The Dissertation Committee for Andrew William Ritchie Certifies that this is the approved version of the following dissertation:

<Title of Dissertation or Treatise, Double-spaced, and centered, may be in ALL CAPITALS or Upper and Lower Case>

Committee:

Lauren J. Webb, Supervisor

Ron Elber

Walter Fast

Graeme Henkelman

Pengyu Ren

<Title of dissertation or treatise, double-spaced, may be in ALL CAPITALS or Upper and Lower Case, and centered>

by

Andrew William Ritchie, B.S. Chem

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctorate of Philosophy

The University of Texas at Austin

August 2015

<Title of dissertation or treatise, double-spaced, and centered>

Andrew William Ritchie, PhD.

The University of Texas at Austin, 2015

Supervisor: Lauren J. Webb

<Abstract: May not exceed 350 words. It should be a continuous description, not disconnected notes or an outline.>

Table of Contents

List of Tables viii

List of Figures ix

List of Code and Parameters x

Chapter 1 Introduction 1

Chapter 2 Simulation Methods 2

2.1 Labeling and Mutating Proteins *in silico* 2

2.2 Enhanced Molecular Dynamics in Amber03: N-Dimensional Umbrella Sampling and Weighted Histogram Analysis Method 6

2.3 Electrostatic Clustering in Vibrational Chromophore Dihedral Space 17

2.4 Probe Parameterization for AMOEBA 22

2.5 Small Molecule Simulations in AMOEBA 34

Chapter 3 Electrostatic Field Methods 35

3.1 Amber03 with Explicit TIP3P Water 35

3.1.1 Reaction Field Electrostatics 35

3.1.2 Hybrid Solvent Reaction Field Electrostatics and Solute Coulomb Field 35

3.2 Amber03 with Poisson-Boltzmann Continuum Solvent 36

3.2.1 Reaction Field Method 36

3.2.2 Grid spacing and size 36

3.2.3 Box Location 36

3.3 Amber03 with Poisson-Boltzmann Continuum Solvent and Select Explicit TIP3P Water Molecules 37

3.3.1 5 Å Water Sphere Around the Vibrational Chromophore 37

3.3.2 Single Water Molecule Nearest the Vibrational Chromophore 37

3.3.3 Water Molecular Hydrogen Bonding to the Vibrational Chromophore 37

3.4 AMOEBA 38

3.4.1 Poisson-Boltzmann Continuum Solvent 38

3.4.2 Explicit AMOEBA Water 38

3.4.3 Charge Penetration Field Corrections 38

Chapter 4 The Role of Electrostatics in Differential Binding of RalGDS to Rap Mutations E30D and K31E Investigated by Vibrational Spectroscopy of Thiocyanate Probes 39

4.1 Introduction 39

4.2 Results 44

4.3 Discussion 54

Chapter 5 Optimizing Electrostatic Field Calculations with the Adaptive Poisson-Boltzmann Solver to Predict Electric Fields at Protein-Protein Interfaces I: Sampling and Focusing 60

5.1 Introduction 60

5.2 Results 60

5.3 Discussion 60

Chapter 6 Optimizing Electrostatic Field Calculations with the Adaptive Poisson-Boltzmann Solver to Predict Electric Fields at Protein-Protein Interfaces II: Explicit Near-Probe and Hydrogen Bonding Water Molecules 61

6.1 Introduction 61

6.2 Results 61

6.3 Discussion 61

Chapter 7 Electrostatic Fields at Protein-Protein Interfaces: Increased Sampling Time and Various Electrostatic Methods: A Case for Simulating in Polarizable Force Fields 62

7.1 Introduction 62

7.2 Results 62

7.3 Discussion 62

Chapter 8 Electrostatic Fields in Small Thiocyanate Molecules with Ensembles Generated using the AMOEBA Force Field 63

8.1 Introduction 63

8.2 Results 63

8.3 Discussion 63

Appendix 64

Glossary 65

References 66

List of Tables

Table 2-1: Sample Parameters for Ryckaert Bellemans dihedral potential function used for validating 2D WHAM code 12

Table 2-2: Umbrella Sampling Parameters, grouped by chapter 13

List of Figures

Figure 2-1: 2D WHAM Validation.....15

Figure 2-2: PB Solvent Reaction Field vs. Solute Analytic Coulomb Field 19

Figure 2-3: Field Values using Clustering Vs. Field Values using All Frames.....20

Figure 2-4: Correlations and Slopes at Various Cutoff Values 21

List of Code and Parameters

Code and Parameters 2-1: Cyanocysteine AMOEBA Parameters 24

Code and Parameters 2-2: Methyl Thiocyanate AMOEBA Parameters.....26

Code and Parameters 2-3: Ethyl Thiocyanate AMOEBA Parameters.....28

Code and Parameters 2-4: Hexyl Thiocyanate AMOEBA Parameters 33

Chapter 1 Introduction

<Body text to begin here.>

Chapter 2 Simulation Methods

2.1 Labeling and Mutating Proteins *in silico*

Simulation parameters for cyanocysteine and GDPNP were obtained from previous studies[1, 2]. Simulations were performed using the Amber03 force field in Gomacs.[3] Ral starting structures came from the 2RGF[4] crystal structure, Ras starting structures came from 1LFD[5], and Rap starting structures came from 1GUA[6].

Generation of Ralβ starting structures has previously been reported.[2] In short, all cysteine sidechain atoms except for Cβ were deleted from 2RGF and the sidechain was renamed to alanine. The C-terminal residues RTFT were taken from the pdb structure 1RAX (deposited in the PDB but unpublished) which ends in KKRTFT, and pasted onto the 2RGF structure, which ends in KKRT, by aligning the backbone atoms of the common KKRT residues and adding the FT coordinates onto the Ral structure file. The first residues in the 2RGF structure are ALA and LEU; these were changed to SER and HIS by renaming the residues in the pdb file, deleting hydrogen atoms, and renaming, using CD1 of LEU as ND1 of HIS. The N-terminal glycine (GLY 4) was modeled using Avogadro, completing the sequence. All missing heavy atoms were added using the tleap utility of Amber Tools.[7] The end resultant sequence is, starting from the N-terminal, GSH+Ral(2RGF,Cys→Ala)+FT.

To make Ras, the GAMGS sequence from chain B of 4K81[8] was used due to previous work which showed it to be the lowest energy conformation of this sequence available in the Protein Data Bank when bound to Ras.[9] The N-terminal methionine of 1LFD was aligned to the methionine immediately following the GAMGS sequence in 4K81, afterwhich the GAMGS coordinates were added to the Ras structure file. The end resultant sequence is GAMGS+Ras(1LFD).

Rap starting structures and mutations have also already been reported.[10] To start, a GSH tag left on the N-terminal methionine after cleavage of the hexa-histidine affinity label during protein purification was modeled onto the N-terminus. This was done by searching the Protein Data Bank for proteins starting with the sequence GSHM (Met is the first Rap residue present in 1GUA). 87 NMR structures and 1 crystal structure were obtained from pdbs 1AQ5 (20 NMR structures)[11], 1W9R (19 NMR structures)[12], 2WCY (48 NMR structures)[13], and 2VKJ (1 crystal structure)[14]. Rap was aligned to the methionine backbone of each GSHM structure using VMD[15], creating 88 structures containing the GSHM N-terminus. After adding hydrogen atoms using the GROMACS utility pdb2gmx[16], an energy minimization was performed and the lowest energy structure from this collection was chosen as the Rap model to be used for further calculations. The end resultant sequence is GSHM+Rap(1GUA). It was observed that in some structures, the N-terminal resides of Ralβ protruded between bonded atoms of Rap. To eliminate this nonphysical steric overlap, heavy atom restraints were placed on all but the Ralβ N-terminal GSHM residues and a 500 step gentle minimization was performed in Gromacs[16].

All side chain mutations were generated in the same manner using Amber Tools.[7] All side chain atoms except for shared heavy atoms were deleted from the mutation residue. The wild type residue was renamed to the desired residue and the resulting structure was passed to the tleap utility in Amber Tools[7] to model back in the missing atoms. In this way, starting from 1LFD Ras D30/E31K and 1GUA Rap E30D/K31E the additional Ras constructs D30/E31, D30E/E31, and D30E/E31K, and the additional Rap constructs E30/K31, E30D/K31, and E30/K31E were each constructed.

To generate cyanylated Ral structures $N27C_{SCN}$, $G28C_{SCN}$, $N29C_{SCN}$, $Y31C_{SCN}$, $K32C_{SCN}$, and $N54C_{SCN}$, residues were renamed to MET in the pdb file (because

methionine contains the same number of heavy atoms as cyanocysteine), retaining and renaming any atoms common to the native residue and methionine. Missing methionine atoms were added using tleap.[7] To complete the mutation to cyanocysteine, the MET atoms CG, SD, and CE were renamed SG, CD, and NE and a short energy minimization was performed.

To dock Rap to Ral, the alpha carbons of the GTPase units of 1LFD and 1GUA were first aligned using the Smith-Waterman algorithm[17] with a gap penalty of -3 (chosen because it gave results most consistent with the STAMP[18] structural alignment in VMD[15]), and the Ral coordinates were merged with the Rap coordinates of 1GUA and saved as a reference structure, Rap(1GUA)+Ral(1LFD). The mutated Rap structure was then aligned to the Rap unit of the Rap(1GUA)+Ral(1LFD) reference structure and the coordinates of the resulting Rap(mutant)+Ral(1LFD) were saved. To dock mutated Ras structures to Ral, the mutated Ras was aligned to the 1LFD Ras and the Ral coordinates were merged with the mutated Ras structure and the coordinates of the resulting Rap(mutant)+Ral(1LFD) were saved. To introduce the probe to the docked system, the cyanylated Ral was aligned to the Ral of each GTPase(mutant)+Ral(1LFD) reference structure, and the GTPase(mutant) coordinates were merged with the cyanylated Ral coordinates to create each GTPase+Probe construct: Ral+N27C_{SCN} , Ral+G28C_{SCN} , Ral+N29C_{SCN} , Ral+Y31C_{SCN} , Ral+K32C_{SCN} , Ral+N54C_{SCN} , Rap E30/K31+N27C_{SCN} , Rap E30/K31+G28C_{SCN} , Rap E30/K31+N29C_{SCN} , Rap E30/K31+Y31C_{SCN} , Rap E30/K31+K32C_{SCN} , Rap E30/K31+N54C_{SCN} , Rap E30/K31E+N27C_{SCN} , Rap E30/K31E+G28C_{SCN} , Rap E30/K31E+N29C_{SCN} , Rap E30/K31E+Y31C_{SCN} , Rap E30/K31E+K32C_{SCN} , Rap E30/K31E+N54C_{SCN} , Rap E30D/K31+N27C_{SCN} , Rap E30D/K31+G28C_{SCN} , Rap E30D/K31+N29C_{SCN} , Rap E30D/K31+Y31C_{SCN} , Rap E30D/K31+K32C_{SCN} , Rap E30D/K31+N54C_{SCN} , Rap E30D/K31E+N27C_{SCN} , Rap

E30D/K31E+G28C_{SCN} , Rap E30D/K31E+N29C_{SCN} , Rap E30D/K31E+Y31C_{SCN} , Rap E30D/K31E+K32C_{SCN} , Rap E30D/K31E+N54C_{SCN} , Ras D30/E31+N27C_{SCN} , Ras D30/E31+G28C_{SCN} , Ras D30/E31+N29C_{SCN} , Ras D30/E31+Y31C_{SCN} , Ras D30/E31+K32C_{SCN} , Ras D30/E31+N54C_{SCN} , Ras D30E/E31+N27C_{SCN} , Ras D30E/E31+G28C_{SCN} , Ras D30E/E31+N29C_{SCN} , Ras D30E/E31+Y31C_{SCN} , Ras D30E/E31+K32C_{SCN} , Ras D30E/E31+N54C_{SCN} , Ras D30/E31K+N27C_{SCN} , Ras D30/E31K+G28C_{SCN} , Ras D30/E31K+N29C_{SCN} , Ras D30/E31K+Y31C_{SCN} , Ras D30/E31K+K32C_{SCN} , Ras D30/E31K+N54C_{SCN} , Ras D30E/E31K+N27C_{SCN} , Ras D30E/E31K+G28C_{SCN} , Ras D30E/E31K+N29C_{SCN} , Ras D30E/E31K+Y31C_{SCN} , Ras D30E/E31K+K32C_{SCN} , and Ras D30E/E31K+N54C_{SCN} .

2.2 Enhanced Molecular Dynamics in Amber03: N-Dimensional Umbrella Sampling and Weighted Histogram Analysis Method

An umbrella sampling strategy was used to obtain a Boltzmann-weighted statistical ensemble of thiocyanate probe orientations for all MD sampling. Through examining simulated protein structures, it became increasingly apparent that a second degree of freedom, χ_1, was relevant to our probe conformational distributions. Therefore, two different umbrella sampling strategies were tested: one-dimensional sampling about the thiocyanate χ_2 dihedral angle and two-dimensional umbrella sampling about the thiocyanate χ_1 and χ_2 dihedral angles, shown in Figure 2-1. All molecular dynamics were completed using the GROMACS[16] software package at 300 K with the AMBER03[19] force field and periodic boundary conditions.

