



UPPSALA
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 255*

Computational Methods for Calculation of Ligand-Receptor Binding Affinities Involving Protein and Nucleic Acid Complexes

MARTIN ALMLÖF



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2007

ISSN 1651-6214
ISBN 91-554-6761-X
urn:nbn:se:uu:diva-7421

Dissertation presented at Uppsala University to be publicly examined in B7:101, BMC, Husargatan 3, Uppsala, Friday, January 19, 2007 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English.

Abstract

Almlöf, M. 2007. Computational Methods for Calculation of Ligand-Receptor Binding Affinities Involving Protein and Nucleic Acid Complexes. Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 255. 53 pp. Uppsala. ISBN 91-554-6761-X.

The ability to accurately predict binding free energies from computer simulations is an invaluable resource in understanding biochemical processes and drug action. Several methods based on microscopic molecular dynamics simulations exist, and in this thesis the validation, application, and development of the linear interaction energy (LIE) method is presented.

For a test case of several hydrophobic ligands binding to P450cam it is found that the LIE parameters do not change when simulations are performed with three different force fields. The nonpolar contribution to binding of these ligands is best reproduced with a constant offset and a previously determined scaling of the van der Waals interactions.

A new methodology for prediction of binding free energies of protein-protein complexes is investigated and found to give excellent agreement with experimental results. In order to reproduce the nonpolar contribution to binding, a different scaling of the van der Waals interactions is necessary (compared to small ligand binding) and found to be, in part, due to an electrostatic preorganization effect not present when binding small ligands.

A new treatment of the electrostatic contribution to binding is also proposed. In this new scheme, the chemical makeup of the ligand determines the scaling of the electrostatic ligand interaction energies. These scaling factors are calibrated using the electrostatic contribution to hydration free energies and proposed to be applicable to ligand binding.

The issue of codon-anticodon recognition on the ribosome is addressed using LIE. The calculated binding free energies are in excellent agreement with experimental results, and further predict that the Leu2 anticodon stem loop is about 10 times more stable than the Ser stem loop in complex with a ribosome loaded with the Phe UUU codon. The simulations also support the previously suggested roles of A1492, A1493, and G530 in the codon-anticodon recognition process.

Keywords: computer simulations, molecular dynamics, linear interaction energy, binding free energy, linear response, protein-protein interactions, structure-based design, point mutations, hot spots, solvation free energy

Martin Almlöf, Department of Cell and Molecular Biology, Box 596, Uppsala University, SE-75124 Uppsala, Sweden

© Martin Almlöf 2007

ISSN 1651-6214

ISBN 91-554-6761-X

urn:nbn:se:uu:diva-7421 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-7421>)

“Come on in, and we'll take this outside!”

Duke of Bourbon

Publications

List of publications included in this thesis

- I **Almlöf, M.**, Brandsdal, B. O., and Åqvist, J. (2004) Binding affinity prediction with different force fields: Examination of the linear interaction energy method, *J. Comput. Chem.* 25, 1242-1254.
- II **Almlöf, M.**, Åqvist, J., Smalås, A. O., and Brandsdal, B. O. (2006) Probing the effect of point mutations at protein-protein interfaces with free energy calculations, *Biophys. J.* 90, 433-442.
- III **Almlöf, M.**, Andér, M., and Åqvist, J. (2006) Energetics of codon-anticodon recognition on the small ribosomal subunit, *Biochemistry*, *in press*.
- IV **Almlöf, M.**, Carlsson, J., and Åqvist, J. (2006) Investigation of the linear response approximation for predicting hydration free energies. *Manuscript*.

Additional publications

- i Brandsdal, B. O., Österberg, F., **Almlöf, M.**, Feierberg, I., Luzhkov, V. B., and Åqvist, J. (2003) Free energy calculations and ligand binding, *Adv. Prot. Chem.* 66, 123-158.
- ii Luzhkov, V. B., **Almlöf, M.**, Nervall, M., and Åqvist, J. (2006) Computational study of the binding affinity and selectivity of the bacterial ammonium transporter AmtB, *Biochemistry* 45, 10807-10814.

Contents

1 Introduction.....	9
2 Computational Methods.....	11
2.1 Force Fields and Molecular Dynamics.....	11
2.1.1 Force Fields	11
2.1.2 Molecular Dynamics.....	11
2.1.3 Q	12
2.2 Free Energy Perturbation.....	13
2.2.1 Single-Step Perturbation.....	15
2.3 Approximate Free Energy Methods	15
2.3.1 Linear Interaction Energy	15
2.3.2 Molecular Mechanics Poisson – Boltzmann Surface Area.....	17
3 Linear Interaction Energy	18
3.1 β	19
3.2 α	24
3.2.1 Protein-Protein Interactions.....	27
3.2.2 LIE and codon-anticodon recognition	31
3.3 γ	31
3.4 Force Field Dependence.....	33
3.5 The Ribosome	35
4 Future Perspectives	41
5 Summary in Swedish	43
6 Acknowledgements.....	46
7 References.....	47

Abbreviations

ASL	Anticodon stem loops
BPTI	Bovine pancreatic trypsin inhibitor
DNA	Deoxyribonucleic acid
EVB	Empirical valence bond
FEP	Free energy perturbation
GBP	Periplasmic glucose/galactose receptor
GTP	Guanosine triphosphate
HIV	Human immunodeficiency virus
LIE	Linear interaction energy
LRA	Linear response approximation
MAP	Mitogen-activated protein
MC	Monte Carlo
MD	Molecular dynamics
MM-PBSA	Molecular-mechanics/Poisson-Boltzmann/surface area
mRNA	Messenger RNA
OMTKY3	3 rd domain of the turkey ovomucoid inhibitor
PB	Poisson-Boltzmann
QSAR	Quantitative structure activity relationships
RDF	Radial distribution function
rms	Root mean square
RNA	Ribonucleic acid
SCAAS	Surface constrained all-atom solvent
tRNA	Transfer RNA
WNDR	Weighted nonpolar desolvation ratio

1 Introduction

Modern drug design involves several phases which, together, can take up to 15 years and expend enormous amounts of resources. The early phases of drug discovery typically consist of determining a (macromolecular) target and identifying compounds that are active against this target. The vast quantity of macromolecular three-dimensional atomic scale structures made available in recent years has made structure-based design a viable alternative to (experimental) high throughput screening where large libraries of compounds are screened against a specific target.

Almost all computational drug design methods aim at calculating the binding strength of a drug candidate and can be divided into non-structure-based and structure-based. Non-structure-based methods such as quantitative structure activity relationships (QSAR) and pharmacophore modeling rely on prior activity knowledge of potent ligands to predict the activity of the drug candidate. Structure-based methods use structural information of the macromolecule to analyze ligand-macromolecule interactions, and from these interactions predict the drug candidate activity.

To date, several structure-based modeling methods have been developed. Some of these are: 3D-scoring functions, free energy perturbation (FEP), the linear interaction energy method (LIE), and the molecular mechanics / Poisson-Boltzmann / surface area method (MM-PBSA). The computational cost and human interaction necessary to perform FEP, LIE, and MM-PBSA calculations make them less applicable for screening of a very large number of compounds and more viable in a later phase of drug design.

The work presented in this thesis is primarily focused on investigation, application, and improvement of the LIE method. LIE uses energies extracted from atom scale simulations of ligand binding to predict the binding affinity in terms of the free energy of binding. In order to obtain meaningful energies to use in the LIE equation, many conformations of the system under investigation need to be sampled. These conformations are generated using molecular dynamics (MD) or Monte Carlo (MC) simulations. Paper I addresses the method's applicability using different force fields. The performance and parameterization of LIE is investigated using three force fields, and some preliminary results regarding LIE's dependence on binding site hydrophobicity are reported. Paper II and III are applications of LIE on systems with ligands which are much larger than drug-like ligands and thus pose some very difficult computational problems due to the sheer size of the en-

ergies involved. The results of Paper II and III shows that by using a new methodology within the LIE framework, important biochemical interactions such as protein-protein interactions can be studied using LIE. Paper IV investigates the electrostatic component of the LIE method in detail and suggests a new formalism for computing the electrostatic contribution to binding.

The thesis is divided into discussions of the three parameters, α , β , and γ which govern the LIE method, and finally a summary of the results and implications of the LIE method applied to the problem of codon-anticodon recognition on the ribosome.

2 Computational Methods

2.1 Force Fields and Molecular Dynamics

2.1.1 Force Fields

All the work presented in this thesis relies on force fields. A force field defines the interactions between all particles in the system of interest. The most common force fields for simulations of biochemical systems use similar potential functions of the form:

$$U_{pot} = \sum_{bonds} K_b (r_{ij} - r_{0ij})^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\varphi [1 + \cos(n\varphi + \delta)] + \\ + \sum_{impropers} K_\xi (\xi - \xi_0)^2 + \sum_{nonbonded} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \sum_{nonbonded} \frac{A_i A_j}{r_{ij}^{12}} + \frac{B_i B_j}{r_{ij}^6} \quad 1$$

Here, the total potential energy of the system is divided into different parts: bond stretching, angle bending, dihedral rotation, improper bending, electrostatic interactions, and van der Waals interactions. The variables r_{ij} , θ , φ , and ξ describe the particles' relative coordinates in the system and change as the system evolves (described below). All the other symbols, K_b , r_{0ij} , K_θ , θ_0 ..., are parameters inherent to the force field and describe a specific type of interaction such as bond stretching, angle bending, etc. These parameters are calibrated to reproduce results such as quantum mechanical calculations, spectroscopic experiments, free energies of solvation, etc. The results and methods to calibrate each force field are often not the same, and thus several force fields exist with different parameter sets. The force fields used in this thesis are AMBER, OPLS-AA, GROMOS, and CHARMM.¹⁻⁴

2.1.2 Molecular Dynamics

In order to obtain a meaningful description of the system under investigation many different configurations of the atoms in the system need to be generated. The reason for this is that all systems are constantly changing due to thermal motion. If we were to take many snapshots of a *real* system, all the snapshots will most likely differ. According to statistical mechanics theory, the different configurations of the system follow a distribution depending on their total potential energy known as the Boltzmann distribution. So in order

to extract *macroscopic* information from a *microscopic* system, we need a method to generate configurations that obey the Boltzmann distribution.

Two commonly used methods to generate configurations, which obey the Boltzmann distribution, are Monte Carlo (MC) and molecular dynamics (MD). The MC method generates configurations by randomly moving atoms and then evaluating the resulting energy change. If the energy change is negative, the new configuration is accepted. If the energy change is positive, the new configuration is only accepted with a probability abiding by the Boltzmann distribution.

The molecular dynamics sampling method generates configurations by moving the atoms according to Newton's laws of motion. Since each atom's potential energy can be calculated from the force field potential function, forces are easily attainable by taking the negative gradient of the potential energy. Through the forces, accelerations and new velocities are calculated and, given a certain time step, new atom positions are obtained.

2.1.3 Q

In this thesis, all the MD simulations were performed using the Q molecular dynamics package.⁵ The MD engine of Q is tailored to perform linear interaction energy (LIE) calculations,⁶ free energy perturbation (FEP) simulations,^{7,8} and empirical valence bond calculations (EVB).^{9,10} Q utilizes the leap-frog version of Verlet's integration algorithm:

$$\begin{aligned}\mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \mathbf{v}\left(t + \frac{\Delta t}{2}\right)\Delta t \\ \mathbf{v}_i\left(t + \frac{\Delta t}{2}\right) &= \mathbf{v}_i\left(t - \frac{\Delta t}{2}\right) + \mathbf{a}(t)\Delta t\end{aligned}\tag{2}$$

where t and Δt are the time and timestep respectively and \mathbf{r} , \mathbf{v} , and \mathbf{a} are the position, velocity, and acceleration respectively. The temperature is controlled by scaling the velocities according to the Berendsen temperature scaling algorithm.¹¹

Spherical and periodic boundaries are available in Q. The spherical boundary implementation in Q uses a combination of a Morse-type boundary attraction potential inside the sphere and a half-harmonic potential outside the sphere. Parameters for the radial restraining function have been calibrated to reproduce the experimental density of bulk water for sphere radii ranging from 12 to 30 Å. In order to avoid boundary polarization effects of water molecules near the surface of the sphere, polarization restraints similar to that of the SCAAS¹² model are also in effect in Q.

2.2 Free Energy Perturbation

Free energy perturbation (FEP) is a commonly used technique based on Zwanzig's formula¹³ (eq 3) to calculate free energy differences.

$$\Delta G = G_B - G_A = -kT \ln \left\langle \exp\left(\frac{-(U_B - U_A)}{kT}\right) \right\rangle_A \quad 3$$

By using Zwanzig's formula, it is possible to calculate the free energy difference between states A and B by evaluating the potential energy difference of state A and B sampled on one of the states' potentials using MD or MC. In eq 3 the free energy difference on going from state A to B , $G_B - G_A$, is calculated; however since the choice of states is arbitrary one could just as well switch the A and B labels to get the free energy difference on going from state B to A . These "forward" and "backward" calculations are often employed and used as a measure of numerical/sampling error. One of the criterions for eq 3 to be useful is that the configurations sampled on one of the potentials are also probable on the other potential. That is, the potential energy difference between state A and state B may not be too large, otherwise large numerical errors are likely.

Since the potential energy difference of different biochemically interesting states, *e.g.* two ligands bound to a macromolecule, is often too large for eq 3 to be directly useful, intermediate states are introduced. The potential energy functions representing the intermediate states are often linear combinations of the potential energy functions representing the original states, *i.e.* linear combinations of U_A and U_B :

$$U_m = (1 - \lambda_m)U_A + \lambda_m U_B \quad 4$$

In eq 4, λ_m varies from 0 to 1 in n steps ($m = 1 \dots n$) thus producing n intermediate potential energy functions, U_m , which smoothly change the potential from state A to state B . Since the free energy difference is uniquely defined by the initial and final states (*i.e.* a state function), the free energy difference between states A and B can be computed as the sum of the free energy differences of the intermediate states U_m . This is known as the free energy perturbation (FEP) method, eq 5.

$$\Delta G = -kT \sum_{m=1}^{n-1} \ln \left\langle \exp\left(\frac{-(U_{m+1} - U_m)}{kT}\right) \right\rangle_m \quad 5$$

In most binding calculations FEP cannot be used in a direct manner to obtain reliable *absolute* binding free energies due to large energy differences, *i.e.* the two states A : ligand free in solution and B : ligand bound to macromolecule are simply energetically too far apart for FEP to be applicable. A workaround is to use a thermodynamic cycle (*Figure 1*) where the binding free energy difference between ligand L and L' is calculated by mutating one ligand into the other in both environments: in solvent and bound to protein (P).¹⁴ Using the FEP methodology, two calculations are made. In the first calculation state A represents ligand L in water and state B represents ligand

L' in water. This is a relatively simple calculation, which yields ΔG_{mut}^w in *Figure 1*, given that the differences between ligand L and L' are not too large.

