

# Designing Ligands to Bind Tightly to Proteins

*George M. Whitesides\*, Phillip W. Snyder,  
Demetri T. Moustakas, and Katherine A. Mirica*

## 1. Introduction

In a scientific reductionist's view, a cell is a compartment containing a large ensemble of reacting molecules. Molecules enter the cell from outside; metabolic waste products leave it; the reactions occurring inside the membrane generate the new molecules required both for the cell to replicate itself, and for it to respond to changes in its environment. The intense activity of the cell at the molecular scale is almost entirely controlled by catalysts (most of which are proteins, but among which are a few very important catalytic nucleic acids, or aggregates of proteins and nucleic acids), or by other proteins that perform specific functions (for example, facilitating passage of ions or molecules across the cell membrane, or sensing the environment outside the cell).

A central task faced by the cell is that of ensuring the correct level of specificity in its reactions. That specificity, in turn, is accomplished through what is called "molecular recognition": loosely, the structure-specific association of one molecule with another.<sup>1-10</sup> In an important sense, molecular recognition is the most fundamental process in biochemistry, as

\*Department of Chemistry and Chemical Biology, Harvard University, Cambridge MA 02138, USA, e-mail: gwhitesides@gmwgroup.harvard.edu.

its fidelity enables the exquisite control the cell exercises over its composition and chemical activity.

Molecular recognition is typically a non-covalent event: the two or more interacting molecules form a structured aggregate without forming a covalent bond. Enzymatic catalysis is also a kind of molecular recognition: that is, the preferential recognition (stabilization) of the transition state relative to the ground state of the reactants. Molecular recognition in biology almost always involves releasing some or all of the water (or other) molecules normally in contact with the interacting surfaces, and bringing these surfaces together; it may also involve changes in the conformation of the molecules, and in other details (for example, in the extent of ionization of functional groups, and in the coordination of ions or molecules).<sup>11-13</sup> The interactions involved are typically individually comparable to thermal energy ( $kT$ ) — from tenths of a kcal/mol to a few kcal/mol.<sup>14,15</sup>

Given the ubiquity and enormous importance of molecular recognition in biology, one might assume that we chemists would have “figured it out”: that is, rationalized the interactions involved, and developed methods of predicting interactions involving new structures. In fact, we have not. The prototypical problems of molecular recognition — understanding how and why a ligand (which might be another protein) binds to a protein, or understanding how a polypeptide folds into a protein — were well-posed but unsolved problems 50 years ago, and largely remain so today.<sup>16-22</sup>

The importance of molecular recognition is abundantly clear: it is the basis for a reductionist understanding of life; it is essential for the rational design of ligands to bind to proteins of known structure, and therefore an important contributor to the early stages of drug design. Molecular recognition in water is also a subject that is technically very difficult. Chemists — in thinking about molecular recognition — have tended to work on the much easier problems posed by systems soluble in aprotic organic solvents (especially chloroform and methylene chloride, chosen largely because they are excellent solvents for NMR spectroscopy and relatively simple liquids).<sup>1-3,8,23,24</sup> These problems have produced an abundance of

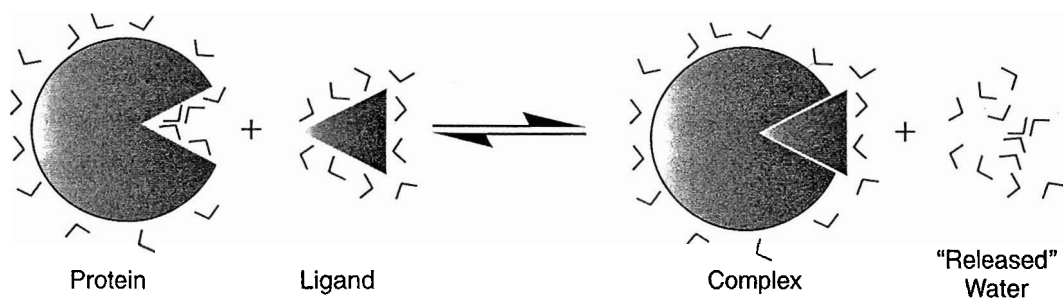
interesting science, but very little that is relevant to molecular recognition in water. In fact, they may have contributed to a misunderstanding of the problem: because it is possible to develop workable rules of thumb for molecular recognition in certain systems (particularly those involving hydrogen bonding) in these solvents that do not explicitly include solvent or solvation, they have given the sense that molecular recognition is a problem involving only two players — “host” and “guest” — rather than three — “host,” “guest,” and “solvent.” In the study of molecular recognition in chloroform (and the like) the solvent is largely ignored, as the penalty of desolvating surface area of the host and guest is negligible. Molecular recognition in water is *much* more complicated, and it remains the case that: (i) we cannot rationalize why a ligand does or does not bind to a particular protein, other than after-the-fact or very qualitatively; (ii) we cannot predict the structures of new ligands that will bind tightly to a protein, other than by close analogy with known ligands; (iii) we cannot even predict the effects of seemingly minor perturbations to the structure of known, characterized ligands on their binding.

## **2. What are the Difficulties?**

One should perhaps simplify the question, and ask “What *is* the difficulty?” to which the short answer would be “water.”<sup>25–27</sup> The story is more complicated than that, of course, and all of the interacting parties — protein, ligand, and medium — contribute to the difficulty of understanding protein-ligand interactions in aqueous media. Still, liquid water is an astonishing material, and after decades of intense study, it remains mysterious, unexplained in many of its properties, and unpredictable in many of its interactions. Water has a number of characteristics that collectively make it uniquely complicated: (i) It is small, has a large dipole moment, and is capable of donating and accepting hydrogen bonds. Together, these characteristics give water a high cohesive energy density and a large heat capacity relative to other molecules of similar molecular weight; (ii) Water has a high cohesive

energy density: it associates with itself through networks of hydrogen bonds in ways that simply are not understood. Phrases such as “structured water,” “ice-like water,” “near-surface water,” and “flickering cluster” speak to the elusive nature of the structure of this network.<sup>27</sup> Even the median number of hydrogen bonds in which a water molecule is involved is still undefined: after years of assuming the number was four, it now seems to be closer to three<sup>14,28,29</sup>; (iii) Its small molar volume and extensive structure underscore the importance of configurational entropy for the structure of water — this structure depends strongly on temperature; (iv) Just as the structure of the hydrogen bond network depends strongly on temperature, so too do its properties: the temperature dependence of the dielectric constant is especially important in estimating electrostatic interactions; (v) It has a very high surface tension, a parameter that probably reflects — in some way that is not well understood — the change between the mean strength of the hydrogen bonds of water molecules at the interface between water and air, and that of water molecules in bulk water, and the order of “surface” and “bulk” water. The surface tension is important in estimating energetics in “cavity” models for solubility, and in rationalizing the hydrophobic effect; (vi) Water is an excellent solvent for ions and polar molecules, species so common within the cellular milieu. It associates strongly with ions by ion-dipole interactions and with polar-protic functionality by hydrogen bonding. In both cases, the associations are enthalpically favorable but entropically unfavorable; (vii) The properties of water still include surprises that are difficult to understand: for example, studies of zeta potential suggest that hydroxide ion apparently preferentially partitions to the water-non-polar interface.<sup>30,31</sup> We do not know why.

