A disruptive dipole-dipole alignment promotes a stable molecular association

L. Ridgway Scott
Institute for Biophysical Dynamics, and
Departments of Computer Science and Mathematics,
University of Chicago, Chicago, Illinois 60637 ridg@uchicago.edu

Ariel Fernández Stigliano
Instituto Argentino de Matemática "Alberto P. Calderón",
CONICET, Buenos Aires 1083, Argentina; and
Collegium Basilea, Institute for Advanced Study,
Hochstrasse 51, Basel, 4053 CH, ariel@afinnovation.com

November 22, 2013

Abstract

An analysis of crystallographically characterized microstates in molecular complexes often reveals metastable dipole-dipole interactions between the binding partners. The statistical mechanical analysis presented in this work demonstrates that these disruptions can paradoxically enhance affinity. The disruption at the microstate level can only be reconciled with the decrease in free energy of association once the full conformational ensemble is considered, yielding both a net decrease in enthalpy and increase in entropy. The full ensemble cannot easily be inferred from crystallographic information. These microstate mismatches are shown to be common in drug-target interactions for recently developed cancer drugs. The reverse engineering of such compounds presented here has suggested that these mismatches at the microstate level may inspire molecular designs that enhance target affinity.

Direct examination of spectroscopically characterized microstates representative of biomolecular complexes can display glaring mismatches at the interface between interacting subunits. By a mismatch, we mean an energetically unfavorable pairwise interaction [1]. Such singularities can be found in drug-target complexes where intermolecular dipole-dipole alignments constitute metastable microstates. For example, a permanent dipole in a ligand may be placed orthogonally to a permanent dipole representing a preformed hydrogen bond in the target protein, as illustrated in Figure 1(a) and shown schematically in Figure 2(a). These singular features emerge from examination of crystal structures of biomolecular complexes (Figure 1(a)). However, a thermodynamic computation subsuming the full statistical ensemble, of which the "anomalous" configuration is only one microstate, reveals a very different physical picture. The ensemble

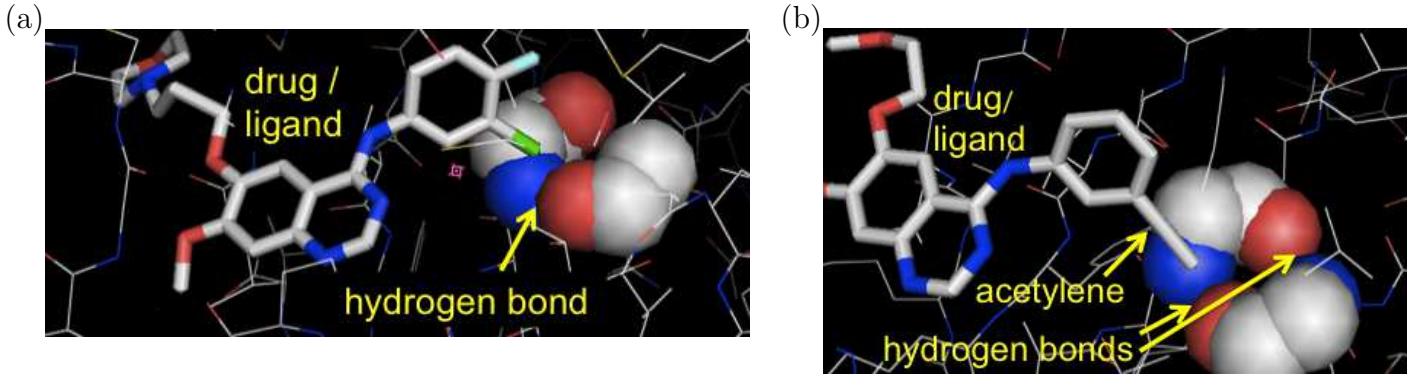


Figure 1: Close-up of disruption in gefitinib (a) and wrapping in erlotinib (b) bound to EGFR in PDB files 1M17 and 2ITY. Oxygens are red, nitrogens are blue, carbons are light grey, fluorine is light blue, and chlorine is green. Space-filling representation is used only for backbone atoms involved in intermolecular hydrogen bonds. The rest of the protein is depicted as lines. The drugs are depicted by sticks.

level analysis reconciles the disruptive intermolecular interaction with a net enhancement in the drug-target affinity. This paradox is resolved by demonstrating that the disruptive interaction is not the representative microstate in the ensemble but simply one microstate out-competed by others that collectively promote an increase in entropy and a decrease in the enthalpy of the complex.

The disruptive dipole in the drug/ligand often is realized as a polarized covalent bond between two atoms of vastly different electrophilicity, a carbon and a halogen (X). A naïve examination of the drug in complex with the target (Figure 1(a)) often reveals the halogen embedded in an environment that deprives it of its hydration layer while the C-X bond is not in the expected energetically favorable antiparallel orientation with respect to the preformed dipole in the target protein (Figure 2(b)). Rather, it is essentially orthogonal to the preformed dipoles in the target protein, as shown in Figure 2(a) and Figure 1(a). This interaction is energetically metastable and the electrostatic potential dictates that the antiparallel orientation represents the stable configuration. However, when the halogen is replaced by a nonpolar group, such as methyl ($-CH_3$) or acetylene ($-C_2H$, Figure 1(b)), the net affinity of the drug for the target protein does not vary significantly relative to the halogen substitution.

This is surprising since the addition of the nonpolar group would be expected to enhance the stability and strength of the preformed hydrogen bond, the so-called wrapping effect [2, 3, 4, 5, 6], while the orthogonal dipole associated with the halogen substitution in the ligand has a disruptive effect on the preformed hydrogen bond upon association. Yet, while the microstate embodying the wrapping is dominant in the case of a ligand with a nonpolar substituent [2, 3, 4, 5, 6], the microstate with the disruptive dipole-dipole interaction is averaged out in the corresponding ensemble and is not representative of the ensemble, as shown in this work. So the two energetic effects may differ at the microstate level but become comparable when regarded as part of their respective statistical ensembles of conformations. In other words, the wrapping state is well represented by the exogenous nonpolar group in the proximity of the preformed hydrogen bond, whereas the dipole-dipole disruption triggers a dynamic process that mitigates the entropic cost of the protein-ligand association, and it is even enthalpically favorable, as demonstrated in this work.

The disruptive microstate is therefore part of an ensemble that represents a far larger explo-

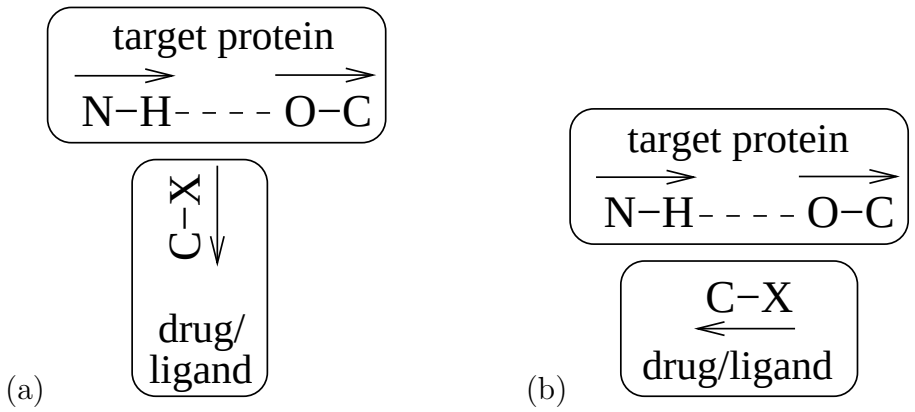


Figure 2: Metastable (a) and energetically favorable (b) dipole orientations for hydrogen bonds and halogen groups. The letter X indicates the halogen (fluorine or chlorine). The arrows indicate the dipole moments for each pair of atoms.

ration of conformation space than the ensemble that contains the wrapping microstate. Thus, while the two microstates described previously can be contrasted in energetic terms, the free-energies associated with the respective ensembles are indeed very similar, reflecting the similar affinities of the two compounds with different substituents.

A lesson to be drawn from our analysis is that the efficacy of a drug cannot be assessed from direct examination of a crystallographic microstate, but requires a full thermodynamic analysis of the corresponding ensemble that represents the drug–target association. Similarly, optimization in drug design cannot be dictated by the examination of a structural microstate, but becomes intimately correlated with a free-energy minimization.

Optimization of free energy is a minimax problem, not a simple minimization problem. As a saddle point problem, free-energy optimization requires more complex strategies to achieve beneficial results [7]. The free-energy change ΔG is related to the enthalpy change ΔH and the entropy change ΔS via

$$\Delta G = \Delta H - T\Delta S, \tag{1}$$

where T is temperature. Thus we can make ΔG smaller either by making ΔH smaller, or by making ΔS larger. The former has been the common strategy in drug design [8, 7, 5]. On the other hand, the latter has been observed in naturally occurring systems [9, 10] and in synthetic systems [1, 11, 12, 13], but it has not been widely disclosed as a drug-design strategy [14].

Although it is unlikely that ΔS will actually be positive due to the loss of conformational freedom upon binding, it may be possible to minimize the impact of the entropic loss by liberating degrees of freedom in the target. In general, there is a substantial correlation between enthalpic decrease and entropic decrease in ligand binding [15, 16, 17], known as enthalpy-entropy compensation, that typically dictates that only small free-energy gains are obtained. However, it is possible to tilt the table in favor of entropic gain in some cases as illustrated in this paper, and this often correlates with improved bioavailability. Thus an interaction that both increases entropy and decreases enthalpy has the potential of great benefit.

