Optimization of Electrostatic Binding Free Energy: Application to Barnase and Barstar

by
Lee-Peng Lee
B. Sc. Massachusetts Institute of Technology
(1993)

Submitted to the Department of Physics
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
at the
Massachusetts Institute of Technology
June 1999

Signiture of Author _____
Department of Physics
May 21, 1999

Certified by _____
Bruce Tidor
Assistant Professor of Chemistry
Thesis Supervisor

Certified by _____
George B. Benedek
Professor of Physics
Thesis Co-supervisor

Accepted by _____
Thomas J. Greytak
Professor of Physics
Associate Department Head for Education

To

Whay

Acknowledgement

This thesis owes its existence first and foremost to Bruce Tidor, who conceived the idea of electrostatic optimization and provided invaluable guidance throughout the course of this work. I still remember Bruce sketching out the question on the blackboard during the first group meeting I attended. I am grateful that he had chosen me to work on the problem. His dedication to science and his many insights have a profound influence on my eduaction and my research. His support of my work has been steadfast and especially felt during the past fourteen months after I gave birth. Bruce and his wife, Young-Mi, have been extremely generous and kind towards my family, and we appreciate it deeply.

I am indebted to the members of my lab for providing me a pleasant and conducive environment for doing research and their help in many things, from stimulating discussions to computer scripts. I want to especially acknowledge Erik Kangas, Zak Hendsch, Justin Caravella, Karl Hanf, David Green, and two former members, Sara Dempster and Cristina Jarque-Uribe.

Michel Baranger, my undergraduate thesis supervisor and my research advisor for the first two years of my graduate studies, is a teacher and a constant advocate of me through thick and thin. I thank him and and his wife Mary Lee for their wonderful friendship.

I am thankful to George Benedek, whose biophysics course has taught me a great deal and who took time to be my thesis co-supervisor. I thank Mehran Kardar for being a reader for this thesis and for the statistical mechanics class he taught.

I want to thank Ying Tan and Hua Yang for their friendship since the first day we entered the Physics graduate program together.

Peggy Berkowitz and Pat Solakoff in the Physics education office have the most postive attitudes towards things and I truly value what they have done for me.

I want to thank Peter Kim, whose encouragement on a couple of occassions has meant a lot to me even though I only know him a little.

My husband and I are blessed with many friends whom we can count on. We want to

especially thank Samuel and Flora Cheng, Wen-Yi and Chien-Ho Tseng, Seng-Hock and Flora Cheah, and Joanne Larrabee.

I am grateful to Marie Chuang for her prayers and her example as a woman of faith.

I thank my parents, my sisters and my brother for their constant love and support.

Finally, I thank my husband, Whay Chiou Lee, for all the love he showers on me, and our fourteen-month-old son, Andrew, who gives us much joy and much to look forward to everyday (not to mention the work).

In all these, I have experienced God's faithfulness and His unfailing love.

When I consider thy heavens, the work of thy fingers,
the moon and the stars, which thou hast ordained;
What is man that thou art mindful of him?
the son of man that thou visitest him?

Psalms 8:3,4

Optimization of Electrostatic Binding Free Energy: Application to Barnase and Barstar

by

Lee-Peng Lee

Submitted to the Department of Physics on May 21, 1999
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Physics

Abstract

A computational method was developed within the continuum electostatic model to optimize the electrostatic interactions in the noncovalent binding of biomolecules in aqueous environment. In a noncovalent association event involving two molecules, the electrostatic optimization method finds for one molecule (the ligand) the optimal charge distribution that minimizes the electrostatic binding free energy when it binds to its partner (the receptor), whose charge distribution is given and fixed. The ligand-charge distribution is contructed to explicitly optimize the balance between unfavorable ligand-desolvation penalty, which is quadratic in ligand charges, and favorable ligand–receptor interaction, which is linear in ligand charges. This method was applied to the tight-binding barnase–barstar protein complex, treating barstar as the ligand. Results reveal the predominance of a relatively few barstar residues at the interface and show that, while sets of optimal barstar charges beyond what is available with amino acids could be found that gave rise to improvement of 10 to 20 kcal/mol in electrostatic binding free energy, the wild type barstar is remarkably well optimized given that the barstar residues are limited to the 20 common amino acids. Binding electrostatics was found to be enhanced both by direct improvement of hydrogen-bond and ion-pair interactions across the binding interface and indirect intramolecular interaction, which decreased the desolvation penalty. Reverse optimization, treating barnase as the ligand, showed that barnase is less optimized electrostatically than barstar for binding, consistent with the perceived functions of these proteins, i.e. barstar as an inhibitor and barnase as an enzyme. A direct by-product of the electrostatic optimization theory is a procedure for illustrating and evaluating electrostatic complementarity of the binding partners, which is demostrated here.

Thesis Supervisor: Bruce Tidor
Title: Assistant Professor of Chemistry

Table of Contents

Chapter I

General Introduction

This thesis describes a computational method of optimizing the electrostatic interactions in the noncovalent association of biomolecules in aqueous environment and its application to the tight-binding barnase–barstar protein complex. Noncovalent binding or association of biomolecules is an important mechanism underlying many biological events such as the action of hormones, DNA transcription, immune system recognition, enzyme catalysis, and the actions of many drugs. While the stronger covalent bonds hold macromolecules in place and intact, it is the much weaker noncovalent interactions or forces which are responsible for the dynamic association and dissociation of biomolecules that give rise to the complex biological processes in living things. Accurate and efficient theoretical and computational methods of analyzing and improving binding will help extract an understanding of principles for tight-binding and will be of practical value in molecular design of drugs or enzymes.

Molecules interact noncovalently in a number of different ways, although all noncovalent forces are fundamentally electrostatic, i.e. they originate from the forces due to the electronic charges and structures of the atoms or molecules.* Rather than lumping all these different types of interactions under one big electrostatic interaction term, biophysicists and biochemists distinguish them from each other for deeper insight in the nature of these interactions. Here electrostatic interactions refer to all screened coulombic interactions (screened

*There are four and only four types of forces in nature: gravitation, electromagnetic, weak and strong.

by the ionic solvent) between the partial atomic charges, including interactions between ionic groups and dipoles, such as salt-bridges and hydrogen bonds, and longer-range interactions. The non-electrostatic terms that enter the energetics of noncovalent binding include van der Waals (or dispersion) forces, the hydrophobic effect, and the various conformational entropic effects (translational, rotational and vibrational).

The magnitudes of the different types of noncovalent bonds have been estimated and can be found in various textbooks.[1] Van der Waals forces drive shape complementarity between two molecules by having each pair of proximate intermolecular atoms contributing up to –0.5 kcal/mol. This relatively modest contribution can add up to a significant number when summed over the contacting atoms of two biomolecules. For instance, the van der Waals energy has been calculated to be –14 kcal/mol between the D subsite of lysozyme and the occupying glucopyranose ring.[2] Hydrophobic effects result from the regaining of entropy when water molecules are freed from ordering themselves around a hydrophobic compound when it is driven into a hydrophobic region. Hydrophobicity has been found to correlate with the surface area of the macromolecule, with an empirical value of 20 to 25 cal/mol per $Å^2$[3] for side chains of hydrophobic amino acids. For typical protein–protein complex with 1000 $Å^2$ or more of buried surface area, that translates into tens of kcal/mol. The conformational entropies are important in binding, too. In the dimerization of insulin, the translational and rotational entropies were calculated to be 13 and 14 kcal/mol while the vibrational was –7 kcal/mol.[4] Finally, each salt bridge or hydrogen bond has been estimated to contribute up to 10 kcal/mol.

Because salt bridges and hydrogen bonds have large free energy contributions and are formed across the binding interface in protein–protein or other complexes, the electrostatic contribution to the free energy of binding was considered favorable and important. However, recent experiments and calculations show that electrostatic interactions actually destabilize binding in most cases.[5–16] The favorable interactions in the bound structure are more than offset by the unfavorable desolvation penalties (i.e. loss of interactions with solvent

and solvent ions) of the polar groups. For instance, the tightly bound barnase–barstar protein complex has a binding constant of 10^{-14} M (or a standard binding free energy of –19 kcal/mol),[17] but the electrostatic contribution is computed to be 14 kcal/mol.[18] The complex forms because the hydrophobic effect is strongly stabilizing. This is only one of the many complexes where the electrostatic contribution is computed to be unfavorable and often large. There is now a growing consensus that electrostatic interactions do not contribute strongly to affinity but are likely to enhance specificity.[5,19–21]

The unfavorable binding electrostatics that occurs in nature leads to the question of whether electrostatic interactions are optimized? Is unfavorable electrostatics a consequence of the governing physical laws, a result of the constraints in the biological world, or simply a choice of the biological systems? Whereas optimization of non-electrostatic interactions is straightforward, the same cannot be said about electrostatic interactions. Optimizing van der Waals effects is very well defined by shape complementarity, while burying hydrophobic regions accounts for optimizing hydrophobicity and having a rigid structure minimizes entropic loss. The electrostatic effects are subtle because of the two competing forces involved in binding: favorable intermolecular interaction and unfavorable desolvation penalties. For instance, increasing the charge at the binding site enhances both its interaction with an opposite charge in the binding partner (tending to favor binding) and its interaction with solvent in the unbound state (tending to disfavor binding).

The key contribution of this thesis is to frame the problem of electrostatic optimization in a manner that leads to a well-defined, computable electrostatic optimum. We have adopted a widely-used continuum electrostatic model for electrostatic free energy calculation and developed a method to optimize the electrostatic interactions in binding. Specifically, in a binding event involving two partners (termed receptor and ligand), our optimization scheme finds the optimal charge distribution for one molecule (the ligand), which minimizes the electrostatic binding free energy when it binds to the other molecule (the receptor) whose charge distribution is given and fixed. The ligand-charge distribution is constructed

to explicitly optimize the balance between unfavorable ligand-desolvation penalty (which is quadratic in ligand charge) and favorable interactions (which is linear in ligand charge) made with a fixed receptor.

This method has been applied to the barnase–barstar complex. We have investigated the properties of the electrostatic optimal ligands to examine the importance that natural proteins place on this and learn general principles for enhancing binding. These lessons and the ligands produced by the optimization method will also be useful in the design of inhibitor drugs. One mechanism operating in many diseases is the undesirable action of a receptor (often a protein) that can be arrested, at least in principle, through the tight binding of a molecular ligand (e.g. by sterically blocking the active site or by preventing a required conformational change). While an effective drug must pass an elaborate array of pharmacological tests for bioavailability and non-toxicity among others, the identification or design of tight-binding ligands is an important step in the discovery of drug molecules. The global electrostatic optimization scheme presented here to find the optimal ligand is very different from the usual ligand design approaches, which tend to be perturbative, i.e. making small improvements instead of locating directly a global optimum.

Summary of the present work

There are two parts to the work described in this thesis: the development of a computational method to optimize binding electrostatics within the continuum electrostatic model and its application to study the tight-binding barnase–barstar protein complex. They are discussed in four self-contained papers (Chapters III through V, and Appendix). Chapter II gives an overview to the theoretical and computational methods employed. The electrostatic optimization method was first formulated in a spherical model of protein binding, where the complex and the ligand are spheres (Chapter III), and applied to a spherical model of the tight-binding barnase-barstar protein complex (Appendix, done in collaboration with Chong *et al.*). This optimization scheme was extended to treat binding using actual molecular shapes and was applied to study the charge complementarity in the barnase–barstar complex (Chapters IV and V). The results show that sets of barstar charges can be found by electrostatic optimization that give rise to improvements of 10 to 20 kcal/mol in electrostatic binding free energy (Chapter IV), but we also found that given the barstar residues are limited to the 20 natural amino acids, the wild type is actually very well optimized (Chapter V). Principles for enhancing electrostatic interactions were extracted from these studies (Chapters IV and V, and Appendix).

REFERENCES

[1] A. R. Fersht. *Enzyme Structure and Mechanism.* W. H. Freeman and Company, New York, second edition (1985).

[2] A. Warhel and M. Levitt. Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **103**: 227–249 (1976).

[3] C. Chothia. Hydrophobic bonding and accessible surface area in proteins. *Nature (London)* **248**: 338–339 (1974).

[4] B. Tidor and M. Karplus. The contribution of vibrational entropy to molecular association. The dimerization of insulin. *J. Mol. Biol.* **238**: 405–414 (1994).

[5] Z. S. Hendsch and B. Tidor. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**: 211–226 (1994).

[6] C. D. Waldburger, J. F. Schildbach, and R. T. Sauer. Are buried salt bridges important for protein stability and conformational specificity? *Nature Struct. Biol.* **2**: 122–128 (1995).

[7] W. C. Wimley, K. Gawrisch, T. P. Creamer, and S. H. White. Direct measurement of salt-bridge solvation energies using a peptide model system: Implications for protein stability. *Proc. Natl. Acad. Sci. U.S.A.* **93**: 2985–2990 (1996).

[8] A.-S. Yang and B. Honig. Free energy determinants of secondary structure formation: I. α-helices. *J. Mol. Biol.* **252**: 351–365 (1995).

[9] L. Wang, T. O'Connell, A. Tropsha, and J. Hermans. Energetic decomposition of the α-helix–coil equilibrium of a dynamic model system. *Biopolymers* **39**: 479–489 (1996).

[10] V. K. Misra, K. A. Sharp, R. A. Friedman, and B. Honig. Salt effects on ligand–DNA binding: Minor groove binding antibiotics. *J. Mol. Biol.* **238**: 245–263 (1994).

[11] V. K. Misra, J. L. Hecht, K. A. Sharp, R. A. Friedman, and B. Honig. Salt effects on protein–DNA interactions: The λcI repressor and EcoRI endonuclease. *J. Mol. Biol.* **238**:

264–280 (1994).

[12] K. Sharp. Electrostatic interactions in hirudin–thrombin binding. *Biophys. Chem.* **61**: 37–49 (1996).

[13] J. Shen and J. Wendoloski. Electrostatic binding energy calculation using the finite difference solution to the linearized Poisson–Boltzmann equation: Assessment of its accuracy. *J. Comput. Chem.* **17**: 350–357 (1996).

[14] R. E. Bruccoleri, J. Novotny, M. E. Davis, and K. A. Sharp. Finite difference Poisson–Boltzmann electrostatic calculations: Increased accuracy achieved by harmonic dielectric smoothing and charge antialiasing. *J. Comput. Chem.* **18**: 268–276 (1997).

[15] J. Novotny, R. E. Bruccoleri, M. Davis, and K. A. Sharp. Empirical free energy calculations: A blind test and further improvements to the method. *J. Mol. Biol.* **268**: 401–411 (1997).

[16] Z. S. Hendsch and B. Tidor. Electrostatic interactions in the gcn4 leucine zipper: Substantial contributions arise from intramolecular interactions enhanced on binding. *Protein Sci.* **In press.** (1999).

[17] G. Schreiber and A. R. Fersht. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry* **32**: 5145–5150 (1993).

[18] L. T. Chong, Z. S. Hendsch, and B. Tidor. In preparation.

[19] C. Tanford, P. K. De, and V. G. Taggart. The role of the α-helix in the structure of proteins. Optical rotatory dispersion of β-lactoglobulin. *J. Am. Chem. Soc.* **82**: 6028–6034 (1960).

[20] C. H. Paul. Building models of globular protein molecules from their amino acid sequences. I. Theory. *J. Mol. Biol.* **155**: 53–62 (1982).

[21] C. V. Sindelar, Z. S. Hendsch, and B. Tidor. Effects of salt bridges on protein structure

and design. *Protein Sci.* **7**: 1898–1914 (1998).

Chapter II

General Methodology

Introduction

This chapter serves as a methodology primer to the rest of the thesis, where each self-contained chapter includes a discussion of the methods relevant and specific to the chapter. There are four sections in this chapter. First, we give an introduction to the continuum electrostatic model widely employed to study electrostatic interactions of biomolecules in ionic aqueous solutions. We next describe the finite difference algorithm underlying the computer package DELPHI[1–3] we used to solve the Poisson-Boltzmann equation and calculate electrostatic free energy. The original DELPHI implementation only allows the source-charge distribution be represented by point charges (monopoles). In the third section, we describe modifications we made to the program that allow treatment of higher-order multipoles as part of the source-charge distribution. Finally, we give an overview of the electrostatic optimization method.

Continuum electrostatics

The basic equation that encompasses all of electrostatics is simple. For a system of charges $\{Q_i\}$ located at $\{\vec{r}_i\}$, the potential at $\vec{r}$ is given by Coulomb's law,

$$\Phi(\vec{r}) = \sum_i \frac{Q_i}{|\vec{r} - \vec{r}_i|}. \tag{1}$$

However, application of this equation requires all charges be explicitly represented. A

biomolecule acts in an aqueous environment with dilute ions present. For a direct application of Coulomb's law, the charge distribution of the biomolecule, the water molecules, and the solvent ions have to be known. But what are these charges and how do we obtain them? At the most fundamental level, the answer is straight-forward to state: this, and all other questions about the physical world, can be answered by solving the Schrödinger equation. In practice, however, most problems are not solved that way. For a system more complicated than a two- or few-body system, solving the Schrödinger equation for the full system is neither computationally feasible nor necessary to provide a useful answer. Often, an approximate theory gives an accurate answer, much deeper insight, and more useful result.

The continuum electrostatic model for macromolecules in ionic solution, first formulated by Debye and Hückel,[4] has been shown to be a very good theoretical framework for its intended purpose. The biomolecule is treated as a low dielectric cavity defined by its molecular surface embedded in a high dielectric continuum solvent. Instead of explicitly representing the charges of the whole system which arise from polarization of all the atoms in the biomolecule and the solvent, only the charge distribution of the biomolecule (to be discussed in detail later) is explicitly represented. The distribution of the ions in the continuum solvent is derived with Boltzmann statistics. The continuum approach abstracts away details of lesser interest and captures the essence of the solvent response, ion reorganization, and the polarizability and flexibility of the biomolecule by treating the solvent and the solute as continuum dielectric media with different dielectric constants. The dielectric constant for the aqueous solvent is taken to be around 80, which is the experimental value for dielectric constant of water at room temperature. The dielectric constant assigned to the biomolecule is usually 2 (the high frequency dielectric constant for most organic liquids) or 4 (which we use in this work to account for the large-scale atomic and dipolar fluctuations expected for a protein compared to a small molecule.)[5]

In inhomogeneous media, the fundamental equation for electrostatics is[6]

$$\vec{\nabla} \cdot \vec{D}(\vec{r}) = 4\pi\rho_{\text{tot}}(\vec{r}). \tag{2}$$

Assuming linear media where

$$\vec{D}(\vec{r}) = \epsilon(\vec{r})\vec{E}(\vec{r}) \tag{3}$$

or

$$\vec{D}(\vec{r}) = -\epsilon(\vec{r})\vec{\nabla}\Phi(\vec{r}) \tag{4}$$

we arrive at the following Poisson's equation in inhomogeneous media,

$$\vec{\nabla} \cdot \left(\epsilon(\vec{r})\vec{\nabla}\Phi(\vec{r})\right) = -4\pi\rho_{\text{tot}}(\vec{r}). \tag{5}$$

In free space, where $\epsilon(\vec{r})$ is 1 everywhere and the source-charge distribution is a set of point charges, the solution to Poisson's equation is just Coulomb's law (Eq. (1)).

In the presence of ions in the continuum solvent, the source-charge distribution in Eq. (5), $\rho_{\text{tot}}(\vec{r})$, consists of two parts: the biomolecule charges, denoted by $\rho(\vec{r})$, and the solvent ion charge distribution, $\rho_{\text{ion}}(\vec{r})$. The mobile solvent ions are treated with Boltzmann statistics (see, for example, ref. 7). The concentration of the ith ion species in the presence of an electric potential $\Phi(\vec{r})$ is Boltzmann-weighted by the electrostatic energy,

$$c_i(\vec{r}) = c_i(\infty)e^{\frac{-z_i e\Phi(\vec{r})}{kT}}, \tag{6}$$

where $c_i(\infty)$ is the bulk concentration and $z_i e$ is the charge of the i-th ion species, k is the Boltzmann constant and T is the absolute temperature. The ion charge distribution is

$$\rho_{\text{ion}}(\vec{r}) = \sum_i z_i e c_i(\vec{r}) \tag{7}$$

$$= \sum_i z_i e c_i(\infty)e^{\frac{-z_i e\Phi(\vec{r})}{kT}}. \tag{8}$$

The Poisson–Boltzmann (PB) equation is obtained by substituting the above equation into Poisson's equation (Eq.(5)),

$$\vec{\nabla} \cdot \left(\epsilon(\vec{r})\vec{\nabla}\Phi(\vec{r})\right) + \sum_i z_i e c_i(\infty)e^{\frac{-z_i e\Phi(\vec{r})}{kT}} = -4\pi\rho(\vec{r}). \tag{9}$$

The linearization of the exponential terms (valid when the magnitude of $e\Phi$ is less than kT), together with electroneutrality ($\sum_i z_i e c_i(\infty) = 0$) results in the linearized Poisson–Boltzmann (LPB) equation,

$$\vec{\nabla} \cdot \left(\epsilon(\vec{r})\vec{\nabla}\Phi(\vec{r})\right) - \bar{\kappa}(\vec{r})^2\Phi(\vec{r}) = -4\pi\rho(\vec{r}), \tag{10}$$

where $\kappa = 8\pi e^2 I(\vec{r})/kT$ and the ionic strength, $I = \frac{1}{2}\sum_i z_i^2 c_i(\infty)$. The LPB equation has been shown to be an excellent approximation to the full PB equation in studies of biological systems.[8] From here on and in the rest of the thesis, only the LPB equation is used.

The LPB equation can be solved analytically only for a simple molecular boundary such as a sphere, infinite cylinder, or slab. Before the advent of modern computers that would facilitate numerical solution, proteins were modeled as spheres (pioneered by Tanford and Kirkwood[9,10]) and DNAs as cylinders so that an analytical solution could be obtained.[11] With fast computers and efficient algorithms, actual molecular shapes can now be used. There are a few definitions of molecular boundary. The simplest surface is the van der Waals surface, defined by the van der Waals radii of the atoms. The solvent-accessible surface is the locus of the center of a finite probe water molecule rolling over the van der Waals surface of the macromolecule. Another commonly used surface and the one used in our work is the molecular surface, which is the locus of the closest approach of the macromolecule to the probe water molecule.[12,13]

The charge distribution of the biomolecule is usually represented by a set of point partial charges $\{Q_i\}$ located at the atomic centers $\vec{r}_i$,

$$\rho(\vec{r}) = \sum_i Q_i \delta(\vec{r}_i). \tag{11}$$

The partial charges, together with the atomic radii, are obtained from various means of parameterization on smaller constituents of the macromolecule such as subunits of amino acid backbone and side chains. One or more of the following are used: *ab initio* quantum mechanical calculations of small molecules in gas phase, Monte Carlo or molecular simulations, and experimental studies of model compounds. Three parameter sets for the radii

and charges commonly used for proteins are CHARMM PARAM19,[14] PARSE,[15] and OPLS.[16] We used predominantly CHARMM PARAM19, but have also tested our results with the other two parameter sets (Chapter IV).

The potential is obtained by solving the LPB equation, and the electrostatic free energy is given by[1]

$$G = \frac{1}{2}\sum_i Q_i \Phi(\vec{r}_i) \ , \tag{12}$$

where $\Phi(\vec{r}_i)$ is the potential at the position of the charge Q_i. This free energy includes entropic contributions from the solvent, including ion reorganization in the case of the linearized Poisson-Boltzmann equation.[1,17,18] The factor of $\frac{1}{2}$ is due to the fact that the free energy arises from the interaction between the charges and their self-induced reaction field.

Solving the LPB equation with finite difference method

There are a number of different numerical algorithms for solving the linearized PB equation. The most popular approach is the finite difference method,[8] while finite element[19,20] and boundary element[21,22] methods have also been used. We perform our calculations using DELPHI,[1–3] a computer package based on the finite difference algorithm. The finite difference method[23] transforms the second-order differential PB equation into a set of finite difference equations on an equally spaced grid of mesh size h. The value of a continuous function $f(\vec{x})$ at the ith grid point is given by

$$f_i = \iiint f(\vec{x}) d^3x \ , \tag{13}$$

where the integration is over a cube of size h centered on the ith grid point. Integrating the linearized PB equation, Eq. (10), over the ith unit cubic volume gives

$$\iiint \vec{\nabla} \cdot \left(\epsilon(\vec{r})\vec{\nabla}\Phi(\vec{r})\right) d^3x - \iiint \bar{\kappa}(\vec{r})^2 \Phi(\vec{r}) d^3x = -4\pi \iiint \rho(\vec{r}) d^3x \ . \tag{14}$$

The first integral can be transformed to a surface integral using Gauss's theorem,

$$\iiint \vec{\nabla} \cdot \left(\epsilon(\vec{r})\vec{\nabla}\Phi(\vec{r})\right) d^3x = \iint \epsilon(\vec{r})\vec{\nabla}\Phi(\vec{r}) \cdot d\vec{A} \ , \tag{15}$$

where $d\vec{A}$ is the surface area normal vector. A forward finite difference formula for the gradient operator further gives

$$\iint \epsilon(\vec{r})\vec{\nabla}\Phi(\vec{r}) \cdot d\vec{A} = \sum_{j=1}^{6} \epsilon_{ij}(\phi_j - \phi_i)h \ , \tag{16}$$

where ϕ_i is the potential at the ith grid point and ϕ_j is the potential at the jth neighboring grid point (there are 6 nearest neighbors for each grid point). ϵ_{ij} is the dielectric constant at the center of the square face bounding the unit volume, i.e. in the middle of the grid line joining the ith and the jth grid points. The second integral in Eq. (14) is approximated by $\bar{\kappa}_i\phi_i h^3$ where $\bar{\kappa}_i$ is $\bar{\kappa}$ at the ith grid point. The RHS becomes $-4\pi q_i$ where q_i is the total fixed charge inside the volume of ith grid point. Eq. (14) becomes

$$h\sum_{j=1}^{6} \epsilon_{ij}(\phi_j - \phi_i) - h^3\bar{\kappa}_i\phi_i = -4\pi q_i \ , \tag{17}$$

which is the discrete version of the LPB equation. We can rearrange this equation to obtain

$$\phi_i = \frac{\sum_j \epsilon_{ij}\phi_j + 4\pi q_i/h}{\sum_j \epsilon_{ij} + \bar{\kappa}_i^2 h^2} \ . \tag{18}$$

The simplest way to solve this set of difference equations is by the Jacobian relaxation algorithm which makes an initial guess $\phi_i^{(0)}$ at all grid points and iterates with

$$\phi_i^{(n)} = \frac{\sum_j \epsilon_{ij}\phi_j^{(n-1)} + 4\pi q_i/h}{\sum_j \epsilon_{ij} + \bar{\kappa}_i^2 h^2} \tag{19}$$

until the difference between $\phi_i^{(n)}$ and $\phi_i^{(n-1)}$ is small. The current DELPHI implementation uses successive over-relaxation[24] to speed the convergence to a solution. The Jacobian relaxation (Eq. (19)) can be written in matrix form as

$$\vec{\phi}^{(n)} = \overleftrightarrow{T}\,\vec{\phi}^{(n-1)} + \vec{C} \ . \tag{20}$$

The successive over-relaxation iterates as follows,

$$\vec{\phi}^{(n)} = \omega\,\overleftrightarrow{T}\,\vec{\phi}^{(n-1)} + (1-\omega)\vec{\phi}^{(n-1)} + \omega\vec{C} \tag{21}$$

where ω is a relaxation parameter approximated by a few Jacobian iterations, as described in ref. 24. The DELPHI program also gives an estimate for the number of iterations needed for the solution to converge. We have found this estimate, which is linear in grid size in one dimension, to be good only for low grid spacing. To ensure convergence, we check at regular intervals if the energy convergence criteria are met.

It is necessary to make assumptions for the potential at the boundary of the grid. The usual boundary conditions are zero or Debye-Hückel.[23] These estimates are only accurate when the macromolecule is far from the edge, i.e. the grid is large relative to the molecule. However, for a fixed mesh size, this means low resolution for the dielectric boundary and the potential inside the macromolecule. To overcome this problem, a technique called "focusing" is used.[23] A 23% filled calculation is first done with the Debye-Hückel boundary condition to determine the potential at the grid boundary for a higher resolution calculation of 92% fill. Another fundamental numerical inaccuracy comes from the use of a discrete grid in place of continuous space. Ten translations relative to the grid are done to attain higher accuracy.[23] The potential calculated is the full potential, which includes a coulombic potential (smoothed by the numerical procedure) and a reaction-field term (due to the inhomogeneous dielectric and solvent ions). The smoothed coulombic potential gives rise to both the usual coulombic energy and a self-energy term. The spurious self-energy term can be dealt with by subtracting from the full potential the smoothed coulombic potential obtained by a reference state calculation where the solvent dielectric is set to the dielectric of the macromolecule and the ionic strength is set to zero. The coulombic energy can be added back in by applying coulomb's law. However, such reference state calculations are not needed in this work. We calculate binding free energy by subtracting the free energy of the unbound state from the bound state. Cancellation of the self-energy takes place when we use a rigid binding model (i.e. atom charges and positions are the same in the bound and unbound states). Once the potentials at the grid points are calculated, DELPHI uses a trilinear interpolation to find the potentials at the charge positions to obtain the

electrostatic free energy.

Multipoles in electrostatic free energy calculations

The finite difference formulation of the PB equation as described in the last section only allows the source-charge distribution to be represented by point charges. The difference equation (Eq. (17)) has a source charge function, $\rho(\vec{r})$, and a higher-order multipole cannot be expressed except by approximating the multipole by point charges. The ligand basis set we use in electrostatic optimization may include both point charges and higher-order multipoles, and the original implementation of DELPHI has to be modified to treat higher-order multipoles. Zhou it et al. have derived a source-free version of the PB difference equation by replacing the source-charge distribution with its coulombic potential.[25] This allows the treatment of dipoles and higher-order multipoles because their coulombic potential functions are readily computed.

The relationship between $\rho(\vec{r})$ and the coulombic potential $\phi_{\mathrm{c}}(\vec{r})$ is given by the Poisson equation,

$$\epsilon_{\mathrm{m}} \nabla^2 \phi_{\mathrm{c}}(\vec{r}) = -4\pi\rho(\vec{r}) \tag{22}$$

where ϵ_{m} is the dielectric of the macromolecule. Applying the finite difference method to the above equation as in Eq. (14) to Eq. (19), we have

$$\phi_{\mathrm{c},i} = \frac{1}{6}\left(\sum_j \phi_{\mathrm{c},j} + \frac{4\pi q_i}{\epsilon_{\mathrm{m}} h}\right) \tag{23}$$

Eq. (10) is now rewritten by substituting $4\pi q_i$ using Eq. (23),

$$\phi_i^{(n)} = \frac{\sum_j \epsilon_{ij}\phi_j^{(n-1)} + \epsilon_{\mathrm{m}}(6\phi_{\mathrm{c},i} - \sum_j \phi_{\mathrm{c},j})}{\sum_j \epsilon_j + \bar{\kappa}_i^2 h^2}. \tag{24}$$

The potential ϕ_i consists of two parts, the coulombic part $\phi_{\mathrm{c},i}$ and the reaction-field part $\phi_{\delta,i}$. $\phi_{\mathrm{c},i}$ can be calculated analytically by Coulomb's law. We only need to solve numerically for $\phi_{\delta,i}$. The finite difference equation for $\phi_{\delta,i}$ is derived as follows:

$$\phi_{\delta,i} = \phi_i^{(n)} - \phi_{\mathrm{c},i} \tag{25}$$

$$= \frac{\sum_j \epsilon_{ij}\phi_j^{(n-1)} + \epsilon_{\mathrm{m}}(6\phi_{\mathrm{c},i} - \sum_j \phi_{\mathrm{c},j})}{\sum_j \epsilon_j + \bar{\kappa}_i^2 h^2} - \phi_{\mathrm{c},i} \tag{26}$$

$$= \frac{\sum_j \epsilon_{ij}(\phi_{\delta,j}^{(n-1)} + \phi_{\mathrm{c},j}) + \epsilon_{\mathrm{m}}(6\phi_{\mathrm{c},i} - \sum_j \phi_{\mathrm{c},j})}{\sum_j \epsilon_j + \bar{\kappa}_i^2 h^2} - \phi_{\mathrm{c},i} \tag{27}$$

If we have a point-charge distribution, we can use Eq. (27) or Eq. (19) to solve the electrostatic problem, but for higher-order multipoles, only Eq. (19) can be used. We also need to generalize the electrostatic free energy expression for higher-order multipoles. Whereas the electrostatic free energy is $G = \frac{1}{2}\sum Q_i\Phi(\vec{r}_i)$ for a monopole, for the lth-order multipole in general we have[26]

$$G = \sum_{m=-l}^{l} \frac{4\pi}{(2l+1)!!} Q_{l,m}^* \mathcal{Y}_{l,m}(\vec{\nabla}) V(\vec{r}) \tag{28}$$

where

$$\mathcal{Y}_{l,m}(\vec{\nabla}) = \frac{(2l)!}{2^l l!}\left[\left(\frac{2l+1}{4\pi}\right)\frac{2^m}{(l+m)!(l-m)!}\right]^{\frac{1}{2}} \nabla_1^m \nabla_0^{l-m} \quad \text{for } m \geq 0. \tag{29}$$

The double-factorial is defined as

$$(2l+1)!! = (2l+1)\cdot(2l-1)\cdot(2l-3)\cdots 3\cdot 1 \tag{30}$$

$$= \frac{(2l+1)!}{2^l l!} \tag{31}$$

and the spherical partial derivatives are

$$\nabla_1 = -\frac{1}{\sqrt{2}}(\nabla_x + i\nabla_y), \qquad \nabla_{-1} = \frac{1}{\sqrt{2}}(\nabla_x - i\nabla_y), \qquad \nabla_0 = \nabla_z. \tag{32}$$

We have modified the original DELPHI code to allow the treatment of dipolar sources (i.e. using Eq. (27) to iterate instead of Eq. (19)). The free energy for each of the x-, y-, z-direction of a dipole is

$$G_i = -\frac{1}{2} p_i \nabla_i \Phi(\vec{r}) \tag{33}$$

where i here is x, y, or z, and we use a finite difference approximation for the partial derivatives,

$$\nabla_x \Phi(\vec{r}) = \frac{\Phi(\vec{r}+\vec{x}) - \Phi(\vec{r}-\vec{x})}{2x} \tag{34}$$

etc. Higher multipoles can also be included with the right finite difference approximation for the higher-order derivatives in Eq. (28).

