Entropy–enthalpy transduction caused by conformational shifts can obscure the forces driving protein–ligand binding

Andrew T. Fenley, Hari S. Muddana, and Michael K. Gilson[1]

Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA 92093-0736

Edited by Barry Honig, Columbia University/The Howard Hughes Medical Institute, New York, NY, and approved October 12, 2012 (received for review July 30, 2012)

Molecular dynamics simulations of unprecedented duration now can provide new insights into biomolecular mechanisms. Analysis of a 1-ms molecular dynamics simulation of the small protein bovine pancreatic trypsin inhibitor reveals that its main conformations have different thermodynamic profiles and that perturbation of a single geometric variable, such as a torsion angle or interresidue distance, can select for occupancy of one or another conformational state. These results establish the basis for a mechanism that we term entropy–enthalpy transduction (EET), in which the thermodynamic character of a local perturbation, such as enthalpic binding of a small molecule, is camouflaged by the thermodynamics of a global conformational change induced by the perturbation, such as a switch into a high-entropy conformational state. It is noted that EET could occur in many systems, making measured entropies and enthalpies of folding and binding unreliable indicators of actual thermodynamic driving forces. The same mechanism might also account for the high experimental variance of measured enthalpies and entropies relative to free energies in some calorimetric studies. Finally, EET may be the physical mechanism underlying many cases of entropy–enthalpy compensation.

calorimetry | heat capacity | control | correlation | causation

The entropic and enthalpic components of free energy can be informative regarding the mechanisms of biophysical processes, like protein folding and protein–ligand binding. Moreover, information on changes in entropy and enthalpy can usefully guide the design of improved drug molecules (1), with advantageous specificity (2) and physical properties (3). However, calorimetric studies of biomolecular binding and folding often reveal unexpected changes in entropy and enthalpy that are difficult to interpret in terms of physical driving forces (4–8). Some of these puzzling results are instances of entropy–enthalpy compensation (9), a common but not universal (10, 11) phenomenon in which perturbations of a system that increase the enthalpy also increase the entropy or vice versa; therefore, the net change in the free energy remains small. Experimental error in measured enthalpies, particularly when obtained by the relatively imprecise van 't Hoff method, can generate spurious entropy–enthalpy compensation (12); the nonphysical operation of Berkson's paradox (selection bias) (13) can also generate this compensation. However, analysis of collected calorimetric data for protein–ligand binding indicates that entropy–enthalpy compensation is a genuine and common physical phenomenon (13).

Today, novel computational technologies, such as graphical processor units (14–17) and the specialized Anton computer (18), are enabling dramatically longer molecular dynamics (MD) simulations than hitherto feasible. The enormous conformational sampling power of these technologies has the potential to open a new window on the molecular mechanisms underlying the thermodynamics of biomolecular systems. For example, it can provide increasingly accurate estimates of thermodynamic quantities that are notoriously difficult to converge, such as configurational entropy (19–21). Such thermodynamic results, coupled with atomistic structural and dynamical data, hold the potential to address some of the conundrums in calorimetric data outlined above. Recently, a 1-ms MD simulation of a relatively simple protein, bovine pancreatic trypsin inhibitor (BPTI), was released to the scientific community (22). This simulation provides a level of sampling unparalleled by any other biomolecular simulation to date, making it an ideal dataset for gaining a deeper understanding of the thermodynamics of proteins.

Here, we present a statistical thermodynamic analysis of the main conformational clusters observed in the 1-ms BPTI simulation using the maximum information spanning tree (MIST) (23, 24) approach to estimate configurational entropy. A central finding is that even this simple protein is delicately poised between states of similar free energy but very different entropy and enthalpy. We also introduce a method, based on the Jensen–Shannon divergence (25, 26), to show that there exists structural degrees of freedom that can easily be perturbed to drive the system into any of the major conformational states. These observations combine to yield a concept of entropy–enthalpy transduction (EET), in which local thermodynamic driving forces are transduced into different global thermodynamics through conformational selection. Finally, we provide a statistical thermodynamic description of EET that does not rely on the idealization of conformational selection. The methods and results presented here bear on the interpretation of simulation and calorimetry data and the physical basis for entropy–enthalpy compensation.

Results

We carried out a thermodynamic analysis of the trajectory of the longest MD simulation of a protein published to date (22), a 1-ms simulation of the small trypsin inhibitor BPTI. The simulation results are available to the scientific community and agree reasonably well with NMR data (27). The previously defined conformational clusters (22) 0, 1, and 2 (C_0, C_1, and C_2) account for 92.3% of the simulation time and include the required number of simulation snapshots for thermodynamic analysis. Cluster C_1 is the most rigid and similar to crystal structures of BPTI bound to trypsin (22), whereas C_2 is the most mobile (Fig. 1, *Upper*).

Although the occupancies (probabilities) of the three conformational clusters vary about fivefold, they are of similar stability, with free energies ranging over only ~0.95 kcal/mol (1 kcal = 4.18 kJ) because of, essentially, the logarithmic relationship between probability and free energy (*Methods*). However, considerably larger entropy and enthalpy differences underlie these small free

Author contributions: A.T.F. and M.K.G. designed research; A.T.F., H.S.M., and M.K.G. performed research; A.T.F., H.S.M., and M.K.G. analyzed data; and A.T.F. and M.K.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: mgilson@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1213180109/-/DCSupplemental.

