

Computer simulation of binding free energy differences for the oppa:kxx system using the
kbff20 force field

by

Trust Onoriode Ekama

B.S., University of Lagos, 2014

M.S., University of Lagos, 2017

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Chemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Prof. Paul E. Smith

Copyright

© Trust Ekama 2023.

Abstract

Free energy calculations are an important part of modern-day drug design processes and molecular dynamics (MD) simulation has become a crucial tool in predicting binding affinities of ligands to their substrates. The quality of a MD simulation depends on the degree of sampling of important events during the simulation, and the accuracy of the force field adopted to describe the interactions between the atoms and molecules included in the system of study. The KBFF20 force field is a recently developed protein force field that continues to be validated.

Here we investigate its capabilities to represent solute-solute interactions in protein-ligand systems by calculating relative binding free energies for the OppA:KXK system and comparing the results with experimental data. OppA is a water-mediated oligopeptide binding protein with the remarkable characteristic of binding to 2-5 residue long peptides with little regard for their composition (amino acid sequence). Experimental binding free energies, enthalpies and entropies are available for the three residue peptides, KXK, where X is all twenty of the naturally occurring amino acid residues. Furthermore, X-ray crystal structures indicate that all twenty peptides bind to OppA with an identical pose.

Here, we determine changes in free energy of binding of several KXK peptides to OppA using the KBFF20 force field. In particular, we address the issues of sampling, the presence of mobile water molecules in the binding site, and analytical corrections to the free energies for charged residues arising due to system size effects. Finally, we also investigate the role of bound ions close to the peptide binding site and their contribution to the free energy differences. In summary, the OppA:KXK system represents a unique system that is available to investigate the accuracy of peptide-protein interactions representative of all twenty amino acid side chains.

However, significant care is required to obtain precise results using current computational approaches.

Table of Contents

List of Figures	vi
List of Tables	viii
Acknowledgements	ix
Chapter 1 - Introduction.....	1
Chapter 2 - Free energy calculations for ligand binding to the OppA protein	21
Chapter 3 - Local electroneutrality and the effect on free energy calculations	46
Chapter 4 - Summary and Future Directions	56
References	57
Appendix A - TI curves for alchemical transformations	73
Appendix B - FEP curves for CMAP transformations	90

List of Figures

Figure 1: A simplified schematic illustrating the basic MD simulation approach.	2
Figure 2: Pictorial representation of the typical components representing a force field.	6
Figure 3 : A thermodynamic free energy cycle using alchemical transformations to determine relative binding free energies. ⁸³	15
Figure 4 : Unliganded OppA structure showing its open conformational state in surface representation drawn with PYMOL.	17
Figure 5 : OppA complex in its closed conformational state in surface representation drawn with PYMOL.....	17
Figure 6: Aligned crystal structures of OppA:KAK and KDK with the ligands in blue and red respectively illustrating the same binding pose drawn with Pymol.....	18
Figure 7 : Excerpt of the result table of calculated binding free energies of OppA:KXK system reported by Tame et al., ⁸⁶	22
Figure 8 : Depiction of a macromolecule (light grey) in a solvent (dark grey) of dielectric constant of 80. ⁹²	24
Figure 9 : Phi & Psi angles in (A) Alanine dipeptide and (B) glycine dipeptide. ⁹⁸	26
Figure 10 : Ramachandran plot showing the allowed secondary structures. ¹⁰¹	27
Figure 11 : Soft-core interactions at $\lambda = 0.5$, with $p = 2$ and $C6A = C12A = C6B = C12B =$ 1 . ¹⁰⁸	29
Figure 12 : A cartoon representation of the correction scheme from MD simulation performed with explicit solvent and periodic images to MD under non periodic boundary condition. ⁹¹	32
Figure 13 : Snapshots showing water molecules within the active site (red and white) and bulk water (blue) at the beginning (Top) and at the end (Bottom) of a 100 ns simulation.....	36
Figure 14 : Alchemical transformation route for KAK to KEK in both ligand in protein and free in solution and the reverse transformation.	37
Figure 15 : Plots of $\partial H \partial \lambda$ and ΔG as a function of λ the transformation from Alanine to Glutamic acid (KAK to KEK) for the free peptide in solution obtained from 10 ns simulations of forward (black lines) and backward (red lines) processes. In both panels,	

dashed red lines are the flipped reverse transformations to compare with forward transformations and evaluate hysteresis.....	39
Figure 16: Plots of $\partial H \partial \lambda$ and ΔG as a function of λ for the transformation from Alanine to Glutamic acid (KAK to KEK) when bound to OppA as obtained from 50 ns simulations of forward (black lines) and backward (red lines) processes. In both panels, dashed red lines are the flipped reverse transformations to compare with forward transformations and evaluate hysteresis.	40
Figure 17 : correction map (CMAP) plot for Alanine to Glutamic acid (free peptide in solution) obtained from analyzing $\lambda = 1$ point of the forward reaction (left) and on the right is CMAP for the backward simulation, the trajectories of $\lambda = 0$ was analyzed.	41
Figure 18: correction map (CMAP) plot for Alanine to Glutamic acid (complex) obtained from analyzing $\lambda = 1$ point of the forward reaction (left) and on the right is CMAP for the backward simulation, the trajectories of $\lambda = 0$ was analyzed.	42
Figure 19: Negative ion surrounded by positively charged ion to illustrate ion atmosphere. ¹²⁹ ..	47
Figure 20 : Charged residues in the active site of the OppA:KKK complex experimental structure (Magenta) aligned with a simulated structure (cyan) showing acetate ion (Magenta & Red) and chloride (Green) ion binding in both structures.	48
Figure 21 : Ion residue distance time histories for the simulation of OppA:KEK in 2 M NaCl solution. The distances for any ions that came within 0.5 nm of any charged pocket residues (p.r) are shown. Black and red represent chloride ions within 0.5 nm of charged pocket residues, while blue represents sodium ion in close contact with a negatively charged Glu acid residue.	51
Figure 22 : Plots of $\partial H \partial \lambda$ (top) and ΔG (bottom) as a function of λ for K=+ to K=0 with (black) and without (red) chloride ion in the active site.	54

List of Tables

Table 1 : Free energy differences (kJ/mol) for alchemical transformation AtoE and the reverse transformation.	38
Table 2: A summary of the charging free energy corrections for charged transformations obtained using analytical scheme.....	38
Table 3 : A summary of calculated relative free binding energies ($\Delta\Delta G_{\text{binding}}$ (kJ/mol)) for all 11 alchemical transformations obtained using the KBFF20 force field compared to experimental data obtained from Ref. 86. Gold represents transformations involving polar residues, Dark Orange are non-polar groups, and Pink and Blue involve negatively charged and positively charged residues respectively.	43
Table 4 : Summary of number of Cl ions in the active site for different systems and force fields.	52

Acknowledgements

My profound appreciation to my advisor, Dr. Paul Smith, for his unwavering support during my master's studies and research and exposing me to the rigors of research. Thank you for giving me the opportunity to carry out research and for the helpful guidance given to me throughout the entire research and thesis writing process.

I would like to thank my thesis committee: Dr. Christine Aikens and Dr. Ping Li for their time valuable time and insightful questions. I am grateful for providing me with the opportunity to learn from you.

My gratitude also goes to Dr. Elizabeth Ploetz for technical assistance through my research.

Last but not least, I appreciate Shin, Farhat, friends and family for their help, support and encouragement during my master's program.

Chapter 1 - Introduction

1.0 Molecular Dynamics

Molecular Dynamics (MD) simulation is a powerful and versatile tool that has been employed to investigate the structure and dynamics of molecules and materials.¹⁻³ This tool has been used in a wide variety of scientific disciplines for many decades, and the development of high-speed computers has made it possible to perform simulations for increasingly longer times and larger systems.^{2,4} The technique is often used to help us understand the movement of atoms/molecules in a system and how these motions play a role in the structural and biological function of many systems.^{5,6}

The first MD simulation was performed by Alder and Wainwright in 1957. Using an IBM 704 computer they simulated the interaction of a system of hard spheres⁷. Gibson et al. simulated radiation damage of solid copper in 1960 and this represented the first realistic simulation of matter.⁸ In 1964, a simulation of liquid argon was performed by Rahman to study the properties of a liquid.⁹ In 1975, the first computer simulation of a biological system (the Bovine Pancreatic Trypsin inhibitor, BPTI protein) was performed, but the simulation time was under 10 picoseconds.¹⁰ Nevertheless, even the simple motions provided during the simulation were an advance on the static crystal structure data of proteins available at that time.

In recent years, MD simulations, and computer simulations in general, have become a popular choice for probing the nature of biological systems. Experimental techniques such as crystallography can be laborious,¹¹ and many times impossible for large proteins, leading to alternative techniques like computation being applied to more and more systems. MD simulations have tremendously transformed the field of biophysics and drug discovery by providing a better understanding of biological systems. This has been achieved through significant improvements in

speed (computer hardware) and accuracy (force fields and computational methods) of modern-day simulations.^{2,12} In particular, MD simulations can be used to validate experimental results and shed light on phenomena and behaviors that are obscure to many experimental techniques.

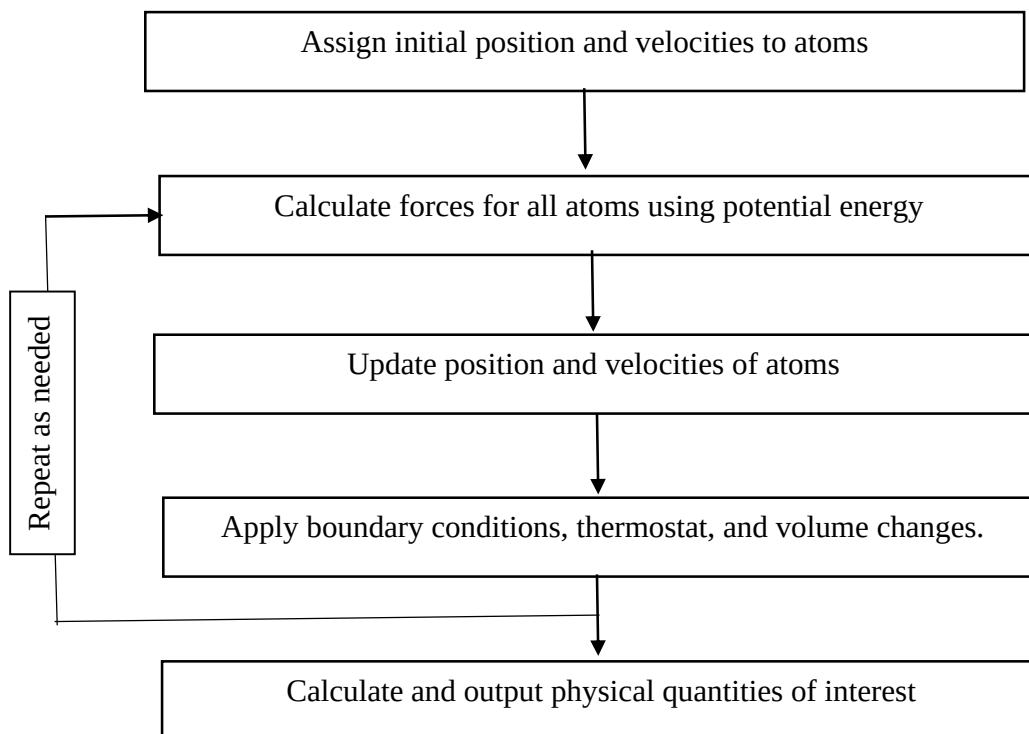


Figure 1: A simplified schematic illustrating the basic MD simulation approach.

Classical MD typically involves the simulation of the movement and interaction of a large number of atoms/molecules in a system by solving the Newton's equation of motion (EOM) for the system.¹³ The classical Hamiltonian, H , provides the total energy of the system and is calculated from the potential, V , and kinetic energy, K , according to,

$$H(\{q\}, \{p\}) = V(\{q\}) + K(\{p\}) \quad 1$$

where the potential and kinetic energies depend on a set of generalized coordinates, q , and their conjugate momenta, p , respectively. The kinetic energy is easily determined according to,

$$K(\{p\}) = \sum_i \frac{p_i^2}{2m_i} \quad 2$$

$$H(\{q\}, \{p\}) = V(\{q\}) + \sum_i \frac{p_i^2}{2m_i} \quad 3$$

However, the potential energy requires a series of equations and parameters, collectively known as the force field, that provides the energy and forces corresponding to a specific set of atomic/molecular coordinates. An accurate classical computer simulation requires an accurate force field. The nature of modern biomolecular force fields will be discussed later.

Hamilton's equations of motion are given by,

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \quad \dot{p}_i = -\frac{\partial H}{\partial q_i} \quad 4\&5$$

For cartesian coordinates these can be written,

$$\dot{r}_i = p_i/m_i \quad \dot{p}_i = m_i \ddot{r}_i = -\frac{\partial V}{\partial r_i} = f_i \quad 6\&7$$

These equations of motion can be implemented using a variety of algorithms, such as the Verlet algorithm,¹⁴ to move the atoms/molecules in time under the influence of forces obtained from the applied force field. The accuracy and reliability of an MD simulation critically depends on the process of generating representative configurations of the system being simulated, referred to as sampling, and an accurate representation of the interactions provided by the force field.^{15,16}

1.1 Applications of MD Simulation

MD simulation is a powerful tool that can be applied to many different problems in science and engineering. Applications of MD include studies of molecular interactions of biological systems,^{3,12,17-20} the kinetics of chemical reactions,^{21,22} the formation of protein structures,²³⁻²⁵ the

structure of crystalline materials,²⁶⁻²⁹ the dynamics of many mechanical systems,³⁰ etc. Some of the more common applications of MD simulation are in biochemical and pharmaceutical research, and material science. This computational tool has been employed to investigate biological processes, with the aim of learning more about how diseases may arise and how they can be effectively treated. In the pharmaceutical sciences, computer simulation is often used, in conjunction with other computational tools, to help discover, design and develop new drugs, and to better understand how the drugs work to treat disease.

The ongoing global pandemic (COVID-19) has scientists all over the world working frantically to find ways to cure and/or manage the coronavirus disease. MD simulation is one of the many tools employed by researchers to investigate the virus, and to screen drugs that might be potent against it. Deganutti et al., used MD simulations to identify small molecules that can effectively prevent the receptor binding domain (RBD) of the spike protein from binding to human angiotensin-converting enzyme (ACE-2).³¹

MD simulations have also been extensively used to study drug loading and controlled release. Poojari et al. used MD to demonstrate that cholesterol and itraconazole, an antifungal drug, do not mix well thereby invalidating the approved clinical formulations of incorporating cholesterol into liposome-itraconazole drugs.³² They showed the presence of cholesterol makes it difficult for the drug to reach the liposomes due to reduced affinity, using an experimental study which supports the findings from MD.

Solubility studies for drugs is another significant area of drug discovery where MD simulations have been very useful. Drugs taken orally are usually crystalline. Hence, for them to be active they need to be dissolved. Das et al. calculated the solvation free energy of a cardiovascular drug using MD simulation to gain insight into the solvation mechanism of the drug

in both water and salt solutions.³³ By performing a radial distribution function analysis, they measured interactions between drug molecules and the different salt solutions, thereby providing information regarding the solubility of the drug in those solutions.³³

Since MD simulation provides detailed information on the atomistic scale, it is often employed to investigate physical, structural, electronic, thermal and chemical properties of materials.^{34,35} In metals or polymers, mechanical properties such as strength, stiffness and toughness can be investigated through MD simulations to gain insight into the structure-property relationships of these materials. Zhu and Averback studied the densification of nanocrystalline copper using MD simulation.³⁶ Their findings show that this material becomes rather unstable at a temperature as high as 1100 K, causing the atoms to reaggregate and grow. These few examples clearly illustrate the role computer simulation plays in modern day research.

1.2 Force Fields

Force fields play an important role in the quality of a molecular dynamics simulation. While the precision of the simulation results is dictated by the degree of sampling achieved, the accuracy of the simulation results depend on the quality of the force field.³⁷⁻⁴⁰ The force field describes exactly how particles interact in an atomistic system. A force field (FF) is a collection of equations and parameters that describe the bonding and interactions between atoms/molecules and thereby provide the energy and forces on the atoms. To accurately represent the interactions within and between particles, many parameters are usually required from experimental data obtained from techniques such as crystallography, spectroscopy, NMR, etc. In addition, accurate quantum mechanical studies are often used to provide additional parameters that might be unavailable via experiment.⁴¹

The potential energy, V , provided by a force field suitable for biomolecular simulations is usually a sum of bonded terms (bonds, angles, dihedral and improper) and non-bonded terms such as electrostatics and dispersion forces, as indicated in Eqs. 8 and 9. This illustrates one of the main underlying assumptions in FF design, known as additivity,⁴² where the total potential energy can be decomposed into a series of additive terms. A second assumption is that of transferability which allows the application of the potential energy terms developed on a small set of test molecules to represent a broader range of molecules with similar functional groups. Both concepts are seen in most biomolecular force fields in use today.

$$V = V_{bond} + V_{angle} + V_{dihedral} + V_{imp} + V_{vdw} + V_{coulomb} \quad 8$$

$$V_{bonded} = V_{bond} + V_{angle} + V_{dihedral} + V_{imp} \quad 9$$

$$V_{nonbonded} = V_{vdw} + V_{coulomb} \quad 10$$

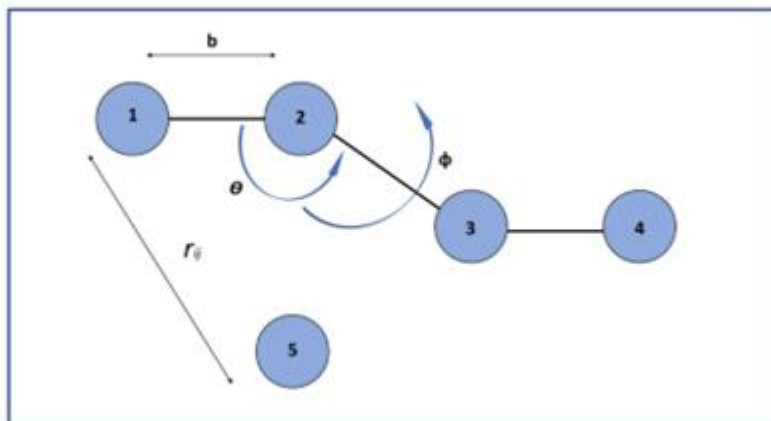


Figure 2: Pictorial representation of the typical components representing a force field.

The harmonic potential is used to represent the quantum mechanical interaction between two atoms with bond length, r , such that,

$$V_{bond} = \frac{1}{2}k_b(r - r_0)^2 \quad 11$$

where the k_b and r_0 parameters represent the force constant and equilibrium bond length, respectively. These parameters are usually obtained from X-ray diffraction, quantum mechanics (QM), infrared or Raman spectra. For large bond displacements compared to the reference bond length, the harmonic potential is a poor approximation, and this simple model does not allow for the breaking of bonds.⁴³

For bond angles, θ , formed between three atoms connected via a central atom, the harmonic potential is used to represent this interaction according to,

$$V_{angle} = \frac{1}{2}k_\theta(\theta - \theta_0)^2 \quad 12$$

where k_θ is the force constant and θ_0 represents the equilibrium angle in radians. Parameters for bond angles are usually from X-ray/neutron crystals, infrared spectroscopy, and quantum mechanical (QM) calculations.

The third term in the total potential equation is the dihedral (torsion) angle potential. This is used to map variations in the conformational energies due to rotation around single (and double) bonds. To describe the rotation around a torsional angle, ϕ , a common potential that represents the interaction between four atoms that are connected via three sequential bonds is provided by,

$$V_{dihedral} = \frac{1}{2}k_\phi[1 + \cos(n\phi - \delta)] \quad 13$$

where k_ϕ is a force constant, n (the multiplicity) is a count of the number of minima or maxima between 0 – 360 degrees of rotation, and δ is a phase shift measured in radians. These parameters are obtained from X-ray/neutron crystals, Infra-red (IR), Nuclear Magnetic Resonance (NMR), and QM.

Another four-atom interaction term is usually added to account for planar and chiral groups, like the aromatic rings, and involves using non sequential sets of four atoms to define an (improper) dihedrals. This potential is given by,

$$V_{imp} = \frac{1}{2}k_{\varphi}(\varphi - \varphi_0)^2 \quad 14$$

where k_{φ} is a force constant and φ_0 an equilibrium improper dihedral angle. Parameters are obtained from IR and QM. The first four terms in eqn. 8 are usually referred to as intramolecular interactions or bonded terms.

The van der Waals (vdws) interactions and Coulomb potential account for non-bonded terms, representing the major intermolecular contributions to the potential energy of a system, and these are usually determined between pairs of atoms, i and j . The Lennard-Jones potential is traditionally used to describe the vdws interaction energy between a pair of atoms when they are separated by a distance r_{ij} . The energy of the non-bonding atoms or molecules is determined as a function of the interatomic separation distance between the two particles.

$$V_{vdw} = 4\varepsilon_{ij} \left[\left[\frac{\sigma_{ij}}{r_{ij}} \right]^{12} - \left[\frac{\sigma_{ij}}{r_{ij}} \right]^6 \right] \quad 15$$

The two terms in the square bracket represent the repulsion and attraction between the particles, respectively. When atoms approach each other, from an infinite distance, there is a force of attraction and the $(1/r)^6$ term dominates until a minimum energy is attained, after which they start to repel and at this point the $(1/r)^{12}$ term dominates.⁴⁴ The prefactor ε_{ij} is an interaction parameter that depends on the atom types, while σ_{ij} is a size parameter. Both ε_{ij} and σ_{ij} can be obtained

from equation of state data for atoms in gas phase, they can also be extracted from QM calculations and x-ray/neutron crystal data.

Usually, ε and σ are known for the same atom types (ii and jj) and therefore combination rules are used to calculate the interaction between different types (ij) of atoms in a molecule. These rules define how the parameters used to describe the interactions between atoms in a force field are combined to obtain the interaction parameters for different types of atom pairs. The most used combination rules are the Lorentz-Berthelot rules,⁴⁵ which state that the Lennard-Jones parameters (ε and σ) for a pair of atoms are obtained by taking the arithmetic mean of the parameters for the two individual atoms.

Alternatively, a geometric combination can also be adopted for σ . Specifically, the rules are:

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \quad \sigma_{ij} = \sqrt{\sigma_{ii}\sigma_{jj}} \quad \text{or} \quad \sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \quad 16,17\&18$$

The Coulomb potential which exists between all ions and polar molecules is used to represent the electrostatic interactions in pairs of atoms. A pair of molecules/ions i and j with partial atomic charges q_i and q_j separated by a distance r_{ij} interact according to,

$$V_{coulomb} = \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}} \quad 19$$

where ε_0 is the permittivity of the medium. For polar molecules, fractional charges that sum up to total charges of the molecule are assigned to each atom or group of atoms, while in ionic compound (like NaCl) the charges are +1 for Na^+ and -1 for Cl^- are typically assigned. These charges are usually either obvious, guessed, or obtained from QM calculations in gas phase. As a result they

can provide an inadequate description of phenomenon like polarization and screening effects observed in the condensed phase. Indeed, choosing the correct effective partial atomic charges is a major challenge for the development of accurate force fields for polar molecules.⁴⁴

There are other functions that have been incorporated into different FFs to help improve accuracy. These include the Morse potential for bonds,⁴⁶ the Urey-Bradley potential for angles,⁴⁷ QM/MM,^{48,49} etc.

There are several FFs currently used to perform simulations of condensed phase systems. They include: GROMOS,⁵⁰ Amber,⁵¹ CHARMM,⁵² Aмеoba (Polarizable FF),⁵³ Martini,⁵⁴ OPLS,⁵⁵ KBFF20,⁵⁶ MM2,⁵⁷ ReaxFF⁵⁸.

In this study, the KBFF20 was chosen as it is relatively new and needs further testing to establish possible strengths and weaknesses. Additional simulations using the Amber and CHARMM force fields were also performed to provide a comparison of some of the results.

1.3 KBFF20

KBFF20⁵⁵ is a non-polarizable, united atom, classical force field developed for the simulation of peptides and proteins that provides an improved description of the interactions in certain systems that are inadequately described by common force fields.⁵⁵ KBFF20 centers on Kirkwood–Buff (KB) theory, an exact theory of solution mixtures. The Kirkwood–Buff integrals (KBIs) for solutes that are analogs of amino acid side chains and/or backbone from simulation and experiment are compared to provide better description of the interactions in the system⁵⁹.

KBFF20 is parametrized by a comparison of KB integrals obtained from experiment and simulation using the KB theory. The KB integrals are expressed as,^{59,60}

$$G_{ij} \equiv G_{ji} = 4\pi \int_0^{\infty} [g_{ij}(r) - 1] r^2 dr \quad 20$$

where, G_{ij} represents the KB integral between species i and j , while $g_{ij}(r)$ represents the corresponding radial distribution function (rdf) in grand canonical ensemble. Experimental KB integrals can be generated by inversion of the KB theory as illustrated by Ben-Naim.⁶¹ These are exact and fundamental quantities describing the intermolecular distributions in solution. KBIs can be used as target data to ensure a protein FF is accurate at reproducing the underlying intermolecular distributions. If not, the parameters (usually partial atomic charges) can be adjusted until they are reproduced.⁵⁹ The KBIs are sensitive to changes in the FF parameters.

During the development of the KBFF20 force field, a variety of kinetic and thermodynamic properties of various systems were studied to determine how well it is reproducing experimental data. KBFF20 has also been used to test structural results which also provides information about the force field's accuracy in modelling systems and describing interactions. Smith et al., reported results for J-couplings for small peptides, simulations to determine the RMSD for larger proteins (residues 20-257 amino acids), together with other physical, kinetics and thermodynamics tests were also performed to evaluate and validate KBFF20.⁶² Mercadante et al., have used the ability of KBFF20 to reproduce experimental results for intrinsically disordered proteins (IDPs) to make a direct comparison with AMBER/CHARMM force fields by performing simulations on peptides rich in PPII-helical content using all three force fields.⁶³ Their results indicated that KBFF20 is a robust force field for sampling both IDPs and well-structured proteins.

Force fields generally require additional tests to be carried out to improve parameters describing molecular interactions in systems with the aim of making the force field more accurate.⁶³ This is one of the goals of this thesis.

1.4 Free Energy Calculations

The Free Energy (ΔG) is a thermodynamic quantity that describes the spontaneity or tendency of a process at a given temperature.⁶⁴ Free energies can be obtained from simulation, but they require additional considerations. These calculations are especially important for use in ligand-complex binding simulations to help predict binding affinities.^{65,66} Currently, it is still difficult to accurately determine free energy changes from simulation as inadequate sampling effects often arise from using finite lengths of simulation. Hence, care must be taken to monitor the quality of the results.

