From KAROLINSKA INSTITUTET
Department of Biosciences at NOVUM
Center for Structural Biochemistry
Stockholm, Sweden

MODELLING BIOMOLECULAR INTERACTIONS

AKADEMISK AVHANDLING

Som för avläggande av teknologie doktorsexamen vid Karolinska Institutet offentligen försvaras på svenska språket i sal 215, Pontus, fredagen 6 juni 2003, kl. 9.00

av

Johan Bredenberg

Civ. Ing.

Huvudhandledare: Professor Lennart Nilsson
Karolinska Institutet, Department of Biosciences,
Stockholm, Sweden

Fakultetsopponent: Professor Johan Åqvist
Uppsala University, Department of Cell and Molecular Biology,
Uppsala, Sweden

Stockholm 2003

To any member of my past, present and future family

ABSTRACT

Computational approaches for understanding and aiding in molecular biology has increased in significance over the last decades, where a wealth of biochemical experiments have provided a solid ground for developing computer models that can be used to predict unresolved issues within biology. Molecular dynamics (MD) is one of these methods, based on classical laws, and suitable for handling large macromolecules in their natural environment, water. A detailed picture at the atomic level can be obtained, and given that the formulation of the computer model is correct, new strategies for experiments and interpretation of real-world results are feasible.

An important target group for treating various diseases is found within the nuclear receptor super family, which is the largest known group of transcription regulators. They are ligand induced and promote gene regulation by recognising a specific DNA-sequence. This recognition is performed by a discrete functional DNA-binding domain (DBD), which consists of two perpendicularly packed amphipathic helix loop regions with eight out of nine invariant cysteine residues formed in two Cys4-zinc fingers.

This thesis is mainly based on molecular dynamics simulations performed on the glucocorticoid receptor DNA-binding domain (GR DBD) and follows two main topics:

Develop and evaluate methods for describing metals within the framework of classical laws and resolve some of the biology behind recognition of protein-DNA assemblies.

Nonbonded models that only include the Coulomb electrostatic and van der Waals interactions between Zn to its ligands were in better agreement with experimental data. Being satisfied with the nonbonded metal description, MD was conducted variant structures of GR DBD. One ligand (C496) in the second zinc finger appeared less rigid than the other ligands in conjunction with conformational changes in this region. In some cases ligand exchange with water was observed, which led to the speculation that P493 acts as a regulator for the conformational switch of C476.

A virtual alanine-scan was done on variant DNA-DBD complexes and on the free GR DBD monomer for defining putative mutations that may be done in vivo and/or in vivo. In particular, P493 mutants appeared to stabilise the free protein monomer GR DBD as well as the dimer GR DBD associated with DNA. MD simulations on variant P493-mutants, suggest that P493 regulates the relative positions of secondary structure elements within the GR DBD, and that mutations involving sidechains with hydrogen bonding capabilities recovers the loss of steric isomerisation that is encountered in the wild type protein.
The methods and results presented in this study are readily applied to other DBDs free or associated with DNA.

Keywords: Computational chemistry, molecular dynamics, nuclear receptors, force field parameterization, zinc fingers, DNA, proteins.

ISBN 91-7349-571-9

Main references

The following references are referred to in the text using roman numerals:

I. Johan Bredenberg and Lennart Nilsson (2001) ; Modeling zinc sulfhydryl bonds in zinc fingers; International Journal of Quantum Chemistry ; **83** ; 230 – 244

II. Johan Bredenberg and Lennart Nilsson (2002) ; Conformational states in the glucocorticoid receptor DNA-binding domain from molecular dynamics simulations; Proteins: Structure, Function and Genetics; **49**; 24 – 36

III. Johan Bredenberg, Pekka Mark and Lennart Nilsson (2003) ; Solvent effects in biomolecular dynamics simulations: A comparison of TIP3P, SPC and SPC/E acting on the glucocorticoid receptor DNA-binding domain; Submitted to Chemical Physics Letters

IV. Johan Bredenberg and Lennart Nilsson (2003) ; Computational alanine scanning on the glucocorticoid receptor DNA-binding domain reveals a biological function for Pro 493. Manuscript

Published and printed by Karolinska University Press
Box 200, SE-171 77 Stockholm, Sweden

ISBN 91-7349-571-9

CONTENTS

LIST OF ABBREVIATIONS

MD	Molecular dynamics
QM	Quantum Chemistry
DFT	Density functional theory
HF	Hartree-Fock
COSMO	Conductor-like screening model
GBMV	Generalized Born molecular volume
bold letter	vector
CHARMM	Chemistry Harvard Molecular Mechanics
NR	Nuclear receptors
DBD	DNA-binding domain
LBD	Ligand-binding domain
GR	Glucocorticoid receptor
ER	Estrogen receptor
RE, GRE, ERE	Response element; glucocorticoid RE; estrogen RE
PDB	Protein Data Bank
SBMD	Stochastic boundary molecular dynamics
PBC	Periodic boundary conditions
PME	Particle-mesh Ewald
μ	Polarizability
k	Boltzmann's constant
LJ	Lennard-Jones
h	Planck's constant
η	$h/2\pi$

1 INTRODUCTION

In theory, everything is possible

Well - not really - but we are not severely restricted to the real world in our minds, which makes computational chemistry a powerful tool for expressing and testing new ideas and thoughts.

Large assemblies of sequential amino acids are encoded in the genes. This code is duplicated. The peptide chain is synthesized and folded into a compact globule, a protein, with one or more functional groups, *domains.* [1] [2]

Today, we know that polymerases duplicate the information, mRNA carries the code in triplet bases, codons, and that each type of amino acid has its own tRNA, which provides the shape complimentary in anti-codons. Each tRNA successively adds the correct amino acid at the ribosome, where rRNA merges the entire sequence into a polypeptide chain [3]. When biosynthesis is completed, the peptide is released and subsequent folding processes commence in various ways to yield a functional three-dimensional structure. Biophysical techniques [4], mainly X-ray, NMR and EM, have characterized a large number of these structures, and consequently a wide diversity of functional domains has been unraveled. Some of these structures are deposited in the Protein Data Bank, [5, 6] with the atoms represented in Cartesian x, y, z *coordinates*. In a broader sense, though, many issues in the chemistry of life are yet poorly understood.

Unfortunately, there is no universal way of gathering the entire cascade of biochemical mechanisms with one single method. Skilled biochemists or molecular biologists and equally skilled structural biologists contribute with a solid experimental ground. Often, however, there is a gap in experimental resolution, owing to different laborative observations or procedures, and this is where computational modeling enters to play a substantial role in linking, explaining or at least mediating these gaps and hopefully suggest new experiments. Sometimes, experiments may be limited for other reasons (i.e. financially or technically), making computational chemistry an alternative. As for the experimentalist, there are many different and more or less established computational approaches with their own limitations and advantages, in resolution or in accuracy.

Structural biology is the science that characterizes the sequence-structure-function relationship for interacting biomolecules (proteins, peptides, DNA, RNA). A synergy of recombinant DNA techniques, structure determining methods and computer modelling has proven powerful in this context and in many other fields such as food-, agricultural-, and pharmaceutical industry, gene therapy and *de novo* design of proteins. Faster and cheaper computers, new or refined computational methods and sophisticated graphic programs for visualizing and manipulating biomolecular structures has made computer modelling a salient feature within the field of structural biology, which now has a devoted mission in interpreting the results from the enormous and accomplished human genome project (HUGO)[7].

Zinc is important for humans, enzymatic reactions [8-10] and appropriate folding of proteins [11, 12], the latter often represented in zinc finger motifs [13] [14], where different amino acid coordination schemes of cysteines and histidines to zinc are encountered, owing to the closed d^{10} electronic valency of this metal.

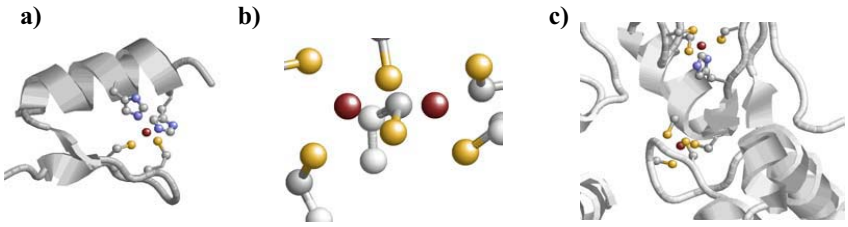

Figure 1. Three commonly found zinc finger motifs: a) classic C2H2 zinc finger, b) di-nuclear zinc coordinated by six cysteines and c) Really Interesting Gene (RING-finger), with C4 coordination and C3H coordination to zinc. The pictures were generated with Rasmol and the structures have pdb accession codes 1tf3.pdb, 3alc.pdb and 1g25.pdb respectively.

These zinc fingers are frequently found in DNA or RNA – binding proteins, which are used for repairing, proofreading, amino acid synthesis and transcription. As implied by the name, the latter class of proteins regulates the transcription machinery either by facilitating and/or preventing the action of polymerases, and is thus termed called *transcription factors* [15]. They are often activated by a cellular signal; e.g. a neurotransmitter, a peptide-hormone, metal or another ligand, to promote gene regulation by recognising and binding to a specific sequence of DNA bases called *response elements*. Naturally, such proteins are in focus of extensive research since they are involved in the primary step of sequence-function relationship between DNA/RNA and proteins.

Molecular dynamics (MD) simulation techniques [16] are based on Newton's motional laws [17], where successive conformations of a molecule are generated in a *trajectory*. A detailed atomic picture of kinetically and time-coupled events can be provided for macromolecules in their natural environment (nicely reviewed in Acc. Chem. Res., 35, 2002). Metals, which are variable in electronic valency, are difficult to treat in MD. Since metals often are incorporated in the protein architecture and used for a diverse set of functions, such as enzymatic, structural build-up and signalling processes, it is of interest to describe metals according to classical laws and, if possible, model the metals to be flexible in ligand coordination geometry.

This thesis is based on protein-metal-DNA-solvent functional interactions in a zinc finger protein from a classical viewpoint and supplies some ingredients needed for modelling, exploring, validating and applying the interaction potential of metals in biomolecules, but it should always be kept in mind that a model is an attempt to describe the reality, not the reality in itself. Complexity does not always mean accuracy, and one has to rationalize the model versus approximations and consistency with experiments. For instance, spending CPU-months in calculating the absolute energy of a hydronium molecule would be somewhat aberrant if the goal is to model the folding of a protein, although both approaches in themselves are important.

The main computational method throughout this study is MD, but higher level of theories, such as *ab initio* quantum mechanics and semi-empirical methods has been visited to provide a basis for modelling zinc to its ligands (cysteins) and energetically resolve observations made from MD simulations.