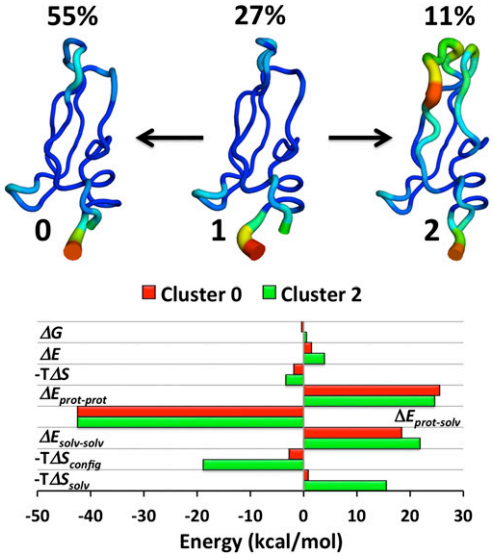

Fig. 1. (*Upper*) Representative backbone traces of the main conformational clusters C_0, C_1, and C_2 as labeled, and their percent occupancies during the simulation; color and thickness of the traces indicate rms fluctuation of the backbone (narrow blue, 0.4 Å; thick red, 3.7 Å) after structural superposition. (*Lower*) Bar graph of the computed thermodynamics (kilocalories per mole) of clusters C_0 (red) and C_2 (green) relative to cluster C_1. ΔE, total potential energy; $\Delta E_{prot\text{-}prot}$, potential energy of protein; $\Delta E_{prot\text{-}solv}$, interaction energy of protein and solvent; $\Delta E_{solv\text{-}solv}$, potential energy of solvent; ΔG, free energy; $-T\Delta S$, total entropic free energy contribution; $-T\Delta S_{config}$, configurational entropy estimated by MIST at second order and corrected by subtraction of matched results with block-shuffled data; $-T\Delta S_{solv}$, solvent entropy computed as $-T(\Delta S - \Delta S_{config})$.

energy differences. For example, although clusters C_1 and C_2 are equistable to within 0.5 kcal/mol, their entropy ($-TS$) and enthalpy differences are, respectively, −3.2 and 3.7 kcal/mol; their configurational entropy differences are estimated at 19 kcal/mol, and the more mobile C_2 has weaker intraprotein (25 kcal/mol) and intrasolvent (22 kcal/mol) interactions, partly balanced by stronger protein-solvent interactions (−42 kcal/mol). The magnitudes of these thermodynamic components reveal an underlying potential for much larger swings in energy and entropy: even slightly imbalanced perturbation of these terms, such as by mutations or changes in solvent composition, could generate larger net entropy and enthalpy differences between the clusters.

Because the three conformational clusters are so closely balanced in free energy, small perturbations of the system might easily shift their probabilities. Such shifts would then be expected to yield significant changes in the entropy–enthalpy balance at minimal cost in free energy. In particular, we hypothesized the existence of control variables, here defined as geometric variables x_i, whose perturbations could select for one of three main conformational clusters and therefore, influence the global entropy and enthalpy. In searching for such variables, one may use the fact that a canonical MD simulation contains information on how the equilibrium distribution of conformations will respond to local perturbations, because perturbing the marginal probability distribution function (pdf) of a subset of variables x_c from $p(x_c)$ to $p'(x_c)$ will cause the equilibrium joint pdf over all variables x to change from $p(x)$ to $p'(x) = p(x|x_c)p'(x_c)$. The conditional pdf thus represents a response function of the system to shifts in the pdf of the control variables (*SI Text* shows the derivation). We focus here on perturbations in which the accessible range of a single variable is reduced, because these perturbations are simple to analyze; also, good statistics are available for the associated conditional probabilities. Thus, for each cluster C_a, where $a = 0$, 1, or 2, we sought a geometric variable x_i, either an interatomic distance or a torsion angle, for which a range $\mathbf{R}$ could be identified such that, to good approximation, $p(C_a|x_i \in \mathbf{R}) = 1$ and $p(x_i \in \mathbf{R}|C_a) = 1$. These conditions are best met when there is little overlap of the pdfs of x_i conditioned on the three clusters, $p(x_i|C_a)$. Because the Jensen–Shannon divergence (25, 26) is zero for identical pdfs and reaches its maximum ($\sqrt{\ln 2}$) for nonoverlapping pdfs, we used it to define a metric M_{ia},

$$0 \le M_{ia} \equiv \sum_{b \neq a}^{N_{clusters}} JSD\left[p(x_i|C_a)||p(x_i|C_b)\right] \le (N_{cluster} - 1)(\ln 2)^{1/2}, \quad [1]$$

which scores the ability of variable x_i to select cluster C_a from all other clusters C_b.

For each cluster, we found multiple carbon α-carbon α distances (x_i) whose M_{ia} scores are near the theoretical maximum of 1.67 (Fig. 2, row 1) and hence, had low overlap of the pdfs $p(x_i|C_0)$, $p(x_i|C_1)$, and $p(x_i|C_2)$ (Fig. 2, row 3). For example, restricting the Thr11–Tyr35 distance to be greater than 6 Å (Fig. 2, row 2) drives the protein almost entirely into cluster C_2. Interestingly, experiments have shown that the Tyr35Gly mutation makes the loop region of BPTI more flexible (28). Note that the significance of certain carbon α-carbon α distances does not derive from direct energetic interactions between the two corresponding residues, but rather from a complex set of couplings of each distance with the structure as a whole. We also used the same analysis to seek other classes of geometric variable that could yield similarly large values of the M_{ia} metric and indeed, found a number of torsional pdfs with similarly low overlaps, as shown in Fig. 2, row 4. This result amplifies the observation that many control variables are available to shift the conformational state of this protein.