Electrostatic optimization in molecular binding

The electrostatic free energy of binding is the difference between the electrostatic free energy in the bound and the unbound state, $\Delta G_{\text{bind}} = G^{\text{bound}} - G^{\text{unbound}}$. Upon binding, the receptor and the ligand lose interaction with solvent but gain favorable intermolecular interaction (which can be assumed to be non-existent in the unbound state because of strong solvent screening). ΔG_{bind} can be written as a sum of three terms, the ligand desolvation penalty, the receptor–ligand interaction in the bound state, and the receptor desolvation penalty,

$$\Delta G_{\text{bind}} = \Delta G_{\text{desolv,L}} + \Delta G_{\text{inter,L-R}} + \Delta G_{\text{desolv,R}} \tag{35}$$

The goal of our electrostatic optimization scheme is to find ligand-charge distributions that minimize ΔG_{bind} given a fixed receptor-charge distribution. We have used two discrete and finite basis sets for the ligand-charge distribution in this thesis, namely point charges and multipoles (truncated according to some convergence criteria). As we now show, ΔG_{bind} is indeed a minimizable function. Specifically, it is quadratic in the ligand-charge distribution (written as $\{Q_{\text{L}}\}$ or $\vec{Q}^{\text{L}}$).[27,28]

The electrostatic free energy of a system of charges is $\frac{1}{2}\vec{Q}^T\vec{\Phi}$ (Eq. (12) in vector notation), while the potential is a linear function of the charges (c.f. LPB equation, Eq. (10)). The ligand desolvation penalty is the difference between the electrostatic free energy of the bound and unbound ligand. Each of these free energy terms has a quadratic dependence on $\vec{Q_{\text{L}}}$ and so does $\Delta G_{\text{desolv,L}}$. The receptor–ligand interaction is the interaction between the potential due to the receptor charges and the ligand charges or vice versa, and is linear in

$\vec{Q}_{\rm L}$ (and also linear in the receptor-charge distribution). The receptor desolvation penalty is quadratic in the receptor-charge distribution, but is independent of $\vec{Q}_{\rm L}$. $\Delta G_{\rm bind}$ as a function of $\vec{Q}_{\rm L}$ is as follows,[27,28]

$$\Delta G_{\rm bind} = \Delta G_{\rm desolv,L} + \Delta G_{\rm inter,L-R} + \Delta G_{\rm desolv,R} \tag{36}$$

$$= \frac{1}{2}\left(\vec{Q}^{\rm L}\right)^T \overleftrightarrow{A}\left(\vec{Q}^{\rm L}\right) + \left(\vec{Q}^{\rm L}\right)^T \vec{B} + C \tag{37}$$

where $\overleftrightarrow{A}$, $\vec{B}$, and C are independent of $\{Q_i^{\rm L}\}$, but depend on the geometry and receptor charges. Because the desolvation matrix $\overleftrightarrow{A}$ is positive definite,[29] $\Delta G_{\rm bind}$ is a multidimensional paraboloid with a minimum, $\Delta G_{\rm bind}^{\rm opt}$, at

$$\vec{Q}^{\rm opt} = -\overleftrightarrow{A}^{-1}\vec{B}. \tag{38}$$

The first implementation of electrostatic optimization is within a spherical model of binding (Chapter III).[27] The ligand is spherical in shape and we choose a set of multipoles at the ligand center to represent the ligand-charge distribution. Any charge distribution inside the ligand sphere can be written as a superposition of multipoles at the ligand center, and we have a complete charge basis set. The bound state is also spherical. $\overleftrightarrow{A}$ and $\vec{B}$ were calculated analytically and the level of multipoles included depends on some convergence criteria.

We next moved on to actual molecular shapes and applied the electrostatic optimization to the barnase–barstar protein complex. Here $\overleftrightarrow{A}$, $\vec{B}$, and C were calculated numerically using DELPHI. Barstar was chosen to be the ligand and the variable ligand-charge basis set was the atomic point charges on barstar. In theory, the ligand desolvation matrix, $\overleftrightarrow{A}$, is symmetric (due to reciprocity) and positive definite,[28] which means its eigenvalues are all positive and $\Delta G_{\rm desolv,L}$ is positive (a desolvation penalty). This is indeed the case for the analytically calculated $\overleftrightarrow{A}$ in the spherical model. When $\overleftrightarrow{A}$ is computed numerically, it is likely that $\overleftrightarrow{A}$ is neither symmetric nor positive definite. The non-symmetric matrix is symmetrized by averaging across the diagonal. To deal with the consequence of limited numerical accuracy of $\overleftrightarrow{A}$ as computed being not positive definite, we use singular value de-

composition (SVD),[30,31] i.e., we construct the optimal ligand charges only from eigenvectors with eigenvalues greater than the absolute value of the most negative eigenvalue.

It is often desirable to impose constraints on the variable charges when we carry out the optimization. For instance, the size of the charges should be constrained to being chemically reasonable. We also constrain the total ligand charge, the total charge of each side chain, and the x-,y-, and z-dipole of each side chain. These constraints are all linear in the variable ligand charges, and since our objective function (i.e., the function to be optimized), ΔG_{bind}, is a quadratic function, the optimization problem is the so-called quadratic programming problem well studied by the operations research community. We use the computer program LOQO[32] to solve our constrained optimization problem.

REFERENCES

[1] K. A. Sharp and B. Honig. Calculating total electrostatic energies with the nonlinear Poisson–Boltzmann equation. *J. Phys. Chem.* **94**: 7684–7692 (1990).

[2] M. K. Gilson, K. A. Sharp, and B. H. Honig. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J. Comput. Chem.* **9**: 327–335 (1988).

[3] K. A. Sharp and B. Honig. Electrostatic interactions in macromolecules: Theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* **19**: 301–332 (1990).

[4] P. Debye and E. Hückel. Zur theorie der electrolyte. *Phyzik. Z.* **24**: 185–206 (1923).

[5] Z. S. Hensdsch, C. V. Sindelar, and B. Tidor. Parameter dependence in continuum electrostatic calculations: A study using protein salt bridges. *J. Phys. Chem. B* **102**: 4404–4410 (1998).

[6] J. D. Jackson. *Classical Electrodynamics.* John Wiley and Sons, New York, second edition (1975).

[7] G. Benedek and F. Villars. *Biological Physics (To be published by American Institute of Physics).*

[8] B. Honig and A. Nicholls. Classical electrostatics in biology and chemistry. *Science (Washington, D.C.)* **268**: 1144–1149 (1995).

[9] J. G. Kirkwood. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *J. Chem. Phys.* **2**: 351–361 (1934).

[10] C. Tanford and J. G. Kirkwood. Theory of protein titration curves. i general equations of impenetrable spheres. *J. Am. Chem. Soc.* **79**: 5333–5339 (1957).

[11] R. M. Fuoss, A. Katchalsky, and S. Lifson. The potential of an infinite rod-like molecule and the distribution of the counter ions. *Proc. Nat. Acad. Sci. U.S.A.* **37**: 579–589 (1951).

[12] F. M. Richards. Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**: 151–176 (1977).

[13] M. L. Connolly. Analytical molecular surface calculation. *J. Appl. Cryst.* **16**: 548–558 (1983).

[14] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**: 187–217 (1983).

[15] D. Sitkoff, K. A. Sharp, and B. Honig. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **98**: 1978–1988 (1994).

[16] W. L. Jorgensen and J. Tirado-Rives. The OPLS potential function for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**: 1657–1666 (1988).

[17] J. T. G. Overbeek. The role of energy and entropy in the electrical double layer. *Colloids and Surfaces* **51**: 61–75 (1990).

[18] F. Fogolari and J. M. Briggs. On the variational approach to Poisson–Boltzmann free energies. *Chem. Phys. Lett.* **281**: 135–139 (1997).

[19] C. M. Cortis and R. A. Friesner. An automatic three-dimensional finite element mesh generation system for the poisson–boltzmann equation. *J. Comp. Chem.* **13**: 1570–1590 (1997).

[20] T. J. You and S. C. Harvey. Finite-element approach to the electrostatics of macromolecules with arbitrary geometries. *J. Comp. Chem.* **14**: 484–501 (1993).

[21] R. J. Zauhar and R. S. Morgan. The rigorous computation of the molecular electric potential. *J. Comput. Chem.* **9**: 171–187 (1988).

[22] R. Bharadwaj, A. Windemuth, S. Sridharan, B. Honig, and A. Nicholls. The fast multipole

boundary element method for molecular electrostatics: An optimal approach for large systems. *J. Comput. Chem.* **16**: 898–913 (1995).

[23] I. Klapper, R. Hagstrom, R. Fine, K. Sharp, and B. Honig. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins: Struct., Funct., Genet.* **1**: 47–59 (1986).

[24] A. Nicholls and B. Honig. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the poisson–boltzmann equation. *J. Comp. Chem.* **12**: 435–445 (1991).

[25] Z. Zhou, P. Payne, M. Vasquez, N. Kuhn, and M. Levitt. Finite-difference solution of the Poisson–Boltzmann equation: Complete elimination of self-energy. *J. Comput. Chem.* **11**: 1344–1351 (1996).

[26] M. E. Rose. The electrostatic interaction of two arbitrary charge distributions. *J. Math. & Phys.* **37**: 215–222 (1958).

[27] L.-P. Lee and B. Tidor. Optimization of electrostatic binding free energy. *J. Chem. Phys.* **106**: 8681–8690 (1997).

[28] E. Kangas and B. Tidor. Optimizing electrostatic affinity in ligand–receptor binding: Theory, computation, and ligand properties. *J. Chem. Phys.* **109**: 7522–7545 (1998).

[29] E. Kangas and B. Tidor. Charge optimization leads to favorable electrostatic binding free energy. *Phys. Rev. E* **59**: 5958–5961 (1999).

[30] G. Strang. *Introduction to Linear Algebra.* Wellesley–Cambridge Press, Wellesley, Massachusetts (1993).

[31] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, Cambridge, second edition (1992).

[32] R. J. Vanderbei. LOQO: An interior-point code for quadratic programming. *Optimization*

methods and Software **To appear.** (1999).

Chapter III

Optimization of Electrostatic Binding Free Energy*

Abstract

An analytic result is derived that defines the charge distribution of the tightest-binding ligand given a receptor charge distribution and spherical geometries. Using the framework of continuum electrostatics, the optimal distribution is expressed as a set of multipoles determined by minimizing the electrostatic free energy of binding. Results for two simple receptor systems are presented to illustrate applications of the theory.

---

*Appears in the Journal of Chemical Physics: Lee-Peng Lee and Bruce Tidor, J. Chem. Phys. **106** (21), 8681 (1997)

I. INTRODUCTION

One mechanism operating in many diseases is the undesirable action of a protein (here termed receptor) that can be arrested, at least in principle, through the tight binding of a molecular ligand (e.g., by sterically blocking the active site or by preventing a required conformational change).[1] To be effective as a drug, such a molecule must possess a number of important pharmacological activities, such as bioavailability and non-toxicity. One step in the discovery of drug molecules is the identification or design of tight-binding ligands. Ligand design is particularly difficult because opposing contributions to the free energy of binding must be properly tuned. For instance, increasing the magnitude of a point charge in a ligand can enhance its interaction with receptor (tending to favor binding), but it will also enhance its interaction with solvent in the unbound state (tending to disfavor binding). What magnitude charge should be chosen to balance these effects and produce the most favorable free energy of binding? The question can be generalized to all multipole terms of the ligand charge distribution. The charge distribution that optimally balances these effects will bind tightest to the receptor.

Here the problem of determining the ligand charge distribution binding tightest to a given receptor is addressed using continuum electrostatic theory. In Section II a solution is presented for the case in which both the free ligand and the bound complex are spherical regions of low dielectric surrounded by aqueous medium of high dielectric and the behavior of the system is governed by the Poisson equation. To facilitate an analytic solution the following assumptions are made: the ligand and receptor do not interact in the unbound state, the ligand charge distribution is the same in the bound and unbound state, and the ligand binds rigidly to the receptor with a unique orientation. The optimal charge distribution is obtained by expressing the ligand charge distribution as an arbitrary set of multipoles and minimizing the free energy of binding with respect to the multipoles. In Section III the theory is applied to a highly symmetric charge distribution devised for test purposes and to a second charge distribution, the terminus of an alpha-helix, present in

some protein binding sites. Discussion and conclusions are presented in Section IV.

II. THEORY

The electrostatic free energy of binding is the difference between the electrostatic free energy in the bound and the unbound state, $\Delta G_{\text{binding}} = G^{\text{bound}} - G^{\text{unbound}}$ (see Fig. 1*a*). Because the dielectric model includes responses that affect the entropy as well as the enthalpy, the electrostatic energy is considered to be a free energy. Here we express the free energy of each state as a sum of coulombic and reaction-field (hydration) terms involving the ligand (L), the receptor (R), and their interaction (L–R)

$$G^{state} = G^{state}_{\text{coul,L}} + G^{state}_{\text{coul,R}} + G^{state}_{\text{coul,L-R}} + G^{state}_{\text{hyd,L}} + G^{state}_{\text{hyd,R}} + G^{state}_{\text{hyd,L-R}}. \tag{1}$$

This results in the following expression for the binding free energy,

$$\Delta G_{\text{binding}} = \Delta G_{\text{coul,L-R}} + \Delta G_{\text{hyd,L-R}} + \Delta G_{\text{hyd,L}} + \Delta G_{\text{hyd,R}} \tag{2}$$

where we have used the fact that the geometry of point charges in the receptor[2] and ligand remain fixed in the model to cancel the coulombic self contribution of ligand and receptor and where the two L–R terms are due only to the bound state because the ligand and receptor are assumed not to interact in the unbound state. Thus, Eq. (2) describes the electrostatic binding free energy as a sum of desolvation contributions of the ligand and the receptor (which are unfavorable) and solvent-screened electrostatic interaction in the bound state (which is usually favorable). Since our goal is to vary the ligand charge distribution to optimize the electrostatic binding free energy and the last term simply adds a constant, we define a relevant variational binding energy,

$$\Delta G_{\text{var}} = \Delta G_{\text{int,L-R}} + \Delta G_{\text{hyd,L}} \tag{3}$$

in which the first two terms on the RHS of Eq. (2) have been combined into a screened interaction term and the constant term has been dropped. Note that

$$\Delta G_{\mathrm{int,L-R}} = \sum_{j\in \mathrm{R}} q_j V_{\mathrm{L}}^{\mathrm{bound}}(\vec{r_j}) = \sum_{j\in \mathrm{R}} q_j \left[V_{\mathrm{coul,L}}^{\mathrm{bound}}(\vec{r_j}) + V_{\mathrm{hyd,L}}^{\mathrm{bound}}(\vec{r_j})\right] \tag{4}$$

and

$$\Delta G_{\mathrm{hyd,L}} = \frac{1}{2}\sum_{i\in \mathrm{L}} q_i V_{\mathrm{hyd,L}}^{\mathrm{bound}}(\vec{r_i}) - \frac{1}{2}\sum_{i\in \mathrm{L}} q_i V_{\mathrm{hyd,L}}^{\mathrm{unbound}}(\vec{r_i}) \tag{5}$$

where V_{L}^{state} is the total electrostatic potential in the indicated state due to the ligand charge distribution only and $V_{term,\mathrm{L}}^{state}$ is the coulombic or reaction-field (hydration) term, as indicated. The summations are over atomic point charges in the ligand ($i \in$ L) or receptor ($j \in$ R). The factor of $\frac{1}{2}$ in Eq. (5) is due to the fact that the ligand charge distribution interacts with the self-induced reaction field.

We proceed by expressing $V_{\mathrm{coul,L}}^{\mathrm{bound}}$, $V_{\mathrm{hyd,L}}^{\mathrm{bound}}$, and $V_{\mathrm{hyd,L}}^{\mathrm{unbound}}$, the three electrostatic potentials in Eqs. (4) and (5), in terms of the given geometry and charge distribution by solving the boundary-value problem shown in Fig. 1*b*. A charge distribution (corresponding to the ligand) is embedded in a sphere of radius R. We take the center of the sphere as the origin of coordinates (unprimed) but expand the charge distribution in multipoles about a second origin (primed) translated a distance d along the z-axis, so that

$$\vec{r}(r,\theta,\phi) = \vec{d}(d,\theta_d = 0,\phi_d = 0) + \vec{r'}(r',\theta',\phi'). \tag{6}$$

The potential everywhere satisfies the Poisson equation. Inside the sphere, it may be written as,

$$V_{\mathrm{in}}(\vec{r}) = \sum_i \frac{q_i}{\epsilon_1|\vec{r} - \vec{r_i}|} + \sum_{l=0}^{\infty}\sum_{m=-l}^{l} A_{l,m} r^l Y_{l,m}(\theta,\phi) \tag{7}$$

where the first term on the RHS is the coulombic and the second is the reaction-field (hydration) potential, and the summation over i corresponds to the ligand point charges. Outside the sphere, the coulombic and reaction-field potential can be combined and written as,

$$V_{\mathrm{out}}(\vec{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \frac{B_{l,m}}{r^{l+1}} Y_{l,m}(\theta,\phi) \tag{8}$$

where $A_{l,m}$ and $B_{l,m}$ are to be determined by the proper boundary conditions and $Y_{l,m}(\theta,\phi)$ are the spherical harmonics. The standard way to proceed is to expand the coulombic term in Eq. (7) in spherical harmonics and multipoles of the charge distribution about the center of the sphere. Here we shift the origin of the multipole expansion to $\vec{d}$,

$$\sum_i \frac{q_i}{\epsilon_1|\vec{r}-\vec{r_i}|} = \sum_i \frac{q_i}{\epsilon_1|(\vec{r}-\vec{d})-\vec{r_i'}|} = \sum_i \frac{q_i}{\epsilon_1|\vec{r'}-\vec{r_i'}|} \tag{9}$$

$$= \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \frac{4\pi}{2l+1} Q_{l,m}'^{*} \frac{Y_{l,m}(\theta',\phi')}{\epsilon_1 r'^{l+1}} \tag{10}$$

where $Q_{l,m}'$ is a spherical multipole expanded about the primed origin, $\vec{d}$,

$$Q_{l,m}'^{*} \equiv \sum_i q_i r_i'^{l} Y_{l,m}^{*}(\theta_i',\phi_i'). \tag{11}$$

Note that throughout this work we adopt the definition of the $Y_{l,m}(\theta,\phi)$ used by Jackson.[3] The expression in Eq. (10) is valid for $r' > r_i'$ (i.e., outside the ligand or, more precisely, outside the sphere whose center is at $\vec{d}$ and whose radius is the distance from $\vec{d}$ to the furthest point charge).

To substitute into Eq. (7) and combine terms involving spherical harmonics, we first expand $Y_{l,m}(\theta',\phi')/r'^{l+1}$ of Eq. (10) in terms of $Y_{l,m}(\theta,\phi)/r^{l+1}$. This is readily done using the results of Greengard,[4] which state that for $r > d$,

$$\frac{Y_{l,m}(\theta',\phi')}{r'^{l+1}} = \sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'} K_{l',m',l,m}\left[\frac{4\pi(2l+1)}{(2l'+1)(2l'+2l+1)}\right]^{\frac{1}{2}} d^{l'} Y_{l',m'}^{*}(\theta_d,\phi_d)\frac{Y_{l'+l,m'+m}(\theta,\phi)}{r^{l'+l+1}} \tag{12}$$

where

$$K_{l',m',l,m} = \left[\frac{(l'+l+m'+m)!(l'+l-m'-m)!}{(l'+m')!(l'-m')!(l+m)!(l-m)!}\right]^{\frac{1}{2}}. \tag{13}$$

Since we have chosen a geometry with $\theta_d = 0$ (Fig. 1*b*), only $m' = 0$ terms in Eq. (12) are non-vanishing, in which case Eq. (10) becomes,

$$\sum_i \frac{q_i}{\epsilon_1|\vec{r}-\vec{r_i}|} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \frac{1}{\epsilon_1}\left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}} Q_{l,m}'^{*} \sum_{l'=0}^{\infty} K_{l',0,l,m} d^{l'}\left(\frac{4\pi}{2l'+2l+1}\right)^{\frac{1}{2}} \frac{Y_{l'+l,m}(\theta,\phi)}{r^{l'+l+1}} \tag{14}$$

in which the multipole distribution is taken about the point $\vec{d}$, but the potential is expressed as a summation of spherical harmonics about the large-sphere center. The above equation can also be written as,

$$\sum_i \frac{q_i}{\epsilon_1|\vec{r}-\vec{r_i}|} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{1}{\epsilon_1}\left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}}\frac{Y_{l,m}(\theta,\phi)}{r^{l+1}}\sum_{l'=|m|}^{l}K_{l-l',0,l',m}d^{l-l'}\left(\frac{4\pi}{2l'+1}\right)^{\frac{1}{2}}Q'^{*}_{l',m} \tag{15}$$

where terms with the same $Y_{l,m}(\theta,\phi)$ are grouped together, as opposed to Eq. (14), where terms with the same $Q'^{*}_{l,m}$ are grouped.

Upon substituting Eq. (15) into Eq. (7) and matching boundary conditions at $r = R$,

$$V_{\text{in}}|_{r=R} = V_{\text{out}}|_{r=R} \tag{16}$$

$$\epsilon_1 \left.\frac{\partial V_{\text{in}}}{\partial r}\right|_{r=R} = \epsilon_2 \left.\frac{\partial V_{\text{out}}}{\partial r}\right|_{r=R} \tag{17}$$

we obtain the hydration (reaction-field) potential inside the sphere,

$$V_{\text{hyd}}(\vec{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}A_{l,m}r^l Y_{l,m}(\theta,\phi) \tag{18}$$

$$= \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}}r^l Y_{l,m}(\theta,\phi)\left(\frac{C_l}{R^{2l+1}}\right)\sum_{l'=|m|}^{l}K_{l-l',0,l',m}d^{l-l'}\left(\frac{4\pi}{2l'+1}\right)^{\frac{1}{2}}Q'^{*}_{l',m} \tag{19}$$

where

$$C_l = \frac{(\epsilon_1-\epsilon_2)}{\epsilon_1[\epsilon_2 + l\epsilon_1/(l+1)]}. \tag{20}$$

We can now write the various V's, with their dependence on the $Q'^{*}_{l,m}$ made explicit. $V'^{\text{bound}}_{\text{coul,L}}$ is given by Eq. (10), $V^{\text{bound}}_{\text{hyd,L}}$ is given by Eq. (19) but rewritten so that the terms with the same $Q'^{*}_{l,m}$ are collected, and $V^{\text{unbound}}_{\text{hyd,L}}$ is given by Eq. (19) with $R = a$ and $d = 0$.

$$V'^{\text{bound}}_{\text{coul,L}}(\vec{r'}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{2l+1}Q'^{*}_{l,m}\frac{Y_{l,m}(\theta',\phi')}{\epsilon_1 r'^{l+1}} \tag{21}$$

$$V^{\text{bound}}_{\text{hyd,L}}(\vec{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=l}^{\infty}\left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}}\left(\frac{4\pi}{2l'+1}\right)^{\frac{1}{2}}Q'^{*}_{l,m}K_{l'-l,0,l,m}d^{l'-l}\left(\frac{C_{l'}}{R^{2l'+1}}\right)r^{l'}Y_{l',m}(\theta,\phi) \tag{22}$$

$$V_{\mathrm{hyd,L}}^{\mathrm{unbound}}(\vec{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{4\pi}{2l+1} \left(\frac{C_l}{a^{2l+1}} \right) Q'^{*}_{l,m} r^l Y_{l,m}(\theta, \phi) \tag{23}$$

Substituting into Eq. (4), we make explicit the dependence of $\Delta G_{\mathrm{int,L-R}}$ on the $Q'^{*}_{l,m}$,

$$\Delta G_{\mathrm{int,L-R}} = \sum_{j \in \mathrm{R}} q_j \left[V_{\mathrm{coul,L}}^{\prime \mathrm{bound}}(\vec{r_j'}) + V_{\mathrm{hyd,L}}^{\mathrm{bound}}(\vec{r_j'}) \right] \tag{24}$$

$$= \sum_{l=0}^{\infty} \sum_{m=-l}^{l} Q'^{*}_{l,m} \sum_{j \in \mathrm{R}} q_j \left[\left(\frac{4\pi}{2l+1} \right) \frac{Y_{l,m}(\theta_j', \phi_j')}{\epsilon_1 r_j'^{l+1}} \right.$$

$$\left. + \sum_{l'=l}^{\infty} \left(\frac{4\pi}{2l+1} \right)^{\frac{1}{2}} \left(\frac{4\pi}{2l'+1} \right)^{\frac{1}{2}} K_{l'-l,0,l,m} d^{l'-l} \left(\frac{C_{l'}}{R^{2l'+1}} \right) r_j^{l'} Y_{l',m}(\theta_j, \phi_j) \right] \tag{25}$$

$$\equiv \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \alpha_{l,m} Q'^{*}_{l,m} \tag{26}$$

where in the last line we have defined the element $\alpha_{l,m}$, which is independent of the $Q'^{*}_{l,m}$, to be the factor multiplying $Q'^{*}_{l,m}$ in Eq. (25). Each $\alpha_{l,m}$ expresses the contribution of a multipole to $\Delta G_{\mathrm{int,L-R}}$ and contains all information concerning the receptor charge distribution required to obtain $\Delta G_{\mathbf{var}}$.

For $\Delta G_{\mathrm{hyd,L}}$ it is useful to re-express Eq. (5) in terms of the $Q'^{*}_{l,m}$, the multipoles describing the ligand charge distribution, rather than the individual charges, q_i. We expand $V(\vec{r})$ around the center of the multipole expansion, $\vec{d}$,

$$\sum_{i \in \mathrm{L}} q_i V(\vec{r_i}) = \sum_{i \in \mathrm{L}} q_i V(\vec{d} + \vec{r_i'}) \tag{27}$$

$$= \sum_{i \in \mathrm{L}} q_i \left[V(\vec{d}) + \vec{r_i'} \cdot \vec{\nabla} V(\vec{d}) + \cdots \right]. \tag{28}$$

It has been shown by Rose that in spherical coordinates the expansion becomes,[5]

$$\sum_{i \in \mathrm{L}} q_i V(\vec{d} + \vec{r_i'}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \frac{4\pi}{(2l+1)!!} Q'^{*}_{l,m} \mathcal{Y}_{l,m}(\vec{\nabla}) V(\vec{d}) \tag{29}$$

where

$$\mathcal{Y}_{l,m}(\vec{r}) \equiv r^l Y_{l,m}(\theta, \phi) \tag{30}$$

and $\mathcal{Y}_{l,m}(\vec{\nabla})$ is the operator obtained by replacing $\vec{r}$ with $\vec{\nabla}$. For positive m and when $\mathcal{Y}_{l,m}(\vec{\nabla})$ operates on a solution of the Laplace equation (i.e., $r^l Y_{l,m}(\theta, \phi)$ or $Y_{l,m}(\theta, \phi)/r^{l+1}$), it has been shown that,[5]

$$\mathcal{Y}_{l,m}(\vec{\nabla}) = \frac{(2l)!}{2^l l!}\left[\left(\frac{2l+1}{4\pi}\right)\frac{2^m}{(l+m)!(l-m)!}\right]^{\frac{1}{2}} \nabla_1^m \nabla_0^{l-m} \quad \text{for } m \geq 0. \tag{31}$$

The double-factorial is defined as

$$(2l+1)!! = (2l+1)\cdot(2l-1)\cdot(2l-3)\cdots 3\cdot 1 \tag{32}$$

$$= \frac{(2l+1)!}{2^l l!} \tag{33}$$

and the spherical partial derivatives are

$$\nabla_1 = -\frac{1}{\sqrt{2}}(\nabla_x + i\nabla_y), \qquad \nabla_{-1} = \frac{1}{\sqrt{2}}(\nabla_x - i\nabla_y), \qquad \nabla_0 = \nabla_z. \tag{34}$$

To compute $\mathcal{Y}_{l,m}(\vec{\nabla})$ for negative m, we use the fact that $Y_{l,-m}(\theta,\phi) = (-1)^m Y^*_{l,m}(\theta,\phi)$ and the definitions of spherical partial derivatives in Eq. (34) to obtain,

$$\mathcal{Y}_{l,-m}(\vec{\nabla}) = \frac{(2l)!}{2^l l!}\left[\left(\frac{2l+1}{4\pi}\right)\frac{2^m}{(l+m)!(l-m)!}\right]^{\frac{1}{2}} \nabla_{-1}^m \nabla_0^{l-m} \quad \text{for } m \geq 0. \tag{35}$$

The hydration energy of the bound ligand is then

$$G^{\text{bound}}_{\text{hyd,L}} = \frac{1}{2}\sum_{i\in \text{L}} q_i V^{\text{bound}}_{\text{hyd,L}}(\vec{d}+\vec{r_i'}) = \frac{1}{2}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'} \frac{4\pi}{(2l'+1)!!} Q'^*_{l',m'} \mathcal{Y}_{l',m'}(\vec{\nabla}) V^{\text{bound}}_{\text{hyd,L}}(\vec{d}) \tag{36}$$

$$= \frac{1}{2}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'} \frac{4\pi}{(2l'+1)!!} Q'^*_{l',m'} \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l''=l}^{\infty} \left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}} \left(\frac{4\pi}{2l''+1}\right)^{\frac{1}{2}}$$
$$\times Q'^*_{l,m} K_{l''-l,0,l,m} d^{l''-l} \left(\frac{C_{l''}}{R^{2l''+1}}\right) \mathcal{Y}_{l',m'}(\vec{\nabla})(r^{l''} Y_{l'',m}(\theta,\phi))\Big|_{\vec{r}=\vec{d}}. \tag{37}$$

To evaluate $\mathcal{Y}_{l,m}(\vec{\nabla})$ in Eq. (37), we use Eq. (31) and the gradient formula,[6]

$$\vec{\nabla}(\Phi(r)Y_{l,m}(\theta,\phi)) = -\left(\frac{l+1}{2l+1}\right)^{\frac{1}{2}}\left(\frac{d\Phi(r)}{dr} - \frac{l}{r}\Phi(r)\right) T_{l,l+1,m}(\theta,\phi)$$
$$+\left(\frac{l}{2l+1}\right)^{\frac{1}{2}}\left(\frac{d\Phi(r)}{dr} + \frac{l+1}{r}\Phi(r)\right) T_{l,l-1,m}(\theta,\phi) \tag{38}$$

where

$$T_{l,l',m}(\theta,\phi) \equiv \sum_{m'\in\{-1,0,1\}} \mathcal{C}(l',1,l;m-m',m') Y_{l',m-m'}(\theta,\phi)\hat{\xi}_{m'} \tag{39}$$

the $\mathcal{C}(l',1,l;m-m',m')$ are the vector addition (or Clebsch–Gordon) coefficients frequently encountered in the study of angular momentum shown in Table I,[6] and $\hat{\xi}_{m'}$ are spherical unit vectors,

$$\hat{\xi}_1 = -\frac{1}{\sqrt{2}}(\hat{x} + i\hat{y}), \qquad \hat{\xi}_{-1} = \frac{1}{\sqrt{2}}(\hat{x} - i\hat{y}), \qquad \hat{\xi}_0 = \hat{z}. \tag{40}$$

It is straightforward to show that

$$\vec{\nabla} = \hat{x}\nabla_x + \hat{y}\nabla_y + \hat{z}\nabla_z = -\hat{\xi}_1\nabla_{-1} - \hat{\xi}_{-1}\nabla_1 + \hat{\xi}_0\nabla_0. \tag{41}$$

From Eqs. (38) through (41), we have

$$\nabla_\mu(r^l Y_{l,m}(\theta,\phi)) = (-1)^\mu [l(2l+1)]^{\frac{1}{2}} \mathcal{C}(l-1,1,l;m+\mu,-\mu) r^{l-1} Y_{l-1,m+\mu}(\theta,\phi). \tag{42}$$

Using Table I, Eq. (31), and Eqs. (37) through (42), we obtain the following intermediate results,

$$\nabla_0^{l'-m'}(r^{l''}Y_{l'',m}) = \left[\frac{(2l''+1)(l''+m)!(l''-m)!}{(2l''-2l'+2m'+1)(l''-m-l'+m')!(l''+m-l'+m')!}\right]^{\frac{1}{2}} r^{l''-l'+m'} Y_{l''-l'+m',m} \tag{43}$$

$$\nabla_1^{m'}(r^{l''-l'+m'}Y_{l''-l'+m',m}) = (-1)^{m'}\left[\frac{(2l''-2l'+2m'+1)(l''-m-l'+m')!}{2^{m'}(2l''-2l'+1)(l''-m-l'-m')!}\right]^{\frac{1}{2}} r^{l''-l'} Y_{l''-l',m+m'} \tag{44}$$

$$\nabla_{-1}^{m'}(r^{l''-l'+m'}Y_{l''-l'+m',m}) = (-1)^{m'}\left[\frac{(2l''-2l'+2m'+1)(l''+m-l'+m')!}{2^{m'}(2l''-2l'+1)(l''+m-l'-m')!}\right]^{\frac{1}{2}} r^{l''-l'} Y_{l''-l',m-m'} \tag{45}$$

and the final expression for the hydration energy of the ligand in the bound state,

$$\begin{aligned} G_{\text{hyd,L}}^{\text{bound}} = \frac{1}{2}\sum_{i\in \text{L}} q_i V_{\text{hyd,L}}^{\text{bound}}(\vec{r}_i) &= \frac{1}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=0}^{\infty} Q_{l,m}'^{*} Q_{l',m}' \sum_{l''=\max(l,l')}^{\infty}\left(\frac{4\pi}{2l+1}\right)^{\frac{1}{2}}\left(\frac{4\pi}{2l'+1}\right)^{\frac{1}{2}}\frac{C_{l''}}{R^{2l''+1}} \\ &\times\frac{(l''+m)!(l''-m)!}{(l''-l)!(l''-l')!}\left[\frac{1}{(l+m)!(l-m)!(l'+m)!(l'-m)!}\right]^{\frac{1}{2}} d^{2l''-l-l'} \qquad (46) \\ &\equiv \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'} \beta_{l,m,l',m'} Q_{l,m}'^{*} Q_{l',m'}' \qquad (47) \end{aligned}$$

where $\beta_{l,m,l',m'}$ is defined by the above two equations; note that $\beta_{l,m,l',m'}$ is zero for $m' \neq m$.