2 NUCLEAR RECEPTORS

Nuclear receptors (NRs) form the largest known family of eukaryotic transcription regulators [18-24]. These receptors are ligand-inducible and involved in development, growth, proliferation, apotosis, homeostasis, metabolism, and cell differentiation [25]. Various diseases and immune-suppressive issues such as inflammatory, osteoporosis, depression, diabetes, skin diseases and different cancers are related to malfunctioning NRs, thus making them a substantial target for pharmaceutical interests. Depending on the type of activating ligand or phylogenicy, NRs can be sub-categorised into six classes; steroid and thyroid hormone receptors, retinoids, vitamin D_3, and orphan receptors with unknown or absent ligands. In general, the overall topology of the NRs consists of three discrete functional domains [26]; the N-terminal transactivation domain, the DNA-binding domain (DBD) and the C-terminal ligand-binding domain (LBD).

2.1 DNA BINDING DOMAIN

In 1990, Härd and co-workers determined the first NMR solution structure [27, 28] of a DBD within the NR superfamily, namely the DBD of the glucocorticoid receptor (GR), and in 1991 Luisi and co-workers presented the first (GR DBD)2-DNA crystal structure[29]. Ever since, DBDs belonging to different NRs has been solved [30-38] unravelling a high degree of homology and structural similarity for the DBDs in the receptor subfamilies; a globular fold of about 70-80 amino acids arranged in two perpendicularly packed amphipathic helix loop regions with eight out of nine invariant cysteine residues formed in two Cys4-zinc fingers. The first finger directs three residues (the P-box, "P" for protein) for specific recognition of the response element [39, 40]. In GR, these residues are G458, S459 and V462, which recognise the palindromic –AG**AA**CA- DNA response element. Site-specific finger P-box swapping mutagenesis in GR to the corresponding estrogen receptor (ER) residues (i.e. the triple mutant G458E, S459G and V459A) results in an altered specificity toward the ER response element, -AG**GT**CA-[40].

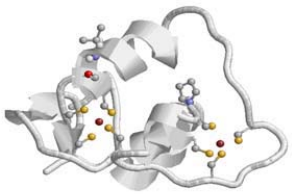

Figure 2. Schematic representation of the free NMR solution structure of GR DBD (pdb accession code 1gdc). The two zinc fingers and residues G458, S459, V462 and P493 are indicated.

The second finger is involved in non-specific phosphate DNA binding and protein-protein contacts formed in the protein dimerisation interface (D-box, "D" for dimerisation). This region is significantly altered in conformation when bound to DNA, such that the free NMR solution structure [27, 28] adopts a conformation perpendicularly oriented relative the crystal structure[29]. However the P493R mutant [41], which is positioned between to the two C-terminal cysteinate ligands in the second zinc finger results in a crystal-like D-box conformation of the free NMR solution structure[37]. This change in conformation is also observed for the S459A mutant [37]. Thus, depending on where a mutation is done in the GR DBD, DNA specificity and/or D-box conformational changes are induced. Luckily, the

DBD appears to remain stable in these mutations and therefore provides a good source for understanding some of the biology behind DBDs interacting with DNA.

3 COMPUTATIONAL CHEMISTRY

Suppose we know everything about the instantaneous interactions (physical, chemical, biological, social and so on) between all particles in the universe. Would that make life boring or extremely interesting?

Well, we cannot, but all these interactions can be condensed into one Law, recognised as the Schrödinger equation.

$$\hat{H}\Psi = E\Psi \qquad [1]$$

In fact, only for two interacting particles an exact (analytical) solution to equation **1** is currently possible, where the quantum mechanical Hamilton operator, $\hat{H}$, act on the wave function Ψ to give the corresponding observable energy in the real world.

In the following sections, some basic concepts and their mathematical interpretations are gathered for an easy review and completeness in the picture of computational chemistry. As mentioned in the introduction, the main method used throughout this thesis is molecular dynamics simulations and naturally, a specific focus on this topic will be done.

3.1 *AB INITIO* QUANTUM MECHANICS

All atoms consist of protons, neutrons and electrons and the two latter forms the nuclei, which is much heavier than the electron. Consequently, the nuclei move much slower than the electrons. Therefore, the nuclei can be approximated being fixed in a field of moving electrons, as stated in the *Born-Oppenheimer* approximation. Thus, the nuclear motion results from the potential energy surface spanned by the surrounding electrons.

While the underlying mathematical concepts of quantum mechanics are rather complex, the main goal is, with no experimental device, to solve equation 1 for electrons belonging to a set of nuclear geometries and therefore the method is called *ab initio* (greek; from the beginning). A trial wave function, Φ , is guessed from the diagonal product of a matrix with N rows of electron coordinates and N columns of electron wave functions (φ), giving a Slater-determinant. This determinant is refined with the variation theorem (finding a minimum) until self-consistency is achieved, subject to the spinorbital ortogonality constraint. This is the Hartree-Fock (HF) method, which forms the basis for subsequent calculations on higher or lower levels of theory. Other methods, such as configuration interactions, perturbation-, coupled cluster and density functional theories are more advanced, but since I am a minor player in this game, I recommend the interested reader to the literature [42].

The basis set is a number of mathematic functions that describe the shape of the electron orbital, and the acronyms of DZ, TZ QZ means that the number of functions is doubled, tripled and quadrupled and so on. Thus larger basis sets provides more space for the electrons, which lowers the energy, and if the number of functions is infinite, the basis set is complete, but this is limited to very small molecules.

The main use of QM in this study is based on density functional theory (DFT). For this purpose, Becke's three parameter functional [43] was used together with the Lee-Yang-Parr local and non-local functional (B3LYP), which to some extent includes electron correlation energy.

Different basis-sets were examined, and for most cases the LANL2DZ was used for geometry optimisations of different zinc-thiolate-water formations. These calculations were prompted by the observed solvent-ligand exchange at zinc in paper **II**. Calculating their zero-point vibrational energy checked the convergence of the optimised structures, and larger basis set was used for subsequent single point energy calculations. These basis-sets covered 6-311G up to 6-311++G(2d,2p), and solvent effects were included using the conductive polarisable continuum method (CPCM). All calculations were performed with the Gaussian98[44] program suite. Typically, about 50 – 70 atoms could be handled with small to moderate basis-sets at the B3LYP level of theory.

3.2 SEMI-EMPIRICAL METHODS

Most reactions in chemistry occur in the outer shell of the electron cloud, and this is the philosophy behind semi-empirical methods. Only valence electrons are included explicitly, while core electrons are parameterised against experimental data. The MOPAC[45] program package and the molecular neglect of differential overlap (MNDO)[46-48] were used for calculations on a zinc finger modelled as a tetra-methiolate-zinc complex (paper **I**). A comparison in geometries obtained with other semi-empirical methods and one QM calculation at the B3LYP/6-31G** level of theory was done. When using semi-empirical methods, the number of atoms that may be treated is in the range of ~500 (which is quite close to smaller proteins).

3.3 MOLECULAR MECHANICS

Merging the nucleus and its electrons into a single particle (i.e. an atom) reduces the complexity significantly and enables energy calculations to be performed on large macromolecules which may contain up to 10^5 atoms. Instead of treating the electrons explicitly, pairs of atoms are held together with springs to mimic a chemical bond. These springs are Hookean with a force constant k, and vibrate harmonically around a reference value to give the *bonded* energy. The Coulomb energy is usually modelled with point charges, which results in an electrostatic potential located at some specific point relative the atom, the *nonbonded electrostatic* energy, while dispersion and electron exchange energies are modelled with an attractive-repulsive term giving the *nonbonded nonpolar* or so-called *van der Waals* energy. Thus, the chemical bond between pairs of atoms seen from a classical view consists of three terms: bonded, non-bonded electrostatic and van der Waals energies.

The basal need for performing a molecular mechanics energy calculation is a library that defines the connectivity between the atoms (i.e. the topology), a source of force field parameters to feed into an empirical energy function and a description of how the atoms are arranged spatially. For the latter, PDB [5, 6] coordinates is a good source, if the subject is a biomolecule. If no coordinates are available, then modelling from scratch is the only option. In a sense one can compare this procedure with the ribosome amino acid chain synthesis (the library) and the folded structure (the coordinates). What goes in between (i.e. the folding process) is subject to intense

research around the world, but I will leave that to the future and to the experimental community.

Once the coordinates and the energy function have been defined, the first or second partial derivatives of the energy function with respect to the atomic coordinates yield the forces that act on each atom, the *force field*. These forces can then be used for locating the most favourable positions for the atoms, using minimisation methods, where $\partial V(\mathbf{r})/\partial \mathbf{r} = 0$ and $\partial^2 V(\mathbf{r})/\partial \mathbf{r}^2 > 0$ for a converged minimum. This is the overall concept of *molecular mechanics*.
Since the CHARMM[49] program was used throughout these studies, the notations and expressions will be based on CHARMM[49] jargon, but for most cases the theory applies to other similar programs and force fields that are used for MM and MD calculations.

3.3.1 The energy function

The empirical energy function for the interactions between atom-pairs in a molecule with N atoms defined in Cartesian coordinates (vector **r**) has the general form:

$$V(\mathbf{r}^N) = \sum_{\text{bonds}} \frac{k_b}{2}(b - b_0)^2 + \sum_{\text{angles}} \frac{k_\theta}{2}(\theta - \theta_0)^2 + \sum_{\text{impropers}} \frac{k_\omega}{2}(\omega - \omega_{i0})^2 + \sum_{\text{dihedrals}} \frac{k_\varphi}{2}[1 + \cos(n\varphi - \delta)] + \sum_{i,j} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{332 \cdot q_i q_j}{\varepsilon r_{ij}} \qquad [2]$$

3.3.2 Bonded terms

Atoms are bonded in pairs, connected with a spring having a force constant "k_x". The harmonic nature of terms 1 – 3 in equation **2** results from the truncation at second order terms in a Taylor expansion series around the natural value with subscript zero. Higher order polynomials can be included in the energy function at the price of increased complexity, but quadratic terms have been demonstrated to reproduce molecular geometries fairly good. The improper, or out-of-plane angle bending, was initially assigned for maintaining the chirality of the molecule for so-called united atom models, which treated the hydrogen atoms implicitly by merging them into the heavy atoms. Today, most force fields apply an all-atom description, and the improper term may therefore be redundant. For torsion angles, corresponding to the fourth term in the energy function, multiple energy minimums can be visited (e.g. gauche and trans), and therefore the form of this term is sinusoidal, having the multiplicity n and phase angle of δ degrees.

Bonded terms are sometimes misquoted as covalent bonds. They are solely bonded and cannot be broken! The penalty in energy is proportional to the square displacement from ideal geometry. This means that when two explicitly bonded atoms are squeezed together or pulled apart, the resulting energy parabolic curve will go to infinite maximum. One way to circumvent this shortcoming is to model the chemical bond according to the Morse potential:

$$V(\mathbf{r}) = D_e\left\{1-\exp\left[-\left(\sqrt{k/2D_e}\right)(r-r_0)\right]\right\}^2 \qquad [3]$$

Which results in a curvature towards zero when the atoms move apart. D_e is the potential well depth (i.e. the dissociation energy) and k the force constant. Despite the more realistic picture obtained for chemical bonds when using Morse potential, this approach is less applied in molecular mechanics, owing to its increased complexity and hence computationally demanding formulation.

