

Copyright

by

Andrew William Ritchie

2015

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

**QUANTIFYING ELECTROSTATIC FIELDS AT PROTEIN  
INTERFACES USING CLASSICAL ELECTROSTATICS  
CALCULATIONS**

**Committee:**

---

Lauren J. Webb, Supervisor

---

Ron Elber

---

Walter Fast

---

Graeme Henkelman

---

Pengyu Ren

**QUANTIFYING ELECTROSTATIC FIELDS AT PROTEIN  
INTERFACES USING CLASSICAL ELECTROSTATICS  
CALCULATIONS**

**by**

**Andrew William Ritchie, B.S.**

**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

**Doctor of Philosophy**

**The University of Texas at Austin**  
**August 2015**

# **QUANTIFYING ELECTROSTATIC FIELDS AT PROTEIN INTERFACES USING CLASSICAL ELECTROSTATICS CALCULATIONS**

Andrew William Ritchie, PhD.

The University of Texas at Austin, 2015

Supervisor: Lauren J. Webb

The functional aspects of proteins are largely dictated by highly selective protein-protein and protein-ligand interactions, even in situations of high structural homology, leaving electrostatic factors as the remaining contributor to selective binding. The vibrational Stark effect (VSE) allows us to measure electrostatic fields in complex environments, such as proteins, by the introduction of a vibrational chromophore whose vibrational absorption energy is linearly sensitive to changes in the local electrostatic field. The works presented here seek to computationally quantify electrostatic fields measured via VSE, with the eventual goal of being able to quantitatively predict electrostatic fields, and therefore Stark shifts, for any given protein-interaction. This is done using extensive molecular dynamics in the Amber03 and AMOEBA force fields to generate large ensembles the GTPase Rap1a docked to RalGDS and <sup>p21</sup>Ras docked to RalGDS. We discuss how side chain orientations help to understand the differential binding of different mutations of Rap1a binding to RalGDS, where it was found that a hydrogen-bonding pocket is disrupted by the mutation of position 31 from lysine to glutamic acid. We then show that multi-dimensional umbrella sampling of the probe orientations yields a wider range of accessible structures, increasing the quality of the

ensembles generated. A large variety of methods for calculating electrostatic fields are presented, with Poisson-Boltzmann electrostatics yielding the most consistent, reliable results. Finally, we explore using AMOEBA for both ensemble-generation as well as the electrostatic description of atoms for field calculations, where early results suggest that the electrostatic field due to the induce dipole moment of the probe is responsible for predicting qualitatively correct Stark shifts.

## Table of Contents

List of Tables .....	x
List of Figures .....	xi
List of Code and Force Field Parameters.....	xv
Chapter 1 Introduction .....	1
Chapter 2 Simulation Methods .....	6
2.1 Introduction .....	6
2.2 Nomenclature .....	6
2.3 Labeling and Mutating Proteins <i>in silico</i> .....	7
2.4 Enhanced Molecular Dynamics in Amber03: N-Dimensional Umbrella Sampling and Weighted Histogram Analysis Method.....	11
2.5 Electrostatic Clustering in Vibrational Chromophore Dihedral Space ...	22
2.6 Side chain orientation: Azimuthal and Polar Angles .....	28
2.7 Probe Parameterization for AMOEBA .....	33
2.7.1 Cyanocysteine parameterization .....	33
2.7.2 GDPNP Parameterization .....	34
2.7.3 Methyl-, Ethyl-, and Hexylthiocyanate Parameterization.....	34
2.8 Small Molecule Simulations in AMOEBA .....	59
Chapter 3 Electrostatic Methods and Theory.....	61
3.1 Introduction.....	61
3.2 Amber03 with Explicit TIP3P Water.....	61
3.2.1 GROMACS Reaction Field Electrostatics .....	61
3.2.2 Hybrid Solvent Reaction Field Electrostatics and Solute Coulomb Field .....	62
3.3 Amber03 with Poisson-Boltzmann Continuum Solvent.....	63
3.3.1 Reaction Field Method.....	64
3.3.2 Grid spacing and size .....	66
3.3.3 Box Location.....	67

3.4 Amber03 with Poisson-Boltzmann Continuum Solvent and Select Explicit TIP3P Water Molecules .....	71
3.4.1 5 Å Water Sphere Around the Vibrational Chromophore .....	71
3.4.2 Single Water Molecule Nearest the Vibrational Chromophore ..	72
3.4.3 Water Molecular Hydrogen Bonding to the Vibrational Chromophore .....	72
3.5 AMOEBA .....	74
3.5.1 Calculating Electrostatic Fields in AMOEBA .....	74
3.5.2 Poisson-Boltzmann Continuum Solvent .....	75
3.5.3 Explicit AMOEBA Water .....	76
3.5.4 Charge Penetration Field Corrections .....	77
3.6 Vibrational Stark Effect .....	80
3.7 Virtual Stark Tuning Rate ( <i>VSTR</i> ) and “Ideal” Dielectrics .....	81
 Chapter 4 The Role of Electrostatics in Differential Binding of RalGDS to Rap Mutations E30D and K31E Investigated by Vibrational Spectroscopy of Thiocyanate Probes .....	82
4.1 Introduction .....	82
4.2 Results .....	87
4.2.1 Dissociation Constant Measurements .....	87
4.2.2 Molecular Dynamics Simulations .....	88
4.2.3 VSE Spectroscopy of the Docked Protein-Protein Complex.....	92
4.3 Discussion .....	98
4.4 Conclusion .....	100
 Chapter 5 Optimizing Electrostatic Field Calculations with the Adaptive Poisson-Boltzmann Solver to Predict Electric Fields at Protein-Protein Interfaces I: Sampling and Focusing .....	117
5.1 Introduction .....	117
5.2 Results and Discussion .....	125
5.2.1 Molecular Dynamics Sampling.....	125
5.2.2 Calculation of absolute electrostatic fields .....	127
Numeric solutions to the LPBE .....	129

Reaction field method solutions to the LPBE.....	132
Comparing the numeric solutions of the LPBE to the reaction field method solutions of the LPBE .....	133
5.2.3 Calculations of relative electrostatic field.....	133
Numeric relative field solutions to the LPBE .....	134
Reaction field method relative fields .....	136
Comparing relative fields calculated from the numeric method to the reaction field method .....	137
5.3 Conclusion .....	138
<b>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 .....</b>	<b>176</b>
6.1 Introduction .....	176
6.2 Results and Discussion .....	178
6.2.1 Calculations of Absolute Fields .....	178
5 Å Sphere.....	178
Nearest Water.....	182
Hydrogen-bonding Water .....	183
6.2.2 Calculations of Field Differences .....	184
5 Å Sphere.....	185
Nearest Water and Hydrogen-bonding Water.....	186
6.3 Conclusion .....	187
<b>Chapter 7 Electrostatic Fields at Protein-Protein Interfaces: Increased Sampling Time and Various Electrostatic Methods: A Case for Simulating in Polarizable Force Fields .....</b>	<b>223</b>
7.1 Introduction.....	223
7.2 Results and Discussion .....	225
7.2.1 Electrostatic Fields Along the Interface of Each GTPase.....	226
7.2.2 Electrostatic Fields at a Single Point on Ral .....	228
7.2.3 Changes in Field upon Docking.....	229
7.2.4 Overall Electrostatic Field Predictions Along Protein Surfaces	230

7.3 PB Implicit Water and Explicit TIP3P Water are Electrostatically Equivalent .....	230
7.4 General Remarks.....	232
7.5 Conclusion .....	234
Chapter 8 Comparing Experimental Absorption Energies to Electrostatic Fields	
Sampled in AMOEBA for Small Thiocyanate-labeled Molecules .....	241
8.1 Introduction.....	241
8.2 Results and Discussion .....	243
8.2.1 Sampling CN Orientations .....	244
8.2.2 Electrostatic Fields.....	245
8.2.3 Correlating Small Molecule Fields to Experiment .....	246
8.3 Conclusion .....	248
References .....	259

## List of Tables

Table 2-1: Experimental Protein Systems Investigated and their Naming Conventions .....	10
Table 2-2: Sample Parameters for Ryckaert Bellemans dihedral potential function used for validating 2D WHAM code .....	17
Table 2-3: Umbrella Sampling Parameters, grouped by chapter .....	18
Table 2-4: Comparisons between AMOEBA and QM Molecular Properties .....	36
Table 3-1: Charge Penetration Parameters .....	79
Table 4-2: Measured vibrational frequencies of SCN-labeled Ral $\beta$ mutants docked with Rap E30/K31; the observed changes in vibrational frequency upon docking each probe to Ras D30/E31 and each Rap mutant .....	116
Table 5-1: Percent Volume of Second-Stage APBS Box Filled with Protein Atoms .....	141
Table 6-1: Average number of Frames Water Molecules Remained within 5 Å of the Thiocyanate N $\epsilon$ Atom .....	189
Table 6-2: Virtual Stark Tuning Rate ( $VSTR$ ), "ideal" Dielectric Constants ( $\epsilon'$ ), Correlation Coefficients (r), and p-values for APBS Calculations of Electrostatic Fields .....	190
Table 6-3: Percentage of Total Frames with a Nitrile-Water Hydrogen Bond....	192
Table 7-2: Correlation Coefficients (R) and Virtual Stark Tuning Rates ( $VSTR^a$ ) for Relative Field Calculations using Various Electrostatic Models....	236
Table 7-3: Correlation Coefficients for Field Standard Deviations Compared to Experimental Full Width at Half Peak Maximum (FWHM) using Various Electrostatic Methods .....	237

## List of Figures

Figure 2-1: Cyanocysteine Side chain Dihedral Definitions .....	19
Figure 2-2: 2D WHAM Validation .....	21
Figure 2-3: PB Solvent Reaction Field vs. Solute Analytic Coulomb Field .....	25
Figure 2-4: Field Values using Clustering Vs. Field Values using All Frames.....	26
Figure 2-5: Correlations and Slopes at Various Cutoff Values .....	27
Figure 2-6: Representation of azimuthal and polar angles .....	31
Figure 3-1: Reaction Field Method Schematic .....	69
Figure 3-2: Representation of Second-Stage Box Positions .....	70
Figure 3-3: Representative Snapshot Showing 5 Å Water Sphere .....	73
Figure 4-1: Rap-Ral Interface and Mutation Locations .....	101
Figure 4-2: Representative Thiocyanate-labeled Protein Spectra.....	102
Figure 4-3: Change in Absorption Energy of Thiocyanate Probes Relative to Rap E30/K31 .....	103
Figure 4-4: Azimuthal and Polar Angles of Simulated CNC Probes.....	104
Figure 4-5: Azimuthal and Polar Angles of Simulated Rap Position 30 and 31 Side chains .....	105
Figure 4-6: Convergence Test Based on CNC Dihedral Distributions .....	107
Figure 4-7: Thiocyanate Probability Distributions .....	109
Figure 4-8: Rap Position 30 Dihedral Distributions .....	111
Figure 4-9: Rap Position 31 Dihedral Distributions .....	113
Figure 5-1: Representative 2D Dihedral Probability Distributions .....	142
Figure 5-2: 2D Dihedral Probability Distributions .....	143
Figure 5-3: 1D Umbrella Sampling $\chi_2$ Dihedral Distributions .....	145

Figure 5-4: 2D Umbrella Sampling $\chi_2$ Dihedral Distribution.....	147
Figure 5-5: Field Convergence Test .....	148
Figure 5-6: Absolute Field Plots for Various Box Locations - Ral Monomer ...	151
Figure 5-7: Absolute Field Plots for Various Box Locations – Rap E30/K31 ...	153
Figure 5-8: Absolute Field Plots for Various Box Locations – Rap E30D/K31 .	155
Figure 5-9: Absolute Field Plots for Various Box Locations – Rap E30/K31E..	157
Figure 5-10: Absolute Field Plots for Various Box Locations – Rap E30D/K31E	159
Figure 5-11: Comparing Absolute and Relative Fields - Rap E30/K31 .....	161
Figure 5-12: Comparing Absolute and Relative Fields - Rap E30D/K31 .....	163
Figure 5-13: Comparing Absolute and Relative Fields - Rap E30/K31E.....	165
Figure 5-14: Comparing Absolute and Relative Fields - Rap E30D/K31E.....	167
Figure 5-15: Comparing Numeric PBE and RFM Relative Fields - Rap E30/K31	169
Figure 5-16: Comparing Numeric PBE and RFM Relative Fields - Rap E30D/K31	171
Figure 5-17: Comparing Numeric PBE and RFM Relative Fields - Rap E30/K31E	173
Figure 5-18: Comparing Numeric PBE and RFM Relative Fields - Rap E30D/K31E	
.....	175
Figure 6-1: PB Absolute Field Plots - Monomer .....	194
Figure 6-2: PB Absolute Field Plots – Rap E30/K31 .....	196
Figure 6-3: PB Absolute Field Plots – Rap E30D/K31 .....	198
Figure 6-4: PB Absolute Field Plots – Rap E30/K31E.....	200
Figure 6-5: PB Absolute Field Plots – Rap E30D/K31E .....	202
Figure 6-6: Probability of Water Lifetimes - Monomer .....	204
Figure 6-7: Probability of Water Lifetimes – Rap E30/K31.....	206
Figure 6-8: Probability of Water Lifetimes – Rap E30D/K31 .....	208
Figure 6-9: Probability of Water Lifetimes – Rap E30/K31E .....	210

Figure 6-10: Probability of Water Lifetimes – Rap E30D/K31E .....	212
Figure 6-11: Average PB Field as a Function of Sampling Time.....	213
Figure 6-12: Average Field Gradient as a Function of Sampling Time .....	214
Figure 6-13: PB Relative Field Plots – Rap E30/K31 .....	216
Figure 6-14: PB Relative Field Plots – Rap E30D/K31 .....	218
Figure 6-15: PB Relative Field Plots – Rap E30/K31E .....	220
Figure 6-16: PB Relative Field Plots – Rap E30D/K31E .....	222
Figure 7-1: Poisson-Boltzmann Solvent Reaction Fields for Methylthiocyanate and Acetophenone in Various Solvents .....	238
Figure 7-2: Methylthiocyanate Solvent Reaction Fields at Various Dielectrics, Modulated by Glycerol and Ethanol .....	239
Figure 7-3: Comparison Between Solvent Reaction Fields Calculated using Explicit TI3P Water and Implicit PB Water for All 54 GTPase/Ral Probe Combinations .....	240
Figure 8-1: Thiocyanante-Labeled Small Molecules.....	250
Figure 8-2: One-Dimensional Dihedral Probability Distributions.....	251
Figure 8-3: Two-Dimensional Dihedral Probability Distributions for Hexylthiocyanate and Capped Cyanocysteine.....	252
Figure 8-4: Dihedral Angles at Each Time Step .....	253
Figure 8-5: Average Electrostatic Field as a Function of Simulation Time .....	254
Figure 8-6: SCN Field Contributions are Constant .....	255
Figure 8-7: Small Molecule AMOEBA Fields Plotted Against Experimental Vibrational Absorption Energies .....	256
Figure 8-8: Comparing Total Field Correlations to Field Less CH <sub>2</sub> SCN Correlations .....	257

Figure 8-9: Small Molecule AMOEBA Field Standard Deviations Plotted Against  
Experimental FWHM.....258

## **List of Code and Force Field Parameters**

Code and Parameters 2-1: Cyanocysteine AMOEBA PRO13 Parameters.....	38
Code and Parameters 2-2: Methyl Thiocyanate AMOEBA PRO13 Parameters ....	40
Code and Parameters 2-3: Ethyl Thiocyanate AMOEBA PRO13 Parameters .....	41
Code and Parameters 2-4: Hexyl Thiocyanate AMOEBA PRO13 Parameters.....	45
Code and Parameters 2-5: GDPNP AMOEBA BIO09 Parameters .....	58

## Chapter 1 Introduction

Accurate and simple calculation of electrostatic fields in proteins is a subject of longstanding interest in the biophysical community. Continuum Poisson-Boltzmann (PB) models have been used to calculate pK<sub>a</sub> values for titratable amino acids.<sup>1-4</sup> Changes in free energy and pK<sub>a</sub> values have been examined using the Protein-Dipoles-Langevin-Dipoles (PDLD) model of Warshel and Levitt, which first introduced solvent polarization by mapping water dipoles to a Langevin grid and including protein dipoles, eliminating the use of a macroscopic dielectric constant in a microscopic system. Comparison to continuum models can be achieved by reintroduction of the concept of solute dielectric through the semi-macroscopic PDLD/S model and PDLD/S-LRA models, which suggest solute dielectrics of at least 2 should be used (higher when solute configurations are not well-considered).<sup>5-7</sup> All-atom electrostatics using fixed charge force fields have also seen use in understanding protein electrostatics,<sup>8-10</sup> specifically in the context of vibrational Stark effect spectroscopy (VSE), allowing for direct measurement of electrostatic fields in complicated systems. Models that include explicit atomic polarization, rather than molecular polarization like PDLD and its variants, have also been used to study protein electrostatics. The fully polarizable force field, atomic multipole optimized energetics for biomolecular applications (AMOEBA) has been used for a variety of electrostatics applications, such as binding free energies<sup>11, 12</sup> and solvation energies.<sup>13, 14</sup> Schnieders *et al.*<sup>15</sup> demonstrated that a polarizable multipole Poisson-Boltzmann continuum model, built upon AMOEBA, reproduces total protein dipole moments calculated with explicit solvent, suggesting it to be an efficient alternative to expensive solvent sampling for other electrostatics problems such as binding free energies and pK<sub>a</sub> prediction. Additionally, mixed QM/MM methods are also being used to investigate protein electrostatics.<sup>16-19</sup>

Water also poses a significant challenge in understanding and calculating protein electrostatics. Hydration water<sup>20-23</sup> is known to behave differently than bulk water; it is less mobile, has a lower dielectric constant,<sup>24</sup> and can play a role in stabilizing protein-protein and protein-ligand interactions.<sup>25,26</sup> Calculations show that the ordered first and second hydration shells have a significant effect on the electrostatic potential energies of solvated proteins which are inadequately addressed by a purely implicit model, particularly when looking at both positively and negatively charged atoms.<sup>27</sup> It has also been shown that these solvation shell effects, and therefore agreement between implicit and explicit solvent models, can be improved by optimizing atomic radii and accounting for charging the solvent in an explicit solvent system.<sup>27</sup> Alternatively, a new implicit solvent model, semi-explicit assembly (SEA), has been developed, which differentiates between bulk and structural water and has been shown to approach implicit solvent calculation speeds with the accuracy of explicit solvent models.<sup>28</sup>

Hydrogen bonding to vibrational chromophores is another difficulty in electrostatics calculations. It has been shown that using *ab initio* absorption frequency calculations on various water clusters and performing a multivariate fit to the calculated electrostatics, can then be used with standard molecular dynamics to yield good agreement with experimentally measured absorption frequencies of small molecules in water clusters.<sup>19,29,30</sup> Corcelli *et al.*<sup>31</sup> does this by finding Stark tuning rates for each x, y, z vector components of the calculated electrostatic field and summing these components, while Choi *et al.* uses a weighted sum of electrostatic potentials around the vibrational chromophore.<sup>29,30</sup> In a QM/MM study of a thiocyanate vibrational chromophore in the enzyme ketosteroid isomerase (KSI), Layfield *et al.* also demonstrated that the explicit consideration of the water molecule nearest to the vibrational probe is important for capturing solvent polarization effects, and therefore calculating vibrational Stark shifts.<sup>16</sup>

In the pK<sub>a</sub> Cooperative, scientists were challenged to calculate protein pK<sub>a</sub> values on a blind protein sample using their computational methods of choice. In the presentation of this challenge, the authors state, “accounting for the heterogeneous response of proteins is generally considered the chief difficulty in modeling pK<sub>a</sub> values in proteins.”<sup>32</sup> This statement can be generalized for *all* experimentally-measured electrostatic properties in proteins. While many electrostatic field calculations have been shown to work well for solvated small molecules,<sup>10</sup> there is an enormous difference in the number of degrees of freedom between a small (7 atom) methyl thiocyanate in a water cluster and the number of degrees of freedom in a protein, which significantly complicate the protein calculations.<sup>8,9,33,34</sup> Obtaining a good ensemble of structures is therefore much more computationally intense for a solvated protein system than for a solvated small molecule.

Our data set of vibrational absorption energies measured on systems for which we have extensive MD-based structural information provide a unique opportunity to test the accuracy of various computational outputs. We have two goals of this work. First, we want to understand the strengths and weaknesses of a variety of computational platforms for accurately predicting our data set: a collection of vibrational absorption energies. Under reasonable assumptions, all computational methods yield results which are physically reasonable, yet in a single system, or even a single MD snapshot, the results can vary significantly depending on how they were calculated. We are interested in determining the amount of physical rigor and computational expense required to obtain *in silico* electrostatic fields which are well correlated to the equivalent *in vitro* vibrational absorption frequencies and comparing the strengths and weaknesses of different electrostatic methods. Understanding why some methods are better or worse for specific problems and specific conditions will help improve understanding of the system in

general. Second, full integration of computational results into our measurements will enhance our ability to interpret experimental observations and design new experiments in a way that enhances molecular-level understanding of our system.

To this end, our group has produced a significant library of protein and small molecule systems labeled with the vibrational chromophore, thiocyanate. We have performed an extensive number of simulations and electrostatic calculations on each of the systems which experimental data exists, allowing a direct comparison between experiment and theory. The ultimate goal of which is to use these sorts of calculations to predict experimentally observed Stark shifts.

Our first and most extensive efforts in this area have been solving the Poisson-Boltzmann equation (PBE) to calculate electric fields at the midpoint of the nitrile bond. PB-based computational strategies are common in the biophysical literature because of their speed and ease of implementation, but have been criticized for the nonphysical implementation of an electrostatic continuum in a highly heterogeneous and anisotropic system. However, given their extensive use throughout the protein-electrostatics community, it is critical to understand the conditions under which PB-based strategies are appropriate and reliable.

We have also examined electrostatics calculated using the polarizable AMOEBA force field, which eliminates the need to assign a solute dielectric via explicit atomic polarization. Our first look at AMOEBA involved sampling in Amber03 and analyzing electrostatics in the AMOEBA force field. More recent ventures have examined sampling small molecules directly in AMOEBA, although the considerable speed difference makes the scalability to a larger system more difficult.

Chapter two will present a detailed description of the simulation protocols used throughout this work, including making protein residue mutations *in silico*,

multidimensional umbrella sampling, electrostatic structure clustering, quantitative orientation definitions, and setting up system in AMOEBA. Chapter three discusses the various methods used for calculating electrostatic fields. Chapter four examines electrostatic and structural interactions responsible for differential binding of RalGDS to Rap, where we observed that the charge-reversal mutation at position 31 from lysine to glutamic acid disrupts a hydrogen-bonding pocket on the RalGDS side of the Rap/Ral interface. Chapter five discusses ways to optimize APBS by changing the positive of a second-stage focusing box and by separating the electrostatic field components into a numerical PB solvent reaction field potential and an analytic Coulomb field, dubbed the reaction field method (RFM). We observed that for small second-stage box sizes, electrostatic fields were very sensitive to the box location with no clear predictability. Furthermore, separating the field using the RFM typically increased correlation with experimental absolute absorption energies, but decreased correlation in changes in absorption energies, relative to using the purely-numeric solutions to the PB equation. Chapter six looks at selectively-including explicit TIP3P water molecules as part of the solute in PB calculations, where it was observed that no reliable improvement could be obtained. Chapter seven compares a wide variety of methods for calculating electrostatic field on a library of protein ensembles. We saw that PB electrostatics were the most consistent electrostatic method, although qualitative Stark shifts are still not reproduced. Chapter 8 moves from sampling configurations in the point-charge Amber03<sup>35,36</sup> force field to the polarizable multipole AMOEBA force field.<sup>13</sup> We looked at small thiocyanate-labeled molecules and observed a qualitatively-correct Stark shift when the electrostatic field due to the induced dipole moment of the probe was included, and a qualitatively incorrect Stark shift when it was excluded.

## Chapter 2 Simulation Methods

### 2.1 INTRODUCTION

Reproducibility is a key aspect of any scientific experiment, and as such, a detailed description of how our simulations were performed and analyzed is vital. In the following chapter, the system naming conventions used, the methods for mutating side chain residues, enhanced molecular dynamics simulations, structural analysis, electrostatic clustering, and probe parameterization are presented. In cases where it is appropriate, a link to the specific code used is provided at the end of the section.

### 2.2 NOMENCLATURE

Throughout this document, we examine two GTPases, proteins p21<sup>Ras</sup> (hereafter Ras) and Rap1A (interchangeably referred to as Rap1a and Rap), the downstream effector Ral guanosine dissociation stimulator (RalGDS). In the case of RalGDS, we are often specifically referring to thiocyanate-labeled mutations of RalGDS where all cysteines have been replaced by alanine. For describing RalGDS, the names Ral, Ral $\beta$ , and RalGDS are all used interchangeably. Specific probe locations on Ral are specified in subscripts; for example, Ral N27C<sub>SCN</sub> refers to RalGDS where all cysteines have been mutated to alanine and position 27 has been mutated to a thiocyanate-labeled cyanocysteine. Mutations to the GTPases are the wild type Rap (WT Rap, Rap E30/K31), wild type Ras (WT Ras, Ras D30/E31), mutations at position 30 only (Rap E30D or Rap E30D/K31, Ras D30E or Ras D30E/E31), mutations at position 31 only (Rap K31E or Rap E30/K31E, Ras E31K or Ras D30/E31K), and mutations at both position 30 and position 31 (Rap E30D/K31E, Ras D30E/E31K).

## 2.3 LABELING AND MUTATING PROTEINS IN SILICO

Simulation parameters for cyanocysteine and GDPNP were obtained from previous studies.<sup>8,37</sup> Simulations were performed using the Amber03 force field in Gomacs.<sup>38</sup> Ral starting structures came from the 2RGF<sup>39</sup> crystal structure, Ras starting structures came from 1LFD,<sup>40</sup> and Rap starting structures came from 1GUA.<sup>41</sup>

Generation of Ral $\beta$  starting structures has previously been reported.<sup>37</sup> In short, all cysteine side chain atoms except for C $\beta$  were deleted from 2RGF and the side chain 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.<sup>35</sup> The end resultant sequence is, starting from the N-terminal, GSH+Ral(2RGF,Cys $\rightarrow$ Ala)+FT.

To make Ras, the GAMGS sequence from chain B of 4K81<sup>42</sup> 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.<sup>43</sup> 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.<sup>44</sup> 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),<sup>45</sup> 1W9R (19 NMR structures),<sup>46</sup> 2WCY (48 NMR structures)<sup>47</sup>, and 2VKJ (1 crystal structure).<sup>48</sup> Rap was aligned to the methionine backbone of each GSHM structure using VMD,<sup>49</sup> creating 88 structures containing the GSHM N-terminus. After adding hydrogen atoms using the GROMACS utility pdb2gmx,<sup>50</sup> 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 residues of Ral $\beta$  protruded between bonded atoms of Rap. To eliminate this nonphysical steric overlap, heavy atom restraints were placed on all but the Ral $\beta$  N-terminal GSHM residues and a 500 step gentle minimization was performed in GROMACS.<sup>50</sup>

All side chain mutations were generated in the same manner using Amber Tools.<sup>35</sup> 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<sup>35</sup> 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<sub>SCN</sub>, G28C<sub>SCN</sub>, N29C<sub>SCN</sub>, Y31C<sub>SCN</sub>, K32C<sub>SCN</sub>, and N54C<sub>SCN</sub>, 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.<sup>35</sup> 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<sup>51</sup> with a gap penalty of -3 (chosen because it gave results most consistent with the STAMP<sup>52</sup> structural alignment in VMD<sup>49</sup>), 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, tabulated in Table 2-1. The end result is 54 different protein systems that are each structurally *identical* to an experimentally measured system.

Table 2-1: Experimental Protein Systems Investigated and their Naming Conventions

		<b>N27C<sub>SCN</sub></b>	<b>G28C<sub>SCN</sub></b>	<b>N29C<sub>SCN</sub></b>
Monomer	<b>Ral</b>	Ral+N27C <sub>SCN</sub>	Ral+G28C <sub>SCN</sub>	Ral+N29C <sub>SCN</sub>
Wild Type	<b>Rap E30/K31</b>	Rap E30/K31+ N27C <sub>SCN</sub>	Rap E30/K31+ G28C <sub>SCN</sub>	Rap E30/K31+ N29C <sub>SCN</sub>
	<b>Ras D30/E31</b>	Ras D30/E31+ N27C <sub>SCN</sub>	Ras D30/E31+ G28C <sub>SCN</sub>	Ras D30/E31+ N29C <sub>SCN</sub>
Position 30 Mutant	<b>Rap E30D/K31</b>	Rap E30D/K31+ N27C <sub>SCN</sub>	Rap E30D/K31+ G28C <sub>SCN</sub>	Rap E30D/K31+ N29C <sub>SCN</sub>
	<b>Ras D30E/E31</b>	Ras D30E/E31+ N27C <sub>SCN</sub>	Ras D30E/E31+ G28C <sub>SCN</sub>	Ras D30E/E31+ N29C <sub>SCN</sub>
Position 31 Mutant	<b>Rap E30/K31E</b>	Rap E30/K31E+ N27C <sub>SCN</sub>	Rap E30/K31E+ G28C <sub>SCN</sub>	Rap E30/K31E+ N29C <sub>SCN</sub>
	<b>Ras D30/E31K</b>	Ras D30/E31K+ N27C <sub>SCN</sub>	Ras D30/E31K+ G28C <sub>SCN</sub>	Ras D30/E31K+ N29C <sub>SCN</sub>
Double Mutant	<b>Rap E30D/K31E</b>	Rap E30D/K31E+ N27C <sub>SCN</sub>	Rap E30D/K31E+ G28C <sub>SCN</sub>	Rap E30D/K31E+ N29C <sub>SCN</sub>
	<b>Ras D30E/E31K</b>	Ras D30E/E31K+ N27C <sub>SCN</sub>	Ras D30E/E31K+ G28C <sub>SCN</sub>	Ras D30E/E31K+ N29C <sub>SCN</sub>
		<b>K32C<sub>SCN</sub></b>	<b>Y31C<sub>SCN</sub></b>	<b>N54C<sub>SCN</sub></b>
Monomer	<b>Ral</b>	Ral+K32C <sub>SCN</sub>	Ral+Y31C <sub>SCN</sub>	Ral+N54C <sub>SCN</sub>
Wild Type	<b>Rap E30/K31</b>	Rap E30/K31+ K32C <sub>SCN</sub>	Rap E30/K31+ Y31C <sub>SCN</sub>	Rap E30/K31+ N54C <sub>SCN</sub>
	<b>Ras D30/E31</b>	Ras D30/E31+ K32C <sub>SCN</sub>	Ras D30/E31+ Y31C <sub>SCN</sub>	Ras D30/E31+ N54C <sub>SCN</sub>
Position 30 Mutant	<b>Rap E30D/K31</b>	Rap E30D/K31+ K32C <sub>SCN</sub>	Rap E30D/K31+ Y31C <sub>SCN</sub>	Rap E30D/K31+ N54C <sub>SCN</sub>
	<b>Ras D30E/E31</b>	Ras D30E/E31+ K32C <sub>SCN</sub>	Ras D30E/E31+ Y31C <sub>SCN</sub>	Ras D30E/E31+ N54C <sub>SCN</sub>
Position 31 Mutant	<b>Rap E30/K31E</b>	Rap E30/K31E+ K32C <sub>SCN</sub>	Rap E30/K31E+ Y31C <sub>SCN</sub>	Rap E30/K31E+ N54C <sub>SCN</sub>
	<b>Ras D30/E31K</b>	Ras D30/E31K+ K32C <sub>SCN</sub>	Ras D30/E31K+ Y31C <sub>SCN</sub>	Ras D30/E31K+ N54C <sub>SCN</sub>
Double Mutant	<b>Rap E30D/K31E</b>	Rap E30D/K31E+ K32C <sub>SCN</sub>	Rap E30D/K31E+ Y31C <sub>SCN</sub>	Rap E30D/K31E+ N54C <sub>SCN</sub>
	<b>Ras D30E/E31K</b>	Ras D30E/E31K+ K32C <sub>SCN</sub>	Ras D30E/E31K+ Y31C <sub>SCN</sub>	Ras D30E/E31K+ N54C <sub>SCN</sub>

## **2.4 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, using the  $\chi_2$  side chain dihedral as the biasing coordinate, labeled in Figure 2-1. Through examining simulated protein structures, it became increasingly apparent that a second degree of freedom,  $\chi_1$ , was relevant to our probe conformational distributions. Therefore, two different umbrella sampling strategies were tested: one-dimensional sampling about the thiocyanate  $\chi_2$  dihedral angle and two-dimensional umbrella sampling about the thiocyanate  $\chi_1$  and  $\chi_2$  dihedral angles, shown in Figure 2-2. All molecular dynamics were completed using the GROMACS<sup>50</sup> software package at 300 K with the AMBER03<sup>53</sup> force field and periodic boundary conditions.

Six probe locations on RalGDS were examined: N27C<sub>SCN</sub>, G28C<sub>SCN</sub>, N29C<sub>SCN</sub>, Y31C<sub>SCN</sub>, K32C<sub>SCN</sub>, and N54C<sub>SCN</sub>, 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. Knowing that the side chain was alkane-like, six structures for each system modeled were generated by fixing the thiocyanate  $\chi_2$  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<sup>-1</sup> rad<sup>-2</sup> 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<sup>45</sup> 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<sup>-1</sup> nm<sup>-2</sup> using PME<sup>54,55</sup> 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<sup>56</sup>, 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  $\chi_1$  dihedral angle on Rap residues 30 and 31 were each individually biased to create a Boltzmann-weighted ensemble of  $\chi_1$  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  $\chi_2$  dihedral angle. The weighted histogram analysis method (WHAM)<sup>57,58</sup> 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):

$$P_i = \frac{e^{-\beta \cdot PMF_i}}{\sum_{j=0}^N e^{-\beta \cdot PMF_j}}, \quad \beta = \frac{1}{k_b T} \quad (2-1)$$

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  $\chi_1$  angle was fixed from  $0^\circ$  to  $300^\circ$  in  $30^\circ$  increments, resulting in 12 structures. Each of these structures then had the  $\chi_2$  angle fixed from  $0^\circ$  to  $300^\circ$  in  $30^\circ$  increments, resulting in 144 total structures. To avoid steric clashes in the starting structures, for each  $\chi_2$  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  $1.5 \text{ \AA}$ , the  $\chi_2$  angle was rotated  $\pm 1.5^\circ$  from the dihedral center and the distances were recalculated. This was done until all non-bonded atoms were at least  $1.5 \text{ \AA}$  from each rotated atom. Next, two harmonic dihedral restraining potentials were generated for each structure, one for the  $\chi_1$  dihedral angle and one for the  $\chi_2$  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  $1000 \text{ kJ mol}^{-1} \text{ rad}^{-2}$ , 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  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  on heavy backbone atoms, dihedral force constants of  $150 \text{ kJ mol}^{-1} \text{ rad}^{-2}$ , and unrestrained side-chain atoms. Each of the 144  $\chi_1$ - $\chi_2$  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  $70 \text{ kJ mol}^{-1} \text{ rad}^{-2}$ , 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  $5184 5^\circ$  by  $5^\circ$  bins. The bins were assigned based on equation (2-2), where  $b_1$  is the one-dimensional  $\chi_1$  bin number,  $b_2$  is the one-dimensional  $\chi_2$  bin number,  $B_i$  is the total number of bins in degree of freedom  $i$  (72 for all two-dimensional sampling done), and  $i_{\text{index}}$  refers to the subscript on  $b$ . Conditional probability was assumed.

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

To validate our 2D WHAM code, we examined a Ryckaert-Bellemans dihedral potential, shown in equation (2-3), where  $\phi$  is some angle and  $n$  and  $C_n$  are some example parameters obtained from the GROMACS manual, shown in Table 2-2. 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,  $p^*(\chi_1, \chi_2)$ , 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 from  $-180^\circ$  to  $180^\circ$ .

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

Next, we constructed the PMF and probability distribution landscapes, shown in Figure 2-2a. 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  $30^\circ \times 30^\circ$ , 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  $p^*(\chi_1, \chi_2)$ .

$$p^*(\chi_1, \chi_2) = \frac{e^{-\beta(V_{rb}(\chi_1) + V_{rb}(\chi_2))}}{\int \int e^{-\beta(V_{rb}(\chi_1) + V_{rb}(\chi_2))} d\chi_1 d\chi_2} \quad (2-4)$$

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

Figure 2-2 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,  $> 17.92 \text{ kJ mol}^{-1}$ ) now quantitatively approach the predictions of the analytical expression. Exact analytical PMF matching of low probability regions is very slow, requiring  $\geq 10000$  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 2-3 summarizes the umbrella sampling setups for each project, listed by chapter.