We obtain the hydration energy of the unbound ligand by setting $d = 0$ and $R = a$ in Eq. (46),

$$G_{\mathrm{hyd,L}}^{\mathrm{unbound}} = \frac{1}{2}\sum_{i\in \mathrm{L}} q_i V_{\mathrm{hyd,L}}^{\mathrm{unbound}}(\vec{r_i}) = \frac{1}{2}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{4\pi}{2l+1}\left(\frac{C_l}{a^{2l+1}}\right) Q_{l,m}'^{*} Q_{l,m}' \tag{48}$$

$$\equiv \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\gamma_{l,m} Q_{l,m}'^{*} Q_{l,m}' \tag{49}$$

where $\gamma_{l,m}$ is defined by Eqs. (48) and (49). We write $\gamma_{l,m}$ as a function of both l and m for notational convenience, although there is no formal dependence on m.

Thus ΔG_{var} has been expressed as a function of the multipoles of the ligand charge distribution, $Q_{l,m}'$ (expanded about the center of the ligand sphere) and the elements $\alpha_{l,m}$, $\beta_{l,m,l',m'}$ and $\gamma_{l,m}$, which do not depend on $Q_{l,m}'$. Combining Eqs. (26), (47), and (49) gives

$$\Delta G_{\mathrm{var}} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\alpha_{l,m}Q_{l,m}'^{*} + \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}\beta_{l,m,l',m'}Q_{l,m}'^{*}Q_{l',m'}' - \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\gamma_{l,m}Q_{l,m}'^{*}Q_{l,m}'. \tag{50}$$

Note that only the $\alpha_{l,m}$ depend on the receptor charges, while the $\beta_{l,m,l',m'}$ and $\gamma_{l,m}$ depend solely on the geometry of the bound and unbound states. While $\Delta G_{\mathrm{var}}^{\mathrm{opt}}$ is a real quantity, the $\alpha_{l,m}$ and $Q_{l,m}'$ are complex and the products $\alpha_{l,m}Q_{l,m}'^{*}$ and $Q_{l,m}'^{*}Q_{l',m}'$ involve summations over terms of the form $Y_{l',m}^{*}(\theta',\phi')Y_{l,m}(\theta,\phi)$; note that the $\beta_{l,m,l'm'}$ and $\gamma_{l,m}$ are real. We rewrite $\Delta G_{\mathrm{var}}^{\mathrm{opt}}$ in terms of the real and imaginary parts of $\alpha_{l,m}$ and $Q_{l,m}'$,

$$\begin{aligned}\Delta G_{\mathrm{var}} = &\sum_{l=0}^{\infty}\left[\alpha_{l,0}Q_{l,0}' + 2\sum_{m=1}^{l}(\mathrm{Re}\alpha_{l,m}\mathrm{Re}Q_{l,m}' + \mathrm{Im}\alpha_{l,m}\mathrm{Im}Q_{l,m}')\right]\\ &+\sum_{l=0}^{\infty}\sum_{l'=0}^{\infty}\left[\beta_{l,0,l',0}Q_{l,0}'Q_{l',0}' + 2\sum_{m=1}^{l}\beta_{l,m,l',m}(\mathrm{Re}Q_{l,m}'\mathrm{Re}Q_{l',m}' + \mathrm{Im}Q_{l,m}'\mathrm{Im}Q_{l',m}')\right]\\ &-\sum_{l=0}^{\infty}\left[\gamma_{l,0}Q_{l,0}'^{2} + 2\sum_{m=1}^{l}\gamma_{l,m}(\mathrm{Re}Q_{l,m}'^{2} + \mathrm{Im}Q_{l,m}'^{2})\right]\end{aligned} \tag{51}$$

(where the summations over m are excluded for $l = 0$) by noting again that $Y_{l,-m}(\theta,\phi) = (-1)^m Y_{l,m}^{*}(\theta,\phi)$ and

$$Y_{l',m}^{*}(\theta',\phi')Y_{l,m}(\theta,\phi) + Y_{l',-m}^{*}(\theta',\phi')Y_{l,-m}(\theta,\phi) = Y_{l',m}^{*}(\theta',\phi')Y_{l,m}(\theta,\phi) + Y_{l',m}(\theta',\phi')Y_{l,m}^{*}(\theta,\phi) \tag{52}$$

$$= 2\left[\mathrm{Re}Y_{l',m}(\theta',\phi')\cdot\mathrm{Re}Y_{l,m}(\theta,\phi) + \mathrm{Im}Y_{l',m}(\theta',\phi')\cdot\mathrm{Im}Y_{l,m}(\theta,\phi)\right]. \tag{53}$$

The new variables $\mathrm{Re}Q_{l,m}'$ and $\mathrm{Im}Q_{l,m}'$ are re-indexed and renamed Q_i as follows,

$$\{Q'_{0,0}, Q'_{1,0}, \mathrm{Re}Q'_{1,1}, \mathrm{Im}Q'_{1,1}, Q'_{2,0}, \mathrm{Re}Q'_{2,1}, \mathrm{Im}Q'_{2,1}, \mathrm{Re}Q'_{2,2}, \cdots\} \leftrightarrow \{Q_1, Q_2, Q_3, Q_4, Q_5, Q_6, Q_7, Q_8, \cdots\}. \tag{54}$$

and similar transformations are used to create α_i, β_{ij}, and γ_i. Eq. (51) can then be written as

$$\Delta G_{\mathrm{var}} = \sum_{i=1}^{\infty} \alpha_i Q_i + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \beta_{ij} Q_i Q_j - \sum_{i=1}^{\infty} \gamma_i Q_i^2 \tag{55}$$

$$= \sum_{i=1}^{\infty} \alpha_i Q_i + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} (\beta_{ij} - \delta_{ij}\gamma_i) Q_i Q_j \tag{56}$$

or in matrix notation,

$$\Delta G_{\mathrm{var}} = \vec{Q}^T \overleftrightarrow{B}\, \vec{Q} + \vec{Q}^T \vec{A} \tag{57}$$

$$= (\vec{Q} + \frac{1}{2}\, \overleftrightarrow{B}^{-1} \vec{A})^T\, \overleftrightarrow{B}\, (\vec{Q} + \frac{1}{2}\, \overleftrightarrow{B}^{-1} \vec{A}) - \frac{1}{4} \vec{A}^T\, \overleftrightarrow{B}^{-1} \vec{A} \tag{58}$$

where $\vec{Q}$ is the vector formed by the Q_i, $\vec{A}$ is the vector formed by the α_i, $\overleftrightarrow{B}$ is the symmetric matrix formed by the $(\beta_{ij} - \delta_{ij}\gamma_i)$, and completion of the square has been used to arrive at Eq. (58). Since $\vec{Q}^T \overleftrightarrow{B} \vec{Q}$ in Eq. (57) corresponds to the ligand desolvation penalty, which must be greater than zero for chemically reasonable geometries, the matrix $\overleftrightarrow{B}$ is positive definite and the extremum of ΔG_{var} is a minimum.[7] From Eq. (58) the optimum values of the multipoles, $\vec{Q}^{\mathrm{opt}}$ and the minimum variational binding energy, $\Delta G_{\mathrm{var}}^{\mathrm{opt}}$ are obtained,

$$\vec{Q}^{\mathrm{opt}} = -\frac{1}{2}\, \overleftrightarrow{B}^{-1} \vec{A} \tag{59}$$

$$\Delta G_{\mathrm{var}}^{\mathrm{opt}} = -\frac{1}{4} \vec{A}^T\, \overleftrightarrow{B}^{-1} \vec{A}. \tag{60}$$

$\Delta G_{\mathrm{var}}^{\mathrm{opt}}$ is always negative because $\overleftrightarrow{B}^{-1}$ is also positive definite.

To solve for the optimal multipole distribution with the monopole (total charge) fixed ($Q_1 = \mathcal{Q}$), the equation for the remaining optimal multipoles ($i \neq 1$) is,

$$2\sum_{j\neq 1} (\beta_{ij} - \delta_{ij}\gamma_i) Q_j^{\mathrm{opt}} + (2\beta_{i1}\mathcal{Q} + \alpha_i) = 0 \tag{61}$$

which is analogous to Eq. (59).

The above matrix equations, with the dimension truncated at $i_{\max} = (l_{\max} + 1)^2$, can be solved numerically by relatively modest computational resources. In practice, since the α_i and β_{ij} contain a summation over an infinite number of terms, a second cutoff value of l_{cut} must be used to truncate the innermost sum in Eqs. (25) and (46). When $l_{\max}$ and l_{cut} are sufficiently large, $\Delta G_{\text{var}}^{\text{opt}}$ converges and the incremental advantage of including more multipoles essentially vanishes.

For any given receptor and geometry, we have thus described a method to determine the charge distribution of the tightest binding ligand as a set of multipoles. The deviation of the binding free energy from the optimum for any test ligand can be calculated by subtracting Eq. (60) from Eq. (58) and using Eq. (59) to eliminate $\vec{A}$,

$$\Delta G_{\text{var}} - \Delta G_{\text{var}}^{\text{opt}} = (\vec{Q} - \vec{Q}^{\text{opt}})^T \overleftrightarrow{B} (\vec{Q} - \vec{Q}^{\text{opt}}). \tag{62}$$

III. RESULTS

A. Implementation

The algorithm described was implemented in a computer program whose input was $l_{\max}$ (which determined the size of the matrix in Eqs. (59) and (61)), l_{cut} (which was used to truncate the innermost summation in Eqs. (25) and (46)), the geometry of the problem, and whether the monopole of the optimum was to be free or fixed at some value. The geometry of the problem included the radius and coordinates of the center of both the bound-state and ligand spheres (on the z-axis) and the coordinates and magnitude of each partial atomic charge in the system. The dielectric constants ϵ_1 and ϵ_2 were chosen to be 4 and 80, respectively. Evaluation of the α_i, β_{ij}, and γ_i was carried out, followed by solution of the matrix equation (Eq. (59) or (61)) using LU decomposition. The eigenvalues of the $\overleftrightarrow{B}$ matrix were obtained to verify that the stationary point was a minimum. All real floating-point values were represented using 64 bits. The matrix algebra was accomplished using increased-precision versions of the appropriate subroutines given by Press *et al.*[8] The

output of the program included the multipoles for the optimal charge distribution, $\Delta G_{\text{var}}^{\text{opt}}$, the nature of the stationary point, and a file recording the α_i, β_{ij} and γ_i. Typical CPU usage for a receptor with $l_{\max} = l_{\text{cut}} = 40$ was 20 minutes on a Hewlett-Packard 9000/735 with the PA–7200 (99 MHz) chip, and the maximum memory used was roughly 22 MB. Because we have used a direct method (i.e., LU decomposition) to solve the matrix equation, where the matrix is of size $(l_{\max}+1)^2 \times (l_{\max}+1)^2$, the time scales as $(l_{\max})^6$ and the memory scales as $(l_{\max})^4$. At this point no attempt has been made to optimize the code. For example, the matrix equation contains a particularly sparse matrix (due to the azimuthal geometry chosen for the problem) that may be used to reduce the necessary computational effort. The optimization problem may also be solved with iterative methods, such as the conjugate gradient method or various relaxation methods.

B. Test Problems

The first test problem consisted of a receptor with four parallel dipolar groups, each containing a negative charge of $-0.55e$ in the $z = 15.50$ plane and a positive charge of $+0.55e$ in the $z = 14.25$ plane. All lengths and distances are given in units of angströms (1 Å=0.1 nm). The (x, y) coordinates of the charges were $(+1.5, +1.5)$, $(-1.5, +1.5)$, $(-1.5, -1.5)$, and $(+1.5, -1.5)$. The bound-state low-dielectric region was bounded by a sphere of radius 24.0 centered at the origin, the ligand sphere was of radius 4.0 and was centered at $(x, y, z) = (0.0, 0.0, 20.0)$ in the bound state.

The second test problem consisted of an idealized alpha-helix as the receptor. The helix was constructed from 18 alanine residues with acetyl and N-methylamide blocking groups at the N- and C-terminus, respectively. Coordinates were generated in the polar-hydrogen representation with the CHARMM PARAM19[9] bond lengths and angles and with $\phi = -57°$ and $\psi = -47°$. The partial atomic charges were adapted from the PARSE parameter set.[10] The axis of the helix coincided with the z-axis of the coordinate system and the nitrogen atom of Ala 10 was closest to the origin. The bound-state low-dielectric region was bounded

by a sphere of radius 24.0 centered at the origin, the ligand sphere was of radius 4.0, and the ligand multipole distribution was centered at $(x, y, z) = (0.0, 0.0, 20.0)$ in the bound state (near the C-terminus of the poly-alanine alpha-helix).

C. Analysis of Results

Each test problem was solved multiple times using different values of $l_{\max}$ (with l_{cut} fixed at 40 for the results shown here, though essentially indistinguishable results were obtained when the value was increased to 80) and with the monopole of the variational distribution either free or fixed at 0 or $+1e$. Fig. 2 shows the convergence of the calculated $\Delta G^{\text{opt}}_{\text{var}}$ as a function of the value of $l_{\max}$ used (part *a* is for the four-dipolar-groups problem and part *b* is for the alpha-helix). In all cases the calculated $\Delta G^{\text{opt}}_{\text{var}}$ was monotonically decreasing for increasing $l_{\max}$, and for any value of $l_{\max}$, $\Delta G^{\text{opt}}_{\text{var}}$ was lower (more favorable) with free than with fixed monopole value, as expected for a variational optimization. For both test problems the value of $\Delta G^{\text{opt}}_{\text{var}}$ appeared to change very little beyond an $l_{\max}$ of 20 for floating or fixed monopole. Fig. 3 shows the magnitude of the low multipole moments of the optimized distribution as a function of $l_{\max}$ (with free monopole value). The magnitude of the 2^l-pole is defined as $|\tilde{Q}_l| \equiv \sqrt{\left(\frac{4\pi}{2l+1}\right) \sum_m (Q_{l,m}/a^l)^2}$, where a is the ligand radius. The magnitudes of the first six multipoles converged by an $l_{\max}$ of 10. Fig. 4 shows the coulombic potential due to the calculated optimal ligand, again as a function of $l_{\max}$; the potential appeared nearly converged at an $l_{\max}$ of 20. The converged coulombic potential of the optimal ligand, plotted in the xy-plane just outside the ligand (at $z = 16.0$) and computed with an $l_{\max}$ of 40, is shown in Fig. 5*a* for the four-dipolar-groups problem and Fig. 5*c* for the alpha-helix. The optimal ligand's potential contained the appropriate four-fold symmetry to match that of the four-dipolar-groups receptor, indicating that such a ligand would interact equally with all four dipoles. However, for the alpha-helix, which presents a coil of dipolar groups receding in the z-direction, it appears that the optimal ligand computed in this manner would interact strongly only with the closest dipolar group.

It is also interesting to note that the coulombic potential due to the optimized multipole distribution calculated in this way is not a simple reflection of the coulombic potential for the isolated receptor. Compare, for instance, the coulombic potentials due to the optimized ligand (Fig. 5*a*) and due to the receptor (Fig. 5*b*) for the four-dipolar-groups problem, both computed in the $z = 16.0$ plane. The peaks in the ligand potential are "inside" those of the receptor potential. This may turn out to be a general feature of electrostatically optimized binding interactions, which are fundamentally asymmetrical, since one distribution is fixed while the other is optimized.

IV. DISCUSSION AND CONCLUSION

Analytic solutions to the Poisson equation have been used to define the multipole distribution of the ligand that produces a minimum for the free energy of binding a spherical ligand to an invariant receptor to form a spherical complex. An algorithm has been developed and implemented using numerical computation to evaluate the analytic theory, and results have been presented for two test cases. In all solutions examined to date, second-derivative analysis has verified that the stationary point is a minimum. In this sense, the multipole distribution is said to be an optimum. An important feature of the theory presented is that, by expressing the optimum as a multipole distribution, it can be solved for directly, without resorting to stochastic searches or other non-deterministic methods of optimization. This characterization of the multipole properties of the optimal charge distribution for a given spherical ligand shape and binding geometry may be useful in understanding complementary interactions in molecular binding and recognition. Such properties may prove particularly applicable to the field of ligand design either by facilitating the construction of individual tight-binding ligands or by providing descriptors that can be used to search compound libraries or aid in the design of combinatorial libraries.

The observation that an optimum can be defined within the continuum model presented here so as to provide the greatest excess of favorable interactions between ligand and receptor

over unfavorable ligand desolvation energy suggests that the successful design of a tight-binding ligand may involve substantially more than the construction of a complementary-shaped molecule that provides compensating interactions for polar and charged groups in the receptor binding site. For example, the electrostatics of compensating a neutral, polar carbonyl group in a receptor with a neutral, polar hydroxyl may be substantially different than complementing it with a positively charged ammonium group. Moreover, due to the effects of longer-range electrostatic interactions, merely discussing the problem in terms of individually compensating pairs of groups may be inappropriate, since each group affects the overall multipole moments of the ligand. To help answer these questions, we are currently studying algorithms for designing sets of point charges, as well as molecules, that have multipole moments corresponding closely to the optimum defined by this algorithm.

The properties of the optimal multipole distribution and binding energy are worthy of further study. Here we note that $\Delta G_{\mathrm{var}}^{\mathrm{opt}}$ is always negative (favorable), but that the overall binding free energy, $\Delta G_{\mathrm{binding}}$, may or may not be positive (unfavorable) due to the receptor desolvation energy.[11] Moreover, it is straightforward to prove that the magnitude of the screened ligand–receptor interaction free energy is twice that of the ligand desolvation energy at the optimum ($\Delta G_{\mathrm{int,L-R}}^{\mathrm{opt}} = -2\Delta G_{\mathrm{hyd,L}}^{\mathrm{opt}}$, so $\Delta G_{\mathrm{var}}^{\mathrm{opt}} = -\Delta G_{\mathrm{hyd,L}}^{\mathrm{opt}} = \frac{1}{2}\Delta G_{\mathrm{int,L-R}}^{\mathrm{opt}}$), and that the same relationship holds for the contribution of each multipole component, Q_i^{opt}.[12] Finally, the relationship between the coulombic potential of the optimized charge distribution and that of the receptor reveals non-trivial features that reflect subtleties of how best to achieve favorable interactions in the bound state relative to ligand desolvation. For the example involving four dipolar groups, this suggests that chemical groups compensating each dipole should lie closer to the azimuthal axis than the corresponding receptor dipole.

The current theory provides a useful starting point for further studies. We are presently investigating extensions to solve the linearized[13] and the non-linear Poisson–Boltzmann equation, which would allow ionic-strength effects of the aqueous medium to be included. Moreover, it may be possible to release the restrictions that both the unbound ligand and

the bound complex have spherical geometry, that the charge distribution of the ligand be the same in the bound and unbound states, and that titratable groups be treated in a fixed protonation state. It should be noted that there is no restriction in the current theory on the shape or charge distribution of the unbound receptor, since its contribution is a constant that has been eliminated in the definition of ΔG_{var}.

REFERENCES

[1] M. Perutz. *Protein Structure: New Approaches to Disease and Therapy.* W. H. Freeman and Company, New York (1992).

[2] The charge distribution for the receptor need not be the same in the bound and unbound states. If they are different, this adds a constant to $\Delta G_{\text{binding}}$ that can be dropped in defining ΔG_{var} in Eq. (3).

[3] J. D. Jackson. *Classical Electrodynamics.* John Wiley and Sons, New York, second edition (1975).

[4] L. Greengard. *The Rapid Evaluation of Potential Fields in Particle Systems.* MIT Press, Cambridge, Massachusetts (1988).

[5] M. E. Rose. The electrostatic interaction of two arbitrary charge distributions. *J. Math. & Phys.* **37**: 215–222 (1958).

[6] M. E. Rose. *Elementary Theory of Angular Momentum.* John Wiley and Sons, New York (1957).

[7] G. Strang. *Introduction to Applied Mathematics.* Wellesley–Cambridge Press, Wellesley, Massachusetts (1986).

[8] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, Cambridge (1988).

[9] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**: 187–217 (1983).

[10] D. Sitkoff, K. A. Sharp, and B. Honig. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **98**: 1978–1988 (1994).

[11] Z. S. Hendsch and B. Tidor. Unpublished.

[12] S. E. Dempster, L.-P. Lee, and B. Tidor. Unpublished.

[13] J. G. Kirkwood. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *J. Chem. Phys.* **2**: 351–361 (1934).

TABLES

TABLE I. $\mathcal{C}(l', 1, l; m - m', m')$.[a]

	$m' = 1$	$m' = 0$	$m' = -1$
$l = l' + 1$	$\left[\frac{(l'+m)(l'+m+1)}{(2l'+1)(2l'+2)}\right]^{\frac{1}{2}}$	$\left[\frac{(l'-m+1)(l'+m+1)}{(2l'+1)(l'+1)}\right]^{\frac{1}{2}}$	$\left[\frac{(l'-m)(l'-m+1)}{(2l'+1)(2l'+2)}\right]^{\frac{1}{2}}$
$l = l'$	$-\left[\frac{(l'+m)(l'-m+1)}{2l'(l'+1)}\right]^{\frac{1}{2}}$	$\frac{m}{[l'(l'+1)]^{\frac{1}{2}}}$	$\left[\frac{(l'-m)(l'+m+1)}{2l'(l'+1)}\right]^{\frac{1}{2}}$
$l = l' - 1$	$\left[\frac{(l'-m)(l'-m+1)}{2l'(2l'+1)}\right]^{\frac{1}{2}}$	$-\left[\frac{(l'-m)(l'+m)}{l'(2l'+1)}\right]^{\frac{1}{2}}$	$\left[\frac{(l'+m+1)(l'+m)}{2l'(2l'+1)}\right]^{\frac{1}{2}}$

[a]from reference 6

FIGURES

FIG. 1. Illustration of problem geometries. (*a*) The binding reaction is shown between a receptor (R) and spherical ligand (L) that dock rigidly to form a spherical bound-state complex. Receptor, ligand, and complex are all low-dielectric media (ϵ_1) that are surrounded by high-dielectric solvent (ϵ_2). (*b*) The boundary-value problem solved here involves a charge distribution in a spherical region of radius R with dielectric constant ϵ_1 surrounded by solvent with dielectric constant ϵ_2. The origin of coordinates is the center of the larger spherical region, but the charge distribution is expanded in multipoles about a point a distance d along the z-axis. The geometric requirement is that the ligand sphere not extend beyond the receptor sphere, $R \geq d + a$, although the case of equality is illustrated in the figure.

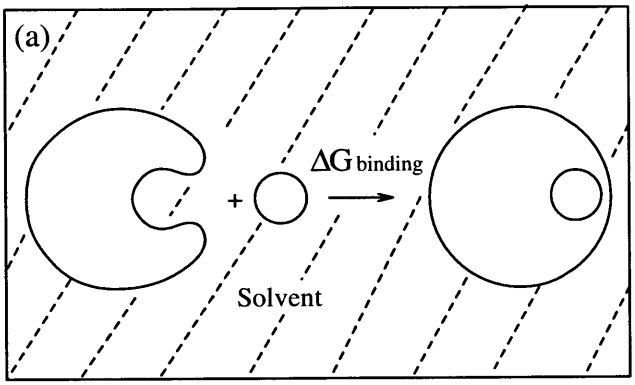
FIG. 2. Convergence of $\Delta G^{\rm opt}_{\rm var}$ as a function of the value of $l_{\rm max}$ used in the calculation. A constant value of $l_{\rm cut} = 40$ was used throughout. Optimizations in which the total charge on the ligand was free are plotted with ($\Box$), fixed at 0 with ($\times$), and fixed at 1 with ($\Diamond$) for (*a*) the four dipolar groups and (*b*) the alpha-helix.

FIG. 3. Convergence of the magnitude of the lowest seven 2^l-poles for the optimal ligand as a function of the value of $l_{\rm max}$ used in the calculation for (*a*) the four dipolar groups and (*b*) the alpha-helix. The optimizations were performed with no constraint on the total ligand charge. In (*a*) note that the $l = 4$ and $l = 5$ lines fall nearly on top of one another.

FIG. 4. Convergence of the coulombic potential due to the optimal ligand, plotted along the line ($y = -1.1, z = 16.0$) for a range of values of $l_{\rm max}$, for (*a*) the four dipolar groups and (*b*) the alpha-helix. The optimizations were performed with no constraint on the total ligand charge. Note that the curves for $l_{\rm max} = 20$ and 40 are nearly identical.

FIG. 5. Contour plot of the coulombic potential in the $z = 16.0$ plane for (*a*) the optimum ligand for the four dipolar groups, (*b*) the four dipolar groups themselves, (*c*) the optimum ligand for the alpha-helix, and (*d*) the alpha-helix itself. The optimizations were performed with no constraint on the total ligand charge. Each plot consists of equally spaced contour levels. Each label marks the closest contour level and is valid to three decimal places (i.e., 0.32 in (*a*) is 0.320 and −0.8 in (*b*) is −0.800), except −1.1 in (*d*), which is −1.090 but was rounded for clarity in the figure.

Fig. 1



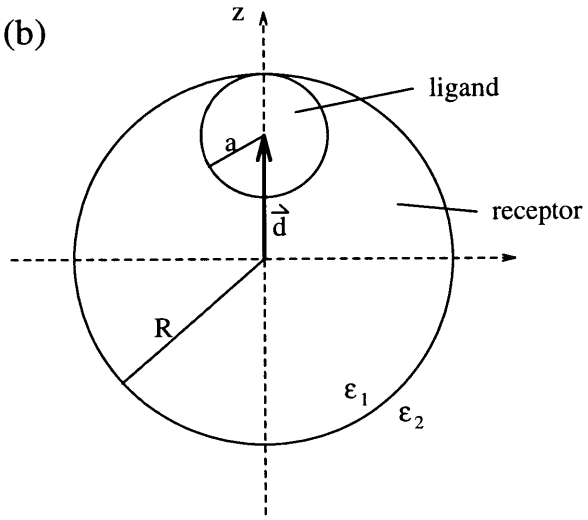


Fig. 2

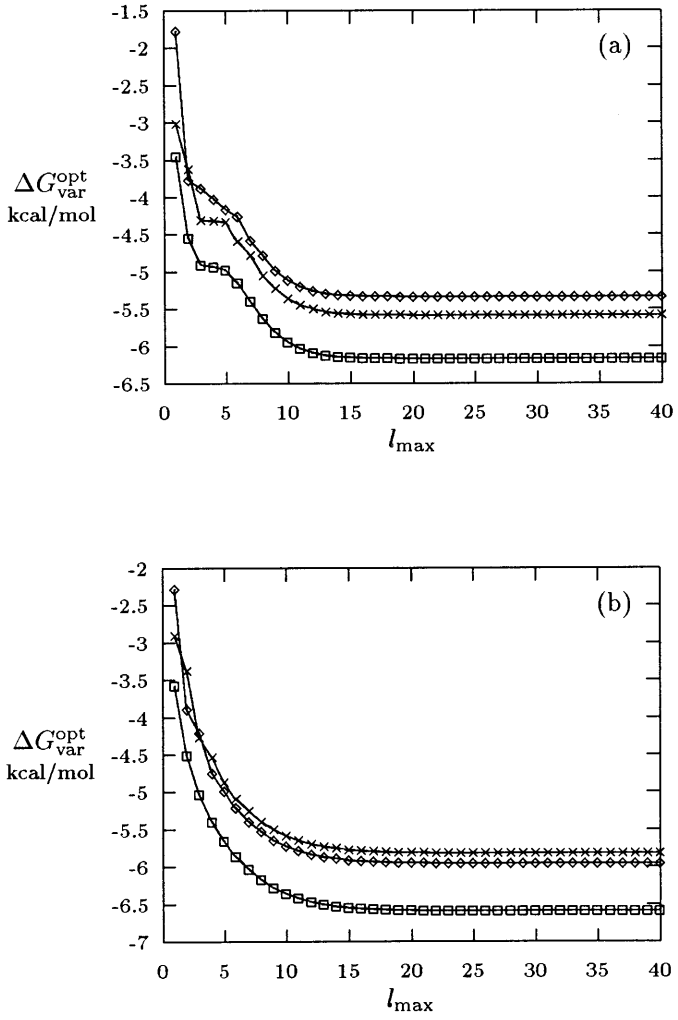

(a)
$\Delta G_{\mathrm{var}}^{\mathrm{opt}}$
kcal/mol
-1.5
-2
-2.5
-3
-3.5
-4
-4.5
-5
-5.5
-6
-6.5
0
5
10
15
20
25
30
35
40
$l_{\max}$
(b)
$\Delta G_{\mathrm{var}}^{\mathrm{opt}}$
kcal/mol
-2
-2.5
-3
-3.5
-4
-4.5
-5
-5.5
-6
-6.5
-7
0
5
10
15
20
25
30
35
40
$l_{\max}$

Fig. 3

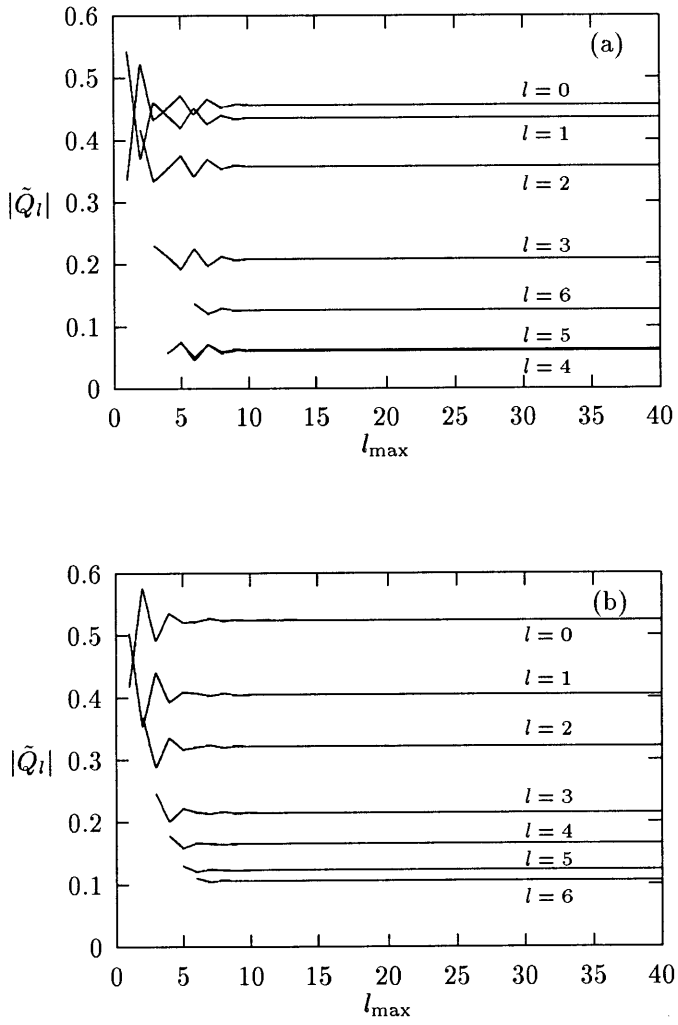

(a)
$|\tilde{Q}_l|$
$l = 0$
$l = 1$
$l = 2$
$l = 3$
$l = 6$
$l = 5$
$l = 4$
$l_{\max}$
(b)
$|\tilde{Q}_l|$
$l = 0$
$l = 1$
$l = 2$
$l = 3$
$l = 4$
$l = 5$
$l = 6$
$l_{\max}$