Six probe locations on RalGDS were examined: $N27C_{SCN}$, $G28C_{SCN}$, $N29C_{SCN}$, $Y31C_{SCN}$, $K32C_{SCN}$, and $N54C_{SCN}$, in the monomeric state and docked to each GTPase system examined. We have therefore examined all probe locations and mutated constructs for which experimental data are available. Six structures for each system modeled were generated by fixing the thiocyanate χ_2 dihedral angle from 0° to 300° in 60° increments. Each structure was sampled with a dihedral potential that was flat within ± 30° of the fixed-dihedral position and quadratic with a force constant of 1000 kJ mol^{-1} rad^{-2} otherwise. These restraining potentials were carried through for the duration of the system set-up and simulation. Each structure was energy minimized with cut-off electrostatics, solvated with tip3p water[11] in a dodecahedron box, charge balanced by randomly replacing the appropriate number of water molecules with sodium or chloride ions using the genion GROMACS utility, and solvent relaxed by sampling for 20 ps with position restraints on all non-solvent heavy atoms with a force constant of 1000 kJ mol^{-1} nm^{-2} using PME[20, 21] electrostatics with a real-space cut-off of 0.9 nm, spacing of 0.12 nm,

and interpolation order 4. Each rotamer of each system was then sampled using the GROMACS stochastic dynamics integrator, constraints on hydrogen-bonds using the LINCS algorithm[22], and PME electrostatics for 3 ns, recording snapshots every 5 ps, for a total of 18 ns of simulation and 3606 frames for each system. To test for structural rearrangements at positions 30 and 31 of Rap that would significantly influence the measured electrostatic field, the χ_1 dihedral angle on Rap residues 30 and 31 were each individually biased to create a Boltzmann-weighted ensemble of χ_1 rotomers. Rotomer libraries of Rap E30/K31, Rap E30D/K31, Rap E30/K31E, and the double mutant Rap E30D/K31E were created using the strategy described above. Once again, structures containing dihedral angles from 0˚ to 300˚ in 60˚ degree increments were generated for each mutant; these were energy minimized, solvated in tip3p water, relaxed, and then sampled for 3 ns each, generating 18 ns of simulation for each biased torsional angle. To summarize, for 24 mutant-containing rotomer structures, MD sampling included 18 ns of simulation biasing the nitrile probe, 18 ns of simulation biasing of Rap position 30, and 18 ns of simulation biasing of Rap position 31, for a total of 1296 ns of MD simulation of this system

Each frame was assigned to one of 72 5˚ bins from -180˚ to 175˚ based on the χ_2 dihedral angle. The weighted histogram analysis method (WHAM)[23, 24] was then used to calculate a torsional potential of mean force (PMF) for each of N bins i, which is related to the torsional probability distribution for each bin i (P_i) described by equation (2-1):

$$\mathrm{P}_i = \frac{e^{-\beta \cdot PMF_i}}{\sum_{j=0}^{N} e^{-\beta \cdot PMF_j}}, \ \beta = \frac{1}{k_b T} \tag{2-1}$$

which is the typical Boltzmann distribution function for a state i divided by the partition function, where T is the temperature in Kelvin, k_b is the Boltzmann constant, and PMF_j is

the PMF for some state *j*. In this way, a Boltzmann-weighted distribution of structures was assembled in order to determine the average orientation of both the thiocyanate residues and the side chains at Rap positions 30 and 31 for comparing to experimental data.

The two-dimensional umbrella sampling was done in much the same way as the one-dimensional sampling, with few minor changes. The χ_1 angle was fixed from 0˚ to 300˚ in 30˚ increments, resulting in 12 structures. Each of these structures then had the χ_2 angle fixed from 0˚ to 300˚ in 30˚ increments, resulting in 144 total structures. To avoid steric clashes in the starting structures, for each χ_2 rotation, the distance between the center of mass coordinate of each rotated atom and every non-rotated atom was calculated. If a distance was found to be less than 1.5 Å, the χ_2 angle was rotated ± 1.5˚ from the dihedral center and the distances were recalculated. This was done until all non-bonded atoms were at least 1.5 Å from each rotated atom. Next, two harmonic dihedral restraining potentials were generated for each structure, one for the χ_1 dihedral angle and one for the χ_2 dihedral angle. Following the same set-up strategy used in the one-dimensional sampling, the system was then energy minimized with cut-off electrostatics and dihedral force constants of 1000 kJ mol^{-1} rad^{-2}, solvated in tip3p water in a dodecahedron box, and charge balanced as described above. The system underwent solvent relaxation using PME electrostatics for 20 ps with a force constant of 1000 kJ mol^{-1} nm^{-2} on heavy backbone atoms, dihedral force constants of 150 kJ mol^{-1} rad^{-2}, and unrestrained side-chain atoms. Each of the 144 χ_1-χ_2 rotamers were then sampled for 400 ps using the GROMACS stochastic dynamics integrator with PME electrostatics (again, with a real-space cut-off of 0.9 nm, spacing of 0.12 nm, and interpolation order 4) and dihedral restraining potentials of 70 kJ mol^{-1} rad^{-2}, for a total of 57.6 ns of simulation and 11664 frames for each system. It is worth reiterating that each step used a progressively

smaller dihedral restraining potential. This was done to ensure that the dihedral angles of the final structure before sampling were as close as possible to the umbrella-sampling window while still allowing nearby residues to relax to orientations that accommodate the inclusion of our probe. Starting with a large restraining potential fixes the probe to a specific location orientation and forcibly moves nearby residues to accommodate the probe to minimize interaction energies. Subsequent weakening of the restraining potential allows the probe to respond to its surroundings in a manner more typical of MD, allowing both the probe as well as the residues near the probe to relax to energy minimized orientations. Without this subsequent weakening, many simulations resulted in dihedral forces becoming larger than the integrator can or is expected to handle, which were usually caused by steric clashes between the probe and side-chain atoms.

A Boltzmann-weighted statistical ensemble of structures obtained from two-dimensional sampling was then assembled using WHAM. Each frame was assigned to one of 5184 5˚ by 5˚ bins. The bins were assigned based on equation (2-2), where b_1 is the one-dimensional χ_1 bin number, b_2 is the one-dimensional χ_2 bin number, B_i is the total number of bins in degree of freedom i (72 for all two-dimensional sampling done), and i_{index} refers to the subscript on b. Conditional probability was assumed.

$$\mathrm{Bin}(b_1,b_2) = \sum_{i=b_1,\,b_2} B_i^{2-i_{\mathrm{index}}} \cdot i \qquad (2\text{-}2)$$

To validate our 2D WHAM code, we examined a Ryckaert-Bellemans dihedral potential, shown in equation (2-3), where ϕ is some angle and n and C_n are some example parameters obtained from the GROMACS manual, shown in Table 2-1. This is a very simple potential function with no contributions from any other source, unlike the potential energy calculation in a protein, which will be influenced by various force field parameters such as bond force constants. However in WHAM, the sources of the

potentials are not distinguished, and we are able to use the simple Ryckaert-Bellemans model to validate the code. We constructed the PMF from the sum of the potentials for a given pair of coordinates and the unbiased probability distribution, $p^{\circ}(\chi_1, \chi_2)$, using the ratio of the Boltzmann distribution function to the partition function, shown in equation (2-4), over an array of 2D dihedral angles ranging form -180° to 180°.

$$V_{rb} = \sum_{n=0}^{5} C_n \left(\cos\left(\phi - \pi\right)\right)^n \tag{2-3}$$

Next, we constructed the PMF and probability distribution landscapes, shown in Figure 2-1a. We then applied biasing potential windows to each dimension in a manner that mimics the methods used in GROMACS. The biased probability, $p(\chi_1, \chi_2)$, is given by equation (2-5). We then performed a Monte Carlo simulation centered on each of 144 biased windows, each with dimensions of 30° x 30°, with a probability $p(\chi_1, \chi_2)$ of sampling a given pair of dihedral angles, which was then analyzed using our 2D WHAM code to return the unbiased PMF and $p^{\circ}(\chi_1, \chi_2)$.

$$p^{\circ}(\chi_1, \chi_2) = \frac{e^{-\beta\left(V_{rb}(\chi_1)+V_{rb}(\chi_2)\right)}}{\int\int e^{-\beta\left(V_{rb}(\chi_1)+V_{rb}(\chi_2)\right)} d\chi_1 d\chi_2} \tag{2-4}$$

$$p(\chi_1, \chi_2) = \frac{e^{-\beta\left(V_{rb}(\chi_1)+V_1+V_{rb}(\chi_2)+V_2\right)}}{\int\int e^{-\beta\left(V_{rb}(\chi_1)+V_1+V_{rb}(\chi_2)+V_2\right)} d\chi_1 d\chi_2} \tag{2-5}$$

Figure 2-1 shows the analytical PMF and probability distributions (a), WHAM PMF and probability distributions on 144 windows each containing 40 frames (b), WHAM PMF and probability distributions on 144 windows each containing 80 frames (c), WHAM PMF and probability distributions on 144 windows each containing 160 frames (d), and WHAM PMF and probability distributions on 144 windows each containing 1000 frames (e). The major features of the probability distribution become clear after only 40 frames; after 80 frames the probability distributions look very similar to the analytical distribution, and after 160 frames very little improvement is seen. We

can also see that the high probability regions, representing the staggered orientations, have PMF landscapes that look like the analytical PMF, although the gauche regions appear to be ill characterized still. After 1000 frames the probability distribution is nearly identical to the analytical distribution and the moderate ranges of the PMF (light blue, > 17.92 kJ mol^{-1}) now quantitatively approach the predictions of the analytical expression. Exact analytical PMF matching of low probability regions is very slow, requiring ≥10000 frames. From these results we conclude that the PMF converges to the analytical expression slowly for regions of low probability and quickly for regions of high probability. Table 2-2 summarizes the umbrella sampling setups for each project, listed by chapter.

Table 2-1: Sample Parameters for Ryckaert Bellemans dihedral potential function used for validating 2D WHAM code

	Constraint (kJ mol^{-1})		**Constraint (kJ mol^{-1})**
$\mathbf{C_0}$	9.28	$\mathbf{C_3}$	-3.06
$\mathbf{C_1}$	12.16	$\mathbf{C_4}$	-26.24
$\mathbf{C_2}$	-13.12	$\mathbf{C_5}$	-31.5

Table 2-2: Umbrella Sampling Parameters, grouped by chapter

Chapters	Biasing Coordinate	Umbrella Window Size	Biasing Potential Shape	Windows per Experiment	Simulation Time per Experiment	Total Simulation Time
4, 5	CNC χ_2	60˚	Quadratic-Flat-Quadratic	6	3.0 ns	18.0 ns
4	Position 30 χ_1	60˚	Quadratic-Flat-Quadratic	6	3.0 ns	18.0 ns
4	Position 31 χ_1	60˚	Quadratic-Flat-Quadratic	6	3.0 ns	18.0 ns
5, 6	CNC χ_1, χ_2	30˚	Quadratic	144	0.4 ns	57.6 ns
7	CNC χ_1, χ_2	30 ˚	Quadratic	144	2.0 ns	288.0 ns
8	N/A	N/A	N/A	N/A	4-7 ns	4-7 ns

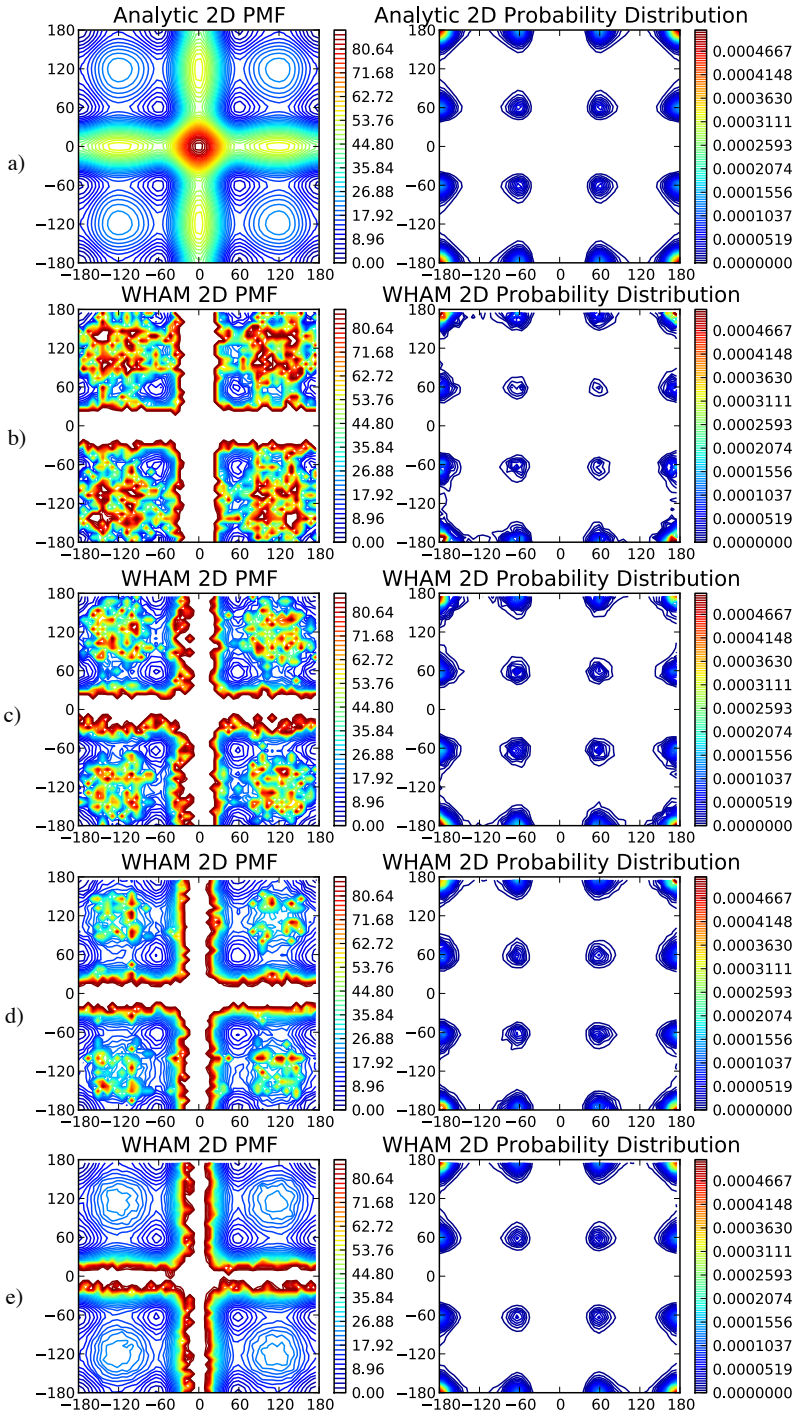

Figure 2-1: 2D WHAM Validation

Comparison between the PMF and probability distributions of a) an analytic Ryckaert-Bellemans dihedral potential and Monte Carlo 2D umbrella sampling for b) 40 frames/biasing window; c) 80 frames/biasing window; d) 160 frames/biasing window; e) 1000 frames/biasing window. Units on the PMF are kJ mol^{-1}.