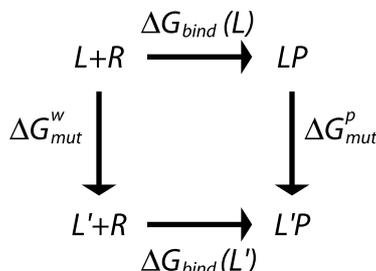


Figure 1: Thermodynamic cycle used to predict relative binding free energies between ligand L and L' using FEP.

The second calculation is that of ΔG_{mut}^p where ligand L is mutated to L' while bound to the protein. The difference in free energy of binding, $\Delta\Delta G$, between ligand L and L' can then be calculated using eq 6.

$$\Delta\Delta G = \Delta G_{bind}(L') - \Delta G_{bind}(L) = \Delta G_{mut}^p - \Delta G_{mut}^w \quad 6$$

Thus we have a rigorous method of calculating relative ligand binding free energies for ligands which do not differ too much in binding modes or chemical composition.

Problems often occur with FEP at the endpoints of a mutation if the mutation involves annihilation or creation of bulky groups. The first problem is a numerical problem which arises when the sampling is performed on a potential in which the atoms to be annihilated have nearly disappeared. In this state the disappearing atoms are almost singularities (or “spikes”) and when other atoms “bump” into them very large potential energies are obtained and the resulting forces become too large for the system to remain stable. Another problem observed at the endpoints is mostly a MD sampling problem; the positions occupied by the disappearing atoms can not be sampled until the atoms have completely disappeared since their Lennard-Jones spikes effectively occlude any other atoms from this space. This becomes particularly problematic when mutating a ligand in a binding site since the space occupied by the vanishing atoms can only be replaced by water (or protein side chains) at the very last λ -step of the simulation. A method to avoid these endpoint problems is the use of an intermediate so called “soft core” potential.^{15,16} This method replaces the standard Lennard-Jones potential term of the intermediate λ -steps with a modified Lennard-Jones term that does not have a singularity at $r_{ij} = 0$. The soft core potential as implemented in the Q molecular dynamics package is shown in eq 7.

$$U_{ij}^{vdW} = \frac{A_{ij}}{(r_{ij}^6 + \alpha_{ij})^2} - \frac{B_{ij}}{r_{ij}^6 + \alpha_{ij}} \quad 7$$

Today, soft core potentials are almost always used when calculating absolute solvation free energies with FEP¹⁷⁻²¹ and calculations of ligand binding free energies using the single-step perturbation technique²²⁻²⁶ relies heavily on the use of soft core potentials.

2.2.1 Single-Step Perturbation

Single-step perturbation is also based on Zwanzig's formula (eq 3) to calculate the free energy difference between two states.²²⁻²⁶ In contrast to the multi-step approach used in traditional FEP, single-step perturbation only requires sampling of one state, the reference state. The "trick" in single-step perturbation is to choose a reference state such that the configurations sampled on the reference state also are relevant for the end state. If the configurations sampled on the reference state are relevant for several end states (*i.e.* several molecules) then the relative free energy difference between all the molecules can be calculated. In practice, the reference state used in single-step perturbation is represented by an unphysical lattice of soft core atoms, thus avoiding singularities and some of the sampling problems mentioned above. For calculations of relative solvation free energies, one simulation of the reference state in water is necessary, while calculations of binding free energies require simulations in water and in the binding site.

2.3 Approximate Free Energy Methods

In theory, the free energy of any system can be calculated using simple sampling techniques such as MD and the relationship $G = H - TS$, where G is the free energy, H the enthalpy, S the entropy, and T the temperature. Thus, to calculate the binding free energy of a ligand to a macromolecule one simply needs to calculate the absolute free energies of the three relevant states: the ligand free in water, the macromolecule free in water, and the ligand bound to the macromolecule. It is, however, quite difficult to obtain reliable estimates of the enthalpy and entropy, since the enthalpy is inherently associated with large error bars and the majority of contributions to the entropy lie in regions of phase space which are seldom explored. Thus alternate methods are necessary to obtain reliable estimates of free energies of binding.

2.3.1 Linear Interaction Energy

In 1994 Åqvist *et al.* introduced a new method, the Linear Interaction Energy (LIE) method (eq 8), to calculate binding free energies from micro-

scopic simulations.⁶ This method relies the electrostatic linear response approximation (LRA) and an empirical relationship between solvation of hydrophobic compounds and their van der Waals energies. LIE basically depicts the ligand as being solvated in two different mediums: water and the macromolecule.

$$\Delta G = \alpha \Delta \langle U_{l-s}^{vdW} \rangle + \beta \Delta \langle U_{l-s}^{el} \rangle + \gamma \quad 8$$

In eq 8, $\langle U_{l-s}^{vdW} \rangle$ and $\langle U_{l-s}^{el} \rangle$ refer to the average ligand surrounding van der Waals or electrostatic interaction energies. The Δ 's denote the difference in interaction energies between the simulations of the ligand bound to the macromolecule and free in solution. α is an empirically derived parameter which relates the change in ligand surrounding van der Waals energies to the effect of hydrophobic solvation and other size correlated contributions to binding. The electrostatic ligand-surrounding interaction energies are scaled by a factor, β , to reproduce the electrostatic contribution to solvation. γ is a medium dependent constant which is needed in some cases to reproduce *absolute* binding free energies.

In the original application of LIE β was set to $\frac{1}{2}$, in accordance with linear response theory, and α was parameterized on a set of five endothiapsin inhibitors to a value of 0.16 in order to minimize the root mean square (rms) deviation to experimental binding free energies.⁶ Including γ in the parameterization did not improve the fit significantly and it was thus omitted from the model. These first results of the LIE method were very encouraging with a mean unsigned error of 0.39 kcal/mol. This model, with $\alpha = 0.16$ and $\beta = \frac{1}{2}$ – hereafter referred to as the *original model*, has been successfully used to predict the free energies of sugar binding to GBP (periplasmic glucose/galactose receptor) and several inhibitors to HIV-1 protease.²⁷⁻³⁰

In a investigation by Åqvist and Hansson³¹ on the validity of linear response it was found that in the charging of neutral dipolar solute molecules in water the solvent does not respond linearly to the electrostatic field exerted by the solute. This departure from linear response was attributed to the solvent hydrogen bond network, and was more pronounced for solute compounds which could hydrogen bond to the solvent. Thus, in 1998 Hansson *et al.*³² tested several different versions of LIE, some of which utilized the new solute dependent β 's, on previously published data.²⁷⁻³⁰ The model which displayed the best statistical figures of merit had solute dependent β 's, $\alpha = 0.18$, and no γ coefficient. The new β coefficient followed a simple scheme where charged compounds had $\beta = \frac{1}{2}$, neutral compounds with one hydroxyl had $\beta = 0.37$, neutral compounds with two or more hydroxyls had $\beta = 0.33$, and all other neutral compounds had $\beta = 0.43$. This model will hereafter be referred to as the *standard model*.

Even though there is an indication of the proper values of α and β to use when calculating binding free energies with LIE, these parameters can just as well be parameterized freely in a QSAR like manner.³³⁻³⁵

2.3.2 Molecular Mechanics Poisson – Boltzmann Surface Area

Molecular Mechanics / Poisson-Boltzmann / Surface Area (MM-PBSA)^{36,37} is a method which calculates the free energy of a state using continuum electrostatic solvation models, surface area, entropy estimates, and molecular mechanics potential energies:

$$\langle G \rangle = \langle G_{PB} \rangle + \langle G_{SA} \rangle + \langle E_{MM} \rangle - T \langle S_{MM} \rangle \quad 9$$

where G_{PB} is the electrostatic component of the solvation free energy calculated using a numerical solution to the Poisson-Boltzmann (PB) equation³⁸ and G_{SA} is the nonpolar contribution of the solvation free energy calculated using a surface area term.³⁹ E_{MM} is the molecular mechanics potential energy (force field potential energy of bonds, angles, torsions, and van der Waals and electrostatic interactions) of the solute, and S_{MM} is the calculated entropy of the solute, often calculated using normal mode or quasi-harmonic analysis. These terms are calculated on snapshots of explicit water MD simulations, and the $\langle \rangle$ denote the averages over the entire trajectory.

MM-PBSA was first used by Kollman and coworkers to explain the free energy differences of the A and B forms of various DNA and RNA fragments.³⁶ In recent years it has also been used to predict ligand binding free energies.^{37,40-46} In order to predict ligand binding free energies using MM-PBSA, the free energy of the ligand, macromolecule, and ligand-macromolecule complex need to be calculated using eq 9. The free energy of binding then becomes:

$$\Delta G_{bind} = \langle G_{complex} \rangle - \langle G_{ligand} \rangle - \langle G_{macromolecule} \rangle \quad 10$$

The obvious approach to calculating ΔG_{bind} using eq 10 is to run MD simulations of the ligand free in solution, the unbound macromolecule in solution, and the ligand-macromolecule complex in solution and then to calculate the energetic components of eq 9 using the three obtained trajectories. However, a major pitfall of this type of procedure is that convergence of the $\langle E_{MM} \rangle$ and $\langle G_{PB} \rangle$ terms for the macromolecule and complex is quite difficult to obtain, and thus large errors are expected.⁴⁷ A proposed solution is to evaluate $\langle G_{ligand} \rangle$ and $\langle G_{macromolecule} \rangle$ from the ligand and macromolecule geometries generated in the ligand-macromolecule complex simulation. In this “single trajectory” scheme, the intramolecular force field energies will cancel and the resulting *numerical* errors will be greatly reduced. It is, of course, quite a strong assumption that the ligand and macromolecule have the same conformations in bound and free states, and any contributions from conformational entropy will be lost in such a scheme. A point in case is an investigation of MM-PBSA on 16 ligands binding to the active site of p38 MAP kinase.⁴⁷ In this investigation the single trajectory scheme yielded much lower numerical errors when compared to a multiple trajectory scheme, however, the agreement with experiment was better for the multiple trajectory scheme.

3 Linear Interaction Energy

The linear interaction energy method is based on the thermodynamic cycle in *Figure 2*. Here the free energy of binding, ΔG_{bind} , is estimated by calculating the three other legs of the cycle, namely ΔG_{free}^{polar} , ΔG_{bound}^{polar} , and $\Delta\Delta G^{nonpolar}$. ΔG_{free}^{polar} and ΔG_{bound}^{polar} are the free energies of charging the ligand in its respective environment while $\Delta\Delta G^{nonpolar}$ is the free energy difference between the nonpolar solvation of the ligand in water and bound to its receptor.

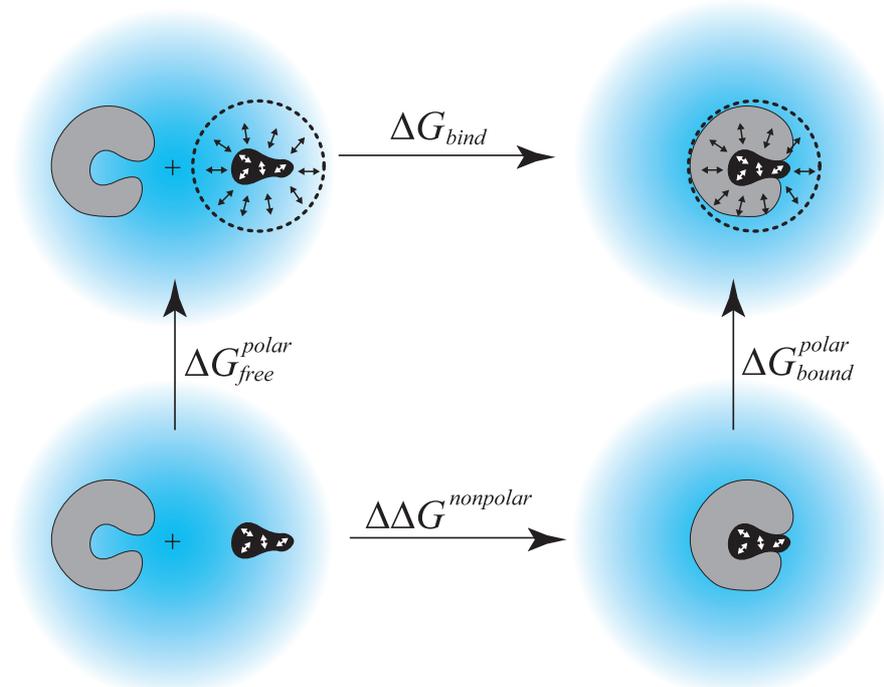


Figure 2. Schematic representation of the underlying free energy components estimated using the LIE method. The white and black arrows represent intra- and inter-molecular electrostatic interactions, respectively.

The polar contributions to the binding free energy are predicted by scaling the ligand-surrounding electrostatic interactions by a parameter, β , while the nonpolar contribution is predicted by scaling the ligand-surrounding van der Waals interactions with the parameter α and a constant γ . The following sections are divided into discussion on each of the three parameters, β , α , and γ .