We have chosen to present the data first from a conventional point of view, both to make it clear that there are not favorable polar contacts of the expected type and to show where the standard thinking fails. We then present a model system whose statistical mechanics can be solved exactly to explain the data. Our assertions are validated in this paper by constructing a

(a) (b)

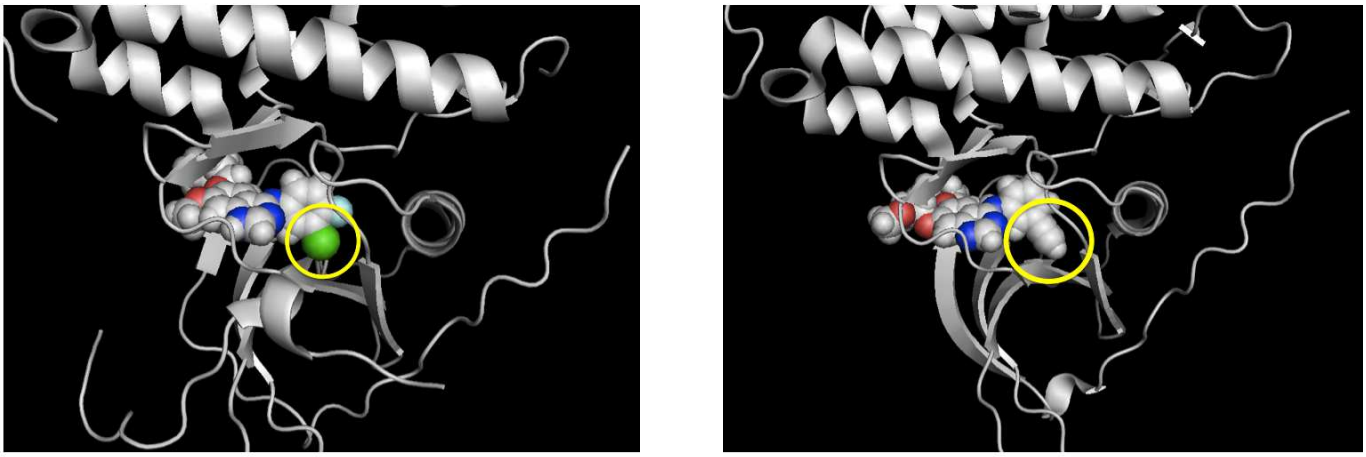


Figure 3: Gefitinib (a) and erlotinib (b) bound to EGFR in PDB files 1M17 and 2ITY. Color scheme as in Figure 1. Yellow circles highlight chlorine in panel (a) and acetylene ($-C\equiv C$-H) in panel (b). Space-filling representation is used only for the drug. The the protein is depicted as a cartoon.

model that enables the thermodynamic evaluation of the dipole-dipole configuration ensemble. In accord with this validation, the paper evolves along the following lines.

First we describe drug/ligands in complex with their respective target proteins to illustrate the dichotomy "disruption versus wrapping." In the subsequent section, we describe mechanistically the emergence of the disruptive metastable microstate by examining relative intermolecular dipole orientations in protein–ligand complexes. After this illustrative example, we develop a statistical mechanical model to compute the thermodynamic parameters determinant of the ensemble that contains the dipole-dipole disruptive state. This is followed by an evaluation of configurations realizing the minimum in potential energy and the mathematical derivation of the thermodynamic parameters as functionals of the potential energy. The results are then contrasted vis-a-vis a systematic survey of structures of drug-protein complexes reported in the Protein Data Bank, the standard repository for protein-ligand structural information. This analysis serves as the basis for a series of considerations on rational drug design that is introduced subsequently. By reversely engineering the "anomalous" mismatches at the protein-ligand interface, a clear molecular design concept emerges, and its impact on rational drug design is assessed.

1 Disruption versus wrapping

From a structural biology perspective, the incorporation of halogens in drugs may be perceived as problematic because the drug–target interfaces reveal mismatches leading to metastable microstates as illustrated in Figure 1(a). The drugs gefitinib and erlotinib provide a visual representation of the contrast between drugs that incorporate halogens and the earlier generation of drugs that does not. The acetylene group attached to the carbon ring in erlotinib (Figure 3(b)) is replaced by a chlorine (Figure 3(a)) in gefitinib. Figure 3 shows that the gross structure of the target EGFR kinase is the same with both ligands (and it is similar to the unbound form of the EGFR kinase in PDB file 1M14, not shown).

The extensive availability of PDB files depicting EGFR both bound and unbound (Table 1)

PDB	mutation	ligand	res.	d_{ON}	d_{OH}	$\angle_{NHO}$	$\angle_{HOC}$	W
1M14	none	unbound	2.60	2.89	1.94	156.3	157.0	28
1M17	none	erlotinib	2.60	2.86	1.89	161.1	147.1	32
2ITY	none	gefitinib	3.42	2.57	1.63	154.4	127.3	34
2ITO	G719S	gefitinib	3.25	2.93	2.20	128.8	146.9	32
2ITZ	L858R	gefitinib	2.72	2.80	1.87	154.9	139.4	31

Table 1: Hydrogen bond quality data for EGFR complexed with various ligands. The designators d_{AB} indicate the distance between atoms A and B in Ångstroms; res. denotes the resolution specified in the corresponding PDB file, in Ångstroms. The designators $\angle_{ABC}$ indicate the angle (in degrees) formed by the three atoms A, B, C. W indicates the number of wrappers coming from the (unbound) protein inside a desolvation domain based at the C-alpha carbons with radii of 6Å. All residues are in chain A in the respective PDB files. The donor for the hydrogen bonds is Lys 721 in 1M14 and 1M17 and Lys 745 in 2ITY, 2ITO and 2ITZ. The acceptor for the hydrogen bonds is Leu 764 in 1M14 and 1M17 and Leu 788 in 2ITY, 2ITO and 2ITZ. (The sequence numbers in 1M14 and 1M17 are 24 less than in 2ITY, 2ITO and 2ITZ.) Donor refers to the mainchain pair N-H; acceptor refers to the mainchain pair O-C.

d_{ON}	d_{OH}	$\angle_{NHO}$	$\angle_{HOC}$
2.91 ± 0.14	1.96 ± 0.16	160 ± 10	150 ± 12

Table 2: Average hydrogen bond quality data for hydrogen bonds in antiparallel beta sheets [18]. Notation and units are the same as in Table 1.

allows us to dissect the effects of various ligands upon binding. There is a conserved double hydrogen bond (between Lys 721 and Leu 764 in PDB file 1M17) closest to the chlorine in gefitinib and the acetylene group in erlotinib, as shown in Figure 3. These hydrogen bonds are in the binding pocket in EGFR, and they display different conformations for different ligands.

The concept of wrapping [5, 2, 6] has been used extensively to estimate the stability and strength of preformed hydrogen bonds. This is done by assessing the number of nonpolar carbonaceous groups CH_n ($n = 0, 1, 2, 3$) in a specified neighborhood of a hydrogen bond. A lack of sufficient carbonaceous groups allows the approach of water molecules that tend to disrupt the hydrogen bond. Such underwrapped hydrogen bonds, dubbed dehydrons [2], are in fact sticky [3] and provide sites for ligand attachment [5, 2]. The mechanism of attachment is based on the change in dielectric environment due to additional carbonaceous groups in the ligand, such as the acetylene group in erlotinib, in the neighborhood of the hydrogen bond, further excluding water and thereby reducing dielectric screening.

The hydrogen bonds in the active site of EGFR nearest to the acetylene group in erlotinib and the chlorine in gefitinib are listed in Table 1. They all appear to be well wrapped within the unbound protein according to a simple count of wrappers located inside a 6 Ångstrom dehydration domain [2], but Figure 3 shows there is a region where further wrapping is inserted by erlotinib, as well as where penetration by the chlorine on gefitinib can also occur. The chlorine has the propensity to attract waters dynamically, and so this pocket near a hydrogen bond is an apparent defect. It is reasonable to speculate that water could also enter such a cavity, disrupting the hydrogen bond. However, the native amount of wrapping is quite high, and this suggests that the change in wrapping upon binding for erlotinib can be compensated by some other mechanism with gefitinib. This is the main focus of the paper.

d_{OX}	d_{CX}	d_{HX}	$\angle_{CO}^{CX}$	$\angle_{HN}^{CX}$	PDB	X	donor	acceptor
3.29	3.85	2.61	121.3	92.9	2ITO	CL	Lys 745	Leu 788
3.18	3.86	3.70	139.6	118.2	2ITY	CL	Lys 745	Leu 788
3.22	3.40	3.68	70.0	87.6	2XP2	F	Gly 1269	Asn 1254
3.22	3.40	3.42	70.0	122.8	2XP2	F	Gly 1270	Asn 1254
3.22	3.53	4.26	76.5	105.2	3BBT	CL	Thr 771	Ala 724
3.25	4.21	3.12	117.5	106.8	3BBT	CL	Lys 726	Leu 769
3.40	3.98	3.67	71.0	42.7	3GCS	F8	Ile 84	Ile 166
3.22	4.08	5.27	108.2	84.4	3G0E	F	Gly 812	Asp 810
3.26	3.62	3.35	70.4	69.6	3OCT	CL(B)	Thr 474	Ala 428
3.40	4.21	3.30	113.5	118.8	3OCT	CL(B)	Lys 430	Ile 472
4.13	5.12	3.13	66.4	55.0	1M52	CL2	Lys 271	Ile 313
4.23	4.57	3.38	61.8	63.8	2ITO	CL	Thr 790	Ala 743
3.77	4.08	3.76	119.1	126.5	2ITZ	CL	Lys 745	Leu 788
3.99	3.93	5.11	149.3	123.3	3CS9	F1	Val 299	Val 379
3.85	4.43	4.14	46.1	20.1	3CS9	F4	Val 299	Val 379

Table 3: Dipole angle data for disruptors. The designators $\angle_{CD}^{AB}$ defined in (2) indicate the angle between the vectors corresponding to the atom pairs AB and CD. All residues are in the A chain of the respective PDB files except for PDB file 3BBT for which it is the B chain. The chlorine in PDB file 3OCT has an alternate position, and the data presented is for the B position. The designator X indicates the halogen of interest; the notation used is taken from the PDB files (so CL or CL2 indicates a particular chlorine atom in the PDB file). For further notation, see Table 1.