Fig. 4

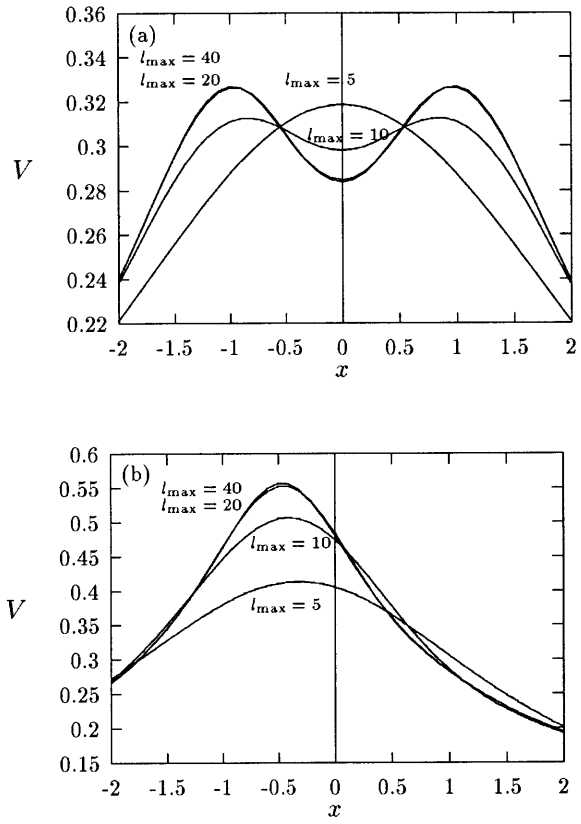
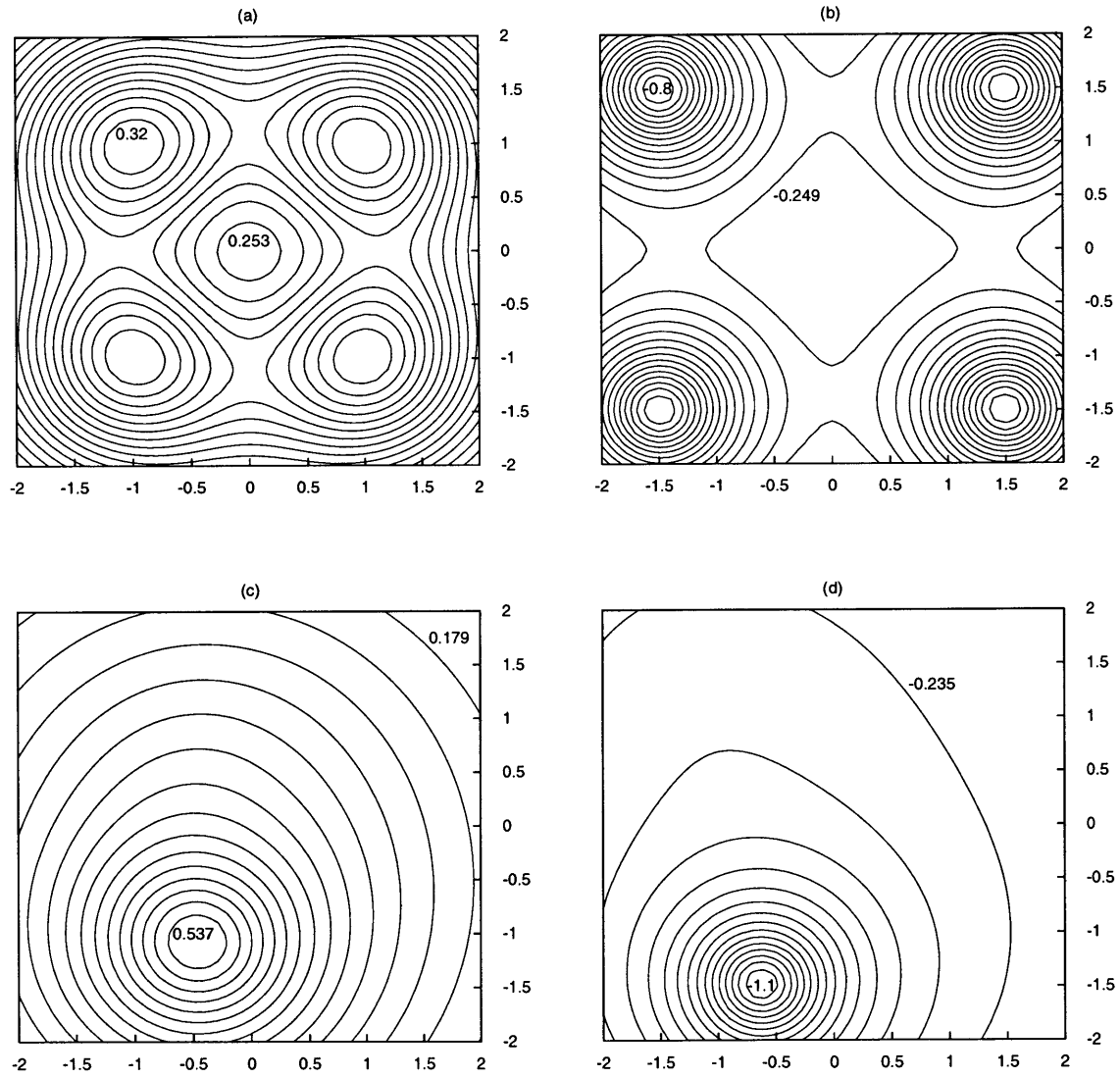

(a)
$l_{max} = 40$
$l_{max} = 20$
$l_{max} = 5$
$l_{max} = 10$
V
x
0.36
0.34
0.32
0.3
0.28
0.26
0.24
0.22
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
(b)
$l_{max} = 40$
$l_{max} = 20$
$l_{max} = 10$
$l_{max} = 5$
V
x
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15

Fig. 5


(a)
0.32
0.253
(b)
-0.8
-0.249
(c)
0.179
0.537
(d)
-0.235
-1.1
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2

Chapter IV

Optimization of Binding Electrostatics: Charge Complementarity in the Barnase–Barstar Protein Complex

Abstract

Theoretical and experimental studies have demonstrated that the large desolvation penalty required for polar and charged groups frequently precludes their involvement in strongly stabilizing electrostatic interactions in the folding or binding of proteins in aqueous solution near room temperature. We have previously developed a theoretical framework for computing optimized electrostatic interactions and illustrated use of the algorithm with simplified geometries. Given a receptor and model assumptions, the method computes the ligand-charge distribution that provides the most favorable balance of desolvation and interaction effects on binding. In this paper, the method has been extended to treat complexes using actual molecular shapes. The barnase–barstar protein complex was investigated with barnase treated as a target receptor. The atomic point charges of barstar were varied to optimize the electrostatic binding free energy. Barnase and natural barstar form a tight complex ($K_d \sim 10^{-14}$ M) with many charged and polar groups near the interface that make this a particularly relevent system for investigating electrostatic effects on binding. The results show that sets of barstar charges (resulting from optimization with different constraints) can be found that give rise to relatively large predicited

improvements in electrostatic binding free energy. Principles for enhancing the stabilizing effect of electrostatic interactions in molecular binding in aqueous environments are discussed in light of the optima. Our findings suggest that substantial enhancements in electrostatic binding free energy in general may result from modification of polar and charged groups that favorably affect the balance of desolvation and interaction effects. Moreover, a recently proposed definition of electrostatic complementarity is demonstrated to be a useful tool for examining binding interfaces. Finally, we found that the wild-type barstar is closer to being optimized than barnase is for their mutual binding, consistent with the perceived functions of these proteins.

I. INTRODUCTION

Due to the large desolvation penalty incurred by polar and charged groups upon protein folding or binding in aqueous solution, theoretical and experimental studies have found electrostatic interactions to be net destabilizing in a variety of contexts near room temperature.[1–12] These interactions form because the hydrophobic effect is strongly stabilizing, outweighing the electrostatic contribution that is computed to be unfavorable and often large. Calculations show that, while salt bridges and hydrogen bonds formed across binding interfaces contribute a favorable *interaction* term to binding free energy, the *desolvation* term due to the loss of electrostatic interaction of charged and polar groups with water and solvent ions upon binding generally more than offsets the favorable bound-state electrostatic interactions. Electrostatic interactions are believed to enhance specificity, however, due to the large penalty for burying but not compensating polar and charged groups.[13]

The unfavorable nature of protein electrostatics is intriguing. Is this a consequence of physics or a choice of biology? In other words, is it possible for two appropriate electrostatically designed molecular binding partners to have favorable intermolecular interactions that compensate and surpass the desolvation penalty? We have developed a theoretical framework for use with the continuum electrostatic model for electrostatic optimization; the charge distribution for a ligand is optimized to give the most favorable balance of these competing factors upon binding a given receptor.[14–16] Electrostatic optimization, which is applicable beyond simple continuum models to general linear response models, is of special interest for three reasons. First, the ability to obtain optimal charge distributions provides a useful point of reference for understanding natural charge distributions as well as the effects of mutations. Second, optimization theory leads to an important measure of electrostatic complementarity demostrated here.[15] Third, it is likely that optimal-ligand computations in general will be a powerful guide to enhancing affinity through mutations and the design of inhibitors.

In previous work, this electrostatic optimization scheme was applied to a range of model

receptors and complexes with idealized geometries of spheres and slabs,[14–16] including a spherical model for barnase.[16] Results have shown that when the ligand-charge distribution is optimized, the electrostatic binding free energy can be favorable with reasonable values for point-charge magnitudes.[15,16] In fact, a proof has been completed showing that the optimization produces favorable electrostatics of binding at zero ionic strength with a few simplifying assumptions.[17] Taken together, these observations suggest that electrostatic interactions in natural complexes are likely to be underoptimized.

We have extended the electrostatic optimization method to treat the actual molecular shapes of real proteins. Here we apply this optimization method to study the barnase–barstar complex, an extremely tight-binding pair with a K_d of 10^{-14} M.[18] Even so, the electrostatic binding free energy is calculated to be somewhat unfavorable (14 kcal/mol).[19] Most other complexes for which the calculations have been done also show even more unfavorable binding electrostatics, consistent with the observation that this is an especially tight complex. In this study, barnase was chosen to be the fixed receptor. Barstar atomic point charges were treated as variable and optimized to produce the net most favorable binding to barnase. The results show that sets of barstar charges (resulting from optimization with different constraints) can be found that give rise to computed improvements of 10 to 20 kcal/mol in electrostatic binding free energy.

In this paper we give a detailed description of our current electrostatic optimization method including molecular shape and investigate the nature of the optimal charge distributions. We extract principles for enhancing electrostatic interactions in molecular binding in aqueous solution. Moreover, we demonstrate the utility of our new measure of electrostatic complementarity,[15] and we also examine the evolutionary question of whether barnase (the enzyme) is more optimized to bind to barstar (the inhibitor) or whether the inhibitor is more optimized to bind the enzyme. In another paper we show that, although charge distributions that give much higher binding affinity than barstar charges can be found, with the constraint that limits barstar residues to the 20 common amino acids, barstar is actu-

ally very well optimized.[20] The implication of these findings for molecular design is exciting: since laboratory chemistry is not limited to amino acids, novel molecules might be designed with substantially improved affinity.

II. METHODS

In the continuum electrostatic model used here, a macromolecule is treated as a low dielectric ($\epsilon = 4$) cavity, defined by its molecular surface, embedded in a high dielectric ($\epsilon = 80$) continuum solvent at an ionic strength near physiological levels. The charge distribution of the macromolecule is a set of point partial charges, $\{Q_i\}$, at the atom centers, $\{\vec{r}_i\}$. In cases where the potential is derived from the Poisson or the linearized Poisson–Boltzmann (PB) equation, the electrostatic free energy is $G = \frac{1}{2}\sum Q_i\Phi(\vec{r}_i)$, where $\Phi(\vec{r}_i)$ is the potential at the position of the charge Q_i.[21] This free energy includes entropic contributions from the solvent, including ion reorganization in the case of the linearized Poisson–Boltzmann equation.[21–23] The factor of $\frac{1}{2}$ is due to the fact that the free energy arises from the interaction between the charges and their self-induced reaction field. In many cases of interest, $\Phi(\vec{r})$ is obtained by solving the linearized Poisson–Boltzmann (PB) equation,

$$\vec{\nabla}\cdot\left(\epsilon(\vec{r})\vec{\nabla}\Phi(\vec{r})\right) - \bar{\kappa}(\vec{r})^2\Phi(\vec{r}) = -4\pi\sum_j Q_j\delta(\vec{r}_j) \tag{1}$$

where $\kappa(\vec{r}) = 8\pi e^2 I(\vec{r})/kT$ with $I(\vec{r})$ the bulk ionic strength, e the unit charge (magnitude of the charge of an electron), k the Boltzmann constant, T the absolute temperature, and $\epsilon(\vec{r})$ the spatially dependent dielectric determined by the molecular surface formed by rolling a probe water molecule of radius 1.4 Å over the macromolecule. In cases where the molecular boundary is a simple shape (e.g., spheres, infinite cylinders, and slabs) analytical solutions exist; otherwise the linearized PB equation is solved numerically. Here we used the actual molecular shapes of barnase and barstar, and numerically computed $\Phi(\vec{r})$ from $\{Q_i\}$, $\{\vec{r}_i\}$, $\kappa(\vec{r})$, and $\epsilon(\vec{r})$. The radii and charges were taken from the CHARMM PARAM19 parameter set.[24] A 2.0-Å Stern layer surrounded the molecule and represents an

ion–exclusion region;[25,26] the salt concentration was 0.145 M. Each calculation was done using "focussing", in which a low grid spacing calculation (using 23% fill and Debye-Huckel boundary conditions[27]) was done to determine the potential at the grid boundary for a higher grid spacing calculation (using 92% fill). Ten translations relative to the grid (65 points in each cartesian direction) were done and the average was used.

The electrostatic free energy of binding is the difference between the electrostatic free energy in the bound and the unbound state, $\Delta G_{\text{bind}} = G^{\text{bound}} - G^{\text{unbound}}$. Because the unbound and bound structures are very similar,[28–31] we use a rigid binding model in which the bound state is taken from the X-ray crystal structure of the complex (C) and the unbound state is simply extracted from the complex of the ligand (L, barstar) and the receptor (R, barnase) (Figure 1). A cap of 12 water molecules from the complex that appears tightly associated with barnase is treated as part of the receptor in the bound and unbound states. A detailed description of the preparation of this structure, including the building of hydrogen atom positions, is found in ref. 19.

The free energy of the unbound state is given by,

$$G^{\text{unbound}} = \frac{1}{2}\sum_{i=1}^{n_{\text{L}}} Q_i^{\text{L}}\Phi^{\text{L}}(\vec{r}_i^{\text{L}}) + \frac{1}{2}\sum_{i=1}^{n_{\text{R}}} Q_i^{\text{R}}\Phi^{\text{R}}(\vec{r}_i^{\text{R}}) \tag{2}$$

where $\Phi^{\text{L}}(\vec{r})$ is the potential due to ligand charges in the unbound state, $\Phi^{\text{R}}(\vec{r})$ is the potential due to receptor charges in the unbound state, and n_{L} and n_{R} are the number of partial atomic charges in the ligand and receptor, respectively. The free energy of the bound state is given by,

$$G^{\text{bound}} = \frac{1}{2}\sum_{i=1}^{n_{\text{L}}} Q_i^{\text{L}}\Phi^{\text{C}}(\vec{r}_i^{\text{L}}) + \frac{1}{2}\sum_{i=1}^{n_{\text{R}}} Q_i^{\text{R}}\Phi^{\text{C}}(\vec{r}_i^{\text{R}}) \tag{3}$$

where $\Phi^{\text{C}}(\vec{r})$ is the potential due to both ligand and receptor charges in the complex. Because of the linearity of Eq. (1), the superposition principle holds and we can write $\Phi^{\text{C}}(\vec{r})$ as a sum of $\Phi^{\text{C}_{\text{L}}}(\vec{r})$, the potential due to ligand charges in the complex, and $\Phi^{\text{C}_{\text{R}}}(\vec{r})$, the potential due to receptor charges in the complex:

$$G^{\text{bound}} = \frac{1}{2}\sum_{i=1}^{n_{\text{L}}} Q_i^{\text{L}}\left(\Phi^{\text{C}_{\text{L}}}(\vec{r}_i^{\text{L}}) + \Phi^{\text{C}_{\text{R}}}(\vec{r}_i^{\text{L}})\right) + \frac{1}{2}\sum_{i=1}^{n_{\text{R}}} Q_i^{\text{R}}\left(\Phi^{\text{C}_{\text{L}}}(\vec{r}_i^{\text{R}_{\text{i}}}) + \Phi^{\text{C}_{\text{R}}}(\vec{r}_i^{\text{R}})\right) \tag{4}$$

We may rearrange these terms to obtain

$$\Delta G_{\text{bind}} = \underbrace{\frac{1}{2}\sum_{i=1}^{n_{\text{L}}} Q_i^{\text{L}} \left(\Phi^{\text{C}_{\text{L}}}(\vec{r}_i^{\,\text{L}}) - \Phi^{\text{L}}(\vec{r}_i^{\,\text{L}})\right)}_{\Delta G_{\text{desolv,L}}} + \underbrace{\frac{1}{2}\left(\sum_{i=1}^{n_{\text{L}}} Q_i^{\text{L}} \Phi^{\text{C}_{\text{R}}}(\vec{r}_i^{\,\text{L}}) + \sum_{i=1}^{n_{\text{R}}} Q_i^{\text{R}} \Phi^{\text{C}_{\text{L}}}(\vec{r}_i^{\,\text{R}})\right)}_{\Delta G_{\text{inter,L-R}}}$$
$$+ \underbrace{\frac{1}{2}\sum_{i=1}^{n_{\text{R}}} Q_i^{\text{R}} \left(\Phi^{\text{C}_{\text{R}}}(\vec{r}_i^{\,\text{R}}) - \Phi^{\text{R}}(\vec{r}_i^{\,\text{R}})\right)}_{\Delta G_{\text{desolv,R}}} \tag{5}$$

In electrostatic optimization, the ligand-charge distribution is varied to optimize ΔG_{bind}; here this is achieved by varying the individual Q_i^{L} values. The potentials $\Phi^{\text{C}_{\text{R}}}(\vec{r})$ and $\Phi^{\text{R}}(\vec{r})$, due to the receptor charges, are computed numerically with a modified version of the computer program DELPHI.[21,32,33] $\Phi^{\text{C}_{\text{L}}}(\vec{r})$ and $\Phi^{\text{L}}(\vec{r})$, due to the unknown ligand charges can be expressed in terms of $\{Q_i^{\text{L}}\}$ and a set of potentials calculated numerically, as we now show. Because of superposition, we may decompose $\Phi(\vec{r})$ as

$$\Phi(\vec{r}) = \sum_j Q_j \hat{\Phi}_j(\vec{r}) \tag{6}$$

where $\hat{\Phi}_j(\vec{r})$ satisfies

$$\vec{\nabla} \cdot \left(\epsilon(\vec{r})\vec{\nabla}\hat{\Phi}_j(\vec{r})\right) - \bar{\kappa}(\vec{r})^2 \hat{\Phi}_j(\vec{r}) = -4\pi\delta(\vec{r}_j) \tag{7}$$

$\hat{\Phi}_j$ is then the potential due to a unit charge at the position $\vec{r}_j$ (i.e., a Green function). We can now write the desolvation penalty of the ligand as

$$\Delta G_{\text{desolv,L}} = \frac{1}{2}\sum_{i=1}^{n_{\text{L}}}\sum_{j=1}^{n_{\text{L}}} Q_i^{\text{L}} Q_j^{\text{L}} \left(\hat{\Phi}_j^{\text{C}_{\text{L}}}(\vec{r}_i^{\,\text{L}}) - \hat{\Phi}_j^{\text{L}}(\vec{r}_i^{\,\text{L}})\right) \tag{8}$$

Furthermore, because of the principle of reciprocity, the two terms in the intermolecular interaction free energy in Eq. (5) are equal and we have

$$\Delta G_{\text{inter,L-R}} = \sum_{i=1}^{n_{\text{L}}} Q_i^{\text{L}} \Phi^{\text{C}_{\text{R}}}(\vec{r}_i^{\,\text{L}}) \tag{9}$$

We have thus expressed ΔG_{bind} as a quadratic function of $\{Q_i^{\text{L}}\}$, and with vector notation,

$$\Delta G_{\text{bind}} = \frac{1}{2}\left(\vec{Q}^{\text{L}}\right)^T \overleftrightarrow{A} \left(\vec{Q}^{\text{L}}\right) + \left(\vec{Q}^{\text{L}}\right)^T \vec{B} + C \tag{10}$$

where

$$A_{ij} = \hat{\Phi}_j^{\mathrm{C_L}}(\vec{r}_i^{\,\mathrm{L}}) - \hat{\Phi}_j^{\mathrm{L}}(\vec{r}_i^{\,\mathrm{L}}) \tag{11}$$

$$B_i = \Phi^{\mathrm{C_R}}(\vec{r}_i^{\,\mathrm{L}}) \tag{12}$$

and

$$C = \frac{1}{2}\sum_{i=1}^{n_{\mathrm{R}}} Q_i^{\mathrm{R}} \left(\Phi^{\mathrm{C_R}}(\vec{r}_i^{\,\mathrm{R}}) - \Phi^{\mathrm{R}}(\vec{r}_i^{\,\mathrm{R}})\right) \tag{13}$$

Eq. (10) describes a paraboloid possessing a minimum corresponding to the optimized electrostatic binding free energy.[14]

Allowing all ligand charges to change is a special case of the more general problem of having a subset of ligand charges fixed and the remainder variable. For instance, in this study we fix the backbone charges, and in the accompanying study (Chapter V)[20] we also fix all charges except for a single side chain. Let $\{Q_i^{\mathrm{Lv}}\}$ denote the set of variable ligand charges and $\{Q_i^{\mathrm{Lf}}\}$ denote the set of fixed ligand charges. It is straightforward to show that ΔG_{bind} can be written as a quadratic function of $\{Q_i^{\mathrm{Lv}}\}$,

$$\Delta G_{\mathrm{bind}} = \frac{1}{2}\left(\vec{Q}^{\mathrm{Lv}}\right)^T \overleftrightarrow{A} \left(\vec{Q}^{\mathrm{Lv}}\right) + \left(\vec{Q}^{\mathrm{Lv}}\right)^T \vec{B} + C \tag{14}$$

where

$$A_{ij} = \hat{\Phi}_j^{\mathrm{C_L}}(\vec{r}_i^{\,\mathrm{Lv}}) - \hat{\Phi}_j^{\mathrm{L}}(\vec{r}_i^{\,\mathrm{Lv}}) \tag{15}$$

$$B_i = \Phi^{\mathrm{C_R}}(\vec{r}_i^{\,\mathrm{Lv}}) + \sum_{j=1}^{n_{\mathrm{Lf}}} \left(\hat{\Phi}_j^{\mathrm{C_L}}(\vec{r}_i^{\,\mathrm{Lv}}) - \hat{\Phi}_j^{\mathrm{L}}(\vec{r}_i^{\,\mathrm{Lv}})\right) Q_j^{\mathrm{Lf}} \tag{16}$$

and

$$C = \frac{1}{2}\sum_{i=1}^{n_{\mathrm{R}}} Q_i^{\mathrm{R}} \left(\Phi^{\mathrm{C_R}}(\vec{r}_i^{\,\mathrm{R}}) - \Phi^{\mathrm{R}}(\vec{r}_i^{\,\mathrm{R}})\right) + \frac{1}{2}\sum_{i=1}^{n_{\mathrm{Lf}}} Q_i^{\mathrm{Lf}} \left(\Phi^{\mathrm{C_L}}(\vec{r}_i^{\,\mathrm{Lf}}) - \Phi^{\mathrm{L}}(\vec{r}_i^{\,\mathrm{Lf}})\right) + \frac{1}{2}\sum_{i=1}^{n_{\mathrm{Lf}}} Q_i^{\mathrm{Lf}} \Phi^{\mathrm{C_R}}(\vec{r}_i^{\,\mathrm{Lf}}) \tag{17}$$

Equations (15), (16), and (17) reduce to equations (11), (12), and (13) when $\{Q_i^{\mathrm{Lv}}\}$ is $\{Q_i^{\mathrm{L}}\}$ and $\{Q_i^{\mathrm{Lf}}\}$ is an empty set.

In theory, the ligand desolvation matrix, $\overleftrightarrow{A}$, is symmetric (due to reciprocity) and positive definite,[15] which means its eigenvalues are all positive and $\Delta G_{\text{desolv,L}}$ is positive (a desolvation penalty). ΔG_{bind} is a multidimensional paraboloid with a minimum, $\Delta G_{\text{bind}}^{\text{opt}}$, at

$$\vec{Q}^{\text{opt}} = -\overleftrightarrow{A}^{-1}\vec{B}. \tag{18}$$

In practice, due to numerical errors inherent in the calculations, it is likely that $\overleftrightarrow{A}$ as computed is non-symmetric; in such cases we symmetrize the matrix by averaging across the diagonal. The more problematic consequence of limited numerical accuracy is that $\overleftrightarrow{A}$ as computed might not be positive definite, which means the extremum computed is not a minimum but a saddle point. The size of the negative eigenvalues relative to the positive ones is some indication of the extent of numerical inaccuracy. Here, there are negative eigenvalues that are quite small, the largest magnitude being 10^{-5} of the largest positive eigenvalue. Our remedy for this numerical problem is to use singular value decomposition,[34,35] i.e., we construct the optimal ligand charges only from eigenvectors with eigenvalues greater than the absolute value of the most negative eigenvalue, indicated by $\sum'$,

$$\vec{Q}^{\text{opt}} = \sum_{i=1}^{n_{\text{Lv}}}{}' \beta_i \vec{v}_i \tag{19}$$

where $\vec{\beta}$ is a function of $\overleftrightarrow{A}$ and $\vec{B}$ (see Chapter Appendix for details).

It is often desirable to impose constraints on the variable charges $\{Q_i^{\text{Lv}}\}$ when we carry out the optimization. For instance, the size of the charges should be constrained to be chemically reasonable. In this study, we limit individual charge values to be between -0.85 and 0.85 (units are in magnitude of an electron charge, e). We also constrain the total ligand charge and the total charge of each side chain. These constraints are all linear in $\{Q_i^{\text{Lv}}\}$, and since our objective function (i.e., the function to be optimized), ΔG_{bind}, is quadratic in $\{Q_i^{\text{Lv}}\}$, the optimization problem is a quadratic programming problem, which has been extensively studied by the operations reseach community. We use the computer program package LOQO[36] to solve this constrained optimization problem. If $\overleftrightarrow{A}$ were positive definite,

we would just use Eq. (14) as the objective function with the various linear constraints. However, with the removal of some eigenvectors, we have to rewrite the objective function. The variable charges $\{Q_i^{\mathrm{Lv}}\}$ are a linear combination of the eigenvectors that remain,

$$\vec{Q}_{\mathrm{Lv}} = \sum_{i=1}^{n_{\mathrm{Lv}}}{}' c_i \beta_i \vec{v}_i \tag{20}$$

When there are no constraints, all the c_i's are 1 for the optimal charge set $\{Q_i^{\mathrm{opt}}\}$. When contraints are imposed, the optimum achievable is worse than $\{Q_i^{\mathrm{opt}}\}$ and the free energy deviation of $\vec{Q}$ from the unconstrained $\Delta G_{\mathrm{bind}}^{\mathrm{opt}}$ is[14]

$$\Delta\Delta G = \frac{1}{2}\left(\vec{Q} - \vec{Q}^{\mathrm{opt}}\right)^T \overleftrightarrow{A} \left(\vec{Q} - \vec{Q}^{\mathrm{opt}}\right) \tag{21}$$

$$= \frac{1}{2}\sum_{i=1}^{n_{\mathrm{Lv}}}{}' (1 - c_i)\beta_i \vec{v}_i^T \overleftrightarrow{A} (1 - c_i)\beta_i \vec{v}_i \tag{22}$$

$$= \frac{1}{2}\sum_{i=1}^{n_{\mathrm{Lv}}}{}' (1 - c_i)^2 E_i \tag{23}$$

where

$$E_i = \beta_i \vec{v}_i^T \overleftrightarrow{A} \beta_i \vec{v}_i \tag{24}$$

$\Delta\Delta G$ is now our objective function. Upon optimization, a set of $\{c_i^{\mathrm{opt}}\}$ will be found such that $\Delta\Delta G$ is minimized and the optimum charge distribution is

$$\vec{Q}^{\mathrm{opt}} = \sum_{i=1}^{n_{\mathrm{Lv}}}{}' c_i^{\mathrm{opt}} \beta_i \vec{v}_i \tag{25}$$

III. RESULTS AND DISCUSSION

Optimal charge distributions have enhanced binding electrostatics. The electrostatic optimization procedure was applied to select sets of barstar atomic charges. Atomic positions and the dielectric boundary for barstar were taken from the X-ray crystal structure of the complex as were the geometry and point-charge distribution for barnase. Optimizations were carried out with different sets of constraints on the atomic point charges, and the results are listed in Table I. When actual (unoptimized) barstar point charges were used

($\{Q_i^{wt}\}$), the electrostatic binding free energy was computed to be unfavorable by about 14 kcal/mol. This reflects the observation made in a number of studies that most complexes pay more in electrostatic dehydration penalty than is recovered in attractive electrostatic interactions across the interface.[1–5,37,6–12] The relatively small electrostatic free energy effect (+14 kcal/mol) for the relatively large amount of surface area buried in the complex (803 and 878 $Å^2$ buried by barnase and barstar, respectively) is consistent with this being a tight-binding pair. The hydrophobic effect is sufficient to overcome the unfavorable electrostatics and configurational, translational, and rotational entropy and drive binding.

In all optimizations the value of each barstar atomic point charge was restricted to the range [−0.85,0.85]. When this was the only constraint applied and all 839 barstar atomic charges in our model were optimized, the resulting charge distribution ($\{Q_i^{opt1}\}$), had an electrostatic binding free energy of −6.1 kcal/mol. This value is favorable and represents a computed enhancement of about 20 kcal/mol over wild type. When the backbone charges were fixed at their standard values and only the 403 side chain atomic point charges were allowed to vary, the computed optimum $\{Q_i^{opt2}\}$ had an electrostatic binding free energy of −0.6 kcal/mol, which is 5.5 kcal/mol less stable than $\{Q_i^{opt1}\}$ but is still favorable. When the further constraint was added that the total charge magnitude of each side chain be less than or equal to 1.5, the resulting charge distribution ($\{Q_i^{opt3}\}$) had an electrostatic binding free energy of −0.1 kcal/mol. These results show that charge optimization produces atomic charge distributions with reasonable charge magnitudes and apparently extremely tight-binding. Moreover, electrostatics can, in principle, have a net favorable effect on binding.[17] It is of interest to examine these charge distributions in greater detail to understand what aspects of them are important and by what mechanisms such electrostatic enhancements to binding are achieved.

Optimization produces enhanced intra- and intermolecular effects on binding. Of fundamental interest are the mechanisms used by optimized ligand-charge distributions to improve the balance of desolvation and interaction effects on binding. To examine this,

the linearity of the continuum treatment was used to dissect the energetics. There are three types of electrostatic terms that contribute to binding here: solvation, intramolecular (or indirect), and intermolecular (or direct) interactions.[12,19,16,38] Solvation represents the loss in electrostatic interaction of a functional group with solvent upon binding, and intermolecular (or direct) terms represent screened coulombic interactions between a functional group and the binding partner in the bound state. Intramolecular (or indirect) terms represent electrostatic interactions between a functional group and other functional groups *within the same binding partner.* Even if the geometry of the interaction is unaltered on binding, an effect on the free energy of binding can be produced by differences in the amount of solvent screening in the bound and unbound states. Generally the effect is to strengthen intramolecular interactions on binding (be they favorable or unfavorable), and the effects are largest when at least one of the functional groups is at the binding interface. It is a striking result that these enhanced intramolecular interactions are computed to be significant in many natural systems, and there is some experimental evidence for their importance as well (see ref. 12).

The free energetic analysis for individual side chains of wild-type barstar and the side-chain optimized charge distribution (opt3) binding to barnase is shown in Tables II and III, respectively. The totals indicate that the optimum achieves binding enhancements through a large dehydration penalty (ΔG_{solv} plus ΔG_{intra}) that is more than offset through a significantly larger improvement in intermolecular electrostatic interactions (ΔG_{inter}). Side chains making moderate and strong intermolecular as well as those making enhanced intramolecular interactions (greater than one kcal/mol) are indicated in the Tables and in Figure 2. The optimal charge distribution utilizes at least three mechanisms to improve binding. There are enhanced intermolecular (direct) interactions (residues 33, 34, 37, 38, 42, 44, 45, 72, 74, 76, and 77), reduced desolvation (35 and 39) and new favorable intramolecular (indirect) interactions (residues 41, 47, and 82). Strong direct intermolecular interactions were achieved by the optimum through the involvement of more side chains at the interface

as well as through a small number of more distant electrostatic interactions. Side chains with enhanced intramolecular interactions generally lie just below the binding interface and make attractive electrostatic interactions with side chains at the interface; these interactions are largely screened by solvent in the unbound state but are enhanced on binding. Coupling of effects is present. For instance, a number of residues have worse intramolecular binding effects that are more than offset by improvements in desolvation or interaction.