2.3 Electrostatic Clustering in Vibrational Chromophore Dihedral Space

The largest bottleneck for these sorts of calculations we do are the electrostatics. A single node on Stampede can generate >10 ns of simulation per day. That number can be increased (logarithmically) by using additional nodes. However, the continuum solvent electrostatics calculations take anywhere from 45-60 seconds (APBS) to $\approx 20 minutes (AMOEBA) per frame. If we keep every 4 ps and collect 250 frames per nanosecond, then the electrostatics require 5-8 ns/day for APBS calculations and approximately 0.3 ns/day for AMOEBA calculations. This can be decreased further by running the serial calculations in parallel. Regardless, it would be convenient to find some method of pruning the total number of frames for continuum electrostatics calculations while ensuring that the average field does not differ significantly from the average using every frame.

For convenience, the vacuum electrostatic field at the nitrile due to solute only was chosen as an indicator of total electrostatic field. In the absence of solvent, this is trivial to calculate for both point charge force fields (Amber03) as well as multipole force fields (AMOEBA). This was chosen because it was 1) intuitive and 2) there is consistently a good correlation between the solute Coulomb field and the PB solvent reaction field, as seen in Figure 2-2,git suggesting that frames which well represent the Coulomb field also well represent the reaction field.

We then took advantage of the weighted averaging over binned data. The Boltzmann weighted average is calculated as in equation (2-6),

$$\langle x \rangle = \sum_{i=1}^{\text{nbins}} \rho_i \sum_{j=1}^{c_i} \frac{x_{ij}}{c_i} \tag{2-6}$$

where the probability of being in each bin i is ρ_i, the number of times bin i is visited is c_i, and each value in bin i is x_{ij} for $j = 1$ to $j = c_i$. There exists some subarray of

values in bin i that has k_i values, where $k_i \geq c_i$ entries and $\left| \sum_{j=1}^{k_i} \frac{x_{ij}}{k_i} - \sum_{j=1}^{c_i} \frac{x_{ij}}{c_i} \right| \leq \chi$, where χ is some threshold. As χ approaches 0, k_i approaches c_i, and the subarray is the full array and the averages are identical. Using the a set of test data where umbrella windows were centered every 120 degrees (at 60˚, 180˚, and 300˚, the expected alkane maximum probability torsions) with a flat biasing potential ± 60˚ of the window center and a force constant of **XXX, the clustered average field for each Ral probe in the monomeric state, docked to each of the four Rap1a mutants and each of the four Ras mutants is plotted against their average field for various values of χ, indicated in the upper-left corner of each subplot, from the full data sets in Figure 2-3. From this, it's clear that the clustered averages are linearly correlated to the full averages. The correlation coefficients and best-fit slopes have also been plotted as a function of the cutoff, χ, in Figure 2-4. Even for a relatively large $\chi = 1$, the clustered correlation coefficient and slope is *very* close to 1.0. In general, as χ approaches zero, the correlation coefficient and slope also approach zero. Moreover, at a $\chi = 0.01$, only approximately 20% of all frames are used, which is a significant decrease in computation requirement. This method also has the advantage of guaranteeing that the property the cluster is based on always has a clustered average nearly identical to the full average, which is a useful sanity check.

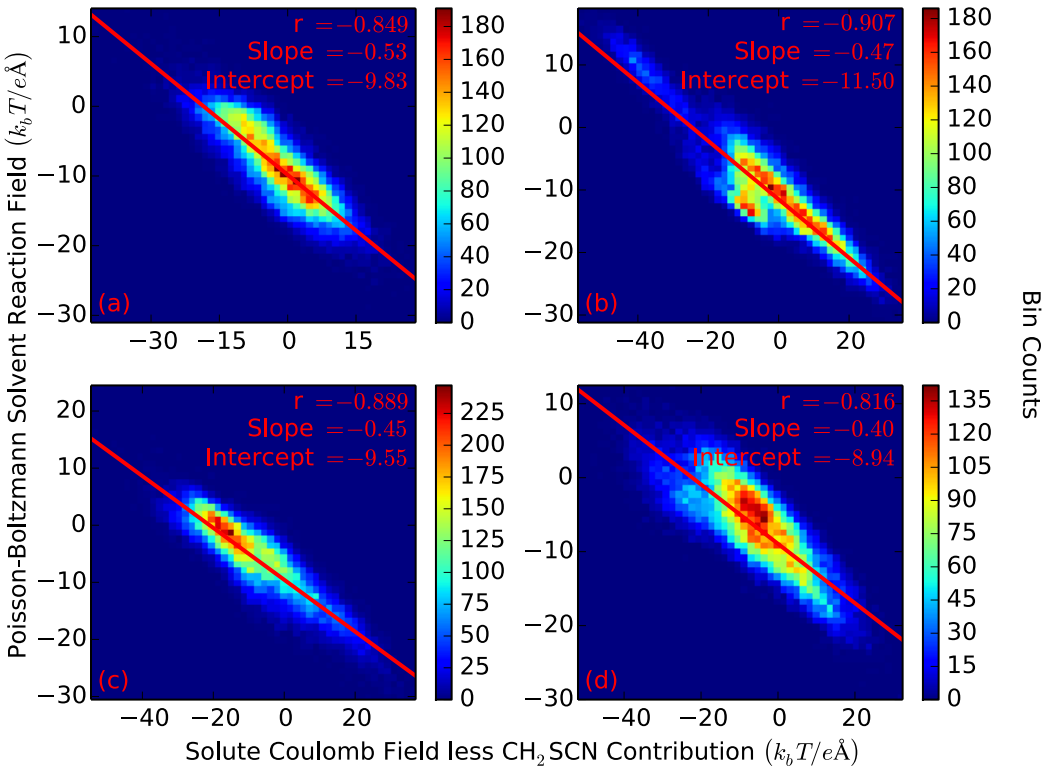

Figure 2-2: PB Solvent Reaction Field vs. Solute Analytic Coulomb Field

Comparison between the analytic Coulomb field at the nitrile bond midpoint due to solute (less the contributions due to the probe itself) (x-axis) and the solvent reaction field at the nitrile bond midpoint (y-axis) using Amber03 point charges for a) Ral $G28C_{SCN}$ monomer; b) Ral $N54C_{SCN}$ monomer; c) Ral $G28C_{SCN}$ docked to wild type Rap; d) Ral $N54C_{SCN}$ docked to wild type Rap. Correlation coefficients (r), slopes (m), and y-intercepts (int), are indicated in the upper-right corner of each figure.

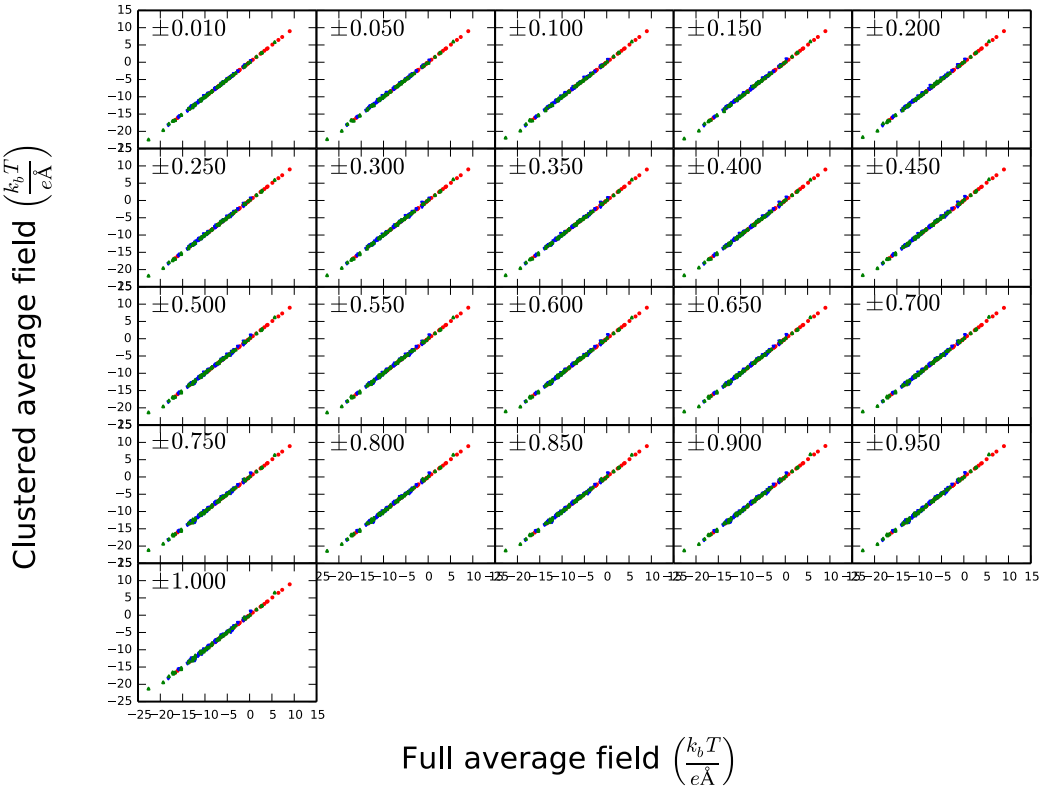

Figure 2-3: Field Values using Clustering Vs. Field Values using All Frames

Average Coulomb electrostatic field (red), solvent reaction field (blue), and the electrostatic field calculated using the AMOEBA force field (green) from clustered frames versus the respective full averages for various cutoff values, χ, indicated in the top left corner of each box.

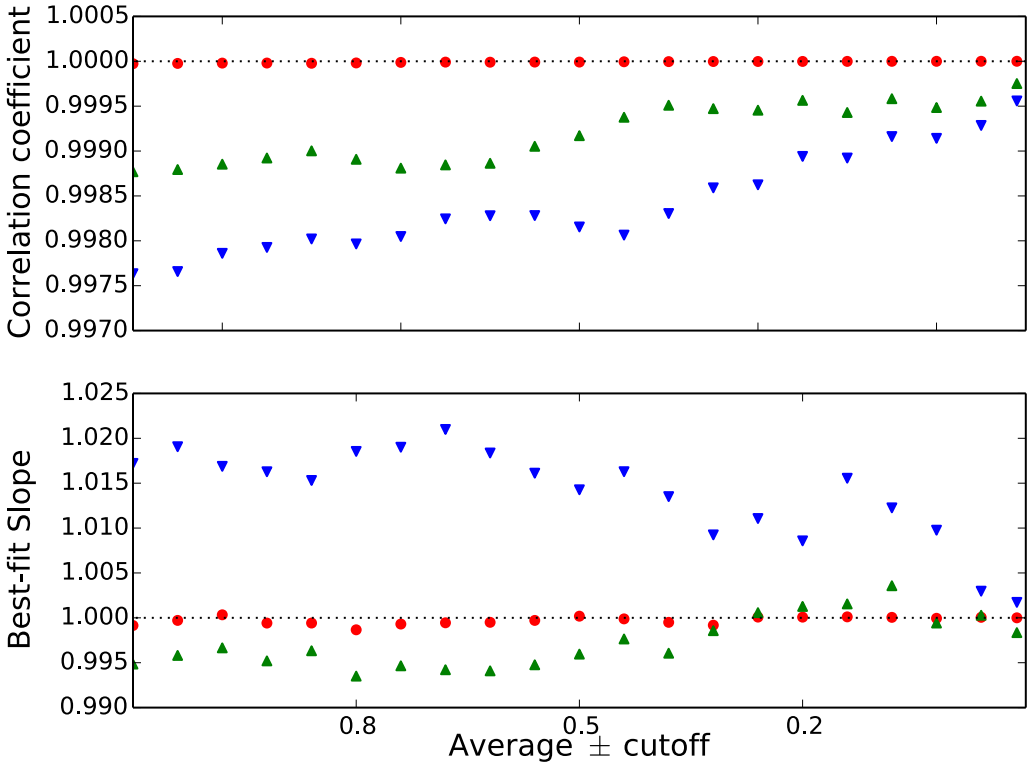

Figure 2-4: Correlations and Slopes at Various Cutoff Values

Correlation coefficients of the Coulomb electrostatic field (red), solvent reaction field (blue), and the electrostatic field calculated using the AMOEBA force field (green) as a function of the cutoff, χ. (Bottom) Best-fit slopes of the Coulomb electrostatic field (red), solvent reaction field (blue), and the electrostatic field calculated using the AMOEBA force field (green) as a function of the cutoff, χ.