The code for performing WHAM analysis can be accessed at:  
<https://github.com/webbgroup-physical-chemistry/wham>

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

	<b>Constraint (kJ mol<sup>-1</sup>)</b>		<b>Constraint (kJ mol<sup>-1</sup>)</b>
<b>C<sub>0</sub></b>	9.28	<b>C<sub>3</sub></b>	-3.06
<b>C<sub>1</sub></b>	12.16	<b>C<sub>4</sub></b>	-26.24
<b>C<sub>2</sub></b>	-13.12	<b>C<sub>5</sub></b>	-31.5

Table 2-3: 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 $\chi_2$	60°	Quadratic-Flat-Quadratic	6	3.0 ns	18.0 ns
4	Position 30 $\chi_1$	60°	Quadratic-Flat-Quadratic	6	3.0 ns	18.0 ns
4	Position 31 $\chi_1$	60°	Quadratic-Flat-Quadratic	6	3.0 ns	18.0 ns
5, 6	CNC $\chi_1$ , $\chi_2$	30°	Quadratic	144	0.4 ns	57.6 ns
7	CNC $\chi_1$ , $\chi_2$	30 °	Quadratic	144	2.0 ns	288.0 ns
8	N/A	N/A	N/A	N/A	4-7 ns	4-7 ns

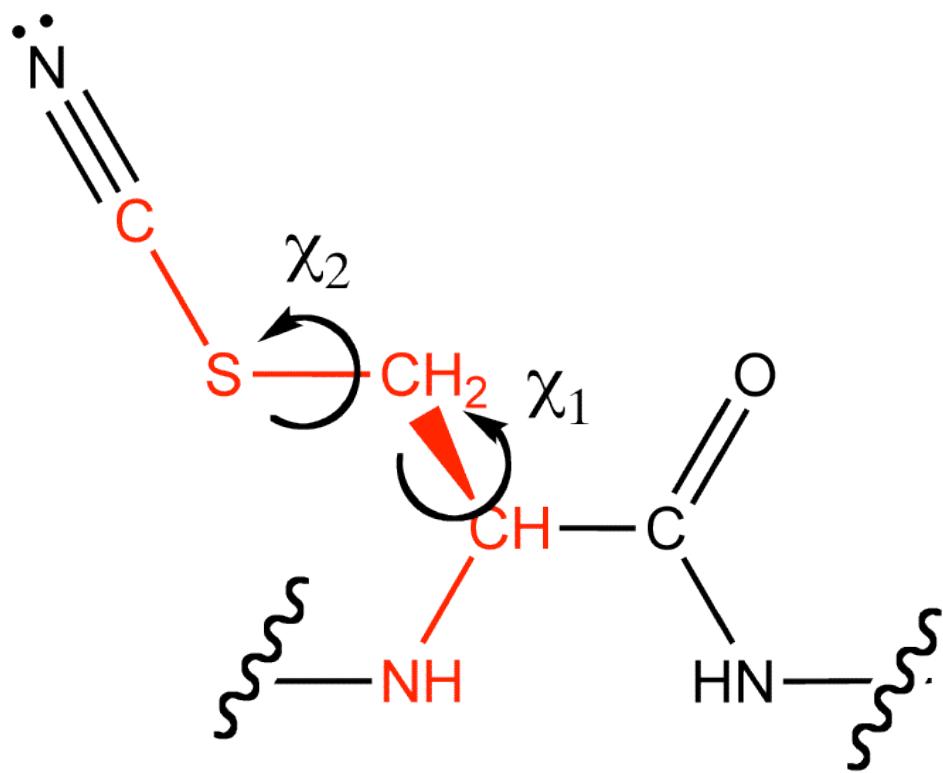


Figure 2-1: Cyanocysteine Side chain Dihedral Definitions

The thiocyanate probe, with the  $\chi_1$  and  $\chi_2$  rotation axes highlighted and the atoms involved shown in red. The  $\chi_1$  dihedral angle is defined as the dihedral of the N-C $\alpha$ -C $\beta$ -S $\gamma$  atoms. The  $\chi_2$  dihedral angle is defined as the dihedral of the C $\alpha$ -C $\beta$ -S $\gamma$ -C $\delta$  atoms.

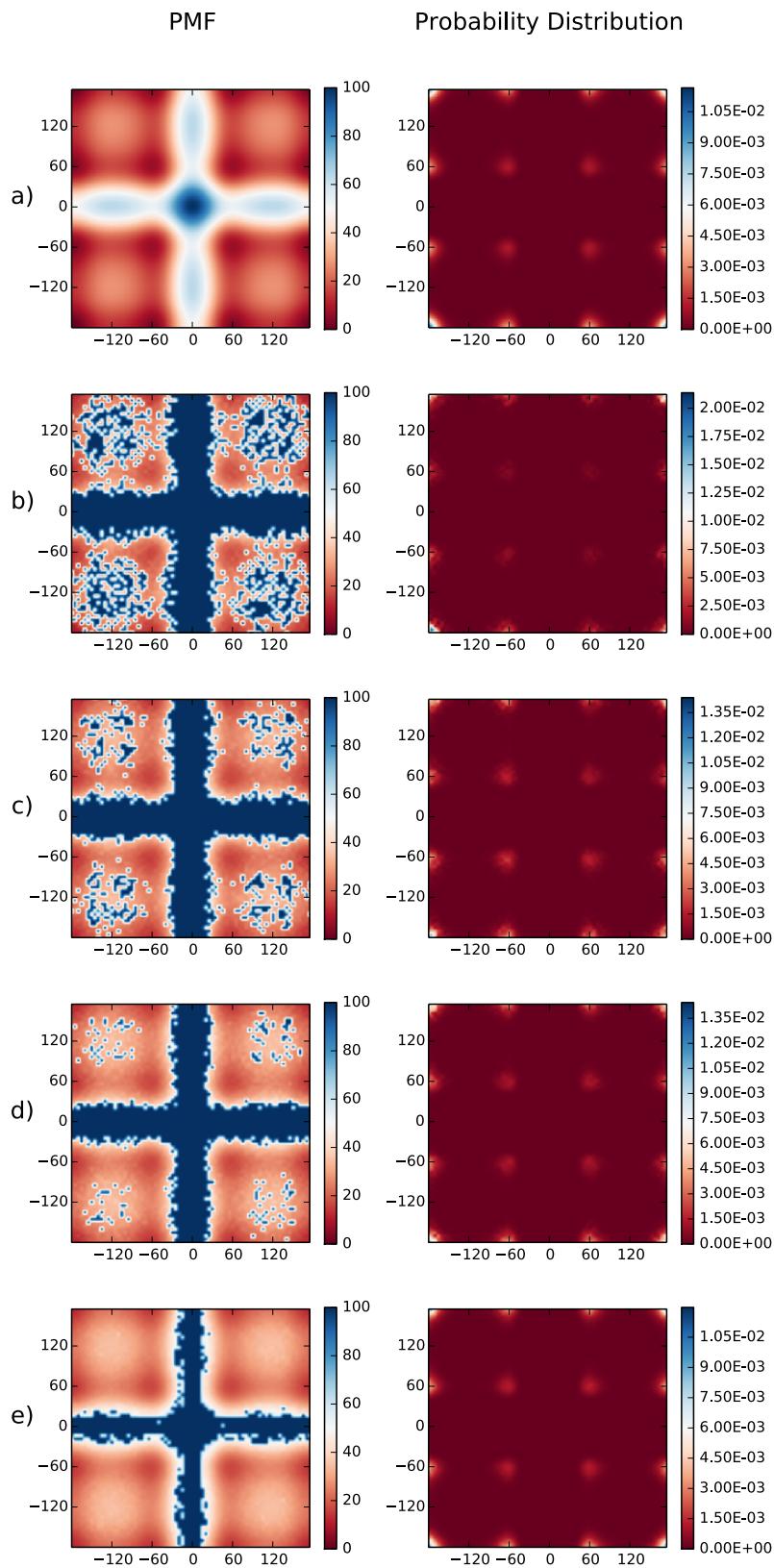


Figure 2-2: 2D WHAM Validation

Comparison between the (left) potential of mean force (PMF) and the (right) dprobability 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  $\text{kJ mol}^{-1}$ .

## **2.5 ELECTROSTATIC CLUSTERING IN VIBRATIONAL CHROMOPHORE DIHEDRAL SPACE**

One of the advantages of an implicit solvent model is that solvent is treated as an ensemble average about some solute structure. However, because we are interested in electrostatics given a complete ensemble of probe orientations, we perform electrostatics calculations on many simulated frames. This dramatically increases the computational cost of an implicit solvent model. A single node on the Stampede supercomputer at the TACC facility can generate >10 ns of simulation in explicit solvent per day (point-charge force field). That number can be increased (logarithmically) by using additional nodes. However, the continuum solvent electrostatics calculations take anywhere from 45-60 seconds (APBS) to approximately 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 2-3, 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),

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

where the probability of being in each bin  $i$  is  $\rho_i$ , 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  $\chi$  is some threshold. As  $\chi$  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  $\pm 60^\circ$  of the window center and a force constant of 30.0 kJ/mol/rad<sup>2</sup>, 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  $\chi$ , indicated in the upper-left corner of each subplot, from the full data sets in Figure 2-4. 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,  $\chi$ , in Figure 2-5. Even for a relatively large  $\chi = 1$ , the clustered correlation coefficient and slope is *very* close to 1.0. In general, as  $\chi$  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. Because this method minimizes the difference between the clustered solute Coulomb field and the full solute Coulomb field, it is trivial to verify that the clustering algorithm is behaving as expected by comparing the two averages, which should be nearly identical.

The code for electrostatic clustering, cluster\_framelist.py, can be accessed at:

[https://github.com/awritchie/python\\_scripts](https://github.com/awritchie/python_scripts)

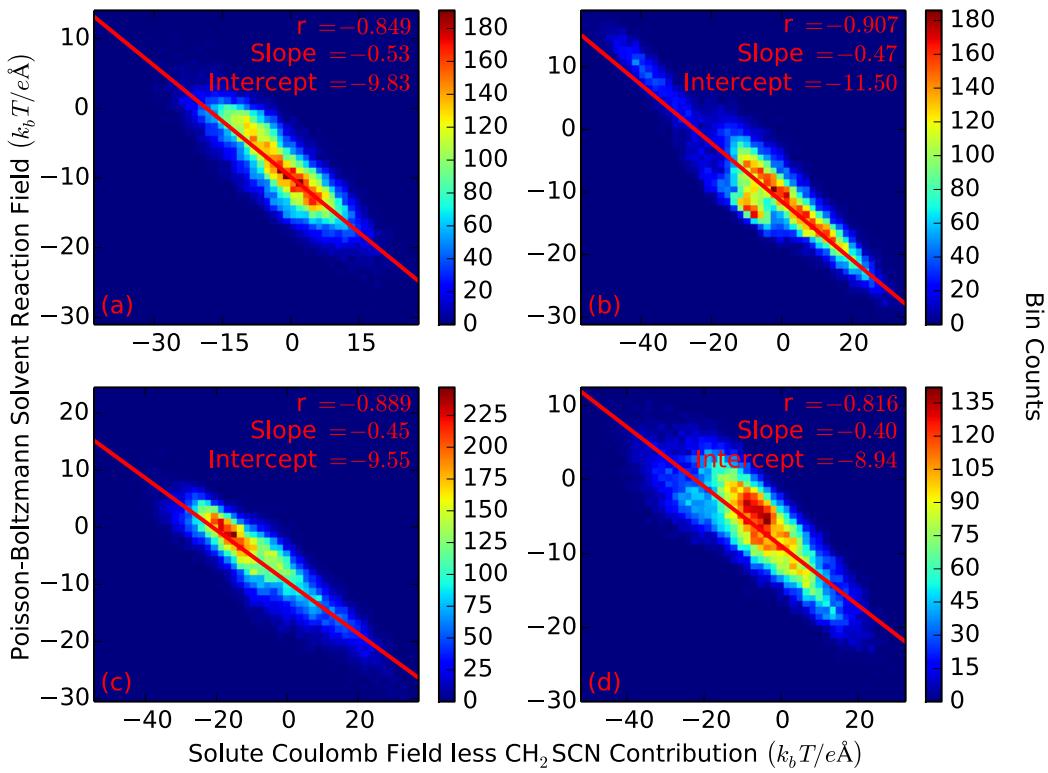


Figure 2-3: PB Solvent Reaction Field vs. Solute Analytic Coulomb Field

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<sub>SCN</sub> monomer; b) Ral N54C<sub>SCN</sub> monomer; c) Ral G28C<sub>SCN</sub> docked to wild type Rap; d) Ral N54C<sub>SCN</sub> docked to wild type Rap. Correlation coefficients (r), slopes (m), and y-intercepts (int), are indicated in the upper-right corner of each figure.

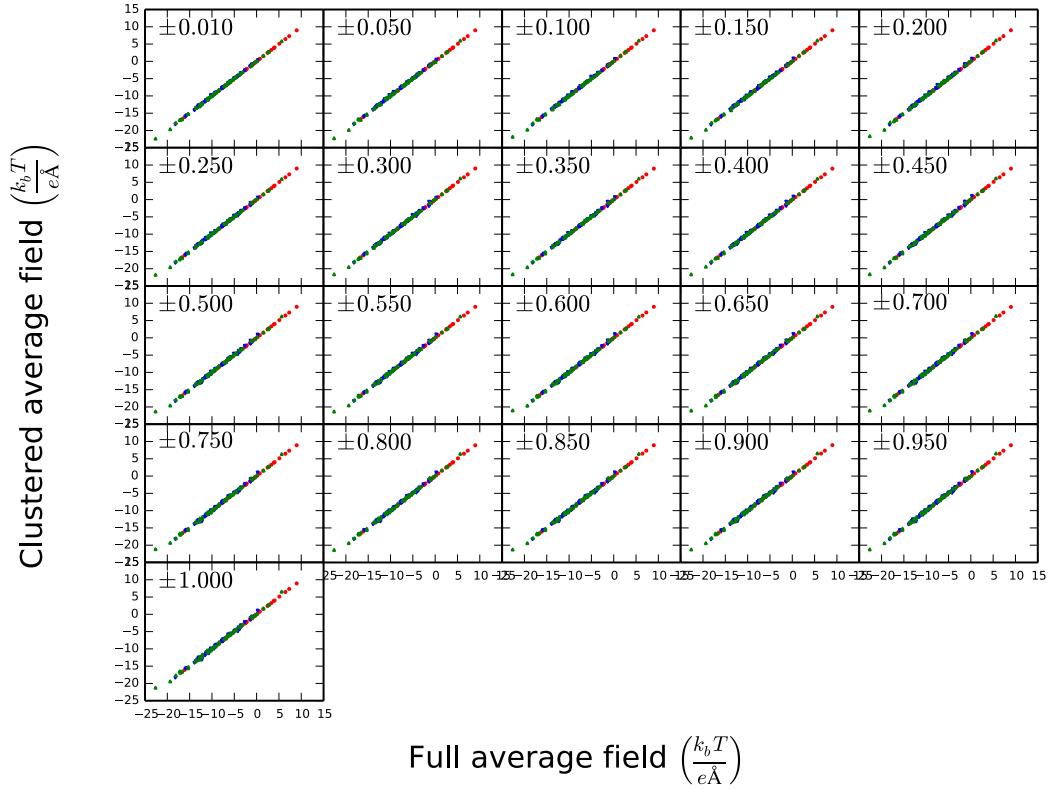


Figure 2-4: Field Values using Clustering Vs. Field Values using All Frames

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,  $\chi$ , indicated in the top left corner of each box.

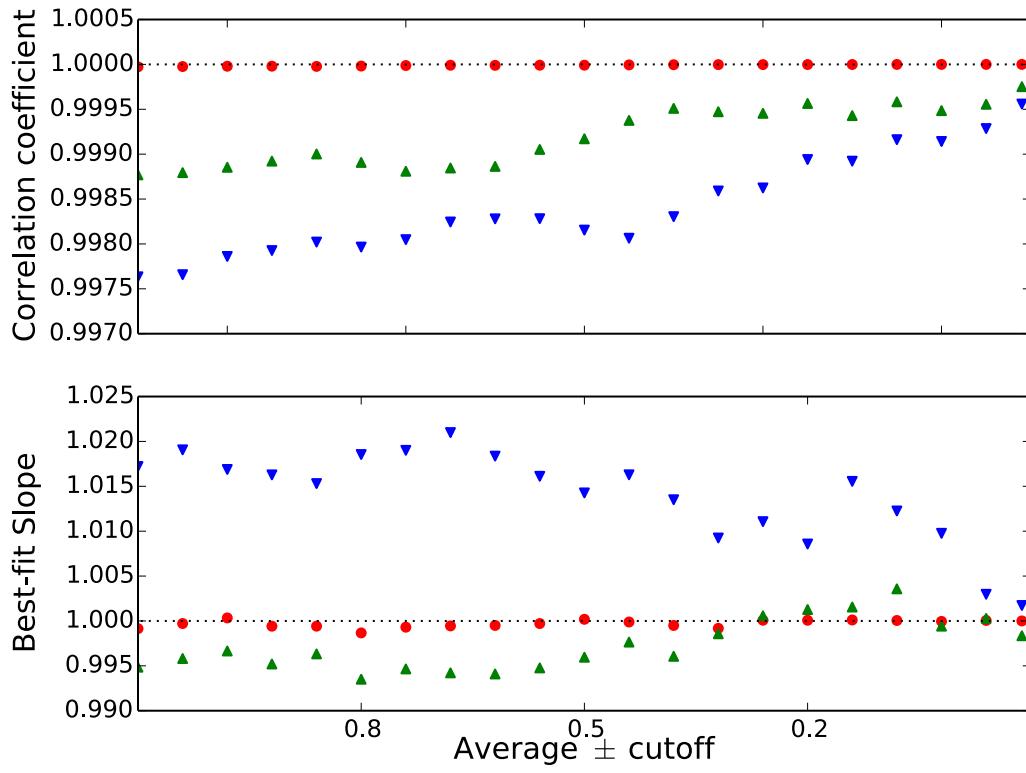


Figure 2-5: Correlations and Slopes at Various Cutoff Values

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,  $\chi$ . (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,  $\chi$ .

## 2.6 SIDE CHAIN ORIENTATION: AZIMUTHAL AND POLAR ANGLES

To analyze our molecular dynamics simulations of the torsional distribution of the thiocyanate residue on each Ral $\beta$  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-Ral $\beta$  surface, which we term azimuthal and polar angles. The azimuthal angle ( $\theta$ ) was defined by measuring the angle between the side chain of interest and the average plane between the Rap and Ral $\beta$  protein surfaces. This plane was generated in the same manner as Ensign *et al.*<sup>43</sup> The positions of the C $\alpha$  atoms of the Rap surface residues Q25, D33, P34, T35, I37, E37, D38, Y40, R41, and K42 from the Rap-Ral $\beta$  starting structure were fit with a least squares regression to obtain the Rap surface plane. Similarly, the Ral $\beta$  surface plane was obtained by a least squares fit of the C $\alpha$  atoms of the Ral $\beta$  surface residues I18, R20, N27, G28, N29, M30, Y31, K32, S33, K52, and N54 in the same structure. These two planes were averaged to define the surface plane of the Rap-Ral $\beta$  interface and a normalized vector,  $\vec{N}$ , orthogonal to this plane. The vector,  $\vec{v}$ , defining the orientation of the side chain was the C $\delta$ -N $\epsilon$  bond vector for cyanocysteine, the C $\alpha$ -C $\gamma$  vector for aspartate, the C $\alpha$ -C $\delta$  vector for glutamate, and the C $\alpha$ -N $\zeta$  vector for lysine. For each MD snapshot, the heavy atoms of each frame were aligned to the starting structure, and the azimuthal angle was calculated from Equation (2-7):

$$\theta = 90^\circ - \cos^{-1}(\vec{N} \cdot \vec{v}) \quad (2-7)$$

To generate the polar angle ( $\phi$ ), a reference axis called the polar axis was first constructed from the intersection of the surface plane defined above with a second plane defined by the least squares fit of all alpha carbons in starting structure (note: this plane is not necessarily orthogonal to the surface plane). The vector orthogonal to both the surface plane as well as this vertical plane was calculated from the cross product. This

vector,  $\vec{X}$ , is the polar axis. The angle between the polar axis and the vector in plane along the x-axis was calculated from the inverse cosine of the dot product, where the sign of the angle was determined by the sign of the y coordinate (since inverse cosine only returns the absolute value of the angle from  $0^\circ$  to  $180^\circ$ ). The angle between the nitrile C $\delta$ -N $\epsilon$  vector and the vector along the x-axis was calculated in the same manner, including sign determination. The polar angle ( $\phi$ ) was then determined by equation (2-8),

$$\phi = \cos^{-1}(\vec{x} \cdot \vec{B}_p) - \cos^{-1}(\vec{X} \cdot \vec{x}) - 90^\circ \quad (2-8)$$

where  $\vec{B}_p$  is the projection of the normalized bond vector onto the surface plane, and  $\vec{x}$  is a reference vector on the average surface plane along the  $x = 0$  axis. Figure 2-6 is a visual representation of these two angles superimposed on the structure of the Rap-Ral $\beta$  interface. Figure 2-6A defines a surface plane at the interface of Rap (above the plane) and Ral (below the plane); azimuthal angles are reported relative to this plane. When the cross hairs on Figure 2-6B are translated to the C $\alpha$  atom of each simulated residue on Rap or Ral $\beta$  (represented as spheres), they become the origin of the polar angles shown in Figure 4-4 and Figure 4-5.

To obtain Boltzmann weighted angle averages, the cyclic boundaries had to be considered to avoid obtaining averages that would be affected by an arbitrary decision for the position of our torsional angle at  $0^\circ$ . We calculated a cyclic average angle for  $n$  equally weighted values of  $\theta$  from equation (2-9). Each frame was assigned to a bin  $j$ , with each bin  $j$  having a probability of  $p_j$ , resulting in 72 bins, each having  $m_j$  equally probable members.

$$\langle \theta \rangle = \text{atan2}\left(\frac{1}{n} \sum_{j=1}^n \sin\theta_j, \frac{1}{n} \sum_{j=1}^n \cos\theta_j\right) \quad (2-9)$$

When all probabilities  $p_j$  are equal, then dividing the sum of values of  $\sin \theta$  expressed in Equation (2-9) by the quantity in each bin ( $m_j$ ) is equivalent to the special case of

summing the products of each value times the probability of that value, shown in equation (2-10):

$$\langle \theta \rangle = \text{atan}2 \left[ \sum_{j=1}^{72} \left( p_j \cdot \frac{1}{m_j} \sum_{i=1}^{m_j} \sin \theta_i \right), \sum_{j=1}^{72} \left( p_j \cdot \frac{1}{m_j} \sum_{i=1}^{m_j} \cos \theta_i \right) \right] \quad (2-10)$$

Mardia, *et al.*<sup>59</sup> have previously demonstrated that the variance for a cyclic average can be described by equation (2-11):

$$\langle \theta^2 \rangle - \langle \theta \rangle^2 = 1 - \frac{1}{n} \sqrt{\left( \sum_{j=1}^n \cos \theta_j \right)^2 + \left( \sum_{j=1}^n \sin \theta_j \right)^2} \quad (2-11)$$

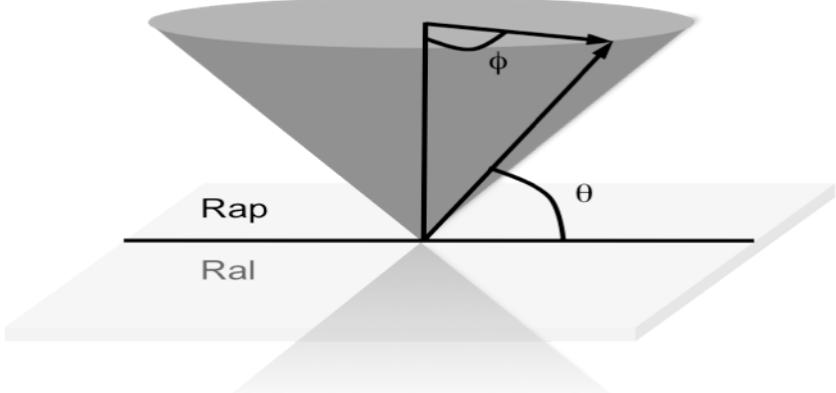
which was then rearranged to accommodate weighting, yielding equation (2-12):

$$\langle \theta^2 \rangle - \langle \theta \rangle^2 = 1 - \sqrt{\left[ \sum_{j=1}^{72} \left( p_j \cdot \frac{1}{m} \sum_{i=1}^{m_j} \sin \theta_j \right) \right]^2 + \left[ \sum_{j=1}^{72} \left( p_j \cdot \frac{1}{m} \sum_{i=1}^{m_j} \cos \theta_j \right) \right]^2} \quad (2-12)$$

The code for calculating azimuthal and polar angles, azimop.py, and the code for performing Boltzmann-weighting over a periodic coordinate, circular\_boltzmann\_Weight.py, can both be accessed at:

[https://github.com/awritchie/python\\_scripts](https://github.com/awritchie/python_scripts)

A



B

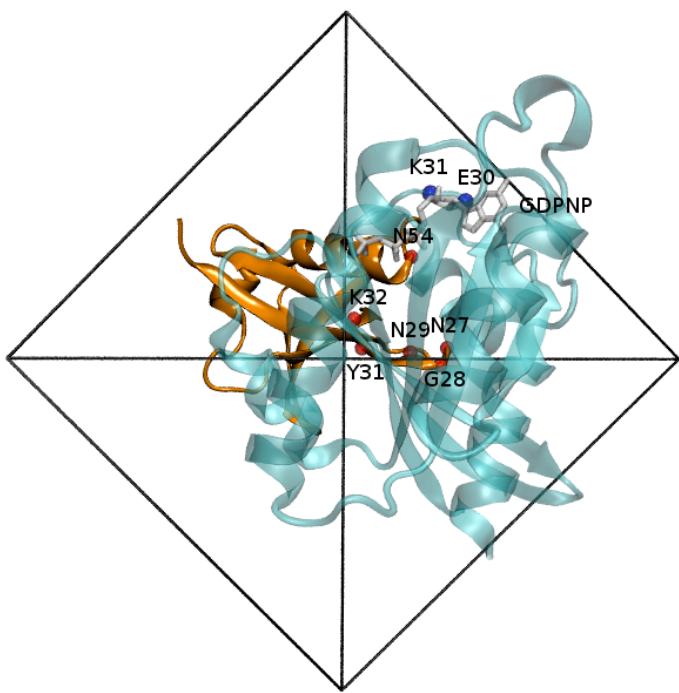


Figure 2-6: Representation of azimuthal and polar angles

Representation of the azimuthal ( $\theta$ ) and polar ( $\phi$ ) angles. (A) The surface plane was defined by the average of the plane fit with all  $C\alpha$  atoms. Azimuthal angles are shown

relative to this plane, where Rap is above the plane and Ral is below the plane. (B) Representation of the polar angle. Translating the black cross hairs to the  $C\alpha$  of each residue (represented by spheres) determines the origin of the polar angles presented in Figure 4-4 and Figure 4-5. Rap is shown in blue and Ral in orange.

## 2.7 PROBE PARAMETERIZATION FOR AMOEBA

All parameterizations were performed using poltype.py using the MP2/6-311++G2d,2p basis set for electrostatic fitting.<sup>60</sup> Starting structures were built using Avogadro and minimized prior to QM minimization using GAFF.<sup>36</sup> AMOEBA parameters used are presented for cyanocysteine (CNC), methylthiocyanate (MeSCN), hexylthiocyanate (HxSCN), ethylthiocyanate (EtSCN), and GDPNP are presented in Code and Parameters 2-1, Code and Parameters 2-2, Code and Parameters 2-3, Code and Parameters 2-4, and Code and Parameters 2-5 respectively.

### 2.7.1 Cyanocysteine parameterization

Cyanocysteine has been parameterized in both amoebabio09 as well as amoebapro13. Although only the parameters for amoebapro13 are presented, the parameterization protocols are identical. CNC was capped with an acetyl group (ACE) on the N-terminus and a N-methylamine (NME) on the C-terminus as an approximation for a continuing peptide backbone, analogous to our previous parameterization in Amber03.<sup>37</sup> This capped structured will be referred to as ACE-CNC-NME hereafter.

After completion of poltype.py, an adjusted CNC key file was generated where the ACE, NME, and backbone atoms were reassigned to the ACE, NME, and peptide backbone parameters already present in AMOEBA. The monopole charge on the C $\beta$  was then adjusted by -0.02632 e<sup>-</sup> to charge-neutralize the molecule after fixing the previously-defined parameters. Using the “fix-monopole” and “potential-fit” keywords, where the indices of the CNC side chain were supplied as arguments to potential-fit, the Tinker “potential” program was used with the QM electrostatic potential file and the adjusted CNC key file to re-optimize the side chain multipole parameters.

Following re-optimization, the MM dipole was calculated to be (5.730, 3.242, -2.811) Debyes, which is in good agreement with the QM dipole (5.7056, 3.3202, -2.8181) Debyes. The room mean square distance (RMSD) between the QM minimized structure and the MM minimized structure was found to be 0.687421 Å<sup>2</sup> and the root mean square potential difference (RMSPD) was found to be 0.3104 e<sup>-2</sup>. All comparisons between the optimized MM calculations and the QM calculations are presented in Table 2-4.

Torsional properties were omitted from the poltype.py run due to size and memory constraints. Instead, the  $\chi_1$  and related torsional properties were ported from cysteine.  $\chi_2$  torsional properties were ported from the ethylthiocyanate parameterization, described below.

## 2.7.2 GDPNP Parameterization

GDPNP was parameterized in amoebabio09. It has not been parameterized in amoebapro13, although because there are no previously defined electrostatic parameters to fit against, the multipole parameters should be the same. Because there was no initial plan to perform MD on a GTPase-GDPNP ligand bound to Ral (due to the size of the resulting protein complex being too large for Tinker with AMOEBA at the time), the torsional parameters were omitted. Since then, memory limits have prevented recent attempts to re-parameterize GDPNP with torsional parameters and a strategy analogous to that used to obtain  $\chi_1$  and  $\chi_2$  parameters for CNC should be used.

## 2.7.3 Methyl-, Ethyl-, and Hexylthiocyanate Parameterization

After initial parameterization of methyl-, ethyl-, and hexylthiocyanate, the SCN multipole parameters were fixed to those obtained for CNC and re-optimized in the manner previously described. This was done to ensure that the SCN permanent contribution to the electrostatic field at a point in space are essentially identical for all the

SCN containing probes, while still optimized to the electrostatic potential obtain from QM calculations. Comparisons between re-optimized parameters and QM calculations are presented in Table 2-4.

Table 2-4: Comparisons between AMOEBA and QM Molecular Properties

		Dipole (Debye)				RMSD (Å <sup>2</sup> )	RMSPD (e <sup>-2</sup> ) <sup>a</sup>
		X	Y	Z	Magnitude		
CNC	MM	5.730	3.242	-2.811	7.158	0.687421	0.3104
	QM	5.7056	3.3202	-2.8181	7.1770		
GDPNP	MM	-42.267	-7.814	3.987	43.168	3.540641	1.3175
	QM	-43.3488	-7.8287	4.1682	44.2468		
MeSCN	MM	4.096	0.269	0.000	4.104	0.026013	0.3785
	QM	4.1485	0.2331	0.0000	4.1551		
EtSCN	MM	-4.031	-1.329	0.324	4.256	0.036127	0.2383
	QM	-4.0267	-1.3264	0.3050	4.2505		
HxSCN	MM	4.905	-0.911	0.001	4.989	0.011824	0.1768
	QM	4.9074	-0.9083	0.0007	4.9908		

<sup>a</sup>Units are square electrons (e<sup>2</sup>), not exp[-2]

1.	atom	259	55	CB	"CNC CB"	6	12.011	4
2.	atom	260	56	SG	"CNC SG"	16	32.066	2
3.	atom	261	57	CD	"CNC CD"	6	12.011	2
4.	atom	262	58	NE	"CNC NE"	7	14.007	1
5.	atom	263	59	HB	"CNC HB"	1	1.008	1
6.								
7.	multipole	259	8	260		-0.15228		
8.						0.07407	0.00000	0.31740
9.						-0.15117		
10.						0.00000	-0.21945	
11.						-0.19282	0.00000	0.37062
12.	multipole	260	259	261		0.33074		
13.						0.44389	0.00000	0.28736
14.						1.22369		
15.						0.00000	-2.16613	
16.						-0.31636	0.00000	0.94244
17.	multipole	261	260	262		0.24556		
18.						0.06457	0.00000	-0.41800
19.						0.15740		
20.						0.00000	0.22633	
21.						0.26601	0.00000	-0.38373
22.	multipole	12	8	259		0.12898		
23.						0.02551	0.00000	0.07014
24.						0.19051		
25.						0.00000	0.17012	
26.						-0.01582	0.00000	-0.36063
27.	multipole	263	259	8		0.09179		
28.						-0.07114	0.00000	-0.02080
29.						-0.16061		
30.						0.00000	-0.02258	
31.						-0.03267	0.00000	0.18319
32.	multipole	262	261	260		-0.58843		
33.						-0.00512	0.00000	-0.24238
34.						0.32332		
35.						0.00000	0.16836	
36.						-0.08202	0.00000	-0.49168
37.								
38.	polarize	259			1.3340	0.3900	263	
39.	polarize	260			3.3000	0.3900	261	
40.	polarize	261			1.3340	0.3900	260	262
41.	polarize	262			1.0730	0.3900	261	
42.	polarize	263			0.4960	0.3900	259	
43.								
44.	vdw	58			3.7100	0.1050		
45.	vdw	55			3.7800	0.1060		
46.	vdw	56			4.0050	0.3550		
47.	vdw	57			3.7800	0.1060		
48.	vdw	59			2.8700	0.0330	0.900	
49.	bond	55	7		323.0000	1.5317		
50.	bond	55	56		235.8000	1.8353		
51.	bond	55	59		341.0000	1.0817		
52.	bond	56	57		235.8000	1.7060		
53.	bond	57	58		450.0000	1.1374		
54.	angle	7	55	56	53.2000	109.6948		

55. angle	7	55	59	42.4400	110.8048													
56. angle	56	55	59	60.2400	108.4887													
57. 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 PRO13 Parameters