Inventory of hydrogen-bonded contacts and ion pairs is only slightly increased through optimization. Examination of individual interactions across the interface for barstar and the $\{Q_i^{\mathrm{opt3}}\}$ charge distribution reveals very similar types of interactions made in both (Tables IV and V). Decreased polarity causes the optimum to remove somewhat distant hydrogen bonds of Asn33 and Trp38 present in barstar. Increased polarity results in a strengthening of interactions in which a barstar side chain provides the hydrogen-bond acceptor. Finally, two new hydrogen-bond type of interactions are added. One is between Arg59 of barnase and Glu76 of barstar through redistributing charge such that all the atoms of the carboxylate bear a partial negative charge. The other is at most a weak hydrogen bond, between Wat29 (an interfacial water molecule) and Trp38 of barstar. Three ion pairing interactions remain similar in opt3, but are somewhat overpolarized relative to the wild type. Additionally, one new ion pair is added, involving Arg59 of barnase and negative charge density added to Gln72 of barstar. The enhanced binding affinity of the optimum relative to the wild type cannot simply be attributed to increases in the inventory of intermolecular interactions. The overpolarization and the new hydrogen bonds are estimated to contribute -5.8 kcal/mol and the new ion pair -1.4 kcal/mol. This is roughly half of the binding free energy improvement of opt3 over the wild type; the remaining -7.1 kcal/mol comes from interactions from other side chains, which are not involved in short-range electrostatic interactions across the interface. Taken together, these results show that simple inventories of short-range interactions present in the bound state neglect important contributions to binding energetics that can be responsible for many

orders of magnitude in a binding constant.

Optimized binding electrostatics are not directly sensitive to total charge or distant functional groups. The nominal charge on barnase is +1 and barstar is −6, assuming unshifted pK_a's at neutral pH. There is a general belief that complementarity is enhanced for cases of oppositely charged binding partners, even though a number of natural counterexamples exist.[27,39–42] Here we show that the electrostatic binding free energy is insensitive to the total molecular charge across a relatively broad range for the case of barstar, in which there are many solvent exposed side chains distant from the binding interface. Table VI gives the results of 3 further optimizations in which side chain atomic charges were constrained to be in the range of [−0.85,0.85], backbone charges were fixed at their ordinary values, and the total molecular charge of the optimum was constrained to be −10, 0, or +10. The electrostatic binding free energy was essentially the same value for each of these cases as that obtained when the total molecular charge was unconstrained ($\{Q_i^{\mathrm{opt2}}\}$). The total molecular charge constraint was met by distributing extra charge to locations away from the interface and that had little effect on binding.

It can be shown that the binding free energy difference between an optimal and non-optimal charge distribution binding to the same receptor with the same geometry is given by,

$$\Delta\Delta G = \frac{1}{2}(\vec{Q} - \vec{Q}^{\mathrm{opt}})^T \overleftrightarrow{A} (\vec{Q} - \vec{Q}^{\mathrm{opt}}) \tag{26}$$

where $\vec{Q}$ is the non-optimal and $\vec{Q}^{\mathrm{opt}}$ is the optimal charge distribution.[14,15] If all point charges but one, Q_k, are optimal, the binding free energy difference is,

$$\Delta\Delta G_k = A_{kk}(Q_k - Q_k^{\mathrm{opt}})^2 \tag{27}$$

and A_{kk} can be viewed as one measure of the sensitivity of the binding free energy to deviations from optimal for charge Q_k. It should be noted that A_{kk} is a diagonal term in the desolvation matrix and will be positive definite, largest for ligand point charges near

the binding interface, and smaller for point charge locations away from the interface. In actual practice multiple point-charge magnitudes will be non-optimal and coupling between point charges will be affected by the entire row or column containing A_{kk}. Such off diagonal terms are especially small when Q_k is solvent exposed in the unbound and bound ligand so its electrostatic interactions are largely screened by solvent. To summarize, large values of A_{kk} are expected to indicate the importance of optimizing partial charge Q_k; small values of A_{kk} may not signal the non-importance of optimizing Q_k due to context-dependent effects, but these effects should be minimal for solvent exposed ligand point charges away from the binding site.

Figure 3A shows a ribbon drawing of barnase and a C_α trace of the ligand barstar, in which the radius of each barstar C_α atom represents the average A_{kk} for that residue's side-chain atoms. The largest values of A_{kk} are at the binding interface. Figure 3B is analogous, but the radius of each C_α represents the r.m.s difference between $\{Q_i^{\mathrm{opt2}}\}$ and $\{Q_i^{\mathrm{opt2a}}\}$ (the unconstrained optimization and that constrained to a total charge of -10) for each side chain. The results show that the constraint was satisfied with very little change to the locations with large A_{kk} and the biggest changes generally at locations with very small A_{kk} values. As a further illustration of this principle, when all point charges in the barstar ligand were fixed to their wild-type value and only side-chain point-charge locations with $A_{kk} \geq 0.1$ (138 atomic locations) were optimized in the range $[-0.85,0.85]$, giving $\{Q_i^{\mathrm{opt2r}}\}$, the electrostatic binding free energy is only 1.6 kcal/mol worse than $\{Q_i^{\mathrm{opt2}}\}$. Figure 4 emphasizes that atoms with small values of A_{kk} tend to be the least restricted through optimization. This is consistent with other studies that show that exposed charges away from the binding interface tend to have negligible effects on the binding free energy.

Improvements through optimization are not parameter restricted. One concern in applying a mathematical optimization procedure such as that employed here is that the optimization may numerically fine-tune the interactions beyond our confidence in the parameters of the model. To test this, we computed electrostatic binding free energies for barstar

binding to barnase using three different charge and radius parameter sets and compared this to binding of the opt1 optimized charges (optimized with CHARMM PARAM19). The results show binding enhancements of –20.3, –17.2, –15.1, and –10.6 kcal/mol with CHARMM PARAM19,[24] OPLS (with 0.0 and 1.25 Å for hydrogen atom radius; all other atomic radii were taken as $2^{-5/6}\sigma$)[43] and PARSE,[44] respectively. That a common optimized charge distribution produced large binding enhancements when used in the context of other parameter sets argues that much of the binding advantage is independent of numerical optimization within a single set of parameters.

The use of complementary electrostatic potentials to assess binding interfaces. A useful graphical representation of the electrostatic properties of a molecule is a plot of the screened coulombic potential at the molecular surface. Such representations are very common in reports of new macromolecular structures and can be conveniently computed and displayed with software packages such as GRASP.[45] The images show the overall effect of the pattern of polar and charged residues at the surface. Figures 5A and B show this representation for the ligand barstar and also for the receptor barnase projected onto the barstar surface for ease of comparison. While the surfaces are generally complementary — barstar is predominantly blue and barnase is red — the detailed distributions do not corrrespond well, particularly given that this is such a tight-binding complex.

However, unbound-state screened coulombic potentials could be unable to give a qualitative representation of electrostatic binding complementarity because they don't explicitly consider the desolvation penalty and the intermolecular interactions recovered in the bound state. We have previously proposed a measure of electrostatic complementarity based on electrostatic optimization theory;[15] the optimization condition can be expressed as,

$$\overleftrightarrow{A}\,\vec{Q}^{\,\mathrm{opt}} + \vec{B} = 0 \tag{28}$$

which can be obtained from Eq. (18). It represents a pair of potentials, one due to the ligand-charge distribution ($\overleftrightarrow{A}\,\vec{Q}^{\,\mathrm{L}}$) and one due to the receptor ($\vec{B}$) that are equal in magnitude and

opposite in sign throughout the ligand volume at an optimum. The first term is called the ligand desolvation potential and represents the bound- minus the unbound-state potential due only to the ligand-charge distribution; the second term is called the receptor interaction potential and represents the bound-state potential due to the receptor-charge distribution. These potentials are plotted at the barstar molecular surface in Figure 5C and D for actual barnase and barstar (not one of the computed optima). The potentials are remarkably complementary, with the pattern and intensity of blue and red being a near quantitative match between the pair. Moreover, the most intense features of the figures are due to residues implicated experimentally as being especially important to binding.[46] For example, the two most intense red regions in Figure 5C are from Asp35 and Asp39 of barstar, which are buried at the binding interface and make a series of interactios with positively charged side chains from barnase that are largely responsible for the corresponding intense blue region in Figure 5D. These observations support the proposal that these potentials represent a useful measure of electrostatic complementarity. Scripts for use with the GRASP[45] program to compute and display these potentials are available from http://mit.edu/tidor.

Relative optimization. An interesting feature of the potentials described above for assessing electrostatic complementarity is that they are not symmetric. For instance, a difference of potentials (bound minus unbound) is plotted for the ligand, whereas simply a single potential is plotted for the receptor. Rather than being awkward, this reflects the fact that different improvements are expected when the enzyme is fixed and the inhibitor is optimized than when the inhibitor is fixed and the enzyme is optimized. Said another way, the enzyme may be further from its optimum than is the ligand. In principle such a situation could arise by chance, or it could be the result of different evolutionary pressures acting on each of the binding partners. For the purpose of illustration, here we describe a relative optimization analysis for barnase and barstar. Two measures were computed to judge which binding partner is closer to its optimum (or in other words, is more complementary to its partner). The first measure is simply to reverse the role of "ligand" and "receptor" and recompute

the complementary potentials described above. These are shown in Figures 5E and F and may be compared with Figures 5C and D. While both sets are quite complementary, parts C and D are somewhat more complementary than parts E and F. This can be confirmed numerically (the correlation coefficient computed over the ligand atom centers is -0.97 for the former pair and -0.92 for the latter; perfect complementarity would be -1.00). The second measure involves substantially more computation but is more accurate. The charge distribution of barstar was fixed while that of barnase was optimized; comparison of energetic improvements was made to calculations carried out with barnase fixed and barstar optimized. The results for a number of optimization procedures are given in Table VII. In each case, optimization of barnase leads to greater binding enhancements than a similar optimization carried out for barstar. Because barstar has fewer atom centers than barnase in the model, care was taken to compare optimizations with the same number of variable charges. Both measures suggest that barstar, the inhibitor, is closer to being optimized for binding barnase than the other way around. This result is consistent with the known functions of these proteins. Barstar is an inhibitor for barnase and its primary function is presumably to fold and bind to barnase; it is reasonable to expect it to optimize binding to barnase. On the other hand, barnase is an enzyme with its own catalytic role; optimal binding to barstar is expected to be less of a priority. It should be noted that there must be many other constraints and considerations on the role of these molecules, many of which we may be unaware. Because of the need to rapidly and efficiently inhibit barnase, further optimization of the kinetics of folding and binding of barstar are also logical.[47–51]

IV. CONCLUSION

Electrostatic optimization methods have been extended to deal efficiently with detailed molecular geometries through numerical computation of the necessary free energetic terms. Application to the barnase–barstar complex has revealed the predominance of a relatively few barstar residues at the interface whose electrostatic charge distributions are computed

to produce especially large effects on the binding free energy. Experiments involving mutational analysis have implicated many of the same residues in binding. Analysis of optimized charge distributions suggests three general mechanisms for enhancing the net electrostatic effect on binding. First, improvements to the hydrogen-bond inventory across the interface are valuable. This includes not only increasing the number of interactions through the introduction of appropriately placed hydrogen-bond donors and acceptors, but also the elimination of ligand interfacial polar groups that are inappropriately placed to make good-geometry intermolecular interactions. Second, the enhancement of intramolecular electrostatic effects on binding through changes in solvent screening appears repeatedly in natural complexes as well as in optima.[16,12,38,19] This frequently takes the form of introducing polar or charged ligand groups that interact favorably with interfacial ligand electrostatic groups. Solvent screening in the unbound state attenuates the interaction, which is correspondingly larger in the bound state due to displacement of high-dielectric solvent with low-dielectric protein. This effect, whose importance has not generally been recognized, suggests that attention must be focused intra- as well as intermolecularly in the analysis and design of binding partners. Third, tuning of particular electrostatic groups can have significant effects on the overall electrostatic contribution to binding. While the details depend on the geometry and charge distributions in the bound and unbound states, it is not yet possible to see whether simple rules can be used to predict the appropriate polarity to place at a site without carrying out a full calculation. The clear implication for combinatorial library design is that a range of polarity should be sampled at electrostatically critical regions of binding pockets. Further studies of electrostatic optimization, both theoretical and experimental, will likely lead to both a focussing of these general guidelines as well as to a catalog of particular functional-group motifs that are effective in enhancing electrostatic binding contributions in a broad class of circumstances. A procedure for illustrating and evaluating electrostatic complementarity is a direct result of charge-optimization theory.[14,15] This new measure of charge complementarity clearly reveals compensatory electrostatics at binding

interfaces, which had generally not been readily apparent using previous methods. Scripts for computing and displaying this new measure with the GRASP[45] computer program are available at http://mit.edu/tidor. Our calculational analysis has shown that barstar is more electrostatically optimized than barnase for binding in this complex, consistent with barnase having at least one other major function, namely to catalyze the hydrolysis of RNA.

V. APPENDIX

$\overleftrightarrow{A}$ can be written as $\overleftrightarrow{V}\overleftrightarrow{\Lambda}\overleftrightarrow{V}^{-1}$ where the columns of $\overleftrightarrow{V}$ are the n_{Lv} independent eigenvectors $\{\vec{v}_i\}$ of $\overleftrightarrow{A}$ and $\overleftrightarrow{\Lambda}$ is a diagonal matrix with the corresponding eigenvalues $\{\lambda_i\}$ as its diagonal entries. Then

$$\vec{Q}^{\mathrm{opt}} = (\overleftrightarrow{V}\overleftrightarrow{\Lambda}\overleftrightarrow{V}^{-1})^{-1}\vec{B} \tag{29}$$

$$\vec{Q}^{\mathrm{opt}} = \overleftrightarrow{V}\overleftrightarrow{\Lambda}^{-1}\ \overleftrightarrow{V}^{-1}\vec{B} \tag{30}$$

and with simple manipulations one can show that

$$\vec{Q}^{\mathrm{opt}} = \sum_i \beta_i \vec{v}_i \tag{31}$$

where

$$\beta_i = \frac{1}{\lambda_i} C_i \tag{32}$$

and

$$\vec{C} = \overleftrightarrow{V}^{-1}\vec{B} \tag{33}$$

REFERENCES

[1] Z. S. Hendsch and B. Tidor. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**: 211–226 (1994).

[2] C. D. Waldburger, J. F. Schildbach, and R. T. Sauer. Are buried salt bridges important for protein stability and conformational specificity? *Nature Struct. Biol.* **2**: 122–128 (1995).

[3] W. C. Wimley, K. Gawrisch, T. P. Creamer, and S. H. White. Direct measurement of salt-bridge solvation energies using a peptide model system: Implications for protein stability. *Proc. Natl. Acad. Sci. U.S.A.* **93**: 2985–2990 (1996).

[4] A.-S. Yang and B. Honig. Free energy determinants of secondary structure formation: I. α-helices. *J. Mol. Biol.* **252**: 351–365 (1995).

[5] L. Wang, T. O'Connell, A. Tropsha, and J. Hermans. Energetic decomposition of the α-helix–coil equilibrium of a dynamic model system. *Biopolymers* **39**: 479–489 (1996).

[6] V. K. Misra, K. A. Sharp, R. A. Friedman, and B. Honig. Salt effects on ligand–DNA binding: Minor groove binding antibiotics. *J. Mol. Biol.* **238**: 245–263 (1994).

[7] V. K. Misra, J. L. Hecht, K. A. Sharp, R. A. Friedman, and B. Honig. Salt effects on protein–DNA interactions: The λcI repressor and EcoRI endonuclease. *J. Mol. Biol.* **238**: 264–280 (1994).

[8] K. Sharp. Electrostatic interactions in hirudin–thrombin binding. *Biophys. Chem.* **61**: 37–49 (1996).

[9] J. Shen and J. Wendoloski. Electrostatic binding energy calculation using the finite difference solution to the linearized Poisson–Boltzmann equation: Assessment of its accuracy. *J. Comput. Chem.* **17**: 350–357 (1996).

[10] R. E. Bruccoleri, J. Novotny, M. E. Davis, and K. A. Sharp. Finite difference Poisson–Boltzmann electrostatic calculations: Increased accuracy achieved by harmonic dielectric

smoothing and charge antialiasing. *J. Comput. Chem.* **18**: 268–276 (1997).

[11] J. Novotny, R. E. Bruccoleri, M. Davis, and K. A. Sharp. Empirical free energy calculations: A blind test and further improvements to the method. *J. Mol. Biol.* **268**: 401–411 (1997).

[12] Z. S. Hendsch and B. Tidor. Electrostatic interactions in the gcn4 leucine zipper: Substantial contributions arise from intramolecular interactions enhanced on binding. *Protein Sci.* **In press.** (1999).

[13] C. V. Sindelar, Z. S. Hendsch, and B. Tidor. Effects of salt bridges on protein structure and design. *Protein Sci.* **7**: 1898–1914 (1998).

[14] L.-P. Lee and B. Tidor. Optimization of electrostatic binding free energy. *J. Chem. Phys.* **106**: 8681–8690 (1997).

[15] E. Kangas and B. Tidor. Optimizing electrostatic affinity in ligand–receptor binding: Theory, computation, and ligand properties. *J. Chem. Phys.* **109**: 7522–7545 (1998).

[16] L. T. Chong, S. E. Dempster, Z. S. Hendsch, L.-P. Lee, and B. Tidor. Computation of electrostatic complements to proteins: A case of charge stabilized binding. *Protein Sci.* **7**: 206–210 (1998).

[17] E. Kangas and B. Tidor. Charge optimization leads to favorable electrostatic binding free energy. *Phys. Rev. E* **59**: 5958–5961 (1999).

[18] G. Schreiber and A. R. Fersht. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry* **32**: 5145–5150 (1993).

[19] L. T. Chong, Z. S. Hendsch, and B. Tidor. In preparation.

[20] L.-P. Lee and B. Tidor. Protein-protein electrostatic recognition:Barstar is optimized for tight-binding to barnase. *In preparation.*

[21] K. A. Sharp and B. Honig. Calculating total electrostatic energies with the nonlinear

Poisson–Boltzmann equation. *J. Phys. Chem.* **94**: 7684–7692 (1990).

[22] J. T. G. Overbeek. The role of energy and entropy in the electrical double layer. *Colloids and Surfaces* **51**: 61–75 (1990).

[23] F. Fogolari and J. M. Briggs. On the variational approach to Poisson–Boltzmann free energies. *Chem. Phys. Lett.* **281**: 135–139 (1997).

[24] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**: 187–217 (1983).

[25] M. K. Gilson and B. H. Honig. Calculation of electrostatic potentials in an enzyme active site. *Nature (London)* **330**: 84–86 (1987).

[26] O'M. Bockris and A. K. N. Reddy. *Modern Electrochemistry.* Plenum, New York (1973).

[27] I. Klapper, R. Hagstrom, R. Fine, K. Sharp, and B. Honig. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins: Struct., Funct., Genet.* **1**: 47–59 (1986).

[28] C. Martin, V. Richard, M. Salem, R. Hartley, and Y. Mauquen. Refinement and structural analysis of barnase at 1.5 angstrom resolution. *Acta. Crystallogr.* **55**: 386–398 (1999).

[29] G. S. Ratnaparkhi, S. Ramachandran, J. B. Udgaonkar, and R. Varadarajan. Discrepancies between the NMR and X-ray structures of uncomplexed barstar: Analysis suggests that packing densities of protein structures determined by NMR are unreliable. *Biochemistry* **37**: 6958–6966 (1998).

[30] V. Guillet, A. Lapthorn, R. W. Hartley, and Y. Mauguen. Recognition between a bacterial ribonuclease, barnase, and its natural inhibitor, barstar. *Structure* **1**: 165–177 (1993).

[31] A. M. Buckle, G. Schreiber, and A. R. Fersht. Protein–protein recognition: Crystal structure analysis of a barnase–barstar complex at 2.0-Å resolution. *Biochemistry* **33**: 8878–

8889 (1994).

[32] M. K. Gilson, K. A. Sharp, and B. H. Honig. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J. Comput. Chem.* **9**: 327–335 (1988).

[33] K. A. Sharp and B. Honig. Electrostatic interactions in macromolecules: Theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* **19**: 301–332 (1990).

[34] G. Strang. *Introduction to Linear Algebra.* Wellesley–Cambridge Press, Wellesley, Massachusetts (1993).

[35] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, Cambridge, second edition (1992).

[36] R. J. Vanderbei. LOQO: An interior-point code for quadratic programming. *Optimization methods and Software* **To appear.** (1999).

[37] J. Novotny and K. Sharp. Electrostatic fields in antibodies and antibody/antigen complexes. *Prog. Biophys. Molec. Biol.* **58**: 203–224 (1992).

[38] J. A. Caravella, J. D. Carbeck, D. C. Duffy, G. M. Whitesides, and B. Tidor. Long-range electrostatic contributions to protein–ligand binding estimated using protein charge ladders, affinity capillary electrophoresis, and continuum electrostatic theory. *J. Am. Chem. Soc.* **In press.** (1999).

[39] S. A. Allison, G. Ganti, and J. A. McCammon. Simulation of the diffusion-controlled reaction between superoxide and superoxide dismutase. I. Simple models. *Biopolymers* **24**: 1323–1336 (1985).

[40] S. A. Allison, R. J. Bacquet, and J. A. McCammon. Simulation of the diffusion-controlled reaction between superoxide and superoxide dismutase. II. Detailed models. *Biopolymers*

27: 251–269 (1988).

[41] M. Marquart, J. Walter, J. Deisenhofer, W. Bode, and R. Huber. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Cryst.* **B39**: 480–490 (1983).

[42] J. Bajorath, D. H. Kitson, J. Kraut, and A. T. Hagler. The electrostatic potential of Escherichia coli dihydrofolate reductase. *Proteins: Struct., Funct., Genet.* **11**: 1–12 (1991).

[43] W. L. Jorgensen and J. Tirado-Rives. The OPLS potential function for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**: 1657–1666 (1988).

[44] D. Sitkoff, K. A. Sharp, and B. Honig. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **98**: 1978–1988 (1994).

[45] A. Nicholls, K. A. Sharp, and B. Honig. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct., Funct., Genet.* **11**: 281–296 (1991).

[46] G. Schreiber and A. R. Fersht. Energetics of protein–protein interactions: Analysis of the barnase–barstar interface by single mutants and double mutant cycles. *J. Mol. Biol.* **248**: 478–486 (1995).

[47] R. R. Gabdoulline and R. C. Wade. Simulation of the diffusional association of barnase and barstar. *Biophys. J.* **72**: 1917–1929 (1997).

[48] G. Schreiber and A. R. Fersht. Rapid, electrostatically assisted association of proteins. *Nature Struct. Biol.* **3**: 427–431 (1996).

[49] G. Schreiber, A. M. Buckle, and A. R. Fersht. Stability and function: Two constraints in the evolution of barstar and other proteins. *Structure* **2**: 945–951 (1994).

[50] B. Nolting, R. Golbik, J. L. Neira, A. S. Soler-Gonzales, G. Schreiber, and A. R. Fersht.

The folding pathway of a protein at high resolution from microseconds to seconds. *Proc. Natl. Acad. Sci. U.S.A.* **94**: 826–830 (1997).

[51] B. Nolting, R. Golbik, A. S. Soler-Gonzales, and A. R. Fersht. Circular dichroism of denatured barstar suggests residual structure. *Biochemistry* **36**: 9899–9905 (1997).

TABLES

TABLE I. Optimized charge distributions and energetics for barstar.[a]

Charge set	$\{Q^{\mathrm{Lv}}\}$	n_{Lv}	Constraints	ΔG_{bind} (kcal/mol)
$\{Q_i^{\mathrm{wt}}\}$	None	0	—	14.2
$\{Q_i^{\mathrm{opt1}}\}$	All barstar charges	839	$-0.85 \leq Q_i \leq 0.85$	−6.1
$\{Q_i^{\mathrm{opt2}}\}$	Side chain charges	403	$-0.85 \leq Q_i \leq 0.85$	−0.6
$\{Q_i^{\mathrm{opt3}}\}$	Side chain charges	403	$-0.85 \leq Q_i \leq 0.85$; $-1.5 \leq Q_n^{\mathrm{sc}} \leq 1.5$	−0.1

[a] $\{Q_i^{\mathrm{wt}}\}$ is the unoptimized wild-type barstar charge distribution for comparison. The remaining three charge distributions represent optimizations the the indicated constraints. $\{Q_i^{\mathrm{Lv}}\}$ describes the set of variable ligand (barstar) charges in each optimization and n_{Lv} is the number of charges in that set, while Q_n^{sc} is the total side-chain charge of the nth residue.

TABLE II. Component analysis for the wild-type barstar binding barnase. [a]

Side chain	ΔG_{solv}	ΔG_{intra}	ΔG_{inter}	ΔG_{total}	Side chain	ΔG_{solv}	ΔG_{intra}	ΔG_{inter}	ΔG_{total}
1	0.02	−0.20	0.41	0.24	45	0.00	0.00	0.00	0.00
2	0.00	−0.02	0.07	0.04	46	0.24	0.13	−0.96	−0.59
3	0.00	0.00	0.00	0.00	47	0.00	0.00	−0.08	−0.08
4	0.00	0.00	0.00	0.00	48	0.00	−0.02	0.04	0.03
5	0.00	0.00	0.00	0.00	49	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	50	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	51	0.00	0.00	0.00	0.00
8	0.00	0.01	−0.02	−0.01	52	0.00	0.02	−0.06	−0.04
9	0.00	0.00	0.00	0.00	53	0.00	0.00	−0.01	−0.01
10	0.00	0.00	0.00	0.00	54	0.00	−0.01	0.03	0.02
11	0.00	0.00	0.00	−0.01	55	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	56	0.00	0.00	0.00	0.00
13	0.00	0.00	0.00	0.00	57	0.00	0.01	−0.03	−0.02
14	0.00	0.00	0.00	0.00	58	0.00	0.00	0.01	0.00
15	0.00	0.02	−0.01	0.01	59	0.00	0.01	−0.01	−0.01
16	0.00	0.00	0.00	0.00	60	0.00	0.00	0.01	0.00
17	0.01	−0.04	0.26	0.23	61	0.00	0.00	0.00	0.00
18	0.00	0.00	0.00	0.00	62	0.00	0.00	0.00	0.00
19	0.00	−0.01	0.03	0.02	63	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	0.00	66	0.00	0.00	0.00	0.00
21	0.04	−0.12	−0.02	−0.10	67	0.00	0.00	0.00	0.00
22	0.00	−0.05	0.09	0.05	68	0.00	0.04	−0.14	−0.10
23	0.00	0.06	−0.15	−0.09	69	0.00	0.00	−0.04	−0.04
24	0.00	0.00	0.00	0.00	70	0.00	0.00	0.00	0.00
25	0.00	0.00	0.00	0.00	71	0.00	0.00	0.00	0.00
26	0.00	0.00	0.00	0.00	72	0.01	−0.03	0.22	0.19
27	0.02	−0.20	0.47	0.29	73	0.00	0.00	0.00	0.00
28	0.01	0.05	−0.02	0.04	74	0.00	0.00	0.00	0.00
29	0.96	0.70	**−2.81**	**−1.15**	75	0.00	−0.06	0.20	0.14
30	0.01	−0.05	0.23	0.19	76	*1.39*	0.26	**−6.95**	**−5.30**
31	0.00	0.00	0.00	0.00	77	0.00	0.00	0.00	0.00
32	0.09	−0.01	0.08	0.16	78	0.00	−0.11	0.21	0.11
33	*1.65*	−0.47	**−1.90**	−0.72	79	0.00	0.00	0.00	0.00
34	0.00	0.00	0.00	0.00	80	0.93	0.55	**−1.90**	−0.42
35	*13.30*	**−2.88**	**−24.49**	**−14.08**	81	0.00	0.00	0.00	0.00
36	0.00	0.00	0.00	0.00	82	0.00	0.00	0.00	0.00
37	0.00	0.00	0.00	0.00	83	0.01	0.13	−0.26	−0.12
38	0.83	−0.02	−0.78	0.03	84	0.00	0.00	0.00	0.00
39	*11.76*	*1.86*	**−27.58**	**−13.96**	85	0.00	−0.01	0.02	0.01
40	0.00	0.00	0.00	0.00	86	0.00	0.00	0.00	0.00
41	0.00	0.00	0.00	0.00	87	0.00	0.00	0.00	0.00
42	0.57	0.49	**−3.15**	**−2.08**	88	0.00	0.00	0.00	0.00
43	0.00	0.00	0.00	0.00	89	0.00	0.00	0.00	0.00
44	0.01	0.09	-0.17	−0.07	Total	31.85	0.11	−69.15	−37.20

[a] ΔG_{solv} is the desolvation penalty for the side chain atoms of the indicated residue, ΔG_{intra} is the intramolecular (indirect) contribution for the indicated side chain with the remainder of barstar, and ΔG_{inter} is the intermolecular (direct) electrostatic interaction between the in-

dicated side chain and barnase. ΔG_{total} is the sum of the contributions. A barstar backbone contribution of -0.71 kcal/mol and a barnase dehydration contribution of 52.1 kcal/mol is to be added to the final ΔG_{total} to give the total electrostatic binding free energy. Bold and italic indicate favorable and unfavorable contributions, respectively, of magnitude exceeding 1 kcal/mol.

TABLE III. Component analysis for the optimal charge distribution ($\{Q^{opt3}\}$).[a]

Side chain	ΔG_{solv}	ΔG_{intra}	ΔG_{inter}	ΔG_{total}	Side chain	ΔG_{solv}	ΔG_{intra}	ΔG_{inter}	ΔG_{total}
1	0.08	0.58	**−1.22**	−0.56	45	0.45	*1.78*	**−4.08**	**−1.85**
2	0.00	0.01	−0.01	−0.01	46	0.48	0.50	**−1.96**	−0.98
3	0.00	0.04	−0.08	−0.04	47	0.33	**−1.97**	*2.75*	*1.11*
4	0.00	0.00	−0.01	0.00	48	0.04	0.24	−0.55	−0.26
5	0.00	−0.02	0.05	0.03	49	0.12	0.94	**−1.97**	−0.92
6	0.00	0.00	0.00	0.00	50	0.00	0.01	−0.02	−0.01
7	0.00	0.00	0.00	0.00	51	0.00	0.07	−0.16	−0.09
8	0.00	0.00	0.00	0.00	52	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	53	0.01	−0.16	0.43	0.28
10	0.00	−0.01	0.03	0.03	54	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	55	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00	56	0.00	−0.05	0.14	0.09
13	0.01	−0.01	0.04	0.04	57	0.00	0.00	0.01	0.01
14	0.05	−0.18	0.07	−0.05	58	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.01	59	0.00	−0.01	0.05	0.04
16	0.02	−0.07	0.31	0.26	60	0.00	−0.01	0.04	0.03
17	0.54	−0.04	0.62	*1.12*	61	0.00	0.00	0.00	0.00
18	0.03	0.03	0.00	0.06	62	0.00	0.00	0.00	0.00
19	0.00	−0.01	0.02	0.01	63	0.01	0.03	0.01	0.04
20	0.00	−0.07	0.20	0.14	66	0.00	0.00	0.00	0.00
21	0.10	0.08	−0.56	−0.38	67	0.00	−0.01	0.06	0.04
22	0.01	0.07	−0.16	−0.08	68	0.01	−0.08	0.22	0.14
23	0.00	0.03	−0.08	−0.05	69	0.17	−0.12	0.08	0.13
24	0.05	0.52	**−1.19**	−0.63	70	0.19	−0.37	*1.04*	0.86
25	0.02	0.15	−0.28	−0.12	71	0.00	−0.05	0.12	0.07
26	0.03	−0.56	0.84	0.31	72	0.44	0.29	**−1.99**	**−1.25**
27	0.13	−0.55	0.92	0.49	73	0.26	−0.76	*1.02*	0.52
28	0.05	−0.05	−0.03	−0.04	74	0.10	0.76	**−1.95**	**−1.09**
29	0.81	−0.02	**−1.80**	**−1.01**	75	0.01	0.25	−0.57	−0.31
30	0.49	−0.74	*1.34*	*1.09*	76	4.92	*1.01*	**−12.30**	**−6.36**
31	0.00	0.00	0.00	0.00	77	0.18	*1.19*	**−2.73**	**−1.36**
32	0.11	−0.19	0.07	−0.01	78	0.00	−0.06	0.10	0.03
33	0.59	−0.70	**−1.35**	**−1.46**	79	0.01	0.25	**−0.53**	−0.27
34	0.34	0.08	**−1.57**	**−1.15**	80	0.33	0.91	**−2.30**	**−1.06**
35	*11.93*	**−2.42**	**−23.81**	**−14.30**	81	0.00	0.00	0.00	0.00
36	0.00	0.02	0.01	0.03	82	0.17	**−1.07**	*1.46*	0.56
37	0.93	0.47	**−4.34**	**−2.94**	83	0.04	−0.39	0.53	0.18
38	*2.23*	0.87	**−6.67**	**−3.57**	84	0.00	0.05	−0.10	−0.05
39	*10.06*	*2.20*	**−24.98**	**−12.72**	85	0.00	−0.02	0.04	0.02
40	0.04	0.37	−0.86	−0.46	86	0.00	0.05	−0.10	−0.05
41	0.23	**−1.92**	*3.46*	*1.77*	87	0.00	0.00	0.00	0.00
42	*2.59*	*1.72*	**−8.75**	**−4.44**	88	0.00	−0.01	0.01	0.01
43	0.00	0.00	0.00	0.00	89	0.00	0.00	0.00	0.00
44	0.54	0.39	**−2.05**	**−1.11**	Total	40.29	3.27	−95.02	−51.46

[a] All ΔG values are in kcal/mol. ΔG_{solv} is the desolvation penalty for the side chain atoms of the indicated residues, ΔG_{intra} is the intramolecular (indirect) contribution for the indicated side chain with the remainder of barstar, and ΔG_{inter} is the intermolecular

(direct) electrostatic interaction between the indicated side chain and barnase. ΔG_{total} is the sum of the contributions. A barstar backbone contribution of -0.71 kcal/mol and a barnase dehydration contribution of 52.1 kcal/mol is to be added to the final ΔG_{total} to give the total electrostatic binding free energy. Bold and italic indicate favorable and unfavorable contributions, respectively, of magnitude exceeding 1 kcal/mol.