3.1 β

The free energy of charging the ligand (turning on the partial charges) in its environment (the vertical legs in *Figure 2*) is calculated using electrostatic linear response theory. A straightforward derivation of the linear response approximation starts at Zwanzig's formula (eq 3). The process of charging the ligand can be considered as a transformation between the two potentials U_{off} and U_{on} , where $U_{off} = U_{s-s} + U^{LJ} + U^{bonded}$ consists of all solvent-solvent interactions, all ligand Lennard-Jones interactions, and all the bonded interactions of the ligand (bonds, torsions, etc.) and $U_{on} = U_{s-s} + U^{LJ} + U^{bonded} + U_{l-l}^{el} + U_{l-s}^{el}$ consists of the same interactions as U_{off} but with the electrostatic ligand-ligand and ligand-solvent interactions turned *on*. Two consecutive Taylor expansions of eq 3, with $U_A = U_{off}$ and $U_B = U_{on}$, yields:

$$\Delta G_{off \rightarrow on} = \langle \Delta U \rangle_{off} - \frac{1}{2kT} \left\langle \left(\Delta U - \langle \Delta U \rangle_{off} \right)^2 \right\rangle_{off} + \dots \quad 11$$

Averaging on the *on* potential yields:

$$\Delta G_{off \rightarrow on} = \langle \Delta U \rangle_{on} + \frac{1}{2kT} \left\langle \left(\Delta U - \langle \Delta U \rangle_{on} \right)^2 \right\rangle_{on} + \dots \quad 12$$

Truncating terms of third order and higher, and assuming that the mean squared fluctuations of the energy gap on the two potentials (*off* and *on*) are equal yields the linear response approximation:

$$\Delta G_{off \rightarrow on} = \frac{1}{2} \left(\left\langle U_{l-l}^{el} + U_{l-s}^{el} \right\rangle_{off} + \left\langle U_{l-l}^{el} + U_{l-s}^{el} \right\rangle_{on} \right) \quad 13$$

In most LIE investigations in the literature, the electrostatic ligand-ligand interactions are not included in the calculation of binding free energies. This can at first seem incorrect since the ligand-ligand terms enter into equation 13. However, the vertical legs of the thermodynamic cycle in *Figure 2* can be split up into two processes: turning on the ligand-ligand electrostatics (ΔG_{l-l}^{el}), and turning on the ligand-surrounding electrostatics (ΔG_{l-s}^{el}) (*Figure 3*). Thus the problem boils down to which nonpolar binding process is best reproduced (described) by $\alpha \Delta \langle U_{l-s}^{vdW} \rangle - \Delta \Delta G'_{nonpolar}$ or $\Delta \Delta G''_{nonpolar}$. Since the $\alpha = 0.18$ of the standard model was parameterized using only ligand-surrounding electrostatics, it is possible that a model which includes the ligand-ligand electrostatics may yield an optimal $\alpha \neq 0.18$.

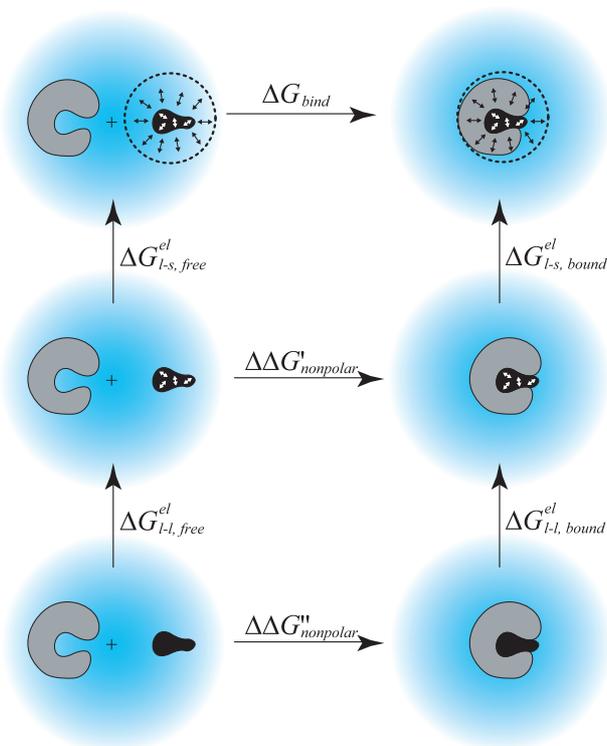


Figure 3. Alternate LIE thermodynamic cycle with the charging process divided into turning on the ligand-ligand electrostatic interactions and turning on the ligand-surrounding electrostatic interactions. The white and black arrows represent intra- and intermolecular electrostatic interactions, respectively.

As shown in Ref. 31 and Paper IV, the $\langle U_{l-s}^{el} \rangle_{off}$ term is near zero for neutral compounds. This is not so surprising as the orientation of solvent dipoles will be (near) random around a solute if the solute-solvent electrostatic interactions are turned off ($\langle \rangle_{off}$). When the solvent is water, the solvent dipole is actually not completely random, and gives a net negative potential at the solute cavity.^{31,48,49} This effect is due to the hydrogen bonding nature of water and the curvature of the solute. Thus, for charged compounds, the $\langle U_{l-s}^{el} \rangle_{off}$ term will not be zero but will be negative for positive ions and positive for negative ions. For neutral solutes the effect of the non random orientation of water cancels at opposite ends of the solute dipole. Even though $\langle U_{l-s}^{el} \rangle_{off}$ is not zero for charged compounds, it is often considered negligible when compared to $\langle U_{l-s}^{el} \rangle_{on}$.

A major finding in Åqvist *et al.* was that when charging solutes in water, the “ $1/2$ ” in eq 13 only applies to solutes with a nonzero formal charge.³¹ For dipolar solutes it was found that the ratio between $\Delta G_{off \rightarrow on}$ and ΔU varied between 0.33 and 0.43, depending on the number of hydroxyls in the solute.

The reason for this nonlinear response was proposed to be the water hydrogen bond network, as aprotic solvents displayed ratios much closer to $\frac{1}{2}$. For solutes which could interact more extensively with the hydrogen bond network than other groups (solutes containing hydroxyls) this nonlinear effect was more pronounced.

Thus a new treatment of the electrostatic contribution to binding in the LIE method was proposed for neutral molecules (net charge = 0),³² where the difference in electrostatic ligand-surrounding energies was not scaled by $\frac{1}{2}$, but by a factor, β , which varied depending on the number of hydroxyls in the solute (eq 14).

$$\Delta G^{polar} = \beta \langle U_{l-s}^{el} \rangle_{on} \quad 14$$

The work of Åqvist and Hansson³¹ does not, however, adequately address several factors which might influence the value of β for predicting the electrostatic contribution to solvation for various solutes, namely the effect of changes in intramolecular electrostatic energies, contributions from several other types of chemical moieties (not just hydroxyls), and the appropriate value of β for solutes containing a combination of chemical moieties. Hence in Paper IV linear response is investigated on a much larger set of compounds than what was done in Ref. 31 and the above issues are addressed.

The approach used in Paper IV is similar to that of Åqvist *et al.*³¹ except that the entire gas phase cycle is taken into account. That is, in Paper IV the changes in intramolecular electrostatic energies between gas and aqueous phase were also included (as derived in eq 13) to yield:

$$\Delta G_{el}^{solv} = \beta \left(\langle U_{l-s}^{el} \rangle_{on} + \Delta \langle U_{l-l}^{el} \rangle_{on} \right) \quad 15$$

β was first calculated on a set of 205 compounds containing one of several types of chemical moieties. From these calculations, 1°-amides, 1°-,2°-amines, alcohols, neutral carboxylic acids, protonated amines, and deprotonated carboxylic acids were identified as the groups which required a $\beta \neq 0.43$ in order to reproduce the ΔG^{polar} as calculated from the FEPs (Table 1). Interestingly, all the chemical moieties with zero net charge which required a $\beta \neq 0.43$ have hydrogen bond donor capabilities, and suggests that their deviation from $\beta = 0.43$ stems from the same underlying principles as was suggested in Ref. 31. That is, the capability of the solute to participate in the hydrogen bond network of the solvent amplifies the nonlinearity. The observed deviation of β from $\frac{1}{2}$ for the ionic moieties stems largely from the exclusion of $\langle U_{l-s}^{el} \rangle_{off}$ when calculating β (eq 15). Since this term is nonzero for ionic moieties, it will impact the calculated β , and inclusion of $\langle U_{l-s}^{el} \rangle_{off}$ does, in fact, yield β 's closer to $\frac{1}{2}$ (data not shown).

Table 1. The optimal parameters of β to use to predict the electrostatic contribution to the free energy of solvation.

Compound	β
Alcohols	0.37
1° Amides	0.41
1°, 2° Amines	0.39
Carboxylic Acids	0.40
Cations	0.52
Anions	0.45
All other compounds	0.43

Two different methods of predetermining β for compounds containing several chemical moieties were proposed (method D and E, in accordance with the nomenclature used in Paper IV). These methods are based on the assumption that each chemical group contributes to the electrostatic solvation free energy in an additive manner, and thus the electrostatic contribution to the free energy of solvation can be calculated as:

$$\Delta G_{el}^{solv} = \sum_i \beta_i \left(\langle U_{l-s}^{el} \rangle + \Delta \langle U_{l-l}^{el} \rangle \right)_i \quad 16$$

where the subscript i denotes the different chemical groups in the compound and β_i are the β 's obtained from the calculations on compounds containing single chemical moieties. From eq 16 a β for any compound can be estimated from:

$$\beta = \frac{\sum w_i \beta_i}{\sum w_i} \quad 17$$

where the energies of eq 16 have been replaced by weighting factors. The two methods of predicting β 's for compounds containing several chemical moieties differ in their treatment of the weighting factors. In method D the weighting factors are based on the energies from the calculations of the compounds containing single chemical moieties. Method E uses the same weight ($w_i = 1$) for all chemical moieties with net charge = 0 and $w_i = 11$ for charged moieties.

Method D and E were tested on a set of 361 compounds containing combinations of different chemical moieties and compared to the scheme devised by Hansson *et al.*³² and to using $\beta = 1/2$ for all compounds. Performance of

Table 2. Performance of the models on predicting the electrostatic contribution to solvation free energies for the 356 non zwitterionic molecules in the test set.

Method	Treatment of β and w_i	rms ^a	$\langle \text{error} \rangle^a$
A	$\beta = 0.5$	3.72	2.88
B	β according to Ref. 32	3.29	1.90
C	one β parameterized for entire training set	3.68	2.91
D	$\beta = \beta_0 + \frac{\sum_i w_i \Delta\beta_i}{\sum_i w_i}$ $w_i \propto \langle U_i^{el} \rangle_B$	1.22	0.76
E	$\beta = \beta_0 + \frac{\sum_i w_i \Delta\beta_i}{\sum_i w_i}$ $w_i \begin{cases} 1 & \text{for net charge} = 0 \\ 11 & \text{otherwise} \end{cases}$	1.26	0.78

^a kcal/mol

the different methods was evaluated in terms of how well the methods predicted (using eq 15 and the methods' predicted β) the electrostatic contribution to the free energy of solvation as calculated using FEP. Methods D and E outperformed the other methods as deemed by the average unsigned error and the root mean square (rms) of the residuals (Table 2). The difference in performance of methods D and E (due to the different handling of the weighting factors, w_i) was very small, and although method D performed slightly better than method E the proposed method to use when predicting new β 's is method E due to the decreased complexity as compared to method D.

The tabulated values for the rms and mean unsigned errors in Table 2 did not include the six zwitterions in the test set. These were analyzed separately since they were expected to not follow the general additivity rule of eq 16, and even though the FEP calculations for the zwitterions consisted of > 82 ns of MD per molecule (> 41 ns *on* \rightarrow *off*, > 41 ns *off* \rightarrow *on*), the uncertainties in ΔG_{el}^{solv} were quite large due to differences in calculated $\Delta G_{off\rightarrow on}$ and $\Delta G_{on\rightarrow off}$. The predicted β for zwitterions using method E was 0.48. This grossly over-predicted the magnitude of the electrostatic contribution to solvation for zwitterionic compounds and gave an rms and a mean unsigned error of 13.1 and 12.7 kcal/mol respectively (for the zwitterions). The optimal β for the zwitterions was much less than the predicted 0.48 and optimization of β for the zwitterions yielded $\beta = 0.39$ with an rms and mean unsigned error of 3.75 and 3.24 kcal/mol respectively. The optimal β for zwitterions

terions suggested the solvent response is similar to that of the dipolar compounds, *i.e.* the electrostatic field of the opposite charges cancels enough so that the solvent behaves as if the zwitterions were “normal” dipolar compounds. This is, however, most likely not the case. Although more calculations on zwitterions are necessary to get a definitive answer, the data in Paper IV suggests that the reason for the abnormal β 's for zwitterions is actually of a completely different nature. The reason for the strange β 's of the zwitterions seems to be due to different β 's in water and gas phase. This may be due to the abnormally large changes in intramolecular electrostatics when charging the compound in the gas phase. However, in order to determine the underlying reasons for the strange behavior of the zwitterions, many more calculations on flexible and rigid zwitterions are necessary.

The application of this new treatment of the electrostatic ligand energies to predict the polar contribution to binding free energies has not yet been tested. Since no extensive investigations have been performed on the validity of linear response in the protein environment, it is difficult to say if the derived β 's of Paper IV will hold for binding free energies. In several studies of binding free energies using LIE, β has been parameterized freely, often yielding very low or even negative values.^{35,50-52} In these cases where the standard β values (Ref. 32 or Paper IV) do not yield meaningful binding free energies it seems more reasonable to use separate β 's for the water and protein environments and only let β for the protein environment assume arbitrary values, as it has been demonstrated quite clearly in Ref. 31 and Paper IV that β in water is not arbitrary. A point in case is the result of And er *et al.* (in manuscript) that shows an optimal $\beta = 0.26$ when scaling the change in electrostatic energies between bound and free state – suggesting LRA collapse. However, when separate scaling of the electrostatic energies of the bound and free state is introduced, the optimal model has FEP-determined scaling of the free state (0.43 in this case) and $\beta = 0.48$ in the bound state – suggesting LRA holds in the protein environment. Anyhow, given the analogy of solvation free energies to binding free energies, the results of Paper IV seem to warrant the inclusion of the change in electrostatic ligand-ligand energies also in calculating binding free energies.

