysis. HINT first calculated Log$P_{o/w}$ for each component (protein and ligand) of the complexes. For proteins, the partition method was dictionary, where HINT was using a lookup table of parameters based on residue type and solvent condition.[14] For most cases, neutral was chosen for the solvent condition (lysine and arginine are protonated, and glutamic acid and aspartic acid are carboxylates); however, when the ionization state of a protein residue was changed from this norm as described above, the inferred option was used to automatically choose the state of each residue in the protein based on its atoms. For the ligands, HINT used the calculate method, an adaptation of the CLOG-P method of Leo.[76,59] HINT also has other important options for calculating Log*P* based on how the hydrogens are accounted for: "united", an approach that calculates Log*P* considering the hydrogen contribution only implicitly as part of the parent heavy atom; "essential", an approach that treats only polar hydrogens explicitly; and "all", an approach that treats all hydrogens explicitly.

We have previously reported (e.g., see refs 20 and 21) that for biomolecular associations dominated by polar interactions, the explicit treatment of only the essential protons is preferable to: (i) ignoring all hydrogens, i.e., united, which causes most hydrogen bonds to be undetectable by the HINT algorithm or (ii) including all hydrogens, which adds complexity to calculations, adds many more potential steric (van der Waals) instabilities to the molecular model, and dilutes the effect of hydrophobic−hydrophobic interactions such that, individually, most fall below the detection thresholds.

For this study, with a wide range of active site polarities in the protein set, HINT scores were calculated using both the essential and the all options. The total HINT score considers all possible interactions between the protein and the ligand. Positive contributions to the HINT score are acid−base, hydrophobic−hydrophobic interactions, and hydrogen bonds, whereas negative contributions are acid−acid, base−base, and hydrophobic−polar interactions. After each calculation, individual contributions to the total HINT score were analyzed and hydropathic interaction maps[13] were prepared to visualize the areas of hydrophobic or polar contacts. For anomalous contributions identified in this way, the optimization procedures described above were applied.

Statistical Analysis. The implementation of partial least squares in the Sybyl QSAR module/molecular spreadsheet was used to generate the reported linear and multilinear regressions. Reported standard errors and correlation coefficients are calculated in the usual way.

Acknowledgment. We gratefully acknowledge the support of the National Institutes of Health (D.J.A., Grant 5R01HL32793-15) and Virginia Commonwealth

University for partial support of this research, the Italian Instruction, University and Research Ministry Grant PRIN01 (A.M.), and the National Institute for the Physics of Matter (A.M).

References

(1) Ajay; Murcko, M. A. Computational Methods to Predict Binding Free Energy in Ligand−Receptor Complexes. *J. Med. Chem.* **1995**, *38*, 8, 4953−4967.

(2) Böhm, H.-J.; Klebe, G. What Can We Learn from Molecular Recognition in Protein−Ligand Complexes for the Design of New Drugs? *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 2588−2614.

(3) Lybrand, T. P. Ligand-Protein Docking and Rational Drug Design. *Curr. Opin. Struct. Biol.* **1995**, *5*, 224−228.

(4) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening—an overview. *Drug Discovery Today* **1998**, *3*, 160−178.

(5) Kellogg, G. E. Ligand Docking and Scoring: New Techniques and Applications in Drug Discovery. *Med. Chem. Res.* **1999**, *9*, 439−442.

(6) Perez, C.; Ortiz, A. R. Evaluation of docking functions for protein−ligand docking. *J. Med. Chem.* **2001**, *44* (23), 3768−3785.

(7) Muegge, I.; Rarey, M. Small molecule docking and scoring. *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 2001; Vol. 17, pp 1−60.

(8) Åqvist, J. Calculation of Absolute Binding Free Energies for Charged Ligands and Effects of Long-Range Electrostatic Interactions. *J. Comput. Chem.* **1996**, *17*, 1587−1597.

(9) Böhm, H.-J. The development of a simple empirical scoring function to estimate the binding constant for a protein−ligand complex of known three-dimensional structure. *J. Comput.-Aided Mol. Des.* **1994**, *8* (3), 243−256.