2.4 PROBE PARAMETERIZATION FOR AMOEBA

<Body text to begin here.>

```
atom         259   55    CB    "CNC CB"                    6    12.011    4
atom         260   56    SG    "CNC SG"                   16    32.066    2
atom         261   57    CD    "CNC CD"                    6    12.011    2
atom         262   58    NE    "CNC NE"                    7    14.007    1
atom         263   59    HB    "CNC HB"                    1     1.008    1

multipole    259    8  260              -0.15228
                                          0.07407   0.00000   0.31740
                                         -0.15117
                                          0.00000  -0.21945
                                         -0.19282   0.00000   0.37062
multipole    260  259  261               0.33074
                                          0.44389   0.00000   0.28736
                                          1.22369
                                          0.00000  -2.16613
                                         -0.31636   0.00000   0.94244
multipole    261  260  262               0.24556
                                          0.06457   0.00000  -0.41800
                                          0.15740
                                          0.00000   0.22633
                                          0.26601   0.00000  -0.38373
multipole     12    8  259               0.12898
                                          0.02551   0.00000   0.07014
                                          0.19051
                                          0.00000   0.17012
                                         -0.01582   0.00000  -0.36063
multipole    263  259    8               0.09179
                                         -0.07114   0.00000  -0.02080
                                         -0.16061
                                          0.00000  -0.02258
                                         -0.03267   0.00000   0.18319
multipole    262  261  260              -0.58843
                                         -0.00512   0.00000  -0.24238
                                          0.32332
                                          0.00000   0.16836
                                         -0.08202   0.00000  -0.49168

polarize     259          1.3340     0.3900    263
polarize     260          3.3000     0.3900    261
polarize     261          1.3340     0.3900    260  262
polarize     262          1.0730     0.3900    261
polarize     263          0.4960     0.3900    259

vdw           58              3.7100     0.1050
vdw           55              3.7800     0.1060
vdw           56              4.0050     0.3550
vdw           57              3.7800     0.1060
vdw           59              2.8700     0.0330      0.900
bond          55    7        323.0000     1.5317
bond          55   56        235.8000     1.8353
bond          55   59        341.0000     1.0817
bond          56   57        235.8000     1.7060
bond          57   58        450.0000     1.1374
angle          7   55   56    53.2000   109.6948
angle          7   55   59    42.4400   110.8048
angle         56   55   59    60.2400   108.4887
angle         59   55   59    39.5700   108.8231
```

```
58. angle         55    7    3    60.0000   108.7411
59. angle         55    7    1    80.0000   112.9036
60. angle         55    7    6    38.0000   109.5706
61. angle         55   56   57    60.4300    98.5327
62. angle         56   57   58    60.0000   178.6763
63. strbnd         7   55   56    18.7000    18.7000
64. strbnd         7   55   59    11.5000    11.5000
65. strbnd        56   55   59    11.5000    11.5000
66. strbnd        55    7    3    18.7000    18.7000
67. strbnd        55    7    1    18.7000    18.7000
68. strbnd        55    7    6    11.5000    18.7000
69. strbnd        55   56   57    -5.7500    -5.7500
70. torsion       56   55    7    3     -
    1.010 0.0 1   1.230 180.0 2   1.000 0.0 3 # CYS 3 7 8 12
71. torsion       56   55    7    1     -0.160 0.0 1   1.080 180.0 2  -
    1.520 0.0 3 # CYS 1 7 8 12
72. torsion       56   55    7    6      0.000 0.0 1   0.000 180.0 2   0.475 0.0 3 #
    CYS 6 7 8 12
73. torsion       59   55    7    3      0.000 0.0 1   0.000 180.0 2   0.180 0.0 3 #
    CYS 3 7 8 9
74. torsion       59   55    7    1      0.000 0.0 1   0.000 180.0 2   0.500 0.0 3 #
    CYS 1 7 8 9
75. torsion       59   55    7    6      0.000 0.0 1   0.000 180.0 2   0.299 0.0 3 #
    CYS 6 7 8 9
76. torsion        7   55   56   57    -0.4400 0.0 1 -
    0.2600 180.0 2  0.6000 0.0 3 # EtSCN C1 C2 S C
77. torsion       59   55   56   57     0.0000 0.0 1  0.0000 180.0 2  0.6600 0.0 3 #
    EtSCN H2 C2 S C
78. torsion       55    7    3    1      0.929 0.0 1   0.328 180.0 2   0.000 0.0 3 #
    CYS 1 3 7 8
79. torsion       55    7    3    5      0.000 0.0 1   0.000 180.0 2   0.000 0.0 3 #
    CYS 5 3 7 8
80. torsion       55    7    1    3      2.576 0.0 1   1.011 180.0 2   0.825 0.0 3 #
    CYS 3 1 7 8
81. torsion       55    7    1    4      0.000 0.0 1   0.000 180.0 2   0.000 0.0 3 #
    CYS 4 1 7 8
82. torsion       55   56   57   58     0.0000 0.0 1  0.0000 180.0 2  0.5000 0.0 3 #
    EtSCN C2 S C N
```

Code and Parameters 2-1: Cyanocysteine AMOEBA Parameters

```
atom         271   67    H1    "MeSCN H1"                    1     1.008    1
atom         272   68    C1    "MeSCN C1"                    6    12.011    4
atom         273   69    S     "MeSCN S"                    16    32.066    2
atom         274   70    C     "MeSCN C"                     6    12.011    2
atom         275   71    N     "MeSCN N"                     7    14.007    1

multipole    272  273  271              -0.22754
                                        0.00000   0.00000  -0.02449
                                        0.72145
                                        0.00000   0.72145
                                        0.00000   0.00000  -1.44290
multipole    273  272  274               0.33074
                                        0.44389   0.00000   0.28736
                                        1.22369
                                        0.00000  -2.16613
                                       -0.31636   0.00000   0.94244
multipole    274  273  275               0.24556
                                        0.06457   0.00000  -0.41800
                                        0.15740
                                        0.00000   0.22633
                                        0.26601   0.00000  -0.38373
multipole    271  272  273               0.07989
                                        0.02678   0.00000  -0.31473
                                        0.66454
                                        0.00000  -0.00998
                                        0.17231   0.00000  -0.65456
multipole    275  274  273              -0.58843
                                       -0.00512   0.00000  -0.24238
                                        0.32332
                                        0.00000   0.16836
                                       -0.08202   0.00000  -0.49168

polarize     271          0.4960     0.3900    272
polarize     272          1.3340     0.3900    271  273
polarize     273          3.3000     0.3900    272  274
polarize     274          1.3340     0.3900    273  275
polarize     275          1.0730     0.3900    274

vdw           68              3.7800     0.1060
vdw           69              4.0050     0.3550
vdw           70              3.7800     0.1060
vdw           67              2.8700     0.0330      0.900
vdw           71              3.7100     0.1050
bond          68   69        235.8000     1.8209
bond          68   67        341.0000     1.0794
bond          69   70        235.8000     1.7068
bond          70   71        450.0000     1.1373
angle         69   68   67    60.2400   110.5747
angle         67   68   67    39.5700   110.6263
angle         68   69   70    60.4300    99.4461
angle         69   70   71    60.0000   178.9042
strbnd        69   68   67    11.5000    11.5000
strbnd        68   69   70    -5.7500    -5.7500
torsion       67   68   69   70     0.0000 0.0 1  0.0000 180.0 2  0.6600 0.0 3
torsion       68   69   70   71     0.0000 0.0 1  0.0000 180.0 2  0.5000 0.0 3
```

Code and Parameters 2-2: Methyl Thiocyanate AMOEBA Parameters

```
1.  atom         264   60    C1    "EtSCN C1"                    6    12.011    4
2.  atom         265   61    H1    "EtSCN H1"                    1     1.008    1
3.  atom         266   62    C2    "EtSCN C2"                    6    12.011    4
4.  atom         267   63    H2    "EtSCN H2"                    1     1.008    1
5.  atom         268   64    S     "EtSCN S"                    16    32.066    2
6.  atom         269   65    C     "EtSCN C"                     6    12.011    2
7.  atom         270   66    N     "EtSCN N"                     7    14.007    1
8.
9.  multipole    264  266  265             -0.18333
10.                                        0.00000   0.00000   0.32541
11.                                       -0.25019
12.                                        0.00000  -0.25019
13.                                        0.00000   0.00000   0.50038
14. multipole    266  264  268             -0.23503
15.                                        0.06819   0.00000   0.15097
16.                                       -0.26732
17.                                        0.00000  -0.08487
18.                                       -0.08743   0.00000   0.35219
19. multipole    267  266  264              0.09244
20.                                       -0.00077   0.00000  -0.03413
21.                                        0.02305
22.                                        0.00000   0.01348
23.                                       -0.00070   0.00000  -0.03653
24. multipole    265  264  266              0.08187
25.                                        0.01969   0.00000  -0.08459
26.                                        0.06966
27.                                        0.00000   0.02253
28.                                       -0.00032   0.00000  -0.09219
29. multipole    268  266  269              0.33074
30.                                        0.44389   0.00000   0.28736
31.                                        1.22369
32.                                        0.00000  -2.16613
33.                                       -0.31636   0.00000   0.94244
34. multipole    269  268  270              0.24556
35.                                        0.06457   0.00000  -0.41800
36.                                        0.15740
37.                                        0.00000   0.22633
38.                                        0.26601   0.00000  -0.38373
39. multipole    270  269  268             -0.58843
40.                                       -0.00512   0.00000  -0.24238
41.                                        0.32332
42.                                        0.00000   0.16836
43.                                       -0.08202   0.00000  -0.49168
44.
45. polarize     264          1.3340     0.3900    265  266
46. polarize     265          0.4960     0.3900    264
47. polarize     266          1.3340     0.3900    264  267
48. polarize     267          0.4960     0.3900    266
49. polarize     268          3.3000     0.3900    269
50. polarize     269          1.3340     0.3900    268  270
51. polarize     270          1.0730     0.3900    269
52.
53. vdw           60              3.8200     0.1040
54. vdw           62              3.7800     0.1060
```

```
vdw          64                4.0050     0.3550
vdw          65                3.7800     0.1060
vdw          63                2.8700     0.0330      0.900
vdw          61                2.9800     0.0240      0.920
vdw          66                3.7100     0.1050
bond         60   62         345.3000     1.5227
bond         60   61         341.0000     1.0855
bond         62   64         323.0000     1.8329
bond         62   63         341.0000     1.0809
bond         64   65         235.8000     1.7068
bond         65   66         450.0000     1.1376
angle        62   60   61     42.4400   111.1890
angle        61   60   61     39.5700   108.4478
angle        60   62   64     53.2000   114.3035
angle        60   62   63     42.4400   111.8003
angle        64   62   63     60.2400   108.0491
angle        63   62   63     45.5700   107.8321
angle        62   64   65     60.4300    99.8280
angle        64   65   66     60.0000   179.1383
strbnd       62   60   61     11.5000    11.5000
strbnd       60   62   64     18.7000    18.7000
strbnd       60   62   63     11.5000    11.5000
strbnd       64   62   63     11.5000    11.5000
strbnd       62   64   65     -5.7500    -5.7500
torsion      61   60   62   64     0.0000 0.0 1  0.0000 180.0 2  0.4750 0.0 3
torsion      61   60   62   63     0.0000 0.0 1  0.0000 180.0 2  0.2990 0.0 3
torsion      60   62   64   65    -0.4400 0.0 1 -0.2600 180.0 2  0.6000 0.0 3
torsion      63   62   64   65     0.0000 0.0 1  0.0000 180.0 2  0.6600 0.0 3
torsion      62   64   65   66     0.0000 0.0 1  0.0000 180.0 2  0.5000 0.0 3
```

Code and Parameters 2-3: Ethyl Thiocyanate AMOEBA Parameters

```
atom         276   72    C1    "HxSCN C1"                    6    12.011    4
atom         277   73    C2    "HxSCN C2"                    6    12.011    4
atom         278   74    H1    "HxSCN H1"                    1     1.008    1
atom         279   75    C3    "HxSCN C3"                    6    12.011    4
atom         280   76    H2    "HxSCN H2"                    1     1.008    1
atom         281   77    H3    "HxSCN H3"                    1     1.008    1
atom         282   78    C4    "HxSCN C4"                    6    12.011    4
atom         283   79    C5    "HxSCN C5"                    6    12.011    4
atom         284   80    H4    "HxSCN H4"                    1     1.008    1
atom         285   81    H5    "HxSCN H5"                    1     1.008    1
atom         286   82    C6    "HxSCN C6"                    6    12.011    4
atom         287   83    H6    "HxSCN H6"                    1     1.008    1
atom         288   84    S     "HxSCN S"                    16    32.066    2
atom         289   85    C     "HxSCN C"                     6    12.011    2
atom         290   86    N     "HxSCN N"                     7    14.007    1

multipole   279  282  277             -0.12665
                                        0.16365   0.00000   0.14302
                                        0.06092
                                        0.00000  -0.43628
                                       -0.27661   0.00000   0.37536
multipole   276  277  278             -0.15938
                                        0.00000   0.00000   0.26734
                                       -0.20136
                                        0.00000  -0.20136
                                        0.00000   0.00000   0.40272
multipole   283  282  286             -0.11656
                                        0.24230   0.00000   0.09285
                                        0.15205
                                        0.00000  -0.43470
                                       -0.43635   0.00000   0.28265
multipole   282  279  283             -0.11327
                                        0.22022   0.00000   0.08479
                                        0.15602
                                        0.00000  -0.44910
                                       -0.36114   0.00000   0.29308
multipole   277  279  276             -0.12195
                                        0.19367   0.00000   0.15795
                                       -0.01138
                                        0.00000  -0.31282
                                       -0.22362   0.00000   0.32420
multipole   286  283  288             -0.18161
                                        0.04843   0.00000   0.18545
                                       -0.01034
                                        0.00000  -0.10804
                                       -0.18618   0.00000   0.11838
multipole   280  277  279              0.05998
                                        0.00441   0.00000  -0.07636
                                        0.04595
                                        0.00000   0.03338
                                       -0.00973   0.00000  -0.07933
```

```
multipole   284  282  279               0.06214
                                       0.02649   0.00000  -0.05387
                                       0.01977
                                       0.00000   0.03305
                                      -0.00324   0.00000  -0.05282
multipole   281  279  282               0.05805
                                       0.00916   0.00000  -0.05550
                                       0.03877
                                       0.00000   0.02937
                                      -0.06123   0.00000  -0.06814
multipole   285  283  282               0.07555
                                       0.01936   0.00000  -0.04776
                                       0.05942
                                       0.00000   0.03376
                                      -0.01589   0.00000  -0.09318
multipole   278  276  277               0.05847
                                       0.01125   0.00000  -0.10159
                                       0.06569
                                       0.00000   0.03435
                                      -0.03952   0.00000  -0.10004
multipole   287  286  283               0.07235
                                       0.02117   0.00000  -0.04358
                                       0.04575
                                       0.00000   0.03869
                                      -0.01006   0.00000  -0.08444
multipole   288  286  289               0.33074
                                       0.44389   0.00000   0.28736
                                       1.22369
                                       0.00000  -2.16613
                                      -0.31636   0.00000   0.94244
multipole   289  288  290               0.24556
                                       0.06457   0.00000  -0.41800
                                       0.15740
                                       0.00000   0.22633
                                       0.26601   0.00000  -0.38373
multipole   290  289  288              -0.58843
                                      -0.00512   0.00000  -0.24238
                                       0.32332
                                       0.00000   0.16836
                                      -0.08202   0.00000  -0.49168

polarize    276          1.3340     0.3900    277  278
polarize    277          1.3340     0.3900    276  280
polarize    278          0.4960     0.3900    276
polarize    279          1.3340     0.3900    281
polarize    280          0.4960     0.3900    277
polarize    281          0.4960     0.3900    279
polarize    282          1.3340     0.3900    284
      polarize    283          1.3340     0.3900    285
      polarize    284          0.4960     0.3900    282
      polarize    285          0.4960     0.3900    283
      polarize    286          1.3340     0.3900    287
      polarize    287          0.4960     0.3900    286
      polarize    288          3.3000     0.3900    289
      polarize    289          1.3340     0.3900    288  290
      polarize    290          1.0730     0.3900    289