The fundamental interactions among species dissolved in water, and their interactions with water, are centrally important in trying to understand phenomena such as the hydrophobic effect, and electrostatic interactions in water.<sup>14,27,32,33</sup> These phenomena — which seem to reflect the ensemble properties of water, large molecular surfaces, and networks of charges — provide conceptual and technical difficulties that are distinct from those



**Fig. 1.** The recognition of a ligand by a protein in water. Formation of a protein-ligand complex requires the dehydration of the molecular surface area of both the ligand and the active site of the protein.

associated with simple pair-wise interactions of molecules. In particular, the current model of the hydrophobic effect as a phenomenon dominated by the entropy of formation of water — whatever its structure — at the interface between non-polar materials and water is probably oversimplified in its neglect of different types of interactions (for example, dispersion interactions) between non-polar but polarizable molecules or functional groups. Similarly, efforts to provide a unified rationalization of biochemical interactions in terms of electrostatic interactions are too simple to be general and predictive.<sup>11,34–36,37</sup> Although these efforts must, in some sense, be correct (chemistry is largely electrostatic interactions operating within the constraints of quantum mechanics), these treatments tend to be very approximate. Biochemical systems tend to be too large to treat in detail using *ab initio* methods, and estimation of entropies requires running molecular dynamics simulations for periods of time that remain impractical.<sup>38–47</sup>

## 2.1. “Swatting Mosquitoes”

The picture that has emerged in thinking about biochemical phenomena is, from the point of view of simplicity, wholly unsatisfactory. It is, in a qualitative sense, a problem of systems containing large numbers of molecules with many small energetic interactions — each poorly defined in

thermodynamic terms. The enthalpy of binding is a sum of both unfavorable terms (desolvation and configurational restriction) and favorable terms (numerous electrostatic interactions between host and guest) of similar magnitude. Entropic terms — which remain even more difficult to estimate — make equally important contributions: differences between the bound and unbound states in translation, rotation, conformation, vibration and solvent ordering are (at least) poorly defined and in all cases impossible to measure directly.<sup>12,13,17,18,48–50</sup> It is a problem that has the feeling of swatting mosquitoes: kill one, and so many remain that it seems to make no difference.

Science moves most rationally toward the solution of a complex problem when that problem can be broken down into parts with a clearly differentiated hierarchy in importance. When there are many contributors of equal importance, it is difficult to know what problem to focus on, and difficult to know when one has been solved until all have been solved simultaneously. The hypothesis is evident: the problem of molecular recognition in water will not resolve itself into one interaction sufficiently dominant to form the basis for simple rules of thumb. Although far from resolving the issue, this realization suggests that computation and simulation will be *required* for solving this problem. Computers do better than people at keeping track of many small numbers in a problem with many parts.

## 2.2. “If You can’t Measure It, You can’t Manage It”

The difficulty in understanding binding of proteins and ligands in water is such that progress requires the application of all plausible tools. Fundamentally, the field of molecular recognition in water requires both reliable, accurate, and interpretable experimental data, and relevant predictive theory. The scientific method — the best template we have for the systematic resolution of a complex problem — proceeds from observation to theory to experimental test of theory. In molecular recognition, we are still in the situation of needing data on which to build and test an adequate theory.

The objective of our program in understanding the molecular recognition of ligands by proteins in aqueous buffer has been to provide good experimental information, against which to calibrate theory and simulation.

A number of experimental and computational techniques have been developed in the last several decades that, collectively, have made major steps toward a process for the rational design of ligands. (i) *Thermodynamics*. On the one hand, many techniques make it possible to measure binding constants — often with high accuracy — and thus to estimate values of  $\Delta G^\circ$ . On the other, separation of  $\Delta G^\circ$  into values of  $\Delta H^\circ$  and  $T\Delta S^\circ$  has been more problematic. Measurement of  $\Delta G^\circ$  as a function of temperature — van't Hoff analysis — is now accepted to be fundamentally too error-prone for use in biochemistry. Too many factors — the structure of proteins, the conformations of ligands, the properties of water — change with temperature to make it possible to isolate parameters that are specific to the molecular recognition event.<sup>51–54</sup> Unfortunately, many thermodynamic data in the literature have been obtained by van't Hoff analysis, and are consequently too unreliable to be useful.<sup>55</sup> Fortunately, a technique that has recently become widely used as a result of the introduction of good commercial apparatus is isothermal titration calorimetry (ITC), which solves many of the problems of estimating enthalpies by measuring heat directly at constant temperature. (We discuss this technique in more detail in the next section). (ii) *Structure*. X-ray crystallography (and to an extent, NMR spectroscopy) have provided enormously useful information concerning the structure of representative soluble proteins, both with and without bound ligands. The development of synchrotron light sources has dramatically increased the quality and decreased the cost of obtaining these structures. High-energy light sources and new NMR techniques have extended the information available using them to include structural information concerning some of the water molecules at the surface of the protein (and particularly in the region of the active site).<sup>12,48,56–58</sup> Site-specific mutagenesis makes it routine to change the amino acid sequence in proteins to test hypotheses that depend on this sequence.<sup>36,59–62</sup> These advances, although invaluable,

still leave critical gaps. Most of the water molecules in a protein are not defined by crystallography, and no information is available concerning the water molecules solvating the complementary binding face of the ligand. Large changes in structure of a protein that might occur on binding a ligand can be characterized by X-ray crystallography, but small ones — especially small ones distributed over many amino acids — cannot be, and even if they could, it is not clear what to do with the information. “Structure” has been extremely useful in defining what *not* to do to get binding (that is, the rational design of ligands that do *not* bind to a protein is now straightforward), but much less useful in designing ligands that *do* bind. Emerging techniques in NMR spectroscopy may someday supplement our understanding of dynamics in molecular recognition, but these techniques are still too highly specialized to be applied broadly.<sup>48,58,63</sup> (iii) “*Molecular Association*”. Association of molecules occurs in a very wide range of processes in chemistry. Phase transitions are forms of association. The properties of materials often reflect association of molecules. Subjects such as molecular recognition have been extensively studied in non-aqueous solutions. All of this information is relevant to the protein-ligand problem, but has not been integrated into a unified theory of molecular association. (iv) “*Water*”. Water has been extensively studied experimentally and theoretically, with results that have provided some (but certainly not complete) clarity.<sup>14,27,37</sup> The structure of water at or close to interfaces is a particularly contentious subject, and complex aqueous media — water containing dissolved organic molecules and ions — is nearly *terra incognita*.<sup>27</sup>