The hydrogen bond in the erlotinib/EGFR complex involving Lys 721 and Leu 764 (Table 1) is similar in quality to the unbound EGFR complex (PDB file 1M14). Indeed, we see a slight improvement of the angular parameters for the hydrogen bond when the stabilizing acetylene group inserts near this hydrogen bond pair. For comparison, we give the average distance and angle data for hydrogen bonds in antiparallel beta sheets [18] in Table 2, since the Lys 721 — Leu 764 hydrogen bond is in an antiparallel beta sheet. By contrast, this hydrogen bond is *disrupted* in all of the gefitinib complexes given in Table 1, corresponding to PDB files 2ITY, 2ITO and 2ITZ. (The sequence numbers in 1M14 and 1M17 are 24 less than in 2ITO, 2ITY and 2ITZ.)

We recall that all data in PDB files have a limited resolution, and the resolution data is given in Table 1 for reference. However, if we take the atom position data at face value, then the data in Table 1 contrast the disruption due to gefitinib versus both the unbound EGFR and the additional wrapping [2] provided by erlotinib.

The intermolecular configuration involving a halogen in close proximity to a preformed hydrogen bond is perplexing for at least two reasons: (a) the hampered hydration of the chlorine in the gefitinib–EGFR kinase complex which is at odds with the high hydration demands of the C-Cl polarized bond, (b) the metastable positioning of C-Cl dipole with regard to the dipole of the preformed hydrogen bond Lys 745 — Leu 788.

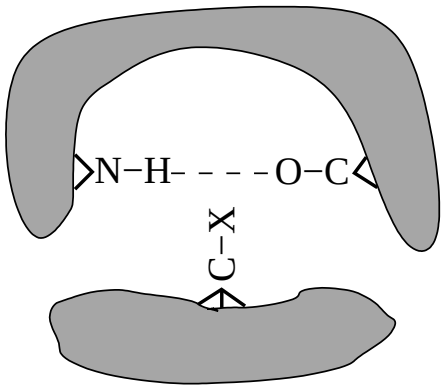


Figure 4: Constraints on a model due to the protein and ligand configurations. The upper shaded region is the protein, with a backbone hydrogen bond emerging from it. The lower shaded region is the ligand, with the halogen rigidly fixed.

2 Disruption analysis

The effect of the gefitinib chlorine on the structure of the target is baffling. Other than the atoms in the Lys 745 and Leu 788 backbone, there are no polar entities within 4.5 Å of the chlorine (in 2ITY) that could form any sort of energetically beneficial interaction. To understand the possible interactions between the chlorine and the Lys 745 and Leu 788 backbone hydrogen bonds, we examined the geometry to see if the chlorine and its neighboring carbon CAX (PDB notation) in gefitinib might be forming some favorable dipole interactions. Indeed, as shown in Table 3, there is nothing favorable about these interactions. The hydrogen-chlorine distance is greater than the oxygen-chlorine distance. Similarly, the chlorine is further from the backbone carbon than the oxygen.

Since there are several angles used to describe a hydrogen bond, the separate angles $\angle_{CO}^{CX}$ and $\angle_{HN}^{CX}$ for the two components (acceptor and donor, respectively) of the hydrogen bonds have been given for completeness, where X denotes the halogen of interest. These are defined via the equation

$$\cos\left(\angle_{CD}^{AB}\right) = \frac{\overline{AB}\cdot\overline{CD}}{|\overline{AB}|\,|\overline{CD}|}. \tag{2}$$

The carbon closest to the halogen in the drug was used to define the CX vector.

As shown in Table 3, the angle of approach of the CAX-chlorine dipole in 2ITY is nearly perpendicular to the Lys 745—Leu 788 hydrogen bond, similar to Figure 2(a), indicating a metastable interaction. We will see in a simple model that this type of interaction can actually be both enthalpically and entropically favorable.

3 A simple model of disruption

In building a model for the ensemble-level effects of disruption, we must respect the relevant constraints. In Figure 4 we present a cartoon of these constraints. The hydrogen bond donor and acceptor groups emanate from different parts of the protein chain. They are free to rotate to some extent. The halogen is fixed on the ligand in a particular orientation. Thus we define our model to reflect these requirements.

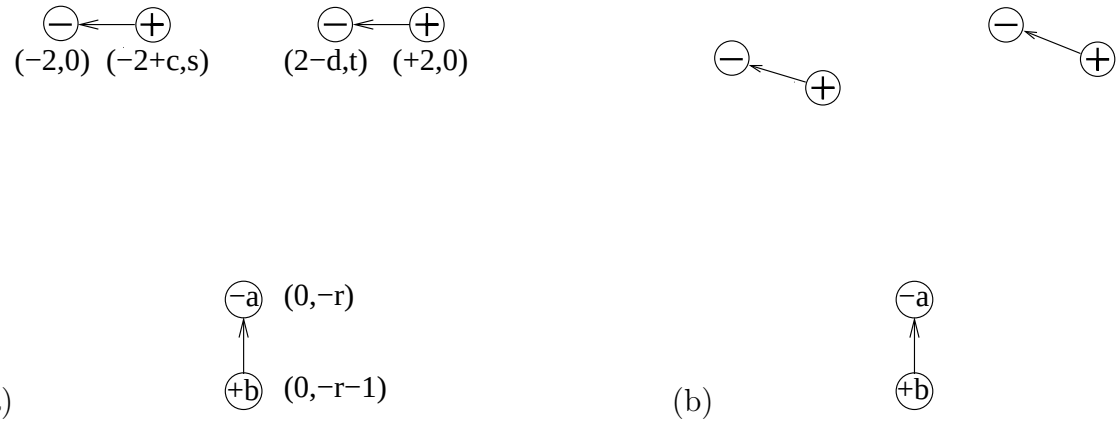


Figure 5: (a) Coordinates for a simple model of a hydrogen bond disrupted by a dipole at a distance r (horizontal axis) from the center of the hydrogen bond. The coefficients: $c = \cos\theta$, $s = \sin\theta$, $d = \cos\phi$, $t = \sin\phi$. (b) Configuration with lower potential energy. The charges marked by a + or − are ± 1 in strength.

We can model the system in Figure 2(a) and Figure 4 as indicated in Figure 5(a). The Cartesian coordinates are chosen with the origin at the midpoint between the two dipoles representing the donor and acceptor of the preformed hydrogen bond. We take the outlying ends of the dipoles to be fixed, with the the x-axis chosen to intersect these two points. The orientation of the dipoles is allowed to vary; we represent these rotational angles in local polar coordinates θ and ϕ based at the respective fixed ends of the dipoles (representing the hydrogen bond donor-antecedent nitrogen and acceptor-antecedent carbon). This is a very simplistic model of a hydrogen bond, ignoring inductive effects, but it is the same as in widely used molecular force fields. Based on these coordinates, we can compute the electrostatic potential energy $V(r,\theta,\phi)$ of system, as follows:

$$\begin{aligned} V(r,\theta,\phi) = b\left(\frac{1}{\sqrt{(2-c)^2+(r+1+s)^2}} - \frac{1}{\sqrt{(2-d)^2+(r+1+t)^2}}\right) \\ + a\left(-\frac{1}{\sqrt{(2-c)^2+(r+s)^2}} + \frac{1}{\sqrt{(2-d)^2+(r+t)^2}}\right) - 1/4 \\ - \frac{1}{\sqrt{(4-d-c)^2+(s-t)^2}} + \frac{1}{\sqrt{(4-d)^2+t^2}} + \frac{1}{\sqrt{(4-c)^2+s^2}}, \end{aligned} \tag{3}$$

where $c = \cos\theta$, $s = \sin\theta$, $d = \cos\phi$, $t = \sin\phi$.

For any metastable system, there are configurations with lower potential energy nearby, and we indicate one in Figure 5(b). Such a configuration does represent the minimum energy state, as we will see in Figure 7, and this suggests that we name this mechanism "disruption" since it involves a perturbation of the original dipole pair symmetry. However, the existence of some configurations with lower potential energy does not imply that the overall average potential energy (enthalpy) will be lower, nor does it hint at the entropic effects of bringing the disruptor dipole from $(0,-\infty)$ to the point $(0,-r)$.

We have allowed the dipole that approaches from below to have charges other than ± 1. Some constituents of protein sidechains involve net charges, such as the residues Asp and Glu, so we have experimented with values of a greater than 1 to include assessment of such interactants.