(10) Wang, J.; Dixon, R.; Kollman, P. A. Ranking ligand binding affinities with avidin: a molecular dynamics-based interaction energy study. *Proteins*: *Struct., Funct., Genet.* **1999**, *34*, 69−81.

(11) Wang, W.; Wang, J.; Kollman, P. A. What determines the van der Waals coefficient β in the LIE (Linear Interaction Energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins*: *Struct., Funct., Genet.* **1999**, *34*, 395−402.

(12) Shoichet, B. K.; Leach, A.; Kuntz, I. Ligand solvation in molecular docking. *Proteins*: *Struct., Funct., Genet.* **1999**, *34*, 4−16.

(13) Wireko, F. C.; Kellogg, G. E.; Abraham, D. J. Allosteric modifiers of hemoglobin. 2. Crystallographically determined binding sites and hydrophobic binding/interaction analysis of novel hemoglobin oxygen effectors. *J. Med. Chem.* **1991**, *34*, 758−767.

(14) Kellogg, G. E.; Joshi, G. S.; Abraham, D. J. New Tools for Modeling and Understanding Hydrophobicity and Hydrophobic Interactions. *Med. Chem. Res.* **1992**, *1*, 444−453.

(15) Kellogg, G. E.; Abraham, D. J. Hydrophobicity: is Log$P_{o/w}$ more than the Sum of its Parts? *Eur. J. Med. Chem.* **2000**, *35*, 651−661.

(16) Meng, E. C.; Kuntz, I. D.; Abraham, D. J.; Kellogg, G. E. Evaluating docked complexes with the HINT exponential function and empirical atomic hydrophobicities. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 299−306.

(17) Kellogg, G. E.; Scarsdale, J. N.; Fornari, F. A., Jr. Identification and Hydropathic Characterization of Structural Features Affecting Sequence Specificity for Doxorubicin Intercalation into DNA Double-Stranded Polynucleotides. *Nucleic Acids Res.* **1998**, *26*, 4721−4732.

(18) Kellogg, G. E.; Scarsdale, J. N.; Cashman, D. J. Ligand Docking and Scoring in DNA Oligonucleotides. Binding of Doxorubicin and Modeled Analogues to Optimize Sequence Specificity. *Med. Chem. Res.* **1999**, *9*, 592−603.

(19) Gussio, R.; Zaharevitz, D. W.; McGrath, C. F.; Pattabiraman, N.; Kellogg, G. E.; Schultz, C.; Link, A.; Kunick, C.; Leost, M.; Meijer, L.; Sausville, E. A. Structure-Based Design Modifications of the Paullone Molecular Scaffold for Cyclin-Dependent Kinase Inhibition. *Anti-Cancer Drug Des.* **2000**, *15*, 53−66.

(20) Abraham, D. J.; Kellogg, G. E.; Holt, J. M.; Ackers, G. K. Hydropathic Analysis of the Non-Covalent Interactions between Molecular Subunits of Structurally Characterized Hemoglobins. *J. Mol. Biol.* **1997**, *272*, 613−632.

(21) Burnett, J. C.; Kellogg, G. E.; Abraham, D. J. Computational Methodology for Estimating Changes in Free Energies of Biomolecular Association upon Mutation. The Importance of Bound Water in Dimer-Tetramer Assembly for β37 Mutant Hemoglobins. *Biochemistry* **2000**, *39*, 1622−1633.

(22) Burnett, J. C.; Botti, P.; Abraham, D. J.; Kellogg, G. E. Accessible Method for Estimating Free Energy Changes Resulting from Site-Specific Mutations of Biomolecules: Systematic Model Building and Structural/Hydropathic Analysis of Deoxy and Oxy Hemoglobins. *Proteins*: *Struct., Funct., Genet.* **2001**, *42*, 355−377.

(23) Marabotti, A.; Balestreri, L.; Cozzini, P.; Mozzarelli, A.; Kellogg, G. E.; Abraham, D. J. HINT Predictive Analysis of Binding between Retinol Binding Protein and Hydrophobic Ligands. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2129−2132.