### 2.3. Theory, Computation, and Simulation

Theory and simulation is not *our* primary business, but understanding the theory with which the experimental data must accord is essential for the design of a useful experimental program. The availability of powerful computers and networks of computers has increased the ability of simulations to handle large numbers of atoms, but the potential functions



used in these calculations are still approximate and of uncertain accuracy. Density functional theory (DFT) — although still a semi-empirical method — provides estimates of energies with high accuracies for certain classes of problems.<sup>64</sup> The many advances in theory and simulation still, however, fall short of what is needed in an important way. Molecular recognition in water is a many-atom problem with contributions of similar magnitude from enthalpy and entropy. The most widely used computational method for estimating entropies in complex systems of many molecules is the quasi-harmonic approximation method.<sup>41</sup> This method uses data from a molecular dynamics trajectory of a biomolecular system to parametrize a quasi-harmonic Hamiltonian for the system, from which the entropy of the system can be estimated. Simulations of realistic systems of protein, ligands, and medium still require so much time that it is impractical to run them over sufficient intervals to generate useful estimates of entropies. (And again, even if they could be run for very long times, or if their speed could be increased by many orders of magnitude, it is not clear how accurate the resulting entropies would be, as everything ultimately depends on potential functions of uncertain accuracy.) Although it is possible in some cases to build empirical effective potential functions that estimate  $\Delta G^\circ$ , the ultimate accuracy of all theoretical methods has to be judged against their ability to predict the results of experiments before they are carried out. By that criterion, for whatever reasons, theory has not yet reached a level where it provides results that are broadly useful to designers of ligands.

### **3. What is New? Back to the Future with Isothermal Titration Calorimetry (ITC)**

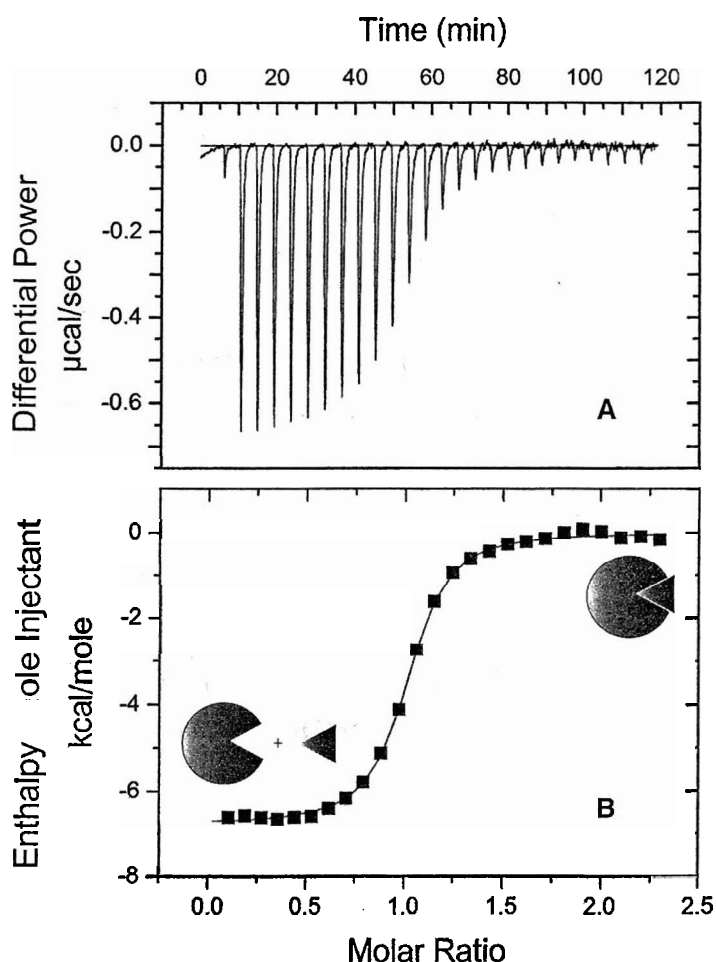
There are several glasses through which to regard molecular recognition: we think of it primarily in terms of thermodynamics — the difference in free energies between protein and ligand separately in solution, and protein and ligand associated in solution. Ultimately, one would wish to predict this difference for new structures and specific media in order to design ligands

that bind tightly; at present, the task is to rationalize the thermodynamic data that are available (or can be obtained) using proteins and ligands of known structure.

Thermodynamics is an approach to the study of molecular recognition that has been used perhaps more than any other.<sup>6,53,65-69</sup> The unreliability of the separation of  $\Delta G^\circ$  into  $\Delta H^\circ$  and  $T\Delta S^\circ$  has, however, been such that it has been difficult to develop a set of interpretable thermodynamic parameters on which to build a theory. Isothermal calorimetry has changed this unfortunate situation for the better (although it remains to be seen if a useful theory will emerge as a consequence).<sup>70,71</sup> ITC is, of course, a technique that has been a part of calorimetry for many decades, but commercial instruments that have the accuracy to measure heats of association of ligands and proteins have been available for only about two decades. These instruments still require a substantial amount of protein (~1 mg for an experiment, and ~50 mg for a complete set of experiments), but preparing pure protein on this scale is no longer impossibly arduous, given the techniques of genetic engineering and protein over-expression.

Commercial instruments now being introduced will provide the ability to generate data entirely automatically, and are thus appropriate for providing the substantial quantities of replicate experiments that may be necessary to obtain estimates of statistical significance for experiments that require reliable identification of trends in small numbers. These data could provide the basis for investigations of phenomena that have previously eluded explanation because the magnitude of their effect on the thermodynamics of binding is small. The thermodynamic solvent isotope effect, for example, which is the difference between thermodynamics of binding in light and heavy waters, has been difficult to address experimentally because of the small magnitude of the effect (typically just larger than the error of measuring enthalpies of binding).<sup>72,73</sup> Similarly, the incremental heat capacity of binding is known to be a hallmark of the hydrophobic effect, but the relationship between this parameter and the structure of water near hydrophobic surface area is far from well established. Resolving these issues will also come from expanded

structure-thermodynamic relationships: automated calorimetry should facilitate the collection of data for large numbers of ligands of different structures, perhaps in different media and at different temperatures. Such aggressive programs directed at understanding the thermodynamics of molecular recognition have, until now, been impractical.