We validated the ability of the control variables to isolate thermodynamically distinct clusters as follows. For each cluster, we chose the control variable with the greatest value of M_{ia} (1.55, 1.52, and 1.59 for clusters C_0, C_1, and C_2, respectively). We then reweighted the entire trajectory by imposing on the associated control variable an infinitely high-walled square-well potential positioned to select the bulk of the conformations associated with the desired cluster, while avoiding values of the control variable shared by any other cluster. The thermodynamic quantities of the reweighted trajectory were compared with those quantities associated with the original cluster. The results for the reweighted trajectories agree with the original cluster thermodynamics to within 0.3 kcal/mol for all ΔG_{ab}, 0.4 kcal/mol for all ΔE_{ab}, and 0.6 kcal/mol for all $-T\Delta S_{ab}$ (Table S1). Thus, the control variables select the intended conformational substates with their respective thermodynamic signatures.

Discussion

Thermodynamic and Structural Analysis of the BPTI Simulation. The present analysis of a long MD simulation indicates that BPTI can exist in multiple conformational states with distinct thermodynamic signatures, and that a variety of control variables are available to drive it into selected conformational states. Given that BPTI is a small protein rigidified by three disulfide bonds and not normally regarded as allosteric, it seems likely that larger, more flexible proteins have even greater scope to shift among global states of different entropy and enthalpy. If so, then various perturbations of biomolecules, such as ligand binding and mutation, should often be accompanied by changes in entropy and entropy that derive from global population shifts rather than local interactions. Even relatively subtle changes, such as chemical changes across a congeneric series of ligands, could produce large