(24) Kellogg, G. E.; Burnett, J. C.; Abraham, D. J. Very Empirical Treatment of Solvation and Entropy: a Force Field Derived from Log $P_{o/w}$. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 381−393.

(25) Cashman, D. J.; Rife, J. P.; Kellogg, G. E. Which Aminoglycoside Ring is most Important for Binding? A Hydropathic Analysis of Gentamicin, Paromomycin, and Analogues. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 119−122.

(26) Bartlett, P. A.; Hanson, J. E.; Giannousis, P. P. Potent Inhibition of Pepsin and Penicillopepsin by Phosphorus-Containing Peptide Analogues. *J. Org. Chem.* **1990**, *55*, 6268−6274.

(27) Berliner, L. T. *Thrombin Structure and Function*; Plenum Press: New York, 1992.

(28) Brandstetter, H.; Turk, H.; Hoeffken, D.; Grosse, H. W.; Stürzebecher, D.; Martin, J.; Edwards, P. A.; Bode, B. F. Refined 2.3 Å X-Ry Cystal Sructure of Bovine Thrombin Complexes Formed with the Benzamidine and Arginine-Based Thrombin-Inhibitors NAPAP, 4-TAPAP and MQPA. A Starting Point for Improving Antithrombotics. *J. Mol. Biol.* **1992**, *226*, 1085−1099.

(29) Engh, R. A.; Brandstetter, H.; Sucher, G.; Eichinger, A.; Baumann, U.; Bode, W.; Huber, R.; Poll, T.; Rudolph, R.; von der Saal, W. Enzyme Flexibility, Solvent and "Weak" Interactions Characterize Thrombin-Ligand Interactions: Implications for Drug Design. *Structure* **1996**, *4*, 1353−1362.

(30) Chirgadze, N. Y.; Sall, D. J.; Briggs, S. L.; Clawson, D. K.; Zhang, M.; Smith, G. F.; Schevitz, R. W. The Crystal Structures of Human α-Thrombin Complexed with Active Site-Directed Diamino Benzo[*b*]thiophene Derivatives: a Binding Mode for a Structurally Novel Class of Inhibitors. *Protein Sci.* **2000**, *9*, 29−36.

(31) Banner, D. W.; Hadváry, P. Crystallographic Analysis at 3.0 Å Resolution of the Binding to Human Thrombin of Four Active Site-Directed Inhibitors. *J. Biol. Chem.* **1991**, *266*, 20085−20093.

(32) Stürzebecher, J.; Walsmann, P.; Voigt, B.; Wagner, G. Inhibition of bovine and human thrombins by derivatives of benzamidine. *Thromb. Res.* **1984**, *36*, 457−465.

(33) Kikumoto, R.; Tamao, Y.; Tezuka, T.; Tonomura, S.; Hara, H.; Ninomiya, K.; Hijikata, A.; Okamoto, S. Selective Inhibition of Thrombin by (2*R*,4*R*)-4-Methyl-1-[N^2-[(3-Methyl-1,2,3,4-Tetrahydro-8-Quinolinyl)Sulfonyl]-L-Arginyl)]-2-Piperidinecarboxylic Acid. *Biochemistry* **1984**, *23*, 85−90.

(34) Sall, D. J.; Bastian, J. A.; Briggs, S. L.; Buben, J. A.; Chirgadze, N. Y.; Clawson, D. K.; Denney, M. L.; Giera, D. D.; Gifford-Moore, D. S.; Harper, R. W.; Hauser, K. L.; Klimkowski, V. J.; Kohn, T. J.; Lin, H.; McCowan, J. R.; Palkowitz, A. D.; Smith, G. F.; Takeuchi, K.; Thrasher, K. J.; Tinsley, J. M.; Utterback, B. G.; Yan, S. B.; Zhang, M. Dibasic benzo[*b*]thiophene derivatives as a novel class of active site-directed thrombin inhibitors. 1. Determination of the serine protease selectivity, structure−activity relationships, and binding orientation. *J. Med. Chem.* **1997**, *40* (22), 3489−3493.