**Fig. 2.** (A) A representative ITC thermogram for the binding of a protein to a ligand. Small quantities of ligand are periodically injected into a protein solution, and the heat evolved by each injection ( $\Delta H$ ) is recorded. As the concentration of ligand in the sample cell approaches saturation, the magnitude of the heat evolved by each injection decreases. (B) A binding curve is obtained by plotting the  $\Delta H$  measured for each injection against the ratio of ligand to protein at the time of the injection. This binding curve yields the  $K_d$ ,  $\Delta H$  and  $n$ , the binding stoichiometry.

ITC is normally carried out by adding sub-stoichiometric aliquots of ligand to a solution of protein, and measuring the heat evolved.<sup>74</sup> Analysis of this heat with a known model of binding provides three useful parameters: (i) a direct measure (the total integrated heat evolved on binding) of the enthalpy of binding at constant temperature; (ii) the binding constant; and (iii) the stoichiometry of binding. Because the experiment is carried out at constant temperature, most of the potential artifacts of van't Hoff analysis disappear.

In ITC, the evolved heat at constant pressure gives the enthalpy of binding ( $\Delta H^\circ$ ) directly. The dissociation constant gives  $\Delta G^\circ$  [Eq. (1)]; this parameter can usually be obtained by some entirely independent measurement (enzymatic activity, or spectrophotometry). The difference between them thus provides  $T\Delta S^\circ$  [Eq. (2)]. ITC offers, for the first time, the ability to generate high-quality thermodynamic data characterizing binding of ligands to proteins under conditions relevant to biological chemistry.

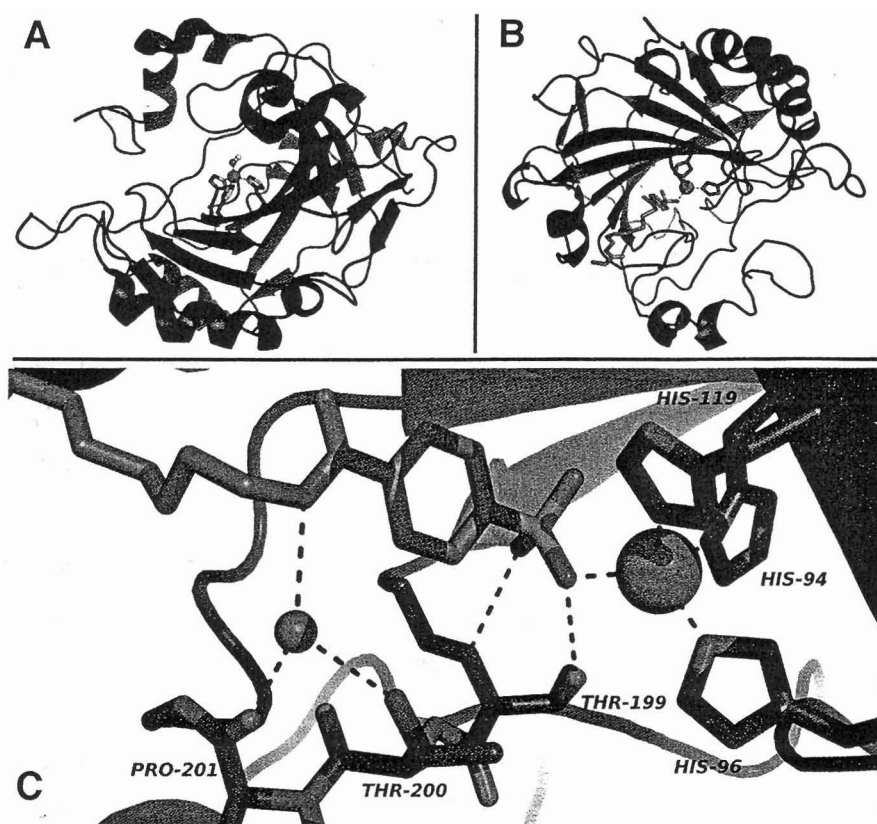
$$\Delta G^\circ = -RT \ln(K_d) \quad (1)$$

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ. \quad (2)$$

#### **4. A Model System: Carbonic Anhydrase and Arylsulfonamides**

Most of our experimental work has involved the enzyme carbonic anhydrase (CA, E.C. 4.2.1.1), and studied the association of derivatives of benzenesulfonamide (as the anion:  $\text{ArSO}_2\text{NH}^-$ ) to the zinc ion in the active site of this enzyme. The choice of this system is based primarily on simplicity. The system of CA and arylsulfonamides offers: (i) an enzyme that is available commercially and inexpensively, and that has high structural stability (and thus few complications from changes in its structure on plasticity — on binding the arylsulfonamide anion); (ii) a class of active-site directed inhibitors — arylsulfonamides — that is easy to prepare synthetically, and whose binding at the active site of the CA is exceptionally

constant and well-characterized in geometry; (iii) an active site that includes a substantial hydrophobic surface, and thus provides a test-bed for studying hydrophobic interactions; (iv) accessible molecular surface proximate to the active site that is non-polar, and thus useful in other examinations of hydrophobic effects. We have summarized the information concerning CA as a model for physical-organic and biophysical studies of protein-ligand binding in an extensive review article.<sup>75</sup>



**Fig. 3.** Structures of carbonic anhydrase II bound to (A) water and to (B) a para-substituted benzenesulfonamide ligand with an oligoethylene glycol "tail."<sup>76</sup> (C) A more detailed view of the polar interactions between the protein and the ligand. The coordination bond between the nitrogen of the sulfonamide and the Zn(II) cofactor is shown as a dashed line. Hydrogen bonds between the nitrogen of the sulfonamide and the side-chain hydroxyl of THR-199 of the protein are shown as dashed lines as well. A network of hydrogen bonds between the amide of the *para*-substituted "tail" and two protein residues (PRO-201 and THR-200) are shown as dashed lines.