1.	atom	271	67	H1	"MeSCN H1"		1	1.008	1
2.	atom	272	68	C1	"MeSCN C1"		6	12.011	4
3.	atom	273	69	S	"MeSCN S"		16	32.066	2
4.	atom	274	70	C	"MeSCN C"		6	12.011	2
5.	atom	275	71	N	"MeSCN N"		7	14.007	1
6.									
7.	multipole	272	273	271		-0.22754			
8.						0.00000	0.00000	-0.02449	
9.						0.72145			
10.						0.00000	0.72145		
11.						0.00000	0.00000	-1.44290	
12.	multipole	273	272	274		0.33074			
13.						0.44389	0.00000	0.28736	
14.						1.22369			
15.						0.00000	-2.16613		
16.						-0.31636	0.00000	0.94244	
17.	multipole	274	273	275		0.24556			
18.						0.06457	0.00000	-0.41800	
19.						0.15740			
20.						0.00000	0.22633		
21.						0.26601	0.00000	-0.38373	
22.	multipole	271	272	273		0.07989			
23.						0.02678	0.00000	-0.31473	
24.						0.66454			
25.						0.00000	-0.00998		
26.						0.17231	0.00000	-0.65456	
27.	multipole	275	274	273		-0.58843			
28.						-0.00512	0.00000	-0.24238	
29.						0.32332			
30.						0.00000	0.16836		
31.						-0.08202	0.00000	-0.49168	
32.									
33.	polarize	271			0.4960	0.3900	272		
34.	polarize	272			1.3340	0.3900	271	273	
35.	polarize	273			3.3000	0.3900	272	274	
36.	polarize	274			1.3340	0.3900	273	275	
37.	polarize	275			1.0730	0.3900	274		
38.									
39.	vdw	68			3.7800	0.1060			
40.	vdw	69			4.0050	0.3550			
41.	vdw	70			3.7800	0.1060			
42.	vdw	67			2.8700	0.0330	0.900		
43.	vdw	71			3.7100	0.1050			
44.	bond	68	69		235.8000	1.8209			
45.	bond	68	67		341.0000	1.0794			
46.	bond	69	70		235.8000	1.7068			
47.	bond	70	71		450.0000	1.1373			
48.	angle	69	68	67	60.2400	110.5747			
49.	angle	67	68	67	39.5700	110.6263			
50.	angle	68	69	70	60.4300	99.4461			
51.	angle	69	70	71	60.0000	178.9042			
52.	strbnd	69	68	67	11.5000	11.5000			
53.	strbnd	68	69	70	-5.7500	-5.7500			
54.	torsion	67	68	69	70	0.0000 0.0 1	0.0000 180.0 2	0.6600 0.0 3	

```
| 55. 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 PRO13 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		
55. vdw	64		4.0050	0.3550		
56. vdw	65		3.7800	0.1060		
57. vdw	63		2.8700	0.0330	0.900	
58. vdw	61		2.9800	0.0240	0.920	
59. vdw	66		3.7100	0.1050		
60. bond	60	62	345.3000	1.5227		
61. bond	60	61	341.0000	1.0855		
62. bond	62	64	323.0000	1.8329		
63. bond	62	63	341.0000	1.0809		
64. bond	64	65	235.8000	1.7068		
65. bond	65	66	450.0000	1.1376		
66. angle	62	60	61	42.4400	111.1890	
67. angle	61	60	61	39.5700	108.4478	
68. angle	60	62	64	53.2000	114.3035	
69. angle	60	62	63	42.4400	111.8003	
70. angle	64	62	63	60.2400	108.0491	
71. angle	63	62	63	45.5700	107.8321	
72. angle	62	64	65	60.4300	99.8280	
73. angle	64	65	66	60.0000	179.1383	
74. strbnd	62	60	61	11.5000	11.5000	
75. strbnd	60	62	64	18.7000	18.7000	
76. strbnd	60	62	63	11.5000	11.5000	
77. strbnd	64	62	63	11.5000	11.5000	
78. strbnd	62	64	65	-5.7500	-5.7500	
79. torsion	61	60	62	64	0.0000 0.0 1 0.0000 180.0 2 0.4750 0.0 3	
80. torsion	61	60	62	63	0.0000 0.0 1 0.0000 180.0 2 0.2990 0.0 3	
81. torsion	60	62	64	65	-0.4400 0.0 1 -0.2600 180.0 2 0.6000 0.0 3	
82. torsion	63	62	64	65	0.0000 0.0 1 0.0000 180.0 2 0.6600 0.0 3	
83. 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 PRO13 Parameters

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

17.	multipole	279	282	277	-0.12665		
18.					0.16365	0.00000	0.14302
19.					0.06092		
20.					0.00000	-0.43628	
21.					-0.27661	0.00000	0.37536
22.	multipole	276	277	278	-0.15938		
23.					0.00000	0.00000	0.26734
24.					-0.20136		
25.					0.00000	-0.20136	
26.					0.00000	0.00000	0.40272
27.	multipole	283	282	286	-0.11656		
28.					0.24230	0.00000	0.09285
29.					0.15205		
30.					0.00000	-0.43470	
31.					-0.43635	0.00000	0.28265
32.	multipole	282	279	283	-0.11327		
33.					0.22022	0.00000	0.08479
34.					0.15602		
35.					0.00000	-0.44910	
36.					-0.36114	0.00000	0.29308
37.	multipole	277	279	276	-0.12195		
38.					0.19367	0.00000	0.15795
39.					-0.01138		
40.					0.00000	-0.31282	
41.					-0.22362	0.00000	0.32420
42.	multipole	286	283	288	-0.18161		
43.					0.04843	0.00000	0.18545
44.					-0.01034		
45.					0.00000	-0.10804	
46.					-0.18618	0.00000	0.11838
47.	multipole	280	277	279	0.05998		
48.					0.00441	0.00000	-0.07636
49.					0.04595		
50.					0.00000	0.03338	
51.					-0.00973	0.00000	-0.07933
52.	multipole	284	282	279	0.06214		
53.					0.02649	0.00000	-0.05387
54.					0.01977		
55.					0.00000	0.03305	
56.					-0.00324	0.00000	-0.05282
57.	multipole	281	279	282	0.05805		
58.					0.00916	0.00000	-0.05550
59.					0.03877		
60.					0.00000	0.02937	
61.					-0.06123	0.00000	-0.06814
62.	multipole	285	283	282	0.07555		
63.					0.01936	0.00000	-0.04776
64.					0.05942		
65.					0.00000	0.03376	
66.					-0.01589	0.00000	-0.09318
67.	multipole	278	276	277	0.05847		
68.					0.01125	0.00000	-0.10159
69.					0.06569		
70.					0.00000	0.03435	
71.					-0.03952	0.00000	-0.10004
72.	multipole	287	286	283	0.07235		
73.					0.02117	0.00000	-0.04358

74.					0.04575
75.				0.00000	0.03869
76.				-0.01006	0.00000
77.	multipole	288	286	289	0.33074
78.				0.44389	0.00000
79.				1.22369	
80.				0.00000	-2.16613
81.				-0.31636	0.00000
82.	multipole	289	288	290	0.24556
83.				0.06457	0.00000
84.				0.15740	
85.				0.00000	0.22633
86.				0.26601	0.00000
87.	multipole	290	289	288	-0.58843
88.				-0.00512	0.00000
89.				0.32332	
90.				0.00000	0.16836
91.				-0.08202	0.00000
92.					-0.49168
93.	polarize	276		1.3340	0.3900
94.	polarize	277		1.3340	0.3900
95.	polarize	278		0.4960	0.3900
96.	polarize	279		1.3340	0.3900
97.	polarize	280		0.4960	0.3900
98.	polarize	281		0.4960	0.3900
99.	polarize	282		1.3340	0.3900
100.	polarize	283		1.3340	0.3900
101.	polarize	284		0.4960	0.3900
102.	polarize	285		0.4960	0.3900
103.	polarize	286		1.3340	0.3900
104.	polarize	287		0.4960	0.3900
105.	polarize	288		3.3000	0.3900
106.	polarize	289		1.3340	0.3900
107.	polarize	290		1.0730	0.3900
108.					289
109.	vdw	75		3.8200	0.1010
110.	vdw	72		3.8200	0.1010
111.	vdw	79		3.8200	0.1010
112.	vdw	78		3.8200	0.1010
113.	vdw	73		3.8200	0.1010
114.	vdw	82		3.7800	0.1060
115.	vdw	84		4.0050	0.3550
116.	vdw	85		3.7800	0.1060
117.	vdw	76		2.9800	0.0240
118.	vdw	80		2.9800	0.0240
119.	vdw	77		2.9800	0.0240
120.	vdw	81		2.9800	0.0240
121.	vdw	74		2.9600	0.0240
122.	vdw	83		2.8700	0.0330
123.	vdw	86		3.7100	0.1050
124.	bond	75	78	453.0000	1.5299
125.	bond	75	73	453.0000	1.5298
126.	bond	75	77	341.0000	1.0885
127.	bond	72	73	323.0000	1.5282
128.	bond	72	74	341.0000	1.0863
129.	bond	79	78	453.0000	1.5312
130.	bond	79	82	345.3000	1.5276

131.	bond	79	81	341.0000	1.0866
132.	bond	78	80	341.0000	1.0879
133.	bond	73	76	341.0000	1.0878
134.	bond	82	84	235.8000	1.8351
135.	bond	82	83	341.0000	1.0810
136.	bond	84	85	235.8000	1.7063
137.	bond	85	86	450.0000	1.1375
138.	angle	78	75	48.2000	113.1072
139.	angle	78	75	38.0000	109.3223
140.	angle	73	75	38.0000	109.2787
141.	angle	77	75	39.5700	106.3036
142.	angle	73	72	42.4400	111.1869
143.	angle	74	72	39.5700	107.7267
144.	angle	78	79	48.2000	111.5264
145.	angle	78	79	38.0000	109.6162
146.	angle	82	79	38.0000	109.5120
147.	angle	81	79	39.5700	106.9493
148.	angle	75	78	48.2000	112.7564
149.	angle	75	78	38.0000	109.3923
150.	angle	79	78	38.0000	109.3217
151.	angle	80	78	39.5700	106.4572
152.	angle	75	73	48.2000	112.9179
153.	angle	75	73	38.0000	109.3170
154.	angle	72	73	38.0000	109.3845
155.	angle	76	73	45.5700	106.3035
156.	angle	79	82	53.2000	109.0812
157.	angle	79	82	42.4400	111.1744
158.	angle	84	82	60.2400	108.2095
159.	angle	83	82	39.5700	108.8985
160.	angle	82	84	60.4300	99.5693
161.	angle	84	85	60.0000	179.0730
162.	strbnd	78	75	18.7000	18.7000
163.	strbnd	78	75	11.5000	18.7000
164.	strbnd	73	75	11.5000	18.7000
165.	strbnd	73	72	11.5000	11.5000
166.	strbnd	78	79	18.7000	18.7000
167.	strbnd	78	79	11.5000	18.7000
168.	strbnd	82	79	11.5000	18.7000
169.	strbnd	75	78	18.7000	18.7000
170.	strbnd	75	78	11.5000	18.7000
171.	strbnd	79	78	11.5000	18.7000
172.	strbnd	75	73	18.7000	18.7000
173.	strbnd	75	73	11.5000	18.7000
174.	strbnd	72	73	11.5000	18.7000
175.	strbnd	79	82	18.7000	18.7000
176.	strbnd	79	82	11.5000	11.5000
177.	strbnd	84	82	11.5000	11.5000
178.	strbnd	82	84	85	-5.7500
179.	torsion	73	75	78	0.576 0.0 1 -
0.017	180.0 2	2.031	0.0 3		
180.	torsion	73	75	78	0.000 0.0 1 0.000 180.0 2 0.000
0.0 3					
181.	torsion	77	75	78	0.000 0.0 1 0.000 180.0 2 0.000
0.0 3					
182.	torsion	77	75	78	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 PRO13 Parameters

1.	atom	412	412	C	"gdppnp	"	6	12.011	3
2.	atom	424	424	N	"gdppnp	"	7	14.007	2
3.	atom	413	413	C	"gdppnp	"	6	12.011	3
4.	atom	411	411	C	"gdppnp	"	6	12.011	3
5.	atom	409	409	N	"gdppnp	"	7	14.007	3
6.	atom	404	404	C	"gdppnp	"	6	12.011	4
7.	atom	426	426	H	"gdppnp	"	1	1.008	1
8.	atom	401	401	C	"gdppnp	"	6	12.011	4
9.	atom	427	427	H	"gdppnp	"	1	1.008	1
10.	atom	403	403	C	"gdppnp	"	6	12.011	4

11. atom	425	425	H	"gdppnp	"	1	1.008	1
12. atom	402	402	C	"gdppnp	"	6	12.011	4
13. atom	428	428	H	"gdppnp	"	1	1.008	1
14. atom	419	419	O	"gdppnp	"	8	15.999	2
15. atom	405	405	C	"gdppnp	"	6	12.011	4
16. atom	421	421	O	"gdppnp	"	8	15.999	2
17. atom	408	408	P	"gdppnp	"	15	30.974	4
18. atom	432	432	O	"gdppnp	"	8	15.999	1
19. atom	422	422	O	"gdppnp	"	8	15.999	2
20. atom	406	406	P	"gdppnp	"	15	30.974	4
21. atom	430	430	O	"gdppnp	"	8	15.999	1
22. atom	410	410	N	"gdppnp	"	7	14.007	3
23. atom	407	407	P	"gdppnp	"	15	30.974	4
24. atom	431	431	O	"gdppnp	"	8	15.999	1
25. atom	418	418	O	"gdppnp	"	8	15.999	2
26. atom	420	420	O	"gdppnp	"	8	15.999	2
27. atom	423	423	N	"gdppnp	"	7	14.007	2
28. atom	436	436	H	"gdppnp	"	1	1.008	1
29. atom	415	415	C	"gdppnp	"	6	12.011	3
30. atom	416	416	N	"gdppnp	"	7	14.007	3
31. atom	414	414	C	"gdppnp	"	6	12.011	3
32. atom	437	437	O	"gdppnp	"	8	15.999	1
33. atom	417	417	N	"gdppnp	"	7	14.007	3
34. atom	434	434	H	"gdppnp	"	1	1.008	1
35. atom	429	429	H	"gdppnp	"	1	1.008	1
36. atom	433	433	H	"gdppnp	"	1	1.008	1
37. atom	439	439	H	"gdppnp	"	1	1.008	1
38. atom	438	438	H	"gdppnp	"	1	1.008	1
39. atom	435	435	H	"gdppnp	"	1	1.008	1
40.								
41. multipole	401	403	404		0.18441			
42.				0.54166	0.00000	-0.13185		
43.				0.46099				
44.				0.00000	-0.48109			
45.				0.35158	0.00000	0.02010		
46. multipole	402	403	405		-0.07186			
47.				0.00570	0.00000	-0.19665		
48.				-0.14501				
49.				0.00000	-0.07747			
50.				-0.18026	0.00000	0.22248		
51. multipole	403	401	402		0.10855			
52.				0.34598	0.00000	-0.09976		
53.				0.01724				
54.				0.00000	-0.42498			
55.				-0.23141	0.00000	0.40774		
56. multipole	404	401	409		-0.10995			
57.				0.11552	0.00000	-0.50485		
58.				0.05753				
59.				0.00000	-0.23286			
60.				-0.06016	0.00000	0.17533		
61. multipole	405	402	421		0.21485			
62.				0.41938	0.00000	0.21068		
63.				0.42325				
64.				0.00000	-0.63492			
65.				-0.51227	0.00000	0.21167		
66. multipole	406	430	-430		1.60715			
67.				0.00000	0.00000	0.38836		

68.			0.65230	
69.			0.00000	-0.16800
70.			0.00000	0.00000
71.	multipole	407 431 -431	1.61701	-0.48430
72.			0.00000	0.00000
73.			-0.24178	0.18477
74.			0.00000	0.11918
75.			0.00000	0.00000
76.	multipole	408 432 -432	1.72326	0.12260
77.			0.00000	0.00000
78.			0.50171	0.11629
79.			0.00000	-0.29704
80.			0.00000	0.00000
81.	multipole	409 404 411	0.13081	-0.20467
82.			-0.12104	0.00000
83.			-0.24418	0.37245
84.			0.00000	-0.92607
85.			-0.26408	0.00000
86.	multipole	410 406 407	-0.59954	1.17025
87.			0.05754	0.00000
88.			0.18757	0.00437
89.			0.00000	-1.10534
90.			-0.00203	0.00000
91.	multipole	411 409 413	0.14282	0.91777
92.			-0.49171	0.00000
93.			0.34909	-0.21409
94.			0.00000	0.12884
95.			-0.36179	0.00000
96.	multipole	412 409 424	0.35661	-0.47793
97.			0.54605	0.00000
98.			-0.19657	0.25474
99.			0.00000	-0.10204
100.			0.18920	0.00000
101.	multipole	413 411 414	-0.03280	0.29861
102.			-0.22277	0.00000
103.			-0.48371	-0.25045
104.			0.00000	-0.03793
105.			0.30907	0.00000
106.	multipole	414 413 416	0.91348	0.52164
107.			-0.17457	0.00000
108.			0.58772	-0.60091
109.			0.00000	0.09596
110.			-0.24587	0.00000
111.	multipole	415 416 417	0.84215	-0.68368
112.			0.13078	0.00000
113.			0.04395	-0.03914
114.			0.00000	-0.63090
115.			-0.36796	0.00000
116.	multipole	416 414 415	-0.43723	0.58695
117.			-0.14280	0.00000
118.			0.82790	-0.18174
119.			0.00000	0.43039
120.			0.25116	0.00000
121.	multipole	417 415 436	-0.38359	-0.39751
122.			0.00000	0.00000
123.			0.01209	-0.13042
124.			0.00000	-0.15222

125.					0.00000	0.00000	0.14013
126.	multipole	418	403	439	-0.45378		
127.					0.02892	0.00000	0.38162
128.					-0.09316		
129.					0.00000	-0.52837	
130.					-0.22732	0.00000	0.62153
131.	multipole	419	402	404	-0.18965		
132.					0.61712	0.00000	0.37792
133.					0.78739		
134.					0.00000	-1.05726	
135.					0.22081	0.00000	0.26987
136.	multipole	420	401	438	-0.47695		
137.					0.01271	0.00000	0.38516
138.					-0.06910		
139.					0.00000	-0.38397	
140.					-0.17741	0.00000	0.45307
141.	multipole	421	405	408	-0.56293		
142.					0.10431	0.00000	0.32156
143.					-0.39433		
144.					0.00000	-0.71106	
145.					0.10017	0.00000	1.10539
146.	multipole	422	406	408	-0.71875		
147.					0.23767	0.00000	-0.09277
148.					0.46399		
149.					0.00000	0.30398	
150.					-0.09204	0.00000	-0.76797
151.	multipole	423	411	415	-0.64948		
152.					0.24801	0.00000	0.16297
153.					0.00530		
154.					0.00000	0.45832	
155.					0.27214	0.00000	-0.46362
156.	multipole	424	412	413	-0.45369		
157.					0.17253	0.00000	0.37845
158.					-0.16847		
159.					0.00000	0.05902	
160.					0.26503	0.00000	0.10945
161.	multipole	425	403	401	0.09431		
162.					0.03599	0.00000	0.01400
163.					0.19957		
164.					0.00000	0.13766	
165.					-0.05022	0.00000	-0.33723
166.	multipole	426	404	401	-0.00163		
167.					-0.02932	0.00000	-0.09982
168.					0.02584		
169.					0.00000	-0.03045	
170.					0.03156	0.00000	0.00461
171.	multipole	427	401	403	0.04804		
172.					-0.02265	0.00000	-0.10526
173.					-0.13798		
174.					0.00000	-0.13074	
175.					-0.12883	0.00000	0.26872
176.	multipole	428	402	403	0.06206		
177.					-0.06054	0.00000	-0.03758
178.					0.00063		
179.					0.00000	0.09500	
180.					0.00384	0.00000	-0.09563
181.	multipole	429	405	402	0.00223		

182.				-0.06041	0.00000	-0.09709
183.				0.03493		
184.				0.00000	0.11917	
185.				-0.34254	0.00000	-0.15410
186.	multipole	430	406	430		
187.				-0.99932		
188.				0.00000	0.00000	-0.07283
189.				-0.06383		
190.				0.00000	-0.36237	
191.	multipole	431	407	431		
192.				-1.04278		
193.				0.00000	0.00000	-0.13648
194.				-0.05248		
195.				0.00000	0.30494	
196.	multipole	432	408	432		
197.				-0.99350		
198.				0.00000	0.00000	-0.12882
199.				-0.00215		
200.				0.00000	-0.08488	
201.	multipole	433	410	406		
202.				0.11310		
203.				-0.04947	0.00000	0.11375
204.				0.35956		
205.				0.00000	-0.32594	
206.	multipole	434	412	409		
207.				0.05761	0.00000	-0.03362
208.				0.04567		
209.				-0.14533	0.00000	-0.04939
210.				-0.01513		
211.	multipole	435	416	414		
212.				0.00000	0.18824	
213.				0.43617	0.00000	-0.17311
214.				0.10193		
215.				0.02972	0.00000	-0.17078
216.	multipole	436	417	415		
217.				-0.10315		
218.				0.00000	-0.07243	
219.				-0.17258	0.00000	0.17558
220.				0.10217		
221.	multipole	437	414	413		
222.				0.04458	0.00000	-0.35375
223.				0.02516		
224.				0.00000	-0.01409	
225.				0.24307	0.00000	-0.01107
226.	multipole	438	420	401		
227.				-0.85364		
228.				-0.03170	0.00000	-0.31120
229.				-0.40976		
230.				0.00000	0.47838	
231.	multipole	439	418	403		
232.				-0.07939	0.00000	-0.06862
233.				0.31216		
234.				-0.07689	0.00000	0.36540
235.				0.15082		
236.				0.00000	0.33844	
237.	polarize	412		0.22936	0.00000	-0.48925
238.	polarize	424		0.28229		
				-0.17802	0.00000	0.31294
				0.07796		
				0.00000	0.06518	
				0.22891	0.00000	-0.14314
				1.750	0.390 424 409 434	
				1.073	0.390 412 413	

239.	polarize	413	1.750	0.390 424 411 414
240.	polarize	411	1.750	0.390 413 409 423
241.	polarize	409	1.073	0.390 412 411
242.	polarize	404	1.334	0.390 426 401 419
243.	polarize	426	0.496	0.390 404
244.	polarize	401	1.334	0.390 404 427 403 420
245.	polarize	427	0.496	0.390 401
246.	polarize	403	1.334	0.390 401 425 402 418
247.	polarize	425	0.496	0.390 403
248.	polarize	402	1.334	0.390 403 428 419
249.	polarize	428	0.496	0.390 402
250.	polarize	419	0.837	0.390 404 402
251.	polarize	405	1.334	0.390 421 429
252.	polarize	421	0.837	0.390 405 408
253.	polarize	408	1.828	0.390 421 432 422
254.	polarize	432	0.837	0.390 408
255.	polarize	422	0.837	0.390 408 406
256.	polarize	406	1.828	0.390 422 430
257.	polarize	430	0.837	0.390 406
258.	polarize	410	1.073	0.390 433
259.	polarize	407	1.828	0.390 431
260.	polarize	431	0.837	0.390 407
261.	polarize	418	0.837	0.390 403 439
262.	polarize	420	0.837	0.390 401 438
263.	polarize	423	1.073	0.390 411 415
264.	polarize	436	0.496	0.390 417
265.	polarize	415	1.750	0.390 423 416 417
266.	polarize	416	1.073	0.390 415 414 435
267.	polarize	414	1.750	0.390 413 416 437
268.	polarize	437	0.837	0.390 414
269.	polarize	417	1.073	0.390 436 415
270.	polarize	434	0.696	0.390 412
271.	polarize	429	0.496	0.390 405
272.	polarize	433	0.496	0.390 410
273.	polarize	439	0.496	0.390 418
274.	polarize	438	0.496	0.390 420
275.	polarize	435	0.496	0.390 416
276.				
277.				
278.	vdw	401	3.6500	0.1010
279.	vdw	402	3.6500	0.1010
280.	vdw	403	3.6500	0.1010
281.	vdw	404	3.6500	0.1010
282.	vdw	405	3.8200	0.1010
283.	vdw	406	4.4500	0.3900
284.	vdw	407	4.4500	0.3900
285.	vdw	408	4.4500	0.3900
286.	vdw	409	3.7100	0.1050
287.	vdw	410	3.7100	0.1050
288.	vdw	411	3.7800	0.1010
289.	vdw	412	3.7800	0.1010
290.	vdw	413	3.7800	0.1010
291.	vdw	414	3.7800	0.1010
292.	vdw	415	3.7800	0.1010
293.	vdw	416	3.7100	0.1050
294.	vdw	417	3.7100	0.1050
295.	vdw	418	3.4050	0.1100

296.	vdw	419	3.4050	0.1100
297.	vdw	420	3.4050	0.1100
298.	vdw	421	3.4050	0.1100
299.	vdw	422	3.4050	0.1120
300.	vdw	423	3.7100	0.1050
301.	vdw	424	3.7100	0.1050
302.	vdw	425	2.8700	0.0240 0.910
303.	vdw	426	2.9800	0.0240 0.940
304.	vdw	427	2.8700	0.0240 0.910
305.	vdw	428	2.9800	0.0240 0.940
306.	vdw	429	2.9800	0.0240 0.940
307.	vdw	430	3.6300	0.1120
308.	vdw	431	3.6300	0.1120
309.	vdw	432	3.6300	0.1120
310.	vdw	433	2.7000	0.0200 0.910
311.	vdw	434	3.0000	0.0240 0.940
312.	vdw	435	2.5900	0.0200 0.910
313.	vdw	436	2.7000	0.0200 0.910
314.	vdw	437	3.3000	0.1120
315.	vdw	438	2.6550	0.0135 0.910
316.	vdw	439	2.6550	0.0135 0.910
317.	bond	401	403	323.0000 1.5279
318.	bond	401	404	323.0000 1.5474
319.	bond	401	420	465.1000 1.3858
320.	bond	401	427	341.0000 1.0849
321.	bond	402	403	323.0000 1.5126
322.	bond	402	405	323.0000 1.5297
323.	bond	402	419	465.1000 1.4600
324.	bond	402	428	341.0000 1.0811
325.	bond	403	418	465.1000 1.4020
326.	bond	403	425	341.0000 1.0703
327.	bond	404	409	400.0000 1.4866
328.	bond	404	419	465.1000 1.3806
329.	bond	404	426	341.0000 1.0809
330.	bond	405	421	465.1000 1.3924
331.	bond	405	429	341.0000 1.0802
332.	bond	406	410	500.0000 1.6791
333.	bond	406	422	450.0000 1.7014
334.	bond	406	430	775.0000 1.4801
335.	bond	407	410	500.0000 1.7654
336.	bond	407	431	775.0000 1.5417
337.	bond	408	421	450.0000 1.6763
338.	bond	408	422	450.0000 1.5941
339.	bond	408	432	775.0000 1.4854
340.	bond	409	411	653.9000 1.3378
341.	bond	409	412	653.9000 1.3697
342.	bond	410	433	520.0000 0.9998
343.	bond	411	413	539.6000 1.3815
344.	bond	411	423	653.9000 1.3693
345.	bond	412	424	653.9000 1.2921
346.	bond	412	434	370.5000 1.0735
347.	bond	413	414	539.6000 1.4238
348.	bond	413	424	653.9000 1.3756
349.	bond	414	416	653.9000 1.4146
350.	bond	414	437	680.0000 1.2095
351.	bond	415	416	653.9000 1.3604
352.	bond	415	417	250.0000 1.3984

353.	bond	415	423	653.9000	1.2691
354.	bond	416	435	467.6000	0.9954
355.	bond	417	436	520.0000	1.0012
356.	bond	418	439	615.9000	0.9919
357.	bond	420	438	615.9000	0.9961
358.	angle	403	401	404	48.2000 102.0140
359.	angle	403	401	420	88.0000 113.2745
360.	angle	403	401	427	42.4400 109.7960
361.	angle	404	401	420	88.0000 108.3134
362.	angle	404	401	427	38.0000 111.8677
363.	angle	420	401	427	60.9900 111.2532
364.	angle	403	402	405	60.0000 118.0537
365.	angle	403	402	419	88.0000 103.6718
366.	angle	403	402	428	42.4400 108.5507
367.	angle	405	402	419	88.0000 110.9297
368.	angle	405	402	428	38.0000 107.9502
369.	angle	419	402	428	60.9900 107.1771
370.	angle	401	403	402	48.2000 102.3474
371.	angle	401	403	418	88.0000 109.2291
372.	angle	401	403	425	42.4400 111.3519
373.	angle	402	403	418	88.0000 113.0692
374.	angle	402	403	425	38.0000 109.7564
375.	angle	418	403	425	60.9900 110.8215
376.	angle	401	404	409	80.0000 114.1295
377.	angle	401	404	419	88.0000 108.7414
378.	angle	401	404	426	42.4400 110.0164
379.	angle	409	404	419	60.0000 107.3638
380.	angle	409	404	426	50.6000 104.7988
381.	angle	419	404	426	60.9900 111.7784
382.	angle	402	405	421	88.0000 111.2538
383.	angle	402	405	429	38.0000 110.3075
384.	angle	421	405	429	60.9900 111.3454
385.	angle	429	405	429	39.5700 108.1476
386.	angle	410	406	422	75.0000 100.7964
387.	angle	410	406	430	75.0000 112.7444
388.	angle	422	406	430	75.8600 105.8936
389.	angle	430	406	430	89.8800 120.5164
390.	angle	410	407	431	75.0000 107.3898
391.	angle	431	407	431	89.8800 112.1177
392.	angle	421	408	422	65.5800 97.8372
393.	angle	421	408	432	75.8600 103.5709
394.	angle	422	408	432	75.8600 113.3530
395.	angle	432	408	432	89.8800 120.0260
396.	angle	404	409	411	65.0000 127.5716
397.	angle	404	409	412	65.0000 126.1994
398.	angle	411	409	412	86.3300 105.8179
399.	angle	406	410	407	65.0000 143.8484
400.	angle	406	410	433	35.0000 108.4789
401.	angle	407	410	433	35.0000 107.5834
402.	angle	409	411	413	69.7800 106.6712
403.	angle	409	411	423	28.7800 126.4981
404.	angle	413	411	423	69.7800 126.8217
405.	angle	409	412	424	28.7800 113.5567
406.	angle	409	412	434	38.1300 119.0165
407.	angle	424	412	434	38.1300 127.4267
408.	angle	411	413	414	64.6700 119.3867
409.	angle	411	413	424	69.7800 109.6158

410.	angle	414	413	424	69.7800	130.9975
411.	angle	413	414	416	69.7800	110.1025
412.	angle	413	414	437	60.0000	132.3816
413.	angle	416	414	437	60.0000	117.5156
414.	angle	416	415	417	60.0000	114.8839
415.	angle	416	415	423	28.7800	124.5114
416.	angle	417	415	423	60.0000	120.5996
417.	angle	414	416	415	86.3300	125.5872
418.	angle	414	416	435	35.2500	114.6460
419.	angle	415	416	435	35.2500	119.2990
420.	angle	415	417	436	35.0000	112.0655
421.	angle	436	417	436	35.0000	109.7768
422.	angle	403	418	439	64.9600	101.1802
423.	angle	402	419	404	88.5000	108.8848
424.	angle	401	420	438	64.9600	103.2471
425.	angle	405	421	408	80.3000	120.6283
426.	angle	406	422	408	80.0000	140.4575
427.	angle	411	423	415	86.3300	113.5733
428.	angle	412	424	413	86.3300	104.3332
429.	strbnd	403	401	404	38.0000	38.0000
430.	strbnd	403	401	420	18.7000	18.7000
431.	strbnd	403	401	427	11.5000	11.5000
432.	strbnd	404	401	420	18.7000	18.7000
433.	strbnd	404	401	427	11.5000	18.7000
434.	strbnd	420	401	427	-4.5000	38.0000
435.	strbnd	403	402	405	18.7000	18.7000
436.	strbnd	403	402	419	18.7000	18.7000
437.	strbnd	403	402	428	11.5000	11.5000
438.	strbnd	405	402	419	18.7000	18.7000
439.	strbnd	405	402	428	11.5000	18.7000
440.	strbnd	419	402	428	-4.5000	38.0000
441.	strbnd	401	403	402	38.0000	38.0000
442.	strbnd	401	403	418	18.7000	18.7000
443.	strbnd	401	403	425	11.5000	11.5000
444.	strbnd	402	403	418	18.7000	18.7000
445.	strbnd	402	403	425	11.5000	18.7000
446.	strbnd	418	403	425	-4.5000	38.0000
447.	strbnd	401	404	409	18.7000	18.7000
448.	strbnd	401	404	419	18.7000	18.7000
449.	strbnd	401	404	426	11.5000	11.5000
450.	strbnd	409	404	419	18.7000	18.7000
451.	strbnd	409	404	426	11.5000	18.7000
452.	strbnd	419	404	426	-4.5000	38.0000
453.	strbnd	402	405	421	18.7000	18.7000
454.	strbnd	402	405	429	11.5000	18.7000
455.	strbnd	421	405	429	-4.5000	38.0000
456.	strbnd	410	406	422	14.4000	14.4000
457.	strbnd	410	406	430	14.4000	14.4000
458.	strbnd	422	406	430	14.4000	14.4000
459.	strbnd	410	407	431	14.4000	14.4000
460.	strbnd	431	407	431	14.4000	14.4000
461.	strbnd	421	408	422	14.4000	14.4000
462.	strbnd	421	408	432	14.4000	14.4000
463.	strbnd	422	408	432	14.4000	14.4000
464.	strbnd	404	409	411	7.2000	7.2000
465.	strbnd	404	409	412	7.2000	7.2000
466.	strbnd	411	409	412	14.4000	14.4000

467.	strbnd	406	410	407	7.2000	7.2000
468.	strbnd	406	410	433	4.3000	7.2000
469.	strbnd	407	410	433	4.3000	7.2000
470.	strbnd	409	411	413	18.7000	18.7000
471.	strbnd	409	411	423	18.7000	18.7000
472.	strbnd	413	411	423	18.7000	18.7000
473.	strbnd	409	412	424	18.7000	18.7000
474.	strbnd	409	412	434	11.5000	11.5000
475.	strbnd	424	412	434	11.5000	11.5000
476.	strbnd	411	413	414	18.7000	18.7000
477.	strbnd	411	413	424	18.7000	18.7000
478.	strbnd	414	413	424	18.7000	18.7000
479.	strbnd	413	414	416	18.7000	18.7000
480.	strbnd	416	415	417	18.7000	18.7000
481.	strbnd	416	415	423	18.7000	18.7000
482.	strbnd	417	415	423	18.7000	18.7000
483.	strbnd	414	416	415	14.4000	14.4000
484.	strbnd	414	416	435	4.3000	4.3000
485.	strbnd	415	416	435	4.3000	4.3000