The free energy of a system can be computed under equilibrium and non-equilibrium conditions, this can be done by several methods.⁶⁷⁻⁶⁹ There are two main approaches to calculate these free energies computationally, namely thermodynamic integration and statistical mechanical perturbation theory.⁷⁰

1.4.1 Thermodynamic Integration (TI)

This approach uses a Hamiltonian that depends on a coupling parameter, λ , such that,

$$H(\lambda) = H_0 + \lambda(H_1 - H_0) = H_0 + \lambda\Delta H \quad 21$$

Hence, when $\lambda = 0$ the Hamiltonian is H_0 and one simulates using a system corresponding to the initial state A, while when $\lambda = 1$ the Hamiltonian is H_1 and one simulates a system corresponding to the final state B, intermediate values of λ correspond to simulations with an intermediate Hamiltonian.⁷¹⁻⁷⁵ This leads to a Hamiltonian, partition function and free energy that all depend on the coupling parameter. Hence, a free energy change between state A and state B can be determined

by performing simulations at a variety of different λ values and then numerically integrating the resulting ensemble averages according to equation 22.⁷¹

$$G^B - G^A = \int_0^1 \left\langle \frac{\partial H}{\partial \lambda} \right\rangle d\lambda \quad 22$$

1.4.2 Statistical Mechanical Perturbation Theory

Statistical mechanical perturbation theory is a method used to calculate the thermodynamic properties of a system by perturbing it from a known reference state.⁷⁶⁻⁷⁸ This technique, developed by Zwanzig,⁷⁶ is based on the principles of statistical mechanics to calculate the free energy difference of one state to another provided they are similar. The result for the Helmholtz free energy change is,

$$\Delta A = -kT \ln \langle \exp[-\beta \Delta E(X^N)] \rangle_0 \quad 23$$

$\beta = \frac{1}{kT}$ where k is the Boltzmann constant and T is absolute temperature. Here, $\Delta E(X^N)$ is the change in the configurational energy and the zero subscript represents configurational averaging over the ensemble of configurations of the initial reference state of the system. Similarly, for averaging using configuration generated for the final state one has,

$$\Delta A = -kT \ln \langle \exp[-\beta \Delta E(X^N)] \rangle_1 \quad 24$$

In principle, this technique can be applied to any change that occurs in the system.⁷⁸ However, for this approach to be numerically accurate, the difference in energy between the initial and final states of the system should be small, such that one state can be considered as a perturbation of the other. For the free energy perturbation method, one simulation is used to determine ΔA (or ΔG) unlike in thermodynamic integration that requires a series of independent

simulations at different coupling parameter values.⁷⁹ There are other approaches to determine/calculate free energy changes from computer simulation. These include Widom Particle Insertion,⁸⁰ weighted histogram analysis method (WHAM),⁸¹ and Potentials of Mean Force (PMF),⁸² etc. However, none of these additional approaches will be used here.

1.4.3 Free Energy Cycles

Free energy cycles are often employed to help circumvent computationally expensive ligand binding simulations. In principle, one could simulate a ligand in the presence of a protein and obtain the fraction of time the ligand was bound to the protein, compared to being free in solution, and that provides the binding (equilibrium) constant, and thereby the free energy of binding. However, that approach is extremely inefficient as the ligand may never “find” the actual binding site during the simulation times typically employed for biomolecular systems ($\approx$ ns- μ s). If one is only interested in relative free energies of binding between a series of similar ligands, the free energy cycle approach provides a way around this problem.

If a protein binds two ligands, A and B, the relative free energies of binding between ligand A and B can be determined using this scheme illustrated in Figure 3.

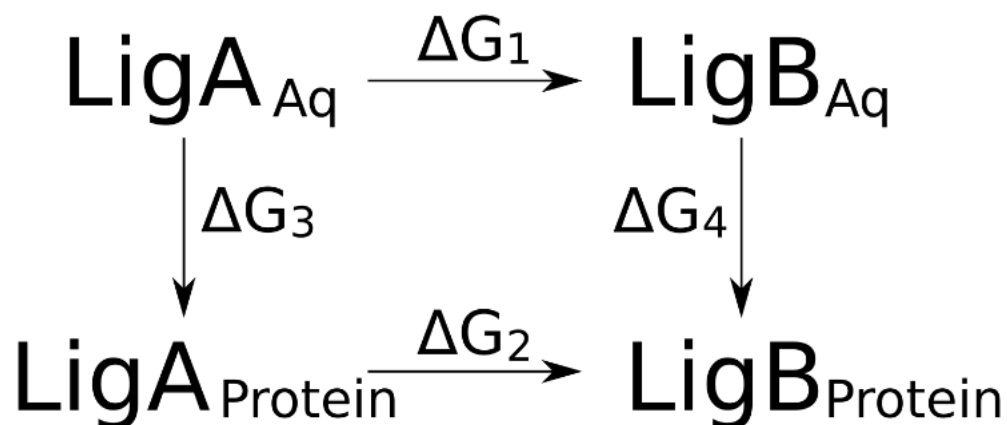


Figure 3 : A thermodynamic free energy cycle using alchemical transformations to determine relative binding free energies.⁸³

The steps corresponding to ΔG_3 and ΔG_4 represent the binding processes (Aq to protein) for Ligand A (LigA) and Ligand B (LigB). These are very difficult to achieve via simulation. Alternatively, the steps corresponding to ΔG_1 and ΔG_2 represent the free energy changes for mutating LigA into LigB in aqueous solution and when attached to the protein, respectively. These latter steps are impossible to do experimentally but can be determined computationally and are referred to as “computational alchemy”. The same approaches as outlined in Sections 1.4.2 and 1.4.3 can be used to determine the required free energy changes. Therefore, free energy for alchemical transformation from state A to B can be calculated by summing to zero around the thermodynamic cycle to give,

$$\Delta G_1 + \Delta G_4 - \Delta G_2 - \Delta G_3 = 0 \quad 25$$

and therefore,

$$\Delta \Delta G = \Delta G_1 - \Delta G_2 = \Delta G_4 - \Delta G_3 \quad 26$$

where $\Delta\Delta G$ is the relative free energy of binding calculated from the simulation of different ligands A and B bound to the protein and in solution, i.e the difference between free energy for transforming P:LigA to P:LigB (bound states) and the free peptides (LigA to LigB) in solution. This approach works best when both ligands are similar in structure.

1.5 OppA:Lys-X-Lys system

The OppA protein is a part of a superfamily of ATP-binding cassette transporters that are usually involved in the uptake of molecules by bacteria cells, and they are known to be virulent.⁸⁴ OppA plays a crucial role in transport of nutrients, minerals, etc., from the periplasm to a transmembrane transporter in Gram-negative bacteria which makes it a good target for antibacterial drugs. In the periplasm of these transporters there is a common feature involving a ligand-binding protein with high substrates specificity.⁸⁵ This provides the OppA protein with the ability to bind to ligands, 2-5 residues long, with highly variable structure. This property makes it a suitable system for correlating binding energy with structure for a wide range of ligands.

1.5.1 The Structure of OppA (ligand and unliganded)

The X-ray crystallographic resolved structure of OppA is available and indicates that the protein consists of three main domains which is unlike other binding proteins with only two domains.⁸⁵ Interestingly, OppA, undergoes conformational changes on ligand binding. It assumes an open conformation in the absence of a ligand where the binding site is fully exposed to solvent. Upon binding, the protein changes conformation to bury the ligand within itself, forming an internal cavity with stabilizing interactions between the ligand and OppA residues, and in the presence of multiple internal water molecules.⁸⁵

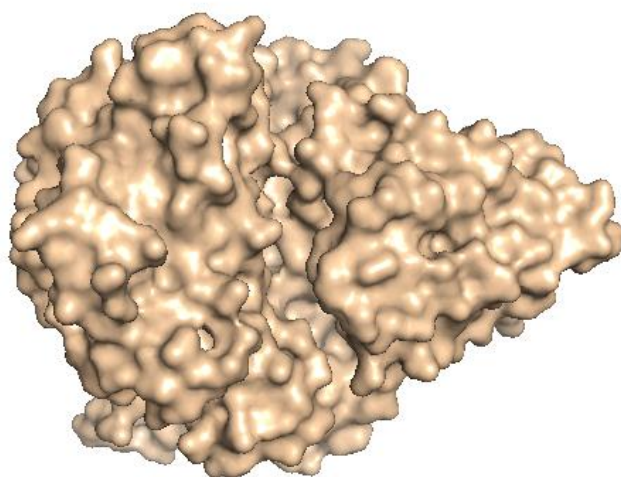


Figure 4 : Unliganded OppA structure showing its open conformational state in surface representation drawn with PYMOL.

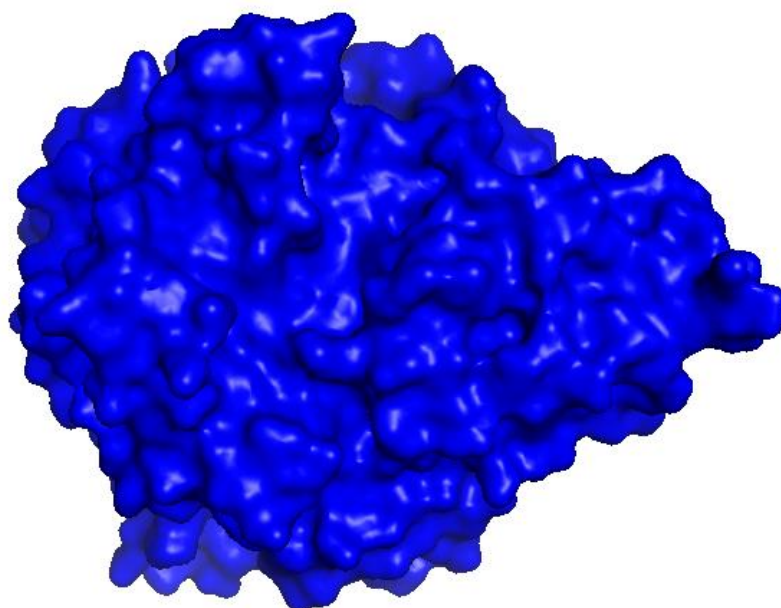


Figure 5 : OppA complex in its closed conformational state in surface representation drawn with PYMOL.

Experimentally, the binding of peptides of residue length between 2-5 to OppA has been studied extensively. Tame et al., reported results of the binding of OppA to a 3-residue long peptide of form LYS-X-LYS, where X is any of the 20 amino acid residues.⁸⁶ This was carried out using the isothermal titration calorimetry technique to determine thermodynamic properties of these protein-ligand complexes and eventually determining binding affinities for all the complexes. Interestingly, all these bound complexes have similar structures, i.e the tripeptide ligands (regardless of chemical composition) bind to same place in the OppA protein as determined by multiple crystal structures of the complexes.⁸⁶ This is unusual as the mutation of the ligand often brings about a change in binding pose of the ligand, or binding at a different site entirely, for most protein-ligand complexes. All crystal structures for the 20 natural amino acid OppA:KXX complexes are available in the protein databank.

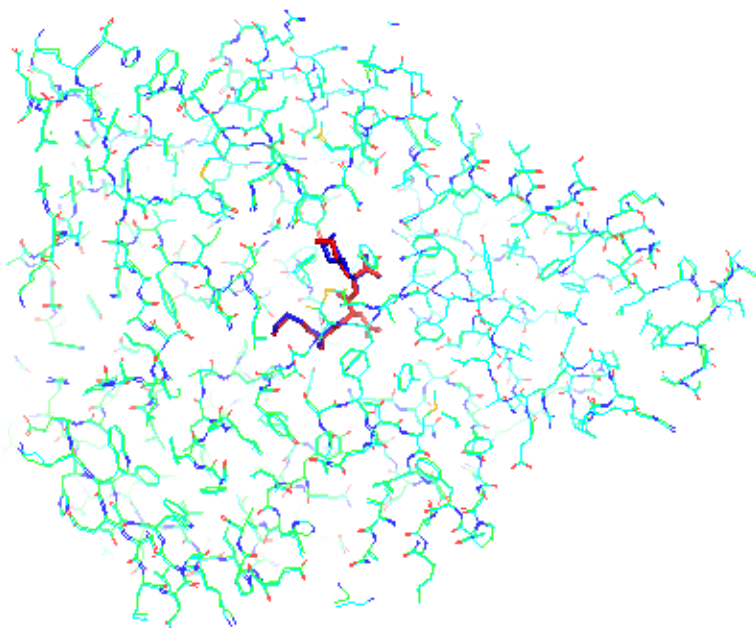


Figure 6: Aligned crystal structures of OppA:KAK and KDK with the ligands in blue and red respectively illustrating the same binding pose drawn with Pymol.

From a computational perspective, the availability of the above data makes this system a good fit for force field validation, and to investigate on some observations experimental techniques fail to capture, such as the role of internal water molecules on the binding of ligands to OppA. Oostenbrink *et al.*, have reported relative free energy calculations for 3 tripeptide ligands (KAK to KSK to KGK) bound to the OppA protein with an emphasis on the sampling of a range of conformations and the corresponding convergence behavior of the results.⁸⁷ They attributed the slow convergence observed for the transformations to the presence of water molecules in the binding site that slowly equilibrate, which they overcame by extending the simulation times until convergence was reached.

Gilson et al., in a bid to improve the convergence of water molecules within binding sites, developed a method that combines Monte Carlo and MD which made the calculated free energy more precise.⁸⁸ In their study the crystal waters were added and removed using a MC/MD method, and then compared the results with when the water molecules were kept fixed using a regular MD approach. The combined MC/MD method was reported to accelerate the equilibration of water molecules in the buried active site and bulk solvent, which led to more precise calculated relative binding free energy, and better hysteresis between forward and backward perturbations.

In summary, this system is an interesting one to study for several reasons. The availability of experimentally determined ΔH , ΔS and ΔG data for OppA bound to KXX, where X can be any of the 20 amino acid residues, provides a large dataset for comparison with simulation. Furthermore, understanding the mechanism involved in binding process of OppA to multiple ligands (2-5 residue long) and the corresponding role played by water molecules would prove beneficial in modern day drug design and discovery. However, this system has significant

complications. The presence of buried waters can be a challenge to sampling as they have been observed to equilibrate slowly.^{87,88} The change of conformation from an ‘open state’ when unliganded to a ‘closed state’ upon binding to a ligand can also cause sampling problems when transforming from one state to another. The free peptides in solution can undergo conformational changes that also have to be sampled adequately. The assignment of protonation states for binding site residues, especially the 11 HIS residues, is a challenge as they are in different environments. Finally, the presence of negatively charged ions observed in the active site of several crystal structures for some mutations (positively charged central residues) can potentially affect the calculated free energies.

1.6 Goals

The aim of this project is to calculate relative free energies of binding for the OppA:KXK system to determine the strengths and weaknesses of KBFF20 in representing unlike solute-solute protein interactions, for example Lys-Glu. We have also investigated ion-binding phenomenon in the system and the effect this may have on the calculated free energy.

Chapter 2 - Free energy calculations for ligand binding to the OppA protein

2.0 Introduction

Protein force fields constantly require testing and updating in an effort to improve their accuracy. KBFF20 is a relatively new protein force field that requires more extensive testing. In particular, while the force field was specifically developed to reproduce experimental data quantifying solute-solute and solute-solvent interactions, the solute-solute interactions studied involved small molecule analogues of the amino acid side chains and only probed the self interactions (equivalent to ASN to ASN, for example). Little to no testing of different amino acid side chain interactions were examined (such as ASN to SER, for example). Here, we provide preliminary studies to rectify that situation. The OppA binding protein is known to bind to multiple ligands (2-5 residue long) of varying compositions. Remarkably, the crystal structures of these complexes have been solved and experimental free energies for the OppA:KXX system have been determined and reported, see Figure 7.⁸⁶ Finally, crystal structure studies indicate that protein-ligand binding pose is same in all these complexes. Hence, this system provides a unique opportunity to study many ligands bound to the same protein in a quantitative manner. As the interactions change for different central ligand amino acids (X), one can also examine many different amino acid side chain interactions to determine if the corresponding binding free energies indicate possible problems with the force field.

In this study, relative binding free energies of OppA:KXX system were calculated by performing MD simulations employing the KBFF20 force field. The central residue 'X' included charged residues Lysine (K), Arginine (R), Aspartic acid (D) and Glutamic acid (E). Other

calculations involved the polar and non-polar residues: Asparagine (N), Cysteine (C), Serine (S) and Alanine (A), Valine (V), Leucine (L), and Phenylalanine (F).

To achieve this, we first need to consider two aspects of the free energy changes that require special attention. The first involves the addition (or loss) of a net charge during the mutation of one side chain into another (ASN to ASP, for instance). The second involves a special two dimensional correction map (CMAP) that is used in the KBFF20 force field to model the free energy surfaces for rotation around the ϕ/ψ dihedrals. This differs between amino acid residues and hence contributes to the overall relative free energy differences.

Free energy calculations involving charged species are an important facet of many fields, including biochemistry, chemistry, drug design and discovery, material science, etc. These charged species, such as ions or charged molecules, can play important roles in their local environment, especially in biological systems. They are known to have strong (electrostatic) interactions, causing significant changes in calculated free energies of the system.^{89,90}

Peptide	K_D (nM)	ΔG (kJ mol ⁻¹)	ΔH (kJ mol ⁻¹)	$T\Delta S$ (kJ mol ⁻¹)
KAK	56 ± 50	-41.1	20.1 ± 2.0	61.1 ± 0.9
KCK	75 ± 3.4	-40.6	7.9 ± 1.3	48.2 ± 1.4
KDK	5900 ± 1200	-29.8	8.1 ± 1.3	37.7 ± 0.8
KEK	150 ± 16	-38.9	11.3 ± 0.9	50.0 ± 1.1
KFK	53 ± 9.9	-41.5	22.0 ± 1.6	63.2 ± 1.6
KGK	1300 ± 460	-33.6	14.1 ± 0.36	47.5 ± 0.8
KHK	130 ± 43	-39.3	20.6 ± 3.0	59.8 ± 3.9
KIK	200 ± 56	-38.3	20.5 ± 3.8	58.6 ± 4.4
KKK	2900 ± 63	-31.6	39.4 ± 0.9	70.8 ± 0.8
KLK	1100 ± 180	-33.9	24.6 ± 0.9	58.3 ± 3.1
KMK	79 ± 17	-40.5	14.6 ± 0.8	54.9 ± 1.2
KNK	90 ± 37	-40.2	7.7 ± 0.2	47.9 ± 1.1
KPK	5200 ± 780	-30.1	16.6 ± 0.6	46.5 ± 0.4
KQK	37 ± 0.9	-42.4	11.4 ± 0.5	53.4 ± 0.5
KRK	1200 ± 490	-33.8	36.0 ± 7.3	69.7 ± 8.3
KSK	43 ± 5.7	-42.0	8.9 ± 0.7	50.6 ± 0.6
KTk	76 ± 4.9	-40.6	17.3 ± 1.7	57.6 ± 1.5
KVK	45 ± 25	-41.9	22.4 ± 0.4	64.3 ± 1.3
KWK	130 ± 11	-39.2	29.3 ± 1.0	68.2 ± 1.0
KYK	260 ± 43	-37.5	20.7 ± 3.9	57.9 ± 4.2

Figure 7 : Excerpt of the result table of calculated binding free energies of OppA:KXXK system reported by Tame et al.,⁸⁶

One common approach used to compute the free energy of a charged species is the thermodynamic integration (TI) method. This involves gradually changing the Hamiltonian of the system by using a coupling parameter, λ , to slowly change from an initial state to a final state, while keeping track of the free energy changes at each step. The free energy difference between the initial and final states can then be calculated by integrating the derivative of the Hamiltonian with respect λ .

A thermodynamic cycle is used to circumvent computationally expensive simulations in MD simulations when calculating relative binding free energies for binding a protein to two or more ligands. For transformations involving charged species (protein and ligands), the nonphysical insertion of the ligand into both the solvated protein and into free solution, and the calculation of the corresponding ligand-insertion free energies in nanoscale computational boxes simulated under periodic boundary conditions in which interactions described by lattice sum methods differs from simulating a macroscopic system under non-periodic boundary conditions with Coulombic electrostatic interactions, which would be ideal in a bulk situation.⁹¹ This results in finite size effects that need to be corrected for when calculating free energy changes in charged species.

Rocklin et al., in their article “Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects” proposed two correction schemes (numerical and analytical) to correct finite size errors on the charging component of the free energy calculation process leading to a more accurate calculated free energy.⁹¹ These finite size errors arise from four physical effects: Periodicity-induced net-charge interactions, periodicity-induced net-charge undersolvation, residual integrated potentials and discrete solvent effects.

2.1 Continuum Electrostatics

Continuum electrostatics represent the theory employed to describe electrostatic behaviors of molecules and macromolecules in solution.⁹³⁻⁹⁵ Many processes in biomolecules involve electrostatic interactions between solutes and solvents. These electrostatic interactions are described by the Poisson-Boltzmann Equation (PBE). The PBE is a partial differential equation that describes the electrostatic potential of a solvated solute (macromolecule) in terms of the charges on the solute, the dielectric constant of the solvent, and ionic distribution.

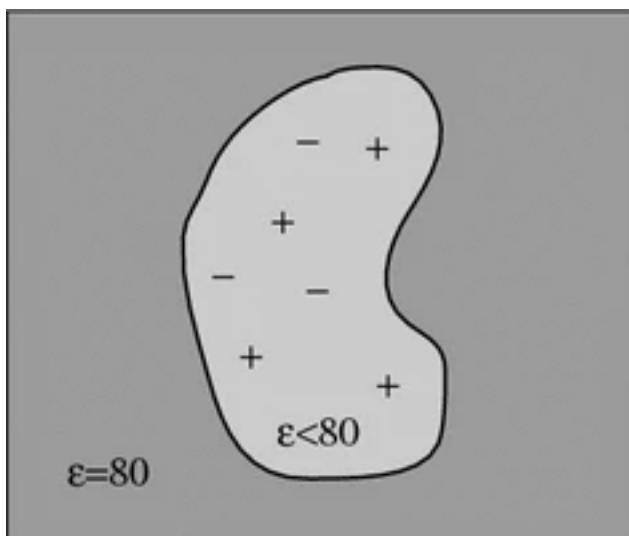


Figure 8 : Depiction of a macromolecule (light grey) in a solvent (dark grey) of dielectric constant of 80.⁹²

Considering a homogenous system with dielectric constant ϵ and no charges, the electrostatic potential $\psi(\vec{r})$ is given by the Laplace equation,⁹⁵

$$\vec{\nabla}[\vec{\nabla}\psi(\vec{r})] = 0 \quad 27$$

In the presence of a charge density $\rho(\vec{r})$ and a difference in dielectric constant $\epsilon(\vec{r})$ of solvent and solute, the electrostatic potential is obtained by the Poisson equation,

$$\epsilon\vec{\nabla}[\vec{\nabla}\psi(\vec{r})] = -4\pi\rho(\vec{r}) \quad 28$$

When considering non-homogenous systems, the above equation becomes modified to account for a dielectric constant that varies spatially throughout the system, especially at the interface between solute and solvent. The electrostatic potential all over the space is then given by,

$$\vec{\nabla}[\epsilon(\vec{r})\vec{\nabla}\psi(\vec{r})] = -4\pi\rho(\vec{r}) \quad 29$$

The Poisson equation can be applied to inherently account for the effect of monovalent ions that move within the solvent according to Debye-Huckel theory.⁹² This results in the (nonlinear) Poisson-Boltzmann equation (PBE),

$$\vec{\nabla}[\epsilon(\vec{r})\vec{\nabla}\phi(\vec{r})] + 4\pi \sum_s c_s(\vec{r})q_s e^{-\beta q_s \phi(\vec{r})} = -4\pi\rho(\vec{r}) \quad 30$$

where c_s is the concentration of salt in the bulk solution, q_s represents the unit charge, and β is the inverse of Boltzmann constant k_b times temperature. The PBE can be linearized by expanding the exponential. This equation can be solved numerically using a variety of methods including finite element methods, finite difference methods, and boundary element methods.⁹⁶

2.2 Correction map (CMAP)

In polypeptides, the central carbon atom ($C\alpha$) is flanked by two bonds with dihedral angles ϕ (φ) and ψ (ψ) that determine the possible rotation of the bonds without sterically clashing.^{97,98} This forms the basis of the Ramachandran Principle which says that secondary structures, such as alpha helices, beta strands and turns, are formed as a result of the conformational energies adopted to avoid steric collisions.^{97,99}

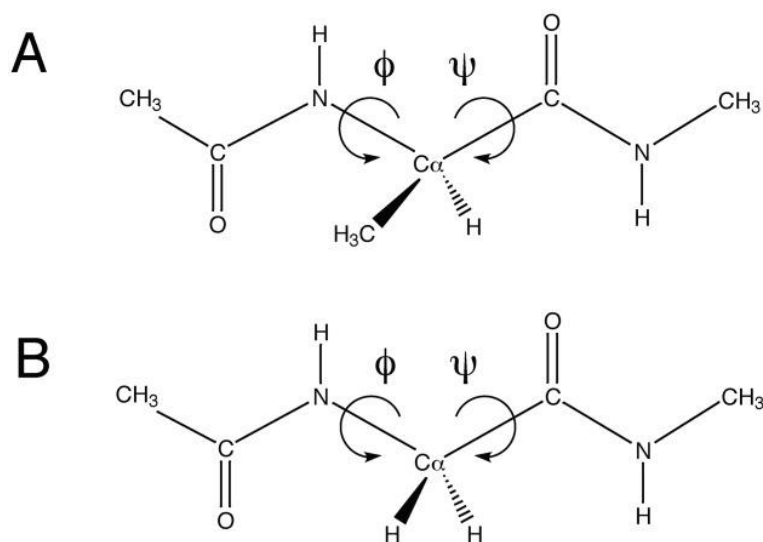


Figure 9 : Phi & Psi angles in (A) Alanine dipeptide and (B) glycine dipeptide.⁹⁸

G.N. Ramachandran and coworkers (in the 60's) computationally determined phi and psi angles by representing atoms as rigid spheres and determining their allowed conformations without steric collisions. Plotting these allowed angles gives a Ramachandran Plot, a 2D plot that shows the energetically allowed and forbidden regions for these backbone dihedral angles ϕ, ψ .¹⁰⁰

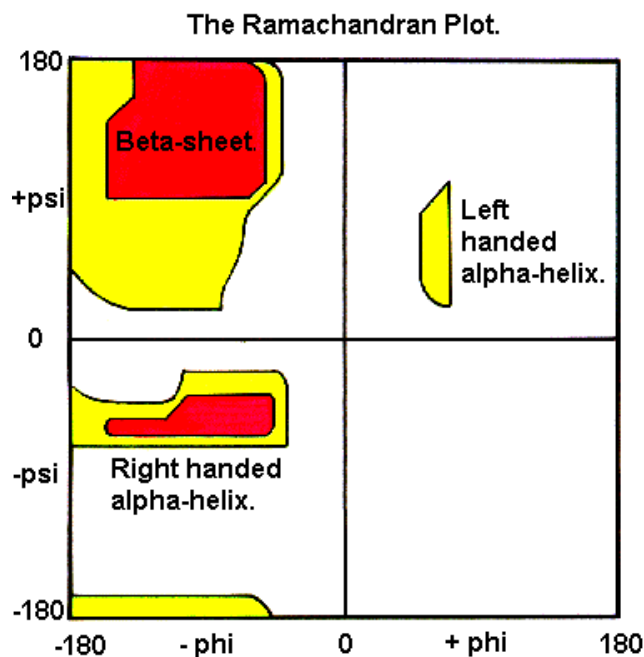


Figure 10 : Ramachandran plot showing the allowed secondary structures.¹⁰¹

Typically, the energy associated with the φ/ψ torsional angles of a peptide backbone is handled using a Fourier series comprising of cosine terms, the energetic effects of φ and ψ are combined.^{104,105} CMAP potentials represents the energies arising from the coupling between backbone dihedrals φ and ψ .^{102,103} Addition of CMAP potentials to force fields has been reported to improve the description of interactions in peptides and proteins in general by obtaining results that agree better with experimental data.¹⁰³ In KBFF20, CMAP changes from residue to residue such that there is a contribution to calculated free energy change of alchemically mutating ligands bound to OppA.