We have used CA as a model in a number of studies. The most relevant to considerations of thermodynamics simplifies the system even further by using it primarily as the basis for perturbational studies. In these studies, we leave the structure of the enzyme and the benzenesulfonamide group unchanged, vary only the structure of the substituents in the para-position of the benzene ring, and examine the influence of those variations on the thermodynamics of binding; these variations reflect interactions of the para-substituents with the surface of the protein proximate to, but not directly in, the active site. We believe this system is the simplest and best defined that can be devised for realistic studies of protein-ligand binding. We have carried out a number of studies using it; probably the most honest description of these studies is that they have consistently produced results that were — to us — *unexpected*: we draw the conclusion that we cannot presently accurately predict the results of even these kinds of perturbational experiments.

We have carried out a number of studies of the binding of arylsulfonamides to CA.<sup>76-85</sup> These include examinations of the influence of Coulombic interactions on binding (using protein charge ladders), of the relation of the structure of the aryl group and binding constant, and of hydrophobic substituents on the aryl ring and binding constants; all have given interesting results that have clarified the interactions that contributed to binding. Detailed interpretation of these results has also remained difficult.

In principle, the most interpretable types of studies should be the perturbation studies that have examined the influence of para-substituents on the aryl ring on the binding. Since these studies concern only binding outside of the immediate active site, and since they do not appear to influence the binding site region itself, we had expected them to minimize changes to the structure of the protein, and thus to provide the simplest case we could develop for studying molecular recognition in a system of protein and ligand. The results of these studies have been uniformly surprising: that is, they have not been in accord with current theory (or at least with our

understanding of current theory). The interpretation of these experiments is often complicated, and we give only one example to show the kinds of information that ITC yields, and the difficulty in reconciling this information with simple rationalizations of molecular recognition in this system. The papers we reference (and in the case of electrostatic effects, a review) give further views of these subjects.<sup>11,75</sup>

#### **4.1. Binding of p-R-C<sub>6</sub>H<sub>4</sub>SO<sub>2</sub>NH<sub>2</sub> to Carbonic Anhydrase**

We prepared a series of benzenesulfonamides with conformationally flexible ("floppy") oligomeric chains in the para-position (oligomers of ethylene glycol, glycine, and sarcosine).<sup>78,79,85</sup> These systems are important in one current approach to the rational design of ligands: this approach identifies a secondary binding site close to the active site (by NMR spectroscopy, by crystallography, and/or by computation), and attempts to construct a tightly binding inhibitor by linking groups that bind simultaneously at the active and the secondary sites. The question of interest is: "What should be the structure of the linker to maximize the value of  $\Delta G^\circ$  for the bivalent ligand?" Answering the question of the nature of bivalent interactions, and of the strategy that maximizes binding through bivalency, is a separate and engaging problem, but one for which it is clearly important to understand the interaction of the primary ligand and the linker (*without* the group targeted to the secondary binding site) with the protein.<sup>49,50</sup> We, therefore, prepared benzenesulfonamides with  $R = (\text{EG})_n$ ,  $(\text{Gly})_n$ , and  $(\text{Src})_n$ , measured values of heat evolved upon binding to CA (and thus  $\Delta H^\circ$ ) and  $\Delta G^\circ$ , and obtained  $T\Delta S^\circ$  from these two. The results of these studies were entirely unexpected.<sup>85</sup>

What we had anticipated was to observe either a trend to tighter binding with longer oligomeric chains (reflecting favorable enthalpy of interaction of the floppy chain with the surface of the protein) or a trend to weaker binding (reflecting loss in conformational entropy due to this interaction, or simply to a restriction in the volume of space accessible to the chain). In fact, we

observed — for all three series — *no* change: the binding constant was independent of the structure of the floppy chains. This independence was *not* due to the absence of interaction of the chains with the surface of the CA: both  $\Delta H^\circ$  and  $T\Delta S^\circ$  varied with the chain length.

This result was thus unexpected in three ways: (i) The binding constant was essentially independent of the length and structure of the floppy chain; (ii) There *were* changes in  $\Delta H^\circ$  and  $T\Delta S^\circ$ , but they precisely cancelled; (iii) The changes in  $\Delta H^\circ$  and  $T\Delta S^\circ$  were opposite to those we had expected: that is, as the chains in each series became long, the contribution to the enthalpy of binding we attributed to the chain became *less favorable*, and the contribution to the entropy of binding *less unfavorable*. How do we explain these very puzzling results? We do not, in fact, presently know. We have a rationalization, but it is very much *post facto*.<sup>85</sup> We are confident that these results are not experimental error, however, and we note that a number of other studies that we have carried out with CA have generated related but equally puzzling results. We conclude that we simply cannot use our current, qualitative, understanding of molecular recognition involving proteins in water in a way that successfully rationalizes magnitudes and contributions of enthalpy and entropy to free energy.

## 4.2. Entropy-Enthalpy Compensation

We also do not understand the apparent entropy-enthalpy compensation (EEC), the tendency for structural changes in ligands and proteins to *have* little or no effect on the free energy of binding because of compensating contributions from enthalpic and entropic terms. The occurrence of EEC is one of the puzzles waiting to be resolved in the study of molecular recognition in water. EEC has been claimed in many experimental investigations of the thermodynamics of biological systems. In studies in which van't Hoff analysis generated the values of  $\Delta H^\circ$  and  $T\Delta S^\circ$ , the potential for compensating errors, and for artifacts in the procedures, was so great that EEC was largely (and we believe correctly) dismissed as a



credible phenomenon.<sup>52,54,55</sup> When ITC is used to generate the data that underpin EEC, the potential for artifact is smaller (although still not zero). So: is EEC a real phenomenon, or is it an artifact? There are plausible, if qualitative, arguments that it should exist, and we are inclined to believe that it is real.<sup>6,86</sup> If it is, it hints at other relationships in the thermodynamics of biochemical systems whose basis we do not presently see clearly.