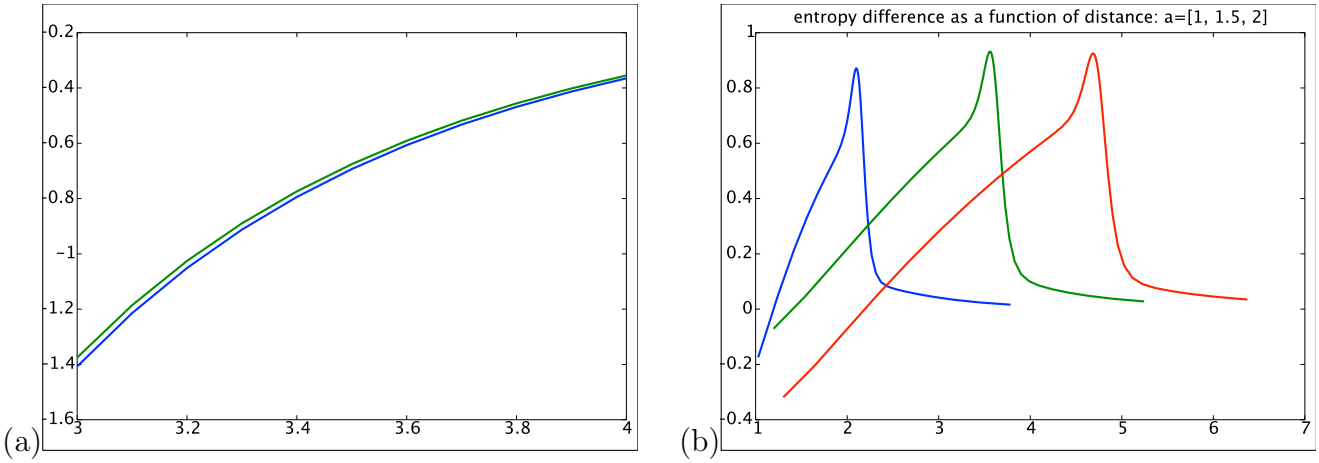


Figure 6: Panel (a) depicts changes in enthalpy (average potential energy, upper curve) and free energy (lower curve) defined in (7) for $a = 1$ as a function of separation distance for a dipole at a distance r (horizontal axis, in Ångstroms) from the center of the hydrogen bond. Panel (b) depicts changes in entropy (in units of $T\Delta S(r)$), for $a = 1$ (left), 1.5 (middle), and 2 (right) as a function of separation distance for a dipole at a distance r (horizontal axis, in Ångstroms) from the center of the hydrogen bond. The vertical scale is in kcal/mole. A temperature of T =303 K was chosen.

For this two-dimensional system, we can evaluate the free energy [19, 20] as

$$\begin{aligned} G(r) &= -k_B T \log Z(r) \\ &= -k_B T \log \left(\int_{[-\pi,\pi]^2} e^{-V(r,\theta,\phi)/k_B T} \, d\theta \, d\phi \right), \end{aligned} \tag{4}$$

where k_B is Boltzmann's constant, T is temperature,

$$Z(r) = \int_{[-\pi,\pi]^2} e^{-V(r,\theta,\phi)/k_B T} \, d\theta \, d\phi, \tag{5}$$

and V denotes the potential (electrostatic) energy of system in Figure 5(a) given in (3). Similarly, the average potential energy (enthalpy) $H(r)$ is given by

$$H(r) = \frac{1}{Z(r)} \int_{[-\pi,\pi]^2} V(r,\theta,\phi) e^{-V(r,\theta,\phi)/k_B T} \, d\theta \, d\phi. \tag{6}$$

In Figure 6(a), we plot the free energy and enthalpy differences

$$\Delta G(r) = G(r) - \lim_{r' \to \infty} G(r') \quad \text{and} \quad \Delta H(r) = H(r) - \lim_{r' \to \infty} H(r') \tag{7}$$

as a function of r at a temperature of $303K$. This figure indicates that the metastable orientation of the halogen disruptors is both enthalpically and entropically favorable, and the change in free energy at a distance of 3.2-3.4 Å is significant. It also shows that the effect is fairly short range, having little residual effect beyond about 5 Å (the free energy difference decreases like r^{-6} for large r as shown in Figure 11).

The amount of angular disruption for the configuration of minimum potential energy is not substantial, as shown in Figure 7, being smaller than 10 degrees for the distances r of interest.

Thus they are smaller than the error bars for angular data in PDB files indicated in Table 2 for the cases of interest here. This perhaps helps to explain why this type of interaction has escaped notice previously.

The entropy S can be defined by

$$S(r) = \frac{H(r) - G(r)}{T}. \tag{8}$$

For most values of r of interest, the entropy change

$$\Delta S(r) = S(r) - \lim_{r' \to \infty} S(r') \tag{9}$$

is positive, as indicted in Figure 6(b). Strikingly, there appears to be an entropy jolt at critical separation values, followed by a decrease in entropy difference as separation is further decreased, ultimately becoming negative.

4 Configuration of minimum potential energy

One might think naively that thermodynamic properties could be inferred by observing the configuration $(\theta(r), \phi(r))$ of minimum potential energy as a function of the interaction distance. However, the depiction in Figure 5(b) is an exaggeration. The actual displacement of the position of minimum potential energy is quite small, as indicated in Figure 7. The angles $\theta(r)$ and $\phi(r)$ decay like r^{-3} for large r (data not shown).

In the range $r > 2.6$ where the model has the most applicability for proteins, the angular deformation is less that 10 degrees. Hydrogen bonds in PDB files are usually not resolved to this accuracy [18]. Thus simply looking at a PDB file would not indicate an evident disruption even though the change in free energy is substantial.

5 Systematic survey of the disruption motif

The motif found in gefitinib associated with the single halogen was found with several other recently designed cancer drugs. The data in Table 3 indicate key hydrogen bonds nearby the halogens in various cancer drugs (Table 4). As expected for a metastable interaction, the angles between the carbon-halogen in the drug and the hydrogen bond components cluster around 90 degrees. Moreover, the halogen is located in a very tight range between 3.2 Å and 3.4 Å from the oxygen acceptors of the hydrogen bonds. The halogen is often closer to the oxygen than the hydrogen of the hydrogen bond, and it is always closer to the oxygen than the carbon of the hydrogen bond. This polar mismatch is the striking feature of this motif. We emphasize that in all cases, the closest atom listed (O or H) is the closest atom to the halogen among all polar atoms in the target protein.

We give in Table 5 data regarding the nearby hydrogen bonds listed in Table 3. We can only talk about normative behavior for hydrogen bonds in secondary structure, but most of these hydrogen bonds are in antiparallel beta sheets, and most of these exhibit some form of defect compared with the expected norm given in Table 2. Thus we will describe the halogens featured in Table 3 as disruptors. However, our main motivation is not based on the fine properties of the interacting hydrogen bonds, but rather on the thermodynamics related to the geometry

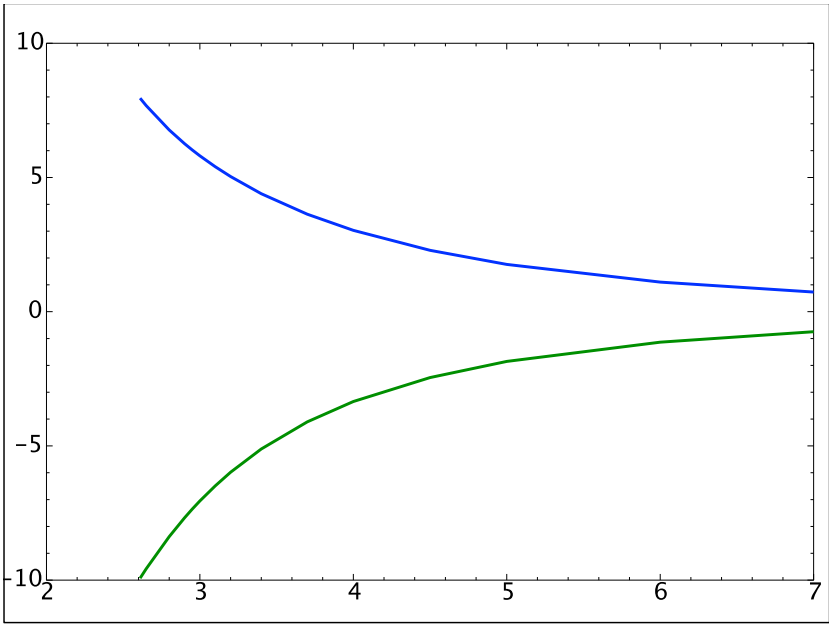


Figure 7: Configuration (θ, ϕ) of minimum potential energy as a function of separation distance r. Plotted are $\theta(r)$ and $\phi(r)$ in degrees for the case $a = b = 1$, T =303K. Note that $\theta < 0 < \phi$.

PDB	target	drug name	PDB code
1M17	EGFR	erlotinib	AQ4
1M52	C-ABL	PD173955	P17
2ITY	EGFR	gefitinib	IRE
2XP2	ALK	crizotinib	VGH
3BBT	ERBB4	lapatinib	FMM
3G0E	c-Kit	sunitinib	B49
3GCS	MAPK14	sorafenib	BAX
3GP0	MAPK11	nilotinib	NIL
3OCT	BCR/ABL	dasatinib	1N1

Table 4: Data for several drug/target complexes. The targets are protein kinases, the signal transducers of the cell which get activated through a phosphorylation event that in turn enables them to phosphorylate their respective targets. The drugs are effectively molecular targeted agents designed to block the signaling pathways by inhibiting specific kinases. We use some shorthand notation such as EGFR to indicate the kinase of the epithelial growth factor receptor. The first column lists the PDB file, the second the protein target, the third the generic drug name, the fourth the PDB code for the drug molecule.

of the dipoles indicating a metastable interaction as opposed to an energetically favorable one. Moreover, the most striking feature of the motif is the lack of any other polar entity within 4 Å of the halogen.

The putative hydrogen bond in the sunitinib complex can be seen from Table 5 to be so disrupted that it is at best a poorly formed hydrogen bond (and note that it is not natively very well wrapped). In addition, the fluorine in 3G0E potentially makes a weak sidechain hydrogen bond with Cys A 809, with the distances from the fluorine to HG and SG being 3.16 Å and 3.83Å, respectively.