2.3 Softcore potentials

Soft core potentials are a type of potential energy function used in MD simulations to model interactions between molecules or atoms that have overlapping or close proximity in space due to

creation or annihilation of atom(s) when alchemically changing from a state A to B.¹⁰⁶⁻¹⁰⁸ In free energy calculations of this nature, there is usually a problem of a ‘singularity’ that arises when atoms disappear, and are therefore not interacting with other atoms, but the $\frac{\partial H}{\partial \lambda}$ appearing in Equation 22 then becomes infinite.¹⁰⁷ This is a well-known problem. One way to overcome the singularity problem is to adopt a shift of the function in the interatomic distance parameter r and a scaling of the non-bonded (vdw and electrostatic) interactions of the atoms as a function of the coupling parameter, λ resulting in λ -dependent potentials.^{106,108}

$$V_{A,soft}(r, \lambda, \alpha_{LJ}, \alpha_C) = \frac{C12_A}{(r^6 + \lambda^2 \cdot \alpha_{LJ} \cdot (C12_A/C6_A))^2} - \frac{C6_A}{(r^2 + \lambda^2 \cdot \alpha_{LJ} \cdot (C12_A/C6_A))} + \frac{q_{i,A} q_{j,A}}{4\pi\epsilon_0 \sqrt{r^2 + \lambda^2 \cdot \alpha_C}} \quad 31$$

where α_{LJ} is an adjustable parameter. In practice, transforming interacting atoms into dummy atoms when destroying atoms requires scaling of the van der Waals potential of the end state to zero.^{106,108}

This method is also referred to as separation-shifted scaling.¹⁰⁸ The effect of this is illustrated in Figure 11.

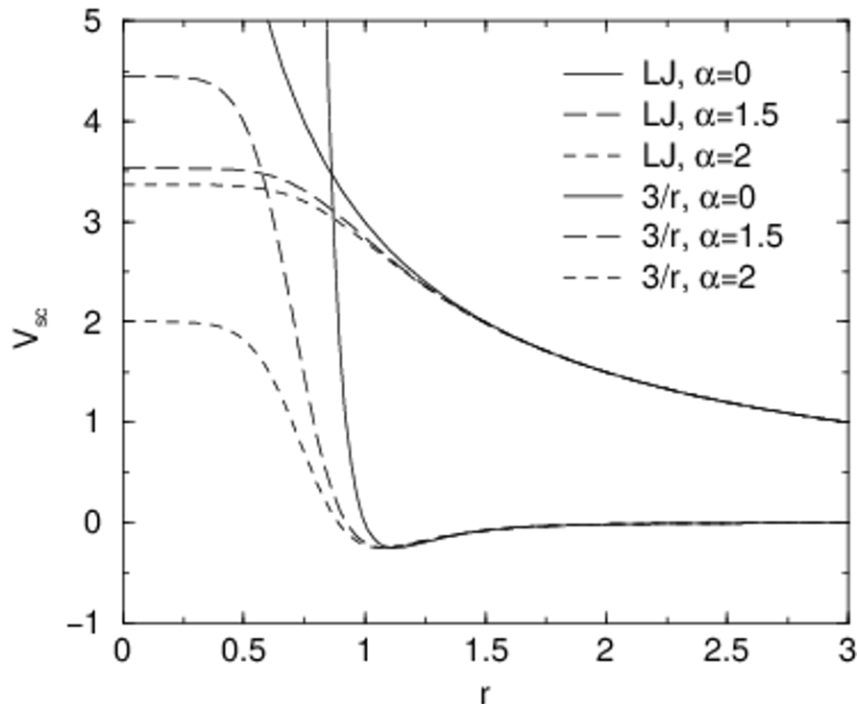


Figure 11 : Soft-core interactions at $\lambda = 0.5$, with $p = 2$ and $C_6^A = C_{12}^A = C_6^B = C_{12}^B = 1$.¹⁰⁸

2.4 Corrections involving charged species

Most biomolecular simulations involving explicit-solvent are performed under periodic boundary conditions (PBC) to limit surface effects.^{109,110} For the computation of electrostatic interactions in these systems, the coulombic interactions are determined using the exact periodicity of the system rather than that of an agreeable cutoff distance.¹⁰⁹ Though the former provides stable trajectories and is less-expensive computationally, it does result in finite size effects due to the enforcement of the system periodicity.^{109,111-112} In simulations where interactions are described by the Ewald summation approach, coupled with images of periodic boxes (as adopted here), this results in the new charge interacting with images of itself and this must be accounted for to minimize finite size effects.¹¹³

There are four main causes of error in the free energies of a charged system simulated with explicit-solvent and periodic boundary conditions.^{114,115}

1. The difference in polarization of the solvent around ion(s) in a system due to electrostatic interaction being described by electrostatic interactions compared to an ideal coulombic system. When cutoff truncation (CT) is used, interactions beyond cutoff parameters are ignored and this results in error in calculated charging free energies. The following equations are used to offset this effect on the charging free energy.

$$\Delta G_{1,1}^{CT} = -(8\pi\epsilon_0)^{-1}q_I^2(1 - \epsilon_s^{-1})R_c^{-1} \quad 32$$

$$\Delta G_{1,2}^{CT} = -(8\pi\epsilon_0)^{-1}q_I^2(1 - \epsilon_s^{-1}) \left[R_I^{-1} - R_c^{-1} + \sum_{k=-1}^{Na} ka_k R_c^{-k-1} \int_{R_I}^{R_c} dr r^{k+1} p(r) \right] \quad 33$$

$$\Delta G_1^{CT} = \Delta G_{1,1}^{CT} + \Delta G_{1,2}^{CT} \quad 34$$

where ϵ_0 is the permittivity of vacuum, ϵ_s is the permittivity of solvent, R_c represents the cut-off distance, r is the atom-based cut-off, k is a scheme dependent dimensionless coefficient, R_I is the ionic radius, and q_I is the ionic charge in e .

2. Periodic images of a system under PBC results in a difference in solvent polarization around an ion(s) compared to an ideal bulk system. An analytical correction is generated when the Born solvation free energy is compared to the solvation free energy of a spherical ion simulated in under PBC with the lattice-sum (LS) approach for electrostatic interactions.

$$\Delta G_2^{LS} = -(8\pi\epsilon_0)^{-1}q_I^2(1 - \epsilon_s^{-1})L^{-1} \left[\xi_{LS} + \frac{4\pi}{3} \left(\frac{R_I}{L} \right)^2 - \frac{16\pi^2}{45} \left(\frac{R_L}{L} \right)^5 \right] \quad 35$$

where $\xi_{LS} = -2.837297$ is the cubic LS (Wigner) integration constant.^{116,117}

3. Depending on the method, the potential at the ion center is affected by other ions present in the solvent, and the computed charging free energies will be different. When LS electrostatics are employed in a simulation under PBC the correction is given by,

$$\Delta G_3^{LS} = -q_I f \Delta \Phi_{ODL} \quad 36$$

$$\Delta \Phi_{ODL} = (6\epsilon_0)^{-1} \rho \gamma \quad 37$$

$$\gamma = \sum_i^{N_s} q_i r_i^2 \quad 38$$

Where ρ represents the number density of the solvent and γ is the quadropole-moment trace of the solvent molecule.

4. As with all MD simulations, force fields determine the accuracy of any calculation as it describes interactions in the systems under investigation.

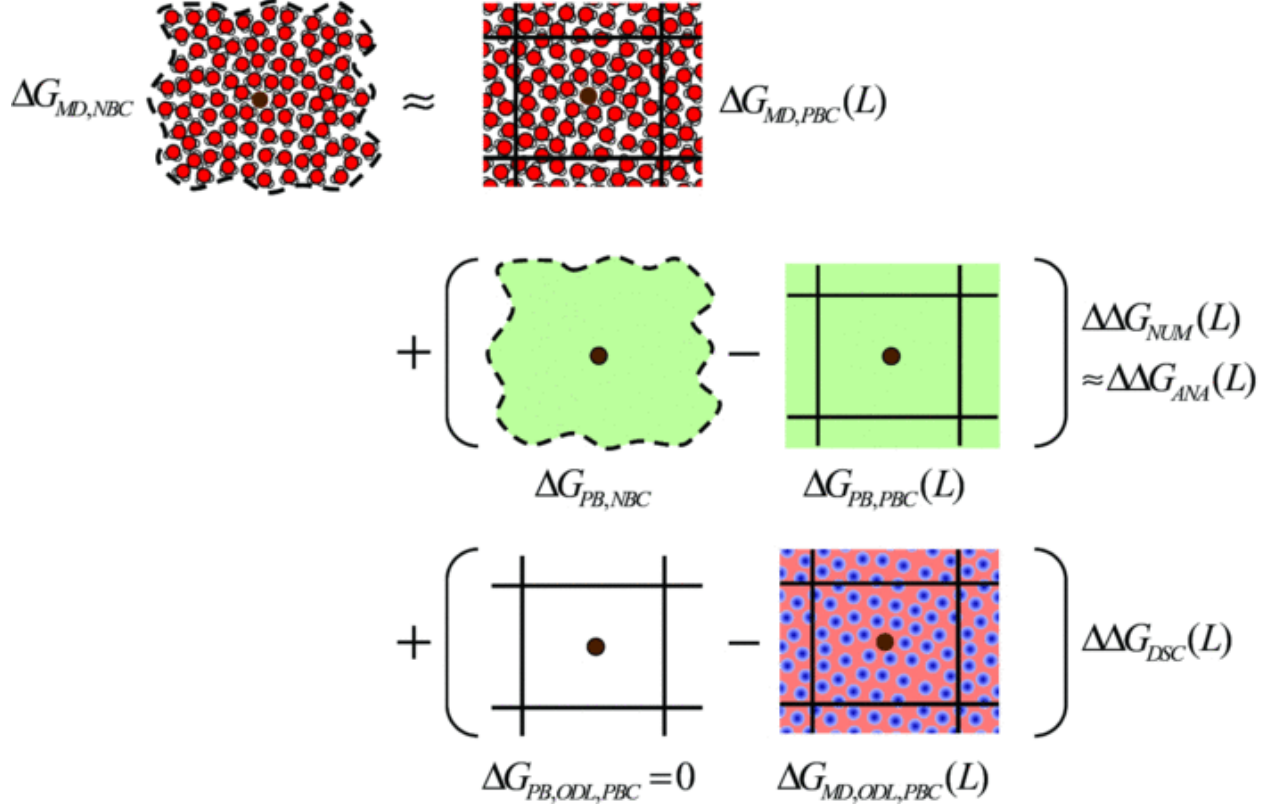


Figure 12 : A cartoon representation of the correction scheme from MD simulation performed with explicit solvent and periodic images to MD under non periodic boundary condition.⁹¹

Mathematically,

$$\Delta G_{MD,NBC} = \Delta G_{MD,PBC} + \Delta \Delta G_{ANA}(L) + \Delta \Delta G_{DSC}(L) \quad 39$$

$$\Delta \Delta G_{ANA}(L) = \Delta \Delta G_{NET} + \Delta \Delta G_{USV}(L) + \Delta \Delta G_{RIP} + \Delta \Delta G_{EMP}(L) \quad 40$$

where $\Delta \Delta G_{NET}$ represents correction for periodicity-induced net charge interactions, $\Delta \Delta G_{USV}$ corrects for periodicity-induced net undersolvation, $\Delta \Delta G_{RIP}$ denotes correction for residual integrated potential effects and $\Delta \Delta G_{EMP}$ is the empirical correction term.

2.5 Methods

All MD simulations were performed with GROMACS 2019.5 software package,¹¹⁸ with the KBFF20⁵⁶ protein force field and the SPC/E water model. Crystal structures of protein complexes (and ligand by extraction from the complex) were obtained from the protein data bank and were used as starting configurations in all MD simulations. The protein complexes and ligands were solvated in replicas of periodic rhombic dodecahedron (12 nm) and cubic (6 nm) boxes respectively. Background salt ions, Na⁺ and Cl⁻, were used to neutralize the charges on the protein complexes and free ligands in solution. These counter-ions were also used to achieve the desired physiological salt concentration of 150 mM. Four (4) of the five (5) charged residues LYS (+1), ARG (+1), ASP (-1) and GLU (-1) have well-defined protonation states at pH 7, as adopted for the simulations. However, the HIS sidechain has two protonation states, neutral and positive, with a pKa of 6.5 in solution, and hence the exact protonation state is dependent on the environment. OppA has 11 HIS residues with some on the surface of the protein being exposed to water while 4 of them are buried within the protein with and 1 His residue being in the active site close to negatively charged carboxylate groups. This informed the decision to protonate His in this study, including the ones on the surface.

In all MD simulations, a time step of 2 fs was employed with the leap-frog algorithm for equilibration and md production runs. Three 500 steps of energy minimizations were performed with steepest descent method while restraining the positions of the solutes and crystal waters. LINCS¹¹⁹ was used for constraining bonds in OppA protein and peptides while SETTLE¹²⁰ was used for the solvent molecules. Heating of the system to 300 K was performed in three steps performed for an equilibration of 1 ns. An additional equilibration of 10 ns was performed with

the weak position restraint on the C-alpha of the protein residues for the complex only, to ensure the simulated structure stays comparable to crystal structure.

For the production runs, NPT simulations of 10 and 50 ns at each λ value were performed for the free peptides in solution and protein-ligand complexes respectively. The simulations were performed for each of 21 λ values of 0.05 spacing between 0 and 1 to gradually change the ligand from one state to another (ligand A to B) alchemically. To evaluate hysteresis, the simulation was performed in both the forward and backward directions. For some of the complexes, 50 ns extensions (making a total of 100 ns simulation time) were done to improve sampling and obtain better hysteresis. The $\frac{\partial H}{\partial \lambda}$ values were extracted and used to generate the corresponding ensemble averages for each λ value, which were then integrated to obtain the final free energies.

A correction map (CMAP) analysis was performed to account for the potential arising from the dihedral coupling of the phi and psi angles of the central residue in the KXK tripeptide ligands. The initial simulations were (mostly) performed using the Alanine map, which is different from maps of other residues, and hence the difference in free energy for changing between maps is required. This correction was obtained using the free energy perturbation (FEP) method outlined in Section 1.4.2, where the change in energy corresponds to the energy difference between the two CMAPs.

For alchemical transformations where the central residue being changed is charged, i.e there is a change in charge resulting from the transformation, the free energy corresponding to the charging or decharging step was corrected for finite size effects. This was done using the analytical scheme proposed by Rocklin et. al which involves performing three Poisson-Boltzmann (PB) calculations under non-periodic boundary conditions with the APBS program,¹²¹ using configurations generated during the current simulations, to correct for periodicity-induced net-

charge interactions, periodicity-induced undersolvation, residual integrated potential effects, discrete solvent effects, and an additional empirical correction term.⁹¹

An alternative approach was employed to correct for finite size effects associated with explicit-solvent simulations where electrostatic interactions are described by lattice-sum methods in one of the systems (A to K). This approach involves simultaneously charging the ligand and discharging a counter-ion in the same computational box but at a large, and therefore non-interacting, distance from the ligand.^{122,123} This is referred to as the simultaneous counter-ion mutation approach.⁹¹

2.6 Results and Discussion

One of the complications of the OppA system is the presence of water molecules within the binding site. These water molecules play an important role in the binding of OppA protein to multiple ligands (2-5 residues) of varying chemical composition, which is a unique characteristic of this system. The buried water molecules equilibrate slowly, which can cause sampling issues that could affect the calculated free energy changes.^{87,88} In our preliminary tests, a 100 ns MD simulation was performed to observe the equilibration of water molecules using KBF20 force field. We observed that, on this timescale, water molecules that were within the active site at the beginning of the simulation were fully exchanged by bulk water molecules by the end of the simulation (see Figure 13). This signifies that water molecules can exchange between the bulk solvent and buried binding site using KBFF20, and the (re)equilibration of water molecules should not be too problematic.

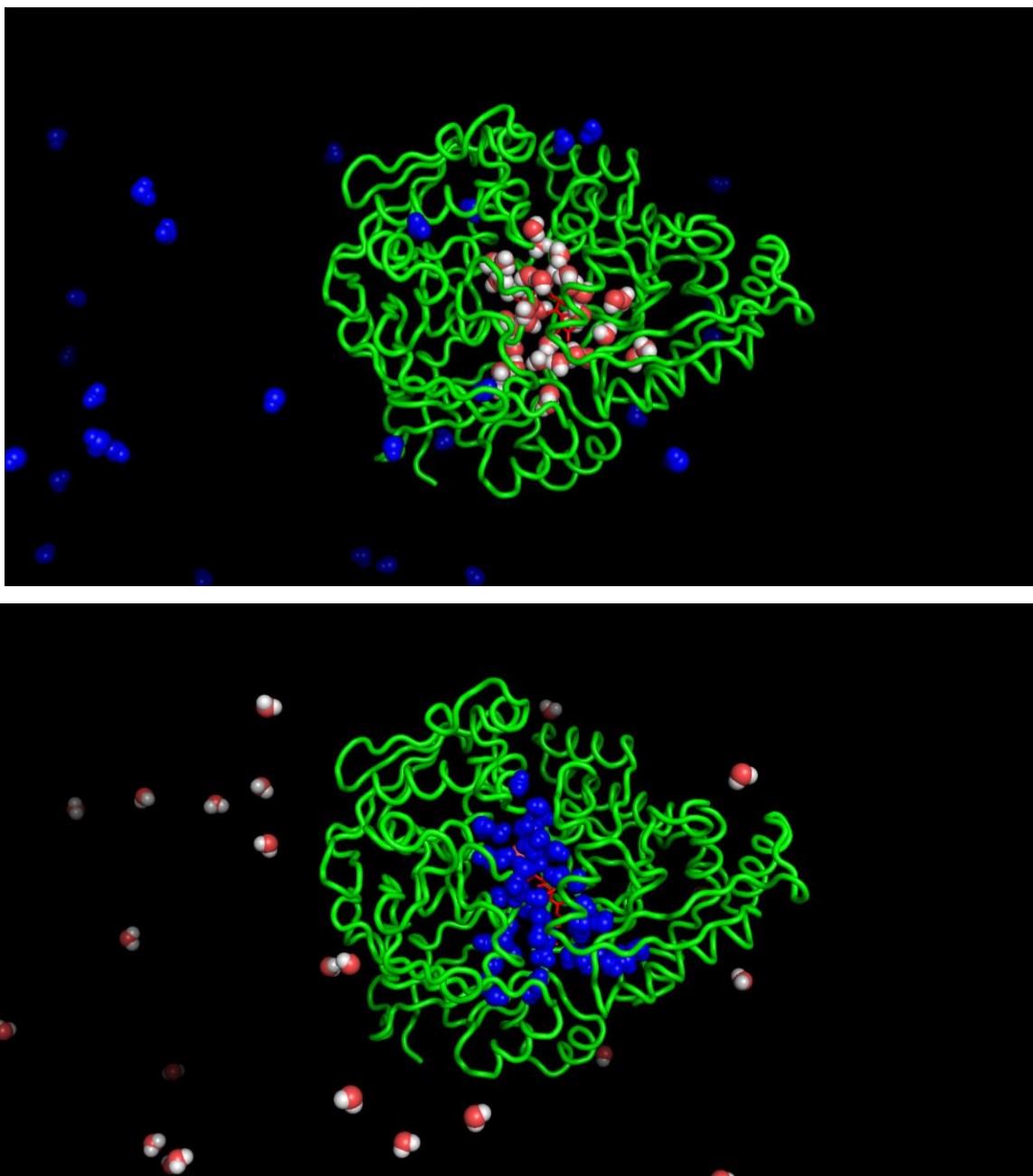


Figure 13 : Snapshots showing water molecules within the active site (red and white) and bulk water (blue) at the beginning (Top) and at the end (Bottom) of a 100 ns simulation.

In this study, alchemical transformations involving polar, non-polar and charged (positive and negative) central residues of the tripeptide ligands were performed using MD simulations. Thermodynamic Integration (TI) calculations were performed by analyzing trajectories of 50 ns length for the complex, and 10 ns length for free peptides in solution, for each λ -point. For some

transformations in the complex, simulations were extended by 50 ns to make a total of 100 ns to have better sampling and reduce the noise observed in these transformations.

For all transformations, forward and reverse simulations were performed to evaluate hysteresis, which provides information about convergence of the system and reproducibility of the KBFF20 force field. It should be noted that the reverse simulations did not start from the final λ -point from the forward simulations, i.e. they are completely independent simulations. The general approach is illustrated in Figure 14 for a mutation from KAK to KEK, i.e. Ala to Glu. The mutations appearing on the top line correspond to the Ala to Glu transformation, while the mutations on the second line correspond to the reverse Glu to Ala transformation. The vdws steps involve introducing (or removing) uncharged atoms corresponding to the additional atoms that form the Glu sidechain. The charge (decharging) steps correspond to free energy changes for mutating from neutral atoms to atoms bearing partial atomic charges characteristic of the charged Glu sidechain. The CMAP mutations correspond to the change in CMAP between Glu and Ala residues. An example of the numbers obtained from such mutations are provided in Table 1. Representative TI curves for the van der Waals steps of transforming KAK to KEK for forward and backward simulations in both complex and free peptide in the solvent are shown in Figures 15 and 16. TI curves for all the transformations studied here are available in the Appendix.

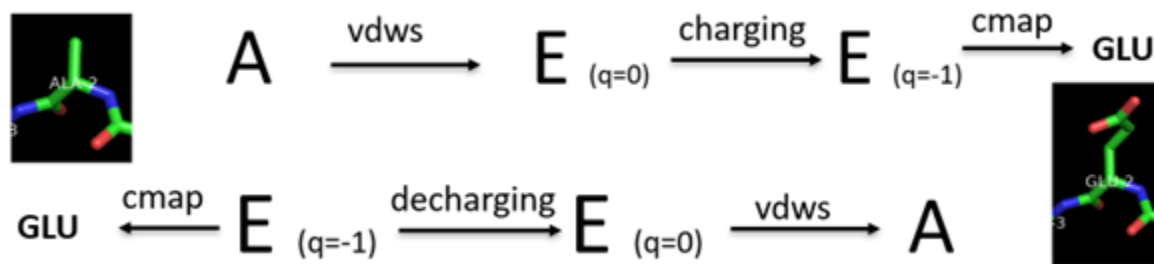


Figure 14 : Alchemical transformation route for KAK to KEK in both ligand in protein and free in solution and the reverse transformation.

MUTATION	ΔG CMAP	ΔG CHARGING	ΔG VDWs	ΔG CHARGING	ΔG CMAP	ΔG TOTAL
Com A $\rightarrow$ E	0	-371.2	10	0	0	-361.2
Com E $\rightarrow$ A	-1	0	-11.4	375.7	0	363.3
Pep A $\rightarrow$ E	0	-386.9	12	0	-0.1	-375
Pep E $\rightarrow$ A	-0.4	0	-12.7	386.2	0	373.1
	ΔG Com	ΔG Pep	$\Delta\Delta G$ MD,PBC	$\Delta\Delta G$ ANA	$\Delta\Delta G$ MD,NBC	
A $\rightarrow$ E	-362.25	-374.05	11.8	-5.1	6.7 ± 3.2	

Table 1 : Free energy differences (kJ/mol) for alchemical transformation AtoE and the reverse transformation.

The charging free energies were corrected as discussed in section 2.4, which involves PB calculations as well as using formulated expressions to obtain corrections due to periodicity-induced net charge interactions $\Delta\Delta G_{NET}$, periodicity-induced net undersolvation $\Delta\Delta G_{USV}$, and discrete solvent effect $\Delta\Delta G_{DSC}$.⁹¹ Table 2 shows the finite-size error corrections to the charging free energies in this study.

MUTATION	ΔG_{ANA} (Com)	ΔG_{ANA} (Pep)	$\Delta\Delta G_{ANA}$
D $\rightarrow$ N	77	82.2	-5.2
A $\rightarrow$ E	77.1	82.2	-5.1
A $\rightarrow$ D	76.8	82.3	-5.5
A $\rightarrow$ R	-76.1	-81.2	5.1
A $\rightarrow$ K	-76.2	-81.2	5

Table 2: A summary of the charging free energy corrections for charged transformations obtained using analytical scheme.

Figures 15 & 16 represent TI curves for the transformation of Alanine to Glutamic acid (negatively charged residue). These are the Leonard-Jones (LJ) steps of the simulation which involve growing and removing atoms only (without charges). The observed trend was better hysteresis in the free peptide in solution compared to the complex, as expected. This was the case for all the transformations, even for the charging and decharging steps involving charged residues.

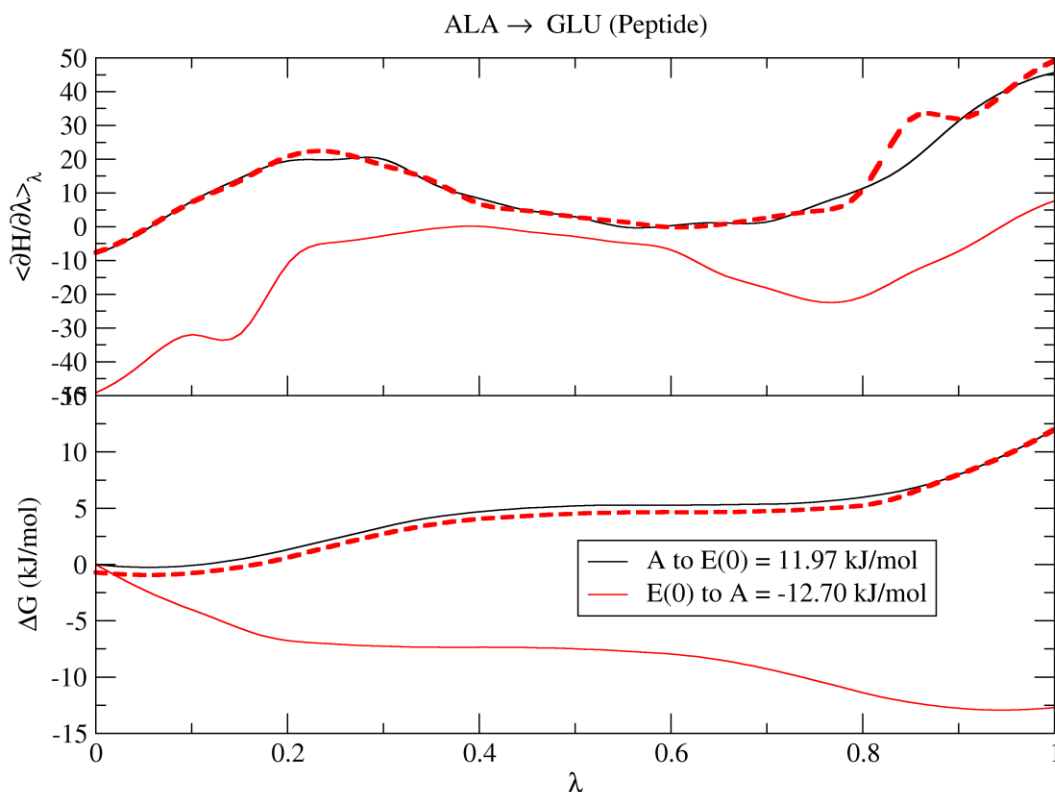


Figure 15 : Plots of $\langle \partial H / \partial \lambda \rangle$ and ΔG as a function of λ the transformation from Alanine to Glutamic acid (KAK to KEK) for the free peptide in solution obtained from 10 ns simulations of forward (black lines) and backward (red lines) processes. In both panels, dashed red lines are the flipped reverse transformations to compare with forward transformations and evaluate hysteresis.