3.2 α

The estimation of the nonpolar contribution to the binding free energy, $\Delta\Delta G^{nonpolar}$, is based on two observations. The first is that the free energy of solvation of hydrocarbons depends approximately linearly on the length of the carbon chain (eq 18).⁵³ The second observation is that the ligand surrounding van der Waals energies of hydrocarbons in solvent also depend approximately linearly on hydrocarbon length (eq 20).⁶ These two observations together with the assumption that in terms of nonpolar solvation the

protein environment behaves as a solvent - only with different proportionality constants (eqs 19,21), leads to the conclusion that the free energy of binding of nonpolar compounds depends approximately linearly on the difference in ligand surrounding van der Waals energies (eq 22).

$$\Delta G_w^{nonpolar} \approx a_w \sigma + b_w \quad 18$$

$$\Delta G_p^{nonpolar} \approx a_p \sigma + b_p \quad 19$$

$$\langle U_{l-s}^{vdW} \rangle_w \approx c_w \sigma + d_w \quad 20$$

$$\langle U_{l-s}^{vdW} \rangle_p \approx c_p \sigma + d_p \quad 21$$

$$\Delta \Delta G^{nonpolar} \approx \frac{a}{c} \left(\Delta \langle U_{l-s}^{vdW} \rangle - (d) \right) + b = \alpha \Delta \langle U_{l-s}^{vdW} \rangle + \gamma \quad 22$$

The four constants, $a = a_p - a_w$, $b = b_p - b_w$, $c = c_p - c_w$, $d = d_p - d_w$, in eq 22 could perhaps be determined theoretically, seeing as the a_w and b_w parameters' association to water's surface tension and free energy of particle insertion is rather straightforward. However, given the more difficult nature of theoretically determining a_p and b_p it seems more tractable to empirically parameterize the combined constants α and γ using experimental binding constants. Thus, once α and γ are parameterized we have a method to predict the bottom leg in *Figure 2*.

When LIE was first introduced, α was parameterized on a set of five endothiapsin inhibitors to a value of 0.16 using $\beta = 1/2$ (original model). Subsequent refinement using the β parameters of Hansson *et al.*³² yielded an optimum α of 0.18 (standard model). The inclusion of a free γ -term did not significantly improve the rms errors of the calculations. What was neglected in this investigation was the possibility that the different binding sites have different γ 's. Parameterization with binding site dependent γ 's would perhaps have led to a different optimal value for α . Nevertheless, $\alpha = 0.18$ has been used in several applications of LIE since then with impressive results.^{32,54-59}

Since the nonpolar solvation characteristics of the binding site enter into the theoretical value of α (eq 22), it would seem reasonable that α will take on different values depending on which binding site it is parameterized on. In several ligand binding studies using LIE on various binding sites this is also the picture that emerges.^{33,34,50,60-63} Wang *et al.*⁶¹ investigated the relationship between α and the weighted non-polar desolvation ratio (WNDR) and found a strong correlation ($r^2 = 0.89$). These results are, however, somewhat inconclusive as the effects of including a constant offset (γ) was not addressed. What these studies do ascertain is that the LIE equation with $\alpha = 0.18$ and $\gamma = 0$ does not work well for the systems studied (and the force fields used, discussed below). Unfortunately, in the report by Wang *et al.*,⁶¹ where the optimal α for several systems is investigated, it is not possible to study the effect of also including a γ -parameter in the LIE equation since the report only deals with one ligand per complex. In order to distinguish be-

tween the effects of letting α take on arbitrary values versus including a non-zero γ -term it is necessary to consider a system for which relatively many experimental binding measurements have been carried out. Another requirement is that uncertainties in the electrostatic contribution and binding orientation are minimized. In this respect, the system examined in Paper I (P450cam) is ideal, as the ligands are relatively small, rigid, and nonpolar, and x-ray crystal structures are available for many of the ligand-protein complexes.

In Paper I, the binding of several ligands (*Figure 4*) to P450cam was investigated. Several different LIE models and three different force fields were used (force fields discussed in 3.4 Force Field Dependence). This particular system was chosen because of the reasons mentioned above and that Paulsen *et al.*⁶⁰ investigated ligand binding to P450cam and discovered that $\alpha = 0.16$ did not reproduce the experimental binding free energies. Their optimized value of $\alpha = 1.043$ was attributed to the different force fields used in their investigation and the original LIE paper (CVFF and GROMOS respectively). A more likely reason, however, is the difference in polar characteristics of the binding sites of P450cam and endothiapsin; P450cam being much more nonpolar.

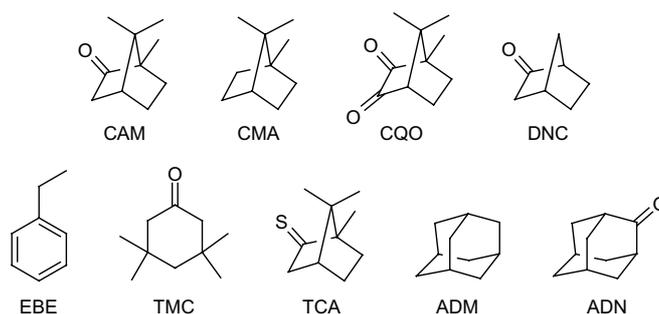


Figure 4. The nine ligands investigated in Paper I

In order to systematically analyze the effects of α , β , and γ on the predictability of the LIE method, five different parameterization models were tested. The performance of the models was judged on their cross-validated R^2 values (Q_{loo}^2) using the ‘leave-one-out’ methodology. Model A uses the standard $\alpha = 0.18$, solute dependent β s, and γ freely optimized. In Model B γ is set to zero, while α is optimized. Model C has both α and γ freely parameterizable. In Model D all three parameters are free, while Model E has $\beta = \frac{1}{2}$ (according to electrostatic linear response theory) and α and γ freely optimizable. The results of the different parameterization schemes for the OPLS-AA force field are shown in Table 3. Even though Model B is the same type of optimization that was performed by Paulsen *et al.*, we were not able to reproduce the high α obtained in their paper (1.043). The parameterized α in Paper I for Model B is however much higher than the standard $\alpha =$

0.18, but it is not the best model. As judged by the Q_{loo}^2 statistical figure of merit, Model A with α fixed at 0.18 and a free γ is the best model.

It is quite remarkable that the optimal model for the P450cam system has the same α as that which was earlier parameterized on binding sites with completely different polar/nonpolar characteristics. This indicates that α is perhaps not system dependent and that the different polar/nonpolar characteristics of a binding site enter into the constant γ term of the LIE method. This implies that the relative binding free energies of various ligands can be calculated without any free parameters.

Table 3. LIE models and statistical figures of merit using the OPLS-AA force field

Model	α	β	γ^a	RMS ^a	$\langle \text{Err} \rangle^a$	R^2	Q_{loo}^2
A	0.18 ^b	0.43 ^b	-4.45	0.27	0.23	0.90	0.88
B	0.60	0.43 ^b	0.00 ^b	0.87	0.65	-0.01	-0.17
C	0.19	0.43 ^b	-4.31	0.27	0.24	0.90	0.81
D	0.16	0.55	-4.62	0.24	0.19	0.92	0.82
E	0.17	0.50 ^b	-4.48	0.25	0.20	0.92	0.85

^a kcal/mol ^b fixed value

3.2.1 Protein-Protein Interactions

In order to test the applicability of LIE on protein-protein interactions, the free energy difference of several point mutations at the P1 residue of three protein inhibitor-proteinase complexes was analyzed in Paper II. Prediction of protein-protein binding free energies has proven extremely difficult, and development of an accurate method to calculate protein-protein binding free energies will be of great use in studying important biochemical interactions. In total, 56 binding free energies of P1 mutations of bovine pancreatic trypsin inhibitor (BPTI) and the third domain of the turkey ovomucoid inhibitor (OMTKY3) binding to α -chymotrypsin, human leukocyte elastase, or bovine β -trypsin were calculated. It was found that the standard LIE model with $\alpha = 0.18$ could not reproduce the observed binding free energies satisfactorily and thus new models were explored. The models which were able to reproduce the binding free energies best for chymotrypsin, elastase, and trypsin had α 's of 0.63, 0.56, and 0.54 respectively and the results of these new models is shown in *Figure 5*. The predictions are in excellent agreement with experimental results and suggest that LIE is indeed applicable to protein-protein interactions.

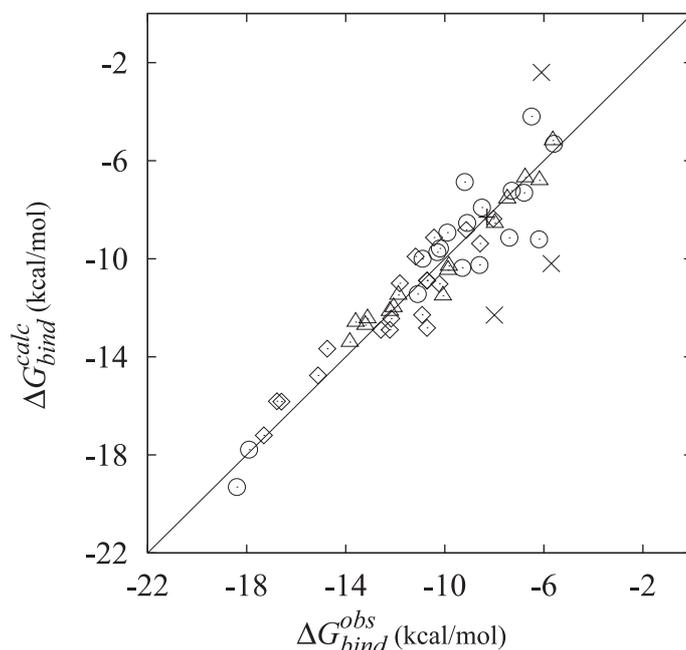


Figure 5. Observed versus predicted binding free energies for P1-variants of OMTKY3 and BPTI binding to chymotrypsin, elastase, and trypsin using new α^2 s. \circ = trypsin (18 points), Δ = elastase (15 points), \diamond = chymotrypsin (19 points), \times = elastase data points which were not included in the parameterization due to the constant secondary interaction assumption not holding, $+$ = P1-Cys ligand binding to trypsin for which no experimental data is available.

This investigation is quite different from Paper I in that the inhibitors in Paper II (BPTI and OMTKY3) are large polypeptides (57 and 50 amino acids respectively). Predicting binding free energies of such large ligands can at first seem as an impossible task. The ligand-surrounding energies of these ligands are most likely in the order of thousands of kcal/mol, and thus convergence of these energies would require very long simulation times. Additionally, in order to avoid errors due to boundary effects, very large simulation systems would be needed. It thus seems inevitable that prohibitively long CPU-times will be needed to reliably calculate binding free energies of protein-protein complexes. To resolve these issues, the protocol used by Brandsdal *et al.*⁶⁴ in which the “ligand” in the LIE framework is only considered to be the atoms which make up the P1 residue, was adopted in Paper II (Figure 6). All interactions between the proteinase and protein inhibitor atoms that did not constitute the P1 residue were assumed to be constant, and thus parameterized into γ in the LIE equation (eq 8). These interactions are referred to as secondary interactions. This assumption seems to be warranted as the deposited crystal structures of ten BPTI P1 variants show hardly any changes to the secondary interactions upon mutating the P1 residue.⁶⁵

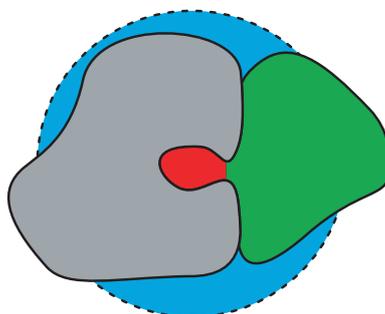


Figure 6. Protocol used in Paper II where the “ligand” in the LIE framework is colored red, the inhibitor green (and red), and the enzyme grey. The secondary interactions are defined as all interactions between the green and grey parts, and where assumed to be constant in the LIE calculations.

In Paper II it was found that part of the reason for the deviation of α from the standard 0.18 is due to a combination of binding site preorganization and a very rigid ligand. The effect of binding site preorganization is reflected in the $\langle \rangle_{off}$ of eq 13, and in most versions of LIE it is assumed to be negligible. The two main advantages of neglecting the *off* term is that for LIE purposes it would require two additional simulations, and that it converges very slowly. As discussed above, for solutes without net charge, the term converges to zero when the environment is water and therefore the simplification of neglecting it is justified. However, it is not apparent beforehand that this simplification will hold for binding sites, which often have a seemingly preorganized electrostatic environment. In order to investigate this, in Paper II we performed the *off* calculations for some small ligands in their binding sites and in water. The ligands consisted of rather hydrophobic ligands to P450cam and some benzamidine-like ligands binding to trypsin. As expected the *off* term in water for the neutral P450cam ligands was zero and for the benzamidine-like ligands it was around -8.5 kcal/mol. Somewhat surprisingly however, the *off* term in the binding site was virtually the same for all ligands as in water. On the contrary, similar calculations on the OMTKY3 P1 variants in water and bound to chymotrypsin revealed a difference in the *off* term between the two environments. In fact, this difference was proportional to the size of the P1 amino acid, and thus scaled rather linearly with $\Delta \langle U_{l-s}^{vdW} \rangle$.

It is thus clear that the *off* term contributes to the binding free energies of OMTKY3 to chymotrypsin and is expected to do so for other protein-protein complexes as well. In *Figure 7* the free energy contribution of the *off* term for small ligands and OMTKY3 P1 variants is plotted against $\Delta \langle U_{l-s}^{vdW} \rangle$. The different α 's obtained for the protein-protein interactions considered in Paper II can be directly attributed to the slopes of the regressions (-0.01 for the small ligands and 0.19 for the OMTKY3 P1 variants). Reparameterizing the

model for chymotrypsin with the *off* term included yields an α of 0.43, as expected ($0.43 \approx 0.63$ (earlier α) $- 0.19$ (slope)). The statistical figures of merit are however worse when including the *off* term which is not so surprising given its inherent poor convergence.