The G719S mutation represented in 2ITO impacts the location of gefitinib substantially. In particular, OG Ser A 719 is in close proximity (about 3.4Å) to CAN and CAP in IRE (gefitinib) in the PDB file 2ITO. On the other hand, the closest Gly A 719 heavy atom (in 2ITY) is about 3.6Å from an IRE heavy atom, whereas the closest backbone atom (CA) of SER A 719 to a heavy atom of IRE (CAN in 2ITO) is 5.03Å. Thus it is the emergence of the sidechain in the mutation Ser A 719 making the contact with gefitinib. This mutation changes the distances of the hydrogen bond proximal to the halogen, but the general character of disruption is unchanged by the mutation.

The L858R mutation represented in 2ITZ does not impact the location of gefitinib in an obvious way. The closest heavy atom in Leu A 858 (in 2ITY) to a heavy atom in IRE is the backbone nitrogen. The distance from N Leu A 858 (in 2ITY) to FAB is 9.85Å. The closest heavy atom in ARG A 858 (in 2ITZ) to a heavy atom in IRE is also the backbone nitrogen. The distance from N Leu A 858 (in 2ITZ) to FAB is 10.03Å. On the other hand, the chlorine in gefitinib is substantially further from the Lys 745 — Leu 788 hydrogen bond in 2ITZ, as indicated in Table 3, than in the native complex 2ITY. More precisely, the mismatch distance between the halogen and the oxygen atom is about half an Ångstrom larger. Correspondingly, the distance and angle data for the hydrogen bond proximal to the halogen are within the normal range for an antiparallel beta sheet.

Not all halogens in these drugs are disrupting a hydrogen bond. Indeed, it would be difficult for all of the fluorines in a trifluorine group to engage separate hydrogen bonds, just due to geometric constraints regarding the locations of hydrogen bonds. On the other hand, chlorines in 3BBT and 3OCT, and the fluorine in 2XP2, are each disrupting two hydrogen bonds.

In all but three of the cases studied so far, each drug containing a halogen has at least one halogen disrupting a nearby hydrogen bond as described here. Only the compounds nilotinib (PDB files 3CS9 and 3GP0) and PD173955 (1M52) do not engage hydrogen bonds as near its halogens as is the common case in Table 3. We have segregated these cases in Table 3 for clarity. Other donor–acceptor interactions are listed there to complete the roster of polar interactions near the indicated halogen.

There is a single (well formed, well wrapped) hydrogen bond in the nilotinib complex (3CS9) proximal to two different halogens. However, these groups have both larger distances and more deviant angles between the dipoles, and the corresponding hydrogen bond has normal angles and distances. There is also a possible sidechain hydrogen bond between F1 and His A 148 in 3GP0; the distance between HE2 and F1 in 3GP0 is 3.22 Å, and the distance between NE2 and F1 in 3GP0 is 3.59 Å.

The chlorine CL2 in compound PD173955 has a nearby (and well formed) hydrogen bond as listed below the line in Table 3, but the distances to the bond are significantly larger. In addition, the chlorine CL2 is possibly making a sidechain hydrogen bond with Thr A 315, with the distances from the chlorine to HG1 and OG1 being 2.93 Å and 3.81Å, respectively. The

PDB	SS	donor	acceptor	d_{ON}	d_{OH}	$\angle_{NHO}$	$\angle_{NOC}$	$\angle_{HOC}$	W
1M52	∦S	Lys A 271	Ile A 313	2.91	1.95	160.1	156.0	155.5	36
2ITO	∦S	Thr A 790	Ala A 743	2.92	2.09	138.9	156.0	144.6	29
2XP2	L	Gly A 1269	Asn A 1254	2.78	1.81	164.5	161.4	165.6	29
2XP2	L	Asp A 1270	Asn A 1254	3.37	2.52	142.2	135.2	135.7	23
3BBT	∦S	Lys B 726	Leu B 769	2.73	1.74	172.5	148.7	146.1	40
3BBT	∦S	Thr B 771	Ala B 724	2.93	2.15	133.6	156.7	147.8	38
3CS9	LS	Val A 299	Val A 379	2.94	1.99	157.1	159.5	158.5	31
3G0E	HL	Gly A 812	Asp A 810	3.63	3.23	105.9	66.7	57.5	22
3GCS	LS	Ile A 84	Ile A 166	2.82	1.88	155.3	160.0	166.7	31
3OCT	∦S	Lys A 430	Ile A 472	2.72	1.72	172.1	157.6	155.4	36
3OCT	∦S	Thr A 474	Ala A 428	3.02	2.05	163.3	158.8	153.6	42

Table 5: Hydrogen bond quality data from `wrappa` [6, 21], a program to identify hydrogen bonds and their wrapping. Abbreviations: donor=hydrogen bond donor, acceptor=hydrogen bond acceptor. SS denotes the secondary structure (H=helix, S=sheet, L=loop); AB means that the donor is in structure A and the acceptor in structure B. The symbol ∦S indicates that the sheet is anti-parallel. For further notation, see Table 1.

environment of chlorine CL2 in compound PD173955 is fairly polar.

We have listed all of the favorable polar interactions found for the listed halogens within a radius of 4 Å (distance between the halogen and the positive component of the dipole). It is notable that the hydrogen bonds more distant from the halogen disruptors (in 1M52 and 3CS9) are better formed than hydrogen bonds closer to their disruptors. Also, all chlorines are interacting with hydrogen bonds in antiparallel beta sheets. The residue pairs for these hydrogen bonds participate in two hydrogen bonds with the roles of donor and acceptor reversed for the complementary hydrogen bonds. These hydrogen bonds are often even more disrupted, even though they are further from the disrupting chlorine, perhaps due to the geometry of the backbone.

For the halogens not listed, there can be favorable polar interactions. For example, the chlorine CL1 in 1M52 makes a potential hydrogen bond with the mainchain N-H group in ASP A 381. The distance between H ASP A 381 and the chlorine CL1 in 1M52 is 2.78 Å, and the distance between N ASP A 381 and the chlorine CL1 in 1M52 is 3.61 Å. In addition, the chlorine in 3BBT may be making a favorable polar interaction [22] with THR B 771, with the distances from the chlorine to HB and CB being 2.61 Å and 3.67Å, respectively. One other possible such interaction is between the chlorine and THR B 790 in 2ITY, but the distances from the chlorine to HB and CB are 3.68 Å and 4.00Å, respectively, so the distances and angle are not ideal. There is a potential sidechain hydrogen bond with His A 148 in 3GCS; the distance between ND1 and F10 in 3GCS is 3.29 Å. However, MolProbity indicates that NE2 is protonated instead of ND1, perhaps due to its proximity to the backbone and sidechain oxygens of Asp A 168.

The hydrogen bonds disrupted by the halogens studied here are found either in sheets or loops or at the end of a helix. None of them are embedded in the interior of a helix.

Data on the individual hydrogen bonds from `wrappa` [6, 21], a program to identify hydrogen bonds and their wrapping, are given in Table 5. Note that all of these bonds are distorted to some extent, compared with the expected parameters given in Table 2, and that the one in 3G0E is a very poor hydrogen bond at best. What is quite surprising, but consistent with our overall

picture, is that each of the disrupted hydrogen bonds is very well wrapped by the native protein structure. Thus the mechanism of disruption is applied to a natively strong and stable hydrogen bond. In Table 4, we list the drug and target for the various PDB files analyzed.

Temperature factors in PDB files provide possible clues about atom mobility. We examined the temperature factors for each of the atoms discussed here. On average, the heavy atoms involved in disrupted hydrogen bonds have slightly elevated temperature factors (about 0.2 standard deviations above the mean), but the temperature factors are not uniformly elevated. (Comparisons were made in each PDB file separately with the average temperature factors of the C_α carbons in helices and sheets.) However, the data in all of the PDB files represents low temperature (about 100 K) behavior, and thus entropic effects are greatly suppressed compared with biological temperatures.

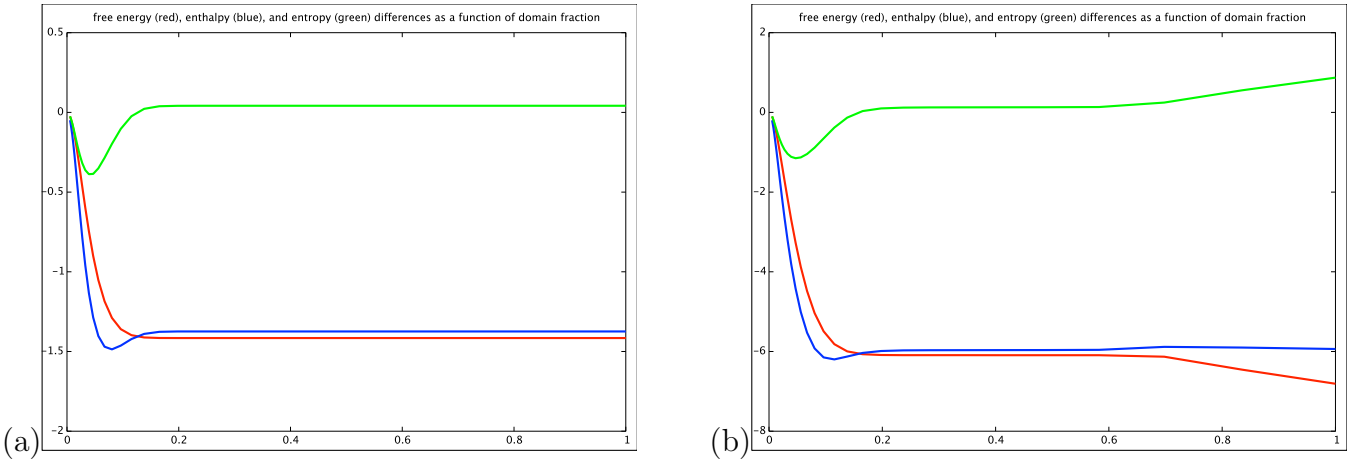
6 Disruption-based drug design

We can describe the drug design process in terms of optimization terminology. Combinatorial chemistry is used to get initial guesses. Then small modifications of the lead compounds are made to achieve higher affinity (and perhaps also greater specificity). The latter step is analogous to exploring potential descent directions in optimization.