486.	strbnd	415	417	436	4.3000	7.2000
487.	strbnd	403	418	439	-4.5000	38.0000
488.	strbnd	402	419	404	38.0000	38.0000
489.	strbnd	401	420	438	-4.5000	38.0000
490.	strbnd	405	421	408	38.0000	38.0000
491.	strbnd	406	422	408	38.0000	38.0000
492.	strbnd	411	423	415	14.4000	14.4000
493.	strbnd	412	424	413	14.4000	14.4000
494.	opbend	404	409	0	0	0.2002
495.	opbend	409	411	0	0	0.2000
496.	opbend	409	412	0	0	0.2000
497.	opbend	411	409	0	0	0.1500
498.	opbend	411	413	0	0	0.2000
499.	opbend	411	423	0	0	0.1500
500.	opbend	412	409	0	0	0.1500
501.	opbend	412	424	0	0	0.1500
502.	opbend	413	411	0	0	0.2000
503.	opbend	413	414	0	0	0.2000
504.	opbend	413	424	0	0	0.1500
505.	opbend	414	413	0	0	0.2000
506.	opbend	414	416	0	0	0.1500
507.	opbend	415	416	0	0	0.1500
508.	opbend	415	423	0	0	0.1500
509.	opbend	416	414	0	0	0.2500
510.	opbend	416	415	0	0	0.2500
511.	opbend	417	415	0	0	0.2002
512.	opbend	423	411	0	0	0.2500
513.	opbend	423	415	0	0	0.2500
514.	opbend	424	412	0	0	0.2000
515.	opbend	424	413	0	0	0.2000
516.	opbend	434	412	0	0	0.2100
517.	opbend	435	416	0	0	0.1500
518.	opbend	437	414	0	0	0.2002
519.	torsion	404	401	403	402	0.1800 0.0 1      0.1700 180.0 2
		0.5200	0.0	3		
520.	torsion	404	401	403	418	-
		1.1500	0.0	1	0.0000	180.0 2      1.2800 0.0 3

521.	torsion	404	401	403	425	0.0000 0.0 1	0.0000 180.0 2
	0.2800 0.0 3						
522.	torsion	420	401	403	402	-	
	1.1500 0.0 1	0.0000 180.0 2			1.2800 0.0 3		
523.	torsion	420	401	403	418	2.2200 0.0 1	-1.3800 180.0 2
	1.1800 0.0 3						-
524.	torsion	420	401	403	425	0.0000 0.0 1	0.0000 180.0 2
	0.3000 0.0 3						
525.	torsion	427	401	403	402	0.0000 0.0 1	0.0000 180.0 2
	0.2800 0.0 3						
526.	torsion	427	401	403	418	0.0000 0.0 1	0.0000 180.0 2
	0.3000 0.0 3						
527.	torsion	427	401	403	425	0.0000 0.0 1	0.0000 180.0 2
	0.2380 0.0 3						
528.	torsion	403	401	404	409	0.0000 0.0 1	0.0000 180.0 2
	0.3000 0.0 3						
529.	torsion	403	401	404	419	-	
	1.1500 0.0 1	0.0000 180.0 2			1.2800 0.0 3		
530.	torsion	403	401	404	426	0.0000 0.0 1	0.0000 180.0 2
	0.2800 0.0 3						
531.	torsion	420	401	404	409	0.0000 0.0 1	0.0000 180.0 2
	0.3000 0.0 3						
532.	torsion	420	401	404	419	2.2200 0.0 1	-1.3800 180.0 2
	1.1800 0.0 3						-
533.	torsion	420	401	404	426	0.0000 0.0 1	0.0000 180.0 2
	0.3000 0.0 3						
534.	torsion	427	401	404	409	0.0000 0.0 1	0.0000 180.0 2
	0.5000 0.0 3						
535.	torsion	427	401	404	419	0.0000 0.0 1	0.0000 180.0 2
	0.3000 0.0 3						
536.	torsion	427	401	404	426	0.0000 0.0 1	0.0000 180.0 2
	0.2380 0.0 3						
537.	torsion	403	401	420	438	-	
	1.3720 0.0 1	0.2320 180.0 2			0.4000 0.0 3		
538.	torsion	404	401	420	438	-	
	1.3720 0.0 1	0.2320 180.0 2			0.4000 0.0 3		
539.	torsion	427	401	420	438	0.0000 0.0 1	0.0000 180.0 2
	0.2660 0.0 3						
540.	torsion	405	402	403	401	0.1800 0.0 1	0.1700 180.0 2
	0.5200 0.0 3						
541.	torsion	405	402	403	418	-	
	1.1500 0.0 1	0.0000 180.0 2			1.2800 0.0 3		
542.	torsion	405	402	403	425	0.0000 0.0 1	0.0000 180.0 2
	0.3400 0.0 3						
543.	torsion	419	402	403	401	-	
	1.1500 0.0 1	0.0000 180.0 2			1.2800 0.0 3		
544.	torsion	419	402	403	418	2.2200 0.0 1	-1.3800 180.0 2
	1.1800 0.0 3						-
545.	torsion	419	402	403	425	0.0000 0.0 1	0.0000 180.0 2
	0.3000 0.0 3						
546.	torsion	428	402	403	401	0.0000 0.0 1	0.0000 180.0 2
	0.2800 0.0 3						
547.	torsion	428	402	403	418	0.0000 0.0 1	0.0000 180.0 2
	0.3000 0.0 3						
548.	torsion	428	402	403	425	0.0000 0.0 1	0.0000 180.0 2
	0.2380 0.0 3						

549.	torsion 0.0000 0.0 3	403	402	405	421	0.0000 0.0 1	0.0000 180.0 2
550.	torsion 0.0000 0.0 3	403	402	405	429	0.0000 0.0 1	0.0000 180.0 2
551.	torsion 0.0000 0.0 3	419	402	405	421	0.0000 0.0 1	0.0000 180.0 2
552.	torsion 0.0000 0.0 3	419	402	405	429	0.0000 0.0 1	0.0000 180.0 2
553.	torsion 0.0000 0.0 3	428	402	405	421	0.0000 0.0 1	0.0000 180.0 2
554.	torsion 0.0000 0.0 3	428	402	405	429	0.0000 0.0 1	0.0000 180.0 2
555.	torsion 1.5000 180.0 2	403	402	419	404	2.0000 0.0 1	-
556.	torsion 1.5000 180.0 2	405	402	419	404	2.0000 0.0 1	-
557.	torsion 0.5970 0.0 3	428	402	419	404	0.0000 0.0 1	0.0000 180.0 2
558.	torsion 1.3720 0.0 1	401	403	418	439	-	
						0.2320 180.0 2	0.4000 0.0 3
559.	torsion 1.3720 0.0 1	402	403	418	439	-	
						0.2320 180.0 2	0.4000 0.0 3
560.	torsion 0.2660 0.0 3	425	403	418	439	0.0000 0.0 1	0.0000 180.0 2
561.	torsion 0.0000 0.0 3	401	404	409	411	0.0000 0.0 1	0.0000 180.0 2
562.	torsion 0.0000 0.0 3	401	404	409	412	0.0000 0.0 1	0.0000 180.0 2
563.	torsion 0.0000 0.0 3	419	404	409	411	0.0000 0.0 1	0.0000 180.0 2
564.	torsion 0.0000 0.0 3	419	404	409	412	0.0000 0.0 1	0.0000 180.0 2
565.	torsion 0.0000 0.0 3	426	404	409	411	0.0000 0.0 1	0.0000 180.0 2
566.	torsion 0.0000 0.0 3	426	404	409	412	0.0000 0.0 1	0.0000 180.0 2
567.	torsion 1.5000 180.0 2	401	404	419	402	2.0000 0.0 1	-
						0.8900 0.0 3	
568.	torsion 1.5000 180.0 2	409	404	419	402	2.0000 0.0 1	-
						0.8900 0.0 3	
569.	torsion 0.5970 0.0 3	426	404	419	402	0.0000 0.0 1	0.0000 180.0 2
570.	torsion 0.0000 0.0 3	402	405	421	408	0.0000 0.0 1	0.0000 180.0 2
571.	torsion 0.0000 0.0 3	429	405	421	408	0.0000 0.0 1	0.0000 180.0 2
572.	torsion 0.0000 0.0 3	422	406	410	407	0.0000 0.0 1	0.0000 180.0 2
573.	torsion 0.0000 0.0 3	422	406	410	433	0.0000 0.0 1	0.0000 180.0 2
574.	torsion 0.0000 0.0 3	430	406	410	407	0.0000 0.0 1	0.0000 180.0 2
575.	torsion 0.0000 0.0 3	430	406	410	433	0.0000 0.0 1	0.0000 180.0 2
576.	torsion 0.0000 0.0 3	410	406	422	408	0.0000 0.0 1	0.0000 180.0 2

577.	torsion 0.0000 0.0 3	430	406	422	408	0.0000 0.0 1	0.0000 180.0 2
578.	torsion 0.0000 0.0 3	431	407	410	406	0.0000 0.0 1	0.0000 180.0 2
579.	torsion 0.0000 0.0 3	431	407	410	433	0.0000 0.0 1	0.0000 180.0 2
580.	torsion 0.0000 0.0 3	422	408	421	405	0.0000 0.0 1	0.0000 180.0 2
581.	torsion 0.0000 0.0 3	432	408	421	405	0.0000 0.0 1	0.0000 180.0 2
582.	torsion 0.0000 0.0 3	421	408	422	406	0.0000 0.0 1	0.0000 180.0 2
583.	torsion 0.0000 0.0 3	432	408	422	406	0.0000 0.0 1	0.0000 180.0 2
584.	torsion 0.0000 0.0 3	404	409	411	413	0.0000 0.0 1	14.0000 180.0 2
585.	torsion 0.0000 0.0 3	404	409	411	423	0.0000 0.0 1	14.0000 180.0 2
586.	torsion 0.0000 0.0 3	412	409	411	413	0.0000 0.0 1	14.0000 180.0 2
587.	torsion 0.0000 0.0 3	412	409	411	423	0.0000 0.0 1	15.0000 180.0 2
588.	torsion 0.0000 0.0 3	404	409	412	424	0.0000 0.0 1	14.0000 180.0 2
589.	torsion 0.0000 0.0 3	404	409	412	434	0.0000 0.0 1	14.0000 180.0 2
590.	torsion 0.0000 0.0 3	411	409	412	424	0.0000 0.0 1	15.0000 180.0 2
591.	torsion 0.0000 0.0 3	411	409	412	434	0.0000 0.0 1	7.0000 180.0 2
592.	torsion 0.0000 0.0 3	409	411	413	414	0.0000 0.0 1	5.4700 180.0 2
593.	torsion 0.0000 0.0 3	409	411	413	424	0.9000 0.0 1	15.0000 180.0 2
594.	torsion 0.0000 0.0 3	423	411	413	414	0.0000 0.0 1	5.4700 180.0 2
595.	torsion 0.0000 0.0 3	423	411	413	424	0.9000 0.0 1	15.0000 180.0 2
596.	torsion 0.0000 0.0 3	409	411	423	415	0.0000 0.0 1	15.0000 180.0 2
597.	torsion 0.0000 0.0 3	413	411	423	415	0.0000 0.0 1	14.0000 180.0 2
598.	torsion 0.0000 0.0 3	409	412	424	413	0.0000 0.0 1	15.0000 180.0 2
599.	torsion 0.0000 0.0 3	434	412	424	413	0.0000 0.0 1	7.0000 180.0 2
600.	torsion 0.0000 0.0 3	411	413	414	416	0.0000 0.0 1	5.4700 180.0 2
601.	torsion 0.0000 0.0 3	411	413	414	437	0.0000 0.0 1	4.4700 180.0 2
602.	torsion 0.0000 0.0 3	424	413	414	416	0.9000 0.0 1	15.0000 180.0 2
603.	torsion 0.0000 0.0 3	424	413	414	437	0.0000 0.0 1	2.5000 180.0 2
604.	torsion 0.0000 0.0 3	411	413	424	412	0.0000 0.0 1	14.0000 180.0 2

605.	torsion	414	413	424	412	0.0000	0.0	1	14.0000	180.0	2
	0.0000	0.0	3								
606.	torsion	413	414	416	415	0.0000	0.0	1	14.0000	180.0	2
	0.0000	0.0	3								
607.	torsion	413	414	416	435	-					
	3.1500	0.0	1	8.0000	180.0	2	0.0000	0.0	3		
608.	torsion	437	414	416	415	0.0000	0.0	1	14.0000	180.0	2
	0.0000	0.0	3								
609.	torsion	437	414	416	435	0.0000	0.0	1	14.0000	180.0	2
	0.0000	0.0	3								
610.	torsion	417	415	416	414	0.0000	0.0	1	14.0000	180.0	2
	0.0000	0.0	3								
611.	torsion	417	415	416	435	0.0000	0.0	1	14.0000	180.0	2
	0.0000	0.0	3								
612.	torsion	423	415	416	414	0.0000	0.0	1	15.0000	180.0	2
	0.0000	0.0	3								
613.	torsion	423	415	416	435	-					
	2.7440	0.0	1	15.0000	180.0	2	0.0000	0.0	3		
614.	torsion	416	415	417	436	0.5200	0.0	1	-		
	0.5100	180.0	2	0.5200	0.0	3					
615.	torsion	423	415	417	436	0.5200	0.0	1	-		
	0.5100	180.0	2	0.5200	0.0	3					
616.	torsion	416	415	423	411	0.0000	0.0	1	15.0000	180.0	2
	0.0000	0.0	3								
617.	torsion	417	415	423	411	0.0000	0.0	1	14.0000	180.0	2
	0.0000	0.0	3								

Code and Parameters 2-5: GDPNP AMOEBA BIO09 Parameters

## 2.8 SMALL MOLECULE SIMULATIONS IN AMOEBA

Simulations performed on ACE-CNC-NME, MeSCN, EtSCN, and HxSCN were all performed in the same manner. Unless otherwise noted, we used: a vdw-cutoff of 12.0 Å, Ewald summation for polarizable multipole interactions, PME for charge-charge interactions with an Ewald-cutoff of 7.0 Å, and the polar-predict directive.

First the molecule was solvated in a  $50^3$  Å<sup>3</sup> periodic, truncated octahedron box of energy-minimized water molecules. The solute atoms were restrained and the solute-solvent system was quickly minimized using the steepest-descent algorithm with an RMS gradient criteria of 1 Å<sup>2</sup>. A second minimization was then performed, still using position restraints on solute atoms, using the L-BFGS quasi-newton nonlinear optimization routine with a gradient criteria of 0.05 Å<sup>2</sup>. The minimization was performed in two stages because there were some instances where the forces became too large to continue with the L-BFGS routine, but once a relatively-minimized structure was obtained using steepest-descent, the L-BFGS routine could continue to completion. Both minimizations used a polar-eps value of 0.0001 D. Next the system was equilibrated for 200 ps under NVT at 298 K with the RESPA integrator,<sup>61</sup> using the Bussi thermostat<sup>62</sup> with a temperature-coupling constant of 0.2 ps and a polar-eps value of 0.0001 D. The system was then equilibrated for 500 ps under NPT using the Berendsen barostat<sup>63</sup> with a pressure-coupling constant of 2.0 ps and all the same parameters are the NVT equilibration. Both equilibrations maintained position restraints on solute atoms and used a 2.5 fs timestep.

Following equilibration, MD was performed at NVT at 298 K using the RESPA integrator and Bussi thermostat with a polar-eps of 0.001 D and 2.0 fs timestep. Snapshots were collected every 0.1 ps, for a total of 10000 frames/ns. To date, a MeSCN and EtSCN have been simulated for 4 ns, EtSCN has been simulated for 5 ns, and ACE-CNC-NME has been simulated for 7 ns.



## Chapter 3 Electrostatic Methods and Theory

### 3.1 INTRODUCTION

As previously described, a plethora of techniques exist for quantifying electrostatic fields. Due to the large library of experimental data our group possesses, we are ideally situated to directly compare experimental measurements of electrostatic field to different theoretical models that attempt to describe those same electrostatic fields. In the following chapter, the different methods for calculating electrostatic fields are presented, as well as the physical relationship between experimentally measured vibrational absorption energies and the local field via the vibrational Stark effect.

### 3.2 AMBER03 WITH EXPLICIT TIP3P WATER

#### 3.2.1 GROMACS Reaction Field Electrostatics

The explicit TIP3P<sup>45</sup> solvent method we used to calculate electrostatic fields uses keyword “coulombtype = reaction-field” in GROMACS,<sup>50,64-66</sup> allowing the system to be treated with periodic boundary conditions. Forces are calculated using equation (3-1), where  $q$  is the charge on an atom,  $\epsilon_{\text{solute}}$  is the dielectric of the short-range interactions and was set to 1,  $\epsilon_{\text{solvent}}$  is the dielectric of the long-range interactions and was set to 78,  $\vec{r}_c$  is the scaling distance where the potential decays to zero and was set to 1.0 nm, and  $\vec{r}_{ij}$  is the distance between two charges. The reaction field interaction is a modification of Coulomb’s law which switches from one dielectric value at short-ranged to a second (typically larger) dielectric value at long ranges.

$$\vec{\text{Force}}_{\text{RF}, i}(\vec{r}_{ij}) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{\epsilon_{\text{solute}}} \left[ \frac{1}{r_{ij}^2} - \frac{2r_{ij}}{r_c^3} \frac{(\epsilon_{\text{solvent}} - \epsilon_{\text{solute}})}{(2\epsilon_{\text{solvent}} + \epsilon_{\text{solute}})} \right] \frac{\vec{r}_{ij}}{r_{ij}} \quad (3-1)$$

To calculate the field at the nitrile, we first placed a test charge ( $q = +1$ ) at the nitrile midpoint for every frame in our MD trajectories. We then used the “rerun” option of the GROMACS molecular dynamics engine, mdrun, to calculate the forces on every atom, including the test charge. Dividing the force on the test charge by its charge (+1) gives the field at the test charge, which was then projected along the nitrile CN bond vector.

### 3.2.2 Hybrid Solvent Reaction Field Electrostatics and Solute Coulomb Field

Because the reaction field coulomb interaction is additive, we are able to break the electrostatic field down into contributions from specific atoms or species.<sup>43</sup> By setting the charge on all solute atoms to zero, we are able to solve for the solvent reaction field of the system. We then calculated the analytical Coulomb field (equation (3-2)) as the bond mid-point, looking only at solute atoms, and summed it with the reaction field electrostatics solvent reaction field to obtain a total field. In this way, water near the probe is treated as structural water with an un-screened dielectric of 1, while water significantly further away from the probe is treated as bulk water with a damped dielectric of 78. In contrast, all solute atoms are treated with an un-screened dielectric of 1, rather than atoms sufficiently far from the probe being screened by more than simple distance.

$$\vec{F}_{\text{Coulomb}}(\vec{r}) = \frac{1}{4\pi\epsilon_0} \frac{1}{\epsilon_{\text{solute}}} \sum_i \frac{q_i \vec{r}}{\vec{r}^3} \quad (3-2)$$

The code for inserting the test charge, gmx\_insert\_dummy\_atom into a GROMACS trajectory file can be accessed at: [https://github.com/awritchie/my\\_gmx](https://github.com/awritchie/my_gmx)

The code for calculating the solute electrostatic field at a point, gmx2pqr, can be accessed at: [https://github.com/awritchie/my\\_gmx](https://github.com/awritchie/my_gmx)

### 3.3 AMBER03 WITH POISSON-BOLTZMANN CONTINUUM SOLVENT

The non-linear Poisson-Boltzmann equation (NLPBE) is given in equation (3-3), where  $\phi(\bar{r})$  is the potential as a function of the position vector  $\bar{r}$ ,  $\varepsilon(\bar{r})$  is the spatially dependent protein interior dielectric,  $\bar{\kappa}^2$  is the ion accessibility coefficient, and  $\rho(\bar{r})$  is the charge distribution density.<sup>67</sup>

$$\nabla \cdot \varepsilon(\bar{r}) \nabla \phi(\bar{r}) = \varepsilon(\bar{r}) \bar{\kappa}^2 \sinh \phi(\bar{r}) - 4\pi \rho(\bar{r}) \quad (3-3)$$

Assuming a small potential,  $\phi \ll k_b T/q$ , the approximation  $\sinh(\phi) \approx \phi$  can be used to formulate the linear Poisson-Boltzmann equation (LPBE), equation (3-4), although it has previously been reported that the solutions to the LPBE and the NLPBE differ only by a few  $k_b T/q$  even when the linearity condition does not hold:<sup>68, 69</sup>

$$\nabla \cdot \varepsilon(\bar{r}) \nabla \phi(\bar{r}) = \varepsilon(\bar{r}) \bar{\kappa}^2 \phi(\bar{r}) - 4\pi \rho(\bar{r}) \quad (3-4)$$

Solutions to the LPBE were calculated using APBS's multigrid solver.<sup>70, 71</sup> For both one-dimensional and two-dimensional umbrella-sampling strategies, each frame was examined using APBS, and a Boltzmann-weighted average field for each system was calculated based on the dihedral probability distributions previously described. A two-stage manual focusing strategy (APBS keyword "mg-manual") was used, in which the system was centered in a 240 Å x 240 Å x 240 Å box with 97 grid points in each dimension and a single Debye-Huckle boundary condition at 300 K. A cubic B-spline was used to map the discretized atomic point charges to the grid. Ion concentrations of 0.150 M +1 charge ions of radius 2.0 Å and 0.150 M -1 charge ions of radius 2.0 Å were included to account for buffer in solution. We used the molecular surface model ("srfm mol") with a surface sphere density of 10.0 ("sdens 10.0"). The solvent dielectric was 78.0, appropriate for bulk water. The choice of protein dielectric is significantly more ambiguous, with values typically ranging from 2 to 40.<sup>72</sup> When all interactions are treated

explicitly (permanent dipole moments, induced dipole moments (polarizability), and explicit solvent), a protein dielectric of one is most appropriate.<sup>73,74</sup> Since we have directly modeled structural fluctuations but not polarizability, we chose an interior dielectric of 2.0. Because the calculated field scales linearly with the interior dielectric chosen, and thus the choice of interior dielectric to be used is by no means well defined,<sup>75</sup> we are less concerned with the value of the protein dielectric, since an “ideal” dielectric value can be extrapolated from the slope of the fit to experimental data (*vida infra*).

The field along the nitrile bond length was calculated by examining the potential gradient along the bond vector. This was done by inserting 11 dummy atoms with charge, mass, and radius of zero, evenly spaced along the nitrile bond length. The potential at each of these dummy atoms was interpolated from the nearest grid points using a linear spline with the “write atompot flat” keyword in APBS. It was observed that the potential changed linearly within ~0.3 Å of the bond midpoint, so the central five dummy atoms were used to determine a first order polynomial of potential *versus* position, and the electrostatic field at the bond midpoint was taken as the negative of the slope.

### 3.3.1 Reaction Field Method

To minimize errors due to using numerical solutions to the LPBE, we have independently solved for both the numeric solvent reaction field potential (SRF), described in equation (3-5) where each  $\phi(\vec{r})$  is a solution to the LPBE (equation (3-4)) at different solvent dielectric, as well as the analytic coulomb field of the solute, then summing the resulting fields to obtain the total electrostatic field of the system. We refer to this method as the reaction field method (RFM), shown in equation (3-7), where the first term is the negative gradient of the solvent reaction field potential at a point in space

$\vec{r}$  and is given in equation (3-6) and the second term is simply the analytical expression for the Coulomb field at the same point in space, shown in equation (3-2).<sup>26,33</sup>

$$SRF = \phi(\vec{r})_{(\epsilon_{\text{solute}} \neq \epsilon_{\text{solvent}})} - \phi(\vec{r})_{(\epsilon_{\text{solute}} = \epsilon_{\text{solvent}})} \quad (3-5)$$

$$\vec{F}_{\text{SRF}}(\vec{r}) = -\nabla \left( \phi(\vec{r})_{(\epsilon_{\text{solute}} \neq \epsilon_{\text{solvent}})} - \phi(\vec{r})_{(\epsilon_{\text{solute}} = \epsilon_{\text{solvent}})} \right) \quad (3-6)$$

$$\text{Reaction Field Method (RFM)} : \vec{F}_{\text{RFM}}(\vec{r}) = \vec{F}_{\text{SRF}}(\vec{r}) + \vec{F}_{\text{Coulomb}}(\vec{r}) \quad (3-7)$$

The RFM requires two numeric PBE solutions and one analytic Coulomb solution for each system and is described schematically in Figure 3-1. The SRF is calculated numerically from the difference between the PBE solution for the solvated potential and the PBE solution for the reference potential, illustrated in Figure 3-1A. The solvated potential,  $\phi(r)_{\epsilon_{\text{solute}} \neq \epsilon_{\text{solvent}}}$ , represented by the leftmost image in Figure 3-1A, is PBE solution for a solute with dielectric  $\epsilon_{\text{solute}}$  solvated in a continuum of dielectric  $\epsilon_{\text{solvent}}$ . The reference electrostatic potential,  $\phi(r)_{\epsilon_{\text{solute}} = \epsilon_{\text{solvent}}}$ , represented by the middle image in Figure 3-1A, is the PBE solution for a solute with dielectric  $\epsilon_{\text{solute}}$  solvated in a continuum also having a dielectric  $\epsilon_{\text{solute}}$ . Removing the reference potential from the solvated potential results in only the potential generated by the solvent polarizing about the solute, represented by the rightmost image in Figure 3-1A. Figure 3-1B shows the SRF (leftmost image) result being added to the analytical solution to the Coulomb field for all protein atoms in a dielectric  $\epsilon_{\text{solute}}$ , (middle image), resulting in a total field which uses a numeric solution for the contribution of the implicit solvent and an analytic solution for the contribution of the explicit solute. In this way, numerical errors in the final result are reduced by only solving for part of the system (the solvent) numerically. This also decreases the calculation sensitivity to input parameters such as box size, grid spacing, box location, charge mapping method, etc. Since our initial work, APBS now has the “lrbpe” keyword, which solves for the SRF directly without requiring a reference calculation.

### 3.3.2 Grid spacing and size

Six sets of box size and grid dimensions were used. The 30 Å box and 60 Å box were centered on system center of mass. The other four boxes were each centered on NC $\delta$ , system center of mass, and RalGDS center of mass, with four different sets of grid spacings (GS), for a total of 14 calculations per frame (excluding the first-stage box calculations): two  $19^3$  Å<sup>3</sup> boxes with 161 and 193 grid points in each dimension with a GS of 0.119 Å/grid point and 0.099 Å/grid point respectively; two  $10^3$  Å<sup>3</sup> boxes with 161 and 193 grid points in each dimension with a GS of 0.062 Å/grid point and 0.052 Å/grid point respectively; a 30 Å x 45 Å x 40 Å box with 129 x 193 x 161 grid points in each dimension with a GS of (0.234, 0.234, 0.250) Å/grid point; and a 60 Å x 45 Å x 40 Å box with 257 x 193 x 161 grid points in each dimension with a GS of (0.234, 0.234, 0.250) Å/grid point. Box sizes will be referred to as the 10 Å box, 19 Å box, 30 Å box, and 60 Å box, respectively, while grid dimensions will be referred to as 161 grid points, 193 grid points, 129 grid points, and 257 grid points, respectively.

The reaction field method was used for the 10 Å box and 19 Å box using 193 grid points for each as well as for the 60 Å box. The system parameters are identical to those previously described for the first term in equation (3-5). For the second term in equation (3-5), the solvent and interior dielectric were both set to 2.0. The analytic Coulomb protein field was taken using the typical Coulomb field expression with a dielectric value of 2.0 and projecting the electrostatic field vector along the nitrile bond vector. This was done for the NC $\delta$  box location, system center of mass box location, and RalGDS box location. In total, 7 additional calculations were carried out per frame (excluding the first-stage box calculations).

### 3.3.3 Box Location

The first-stage box provided the boundary conditions for the second-stage box, and various strategies for constructing second-stage boxes were examined. Different combinations of box center locations, box dimensions, and the number of grid points were examined, as described in the Introduction. By translating the box away from the solvent and towards the protein, we hoped to reduce how often the discontinuity in dielectric environment occurs in the second-stage calculation. The remaining parameters used were identical to those of the first-stage box.

Three different box centers were used: the C $\delta$  atom of the cyanocysteine side chain (i.e nitrile carbon), the center of mass of the entire system (excluding water), and the center of mass of the RalGDS protein containing the thiocyanate probe (excluding water). These boxes will be referred to as NC $\delta$ , system center of mass, and RalGDS center of mass, respectively. Unless otherwise specified, the NC $\delta$  box was used. Representative snapshots of each box are shown in Figure 3-2 for the nitrile probe positioned at RalGDS N27C<sub>SCN</sub> (black: 60 Å system center of mass; red: 10 Å NC $\delta$ ; green: 10 Å system center of mass; and blue: 10 Å RalGDS center of mass). The position of the nitrile C $\delta$  was selected as the center of the NC $\delta$  box. The center of the system center of mass box was selected as a point along the vector between the furthest of the nitrile C $\delta$  and N $\epsilon$  and the center of mass of the Rap/RalGDS system. If the distance to the center of mass of the system and the furthest of the nitrile atoms was less than 45% of the smallest box dimension (90% of half the smallest dimension), then system center of mass was the center of mass of Rap/RalGDS system. Otherwise, the box center was translated 45% of the smallest box dimension toward the center of mass of the Rap/RalGDS system. This treatment ensures that the entirety of the C $\delta$ -N $\epsilon$  bond vector is always within the box. RalGDS center of mass was treated in the same manner, except