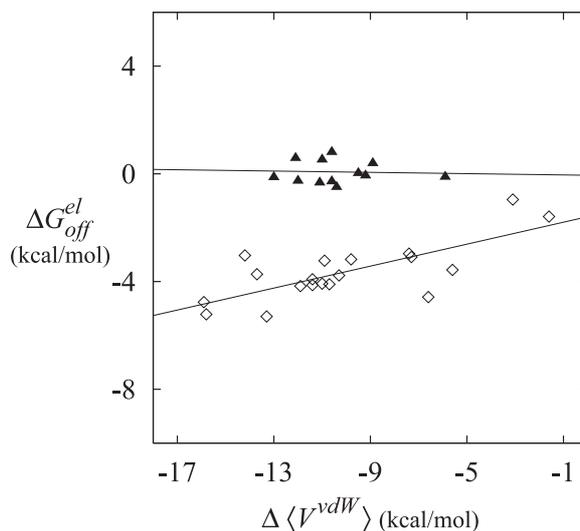


Figure 7. The contribution to the free energy of binding of $\Delta\langle U_{l-s}^{el} \rangle_{off}$, plotted against the change in $\langle U_{l-s}^{vdW} \rangle$ for small ligands (triangles) and larger protein ligands (diamonds). The slopes of the regressions are -0.01 and 0.19 respectively.

Out of the 56 calculated binding free energies, 52 were found to reproduce the observed binding constants. The calculations involving P1 variants of OMTKY3 with large side chains binding to elastase were problematic since the S1 site in elastase is much narrower compared to trypsin and chymotrypsin. Docking of the OMTKY3 variants containing Phe, Trp, and Tyr to elastase was therefore not possible without steric clashes. This suggests that the secondary interactions involving non P1 OMTKY3 residues and elastase are disturbed compared to the OMTKY3 variants with smaller side chains. Thus the methodology of only considering the P1 residue as a ligand in the LIE framework will not work for these cases. The observed binding free energies of these larger side chains also seem to confirm this as they are more positive than the observed binding free energy for the P1-Gly OMTKY3 variant. The effect of squeezing the larger OMTKY3 P1 residues into the S1 site may also be loss of binding site preorganization, which in turn would have the effect of losing the ΔG_{off}^{el} contribution.

Although the uncertainty in the dependence of ΔG_{off}^{el} to $\Delta\langle U_{l-s}^{vdW} \rangle$ is rather large, there still seems to be an unaccounted difference in the α 's obtained for protein-protein complexes compared to small ligand binding. A possible explanation could be relative differences in the entropy of binding. If we consider the change in ligand rotational and translational entropy upon bind-

ing when adding a small moiety to a large protein ligand and adding the same moiety to a small ligand it is reasonable to assume that the small ligand’s binding entropy will be affected more than the protein ligand’s binding entropy (this is formalized in Ref. 66). That is, the difference in the entropic contribution to the free energy of binding is less between very large ligands than between small drug like ligands. Since the relative ligand rotational and translational entropic contribution to binding is anticorrelated to $\Delta\langle U_{l-s}^{vdW} \rangle$ (*i.e.* it lowers α), the proposed diminished relative contribution to binding from translational and rotational entropy for very large ligands will serve to raise α .

3.2.2 LIE and Codon-Anticodon Recognition

Paper III presents an analysis of codon recognition on the small (30S) subunit of the ribosome. This study is similar to that of Paper II in that it deals with the differences in binding free energies of large ligands which are relatively similar. The ligands are anticodon stem loops (ASL) which consist of 17 nucleotides. In order to avoid convergence errors which would arise if the ligand-surrounding energies of the entire ASLs were used in calculating the binding free energies, the same procedure is used as in Paper II; *i.e.* the “ligand” in the LIE framework is only a small portion (the three nucleotide bases of the anticodon) of the entire ASL and the binding contributions from the secondary interactions not involving the anticodon are modeled by a constant γ . The LIE parameters used in Paper III are the standard LIE parameters with $\alpha = 0.18$, β dependent on solute (in this case 0.43 for all ligands) and γ freely parameterizable. From a methodological point of view it would have been of great interest to see if α for this system also takes on a different optimal value. However, given the numerical uncertainties of the simulations and the scarce experimental binding data available (four measured ΔG ’s) it was not warranted to parameterize α freely.

3.3 γ

Unlike several other LIE investigations, the γ ’s obtained in the parameterizations of Paper II reflect the contribution to binding from the secondary interactions of the protein-protein interface. This can be easily verified by comparing the γ ’s obtained with the observed binding free energies for the P1 Gly variants. The γ ’s obtained for elastase, chymotrypsin, and trypsin are -9.1 , -8.0 , and -3.6 kcal/mol which can be compared to the observed free energies of their respective ligand’s P1 Gly variants which are -9.8 , -9.1 , and -5.6 kcal/mol. Many reports using LIE to calculate small ligand binding free energies however require a nonzero γ in order to reproduce absolute binding free energies even though secondary interactions of the kind in Pa-

per II don't exist. In terms of predicting binding free energies, it would be very useful to be able to determine γ beforehand without having to parameterize it using experimental binding data.

The origin of γ can be traced back to the derivation of the nonpolar contribution to the binding free energy (eqs 18-22). In this notation, γ is a complex relationship between the slopes and intercepts of eqs 18-21:

$$\gamma = b_p - b_w - \frac{(a_p - a_w)(d_p - d_w)}{c_p - c_w} \quad 23$$

Alternatively, if the nonpolar solvation contributions for each phase (water or protein) are rewritten as:

$$\Delta G_w^{nonpolar} = \alpha_w \langle U_{l-s}^{vdW} \rangle_w + \gamma_w \quad 24$$

$$\Delta G_p^{nonpolar} = \alpha_p \langle U_{l-s}^{vdW} \rangle_p + \gamma_p \quad 25$$

the resulting nonpolar contribution to the free energy of binding becomes:

$$\Delta \Delta G_{bind}^{nonpolar} = \alpha_p \langle U_{l-s}^{vdW} \rangle_p - \alpha_w \langle U_{l-s}^{vdW} \rangle_w + \gamma_p - \gamma_w \quad 26$$

where

$$\gamma_p - \gamma_w = b_p - b_w - \left(\frac{a_p d_p}{c_p} + \frac{a_w d_w}{c_w} \right) \quad 27$$

Thus γ has different underlying meanings depending on how the nonpolar contribution to binding is calculated (eq 22 or 26). However, if the reasonable assumption that $d_p \approx d_w \approx 0$ is made,⁶⁶ the γ 's of eqs 22 and 26 (γ and $\gamma_p - \gamma_w$ respectively) are equal. At any rate, if we consider eq 26 we see that the free energy of solvation extrapolated to zero van der Waals interaction need to be equal in water (γ_w) and protein phases (γ_p) in order to avoid a constant term in the total LIE equation (eq 8).

Previous reports using LIE which required a nonzero γ to reproduce absolute binding free energies have all been conducted on binding sites which are considerably more hydrophobic than the investigations which did not need a γ . Thus, there seems to be a correlation between γ_p and binding site hydrophobicity and in Paper I a preliminary analysis of this is reported.

In Paper I the binding site hydrophobicity is calculated as the ratio of binding site hydrophobic area to total binding site area. This hydrophobicity measure is then plotted as a function of distance from the binding site center and gives a useful hydrophobicity profile for each binding site. As can be seen from *Figure 8* there is a clear correlation between binding site hydrophobicity and the γ necessary to reproduce absolute binding free energies.

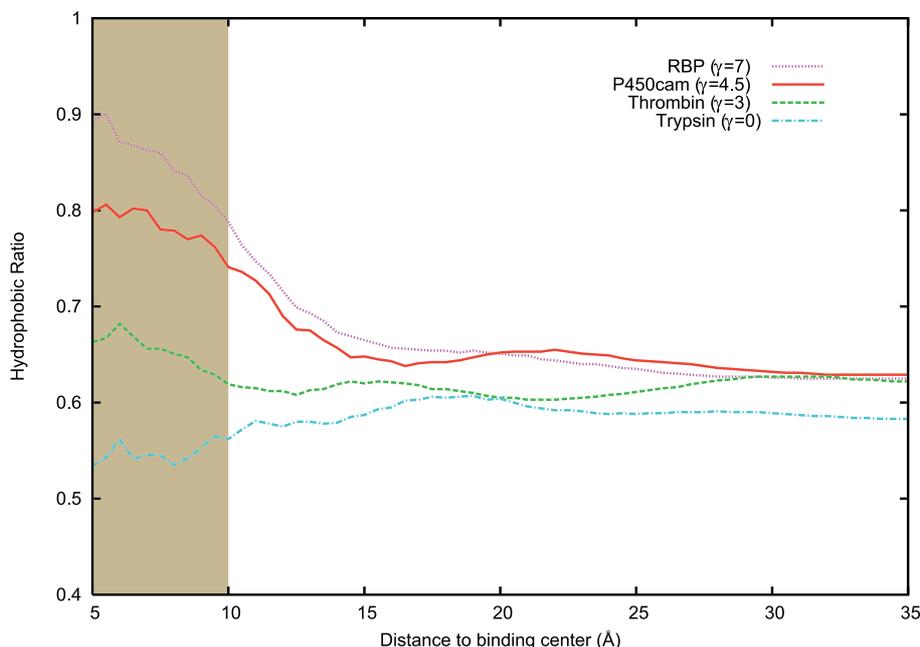


Figure 8. Hydrophobicity of the protein (ratio of nonpolar surface area to total surface area) as a function of the distance from binding site center. Close to the binding site center (shaded in brown – arbitrarily chosen to be distances less than 10 Å) the proteins differ substantially in their hydrophobic profiles and show a clear correlation to the γ used in the LIE calculations. At large radii (when the entire protein is included in the calculation) the hydrophobicity measures for the different proteins converge to about the same value.

The hydrophobicity vs γ analysis in Paper I is only qualitative in nature and does not attempt to describe any quantitative relationships between the hydrophobicity measure and γ . Perhaps with more LIE investigations and alternative hydrophobicity measures, a quantitative relationship between γ and hydrophobicity is feasible.

3.4 Force Field Dependence

It has been suggested that α may be dependent on the force field potential used^{33,34,50,60,62,63} since b_w and b_p (eqs 20 and 21) would seem to be force field dependent. In a report by Paulsen and Ornstein⁶⁰ on the binding of several small compounds to P450cam very good correlation with experiment is obtained using the LIE parameters $\alpha = 1.043$, $\beta = 0.5$ and $\gamma = 0$. The vastly different parameterized value of 1.043 as compared to the standard 0.18 or 0.16 was attributed to the different force fields used in that work (CVFF⁶⁷) and the original LIE paper (GROMOS). It is however unclear if a constant

offset (γ) would yield a better model. Unfortunately, the ligand surrounding energies are not provided in the Paulsen and Ornstein work, and therefore different models are not possible to explore using their data.

Thus, in Paper I the binding of nine ligands to P450cam was investigated. Three different force fields were used (AMBER, OPLS-AA, GROMOS) and the differences between them investigated. It was found that for all three force fields, the standard LIE model with $\alpha = 0.18$ along with a nonzero γ (-4 to -5 kcal/mol in all three cases) was the most appropriate model to use (Model A in Table 4). Thus, for these force fields, the optimal parameters in the LIE method do not seem to depend on the force field used. It should be noted that the *a priori* model was the standard LIE model and thus imposed a slight bias towards this model. Also noteworthy is that the CVFF force field was not tested in this report and thus one can not rule out differences in parameterizations between CVFF and the force fields considered in Paper I.

Table 4. *Effects of different force fields on the optimal model for the P450cam system studied in Paper I*

Model	α	β	γ^a	RMS ^a	$\langle \text{Err} \rangle^a$	R^2	Q_{loo}^2
A _{OPLS-AA}	0.18	0.43	-4.45	0.27	0.23	0.90	0.88
A _{GROMOS}	0.18	0.43	-4.19	0.51	0.45	0.65	0.56
A _{AMBER}	0.18	0.43	-4.82	0.62	0.50	0.48	0.34
C _{OPLS-AA}	0.19	0.43	-4.31	0.27	0.24	0.90	0.81
C _{GROMOS}	0.17	0.43	-4.26	0.51	0.45	0.65	0.36
C _{AMBER}	0.27	0.43	-3.84	0.59	0.49	0.53	-0.15

^a kcal/mol

GROMOS differs the most from the other force fields in this investigation in that it is a united atom force field and is known to be too nonpolar in nature.⁶⁸ This nonpolar nature is also shown in Paper I where GROMOS consistently has the lowest magnitude electrostatic energies and the highest magnitude van der Waals energies. However, the differences between the force fields cancel when taking the differences between the water and protein simulations, and thus the optimal model for GROMOS is the same as for the other force fields. This observation has also been made earlier by Kollman and coworkers⁶² with the benzamidine-trypsin complex.

In Paper IV the validity of linear response of several solutes is investigated. This report is very similar to that of Åqvist *et al.*³¹ except that the

entire gas phase cycle is considered, intramolecular energies are included, and the OPLS-AA force field is used instead of GROMOS in Ref. 31. The derived β 's for nonpolar and hydroxyl containing solutes of 0.43 and 0.37 respectively is the same as that obtained by Åqvist *et al.*³¹ This indicates that the optimal β parameters are not dependent on the force field used.

3.5 The Ribosome

The ribosome's role in living cells is to translate the genetic code into working proteins. It does this by reading trinucleotide segments of mRNA (the codons) and matching them with the appropriate tRNA (at the anticodons). In order to sustain cell growth, the ribosome must do its task quickly and with a very low error frequency (high fidelity). Investigations have shown that the ribosome incorporates erroneous amino acids into the growing peptide chain in the order of 1 per 10^4 amino acids.⁶⁹ The origin of this high fidelity has for a long time eluded the scientific community because the required stability difference between an incorrectly matched tRNA (non/near cognate) and a correctly matched tRNA (cognate) would need to be about 7 kcal/mol to achieve the measured fidelity assuming binding equilibrium is achieved. However, the observed free energy difference for helix formation in solution on changing, for example, an A-U pair to a G-U mismatch is only 0-3 kcal/mol.^{70,71} Thus it seems that the ribosome somehow enhances the free energy differences between cognate base pairs and non/near cognate base pairs compared to solution. Several models of this enhancement have been proposed⁷²⁻⁷⁵ where the latter, proofreading,^{74,75} has been verified experimentally.^{76,77} According to the proofreading concept, the tRNA-mRNA-ribosome interactions are evaluated twice, separated by an irreversible GTP hydrolysis step, and thus the differences in base-pairing thermodynamic stability are exploited two times.