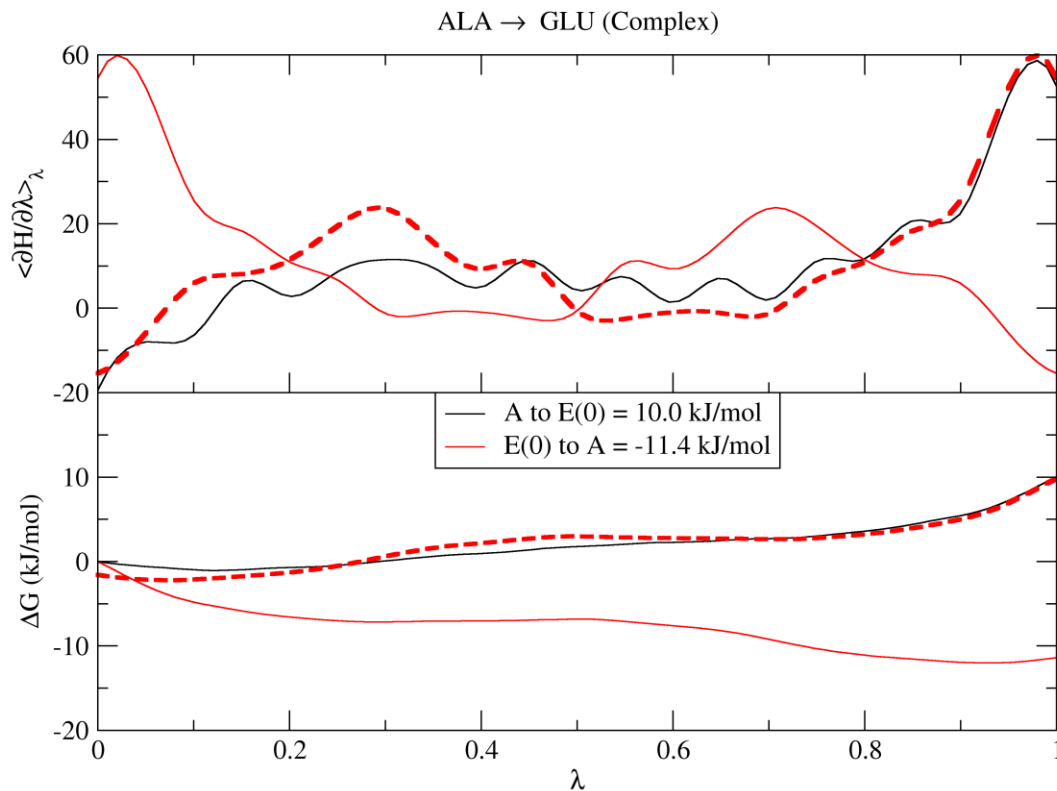


Figure 16: Plots of $\langle \partial H / \partial \lambda \rangle$ and ΔG as a function of λ for the transformation from Alanine to Glutamic acid (KAK to KEK) when bound to OppA as obtained from 50 ns simulations of forward (black lines) and backward (red lines) processes. In both panels, dashed red lines are the flipped reverse transformations to compare with forward transformations and evaluate hysteresis.

MD simulations for all the alchemical transformations were performed starting with the CMAP of one of the residues involved in the transformation. A FEP analysis performed on the resulting trajectories determine the free energy associated with mutating the CMAPs, which is then included in the calculated relative free energy for that transformation. This is performed for both bound tripeptides and free ligands in solution.

Figures 17 and 18 represent CMAP contributions to the relative free energy of alchemically transforming A to E for both forward and reverse simulations in peptide and the complex. These potentials are different from residue to residue, thus contributing to calculated free energies of mutating one amino acid residue to another. In this study, CMAP contributions range from -2.9 to 2.7 kJ/mol.

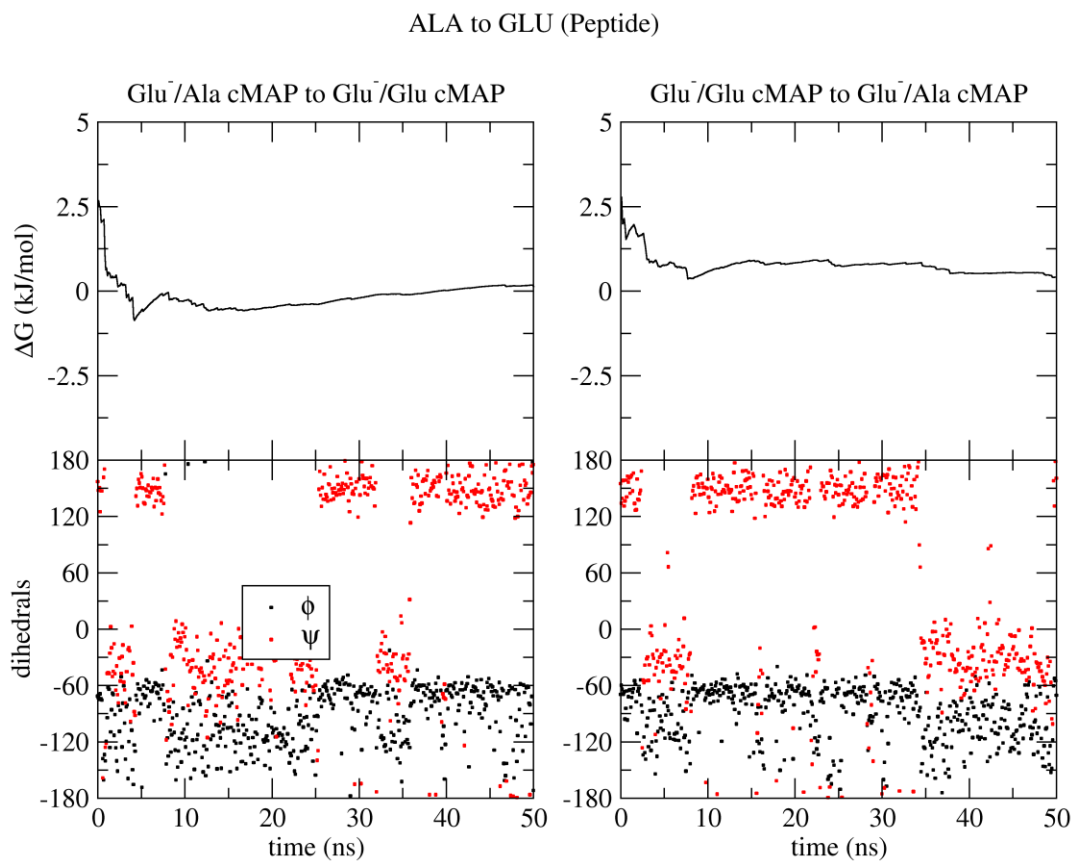


Figure 17 : correction map (CMAP) plot for Alanine to Glutamic acid (free peptide in solution) obtained from analyzing $\lambda = 1$ point of the forward reaction (left) and on the right is CMAP for the backward simulation, the trajectories of $\lambda = 0$ was analyzed.

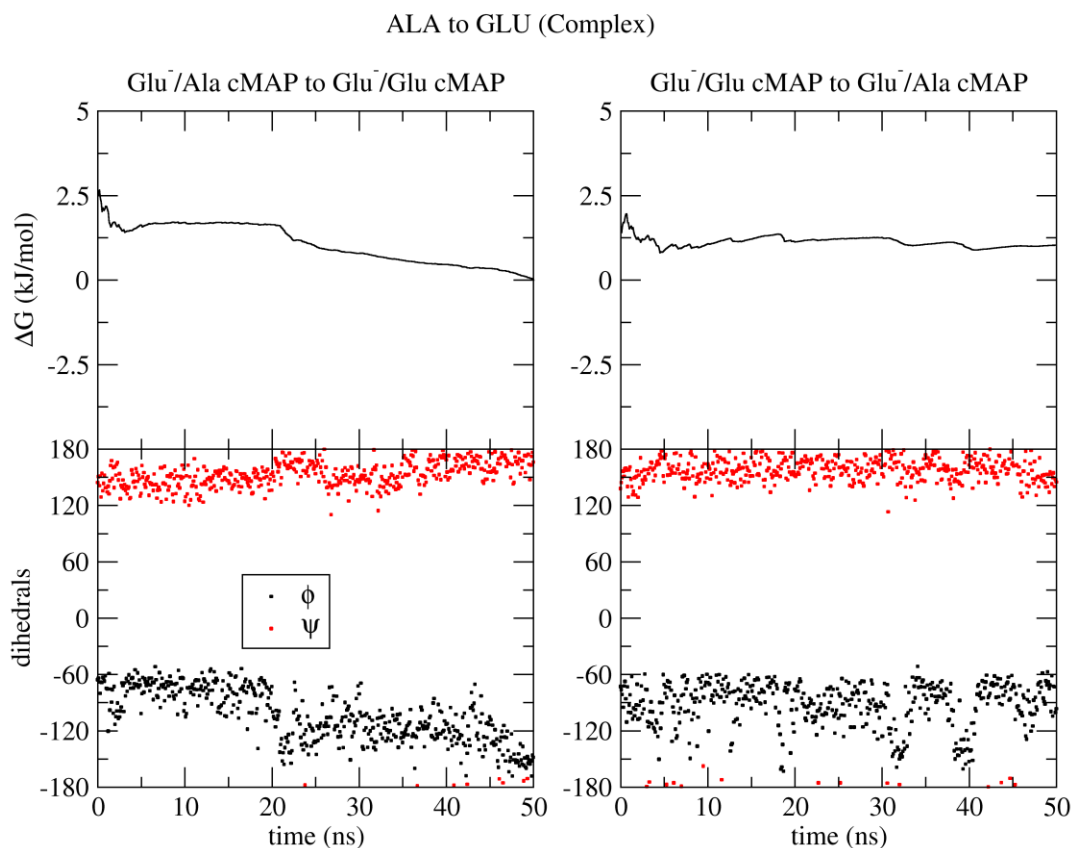


Figure 18: correction map (CMAP) plot for Alanine to Glutamic acid (complex) obtained from analyzing $\lambda = 1$ point of the forward reaction (left) and on the right is CMAP for the backward simulation, the trajectories of $\lambda = 0$ was analyzed.

All these different contributions and corrections (for charged transformations) are summed up and averaged for forward and backward simulations in both free peptide in solution and when bound to the protein as shown in Table 1.

MUTATION	ΔG_{Com}	ΔG_{Pep}	$\Delta\Delta G_{\text{Sim}}$	$\Delta\Delta G_{\text{Exp}}$
A $\rightarrow$ N	-132.6	-140.9	8.3 ± 1.4	1.2 ± 2.4
A $\rightarrow$ S	54.9	51.1	3.8 ± 2.7	-0.6 ± 2.2
D $\rightarrow$ N	167.4	172.9	-10.7 ± 2.3	-10.4 ± 1.1
C $\rightarrow$ S	28.6	28.3	0.3 ± 0.5	-1.3 ± 0.4
A $\rightarrow$ V	6.6	6.7	-0.1 ± 0.9	-0.5 ± 2.6
A $\rightarrow$ L	9.6	6.8	2.8 ± 1.7	7.4 ± 2.2
A $\rightarrow$ F	29.9	29.1	0.8 ± 1.6	-0.1 ± 2.2
A $\rightarrow$ D	-299.9	-311.2	5.8 ± 2.4	11.6 ± 2.3
A $\rightarrow$ E	-364.3	-374.1	6.7 ± 3.2	2.5 ± 2.2
A $\rightarrow$ R	-330.6	-330.9	5.4 ± 2.1	7.6 ± 2.4
A $\rightarrow$ K	-187.3	-193.9	16.6 ± 1.7	9.8 ± 2.2

Table 3 : A summary of calculated relative free binding energies ($\Delta\Delta G_{\text{binding}}$ (kJ/mol)) for all 11 alchemical transformations obtained using the KBFF20 force field compared to experimental data obtained from Ref. 86. Gold represents transformations involving polar residues, Dark Orange are non-polar groups, and Pink and Blue involve negatively charged and positively charged residues respectively.

The final results for the transformations considered here are presented in Table 3. All free energies were calculated from the average of the forward and backward simulations. The results are comparable to experiment in a few cases. Transformations involving nonpolar residues (except for A to F) seem to be in agreement with experiment and significantly better than for polar residues, the latter of which may be problematic due to many polar interactions between water in the active sites and side chains of ligands and protein residues close to the binding sites. Interestingly, The

D to N free energy difference, involving a polar residue transformation, agrees excellently with experimental data.

Surprisingly, there is large difference in experimental free energies of transforming A to D and A to E, despite the similarities between the E and D residues. The simulation results are in close agreement (though different than observed in experiment) as we would expect as both residues use the same σ , ϵ and q values for the carboxyl group. For positively charged residues, A to K is very different compared to experiment while A to R is comparable to experiment. The reason for this is unknown, but it should be noted that the experimental result for some transformations have large errors associated with them.

2.7 Conclusion

OppA-KXX is a very interesting and complicated system. How the oligopeptide binding protein A binds to multiple ligands (2-5 residues) of varying composition has been extensively studied and experimental data are available for thermodynamic properties of these systems. In this study, the relative binding free energies of OppA binding to 11 tripeptide ligands were calculated using MD simulation and employing TI method of free energy calculations to evaluate and test the KBFF20 force field parameters with respect to solute-solute interactions. Several tripeptide ligands of the form LYS-X-LYS were evaluated, where X was varied between polar, non-polar and charged residues. This system is known for water-mediated binding resulting in slow equilibration of water molecules in the active site, but with longer simulations this challenge was overcome. This could be said to have improved sampling as hysteresis between the forward and backwards simulations were reduced on the extension of some of the simulations.

The final results are mixed with some transformations giving good agreement with experiment, while several are in poor agreement. However, these results should be considered exploratory as we have not addressed important factors such as the protonation state of the buried HIS residues, and the possible role of ion binding close to the binding site. In conclusion, the calculated relative binding free energies for OppA:KXX system studied using KBFF20 protein force field is presented in this report. However, for full comparison with experimental data further work still needs to be done.

Chapter 3 - Local electroneutrality and the effect on free energy calculations

3.0 Introduction

Electroneutrality is a fundamental principle that refers to the balance of positive and negative charges within a molecule or system, i.e. the net charge of a system is zero. Electroneutrality is attained by charged species (ions) in solution. In biological systems, ions play a crucial role in maintaining electroneutrality as well as in the general functionality of these systems such as in protein folding and stability,¹²⁴ membrane transport,¹²⁵ enzyme catalysis,¹²⁶ and many more processes.

The Debye-Hückel theory provides a theoretical framework for understanding the behavior of electrolytes in solution but the theory is limited to low ionic strengths,¹²⁷ much lower than physiological salt concentrations of 0.15 M. The theory explains the deviation from ideality in ionic solutions due to strong electrostatic interactions between ions and was foremost in the development of our current picture of ion atmospheres. The ion atmosphere, or related electrical double layer, refers to the phenomenon where an ion (or surface) is surrounded by a “cloud” of counterions of oppositely charged particles due to the electrostatic interactions between the ion and surrounding charged species. The extent of the ion atmosphere is determined by the Debye length which is related to the total ionic strength. A consequence of the ion atmosphere is neutralization of the ion charges over finite distances.¹²⁸

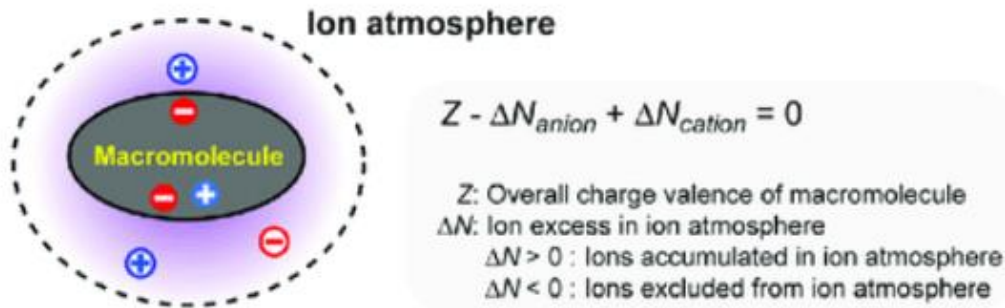


Figure 19: Negative ion surrounded by positively charged ion to illustrate ion atmosphere.¹²⁹

Electroneutrality can be considered in two ways. First, one has global electroneutrality where the number of positively charged ions (cations) matches the number of negatively charged ions (anions) in the whole system of interest. Second, one also has local electroneutrality where small regions of space, larger than the Debye length, are neutral on average.¹³⁰ The latter indicates that biomolecules are surrounded by local ion atmospheres that result in overall net neutral regions of solutions centered on the biomolecule. Consequently, when the charge on a protein or ligand changes there will be a rearrangement of the local ion atmosphere to regain local electroneutrality.

In the previous chapter we performed free energy calculations for the OppA:KXX system. In several cases, the charge of the ligand was mutated to determine relative free energies of binding. Consequently, the ion atmosphere will have changed accordingly. For the mutations performed in bulk solvent and 0.15 M NaCl we can imagine that the changes in ion atmosphere will occur quickly and easily. However, in the case of the protein-ligand complexes the situation is more subtle. The ligand is buried in the bins site and, while there are many solvent molecules also in the binding site cavity, it is not clear that any ions are also present inside the cavity. Hence, the rearrangement of the ion atmosphere on changing the charge of the ligand could be severely limited. It is unclear to what extent this affects the simulated free energy changes.

An examination of the experimental crystal structure of the OppA:KKK system obtained in the presence of background salt clearly indicates an has several acetate ions close to the active site. The acetate anions interact with several residues of OppA, including HIS-371 (Histidine), and also with the positively charged tripeptide ligand KKK. Analysis of our previous simulations (see Chapter 2) performed for this same system, but using a NaCl electrolyte, indicated binding of the chloride ions in the same region. An illustrative snapshot of the binding of acetate to OppA:KKK observed experimentally and the binding of chloride ions observed by simulation is provided in Figure 20.

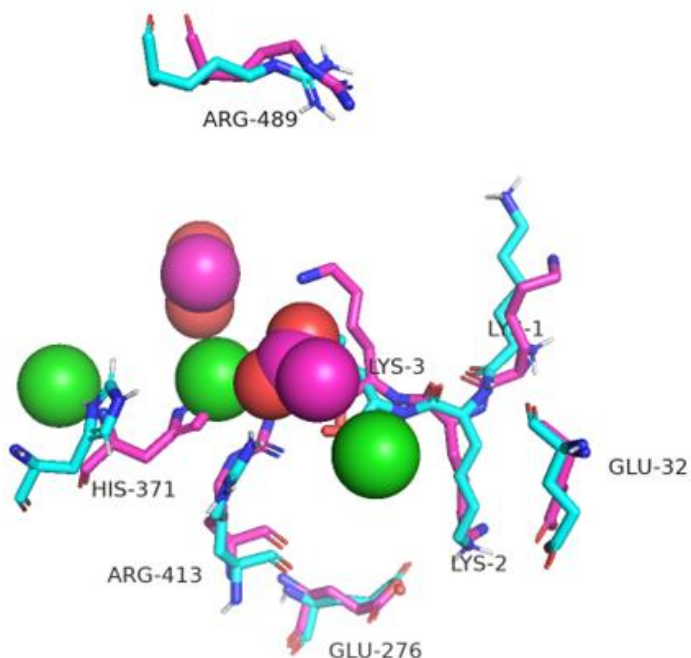


Figure 20 : Charged residues in the active site of the OppA:KKK complex experimental structure (Magenta) aligned with a simulated structure (cyan) showing acetate ion (Magenta & Red) and chloride (Green) ion binding in both structures.

Considering the above observations, we decided to investigate the nature of the ion atmosphere and possible binding to OppA in the presence of different charged ligands KXX using simulation. As ion binding to proteins may be force field dependent, we also performed simulations with two alternative force fields, in addition to KBFF20, to establish if any such dependence exists. Finally, we examined the if, and/or to what extent, the free energy changes determined in Chapter 2 may depend on the presence of nearby charged ions.

3.1 Methods

MD simulations were carried out on the systems indicated in Table 4. The primary force field used in this study was KBFF20⁵⁶ as implemented in the GROMACS software package. The corresponding OppA:KXX crystal structures were used for the starting structure in all simulations. For the simulation of the unliganded protein, the coordinates of the ligand were removed from the OppA:KGK complex. A relatively high salt concentration of 2 M was used in all simulations except for one. A simulation of OppA:KKK in 0.15 M NaCl was used as a control to determine if the ion-binding observed was due to the high number of ions (Na⁺ and Cl⁻) present in our systems.

In all MD simulations, a time step of 2 fs was employed with the leap-frog integrators used for the equilibration and production runs. Three 500 steps of energy minimizations were performed with steepest descent method while restraining the positions of the solutes and crystal waters. LINCS¹¹⁹ was used for constraining bonds in OppA protein and peptides while SETTLE¹²⁰ was used for the solvent molecules. Gradual heating of the system to 300 K was achieved through three steps equilibration of 1 ns at successively higher temperatures. An additional equilibration of 10 ns was performed with the weak position restraint on the C-alpha of the protein residues, to maintain the simulated structure close to the crystal structure.

Simulations of the OppA:KKK complex were also performed with two other protein force fields to provide a direct comparison with KBFF20. All simulations were performed for 100 ns and simple distance calculations between all free anions and cations to all the charged residues on the protein were carried out using GROMACS.

3.2 Ion Binding to OppA:KXX Complexes Using KBFF20

The analysis of ion binding involved the distance measurements of all free ions (positive and negative) that were within 0.5 nm of the charged pocket residues of OppA. These “pocket residues” were determined by following the network of water molecules in the active site and identifying the residues in contact with these water molecules using two separate Fortran codes. Using this approach, seven positively charged residues (ARG-17, ARG-41, HIS-161, HIS-371, ARG-404, ARG-413, and ARG-489) and five negatively charged residues (GLU-32, GLU-36, ASP-276, ASP-419 and ASP-505) were among the residues identified as pocket residues. This indicates an overall +2 charged binding pocket.

In Figure 21 we illustrate the time histories for several ions that contact the pocket residues during the simulation of OppA:KEK in 2 M NaCl solution. We observed one chloride bound to HIS-371 for the duration of the simulation, while a second chloride ion binding more transiently to residue K-3 of the bound ligand. Binding of sodium cations was also observed to Glu-2 of the ligand. Clearly, ions binding is possible under these conditions using the KBFF20 force field. Subsequently, we simulated a series of ligands in an attempt to establish if there were significant differences between them, especially if the total ligand charge varied. A summary of the results of this study is presented in Table 4. Though the numbers of chloride ion(s) in the active sites of these complexes differ, they all reflect the need to provide a neutral cavity to some extent. The number of chloride ion(s) present close to our in the active site do not match exactly the total charge on

the systems in most cases. This could be due to local electroneutrality where other charged residues on the protein could be involved. Nevertheless, the presence of significant ion binding, especially by chloride ions, suggest that the local ion distribution may affect the binding free energies to some (unknown) degree.

A simulation of the OppA:KKK complex in a lower (0.15 M) salt solution resulted in less ion binding. This could be due to the small number of ions in the system which would increase the time required for ions to find any favorable binding sites. Alternatively, the lower ion concentration indicates a larger Debye length which would indicate electroneutrality would need to be satisfied over larger distances and could therefore involve more bulk solvent ions.

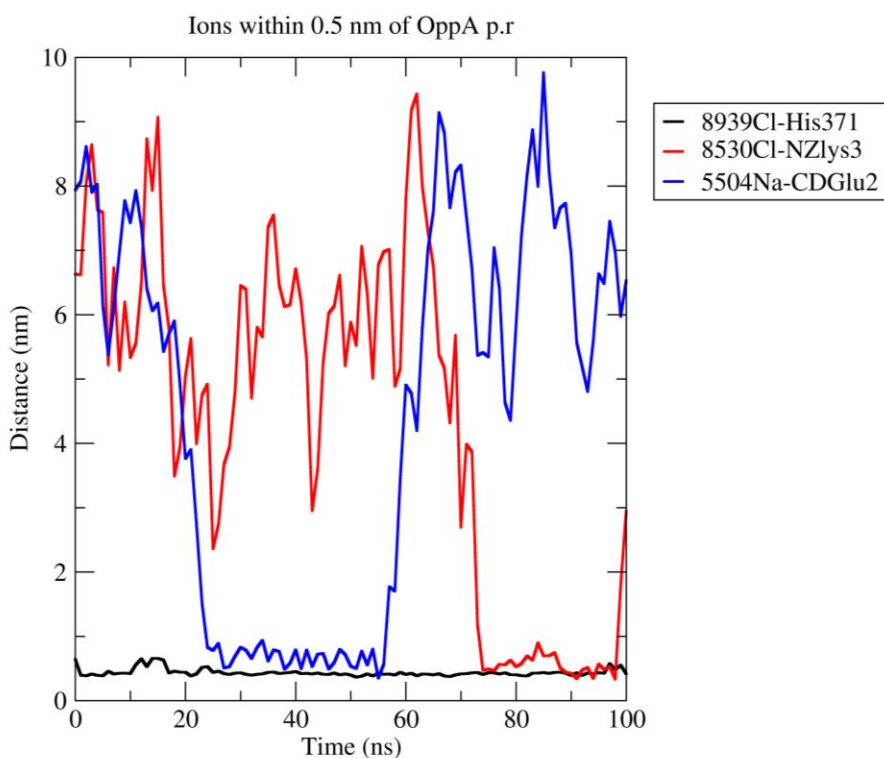


Figure 21 : Ion residue distance time histories for the simulation of OppA:KEK in 2 M NaCl solution. The distances for any ions that came within 0.5 nm of any charged pocket residues (p.r) are shown. Black and red represent chloride ions within 0.5 nm of charged pocket residues, while blue represents sodium ion in close contact with a negatively charged Glu acid residue.

System	Binding pocket charge	Ligand Charge	Total Charge	# of Cl ions from Sim
KGK (no ligand)	2	0	2	2
KGK	2	2	4	1
KDK	2	1	3	2
KEK	2	1	3	1
KKK_KBFF20	2	3	5	3
KKK_Charmm	2	3	5	1
KKK_Amber	2	3	5	0
KKK_0.15M	2	3	5	1
KRK	2	3	5	2

Table 4 : Summary of number of Cl ions in the active site for different systems and force fields.

3.3 Ion Binding to OppA:KXK Complexes Using Different Force Fields

Since ion binding was observed in our simulations using the KBFF20 force field, this suggests that ions binding may play a role in the ligand binding free energies. Before examining this issue, we wanted to determine if other common protein force fields also display ion binding. Additional simulations of the OppA:KKK complex in 2 M NaCl were performed using the Amber and CHARMM force fields. The results are also included in Table 4. In all systems except OppA:KKK simulated with Amber, ion binding was observed after a timescale of 100 ns. In the systems simulated using KBFF20, there is an appreciable number of negatively charged ion compared to CHARMM and Amber. This can be attributed to the method of parametrization of the ion parameters in the different force fields. In KBFF20, the ion parameters are obtained in a similar fashion to the amino acid residue parameters by using experimentally extracted KB integrals. This provides some consistency between the protein and ion force fields, which is not

the case for CHARMM and Amber approaches. As the observed chloride ion(s) in the active site of the simulations performed with KBFF20 agrees with the experimental crystal structure of the OppA:KKK complex, where acetate ions was captured in the active site, it seems possible that the KBFF20 results are more representative of the real system.

3.4 The Effect of Bound Ions on Free Energy Calculations

The presence of ion(s) in the binding site of the complexes, especially the OppA:KKK system, using the KBFF20 force field led us to investigate the possible effect of bound ions on calculated free energy change. MD simulations were performed for decharging the OppA:KKK system in the presence and absence of a chloride ion in proximity (within 0.5 nm) of the active site of the complex. This was achieved by restraining a chloride ion in contact with HIS-371 which is close to the active site. This structure was then used to populate the 21 λ windows from which the free energy change is obtained. The simulation was set up as discussed previously in the ‘Methods’ section (2.5).

The resulting $\frac{\partial H}{\partial \lambda}$ plot is shown in Figure 22. The results look very similar as the free energy for decharging the central Lys residue is so large and positive. The effect of the presence of ion close to the active site of the OppA:KKK complex resulted in a 4.3 kJ/mol difference in comparison to when the active site is free of any ions. This is less favorable compared to when there are no ions in the binding site. This can be considered a substantial difference in the calculated relative free energy for removing the charge in this system, attributed to the presence of the chloride in the active site, especially as many transformations result in an overall free energy difference that are less than this in magnitude (see Table 3).