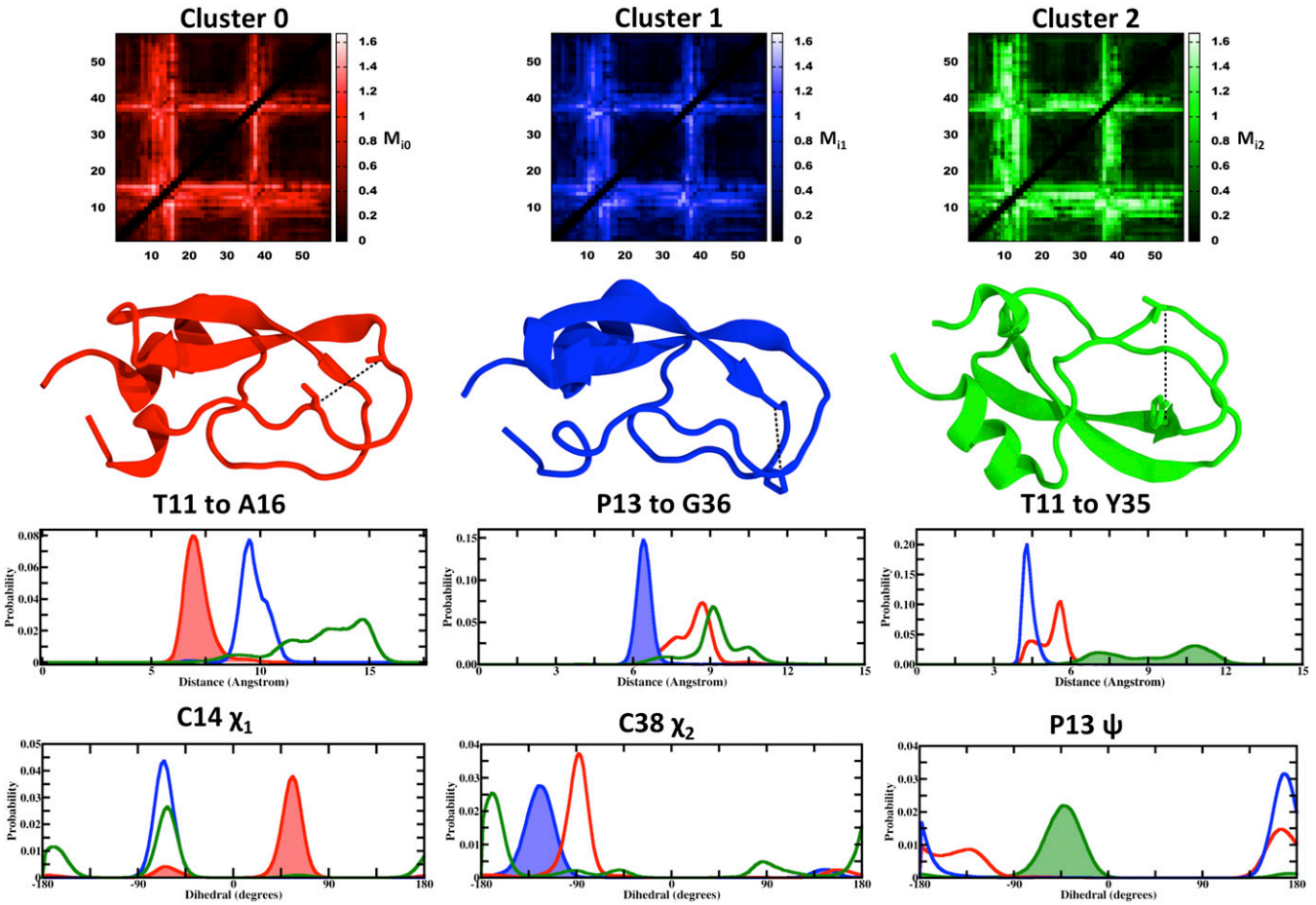

Fig. 2. Identification and analysis of control variables. (Row 1) Heat maps of the M_{ia} values for each cluster a, where the possible control variables x_i are Cα–Cα distances among all 58 residues. The brightest pixels correspond to distances that can best isolate the respective cluster. (Row 2) Representative structure of the respective clusters, with the Cα–Cα distance, x_i, corresponding to the largest M_{ia} value shown as a dashed line. (Row 3) 1D pdfs of the Cα–Cα distances labeled in the row 2 for all three clusters; the pdf corresponding to the largest M_{ia} value for a particular cluster is shaded. (Row 4) Analogous pdfs corresponding to the largest M_{ia} value for each cluster when torsions are used as the x_i instead of Cα–Cα distances.

changes in entropy and enthalpy of binding if the global states are of similar free energy. Furthermore, because the local forces driving, for example, ligand binding need not be the same as the global thermodynamic consequences of binding, one may view the biomolecule as transducing the local driving force into a quite different global thermodynamic signature. This phenomenon, essentially a manifestation of thermodynamic linkage (29), may be termed entropy–enthalpy transduction (EET).

Illustration and Implications of EET. Some of the implications of EET may be illustrated quantitatively by considering a protein P, which can occupy either a low-entropy state, A, or a high-entropy state, B, where the intrinsic free energy difference between the two states is x kcal/mol (Fig. 3, *Upper*). The ligand L binds exclusively to state B through mainly enthalpic forces. (Note that, as in the BPTI analysis above, the changes in entropy and enthalpy in this model are considered to include contributions from the protein, ligand, and solvent.) When x exceeds thermal energy (RT = 0.6 kcal/mol, and R is the gas constant), the free protein exists primarily in state B, and the overall binding thermodynamics are simply the thermodynamics of the ligand's binding to B and hence are seen to be enthalpy-driven (Fig. 3, *Lower*). When $x < -RT$, the free protein exists primarily in state A, but it can switch to the high-entropy state B when L binds; therefore, the binding process now seems entropy-driven. In this way, the protein's ability to switch states on binding enables transduction of the ligand's binding enthalpy into entropy. Conversely, the Tyr35Gly mutant of BPTI, which makes its loops more flexible without visibly perturbing the trypsin-bound conformation (30), causes trypsin binding to become more favorable enthalpically but less favorable entropically (30), which was expected for a mutation that selectively stabilizes a conformational state similar in character to cluster C_2. The EET mechanism can also reinforce or amplify the driving forces of binding and other reactions. In the present example, this would occur if state B, rather than state A, were more stable enthalpically and less stable entropically. The enthalpic driving force for binding would then seem much stronger than the actual enthalpic forces driving ligand binding to the protein. In general, EET can generate shifts in enthalpy and entropy that are opposite to or far larger than expected based on local considerations. This finding implies that calorimetric data alone cannot be relied on to report the actual thermodynamic forces driving binding.

Especially when unexpected shifts are observed, a structure-based search for substantial and potentially functionally relevant conformational changes may be fruitful, as illustrated by prior studies of drug binding to various sequences of DNA (31) and the binding of agonists and antagonists to the A_{2B} adenosine receptor (32). Changes in NMR-order parameters (33) can be informative in this regard. In some cases, they indicate increases in protein flexibility—and hence, presumably in configurational entropy—caused by protein–ligand binding (34–36). However,

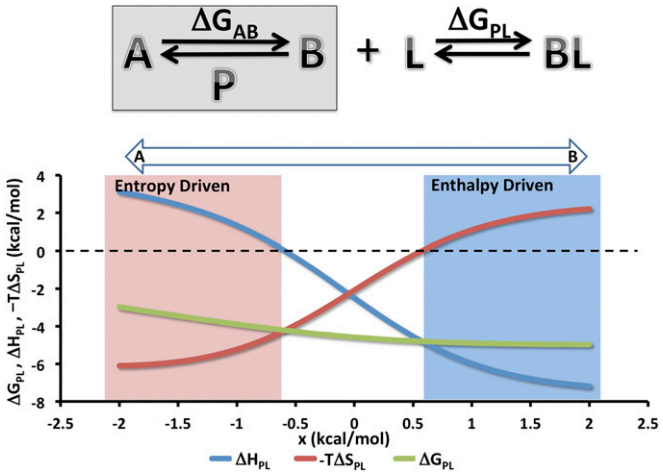

Fig. 3. Two-state model system illustrating EET. (*Upper*) The protein P has two states A and B with a free energy difference of ΔG_{AB}. By construction, state B is higher in entropy and lower in enthalpy than A. Ligand L binds only state B and hence, drives P from a two- to one-state system. The overall binding free energy is ΔG_{PL}. (*Lower*) Overall thermodynamics of ligand–protein binding as a function of a small free energy bias, *x*, added to ΔG_{AB} to shift the preferred unbound state of P from state A ($x < 0$) to state B ($x > 0$). Here, $\Delta H_{BL} = -7.5$, $-T\Delta S_{BL} = 2.5$, $\Delta H_{AB} = 10.0$, $-T\Delta S_{AB} = -10.0$, and $RT = 0.6$, where ΔH_{BL} and $-T\Delta S_{BL}$ correspond to the enthalpy and entropy of the ligand L binding to state B. All quantities are in kilocalories per mole.