it used only the center of mass of the RalGDS part of the protein. In the case of calculations conducted on SCN-containing RalGDS monomer structures, RalGDS center of mass and system center of mass were identical.

The code for calculating generating .pqr files for APBS, gmx2pqr, can be accessed at: [https://github.com/awritchie/my\\_gmx](https://github.com/awritchie/my_gmx)

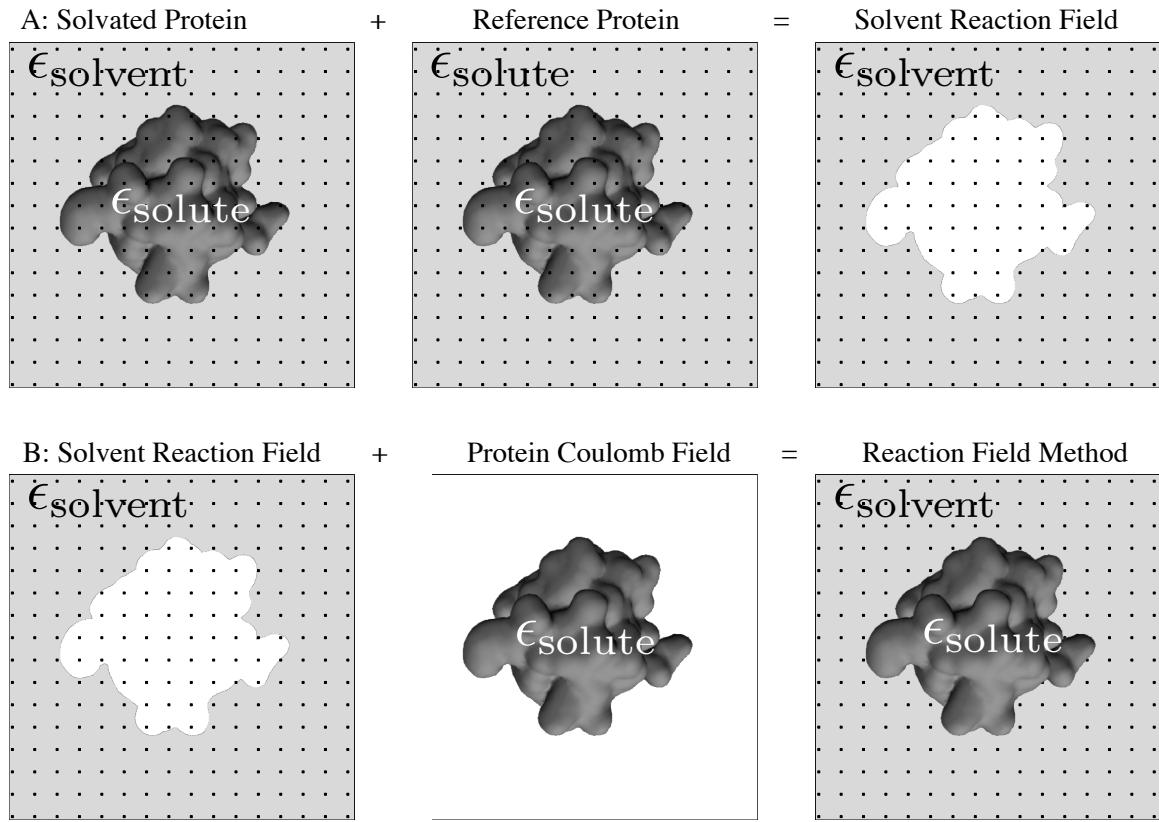


Figure 3-1: Reaction Field Method Schematic

A schematic diagram of the reaction field method (RFM). A) Method for solving the solvent reaction field potential numerically using the Poisson-Boltzmann equation. Top left: The solvated electrostatic potential. Top middle: The reference electrostatic potential. Top right: The reference electrostatic potential is removed from the solvated electrostatic potential, leaving only the electrostatic potential due to solvent—the solvent reaction field potential. B) Method for combining the numerical solution to the solvent reaction field potential to the analytic coulomb field. Bottom left: The numeric solvent reaction field potential. Bottom middle: The analytic field obtained from Coulomb's Law. Bottom right: The total electrostatic field is the sum of the numeric solvent reaction field and the analytic coulomb field.

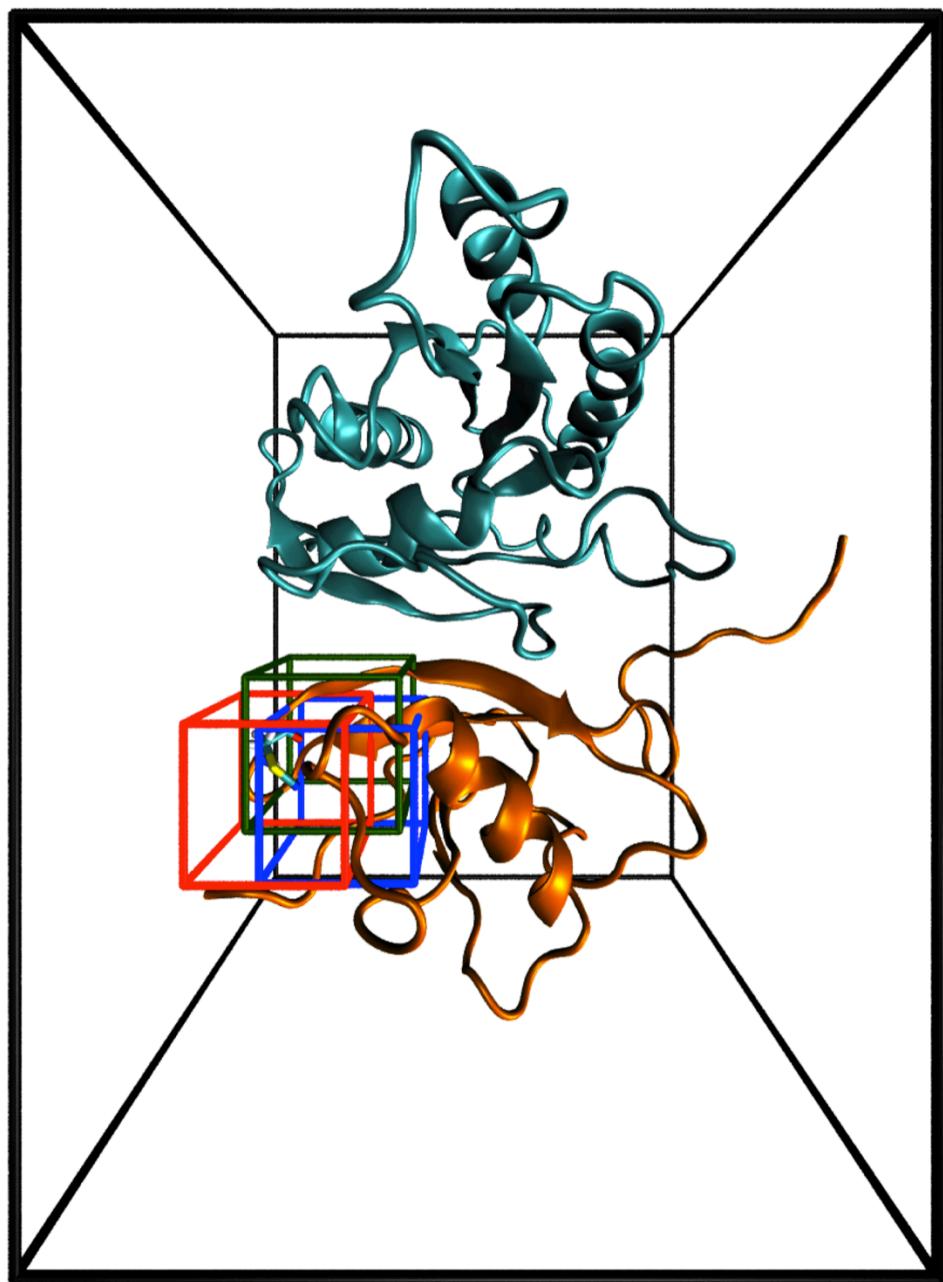


Figure 3-2: Representation of Second-Stage Box Positions

A representative snapshot of the Rap E30/K31-RalGDS N27C<sub>SCN</sub> protein complex demonstrating the position and sizes of the boxes used to calculate electrostatic fields in APBS. Black: 60 Å system center of mass; red: 10 Å NC $\delta$ ; green: 10 Å system center of mass; blue: 10 Å RalGDS center of mass.

### **3.4 AMBER03 WITH POISSON-BOLTZMANN CONTINUUM SOLVENT AND SELECT EXPLICIT TIP3P WATER MOLECULES**

For the following, all electrostatics were done using the reaction field method and the numeric solution to the LPBE. In all cases, water molecules selected are explicitly treated as part of the solute calculation. Here we are merely addressing the water selection criteria.

We examined three different methods for selecting individual water molecules to be included in the explicit portion of the calculation: 1) all water within 5 Å of the nitrile N $\epsilon$  atom, 2) the water molecule closest to the nitrile N $\epsilon$  atom, and 3) any water molecules hydrogen-bonding to the nitrile N $\epsilon$  atom. Frames were prepared using the GROMACS g\_select utility in conjunction with trjconv to remove all solvent that was more than 15 Å away from the nitrile N $\epsilon$ . Each frame was further processed using Python scripts to generate a .pqr file containing explicit water molecules selected through each of the three methods.

#### **3.4.1 5 Å Water Sphere Around the Vibrational Chromophore**

Figure 3-3 demonstrates a 5 Å solvent sphere for a snapshot of RalG28C<sub>SCN</sub> docked to Rap E30/K31. We chose a 5 Å radius sphere because it is large enough that it will include any waters significantly interacting with the nitrile. The sphere was centered on the N $\epsilon$  rather than the C $\delta$  because the nitrogen can act as a hydrogen-bond acceptor and, should this be occurring in the simulation, we want to be sure to capture those water molecules. Each 5 Å sphere frame could contain zero (~1.4% of frames), one (~2.7% of frames), or more (~95.9% of frames) water molecules in the frame.

### **3.4.2 Single Water Molecule Nearest the Vibrational Chromophore**

The water molecule with the closest atom (oxygen or either hydrogen) was considered the nearest water molecule, where or not that water molecule met hydrogen bonding criteria. Each nearest water frame was guaranteed to have exactly one water molecule.

### **3.4.3 Water Molecular Hydrogen Bonding to the Vibrational Chromophore**

We considered a water molecule to be hydrogen-bonded to the nitrile if the  $N\epsilon$ --H separation was  $\leq 2.25 \text{ \AA}$  and the  $N\epsilon$ --H-O smallest angle was  $\geq 138^\circ$ , which were chosen to reach the extremes of the nitrile-water parameters reported by Le Questel *et al.*<sup>76</sup> For both the 5 Å sphere and hydrogen-bonding methods, if a frame did not have a water molecule fitting the respective criteria, no explicit water molecules were included and the LPBE solutions from the original, entirely implicit calculations, were used for that frame. Each hydrogen-bonding water frame could have exactly zero (~89.3% of frames) or exactly one (~10.7% of frames) water molecule in each frame.

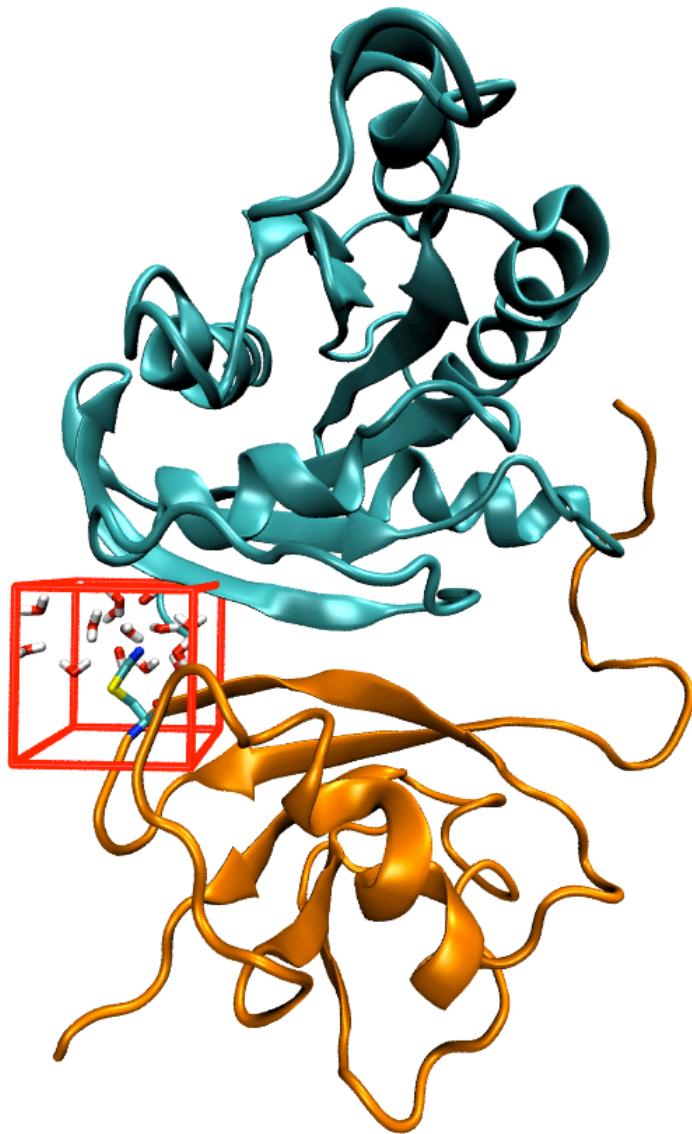


Figure 3-3: Representative Snapshot Showing 5 Å Water Sphere

A representative snapshot of RalGDS G28C<sub>SCN</sub> (orange) docked to Rap E30/K31 (cyan) demonstrating the size of the 10<sup>3</sup> Å<sup>3</sup> second-stage box (red) used to calculate the electrostatic field in APBS and showing water molecules within 5 Å of the nitrile Nε.

## 3.5 AMOEBA

### 3.5.1 Calculating Electrostatic Fields in AMOEBA

We have calculated the electrostatic field experienced by the vibrational chromophore using two methods. The first method is to consider the field at the bond midpoint as the average of the fields at the nitrile carbon and the nitrile nitrogen. This method is relatively simple and trivial to implement with Tinker. The induced dipole of an atom or molecule,  $\vec{\mu}_{\text{induced}}$ , is given by equation (3-8), where  $\alpha$  is the isotropic polarizability and  $\vec{F}$  is the local electrostatic field. This can be trivially rearranged to solve for the electrostatic field. The Tinker “analyze” program was modified to report the induced dipole moment of every atom (or in the case of performing the molecular dynamics in Tinker with the AMOEBA force field, the “save-induced” keyword was used), and the field was calculated using equation (3-9). Field calculated in this way will hereafter be referred to as the induced method (IM).

$$\vec{\mu}_{\text{induced}} = \alpha \vec{F} \quad (3-8)$$

$$\vec{F}_{\mu_{\text{ind}}} = \frac{1}{2} \left( \frac{\mu_{C, \text{induced}}}{\alpha_C} + \frac{\mu_{N, \text{induced}}}{\alpha_N} \right) \quad (3-9)$$

The other method is to calculate the electrostatic field exactly at the bond midpoint. While this requires some code modifications, it has the advantage of being trivial to examine or remove the electrostatic field contributions of specific atom or atom fragments in the system without removing information about the system’s response (via induced dipole moments) to the removed parts. This is useful, for example, for removing the contributions of the SCN atoms to the electrostatic field since they should, on average, always have the same geometry relative to each other and therefore always have the same monopole, permanent dipole, and quadrupole field values. In other words, we

do not care about the field created by the SCN atoms, but rather the field external to the SCN, which is created by the atoms surrounding the SCN. Likewise, we do not want to simply zero the multipoles on the SCN atoms because the solvent and non-probe solute atoms *do* respond to the presence of the SCN, and that is something that should be detectable via VSE.

$$\begin{aligned}
 \vec{f}_{i, \text{monopole}}(\vec{r}_i) &= \frac{q_i}{r_i^3} \vec{r}_i \\
 \vec{f}_{i, \text{dipole}}(\vec{r}_i) &= \frac{3(\vec{p} \cdot \vec{r}_i)}{r_i^5} \vec{r}_i - \frac{\vec{p}}{r_i^3} \\
 \vec{f}_{i, \text{quadrupole}}(\vec{r}_i) &= \frac{15}{2r_i^7} \left[ \sum_{j=(x,y,z)} \sum_{k=(x,y,z)} \bar{r}_{ij} \bar{r}_{ik} Q_{ijk} \right] \vec{r}_i - \frac{6}{r_i^5} \sum_{j=(x,y,z)} Q_{i\hat{j}j} \bar{r}_{ij} \\
 \vec{F}_{\text{mdq}}(\vec{r}) &= \frac{1}{4\pi\epsilon_0} \sum_i^{N_{\text{atoms}}} \vec{f}_{i, \text{monopole}}(\vec{r}_i) + \vec{f}_{i, \text{dipole}}(\vec{r}_i) + \vec{f}_{i, \text{quadrupole}}(\vec{r}_i)
 \end{aligned} \tag{3-10}$$

Equation (3-10) is the multipole expansion for an electrostatic field, where  $\vec{r}$  is some position,  $\epsilon_0$  is the permittivity of free space,  $q$  is the monopole charge of an atom,  $\vec{p}$  is the total dipole vector of an atom and is the sum of  $\vec{\mu}_{\text{permanent}}$  and  $\vec{\mu}_{\text{induced}}$ , and  $Q_{jk}$  is the traceless tensor describing the quadrupole of an atom. For every frame,  $\vec{F}_{\text{mdq}}(\vec{r})$  was solved at  $\vec{r}$  equals the nitrile bond midpoint. This is solved *after* the induced dipoles are solved to self consistency, meaning the contributions of all atom in the structure file are allowed to mutually polarize each other and the removal of any single atom does not induce a many-body effect on the resultant field calculation. This method will hereafter be referred to as the midpoint method (MPM).

### 3.5.2 Poisson-Boltzmann Continuum Solvent

All non-solvent atoms were kept and a Tinker .xyz file was generated for each frame using custom code written in C++ and linked against the GROMACS 4.6.x libraries. The “PB-HPMF” keyword was used, which uses APBS to iteratively solve for

the induced dipole to self consistency.<sup>73</sup> Induced dipoles were solved to self-consistency of 0.01 D, which typically required 4 iterations of the SRF (or 8 PB solutions) for the protein systems. As before, we used ion concentrations of 0.150 M +1 charge ions of radius 2.0 Å and 0.150 M -1 charge ions of radius 2.0, the molecular surface model (“srfm mol”) with a surface sphere density of 10.0 (“sdens 10.0”) and a solvent dielectric of 78.0. A solute dielectric of 1.0 was used since polarizability is directly modeled in the force field. Once the induced dipoles were solved to self-consistency, equation (3-10) was used to find the field at the nitrile bond midpoint. Because focusing was not implemented in APBS-linked Tinker, a multiple Debye-Hückel boundary condition was used and the box size and grid dimensions were those automatically chosen by Tinker (guaranteed to fully encompass the solute with a grid spacing of 0.5 Å between grid points).

### 3.5.3 Explicit AMOEBA Water

All atoms were kept and a Tinker .xyz file was generated for each frame using custom code written in C++ and linked against the GROMACS 4.6.x libraries. Aside from TIP3P water (having AMOEBA water parameters) and therefore excluding “PB-HPMF” and its associated keywords, these calculations were performed identically to the AMOEBA with PB calculations with one special consideration. Because a dodecahedron box was used for the simulations in Amber03 and Tinker does not (currently) support a dodecahedron periodic box, periodicity was included in neither the IM nor the MPM field calculations. In simulations performed in Tinker with AMOEBA, a truncated octahedron box was used and periodicity is included for the self-consistent field calculations as well as the IM, although for the final calculation of  $\vec{F}(\vec{r})$  using the MPM, only the minimum-distance images were used. In either case, because electrostatic fields go to zero quickly

as a function of distance and the box sizes were sufficiently large, the differences were negligible.

### 3.5.4 Charge Penetration Field Corrections

Electrostatic interactions have focused heavily on long-range interactions. However, at *very* short distances, spherical approximations of atomic charge distributions become non-Coulombic due to overlapping atomic orbitals. This phenomenon is known as charge penetration and the electrostatic potential is no longer described as a  $1/r$  function. This charge penetration typically leads to more negative electrostatic interaction energy as compared to a Coulombic point charge representation. Typical charge penetration corrections involve breaking the atom-centered point charge into an effective core charge and a valence electron density.<sup>77,78</sup> The method implemented in AMOEBA by Wang *et al.* is presented in equations (3-11) and (3-12), which breaks down each point charge into a positive effective core  $Z$  (set to the number of valence electrons) and a damped electron distribution, controlled by the  $\alpha$  when interacting with the core and  $\beta$  when interacting with the electron of another atom, where  $q$  is the net partial charge of the atom and  $r$  is the distance between atoms.<sup>79</sup> Electrostatic interaction potential energies computed with and without this charge penetration correction were compared to those calculated using symmetry-adapted perturbation theory (SAPT),<sup>80</sup> and it was found that the charge penetration correction significantly improved agreement between the quantum and classical model, while the non-corrected energies were consistently over-approximated.

$$E_{\text{cp}}(r_{ij}) = \begin{bmatrix} Z_i Z_j \\ -Z_i (Z_j - q_j) (1 - \exp(-\alpha_j r_{ij})) \\ -Z_j (Z_i - q_i) (1 - \exp(-\alpha_i r_{ij})) \\ +(Z_i - q_i) (Z_j - q_j) (1 - \exp(-\beta_i r_{ij})) (1 - \exp(-\beta_j r_{ij})) \end{bmatrix} \quad (3-11)$$

$$V_{\text{cp}}(\vec{r}) = \sum_i^{N \text{ atoms}} [Z_i - (Z_i - q_i) \cdot (1 - e^{-\alpha_i r_i})] / r_i \quad (3-12)$$

Taking the negative gradient of equation (3-12) results in equation (3-13). Two methods for obtain  $\alpha$  and  $\beta$  parameters have been presented in Wang *et al.*. The first is referred to here as charge penetration fitted (CPf), where the  $\alpha$  parameter is fitted to the QM electrostatic potential at short range using equation (3-12), after which equation (3-11) is used to solve for  $\beta$ , although this is irrelevant for electrostatic field and potential calculations, as the  $\beta$  term is not present in either. The second method, referred to as only charge penetration (CP), is to intuitively set  $\alpha$  to be the same as the number of valence electrons (with the exception of hydrogen, where  $\alpha$  was set to  $2.0 \text{ \AA}^{-1}$ ) and again use equation (3-11) to solve for  $\beta$ . The  $\alpha$  parameters used are presented in Table 3-1.

$$\vec{F}_{\text{cp}}(\vec{r}) = \sum_i^{N \text{ atoms}} \left( \frac{q_i}{r_i^2} - \frac{q_i e^{-\alpha_i r_i}}{r_i} - \frac{\alpha_i q_i e^{-\alpha_i r_i}}{r_i} + \frac{Z_i e^{-\alpha_i r_i}}{r_i^2} + \frac{\alpha_i Z_i e^{-\alpha_i r_i}}{r_i} \right) \frac{\vec{r}_i}{r_i} \quad (3-13)$$

The code to convert a GROMACS trajectory into a series of Tinker .xyz files, gmx2xyz, can be accessed at: [https://github.com/awritchie/my\\_gmx](https://github.com/awritchie/my_gmx)

The code to calculate the electrostatic field at a point in Tinker, calc\_field.f, can be accessed at: [https://github.com/awritchie/tinker\\_tools](https://github.com/awritchie/tinker_tools)

The code to calculate the charge penetration corrects, gmx\_gmx2amoeba\_cp, can be accessed at: [https://github.com/awritchie/my\\_gmx](https://github.com/awritchie/my_gmx)

Table 3-1: Charge Penetration Parameters

	Z (e <sup>-</sup> )	CP $\alpha$ ( $\text{\AA}^{-1}$ )	CPf $\alpha$ ( $\text{\AA}^{-1}$ )
H	1	2	3.3
C	4	4	3.8
N	5	5	3.1
O	6	6	3.5
P	5	5	N/A <sup>a</sup>
S	6	6	N/A <sup>a</sup>

<sup>a</sup>Parameters not presented in original literature, so values of 0 were used to reproduce the Coulomb interaction energy expression

### 3.6 VIBRATIONAL STARK EFFECT

The vibrational Stark effect (VSE) relates the change in vibrational absorption frequency of a probe harmonic oscillator to a change in electric field, as shown in equation (3-14), where  $\Delta\vec{\mu}$ , the Stark tuning rate, is the difference in dipole moments from the ground state to the first vibrationally excited state of the probe,  $\Delta\vec{v}$  is the change in vibrational absorption energy, and  $\Delta\vec{F}$  is the change in external electric field caused by a perturbation to the system:<sup>37, 81-87</sup>

$$\begin{aligned}\Delta E &= hc\Delta\tilde{v} = -\Delta\vec{\mu} \cdot \Delta\vec{F}_{\overline{\text{NC}}} \\ \Delta E &= hc\Delta\tilde{v} = \Delta\vec{\mu} \cdot \Delta\vec{F}_{\overline{\text{CN}}} \\ \Delta\vec{F}_{\overline{\text{NC}}} &= -\Delta\vec{F}_{\overline{\text{CN}}}\end{aligned}\tag{3-14}$$

The defined direction of the field is indicated as subscripts, where  $\overline{\text{NC}}$  indicates the positive direction of the field points from the nitrile nitrogen to the nitrile carbon, and  $\overline{\text{CN}}$  indicates that the positive direction of the field points from the nitrile carbon to the nitrile nitrogen.

Additional higher order terms, such as difference polarizability, are also present in the Stark effect; however, in VSE spectroscopy, their contribution is small relative to the difference dipole and thus are not considered in our analysis.<sup>85, 86</sup> The sign of the Stark effect is dependent on what is taken as the positive field direction, and the case where the positive direction is considered N to C as well as the case where the positive direction is considered C to N are both listed in equation (3-14). In general, we treat the positive direction to be pointing from C to N, and therefore we would expect a positive correlation between calculated fields and experimentally measured Stark shifts.

### 3.7 VIRTUAL STARK TUNING RATE (*VSTR*) AND “IDEAL” DIELECTRICS

Fitting the calculated electric fields to the experimental vibrational absorption frequencies yields a linear equation in which the fitted slope is related to the inverse of the Stark tuning rate (3-15), where we substituted in the vibrational Stark effect for  $\Delta\tilde{\nu}$ . In this way we can calculate an observed or virtual Stark tuning rate (*VSTR*) from our calculations, which can then be compared to the known, experimental Stark tuning rate for a thiocyanate probe of  $0.77 \text{ cm}^{-1}/(\text{MV cm}^{-1}) = 1.99 \text{ cm}^{-1}/(k_b T/e\text{\AA})$ .

$$\begin{aligned}\vec{\Delta F}_{\text{NC}} &= m_{\text{NC}}(\Delta\tilde{\nu}) = m_{\text{NC}}(-\Delta\bar{\mu}'\vec{\Delta F}_{\text{NC}}) \\ \vec{\Delta F}_{\text{CN}} &= m_{\text{CN}}(\Delta\tilde{\nu}) = m_{\text{CN}}(\Delta\bar{\mu}'\vec{\Delta F}_{\text{CN}}) \\ \Delta\bar{\mu}' &= -\frac{1}{m_{\text{NC}}} = \frac{1}{m_{\text{CN}}} = \text{VSTR}\end{aligned}\quad (3-15)$$

We can also calculate an “ideal” dielectric,  $\varepsilon'(\bar{r})$ , required to force the *VSTR* to be equal to the known Stark tuning rate. The “ideal” dielectric can be calculated using the *VSTR*,  $\Delta\bar{\mu}'$ , the experimental Stark tuning rate,  $\Delta\bar{\mu}$ , and solute dielectric,  $\varepsilon(\bar{r})$ , using equation (3-16).<sup>75,88</sup>

$$\begin{aligned}\vec{\Delta F}(\bar{r})\varepsilon(\bar{r}) &= \vec{\Delta F}'(\bar{r})\varepsilon'(\bar{r}) \\ \frac{\varepsilon(\bar{r})}{\Delta\bar{\mu}} &= \frac{\varepsilon'(\bar{r})}{\Delta\bar{\mu}'}\end{aligned}\quad (3-16)$$

## **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<sup>Ras</sup> (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.<sup>89</sup> 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,<sup>41</sup> 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.<sup>90-92</sup> The biological function of Rap is less well understood, but it has recently been shown to play a role in cellular adhesion<sup>93</sup> and has been implicated in cancer metastasis.<sup>94,95</sup> 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,<sup>40,41</sup> measurements of thermodynamic<sup>37,96-98</sup> parameters of the protein-protein binding events ( $\Delta G_{dock}$  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<sup>8,37</sup> 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<sup>98,99</sup> 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 E30/K31E and the double mutation Rap E30D/K31E cause the dissociation constant with the downstream effector Raf to resemble Ras D30/E31 more closely than Rap E30/K31.<sup>99</sup> Furthermore, the double mutant Rap E30D/K31E co-crystallized with the downstream effector Raf, which usually binds more strongly to Ras. Similarly, the mutation Ras D30/E31K was used to co-crystallize Ras with Ral, normally Rap's immediate downstream effector,<sup>40</sup> and the double reversion mutant Ras D30E/E31K has significantly reduced binding affinity with Raf.<sup>100</sup> 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.<sup>2, 3, 7, 101-103</sup> Vibrational Stark effect (VSE) spectroscopy is a recently developed experimental technique capable of measuring electrostatic fields in proteins.<sup>43, 59, 81-86</sup> In VSE spectroscopy, the intrinsic response of a probe vibrational oscillator to its local electrostatic environment is measured spectrally 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 identified.<sup>86</sup> Several characteristics of the nitrile oscillator make it particularly attractive; its absorption energy of ~2100-2250 cm<sup>-1</sup> 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.<sup>104</sup>. There are a growing number of reports describing VSE spectroscopy of nitrile probes to study the function of electrostatic fields in enzyme active sites.<sup>43, 59, 81-83</sup> 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 Ras D30/E31 or Rap E30/K31.<sup>37</sup> This study determined that at some probe locations the change in absorption upon binding to Ras D30/E31 and Rap E30/K31 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<sup>8</sup>. 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 4-1. Positions N27 and Y31 were selected because at these probe locations, we previously measured a significant difference between docking of Ras D30/E31 versus Rap E30/K31,<sup>37</sup> 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<sup>8</sup>. 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<sup>40</sup> and 1GUA<sup>41</sup> 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 Rap E30/K31, Rap E30D/K31, Rap E30/K31E, and the double mutant Rap E30D/K31E. The difference in vibrational absorption energy,  $\Delta\nu_{obs}$ , 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 Ras D30/E31. 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

All experimental measurements presented were performed by Christina Ragain, Ph.D.<sup>44</sup>