(35) Kurinov, I. V.; Harrison, R. W. Prediction of New Serine Proteinase Inhibitors. *Struct. Biol.* **1994**, *1*, 735−743.

(36) Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R. The geometry of the reactive site of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr.* **1983**, *B39*, 480−490.

(37) Turk, D.; Stürzebecher, J.; Bode, W. Geometry of binding of the Nα-tosylated piperidides of *m*-amidino-, *p*-amidino- and *p*-guanidino phenylalanine to thrombin and trypsin. *FEBS Lett.* **1991**, *287*, 133−138.

(38) Mares-Guia, M.; Shaw, E. Studies on the active center of trypsin. *J. Biol. Chem.* **1965**, *240* (4), 1579−1585.

(39) Stürzebecher, J.; Markwardt, F.; Voigt, B.; Wagner, G.; Walsmann, P. Cyclic amides of Nα-arylsulfonylaminoacylated 4-amidinophenylalanine − tight binding inhibitors of thrombin. *Thromb. Res.* **1983**, *29*, 635−642.

(40) Rhee, S.; Parris, K. D.; Hyde, C. C.; Ahmed, S. A.; Miles, E. W.; Davies, D. R. Crystal structures of a mutant (βK87T) tryptophan synthase $\alpha_2\beta_2$ complex with ligands bound to the active site of the α- and β-subunits reveal ligand-induced conformational changes. *Biochemistry* **1997**, *36*, 7664−7680.

(41) Schneider, T. R.; Gerhardt, E.; Lee, M.; Liang, P.-H.; Anderson, K. S.; Schlichting, I. Loop closure and intersubunit communication in tryptophan synthase. *Biochemistry* **1998**, *37*, 5394−5406.

(42) Weyand, M.; Schlichting, I. Crystal structure of wild-type tryptophan synthase complexed with the natural substrate indole-3-glycerol phosphate. *Biochemistry* **1999**, *38*, 16469−16480.

(43) Sachpatzidis, A.; Dealwis, C.; Lubetsky, J. B.; Liang, P.-H.; Anderson, K. S.; Lolis, E. Crystallographic studies of phosphonate-based α-reaction transition state analogues complexed to tryptophan synthase. *Biochemistry* **1999**, *38*, 12665−12674.