3.3.3 Non-bonded terms

The two last terms in equation 2 represents the sum over all non-bonded pairs of atoms for the van der Waals repulsive (12) and the attractive dispersion term (6) and the electrostatic Coulomb interaction energies, where the two former terms are called Lennard-Jones 6,12 interactions. The collision diameter between two atoms is $\sigma_{ij}=(\sigma_i+\sigma_j)/2$ and $\varepsilon_{ij}=(\varepsilon_i\varepsilon_j)^{1/2}$ is the potential well depth. The 6,12 Lennard-Jones term is widely used because of its computational efficiency. The atoms are separated by the distance r_{ij}, carrying charges q_i and q_j, and embedded in a dielectric medium, ε (usually, $\varepsilon=1$). $R_{min} = 2^{1/6}\sigma$ and $r_{min,\,ij}=r_{min,\,i}+ r_{min,\,j.}$ Electrostatic and van der Waals interactions are omitted for atoms that are connected with one or two bonds. For different molecules only nonbonded interactions (i.e. intermolecular) are present.

3.4 FORCE FIELD PARAMETERIZATION

Defining the parameters for the energy function is in many cases a state-of-the-art procedure, which is based on experimental data, *ab initio* calculations, frustration and sometimes luck. Parameterisations are usually initiated on small molecules, where the force constants can be derived from spectroscopic data and geometries from highly resolved crystals of small compounds.

3.4.1 Bonded terms – reference geometries and force constants

One advantage of using *ab initio* methods for determining force constants and geometries is the cancellation of experimental biases. One source or molecule – one set of parameters for fitting in the empirical energy function, which in turn may be modified to agree with the quantum mechanical calculation. Given a small molecule, the geometry is fully optimised to rest in its lowest energy conformation. This is the reference state for the energy, bonds, angles etc, etc. When the geometry is manipulated without subsequent or only partial optimisation in geometry, the second energy derivative yield the force constants and vibrations of the molecule. These numbers are fitted empirically to agree with the *ab initio* potential energy surface. In such a way, bonds, angles etc, etc, are varied systematically for a set of small compounds, which then are combined to larger molecules. At this point comparisons with experiments provide a good validation of the parameterisation procedure. For instance, would mixing of parameters obtained from *ab initio* with different basis-sets and/or methods influence the performance of the force field?

3.4.2 Non-bonded terms – electrostatic charges and Lennard-Jones parameters

Various schemes for calculating the charges have been suggested over the years [50-53] since the quantity of charges cannot be devised experimentally. Usually, the wave function is used for calculating the polar moment (of order 2 or higher) and the charge distribution around the nuclei. In the subsequent force field fitting the charge may be centered at the nucleus or distributed around the nucleus and/or be assigned to reproduce the polarity moment only, ignoring the resulting numbers of the charges from the QM calculation. The latter procedure will only be valid if the net charge of the molecule is zero.
An initial guess for the ω term is normally taken from the periodic system, or may derived from crystal packing data. Given a reference molecule, e.g. water (model), the solvent radial distribution around the atom corresponds to ω and the hydration free energies (for κ) are calculated in an iterative way to be consistent with experimental data (if available). Such calculations are mainly carried out with Monte Carlo or molecular dynamics simulations, and uses statistical thermodynamics for the parameterisation of κ (corresponding to the 6:th term in equation **2**), where the relative differences in hydration free energies, ÷÷G, are determined.

3.5 METALS IN FORCE FIELDS

Metals are not encoded in the genes. Their variance in electronic structure enables diverse functions in the biomolecules. This flexibility in electronic valence makes metals, and in particular those with higher valence and transition row metals, difficult to describe in standard force fields without sacrificing some other property.

In the beginning of the simulation era, metals were often left out in the calculations and the biomolecule was treated as being in the apo-state. Not very exciting, although a calculation in the absence of a metal may provide information of local folding/unfolding processes.

3.5.1 Bonded models

A bonded [48, 54-60] representation of the metal interaction benefits the preservation of local geometry around the metal, but will also restrict the ligand coordination flexibility, which may strain the biomolecule and eventually lead to a propagation of atomic displacements.

3.5.2 Non-bonded models

The use of a non-bonded model allows for conformational relaxation and ligand coordination flexibility, but may result in anomalous geometries. Another difficulty is to treat the non-bonded interactions appropriately, such that solvent-ligand-zinc interactions are balanced and fairly reasonable in magnitude. Åqvist and coworkers[61] devised an elegant model to mimic the valence electrons of ionic zinc by placing six symmetrical charges around the nucleus, which carries a charge of –1, while the overall charge on the zinc ion is +2. Later, very similar models for zinc appeared in the literature[62].

In paper **I**, a non-bonded model based on fractional charges centered on the nucleus was constructed for the thiolate-zinc formation. In this model, the charges were obtained from a semi-empirical MNDO[48]/ESP[63] calculation, followed by fitting of Lennard-Jones parameters and a charge-constraining scheme. This model performed well in MD simulations, but needed additional support for solvent interactions. Therefore, a formal +2 charge zinc model from Stote [64] was reconsidered. This model is parameterised in accordance with CHARMM parameters and claimed to be "simple but accurate"[64], a statement, which in my eyes, has become less convincing by the time.

However, using Stote's zinc model together with standard CHARMM [65] parameters for the cysteinate ligands, turned out to be pretty much like fitting an elephant into a bikini, and therefore it was necessary to modify the interaction potential of the thiolate ligands to zinc for maintaining the tetrahedral coordination geometry.

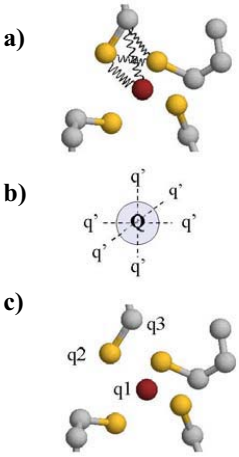

Figure 3. Schematic drawing of different approaches for modelling metals in proteins. The ligands (cysteins) are yellow and the metal (zn) is in red.
a) a bonded formalism.
b) Åqvist's fractional charge model.
c) a non-bonded charge-centered model.

In most cases, one or more properties has to be sacrificed when modelling metals in classical ways, but for a successful parameterisation, one has to use at least some chemical intuition, stick with reasonable interactions and targeting acceptable geometries together with some patience. A centered formal charge of +2 at zinc is only true for an isolated ion, whereas for interacting particles, fractional charges often results of numbers close to one, irrespective of the valence state of the metal. The present model for the cys_4 zinc finger assumes that Stote's zinc model is appropriate in the sense of solute-solvent interactions seen from a thermodynamic perspective.

Papers **I** and **II** describes a comparative study of how the force field and protein dynamics are influenced when the thiolate-zinc interaction potential is altered for a cys_4 - zinc finger and the perturbation in coordination ligand geometry associated with conformational changes in the biomolecule respectively.

3.5.3 Mixed potentials

A way to model the metal, perhaps in a somewhat more realistic fashion, is to partition the metal-ligand and rest of the biomolecule in a QM and a MM part respectively.

The generic Hamiltonian is then

$$\hat{H} = \hat{H}_{QM} + \hat{H}_{QM/MM} + \hat{H}_{MM} \qquad [4]$$

Where the linking part between the molecular mechanics and the quantum region is defined with the QM/MM Hamiltonian. A drawback with this method is to define the QM/MM boundary, and for the less experienced user it might be safer to preserve the intactness of the force field. Nonetheless, accurate calculations of these formulations would certainly provide important information in developing novel metal parameters for classical descriptions.

3.6 MOLECULAR DYNAMICS

3.6.1 In a nutshell

Take a coordinate system of N atoms, assign these atoms initial velocities or equally temperature, T, of known magnitude ($3NkT=\Sigma m_i v_i^2$) and use finite time steps to calculate the resulting forces on each atom at time t + δt.

According to Newton's second law[17] (**F**=m**a**), and the ordinary second order differential equation

$$\frac{d^2\mathbf{r}}{dt^2} = -\nabla\frac{V(\mathbf{r})}{m} = \frac{\mathbf{F}}{m} = \mathbf{a} \qquad [5]$$

which can be solved numerically *and* simultaneously for each atom to obtain their new velocities and positions if the initial potential energy is known is known.

If the potential is continuos, then the integration of **r,** can be approximated with a Taylor series expansion

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \frac{d\mathbf{r}(t)}{dt}\delta t + \frac{1}{2}\frac{d^2\mathbf{r}(t)}{dt^2}\delta t^2 + \ldots + \frac{1}{n!}\frac{d^n\mathbf{r}(t)}{dt^n}\delta t^n \qquad [6]$$

Which gives the finite difference approximation for numerical integration of Newton's motional laws. Various algorithms [16, 66] have evolved over the years, such as the leapfrog [67], the predictor-corrector [68] and the velocity Verlet [69] algorithms and most of them originate from the Verlet [70] algorithm:

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t+\delta t) + \frac{d^2\mathbf{r}}{dt^2}\delta t^2 \qquad [7]$$

The acceleration is recognised as the third right-hand term and no velocities are included explicitly (one reason for the birth of similar algorithms).

3.6.2 Ensembles, stochastic simulations

Molecular dynamics is a statistical method that samples low energy regions (i.e. regions with barrier heights less than the kinetic energy; $T \geq E\text{-}V$). Irrespective of initial coordinates in the simulation, the trajectory is assumed to converge towards a "true" value for a specific property (i.e. internal energy or conformation of the molecule). The ensemble average is the corresponding property in the real world, a cornerstone in statistical thermodynamics, known as the ergodic hypothesis.

The choice of ensemble determines which thermodynamic property (normally Boltzmann distributed) that is associated with the macroscopic quantity. For a micro-canonical ensemble with N particles comprising a volume V and the conservation of energy we have $(NVE)_{constant}$ and fluctuations in temperature. This ensemble will give the maximum entropy, S_{max} at equilibrium. In a similar way, the canonical $(NVT)_{constant}$, isothermal-isobaric $(NPT)_{constant}$ and grand canonical $(\mu VT)_{constant}$ ensembles in equilibrium corresponds to the minimum Helmholtz free energy, A_{min}, Gibbs minimum free energy, G_{min} and maximum of the PV-term respectively. Simulations conducted in the NVT ensemble are coupled to an external temperature source, a heat bath. This heat is associated with a temperature "mass", a piston that absorb or release kinetic energy to maintain the target temperature. In a similar way, NPT simulations are performed with a pressure piston (in atomic mass units). Other approaches in this aspect involve coupling parameters of unit ps or reciprocal pressure atm^{-1}.
The grand canonical ensemble, with constant chemical potential, μ, is a challenge within MD simulation techniques, since the number of particles is variable. In MD simulations, difficulties arise because of the particle creation/annihilation end-point-sampling catastrophe (i.e. the ergodic breakdown).