an increase in configurational entropy identified in this way may be balanced by a decrease in solvent entropy, as in the present calculations and further considered below, leading to either a positive or negative change in the net entropy. It is also worth noting that, although likely cases of EET amounting to 10 kcal/mol or more have been uncovered experimentally (31, 32, 35, 37, 38), identifying them required a combination of calorimetric and noncalorimetric experiments. Thus, there are presumably many systems where EET affects calorimetric results but has not been brought to light. Similarly, although measured changes in heat capacity are often used to elucidate the polar or nonpolar character of a protein–protein binding interface (39), if binding induces a significant conformational change in either protein, then the measured heat capacity may not accurately indicate the nature of the binding interface. The potential significance of this scenario is supported by the present calculations, because the computed heat capacity of conformational cluster C_2 is ∼0.09 and 0.13 kcal/mol per Kelvin greater than the computed heat capacities of clusters C_0 and C_1, respectively (Table S2).

A biomolecular system may access a high-entropy state not only through increased protein motion, but also through a conformational shift that leads to release of water or other solvent molecules far from the site of the inducing perturbation. The potential role of solvent in EET is illustrated by thrombin, whose affinity for dissolved Na^+ increases when substrate binds (40), leading to possible cobinding of substrate and Na^+. Because binding of Na^+ to the enzyme–substrate complex at 300 K is enthalpically favorable (−18 kcal/mol) and entropically unfavorable (15.6 kcal/mol) (40), cobinding of substrate with Na^+ could cause a substrate with intrinsic binding that is entropy-driven to seem enthalpy-driven. More generally, subtle changes in solvent, or indeed, in other experimental conditions, could lead to marked changes in measured entropy and enthalpy, with minimal change in free energy. Thus, in the scenario of Fig. 3, small changes in experimental conditions may produce small changes in the balance between states A and B (i.e., changes in *x*) and hence, large variations in the apparent thermodynamics of binding. Such sensitivity to experimental conditions could help explain cases where the variance of calorimetric enthalpies exceeds the variance of free energy (41), although other factors can also contribute (42).

Finally, the EET mechanism could underlie many cases of entropy–enthalpy compensation, a phenomenon in which related processes, such as a drug-like molecule binding various similar biomolecules, are associated with similar free energy changes but seemingly disproportionately large entropy and enthalpy changes that correlate near-linearly with a slope or compensation temperature, $T_c \equiv \frac{d\Delta S}{d\Delta H} \approx T$. Entropy–enthalpy compensation is often explained on physical grounds by arguing that deeper energy wells are narrower and hence lower in entropy than shallow ones (43, 44). Although this mechanism may indeed play a role, it is not known whether the depths and widths of energy wells in fact anticorrelate or show no correlation at all (45). Moreover, this mechanism would not explain why compensation is usually near-linear with slope $T_c \approx T$. The present analysis of BPTI suggests that a series of related proteins with similar binding sites could readily access different conformations of similar free energy but different enthalpy and entropy. As a consequence, the same ligand could bind the series of proteins with similar free energies but quite different entropies and enthalpies. In this way, the EET mechanism can lead to highly linear entropy–enthalpy compensation. This observation is broadly consistent with prior suggestions, derived from plausible mathematical models rather than detailed simulation, regarding the importance of delicately poised equilibria (46), large fluctuations, and a high density of states (47–51). Although the EET mechanism can, in this manner, generate a highly linear correlation between entropy and enthalpy, it can also generate other thermodynamic patterns. For example, if the various drug-like molecules that bind a protein drive it into global conformational states with free energies that vary substantially, one may still observe a degree of entropy–enthalpy compensation, but it will not be as clear cut as in a system where the protein instead moves among states of very similar free energy.

General Formalism for EET. To unify and generalize the ideas above, we describe here how the concepts of control variables and response functions allow partitioning of the thermodynamics of protein–ligand binding into an intrinsic part, due to the ligand's direct interactions with the protein, and a transduced part, due to the long-ranged conformational response to the ligand's manipulation of control variables in the binding site: $\Delta E = \Delta E_{int} + \Delta E_t$, $\Delta S^o = \Delta S^o_{int} + \Delta S_t$, and $\Delta G^o = \Delta G^o_{int} + \Delta G_t$. (A detailed derivation is provided in *SI Text*.) Given the full set of spatial coordinates, *x*, the coordinates of the atoms local to the binding site are defined as the control variables, x_c, which will be manipulated by the ligand when it binds. The coordinates of the remaining atoms are termed the transducing variables, x_t, which will respond to the manipulation of the control variables. For the free protein, the potential energy function of the whole system $E_p(x_c,x_t)$ is expressed as the sum of a binding site part that depends only on the control variables and a transducing part that depends on the transducing variables as well. For the energy of the ligand–protein complex, $E_q(x_c,x_t)$, the binding site part also depends on the conformation, position, and orientation of the ligand, (x_l,r,ω), whereas the ligand's long-ranged energetic interactions with atoms beyond the binding site may be approximated as a small constant value, E_{lt}. Thus, binding of the ligand is approximated as directly affecting only the control variables, while the energy landscape of the transducing variables remains unchanged for any conformation of the control variables. This approximation should be good in a number of cases, such as when a small ligand binds a large protein. Hence,

$$E_p(x_c,x_t) = E_c(x_c) + E_t(x_c,x_t) \quad [2]$$

and

$$E_q(x_l, r, \omega, x_c, x_t) = \left[E_{lc}(x_l, r, \omega, x_c) + E_{lt}\right] + E_t(x_c, x_t). \quad [3]$$