TABLE IV. Ion pairs in wild-type and optimized (opt3) charge distribution.

Ion pair		Distance (Å)	Optimal residue charge	Charges of barstar side chain		
				Atom	Q_i^{wt}	Q_i^{opt3}
Asp39b*	Arg83bn	1.6	−0.69	CB	−0.16	−0.36
	Arg87bn	1.9		CG	0.36	0.44
	Lys27bn	3.9		OD1	−0.60	−0.65
				OD2	−0.60	−0.85
Glu76b*	Arg59bn	2.1	−1.15	CB	0.00	0.45
				CG	−0.16	0.85
				CD	0.36	−0.85
				OE1	−0.60	−0.75
				OE2	−0.60	−0.85
Asp35b*	Lys62bn	4.6	−0.90	CB	−0.16	−0.20
				CG	0.36	0.85
				OD1	−0.60	−0.79
				OD2	−0.60	−0.75
Gln72b*	**Arg59bn**	4.8	−1.50	CB	0.00	0.49
				CG	0.00	−0.53
				CD	0.55	−0.85
				OE1	−0.55	0.08
				NE2	−0.60	−0.69
				HE21	0.30	−0.85
				HE22	0.30	0.85

TABLE V. Hydrogen bonds in wild-type and optimized (opt3) charge distribution.

Donor		Charge		Acceptor		Charge		Distance (Å)
Residue	Atom	Q_i^{wt}	Q_i^{opt3}	Residue	Atom	Q_i^{wt}	Q_i^{opt3}	
Lys27bn	HZ1	0.35		Thr42b*	OG1	−0.65	−0.85	1.94
Arg59bn	H	0.25		Asp35b*	OD1	−0.60	−0.79	1.98
Arg59bn	HH22	0.35		Glu76b*	OE1	−0.60	−0.75	2.07
Glu60bn	H	0.25		Asp35b*	OD2	−0.60	−0.75	2.45
Arg83bn	HH11	0.35		Gly43b*	O	−0.55		2.11
Arg83bn	HH12	0.35		Asp39b*	OD1	−0.60	−0.65	2.11
Arg83bn	HH12	0.35		Asp39b*	O	−0.55		2.93
Arg83bn	HH21	0.35		Asp39b*	OD1	−0.60	−0.65	1.58
Arg87bn	HH21	0.35		Asp39b*	OD2	−0.60	−0.85	1.94
His102bn	HE2	0.30		Asp39b*	OD2	−0.60	−0.85	1.84
Tyr29b*	HH	0.40	0.30	Arg83bn	O	−0.55		1.88
Gly31b*	H	0.25		His102bn	ND1	−0.40		2.12
Asn33b*	HD22	0.30	0.09	His102bn	O	−0.55		2.29
Leu34b*	H	0.25		Glu60bn	OE2	−0.60		1.86
Trp38b*	HE1	0.30	0.04	Wat48bn	OH2	−0.834		2.45
Trp38b*	HE1	0.30	0.04	Wat56bn	OH2	−0.834		2.63
Trp38b*	HE1	0.30	0.04	Wat60bn	OH2	−0.834		2.90
Wat14bn	H2	0.417		Asp35b*	O	−0.55		2.17
Wat22bn	H2	0.417		Asp35b*	OD2	−0.60	−0.75	1.87
Wat29bn	H2	0.417		Asp35b*	OD1	−0.60	−0.79	1.86
Wat33bn	H1	0.417		Asp39b*	OD1	−0.60	−0.65	2.29
Wat36bn	H1	0.417		Val45b*	O	−0.55		2.39
Wat36bn	H2	0.417		Val45b*	O	−0.55		2.14
Wat128bn	H2	0.417		Asp35b*	OD2	−0.60	−0.75	1.97
Wat155bn	H2	0.417		Gly43b*	O	−0.55		1.97
Arg59bn	**HH12**	0.35		**Glu76b***	**CD**	0.36	−0.85	2.24
Wat29bn	**H1**	0.417		**Trp38b***	**CB**	0.00	−0.24	2.93

TABLE VI. Dependence of electrostatic binding free energy on total charge.

Charge set	Total charge	ΔG_{bind} (kcal/mol)
$\{Q_i^{\text{opt2}}\}$	−0.3[a]	−0.56
$\{Q_i^{\text{opt2a}}\}$	−10.0	−0.52
$\{Q_i^{\text{opt2b}}\}$	0.0	−0.56
$\{Q_i^{\text{opt2c}}\}$	10.0	−0.52

[a] Opt2 had no constraint on the total charge. The remaining charge sets had the total charge constrained to the listed values.

TABLE VII. Improvement in ΔG_{bind}: A comparison between barnase and barstar optimization

	Barnase optimization		Barstar optimization	
Degree of freedom(Constraints)[a]	n_{Lv}	$\Delta\Delta G_{\text{bind}}$[b]	n_{Lv}	$\Delta\Delta G_{\text{bind}}$[b]
All charges	1071	-30.8	839	-20.3
Side chain charge	533	-27.9	403	-14.8
Side chain charge; $-1.5 \leq \sum_{\text{sidechain}} Q_i \leq 1.5$	533	-27.0	403	-14.3
Side chain charges with $A_{kk} > 0.1$	179	-25.9	138	-13.3
100 side chain charges with the greatest A_{kk}	100	-21.4	100	-12.0
200 side chain charges with the greatest A_{kk}	200	-26.5	200	-14.6
300 side chain charges with the greatest A_{kk}	300	-27.8	300	-14.8

[a] The constraints are on top of the individual charge range constraint $-0.85 \leq Q_i \leq 0.85$.

[b] $\Delta\Delta G_{\text{bind}}$ is the improvement of the optimized charge distribution over wild type in kcal/mol.

FIGURES

FIG. 1. Ribbon drawing of the bound structure of barnase (blue) and barstar (yellow) with bound water molecules (gray). Side chains found to be important for binding in mutagenesis experiments are highlighted.

FIG. 2. Ribbon drawing of the bound structure of barnase (blue) and barstar (yellow), highlighting barstar side chains with ΔG_{inter} more than 1 kcal/mol for the $\{Q^{\text{wt}}\}$ (A) and $\{Q^{\text{opt3}}\}$ (B) and side chains with ΔG_{intra} more than 1 kcal/mol for the $\{Q^{\text{wt}}\}$ (C) and $\{Q^{\text{opt3}}\}$ (D)

FIG. 3. Ribbon drawing of barnase and a C_α trace of the ligand barstar, in which the radius of each barstar C_α atom represents the average A_{kk} for that residue's side-chain atoms (A) and the r.m.s difference between $\{Q_i^{\text{opt2}}\}$ and $\{Q_i^{\text{opt2a}}\}$ (the unconstrained optimization and that constrained to a total charge of -10) (B).

FIG. 4. Scatter plot of the A_{ii} versus the difference between Q_i^{opt2} and Q_i^{opt2r} (ΔQ_i). Only charges with small A_{ii} have significant ΔQ_i.

FIG. 5. Potential plots displayed using GRASP: screened coulombic potential for barstar (A) and barnase (B) on barstar's molecular surface, barstar's desolvation potential (C) and interaction potential (D) on barstar's molecular surface, barnase's desolvation potential (E) and interaction potential (F) on barnase's molecular surface.