Most simulations reported in this thesis were performed in the NVT-ensemble and using stochastic boundary molecular dynamics simulations (SBMD)[71] accordingly to randomly fluctuating forces (with average zero) that adds energy to the system and a frictional term (γ) that removes excess energy. This energy is transferred across a boundary zone, which in this case is the outermost solvent layer ~3 Å in thickness. Solvent molecules that cross the boundary are treated as Langevin particles, whereas all other atoms being in the inner region are treated as classical particles. To maintain the spherical shape of the droplet, solvent molecules "see" a mean field interaction potential of fictitious water beyond the sphere edge.[72]

3.6.3 Increasing the time-step

The magnitude of the time-step will be dictated by the highest bond frequencies, $\delta t << \nu\ s^{-1}$, which correspond to the motions of light atoms (i.e. hydrogen atoms with frequencies ~ 10 femtoseconds). Therefore it is of interest to cover the phase space by maximising the time-step without violating the ergodicity of the system, that is, in principle, when atomic motions are ill-behaved due to numerical errors from the integration step size, δt.

Fast motions of these light atoms can be filtered out with constraining schemes on the internal geometries involving explicit bonds between heavy atoms and hydrogen

atoms [73], and permit larger time-steps, usually 1 – 2 femtoseconds. Recent methods employ multiple time-step algorithms [74] and multi-bond schemes [75], leading to larger portions of sampling without sacrificing the integration accuracy. Hence, progress in this field is at hand, and possibly MD simulations are reaching time-scales of seconds or even minutes in the future. Currently, simulations are normally in the range of a few up to hundred nanoseconds, depending on the system size.

3.6.4 Long-range electrostatic interactions and boundary conditions

Non-bonded electrostatic interaction is the most CPU-hungry issue and scales as $1/\mathbf{r}$. Truncation between atom-pairs can be done by using truncation schemes, i.e. force-shifting or switched scaling-factor for the non-bonded interactions and forces [76]. This means that atoms outside a specific distance, $\mathbf{r}_{cut}$, are omitted in the non-bonded evaluation. The cut-off is either group-based or atom-based, where the former means that the all atoms within the same group is included in the evaluation of electrostatic energies and forces whenever it has an atom within $\mathbf{r}_{cut}$, whereas the latter only concerns individual atom-pairs where $\mathbf{r}_{ij} \leq \mathbf{r}_{cut}$. However, atoms outside $\mathbf{r}_{cut}$ may still be accounted for in a neighbour-list, which is regenerated either by keeping track on the distance the atom has moved or with a pre-defined step frequency. This list usually expands two Angstroms beyond $\mathbf{r}_{cut}$. Such approaches liberate CPU-demanding calculations, but may in cases where $\mathbf{r}_{cut}$ is too small, or the step size for updating the pair-list is too long, lead to numerical instabilities.

Advances in appropriate modelling of long-range interactions involves multipole expansions [77] (which scales as $1/^{(n+m-1)}$ for molecules having polarities m and n respectively). Ewald [78] summation techniques [79, 80], which includes all atoms for handling long-range electrostatic interactions, and progress in this field involve the multipole-Ewald combination [81].

Another source of increased CPU-times is the choice of boundary conditions [16]. Periodic boundary conditions means that a system (biomolecule, solvent, ions etc, etc) is surrounded by mirror images, replicas. Each replica is surrounded by identical systems (e.g. biomolecule in a box of solvent), subject to the constraint that the symmetry of the system is applicable in the x, y and z-direction, giving a pseudo-infinite array of molecules surrounding the central box. When the biomolecule moves outside the boundary of the box (or a similar geometry), it will simultaneously enter at the other side of the box, in a similar spirit as for those TV-games that appeared in the early 70's. Compared to non-periodic boundaries (e.g. a biomolecule in a sphere droplet of solvent) the gain in CPU-time is ~ 2 - 3 times faster for non-PBC than PBC. Papers **II** and **III** compare different ensembles, boundary conditions and CPU-times from a MD viewpoint.

3.6.5 Explicit treatment of solvent

A key feature for performing a realistic MD simulation is to model the environment properly by means of solvent (e.g. water). Including solvent explicitly is probably the most appropriate way of describing the solvent effects since both bulk properties and direct interactions, i.e. solute-solvent hydrogen bonds; charge

screening and structurally or chemically important effects seen from the biomolecule are treated in the simulation, which is carried out in the condensed phase. However, explicit treatment of solvent requires an adequately parameterised water model and in most cases such model fails in describing all desired solvent-properties simultaneously. Despite its rather simple form, three-centered water models are commonly used in the simulation community, where the mostly used models are the transferable interaction potential three-point (TIP3P)[82, 83], the simple point charge (SPC)[84, 85] and simple point extended (SPC/E)[86] charge models. The latter is an updated version of SPC. These models differ in Lennard-Jones parameters, charges and internal geometries, and given the available computational resources by the time they were parameterised, they all provide a fairly realistic environment for condensed phase simulations. In paper **III** a comparison between these three water models was done. It was found that experimentally related observations were mildly dependent on the choice of model and that a main source of differences was the rotational diffusion of the biomolecule. This difference was accentuated by the choice of boundary conditions, such that the use of stochastic simulations, a frictional constant on the water oxygen atoms and non-PBC moderated the rotation of the solute.

3.6.6 Implicit treatment of solvent

Because the use of explicit solvent requires heavier computations, people have developed methods for modelling the environment implicitly, such that the solvent perturbs the gas-phase behaviour of the system. These approaches are called continuum solvent models (CSM), and significantly reduce the calculation, since no water molecules are explicitly involved in the simulation. The main basis of CSM is derived on thermodynamic ground, where the free energy of solvation, ΔG, can be partitioned into electrostatic, van der Waals and cavitation free energies giving

$$\Delta G_{solv} = \Delta G_{elec} + \Delta G_{vdW} + \Delta G_{cav} \qquad \textbf{[8]}$$

In 1920, Born derived an expression for the work done for transferring an ion with a radius a and charge q from vacuum to a medium with dielectric constant, ε, for the electrostatic contribution to the free energy of solvation [87]:

$$\Delta G_{elec} = -\frac{q^2}{2a}\left(1-\frac{1}{\varepsilon}\right) \qquad \textbf{[9]}$$

This formulation provides a basis for the so-called generalised Born models for pair-wise interacting particles with radius a:

$$\Delta G_{elec} = -\frac{1}{2}\left(1-\frac{1}{\varepsilon}\right)\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{f(r_{ij}, a_{ij})} \qquad \textbf{[10]}$$

Still and co-workers devised the formulation; $f(r_{ij}, a_{ij}) = \sqrt{(r_{ij}^2 + a_{ij}^2 \exp[D])}$; $a_{ij} = \sqrt{a_i a_j}$; $D = r_{ij}^2/(2a_{ij})^2$ which is incorporated in both MM and in QM methods[88].

In the same spirit, the generalised Born molecular volume (GBMV) approach was developed [89]. This method was used in paper **IV**.

Born's original equation is only appropriate to molecules with a formal charge. In 1936, Onsager modified Born's equation to a dipole representation, the so-called reaction field model [90].

$$\Phi = \frac{2(\varepsilon - 1)}{(2\varepsilon + 1)a^3}\mu \qquad \textbf{[11]}$$

where μ is the dipole moment and the dipole-energy within an electric field is $-\Phi\mu$.
If the dipole is polarisable, an additional term is added which gives the electrostatic contribution of the free energy as

$$\Delta G_{elec} = -\frac{1}{2}\Phi\mu \qquad \textbf{[12]}$$

In QM, the philosophy of reaction field is mostly used in conjunction with some correction terms or variant mathematical formulations of the Onsager model, i.e. the conductive like screening model (COSMO, CPCM)[91].

A third possibility is to use the Poisson-Boltzmann (PB) equation for modelling the surrounding solvent.

$$\nabla\varepsilon(\mathbf{r})\nabla\Phi(\mathbf{r}) - \kappa'\Phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \qquad \textbf{[13]}$$

This is the linearised form of the PB equation, where Φ is the electrostatic potential in a dielectric medium, ε; κ is related to the salt concentration and ρ is the charge density and for a molecule with N charges and the electrostatic contribution to the solvation free energy is

$$\Delta G_{elec} = \frac{1}{2}\sum_i q_i(\Phi_i wtr - \Phi_i vac) \qquad \textbf{[14]}$$

The non-electrostatic contributions to the free energy (i.e. ΔG_{vdW} and ΔG_{cav}) are often described with a solvent accessible area (SA) correction and some additional constant:

$$\Delta G_{vdW} + \Delta G_{cav} = \gamma A + \beta \qquad \textbf{[15]}$$

In summary, QM uses reaction field techniques based on Onsager's dipole model [90], whereas classical methods mainly rely on methods based on Born's charge-cavity model [87], or PB approaches.

4 APPLICATIONS OF MOLECULAR DYNAMICS

4.1 RUNNING MD

Before performing a MD simulation, one has to consider whether the subject is possible to determine with the available method. It should be kept in mind that one single simulation describes the event of a limited set of atoms, whereas wet-chemical experiments reflect the average of a large ensemble of interacting molecules. Therefore, it is useful to conduct a series of simulations, starting from different initial coordinates or by assigning the molecule randomly selected initial velocities and with different system setups (i.e. PBC, NVT, NPT and so on). If repeated events are seen, then it is likely that the results from a simulation are unambiguous. Unfortunately, this does not always mean that the observed event happens in the real world. In this aspect, comparisons with and interpretation of experiments are invaluable. However, if the simulations show consistency with experiments, a solid ground for establishing new approaches, insights and predictions are feasible. It may also be useful to perform at least one costly simulation of substantial length and use this trajectory as a benchmark for rationalising which approximations that could be done for faster computations.
The general scheme for conducting a MD simulation involve hydration of the biomolecule, adding prosthetic groups (e.g. sodium or chloride ions for charge neutrality), minimisation, heating and equilibration of the entire system. The minimisation step may be replaced by adding soft harmonic constraints on the solute atoms, scale these constraints in subsequent windows as the system is heated, until the constraint forces are zero. If PBC is used, the equilibrium portion of the simulation can be done in the NPT ensemble, which relaxes the volume, followed by NVT conditions when the production part commences.

4.2 ANALYSIS

The decision of where the production part of the simulation begins is somewhat illusive, but as a rule of thumb the target temperature should be stable (and consequently the total energy of the system). It is also useful to use graphics and manually inspect extracted snapshots from the simulation, and not only simulate blindly.

We can calculate the time evolution of a specific property, p (t), or the time-average of this property. The latter denoted as <p>, and the average is calculated over blocks of the trajectory, where the statistical uncertainty for M points scales as $1/\sqrt{M}$.

4.2.1 Dynamics and experimental connectivity

By monitoring the root mean square deviation (rmsd) of a property (**p**) evolving in time relative a reference, well-behaved MD should result in a curve that levels off and fluctuates around an average value.

$$\mathrm{rmsd}(t) = \sum_i \sqrt{\frac{1}{N}\left(\mathbf{p_i}(t) - \mathbf{p_i}(t_{ref})\right)^2} \qquad \textbf{[16]}$$

If the property involves Cartesian displacement, translational and rotational motions are removed by superposition of the instantaneous frame onto a reference set.

However, if the interest is to monitor conformational sampling, superposition of multiple coordinates, or different parts of the molecule, can be used for reference. In such cases, the curvature may look different, picturing the time-dependent molecular flexibility and conformational accessibility.