The energetic and entropic contributions of the transducing part of the system may now be expressed as functions of the control variables, $\bar{E}_t(x_c)$ and $\bar{S}_t(x_c)$, respectively, and these functions can, furthermore, be combined to yield the contribution of the transducing variables to the potential of mean force over the control variable $W_t(x_c)$:

$$\bar{E}_t(x_c) = \int p(x_t|x_c) E_t(x_c, x_t) dx_t, \quad [4]$$

$$\bar{S}_t(x_c) = -R \int p(x_t|x_c) \ln\, p(x_t|x_c) dx_t, \quad [5]$$

where R is the gas constant, and

$$W_t(x_c) = \bar{E}_t(x_c) - T\bar{S}_t(x_c). \quad [6]$$

These thermodynamic response functions, which apply to both the free protein and the complex, are mediated by $p(x_t|x_c)$, the pdf of the transduced variables conditioned on the control variables. The net contribution of the transducing variables to the binding thermodynamics may now be written as

$$\begin{aligned} \Delta E_t &= \int \left[p_{q,c}(x_c) - p_{p,c}(x_c)\right] \bar{E}_t(x_c) dx_c \\ \Delta S_t &= \int \left[p_{q,c}(x_c) - p_{p,c}(x_c)\right] \bar{S}_t(x_c) dx_c \\ \Delta G_t &= \int \left[p_{q,c}(x_c) - p_{p,c}(x_c)\right] W_t(x_c) dx_c \end{aligned} \quad [7]$$

where $p_{q,c}(x_c)$ and $p_{p,c}(x_c)$ are the marginal pdfs over the control variables for the complex and free protein, respectively, at equilibrium.

The intrinsic binding terms, which make up the remainder of the binding thermodynamics, are as follows:

$$\Delta E_{int} = \langle E_{lc} \rangle_q - \langle E_c \rangle_p - \langle E_l \rangle_l, \quad [8]$$

$$\Delta S^o_{int} = -R \ln \frac{8\pi^2}{C^o} + S_{q,lc} - S_{p,c} - S_l, \quad [9]$$

and

$$\Delta G^o_{int} = \Delta E_{int} - T\Delta S^o_{int}. \quad [10]$$

Here, the angle brackets signify Boltzmann averages for the complex, free protein, and free ligand, respectively; E_l is the potential function of the free ligand, and S_x signifies the Gibbs–Shannon entropies of the subscripted variables for the complex, free protein, and free ligand, respectively. Note that, although the intrinsic and transduced quantities are formally separated, the control and transducing variables are linked in the sense that the potential of mean force of the control variables, which influences both the intrinsic and transduced quantities, is determined jointly by both parts of the system. Thus, the control variables may be viewed as a boundary between two parts of the system, the ligand and transducing part of the protein. The present analysis, thus, is reminiscent of the fluctuating boundary ensemble (48).

The transduction mechanism, then, operates as follows. When a ligand binds, it changes the pdf of the control variables from $p_{p,c}(x_c)$ to $p_{q,c}(x_c)$, driving the transducing variables to adopt a new pdf dictated by the response function $p(x_t|x_c)$ and thus, altering the thermodynamic contributions of the transducing part of the protein according to the above expressions. Because there is no requirement that this conformational distribution match any state of the free protein, conformational selection is not an issue. Also note that ΔG^o_t can be negative in sign. For example, stabilizing interactions internal to the binding site of the free protein might hold it in a closed conformation that keeps the transducing part in an unstable state; then, binding of the ligand could open up the binding site and allow the transducing part to relax.

Entropy–enthalpy compensation occurs when $W_t(x_c)$, the transducing variables' contribution to the potential of mean force over the control variables, varies little (e.g., $\sim RT$ or less) over some conformational region **R**. In this case, the transducing variables permit the conformation of the binding site to rearrange freely within the bounds of **R** in response to a variety of ligands. However, saying that $W_t(x_c)$ varies little with x_c places no constraint on either $\bar{E}_t(x_c)$ or $\bar{S}_t(x_c)$ individually; therefore, these components can still vary substantially over **R**. As a consequence, there can be many possible changes in $p(x_c)$ that are associated with small changes in ΔG_t but large changes in ΔE_t and ΔS_t. Such changes will lead directly to entropy–enthalpy compensation across the series of ligands, with compensation temperatures, T_c, near unity, as commonly observed experimentally. Deviations from such perfect (or strong) (50) entropy–enthalpy compensation will occur, however, if some ligands drive the binding site conformation (i.e., the control variables) into regions of conformational space associated with large changes in $W_t(x_c)$ and hence, generate different values of ΔG_t or if the intrinsic energy and entropy changes vary substantially across the set of ligands.