## 5. Assessment

Our studies of benzenesulfonamides binding to carbonic anhydrase provide the simplest biophysical system that we can construct with which to study the physical-organic chemistry of binding of a low molecular weight ligand to a globular protein. The combination of ITC (for thermodynamics) and X-ray crystallography (for structure) provides a solid basis for characterizing the entropies and enthalpies of binding, and for trying to relate trends in these values to trends in structure, and for comparisons with the results of simulations. The situation from these studies suggests that there is still a substantial amount of work — both experimental and conceptual — to be done before this simplest system can be claimed to be understood. We can rationalize some of its features.<sup>75</sup> For example: (i) The largest single contribution to binding is probably the interaction of the  $\text{Zn}^{+2}$  ion in the active site, the  $^-\text{HNSO}_3\text{-R}$  group to which it binds, and the network of hydrogen bonds — some directly between protein and sulfonamide anion, and some involving water — in the active site. (ii) The major influence of the structure of the arylsulfonamide on the binding is through its influence on the  $pK_a$  of the  $\text{H}_2\text{NSO}_2\text{Ar}$  group. (iii) Strongly hydrophobic substituents in the para position of the benzene ring increase binding through the hydrophobic effect. Others we cannot rationalize: (i) We do not understand the contributions of substituents to the enthalpy and entropy of binding; (ii) We do not understand the apparent importance of EEC in these (and other biochemical) systems; (iii) Our estimations of entropy and enthalpy in interactions that would plausibly be ascribed to the hydrophobic effect do

not suggest a system dominated by entropy (as one would expect from the most commonly accepted explanations of the hydrophobic effect; (iv) We do not understand the role of water — its release on binding of inhibitor, its role in forming hydrogen bonds in unbound and bound states, its role in mediating electrostatic effects — at all; and (v) We do not understand how to take into account the conformational mobility of the ligand (free in solution, and bound to the protein), or the plasticity of the protein (although we speculate that protein plasticity is small — or at least smaller than for most other proteins — for CA).

So, although we (and others) are making good progress in rationalizing the binding of arylsulfonamides to CA, the progress is more at the level of understanding what it is that we do not understand, than it is in understanding. We are — in the phraseology of the military — converting “unknown unknowns” into “known unknowns,” but we still have a long road to travel before converting “known unknowns” into “knowns.”

## 6. What Do We Need?

We close this chapter with a wish list. What is it that could make an important contribution to the rational design of ligands? This list is not unique, and other researchers in this field might come up with other subjects, but there would be broad agreement on most of these items. (i) *Reliable experiments in interpretable systems*. Although we have spoken enthusiastically of the virtues of ITC, it is still too difficult (slow and labor-intensive) to provide the large numbers of data that would be helpful in unraveling the relationships between structure and binding in this difficult field. It would also be extremely useful to have new techniques to measure thermodynamic parameters. There are, for example, no good ways of measuring thermodynamic parameters for materials with very low solubilities (e.g., hydrophobic materials — often the ones of the greatest interest). (ii) *Simulation*. It seems inevitable that computation and simulation will play a very important part in the future of this field: computers are the only practical way to add many small

contributions — often with opposite signs — with reasonable accuracy. Computation is still too slow, and especially too slow to provide estimates of entropies. (iii) *Theory*. There is very little about molecular recognition in water that we can truly claim to understand theoretically. The plasticity of proteins. The role and structure of water. The nature of the hydrophobic effect. The character of electrostatic effects. The tradeoff between exact fits (“lock-in-key”) with favorable enthalpy and unfavorable entropy, and loose fits (“horse-in-stall”) with less favorable enthalpy and less unfavorable configurational entropy. All, and more, are understood in only general terms. (iv) *Experiments in realistic media*. Much of biophysical chemistry is carried out in simple, dilute buffers. The inside of the cell — where water rarely achieves a thickness of more than a few monolayers — is a much more complex medium: approximately 300 g/L of organic materials, approximately 1 M ionic strength, high concentrations of zwitterions.<sup>27</sup> It would be most valuable to have some comparison between thermodynamic parameters in simple and complex media. (v) *Understanding of extra-thermodynamic relationships*. Is EEC real? How do we tell? If so, what is its origin, and why is it so common in biochemical systems. (vi) *Understanding of protein plasticity*. It is evident that the conceptual model of a protein as a rigid object, with a definite shape, is — in general — wrong. Many proteins are probably more like structured liquid crystals than the rigidly defined structures so elegantly suggested by crystallography. How does this plasticity influence binding? Time-resolved spectroscopic methods (electron spin resonance and infrared spectroscopies, for example) might give insight into the problems associated with protein dynamics on very short time scales, but these techniques are far from being common approaches in the biophysics community.<sup>57,58,87–89</sup> (vii) *New types of experiments*. New tools open new doors. Molecular recognition involving proteins is still starved for relevant data. We need everything: tools to define the structure of water, and to watch proteins at atomic resolution in real time, and to measure entropy changes, and to explore complex media. The list of measurements we would like to make is almost endless. (viii) *Better understanding of fundamentals*.

Finally, at the core, we need to understand the fundamentals of non-covalent interactions in water better. What is a hydrogen bond? The hydrophobic effect? Structured water at interfaces? Ionic atmospheres around proteins? Protein conformational mobility/plasticity? It may or may not be necessary to have a profound understanding of these and other fundamental interactions to build working engineering solutions to the problem of designing ligands to fit tightly to proteins, but it cannot hurt to have them.

## Acknowledgments

This work is supported by the National Institutes of Health (GM030367, GM051559). D.T.M. was supported by fellowship AI068605. K.A.M. acknowledges support from Eli Lilly.

## References

1. Lehn JM. (1990) Perspectives in supramolecular chemistry — from molecular recognition towards molecular information-processing and self-organization. *Angew Chem* **29**: 1304–1319.
2. Cram DJ. (1992) Molecular container compounds. *Nature* **356**: 29–36.
3. Hof F, Craig SL, Nuckolls C, Rebek J. (2002) Molecular encapsulation. *Angew Chem Int Ed* **41**: 1488–1508.
4. Houk KN, Leach AG, Kim SP, Zhang XY. (2003) Binding affinities of host-guest, protein-ligand, and protein-transition-state complexes. *Angew Chem Int Ed* **42**: 4872–4897.
5. Meyer EA, Castellano RK, Diederich F. (2003) Interactions with aromatic rings in chemical and biological recognition. *Angew Chem Int Ed* **42**: 1210–1250.
6. Williams DH, Stephens E, O'Brien DP, Zhou M. (2004) Understanding noncovalent interactions: Ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes. *Angew Chem Int Ed* **43**: 6596–6616.
7. Hof F, Diederich F. (2004) Medicinal chemistry in academia: Molecular recognition with biological receptors. *Chem Commun* 477–480.