A direct route to experimental X-ray B-factors is given for the mean square displacement of the *i*th atom.

$$8\pi^2\langle \Delta \mathbf{r}_i^2 \rangle = 3B_i \qquad \textbf{[17]}$$

The B-factor indicates the thermal vibration of an atom in the crystal structure. Disagreement between simulated and experimental B-factors could result from a non-equilibrated trajectory, crystal packing defects or differences between the crystal and simulated environments. Normally, however, consistency with experimental B-factors indicates a well-behaved simulation.

To access the structural properties of a protein that has been subjected to MD, so-called Ramachandran plots is useful to inspect, since Φ and Ψ dihedral angles in the backbone atoms are restricted to adopt certain values within a specific range in defined secondary elements. This analysis was done in paper **III**.

NMR characterises various biomolecular properties in terms of spectral density functions, J(ω), which These can be calculated from a MD trajectory and provide a validation of the force field. One such function is the 'model free' Lipari-Szabo formalism [92, 93], which enables the generalised order parameter (S^2), i.e. the amplitude of intra-molecular motion, to be calculated. Basically, the relaxation from an external magnetic field perturbation of the ^{15}N-^{1}H vectors is monitored.

S^2 varies in the range 0 to 1, where the latter value means that the amide-hydrogen bond vector is totally restricted in motion.

The corresponding generalised S^2 order parameter can be calculated from a MD trajectory[94]

$$C_{2i}(t) = < P_2(\mathbf{h}_i(\tau + t) \cdot \mathbf{h}_i(\tau) > \qquad \textbf{[18]}$$

P_2 is the second order Legendre polynomial ($1.5x^2$-0.5) and $\mathbf{h}_i$ is a unit vector (in this case along the N-H bond) in the molecules fixed frame. On 'infinite' simulation length we have for residue "i"

$$S_i^2 = \lim_{t \to \infty} C_{2i}(t) \qquad \textbf{[19]}$$

The effective internal correlation time, τ_e, for each residue can be evaluated as

$$\tau_{e,i} = \frac{1}{(1-S_i^2)} \int_{t=0}^{\infty} [C_{2i}(t) - S_i^2]dt \qquad \textbf{[20]}$$

However, the simulation finite in length, and in most cases, the auto-correlation curve of the vector converges, which gives the S^2 value.

In paper **III**, an estimate of the overall re-orientation time was made, based on three different vectors within the protein frame. These correlation times were calculated as

$$C(t_m) = \left\langle P_2\left[\boldsymbol{\mu}(0)\cdot\boldsymbol{\mu}(t_m)\right]\right\rangle \approx \frac{1}{N-m}\sum_{n=1}^{N-m} P_2\left[\boldsymbol{\mu}(t_n)\cdot\boldsymbol{\mu}(t_n+t_m)\right] \qquad \textbf{[21]}$$

N is the total number of time steps in the simulation and m is the number of time steps passed at time t_m. The re-orientation time is then estimated from the inverse slope of $\ln(Ct_m)$ versus t_m.

Experimental correlation and generalised S^2 order parameters have been determined for GR DBD and ER DBD [95, 96], and therefore a direct evaluation of the force field and the simulation technique is feasible.

Another important property that can be calculated is the translational diffusion or self-diffusion coefficient (D), which is available trough the Einstein relation:

$$\lim_{t\to\infty}\frac{d}{dt}\left\langle\left|\mathbf{r}(t+\tau)-\mathbf{r}(t)\right|^2\right\rangle = 6D \qquad \textbf{[22]}$$

Where **r** is the vector for a molecule or an atom starting at time t. Plotting the mean square displacement (i.e. $<\Delta\mathbf{r}^2>$) versus time (τ) gives 6D as the slope.

Solvent accessible surfaces areas (ASA) can be obtained by probing the contact surface with a hypothetical molecule[97], typically water, with a radius of 1.4 Å. For hydrogen bonds, a distance/angle criterion between the acceptor-hydrogen is usually ≤ 2.4 Å and the donor-hydrogen-acceptor angle ~ $\geq 130°$.

Atomic co-variances, C_{ij}, can provide useful information about the correlation and function of the biomolecule, as well as the force field parameterisation effects (paper **I**).

The general formula for such estimates is

$$C_{ij} = \frac{< k_i l_j > - < k_i >< l_j >}{[(< k_i^2 > - < k_i >^2)(< l_j^2 > - < l_j >^2)]^{1/2}} \qquad \textbf{[23]}$$

Where the denominator ensures normalisation, k and l are some arbitrary properties for residues "i" and "j". In cases where k=l and i=j the auto-correlation coefficient is obtained, otherwise C_{ij} is the (normalised) cross-correlation coefficient.

4.3 CALCULATING FREE ENERGIES

Quantitative energies from a MD simulation are often difficult to obtain, owing to some extent of the force field but in particular to the reduction in phase space sampling. However, there are some approaches that permit relative free energy differences, $\Delta\Delta G$, to be calculated within chemical accuracy (i.e. ~ 1kcal/mol). Since ΔG is a state-function, a closed thermodynamic cycle can be used, which means that molecule A can be transformed to molecule B in different environments, and $\Delta\Delta G$ would be the energy difference between those two calculations. The transformation in itself may be non-physical (alchemical) since only the final states are considered.

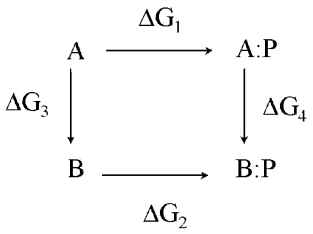

Figure 4. Thermodynamic perturbation cycle, for the transformation of A to B in two different environment, e.g. solvent and protein $\Delta G_2-\Delta G_1=\Delta G_4-\Delta G_3=\Delta\Delta G$.

4.3.1 Thermodynamic perturbation

The energy, E, in the alchemical transformation from A to B can be described as a function of a coupling parameter, λ, such that $E(\lambda) = (1-\lambda)E_0 + \lambda E_1$. E0 is the initial state potential energy (i.e. when $\lambda=0$) and E_1 is the potential energy of the final state. The transformation path from A to B is done in small steps, which gives the thermodynamic corresponding ensemble (ΔA if the calculations were performed under NVT conditions) free energy as:

$$\Delta G = -RT \sum_{\lambda=0}^{1} \ln\left\langle \exp\left(\frac{\Delta E(\lambda)}{RT} \right) \right\rangle \qquad \textbf{[24]}$$

This is the so-called *thermodynamic perturbation* formula [98], where energy-differences are averaged over finite differences in energy functions. If the energy function instead is differentiated, we have the *thermodynamic integration* method:

$$\Delta G = \int_0^1 \left\langle \frac{\partial E(\lambda)}{\partial \lambda} \right\rangle d\lambda \approx \sum_i \left\langle \frac{\partial E(\lambda)}{\partial \lambda} \right\rangle \Delta\lambda_i \qquad \textbf{[25]}$$

Usually, the λ step-size should result in energy differences less than kT between two adjacent windows for reasonable results.

4.3.2 Potential of mean force

When the reaction path only involves changes in geometry (i.e. a distance or an angle) the free energy surface can be obtained as a potential of mean force (PMF). This means that a biasing potential, often harmonic $W(\mathbf{r})=k_w(\mathbf{r}-\mathbf{r}_0)^2$, is applied on the atoms of interest and conformational sampling of is confined to satisfy this biasing potential [99]. The method is called umbrella sampling, and has the general formula

$$V'(\mathbf{r}^N) = V(\mathbf{r}^N) + W(\mathbf{r}^N) \qquad \textbf{[26]}$$

Where $V'(\mathbf{r}^N)$ is the resulting potential for the atoms being perturbed from $\mathbf{r}_0$, and this will result in a non-Boltzmann distribution of energies.

The key issue is to extract the probability based Boltzmann averages and a method for doing this was introduced by Torrie and Valleau [100].

$$\langle A \rangle = \frac{\langle A(\mathbf{r}^N)P(\mathbf{r}^N)\rangle_W}{\langle P(\mathbf{r}^N)\rangle_W} \quad ; \quad P(\mathbf{r}^N) = \exp\left(\frac{W(\mathbf{r}^N)}{kT}\right) \qquad \textbf{[27]}$$

In paper **IV** a PMF calculation was done on the Ψ dihedral angle of one cysteinate zinc ligand in the second zinc finger. The PMF was constructed with the weighted histogram analysis method.[101]

4.3.3 Linear interaction energy method

Estimation of absolute binding free energies, ΔG, within the framework of classical dynamics is difficult, but the linear response (also termed linear interaction energy) method can provide a route to access such energies[102]. This method was initially devised by Åqvist and co-workers and is based on electrostatic and van der Waals scaling of the difference in non-bonded energy between a ligand free in solvent and the same ligand bound to a biomolecule (i.e. a protein). Thus, two MD simulations are conducted and four energy averages are calculated.

$$\Delta G_{bind} = \beta(\langle E\rangle_{lig/prot} - \langle E\rangle_{lig/sol})^{elec} + \alpha(\langle E\rangle_{lig/prot} - \langle E\rangle_{lig/sol})^{vdW} \qquad \textbf{[28]}$$

The main issue is to determine α and β and possibly add another correction term γ [103]. All these terms are initially calibrated against experimental data, although the β-term should be exactly 0.5 for a linear response in electrostatic interactions upon charging a neutral atom in solvent[104]. In current MD simulations of protein-DNA complexes, this method is applied for estimating the binding energy of DBD to RE.

4.3.4 Generalized Born molecular volume

The free energy of solvation, including the solvent reorganisation entropy and enthalpy, can also be estimated on a continuum basis, which may serve as a guide for energy ranking of compounds and targeting more elaborate calculations. The main

advantage of such approaches is that they are fast and not delimited to the size or shape of the surrounding solvent and

$$G^{tot} = G^{nonpolar} + G^{solv} + G^{conf} \qquad \textbf{[29]}$$

$$G^{nonpolar} \approx ASA \cdot \gamma + \beta$$

$$G^{solv} = H^{solv} - TS^{solv} \approx G^{elec} = G^{GBMV}$$

$$G^{conf} = E^{force\ field} - TS^{conf}$$

ASA is the solvent accessible surface area (Å^2), γ = 0.00592 and β=0.92 have been derived from solvation free energies of hydrocarbons[105].
This approach was used in paper **IV**.

4.3.5 Entropies

While solvent re-organisation entropies are included in the GBMV[89] term, the configurational entropy of the solute can be obtained from the generalised S^2 order parameters [106, 107]

$$S_{conf,i} = A + k\ln\left[\pi\left(3 - \sqrt{1 + 8S_i}\right)\right] \qquad \textbf{[30]}$$

and the sum running over all residues gives the configurational entropy of the solute. A is a constant and $S_i = \sqrt{S_i^2}$.