Methods

The numerical thermodynamic analysis presented above and detailed in *SI Text* is based on the recent 1-ms BPTI simulation (22). Because the simulation yields a canonical distribution, the relative free energy ΔG_{ab} of two clusters a and b can be estimated as $-RT\, ln(n_a/n_b)$, where n_i is the number of simulation snapshots in cluster i. By the same token, the Boltzmann-averaged potential energy of cluster a, E_a, can be computed as a simple mean over the snapshots in each cluster. The entropy difference between clusters a and b then is given by $-T\Delta S_{ab} = \Delta G_{ab} - \Delta E_{ab}$. Furthermore, we decomposed the total energy into protein–protein, water–water, and protein–water contributions and estimated the protein configurational (19, 20) entropy of each cluster by means of the MIST approach (23, 24). Issues with noise and convergence were minimized by subtracting out the results for block-permuted data (*SI Text*). Solvation entropy differences were then estimated as differences between total entropy differences and configurational entropy differences (52). Finally, the heat capacity of each conformational cluster was computed by application of existing formulae (53). Fig. 1 and Table S2 summarize the results, and Figs. S1, S2, and S3 provide convergence plots of the various thermodynamic quantities.

ACKNOWLEDGMENTS. We thank Drs. C. McClendon and L. Pierce for comments, and D. E. Shaw Research for providing the BPTI trajectory and their assignment of snapshots to conformational clusters. This publication was supported by National Institute of General Medical Sciences of the National Institutes of Health Grant GM61300.