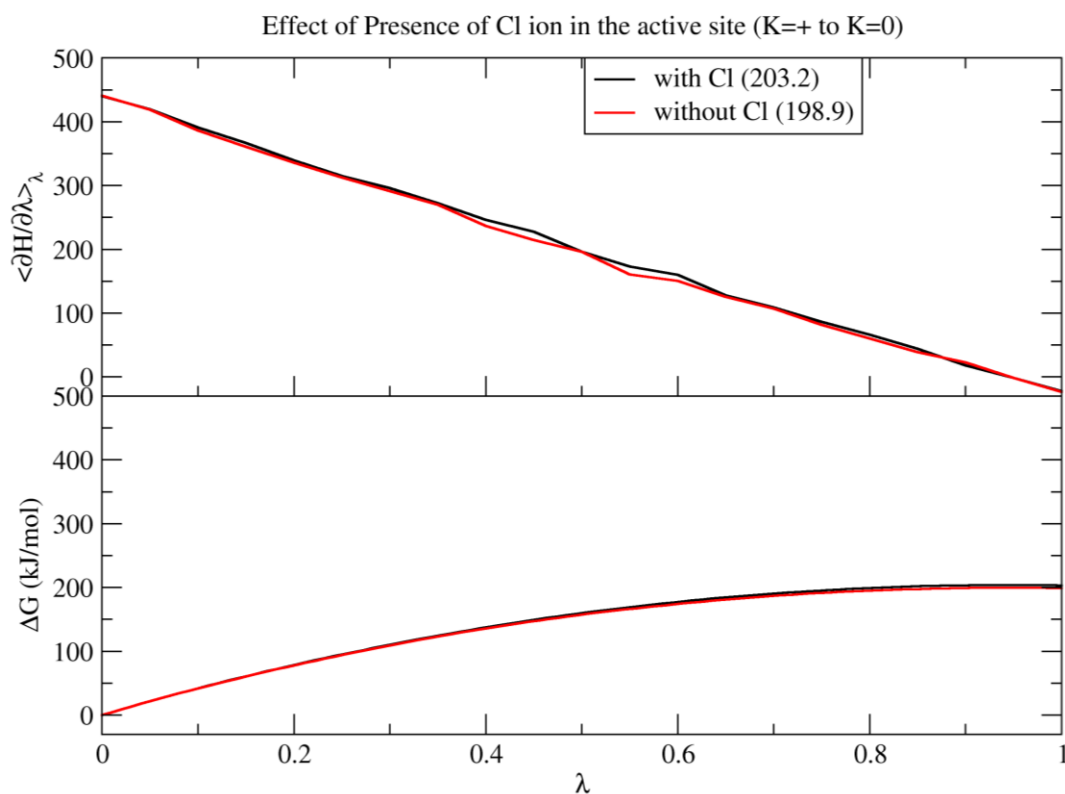


Figure 22 : Plots of $\langle \partial H / \partial \lambda \rangle$ (top) and ΔG (bottom) as a function of λ for K=+ to K=0 with (black) and without (red) chloride ion in the active site.

3.5 Conclusions

In this study, we observed the binding of chloride ion from bulk solution to the positively charged pocket residues. Analysis of trajectories from the MD simulations also showed that chloride and sodium ions migrate to charged residues/regions in the active site of many of the OppA:KXX complexes investigated. These observations can be attributed to the probable interaction of ions with charged residues in the active site to achieve local electroneutrality. Hence, local electroneutrality is associated with the chloride ions present in the overall positively charged pockets of the OppA:KXX complexes which was observed after 100 ns in all the simulations

except the one using the Amber force field. Ion binding results in less favorable binding of the ligand to protein in the OppA:KKK system upon investigating the effect on free energy for decharging the system resulting in a substantial difference in comparison to ions not being in the active site. Clearly, further work needs to be done to analyze the results of these simulations to better understand how electroneutrality is achieved in these systems. However, the binding of ions represents a new complication for this system, which could hinder the comparison with experiment of free energy differences determined using any force field.

Chapter 4 - Summary and Future Directions

The free energy calculations in the previous chapters can be considered exploratory in nature. The results indicate that reasonably precise free energy calculations can be obtained for this system. However, the accuracy of the KBFF20 results cannot be completely ascertained without additional studies. In particular, a deeper analysis of the convergence properties of the different free energy changes is required. Furthermore, the role of the His residue protonation states in the active site has not been investigated, while the presence of ions in the vicinity of the ligand binding site raises questions regarding the nature and representation of the ion atmosphere in this system. Nevertheless, while the OppA:KXX systems represents a challenge for current free energy approaches, the ability to study a wide variety of different amino acid side chain interactions makes this a very attractive system for testing the accuracy of current force fields, and the role of water in protein-ligand binding.

References

1. Hansson, T., Oostenbrink, C., & van Gunsteren, W. (2002). Molecular dynamics simulations. *Current opinion in structural biology*, 12(2), 190-196.
2. Hospital, A., Goñi, J. R., Orozco, M., & Gelpí, J. L. (2015). Molecular dynamics simulations: advances and applications. *Advances and applications in bioinformatics and chemistry*, 37-47.
3. Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature structural biology*, 9(9), 646-652.
4. Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O., & Shaw, D. E. (2009). Long-timescale molecular dynamics simulations of protein structure and function. *Current opinion in structural biology*, 19(2), 120-127.
5. Rapaport, D. C. (2014). Molecular dynamics simulation: a tool for exploration and discovery using simple models. *Journal of Physics: Condensed Matter*, 26(50), 503104.
6. Hernández-Rodríguez, M., C Rosales-Hernández, M., E Mendieta-Wejbe, J., Martínez-Archundia, M., & Correa Basurto, J. (2016). Current tools and methods in molecular dynamics (MD) simulations for drug design. *Current medicinal chemistry*, 23(34), 3909-3924.
7. Alder BJ, Wainwright T (1959) Studies in Molecular Dynamics. I. General Method. *The Journal of Chemical Physics*. **31** (2): 459-466. [Bibcode:1959JChPh..31..459A](#). [doi:10.1063/1.1730376](#).

8. Gibson JB, Goland AN, Milgram M, Vineyard G (1960). "Dynamics of Radiation Damage". *Phys. Rev.* **120** (4): 1229–1253. [Bibcode:1960PhRv..120.1229G](#). [doi:10.1103/PhysRev.120.1229](#).
9. [Rahman A](#) (1964) Correlations in the Motion of Atoms in Liquid Argon. *Physical Review*. **136** (2A): A405–A411. [Bibcode:1964PhRv..136..405R](#). [doi:10.1103/PhysRev.136.A405](#).
10. Patodia S, Bagaria A, Chopra D (2014) Molecular Dynamics Simulation of Proteins: A Brief Overview. *J Phys Chem Biophys* 4: 166. doi: 10.4172/2161- 0398.1000166
11. Durrant, J.D., McCammon, J.A. Molecular dynamics simulations and drug discovery. *BMC Biol* 9, 71 (2011). <https://doi.org/10.1186/1741-7007-9-71>
12. Hollingsworth, S. A., & Dror, R. O. (2018). Molecular dynamics simulation for all. *Neuron*, 99(6), 1129-1143.
13. Ladd, A. J. (1984). Equations of motion for non-equilibrium molecular dynamics simulations of viscous flow in molecular fluids. *Molecular Physics*, 53(2), 459-463.
14. Grubmüller, H., Heller, H., Windemuth, A., & Schulten, K. (1991). Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation*, 6(1-3), 121-142.
15. Best, R. B. (2012). Atomistic molecular simulations of protein folding. *Current opinion in structural biology*, 22(1), 52-61.
16. Prigozhin, M. B., & Gruebele, M. (2013). Microsecond folding experiments and simulations: a match is made. *Physical Chemistry Chemical Physics*, 15(10), 3372-3388.

17. Sagui, C., & Darden, T. A. (1999). Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annual review of biophysics and biomolecular structure*, 28(1), 155-179.
18. Xu, D., Williamson, M. J., & Walker, R. C. (2010). Advancements in molecular dynamics simulations of biomolecules on graphical processing units. In *Annual Reports in Computational Chemistry* (Vol. 6, pp. 2-19). Elsevier.
19. Wang, A. H., Zhang, Z. C., & Li, G. H. (2019). Advances in enhanced sampling molecular dynamics simulations for biomolecules. *Chinese Journal of Chemical Physics*, 32(3), 277.
20. Norberg, J., & Nilsson, L. (2003). Advances in biomolecular simulations: methodology and recent applications. *Quarterly Reviews of Biophysics*, 36(3), 257-306.
21. Ilyin, D. V., Goddard III, W. A., Oppenheim, J. J., & Cheng, T. (2019). First-principles-based reaction kinetics from reactive molecular dynamics simulations: Application to hydrogen peroxide decomposition. *Proceedings of the National Academy of Sciences*, 116(37), 18202-18208.
22. Gissinger, J. R., Jensen, B. D., & Wise, K. E. (2017). Modeling chemical reactions in classical molecular dynamics simulations. *Polymer*, 128, 211-217.
23. Smith, G. R., Sternberg, M. J., & Bates, P. A. (2005). The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking. *Journal of molecular biology*, 347(5), 1077-1101.
24. Lindorff-Larsen, K., Trbovic, N., Maragakis, P., Piana, S., & Shaw, D. E. (2012). Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *Journal of the American Chemical Society*, 134(8), 3787-3791.

25. Skolnick, J., Kolinski, A., Brooks III, C. L., Godzik, A., & Rey, A. (1993). A method for predicting protein structure from sequence. *Current Biology*, 3(7), 414-423.
26. Weinan, E., & Huang, Z. (2002). A dynamic atomistic–continuum method for the simulation of crystalline materials. *Journal of Computational Physics*, 182(1), 234-261.
27. Sholl, D. S. (2006). Understanding macroscopic diffusion of adsorbed molecules in crystalline nanoporous materials via atomistic simulations. *Accounts of Chemical Research*, 39(6), 403-411.
28. Stukowski, A. (2012). Structure identification methods for atomistic simulations of crystalline materials. *Modelling and Simulation in Materials Science and Engineering*, 20(4), 045021.
29. Mendelev, M. I., Kramer, M. J., Becker, C. A., & Asta, M. (2008). Analysis of semi-empirical interatomic potentials appropriate for simulation of crystalline and liquid Al and Cu. *Philosophical Magazine*, 88(12), 1723-1750.
30. Ciccotti, G., & Ryckaert, J. P. (1986). Molecular dynamics simulation of rigid molecules. *Computer Physics Reports*, 4(6), 346-392.
31. Deganutti, G., Prischi, F., & Reynolds, C. A. (2021). Supervised molecular dynamics for exploring the druggability of the SARS-CoV-2 spike protein. *Journal of computer-aided molecular design*, 35, 195-207.
32. Poojari, C., Zak, A., Dzieciuch-Rojek, M., Bunker, A., Kepczynski, M., and Róg, T. (2020). Cholesterol reduces partitioning of antifungal drug itraconazole into lipid bilayers. *J. Phys. Chem. B* 124, 2139–2148. doi: 10.1021/acs.jpcc.9b11005

33. Dasari, S., & Mallik, B. S. (2020). Solubility and solvation free energy of a cardiovascular drug, LASSBio-294, in ionic liquids: A computational study. *Journal of Molecular Liquids*, 301, 112449.
34. Arbe, A., Alvarez, F., & Colmenero, J. (2012). Neutron scattering and molecular dynamics simulations: Synergetic tools to unravel structure and dynamics in polymers. *Soft Matter*, 8(32), 8257-8270.
35. Piana, S., Klepeis, J. L., & Shaw, D. E. (2014). Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Current opinion in structural biology*, 24, 98-105.
36. Zhu, H., & Averbach, R. S. (1995). Molecular dynamics simulations of densification processes in nanocrystalline materials. *Materials Science and Engineering: A*, 204(1-2), 96-100.
37. Zhu, H., & Averbach, R. S. (1995). Molecular dynamics simulations of densification processes in nanocrystalline materials. *Materials Science and Engineering: A*, 204(1-2), 96-100.
38. González, M. A. (2011). Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique*, 12, 169-200.
39. Lopes, P. E., Guvench, O., & MacKerell, A. D. (2015). Current status of protein force fields for molecular dynamics simulations. *Molecular modeling of proteins*, 47-71.
40. Betz, R. M., & Walker, R. C. (2015). Paramfit: Automated optimization of force field parameters for molecular dynamics simulations. *Journal of computational chemistry*, 36(2), 79-87.

41. Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F., *European Biophysics Journal* **2011**, *40*, 843-856.
42. Monticelli, L. and Tieleman, D.P., 2013. Force fields for classical molecular dynamics. *Biomolecular simulations*, pp.197-213.
43. González, M.A., (2011). Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique*, *12*, pp.169-200
44. Guvench, O., & MacKerell, A. D. (2008). Comparison of protein force fields for molecular dynamics simulations. *Molecular modeling of proteins*, 63-88.
45. Lorentz, J.H.A., & Berthelot, M. (1881). *La Théorie Mécanique de la Chaleur*. Gauthier-Villars, Paris.
46. Brameld, K., Dasgupta, S., & Goddard, W. A. (1997). Distance dependent hydrogen bond potentials for nucleic acid base pairs from ab initio quantum mechanical calculations (LMP2/cc-pVTZ). *The Journal of Physical Chemistry B*, *101*(24), 4851-4859.
47. Suzuki, I., & Overend, J. (1981). The anharmonic Urey-Bradley force field; the computational algorithm and its application to CF₄. *Spectrochimica Acta Part A: Molecular Spectroscopy*, *37*(12), 1093-1098.
48. Wen, S., & Beran, G. J. (2011). Accurate molecular crystal lattice energies from a fragment QM/MM approach with on-the-fly ab initio force field parametrization. *Journal of Chemical Theory and Computation*, *7*(11), 3733-3742.
49. Cisneros, G. A., Tholander, S. N. I., Parisel, O., Darden, T. A., Elking, D., Perera, L., & Piquemal, J. P. (2008). Simple formulas for improved point-charge electrostatics in classical force fields and hybrid quantum mechanical/molecular mechanical embedding. *International journal of quantum chemistry*, *108*(11), 1905-1912.

50. Scott, W. R., Hünenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., ... & Van Gunsteren, W. F. (1999). The GROMOS biomolecular simulation program package. *The Journal of Physical Chemistry A*, 103(19), 3596-3607.
51. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3), 712-725.
52. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. A., & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, 4(2), 187-217.
53. Shi, Y., Xia, Z., Zhang, J., Best, R., Wu, C., Ponder, J. W., & Ren, P. (2013). Polarizable atomic multipole-based AMOEBA force field for proteins. *Journal of chemical theory and computation*, 9(9), 4046-4063.
54. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., & De Vries, A. H. (2007). The MARTINI force field: coarse grained model for biomolecular simulations. *The journal of physical chemistry B*, 111(27), 7812-7824.
55. Damm, W., Frontera, A., Tirado-Rives, J., & Jorgensen, W. L. (1997). OPLS all-atom force field for carbohydrates. *Journal of computational chemistry*, 18(16), 1955-1970.
56. <http://kbff.chem.k-state.edu>
57. Allinger, N. L., Yuh, Y. H., & Lii, J. H. (1989). Molecular mechanics. The MM3 force field for hydrocarbons. 1. *Journal of the American Chemical Society*, 111(23), 8551-8566.
58. Senftle, T. P., Hong, S., Islam, M. M., Kylasa, S. B., Zheng, Y., Shin, Y. K., ... & Van Duin, A. C. (2016). The ReaxFF reactive force-field: development, applications and future directions. *npj Computational Materials*, 2(1), 1-14.

59. Ploetz, E. A., Karunaweera, S., Benteinitis, N., Chen, F., Dai, S., Gee, M. B., ... & Smith, P. E. (2021). Kirkwood–Buff-derived force field for peptides and proteins: philosophy and development of KBFF20. *Journal of Chemical Theory and Computation*, 17(5), 2964-2990.
60. Kirkwood, J. G., & Buff, F. P. (1951). The statistical mechanical theory of solutions. I. *The Journal of chemical physics*, 19(6), 774-777.
61. Ben-Naim, A. (1977). Inversion of the Kirkwood–Buff theory of solutions: Application to the water–ethanol system. *The Journal of Chemical Physics*, 67(11), 4884-4890.
62. Ploetz, E. A., Karunaweera, S., & Smith, P. E. (2021). Kirkwood–Buff-Derived Force Field for Peptides and Proteins: Applications of KBFF20. *Journal of Chemical Theory and Computation*, 17(5), 2991-3009.
63. McIvor, J. A., Larsen, D. S., & Mercadante, D. (2022). Simulating Polyproline II-Helix-Rich Peptides with the Latest Kirkwood–Buff Force Field: A Direct Comparison with AMBER and CHARMM. *The Journal of Physical Chemistry B*, 126(40), 7833-7846.
64. Hall, R., Dixon, T., & Dickson, A. (2020). On calculating free energy differences using ensembles of transition paths. *Frontiers in molecular biosciences*, 7, 106.
65. Ballester, P. J., & Mitchell, J. B. (2010). A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics*, 26(9), 1169-1175.
66. Michel, J., & Essex, J. W. (2010). Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *Journal of computer-aided molecular design*, 24, 639-658.

67. Jarzynski, C. (1997). Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E*, 56(5), 5018.
68. Chen, L. Y., Bastien, D. A., & Espejel, H. E. (2010). Determination of equilibrium free energy from nonequilibrium work measurements. *Physical Chemistry Chemical Physics*, 12(25), 6579-6582.
69. Rodriguez-Gomez, D., Darve, E., & Pohorille, A. (2004). Assessing the efficiency of free energy calculation methods. *The Journal of chemical physics*, 120(8), 3563-3578.
70. Guest, E. E., Cervantes, L. F., Pickett, S. D., Brooks III, C. L., & Hirst, J. D. (2022). Alchemical free energy methods applied to complexes of the first bromodomain of BRD4. *Journal of Chemical Information and Modeling*, 62(6), 1458-1470.
71. Straatsma, T. P., & Berendsen, H. J. C. (1988). Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *The Journal of chemical physics*, 89(9), 5876-5886.
72. Kirkwood, J. G. (1935). Statistical mechanics of fluid mixtures. *The Journal of chemical physics*, 3(5), 300-313.
73. Mitchell, M. J., & McCammon, J. A. (1991). Free energy difference calculations by thermodynamic integration: difficulties in obtaining a precise value. *Journal of computational chemistry*, 12(2), 271-275.
74. Habershon, S., & Manolopoulos, D. E. (2011). Thermodynamic integration from classical to quantum mechanics. *The Journal of chemical physics*, 135(22), 224111
75. Pohorille, A. (2007). *Free energy calculations: theory and applications in chemistry and biology*. Springer.

76. Zwanzig, R. W. (1954). High-temperature equation of state by a perturbation method. I. Nonpolar gases. *The Journal of Chemical Physics*, 22(8), 1420-1426
77. Zwanzig, R. W. (1955). High-temperature equation of state by a perturbation method. II. polar gases. *The Journal of Chemical Physics*, 23(10), 1915-1922.
78. Beveridge, D. L., & Dicapua, F. M. (1989). Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annual review of biophysics and biophysical chemistry*, 18(1), 431-492.
79. Shirts, M. R., & Mobley, D. L. (2013). An introduction to best practices in free energy calculations. *Biomolecular Simulations: Methods and Protocols*, 271-311.
80. Widom, B. (1963). Some topics in the theory of fluids. *The Journal of Chemical Physics*, 39(11), 2808-2812.
81. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., & Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry*, 13(8), 1011-1021.
82. Roux, B. (1995). The calculation of the potential of mean force using computer simulations. *Computer physics communications*, 91(1-3), 275-282.
83. http://pmx.mpibpc.mpg.de/summerSchool2020_tutorial2/index.html
84. Zheng, Feng; Shao, Zhu-Qing; Hao, Xina; Wu, Qianqian; Li, Chaolong; Hou, Hongfen; Hu, Dan; Wang, Changjun; Pan, Xiuzhen (2018). *Identification of oligopeptide-binding protein (OppA) and its role in the virulence of Streptococcus suis serotype 2. Microbial Pathogenesis*, 118(), 322–329.

85. Tame, J. R., Dodson, E. J., Murshudov, G., Higgins, C. F., & Wilkinson, A. J. (1995). The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands. *Structure*, 3(12), 1395-1406.
86. Sleigh, S. H., Seavers, P. R., Wilkinson, A. J., Ladbury, J. E., & Tame, J. R. H. (1999). Crystallographic and calorimetric analysis of peptide binding to OppA protein. *Journal of molecular biology*, 291(2), 393-415.
87. Maurer, M., de Beer, S. B., & Oostenbrink, C. (2016). Calculation of relative binding free energy in the water-filled active site of oligopeptide-binding protein A. *Molecules*, 21(4), 499.
88. Ben-Shalom, I. Y., Lin, Z., Radak, B. K., Lin, C., Sherman, W., & Gilson, M. K. (2020). Accounting for the central role of interfacial water in protein–ligand binding free energy calculations. *Journal of chemical theory and computation*, 16(12), 7883-7894.
89. Hünenberger, P. H., Börjesson, U., & Lins, R. D. (2001). Electrostatic interactions in biomolecular systems. *Chimia*, 55(10), 861-861.
90. Cheatham, T. I., Miller, J. L., Fox, T., Darden, T. A., & Kollman, P. A. (1995). Molecular dynamics simulations on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA, and proteins. *Journal of the American Chemical Society*, 117(14), 4193-4194.
91. Rocklin, G. J., Mobley, D. L., Dill, K. A., & Hünenberger, P. H. (2013). Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: An accurate correction scheme for electrostatic finite-size effects. *The Journal of chemical physics*, 139(18), 11B606_1.

92. Radhakrishnan, M. L. (2012). Designing electrostatic interactions in biological systems via charge optimization or combinatorial approaches: insights and challenges with a continuum electrostatic framework. *Theoretical Chemistry Accounts*, 131, 1-24.
93. Gilson, M. K., McCammon, J. A., & Madura, J. D. (1995). Molecular dynamics simulation with a continuum electrostatic model of the solvent. *Journal of computational chemistry*, 16(9), 1081-1095.
94. Hünenberger, P. H. (1999, November). Lattice-sum methods for computing electrostatic interactions in molecular simulations. In *AIP Conference Proceedings* (Vol. 492, No. 1, pp. 17-83). American Institute of Physics.
95. Fogolari, F., Brigo, A., & Molinari, H. (2002). The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *Journal of Molecular Recognition*, 15(6), 377-392.
96. Lu, B. Z., Zhou, Y. C., Holst, M. J., & McCammon, J. A. (2008). Recent progress in numerical methods for the Poisson-Boltzmann equation in biophysical applications. *Commun Comput Phys*, 3(5), 973-1009.
97. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J Mol Biol* 1963, 7, 95.
98. Mackerell Jr, A. D., Feig, M., & Brooks III, C. L. (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of computational chemistry*, 25(11), 1400-1415.
99. [https://proteopedia.org/wiki/index.php/Tutorial:Ramachandran principle and phi psi angles](https://proteopedia.org/wiki/index.php/Tutorial:Ramachandran_principle_and_phi_psi_angles)

100. Ramachandran, G. N., Sasisekharan V. Conformation of polypeptides and proteins. Adv Protein Chem. 1968;23:283-438. [PMID:4882249](#)
101. https://www.cryst.bbk.ac.uk/PPS95/course/3_geometry/rama.html
102. Perez, A., MacCallum, J. L., Brini, E., Simmerling, C., & Dill, K. A. (2015). Grid-based backbone correction to the ff12SB protein force field for implicit-solvent simulations. *Journal of chemical theory and computation*, 11(10), 4770-4779.
103. Kang, W., Jiang, F., & Wu, Y. D. (2018). Universal implementation of a residue-specific force field based on CMAP potentials and free energy decomposition. *Journal of chemical theory and computation*, 14(8), 4474-4486.
104. MacKerell Jr, A. D., Feig, M., & Brooks, C. L. (2004). Improved treatment of the protein backbone in empirical force fields. *Journal of the American Chemical Society*, 126(3), 698-699.
105. Brooks III, C. L. (1988). III; Karplus, M.; Pettitt, BM. Proteins: A theoretical perspective of dynamics, structure, and thermodynamics. *Advances in Chemical Physics*, 71.
106. Pitera, J. W., & van Gunsteren, W. F. (2002). A Comparison of Non-Bonded Scaling Approaches for Free Energy Calculations. *Molecular Simulation*, 28(1-2), 45-65.
107. Beutler, T. C., Mark, A. E., van Schaik, R. C., Gerber, P. R., & van Gunsteren, W. F. (1994). Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chemical Physics Letters*, 222(6), 529-539.
108. H. Bekker, H.J.C. Berendsen, E.J. Dijkstra, S. Achterop, R. van Drunen, D. van der Spoel, A. Sijbers, and H. Keegstra et al., "Gromacs: A parallel computer for molecular

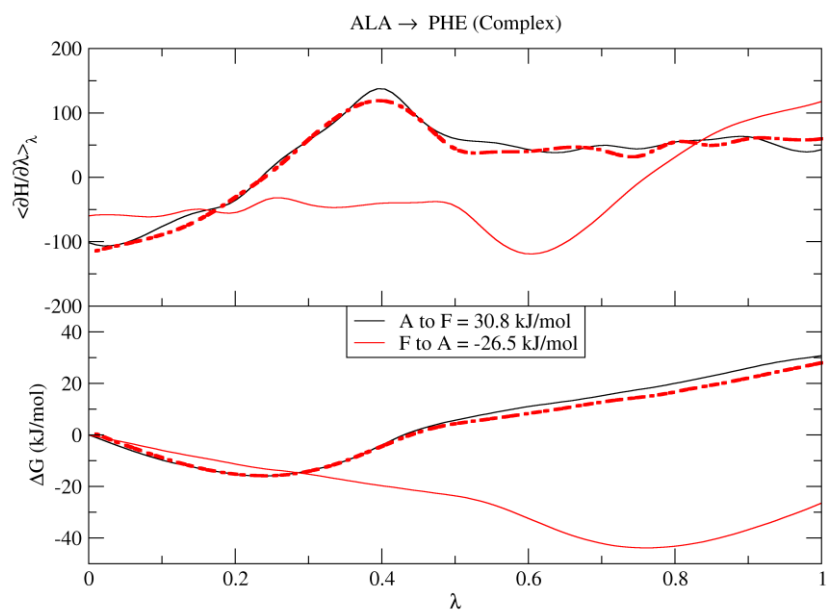
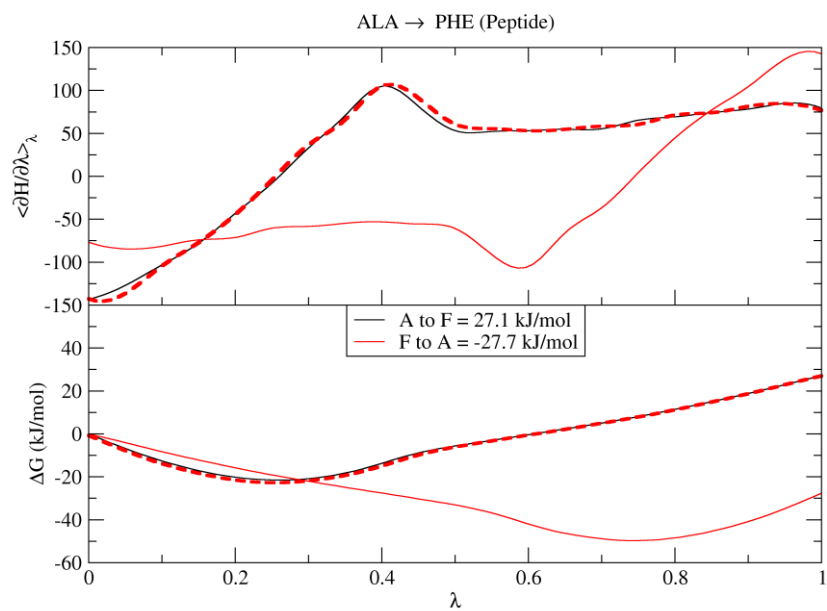
dynamics simulations”; pp. 252–256 in *Physics computing 92*. Edited by R.A. de Groot and J. Nadrchal. World Scientific, Singapore, 1993.