Measured dissociation constants of both WT and SCN-labeled Ral $\beta$  mutants docking with all GTPases studied here are reported in Table 4-1. Binding of WT Ral to Rap E30/K31 was approximately 10-fold faster than binding to Ras D30/E31, as has been reported before.<sup>37, 98, 99</sup> Any effect of the thiocyanate probe on the formation of the interface can be investigated by comparing  $K_d$  values obtained for wild type (WT) Ral versus the SCN-labeled Ral $\beta$  mutants docking to Ras D30/E31 and Rap E30/K31 in Table 4-1. 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 Ral $\beta$  G28C<sub>SCN</sub> (reduced  $K_d$  by an order of magnitude) and Ral $\beta$  Y31C<sub>SCN</sub> (increased  $K_d$  by an order of magnitude), still showed an order of magnitude increase in  $K_d$  when binding to Ras D30/E31 as opposed to Rap E30/K31, as is expected from our previous work. The dissociation constants presented in Table 4-1 demonstrate that the presence of the SCN probe on Ral $\beta$  mutants did not substantially affect binding to either Ras D30/E31 or Rap E30/K31, as has been observed before.<sup>37</sup> 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 Ral $\beta$  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 4-1 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 Ral $\beta$  mutants interacted with Rap E30/K31E with a  $K_d$  10-fold higher than with Rap E30/K31, and this effect was preserved in the double mutant Rap E30D/K31E. The single mutation Rap E30D/K31 had no effect on binding, and all  $K_d$  values measured with that construct were essentially identical to Rap E30/K31. 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.<sup>40,41</sup> 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 Ral $\beta$  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 Ral $\beta$  mutant docked with Rap E30/K31 and the Rap mutants E30D/K31, E30/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  $\chi_2$  torsional probability distribution for all thiocyanate groups and  $\chi_1$  torsional probability distribution side chains at Rap positions 30 and 31 are shown in Figure 4-7, Figure 4-8, and Figure 4-9. Similar to what we have seen previously,<sup>8</sup> the torsional distribution for all studied side chains was characteristic of an unhindered alkyl group, with three probability maxima separated by  $\sim 120^\circ$  and essentially no significant difference depending on the chemical identity of either the SCN-labeled Ral $\beta$  mutant in the docked complex (Figure 4-7) or the Rap variant (Figure 4-8 and Figure 4-9). Our MD sampling strategy therefore provides us with a comprehensive ensemble of structures of each docked Rap-Ral $\beta$  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 Ral $\beta$  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-Ral $\beta$  surface, which we term azimuthal and polar angles. These two angles are shown schematically in Figure 2-6. When the cross hairs on Figure 2-6 are translated to the C $\alpha$  atom of each simulated residue on Rap or Ral $\beta$  (represented as spheres), they become the origin of the polar angles shown in Figure 4-4 and Figure 4-5. These simulations were compared with previous simulations of the Ras-Ral $\beta$  binding interface,<sup>8</sup> and in all cases, the orientation of the nitrile at the docked interface for Ras D30/E31 and Rap E30/K31 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 Ral $\beta$  mutants dock with Ras D30/E31 and Rap with the same dissociation constant as WT Ral. Furthermore, as seen in Figure 4-7, 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 Rap E30/K31, 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 does not 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 4-5. When looking at position 30 (Figure 4-5A), the azimuthal angles showed essentially no difference depending on either the Rap mutant or the SCN-labeled Ral $\beta$  mutant to which it was docked. In most cases, the aspartate (for E30D and E30D/K31E) or glutamate vector (for WT and Rap E30/K31E) pointed approximately parallel and slightly below the Rap-Ral surface plane. The single exception to this was when Rap E30D/K31 was docked with Ral $\beta$  K32C<sub>SCN</sub>, where it was observed that the residue oriented itself 5° above the Rap-Ral $\beta$  K32C<sub>SCN</sub> surface plane. The polar angles of the side chain at position 30 (Figure 4-5B) 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 4-5C) were pointed significantly further below the Rap-Ral $\beta$  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 Ral $\beta$  Y31C<sub>SCN</sub>, 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 4-5D). 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 Ral $\beta$  N54C<sub>SCN</sub>, Rap E30/K31 and the Rap E30D/K31 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 Rap E30/K31 or Rap E30/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 Ral $\beta$  D51, N54, and E58. When this side chain was mutated to Rap E30/K31E in either the single or double mutant, the negatively charged Glu reoriented to avoid electrostatic repulsion with the Ral $\beta$  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-Ral $\beta$  binding and Ras-Ral $\beta$  binding described here. As discussed above, this behavior was not observed when the Rap constructs were docked with Ral $\beta$  N54C<sub>SCN</sub>; investigation of MD structures showed that this probe location disrupted the hydrogen bonding pocket sufficiently to cause this portion of the Ral $\beta$  surface retracted slightly from the Rap-Ral $\beta$  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.

*Ral $\beta$  N27C<sub>SCN</sub> and N29C<sub>SCN</sub>:* Selection of each of the 6 SCN-labeled Ral $\beta$  probes was based on consideration of each probes' orientation compared to the Rap-Ral $\beta$  interface, proximity to the Rap positions 30 and 31, and large differences in vibrational absorption energy upon binding to Ras D30/E31 and Rap E30/K31 measured in a

previous study.<sup>37</sup> N27C<sub>SCN</sub> and N29C<sub>SCN</sub> 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 4-4A).<sup>8</sup> 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 4-5C, 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 4-2.<sup>44</sup> The Ralβ N29C<sub>SCN</sub> 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 Rap E30/K31 from a previous study.<sup>37</sup> When docked with Rap E30D/K31, the thiocyanate absorption energy was 2160.8 cm<sup>-1</sup>, identical to the observed absorption energy when Ralβ N29C<sub>SCN</sub> was bound to Rap E30/K31. When Ralβ N29C<sub>SCN</sub> was incubated with Rap E30/K31E, the absorption energy was 0.6 cm<sup>-1</sup> higher in energy than when docked with Rap E30/K31, 2161.4 cm<sup>-1</sup>. However, when docked with the double mutant, Rap E30D/K31E, the absorption energy of the thiocyanate shifted 0.8 cm<sup>-1</sup> higher in energy (2161.6 cm<sup>-1</sup>). Both Rap mutants containing K31E were therefore more similar to the observed absorption energy when Ralβ N29C<sub>SCN</sub> is docked with Ras D30/E31 (2161.1 cm<sup>-1</sup>) than with Rap E30/K31. This means that the probe is

experiencing an electrostatic environment in the double mutant E30D/K31E that is more like that of Ras D30/E31 than to Rap E30/K31, although the effect is small.

Changes in the absorption energy ( $\Delta\nu_{\text{obs}}$ ) of the thiocyanate on each SCN-labeled Ral $\beta$  mutant due to binding for each SCN-labeled Ral $\beta$  mutant studied here are summarized in Table 4-2 and Figure 4-3. In these figures, all absorption energies are referenced to the absorption energy of the nitrile probe when docked with Rap E30/K31 (i.e.  $\Delta\nu_{\text{obs}} = 0$  represents no change from the thiocyanate absorption energy when docked with Rap E30/K31 reported previously).<sup>37</sup> These shifts are related to the change in electrostatic field through the known value of the Stark tuning rate of the cyanocysteine probe, of  $0.7 \text{ cm}^{-1}/(\text{MV/cm})$ ; the values of  $\Delta F$  determined from equation (3-14) are reported in Table 4-2.

When the probe is located at Ral $\beta$  N27C<sub>SCN</sub> and N29C<sub>SCN</sub>, the data in Figure 4-3 clearly show that both the single and double reversion mutants Rap E30/K31E and E30D/K31E have VSE absorption energies that are similar to Ras D30/E31, not Rap E30/K31. In both cases, the single mutation Rap E30D/K31 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 Ras D30/E31 and the double mutant Rap E30D/K31E than when bound to Rap E30/K31 are clear. The relatively small effect of mutations on the magnitude of the absorption energy of Ral $\beta$  N29C<sub>SCN</sub> 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.<sup>8</sup> This effect is currently being investigated in our laboratory and will be described in a future report.

*K32C<sub>SCN</sub> and N54C<sub>SCN</sub>:* 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 4-3, the two single mutations E30D and K31E have small but opposite effects on the absorption energy of N54C<sub>SCN</sub> when compared to Rap E30/K31; Rap E30D/K31 is 0.4 cm<sup>-1</sup> higher in absorption energy, while Rap E30/K31E is 0.1 cm<sup>-1</sup> lower in absorption energy. The combined effect of the double mutant Rap E30D/K31E, however, gave a VSE shift 0.2 cm<sup>-1</sup> higher in energy than Rap E30/K31, resulting in an absorption energy that was approximately the sum of the behavior of the two single mutations. As discussed above, because N54C<sub>SCN</sub> 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<sub>SCN</sub> because of its proximity to positions 30 and 31 on the GTPase in the docked complex. As can be seen in Table 4-2 and Figure 4-3, although the measured error in  $\Delta\nu_{obs}$  was larger than other positions, there was a dramatic effect of the mutation Rap E30/K31E on the absorption energy of the thiocyanate compared to Rap E30/K31 (-1.0 cm<sup>-1</sup>), while Rap E30D/K31 caused only a small perturbation (-0.2 cm<sup>-1</sup>). The behavior of the double mutant Rap E30D/K31E was the sum of these two shifts, -1.2 cm<sup>-1</sup>. This was very different from the response on binding to Ras D30/E31, which showed a shift in absorption energy of +0.7

$\text{cm}^{-1}$  compared to Rap E30/K31. Thus, again, the differences in the electrostatic fields between Ras D30/E31 and Rap E30/K31 experienced by the probe at Ral $\beta$  K32C<sub>SCN</sub> 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<sub>SCN</sub>, the mutation E30D caused an increase in absorption energy relative to Rap E30/K31, 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 Ral $\beta$  K32C<sub>SCN</sub>, 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<sub>SCN</sub> and Y31C<sub>SCN</sub>:* We chose Ral $\beta$  G28C<sub>SCN</sub> 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 Ras D30/E31 versus Rap E30/K31,<sup>37</sup> and a position in the docked complex that was more consistently parallel to the surface than either N27C<sub>SCN</sub> or N29C<sub>SCN</sub>.<sup>8</sup> Even still, this probe did respond strongly to mutants Rap E30D/K31 and K31E ( $+1.0\text{ cm}^{-1}$  and  $+0.5\text{ cm}^{-1}$  versus Rap E30/K31, respectively), compared to a shift of  $-0.8\text{ cm}^{-1}$  when bound to Ras D30/E31. The double mutant, Rap E30D/K31E, however, only demonstrated a shift of  $+0.2\text{ cm}^{-1}$  compared to Rap E30/K31. The probe at this location thus reacted in a manner in which the two single mutations appear to cancel each other out.

Position Ral $\beta$  Y31C<sub>SCN</sub> was selected for study because the change in electrostatic field upon docking to Rap E30/K31 differed from Ras D30/E31 by a large amount, 0.8 cm<sup>-1</sup>, and resulted in changes in field in the opposite direction. As shown in Table 4-2, both single mutants Rap E30D/K31 and Rap E30/K31E shifted the absorption energy of the nitrile probe on Ral $\beta$  Y31C<sub>SCN</sub> to lower energy, by -2.0 cm<sup>-1</sup> and -1.0 cm<sup>-1</sup> respectively. However, the combined effect of the double mutant was to shift the absorption energy lower by only -0.2 cm<sup>-1</sup> compared to Rap E30/K31, 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) Ral $\beta$  Y31C<sub>SCN</sub> 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, Ral $\beta$  G28C<sub>SCN</sub> and Ral $\beta$  Y31C<sub>SCN</sub>, although each single Rap mutant caused a large change in absorption energy (higher in energy at Ral $\beta$  G28C<sub>SCN</sub>, and lower in energy at Ral $\beta$  Y31C<sub>SCN</sub>), the double mutant effectively canceled out those changes in both cases. While MD sampling of position Ral $\beta$  Y31C<sub>SCN</sub> makes it clear that this is likely due to a large structural reorientation of the Rap side chain position 31, the influences of the probe at position Ral $\beta$  G28C<sub>SCN</sub> are much less clear. After extensive inspection of our MD simulations, we have found no significant structural differences near G28C<sub>SCN</sub> 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 Ral $\beta$  G28C<sub>SCN</sub> 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 Ral $\beta$  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 Ral $\beta$  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 Ral $\beta$  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<sub>SCN</sub> and N29C<sub>SCN</sub>, 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 Ras D30/E31, 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<sub>SCN</sub> and N54C<sub>SCN</sub>. 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<sub>SCN</sub> and Y31C<sub>SCN</sub>, while each single Rap mutation caused a large shift in vibrational absorption energy, the double mutant Rap E30D/K31E behaved essentially identically to Rap E30/K31. 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.

#### **4.4 CONCLUSION**

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 Ral $\beta$  N27C<sub>SCN</sub> and N29C<sub>SCN</sub> 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.

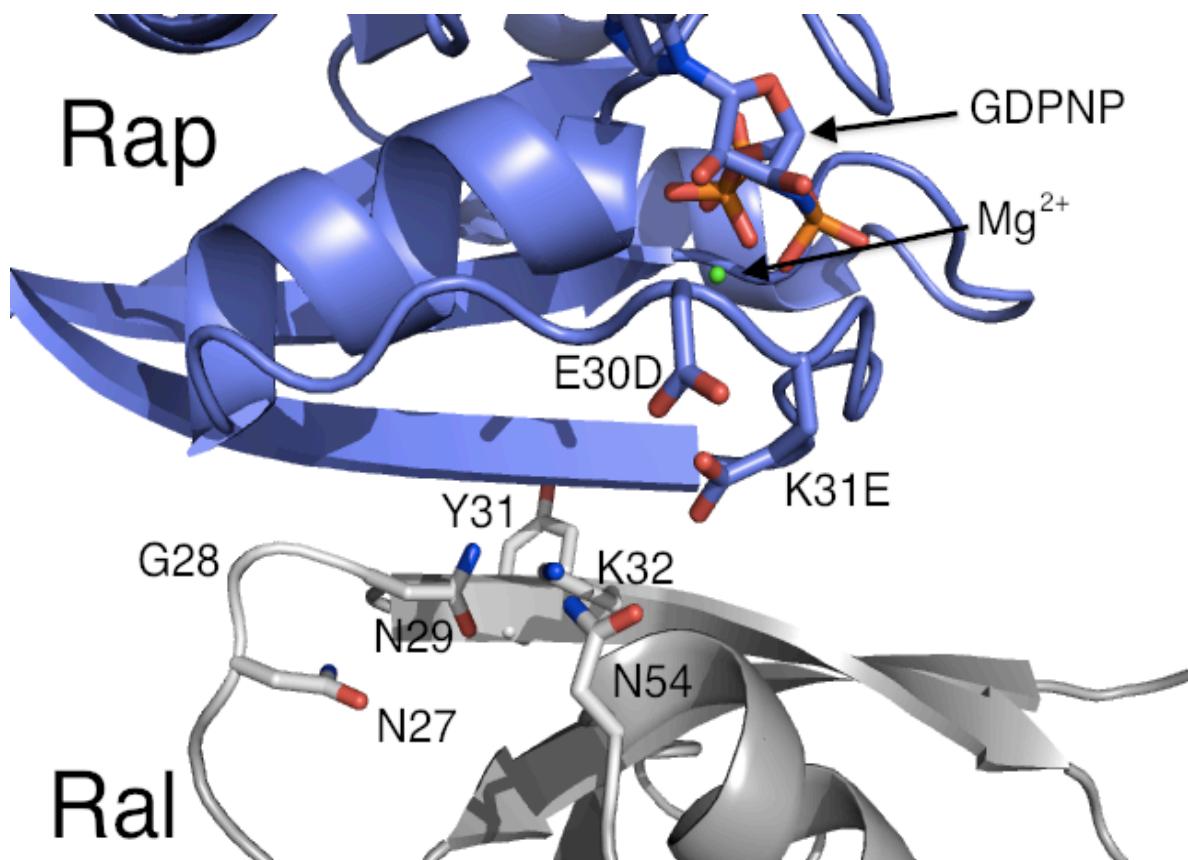


Figure 4-1: Rap-Ral Interface and Mutation Locations

The Rap (blue)-Ral (gray) interface, highlighting Rap E30D/K31 and K31E and six amino acids on Ral that were selected for positioning the thiocyanate VSE probe in this study. The figure was prepared by aligning Rap from 1GUA<sup>41</sup> (containing the mutations E30D/K31E) to Ras docked with Ral from 1LFD<sup>39</sup>.

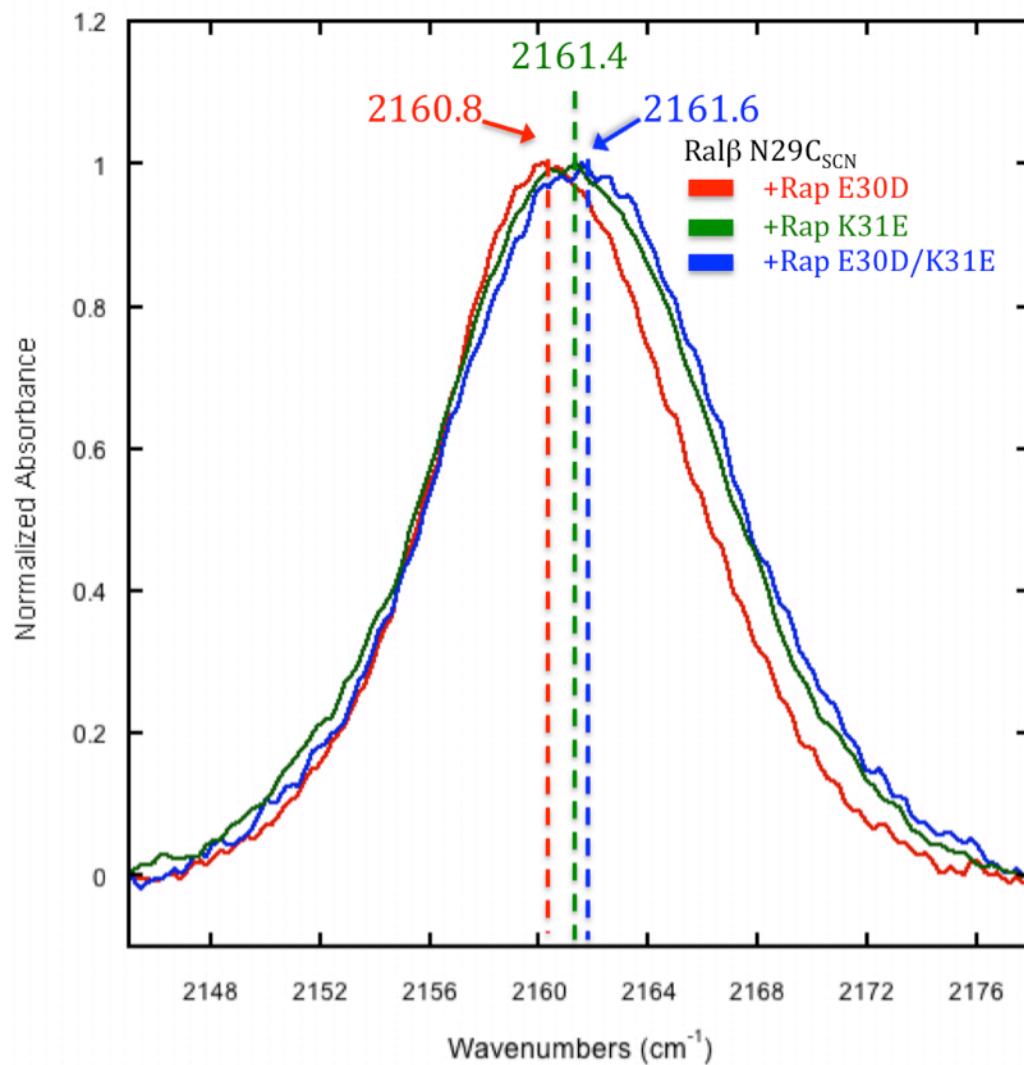


Figure 4-2: Representative Thiocyanate-labeled Protein Spectra

Normalized absorbance of thiocyanate on Ral $\beta$  N29C<sub>SCN</sub> measured when docked with Rap E30D/K31 (red,  $2160.8 \text{ cm}^{-1}$ ), Rap E30/K31E (green,  $2161.4 \text{ cm}^{-1}$ ), and Rap E30D/K31E (blue,  $2161.6 \text{ cm}^{-1}$ ).

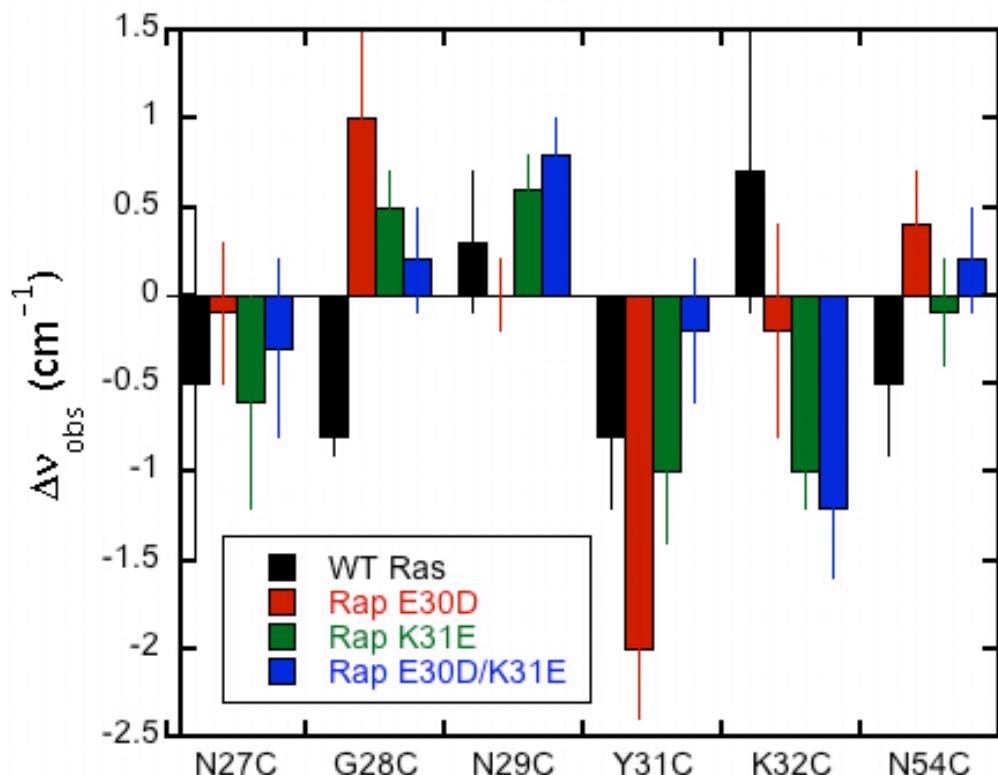


Figure 4-3: Change in Absorption Energy of Thiocyanate Probes Relative to Rap E30/K31

Change in absorption energy compared to Rap E30/K31,  $\Delta V_{obs}$ , of a nitrile probe on SCN-labeled Ral $\beta$  mutants when bound to Ras D30/E31 (blue), Rap E30D/K31 (red), Rap E30/K31E (green), and Rap E30D/K31E (blue) where  $\Delta V_{obs} = 0$  represents no change from the thiocyanate absorption energy when bound to Rap E30/K31 reported in Table 4-2. Error bars represent propagation of error of  $V_{obs}$ .

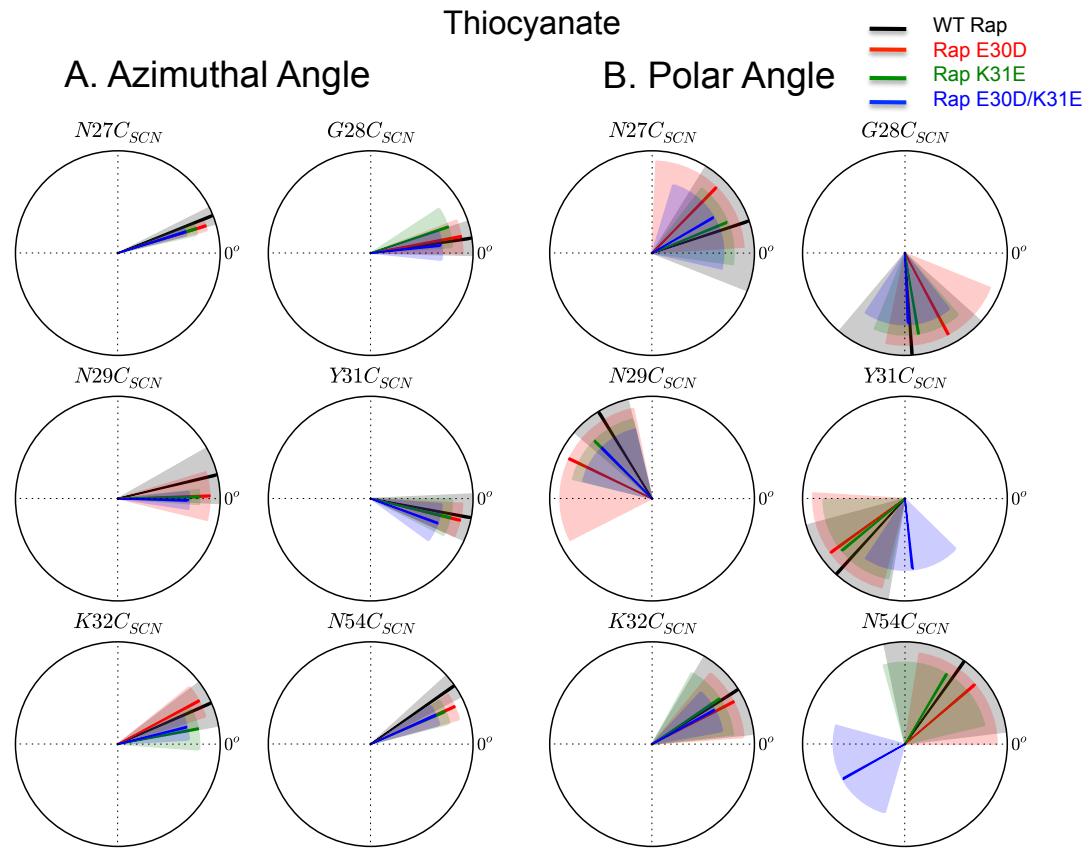


Figure 4-4: Azimuthal and Polar Angles of Simulated CNC Probes

(A) Azimuthal and (B) Polar angles of the thiocyanate on SCN-labeled Ral $\beta$  mutants calculated from each mutant docked with Rap E30/K31 (black), Rap E30D/K31 (red), Rap E30/K31E (green), and Rap E30D/K31E (blue). Azimuthal angles are shown relative to the Rap-Ral $\beta$  surface plane. Polar angles are shown relative to the coordinate system described in Figure 2-6B. The shaded area represents one standard deviation on the calculated angle from the Boltzmann-weighted ensemble of structures derived from Equation (2-12).

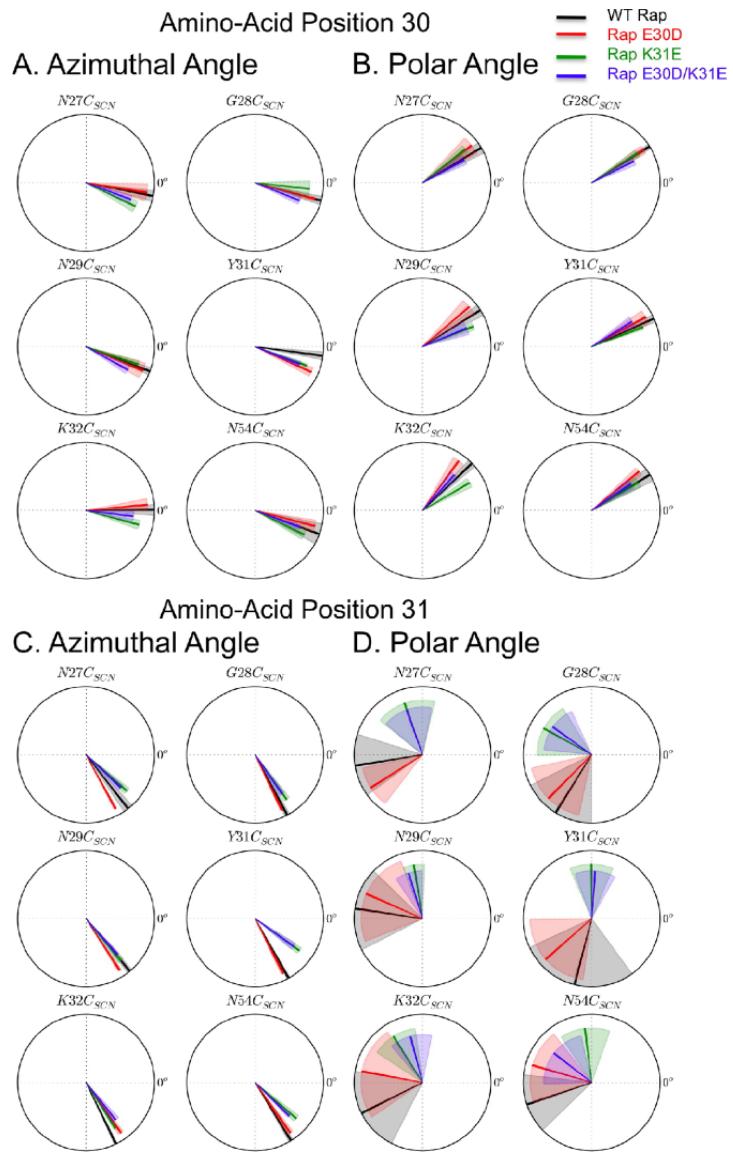


Figure 4-5: Azimuthal and Polar Angles of Simulated Rap Position 30 and 31 Side chains

Azimuthal and polar angles of the side chain at Rap positions 30 and 31 in Rap E30/K31 (black), Rap E30D/K31 (red), Rap E30/K31E (green), and Rap E30D/K31E (blue). (A) Azimuthal angle at Rap position 30; (B) polar angle at Rap position 30; (C) azimuthal angle at Rap position 31; (D) polar angle at Rap position 31. Azimuthal angles are shown relative the Rap-Ral surface plane. Polar angles are shown relative to the

coordinate system shown in Figure 2-6. Dashed lines represent one standard deviation on the calculated angle from the Boltzmann-weighted ensemble of structures.

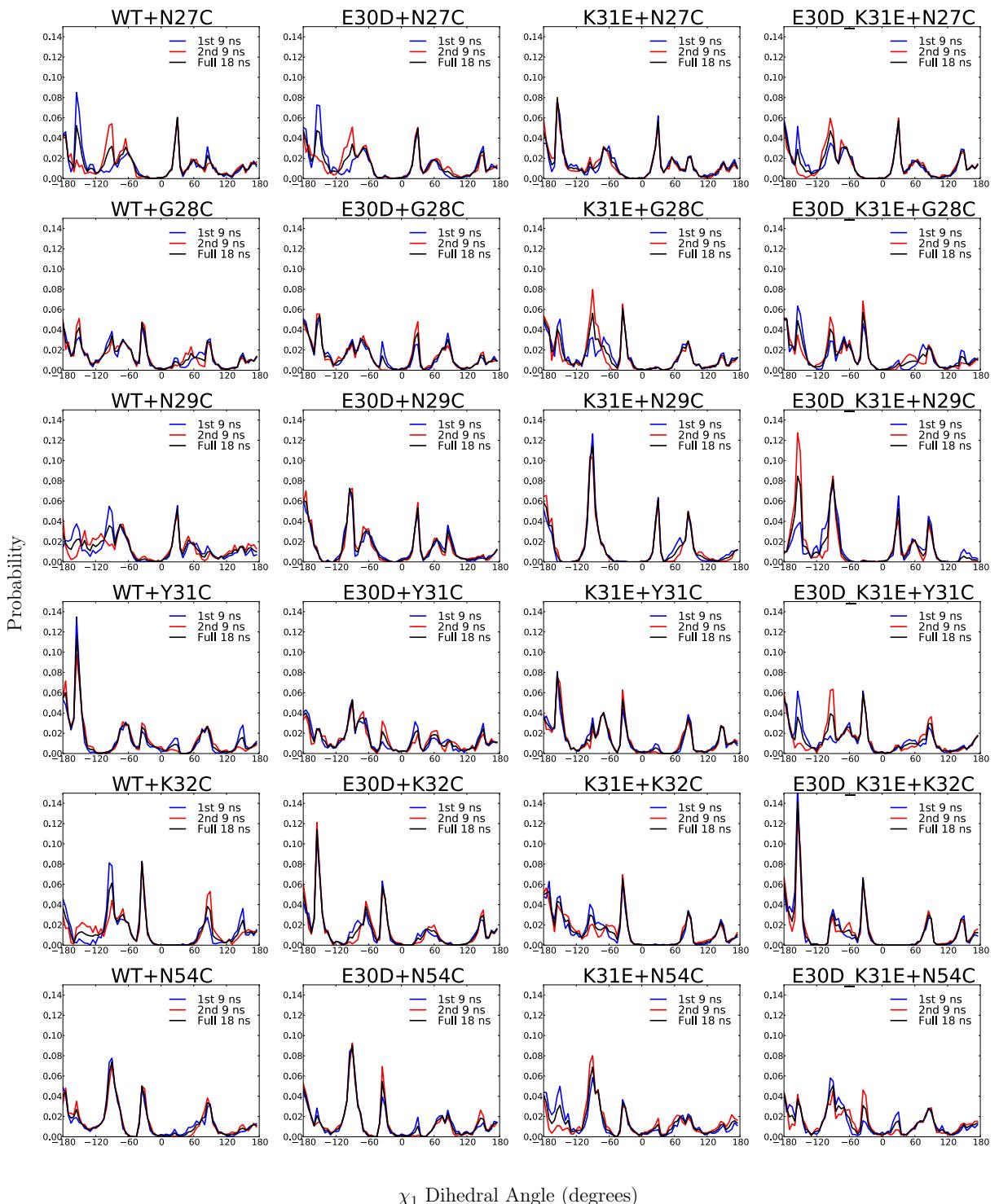


Figure 4-6: Convergence Test Based on CNC Dihedral Distributions

Probability distributions of the cyanocysteine  $\chi_2$  dihedral angle derived from WHAM of the first half (blue), second half (red), and total simulation time (black) for each Ral $\beta$  mutant docked to Rap E30/K31 and each of its mutants.

### Thiocyanate Dihedral Angles

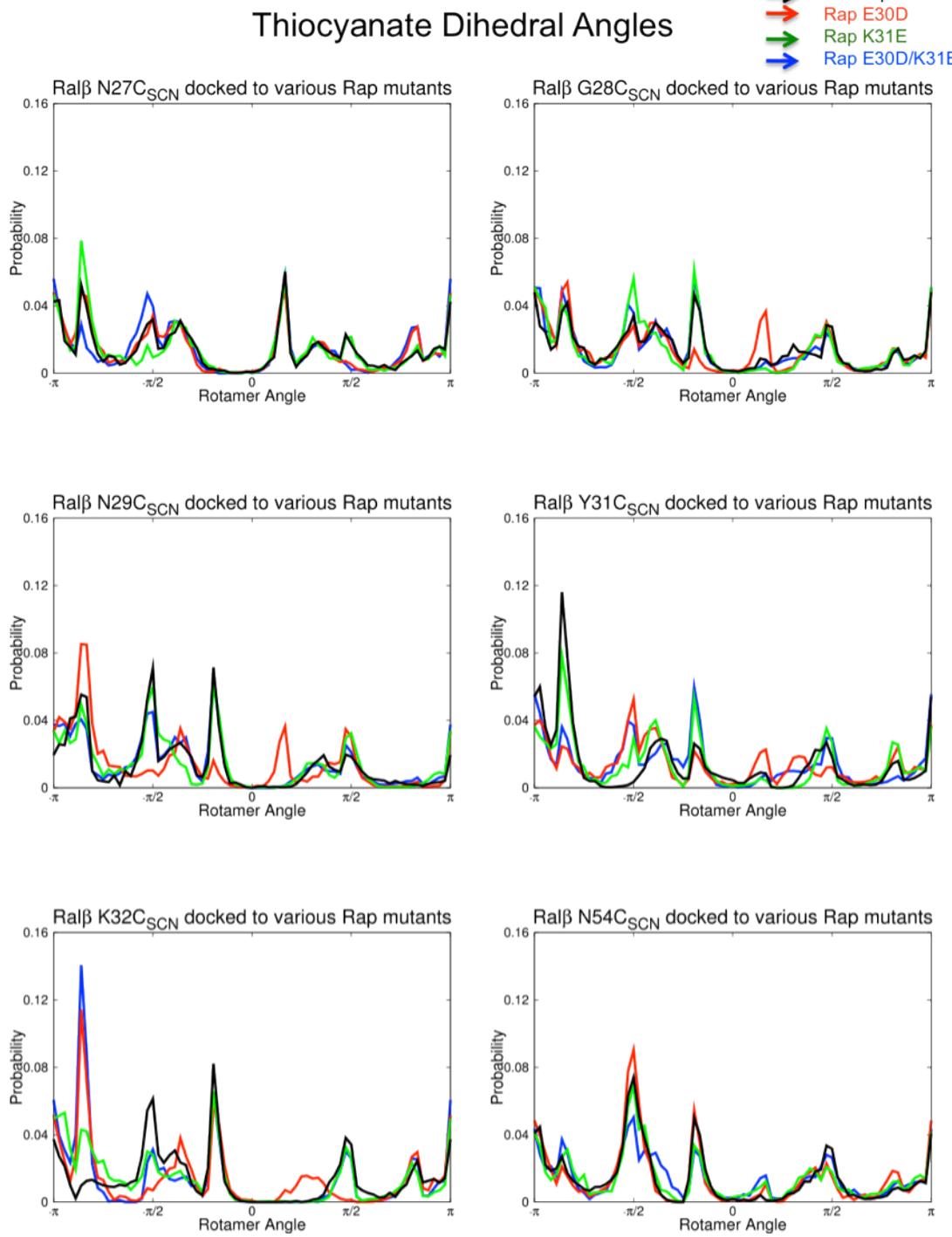


Figure 4-7: Thiocyanate Probability Distributions

Probability distributions, derived from WHAM, of the thiocyanate probe  $\chi_2$  dihedral torsion for SCN-labeled Ral $\beta$  mutants docked with Rap E30/K31 (black), Rap E30D/K31 (red), Rap E30/K31E (green), and Rap E30D/K31E (blue).

### Residue 30 Dihedral Angles


 WT Rap  
 Rap E30D  
 Rap K31E  
 Rap E30D/K31E

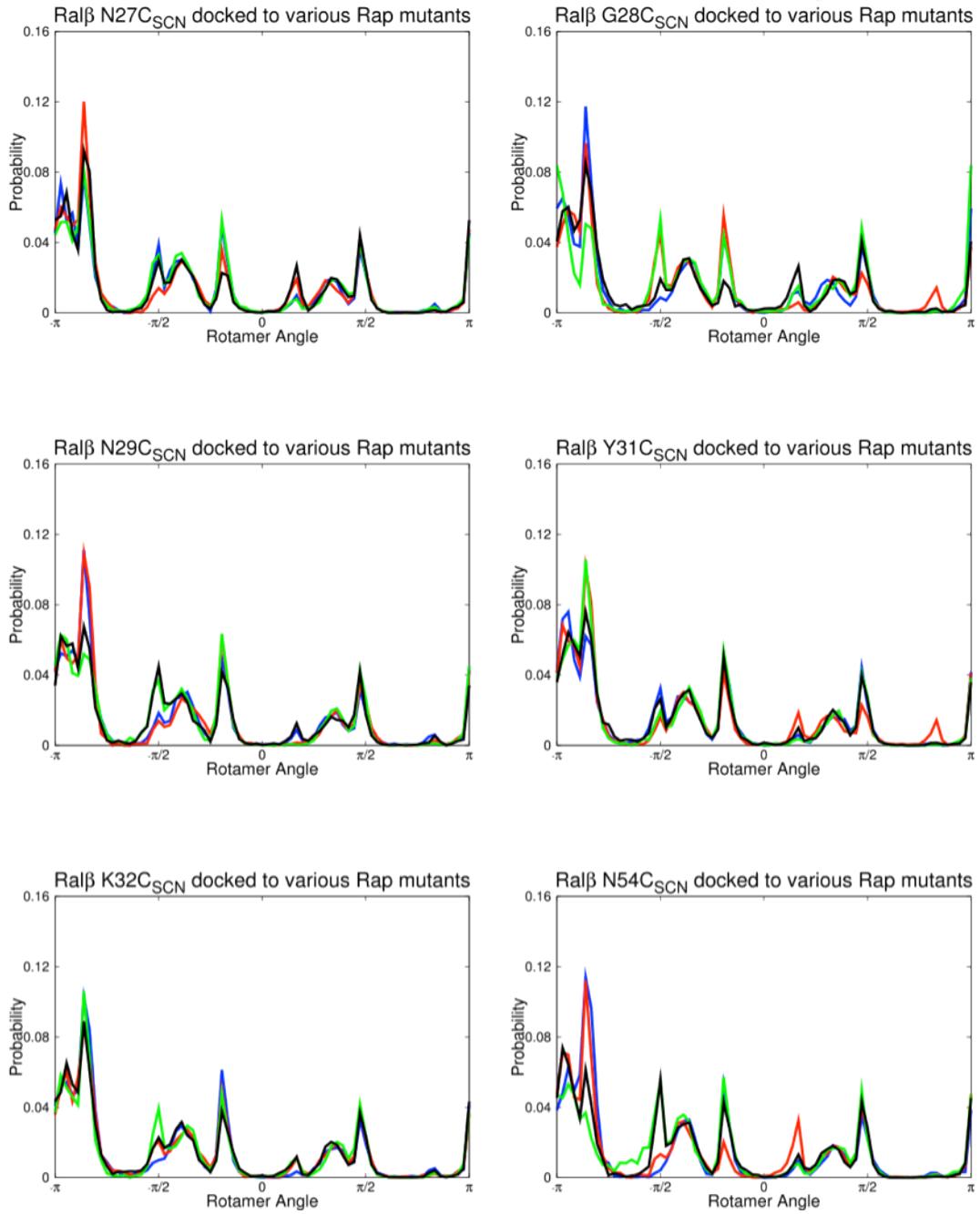


Figure 4-8: Rap Position 30 Dihedral Distributions

Probability distributions, derived from WHAM, of the side chain  $\chi_1$  dihedral torsion for Rap position 30 on SCN-labeled Ral $\beta$  mutants docked with Rap E30/K31 (black), Rap E30D/K31 (red), Rap E30/K31E (green), and Rap E30D/K31E (blue).

### Residue 31 Dihedral Angles

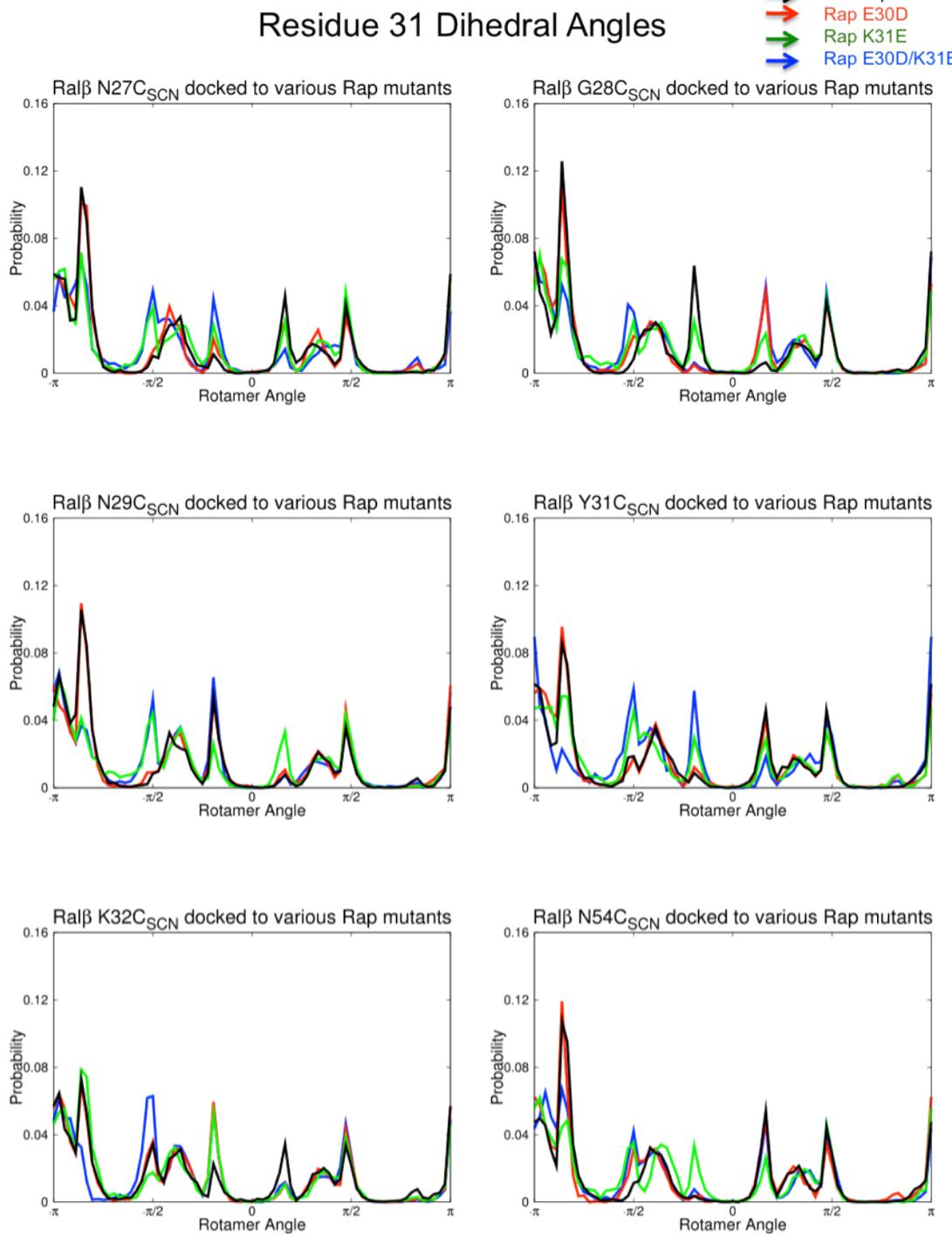


Figure 4-9: Rap Position 31 Dihedral Distributions

Probability distributions, derived from WHAM, of the side chain  $\chi_1$  dihedral torsion for Rap position 31 on SCN-labeled Ral $\beta$  mutants docked with Rap E30/K31 (black), Rap E30D/K31 (red), Rap E30/K31E (green), and Rap E30D/K31E (blue).

Table 4-1: Dissociation constant of the formation of docked complexes of SCN-labeled Ral $\beta$  mutants with GTPases

		GTPase				
		Rap	Ras	Rap	Rap	Rap
		E30/K31	D30/E31	E30D/K31	E30/K31E	E30D/K31E
Ral $\beta$ Mutation K <sub>d</sub> ( $\mu$ M)	WT	0.26 $\pm$ 0.09	1.4 $\pm$ 0.2	0.24 $\pm$ 0.06	1.8 $\pm$ 0.7	1.0 $\pm$ 0.1
	N27C <sub>SCN</sub>	0.30 $\pm$ 0.05	7.3 $\pm$ 2.2	0.12 $\pm$ 0.01	2.9 $\pm$ 0.7	5.4 $\pm$ 1.0
	G28C <sub>SCN</sub>	0.04 $\pm$ 0.004	4.8 $\pm$ 1.1	0.03 $\pm$ 0.015	8.1 $\pm$ 1.5	2.5 $\pm$ 0.3
	N29C <sub>SCN</sub>	0.15 $\pm$ 0.08	3.1 $\pm$ 0.6	0.10 $\pm$ 0.02	6.0 $\pm$ 1.1	5.0 $\pm$ 1.7
	Y31C <sub>SCN</sub>	1.0 $\pm$ 0.1	12.9 $\pm$ 4.8	0.89 $\pm$ 0.35	6.2 $\pm$ 2.0	2.0 $\pm$ 0.2
	K32C <sub>SCN</sub>	0.69 $\pm$ 0.18	6.0 $\pm$ 2.3	0.34 $\pm$ 0.09	1.9 $\pm$ 0.8	1.8 $\pm$ 0.5
	N54C <sub>SCN</sub>	0.91 $\pm$ 0.20	4.7 $\pm$ 0.4	0.79 $\pm$ 0.10	2.2 $\pm$ 1.5	1.1 $\pm$ 0.3

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

		Ral $\beta$ Mutant		
GTPase		N27C	G28C	N29C
Rap E30/K31	(cm $^{-1}$ )	2162.6 (0.4)	2161.8 (0.1)	2160.8 (0.2)
Difference Compared to Rap E30/K31				
Ras D30/E31	(cm $^{-1}$ )	-0.5 (1.0)	-0.8 (0.1)	0.3 (0.4)
Rap E30D/K31	(MV/cm)	0.7 (1.4)	1.1 (0.1)	-0.4 (0.6)
Rap E30/K31E	(cm $^{-1}$ )	-0.1 (0.4)	1.0 (0.5)	0.0 (0.2)
Rap E30D/K31E	(MV/cm)	0.1 (0.6)	-1.4 (0.7)	0.0 (0.3)
Rap E30/K31E	(cm $^{-1}$ )	-0.6 (0.6)	0.5 (0.2)	0.6 (0.2)
Rap E30D/K31E	(MV/cm)	0.9 (0.9)	-0.7 (0.3)	-0.9 (0.3)
Rap E30D/K31E	(cm $^{-1}$ )	-0.3 (0.5)	0.2 (0.3)	0.8 (0.2)
Rap E30D/K31E	(MV/cm)	0.4 (0.7)	-0.3 (0.4)	-1.1 (0.3)

		Ral $\beta$ Mutant		
GTPase		Y31C	K32C	N54C
Rap E30/K31	(cm $^{-1}$ )	2161.5 (0.2)	2160.9 (0.2)	2161.4 (0.2)
Difference Compared to Rap E30/K31				
Ras D30/E31	(cm $^{-1}$ )	-0.8 (0.4)	0.7 (0.8)	-0.5 (0.4)
Rap E30D/K31	(MV/cm)	1.1 (0.6)	-1.0 (1.1)	0.7 (0.6)
Rap E30/K31E	(cm $^{-1}$ )	-2.0 (0.4)	-0.2 (0.6)	0.4 (0.3)
Rap E30D/K31E	(MV/cm)	2.9 (0.6)	0.3 (0.8)	-0.6 (0.4)
Rap E30/K31E	(cm $^{-1}$ )	-1.0 (0.4)	-1.0 (0.2)	-0.1 (0.3)
Rap E30D/K31E	(MV/cm)	1.4 (0.6)	1.4 (0.3)	0.1 (0.4)
Rap E30D/K31E	(cm $^{-1}$ )	-0.2 (0.4)	-1.2 (0.4)	0.2 (0.3)
Rap E30D/K31E	(MV/cm)	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**

Electrostatic interactions are an important force in all biological macromolecules, influencing fundamental processes such as protein folding, ligand binding, signal transduction, macromolecular interactions, and more.<sup>1, 105-116</sup> Experimental measurement of electrostatic forces in biomolecules is a difficult and long-standing problem in biophysics, and so electrostatic effects on numerous biological structures and processes are usually investigated *in silico* through a variety of computational strategies. Many computational techniques exist for quantitatively predicting the strength and direction of biomolecular electrostatic fields, such as combined quantum mechanical / molecular mechanics (QM/MM) calculations,<sup>17, 18, 29, 117-119</sup> the protein dipole-Langevin dipoles (PDLD) method of Warshel and Levitt,<sup>6</sup> the continuum Poisson-Boltzmann (PB) model,<sup>47, 67, 70, 71, 73, 120-124</sup> and Coulomb potentials using both fixed-charge<sup>43</sup> and polarizable force fields.<sup>13, 73, 106, 125-129</sup> All of these methods vary tremendously in their physical basis, scope, computational expense, and application.

Within this diverse set of computational strategies, PB approaches have become common for addressing electrostatic questions of importance across many areas of biophysics and biochemistry because they are quick, easy to implement, and have a low barrier to application both in technical difficulty and in infrastructure requirements.<sup>68</sup>

Solutions to the LPBE are now used extensively throughout biophysical and biochemical studies to evaluate important protein properties such as free energy of ligand binding<sup>130-132</sup> or protein-protein docking.<sup>133-135</sup> However, it has been difficult to bridge

theoretical electrostatic calculations with experimental considerations by validating the results of the LPBE with direct experimental measurements of electrostatic fields in biomolecules. Attempts to do this have often relied on experimental observables that are indirectly related to the electrostatic environment, such as  $^{19}\text{F}$  resonance studies,  $\text{pK}_\text{a}$  titrations,<sup>74, 136-141</sup> or isothermal titration calorimetry,<sup>142-145</sup> making quantitative interpretation of purely electrostatic effects difficult.<sup>95</sup> Because of this, there is an important need for new data sets that measure electrostatic fields directly and against which the results of different electrostatics calculation strategies can be compared and contrasted.

Our laboratory has approached this difficulty by using vibrational Stark effect spectroscopy (VSE)<sup>82, 84-86</sup> to measure electrostatic fields in proteins directly. We have applied VSE spectroscopy to nitrile vibrational probes placed throughout the protein-protein interface of Ras-like guanosine triphosphatases (GTPases) docked with downstream effector protein domains. The Ras superfamily of GTPases, which have over 150 variants in human cells,<sup>146</sup> plays a significant role in a number of different signal transduction cascades in healthy cells.<sup>89</sup> The human oncoprotein p21<sup>Ras</sup> (hereafter “Ras”) and a related GTPase, Rap, share 50% amino acid identity,<sup>41</sup> 80% amino acid homology, and have a RMSD of 0.7 Å, and yet show significant differences in their binding specificity to their respective downstream effector proteins *in vitro*.<sup>37, 96-98</sup> It has been postulated by our laboratory that due to their large structural similarities, these functional differences are due to electrostatic interactions at the docked protein-protein interface. Our laboratory has investigated this hypothesis with VSE spectroscopy, and several publications have provided an extensive set of experimental data against which to compare the results of electrostatic calculations.<sup>37, 44</sup>

In the experimental work previously done in our laboratory, a thiocyanate probe was introduced to the Rap downstream effector RalGDS through site-directed mutagenesis to add a cysteine side chain and subsequent chemical labeling to the nitrile-containing cyanocysteine side chain,<sup>81</sup> and VSE spectroscopy has been used to measure electrostatic fields in the docked Rap-RalGDS construct.<sup>37,44</sup> We have focused on the thiocyanate probe for these experiments because it has a large difference dipole moment or Stark tuning rate,  $\Delta\vec{\mu}$ , ( $0.77 \text{ cm}^{-1}/(\text{MV cm}^{-1}) = 1.99 \text{ cm}^{-1}/(k_b T/e\text{\AA})$ ), absorbs in an unpopulated region of the infrared spectrum (typically  $2100\text{-}2300 \text{ cm}^{-1}$ ), and is small, minimizing perturbations to the probed system.<sup>82</sup> The difference in vibrational absorption energies of the nitrile probe on the surface of RalGDS in the docked complexes was compared to that on the monomeric protein to obtain an experimental value of  $\Delta\tilde{v}$  in equation (3-14). Because  $\Delta\vec{\mu}$  is parallel to the nitrile bond axis, the dot product is equal to the scalar product and equation (3-14) is simplified to equation (5-1):

$$\Delta E = hc\Delta\tilde{v} = -\Delta\vec{\mu}\Delta\vec{F}_{\text{NC}} \quad (5-1)$$

These experimental data provide a unique opportunity to test the predictions from electrostatic computational strategies, including LPBE methods, by attempting to reproduce the observed vibrational energy shifts due to the binding of RalGDS to the GTPase Rap and rationalizing the successes and failures of these attempts based on the physics of the system and the capabilities of the calculation. This system therefore serves as a tool for comparing our computational methods to experimental results.