Another robust method for estimating the configurational entropy is quasi-harmonic analysis [108] based on the diagonalization of the mass-weighted co-variance matrix of vector displacement for solute atoms having 3n-6 degrees of freedom:

$$\sigma_{ij} = \frac{\left\langle \left(\mathbf{r}_i - \langle \mathbf{r}_i \rangle\right)\left(\mathbf{r}_j - \langle \mathbf{r}_j \rangle\right)\right\rangle}{\sqrt{\left(\langle \mathbf{r}_i^2 \rangle - \langle \mathbf{r}_i \rangle^2\right)\left(\langle \mathbf{r}_j^2 \rangle - \langle \mathbf{r}_j \rangle^2\right)}}; \qquad \sigma' = M^{1/2}\sigma_{ij}M^{1/2} = \mathbf{M\sigma}$$

$$S_{conf} = k\sum_{i}^{3n-6} \frac{\eta\omega_i/kT}{\exp(\eta\omega_i/kT) - 1} - \ln\left[1 - \exp(-\eta\omega_i/kT)\right]; \quad \omega_i = \sqrt{\frac{kT}{\lambda_i}} \qquad \textbf{[31]}$$

and λ is the *i*th eigenvalue in the co-variance matrix.

Both these methods will depend on the orientation of the instantaneous frames onto to the fixed reference set, and therefore these calculations must be preceded by removal of translational and rotational degrees of freedom.

4.4 VIRTUAL MUTAGENESIS

Besides calculating some physical properties of the biomolecule, it is interesting to estimate and predict putative sites for mutation. Such information may be useful in designing drugs or proteins or understanding biomolecular mechanism for humans with a specific mutation in their genes. A rapid way of preliminary mapping the protein stability is a so-called virtual alanine-scan [109] (paper IV). Putative mutation sites may then be resolved computationally by more elaborate calculations, for instance free energy perturbations.

A somewhat simplified view, for a protein with i residues in an ensemble of j structures, is given in the following pseudo-code:

```
DO I=1, NRES
  "CALL LIBRARY REFENER"
   DO J=1, NSTRUC
        READ STRUCTURE
          "CALCULATE ENERGY = E1"
          MUTATE RESIDUE "I"
         "CALCULATE ENERGY = E2"
          "CALC ENERGY DIFF = (E2 – E1) – REFENER"
   -----------
     ENDDO
------------
 ENDDO
```

The wild-type ensemble of structures can be generated with a MD or Monte-Carlo simulation.

5 RESULTS AND DISCUSSION

5.1 SOLVENT – CYSTEINATE LIGAND EXCHANGE (PAPERS I AND II)

The use of a non-bonded thiolate-zinc interaction potential and a zinc model[64] parameterized against TIP3P [82, 83] water, allows ligand solvent-thiolate exchange. This observation, and recent experimental [110-112] and theoretical [113-115] reports prompted a QM calculation on variant methiolate-Zinc-H2O complexes with an overall charge of –2.0

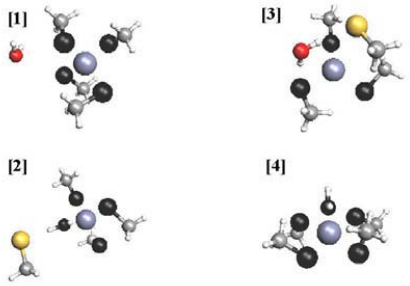

Figure 5. Model compounds of different methiolate-Zinc-H2O complexes. Atoms "bonded" to zinc are colored in black. The net charge on all systems was –2.0.

The potential energy surface from these calculations indicate that the relative energy-differences between the complexes are quite low, and thus support the findings of possible ligand exchange at the zinc.

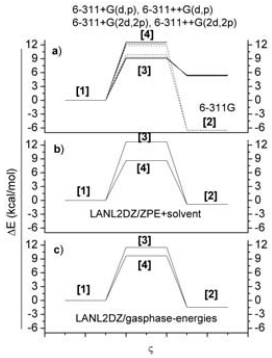

Figure 7. Potential energy surface of the optimized methiolate-Zn-H2O complexes relative the lowest energy (formation **[1]**). All optimizations were performed with the B3LYP functional and the LANL2DZ basis set.

5.2 THE REGULATORY ROLE OF PROLINE IN ZINC FINGERS (PAPERS II AND IV)

Despite the simple form of the zinc-cysteinate interaction potential, novel dynamics within the Cys4-zinc finger type has been observed. From a biological view, the dynamics in the second finger of the GR DBD (and similar systems) governs the orientation of the D-box and provides a route for the release/uptake of zinc possibly via residue C476, which exhibited increased side chain flexibility in contrast to the other cysteinate ligands. This flexibility was repeatedly seen in different simulations

with variant structures and system setups. MD simulations performed on a smaller system, methionyl-tRNA-syntethase (pdb accession code 1mea), showed no increased flexibility of the cysteinate ligands, despite the more loosely packed character of this protein fragment. Why?

One possible reason is the isomeric form of a following proline residue to one of the cysteinate ligands, thus forming a cys-pro residue pair. In this pair, the backbone torsion angle is very similar to a variant NMR structure in the GR DBD ensemble (structure number 9), and MD on this structure showed no significant flexibility in the C476 side chain either. These findings imply that proline is involved in regulating D-box conformations in GR DBD (and possibly in other proteins with a proline connected to a residue that is coordinated to a metal). It would thus be interesting to explore other Cys4-zinc fingers and in particular those with a proline linked to the zinc-ligand. A quick survey of the DBD phylogenicy within the NRs revealed that the cys-pro residue pair appears frequently in the first *or* in the second finger.

The virtual alanine-scan done in paper IV indicated that a mutation of P493 results in a stabilized protein structure for the free and the DNA-bound GR DBD.
MD dynamics was performed on "in silico" P493 mutants suggest that the P493R mutant recovers stabilisation by forming hydrogen bonds between the arginine sidechain and the backbone of Y474 and L475, which in turn induces the D-box to adopt a crystal like conformation in the NMR monomer solution structure. Such hydrogen pattern is not possible in the wild type structure, but the chemical nature of the imidazole ring stabilises the internal orientation of helices III and I. When the C492 Ψ dihedral angle was perturbed towards the trans isomer an increased mobility of the side chain of C476 appeared.

5.3 SOLVENT EFFECTS ON PROTEIN DYNAMICS (PAPER III)

None of the three water models appeared to influence the structural properties of the solute, but the number of hydrogen bonds and re-orientational correlation times increased in the order of TIP3P < SPC < SPC/E. The calculated generalized S^2 order S2 and the mean squared atomic fluctuations showed a similar trend, although not very markedly, and the choice of initial coordinates along a 20 ns trajectory for shorter simulations appeared very similar to each other.

5.4 VIRTUAL MUTAGENESIS IN GR DBD (PAPER IV)

In this work, a virtual alanine-scan was done on variant structures for the free monomer GR DBD and for a mutant EGA GR DBD and wild type GR DBD in complex with the glucocorticoid and estrogen response elements. Thus five different scans on all non-alanine residues in the GR DBD were performed. In all cases, mutation of P493 appeared to stabilize the free as well as the DNA-bound protein. MD dynamics was performed on "in silico" P493 mutants (Ala, Arg, Gln) done on structure number 14 in the 2gda NMR ensemble.
The method in itself provides a rather crude energy surface for mutated residues, but as a first approximation it offers a fast route for probing sites that may be interesting for biological or pharmaceutical communities.

6 CONCLUDING GLIMPSES INTO THE FUTURE

Given all these methods for calculating the physical properties of a biomolecule, it would be interesting to develop a self-consistent force field parameterisation scheme, for instance, calibrating α, β and γ linear interaction scaling parameters against FEP calculations. Or using the QM reaction field solvation free energies to calibrate force field Lennard-Jones parameters within the framework of GBMV or Poisson-Boltzmann electrostatic calculations. In such a way, force field parameters could be improved and honour the field of MD simulations with an increased predictive power, while maintaining the force field intactness, simplicity and computational efficiency.

The use of implicit solvation models will never result in a solvent ligand exchange, but an alternative is to post-solvate structures generated under continuum conditions.

An alternative approach in the alanine-scan would be to model a pseudo-sidechain, with variant charges (or a dipole) to provide an electrostatic picture of the mutant effect i.e. an electrostatic potential scan (EPS). Hopefully, I will be able to tackle some of these topics soon, but at this writing-moment, time is far too limited

7 TACK

När jag sitter här med tidsbrist och försöker formulera den kanske viktigaste delen av denna avhandling, inser jag att nanosekunderna har gått fort under dessa år på NOVUM, vilket förmodligen innebär att jag har haft skoj.

Som brukligt är tackar jag först och främst Lennart Nilsson för en *mycket* omdömesgill handledning, en stor generositet med tid, tålamod och kunskaper, de konferenser jag fick åka på samt ett och annat julbord.

Pekka ”går det bra?” Mark, Erik ”köttbjugg” Lundgren, Jonas ”törnrosa” Sjöström och Anders Lindholm för ett antal flin och en hel del gapskratt.

Övriga medlemmar i molekyl-böjar gruppen,

Arne, Janne, Mats, Ana, Peter, Jianxin, Boel, Katarina, Lotta och Sofia för intressanta diskussioner.

Kristallerna, med Rudolph Ladenstein i spetsen, Micke, Winnifried, Linda, Gudrun

Elekronmikropskopisterna, Hans, Peter, Matthew, Ingrid

Nuvarande kärnspinnare, Kurt, Tobias, Johan och

Avflyttade kärnspinnare, Torleif, Henke, Magnus W., Mange, Susanne, Helena, Peter, Niklas, Esmeralda

Kvantmekanikerna, Leif, Margareta, Arianna, Markus och Tore som alla lyckats med konststycket att lära mig en smula kvantmekanik.

Tre personer som förutom Lennart haft en direkt avgörande influens på min teoretiska böjelse:

Peter Politzer, Karl Hult och Pär Nordlund

Ursprungliga NOV2k-gänget och I synnerhet Sara, Lotta, Jenny och Magnus, samt stöd från Bo Lambert, Lotta Wikström och Jan-Åke Gustafsson.

Anette W., Per A., ekonomiavdelningen, Ingvar, Södertörns Högskola, service enheten, Tango

Peter Sand för ett antal goda skratt i ämnet ”beoretisk” kemi.

I vardagslivet vill jag tacka mina föräldrar, Gertie och Håkan, syskonen Jenny och Max, mormor och farmor, onklar, kusiner, halvsyskon, halvpappor och halvmammor för att…tja…för allt möjligt.

Goda grannar, som familjen Angerbjörn-Lidén och familjen Lidby.

Vänner och klubbar över hela världen, som *trots* sin existens ändå resulterade i denna avhandling.

Familjen Munshi – och i synnerhet Amy för kloka råd, Neena, Ingegerd, Sunil, Isabel och Yasmine.

Sist och endast i en mening minst:
Josephine, som fick utstå mitt frenetiska knackande på tangentbordet i kombination med en och annan frustrerad suck. Du är bäst!

Jag har säkert glömt någon, men jag tackar Dig också!