If all possible chemical modifications of lead compounds are considered, then there is a combinatorial explosion of possibilities to be explored. However, recently a more limited set of search directions has been introduced which leads to a smaller search space via wrapping technology [5, 4]. The concept of wrapping has been used to redesign imatinib effectively [23]. However, wrapping as a design strategy tends to diminish the solubility of candidate drugs, and thus may restrict the bioavailability of novel ligands. By contrast, the introduction of halogens enhances solubility, and it has an effect opposite to wrapping at the microstate level but not the ensemble level. Wrapping tends to stabilize hydrogen bonds [6] at the microstate level, whereas the halogen substitutions studied here appear to disrupt hydrogen bonds. Thus the addition of halogens at strategic locations provides a new mechanism to modify lead compounds while increasing bioavailability. The free-energy change due to the disruption of a hydrogen bond by a halogen may be too small to be the dominant factor in ligand binding, but the enhanced solubility of the ligand makes it an appealing option.

In the current paper, we have explored a nonconventional local optimization step, whereby the halogen substitution enhances the affinity for its target in a context dependent manner. To avoid unproductive steps, the local optimization needs to be understood in terms of the enhancement of the protein/ligand association. Although the initial combinatorial chemistry step makes a huge reduction in the search space, the remaining local variations still suffer combinatorial blowup. With a rational way to limit local searches, such as revealed by this dynamics-based study, the process becomes more manageable. Moreover, the initial combinatorial chemistry step can only explore a small fraction of the potential lead compounds, so better local searches can allow the extension of less promising lead compounds at reasonable expense.

Finally, the full optimization problem of drug design is a constrained one. We seek the lowest free energy, but the search space must be constrained to avoid side effects [5] as well as insure bioavailability. The number of possible side effects is so large that an enhanced rational approach becomes imperative.



(a) (b)

Figure 8: Differences of free energy $\Delta G_f(r)$ (lowest curve), average potential energy $\Delta H_f(r)$ (middle curve), and temperature-times-entropy $T\Delta S_f(r)$ (upper curve) as defined in (11) as a function of the fraction f (horizontal axis) of the domain of integration as in (10), for the case $a = b = 1$, T =303K. (a) $r = 3$. (b) $r = 2.1$. Units for the vertical axis are kcal/mole.

7 Potential energy locality

One concern about the model presented here relates to the range of angles ϕ and θ allowed. Computations were done allowing then to range over the full possibilities $[-\pi, \pi]$. This was done in part for convenience, since the resulting integrals are periodic and the quadrature becomes simpler (see Section 9), but also because this removes the need for another parameter in the model. However, we will see that the conclusions drawn here about the model do not depend on the large range that was allowed. Rather, the range can be made quite small.

The free energy, average potential energy, and entropy are all functionals of the potential energy. We have argued that it is not sufficient to consider the potential at isolated points. Moreover, these functionals are not local by definition. But the potentials themselves can be quite localized in many cases. To examine the locality quantitatively, we can restrict the domain of integration via

$$\begin{aligned} Z_f(r) &= \int_{[-f\pi, f\pi]^2} e^{-V(r,\theta,\phi)/k_B T}\, d\theta\, d\phi \\ H_f(r) &= \frac{1}{Z_f(r)} \int_{[-f\pi, f\pi]^2} V(r,\theta,\phi) e^{-V(r,\theta,\phi)/k_B T}\, d\theta\, d\phi \\ G_f(r) &= -k_B T\, \log Z_f(r) \\ S_f(r) &= \frac{H_f(r) - G_f(r)}{T}, \end{aligned} \tag{10}$$

where f denotes the fraction of the domain of integration: $0 < f \leq 1$. Further, differences are defined via

$$\Delta\alpha_f(r) = \alpha_f(r) - \lim_{r' \to \infty} \alpha_f(r') \text{ where } \alpha = G, H, S. \tag{11}$$

In Figure 8 we show the free energy, average potential energy, and entropy differences as a function of f for the case $a = b = 1$, T =303K, and for two values of r. For $r = 3$, we see that for f greater than 0.15, the resulting values are unchanged. Thus the behavior of the potential for extreme values of the configuration parameters is not influential.

One interpretation of the results in Figure 8(a) is that we could restrict the angular variables in the model to be in the range ±30 degrees, and the results would be unchanged. Thus the primary affects observed here depend only on small angular fluctuations, but the fluctuations (dynamics) are essential. Of course, the behavior near the separation distances corresponding to maximum entropy in Figure 6(b) (see the contour plots of the corresponding potentials in Figure 10) require a more careful analysis to clarify the effects of angular limitations, as shown in Figure 8(b). But if we choose f between about 0.2 and 0.6, we get consistent values for the thermodynamic variables.

We can visualize changes in the thermodynamic functionals in (10) by plotting the potential energy directly. In Figure 9, we see the contour plots for the Boltzmann distribution

$$\frac{1}{Z(r)} e^{-V(r,\theta,\phi)/k_B T} \tag{12}$$

for six different values of r. For very large r, we obtain the Boltzmann distribution for a dipole-dipole interaction, as indicated in Figure 9(i). For $r = 5$, this distribution is changed very little, the main difference being a small translation of the maximal point for the distribution, as indicated in Figure 9(ii). But for $r = 3.4$, a deformation in the distribution appears, as indicated in Figure 9(iii).

In the second line of figures (iv–vi) in Figure 9, we see the contour plots for the Boltzmann distribution (12) for smaller values of r. Note that the figures in Figure 9 have been automatically scaled, so there are differences in scale for the first group (i–iii) and the second group (iv–vi). For $r = 3.2$, depicted in Figure 9(iv), another local maximum emerges in the distribution, although it should be remembered that the height of this is 10^{-9} smaller than the main peak. As r decreases, the bulge in the distribution becomes more marked, as seen in Figure 9(v–vi). Similarly, Figure 10(i), depicts the same configuration as in Figure 9 but with the smaller value $r = 2.1$. Although it is useful to see the changes in the Boltzmann distribution (12) as a function of r, the eye has difficulty telling whether entropy and enthalpy are increasing or decreasing as a function of r without carrying out the integration.

The three panels in Figure 10 depict the potential energy distributions corresponding to the pairs of values a, r near the entropy maxima depicted in Figure 6(b). We see that the extreme entropy values correspond to radical changes in the potential energy in which extensive energy is found far away from the configurations for which the model was designed.

8 Asymptotic dependence of disruption

In Figure 11, we plot the (negative of) free energy and enthalpy differences as a function of r at a temperature of $303K$, where the energy unit is kcal/mole. The partial charges in the upper dipoles were all chosen to be ± 1 the charge of an electron, but the charges in the lower group are allowed to vary. The distances in the figure are in Ångstroms.

We see in Figure 11 the decrease in both enthalpy and free energy as a function of dipole system separation. For the case $a = b = 1$, the decay rate is r^{-6}, consistent with other dipole systems which become perturbed by another dipole. But for the case $a = 1.5$ and $b = 1$, the decay rate is r^{-4}. Note that both enthalpy and free energy tend to a finite limit as $r \to 0$.

Figure 11 indicates that the metastable dipole orientation has a significant change in free energy at a distance of 3–4 Å, and it also shows that the effect is fairly short range, having little residual effect beyond about 5 Å.