8. Reed CA. (2005) Molecular architectures — guest editorial. *Acc Chem Res* **38**: 215–216.
9. Butlin NG, Meares CF. (2006) Antibodies with infinite affinity: Origins and applications. *Acc Chem Res* **39**: 780–787.
10. Hirsch AKH, Fischer FR, Diederich F. (2007) Phosphate recognition in structural biology. *Angew Chem Int Ed* **46**: 338–352.
11. Gitlin I, Carbeck JD, Whitesides GM. (2006) Why are proteins charged? Networks of charge-charge interactions in proteins measured by charge ladders and capillary electrophoresis. *Angew Chem Int Ed* **45**: 3022–3060.
12. Homans SW. (2007) Water, water everywhere — except where it matters? *Drug Discov Today* **12**: 534–539.
13. Homans SW. (2007) Dynamics and thermodynamics of ligand-protein interactions, Bioactive conformation i, Vol. 272, Topics in Current Chemistry, pp. 51–82.
14. Dill KA, Truskett TM, Vlachy V, Hribar-Lee B. (2005) Modeling water, the hydrophobic effect, and ion solvation. *Annu Rev Biophys Biomol Struct* **34**: 173–199.
15. Paulini R, Muller K, Diederich F. (2005) Orthogonal multipolar interactions in structural chemistry and biology. *Angew Chem Int Ed Engl* **44**: 1788–1805.
16. Kauzmann W. (1959) Some factors in the interpretation of protein denaturation. *Adv Protein Chem* **14**: 1–63.
17. Jencks WP, Page MI. (1974) Orbital steering, entropy, and rate accelerations. *Biochem Biophys Res Commun* **57**: 887–892.
18. Page MI, Jencks WP. (1971) Entropic contributions to rate accelerations in enzymic and intramolecular reactions and chelate effect. *Proc Natl Acad Sci USA* **68**: 1678–1683.
19. Tanford C. (1962) Contribution of hydrophobic interactions to stability of globular conformation of proteins. *J Am Chem Soc* **84**: 4240–&.
20. Mirsky AE, Pauling L. (1936) On the structure of native, denatured, and coagulated proteins. *Proc Natl Acad Sci USA* **22**: 439–447.
21. Makhatadze GI, Privalov PL. (1993) Contribution of hydration to protein-folding thermodynamics. 1. The enthalpy of hydration. *J Mol Biol* **232**: 639–659.

36. Kangas E, Tidor B. (2000) Electrostatic specificity in molecular ligand design. *J Chem Phys* **112**: 9120–9131.
37. Sheinerman FB, Norel R, Honig B. (2000) Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol* **10**: 153–159.
38. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE, 3rd. (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc Chem Res* **33**: 889–897.
39. Massova I, Kollman P. (2000) Combined molecular mechanical and continuum solvent approach (mm-pbsa/gbsa) to predict ligand binding. *Persp Drug Discov* **18**: 113–135.
40. Swanson JM, Henchman RH, McCammon JA. (2004) Revisiting free energy calculations: A theoretical connection to mm/pbsa and direct calculation of the association free energy. *Biophys J* **86**: 67–74.
41. Schafer H, Mark AE, Gunsteren WFv. (2000) Absolute entropies from molecular dynamics simulation trajectories. *J Chem Phys* **113**: 7809–7817.
42. Karplus M, Kushick JN. (1981) Method for estimating the configurational entropy of macromolecules. *Macromolecules* **14**: 325–332.
43. Gilson MK, Given JA, Bush BL, McCammon JA. (1997) The statistical-thermodynamic basis for computation of binding affinities: A critical review. *Biophys J* **72**: 1047–1069.
44. Gilson MK, Zhou HX. (2007) Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct* **36**: 21–42.
45. Ruvinsky AM, Kozintsev AV. (2005) New and fast statistical-thermodynamic method for computation of protein-ligand binding entropy substantially improves docking accuracy. *J Comput Chem* **26**: 1089–1095.
46. Wang J, Deng Y, Roux B. (2006) Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys J* **91**: 2798–2814.
47. Woo HJ, Roux B. (2005) Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc Natl Acad Sci USA* **102**: 6825–6830.

48. Homans SW. (2005) Probing the binding entropy of ligand-protein interactions by nmr. *ChemBioChem* **6**: 1585–1591.
49. Lundquist JJ, Toone EJ. (2002) The cluster glycoside effect. *Chem Rev* **102**: 555–578.
50. Mammen M, Choi SK, Whitesides GM. (1998) Polyvalent interactions in biological systems: Implications for design and use of multivalent ligands and inhibitors. *Angew Chem Int Ed* **37**: 2755–2794.
51. Wold S, Exner O. (1973) Statistics of enthalpy-entropy relationship. 4. Temperature-dependent activation parameters. *Chemica Scripta* **3**: 5–11.
52. Krug RR, Hunter WG, Grieger RA. (1976) Statistical interpretation of enthalpy-entropy compensation. *Nature* **261**: 566–567.
53. Liu YF, Sturtevant JM. (1997) Significant discrepancies between van't Hoff and calorimetric enthalpies. 3. *Biophys Chem* **64**: 121–126.
54. Cornish-Bowden A. (2002) Enthalpy-entropy compensation: A phantom phenomenon. *J Biosci* **27**: 121–126.
55. Beasley JR, Doyle DF, Chen LX, Cohen DS, Fine BR, Pielak GJ. (2002) Searching for quantitative entropy-enthalpy compensation among protein variants. *Proteins: Struct Funct Genet* **49**: 398–402.
56. Mukherjee M, Dutta K, White MA, Cowburn D, Fox RO. (2006) Nmr solution structure and backbone dynamics of domain iii of the e protein of tick-borne langat flavivirus suggests a potential site for molecular recognition. *Protein Sci* **15**: 1342–1355.
57. Frederick KK, Marlow MS, Valentine KG, Wand AJ. (2007) Conformational entropy in molecular recognition by proteins. *Nature* **448**: 325–U3.
58. Igumenova TI, Frederick KK, Wand AJ. (2006) Characterization of the fast dynamics of protein amino acid side chains using nmr relaxation in solution. *Chem Rev* **106**: 1672–1699.
59. Horovitz A, Fersht AR. (1990) Strategy for analyzing the cooperativity of intramolecular interactions in peptides and proteins. *J Mol Biol* **214**: 613–617.
60. Hunt JA, Ahmed M, Fierke CA. (1999) Metal binding specificity in carbonic anhydrase is influenced by conserved hydrophobic core residues. *Biochemistry* **38**: 9054–9062.