First Letter	Second Letter						Third Letter		
	U		C		A			G	
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA		UCA		UAA	Stop	UGA	Stop	A
	UUG	Leu	UCG	Ser	UAG		UGG	Trp	G
	CUU		CCU	Pro	CAU	His	CGU	Arg	U
	CUC	Leu	CCC		CAC		CGC		C
C	CUA		CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG		CAG		CGG		G
	AUU		ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC	Ile	ACC		AAC		AGC		C
A	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG	Met	ACG		AAG		AGG		G
	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
G	GUC		GCC		GAC		GGC		C
	GUA		GCA	Ala	GAA		GGA		A
	GUG	Val	GCG		GAG	Glu	GGG	Gly	G

Figure 9. The 64 codons and their corresponding amino acids. The red, green, and blue colors signify charged, nonpolar, and polar amino acids respectively.

Recently, Rodnina and coworkers suggested that the tRNA selection at the initial selection step is entirely kinetically controlled and not governed by thermodynamic stability.⁷⁸ They find that the dissociation rates (k_{-2}) for near cognate tRNAs is 350 times slower for cognate tRNAs than for near cognate tRNAs (k_{-2} rates of 0.23 s^{-1} and 85 s^{-1} respectively). This in itself would suggest that discrimination between correct and incorrect tRNAs at the initial selection step is thermodynamically controlled. However, they also report a 650 fold difference in GTPase activation rates (k_3), and more importantly the GTPase activation rate for cognate tRNAs is faster than the observed association rate of codon recognition (k_2) (for cognate tRNAs $k_2 = 190 \text{ s}^{-1}$ and $k_3 = 260$). Thus for cognate tRNAs the GTPase activation is not rate limiting, and thermodynamic equilibrium will never be reached for bound/unbound cognate tRNA. What is neglected in Ref. 78 is that for near cognate tRNAs, GTPase activation becomes rate limiting, and thus differential free energy of binding near cognate tRNAs will affect the rate of near cognate tRNA initial selection.

Recently, several structures and binding free energies of the small ribosomal subunit in complex with various anticodon stem loops (ASLs) were published.^{79,80} This enabled us to perform atomic scale computer simulations on the process of codon recognition and determine the underlying energetic factors of discrimination (Paper III). The complexes considered were the three ASLs: ASL^{Phe} (anticodon 3'-AAG-5'), ASL^{Leu2} (anticodon 3'-GAG-5'), and ASL^{Ser} (anticodon 3'-AGG-5') in complex with ribosomes loaded with mRNA with either the UUU or UUC codon. Thus six ribosomal complexes were analyzed and these were compared to the energetics of the ASLs in water and with the ASLs in complex with the mRNAs in water.

The results of the simulations support the proposed roles of the ribosomal groups A1492, A1493, and G530 in discriminating between cognate and near cognate tRNAs in initial selection.^{79,81} In the first codon-anticodon position, the base pairing is monitored by A1493. In switching from a cognate base pair (U-A, ASL^{Phe} and ASL^{Ser}) at the first codon-anticodon position to a near cognate base pair (U-G, ASL^{Leu2}), the simulations showed that the near cognate base pair loses some of its interactions with A1493. More specifically, the codon uracil is displaced into the major groove in order to retain its base pairing to the anticodon guanine and thus loses the codon 2' hydroxyl – A1493 2' hydroxyl base pair. Thus A1493 amplifies the free energy difference between cognate and near cognate base pairs at the first position. At the second codon position A1492 spans the minor groove and monitors base pairing at the second position through hydrogen bonds with the codon and anticodon bases at the second position. According to the crystallographically determined structures, it only spans the minor groove when cognate tRNAs are bound.⁸⁰ Simulations of a near cognate base pair (U-G, ASL^{Ser}) at the second position with A1492 spanning the minor groove showed that the hydrogen bond interaction between the codon 2' hydroxyl

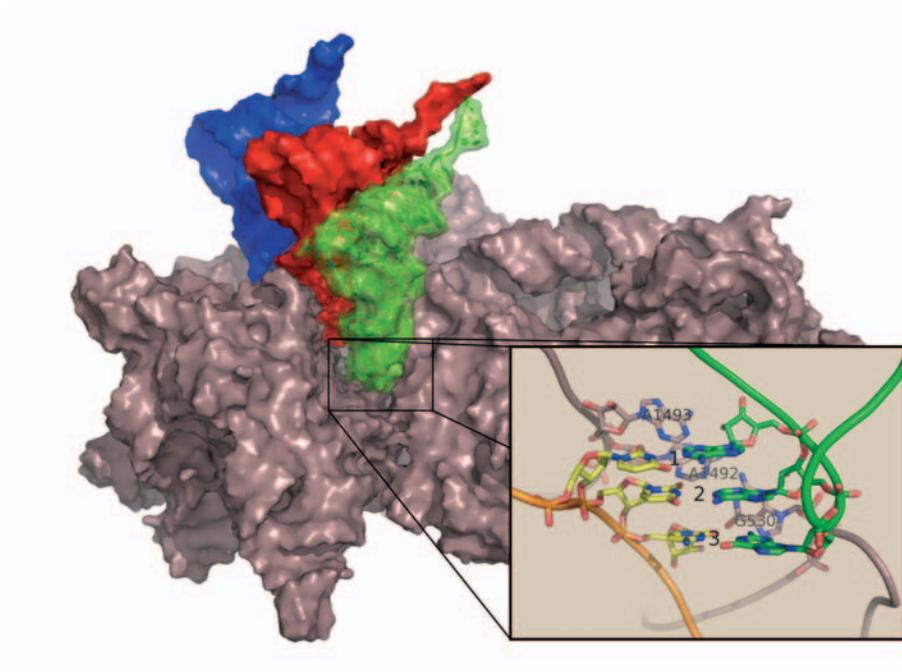


Figure 10. The small subunit (30S) of the ribosome with A-, P-, and E-site tRNAs (green, red, and blue respectively). The insert shows the decoding center of the A-site with the most important residues involved in codon-anticodon recognition. Numbered 1, 2, and 3 are codon-anticodon positions referenced in the text. The yellow, green and magenta colors in the insert depict the codon, anticodon, and ribosomal nucleotides respectively.

and the N3 of A1492 present in the cognate case (ASL^{Phe}) is lost when there is a non Watson-Crick base pair at the second position (U-G, ASL^{Ser}). This gives some indication as to why A1492 does not span the minor groove for near cognate ASLs. At the third position, it was found through the simulations that the base pair is shielded from solvent by G530. This was quantified by a radial distribution function (RDF) for one of the guanine N2 hydrogens to all hydrogen bond acceptors (*Figure 11*). As expected, when in a wobble base pair (U-G) the N2 hydrogen on the guanine is not in as favorable hydrogen bonding environment as when in a Watson-Crick base pair (C-G). However, also evident from the RDF is that the guanine N2 hydrogen in the wobble base pair is in an even less advantageous hydrogen bonding environment than water. Thus the hydrogen bond lost when switching from a third position codon cytosine to uracil is not compensated for. This is primarily attributed to the solvent shielding effect of G530.

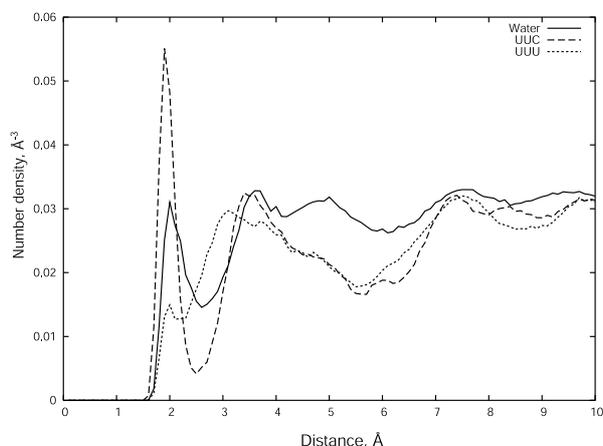


Figure 11. Radial Distribution Function for one of the N2 hydrogens on the anti-codon guanine at the third position to all hydrogen bond acceptors for the water simulation, ASL^{Phe} bound to the ribosome with a UUC codon, and ASL^{Phe} bound to the ribosome with a UUU codon.

The absence of A1492 in the minor groove when near cognate ASLs are bound is also supported by the calculated binding free energies with and without A1492 in the minor groove. These are about 0.5-1 kcal/mol more positive (for ASL^{Leu2} and ASL^{Ser} binding to the ribosome with UUU and UUC codons) when A1492 spans the minor groove compared to when A1492 does not span the minor groove.

Table 5. Energetics of codon-anticodon recognition on the ribosome^a

Anticodon-Codon (3'-5') - (5'-3')	$\langle U_{l-s}^{vdW} \rangle_f$	$\langle U_{l-s}^{vdW} \rangle_b$	$\langle U_{l-s}^{el} \rangle_f$	$\langle U_{l-s}^{el} \rangle_b$	ΔG_{calc}	σ_M^b	ΔG_{obs}^c
AAG-UUC	-46.9	-65.7	-144.3	-138.5	-7.5	0.46	-8.3
AAG-UUU	-46.9	-64.4	-144.3	-137.1	-6.7	0.47	-7.0
GAG-UUC	-46.9	-68.6	-170.1	-158.8	-5.7	0.83	-5.0
GAG-UUU	-46.9	-67.8	-170.1	-157.4	-5.0	0.87	N/A ^d
AGG-UUC	-46.4	-64.9	-174.3	-164.0	-5.5	0.60	-5.2
AGG-UUU	-46.4	-62.7	-174.3	-160.0	-3.4	0.51	N/A ^d

^a All values are kcal/mol. $\langle \rangle_f$ and $\langle \rangle_b$ denote the free and bound states of the ASL. ^b σ_M is the standard error of the mean ΔG_{calc} . ^c Reference 80. ^d No experimental data available.

The calculated binding free energies for the three ASLs binding to ribosome loaded with mRNA with codons UUU or UUC are shown in Table 5 and *Figure 12*. The predicted binding free energies for each ASL are averages of nine 1 ns simulations of the ASL in water and in complex with the ribosome. This yielded $6 \times 9 \times 1$ ns = 54 ns of production phase MD of the bound states and $3 \times 9 \times 1$ ns = 27 ns of the free states. The calculated binding free energies are in good agreement with experiment (rms = 0.55 kcal/mol). The predicted binding free energies of ASL^{Leu2} and ASL^{Ser} binding to ribosomes with a UUU codon in the A-site are -5.0 and -3.4 kcal/mol respectively, indicating that multiple mismatches involving the second codon-anticodon position are more severe than multiple mismatches involving the first position. This is somewhat corroborated by the standard codon table (*Figure 9*), from which it is evident that translation of tRNAs mismatched at the second position yields amino acids of considerably more different nature than translation of tRNAs mismatched at the first position compared to the cognate case.

In order to establish a protocol for future simulations of the other codon-anticodon combinations (there are 4096 possible combinations) the effects of

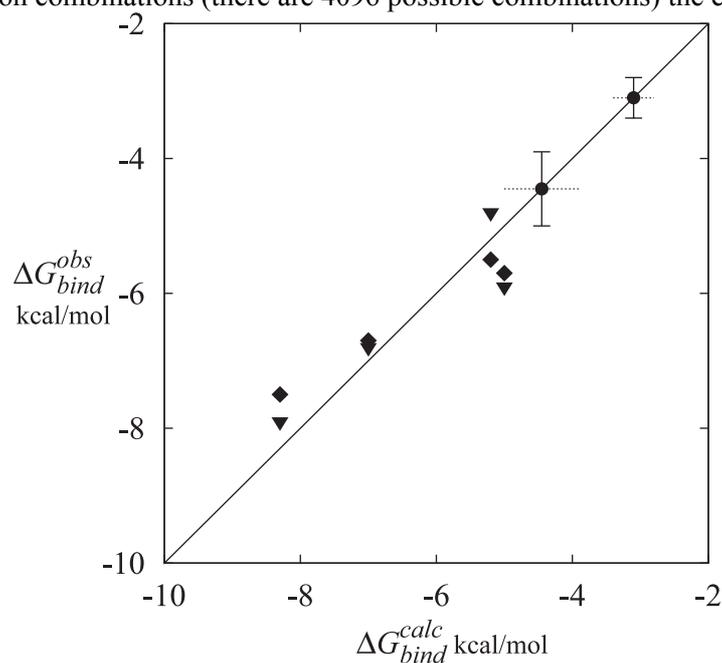


Figure 12. Calculated and observed⁸⁰ binding free energies (kcal/mol) for four ASLs binding to the A-site of the 30s ribosomal subunit as calculated from 1ns (diamonds) and 50ps (triangles) of production phase molecular dynamics. Shown as circles are the predicted binding free energies of the two ASLs for which no experimental data was available. The error bars indicate the difference between the long and short simulations

shortening the production phase MD from 1 ns to 50 ps was also investigated (*Figure 12*). It was found that the agreement with experiment was of similar quality (rms = 0.53 kcal/mol) but with increased error bars.

In order to discern which energetic factors are responsible for the ribosome's role in discriminating between cognate and near cognate ASLs, simulations were carried out with the ASLs bound to mRNAs in solution. The starting structures for these simulations were identical to that of the ribosome-mRNA-tRNA complexes except that the ribosomal parts (and appropriate counter ions) had been removed. The results of these calculations are displayed in Table 6. Comparison of the energetics of ASL-mRNA binding in solution and on the ribosome showed that the ribosome contributes about 1 kcal/mol to discrimination. The enhanced discrimination by the ribosome was primarily attributed to differences in electrostatic interactions between the ASL-mRNA complexes in water and on the ribosome.