Figure 1

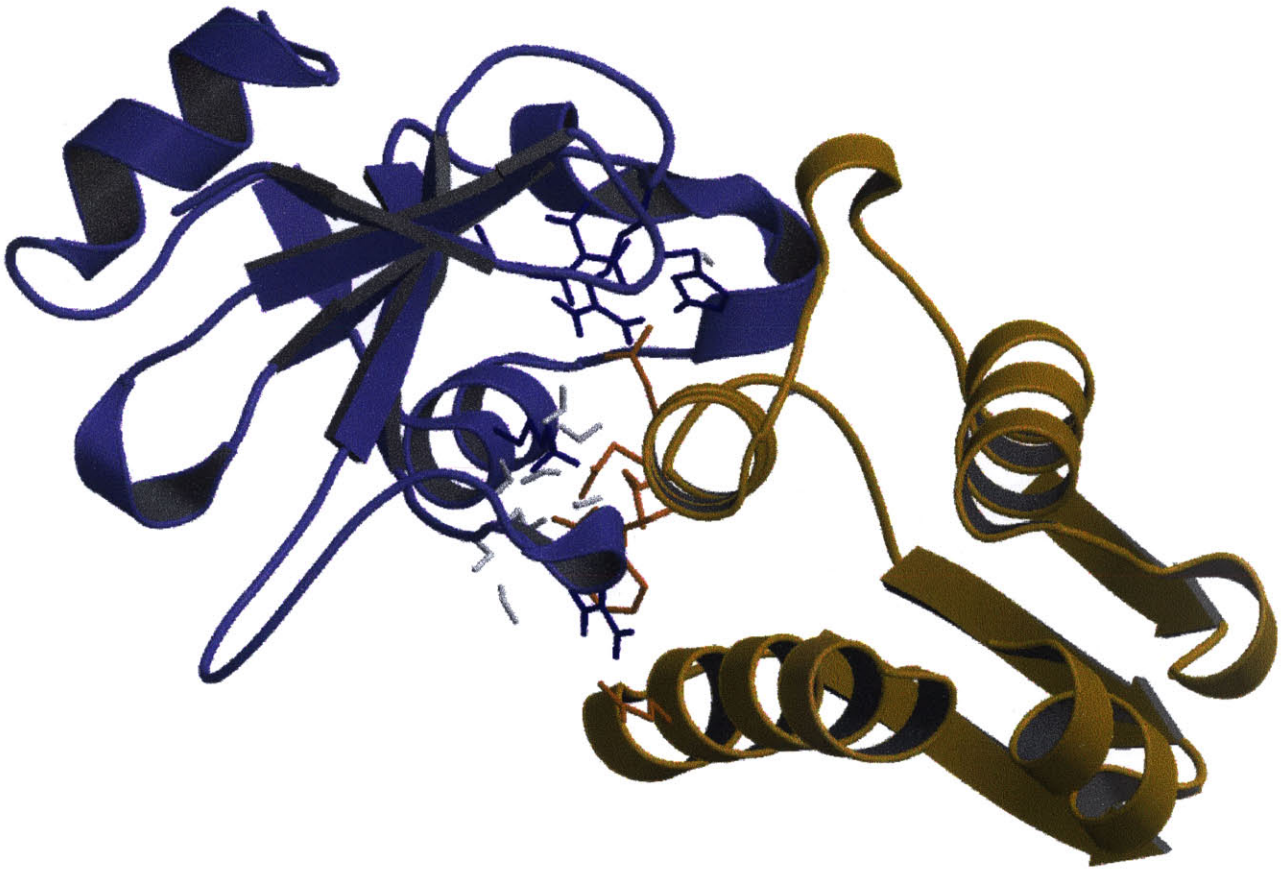


Figure 2A

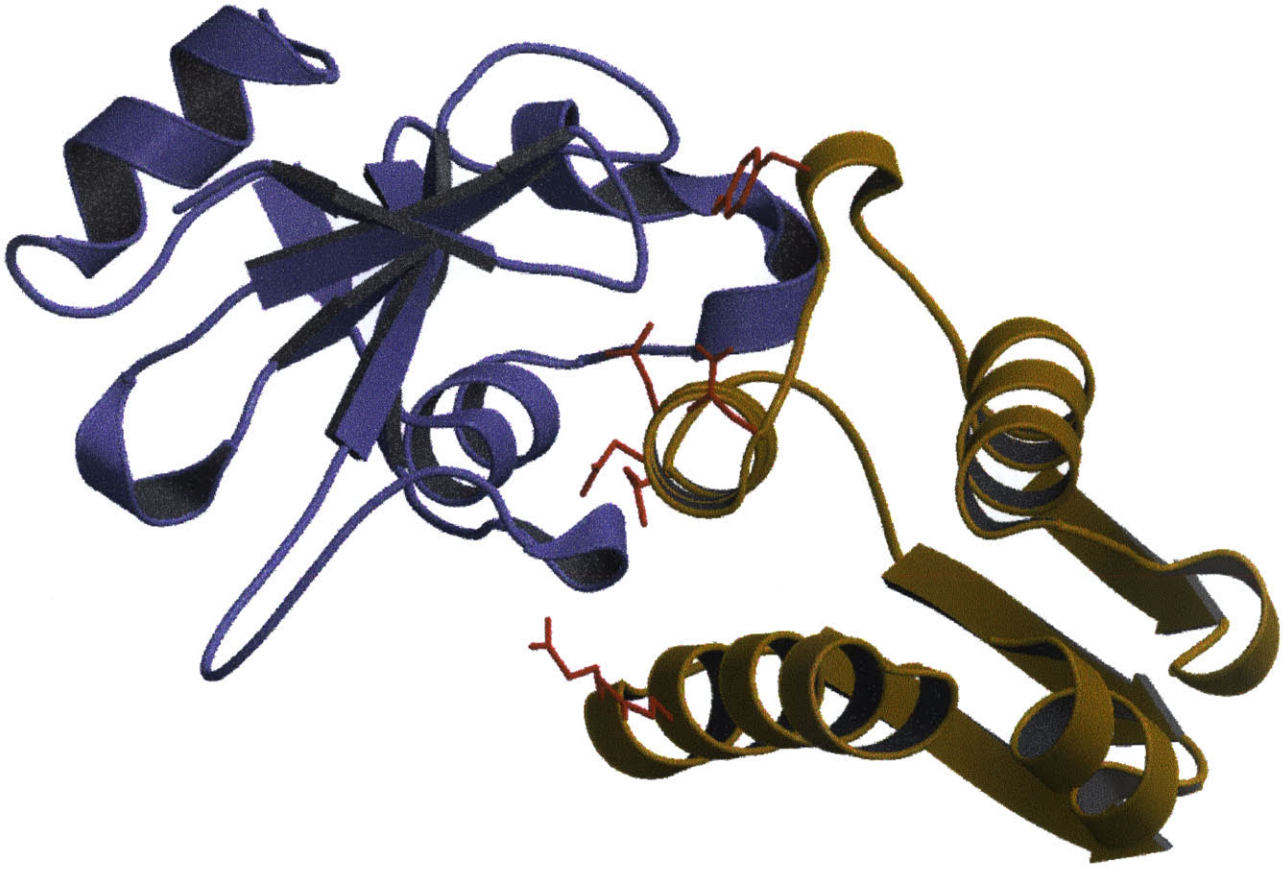


Figure 2B

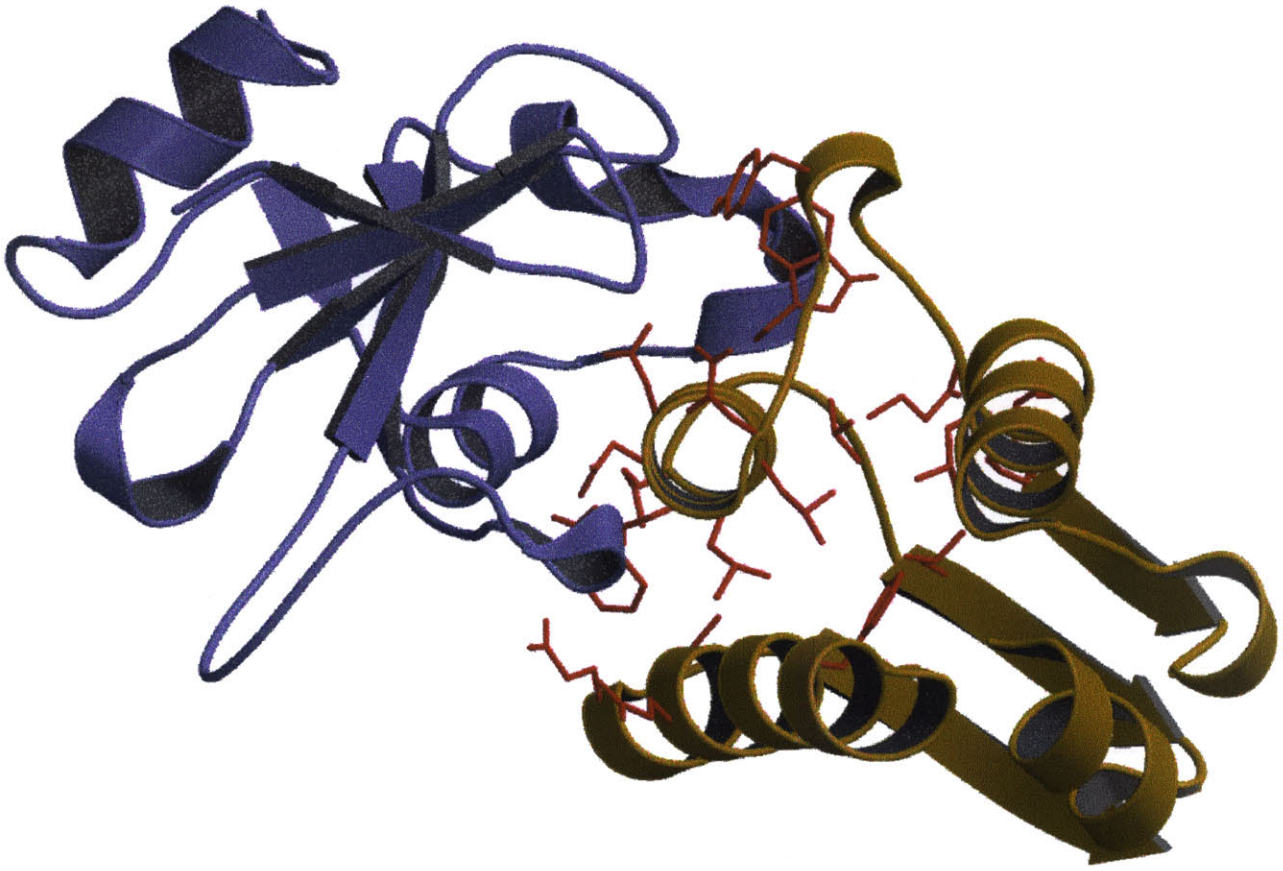


Figure 2C

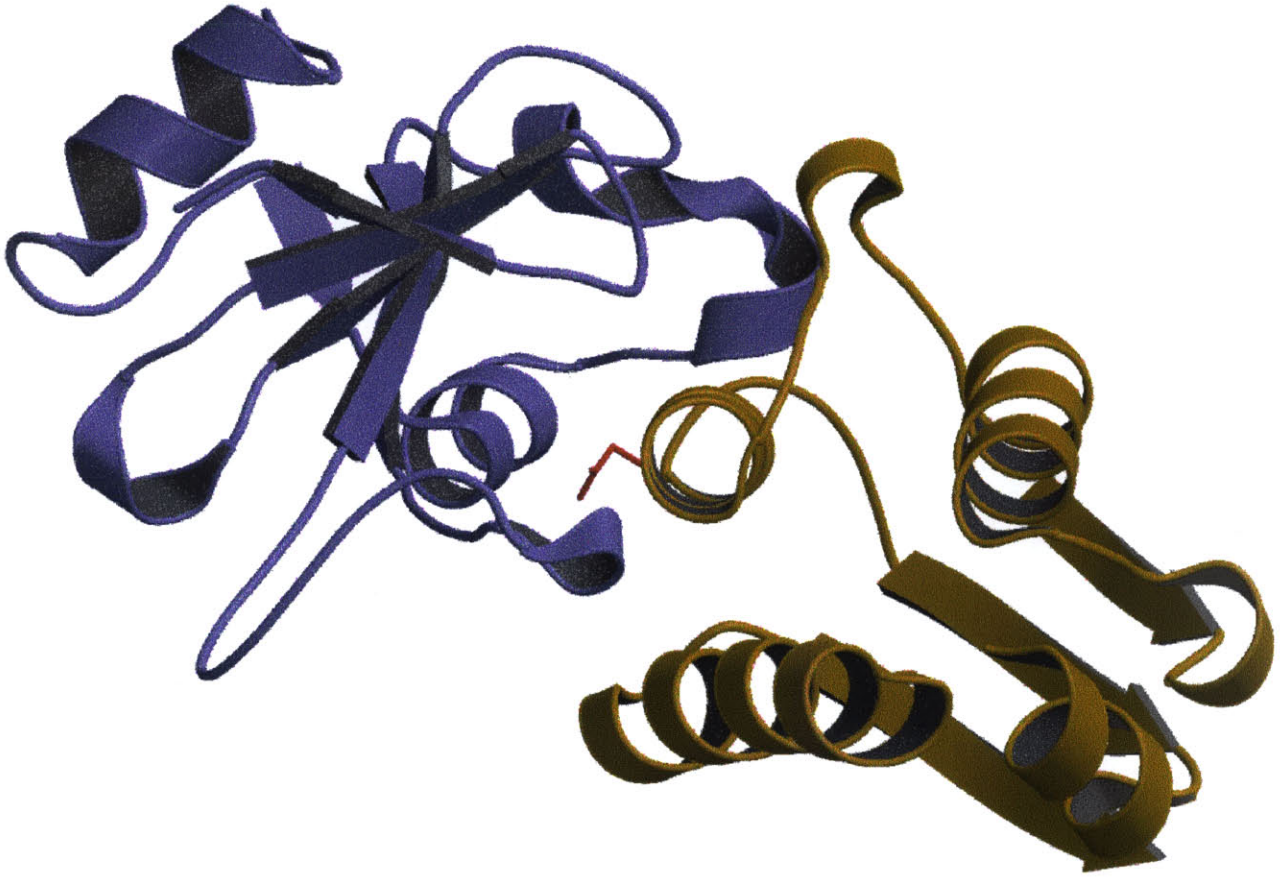


Figure 2D

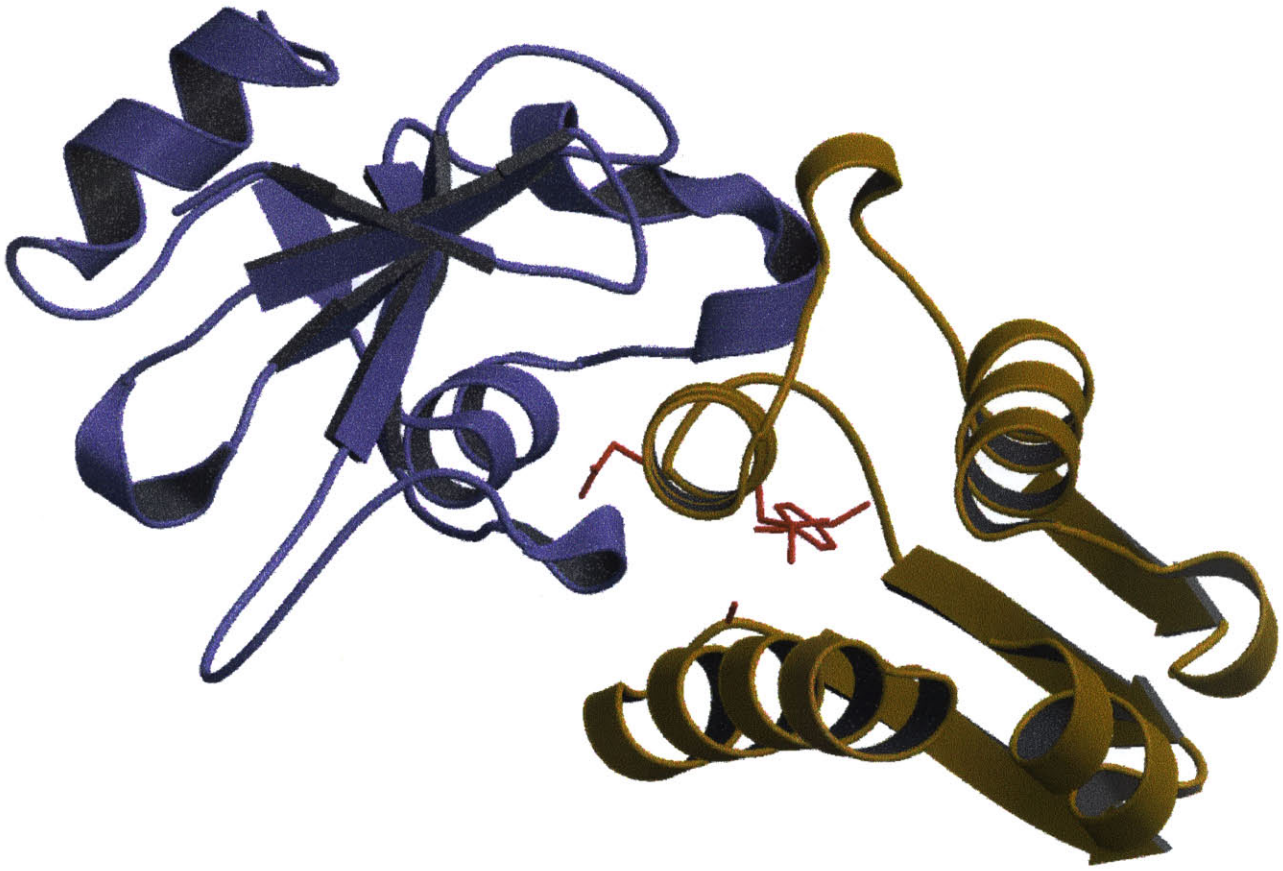


Figure 3A

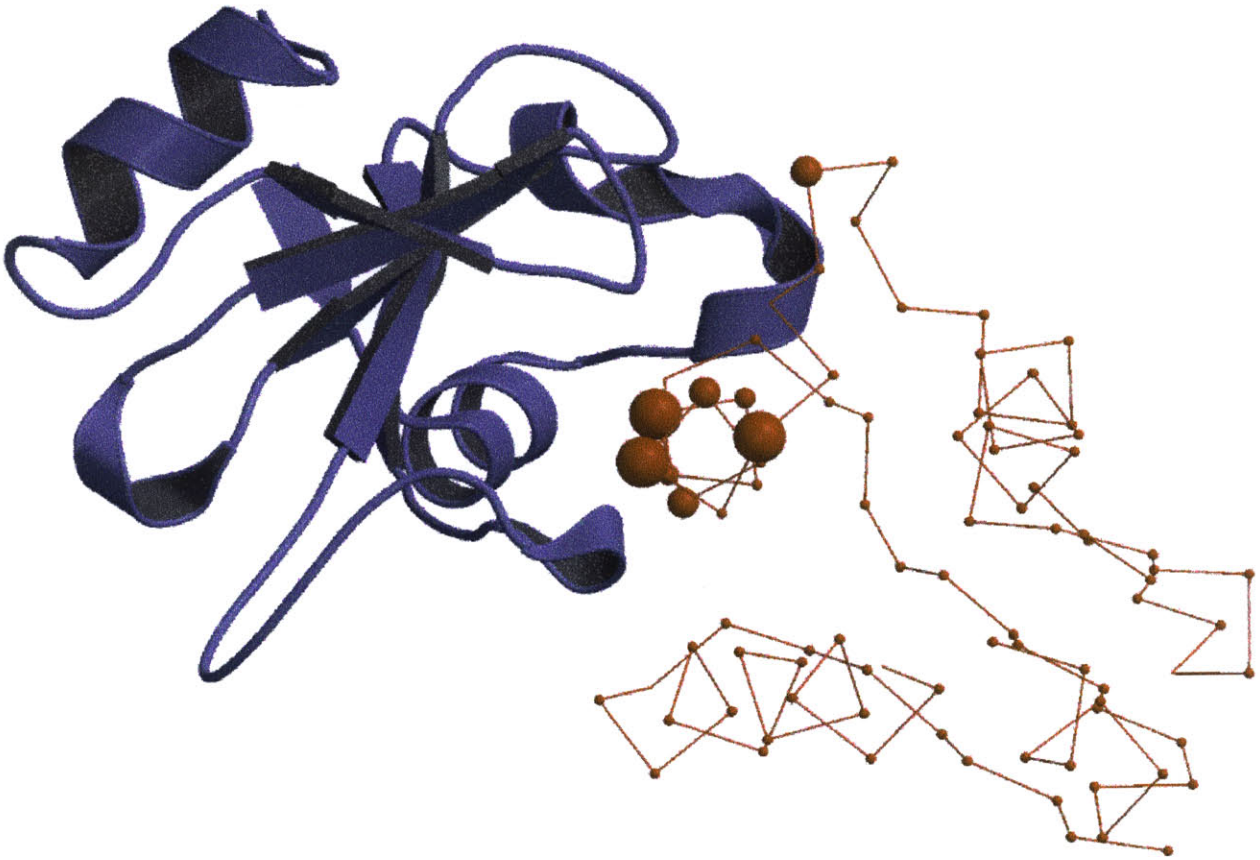


Figure 3B

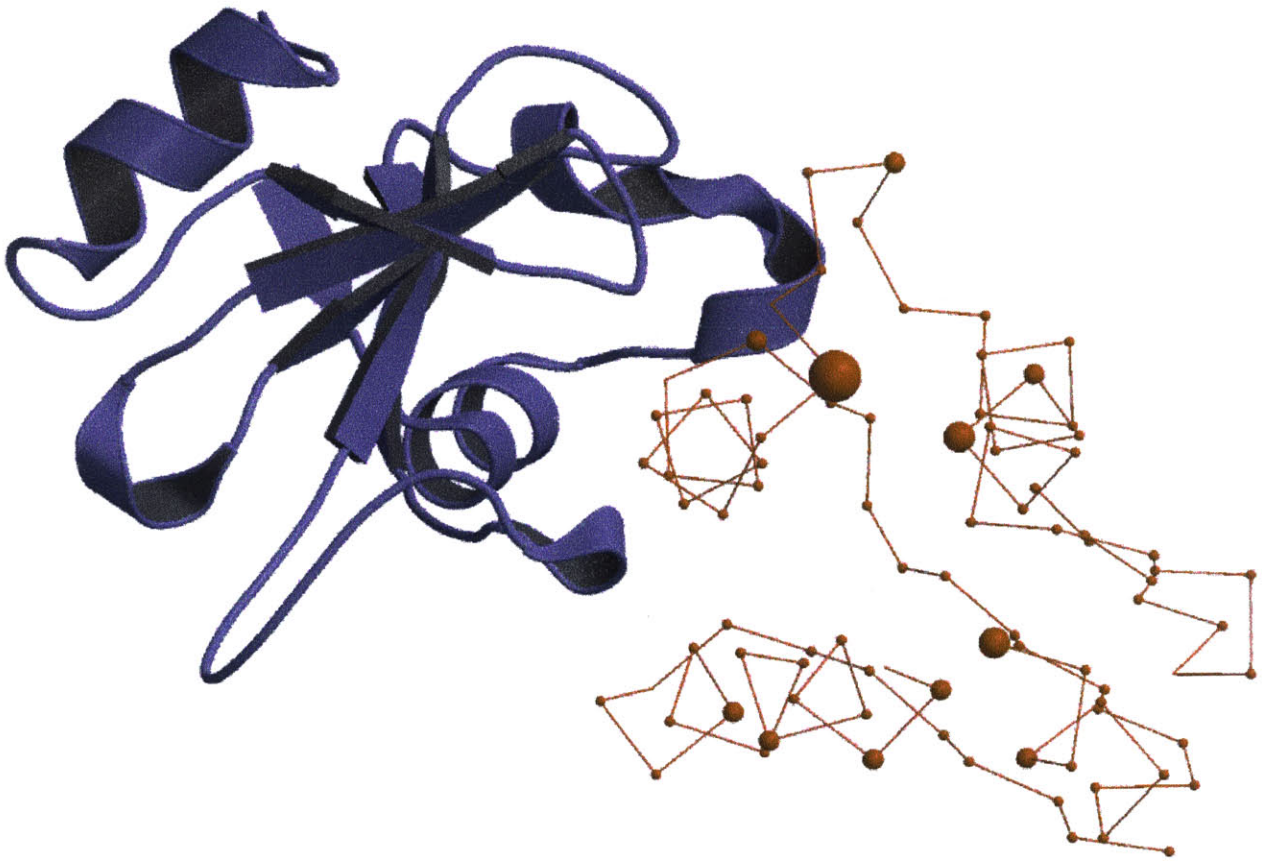


Figure 4

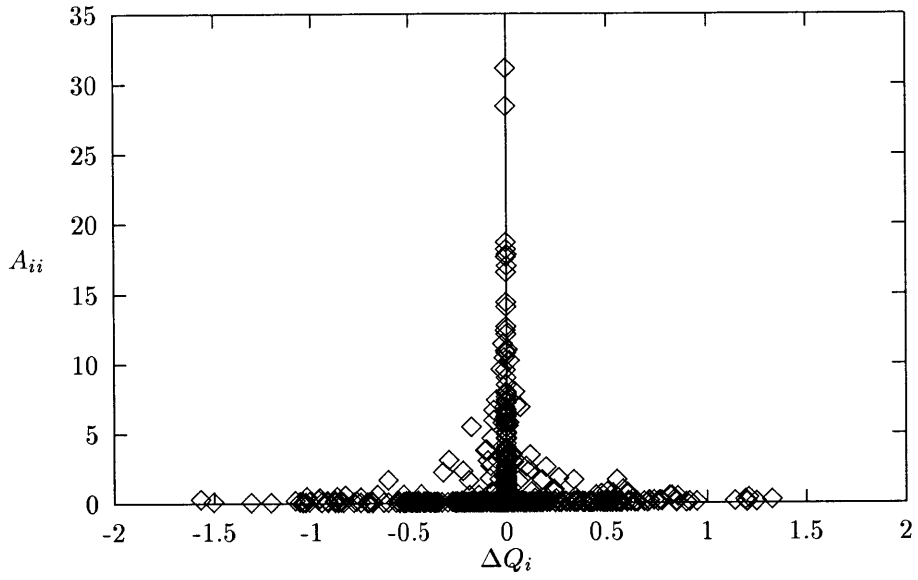


Figure 5A

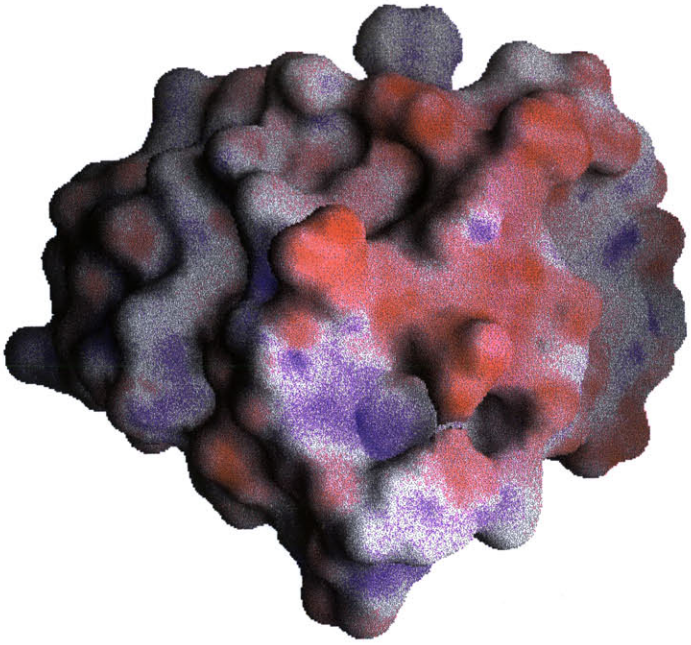


Figure 5B

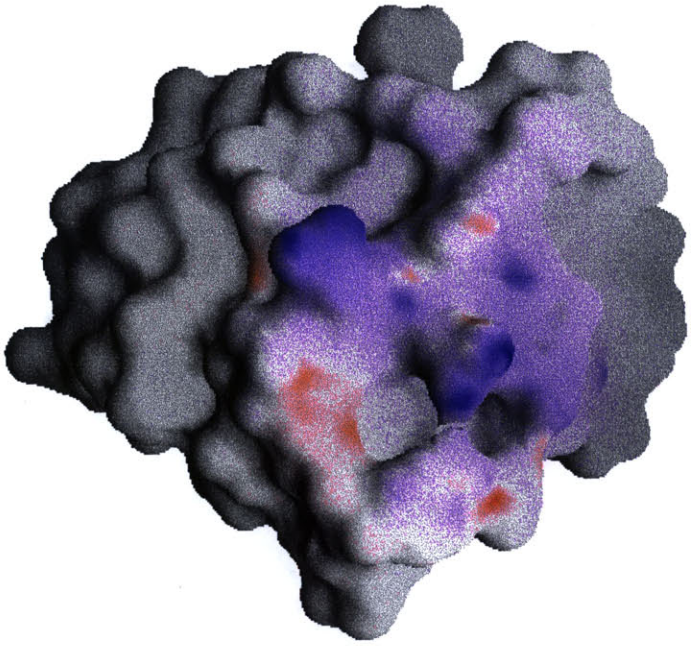


Figure 5C

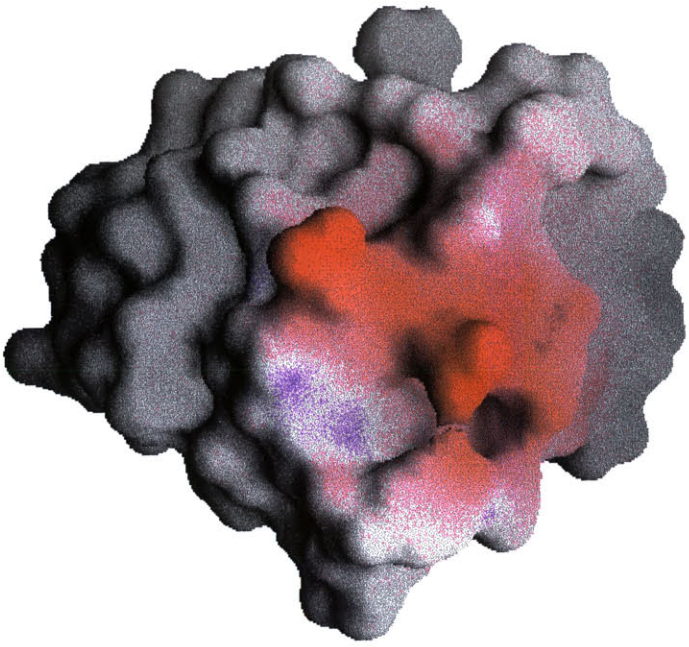


Figure 5D

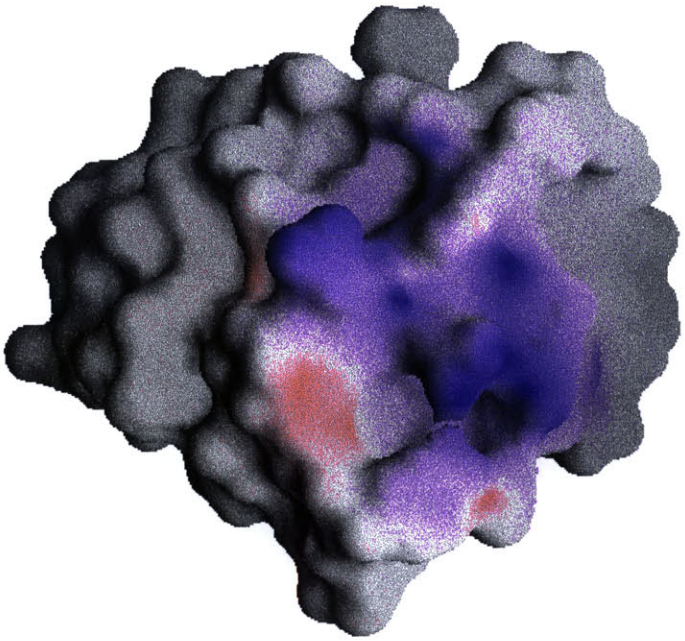


Figure 5E

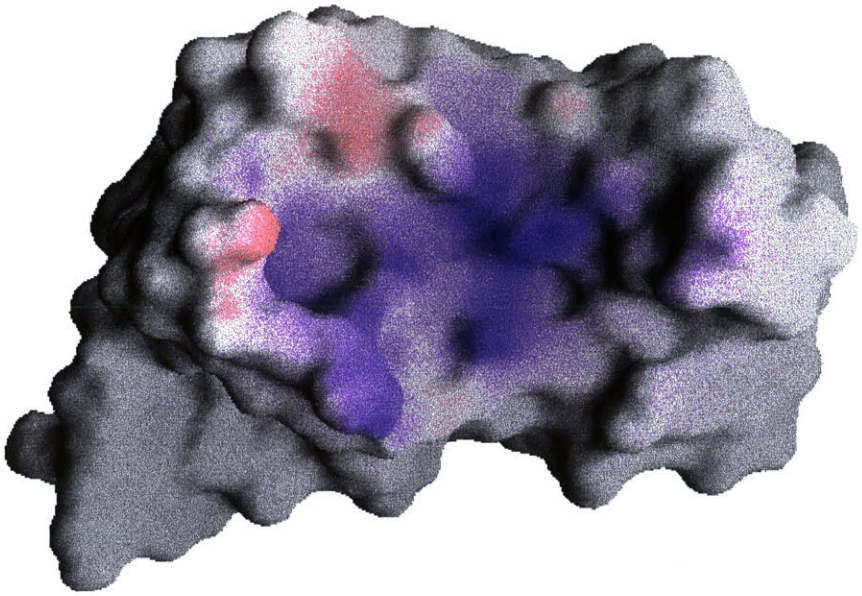
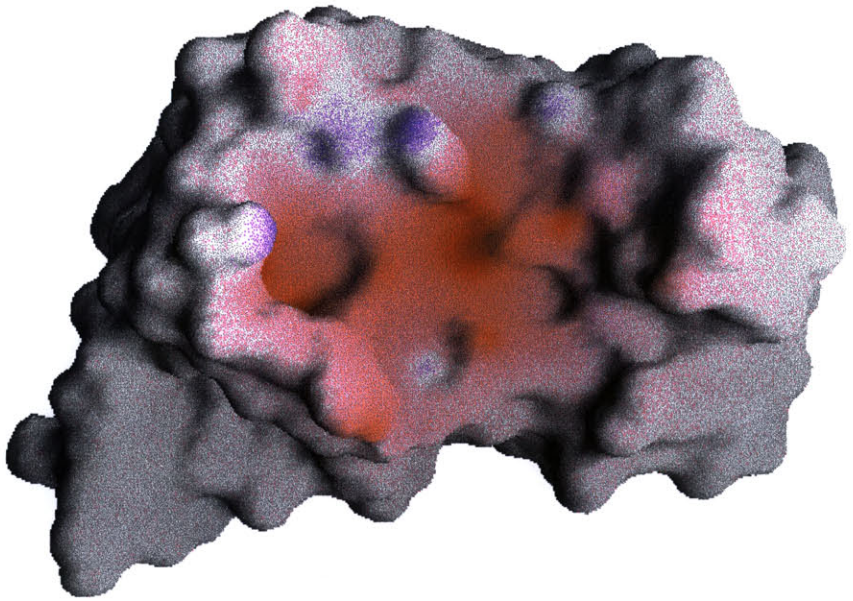


Figure 5F



Chapter V

Protein-Protein Electrostatic Recognition: Barstar Is Optimized for Tight-Binding to Barnase

Abstract

A novel method for computing ligand-charge distributions that optimize the electrostatic binding free energy has been applied to the small ribonuclease barnase and the results compared to the enzyme's natural inhibitor, the protein barstar. Previous studies have shown that electrostatic steering enhances the kinetic binding rate between these two proteins. Here we present evidence that barstar's electrostatics are also thermodynamically optimized for binding to barnase. Specifically, we found a group of residues important and optimized for tight-binding to barnase, which overlap substantially with those found important in fast-binding of the complex. Taken together, these results show that a limited set of side chains can be co-optimized for both kinetic and thermodynamic aspects of binding.

Electrostatic interactions are important in molecular binding and recognition. Analysis of atomic-resolution structures has shown that protein–protein interfaces are rich in polar and charged groups and that these groups generally make compensating hydrogen-bond interactions.[1] Nevertheless, the role of these electrostatic groups remains unclear. Previously it was thought that electrostatic interactions are strongly stabilizing due both to the obvious screened coulombic attraction and to the substantial reduction in stability that accompanies removal of part of a hydrogen-bonded network.[2–5] However, there is now a growing consensus that electrostatic interactions do not contribute strongly to affinity due to large desolvation penalties incurred by charged and polar groups on binding;[6–12] however, electrostatic effects are likely to enhance specificity.[6,13–21] To understand better the nature of binding electrostatics, we have developed an algorithm to design ligand-charge distributions that optimize the electrostatic binding free energy. The ligand-charge distribution is constructed to explicitly optimize the balance between unfavorable ligand-desolvation penalty (which is quadratic in ligand-charge distribution) and favorable interactions (which is linear in ligand-charge distribution) made with a fixed receptor.

Here we apply this procedure to investigate electrostatic interactions in the barnase–barstar protein complex, which is extremely tight-binding ($K_d \approx 10^{-14}$)[22] with many salt bridges and hydrogen bonds at the interface.[23] Mutational studies have implicated charged interfacial residues as large contributors to the thermodynamics of binding.[24] In addition, experimental and theoretical investigations have shown the importance of electrostatic steering in achieving rapid binding kinetics in this system.[25–30] In the present work, the focus is on the binding thermodynamics. In this report we show that only a small number of barstar side chain positions at the interface are electrostatically critical to binding. The side chain at each of these positions in the wild-type barstar matches the charge computed through optimization; moreover the detailed charge distributions are remarkably similar between actual and optimal for the most critical positions. These results indicate that barstar is very well optimized for tight binding *given that the residues have to be common*

amino acids. The side chains that are important and optimized for binding thermodynamics overlap substantially with those that are important for rapid binding kinetics.

Individual side chains of barstar, each constrained to have a total side chain charge of either $-1e$, $0e$, or $+1e$, were electrostatically optimized for tight binding to barnase. The atomic point charge values were contrained to the range $-0.85e$ to $+0.85e$. Each side chain was individually optimized, except for alanine and glycine side chains (glycine contains no side chain atoms in the polar hydrogen model used here, and alanine has only one side chain atom and so could not simultaneously satisfy the unit charge constraints and the atomic constraint) and residues 58, 60, 64, and 65 (missing side chain atoms in the X-ray crystal structure).[23] The results are illustrated in Figure 1 and are expressed as enhancements in the binding free energy over wild type computed for each optimization. Optimization leads only to relatively small improvement (less than one kcal/mol) for all but three of the side chains (Trp38, Gln72, and Glu78, discussed below). Moreover, the computed affinity is surprisingly insensitive to the total charge of most side chains. The exceptions, those most sensitive, are Asn33, Asp35, Trp38, Asp39, Thr42, Val73, and Glu76. It is rather remarkable that the optimized total charge computed for these seven side chains matches the actual total charge in barstar in each case.

Experimental studies have shown many of the sensitive side chains to be important in binding. The two most senstive sites, Asp35 and Asp39, are found to be especially important experimentally; Trp38, Thr42, and Glu76 are also important, but less sensitive to mutation. The remaining two positions, Asn33 and Val73, appear not to have been tested in mutagenesis studies.[24] Here we note that the senstivity of a site to mutation is not necessarily the same as its contribution to binding affinity. We have previously shown that this sensitivity correlates with the desolvation of a side chain upon binding, which only depends on the receptor binding geometry and not its charge distribution, but the affinity contribution of a side chain depends also on the charges.[31] However, using the definition for contribution described by Hendsch & Tidor,[32] we found that these two sets of side chains

overlap. Specifically, Asp35, Asp39, and Glu76 are major contributors, while Tyr29 and Thr42 contribute to a lesser extent.

The view afforded by these studies is that the binding interface consists of a small number of "hot spots" that are most sensitive in affecting the binding free energy. This scenerio of binding is similar to that resulting from the thorough alanine mutagensis experiments of Clackson & Wells, who found that only a small set of contact residues in a hormone-receptor interface maintains binding affinity.[33] Here the hot spots are Asp35 and Asp39, and to a lesser extent, the rest of the seven most sensitive residues. These residues are shown in Figure 2 and do not cover the whole region buried at the binding interface. The computations here focus exclusively on electrostatic sensitvity to binding; alanine scanning mutagensis includes electrostatic and hydrophobic effects, the latter of which can be especially significant for larger side chains.

Comparison of the detailed charge distributions for each optimum with its wild-type values (in the CHARMM19 parameter set)[34] reveals remarkable similarity for the charged side chains, especially Asp35 and Asp39, good similarity for the polar side chains Thr42 and Asn33, and rather poor similiarity for the hydrophobic side chains, Trp38 and Val73 (Table I). Moreover, the differences beween actual and optimal charge distributions result in only small (less than one kcal/mol) improvements in computed binding free energy for most of the charge-sensitive positions. This is direct evidence that wild-type charges of barstar are very well optimized for binding barnase, especially at the critical residues. Two exceptions (Trp38 and Glu76) and a third side chain (Gln72), identified from Figure 1 because optimization leads to a binding enhancement of over 1 kcal/mol, are the most underoptimzed residues and will be discussed next.

The most underoptimized residues, Trp38, Gln72, and Glu76, with -2.20, -1.04, and -1.84 kcal/mol improvement through optimization, might be thought of as potential sites for mutations to enhance binding affinity. To investigate this possibility, we computed the effects of single mutations in these side chains to all polar and charged residues. The optimal

charge distribution for residue 76 has a total side-chain charge of −1, the same as wild type; constraining the total charge to 0 gave rise to an optimal charge distribution with unchanged electrostatic binding free energy while the optimum achievable with total charge of +1 was 3.2 kcal/mol worse than wild type. Our mutation computation results are consistent with this; while all mutations of this site produced increased electrostatic binding free energy, the increase was most severe for the mutations to Arg or Lys. Mutation to Asp, the only other negatively charged amino acid, did not result in more favorable binding electrostatics. Although a set of optimal charges could be found from electrostatic optimization to improve binding affinity, it could not be readily incorporated into the 20 amino acids. That is, given the natural amino acids, Glu appears the optimal choice residue 76.

Next, the optimal charge distribution for residue 72 carries a negatve unit charge and is 0.2 kcal/mol better than the neutral optimum. A mutation of Gln72 to Glu or Asp resulted in about the same computed binding free energy as the wild type, both electrostatic and non-electrostatic, while other mutations yielded worse binding affinity. The optimum for Trp38 is neutral and 0.2 kcal/mol better than the negative optimum. We found a handful of mutations that improved binding electrostatics. However, because Trp has a large side chain, the mutated structures lost more non-electrostatic (hydrophobic) binding affinity than they gained in electrostatic. Although we found no single mutation to enhance binding, it might be possible to carry out double or triple mutations that include Trp38 and neighboring sites to compensate for size.

Whereas the calculations described above were carried out by locally optimizing each side chain individually with all others constant, when full optimization is carried out on all side chains simultaneously, a similar picture emerges. We carried out the full optimization with each side chain charge constrained to between $-1.5e$ and $1.5e$ and then reduced each side chain's total charge to the nearest integer in the optimum. Again, the results show an exact match in overall side chain charge with those of barstar for the side chains important in binding electrostatics, and the calculated charge distributions are close to those obtained

by single side chain optimization.

We have shown that barstar is extraordinarily well optimized electrostatically for binding barnase given the backbone fold; with the contraint of selecting from the natural amino acids, there is no straight-forward way to improve even the most underoptimized sites (though combinatorial mutagenesis involving Trp38 and its neighbors appears a good possibility.) The results presented here confirm experimental studies involving mutagenesis to identify residues important for binding. Our work goes beyond previous studies in showing that residues that are important are electrostatically optimized. To our knowledge this is the first demonstration that nature has made binding interfaces in which optimization of the balance of desolvation and interaction is a major consideration. It will be important to see the extent of electrostatic optimization present in weaker binding complexes, though we expect it to be less. We note that in other work, we have shown that further tuning of barstar electrostatics, beyond what is available with the 20 common amino acids, is computed to result in substantial binding enhancements.[31]

Although the present work is on the thermodymics of binding, the results have implications in the wider context of binding kinetics. The side chains identified here to be important and optimized for binding overlap with the four negatively charged residues (Asp35, Asp39, Glu76, and Glu80) that when mutated to Ala reduced the association rate constant by two-fold and increased the dissociation rate constant from five orders of magnitude for Asp39 to no significant change for Glu80.[26] We thus see a common subset of residues (Asp35, Asp39 and Glu76) that are important and optimized for both binding thermodynamics and kinetics. Indeed, it is likely that similar electrostatic effects are responsible for both.[30]

Methods

Electrostatic optimization

The electrostatic free energy of binding is the difference between the electrostatic free energy in the bound and unbound state. Here only the charges of a ligand side chain (denoted by $\{Q_{\mathrm{Lv}}\}$) were variable subject to optimization, while the rest were fixed. It has been shown that ΔG_{bind} can be written as a function of $\{Q_{\mathrm{Lv}}\}$ as follows,

$$\Delta G_{\mathrm{bind}} = \frac{1}{2}\left(\vec{Q}^{\mathrm{Lv}}\right)^T \overleftrightarrow{A}\left(\vec{Q}^{\mathrm{Lv}}\right) + \left(\vec{Q}^{\mathrm{Lv}}\right)^T \vec{B} + C \qquad (1)$$

where $\overleftrightarrow{A}$, $\vec{B}$, and C are independent of $\{Q_i^{\mathrm{Lv}}\}$, but depend on the geometry, the fixed ligand charges, and the receptor charges.[35,36,31] The matrix element A_{ij} is the difference in potential at the position of Q_i^{Lv} in the bound and unbound states when Q_j^{Lv} is set to a unit charge and all other charges are set to zero. The vector component B_i is the sum of potential at the position of Q_i when the receptor is charged as well as the difference in potential when the fixed ligand charges are charged in the bound and unbound state. C is the sum of three terms: the desolvation energy of the receptor, the desolvation energy of the ligand when only the fixed ligand charges are charged, and the interaction between the fixed ligand charges and the receptor. All of these coefficients were calculated by solving the linearized Poisson–Boltzmann equation using the computer program DELPHI and the CHARMM PARAM19 parameter set[34] for charges and radii.[37–39] The reader is referred to ref. 31 for the details of the DELPHI calculations and ref. 40 for the preparation of the protein structure used in these computations. Eq. (1), together with the constraints on individual charges and the total side-chain charge, was solved with the computer program LOQO.[41]

Mutation computations

To construct the mutated bound structure, rotamers from the Dunbrack and Karplus backbone-dependent rotamer library were used as initial structures.[42] Each structure was minimized using CHARMM.[34] The minimized structure was used both in the bound and

unbound states. Electrostatic contribution to the binding free energy was calculated with DELPHI, while the non-electrostatic with CHARMM.

REFERENCES

[1] J. Janin and C. Chothia. The structure of protein–protein recognition sites. *J. Biol. Chem.* **265**: 16027–16030 (1990).

[2] A. R. Fersht. Conformational equilibria and the salt bridge in chymotrypsin. *Cold Spring Harbor Symp. Quant. Biol.* **36**: 71–73 (1971).

[3] A. R. Fersht. Conformational equilibria in α- and δ-chymotrypsin. The energetics and importance of the salt bridge. *J. Mol. Biol.* **64**: 497–509 (1972).

[4] A. Horovitz, L. Serrano, B. Avron, M. Bycroft, and A. R. Fersht. Strength and co-operativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol.* **216**: 1031–1044 (1990).

[5] D. E. Anderson, W. J. Becktel, and F. W. Dahlquist. pH-induced denaturation of proteins: A single salt bridge contributes 3–5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* **29**: 2403–2408 (1990).

[6] Z. S. Hendsch and B. Tidor. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**: 211–226 (1994).

[7] A.-S. Yang and B. Honig. Free energy determinants of secondary structure formation: I. α-helices. *J. Mol. Biol.* **252**: 351–365 (1995).

[8] A.-S. Yang and B. Honig. Free energy determinants of secondary structure formation: II. Antiparallel β-sheets. *J. Mol. Biol.* **252**: 366–376 (1995).

[9] A.-S. Yang, B. Hitz, and B. Honig. Free energy determinants of secondary structure formation: III. β-turns and their role in protein folding. *J. Mol. Biol.* **259**: 873–882 (1996).

[10] L. Wang, T. O'Connell, A. Tropsha, and J. Hermans. Energetic decomposition of the α-helix–coil equilibrium of a dynamic model system. *Biopolymers* **39**: 479–489 (1996).

[11] C. D. Waldburger, J. F. Schildbach, and R. T. Sauer. Are buried salt bridges important for protein stability and conformational specificity? *Nature Struct. Biol.* **2**: 122–128 (1995).

[12] W. C. Wimley, K. Gawrisch, T. P. Creamer, and S. H. White. Direct measurement of salt-bridge solvation energies using a peptide model system: Implications for protein stability. *Proc. Natl. Acad. Sci. U.S.A.* **93**: 2985–2990 (1996).

[13] C. Tanford, P. K. De, and V. G. Taggart. The role of the α-helix in the structure of proteins. Optical rotatory dispersion of β-lactoglobulin. *J. Am. Chem. Soc.* **82**: 6028–6034 (1960).

[14] C. H. Paul. Building models of globular protein molecules from their amino acid sequences. I. Theory. *J. Mol. Biol.* **155**: 53–62 (1982).

[15] C. V. Sindelar, Z. S. Hendsch, and B. Tidor. Effects of salt bridges on protein structure and design. *Protein Sci.* **7**: 1898–1914 (1998).

[16] S. A. Potekhin, V. N. Medvedkin, I. A. Kashparov, and S. Y. Venyaminov. Synthesis and properties of the peptide corresponding to the mutant form of the leucine zipper of the transcriptional activator GCN4 from yeast. *Protein Eng.* **7**: 1097–1101 (1994).

[17] L. Gonzalez, Jr., J. J. Plecs, and T. Alber. An engineered allosteric switch in leucine-zipper oligomerization. *Nature Struct. Biol.* **3**: 510–515 (1996).

[18] L. Gonzalez, Jr., R. A. Brown, D. Richardson, and T. Alber. Crystal structures of a single coiled-coil peptide in two oligomeric states reveal the basis for structural polymorphism. *Nature Struct. Biol.* **3**: 1002–1010 (1996).

[19] L. Gonzalez, Jr., D. N. Woolfson, and T. Alber. Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nature Struct. Biol.* **3**: 1011–1018 (1996).

[20] E. K. O'Shea, R. Rutkowski, and P. S. Kim. Mechanism of specificity in the Fos–Jun oncoprotein heterodimer. *Cell* **68**: 699–708 (1992).

[21] M. John, J. Briand, M. Granger-Schnarr, and M. Schnarr. Two pairs of oppositively charged amino acids from Jun and Fos confer heterodimerization to GCN4 leucine zipper. *J. Biol. Chem.* **269**: 16247–16253 (1994).

[22] G. Schreiber and A. R. Fersht. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry* **32**: 5145–5150 (1993).

[23] A. M. Buckle, G. Schreiber, and A. R. Fersht. Protein–protein recognition: Crystal structure analysis of a barnase–barstar complex at 2.0-Å resolution. *Biochemistry* **33**: 8878–8889 (1994).

[24] G. Schreiber and A. R. Fersht. Energetics of protein–protein interactions: Analysis of the barnase–barstar interface by single mutants and double mutant cycles. *J. Mol. Biol.* **248**: 478–486 (1995).

[25] G. Schreiber and A. R. Fersht. Rapid, electrostatically assisted association of proteins. *Nature Struct. Biol.* **3**: 427–431 (1996).

[26] G. Schreiber, A. M. Buckle, and A. R. Fersht. Stability and function: Two constraints in the evolution of barstar and other proteins. *Structure* **2**: 945–951 (1994).

[27] R. R. Gabdoulline and R. C. Wade. Simulation of the diffusional association of barnase and barstar. *Biophys. J.* **72**: 1917–1929 (1997).

[28] M. Vijayakumar, K. Y. Wong, G. Scheiber, A. R. Fersht, A. Szabo, and H. X. Zhou. Electrostatic enhancement of diffusion-controlled protein–protein association: Comparison of theory and experiment on barnase and barstar. *J. Mol. Biol.* **278**: 1015–1024 (1998).

[29] C. J. Camacho, Z. Weng, S. Vadja, and C. DeLisi. Free energy landscapes of encounter complexes in protein–protein association. *Biophys. J.* **76**: 1166–1178 (1999).

[30] T. Selzer and G. Schreiber. Predicting the rate enhancement of protein complex formation from the electrostatic energy of interaction. *J. Mol. Biol.* **287**: 409–419 (1999).

[31] L.-P. Lee and B. Tidor. Optimization of binding electrostatics: Charge complementarity in the barnase–barstar protein complex. *In preparation.*

[32] Z. S. Hendsch and B. Tidor. Electrostatic interactions in the gcn4 leucine zipper: Substantial contributions arise from intramolecular interactions enhanced on binding. *Protein Sci.* **In press.** (1999).

[33] T. Clackson and J. A. Wells. A hot spot of binding energy in a hormone–receptor interface. *Science (Washington, D.C.)* **267**: 383–386 (1995).

[34] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**: 187–217 (1983).

[35] L.-P. Lee and B. Tidor. Optimization of electrostatic binding free energy. *J. Chem. Phys.* **106**: 8681–8690 (1997).

[36] E. Kangas and B. Tidor. Optimizing electrostatic affinity in ligand–receptor binding: Theory, computation, and ligand properties. *J. Chem. Phys.* **109**: 7522–7545 (1998).

[37] K. A. Sharp and B. Honig. Calculating total electrostatic energies with the nonlinear Poisson–Boltzmann equation. *J. Phys. Chem.* **94**: 7684–7692 (1990).

[38] M. K. Gilson, K. A. Sharp, and B. H. Honig. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J. Comput. Chem.* **9**: 327–335 (1988).

[39] K. A. Sharp and B. Honig. Electrostatic interactions in macromolecules: Theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* **19**: 301–332 (1990).

[40] L. T. Chong, Z. S. Hendsch, and B. Tidor. In preparation.

[41] R. J. Vanderbei. LOQO: An interior-point code for quadratic programming. *Optimization methods and Software* **To appear.** (1999).

[42] R. L. Dunbrack and M. Karplus. Backbone-dependent rotamer library for proteins: Application to side-chain prediction. *J. Mol. Biol.* **230**: 543–574 (1993).

TABLES

TABLE I. Comparison of wild-type (wt) and optimized (opt) side-chain atomic charge values.[a]

Asp35	CB	CG	OD1	OD2							
opt	0.08	0.50	−0.72	−0.85							
wt	−0.16	0.36	−0.60	−0.60							
Asp39	CB	CG	OD1	OD2							
opt	−0.05	0.85	−0.85	−0.78							
wt	−0.16	0.36	−0.60	−0.60							
Glu76	CB	CG	CD	OE1	OE2						
opt	0.46	0.85	−0.85	−0.85	−0.61						
wt	0.00	−0.16	0.36	−0.60	−0.60						
Thr42	CB	OG1	HG1	CG2							
opt	0.27	−0.85	0.33	0.25							
wt	0.25	−0.65	0.40	0.00							
Asn33	CB	CG	OD1	ND2	HD21	HD22					
opt	0.21	−0.33	−0.09	0.09	0.01	0.11					
wt	0.00	0.55	−0.55	−0.60	0.30	0.30					
Val73	CB	CG1	CG2								
opt	0.85	−0.46	−0.39								
wt	0.00	0.00	0.00								
Trp38	CB	CG	CD2	CE2	CE3	CD1	NE1	HE2	CZ2	CZ3	CH2
opt	0.27	−0.66	−0.78	−0.43	0.62	0.67	0.04	0.06	0.37	−0.85	0.67
wt	0.00	−0.03	0.10	−0.04	−0.03	0.06	−0.36	0.30	0.00	0.00	0.00
Gln72	CB	CG	CD	OE1	NE2	HE21	HE22				
opt	−0.67	−0.85	0.85	0.85	−0.18	−0.85	0.85				
wt	0.00	0.00	0.55	−0.55	−0.60	0.30	0.30				

[a] All charge magnitudes are in units of the magnitude of charge on the electron, e.

FIGURES

FIG. 1. Enhancements in the binding free energy for barnase to barstar computed in individual optimizations of each side chain with the total side chain charge constrained to be neutral, $+1e$, or $-1e$.

FIG. 2. Space-filling model of barstar side chains at binding interface to barnase. Light color indicates little or no sensitivity to overall side chain charge; dark color indicates the seven residues with strong sensitivity to side chain charge.