```

```
vdw          75                3.8200     0.1010
vdw          72                3.8200     0.1010
vdw          79                3.8200     0.1010
vdw          78                3.8200     0.1010
vdw          73                3.8200     0.1010
vdw          82                3.7800     0.1060
vdw          84                4.0050     0.3550
vdw          85                3.7800     0.1060
vdw          76                2.9800     0.0240      0.940
vdw          80                2.9800     0.0240      0.940
vdw          77                2.9800     0.0240      0.940
vdw          81                2.9800     0.0240      0.940
vdw          74                2.9600     0.0240      0.920
vdw          83                2.8700     0.0330      0.900
vdw          86                3.7100     0.1050
bond         75   78         453.0000     1.5299
bond         75   73         453.0000     1.5298
bond         75   77         341.0000     1.0885
bond         72   73         323.0000     1.5282
bond         72   74         341.0000     1.0863
bond         79   78         453.0000     1.5312
bond         79   82         345.3000     1.5276
bond         79   81         341.0000     1.0866
bond         78   80         341.0000     1.0879
bond         73   76         341.0000     1.0878
bond         82   84         235.8000     1.8351
bond         82   83         341.0000     1.0810
bond         84   85         235.8000     1.7063
bond         85   86         450.0000     1.1375
angle        78   75   73     48.2000   113.1072
angle        78   75   77     38.0000   109.3223
angle        73   75   77     38.0000   109.2787
angle        77   75   77     39.5700   106.3036
angle        73   72   74     42.4400   111.1869
angle        74   72   74     39.5700   107.7267
angle        78   79   82     48.2000   111.5264
angle        78   79   81     38.0000   109.6162
angle        82   79   81     38.0000   109.5120
angle        81   79   81     39.5700   106.9493
angle        75   78   79     48.2000   112.7564
angle        75   78   80     38.0000   109.3923
angle        79   78   80     38.0000   109.3217
angle        80   78   80     39.5700   106.4572
angle        75   73   72     48.2000   112.9179
angle        75   73   76     38.0000   109.3170
angle        72   73   76     38.0000   109.3845
angle        76   73   76     45.5700   106.3035
angle        79   82   84     53.2000   109.0812
angle        79   82   83     42.4400   111.1744
angle        84   82   83     60.2400   108.2095
angle        83   82   83     39.5700   108.8985
angle        82   84   85     60.4300    99.5693
angle        84   85   86     60.0000   179.0730
strbnd       78   75   73     18.7000    18.7000
strbnd       78   75   77     11.5000    18.7000
strbnd       73   75   77     11.5000    18.7000
strbnd       73   72   74     11.5000    11.5000
```

```
166.        strbnd        78   79   82    18.7000    18.7000
167.        strbnd        78   79   81    11.5000    18.7000
168.        strbnd        82   79   81    11.5000    18.7000
169.        strbnd        75   78   79    18.7000    18.7000
170.        strbnd        75   78   80    11.5000    18.7000
171.        strbnd        79   78   80    11.5000    18.7000
172.        strbnd        75   73   72    18.7000    18.7000
173.        strbnd        75   73   76    11.5000    18.7000
174.        strbnd        72   73   76    11.5000    18.7000
175.        strbnd        79   82   84    18.7000    18.7000
176.        strbnd        79   82   83    11.5000    11.5000
177.        strbnd        84   82   83    11.5000    11.5000
178.        strbnd        82   84   85    -5.7500    -5.7500
179.        torsion       73   75   78   79      0.576 0.0 1  -
    0.017 180.0 2   2.031 0.0 3
180.        torsion       73   75   78   80      0.000 0.0 1   0.000 180.0 2   0.000
    0.0 3
181.        torsion       77   75   78   79      0.000 0.0 1   0.000 180.0 2   0.000
    0.0 3
182.        torsion       77   75   78   80      0.000 0.0 1   0.000 180.0 2   0.000
    0.0 3
183.        torsion       78   75   73   72      0.484 0.0 1   0.014 180.0 2   2.221
    0.0 3
184.        torsion       78   75   73   76      0.000 0.0 1   0.000 180.0 2   0.000
    0.0 3
185.        torsion       77   75   73   72      0.000 0.0 1   0.000 180.0 2   0.000
    0.0 3
186.        torsion       77   75   73   76      0.000 0.0 1   0.000 180.0 2   0.000
    0.0 3
187.        torsion       74   72   73   75     0.0000 0.0 1  0.0000 180.0 2  0.3410
    0.0 3
188.        torsion       74   72   73   76     0.0000 0.0 1  0.0000 180.0 2  0.2990
    0.0 3
189.        torsion       82   79   78   75      0.364 0.0 1  -
    0.024 180.0 2   1.958 0.0 3
190.        torsion       82   79   78   80      0.000 0.0 1   0.000 180.0 2   0.000
    0.0 3
191.        torsion       81   79   78   75      0.000 0.0 1   0.000 180.0 2   0.000
    0.0 3
192.        torsion       81   79   78   80      0.000 0.0 1   0.000 180.0 2   0.000
    0.0 3
193.        torsion       78   79   82   84     -
    0.688 0.0 1   0.489 180.0 2   1.957 0.0 3
194.        torsion       78   79   82   83      0.000 0.0 1   0.000 180.0 2   0.000
    0.0 3
195.        torsion       81   79   82   84      0.000 0.0 1   0.000 180.0 2   0.000
    0.0 3
196.        torsion       81   79   82   83      0.000 0.0 1   0.000 180.0 2   0.000
    0.0 3
197.        torsion       79   82   84   85     -
    2.643 0.0 1   0.709 180.0 2   1.072 0.0 3
198.        torsion       83   82   84   85      0.000 0.0 1   0.000 180.0 2   0.000
    0.0 3
199.        torsion       82   84   85   86     0.0000 0.0 1  0.0000 180.0 2  0.5000
    0.0 3
```

Code and Parameters 2-4: Hexyl Thiocyanate AMOEBA Parameters

2.5 SMALL MOLECULE SIMULATIONS IN AMOEBA

<Body text to begin here.>

Chapter 3 Electrostatic Field Methods

3.1 AMBER03 WITH EXPLICIT TIP3P WATER

3.1.1 Reaction Field Electrostatics

3.1.2 Hybrid Solvent Reaction Field Electrostatics and Solute Coulomb Field

<Body text to begin here.>

3.2 AMBER03 WITH POISSON-BOLTZMANN CONTINUUM SOLVENT

3.2.1 Reaction Field Method

<Body text to begin here.>

3.2.2 Grid spacing and size

<Body text to begin here.>

3.2.3 Box Location

<Body text to begin here.>

3.3 AMBER03 WITH POISSON-BOLTZMANN CONTINUUM SOLVENT AND SELECT EXPLICIT TIP3P WATER MOLECULES

3.3.1 5 Å Water Sphere Around the Vibrational Chromophore

<Body text to begin here.>

3.3.2 Single Water Molecule Nearest the Vibrational Chromophore

<Body text to begin here.>

3.3.3 Water Molecular Hydrogen Bonding to the Vibrational Chromophore

<Body text to begin here.>

3.4 AMOEBA

3.4.1 Poisson-Boltzmann Continuum Solvent

<Body text to begin here.>

3.4.2 Explicit AMOEBA Water

<Body text to begin here.>

3.4.3 Charge Penetration Field Corrections

<Body text to begin here.>

Chapter 4 The Role of Electrostatics in Differential Binding of RalGDS to Rap Mutations E30D and K31E Investigated by Vibrational Spectroscopy of Thiocyanate Probes

4.1 Introduction

The human proteins $p21^{Ras}$ (hereafter Ras) and Rap1A (hereafter Rap) are members of the Ras family of guanosine triphosphate (GTP)-hydrolyzing proteins that switch between an ON state when bound to GTP and an OFF state when bound to guanosine diphosphate (GDP) in the regulation of signal transduction pathways[25]. When in the GTP-bound ON state, both proteins dock to the Ras binding domain (RBD) of downstream effector proteins to propagate a signal transduction cascade. As members of the Ras superfamily of GTPases, Ras and Rap are similar in both sequence and structure; they share 50% amino acid identity[6], 80% amino acid homology, and have nearly identical structure and effector binding surfaces (rmsd of 0.7 Å for homologous residues).

Despite these similarities, Ras and Rap have very different functions within the cell. Ras is involved in the propagation of chemical signals that regulate cell division, cell survival and apoptosis, and has been a focus of research in molecular oncology for many years[26-28]. The biological function of Rap is less well understood, but it has recently been shown to play a role in cellular adhesion[29] and has been implicated in cancer metastasis[30, 31]. The origin of the functional specificity of these two similar proteins lies in the ability of Ras and Rap to bind to different downstream effectors while in the GTP-bound ON state, thus initiating different signaling cascades. One of the best studied downstream effector proteins of Ras is the human protein c-Raf-1 (hereafter Raf), while Rap is thought to interact most strongly with the Ras binding domain (RBD) of the downstream effector Ral guanosine dissociation stimulator (RalGDS, hereafter Ral). While the structures of the RBD of the downstream effector proteins are very similar[5, 6],

measurements of thermodynamic [2, 32-34] parameters of the protein-protein binding events ($D G_{dock}$ and K_d) indicate high specificity between the two GTPases and their appropriate downstream effector. Given the structural similarities of these protein-protein interfaces, a possible mechanism for the binding specificity that each GTPase shows for its appropriate downstream effector could be based on differences in electrostatic complementarity at the protein-protein interface. To investigate this hypothesis, the binding of these two GTPases to a variant of Ral have previously been investigated by our laboratory through both experimental and computational methods[1, 2] in order to explore the contribution of electrostatic fields to interface formation in a system in which structural factors are controlled as much as possible.

In 1995, Herrmann and coworkers[34, 35] demonstrated the importance of the amino acids at positions 30 and 31 on Ras and Rap in discriminating downstream effector partners. These amino acids, Ras Asp 30 and Glu 31 and Rap Glu 30 and Lys 31, are positioned at the protein-protein binding interface, and studies involving reversion mutations of Rap to Ras have shown that these residues are involved with the interactions with downstream binding partners. Those studies found that the charge reversion mutation Rap K31E and the double mutation Rap E30D/K31E cause the dissociation constant with the downstream effector Raf to resemble WT Ras more closely than WT Rap[35]. Furthermore, the double mutant Rap E30D/K31E co-crystallized with the downstream effecter Raf, which usually binds more strongly to Ras. Similarly, the mutation Ras E31K was used to co-crystallize Ras with Ral, normally Rap's immediate downstream effector[5], and the double reversion mutant Ras D30E/E31K has significantly reduced binding affinity with Raf[36]. The role of these mutations in creating strong binding affinity for an alternative downstream effector suggested an electrostatic mechanism caused by altering the charge of the residue at position 31 on either GTPase.

The arrangement of amino acid charges from secondary and tertiary protein structure can generate large and heterogeneous electrostatic fields that affect all protein function, including protein-protein interactions.[37-42] Vibrational Stark effect (VSE) spectroscopy is a recently developed experimental technique capable of measuring electrostatic fields in proteins[9, 43-49]. In VSE spectroscopy, the intrinsic response of a probe vibrational oscillator to its local electrostatic environment is measured spectroscopically and is used to quantitate the magnitude and direction of the local electric field to which the probe is exposed. After calibration, the probe is inserted into a known position of a protein where it becomes a highly local, sensitive, and directional reporter of fluctuations of the protein's electrostatic field caused by structural or chemical perturbations to the protein. For example, the formation of a protein-protein interface through a docking interaction may induce changes in the absorption energy of a probe located on each protein surface as that probe is exchanged from an electrostatic field generated by the protein-water interface to one generated by the protein-protein interface.

The nitrile stretching vibration is one of several useful VSE probes that have been indentified[46]. Several characteristics of the nitrile oscillator makes it particularly attractive; its absorption energy of ~2100-2250 cm^{-1} is in a region of the spectrum that is outside the vibrational background of a complex biomolecule, it is reasonably sensitive to fields of the magnitude thought to be present in proteins, and it can be incorporated into proteins with relative ease[50]. There are a growing number of reports describing VSE spectroscopy of nitrile probes to study the function of electrostatic fields in enzyme active sites[9, 45, 47-49]. Recently in our laboratory, VSE spectroscopy has been used to study the change in absorption energy of the nitrile probe incorporated at eleven locations on the surface of Ral that become buried in the protein-protein interface upon binding to either WT Ras or WT Rap[2]. This study determined that at some probe locations the

change in absorption upon binding to WT Ras and WT Rap was the same, while at other locations the change was different, suggesting that these latter amino acids may participate in an electrostatic mechanism that enables Ral to distinguish Ras from Rap. Furthermore, measurement of the dissociation constant of the docking of the GTPases with nitrile-containing Ral mutants showed no deleterious effect of the spectroscopic probe on the formation of the docked complex. These measurements were later confirmed through extensive molecular dynamics sampling of the protein-protein complex[1]. We hypothesized that several of these Ral-based nitrile probes would be ideal for investigating our hypothesis of an electrostatic mechanism for binding discrimination mediated by the charge on position 31 of these two GTPases.