(44) Cheng, Y.; Prusoff, W. H. Relationship between the Inhibition Constant (Ki) and the Concentration of Inhibitor which Causes 50% Inhibition (I_{50}) of an Enzymatic Reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099−3108.
(45) Kirschner, K.; Wiskocil, R. L.; Foehn, M.; Rezeau, L. The Tryptophan Synthase from *E. Coli*. An Improved Purification Procedure for the α-Subunit and Binding Studies with Substrate Analogues. *Eur. J. Biochem.* **1975**, *60*, 513−523.
(46) Finn, J.; Langevine, C.; Birk, I.; Birk, J.; Nickerson, K.; Rodaway, S. Rational Herbicide Design by Inhibition of Tryptophan Biosynthesis. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2297−2302.
(47) Weischet, W. O.; Kirschner, K. Steady-state kinetic studies of the synthesis of the indoleglycerol phosphate catalyzed by the α subunit of tryptophan synthase from *Escherichia Coli*. Comparison with the $\alpha_2\beta_2$-complex. *Eur. J. Biochem.* **1976**, *65*, 375−385.
(48) Marabotti, A.; Cozzini, P.; Mozzarelli, A. Novel Allosteric Effectors of the Tryptophan Synthase $\alpha_2\beta_2$ Complex Identified by Computer-Assisted Molecular Modeling. *Biochim. Biophys. Acta* **2000**, *1476*, 287−299.
(49) Khan, A. R.; Parrish, J. C.; Fraser, M. E.; Smith, W. W.; Bartlett, P. A.; James, M. N. G. Lowering the Entropic Barrier for Binding Conformationally Flexible Inhibitors to Enzymes. *Biochemistry* **1998**, *37*, 16839−16845.
(50) Fraser, M. E.; Strynadka, N. C. J.; Bartlett, P. A.; Hanson, J. E.; James, M. N. G. Chrystallographic Analysis of Transition State Mimics Bound to Penicillopepsin: Phosphorus Containing Peptide Analogues. *Biochemistry* **1992**, *31*, 5201−5214.
(51) James, M. N. G.; Sielecki, A. R.; Hayakawa, K.; Gelb, M. H. Crystallographic Analysis of Transition State Mimics Bound to Penicillopepsin: Difluorostatine- and Difluorostatone-Containing Peptides. *Biochemistry* **1992**, *31*, 3872−3886.
(52) Cronin, N. B.; Badasso, M. O.; Tickle, I. J.; Dreyer, T.; Hoover, D. J.; Rosati, R. L.; Humblet, C. C.; Lunney, E. A.; Cooper, J. B. X-ray Structures of Five Renin Inhibitors Bound to Saccharopepsin: Exploration of Active-Site Specificity. *J. Mol. Biol.* **2000**, *303*, 745−760.
(53) Zuegg, J.; Gruber, K.; Gugganig, M.; Wagner, U. G.; Kratky, C. Three- dimensional structures of enzyme−substrate complexes of the hydroxynitrile lyase from *hevea brasiliensis*. *Protein Sci.* **1999**, *8*, 1990−2000.
(54) LaLonde, J. M.; Levenson, M. A.; Roe, J. J.; Bernlohr, D. A.; Banaszak, L. J. Adipocyte Lipid Binding Protein Complexed with Arachidonic Acid. *J. Biol. Chem.* **1994**, *269*, 25339−25347.
(55) LaLonde, J. M.; Bernlohr, D. A.; Banaszak, L. J. X-ray Crystallographic Structures of Adipocyte Lipid-Binding Protein Complexed with Palmitate and Hexadecanesulfonic Acid. Properties of Cavity Binding Sites. *Biochemistry* **1994**, *33*, 4885−4895.
(56) Xu, Z.; Bernlohr, D. A.; Banaszak, L. J. The adipocyte lipid-binding protein at 1.6 Å resolution. Crystal structures of the apoprotein and with bound saturated and unsaturated fatty acids. *J. Biol. Chem.* **1993**, *268*, 7874−7884.
(57) Buelt, M. K.; Bernlohr, D. A. Modification of the Adipocyte Lipid Binding by Sulfhydryl Reagents and Analysis of the Fatty Acid Binding Domain. *Biochemistry* **1990**, *29*, 7408−7413.
(58) Richieri, G. V.; Ogata, R. T.; Kleinfeld, A. M. Equilibrium constants for the binding of fatty acids with fatty acid-binding proteins from adipocyte, intestine, heart and liver measured with the fluorescent probe ADIFAB. *J. Biol. Chem.* **1994**, *269*, 23918−23930.
(59) Abraham, D. J.; Leo, A. J. Extension of the Fragment Method to Calculate Amino Acid Zwitterion and Side Chain Partition Coefficients. *Proteins*: *Struct., Funct., Genet.* **1987**, *2*, 130−152.
(60) Mitchell, J. B. O.; Laskowski, R. A.; Alex, A.; Thornton, J. M. Bleep-Potential of mean force describing protein−ligand interactions: I. Generating Potential. *J. Comput. Chem.* **1999**, *20* (11), 1165−1176.