8 REFERENCES

1. Branden, C., Tooze, J., *Introduction to protein structure.* 1991, New York: Garland publishing Inc., New York.
2. Nishikawa, K., Ooi, T., Isogai, Y., Saito, N., *Tertiary structure of proteins. I. Representation and computation of the conformations.* J. Phys. Soc. (Jpn), 1972. **32**: p. 1331-1337.
3. Dalberg, A.E., *Ribosome structure: The ribosome in action.* Science, 2001. **292**: p. 868-896.
4. Cantor, C.R., Schimmel, P. R., *Biophysical chemistry.* II. 1980: W.H. Freeman and company, N.Y., USA.
5. Bernstein, F.C., Koetzle, T. F., Williams, G. J. B., Mayer, E. F., Brice, J. M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., Tasumi, M., *The protein data bank. A computer-based archival file for macromolecular structures.* J. Mol. Biol., 1977. **112**: p. 535-542.
6. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E., *The protein data bank.* Nucleic Acids Res., 2000. **28**: p. 235-242.
7. HUGO, *The human genome project.* Nature, 2001. **409**: p. 745-964.
8. Lipscomb, W.N., Sträter, N., *Recent advances in zinc enzymology.* Chem. Rev., 1996. **96**: p. 2375-2433.
9. *Zinc and health: Current status and future directions.* J. Nutr., 2000. **130**.
10. Berg, J.M. and Y. Shi, *The galvanization of biology: A growing appreciation for the roles of zinc.* Science, 1996. **271**: p. 1081-1085.
11. Pan, T., Freedman, L. P., Coleman, J. E., *Cadmium-113 NMR studies of the DNA binding domain of the mammalian glucocorticoid receptor.* Biochemistry, 1990. **29**(39): p. 9218-9225.
12. Low, Y.L., Hernández, H., Robinson, C. V. O'Brien, R., Grossmann, J. G. John E. Ladbury, J. E. and Luisi, B., *Metal-dependent folding and stability of nuclear hormone receptor DNA-binding domains.* J. Mol. Biol., 2002. **319**: p. 87-106.
13. Harrison, S.C., *A structural taxonomy of DNA-binding domains.* Nature, 1991. **353**: p. 715-719.
14. Kaptein, R., *Zinc fingers.* Curr. opin. struct. biol., 1991. **1**: p. 63-70.
15. Burley, S.K., Katsuhiko, k., *Transcription factor complexes.* Curr. Opp. Struct. Biol., 2002. **12**: p. 225-230.
16. Allen, M.P., Tildesley, D. J., *Computer simulations of liquids.* 1987: Oxford university press, Oxford, USA.
17. Newton, I., *Philosophiae naturalis principia mathematica.* 1687, Trinity College, Cambrigde, England.
18. Renaud, J.P., Moras, D., *Structural studies on nuclear receptors.* Cell. Mol. Life Sci., 2000. **57**: p. 1748-1769.
19. Tsai, M.-J., O'Malley, B. W., *Molecular mechanisms of action of steroid/thyroid receptor superfamily members.* Annu. Rev. Biochem., 1994. **63**: p. 451-486.
20. Umesono, K., Evans, R. M., *Determinants of target gene specificity for steroid/thyroid hormone receptors.* Cell, 1989. **57**: p. 1139-1146.
21. Kumar, R., Thompson, E. B., *The structure of the nuclear hormone receptors.* Steroids, 1999. **64**: p. 310-319.

22. Di Croce, L., Okret, S., Kersten, S., Gustafsson, J-Å., Parker, M., Wahli, W., and Beato, M., *Steroid and nuclear receptors Villefranche-sur-Mer, France, May 25-27, 1999.* EMBO Journal, 1999. **18**: p. 6201-6210.
23. Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schütz, G., Umeosono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., and Evans, R. M., *The Nuclear Receptor Superfamily: The Second Decade.* Cell, 1995. **83**: p. 835-839.
24. Mangelsdorf, D.J., Ewans, R. M., *The RXR heterodimers and orphan receptors.* Cell, 1995. **83**: p. 841-850.
25. Steinmetz, A.C.U., Renaud, J.P., Moras, D., *Binding of ligands and activation of transcription by nuclear receptors.* Annu. Rev. Biophys. Biomol. Struct., 2001. **30**(1): p. 329-359.
26. Giguère, V., Hollenberg,S. M., Rosenfeld, M. G., Evans R. M., *Functional domains of the human glucocorticoid receptor.* Cell., 1986. **46**: p. 645-652.
27. Härd, T., Kellenbach, E., Boelens, R., Maler, A. B., Dahlman, K., Freedman, L. P., Carlstedt-Duke, J., Yamamoto, K. R., Gustafsson, J-Å., Kaptein, R., *Solution structure of the glucocorticoid receptor DNA-binding domain.* Science, 1990. **249**: p. 157-160.
28. Baumann, H., Paulsen, K., Kovacs, H., Berglund, H., Wright, A. .H., Gustafsson, J-Å., Härd, T., *Refined solution structure of the glucocorticoid receptor DNA-binding domain.* Biochemistry, 1993. **32**(49): p. 13463-13471.
29. Luisi, B.F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K.R., Sigler, P. B., *Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA.* Nature, 1991. **352**: p. 497-505.
30. Schwabe, J.W.R., Neuhaus, D., Rhodes, D., *Solution structure of the oestrogen receptor.* Nature, 1990. **348**: p. 458-461.
31. Schwabe, J.W.R., Chapman, L., Finch, J. T., Rhodes, D., *The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: How receptors discriminate between their response elements.* Cell, 1993(75): p. 657-578.
32. Holmbeck, M.A.S., Foster, M.P., Casimiro, D.R., Sem, D.S., Dyson, J., Wright, P.E., *High-resolution solution structure of the retinoid X receptor DNA-binding domain.* J. Mol. Biol., 1998. **281**: p. 271-284.
33. Knegtel, R.M.A., Katahira, M., Schilthuis, J. G., Bonvin, A. M. J. J., Boelens, R., Eib, D., van der Saag, P. T., Kaptein, R., *The solution structure of the human retinoic acid receptor-β DNA-binding domain.* J. Biomol. NMR, 1993. **3**: p. 1-17.
34. Lee, M.S., Kliewer, S. A., Provencal, J., Wright, P. E., Evans, R. M., *Structure of the retinoid X receptor a DNA-binding domain: a helix required for homodomreic DNA-binding.* Science, 1993. **260**: p. 1117-1121.
35. Schwabe, J.W.R., Chapman, L., Finch, J. T., Rhodes, D., Neuhaus, D., *DNA recognition by the oestrogen receptor from solution to the crystal.* Structure, 1993. **1**: p. 187-204.
36. Gewirth, D.T., Sigler, P.B., *The basis for half-site specificity explored through a non-cognate steroid receptor-DNA complex.* Struc. Biol., 1995. **2**(5): p. 386-394.
37. van Tilborg, M.A.A., Lefstin, J. A., Kruiskamp, M., Teuben, J.-M., Boelens, R., Yamamoto, K. R., Kaptein, R., *Mutations in the glucocorticoid receptor DNA-binding domain mimic an allosteric effect of DNA.* J. Mol. Biol., 2000. **301**: p. 947-958.
38. van Tilborg, M.A.A., Bonvin, A. M. J., Hård, K., Davis, A. L., Maler, B., Boelens, R., Yamamoto, K. R., Kaptein, R., *Structure refinement of the*

glucocorticoid receptor DNA-binding domain from NMR data by relaxation matrix calculations. J. Mol. Biol., 1995. **247**: p. 689-700.

39. Anuradha Ray, K., LaForge, S., Pravinkumar, B.S., *Repressor to activator switch by mutations in the first Zn finger of the glucocorticoid receptor:Is direct DNA binding necessary?* Proc. Natl. Acad. Sci., 1991. **88**: p. 7086-7090.
40. Zillicaus, J., *Structural determinants of DNA-binding specificity by the glucocorticoid receptor*, in *Center for Biotechnology and Department of Medical Nutrition at NOVUM*. 1994, Karolinska Institute: Stockholm.
41. Lefstin, J.A., Yamamoto, K. R., *Allosteric effects of DNA on transcriptional regulators.* Nature, 1998. **392**: p. 885-888.
42. Szcabo, A., Ostlund, N. S., *Modern quantum chemistry.* 1996: Dover publications, inc. Minneola, N. Y., USA.
43. Becke, A.D., *Density-functional thermochemistry. III. The role of exact change.* J. Chem. Phys., 1993. **98**: p. 5648-5652.
44. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1998.
45. Steward, J.J.P., *MOPAC: a semi-empirical molecular orbital program.* J. Comp. Aid. Mol. Des., 1990. **4**: p. 1-45.
46. Dewar, M.J.S. and W.J. Thiel, *Ground states of molecules.38. The MNDO method. Approximations and parameters.* J. Am. Chem. Soc, 1977 a. **99**: p. 4899-4906.
47. Dewar, M.J.S., Thiel, W. J., *Ground states of molecules.39. MNDO results for molecules containing hydrogen, carbon, nitrogen and oxygen.* J. Am. Chem. Soc., 1977b. **99**: p. 4907-4917.
48. Dewar, M.J.S., Merz, K. M. Jr., *MNDO Calculations for compounds containing zinc.* Organometallics, 1986. **5**: p. 1494-1496.
49. Brooks, B.R., Bruccoleri E., Olafson B.D., States D.J., and S. Swaminathan, Karplus, M., *CHARMM: A Program for macromolecular energy minimization and dynamics calculations.* J. Comp. Chem., 1983. **4**: p. 187-217.
50. Mulliken, R.S., *Electronic population analysis on LCAO-MO molecular wave functions. I.* J. Chem. Phys., 1955. **23**: p. 1833-1846.
51. Löwdin, P.-Q., *On the orthogonality problem.* Adv. quant. chem., 1970. **5**: p. 185-199.
52. Politzer, P., Murray, J. S., *Molecular electrostatic potentials and chemical reactivity*, in *Reviews in Computational Chemistry*, K.B. Lipkowitz, Boyd, D. B, Editor. 1991, VCH Publishers: New York. p. 273-312.
53. Wiberg, K.B., Rablen, P. R, *Comparison of atomic charges derived via different procedures.* J. Comp. Chem., 1993. **14**: p. 1504-1518.
54. Eriksson, M.A.L., Härd, T., Nilsson, L., *Molecular dynamics simulations of the glucocorticoid receptor DNA-binding domain in complex with DNA and free in solution.* Biophys. J., 1995. **68**: p. 402-426.