Figure 1

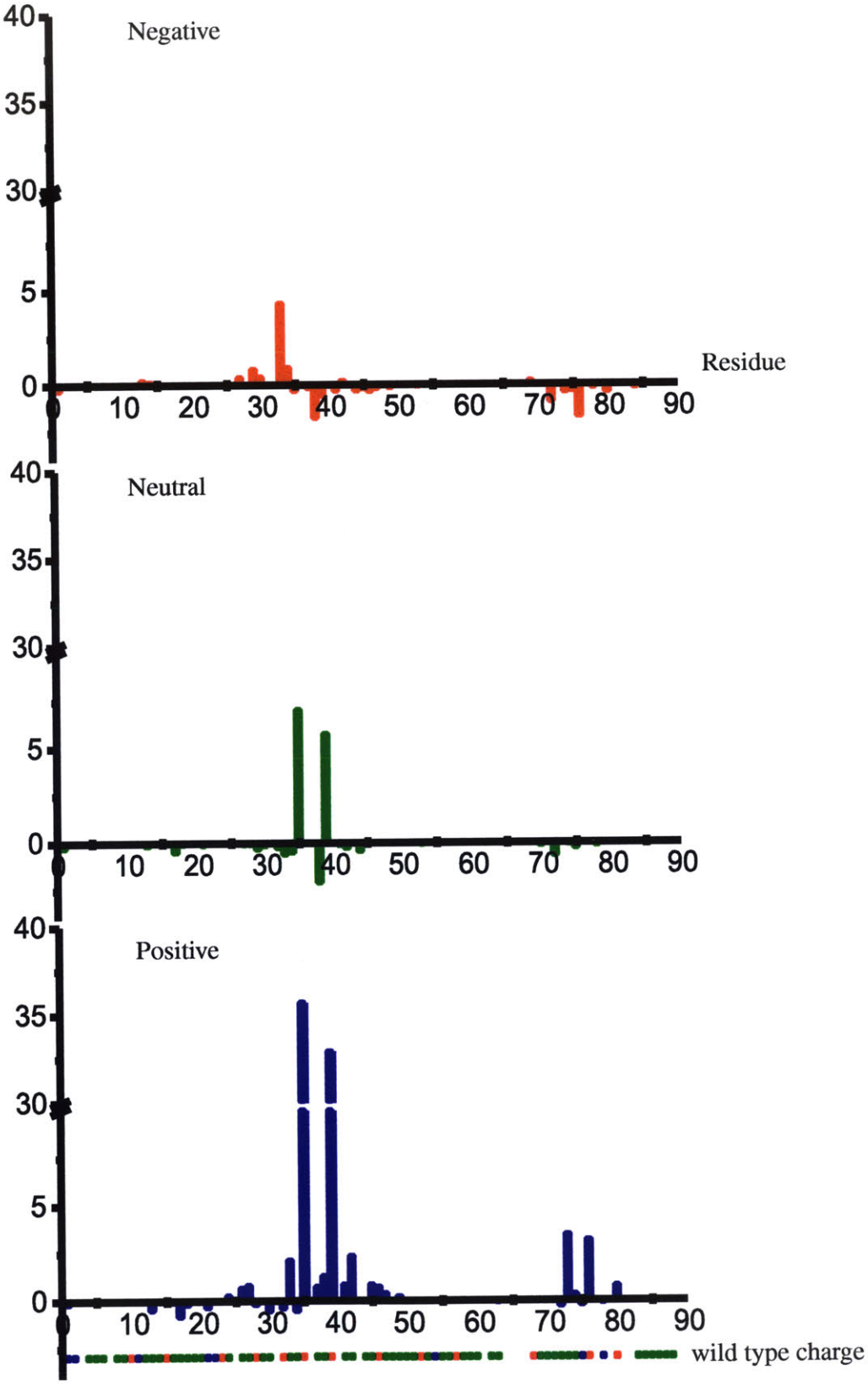
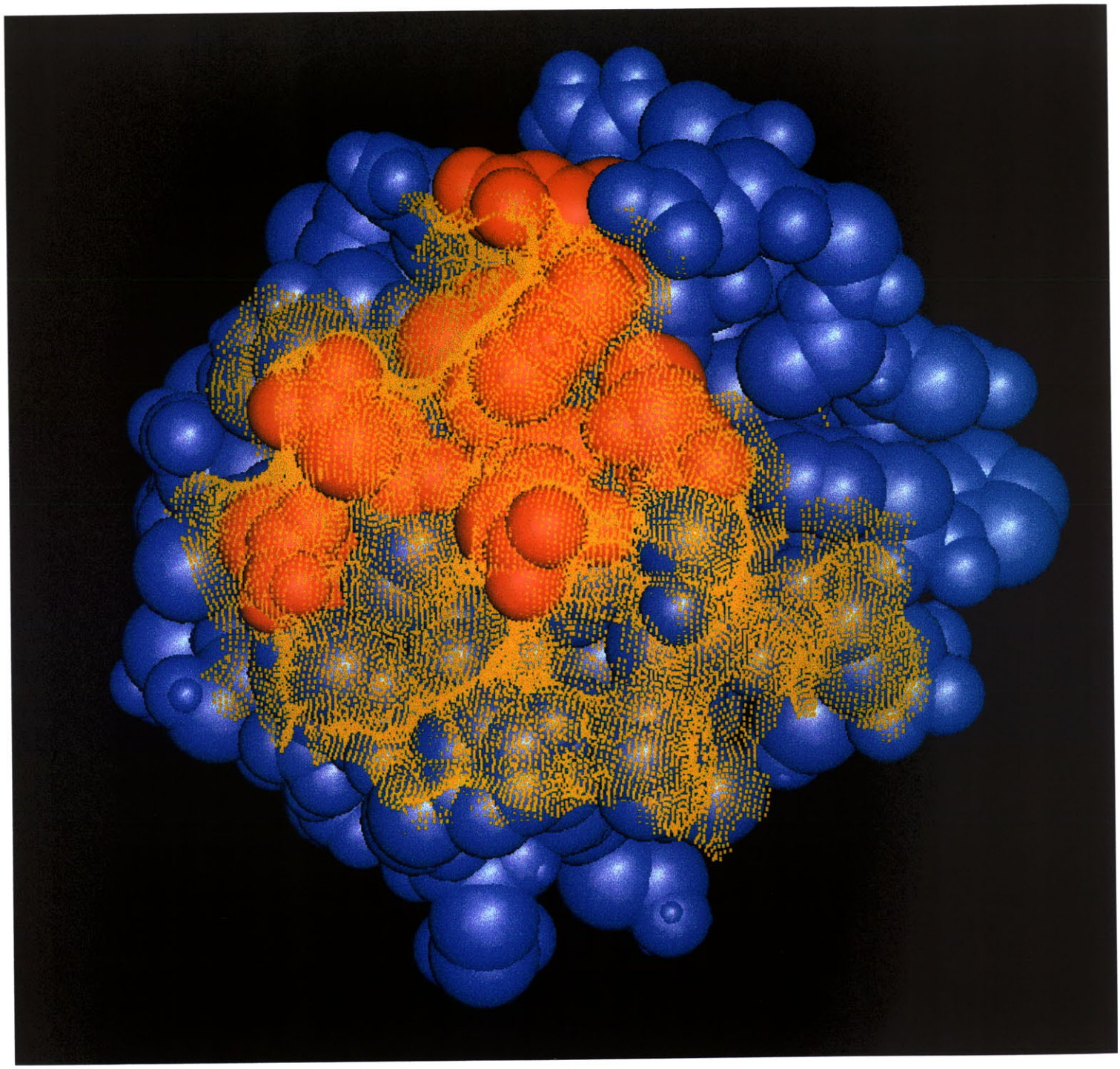

Negative
Neutral
Positive
Residue
wild type charge
0
5
30
35
40
10
20
30
40
50
60
70
80
90

Figure 2



Chapter VI

General Conclusion

The present work might be seen as an attempt to apply physics to an important biological mechanism, namely the noncovalent binding of two biomolecules. Biological systems, as complex as they are, obey the same laws of physics as simpler matters. Although the fundamental physical laws now known more than cover that required for biological processes, understanding of much of biology in terms of physics remains a challenge. Due to the large number of non-trivially interacting particles that make up a biological sytem, a direct application of basic equations from physics is often not possible. Approximate theories and models that capture the essence of the aspects of interests are needed, and in the last few decades, many useful theoretical and computational models have emerged. An example is the continuum electrostatic model for biomolecules in aqueous environment which is both efficient and accurate in calculating electrostatic free energy.

We adopted the continuum model for calculating the electrostatic binding free energy and wrote a quadratic equation relating the electrostatic binding free energy and the charge distribution of one molecule (the ligand) given that of the other (the receptor) is known. The electrostatic binding free energy was then minimized to find an optimal ligand-charge distribution. While the electrostatic optimization scheme has been derived in the basic continuum model described by the linearized Poisson–Boltzmann equation, it is applicable to general linear response models, where the dependence on charges is linear for electrostatic potential and quadratic for electrostatic free energy. This is noted because the on-going efforts to derive more efficient continuum models to replace the Poisson–Boltzmann equation

may bring about a faster way for calculating the electrostatic free energy, but electrostatic optimization will likely remain applicable. For examples, new models in the spirit of the generalized Born model are generally linear response models. In fact, given the accuracy of linear response models for biomolecular systems, linear response is likely to remain in use.

Application of the electrostatic optimization method to the barnase–barstar protein complex demonstrates its ready application in real biomolecular systems. We have seen its scope of use both as a tool to understand the degree of optimality of electrostatic interactions in natural complexes and as a guide for designing more favorable interactions. In the barnase–barstar study, we found that the wild-type barstar is very well optimized given that the residues have to come from the set of 20 common amino acids, but in the laboratory, where chemical compounds can be made *not* bound by this limitation, tighter-binding ligands could be found to inhibit the binding site. This study is just one of the first examples in using electrostatic optimization to investigate a diverse range of problems. For instance, it has been applied to study the catalytic mechanism of glutaminyl-tRNA synthetase, which plays a key role in protein synthesis. Electrostatic optimization is also being incorporated into rational drug design, while its by-product, a measure of electrostatic complementarity, is being developed to analyse genomic data. It is hoped that the use of the electrostatic optimization scheme presented here will help lead to greater understanding of the principles underlying molecular binding and better formulation of molecular design methodology.

Appendix

Computation of Electrostatic Complements to Proteins: A Case of Charge Stabilized Binding

Lillian T. Chong, Sara. E. Dempster, Zachary S. Hendsch,

Lee-Peng Lee and Bruce Tidor

Reprint from *Protein Science* **7:** 206-210 (1998)

Protein Science (1998), 7:206–210. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Computation of electrostatic complements to proteins: A case of charge stabilized binding

LILLIAN T. CHONG, SARA E. DEMPSTER, ZACHARY S. HENDSCH, LEE-PENG LEE, and BRUCE TIDOR

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307

(RECEIVED August 28, 1997; ACCEPTED October 16, 1997)

Abstract: Recent evidence suggests that the net effect of electrostatics is generally to destabilize protein binding due to large desolvation penalties. A novel method for computing ligand-charge distributions that optimize the tradeoff between ligand desolvation penalty and favorable interactions with a binding site has been applied to a model for barnase. The result is a ligand-charge distribution with a favorable electrostatic contribution to binding due, in part, to ligand point charges whose direct interaction with the binding site is unfavorable, but which make strong intra-molecular interactions that are uncloaked on binding and thus act to lessen the ligand desolvation penalty.

Keywords: continuum electrostatics; electrostatic complement; protein binding; rational ligand design

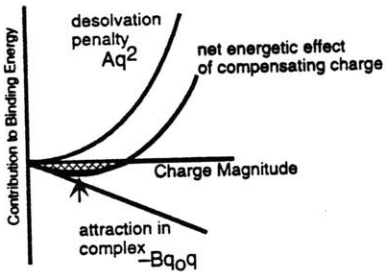
The relative strengths of interactions involved in protein folding and binding are of fundamental importance for our understanding of these processes and for our ability to design modified or novel proteins and tight-binding ligands. Recent theoretical (Hendsch & Tidor, 1994; Yang & Honig, 1995; Wang et al., 1996) and experimental (Waldburger et al., 1995; Wimley et al., 1996) studies have emphasized the large electrostatic desolvation penalty due to polar and charged groups incurred in protein folding and have led to the realization that the desolvation penalty is not generally recovered in favorable interactions created in the folded state (although exceptions have been noted: Hendsch & Tidor, 1994; Lounnas & Wade, 1997; and, in the context of binding, Xu et al., 1997). For example, Sauer and co-workers replaced a triad of hydrogen-bonded, salt bridging groups with hydrophobic groups (which pay essentially no electrostatic desolvation penalty but recover negligible electrostatic interactions on folding) in Arc repressor and measured enhanced stabilities of 1–2$\frac{1}{2}$ kcal/mol per monomer (Waldburger et al., 1995). Although electrostatic interactions may not generally contribute to stability, they are likely to have a substantial effect on specificity due to the cost of burying but not compensating polar or charged groups (Tanford et al., 1960; Paul, 1982; Hendsch & Tidor, 1994). Similar results are emerging for protein binding as well. Calculations for a number of protein complexes show that the net effect of electrostatic interactions is generally to destabilize the docking of two pre-conformed molecules (Novotny & Sharp, 1992; Misra et al., 1994a, 1994b; Sharp, 1996; Shen & Wendoloski, 1996; Bruccoleri et al., 1997; Novotny et al., 1997). This raises the important question of whether it is possible to design a binding partner for a given site such that the electrostatic contribution to binding is favorable. Here we utilize a recently introduced electrostatic optimization strategy to address this question (Lee & Tidor, 1997).

We have developed a scheme (outlined in Fig. 1) to optimize the electrostatic component of the binding free energy within the continuum model (Lee & Tidor, 1997). When a ligand binds to a receptor, there are two competing forces: the desolvation or dehydration penalty (the loss of favorable interaction between each molecule and the solvent) acts against the favorable interaction between the two molecules. Viewing the charge distribution of the receptor as fixed and that of the ligand as variable, the unfavorable ligand dehydration penalty increases with the square of the ligand-charge distribution, whereas the favorable interactions increase only linearly. This results in the existence of an optimal charge distribution that represents the most favorable tradeoff of dehydration and interaction. For the special case in which the ligand and the complex can be considered spherical regions of low dielectric embedded in a higher dielectric solvent, we represent the ligand-charge distribution by a set of variable multipoles whose values are chosen to minimize the electrostatic binding free energy. The monopole component (i.e., the total ligand charge) may either be variable or constrained (e.g., to an integer) to facilitate subsequent molecular construction (Lee & Tidor, 1997).

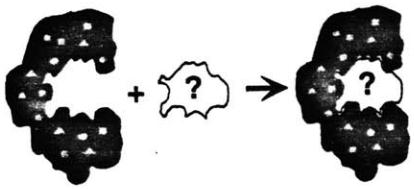
We have applied this method to the small bacterial ribonuclease from *Bacillus amyloliquefaciens*, barnase. The point-charge distribution from barnase was used for the receptor, and locations for the docking of a series of spherical low-dielectric ligands (of radius 8, 10, and 11 Å) were chosen using molecular graphics. The bound complexes were modeled using the point-charge representation for barnase, the variable multipole distribution for the ligand, and a 31-Å radius spherical dielectric boundary.

Reprint requests to: Bruce Tidor, Department of Chemistry, Room 6-135, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307; e-mail: tidor@mit.edu.

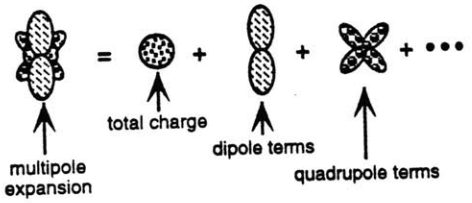
If Coulomb's law interactions between ligand and receptor were the dominant electrostatic effect, then tight-binding ligands would be constructed with very large charges to compensate charges in the active site. For instance, the ligand with charge of −7 would bind better than that with −2, which would bind better than that with −1.



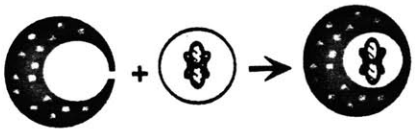
Because ligands pay a large desolvation penalty that increases as the square of the charge (q^2), highly charged ligands are not necessarily tight binders. In fact, because the favorable interaction is only linear in the charge (q), the net contribution to the binding energy as a function of ligand charge is a shifted parabola with a minimum (optimal) binding energy indicated by the arrow.



The primary methodological advance applied here is the development of a technique to solve for the optimal ligand charge distribution — that is, the charge distribution resulting in the most favorable electrostatic contribution to binding — given an arbitrary receptor shape and charge distribution (indicated by the blue region and red symbols, respectively).



For a spherical ligand, this can be achieved by using an arbitrary set of multipoles to represent any ligand charge distribution and solving for that set of multipoles that minimizes the electrostatic contribution to binding. The multipoles describe the total charge, dipolar terms, quadrupolar terms, etc. of the charge distribution.



The theory has been developed and implemented for the special case in which the ligand and the complex are perfect spheres. Work is currently under way to extend the methodology to arbitrarily shaped ligands and complexes.

Fig. 1. Rationale for electrostatic optimization.

Optimal ligand-charge distributions were determined as sets of multipoles. The total charges of the optimal ligands (for the 8, 10, and 11-Å ligands, respectively) were $-1.16e$, $-1.30e$, and $-1.36e$. A separate computation was carried out in which the total charge was constrained to the integer value of $-1e$ for the 8-Å ligand, whose unconstrained total charge was closest to an integer value. The binding energy converged relatively quickly as a function of multipole order and was dominated by the monopole and dipole terms (see Fig. 2A).

To determine the total electrostatic contribution to binding, we computed the ligand dehydration penalty and ligand–receptor interaction energy analytically (because in each case the dielectric boundary was spherical) and computed the receptor dehydration penalty numerically with the DELPHI computer program (Gilson et al., 1988; Gilson & Honig, 1988; Sharp & Honig, 1990). In each of the four cases, the overall electrostatic binding free energy was favorable, including the docking of the ligand with the total charge constrained to $-1e$; this is in contrast to computations on many natural complexes, for which the electrostatic contribution to the binding free energy is unfavorable (Novotny & Sharp, 1992; Misra et al., 1994a, 1994b; Sharp, 1996; Shen & Wendoloski, 1996; Bruccoleri et al., 1997; Novotny et al., 1997). Table 1 shows the

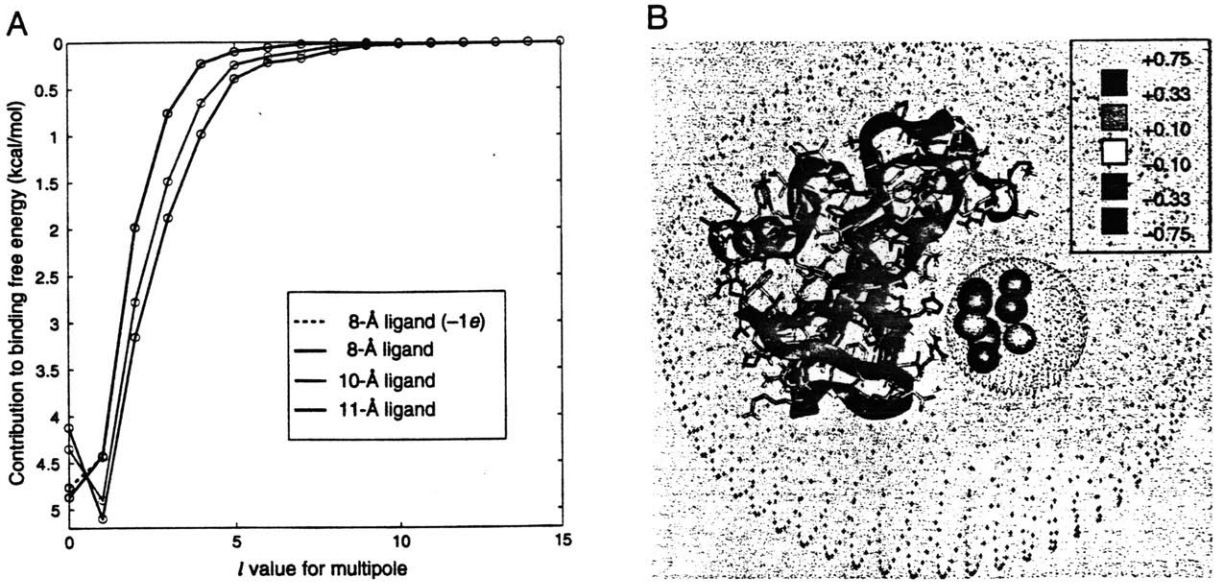


Fig. 2. A: The contribution of the first 16 multipoles to the electrostatic binding free energy is shown for each of the optimized ligand-charge distributions, where $l = 0$ corresponds to the monopole contribution, $l = 1$ corresponds to the dipole contribution, $l = 2$ to the quadrupole contribution, etc. **B:** The structure of barnase and the 11 point-charge distribution fit to the optimal 8-Å ligand (purple dotted sphere) with total charge constrained to $-1e$ is shown. Values of the point charges are indicated by their color (see legend). The dielectric boundary for the complex is indicated by the orange dotted sphere. This figure was prepared with the QUANTA program (Molecular Simulations Inc., San Diego).

contributions to the electrostatic binding free energy for each ligand. The docking of the smallest ligand had the most favorable electrostatic binding free energy (−4.8 kcal/mol), which was diminished by only a tenth of a kcal/mol when the total charge was constrained to $-1e$, due in this case essentially only to changes in the contribution of the monopole term.

The multipole distribution for the 8-Å ligand with the total charge constrained to $-1e$ could be well represented by a set of only 11 point charges (Fig. 2B). These partial charges were fit from a 2-Å cubic grid and ranged from $-0.62e$ to $+0.42e$. The charge locations were restricted to be at least 1 Å from the ligand dielectric boundary because atom-centered point charges must be at least an atomic radius from the molecular surface. The fit between the point charges and the optimal multipole distribution was quite good, as gauged by the small difference (0.2 kcal/mol) between their binding free energies.

The point-charge distribution fit reveals a number of interesting features of ligand design. First, the barnase enzyme carried a net charge of $+1e$, as modeled here, and presented a large number of positive charges facing the ligand binding site (Fig. 2B). The ligand point-charge distribution (which was chosen to have a total charge of $-1e$, but unconstrained optimizations yielded total charges of $-1.16e$ to $-1.36e$) had seven negative partial charges ($-0.13e$ to $-0.62e$) arranged to make strong, favorable electrostatic inter-

Table 1. *Electrostatically optimized ligand-charge distributions*

Ligand radius (Å)	Total charge (e)	$\Delta G_{hyd.L}$[a] (kcal/mol)	$\Delta G_{int.L\text{-}R}$[b] (kcal/mol)	$\Delta G_{hyd.R}$[c] (kcal/mol)	$\Delta G_{binding}$[d] (kcal/mol)
8	−1.16	12.45	−24.91	7.67 (0.02)	−4.78
8	−1.00[e]	11.20	−23.56	7.67 (0.02)	−4.69
10	−1.30	14.74	−29.48	10.23 (0.02)	−4.51
11	−1.36	16.21	−32.42	11.71 (0.02)	−4.50

[a]Ligand desolvation contribution to electrostatic binding free energy.
[b]Screened electrostatic ligand–receptor interaction contribution.
[c]Receptor desolvation contribution. The values are the average of 10 translations on the finite-differences grid. The numerical uncertainty, given in parentheses, is twice the standard deviation of the mean.
[d]Overall electrostatic binding free energy.
[e]The total charge was constrained to $-1e$.

actions (−1.6 to −9.4 kcal/mol) with barnase. Second, the point-charge distribution also contained four positive partial charges. Each of these charges (+0.17e to +0.42e) made unfavorable interactions with barnase (+1.4 to +3.3 kcal/mol) and had an unfavorable desolvation penalty (+0.3 to +1.3 kcal/mol) upon binding. However, each of the four positive partial charges made improved electrostatic interactions within the ligand (−1.7 to −4.6 kcal/mol) due to reduced solvent screening upon binding. That is, none of these four positive partial charges made favorable interactions with the receptor, yet leaving any one out diminished the binding free energy by up to 0.8 kcal/mol and leaving all four out diminished binding by 4.7 kcal/mol. This binding enhancement can be traced to favorable intra-ligand interactions that are weaker in the unbound state due to screening by solvent and stronger in the bound state due to exclusion of solvent by the receptor. An equally valid description of the role of these four positive partial charges is that they reduce the ligand dehydration penalty by more than the repulsion they introduce with the receptor. Although favorable intramolecular electrostatic interactions are frequently observed within individual binding partners, their role in enhancing molecular association due to a reduced effective dielectric constant on binding has not generally been appreciated.

One limitation of the geometries used in this study is that the ligand binding site is enclosed within the receptor sphere, which may artificially reduce the receptor dehydration penalty. This penalty was recomputed for a number of geometries in which a tunnel of low dielectric was removed from the unbound receptor (tunnel diameters up to 16 Å were computed). This increased the receptor dehydration contribution by up to 1.1 kcal/mol, which resulted in a total electrostatic binding free energy that was still favorable by more than 3.5 kcal/mol.

Another result of the geometries chosen here is that ligand "atoms" could not approach receptor atoms closely enough to make direct hydrogen-bond interactions. This is largely a result of the use of a spherical ligand, although we purposely chose to maintain a substantial distance of closest approach between receptor point charges and the receptor surface (2.3 Å) so that the finite-difference computation of the receptor dehydration penalty could be done especially accurately (Gilson et al., 1988). The full consequences of using a spherical ligand and a spherical bound complex are unknown at this point. It seems likely that the general picture of electrostatic complementarity will be similar for flat, curved, and bumpy interfaces, but the details might very well be different in important ways. As the algorithm is extended to treat actual molecular shapes, it will be important to see whether electrostatically favorable binding energies can still be achieved.

In summary, we have applied a novel charge optimization scheme to compute electrostatic complements for the protein barnase and found that these charge distributions (1) produce favorable electrostatic binding free energies, and (2) can be fit to a relatively small number of partial charges that are of small enough magnitude (under 0.7e here) and of sufficient spacing (at least 2 Å apart) that molecules may be conceived with similar point-charge distributions. By using these electrostatic complements as a guide, it is possible that this scheme will prove useful in the design of tight-binding ligands, where significant improvements in electrostatic interactions may be realizable. The covalent and non-covalent constraints of chemistry may make it impossible to construct molecules as effective as the complements described here, but it may be possible to approach this optimum. The algorithm presented here is rapid (multipole distributions can be determined in about an hour on a laboratory workstation) because the ligand and complex were chosen to have spherical geometries, allowing analytic solutions to the electrostatic equations to be used (Lee & Tidor, 1997). Extensions to realistic molecular shapes are in progress; such computations are substantially slower due to the need to repeatedly solve the linearized Poisson–Boltzmann equation numerically for all members of a representative charge basis set, but we anticipate that they may be even more useful in molecular design studies. A further result of this study is the observation that strong intramolecular electrostatic interactions that are screened by solvent in the unbound state can be uncloaked by the exclusion of solvent in the bound state. The consequent enhancement can contribute substantially to binding free energy and is likely to be effective for arbitrarily shaped ligands and complexes.

Methods: The coordinates for the barnase receptor were taken from the crystal structure of the barnase–barstar complex (chain A of 1BRS in the Brookhaven Protein Data Bank; Bernstein et al., 1977; Buckle et al., 1994). Polar-hydrogen positions were built using the HBUILD facility of CHARMM (Brooks et al., 1983; Brünger & Karplus, 1988). Partial atomic charges for barnase were taken from the CHARMM PARAM19 parameter set (Brooks et al., 1983). In other work (ZSH, CV Sindelar, & BT, unpubl. obs.) we have found this parameter set to give results similar to the PARSE parameter set (Sitkoff et al., 1994). The spherical boundary defining the complex was centered at (−6.781, 66.603, −38.750), enclosing all point charges and radii in the barnase–barstar complex. Ligand spheres were chosen such that they fit wholly within the position of the barstar ligand. Ligand spheres of radius 8, 10, and 11 Å were centered at (−13.788, 72.885, −41.659), (−15.317, 74.256, −42.294), and (−16.0288, 74.8939, −42.5893), respectively. The optimal multipole expansion for each ligand was computed with the algorithm of Lee and Tidor (1997). Multipole expansions were evaluated to 58 poles ($l_{max} = l_{cut} = 57$). The ligand and complex interior dielectric constant (ϵ_{int}) and the solvent exterior dielectric constant (ϵ_{ext}) were 4 and 80, respectively.

The receptor dehydration penalty was computed numerically with the DELPHI computer program (Gilson et al., 1988; Gilson & Honig, 1988; Sharp & Honig, 1990) as the difference in the total electrostatic energy for the barnase charge distribution embedded in a low-dielectric region representing the bound and unbound state. The Poisson equation was solved (corresponding to zero ionic strength) with the low-dielectric ($\epsilon_{int} = 4$) region for the receptor defined as the spherical region for the complex with the spherical ligand region (and, for some computations, a tunnel leading from the spherical ligand to the exterior) subtracted. The remaining volume had $\epsilon_{ext} = 80$. Each calculation was run using a finite-difference grid consisting of 131 points in each of the three Cartesian directions. A dual-level focusing approach filling 23% and then 92% of the grid was used. The results presented are the average of ten translations at the fine grid spacing (corresponding to 0.52 Å per grid unit).

Point charges were fit to the multipole distribution (up to $l = 27$) on a 2-Å Cartesian grid of point-charge centers using singular-value decomposition (Press et al., 1992; Strang, 1993). Iterative cycles were used in which point-charge locations were successively removed from the basis set to achieve a good fit with a small number of point charges. In each iteration singular-value decomposition was used to perform a least-squares fit of the multipoles to the current basis set, point charges with very small magnitudes

were removed from the basis, and the multipoles were re-fit to the new basis.

Acknowledgments: The authors thank Barry Honig for making the DELPHI program package available. We also thank Barry Honig, Wayne L. Hubbell, Peter S. Kim, Robert T. Sauer, Lawrence J. Stern, Gilbert Strang, and James R. Williamson for helpful discussions, comments, and suggestions. This work was supported by the National Institutes of Health (GM47678 and GM56552) and the MIT Science Partnership Fund.

References

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem 4*:187–217.

Bruccoleri RE, Novotny J, Davis ME, Sharp KA. 1997. Finite difference Poisson-Boltzmann electrostatic calculations: Increased accuracy achieved by harmonic dielectric smoothing and charge antialiasing. *J Comput Chem 18*:268–276.

Brünger AT, Karplus M. 1988. Polar hydrogen positions in proteins: Empirical energy placement and neutron diffraction comparison. *Proteins Struct Funct Genet 4*:148–156.

Buckle AM, Schreiber G, Fersht AR. 1994. Protein–protein recognition: Crystal structure analysis of a barnase–barstar complex at 2.0-Å resolution. *Biochemistry 33*:8878–8889.

Gilson MK, Honig B. 1988. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Proteins Struct Funct Genet 4*:7–18.

Gilson MK, Sharp KA, Honig BH. 1988. Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J Comput Chem 9*:327–335.

Hendsch ZS, Tidor B. 1994. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci 3*:211–226.

Lee L-P, Tidor B. 1997. Optimization of electrostatic binding free energy. *J Chem Phys 106*:8681–8690.

Lounnas V, Wade, RC. 1997. Exceptionally stable salt bridges in cytochrome P450cam have functional roles. *Biochemistry 36*:5402–5417.

Misra VK, Hecht JL, Sharp KA, Friedman RA, Honig B. 1994a. Salt effects on protein-DNA interactions: The λcI repressor and EcoRI endonuclease. *J Mol Biol 238*:264–280.

Misra VK, Sharp KA, Friedman RA, Honig B. 1994b. Salt effects on ligand–DNA binding: Minor groove binding antibiotics. *J Mol Biol 238*: 245–263.

Novotny J, Bruccoleri RE, Davis M, Sharp KA. 1997. Empirical free energy calculations: A blind test and further improvements to the method. *J Mol Biol 268*:401–411.

Novotny J, Sharp K. 1992. Electrostatic fields in antibodies and antibody/antigen complexes. *Prog Biophys Mol Biol 58*:203–224.

Paul CH. 1982. Building models of globular protein molecules from their amino acid sequences. I. Theory. *J Mol Biol 155*:53–62.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 1992. *Numerical recipes in C: The art of scientific computing*, 2nd ed. Cambridge: Cambridge University Press.

Sharp KA. 1996. Electrostatic interactions in hirudin–thrombin binding. *Biophys Chem 61*:37–49.

Sharp KA, Honig B. 1990. Electrostatic interactions in macromolecules: Theory and applications. *Annu Rev Biophys Biophys Chem 19*:301–332.

Shen J, Wendoloski J. 1996. Electrostatic binding energy calculation using the finite difference solution to the linearized Poisson–Boltzmann equation: Assessment of its accuracy. *J Comput Chem 17*:350–357.

Sitkoff D, Sharp KA, Honig B. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J Phys Chem 98*:1978–1988.

Strang G. 1993. *Introduction to linear algebra*. Wellesley, Massachusetts: Wellesley-Cambridge Press.

Tanford C, De PK, Taggart VG. 1960. The role of the α-helix in the structure of proteins. Optical rotatory dispersion of β-lactoglobulin. *J Am Chem Soc 82*:6028–6034.

Waldburger CD, Schildbach JF, Sauer RT. 1995. Are buried salt bridges important for protein stability and conformational specificity? *Nat Struct Biol 2*:122–128.

Wang L, O'Connell T, Tropsha A, Hermans J. 1996. Energetic decomposition of the α-helix–coil equilibrium of a dynamic model system. *Biopolymers 39*:479–489.

Wimley WC, Gawrisch K, Creamer TP, White SH. 1996. Direct measurement of salt-bridge solvation energies using a peptide model system: Implications for protein stability. *Proc Natl Acad Sci USA 93*:2985–2990.

Xu D, Lin SL, Nussinov R. 1997. Protein binding versus protein folding: The role of hydrophilic bridges in protein associations. *J Mol Biol 265*: 68–84.

Yang A-S, Honig B. 1995. Free energy determinants of secondary structure formation: I. α-helices. *J Mol Biol 252*:351–365.