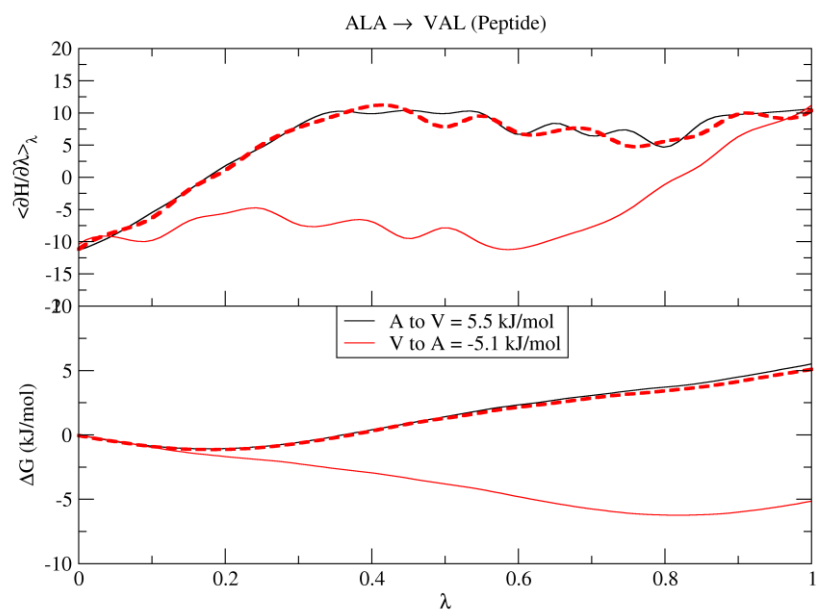
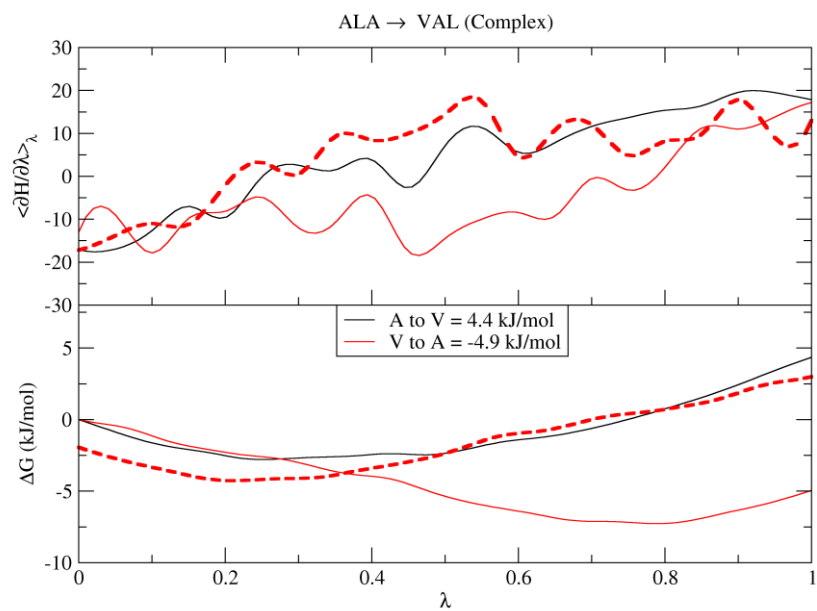
109. Hünenberger, P. H., & McCammon, J. A. (1999). Ewald artifacts in computer simulations of ionic solvation and ion–ion interaction: a continuum electrostatics study. *The Journal of chemical physics*, 110(4), 1856-1872.
110. Allen, M. P. (1987). *DJ Tildesley Computer simulation of liquids*. Clarendon, Oxford.
111. Smith, P. E., & Pettitt, B. M. (1991). Peptides in ionic solutions: a comparison of the Ewald and switching function techniques. *The Journal of chemical physics*, 95(11), 8430-8441.
112. Schreiber, H., & Steinhauser, O. (1992). Cutoff size does strongly influence molecular dynamics results on solvated polypeptides. *Biochemistry*, 31(25), 5856-5860.
113. Schreiber, H., & Steinhauser, O. (1992). Molecular dynamics studies of solvated polypeptides: why the cut-off scheme does not work. *Chemical physics*, 168(1), 75-89.
114. Kastenholz, M. A., & Hünenberger, P. H. (2006). Computation of methodology-independent ionic solvation free energies from molecular simulations. II. The hydration free energy of the sodium cation. *The Journal of chemical physics*, 124(22), 224501.
115. Kastenholz, M. A., & Hünenberger, P. H. (2006). Computation of methodology-independent ionic solvation free energies from molecular simulations. I. The electrostatic potential in molecular liquids. *The Journal of chemical physics*, 124(12), 124106.
116. Nijboer, B. R. A., & Ruijgrok, T. W. (1988). On the energy per particle in three- and two-dimensional Wigner lattices. *Journal of statistical physics*, 53, 361-382.

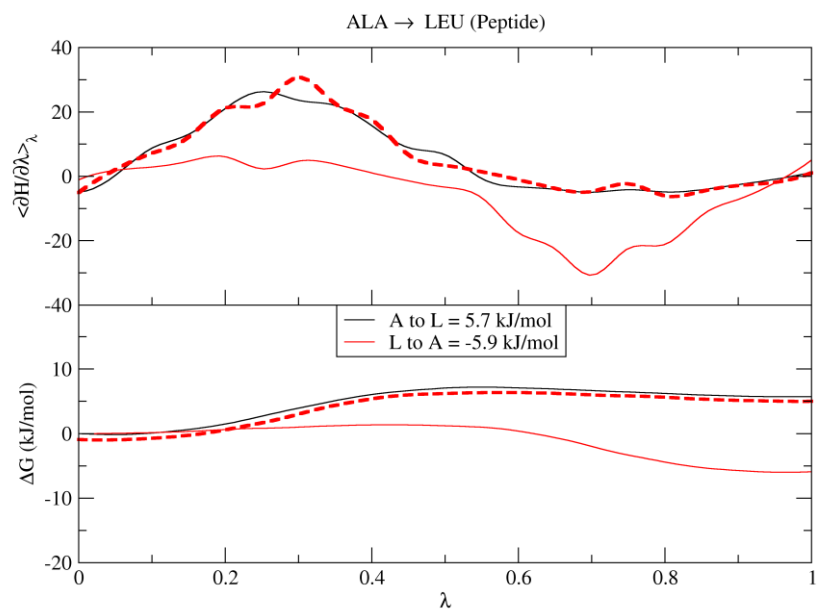
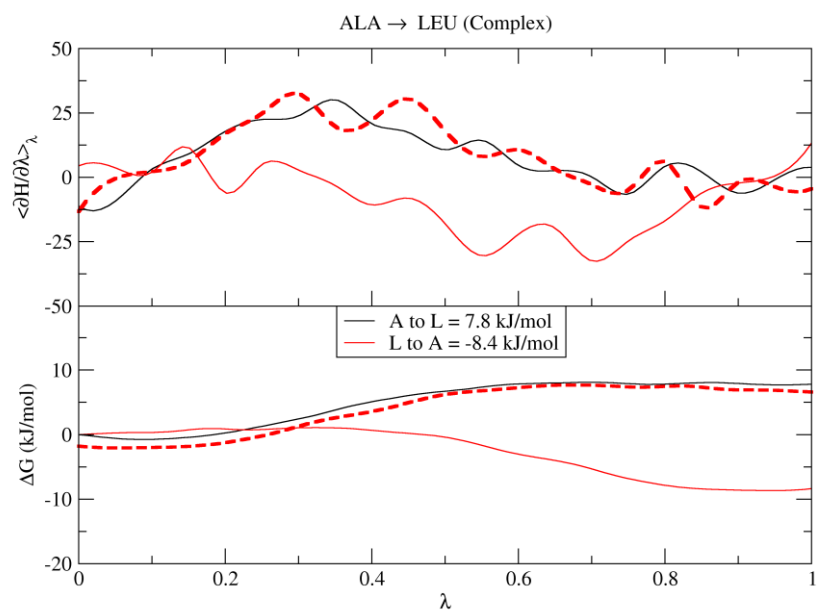
117. Cichocki, B., Felderhof, B. U., & Hinsen, K. (1989). Electrostatic interactions in periodic Coulomb and dipolar systems. *Physical Review A*, 39(10), 5350.
118. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *Journal of computational chemistry*, 26(16), 1701-1718.
119. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J., LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* 1997, 18 (12), 1463-1472.
120. Miyamoto, S., & Kollman, P. A. (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of computational chemistry*, 13(8), 952-962.
121. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, 98(18), 10037-10041.
122. Dixit, S. B., & Chipot, C. (2001). Can absolute free energies of association be estimated from molecular mechanical simulations? The biotin– streptavidin system revisited. *The Journal of Physical Chemistry A*, 105(42), 9795-9799.
123. Donnini, S., Mark, A. E., Juffer, A. H., & Villa, A. (2005). Incorporating the effect of ionic strength in free energy calculations using explicit ions. *Journal of computational chemistry*, 26(2), 115-122.
124. Li, L., & Liang, J. (2015). Electrostatic contribution to protein stability: insights from long-range electrostatic theory and molecular dynamics simulations. *Current Opinion in Structural Biology*, 30, 27-34.

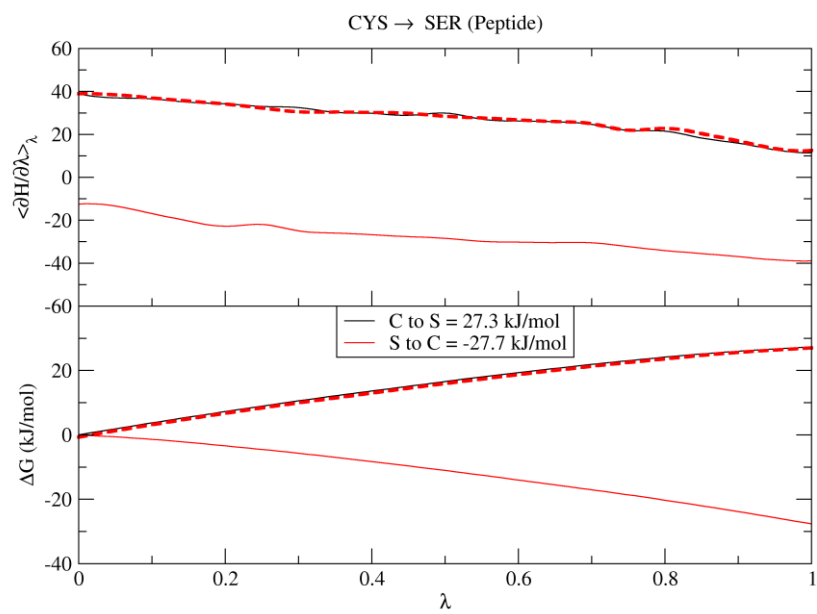
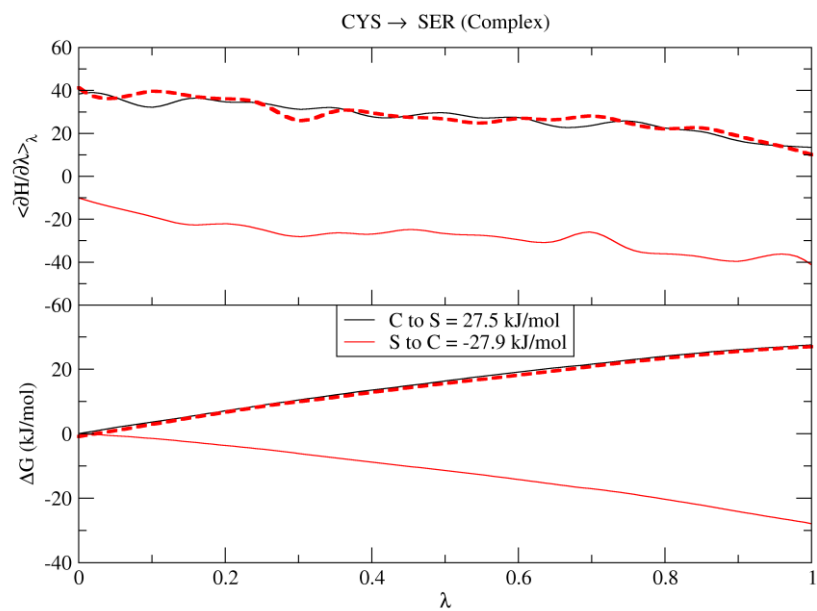
125. Lang, F., Föller, M., Lang, K. S., Lang, P. A., Ritter, M., Gulbins, E., ... & Huber, S. M. (2005). Ion channels in cell proliferation and apoptotic cell death. *The Journal of membrane biology*, 205, 147-157.
126. Fersht, A. (1985). Measurement and magnitude of enzymatic rate constants. *Enzyme structure and mechanism*, 2, 121-124.
127. Huckel, E., & Debye, P. (1923). Zur theorie der elektrolyte. i. gefrierpunktserniedrigung und verwandte erscheinungen. *Phys. Z*, 24, 185-206.
128. Lipfert, J., Doniach, S., Das, R., & Herschlag, D. (2014). Understanding nucleic acid-ion interactions. *Annual review of biochemistry*, 83, 813-841.
129. da Rocha, L., Baptista, A. M., & Campos, S. R. (2022). Computational Study of the pH-Dependent Ionic Environment around β -Lactoglobulin. *The Journal of Physical Chemistry B*, 126(45), 9123-9136.
130. Lee, C. C. (2014). The charge conserving Poisson-Boltzmann equations: Existence, uniqueness, and maximum principle. *Journal of Mathematical Physics*, 55(5), 051503.
131. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., ... & MacKerell Jr, A. D. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods*, 14(1), 71-73.
132. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., & Shaw, D. E. (1950). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*. 2010. V. 78.

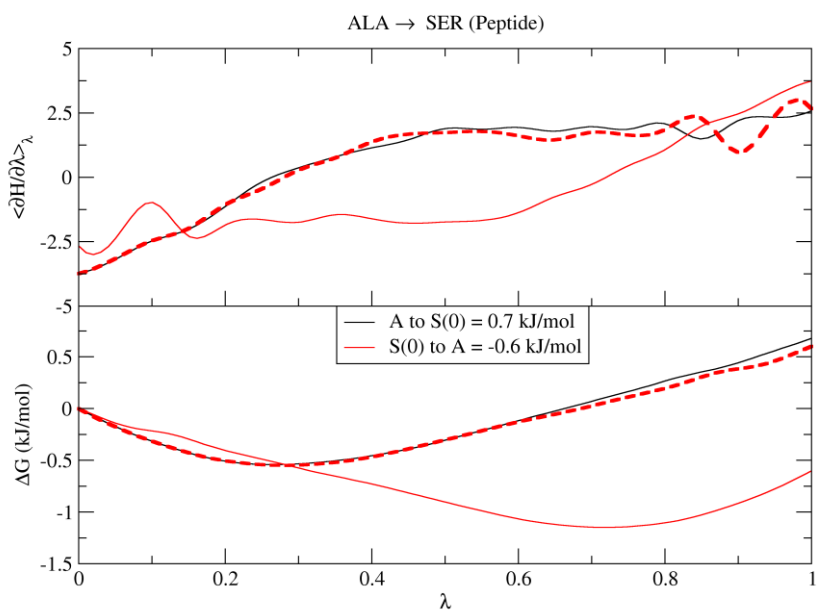
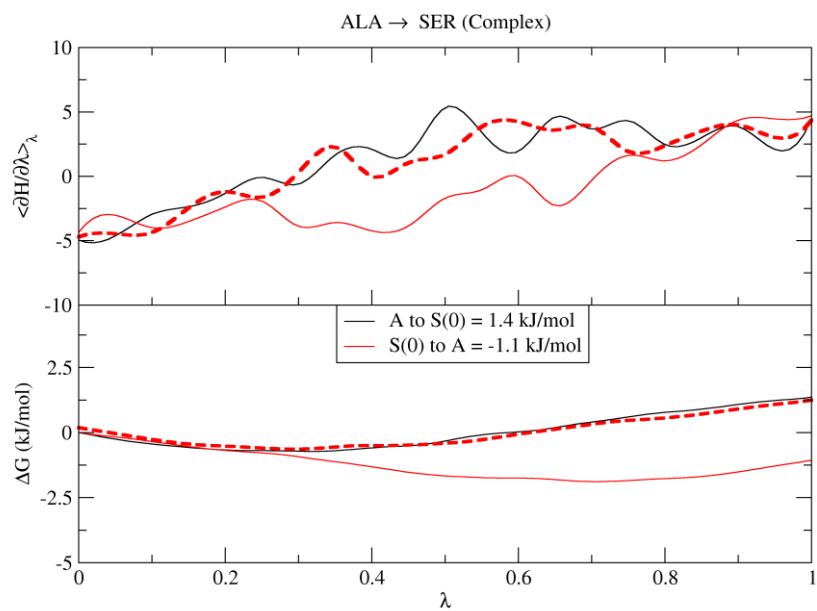
Appendix A - TI curves for alchemical transformations

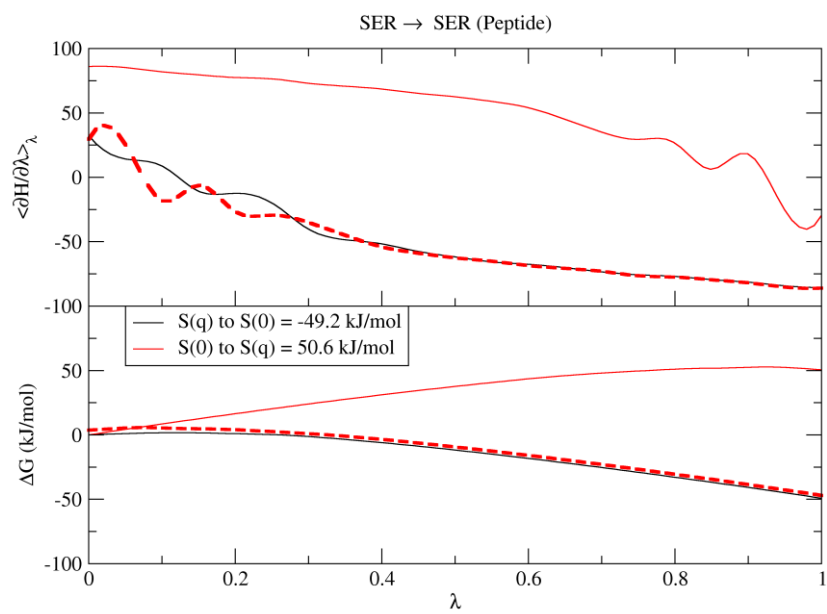
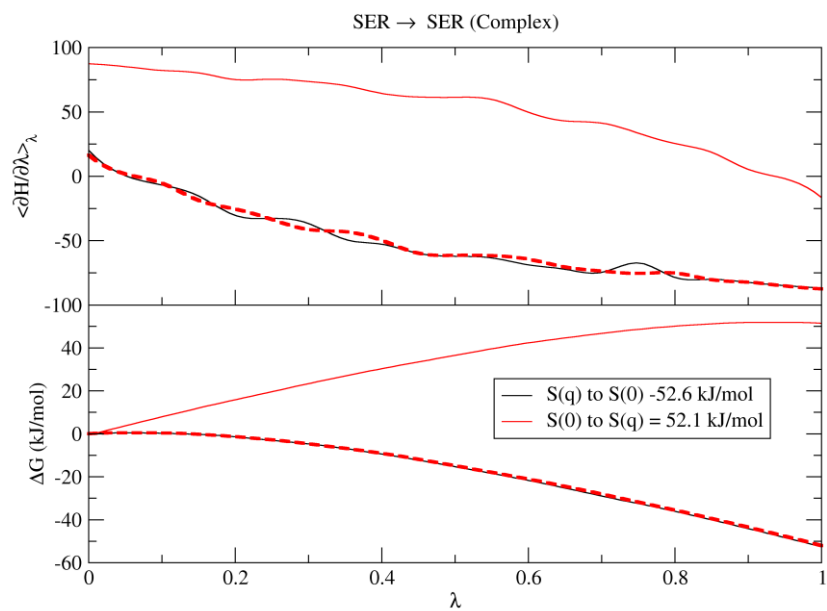


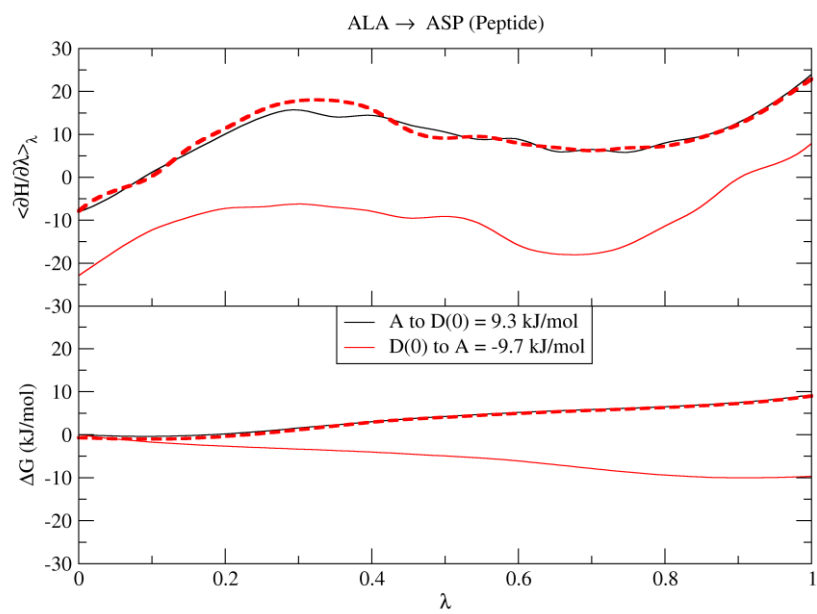
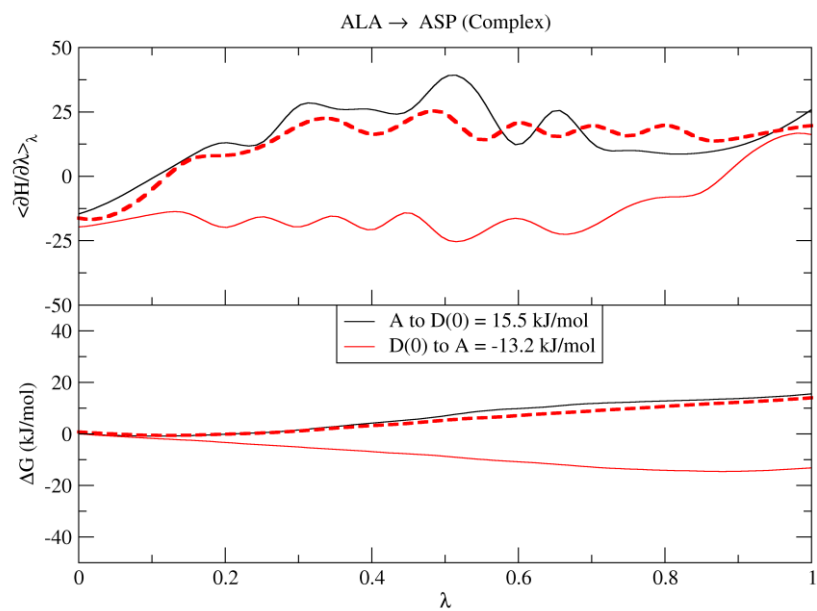


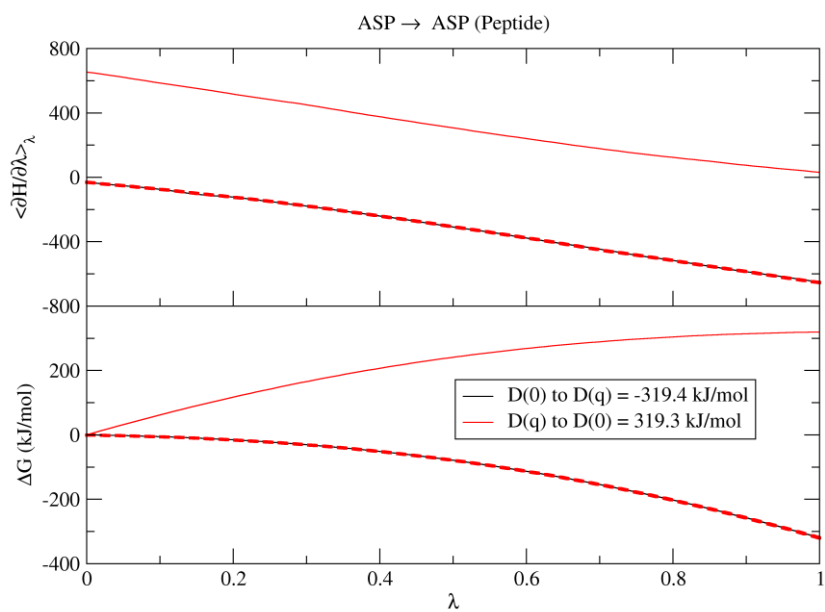
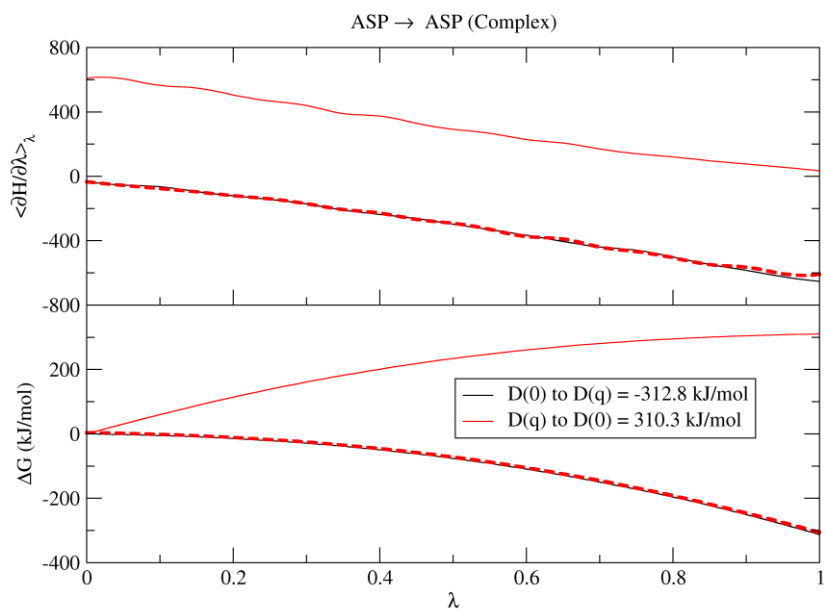