Previous computational work examining the electrostatic interactions of Ras docked to RalGDS has been done in our laboratory using two electrostatic methods: Coulomb interactions with explicit waters<sup>43</sup> and the implicit solvent, continuum-model LPBE.<sup>75</sup> To solve the LPBE, we chose to use the Adaptive Poisson-Boltzmann Solver

(APBS) software package.<sup>111</sup> APBS is a Poisson-Boltzmann equation solver which can use both multigrid<sup>70, 71</sup> and finite-element methods<sup>147, 148</sup> to solve the PB equation, and has been widely used for solvation free energy calculations.<sup>6, 123</sup> It is highly tunable, well documented, easy to use, free, and has the active and continuing support of a development community to ensure reliability and reproducibility of results.<sup>149</sup> Furthermore, APBS has already been successfully integrated with the AMOEBA polarizable force field for solvation free energy calculations, and is being used in our laboratory to test the benefits of polarizable *versus* point-charge force fields.<sup>73</sup>

Work in our group has previously used APBS to test the reliability of LPBE solutions for predicting experimental vibrational spectroscopy data on Ras docked to nitrile-labeled RalGDS mutants.<sup>37, 43</sup> These studies employed a two-stage focusing strategy where an initial calculation on a large box filled with a coarse grid (“first stage”) provides the boundary conditions for a second calculation within a smaller box filled with a significantly finer grid (“second stage”).<sup>1</sup> We have shown that decreasing the second stage calculation box sizes and increasing grid density led APBS calculations to converge to a particular numerical value for an individual structure.<sup>75</sup> This work raised two outstanding questions that are the subject of the current investigation: 1) what is the most productive focusing strategy available for balancing accuracy of the results with time and resources used; and 2) what computational strategy provides the most accurate results for both absolute vibrational absorption energies measured on both the monomeric RalGDS surface and the docked Rap-RalGDS construct. To address these questions, in the current report we are using a recently published data set in which nitrile-labeled RalGDS constructs were bonded both with Rap E30/K31 and three Rap mutants, E30D, K31E, and E30D/K31E, that had been generated to revert the function of Rap to that of the structurally similar but functionally distinct GTPase Ras.<sup>44</sup> Apart from addressing the

questions stated above, this also presents an opportunity to test APBS against an entirely new data set based on VSE experiments, to determine the generality of our observations and conclusions.

To address our first question, in the work reported here we have focused primarily on the two smallest boxes and the two highest grid densities previously tested: a  $19^3 \text{ \AA}^3$  box with either 193 and 161 grid points in each dimension, and a  $10^3 \text{ \AA}^3$  box with either 193 and 161 grid points in each dimension. We have also looked at larger boxes with lower grid density to determine whether a high-resolution field at the bond vector was more or less important than explicitly considering a larger number of protein atoms in the second stage calculation. This was done with a  $30 \text{ \AA} \times 45 \text{ \AA} \times 40 \text{ \AA}$  box and a  $60 \text{ \AA} \times 45 \text{ \AA} \times 40 \text{ \AA}$  box, each with approximately  $0.25 \text{ \AA}$  between grid points in each dimension. Increasing the grid density of the larger boxes to match the finest grid spacing of the smaller boxes requires more than a 100-fold increase in memory and almost 100-fold increase in the number of processors, as estimated by the psize.py utility included with APBS. Because of this increase in computational cost and the large number of structural frames in which we need to solve the LPBE, we chose to limit the resolution of the  $30 \text{ \AA}$  and  $60 \text{ \AA}$  box calculations.

To address our second outstanding question, determining methods that give the most accurate results for both absolute and difference measurements, we have studied the effect of the location of the second stage box to investigate whether decreasing the volume of the box occupied by implicit solvent would have a significant impact on the quality of the calculations. To solve the Poisson-Boltzmann equation to determine an electrostatic potential, APBS divides the system into distinct dielectric regions, based on the atom positions and the van der Waals radii of the atoms and solvent molecules. The high dielectric region is defined as the total region in which spheres with the radius of the

solvent ( $\text{H}_2\text{O}$ ) molecules exist without overlapping with the van der Waals radii of the explicit (protein) molecules. All other space is considered part of the low dielectric system. This creates a discontinuity in the dielectric at the protein-solvent high dielectric-low dielectric interface.<sup>67</sup> Setting aside that fact that dielectric is a macroscopic property that is a function of atomic polarizabilities, and thus fails to capture the local molecular environment within a region of the protein, we wanted to examine whether trying to minimize the surface area of this discontinuity in the calculation could be an easy and reproducible way to improve the accuracy of our results when compared to experiment. To do this we translated the second stage box away from the solvent to define three unique box locations previously described on page 67 and shown in Figure 3-2. In all cases, the volume of the box filled with explicit protein atoms was quantified and compared to the accuracy of the results, tested against our experimental data.

Additionally, we have compared the numeric result of the LPBE to the result obtained by taking the negative gradient of the numeric solvent reaction field potential (SRF), shown in equation (3-5) (where  $\phi(\vec{r})$  is the potential,  $\epsilon_{protein}$  is the protein dielectric and, and  $\epsilon_{solvent}$  is the solvent dielectric), and adding the analytic Coulomb protein field, shown in equation (3-2) (where  $\epsilon_0$  is the permittivity of free space,  $\epsilon_{protein}$  is the protein dielectric,  $q_i$  is the charge on atom  $i$ ,  $\vec{r}$  is the vector from atom  $i$  to the location of interest, and  $\hat{r}$  is the unit vector of  $\vec{r}$ ), projected along the bond vector at a dielectric of 2, hereafter referred to as the reaction field method.<sup>27</sup>

In the reaction field method, the LPBE is solved twice for each frame. The first term in equation (3-5) is the numeric solution to the LPBE equation and the second term is the numeric solution of the LPBE in a reference dielectric. Taking the difference, the numeric, low-resolution protein field is removed, leaving the only low-resolution electrostatic potential due to solvent. By using the exact analytic field due to protein

atom partial charges with the low-resolution SRF, errors due to grid spacing and potential mapping to the dummy atoms should be reduced.

Many crystal structures of the GTPase-downstream effector docked complex have generated consensus on the structure of the Rap E30/K31-WT RalGDS interface. However, the protein constructs employed in our experiments contain a number of small perturbations to these WT molecules (mutations to Rap and thiocyanate VSE probes on RalGDS) for which there is no structural data. Because of this, it is necessary for these differences in the molecular structure of the protein-protein interface to be modeled with molecular dynamics (MD) simulations to determine structural changes in the system due to the probe, mutations, and protein-protein docking. Furthermore, other studies looking at electrostatic effects have shown that allowing for structural reorientation improves the reliability of these sorts of calculations—not surprising for experimental data collected in a steady-state ensemble.<sup>104, 150, 151</sup> In our previous simulations, a single degree of freedom, the  $\chi_2$  dihedral angle (shown in Figure 2-1), was biased in an enhanced MD sampling strategy to generate a Boltzmann-weighted statistical ensemble of structures. This was a logical choice for determining the orientation of the nitrile bond vector within the protein-protein complex in order to calculate the magnitude and direction of  $\Delta\vec{F}$  through equation (3-14). However, close examination of the structures generated during these simulations revealed that the  $\chi_2$  probability distributions were functions of the  $\chi_1$  dihedral angle (Figure 2-1), and potentially sensitive to the  $\chi_1$  angle in the starting structure. Furthermore, examination of the structures revealed that the  $\chi_1$  dihedral was relatively fixed at a single local minimum over the course of a 3 ns simulation. This observation is not unreasonable; umbrella sampling of  $\chi_2$  was used to allow the probe to visit energetic states that have high transition barriers in that degree of freedom and thus have a low probability of being observed over the course of an unenhanced MD simulation, but

would not effectively sample other important degrees of freedom. In light of this, we decided to test a two-dimensional umbrella sampling technique to obtain probabilities as a function of  $\chi_1$  and  $\chi_2$ , and compare to previous results based on one-dimensional sampling. The goal of this work was to find an algorithm of MD sampling followed by Poisson-Boltzmann calculations that adequately reproduced our experimental data.

\*\*\*This was already introduced, but I think it does a good job here of elaborating within the context of this chapter's goals. I think moving it to the umbrella sampling section adds a lot of required background to that chapter, while here we have the background of umbrella sampling in 2 dimension and context is added.

In the work reported here, electrostatics calculations using a continuum-solvent model were done for six nitrile-labeled RalGDS proteins both in a monomeric state and when docked to each of four Rap constructs using APBS. We 1) investigated the effects of different box locations relative to the probe site and 2) compared numeric solutions to the LPBE to the reaction field method to optimize the accuracy of the calculations in APBS. To model structural fluctuations, one- and two-dimensional umbrella sampling was used to generate an ensemble of structures, appropriately weighted based on the potential of mean force landscape in the appropriate number of dimensions. We have determined that APBS is a viable strategy for calculating electrostatic fields in proteins using the reaction field method as well as for comparing changes in electrostatic fields to changes in vibrational absorption frequencies using the numeric solutions. Quantitative correlation between absolute values of the numeric LPBE calculations to experiments may be increased by careful consideration of second-stage box locations when using a two-stage focusing strategy, but in the reaction field method the second-stage box should be centered on the point of interest.

## 5.2 RESULTS AND DISCUSSION

### 5.2.1 Molecular Dynamics Sampling

One of the central challenges in any dynamic system, such as the protein-protein complexes investigated here, is determining the most important degrees of freedom relevant to the experimental observables and then sampling them appropriately and completely. Previous work in this laboratory has consistently observed that in the one-dimensional umbrella sampling of the cyanocysteine  $\chi_2$  dihedral angle, the  $\chi_1$  dihedral of the side chain very rarely moved more than a few degrees. There are two possible reasons for this observation: 1) the system had a significant preference for the starting  $\chi_1$  dihedral over all other possibilities; or 2) the system had insufficient time and energy to transition to another  $\chi_1$  dihedral with any significant probability. To test for the first possibility, the  $\chi_1$  dihedral angles of the six probes on the RalGDS monomers (as well as three additional probe sites not previously mentioned, I18C<sub>SCN</sub>, R20C<sub>SCN</sub>, and S33C<sub>SCN</sub>) were rotated to different expected minima based on the alkane-like structure of the side chain, and sampled again. The results were similar to all previous observations; the new  $\chi_1$  dihedral remained fixed and did not explore alternative alkane-like values. This result suggested that the second possibility is correct. To overcome this, we turned to two-dimensional sampling of both  $\chi_1$  and  $\chi_2$  in our MD simulations. Figure 5-1 shows a representative log-scale probability distribution for RalGDS N27C<sub>SCN</sub> and RalGDS N54C<sub>SCN</sub> in the monomeric form (Figure 5-1A and B) and docked to Rap E30D/K31E (Figure 5-1C and D), and Figure 5-2 shows the log-scale probability distributions for all 30 structures sampled. The gauche-regions of an alkane-like molecule were found to have low probabilities of being populated in both degrees of freedom, while the staggered-regions of an alkane-like molecule, at 60°, 180°, and -60°, had a high probability of being sampled. Additionally, there was not a single, dominant  $\chi_1$  angle for

all probe sites; multiple  $\chi_1$  angles were found with similar probabilities. Finally, the two-dimensional distribution was the product of two one-dimensional alkane-like distributions. These observations strongly support the use of two-dimensional sampling over one-dimensional sampling to completely explore reasonable probe orientations by revealing that the barrier between minimum energy regions is sufficiently large that, in reasonable timescales, an insignificant number of transitions will occur to represent the nitrile probes' orientations about the amide backbone accurately with sampling only one degree of freedom.

The one-dimensional  $\chi_2$  dihedral probability distributions, shown in Figure 5-3, were also compared to the two-dimensional dihedral probability distributions integrated over  $\chi_2$ , shown in Figure 5-4. In general, while they both display alkane-like structure with high probability of being found at  $180^\circ$ ,  $60^\circ$ , and  $-60^\circ$ , the one-dimensional distributions were more prone to having regions of expected-high probability being under- or over-represented. Two examples of this are the  $60^\circ$  rotamer of Rap E30D/K31E docked to RalGDS Y31C<sub>SCN</sub> and the  $-60^\circ$  rotamer of Rap E30/K31 docked to RalGDS Y31C<sub>SCN</sub>. It is not immediately clear whether those specific rotamers are actually improbable, or if they are just improbable for the chosen  $\chi_1$  dihedral angle. By examining the two-dimensional dihedral distributions, however, it becomes clear that the former was merely underrepresented in the chosen  $\chi_1$ , while the latter truly is improbable given adequate sampling. In summary, using a two-dimensional sampling strategy allows a greater degree of certainty that observed features are real and not artifacts of a relevant, unexplored degree of freedom.

Increasing the number of degrees of freedom in a simulation typically involves an exponential increase in simulation time. Since we are not looking for rare transition-state structures and instead want a representative ensemble where high-

probability conformations exceed low-probability conformations by at least six orders of magnitude, we have only performed a three-fold increase in the length of the simulation. This was sufficient to ensure that low probability regions only represented a small fraction of the overall statistical distribution, and thus did not need to be fully sampled to generate an accurate Boltzmann-weighted ensemble average.

An additional criterion of structural convergence was used based on our LPBE-based calculations of the electrostatic field at the nitrile midpoint. No matter how reliable the accuracy of these calculations are compared to experimental data (the subject of extensive discussion below), the magnitude of the calculated field at the midpoint of the nitrile bond should converge on a single value as the MD structural ensemble converges on the highest probability structures. To test this, an updated probability distribution was calculated for every  $n\Delta t$ , where  $\Delta t = 0.72$  ns and  $n$  is an integer from 1 to 80 (the total number of frames for a given trajectory), to determine the probability distribution as a function of simulation time. This was then used with the calculated numeric PBE field values to determine the Boltzmann-weighted average electrostatic field as a function of simulation time. This is shown in Figure 5-5. In all simulations, the field stopped changing significantly relative to the standard deviation after the first half of the simulation, indicating that each additional time step is not significantly altering the average field, and an adequate simulation time has been achieved.

### 5.2.2 Calculation of absolute electrostatic fields

APBS solves the LPBE by mapping force field potentials to discrete grid points inside a sample box defined by the user by means of box dimensions, grid spacing, and location. These parameters therefore play a significant role in the values of calculated electrostatic fields. Here we have examined two  $19^3$  Å<sup>3</sup> boxes with a GS of 0.119 Å/grid

point and 0.099 Å/grid point, two  $10^3$  Å<sup>3</sup> boxes with a GS of 0.062 Å/grid point and 0.052 Å/grid point, a 30 Å x 45 Å x 40 Å box with a GS of (0.234, 0.234, 0.250) Å/grid point, and a 60 Å x 45 Å x 40 Å box with a GS of (0.234, 0.234, 0.250) Å/grid point. As a reminder, boxes will be referred to as the 10 Å box, 19 Å box, 30 Å box, and 60 Å box, respectively.

Box positions were adjusted in order to minimize the volume of the box filled with the high-dielectric continuum (solvent) and maximize the volume filled with protein atoms. Table 5-1 shows the Boltzmann-weighted average of the percent of the box volume filled by protein atoms, treated as spheres with radii equal to their Amber99 van der Waals radii, for the NC $\delta$  box strategy in the left column, and the change relative to the NC $\delta$  box for RalGDS center of mass and system center of mass in the right two columns. From Table 5-1, it is clear that shifting the box towards the center of mass of RalGDS (RalGDS center of mass) universally increased the percentage of the box filled with protein atoms. In some cases this change was by a factor of two or more, such as for probe sites G28C<sub>SCN</sub>, Y31C<sub>SCN</sub>, N54C<sub>SCN</sub>, which showed an average increase in the volume of the box filled with protein atoms of 82.3% from NC $\delta$ .

Figure 5-6 and Figure 5-7 show the calculated fields for the six probes on the RalGDS monomer and the six probes on RalGDS docked to Rap E30/K31, respectively, compared to the experimental frequencies measured by Ragain *et al.* Results for RalGDS-based probes docked with Rap E30D/K31, K31E, and E30D/K31E are shown in Figure 5-8, Figure 5-9, and Figure 5-10. In each figure, the top row shows the 60 Å box, the middle row the 19 Å box, and the bottom row the 10 Å box. The grid spacing (GS) values are listed in the row titles. The left column shows the results obtained from the numeric solutions to the LPBE and the right column shows the results obtained from the reaction field method. Colors indicate which box center the data describes: red (NC $\delta$ ),

green (system center of mass), blue (RalGDS center of mass), or black (system center of mass; 60 Å boxes only). In Figure 5-6, results for the system center of mass are excluded because they are identical to the RalGDS center of mass on the nitrile-containing RalGDS monomers. The correlation coefficients between the calculated absolute field projected along the bond vector, in  $k_b T/e\text{\AA}$ , and the experimental vibrational absorption energy, in  $\text{cm}^{-1}$ , are listed in the corresponding colors in the bottom-left corners. Numeric PBE absolute fields (left) and numeric PBE relative fields (right) for the 161 grid point calculations the 30 Å box are shown in Figure 5-11, Figure 5-12, Figure 5-13, and Figure 5-14. The 161 grid point calculations and the 30 Å box are either worse than or not appreciably different results from the 193 grid point calculations or the 60 Å box calculations, and therefore are not the topic of any further discussion. The error bars have been excluded for clarity, with the vibrational absorption standard deviations ranging from 0.0 to 1.04  $\text{cm}^{-1}$  and the calculated field standard deviations ranging from 1.2 to 8.2  $k_b T/e\text{\AA}$ .

### ***Numeric solutions to the LPBE***

The left columns of Figure 5-6, Figure 5-7, Figure 5-8, Figure 5-9, and Figure 5-10 show the calculated fields obtained by taking the numeric solutions to the LPBE. In general, the 60 Å box for the numeric PBE fields cannot reliably replicate experimental results. It is possible that if the large box had a grid spacing comparable to the 19 Å or 10 Å boxes, correlation could increase; however, that would require an increase in memory usage and number of processors by approximately two orders of magnitude, and thus was not tested. The 19 Å boxes with coarse and fine grid spacing were fairly insensitive to box location, so further discussion about box location will be directed at the 10 Å boxes (bottom left) only.

From Figure 5-6, Figure 5-7, Figure 5-8, Figure 5-9, and Figure 5-10, calculations conducted in the RalGDS center of mass box had the highest correlation to experimental data, while system center of mass, with the exception of the undocked monomer (since system center of mass and RalGDS center of mass are the same), is consistently either approximately the same as NC $\delta$  (E30/K31-Figure 5-7; E30/K31E-Figure 5-9) or significantly worse (E30D/K31-Figure 5-8; E30D/K31E-Figure 5-10). The most striking example of the effect of changing the position of the box was seen for the RalGDS monomer in Figure 5-6, which has a correlation coefficient absolute value less than 0.2 for the NC $\delta$  calculation, but a correlation coefficient absolute value greater than 0.6 in the RalGDS center of mass/system center of mass calculation.

Examining the nitrile-containing RalGDS mutants docked with Rap E30/K31 (Figure 5-7), the correlation with experiment was approximately the same for system center of mass as for NC $\delta$ , while there was a significant increase for the RalGDS center of mass box location. These results can be justified by the data in Table 5-1, which shows that in most cases, shifting the box towards the system center of mass did not substantially increase the volume of the box occupied by protein, with only G28C<sub>SCN</sub> showing more than a 50% increase in occupied volume compared to NC $\delta$ . As a whole, system center of mass only increased the volume of the box occupied by protein atoms by 30.8%, while RalGDS center of mass more than doubled that with an average of 64.5% increase from NC $\delta$ .

The correlation between the absolute field calculation and the measured vibrational absorption energy on the nitrile-labeled RalGDS monomers and on the Rap E30/K31-RalGDS docked complex clearly increased as the volume of the box occupied by protein atoms increased. When nitrile probes were placed throughout the Rap mutants-RalGDS interface, the results were less straightforward. For example, Rap E30D/K31

(Figure 5-8) showed a decrease in correlation for the system center of mass strategy, which only has an average increase in percent of the box volume occupied of 19.6%. Many of the probes showed very little increase in percent box volume occupied, and in fact some showed a small decrease by moving the box. In the RalGDS center of mass strategy, the volume increases by 45.0% relative to NC $\delta$ , yet there is only a small increase in correlation in the coarser 0.062 Å/grid point GS and a small decrease in correlation in the finer 0.052 Å/grid point GS strategy. In Rap E30/K31E (Figure 5-9) and E30D/K31E (Figure 5-10), the RalGDS center of mass strategy increased the percent volume occupied by protein atoms by 47% and 54% respectively, yet also showed a decrease in correlation, relative to NC $\delta$ , in the finest grid calculations.