In this report, we describe the systematic investigation of the effect of the E30D and K31E mutations to Rap on the local electrostatic fields formed by docking to the downstream effector and measured by the nitrile probe at six locations on the surface of the RBD of Ral using VSE, molecular dynamics (MD) simulations, and dissociation constant (K_d) measurements. We selected 6 amino acids on Ral that are positioned in the protein-protein interface near positions 30 and 31 of Ras and Rap when the docked complex is formed: N27, G28, N29, Y31, K32, and N54. The locations of these amino acids within the Rap-Ral interface, as well as Rap positions 30 and 31, are shown in Figure 1. Positions N27 and Y31 were selected because at these probe locations, we previously measured a significant difference between docking of WT Ras versus WT Rap[2], and wanted to investigate whether these observations were due to the different amino acid identity at positions 30 and 31. Position N29 was chosen because, along with position N27, molecular dynamics (MD) structural sampling of the thiocyanate side chain within the docked complex showed that the nitrile group has the largest angle with respect to the Ras-Ral interfacial plane of the collection of Ral-based probes we have

investigated, approximately 45° from the plane of the Ras-Ral interface[1]. Because the effect of the change in electrostatic field is on the projection of the field vector onto the nitrile probe bond axis, these positions seemed the most likely to feel the effect of the field vector at positions 30 and 31 of Rap, which appear from the crystal structures 1LFD[5] and 1GUA[6] to be approximately perpendicular to the interface. Gly28 was chosen as intermediate between these two residues. Finally, positions K32 and N54 were selected based on their physical proximity to the side chains of positions 30 and 31 on the GTPase once the docked complex formed.

Each of the six Ral amino acids were mutated to cysteine, then chemically modified to introduce the nitrile VSE probe in the form of the thiocyanate. These mutants were then bound to WT Rap, Rap E30D, Rap K31E, and the double mutant Rap E30D/K31E. The difference in vibrational absorption energy, Dn_{obs}, between the Ral monomer and the docked complex was measured by Fourier transform infrared spectroscopy (FTIR). The dissociation constant of the docked complex, K_d, was determined through a fluorescence assay. Finally, extensive MD simulations on docked complexes of all Rap and Ral variants were conducted to determine Boltzmann-weighted orientational data for the Ral-based nitrile probes and for the side chains at Rap positions 30 and 31. These simulations demonstrate that the mutation K31E is almost exclusively responsible for changes in side chain orientations at Rap positions 30 and 31 that cause the observed change in K_d. VSE spectroscopy demonstrates that only two of the probe locations examined, at N27C and N29C displayed a change in the absorption energy upon binding the Ras-like Rap double mutants that strongly resembled the change for WT Ras. However, several of these probes did respond in an additive manner to the individual single mutations. In some cases, these effects are replicated in subtle structural changes at positions 30 and 31 observed in molecular dynamics sampling. These studies support

both a structural and electrostatic mechanism to explain observed differences in GTPase-effector binding.

4.2 RESULTS

4.2.1 Dissociation Constant Measurements

Measured dissociation constants of both WT and SCN-labeled Ralβ mutants docking with all GTPases studied here are reported in Table 1. Binding of WT Ral to WT Rap was approximately 10-fold faster than binding to WT Ras, as has been reported before[2, 34, 35]. Any effect of the thiocyanate probe on the formation of the interface can be investigated by comparing K_d values obtained for WT Ral versus the SCN-labeled Ralβ mutants docking to WT Ras and WT Rap in Table 1. Although there are some variations in the absolute value of these numbers, they are for the most part small. The two largest deviations, caused by Ralβ G28C_{SCN} (reduced K_d by an order of magnitude) and Ralβ Y31C_{SCN} (increased K_d by an order of magnitude), still showed an order of magnitude increase in K_d when binding to WT Ras as opposed to WT Rap, as is expected from our previous work. The dissociation constants presented in Table 1 demonstrate that the presence of the SCN probe on Ralβ mutants did not substantially affect binding to either WT Ras or WT Rap, as has been observed before[2]. Along with results from molecular dynamics sampling of this system, described below, this is strong circumstantial evidence that the nitrile VSE probe does not significantly alter the interface formed between Rap and the SCN-labeled Ralβ mutants compared to the WT interaction. Our experimental mutagenesis and chemical labeling strategy therefore does not destroy the protein-protein interaction that we are attempting to measure.

The results in Table 1 confirm previous reports that the reversion mutation at Rap position 31 alters the binding interaction between Rap and Ral to resemble that of

Ras and Ral. All SCN-labeled Ralβ mutants interacted with Rap K31E with a K_d 10-fold higher than with WT Rap, and this effect was preserved in the double mutant Rap E30D/K31E. The single mutation Rap E30D had no effect on binding, and all K_d values measured with that construct were essentially identical to WT Rap. It is therefore clear that the amino acid located at position 31 of the GTPase is critical in the mechanism that enables Ral to distinguish structurally similar but functionally distinct GTPases for appropriate binding. Exploring the structural and electrostatic components of that mechanism is the subject of the MD sampling and VSE spectroscopy discussed here.

4.2.2 Molecular Dynamics Simulations

Structural details of the interface formed between Rap and Ral have been investigated with two high-quality crystal structures, 1GUA and 1LFD[5,6]. These are crucial starting points for any comprehensive biochemical understanding of the Rap-Ral interactions, but the static structures of these proteins might not represent the full range of side chain conformations of either our SCN probe or positions 30 and 31 of the GTPase, and so are not sufficient for our purposes. Furthermore, our experiments are on an unnatural variant of the system, the SCN-labeled Ralβ mutant. Although our K_d measurements indicate that the presence of the thiocyanate probe does not significantly disrupt interface formation, knowledge of the structure and orientation of the probe within the interface is necessary to interpret vibrational absorption data. To address these issues, we performed extensive molecular dynamics sampling of each SCN-labeled Ralβ mutant docked with WT Rap and the Rap mutants E30D, K31E, and E30D/K31E to accumulate a Boltzmann-weighted ensemble of the orientations of the thiocyanate group and the side chains at Rap positions 30 and 31. The WHAM-derived χ_2 torsional probability distribution for all thiocyanate groups and χ_1 torsional probability distribution

side chains at Rap positions 30 and 31 are shown in Figures S1-S3. Similar to what we have seen previously[1], the torsional distribution for all studied side chains was characteristic of an unhindered alkyl group, with three probability maxima separated by ~120˚ and essentially no significant difference depending on the chemical identity of either the SCN-labeled Ralβ mutant in the docked complex (Figure S1) or the Rap variant (Figures S2 and S3). Our MD sampling strategy therefore provides us with a comprehensive ensemble of structures of each docked Rap-Ralβ variant studied with VSE spectroscopy to aid in interpreting our spectroscopic results.

To analyze our molecular dynamics simulations of the torsional distribution of the thiocyanate residue on each Ralβ mutant and on the side chains at Rap positions 30 and 31, we defined two angles for each side chain with respect to the Rap-Ralβ surface, which we term azimuthal and polar angles. These two angles are described in the Supplemental Materials, and are shown schematically in Figure 2. When the cross hairs on Figure 2B are translated to the Cα atom of each simulated residue on Rap or Ralβ (represented as spheres), they become the origin of the polar angles shown in Figures 3 and S4. The average azimuthal and polar angles of the nitrile probe relative to the Rap-Ralβ binding interface are shown in Supplemental Figure S4. These simulations were compared with previous simulations of the Ras-Ralβ binding interface[1], and in all cases, the orientation of the nitrile at the docked interface for WT Ras and WT Rap are identical, within the error distribution of the Boltzmann-weighted ensemble. This indicates that the structure of the probe at the interface of these two GTPases is very similar, as expected by the structural similarities of the crystal structures, the amino acid homology of the GTPases, and the fact that the measured K_d values of the SCN-labeled Ralβ mutants dock with WT Ras and Rap with the same dissociation constant as WT Ral. Furthermore, as seen in Figure S4, mutations to positions 30 and 31 of Rap did not

substantially alter the orientation of the nitrile probe at the interface, with all 6 nitrile positions remaining approximately constant with the same dihedral distribution for WT Rap, or the E30D, K31E, E30D/K31E mutations. The consistency of the probe orientation in these Boltzmann-weighted ensembles is further indication of the structural robustness of these interacting proteins, making them suitable for systematic measurements of electrostatic fields at the protein-protein interface without compromising structural integrity of the interface. This observation is important to justify using VSE spectroscopy to explore the biochemistry of Rap-Ral binding with a spectroscopic technique that doesn't compromise the very interaction we are interesting in studying.

The azimuthal and polar angles of the side chains at Rap positions 30 and 31 for each mutant studied here are shown in Figure 3. When looking at position 30 (Figure 3A), the azimuthal angles showed essentially no difference depending on either the Rap mutant or the SCN-labeled Ralβ mutant to which it was docked. In most cases, the aspartate (for E30D and E30D/K31E) or glutamate vector (for WT and Rap K31E) pointed approximately parallel and slightly below the Rap-Ral surface plane. The single exception to this was when Rap E30D was docked with Ralβ K32C_{SCN}, where it was observed that the residue oriented itself 5˚ above the Rap-Ralβ K32C_{SCN} surface plane. The polar angles of the side chain at position 30 (Figure 3B) showed significantly more fluctuations, but these fluctuations were essentially random and all demonstrated very large WHAM-derived standard deviations. This suggests that both aspartate and glutamate side chains at this position undergo large polar motions that are not influenced by the chemical identity of position 31 or the location of the thiocyanate probe.

The side chains at Rap position 31 behaved very differently. The azimuthal angles of the side chain at Rap position 31 (Figure 3C) were pointed significantly further

below the Rap-Ralβ mutant plane than at position 30 (55-60˚ below the surface plane for position 31 versus < 25˚ below the surface plane for position 30), but with the exception of docked to Ralβ Y31C_{SCN}, all Rap mutants behaved identically. The distribution of polar angles of mutants of Rap position 31, however, displayed a significant dependence on the identity of the residue at position 31 (Figure 3D). Although again the distribution of observed angles for each Rap mutant was large, an obvious trend appeared in the data. With the sole exception Rap mutants docked to Ralβ N54C_{SCN}, WT Rap and the Rap E30D behaved identically, while the Rap mutants K31E and E30D/K31E also behaved identically. This strongly supports the evidence that mutations at Rap K31 are most important for determining binding to the downstream effector Ral, and that mutations to Rap E30 do not significantly affect the behavior of either WT Rap or Rap K31E. Both of these observations are consistent with our observations of differences in dissociation constants between WT and mutated Rap binding to the downstream effector Ral.

We investigated the causes of the difference in orientation between the lysine and glutamate side chains at Rap position 31 through inspection of representative MD snapshots. This clearly revealed that when a lysine is at position 31, it pointed towards a hydrogen bonding acceptor pocket formed by Ralβ D51, N54, and E58. When this side chain was mutated to Rap K31E in either the single or double mutant, the negatively charged Glu reoriented to avoid electrostatic repulsion with the Ralβ surface caused by the hydrogen bond accepting pocket created by these three residues. This appears to be the central cause for both the dissociation constant and electrostatic differences between Rap-Ralβ binding and Ras-Ralβ binding described here. As discussed above, this behavior was not observed when the Rap constructs were docked with Ralβ N54C_{SCN}; investigation of MD structures showed that this probe location disrupted the hydrogen bonding pocket sufficiently to cause this portion of the Ralβ surface retracted slightly

from the Rap-Ralβ interface, leaving K31 without the hydrogen bonding pocket. Although the importance of this hydrogen bonding pocket can be hypothesized from the 1LFD crystal structure (of Ras variant E31K docked with RalGDS), our MD sampling has provided the first confirmation of the observed biochemical behavior of Ras and Rap with direct structural evidence.

4.2.3 VSE Spectroscopy of the Docked Protein-Protein Complex

A principle goal of the work described here is to deconvolute the effects of structural and electrostatic influences on protein-protein interface formation. Our systematic molecular dynamics simulations allowed us to highlight differences between WT and mutant Rap that appeared to be due entirely to subtle changes in structure. With that information, a systematic exploration of electrostatic factors was conducted by measuring changes in electrostatic interactions between the GTPase and the downstream effector Ral by VSE spectroscopy of vibrational probes placed at the GTPase-effector interface. In this section we systematically review the spectroscopic results from each VSE probe, then draw general conclusions in the Discussion section.

Ralβ $N27C_{SCN}$ and $N29C_{SCN}$: Selection of each of the 6 SCN-labeled Ralβ probes was based on consideration of each probes' orientation compared to the Rap-Ralβ interface, proximity to the Rap positions 30 and 31, and large differences in vibrational absorption energy upon binding to WT Ras and WT Rap measured in a previous study[2]. $N27C_{SCN}$ and $N29C_{SCN}$ were selected for probe placement because our Boltzmann-weighted molecular dynamics simulations of the orientations of the 6 thiocyanate probes when docked with each of the Rap constructs determined that thiocyanate group on these two Ralβ mutants had some of the largest angles with respect to the plane of the Ras-Ralβ interface of any of our probes, approximately 20-30° above the surface plane

(Figure S4A)[1]. Because VSE spectroscopy is only sensitive to changes in the electrostatic field vector projected onto the nitrile bond axis, and because, as shown in Figure 3C, mutations to Rap K31 were themselves close to perpendicular to the Rap-Ralβ surface plane, probes perpendicular to the Rap-Ralβ plane would be most sensitive to mutations to Rap K31. A Ralβ-based SCN probe that was itself as close to perpendicular to the Rap-Ralβ surface plane would therefore be in an appropriate position to observe changes caused by the charge reversion mutation at position 31.