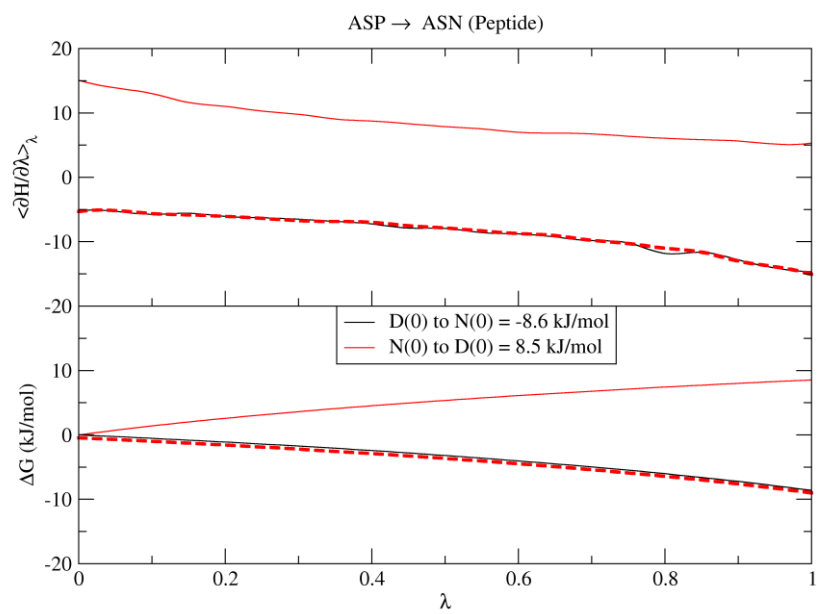
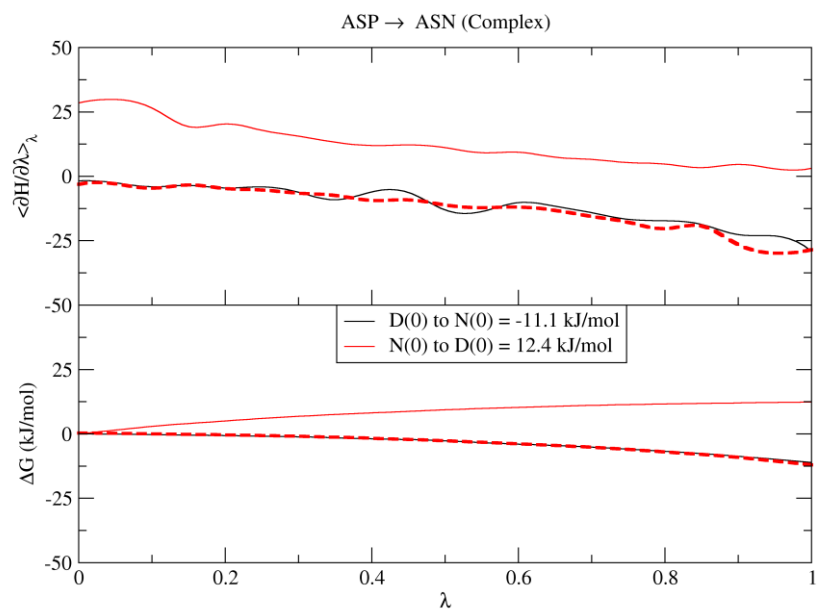


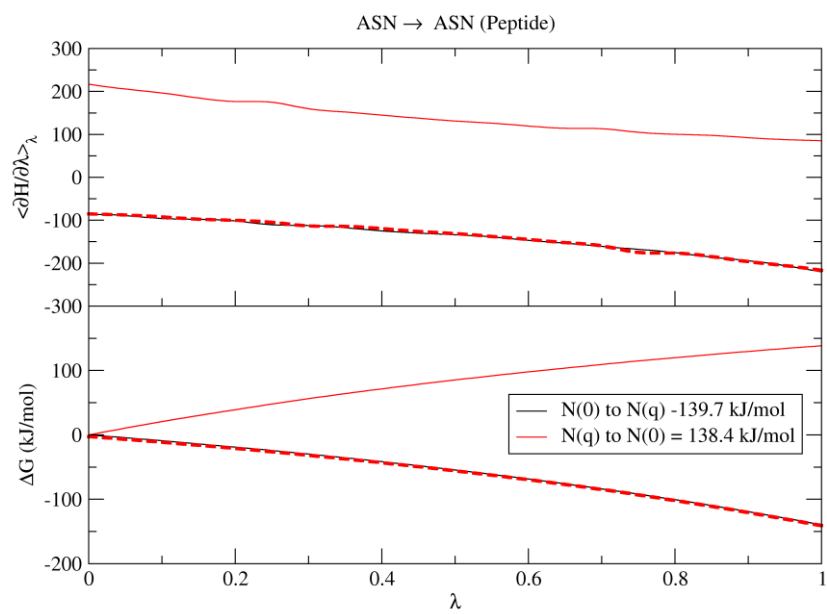
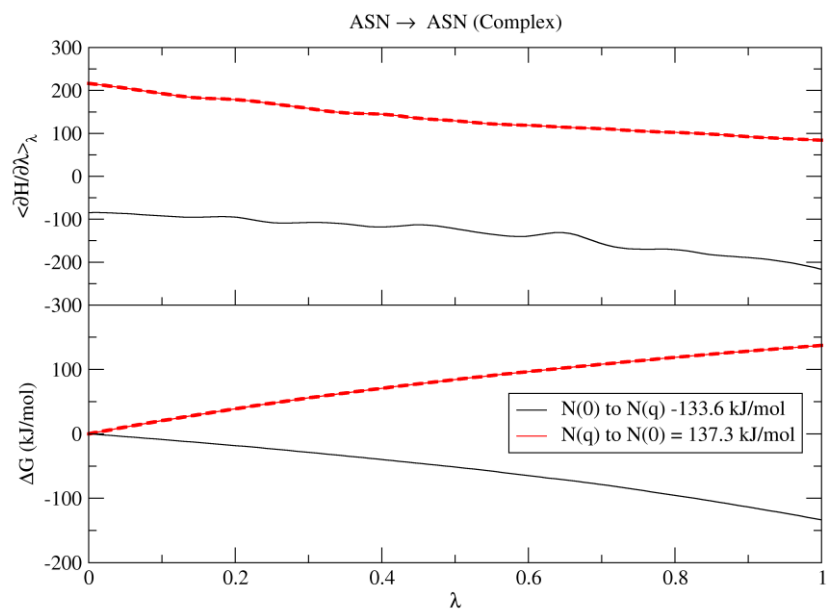


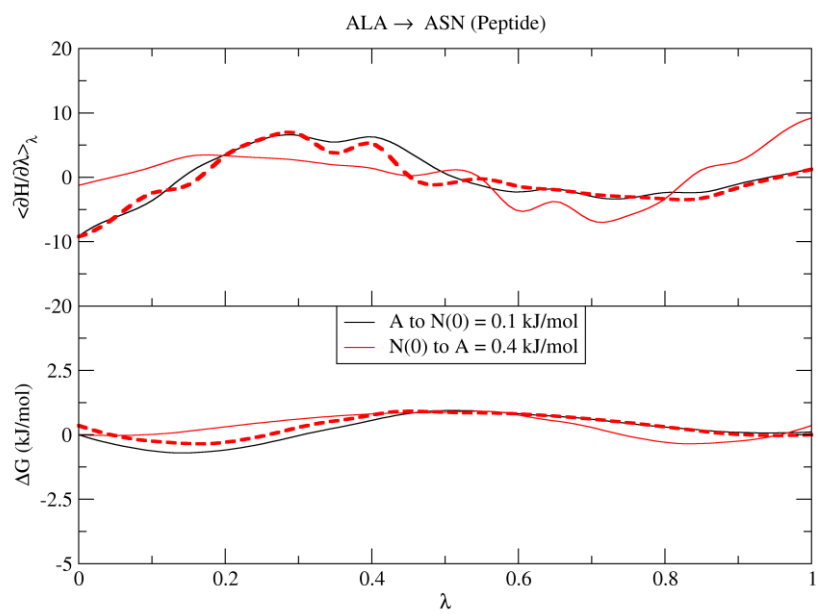
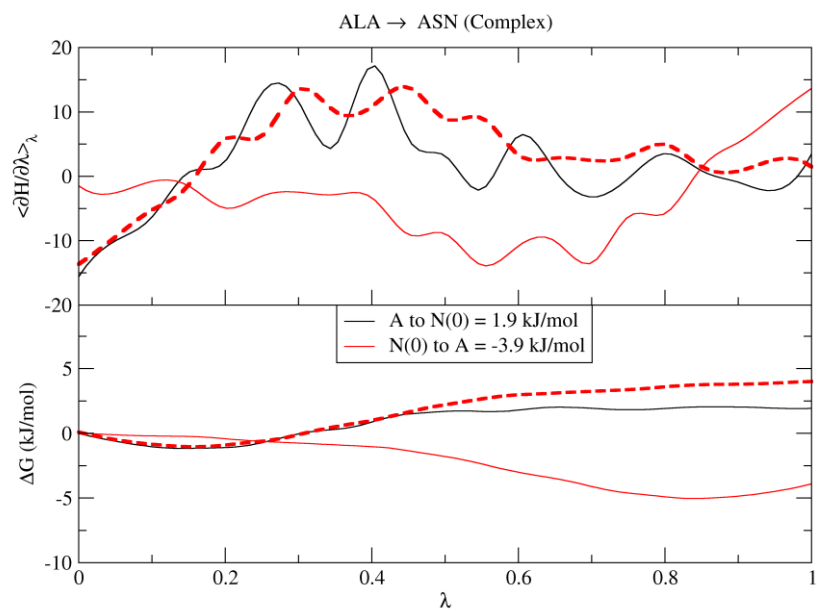


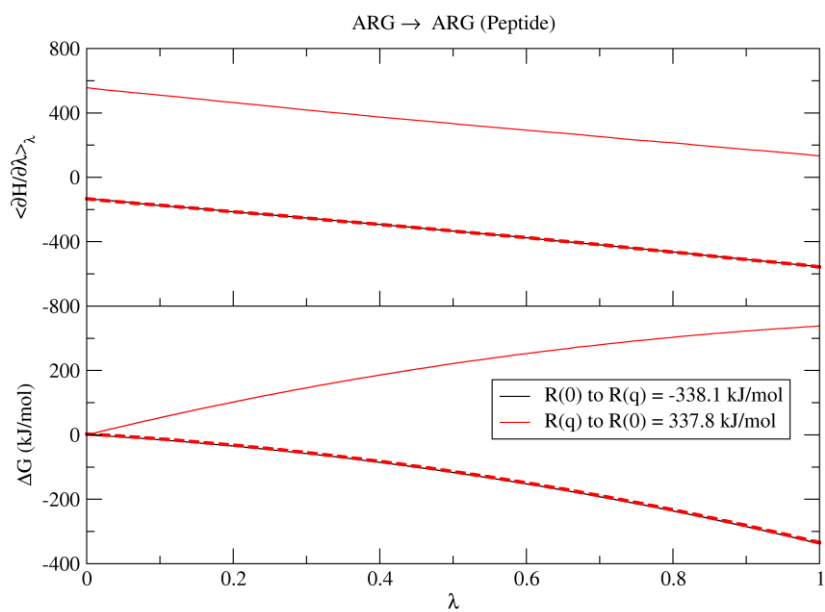
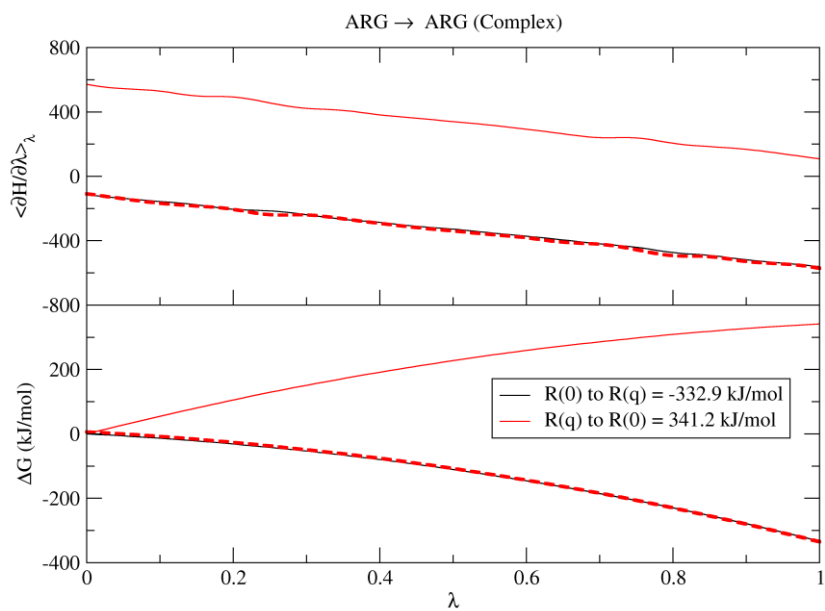


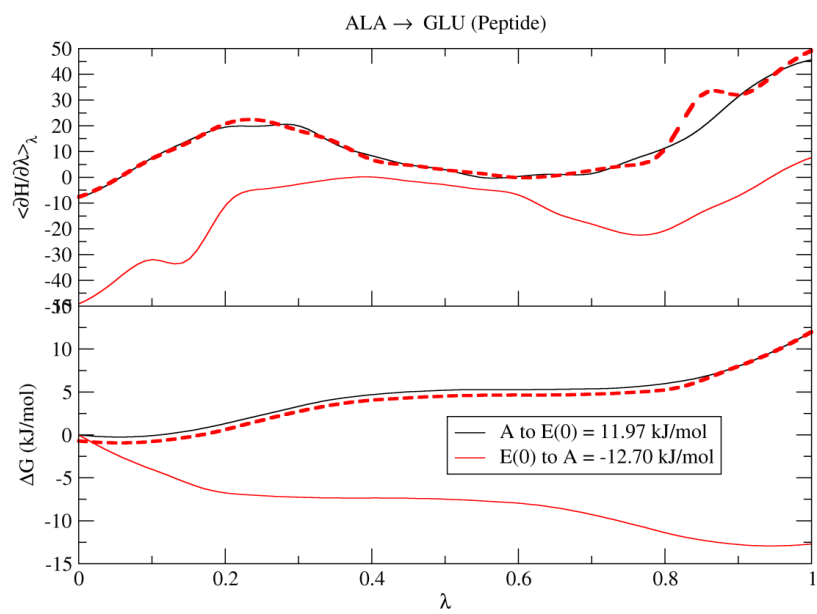
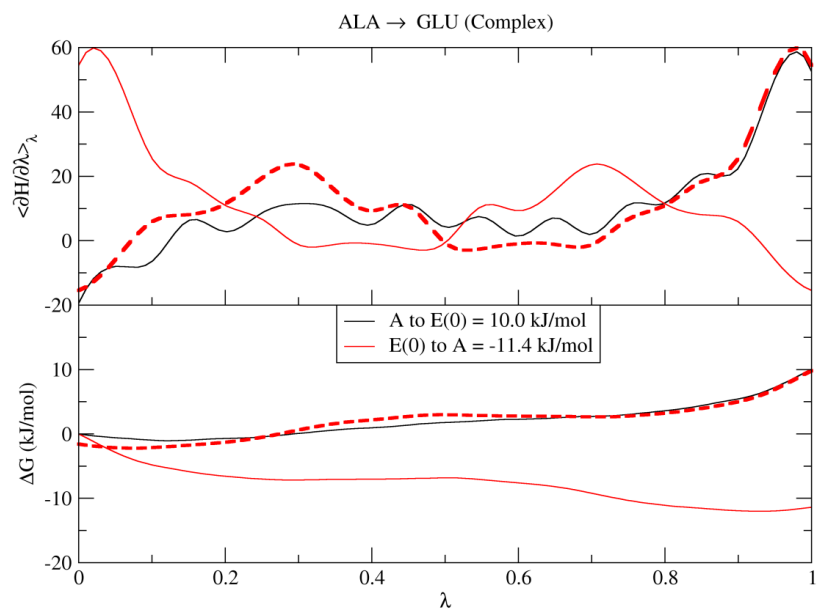


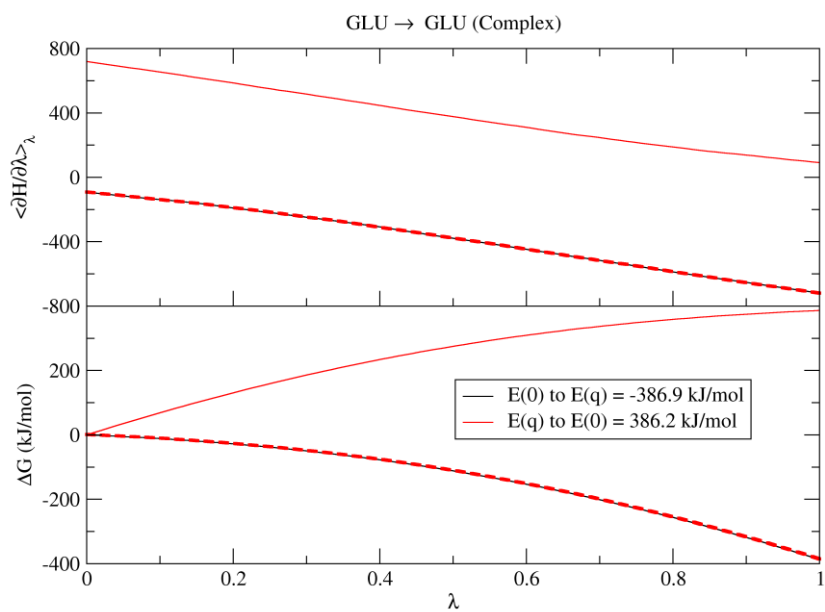
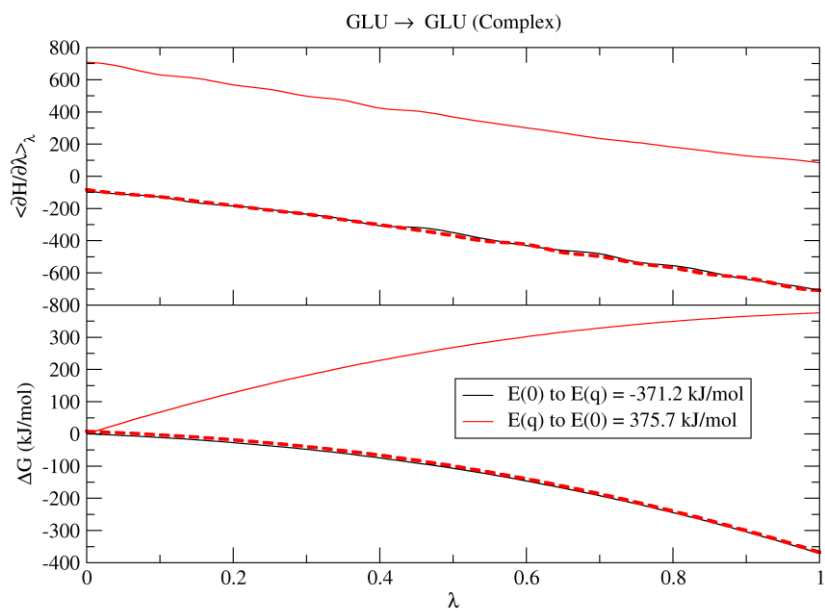


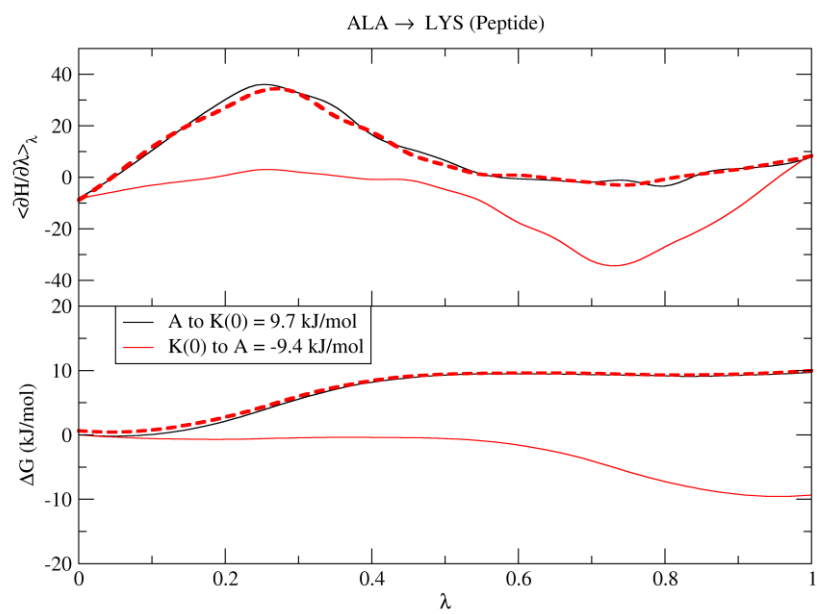
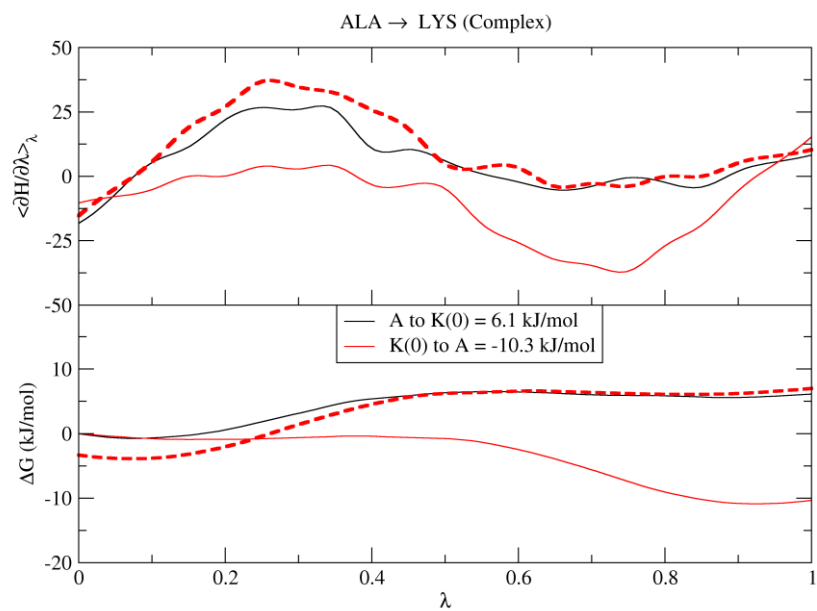


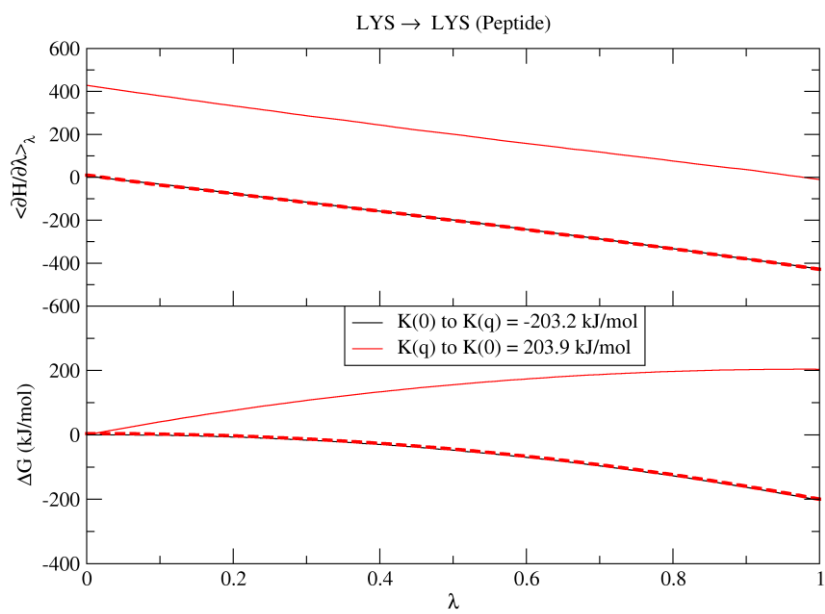
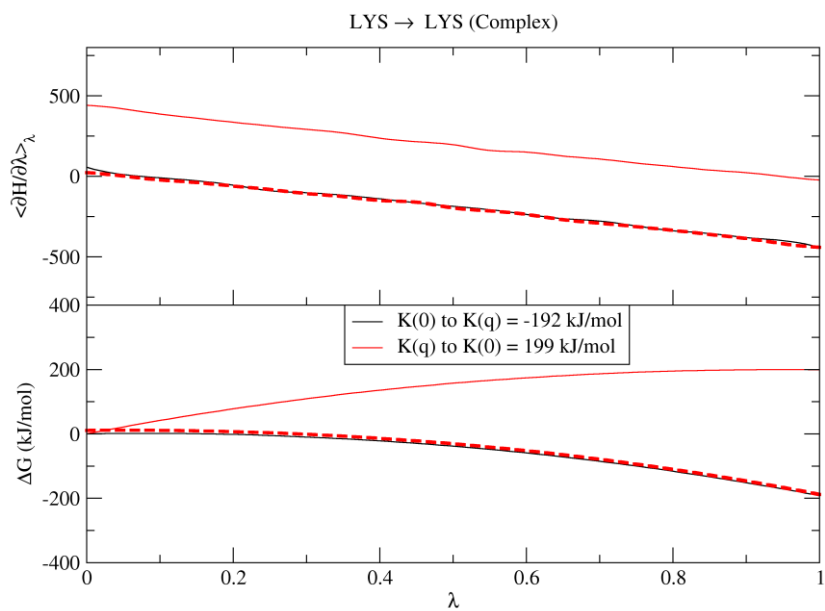




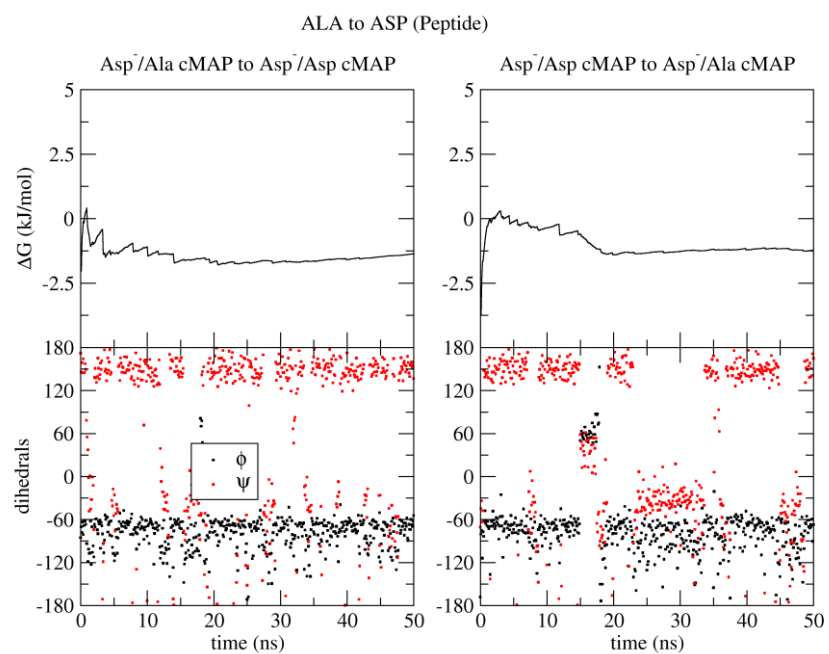
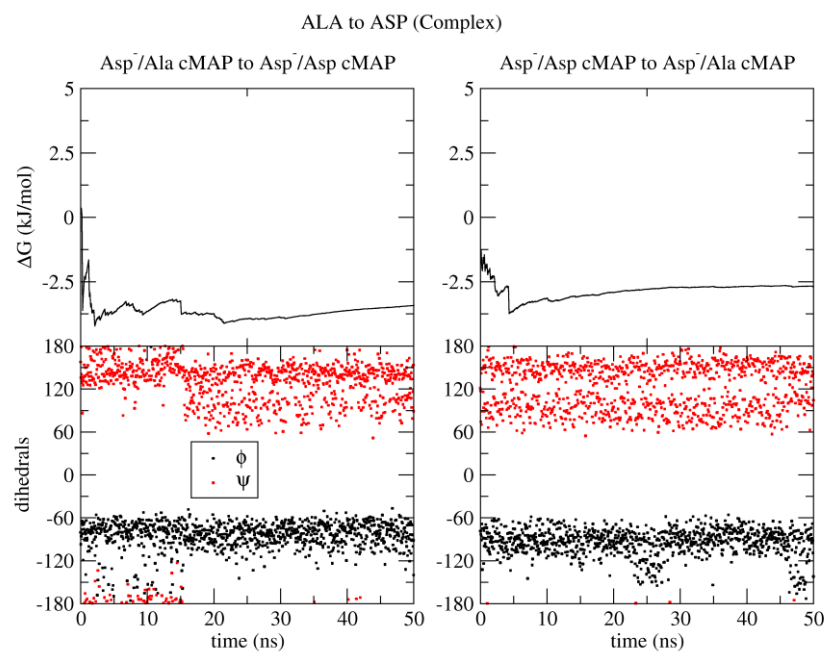