55. Hoops S., C., Anderson, K., W., Merz Jr , K. M., *Force field design for metalloproteins.* J. Am. Chem. Soc., 1991(113): p. 8262-8270.
56. Hæffner, F., Brinck, Tore, Hæberlin, M., Moberg, K., *Force field parameterization of copper(I)-olefin systems from density functional calculations.* J. Mol. Struc. (Theochem), 1997. **397**: p. 39-50.
57. Merz, J.K.M., *Insights into the function of the zinc hydroxide-thr199-glu106 hydrogen bonding network in carbonic anhydrases.* J. Mol. Biol., 1990(214): p. 799-802.
58. Ryde, U., *Molecular dynamics simulations of alcohol dehydrogenase with a four- or five-coordinate catalytic zinc ion.* Proteins, 1995(21): p. 40-56.
59. Vedani, A., Huhta D.W., *A new force field for modeling metalloproteins.* J. Am. Chem. Soc., 1990(112): p. 4759-4767.
60. Gajewski, J.J., Gilbert, K. E., Kreek, T. W., *General molecular mechanics approach to transition metal complexes.* J. Comp. Chem., 1998. **19**: p. 1167-1178.
61. Åqvist, J., Warshel, A., *Computer simulation of the initial proton transfer step in human carbonic anhydrase I.* J. Mol. Biol., 1992. **224**: p. 7-14.
62. Pang, Y.-P., *Novel zinc protein molecular dynamics simulations: steps toward antiangiogenesis for cancer treatment.* J. Mol. Mod., 1999. **5**: p. 196-202.
63. Besler, B.H., Merz, K.M. Jr., Kollman, P.A., *Atomic charges derived from semi-empirical methods.* J. Comp. Chem., 1990. **11**: p. 431-439.
64. Stote, R.H., Karplus, M., *Zinc binding in proteins and solution: A simple but accurate nonbonded representation.* Proteins, 1995. **23**: p. 12-31.
65. McKerrel, Jr., A. D., Bashford, D., Bellot, M., Dunbrack, Jr R.L., Evanseck, J., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Ngyen, D.T., Prodholm, B., Reiher, I.W.E, Roux, B., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus. M., *All-hydrogen empirical potential for molecular modeling and dynamics using the CHARMM22 force field.* J. Phys. Chem. B, 1998. **102**: p. 3586-3616.
66. Leach, A.R., *Molecular modelling - principles and applications.* 2001: Pearson education limited, England.
67. Hockney, R.W., *The potential calculation and some applications.* Meth. Comp. Phys., 1970. **9**: p. 136-211.
68. Gear, C.W., *Numerical initial value problems in ordinary differential equations.* 1971: Englewood Cliffs, NJ, Prentice Hall.
69. Swope, W.C., Anderson, H. C., Berens, P. H., Wilson, K. R, *A computer simulation method for the formation of physical clusters of molecules: Application to small water clusters.* J. Chem. Phys., 1982. **76**: p. 637-649.
70. Verlet, L., *Computer 'experiments' on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules.* Phys. Rev., 1967. **159**: p. 98-103.
71. Brooks, C.L.I., Brunger, A.T., Karplus, M., *Active site dynamics in protein molecules: A stochastic boundary molecular-dynamics approach.* Biopolymers, 1985. **24**: p. 843-865.
72. Brooks III, C.L., Karplus, M., *Deformable stochastic boundaries in molecular dynamics.* J. Chem. Phys., 1983. **79**: p. 6312-6325.
73. Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C., *Numerical integration of the cartesian equitaions of motion of a system with constraints: Molecular dynamics of n-alkanes.* J. Comp. Chem., 1977. **23**: p. 327-341.
74. Fihuerido, F., Levy, R. M., Zhou, R., Berne, B.J., *Large scale simulation of macromolecules in solution: Combining the periodic fast multipole method with multiple time step integrators.* J. Chem. Phys., 1997. **106**: p. 9835-9849.

75. M. Chun, H.M., Padilla, C. E., Chin, D. N., Watanabe, M., Karlov, V.I., Alper, H. E., Soosaar, K., Blair, K. M., Becker, O. M., Caves, L. S. D., Nagle, R., Haney, D. N., Farmer, B. L., *MBO(N)D: A multibody method for long-time molecular dynamics simulations.* J. Comp. Chem., 2000. **21**: p. 159-184.
76. Steinbach, P.J., Brooks, B. R., *New spherical-cutoff methods for long range forces in macromolecular simulation.* J. Comp. Chem., 1994. **15**(7): p. 667-683.
77. Petersen, H.G., Soelvason, D., Perram, J. W., Smith, E. R., *The very fast multipole method.* J. Chem. Phys., 1994. **101**: p. 8870-8876.
78. Ewald, P., *Due berechnung optischer und elektrostatischer gitterpotentiale.* Annalen der physik, 1921. **64**: p. 253-287.
79. Darden, T., York, D., Pedersen, L., *Particle Mesh Ewald: An N·log(N) method for Ewald sums in large systems.* J. Chem. Phys., 1993. **98**: p. 10089-10092.
80. Sagui, C., Darden, T.A., *Molecular dynamics simulations of biomolecules: Long-range electrostatic effects.* Annu. Rev. Biophys.Biomol. Struct., 1999. **28**: p. 155-179.
81. Amisaki, T., *Precise and efficient Ewald summation for periodic fast multipole method.* J. Comp. Chem, 2000. **21**: p. 1075-1087.
82. Jorgensen, W.L., Chandrasekar, J., Madura, J.D, Impey, R.W., Klein, M.L., *Comparison of simple potential functions for simulating liquid water.* J. Chem. Phys., 1983. **79**: p. 926-935.
83. Nera, E., Fischer, S., Karplus, M., *Simulation of activation free energies in molecular systems.* J. Chem. Phys, 1996. **105**: p. 1902-1921.
84. Berendsen, H.J.C., Postma, J. P. M., van Gunsteren, W. F., Hemans, J., *Intermoleular Forces*. 1981: Pullman, B, Ed.; Reidel: Dordrecht,. 331.
85. Berweger, C.D., van Gunsteren, W. F., Müller-Plathe, F., *Force field parameterization by weak coupling: Re-enginering SPC water.* Chem. Phys. Lett., 1995. **232**: p. 429-436.
86. Berendsen, H.J.C., Grigera, J. R., Straatsma, T. P., *The missing term in effective pair potentials.* J. Phys. Chem., 1987(91): p. 6269-6271.
87. Born, M., *Volumen and hydratationzwärme der ionen.* Zeitschrift für physik, 1920. **1**: p. 45-48.
88. Still, C.W., Tempczyrk, A., Hawley, R. C., Hendrickson, T., *Semianalytical treatment of solvation for molecular mechanics and dynamics.* J. Am. Chem. Soc., 1990. **112**: p. 6127-6129.
89. Lee, M.S., J. Salsbury, Freddie R., and C.L. Brooks III , *Novel generalized Born methods.* J. Chem. Phys., 2002. **116**(24): p. 10606-10614.
90. Onsager, L., *Electric moments of molecules in liquids.* J. Am. Chem. Soc., 1936. **58**: p. 1486-1493.
91. Barone, V.a.C., M., *Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model.* J. Phys. Chem. A, 1998. **102**: p. 1995-2001.
92. Lipari, G., Szabo, A., *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules.1. Theory and range of validity.* J. Am. Chem. Soc, 1982a. **104**: p. 4546-4559.
93. Lipari, G., Szabo, A., *Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules.2. Analysis of experimental results.* J. Am. Chem. Soc., 1982b. **104**: p. 4559-4570.
94. Eriksson, M.A.L., Berglund, H., Härd, T., Nilsson, L., *A comparison of ^{15}N NMR relaxation measurements with a molecular dynamics simulation:*

Backbone dynamics of the glucocorticoid receptor DNA-binding domain. Proteins, 1993. **17**: p. 375-390.
95. Berglund, H., Kovacs, H. Dahlman-Wright, K., Gustafsson, J-Å., Härd, T., *Backbone dynamics of the glucocorticoid receptor DNA-binding domain.* Biochemistry, 1992. **31**: p. 12001-12011.
96. Wikström, A., Beglund, H., Hambræus, C., van den Berg, S., Härd, T., *Conformational dynamics and molecular recognition: Backbone dynamics of the estrogen receptor DNA-binding domain.* J. Mol. Biol., 1999. **289**: p. 963-979.
97. Lee, B., Richards, F. M., *The interpretation of protein structures: estimation of static accesibility.* J. Mol. Biol., 1971(55): p. 379-400.
98. Zwanzig, R.W., *High-temperature equation of state by a perturbation method. I. Nonpolar gases.* J. Chem. Phys., 1954. **22**: p. 1420-1426.
99. Tobias, D.J., Brooks III, C. L., *Molecular dynamics with internal coordinate constraints.* J. Chem. Phys., 1988. **89**: p. 5115-5126.
100. Torrie, G.M., Valleu, J.P., *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling.* J. Comp. Phys., 1977. **89**: p. 187-199.
101. Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A, Rosenberg, J. M., *The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method.* J. Comp. Chem., 1992. **13**: p. 1011-1021.
102. Åqvist, J., Medina, C., Samuelsson, J.E, *A new method for predicting binding affinity in computer-aided drug design.* Prot. Eng., 1994. **7**: p. 385-391.
103. Hansson, T., Marelius, J., Åqvist, J., *Ligand binding affinity prediction by linear interaction methods.* J. Comp.-Aid. Mol. Des., 1998. **12**: p. 27-35.
104. Hansson, T., Åqvist, J., *Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations.* Prot. Eng., 1995. **8**: p. 1137-1144.
105. Sitkoff, D., Sharp, K. A., Honig, B., *Accurate calculation of hydration free energies using macroscopic solvent models.* J. Phys. Chem., 1994. **98**: p. 1978-1988.
106. Stone, M.J., *NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding.* Acc. Chem. Res., 2001. **34**: p. 379-388.
107. Yang, D., Kay, L. E., *Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: Application to protein folding.* J. Mol. Biol., 1999. **263**: p. 369-382.
108. Andricioaei, I., Karplus, M., *On the calculation of entropy from covariance matrices of the atomic fluctuations.* J. Chem. Phys., 2001. **115**: p. 6289-6292.
109. Huo, S., H., Massova, I., Kollman, P. A., *Computational alanine scanning of the 1:1 human growth hormone-receptor complex.* J. Comp. Chem., 2002. **23**: p. 15-27.
110. Fabris, D., Hathout, Y., Fenselau, C., *Inverstigation of zinc chelation in zinc-finger arrays by electrosspray mass spectrometry.* Inorg. Chem., 1999. **38**: p. 1322-1325.
111. Cano-Gauci, D.F., Sarkar, B., *Reversible zinc exchange between metallothionein and the estrogen receptor zinc finger.* FEBS Lett., 1996. **386**: p. 1-4.
112. Freedman, L.P., Luisi, B. F., Korzun, R., Basavappa, R., Sigler, P. B., Yamamoto, K. R., *The function and structure of the metal coordination sites within the glucocorticoid receptor DNA binding domain.* Nature, 1988. **334**: p. 543-546.

113. Maynard, A.T., Covell, D. G., *Reactivity of zinc finger cores: Analysis of protein packing and electrostatic screening.* J. Am. Chem. Soc., 2001. **123**: p. 1047-1058.
114. Simonson, T., Calimet, N., *Cys_xHis_y-Zn^{2+} interactions: Thiol vs. thiolate coordination.* Proteins, 2002. **49**: p. 37-48.
115. Deerfield II, D.W., Carter Jr, C. W., Pedersen, L. G., *Models for protein-zinc ion binding sites. II. The catalytic sites.* Int. J. Quant. Chem., 2001. **83**: p. 150-165.