(61) Mitchell, J. B. O.; Laskowski, R. A.; Alex, A.; Forster, M. J.; Thornton, J. M. Bleep-Potential of mean force describing protein−ligand interactions: II. Calculation of binding energies and comparison with experimental data. *J. Comput. Chem.* **1999**, *20* (11), 1177−1185.
(62) Beveridge, D. L.; Di Capua, F. M. Free Energy via Molecular Simulation: Applications to Chemical and Biomolecular Systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−492.
(63) Kollman, P. Free energy calculations-applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.
(64) Åqvist, J.; Medina, C.; Samuelsson, J. E. A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng.* **1994**, *7* (3), 385−391.
(65) Hansson, T.; Marelius, J.; Åqvist, J. Ligand Binding Affinity Prediction by Linear Interaction Energy Methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27−35.
(66) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.
(67) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein−ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791−804.
(68) Terp, G. E.; Johansen, B. N.; Christensen, I. T.; Jorgensen, F. S. A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein−ligand binding affinities. *J. Med. Chem.* **2001**, *44*, 2333−2343.
(69) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein−ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337−356.
(70) Head, R. D.; Smythe, M. L.; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R. VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. *J. Am. Chem. Soc.* **1996**, *118*, 3959−3969.
(71) Böhm, H.-J. Prediction of Binding Constants of Protein Ligands: a Fast Method for the Prioritization of Hits Obtained from *De Novo* Design or 3D Database Search Programs. *J. Comput.-Aided Mol. Des.* **1998**, *12* (4), 309−323.
(72) Weber, P. C.; Pantoliano, M. W.; Simons, D. M.; Salemme, F. R. Structure-based design of synthetic azobenzene ligands for streptavidin. *J. Am. Chem. Soc.* **1994**, *116*, 2717−2724.
(73) Finer-Moore, J.; Fauman, E. B.; Foster, P. G.; Perry, K. M.; Santi, D. V.; Stroud, R. M.. Refined Structures of Substrate-Bound and Phosphate-Bound Thymidylate Synthase from Lactobacillus Casei. *J. Mol. Biol.* **1993**, *232* (4), 1101−1116.
(74) Dill, K. A. Additivity Principles in Biochemistry. *J. Biol. Chem.* **1997**, *272*, 701−704.
(75) Berman, H. M.; Bhat, T. N.; Bourne, P. E.; Feng, Z.; Gilliland, G.; Weissig, H.; Westbrook, J. The Protein Data Bank and the Challenge of Structural Genomics. *Nat. Struct. Biol.* **2000**, *7*, 957−959.
(76) Hansch, C.; Leo, A. J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley and Sons, Inc.: New York, 1979.
(77) Zanotti, G.; Berni, R.; Monaco, H. Crystal structure of liganded and unliganded forms of bovine plasma retinol binding protein. *J. Biol. Chem.* **1993**, *268*, 10728−10738.
(78) Zanotti, G.; Malpeli, G.; Berni, R. The interaction of *N*-ethyl retinamide with plasma retinol-binding protein (RBP) and the crystal structure of the retinoid-RBP complex at 1.9 Å resolution. *J. Biol. Chem.* **1993**, *268*, 24873−24879.
(79) Zanotti, G.; Marcello, M.; Malpeli, G.; Folli, C.; Sartori, G.; Berni, R. Crystallographic studies on complexes between retinoids and plasma retinol-binding protein. *J. Biol. Chem.* **1994**, *26*, 29613−29620.
(80) Liaw, S. H.; Jun, G.; Eisenberg, D. Interactions of Nucleotides with Fully Unadenylylated Glutamine Sinthetase from S*almonella Typhimurium*. *Biochemistry* **1994**, *33*, 11184−11188.
(81) Lolis, E.; Petsko, G. A. Crystallographic Analysis of the Complex between Triosephosphate Isomerase and 2-Phosphoglycolate at 2.5 Å Resolution: Implication for Catalysis. *Biochemistry* **1990**, *29*, 6619−6625.
(82) Scapin, G.; Blanchard, J. S.; Sacchettini, J. C. Three-dimensional structure of *Escherichia Coli* dihydrodipicolinate reductase. *Biochemistry* **1995**, *34*, 3502−3512.
(83) Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Hosur, M. V.; Sowder, R. C.; Cachau, R. E.; Collins, J.; Silva, A. M.; Erickson, J. W. Crystal Structures of Native and Inhibited Forms of Human Cathepsin D: Implications for Lysosomal Targeting and Drug Design. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6796−6800.

JM0200299