Table 6. *Energetics of mRNA-ASL binding in solution*^a

Anticodon-Codon (3'-5') - (5'-3')	$\langle U_{l-s}^{vdW} \rangle_f$	$\langle U_{l-s}^{vdW} \rangle_b$	$\langle U_{l-s}^{el} \rangle_f$	$\langle U_{l-s}^{el} \rangle_b$	ΔG_{calc}	σ_M ^b
AAG-UUC	-46.9	-60.2	-144.3	-145.8	-3.1	0.15
AAG-UUU	-46.9	-57.8	-144.3	-146.8	-3.0	0.05
GAG-UUC	-46.9	-61.6	-170.1	-169.4	-2.4	0.28
GAG-UUU	-46.9	-58.8	-170.1	-169.4	-1.8	0.06
AGG-UUC	-46.4	-59.2	-174.3	-172.8	-1.6	0.08
AGG-UUU	-46.4	-57.9	-174.3	-172.7	-1.4	0.23

^aAll values are kcal/mol. $\langle \rangle_f$ and $\langle \rangle_b$ denote the free and bound states of the ASL. ^b σ_M is the standard error of the mean ΔG_{calc} . Note that the constant γ in eq 8 is set to zero here, since these calculations represent the association of triplets in water.

4 Future Perspectives

With ever increasing computer power, improvements in molecular simulation protocols and simulation program packages, the future of free energy methods such as LIE looks bright. Increased sampling and improved force field potentials will serve to reduce the simulation errors, which make accurate prediction of the empirical parameter α difficult today. The issue of whether or not LRA holds for all protein environments will, however, not be solved by increased sampling or improved force field parameters. Although the successes of LIE and similar methods which rely on LRA published in the literature do indeed suggest that LRA holds for the protein environment, there are several examples of LIE calculations where abnormal scaling of the electrostatics has been necessary. The question of whether or not these scaling factors indicate LRA failure in the protein environment cannot be answered until rigorous calculations of LRA on the protein environment are performed. This will be quite a daunting task, as several types of active sites will need to be investigated, and the interpretation of the *off* term in the protein environment is somewhat unclear.

The correct interpretation and parameterization of the electrostatic scaling factor, β , for the bound state is a matter which may be solved by theoretically rigorous approaches. The correct treatment of the nonpolar contribution, however, is an issue which is inherently more complex to address. Since α and γ are parameterized on experimental binding constants, in their current form they represent everything about the calculated binding free energy which the calculated electrostatic contribution does not. This includes hydrophobic effects, solute translational and rotational entropies (and to some extent conformational entropy), van der Waals enthalpies, and perhaps more importantly – force field and simulation errors. Most of these quantities are very difficult to assess through molecular dynamics simulations, which is precisely the reason for parameterizing α and γ on experimental binding data. Thus, in order to properly account for (or rule out) any binding site dependence of α and γ , it is necessary to perform LIE calculations on many datasets of ligands binding to many binding sites. Preferably, experimental structures should be available for all ligand-macromolecular complexes in order to minimize possible simulation errors entering into the optimal parameterization of α and γ . Then perhaps, along with a proper size measure (SASA, U^{vdW} , number of atoms, etc.) and binding site hydrophobic-

ity measure (various surface area ratios, ratio between U^{vdW} and U^{el} , etc) the final word can be said on α and γ .

Nevertheless, the validation, application, and improvement of the LIE-method presented in this thesis has served to improve the understanding of the underlying mechanisms of LIE and as such will make more improvement possible. The exponential increase in computer power will make LIE an important tool in the arsenal of methods available in computer aided drug design and improve the outlook of this field.

5 Summary in Swedish

I denna avhandling presenteras utveckling och tillämpning av en datorbaserad metod vars syfte är att beräkna bindningskonstanter mellan ligander och makromolekyler. Metoden kallas ”linear interaction energy” (LIE) och använder sig av datorsimuleringar på atomär nivå för att beräkna hur hårt en ligand binder till sitt makromolekylära mål (receptor). Det finns en uppsjö av datorbaserade metoder för att beräkna bindningskonstanter och alla har sina fördelar och nackdelar. Några fördelar med LIE är:

- LIE har relativt få empiriska parametrar
- Ligandens vattenmiljö tas hand om explicit i LIE
- Receptorn kan anpassa sig till ligandens form i LIE
- LIE kräver mindre beräkningstid än de mest rigorösa metoderna

I en LIE-beräkning utförs två datorsimuleringar för varje ligand: liganden i vatten och liganden bunden till sin receptor. Datorsimuleringarna utförs med hjälp av s.k. kraftfält, vilka bestämmer med vilken energi atomerna interagerar och således styr hur atomerna rör sig. Det är skillnaden i ligandens interaktionsenergi när den är fri i vatten och bunden till receptorn som LIE använder sig av för att beräkna bindningskonstanter. Dessa interaktionsenergier består av elektrostatiske och van der Waals-komponenter som, i LIE, skalas med konstanterna β och α (ekv. 28)

$$\Delta G_{bind}^{LIE} = \alpha \Delta U^{vdW} + \beta \Delta U^{el} + \gamma \quad 28$$

I ekv. 28 är ΔG_{bind}^{LIE} den beräknade styrkan på ligandens bindning, ΔU^{vdW} är skillnaden i van der Waals komponenten av interaktionsenergin, och ΔU^{el} är skillnaden i den elektrostatiske komponenten av interaktionsenergin. Parametrarna α , β , och γ bestämmer hur stor del av de olika komponenternas interaktionsenergi bidrar till styrkan på bindningen.

Grunden i LIE baseras på två observationer. Den första är att det elektrostatiske bidraget till solvatiseringsenergin för molekyler i ett dielektriskt medium är hälften av dess interaktionsenergi med mediet. Den andra observationen är att solvatiseringsenergin för alifatiska molekyler är proportionell mot deras storlek. Denna observation i kombination med att van der Waals komponenten av interaktionenergin också är proportionell med molekylstor-

leken ger att den s.k. icke-polära solvatiseringsenergin är proportionell mot van der Waals-komponenten av interaktionsenergin. Dessa linjära förhållanden till de olika interaktionsenergikomponenterna ger slutligen LIE-förhållandet till bindningsstyrkan om man föreställer sig receptorn som ett solvatiseringsmedium. Således bestämmer parametrarna α och γ det icke-polära bidraget till bindningsstyrkan och β det polära bidraget till bindningsstyrkan. Det är dessa parametrars beroende på kraftfält, receptor och ligand som är fokus i denna avhandling.

Som sagt så baserar sig LIE delvis på observationen att det elektrostatiska bidraget till solvatiseringsenergin för molekyler i ett dielektriskt medium är hälften av dess elektrostatiska interaktionsenergi. Detta borde ge $\beta = 1/2$ i ekv. 28. Ett arbete från 1996 visade dock att β varierar något beroende på laddningsfördelningen i molekylen. För oladdade molekyler, och i synnerhet alkoholer, så var β mindre än $1/2$. Denna undersökning var dock något begränsad i storlek med avseende på antal molekyler och variationen i molekylernas sammansättning. I avhandlingen presenteras en undersökning på ett stort antal molekyler med varierande sammansättning där resultaten bekräftar den tidigare uppvisade avvikelser av β för oladdade molekyler. Studien föreslår nya värden på β för amider, aminer, karboxylsyror, negativt laddade grupper och positivt laddade grupper. En ny metod att prediktera β för grupper med flera funktionella grupper föreslås också.

För att utröna om olika kraftfält påverkar parametern α analyserades icke-polära liganders bindning till proteinet P450cam med tre olika kraftfält: GROMOS, AMBER och OPLS-AA. En tidigare studie av en annan forskargrupp visade att α varierade stort från det tidigare förslagna $\alpha = 0,16$. Deras slutsats var att detta berodde på de olika kraftfälten som hade använts (CVFF och GROMOS). Enligt arbetet som presenteras i avhandlingen så varierade inte α nämnvärt mellan de olika kraftfälten, och den bästa modellen för alla tre kraftfält var med $\alpha = 0,18$ och $\gamma \approx -4,5$ kcal/mol. Denna konstant, γ , var något ny i sammanhanget och visade sig vara beroende på hur icke-polär receptorn var.

I avhandlingen presenteras också ett arbete på tillämpning av LIE-metoden på protein-proteininteraktioner. Protein-proteinkomplexen består av olika P1-mutanter av "bovine pancreatic trypsin inhibitor" (BPTI) och "the third domain of the turkey ovomucoid inhibitor" (OMTKY3) bundna till α -chymotrypsin, mänskligt leukocyt-elastas, eller bovint β -trypsin. Ett relativt nytt protokoll testas, där "liganden" i LIE-sammanhang består endast av den aminosyra som skiljer sig mellan de olika inhibitorerna (P1 aminosyran). Beräkningarna visar god överensstämmelse med experiment – dock med ett annat värde på α . En stor del av avvikelserna från det tidigare förslagna värdet på $\alpha = 0,18$ härrör från elektrostatiska preorganisationseffekter. Det vill säga, till skillnad från "vanliga" ligand-makromolekyl-interaktioner så finns det ett storleksberoende i preorganisationsenergin för protein-proteininteraktioner.

Slutligen presenteras en undersökning om vilka faktorer som bidrar till att ribosomer uppvisar så hög noggrannhet i proteinsyntesen. Ribosomens uppgift är att översätta den genetiska koden till fungerande proteiner. Detta görs genom att, i ribosomen, para ihop kodon på budbärar-RNA (mRNA) med antikodon på transport-RNA (tRNA). Om fel tRNA paras ihop med ett mRNA i ribosomen så kommer fel aminosyra att inkorporeras i det tillverkade proteinet. Tidigare undersökningar på RNA-RNA komplex i vatten visar att den uppmätta noggrannheten i proteinsyntesen inte enbart kan förklaras av den inneboende skillnaden i stabilitet mellan rätt och fel kodon-antikodonkombination.

I arbetet som presenteras i avhandlingen visas att ribosomen bidrar med ca 1 kcal/mol till diskriminering mellan rätt och fel kodon-antikodonkombination. Detta beror främst på elektrostatiske interaktioner mellan kodon-antikodonkomplexet och ribosomen. De för ribosomen viktiga nukleotiderna G530, A1492 och A1493 spelar alla en roll i ribosomens förmåga att diskriminera mellan rätt och fel kodon-antikodonkombination. Mekanismen för den ökade diskrimineringen är både igenkänning av s.k. Watson-Crick baspar (genom A1492 och A1493) och den vattenavskärmande effekten av G530.

6 Acknowledgements

Johan Åqvist – Thank you. Your straightforward attitude and trust have made these past years very enjoyable and enlightening. From sodium ions to ribosomes, you always know where to start looking for answers.

Past and present members of the Åqvist group – **Martin Andér, Sinisa Bjelic, Bjørn Brandsdal, Jens Carlsson, Isabella Feierberg, Hugo Guterrez de Teran, Martin Nervall, Viktor Luzhkov, Lars Persson, Stefan Trobro, and Fredrik Österberg**. Lots of good times in and out of the lab: San Diego, MM, Grövelsjön, Norreda, Antalus, Tromsø, Istanbul, Lund, AFM, Domarringen

Bjørn Brandsdal – For helping me get started, inviting me to Tromsø (!), always taking the four hour plane trip from Tromsø as soon as someone mentions “julbord”, and putting up a good fight in UT.

Jens Carlsson – The (sometimes fervent) scientific discussions with you have contributed a lot to my understanding in this field. And thank you for all the help with Paper IV and proofreading of the thesis.

Christian Sköld – What would lunch at Bihålan be without our meaningless theorycrafting? Thanks for the snus.

Beer Club regents – **Al MacKenzie-Hose, Jimmy Lindberg, Andrea Hinas, Martin Welin, Daniel Ericsson, Annette Roos, Lotta Aveson, and Agata Naworyta**

everyoneelse@xray

Sven, Cartman, Kournikova_, Bugs, Vanilla, Wilander, Trasgu, FOE – Better to have tried and failed than to never have tried at all.

Anna Maria Lundin – Thank you for the financial support. R.I.P.

Helena – Thank you for putting up with me during the past weeks. Du är underbar!

7 References

1. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *Journal of the American Chemical Society* 117, 5179-5197.
2. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. (1996) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, *Journal of the American Chemical Society* 118, 11225-11236.
3. van Gunsteren, W. F.; Berendsen, H. J. C. Groningen Molecular Simulation (GROMOS) Library Manual; Biomos B.V.: Nijenborgh, Groningen, The Netherlands, 1987.
4. MacKerell, A. D.; Wiorkiewicz-Kuczera, J.; Karplus, M. (1995) An all-atom empirical energy function for the simulation of nucleic-acids, *Journal of the American Chemical Society* 117, 11946-11975.
5. Marelus, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. (1998) Q: A molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems, *Journal of Molecular Graphics & Modelling* 16, 213-225.
6. Åqvist, J.; Medina, C.; Samuelsson, J. E. (1994) A new method for predicting binding affinity in computer-aided drug design, *Protein Engineering* 7, 385-391.
7. Kollman, P. (1993) Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena, *Chemical Reviews* 93, 2395-2417.
8. Beveridge, D. L.; Dicapua, F. M. (1989) Free-Energy Via Molecular Simulation - Applications to Chemical and Biomolecular Systems, *Annual Review of Biophysics and Biophysical Chemistry* 18, 431-492.
9. Warshel, A.; Weiss, R. M. (1980) An empirical valence bond approach for comparing reactions in solutions and in enzymes, *Journal of the American Chemical Society* 102, 6218-6226.
10. Åqvist, J.; Warshel, A. (1993) Simulation of Enzyme-Reactions Using Valence-Bond Force-Fields and Other Hybrid Quantum-Classical Approaches, *Chemical Reviews* 93, 2523-2544.
11. Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *Journal of Chemical Physics* 81, 3684-3690.