An example of the VSE data collected here is shown in Figure 4. The Ralβ $N29C_{SCN}$ mutant was incubated with each Rap mutant, concentrated, and the absorption energy of the nitrile probe was recorded and compared with the measured absorption energy when bound to WT Rap from a previous study[2]. When docked with Rap E30D, the thiocyanate absorption energy was 2160.8 cm^{-1}, identical to the observed absorption energy when Ralβ $N29C_{SCN}$ was bound to WT Rap. When Ralβ $N29C_{SCN}$ was incubated with Rap K31E, the absorption energy was 0.6 cm^{-1} higher in energy than when docked with WT Rap, 2161.4 cm^{-1}. However, when docked with the double mutant, Rap E30D/K31E, the absorption energy of the thiocyanate shifted 0.8 cm^{-1} higher in energy (2161.6 cm^{-1}). Both Rap mutants containing K31E were therefore more similar to the observed absorption energy when Ralβ $N29C_{SCN}$ is docked with WT Ras (2161.1 cm^{-1}) than with WT Rap. This means that the probe is experiencing an electrostatic environment in the double mutant E30D/K31E that is more like that of WT Ras than to WT Rap, although the effect is small.

Changes in the absorption energy ($\Delta\nu_{obs}$) of the thiocyanate on each SCN-labeled Ralβ mutant due to binding for each SCN-labeled Ralβ mutant studied here are summarized in Table 2 and Figure 5. In these figures, all absorption energies are referenced to the absorption energy of the nitrile probe when docked with WT Rap (i.e.

$\Delta\nu_{obs} = 0$ represents no change from the thiocyanate absorption energy when docked with WT Rap reported previously)[2]. These shifts are related to the change in electrostatic field through the known value of the Stark tuning rate of the cyanocysteine probe, [ ], of 0.7 cm^{-1}/(MV/cm); the values of [ ] determined from equation 1 are reported in Table 2.

When the probe is located at Ralβ $N27C_{SCN}$ and $N29C_{SCN}$, the data in Figure 5 clearly show that both the single and double reversion mutants Rap K31E and E30D/K31E have VSE absorption energies that are similar to WT Ras, not WT Rap. In both cases, the single mutation Rap E30D appears to have a negligible effect on the electrostatic environment of the probe. At these particular probe locations, therefore, the double reversion mutation does indeed revert the electrostatic field of Rap back to that found in Ras, supporting our hypothesis. Although these energy shifts are small, the trend towards higher absorption energies when bound to WT Ras and the double mutant Rap E30D/K31E than when bound to WT Rap are clear. The relatively small effect of mutations on the magnitude of the absorption energy of Ralβ $N29C_{SCN}$ in particular may be due to the fact that calculations have indicated that the nitrile at this location is actually most sensitive to changes in electrostatic field caused by amino acid mutations at the Ral surface, thus screening the effect of mutations made to the Ras surface[1]. This effect is currently being investigated in our laboratory and will be described in a future report.

$K32C_{SCN}$ and $N54C_{SCN}$: The closest Ralβ amino acid to positions 30 and 31 is N54, (a distance from backbone atoms of approximately 11-16 Å in our Boltzmann-weighted ensemble of structures). Because MD simulations indicated that this also was oriented above the Rap-Ralβ surface plane, this probe location was also selected for this study. As shown in Figure 5, the two single mutations E30D and K31E have small but opposite

effects on the absorption energy of N54C_{SCN} when compared to WT Rap; Rap E30D is 0.4 cm^{-1} higher in absorption energy, while Rap K31E is 0.1 cm^{-1} lower in absorption energy. The combined effect of the double mutant Rap E30D/K31E, however, gave a VSE shift 0.2 cm^{-1} higher in energy than WT Rap, resulting in an absorption energy that was approximately the sum of the behavior of the two single mutations. As discussed above, because N54C_{SCN} visibly disrupted the hydrogen-bonding interaction with Rap K31 in the docked complex, it is likely that the structural disruption of this interface means that this position is not appropriate for deconvoluting structural and electrostatic effects. This observation demonstrates the importance of investigation both structural and electrostatic contributions to the biochemical question of interest.

The nitrile vibrational probe was also placed at Ralβ K32C_{SCN} because of its proximity to positions 30 and 31 on the GTPase in the docked complex. As can be seen in Table 2 and Figure 5, although the measured error in $\Delta\nu_{obs}$ was larger than other positions, there was a dramatic effect of the mutation Rap K31E on the absorption energy of the thiocyanate compared to WT Rap (-1.0 cm^{-1}), while Rap E30D caused only a small perturbation (-0.2 cm^{-1}). The behavior of the double mutant Rap E30D/K31E was the sum of these two shifts, -1.2 cm^{-1}. This was very different from the response on binding to WT Ras, which showed a shift in absorption energy of +0.7 cm^{-1} compared to WT Rap. Thus, again, the differences in the electrostatic fields between WT Ras and WT Rap experienced by the probe at Ralβ K32C_{SCN} did not appear to be caused by the reversion mutations at positions 30 and 31 of Rap.

These two probe locations responded to the double mutant Rap E30D/K31E essentially as the addition of each single E30D and K31E mutation. When the probe was at N54C_{SCN}, the mutation E30D caused an increase in absorption energy relative to WT Rap, while K31E caused a decrease in absorption energy. The double mutant

E30D/K31E shifted approximately halfway between these two extremes. In the case of the probe at location Ralβ K32C_{SCN}, both single mutations caused a reduction in the absorption energy of the nitrile probe, while the double mutation caused a shift that was the sum of the two. In these two cases, it appears that we are measuring additive electrostatic effects as would be predicted by a simple model of adding electrostatic fields together. However, in neither case did the double reversion mutation to Rap produce an electrostatic environment measured by the probe that was similar to Ras.

G28C_{SCN} and Y31C_{SCN}: We chose Ralβ G28C_{SCN} as a probe location because it was between positions N27 and N29, even though it did not meet other desirable criteria. Previous studies had shown very little sensitivity to differences in binding to WT Ras versus WT Rap,[2] and a position in the docked complex that was more consistently parallel to the surface than either N27C_{SCN} or N29C_{SCN}[1]. Even still, this probe did respond strongly to mutants Rap E30D and K31E (+1.0 cm^{-1} and +0.5 cm^{-1} versus WT Rap, respectively), compared to a shift of -0.8 cm^{-1} when bound to WT Ras. The double mutant, Rap E30D/K31E, however, only demonstrated a shift of +0.2 cm^{-1} compared to WT Rap. The probe at this location thus reacted in a manner in which the two single mutations appear to cancel each other out.

Position Ralβ Y31C_{SCN} was selected for study because the change in electrostatic field upon docking to WT Rap differed from WT Ras by a large amount, 0.8 cm^{-1}, and resulted in changes in field in the opposite direction (sign of ☐). As shown in Table 2, both single mutants Rap E30D and Rap K31E shifted the absorption energy of the nitrile probe on Ralβ Y31C_{SCN} to lower energy, by -2.0 cm^{-1} and -1.0 cm^{-1} respectively. However, the combined effect of the double mutant was to shift the absorption energy lower by only -0.2 cm^{-1} compared to WT Rap, substantially less than either single mutant. As shown in our molecular dynamics simulations, there is a significant reorientation of

the side chain at position Rap 31 above (azimuthal angle) and away from (polar angle) Ralβ $Y31C_{SCN}$ in the mutants Rap K31 and E30D/K31E which are likely causing this probe to become significantly less sensitive to change in electrostatic field in the binding region in which we are interested.

At both of these locations on the protein surface, Ralβ $G28C_{SCN}$ and Ralβ $Y31C_{SCN}$, although each single Rap mutant caused a large change in absorption energy (higher in energy at Ralβ $G28C_{SCN}$, and lower in energy at Ralβ $Y31C_{SCN}$), the double mutant effectively canceled out those changes in both cases. While MD sampling of position Ralβ $Y31C_{SCN}$ makes it clear that this is likely due to a large structural reorientation of the Rap sidechain position 31, the influences of the probe at position Ralβ $G28C_{SCN}$ are much less clear. After extensive inspection of our MD simulations, we have found no significant structural differences near $G28C_{SCN}$ to explain this observation. It could be that the distance between G28 and the region of the binding surface we are investigating by mutations at Rap E30 and K31 make Ralβ $G28C_{SCN}$ an ineffective probe for this study. This result is noteworthy, however, because the distance over which linear VSE effects can be accurately measured has still not been experimentally established.

4.3 DISCUSSION

This study was motivated by functional observations that positions 30 and 31 of Ras and Rap, which are among the few chemical differences between the Ras and Rap interfaces, could help discriminate appropriate downstream effectors for each GTPase. Because of the structural similarities of these two protein surfaces, it is possible that downstream effector binding selectivity could be caused by changes in the electrostatic fields at the GTPase-effector interface caused by mutations at position 30 and 31, and in particular by the reversal of charge caused by the K31E mutation. Measuring this effect

from the perspective of the downstream effector Ral, which can bind to both GTPases, proved useful to explore this aspect of GTPase function. Previous experimental work and extensive MD sampling provided us with criteria to guide the selection of useful places to position the nitrile probe on the Ralβ surface. Three criteria that were used were 1) angle of the nitrile with respect to the GTPase-effector interface, with probes perpendicular to the plane of the interface preferred ($N27C_{SCN}$ and $N29C_{SCN}$); 2) proximity to positions 30 and 31 when the docked complex is formed ($K32C_{SCN}$ and $N54C_{SCN}$), and 3) previously measured discrimination in absorption energy when docking to Ras as opposed to Rap, indicating those probes end up in significantly different electrostatic environments after the docked complex is formed ($N27C_{SCN}$, $N29C_{SCN}$, and $Y31C_{SCN}$). The location $G28C_{SCN}$ was chosen because of its position between $N27C_{SCN}$ and $N29C_{SCN}$, not because of any favorable selection criteria, and could be considered as a control location on the Ralβ surface. No probe displayed ideal behavior in all three selection criteria, and some probes were favorable from one aspect but unfavorable from another. For example, $N27C_{SCN}$ and $N29C_{SCN}$ were the most perpendicular of all of the Ralβ mutant, but were far removed from Rap positions 30 and 31 in the docked complex. Thus even with detailed structural and electrostatic studies we have carried out on these GTPases, working within the limitations of the biological system does not allow us to design a perfect experiment. All probes will experience a convolution of favorable and unfavorable interactions, and all probes will be unique. Electrostatic fields must therefore be explored from the perspectives of multiple probes, and information taken from the aggregation of the data.

We observed three general trends from the six probe locations studied here. The first was seen with $N27C_{SCN}$ and $N29C_{SCN}$, which displayed the Rap-to-Ras reversion behavior caused by the double mutation Rap E30D/K31E. The absorption energy of this double mutant essentially matched that of WT Ras, and in both cases was caused almost

exclusively by the mutation K31E, not E30D. This is direct confirmation of hypotheses proposed from previous crystallographic and docking studies that the difference between Ras and Rap at these two positions leads to the functional discrimination these two GTPases have for different downstream effectors. The second observed trend was that of the double mutant displaying a difference in absorption energy that was simply the sum of the effect caused by the two single mutations. This was seen at positions $K32C_{SCN}$ and $N54C_{SCN}$. This is strong experimental evidence that these mutations cause very little disruption in the docked complex, and experience additive changes in electrostatic field that can be measured by VSE spectroscopy of appropriately placed probes. However, it is important to study these effects from as many different probe locations as possible, because the third trend was less easily interpreted. For two probe locations, $G28C_{SCN}$ and $Y31C_{SCN}$, while each single Rap mutation caused a large shift in vibrational absorption energy, the double mutant Rap E30D/K31E behaved essentially identically to WT Rap. We have found no structural cause for this behavior. These residues may simply be unimportant for the formation of the Rap-Ral interface, and changes in electrostatic field measured by probes at these locations may not be correlated with changes in Rap-Ral interface formation.

In summary, this study demonstrates that while VSE spectroscopy is a useful tool for examining molecular-level mechanism of electrostatic events in complex biological systems, the convolution of distance, orientation, and change in determining the change in local electrostatic field actually experienced by the nitrile probe needs to be carefully interpreted. Observation of reversion behavior in the double mutant Rap E30D/K31E with the probes Ralβ $N27C_{SCN}$ and $N29C_{SCN}$ strongly supports the role these two positions play in supporting an electrostatic mechanism of functional discrimination in GTPase-effector binding. Although these probe locations displayed the reversion mutation

behavior that we were predicting when designing these experiments, the convolution of structure and sensitivity to mutations on the surface of Rap demonstrate that unambiguous measurement of electrostatic effects at the Rap-Ral interface will be difficult to achieve. Molecular dynamics sampling appears to be particularly useful in selection of appropriately placed VSE probes. Further experiments on WT and mutant Rap will be guided by extensive MD and continuum electrostatic calculations.