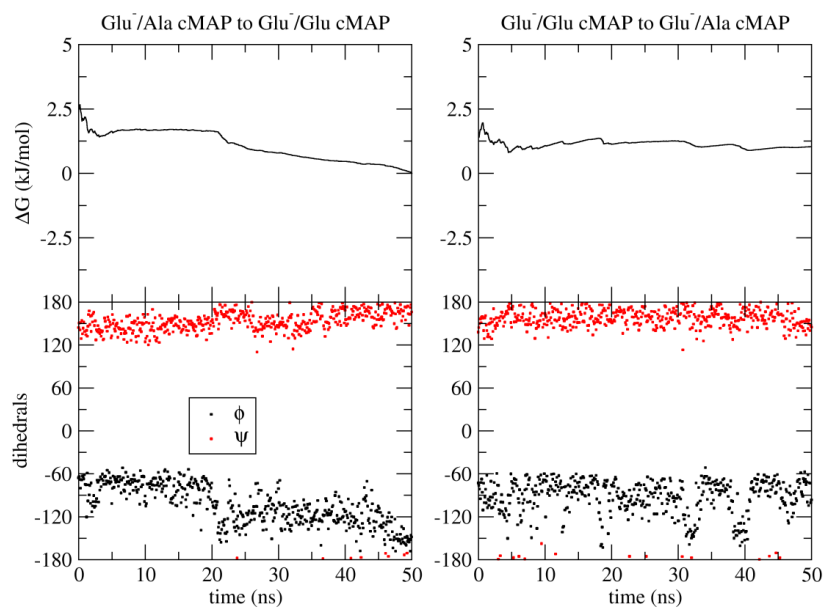




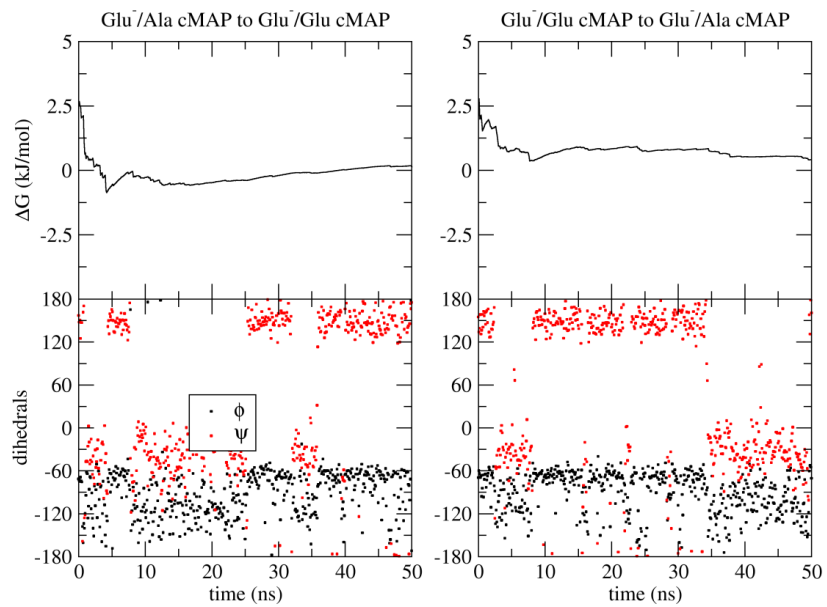
Appendix B - FEP curves for CMAP transformations



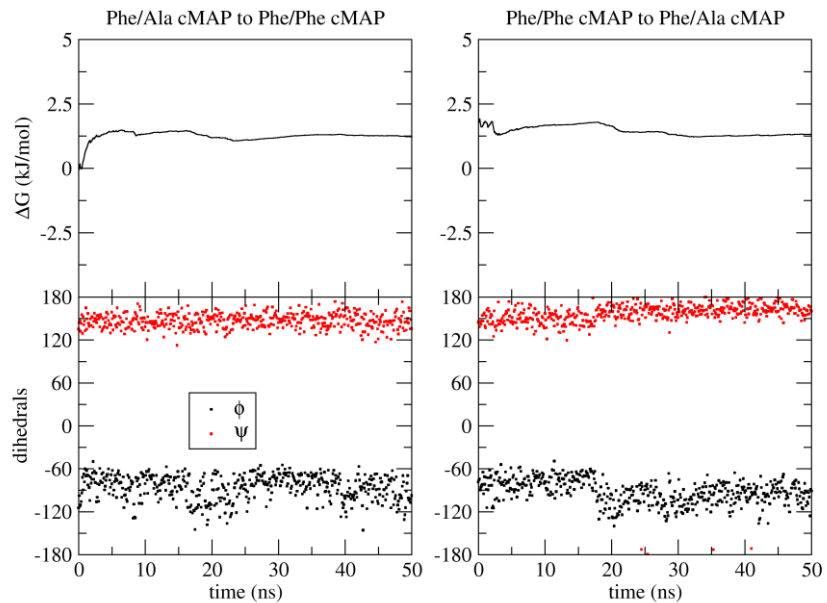
ALA to GLU (Complex)



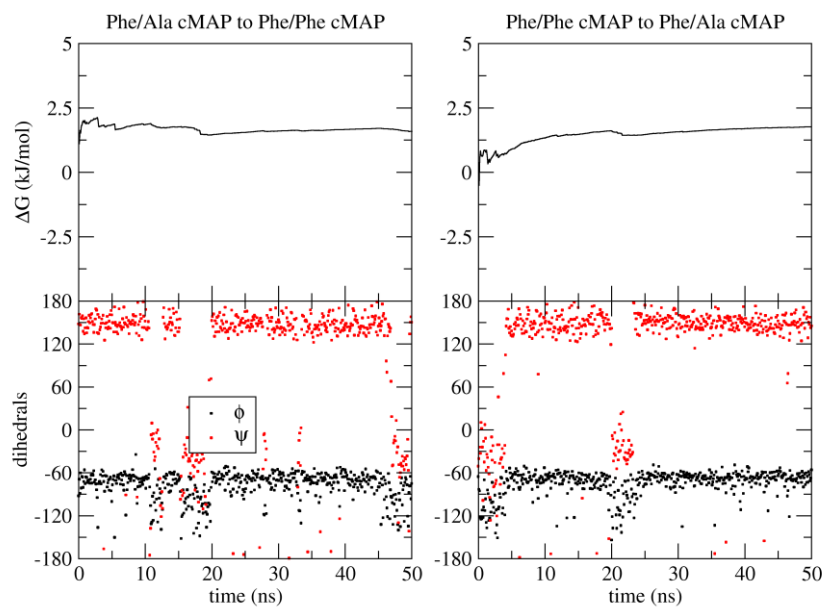
ALA to GLU (Peptide)



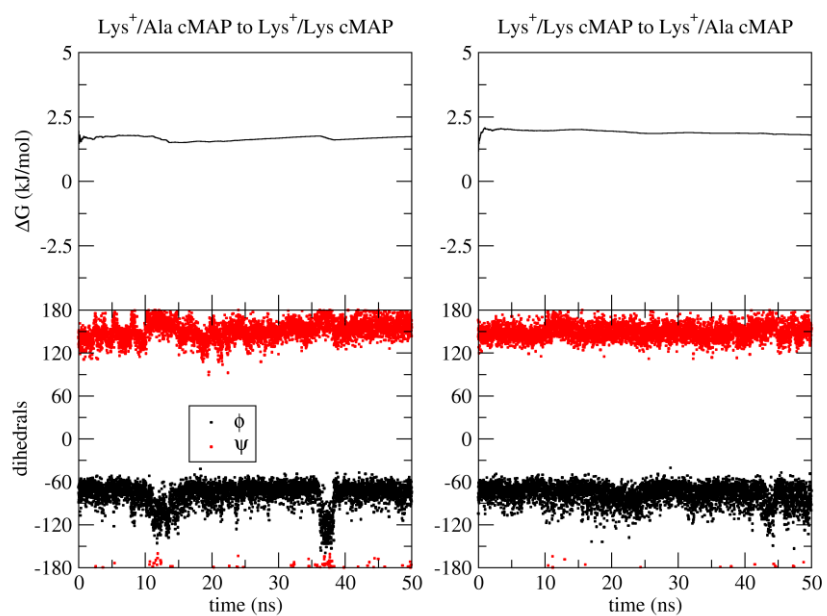
ALA to PHE (Complex)



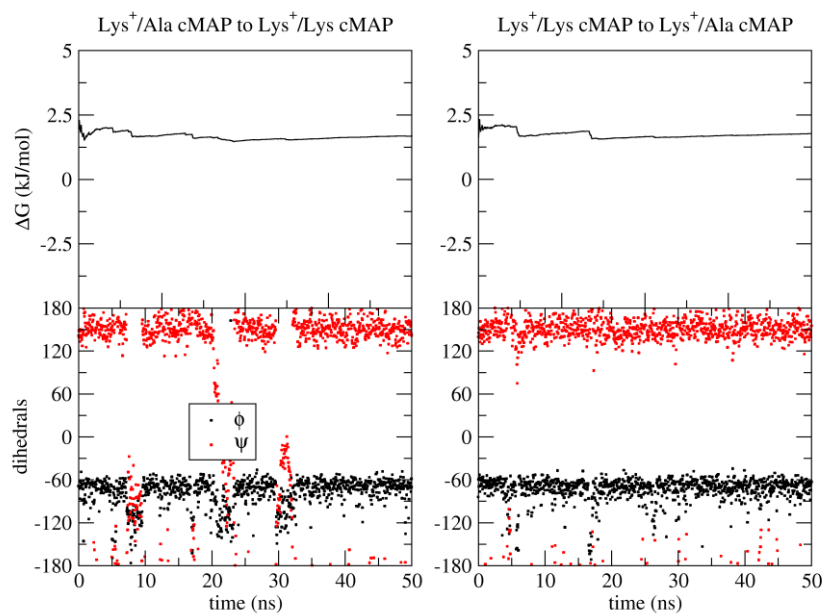
ALA to PHE (Peptide)



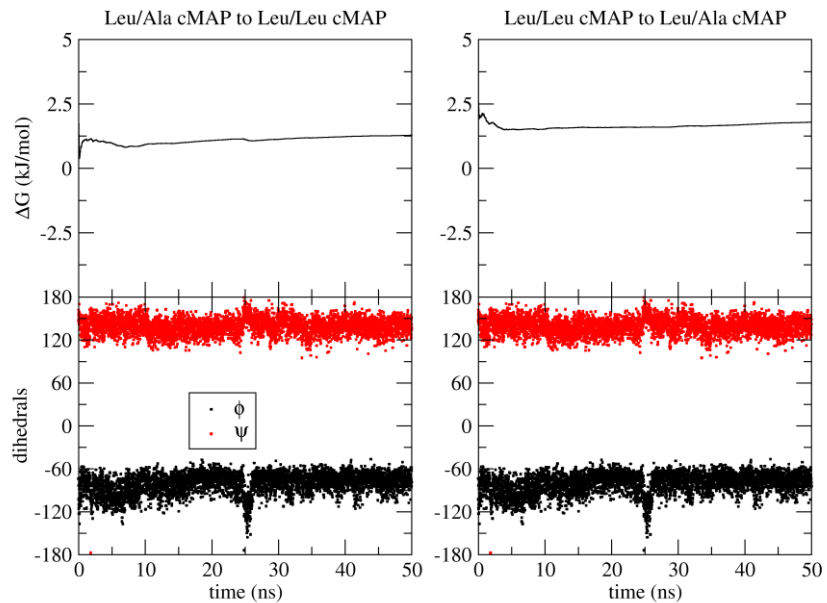
ALA to LYS (Complex)



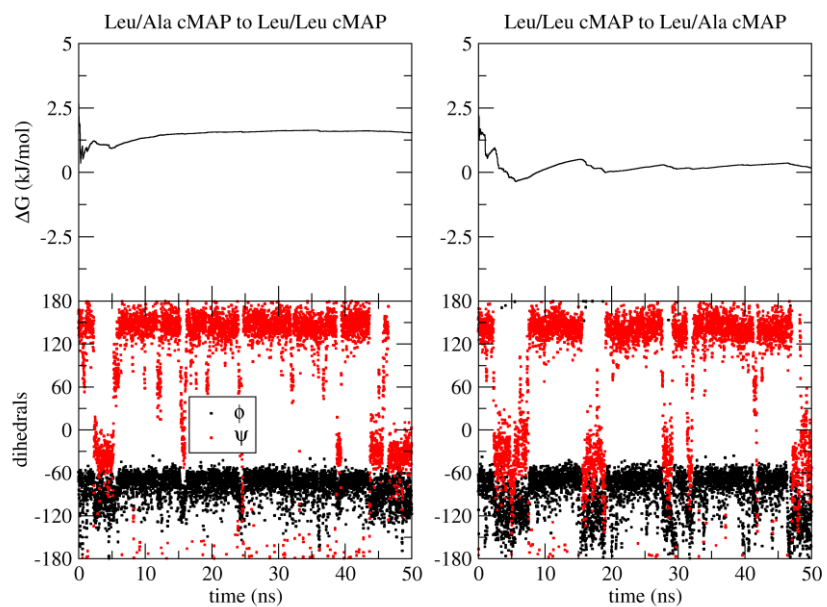
ALA to LYS (Peptide)



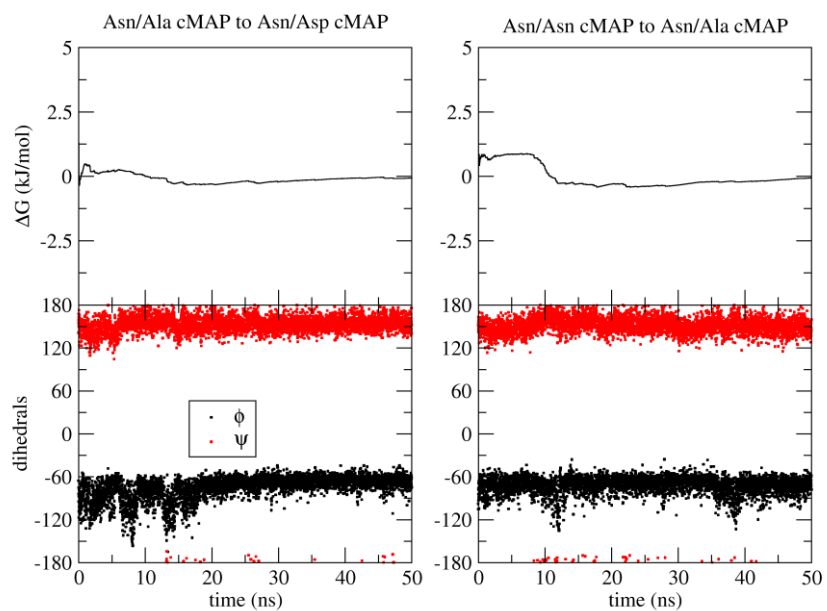
ALA to LEU (Complex)



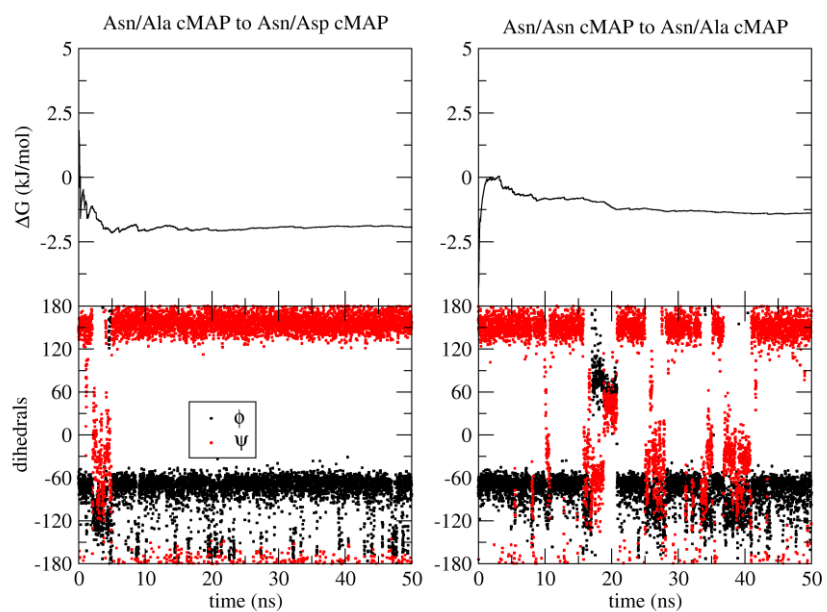
ALA to LEU (Peptide)



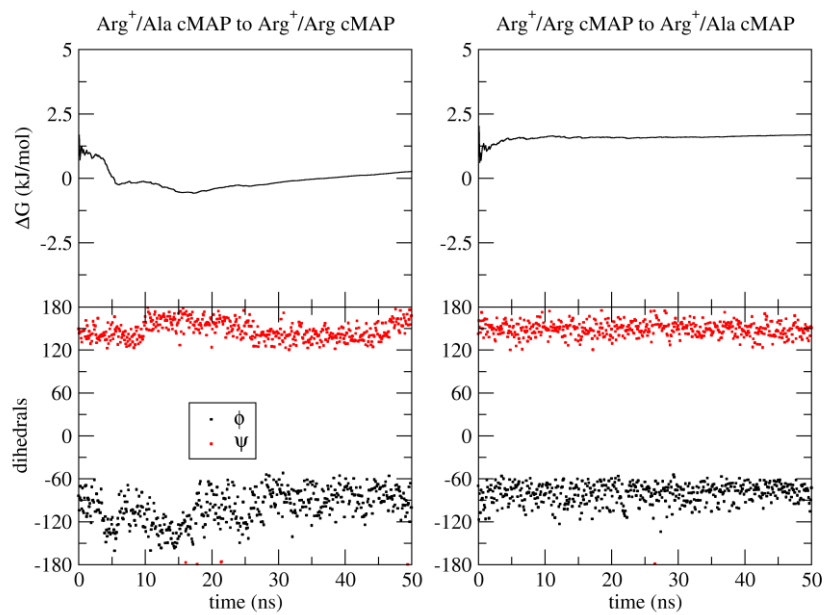
ALA to ASN (Complex)



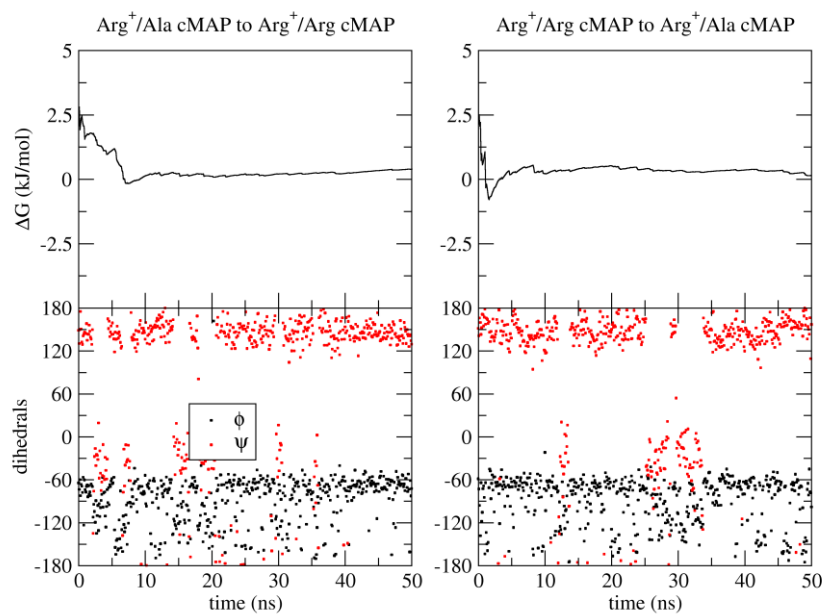
ALA to ASN (Peptide)



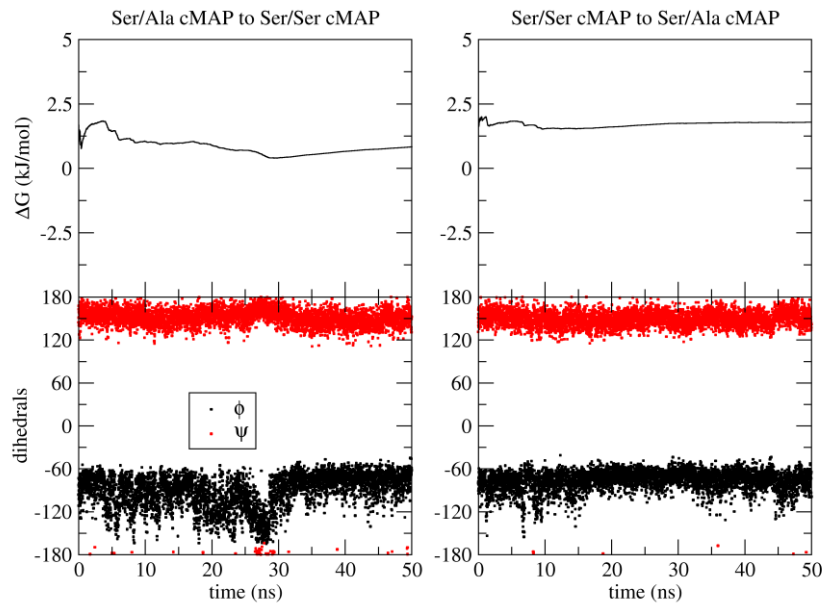
ALA to ARG (Complex)



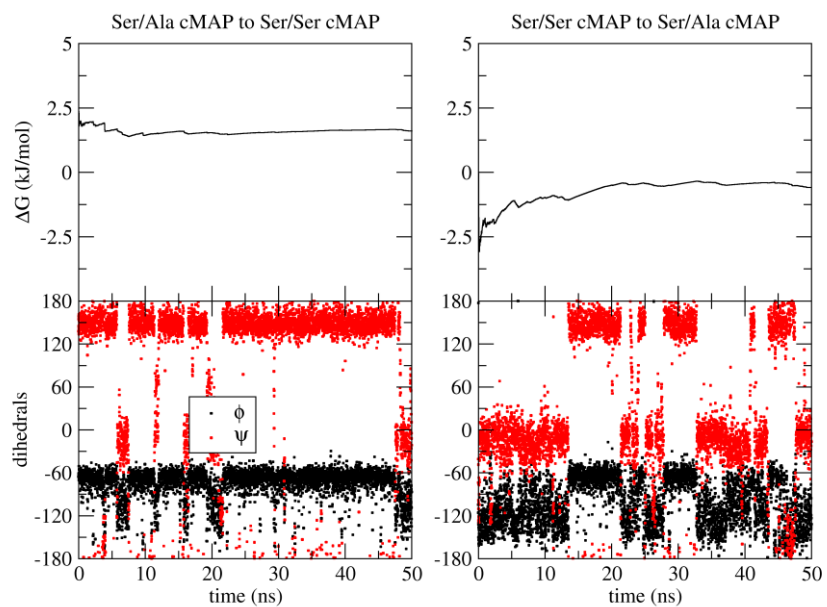
ALA to ARG (Peptide)



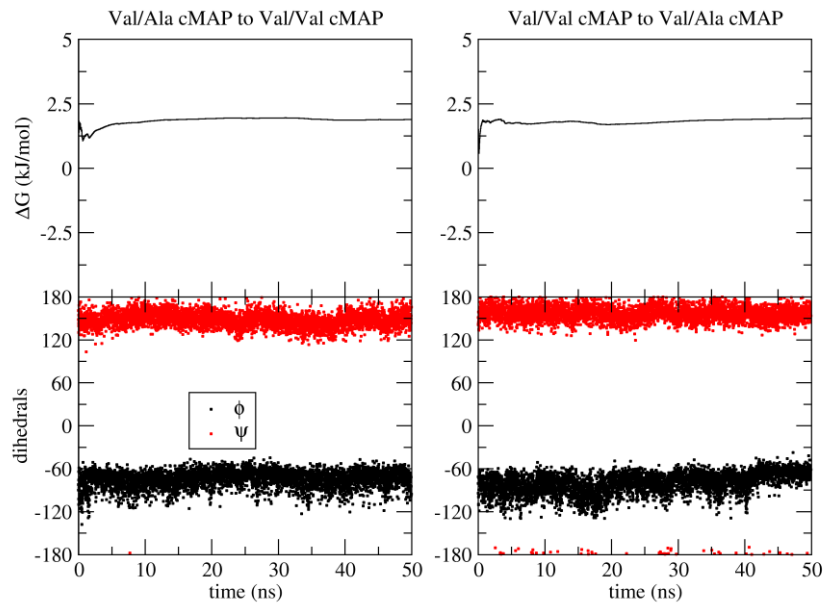
ALA to SER (Complex)



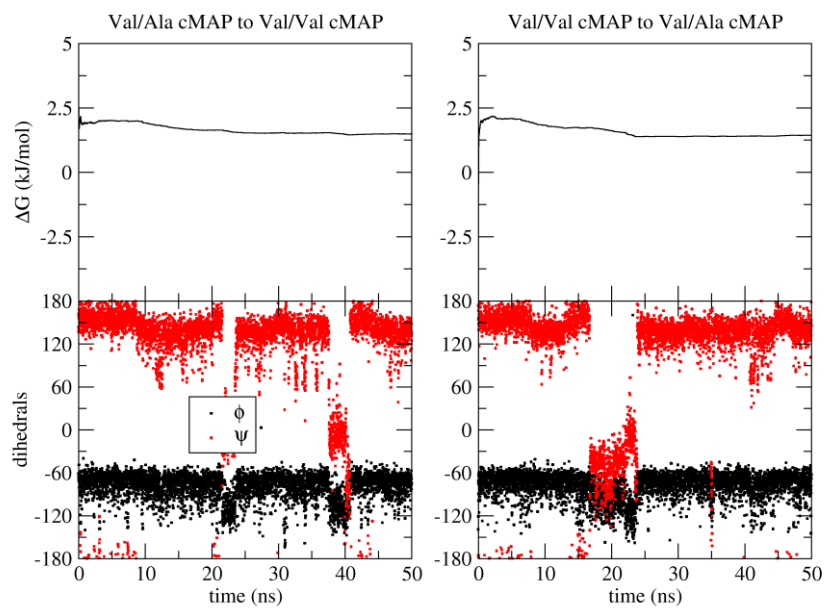
ALA to SER (Peptide)



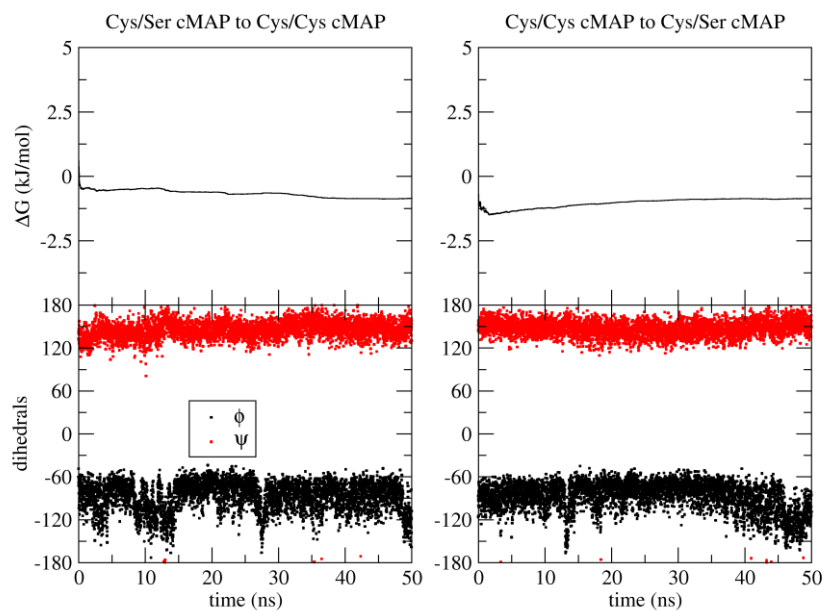
ALA to VAL (Complex)



ALA to VAL (Peptide)



SER to CYS (Complex)



SER to CYS (Peptide)