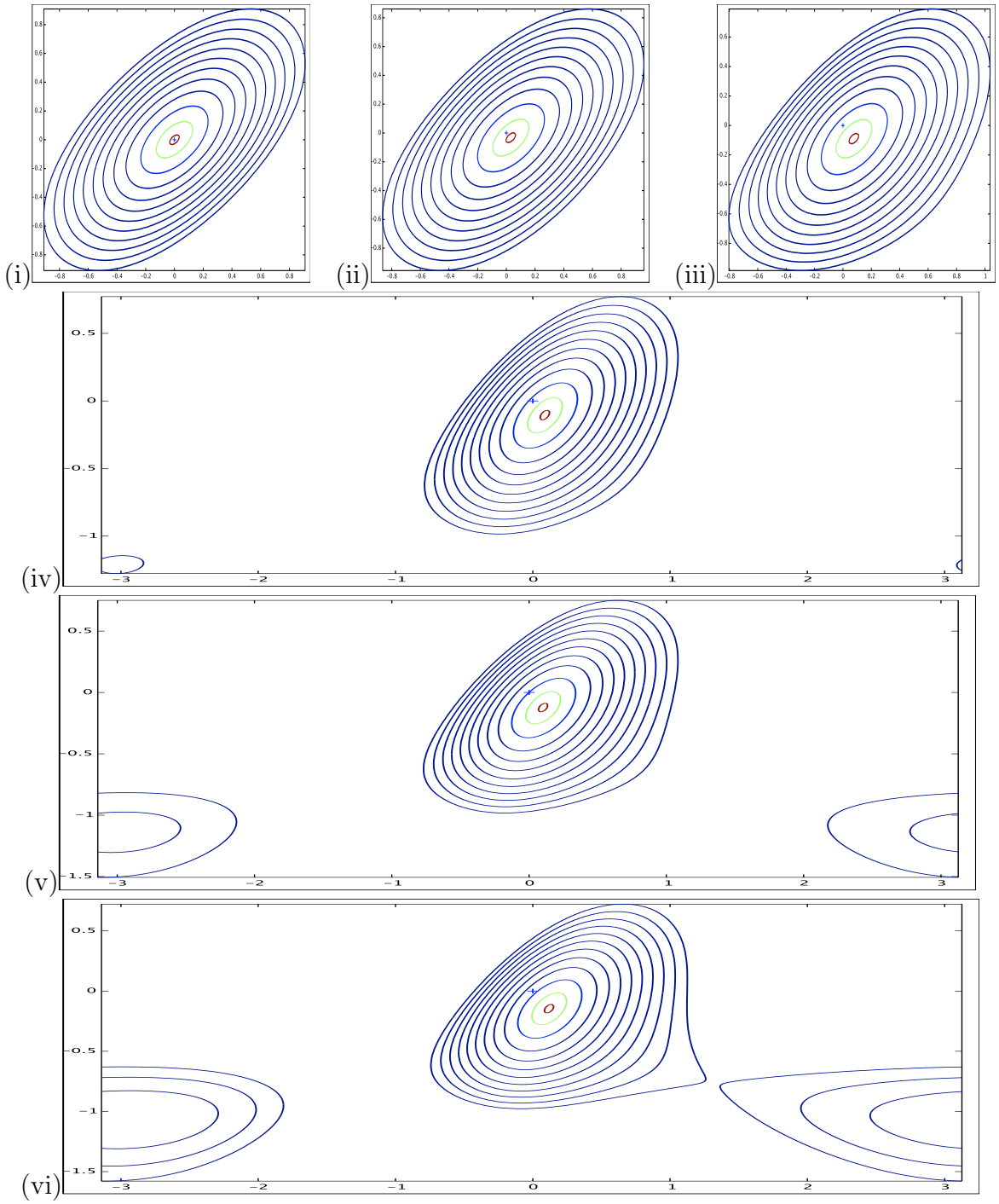


Figure 9: Potential energy surface for (i) $r = 10^5$, (ii) $r = 5$, (iii) $r = 3.4$, (iv) $r = 3.2$, (v) $r = 3$ and (vi) $r = 2.8$, with $a = b = 1$, T =303K. The expression (12) is plotted as a function of θ, ϕ in radians. Contours are at $0.95m$, $0.5m$, and $10^{-i}m$, for $i = 1, \ldots, 9$, where m is the maximum value of the distribution. The small '+' indicates the point $\phi = \theta = 0$ for reference.

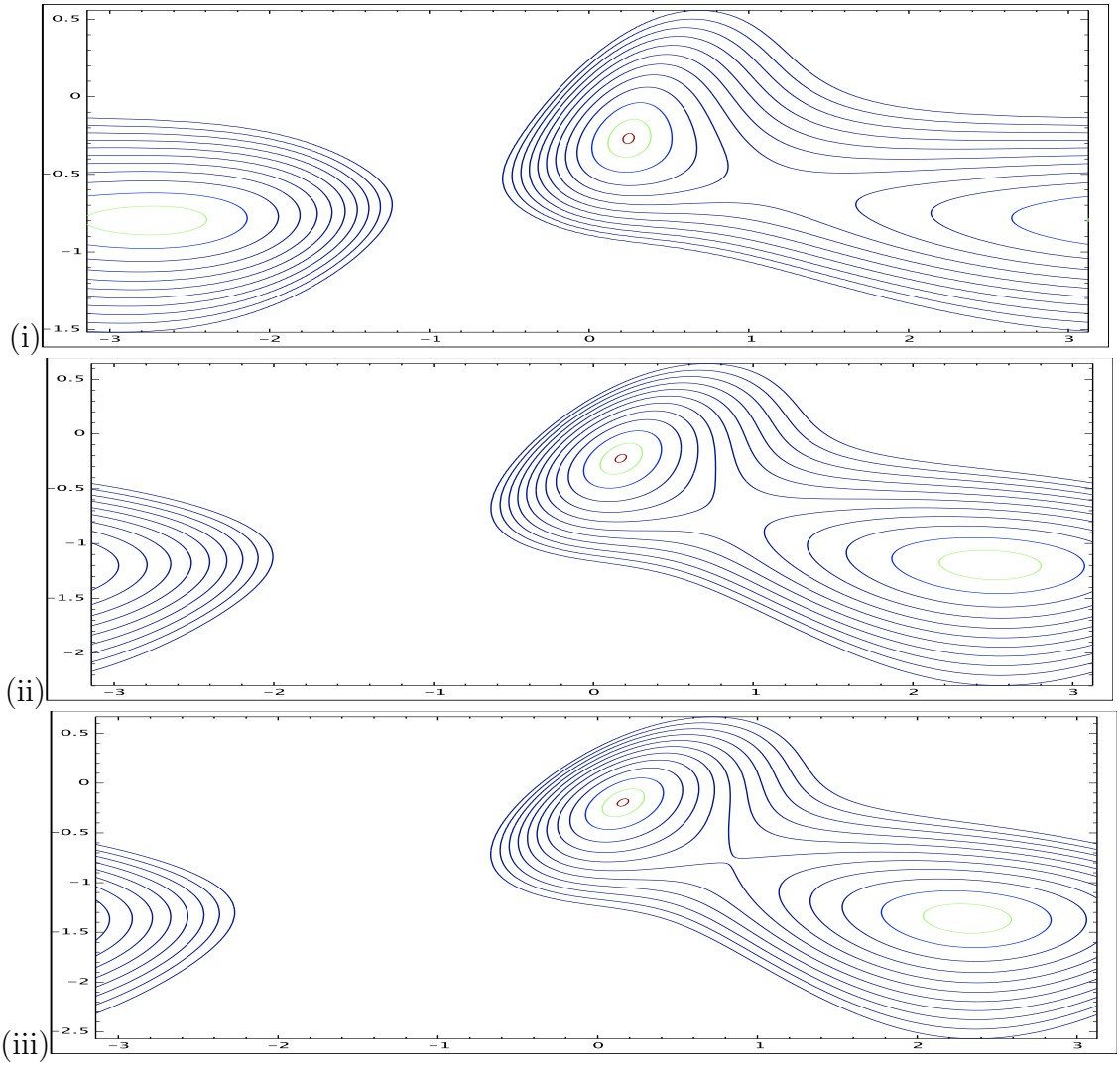


Figure 10: Potential energy surface for (i) $a = 1$ and $r = 2.1$, (ii) $a = 1.5$ and $r = 3.56$ and (iii) $a = 2$ and $r = 4.687$, with $b = 1$, T =303K. The values of r correspond to the point of entropy maximum. The expression (12) is plotted as a function of θ, ϕ in radians. Contours are at $0.95m$, $0.5m$, and $10^{-i}m$, for $i = 1, \ldots, 9$, where m is the maximum value of the distribution.

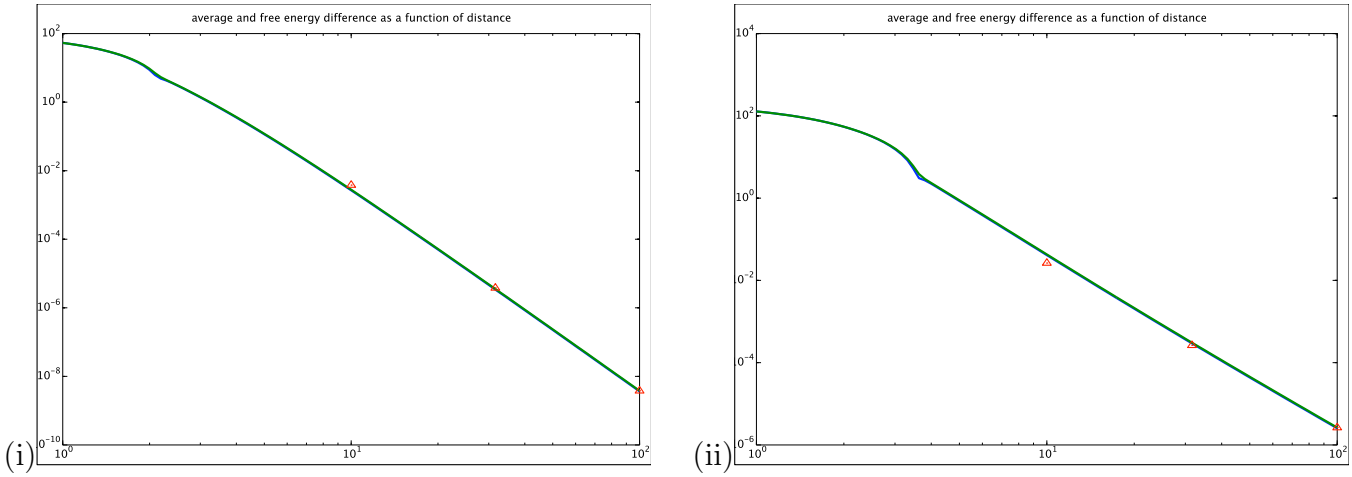


Figure 11: Negative of enthalpy and free energy differences (vertical axis) as a function of distance of separation r (horizontal axis). (i) $a = b = 1$; triangles indicate decay rate of r^{-6}. (ii) $a = 1.5$ and $b = 1$; triangles indicate decay rate of r^{-4}. Units are kcal/mole. T =303K.