Table 4-1: Dissociation Constant of the formation of docked complexes of WT and SCN labeled Ralβ mutants with GTPases

	Ralβ Mutation K_d (μM)						
GTPase	WT	$N27C_{SCN}$	$G28C_{SCN}$	$N29C_{SCN}$	$Y31C_{SCN}$	$K32C_{SCN}$	$N54C_{SCN}$
WT Rap	0.26 ± 0.09	0.30 ± 0.05	0.04 ± 0.004	0.15 ± 0.08	1.0 ± 0.1	0.69 ± 0.18	0.91 ± 0.20
WT Ras	1.4 ± 0.2	7.3 ± 2.2	4.8 ± 1.1	3.1 ± 0.6	12.9 ± 4.8	6.0 ± 2.3	4.7 ± 0.4
Rap E30D	0.24 ± 0.06	0.12 ± 0.01	0.03 ± 0.015	0.10 ± 0.02	0.89 ± 0.35	0.34 ± 0.09	0.79 ± 0.10
Rap K31E	1.8 ± 0.7	2.9 ± 0.7	8.1 ± 1.5	6.0 ± 1.1	6.2 ± 2.0	1.9 ± 0.8	2.2 ± 1.5
Rap E30D/K31E	1.0 ± 0.1	5.4 ± 1.0	2.5 ± 0.3	5.0 ± 1.7	2.0 ± 0.2	1.8 ± 0.5	1.1 ± 0.3

Table 4-2: Measured vibrational frequencies of SCN-labeled Ralβ mutants docked with Rap E30/K31; the observed changes in vibrational frequencies upon docking each probe to Ras D30/E31 and Rap mutants

		Ralβ Mutant					
GTPase		N27C	G28C	N29C	Y31C	K32C	N54C
WT Rap	$\tilde{\nu}$ (cm^{-1})	2162.6 (0.4)	2161.8 (0.1)	2160.8 (0.2)	2161.5 (0.2)	2160.9 (0.2)	2161.4 (0.2)
		Difference Compared to WT Rap					
WT Ras	$\Delta\tilde{\nu}$ (cm^{-1})	-0.5 (1.0)	-0.8 (0.1)	0.3 (0.4)	-0.8 (0.4)	0.7 (0.8)	-0.5 (0.4)
	[ ] (MV/cm)	0.7 (1.4)	1.1 (0.1)	-0.4 (0.6)	1.1 (0.6)	-1.0 (1.1)	0.7 (0.6)
Rap E30D	[ ] (cm^{-1})	-0.1 (0.4)	1.0 (0.5)	0.0 (0.2)	-2.0 (0.4)	-0.2 (0.6)	0.4 (0.3)
	[ ] (MV/cm)	0.1 (0.6)	-1.4 (0.7)	0.0 (0.3)	2.9 (0.6)	0.3 (0.8)	-0.6 (0.4)
Rap K31E	[ ] (cm^{-1})	-0.6 (0.6)	0.5 (0.2)	0.6 (0.2)	-1.0 (0.4)	-1.0 (0.2)	-0.1 (0.3)
	[ ] (MV/cm)	0.9 (0.9)	-0.7 (0.3)	-0.9 (0.3)	1.4 (0.6)	1.4 (0.3)	0.1 (0.4)
Rap E30D/K31E	[ ] (cm^{-1})	-0.3 (0.5)	0.2 (0.3)	0.8 (0.2)	-0.2 (0.4)	-1.2 (0.4)	0.2 (0.3)
	[ ] (MV/cm)	0.4 (0.7)	-0.3 (0.4)	-1.1 (0.3)	0.3 (0.6)	1.7 (0.6)	-0.3 (0.4)

Chapter 5 Optimizing Electrostatic Field Calculations with the Adaptive Poisson-Boltzmann Solver to Predict Electric Fields at Protein-Protein Interfaces I: Sampling and Focusing

5.1 Introduction

<Body text to begin here.>

5.2 Results

<Body text to begin here.>

5.3 Discussion

<Body text to begin here.>

Chapter 6 Optimizing Electrostatic Field Calculations with the Adaptive Poisson-Boltzmann Solver to Predict Electric Fields at Protein-Protein Interfaces II: Explicit Near-Probe and Hydrogen Bonding Water Molecules

6.1 Introduction

<Body text to begin here.>

6.2 Results

<Body text to begin here.>

6.3 Discussion

<Body text to begin here.>

Chapter 7 Electrostatic Fields at Protein-Protein Interfaces: Increased Sampling Time and Various Electrostatic Methods: A Case for Simulating in Polarizable Force Fields

7.1 Introduction

<Body text to begin here.>

7.2 Results

<Body text to begin here.>

7.3 Discussion

<Body text to begin here.>

Chapter 8 Electrostatic Fields in Small Thiocyanate Molecules with Ensembles Generated using the AMOEBA Force Field

8.1 Introduction

<Body text to begin here.>

8.2 Results

<Body text to begin here.>

8.3 Discussion

<Body text to begin here.>

Appendix

Glossary

This page is optional—must be placed in this order if it is included in the dissertation. If you don't want to include a glossary, then delete the entire page and the following page break.

References

1. Ensign, D. L.; Webb, L. J., Statistical, Geometric, and Physical Factors Determining Electrostatic Fields at the Ras/Effector Interface. **2010,** *in preparation*.
2. Stafford, A. J.; Ensign, D. L.; Webb, L. J., Vibrational Stark Effect Spectroscopy at the Interface of Ras and Rap1A Bound to the Ras Binding Domain of RalGDS Reveals an Electrostatic Mechanism for Protein-Protein Interaction. *J. Phys. Chem. B* **2010,** *114*, 15331-15344.
3. Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., Gromacs: Fast, Flexible, and Free. *J Comput Chem* **2005,** *26* (16), 1701-1718.
4. Geyer, M.; Herrmann, C.; Wohlgemuth, S.; Wittinghofer, A.; Kalbitzer, H. R., Structure of the Ras-binding domain of RalGEF and implications for Ras binding and signalling. *Nat. Struct. Biol.* **1997,** *4*, 694-699.
5. Huang, L.; Hofer, F.; Martin, G. S.; Kim, S. H., Structural basis for the interaction of Ras with RalGDS. *Nat. Struct. Biol.* **1998,** *5*, 422-426.
6. Nassar, N.; Horn, G.; Herrmann, C.; Scherer, A.; McCormick, F.; Wittinghofer, A., The 2.2 A crystal structure of the Ras-binding domain of the serine/threonine kinase c-Raf1 in complex with Rap1A and a GTP analogue. *Nature* **1995,** *375* (6532), 554-560.
7. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M., et al., *AMBER 11*. University of California, San Francisco: 2010.
8. Qamra, R.; Hubbard, S., - Structural basis for the interaction of the adaptor protein grb14 with activated. - *PLoS One. 2013 Aug 13;8(8):e72473. doi: 10.1371/journal.pone.0072473. eCollection* (- 1932-6203 (Electronic)).
9. Ensign, D. L.; Webb, L. J., Factors Determining Electrostatic Fields at the Ras/Effector Interface. *Proteins* **2011,** *79*, 3511-3524.
10. Ragain, C. M.; Newberry, R. W.; Ritchie, A. W.; Webb, L. J., Role of Electrostatics in Differential Binding of RaIGDS to Rap Mutations E30D and K31E Investigated by Vibrational Spectroscopy of Thiocyanate Probes. *J Phys Chem B* **2012,** *116* (31), 9326-9336.
11. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **1983,** *79* (2), 926-935.
12. Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A., PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* **2007,** *35*, W522-W525.
13. Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A., PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **2004,** *32*, W665-W667.

14. McCleverty, C. J.; Columbus, L.; Kreusch, A.; Lesley, S. A., Structure and ligand binding of the soluble domain of a Thermotoga maritima membrane protein of unknown function TM1634. *Protein Sci* **2008,** *17*, 869.
15. Humphrey, W.; Dalke, A.; Schulten, K., VMD - Visual Molecular Dynamics. *J. Molec. Graphics* **1996,** *14*, 33-38.
16. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C., GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005,** *26*, 1701-1718.
17. Smith, T. F.; Waterman, M. S., Identification of Common Molecular Subsequences. *J Mol Biol* **1981,** *147* (1), 195-197.
18. Russell, R. B.; Barton, G. J., Multiple protein sequence alignment from tertiary structure comparison. *PROTEINS: Struct. Funct. Genet.* **1992,** *14*, 309-323.
19. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T., et al., A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003,** *24*, 1999-2012.
20. Darden, T.; York, D.; Pedersen, L. G., Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.
21. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103*, 8577-8593.
22. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., - LINCS: A linear constraint solver for molecular simulations. **1997,** - *18* (- 12), - 1472.
23. Roux, B., The calculation of the potential of mean force using computer simulations. *Comp. Phys. Commun.* **1995,** *91*, 275-282.
24. Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M., Temperature Weighted Histogram Analysis Method, Replica Exchange, and Transition Paths. *J. Phys. Chem. B* **2005,** *109*, 6722-6731.
25. Krauss, G., *Biochemistry of Signal Transduction and Regulation*. 3 ed.; WILEY-VCH Verlag: Weinheim, Germany, 2003.
26. Cox, A. D.; Der, C. J., The dark side of Ras: regulation of apoptosis. *Oncogene* **2003,** *22*, 8999-9006.
27. Downward, J., Targeting Ras Signalling Pathways in Cancer Therapy. *Nature Canc. Rev.* **2002,** *3*, 11-22.
28. Repasky, G. A.; Chenette, E. J.; Der, C. J., Renewing the conspiracy theory debate: does Raf function alone to mediate Ras oncogenesis? *Trends Cell Biol.* **2004,** *14*, 639-647.
29. Thomas, C.; Fricke, I.; Scrima, A.; Berken, A.; Wittinghofer, A., Structural Evidence for a Common Intermediate in Small G Protein-GEF Reactions. *Molecular Cell* **2007,** *25* (1), 141-149.
30. Wu, C.; Lai, C.-F.; Mobley, W. C., Nerve Growth Factor Activates Persistent Rap1 Signaling in Endosomes. *The Journal of Neuroscience* **2001,** *21* (15), 5406-5416.
31. Alexov, E.; Mehler, E. L.; Baker, N.; Baptista, A. M.; Huang, Y.; Milletti, F.; Nielsen, J. E.; Farrell, D.; Carstensen, T.; Olsson, M. H. M., et al., Progress in the

prediction of pK(a) values in proteins. *Proteins-Structure Function and Bioinformatics* **2011,** *79* (12), 3260-3275.
32. Herrmann, C., Ras-effector interactions: after one decade. *Curr. Opinion Struct. Biol.* **2003,** *13*, 122-129.
33. Rudolph, M. G.; Linnemann, T.; Grunewald, P.; Wittinghofer, A.; Vetter, I. R.; Herrmann, C., Thermodynamics of Ras/Effector and Cdc42/Effector Interactions Probed by Isothermal Titration Calorimetry. *J. Biol. Chem.* **2001,** *276*, 23914-23921.
34. Herrmann, C.; Horn, G.; Spaargaren, M.; Wittinghofer, A., Differential Interactions of the Ras Family of GTP-binding Proteins H-Ras, Rap1A, and R-Ras with the Putative Effector Meolecules Raf Kinase and Ral-Guanine Nucleotide Exchange Factor. *J. Biol. Chem.* **1996,** *271*, 6794-6800.
35. Nassar, N.; Horn, G.; Herrmann, C.; Block, C.; Janknecht, R.; Wittinghofer, A., Ras/Rap effector specificity determined by charge reversal. *Nat. Struct. Biol.* **1996,** *3*, 723-729.
36. Davis-Searles, P. R.; Saunders, A. J.; Erie, D. A.; Winzor, D. J.; Pielak, G. J., INTERPRETING THE EFFECTS OF SMALL UNCHARGED SOLUTES ON PROTEIN-FOLDING EQUILIBRIA. *Annual Review of Biophysics and Biomolecular Structure* **2001,** *30* (1), 271-306.
37. Honig, B.; Nicholls, A., Classical Electrostatics in Biology and Chemistry. *Science* **1995,** *268*, 1144-1149.
38. Gunner, M. R.; Nicholls, A.; Honig, B., Electrostatic Potentials in Rhodopseudomonas viridis Reaction Centers: Implications for the Driving Force and Directionality of Electron Transfer. *J. Phys. Chem.* **1996,** *100*, 4277-4291.
39. Lee, L. P.; Tidor, B., Optimization of Binding Electrostatics: Charge Complementarity in the Barnase-Barstar Protein Complex. *Protein Science* **2001,** *10*, 362-377.
40. Simonson, T., Macromolecular electrostatics: continuum models and their growing pains. *Curr. Opin. Struct. Biol.* **2001,** *11*, 243-252.
41. Villa, J.; Warshel, A., Energetics and Dynamics of Enzymatic Reactions. *J. Phys. Chem. B* **2001,** *105*, 7887-7907.
42. Warshel, A.; Papazyan, A., Electrostatic effects in macromolecules: fundamental concepts and practical modeling. *Curr. Opin. Struct. Biol.* **1998,** *8*, 211-217.
43. Andrews, S. S.; Boxer, S. G., Vibrational Stark Effects of Nitriles I. Methods and Experimental Results. *J. Phys. Chem. A* **2000,** *104*, 11853-11863.
44. Andrews, S. S.; Boxer, S. G., Vibrational Stark Effects of Nitriles II. Physical Origins of Stark Effects from Experiment and Perturbation Models. *J. Phys. Chem. A* **2002,** *106*, 469-477.
45. Fafarman, A. T.; Webb, L. J.; Chuang, J. I.; Boxer, S. G., Site-Specific Conversion of Cysteine Thiols into Thiocyanate Creates an IR Probe for Electric Fields in Proteins. *J. Am. Chem. Soc.* **2006,** *128*, 13356-13357.
46. Suydam, I. T.; Boxer, S. G., Vibrational Stark Effects Calibrate the Sensitivity of Vibrational Probes for Electric Fields in Proteins. *Biochemistry* **2003,** *42*, 12050-12055.

47. Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G., Electric Fields at the Active Site of an Enzyme: Direct Comparison of Experiment with Theory. *Science* **2006,** *313*, 200-204.
48. Webb, L. J.; Boxer, S. G., Electrostatic Fields Near the Active Site of Human Aldose Reductase: 1. New Inhibitors and Vibrational Stark Effect Measurements. *Biochemistry* **2008,** *47*, 1588-1598.
49. Mardia, K. V., Statistics of Directional Data. *J. Roy. Statist. Soc. Ser. B (Methodol.)* **1975,** *37* (3), 349-393.
50. Nielsen, J. E.; Gunner, M. R.; Garcia-Moreno, E. B., The pK(a) Cooperative: A collaborative effort to advance structure-based calculations of pK(a) values and electrostatic effects in proteins. *Proteins-Structure Function and Bioinformatics* **2011,** *79* (12), 3249-3259.