1. Freire E (2008) Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov Today* 13(19–20):869–874.
2. Ohtaka H, et al. (2004) Thermodynamic rules for the design of high affinity HIV-1 protease inhibitors with adaptability to mutations and high selectivity towards unwanted targets. *Int J Biochem Cell Biol* 36(9):1787–1799.
3. Ferenczy GG, Keserü GM (2010) Thermodynamics guided lead discovery and optimization. *Drug Discov Today* 15(21–22):919–932.
4. Talhout R, Villa A, Mark AE, Engberts JBFN (2003) Understanding binding affinity: A combined isothermal titration calorimetry/molecular dynamics study of the binding of a series of hydrophobically modified benzamidinium chloride inhibitors to trypsin. *J Am Chem Soc* 125(35):10570–10579.
5. 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(17): 5802–5812.
6. Bissantz C, Kuhn B, Stahl M (2010) A medicinal chemist's guide to molecular interactions. *J Med Chem* 53(14):5061–5084.
7. DeLorbe JE, et al. (2009) Thermodynamic and structural effects of conformational constraints in protein-ligand interactions. Entropic paradoxy associated with ligand preorganization. *J Am Chem Soc* 131(46):16758–16770.
8. Reynolds CH, Holloway MK (2011) Thermodynamics of ligand binding and efficiency. *ACS Med Chem Lett* 2(6):433–437.
9. Lumry R, Rajender S (1970) Enthalpy-entropy compensation phenomena in water solutions of proteins and small molecules: A ubiquitous property of water. *Biopolymers* 9(10):1125–1227.
10. Gallicchio E, Kubo MM, Levy RM (1998) Entropy-enthalpy compensation in solvation and ligand binding revisited. *J Am Chem Soc* 120(18):4526–4527.
11. Rekharsky MV, et al. (2007) A synthetic host-guest system achieves avidin-biotin affinity by overcoming enthalpy-entropy compensation. *Proc Natl Acad Sci USA* 104(52): 20737–20742.
12. Krug RR, Hunter WG, Grieger RA (1976) Statistical interpretation of enthalpy-entropy compensation. *Nature* 261(5561):566–567.
13. Olsson TSG, Ladbury JE, Pitt WR, Williams MA (2011) Extent of enthalpy-entropy compensation in protein-ligand interactions. *Protein Sci* 20(9):1607–1618.
14. Friedrichs MS, et al. (2009) Accelerating molecular dynamic simulation on graphics processing units. *J Comput Chem* 30(6):864–872.
15. Götz AW, et al. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J Chem Theory Comput* 8(5):1542–1555.
16. Pierce LCT, Salomon-Ferrer R, Augusto F de Oliveira C, McCammon JA, Walker RC (2012) Routine access to millisecond time scale events with accelerated molecular dynamics. *J Chem Theory Comput* 8(9):2997–3002.
17. Stone JE, Hardy DJ, Ufimtsev IS, Schulten K (2010) GPU-accelerated molecular modeling coming of age. *J Mol Graph Model* 29(2):116–125.
18. Shaw DE, et al. (2007) Anton, a special-purpose machine for molecular dynamics simulation. *SIGARCH Comput Archit News* 35(2):1–12.
19. Killian BJ, Yundenfreund Kravitz J, Gilson MK (2007) Extraction of configurational entropy from molecular simulations via an expansion approximation. *J Chem Phys* 127 (2):024107.
20. Hnizdo V, Tan J, Killian BJ, Gilson MK (2008) Efficient calculation of configurational entropy from molecular simulations by combining the mutual-information expansion and nearest-neighbor methods. *J Comput Chem* 29(10):1605–1614.
21. Numata J, Knapp E-W (2012) Balanced and bias-corrected computation of conformational entropy differences for molecular trajectories. *J Chem Theory Comput* 8(4): 1235–1245.
22. Shaw DE, et al. (2010) Atomic-level characterization of the structural dynamics of proteins. *Science* 330(6002):341–346.
23. King BM, Tidor B (2009) MIST: Maximum Information Spanning Trees for dimension reduction of biological data sets. *Bioinformatics* 25(9):1165–1172.
24. King BM, Silver NW, Tidor B (2012) Efficient calculation of molecular configurational entropies using an information theoretic approximation. *J Phys Chem B* 116(9): 2891–2904.
25. Lin J (1991) Divergence measures based on the Shannon entropy. *IEEE Trans Inf Theory* 37(1):145–151.
26. Endres DM, Schindelin JE (2003) A new metric for probability distributions. *IEEE Trans Inf Theory* 49(7):1858–1860.
27. Xue Y, Ward JM, Yuwen T, Podkorytov IS, Skrynnikov NR (2012) Microsecond timescale conformational exchange in proteins: Using long molecular dynamics trajectory to simulate NMR relaxation dispersion data. *J Am Chem Soc* 134(5):2555–2562.
28. Hanson WM, Beeser SA, Oas TG, Goldenberg DP (2003) Identification of a residue critical for maintaining the functional conformation of BPTI. *J Mol Biol* 333(2): 425–441.
29. Wyman JJ, Jr. (1948) Heme proteins. *Adv Protein Chem* 4:407–531.
30. Hanson WM, Domek GJ, Horvath MP, Goldenberg DP (2007) Rigidification of a flexible protease inhibitor variant upon binding to trypsin. *J Mol Biol* 366(1):230–243.
31. Breslauer KJ, et al. (1987) Enthalpy-entropy compensations in drug-DNA binding studies. *Proc Natl Acad Sci USA* 84(24):8922–8926.
32. Gessi S, et al. (2008) Thermodynamics of A2B adenosine receptor binding discriminates agonistic from antagonistic behaviour. *Biochem Pharmacol* 75(2):562–569.
33. Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* 104(17):4546–4559.
34. Zídek L, Novotny MV, Stone MJ (1999) Increased protein backbone conformational entropy upon hydrophobic ligand binding. *Nat Struct Biol* 6(12):1118–1121.
35. Arumugam S, et al. (2003) Increased backbone mobility in β-barrel enhances entropy gain driving binding of N-TIMP-1 to MMP-3. *J Mol Biol* 327(3):719–734.
36. Diehl C, et al. (2010) Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. *J Am Chem Soc* 132(41): 14577–14589.
37. Edink E, Jansen C, Leurs R, de Esch IJP (2010) The heat is on: Thermodynamic analysis in fragment-based drug discovery. *Drug Discov Today Technol* 7(3):189–201.
38. Tzeng S-R, Kalodimos CG (2012) Protein activity regulation by conformational entropy. *Nature* 488(7410):236–240.
39. D'Aquino JA, et al. (1996) The magnitude of the backbone conformational entropy change in protein folding. *Proteins* 25(2):143–156.
40. Wells CM, Di Cera E (1992) Thrombin is a Na(+)-activated enzyme. *Biochemistry* 31 (47):11721–11730.
41. Myszka DG, et al. (2003) The ABRF-MIRG'02 study: Assembly state, thermodynamic, and kinetic analysis of an enzyme/inhibitor interaction. *J Biomol Tech* 14(4):247–269.
42. Tellinghuisen J, Chodera JD (2011) Systematic errors in isothermal titration calorimetry: Concentrations and baselines. *Anal Biochem* 414(2):297–299.
43. Dunitz JD (1995) Win some, lose some: Enthalpy-entropy compensation in weak intermolecular interactions. *Chem Biol* 2(11):709–712.
44. Searle MS, Westwell MS, Williams DH (1995) Application of a generalised enthalpy-entropy relationship to binding cooperativity and weak associations in solution. *J Chem Soc, Perkin Trans* 2(1):141–151.
45. Ford DM (2005) Enthalpy-entropy compensation is not a general feature of weak association. *J Am Chem Soc* 127(46):16167–16170.
46. Eftink MR, Anusiem AC, Biltonen RL (1983) Enthalpy-entropy compensation and heat capacity changes for protein-ligand interactions: General thermodynamic models and data for the binding of nucleotides to ribonuclease A. *Biochemistry* 22(16):3884–3896.
47. Weber G (1995) van 't Hoff revisited: Enthalpy of association of protein subunits. *J Phys Chem* 99(3):1052–1059.
48. Qian H (1998) Entropy-enthalpy compensation: Conformational fluctuation and induced-fit. *J Chem Phys* 109(22):10015–10017.
49. Qian H, Hopfield JJ (1996) Entropy-enthalpy compensation: Perturbation and relaxation in thermodynamic systems. *J Chem Phys* 105(20):9292–9298.
50. Sharp K (2001) Entropy-enthalpy compensation: Fact or artifact? *Protein Sci* 10(3): 661–667.
51. Starikov EB, Nordén B (2007) Enthalpy-entropy compensation: A phantom or something useful? *J Phys Chem B* 111(51):14431–14435.
52. Zhou H-X, Gilson MK (2009) Theory of free energy and entropy in noncovalent binding. *Chem Rev* 109(9):4092–4107.
53. Prabhu NV, Sharp KA (2005) Heat capacity in proteins. *Annu Rev Phys Chem* 56: 521–548.