9 Methods

The free energy and enthalpy calculations were done on a uniform $n \times n$ mesh using the midpoint rule for numerical integration. Due to the periodicity of the integrands, such rules are of arbitrary order similar to the Euler-Maclaurin formula (see Exercise 13.17 in [24]). Most computations were fully converged with $n = 30$, but contour lines were plotted with $n = 200$ to eliminate wiggles. Similarly, computation of the angular dependence of the point of minimum potential energy depicted in Figure 7 required $n = 360$ together with a quadratic fitting algorithm to estimate the position of the minimum from the grid data. The programs were executed in octave. To estimate the limiting values for $r' \to \infty$, a value of $r' = 10^5$ was used.

The character of the interaction potential for dipoles can be deduced from the well known formula

$$V(\mathbf{r}, \mathbf{u}, \mathbf{v}) \approx \frac{4\mathbf{u} \cdot \mathbf{v} - 12(\tilde{\mathbf{r}} \cdot \mathbf{u})(\tilde{\mathbf{r}} \cdot \mathbf{v})}{r^3} + \mathcal{O}(r^{-5}), \tag{13}$$

where $\mathbf{u}$ and $\mathbf{v}$ denote unit vectors representing any two dipoles, $\mathbf{r}$ is the vector from one dipole to the other, and $\tilde{\mathbf{r}} = (1/r)\mathbf{r}$ is the unit vector in the direction of $\mathbf{r}$. The potential energy of the dipole system in Figure 2(b) can be determined by two applications of (13), and it is negative since each of the top two dipoles is in the opposite direction from the one below. The terms $\tilde{\mathbf{r}} \cdot \mathbf{u}$ and $\tilde{\mathbf{r}} \cdot \mathbf{v}$ are both of size $\mathcal{O}(1/r)$ since $\mathbf{r}$ is nearly orthogonal to the dipoles. The contribution to the potential energy from the top two dipoles is also negative as is easily verified.

A Python code was developed that parses PDB files and determines the atoms closest to a specified atom, rank-ordered by distance. This was applied with the specified atom being each of the halogens in the various drug compounds studied. A cut-off radius can be specified to limit the output to a readable amount.

In addition, the software assesses polar interactions within a cut-off radius, as follows. Each of the halogens studied is covalently bonded to a carbon, and the locations of these two atoms defines a polarity vector. All amide (NH) and carbonyl (CO) pairs in the protein backbone were examined and their polarity vectors were compared with those of the halogens.

All PDB files were preprocessed to add hydrogens via the MolProbity web site (i.e., using the

program `reduce`).

Hydrogen bond quality data and the amount of wrapping for each bond was determined by the code `wrappa` [6].

10 Conclusions

The interfaces for protein associations are not well understood from a physical point of view. This is a major impediment to the rational design of drugs of therapeutic value. This is partly because the structural analysis of target–drug complexes is limited to interpreting a single microstate often identified by X-ray diffraction. In this work we identified an interactive motif that can only be understood by an ensemble-level dynamic analysis.

A direct interpretation of a crystal structure, which represents only one microstate of the complex, can underestimate the constructive role of certain unfavorable interactions between polar constituents. We have described such a type of three-body intermolecular interaction, involving a preformed hydrogen bond in the targeted protein and a permanent dipole in the ligand. The interaction is disruptive at the microstate level, but when examined in a statistical ensemble context it proves to be an enhancer of the affinity of the ligand for its binding partner. The role of such interactions in protein systems could be quite widespread.

The results of this paper suggest that this three-body motif could be factored into drug design. This would imply that drugs could be optimized in a larger space than previously disclosed. These observations herald the advent of an improved methodology that incorporates dynamic information into drug design. The reverse engineering of the interactive context of approved cancer drugs has revealed a design principle that exploits unconventional features of the drug–target association. These features become apparent only in a dynamics-based study of three-body interactions, as suggested in this contribution. Thus a novel paradigm emerges as the results of diverse drug-discovery efforts are critically contrasted using the concepts presented in this work.

Hopefully, the work described in this contribution will inspire future efforts to identify the conformational ensemble of a drug-target complex using NMR techniques [25]. Such studies are likely to provide valuable means of validating the disruption concept.

A major challenge will be to identify hydrogen bonds capable of being exogenously disrupted. We could call these design features “disruptons” by analogy with the term “dehydrons.” In forthcoming work, we hope to explore these directions.

11 Acknowledgments

The research of A.F.S. was partially supported by the National Research Council of Argentina (CONICET) and the Institute of Biophysical Dynamics and the Computer Science Department at The University of Chicago. L.R.S. acknowledges partial support from NSF grants DMS-0920960 and DMS-1226019.

References

[1] Ariel Fernández, Christopher M. Fraser, and L. Ridgway Scott. Purposely engineered drug-target mismatches for entropy-based drug optimization. *Trends in Biotechnology*, 30(1):1–7, 2012.

[2] Ariel Fernández and L. Ridgway Scott. Dehydron: a structurally encoded signal for protein interaction. *Biophysical Journal*, 85:1914–1928, 2003.

[3] Ariel Fernández and L. Ridgway Scott. Adherence of packing defects in soluble proteins. *Physical Review Letters*, 91:18102(4), 2003.

[4] Ariel Fernández. Incomplete protein packing as a selectivity filter in drug design. *Structure*, 13:1829–1836, 2005.

[5] Ariel Fernández and L. Ridgway Scott. Modulating drug impact by wrapping target proteins. *Expert Opinion on Drug Discovery*, 2:249–259, 2007.

[6] Christopher M. Fraser, Ariel Fernández, and L. Ridgway Scott. Dehydron analysis: quantifying the effect of hydrophobic groups on the strength and stability of hydrogen bonds. In Hamid R. Arabnia, editor, *Advances in Computational Biology*, New York, 2010. Springer.

[7] G.M. Whitesides and V.M. Krishnamurthy. Designing ligands to bind proteins. *Quarterly Reviews of Biophysics*, 38(04):385–395, 2005.

[8] Ernesto Freire. The propagation of binding interactions to remote sites in proteins: Analysis of the binding of the monoclonal antibody D1.3 to lysozyme. *Proceedings of the National Academy of Sciences, USA*, 96:10118–10122, 1999.

[9] M.J. Stone. NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding. *Accounts of Chemical Research*, 34(5):379–388, 2001.

[10] P. Sörme, P. Arnoux, B. Kahl-Knutsson, H. Leffler, J.M. Rini, and U.J. Nilsson. Structural and thermodynamic studies on cation-π interactions in lectin-ligand complexes: high-affinity galectin-3 inhibitors through fine-tuning of an arginine-arene interaction. *Journal of the American Chemical Society*, 127(6):1737–1743, 2005.

[11] Alejandro Crespo and Ariel Fernández. Induced disorder in protein–ligand complexes as a drug-design strategy. *Molecular Pharmaceutics*, 5(3):430–437, 2008.

[12] I.H. Park and C. Li. Characterization of molecular recognition of STAT3 SH2 domain inhibitors through molecular simulation. *Journal of Molecular Recognition*, 24:254265, 2011.

[13] K. K. Frederick, M. S. Marlow, K. G. Valentine, and A. J. Wand. Conformational entropy in molecular recognition by proteins. *Nature*, 448(7151):325–329, 2007.

[14] Ariel Fernández. *Transformative concepts for drug design. Target wrapping*. Springer, Berlin, 2010.

[15] R. Lumry and S. Rajender. Enthalpy–entropy compensation phenomena in water solutions of proteins and small molecules: A ubiquitous property of water. *Biopolymers*, 9(10):1125–1227, 1970.

[16] Dudley H. Williams, Elaine Stephens, Dominic P. O'Brien, and Min Zhou. Understanding noncovalent interactions: Ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes. *Angewandte Chemie International Edition*, 43(48):6596–6616, 2004.

[17] Charles H. Reynolds and M. Katharine Holloway. Thermodynamics of ligand binding and efficiency. *ACS Medicinal Chemistry Letters*, 2(6):433–437, 2011.

[18] E. N. Baker and R. E. Hubbard. Hydrogen bonding in globular proteins. *Progress in Biophysics and Molecular Biology*, 44:97–179, 1984.

[19] T. Lelièvre, G. Stoltz, and M. Rousset. *Free energy computations: A mathematical perspective*. Imperial College Press, London, 2010.

[20] Roger Balian. *From Microphysics to Macrophysics*, volume 1. Springer, Berlin, 1991.

[21] Christopher M. Fraser, Ariel Fernández, and L. Ridgway Scott. Wrappa: A screening tool for candidate dehydron identification. Research Report UC/CS TR-2011-5, Dept. Comp. Sci., Univ. Chicago, 2011.

[22] V. Balamurugan, Maninder Singh Hundal, and Rabindranath Mukherjee. First systematic investigation of C—H$\cdots$Cl hydrogen bonding using inorganic supramolecular synthons: Lamellar, stitched stair-case, linked-ladder, and helical structures. *Chemistry – A European Journal*, 10(7):1683–1690, 2004.

[23] A. Fernández, A. Sanguino, Z. Peng, E. Ozturk, J. Chen, A. Crespo, S. Wulf, A. Shavrin, C. Qin, J. Ma, et al. An anticancer C-Kit kinase inhibitor is reengineered to make it more active and less cardiotoxic. *Journal of Clinical Investigation*, 117(12):4044, 2007.

[24] L. Ridgway Scott. *Numerical Analysis*. Princeton Univ. Press, Princeton, NJ, 2011.

[25] Kresten Lindorff-Larsen, Robert B Best, Mark A DePristo, Christopher M Dobson, and Michele Vendruscolo. Simultaneous determination of protein structure and dynamics. *Nature*, 433(7022):128–132, 2005.