12. King, G.; Warshel, A. (1989) A Surface Constrained All-Atom Solvent Model for Effective Simulations of Polar Solutions, *Journal of Chemical Physics* 91, 3647-3661.
13. Zwanzig, R. (1954) High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases, *The Journal of Chemical Physics* 22, 1420-1426.
14. Tembe, B. L.; McCammon, J. A. (1984) Ligand-receptor interactions, *Computers & Chemistry* 8, 281-283.
15. Beutler, T. C.; Mark, A. E.; Vanschaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. (1994) Avoiding Singularities and Numerical Instabilities in Free-Energy Calculations Based on Molecular Simulations, *Chemical Physics Letters* 222, 529-539.
16. Zacharias, M.; Straatsma, T. P.; Mccammon, J. A. (1994) Separation-Shifted Scaling, a New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration, *Journal of Chemical Physics* 100, 9025-9031.
17. Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. (2003) Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins, *Journal of Chemical Physics* 119, 5740-5761.
18. Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6, *Journal of Computational Chemistry* 25, 1656-1676.
19. Shirts, M. R.; Pande, V. S. (2005) Solvation free energies of amino acid side chain analogs for common molecular mechanics water models, *Journal of Chemical Physics* 122, 134508.
20. Oostenbrink, C.; Juchli, D.; van Gunsteren, W. F. (2005) Amine hydration: A united-atom force-field solution, *ChemPhysChem* 6, 1800-1804.
21. Hess, B.; van der Vegt, N. F. A. (2006) Hydration thermodynamic properties of amino acid analogues: A systematic comparison of biomolecular force fields and water models, *Journal of Physical Chemistry B* 110, 17616-17626.
22. Liu, H. Y.; Mark, A. E.; van Gunsteren, W. F. (1996) Estimating the relative free energy of different molecular states with respect to a single reference state, *Journal of Physical Chemistry* 100, 9485-9494.
23. Schäfer, H.; Van Gunsteren, W. F.; Mark, A. E. (1999) Estimating relative free energies from a single ensemble: Hydration free energies, *Journal of Computational Chemistry* 20, 1604-1617.
24. Oostenbrink, B. C.; Pitera, J. W.; van Lipzig, M. M. H.; Meerman, J. H. N.; van Gunsteren, W. F. (2000) Simulations of the estrogen receptor ligand-binding domain: Affinity of natural ligands and xenoestrogens, *Journal of Medicinal Chemistry* 43, 4594-4605.
25. Oostenbrink, C.; Van Gunsteren, W. F. (2003) Single-step perturbations to calculate free energy differences from unphysical reference

- states: Limits on size, flexibility, and character, *Journal of Computational Chemistry* 24, 1730-1739.
26. Oostenbrink, C.; van Gunsteren, W. F. (2005) Free energies of ligand binding for structurally diverse compounds, *Proceedings of the National Academy of Sciences of the United States of America* 102, 6750-6754.
 27. Åqvist, J.; Mowbray, S. L. (1995) Sugar Recognition by a Glucose/Galactose Receptor - Evaluation of Binding Energetics from Molecular-Dynamics Simulations, *Journal of Biological Chemistry* 270, 9978-9981.
 28. Hansson, T.; Åqvist, J. (1995) Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations, *Protein Engineering* 8, 1137-1144.
 29. Åqvist, J. (1996) Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions, *Journal of Computational Chemistry* 17, 1587-1597.
 30. Hultén, J.; Bonham, N. M.; Nillroth, U.; Hansson, T.; Zuccarello, G.; Bouzide, A.; Åqvist, J.; Classon, B.; Danielson, U. H.; Karlén, A.; Kvarnström, I.; Samuelsson, B.; Hallberg, A. (1997) Cyclic HIV-1 protease inhibitors derived from mannitol: Synthesis, inhibitory potencies, and computational predictions of binding affinities, *Journal of Medicinal Chemistry* 40, 885-897.
 31. Åqvist, J.; Hansson, T. (1996) On the Validity of Electrostatic Linear Response in Polar Solvents, *Journal of Physical Chemistry* 100, 9512-9521.
 32. Hansson, T.; Marelius, J.; Åqvist, J. (1998) Ligand binding affinity prediction by linear interaction energy methods, *Journal of Computer-Aided Molecular Design* 12, 27-35.
 33. Jones-Hertzog, D. K.; Jorgensen, W. L. (1997) Binding Affinities for Sulfonamide Inhibitors with Human Thrombin Using Monte Carlo Simulations with a Linear Response Method, *Journal of Medicinal Chemistry* 40, 1539-1549.
 34. Wall, I. D.; Leach, A. R.; Salt, D. W.; Ford, M. G.; Essex, J. W. (1999) Binding Constants of Neuraminidase Inhibitors: An Investigation of the Linear Interaction Energy Method, *Journal of Medicinal Chemistry* 42, 5142-5152.
 35. Stjernschantz, E.; Marelius, J.; Medina, C.; Jacobsson, M.; Vermeulen, N. P. E.; Oostenbrink, C. (2006) Are automated molecular dynamics simulations and binding free energy calculations realistic tools in lead optimization? An evaluation of the linear interaction energy (LIE) method, *Journal of Chemical Information and Modeling* 46, 1972-1983.
 36. Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. (1998) Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate - DNA helices, *Journal of the American Chemical Society* 120, 9401-9409.

37. Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. (2000) Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models, *Accounts of Chemical Research* 33, 889-897.
38. Honig, B.; Nicholls, A. (1995) Classical Electrostatics in Biology and Chemistry, *Science* 268, 1144-1149.
39. Sitkoff, D.; Sharp, K. A.; Honig, B. (1994) Accurate Calculation of Hydration Free-Energies Using Macroscopic Solvent Models, *Journal of Physical Chemistry* 98, 1978-1988.
40. Kuhn, B.; Kollman, P. A. (2000) A ligand that is predicted to bind better to avidin than biotin: Insights from computational fluorine scanning, *Journal of the American Chemical Society* 122, 3909-3916.
41. Kuhn, B.; Kollman, P. A. (2000) Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models, *Journal of Medicinal Chemistry* 43, 3786-3791.
42. Wang, J. M.; Morin, P.; Wang, W.; Kollman, P. A. (2001) Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA, *Journal of the American Chemical Society* 123, 5221-5230.
43. Wang, W.; Lim, W. A.; Jakalian, A.; Wang, J.; Wang, J. M.; Luo, R.; Bayly, C. T.; Kollman, P. A. (2001) An analysis of the interactions between the Sem-5 SH3 domain and its ligands using molecular dynamics, free energy calculations, and sequence analysis, *Journal of the American Chemical Society* 123, 3986-3994.
44. Huo, S. H.; Wang, J. M.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D. (2002) Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: Insight into structure-based ligand design, *Journal of Medicinal Chemistry* 45, 1412-1419.
45. Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. (2005) Validation and use of the MM-PBSA approach for drug discovery, *Journal of Medicinal Chemistry* 48, 4040-4048.
46. Weis, A.; Katebzadeh, K.; Soderhjelm, P.; Nilsson, I.; Ryde, U. (2006) Ligand affinities predicted with the MM/PBSA method: Dependence on the simulation method and the force field, *Journal of Medicinal Chemistry* 49, 6596-6606.
47. Pearlman, D. A. (2005) Evaluating the molecular mechanics Poisson-Boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase, *Journal of Medicinal Chemistry* 48, 7796-7807.
48. Vorobjev, Y. N.; Hermans, J. (1999) A critical analysis of methods of calculation of a potential in simulated polar liquids: Strong arguments in favor of "molecule-based" summation and of vacuum boundary

- conditions in Ewald summation, *Journal of Physical Chemistry B* 103, 10234-10242.
49. Kastenholz, M. A.; Hünenberger, P. H. (2006) Computation of methodology-independent ionic solvation free energies from molecular simulations. I. The electrostatic potential in molecular liquids, *Journal of Chemical Physics* 124, 124106.
 50. Lamb, M. L.; Tirado-Rives, J.; Jorgensen, W. L. (1999) Estimation of the binding affinities of FKBP12 inhibitors using a linear response method, *Bioorganic & Medicinal Chemistry* 7, 851-860.
 51. Tounge, B. A.; Reynolds, C. H. (2003) Calculation of the binding affinity of beta-secretase inhibitors using the linear interaction energy method, *Journal of Medicinal Chemistry* 46, 2074-2082.
 52. Tounge, B. A.; Rajarnani, R. T.; Baxter, E. W.; Reitz, A. B.; Reynolds, C. H. (2006) Linear interaction energy models for beta-secretase (BACE) inhibitors: Role of van der Waals, electrostatic, and continuum-solvation terms, *Journal of Molecular Graphics & Modeling* 24, 475-484.
 53. Ben-Naim, A.; Marcus, Y. (1984) Solvation thermodynamics of non-ionic solutes, *Journal of Chemical Physics* 81, 2016-2027.
 54. Ersmark, K.; Feierberg, I.; Bjelic, S.; Hultén, J.; Samuelsson, B.; Åqvist, J.; Hallberg, A. (2003) C-2-symmetric inhibitors of Plasmodium falciparum plasmepsin II: Synthesis and theoretical predictions, *Bioorganic & Medicinal Chemistry* 11, 3723-3733.
 55. Marelius, J.; Graffner-Nordberg, M.; Hansson, T.; Hallberg, A.; Åqvist, J. (1998) Computation of affinity and selectivity: Binding of 2,4-diaminopteridine and 2,4-diaminoquinazoline inhibitors to dihydrofolate reductases, *Journal of Computer-Aided Molecular Design* 12, 119-131.
 56. Ersmark, K.; Feierberg, I.; Bjelic, S.; Hamelink, E.; Hackett, F.; Blackman, M. J.; Hultén, J.; Samuelsson, B.; Åqvist, J.; Hallberg, A. (2004) Potent inhibitors of the Plasmodium falciparum enzymes plasmepsin I and II devoid of cathepsin D inhibitory activity, *Journal of Medicinal Chemistry* 47, 110-122.
 57. Graffner-Nordberg, M.; Kolmodin, K.; Åqvist, J.; Queener, S. F.; Hallberg, A. (2001) Design, synthesis, computational prediction, and biological evaluation of ester soft drugs as inhibitors of dihydrofolate reductase from Pneumocystis carinii, *Journal of Medicinal Chemistry* 44, 2391-2402.
 58. Leiros, H. K. S.; Brandsdal, B. O.; Andersen, O. A.; Os, V.; Leiros, I.; Helland, R.; Otlewski, J.; Willassen, N. P.; Smalås, A. O. (2004) Trypsin specificity as elucidated by LIE calculations, X-ray structures, and association constant measurements, *Protein Science* 13, 1056-1070.
 59. Luzhkov, V. B.; Åqvist, J. (2001) Mechanisms of tetraethylammonium ion block in the KcsA potassium channel, *FEBS Letters* 495, 191-196.

60. Paulsen, M. D.; Ornstein, R. L. (1996) Binding free energy calculations for P450cam-substrate complexes, *Protein Engineering* 9, 567-571.
61. Wang, W.; Wang, J.; Kollman, P. A. (1999) What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations?, *Proteins* 34, 395-402.
62. Wang, J.; Dixon, R.; Kollman, P. A. (1999) Ranking ligand binding affinities with avidin: a molecular dynamics-based interaction energy study, *Proteins* 34, 69-81.
63. Chen, J.; Wang, R.; Taussig, M.; Houk, K. N. (2001) Quantitative Calculations of Antibody-Antigen Binding: Steroid-DB3 Binding Energies by the Linear Interaction Energy Method, *Journal of Organic Chemistry* 66, 3021-3026.
64. Brandsdal, B. O.; Åqvist, J.; Smalås, A. O. (2001) Computational analysis of binding of P1 variants to trypsin, *Protein Science* 10, 1584-1595.
65. Helland, R.; Otlewski, J.; Sundheim, O.; Dadlez, M.; Smalås, A. O. (1999) The crystal structures of the complexes between bovine beta-trypsin and ten P-1 variants of BPTI, *Journal of Molecular Biology* 287, 923-942.
66. Carlsson, J.; Åqvist, J. (2006) Calculations of solute and solvent entropies from molecular dynamics simulations, *Physical Chemistry Chemical Physics* 8, 5385-5395.
67. Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. (1988) Structure and energetics of ligand binding to proteins: Escherichia coli dihydrofolate reductase-trimethoprim, a drug-receptor system, *Proteins* 4, 31-47.
68. Åqvist, J.; Alvarez, O.; Eisenman, G. (1992) Ion-Selective Properties of a Small Ionophore in Methanol Studied by Free-Energy Perturbation Simulations, *Journal of Physical Chemistry* 96, 10019-10025.
69. Kurland, C. G.; Hughes, D.; Ehrenberg, M. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*; Neidhart, F. C.; Curtiss III, R.; Lin, E. C. C.; Low, K. B.; Magasanik, B., Eds.; American Society for Microbiology Press: Washington, D.C., 1996, p 979-1004.
70. Sugimoto, N.; Kierzek, R.; Freier, S. M.; Turner, D. H. (1986) Energetics of internal GU mismatches in ribooligonucleotide helices, *Biochemistry* 25, 5755-5759.
71. Meroueh, M.; Chow, C. S. (1999) Thermodynamics of RNA hairpins containing single internal mismatches, *Nucleic Acids Research* 27, 1118-1125.
72. Kurland, C. G.; Rigler, R.; Ehrenberg, M.; Blomberg, C. (1975) Allosteric mechanism for codon-dependent tRNA selection on ribosomes, *Proceedings of the National Academy of Sciences of the United States of America* 72, 4248-4251.

73. Potapov, A. P. (1982) A stereospecific mechanism for the aminoacyl-tRNA selection at the ribosome, *FEBS Letters* 146, 5-8.
74. Hopfield, J. J. (1974) Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity, *Proceedings of the National Academy of Sciences of the United States of America* 71, 4135-4139.
75. Ninio, J. (1975) Kinetic amplification of enzyme discrimination, *Biochimie* 57, 587-595.
76. Thompson, R. C.; Stone, P. J. (1977) Proofreading of the codon-anticodon interaction on ribosomes, *Proceedings of the National Academy of Sciences of the United States of America* 74, 198-202.
77. Ruusala, T.; Ehrenberg, M.; Kurland, C. G. (1982) Is there proofreading during polypeptide synthesis?, *EMBO Journal* 1, 741-745.
78. Gromadski, K. B.; Rodnina, M. V. (2004) Kinetic determinants of high-fidelity tRNA discrimination on the ribosome, *Molecular Cell* 13, 191-200.
79. Ogle, J. M.; Brodersen, D. E.; Clemons, W. M.; Tarry, M. J.; Carter, A. P.; Ramakrishnan, V. (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit, *Science* 292, 897-902.
80. Ogle, J. M.; Murphy, F. V.; Tarry, M. J.; Ramakrishnan, V. (2002) Selection of tRNA by the ribosome requires a transition from an open to a closed form, *Cell* 111, 721-732.
81. Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics, *Nature* 407, 340-348.

Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Science and Technology 255*

Editor: The Dean of the Faculty of Science and Technology

A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-7421



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2007