61. McCall KA, Fierke CA. (2004) Probing determinants of the metal ion selectivity in carbonic anhydrase using mutagenesis. *Biochemistry* **43**: 3979–3986.
62. Nair SK, Calderone TL, Christianson DW, Fierke CA. (1991) Altering the mouth of a hydrophobic pocket — structure and kinetics of human carbonic anhydrase-ii mutants at residue val-121. *J Biol Chem* **266**: 17320–17325.
63. Agarwal PK, Billeter SR, Rajagopalan PTR, Benkovic SJ, Hammes-Schiffer S. (2002) Network of coupled promoting motions in enzyme catalysis. *Proc Natl Acad Sci USA* **99**: 2794–2799.
64. Parr RG, Yang W. (1995) Density-functional theory of the electronic structure of molecules. *Annu Rev Phys Chem* **46**: 701–728.
65. Bingham RJ, Findlay JBC, Hsieh SY, Kalverda AP, Kjeliberg A, Perazzolo C, Phillips SEV, Seshadri K, Trinh CH, Turnbull WB, Bodenhausen G, Homans SW. (2004) Thermodynamics of binding of 2-methoxy-3-isopropylpyrazine and 2-methoxy-3-isobutylpyrazine to the major urinary protein. *J Am Chem Soc* **126**: 1675–1681.
66. Freire E. (1994) Statistical thermodynamic analysis of differential scanning calorimetry data — structural deconvolution of heat-capacity function of proteins, Numerical computer methods, pt b, Vol. 240, *Methods in Enzymology*, pp. 502–530.
67. Chervenak MC, Toone EJ. (1994) A direct measure of the contribution of solvent reorganization to the enthalpy of ligand-binding. *J Am Chem Soc* **116**: 10533–10539.
68. Christensen T, Gooden DM, Kung JE, Toone EJ. (2003) Additivity and the physical basis of multivalency effects: A thermodynamic investigation of the calcium edta interaction. *J Am Chem Soc* **125**: 7357–7366.
69. Malham R, Johnstone S, Bingham RJ, Barratt E, Phillips SEV, Laughton CA, Homans SW. (2005) Strong solute-solute dispersive interactions in a protein-ligand complex. *J Am Chem Soc* **127**: 17061–17067.
70. Lovitt S, Freire E. (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr Opin Struct Biol* **11**: 560–566.



71. Ladbury JE, Chowdhry BZ. (1996) Sensing the heat: The application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chem Biol* **3**: 791–801.
72. Connelly PR, Thomson JA, Fitzgibbon MJ, Bruzzese FJ. (1993) Probing hydration contributions to the thermodynamics of ligand-binding by proteins — enthalpy and heat-capacity changes of tacrolimus and rapamycin binding to fk506 binding-protein in d<sub>2</sub>O and h<sub>2</sub>O. *Biochemistry* **32**: 5583–5590.
73. Oas TG, Toone EJ. (1997) Thermodynamic solvent isotope effects and molecular hydrophobicity. *Adv Biophys Chem* **6**: 1–52.
74. Wiseman T, Williston S, Brandts JF, Lin LN. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* **179**: 131–137.
75. Krishnamurthy VM, Kaufman GK, Urbach AR, Gitlin I, Gudiksen KL, Weibel DB, Whitesides GM. (In press) Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding. *Chem Rev*.
76. Boriack PA, Christianson DW, Kingery-Wood J, Whitesides GM. (1995) Secondary interactions significantly removed from the sulfonamide binding pocket of carbonic anhydrase ii influence inhibitor binding constants. *J Med Chem* **38**: 2286–2291.
77. Gao JM, Qiao S, Whitesides GM. (1995) Increasing binding constants of ligands to carbonic-anhydrase by using greasy tails. *J Med Chem* **38**: 2292–2301.
78. Jain A, Huang SG, Whitesides GM. (1994) Lack of effect of the length of oligoglycine-derived and oligo(ethylene-glycol)-derived para-substituents on the affinity of benzenesulfonamides for carbonic-anhydrase-ii in solution. *J Am Chem Soc* **116**: 5057–5062.
79. Jain A, Whitesides GM, Alexander RS, Christianson DW. (1994) Identification of 2 hydrophobic patches in the active-site cavity of human carbonic-anhydrase-ii by solution-phase and solid-state studies and their use in the development of tight-binding inhibitor. *J Med Chem* **37**: 2100–2105.
80. Carbeck JD, Colton IJ, Gao J, Whitesides GM. (1998) Protein charge ladders, capillary electrophoresis, and the role of electrostatics in biomolecular recognition. *Acc Chem Res* **31**: 343–350.

81. Grzybowski BA, Ishchenko AV, Kim C-Y, Topalov G, Chapman R, Christianson DW, Whitesides GM, Shakhnovich EI. (2002) Combinatorial computational method gives new picomolar ligands for a known enzyme. *Proc Natl Acad Sci USA* **99**: 1270–1273.
82. Krishnamurthy VM, Bohall BR, Kim C-Y, Moustakas DT, Christianson DW, Whitesides GM. (2007) Thermodynamic parameters for the association of fluorinated benzenesulfonamides with bovine carbonic anhydrase ii. *Chem-Asia J* **2**: 94–105.
83. Krishnamurthy VM, Bohall BR, Kim CY, Moustakas DT, Christianson DW, Whitesides GM. (2007) Thermodynamic parameters for the association of fluorinated benzenesulfonamides with bovine carbonic anhydrase ii. *Chem-Asia J* **2**: 94–105.
84. Krishnamurthy VM, Semetey V, Bracher PJ, Shen N, Whitesides GM. (2007) Dependence of effective molarity on linker length for an intramolecular protein-ligand system. *J Am Chem Soc* **129**: 1312–1320.
85. Krishnamurthy VM, Bohall BR, Semetey V, Whitesides GM. (2006) The paradoxical thermodynamic basis for the interaction of ethylene glycol, glycine, and sarcosine chains with bovine carbonic anhydrase ii: An unexpected manifestation of enthalpy/entropy compensation. *J Am Chem Soc* **128**: 5802–5812.
86. Dunitz JD. (1995) Win some, lose some — enthalpy-entropy compensation in weak intermolecular interactions. *Chem Biol* **2**: 709–712.
87. Ishikawa H, Finkelstein IJ, Kim S, Kwak K, Chung JK, Wakasugi K, Massari AM, Fayer MD. (2007) Neuroglobin dynamics observed with ultrafast 2d-ir vibrational echo spectroscopy. *Proc Natl Acad Sci USA* **104**: 16116–16121.
88. Ishikawa H, Kim S, Kwak K, Wakasugi K, Fayer MD. (2007) Disulfide bond influence on protein structural dynamics probed with 2d-ir vibrational echo spectroscopy. *Proc Natl Acad Sci USA* **104**: 19309–19314.
89. Park S, Kwak K, Fayer MD. (2007) Ultrafast 2d-ir vibrational echo spectroscopy: A probe of molecular dynamics. *Laser Phys Lett* **4**: 704–718.