To summarize, increasing the percent of the second-stage box explicitly filled with protein atoms by moving the focusing center away from the object of interest can be both beneficial and deleterious to the accuracy of the calculation. On one hand, it may increase the correlation to experiment by focusing on a location as remote as possible from the boundary between low and high dielectric media. However, occasionally moving the focusing box may also cause the loss of information that is vital to the calculation of the electrostatic field by shifting atoms previously treated explicitly in the second-stage calculation outside of the box, where they are treated implicitly as a contribution to the boundary condition. Ensign *et al.*<sup>43</sup> previously showed that in a similar system, Ras bound to various thiocyanate-labeled RalGDS, different probe locations were most sensitive to different contributions to the electrostatic field—solvent, charges on other RalGDS amino acids, charges on Ras, etc. Here, no consideration was given to how translating the box affected which protein atoms were inside the box, only that the nitrile stayed within the confines. In the monomer, losing nearest-neighbor residues was unlikely; the protein can be adequately described as a sphere with the probe

on the surface, and translating a cubic box towards the center of mass of a sphere would not remove nearby residues from the second-stage calculation. In the docked complexes, however, moving the box towards the system center of mass would move it away from RalGDS and thus remove nearby residues on RalGDS, while moving towards the RalGDS center of mass would move it away from Rap and thus remove nearby residues on Rap. In these cases, the loss of important, nearby residues partially or entirely negates the benefit of removing part of the dielectric discontinuity. Increasing the amount of protein in the box does not automatically help the calculation if, in doing so, the most important parts of the system are removed from the calculation. The selection of a focusing strategy and location must therefore be considered carefully when using focusing in these contexts, and the most ideal focusing strategy will most likely be unique to the chemical identity of the system under investigation and the desired information.

### ***Reaction field method solutions to the LPBE***

The right column of Figure 5-6, Figure 5-7, Figure 5-8, Figure 5-9, and Figure 5-10 show the calculated fields obtained by using the reaction field method solutions to the LPBE. In general, there is no appreciable difference between the box sizes or GS used; all perform comparably well. This is not a particularly surprising result given the primary contribution to the electrostatic field (magnitudes  $\geq 600 k_b T/e\text{\AA}$ ) should be the protein atoms and not the SRF (magnitudes  $\leq 10 k_b T/e\text{\AA}$ ).

From Figure 5-6, Figure 5-7, Figure 5-8, Figure 5-9, and Figure 5-10, calculations in the NC $\delta$  box showed either the highest correlations (Monomer-Figure 5-6; E30/K31-Figure 5-7, E30/K31E-Figure 5-9) or comparable correlations (E30D/K31-Figure 5-8, E30D/K31E-Figure 5-10) to the system center of mass and the RalGDS center of mass. Because the exact analytic protein field was used in these calculations, only the SRF is

affected by changing the box location. In this regard, these calculations are less sensitive to moving atoms from inside to outside the second-stage box than in the numeric PBE fields. Moving the box, regardless of how it affects the percent volume of the box occupied by atoms, does not improve the correlation to experiment. The best results were obtained by centering the second-stage box on the location of interest *or* by using a box large enough to include the entire protein, such as the 60 Å box used here.

#### ***Comparing the numeric solutions of the LPBE to the reaction field method solutions of the LPBE***

On the one hand, the 60 Å box performed significantly better in the reaction field method when using the numeric PBE solutions. Using the NC $\delta$  box, the reaction field also nearly always reproduced experimental results better than the numeric PBE solutions. On the other hand, intelligently changing the second-stage box location can cause the numeric solutions to outperform the reaction field (monomer-Figure 5-6; E30/K31-Figure 5-7, E30/K31E-Figure 5-9). Given the unreliability of predicting how changing the box location will alter the correlation with experiment previous discussed, we conclude that the most consistent, reliable method for calculating protein electrostatic fields is to use the reaction field method and a high grid point density near the location of interest.

#### **5.2.3 Calculations of relative electrostatic field**

The VSE experiment defined by equation (5-1) compares changes in electric fields around the probe,  $\Delta\vec{F}$ , to changes in the absorption energy of that probe,  $\Delta\nu$ . Because of this, we can make a direct comparison between computational predictions and experimental results is obtained from the difference in results between the docked Rap-RalGDS complex and the monomeric RalGDS. Figure 5-15 shows the change in the

calculated electric field versus change in absorption frequency due to docking the SCN-labeled RalGDS mutant to Rap E30/K31, formatted identically to the previously presented figures. Figure 5-16, Figure 5-17, and Figure 5-18 show the change in calculated fields versus the change in absorption frequencies for RalGDS docked to Rap E30D/K31, Rap E30/K31E, and Rap E30D/K31E respectively.

### ***Numeric relative field solutions to the LPBE***

There are three interesting features in numeric relative fields in the left column of Figure 5-15 that are also present for the other Rap constructs. The first and most pronounced is the increase in correlation between experiment and calculation upon subtracting the monomer field and frequencies values compared to any of the absolute field calculations. Almost universally, the correlation to experiment increased in the difference calculations compared to the results of the absolute calculation, even when including information from data sets with poor correlation to begin with, such as the NC $\delta$  monomer calculations (Figure 5-6) or the system center of mass Rap E30D/K31E calculations (Figure 5-10). This was even true for the 19 Å boxes, which showed no significant improvement on moving the center of the box. This suggests that errors introduced into the absolute calculations, either from non-physical effects included in the LPBE or important physical effects excluded from consideration, cancel out reliably well in the difference calculations. The absolute calculations manage to capture some degree of the physics of the system and can be improved by logical decisions about the position and size of the calculation box, but taking the difference calculation significantly improves the correlation to experimental results.

The second interesting feature is the significant improvement in correlation seen in the 19 Å boxes compared to the absolute calculations. In the absolute calculations

(Figure 5-6, Figure 5-7, Figure 5-8, Figure 5-9, and Figure 5-10), the 19 Å boxes had very poor correlations to experiment compared to the absolute calculations (typically  $|r| \leq 0.4$ ), and could not be noticeably improved by changing the size or location of the calculation box. In the difference calculations, however, the calculated correlations ranged from  $r = 0.628$  (NCδ Rap E30D/K31-Figure 5-6) to 0.843 (RalGDS center of mass Rap E30D/K31E-Figure 5-18). When these results are compared to the smallest and largest magnitude correlation coefficients for the 10 Å box of 0.283 and 0.916 (RalGDS center of mass and NCδ respectively for Rap E30D/K31-Figure 5-6), it is apparent that although the smaller boxes have the possibility for a higher correlation to experiment, the larger boxes are still quite good. The 19 Å boxes are also significantly less sensitive to box position than the 10 Å boxes, discussed previously and shown in Figure 5-6, Figure 5-7, Figure 5-8, Figure 5-9, and Figure 5-10, suggesting that the large box would be ideal to use when computational time is short or expensive, since refining the box position is fairly insignificant and the correlations in the difference calculation are still comparable to or occasionally better than the 10 Å boxes. In general, correlations between the changes in fields compared to the changes in frequencies are consistently higher than the correlations for the absolute calculations and less sensitive to the specific Rap mutant.

The 10 Å boxes, however, are still significantly more sensitive to the box position than the 19 Å boxes. There is no general “best” box location, however, and for different Rap constructs, different box positions yield the highest correlation. On average, however,  $\text{NC}\delta > \text{system center of mass} > \text{RalGDS center of mass}$ , which is very unlike the absolute field calculation’s hierarchy of  $\text{RalGDS center of mass} > \text{NC}\delta > \text{system center of mass}$ . This is further evidence that very careful consideration about the position of multiple-state boxes is necessary when designing LPBE strategies.

The third interesting feature comes from examining the 60 Å boxes, which showed a greater correlation with experiment than in the previous, absolute field calculations for RalGDS bound to Rap E30/K31 (Figure 5-6). In fact, the 60 Å boxes only show an increased correlation with experiment in the E30/K31 (Figure 5-15), E30D /K31(Figure 5-16), and the double Rap mutant E30D/K31E (Figure 5-18) upon subtraction of the monomer-field, which is inconsistent with the observation that correlations are better when examining the differences between systems. This is likely due to the coarseness of the calculation resulting in too few grid points in the proximity of the bond vector to interpolate a potential along. For example, the coarsest 19 Å box calculation had a grid GS of 0.119 Å, while the 60 Å boxes have a GS of approximately 0.25 Å. This means that, in the best-case-scenario where the bond lies exactly parallel to one of the box sides, the number of grid points to interpolate a potential from for each of the 10 dummy atoms is fewer than the number of dummy atoms by nearly a factor of four (CN bond length is approximately 1.14 Å). Since the field is related to the gradient of the potential, a coarser GS gives an average field over a larger area, and thus resulting in a coarser field.

### ***Reaction field method relative fields***

The right columns of Figure 5-15, Figure 5-16, Figure 5-17, and Figure 5-18 show the calculated change in field, computed using the reaction field method, compared to the experimental change in field of the nitrile-labeled RalGDS upon docking to the various Rap constructs. As in the absolute field calculations, the box size and GS were largely irrelevant and the NC $\delta$  box locations gave the highest correlation to results. However, with the exception of Rap E30D/K31E (Figure 5-18), all correlations for the

difference calculations were lower than the absolute field correlations. We did not observe the sort of cancellation of error observed using the numeric solutions.

### ***Comparing relative fields calculated from the numeric method to the reaction field method***

In all cases of the NC $\delta$  box locations, the numeric method relative correlations are better than the reaction field method relative correlations. Regardless of correlation in the absolute calculations, the numeric method gives changes in fields that are well correlated to experiment, indicating that the sources of error in the absolute calculations being subtracted are also being subtracted, i.e.: the cause of low correlation in the NC $\delta$  monomer (Figure 5-6) and the cause of low correlation in the NC $\delta$  Rap E30/K31E (Figure 5-9) are due to the same source of systematic error in the calculation, which is removed upon subtracting the two values. This is contrary to the reaction field method, where the cause of low correlation in the NC $\delta$  monomer (Figure 5-6) is in no way related to a cause of a magnitude of correlation less than one in NC $\delta$  Rap E30/K31E (Figure 5-9), and no such subtraction of error occurs.

A reasonable hypothesis is the source of error is the low-resolution of the numeric solutions, which is removed upon taking the difference in the numeric solutions. In the reaction field method, however, there is a LPBE numeric solution term,  $F_{reference}^{(2)}$ , added back to the final answer:

$$\begin{aligned}
 F_{numeric} &= F_{protein} \\
 F_{reaction\ field} &= F_{protein} - F_{reference} + F_{Coulomb} \\
 \Delta F_{numeric} &= F_{protein}^{(2)} - F_{protein}^{(1)} \\
 \Delta F_{reaction\ field} &= F_{protein}^{(2)} - F_{reference}^{(2)} + F_{Coulomb}^{(2)} - \left( F_{protein}^{(1)} - F_{reference}^{(1)} + F_{Coulomb}^{(1)} \right) \quad (5-2) \\
 \Delta F_{reaction\ field} &= +F_{protein}^{(2)} - F_{protein}^{(1)} - F_{reference}^{(2)} + F_{reference}^{(1)} + F_{Coulomb}^{(2)} - F_{Coulomb}^{(1)} \\
 \Delta F_{reaction\ field} &= \Delta F_{numeric} - F_{reference}^{(2)} + F_{reference}^{(1)} + F_{Coulomb}^{(2)} - F_{Coulomb}^{(1)}
 \end{aligned}$$

In equation (5-2), the superscripts indicate the index of the system,  $F_{protein}$  is the field obtained from the numeric solution to the LPBE with  $\epsilon_{interior} \neq \epsilon_{exterior}$ ,  $F_{reference}$  is the field obtained from the numeric solution to the LPBE with  $\epsilon_{interior} = \epsilon_{exterior}$ , and  $F_{Coulomb}$  is the analytic Coulomb field due to the force field charges on the protein atoms. As the final line of equation (5-2) clearly shows, the  $\Delta F_{reaction\ field}$  includes the numeric field as well as additional terms that include two low-resolution LPBE solutions. It may be that error removed in the  $\Delta F_{numeric}$  term is reintroduced by the two  $F_{reference}$  terms. In short, the best results are obtained from a single subtraction between a single pair of LPBE solutions, as in the absolute fields from the reaction field method and in the relative fields in the numeric PBE method.

It should be noted that other works have demonstrated the importance of water in the nitrile stretching mode,<sup>117</sup> which we appear to have treated in an entirely implicit manner when calculating electrostatic fields. This is not entirely true, because we treated the water explicitly in the MD sampling procedure. Hydrogen bonding with the nitrile nitrogen can lead to carbon-nitrogen bond length changes as well as increased stability of hydrogen-bonding probe orientations, which is accounted for in the MD electrostatic force calculations during sampling. Orientation stabilizations are included in our WHAM PMF while bond-length changes are included in the structural information used in electrostatic calculations. So although we have implicit solvent in our APBS electrostatic calculations, solvent effects on the nitrile vibrational frequency are not ignored.

### 5.3 CONCLUSION

Our results show that although the one-dimensional umbrella sampling showed the expected alkane-like torsional distributions along the  $\chi_2$  dihedral coordinate, the probe orientations in space are clearly functions of both the  $\chi_2$  dihedral as well as the  $\chi_1$

dihedral angle. Therefore, at minimum both degrees of freedom must be sampled in order to construct a complete ensemble of probable structures. These structures were used to calculate solutions to the LPBE over various grid spacings (GS) and box locations as well as using two different computational methods. It was found that APBS correlated to experimental absorption energies using an unsophisticated, single protein dielectric value in both the absolute and difference field calculations using both methods. It was observed that the absolute field calculations using the reaction field method yielded higher correlation to experimental results than the numeric method, while the relative field calculations showed higher correlation when using the numeric method. This indicates that some limiting aspect of solution to the LPBE is removed upon taking the difference between two solutions, but if we subtract another pair of LPBE solutions, we lose the cancellation of error previously seen. Regardless, the observed correlation between calculated fields and experimental absorption energies indicate that, although electric fields play a role in the nitrile absorption frequency, there are other physical properties involved that are not accounted for in our model. In short, the reaction field method should be used to calculate *absolute* protein electrostatic fields, while the numeric method should be used to calculate *differences* in protein electrostatic fields.

For all systems, the location of the second-stage calculation box was also examined. It was found that, in both the absolute as well as the difference field calculations using the numeric solutions to the LPBE, moving the box to explicitly include more protein and less implicit solvent tended to increase the correlation with experimental frequency values only for the smaller 10 Å boxes; the 19 Å boxes were invariant to box location. Among the absolute and difference field calculations, no agreement was found as to which box strategy yielded the best correlation to experimental data, indicating that, when using a small second-stage box in an APBS

focusing strategy, significant care must be taken in deciding the location of the box. However, the location of the second-stage box, while important, is consistently best when centered on the region of interest *or* when large enough to fully encompass the entire system.

These results demonstrate that solving the LPBE through APBS is a viable strategy for quantifying protein electrostatic fields and changes in protein electrostatic field due to docking. For the best results, careful consideration of the chemical identity and structure of the system being studied and the purpose of the calculations is necessary to determine which strategy to use.

Table 5-1: Percent Volume of Second-Stage APBS Box Filled with Protein Atoms

Probe/Construct	Volume (% of box)		$\Delta(\text{Volume})$ (% of box) <sup>a</sup>			
	NC $\delta$		system center of mass - NC $\delta$		RaLGDS center of mass - NC $\delta$	
	10 Å	19 Å	10 Å	19 Å	10 Å	19 Å
N27C <sub>SCN</sub> /Monomer	42 ± 4	39 ± 3	30 ± 5	39 ± 4	-- <sup>b</sup>	-- <sup>b</sup>
G28C <sub>SCN</sub> /Monomer	24 ± 3	22 ± 2	29 ± 4	39 ± 4	-- <sup>b</sup>	-- <sup>b</sup>
N29C <sub>SCN</sub> /Monomer	63 ± 3	52 ± 2	22 ± 4	33 ± 3	-- <sup>b</sup>	-- <sup>b</sup>
Y31C <sub>SCN</sub> /Monomer	36 ± 3	39 ± 3	31 ± 4	44 ± 4	-- <sup>b</sup>	-- <sup>b</sup>
K32C <sub>SCN</sub> /Monomer	45 ± 3	47 ± 2	34 ± 3	34 ± 3	-- <sup>b</sup>	-- <sup>b</sup>
N54C <sub>SCN</sub> /Monomer	28 ± 3	30 ± 2	30 ± 4	35 ± 3	-- <sup>b</sup>	-- <sup>b</sup>
N27C <sub>SCN</sub> /WT	32 ± 3	32 ± 2	15 ± 5	30 ± 3	30 ± 4	36 ± 4
G28C <sub>SCN</sub> /WT	27 ± 3	34 ± 3	20 ± 4	33 ± 4	28 ± 4	34 ± 4
N29C <sub>SCN</sub> /WT	65 ± 3	70 ± 2	6.4 ± 3.7	9.1 ± 2.5	20 ± 4	17 ± 3
Y31C <sub>SCN</sub> /WT	49 ± 2	60 ± 2	21 ± 3	23 ± 3	24 ± 4	22 ± 3
K32C <sub>SCN</sub> /WT	63 ± 2	80 ± 1	4.6 ± 3.3	1.1 ± 1.8	14 ± 3	6.5 ± 1.9
N54C <sub>SCN</sub> /WT	38 ± 3	48 ± 3	2.5 ± 4.2	24 ± 3	32 ± 4	31 ± 3
Y31C <sub>SCN</sub> /E30D	80 ± 3	64 ± 2	-0.66 ± 3.75	13 ± 2	5.9 ± 3.6	26 ± 3
N27C <sub>SCN</sub> /E30D	28 ± 3	37 ± 2	19 ± 4	34 ± 3	21 ± 4	28 ± 4
G28C <sub>SCN</sub> /E30D	55 ± 3	59 ± 3	9.3 ± 4.1	19 ± 4	24 ± 4	25 ± 4
N29C <sub>SCN</sub> /E30D	46 ± 3	56 ± 2	20 ± 4	22 ± 3	27 ± 4	27 ± 3
K32C <sub>SCN</sub> /E30D	63 ± 2	80 ± 1	-0.9 ± 3.16	1.1 ± 1.7	16 ± 3	5.5 ± 1.7
N54C <sub>SCN</sub> /E30D	44 ± 2	45 ± 2	-4.0 ± 3.6	27 ± 3	26 ± 4	33 ± 3
N27C <sub>SCN</sub> /K31E	39 ± 3	42 ± 3	20 ± 5	30 ± 4	30 ± 5	39 ± 4
G28C <sub>SCN</sub> /K31E	35 ± 3	42 ± 3	25 ± 5	34 ± 4	20 ± 5	28 ± 4
N29C <sub>SCN</sub> /K31E	56 ± 3	68 ± 2	3.6 ± 4.5	10 ± 2	23 ± 4	19 ± 3
Y31C <sub>SCN</sub> /K31E	51 ± 3	62 ± 2	25 ± 4	22 ± 2	15 ± 4	17 ± 3
K32C <sub>SCN</sub> /K31E	72 ± 2	84 ± 1	3.5 ± 3	-0.51 ± 1.89	5.6 ± 3	3.3 ± 1.9
N54C <sub>SCN</sub> /K31E	26 ± 2	32 ± 2	-7.0 ± 3.3	29 ± 3	18 ± 4	29 ± 3
N27C <sub>SCN</sub> /E30D						
K31E	40 ± 4	41 ± 3	16 ± 6	25 ± 4	30 ± 5	35 ± 4
G28C <sub>SCN</sub> /E30D						
K31E	32 ± 3	42 ± 3	26 ± 4	36 ± 4	18 ± 4	25 ± 3
N29C <sub>SCN</sub> /E30D						
K31E	62 ± 3	70 ± 2	3.8 ± 3.7	13 ± 3	19 ± 4	19 ± 3
Y31C <sub>SCN</sub> /E30D						
K31E	42 ± 3	54 ± 2	21 ± 5	22 ± 3	26 ± 5	26 ± 4
K32C <sub>SCN</sub> /E30D						
K31E	74 ± 3	78 ± 1	-6.7 ± 3.3	0.21 ± 1.86	14 ± 3	11 ± 2
N54C <sub>SCN</sub> /E30D						
K31E	36 ± 2	41 ± 2	10 ± 3	28 ± 3	28 ± 4	34 ± 3

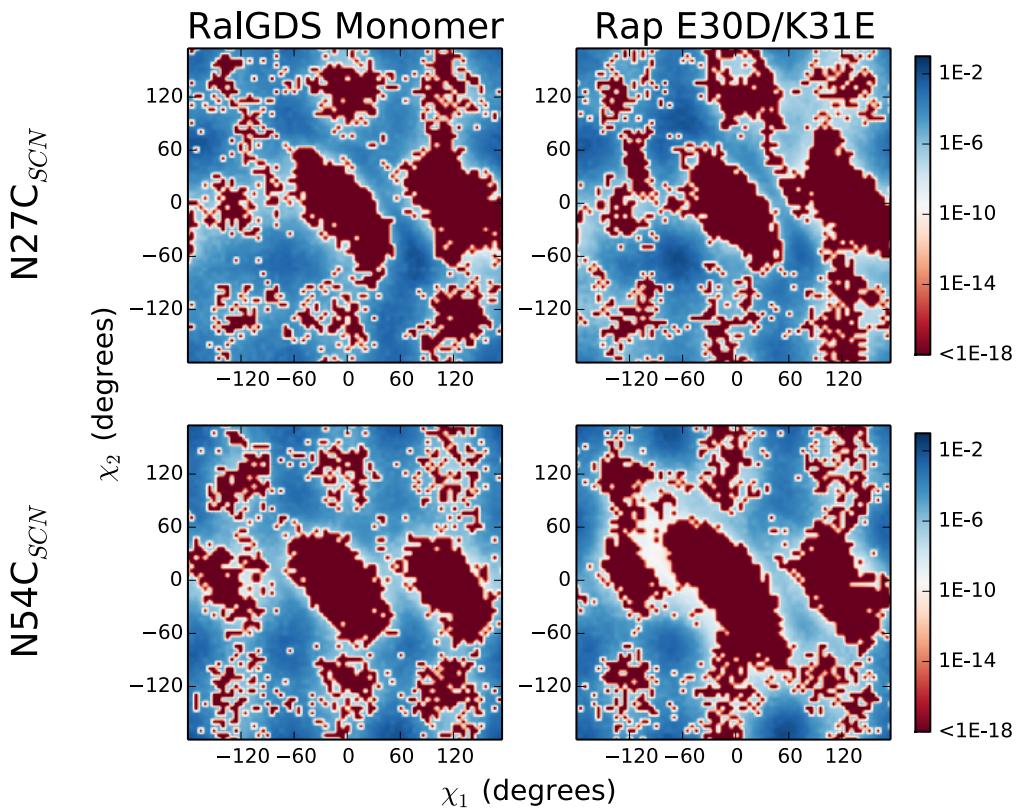


Figure 5-1: Representative 2D Dihedral Probability Distributions

A representative two-dimensional probability distribution for  $N27C_{SCN}$  (top row) and  $N54C_{SCN}$  (bottom row) in the monomeric state (left column) and bound to Rap E30D/K31E (right column). On the x-axis is the  $\chi_1$  dihedral angle and on the y-axis is the  $\chi_2$  dihedral angle. The log scale shown on the right allows easy visualization of lower-probability states.

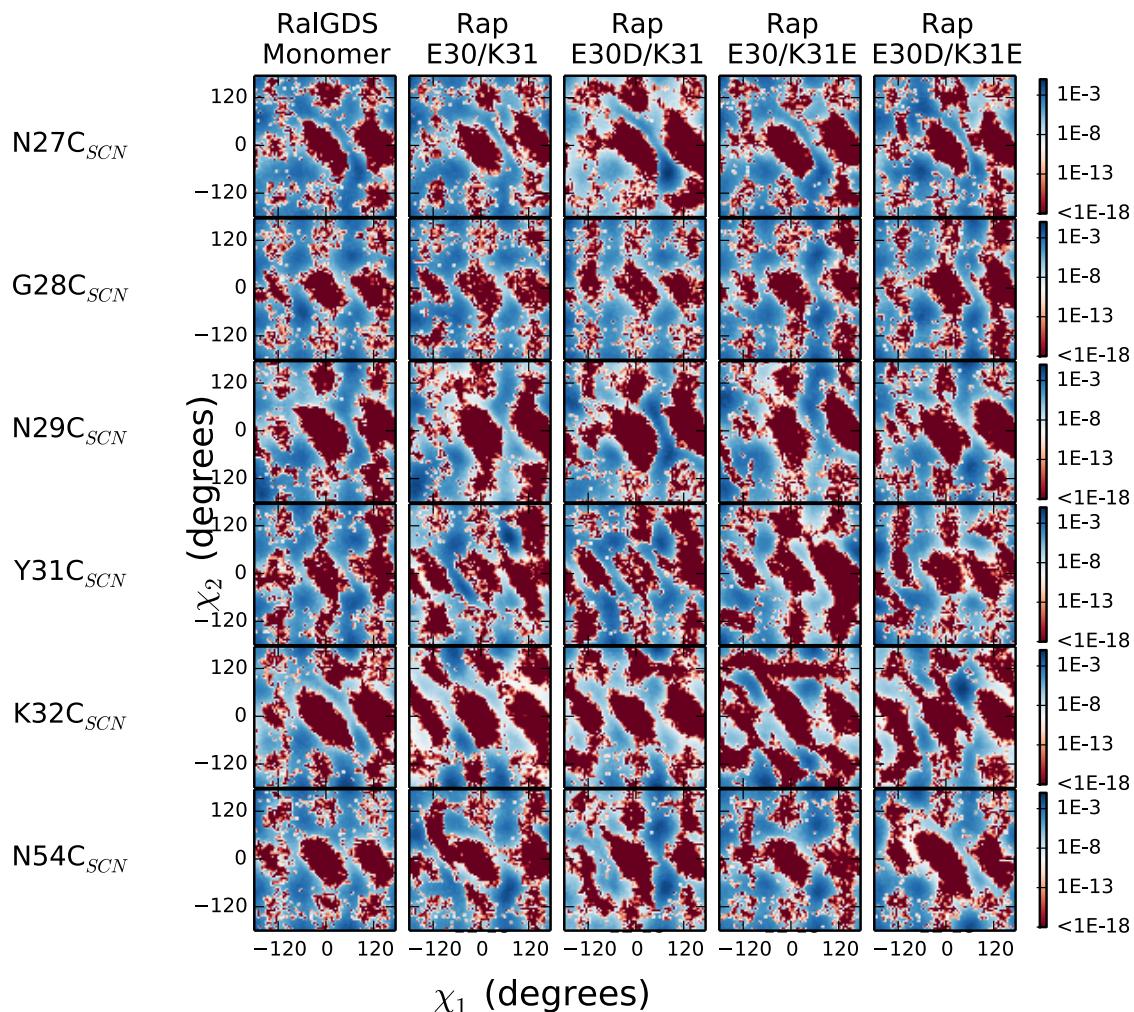


Figure 5-2: 2D Dihedral Probability Distributions

Two-dimensional probability distribution of the six probes on monomeric RalGDS and docked to the various Rap constructs. On the  $x$ -axis is the  $\chi_1$  dihedral angle and on the  $y$ -axis is the  $\chi_2$  dihedral angle. Each row represents an individual nitrile-labeled RalGDS mutant (identified to the left of the figure) either in a monomeric form or docked with a Rap construct (identified on the top of the figure). The log scale shown on the right allows easy visualization of lower-probability states.

## $\chi_2$ Probability Distributions from 1-Dimensional Umbrella Sampling

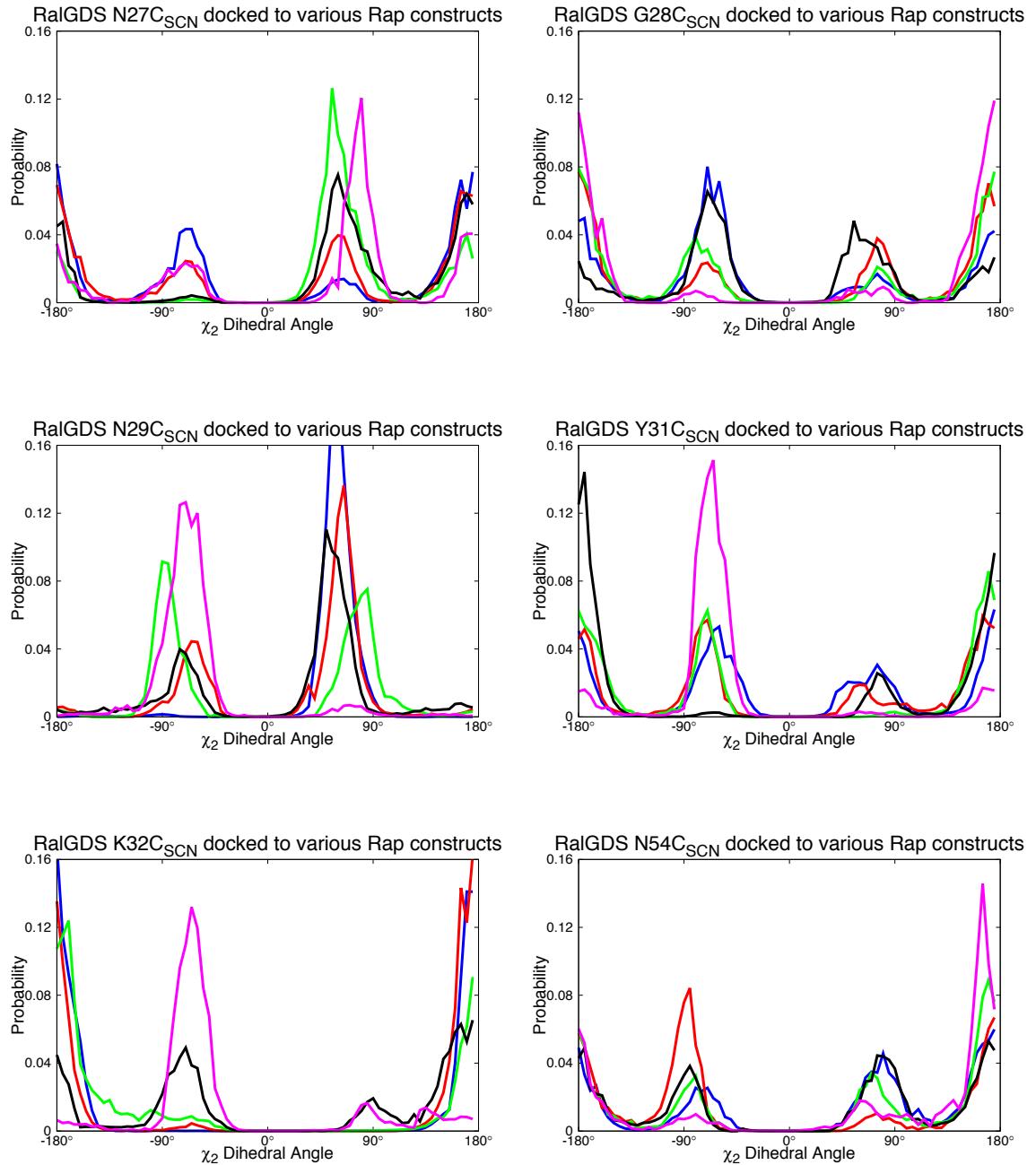


Figure 5-3: 1D Umbrella Sampling  $\chi_2$  Dihedral Distributions

The  $\chi_2$  probability distributions of nitrile-labeled RalGDS mutants in various states determined by one-dimensional umbrella sampling. Magenta: RalGDS monomer; black: Rap E30/K31; red: Rap E30D/K31; green: Rap E30/K31E; blue: Rap E30D/K31E.

## $\chi_2$ Probability Distributions from 2-Dimensional Umbrella Sampling

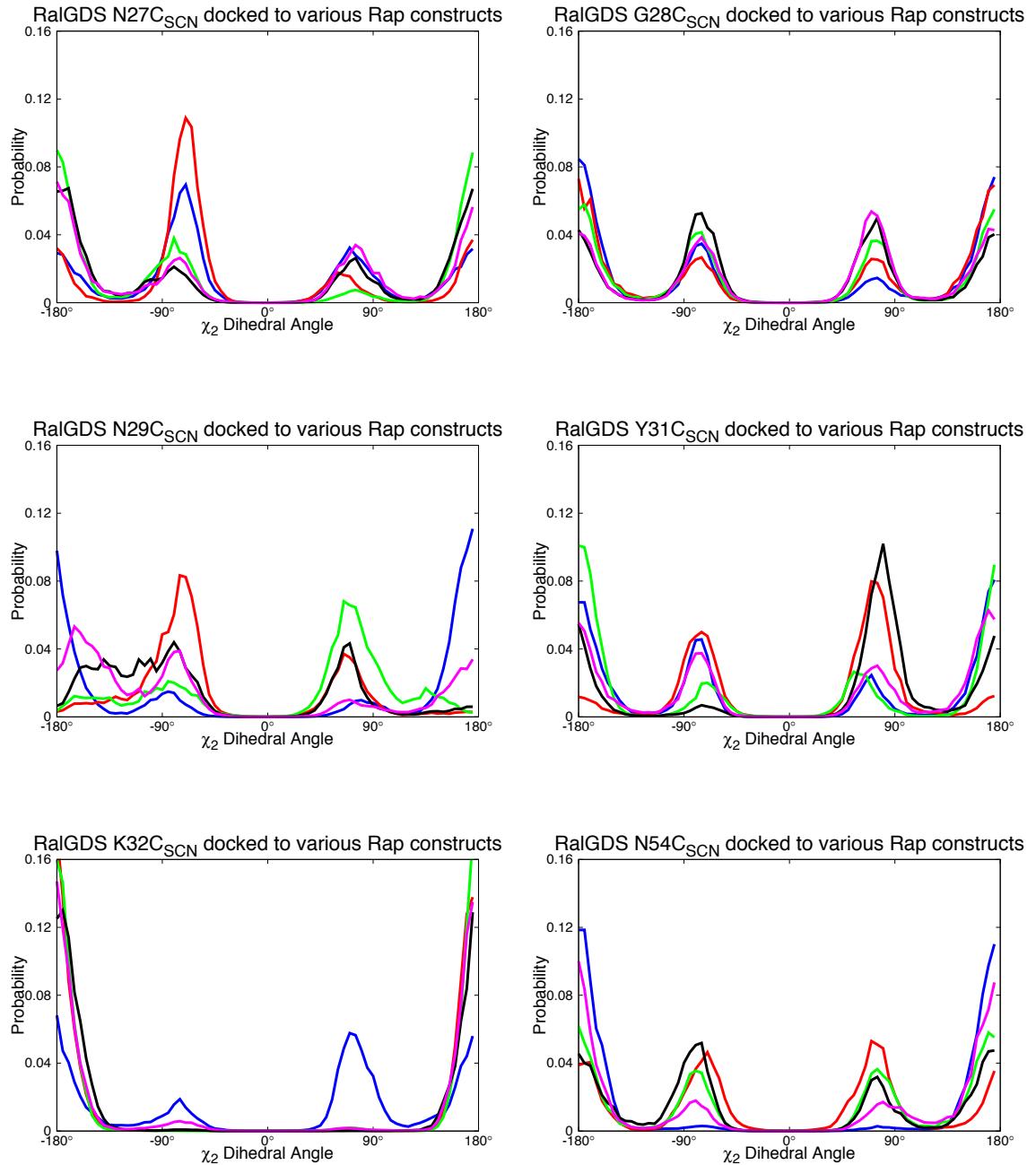


Figure 5-4: 2D Umbrella Sampling  $\chi_2$  Dihedral Distribution

The  $\chi_2$  probability distributions of nitrile-labeled RalGDS mutants in various states determined by two-dimensional umbrella sampling. Magenta: RalGDS monomer; black: Rap E30/K31; red: Rap E30D/K31; green: Rap E30/K31E; blue: Rap E30D/K31E.

## Boltzmann-weighted average electrostatic field as a function of sampling time

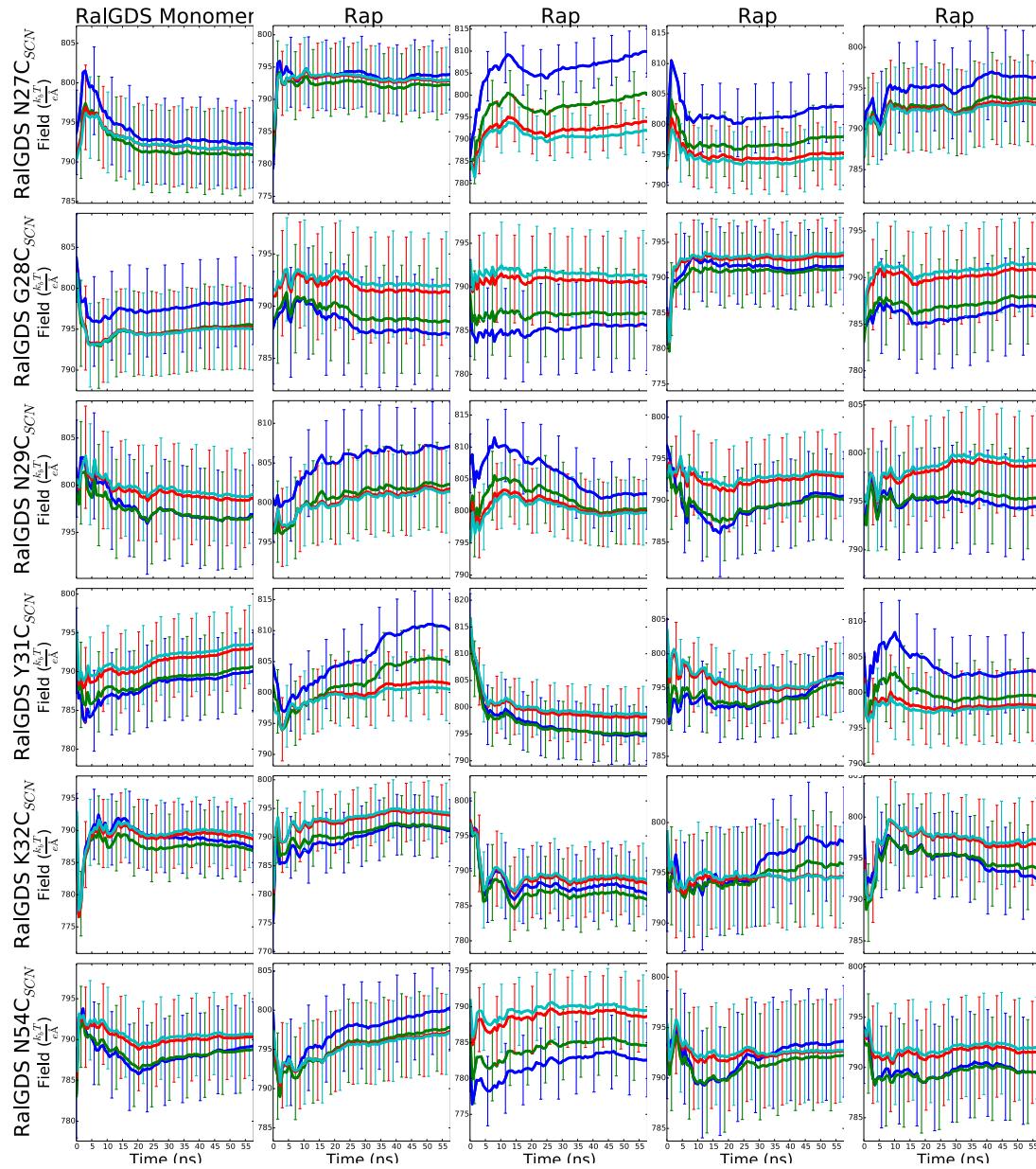


Figure 5-5: Field Convergence Test

The Boltzmann-weighted average field as a function of simulation time. Blue: GS = 0.119 Å/grid point; green: GS = 0.099 Å/grid point; red: GS = 0.062 Å/grid point; cyan: GS = 0.052 Å/grid point. Error bars show one standard deviation from calculations conducted on the Boltzmann-weighted ensemble of snapshots.

### SCN-labeled RaIGDS monomers

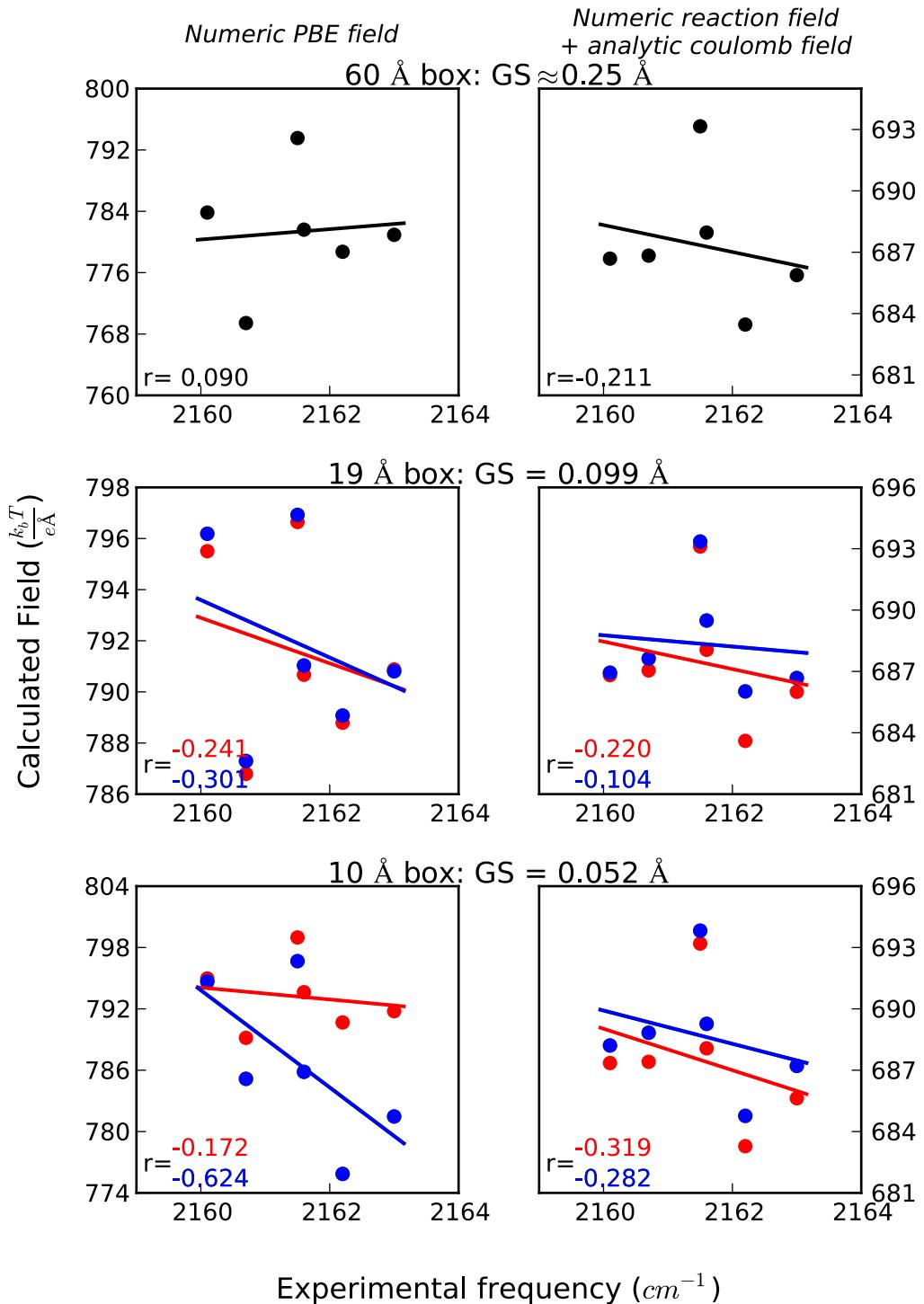


Figure 5-6: Absolute Field Plots for Various Box Locations - Ral Monomer

The absolute field calculations from solving numerically solving one LPBE (left) and from summing the numeric reaction field obtained from two LPBE calculations with the analytic Coulomb field (right) compared to the experimental frequencies for various nitrile probe locations on the RalGDS monomer. Black: system center of mass ( $60 \text{ \AA}$  boxes); red: NC $\delta$ ; blue: system center of mass = RalGDS center of mass. Correlation coefficients are labeled in the corresponding color on each plot.

### SCN-labeled RalGDS mutants docked with WT Ral

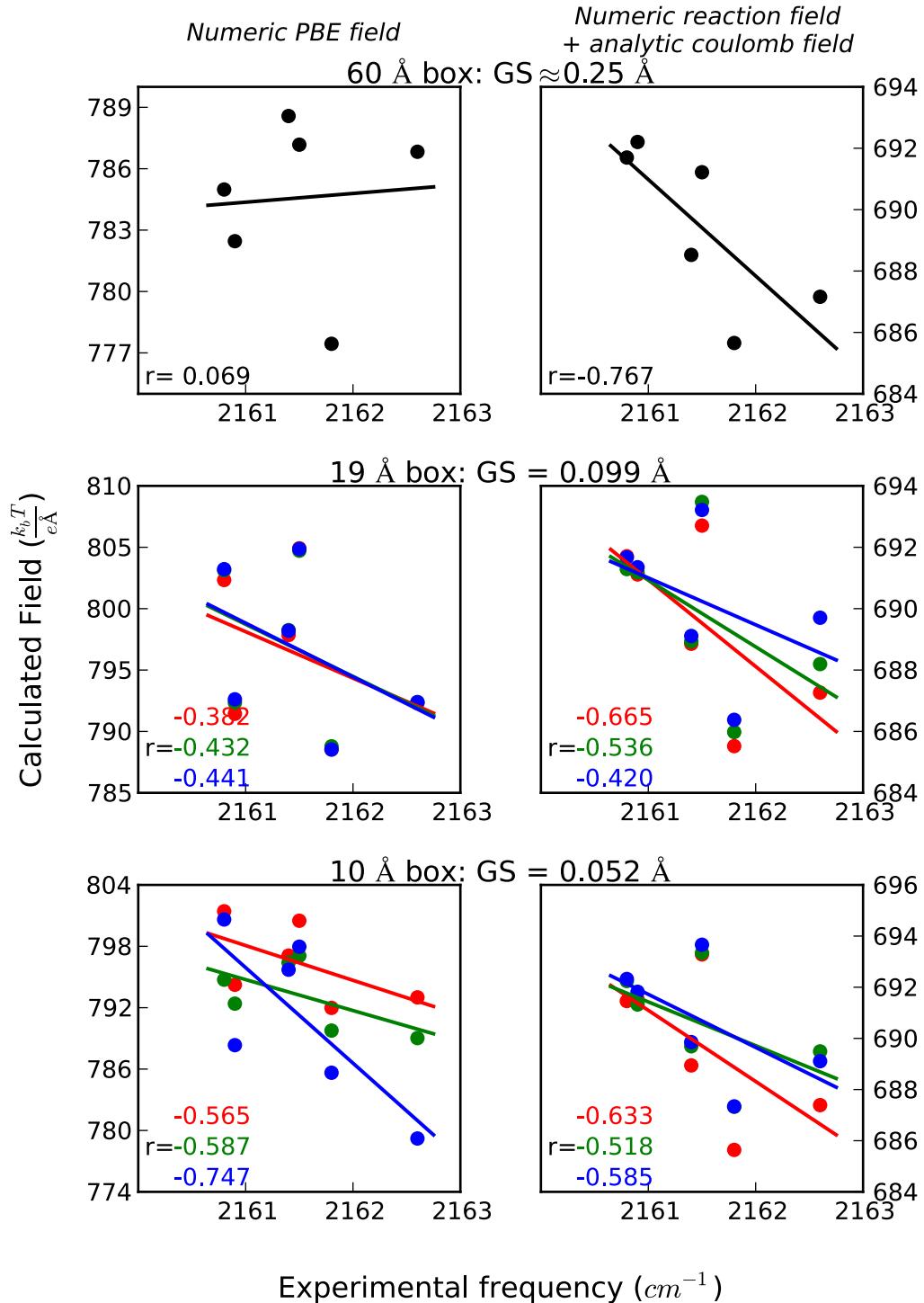


Figure 5-7: Absolute Field Plots for Various Box Locations – Rap E30/K31

The absolute field calculations from solving numerically solving one LPBE (left) and from summing the numeric reaction field obtained from two LPBE calculations with the analytic Coulomb field (right) compared to the experimental frequencies for various nitrile probe locations on RalGDS docked to Rap E30/K31. Black: system center of mass (60 Å boxes); red: NC $\delta$ ; green: system center of mass; blue: RalGDS center of mass. Correlation coefficients are labeled in the corresponding color on each plot.

## SCN-labeled RaIGDS mutants docked with Rap E30D

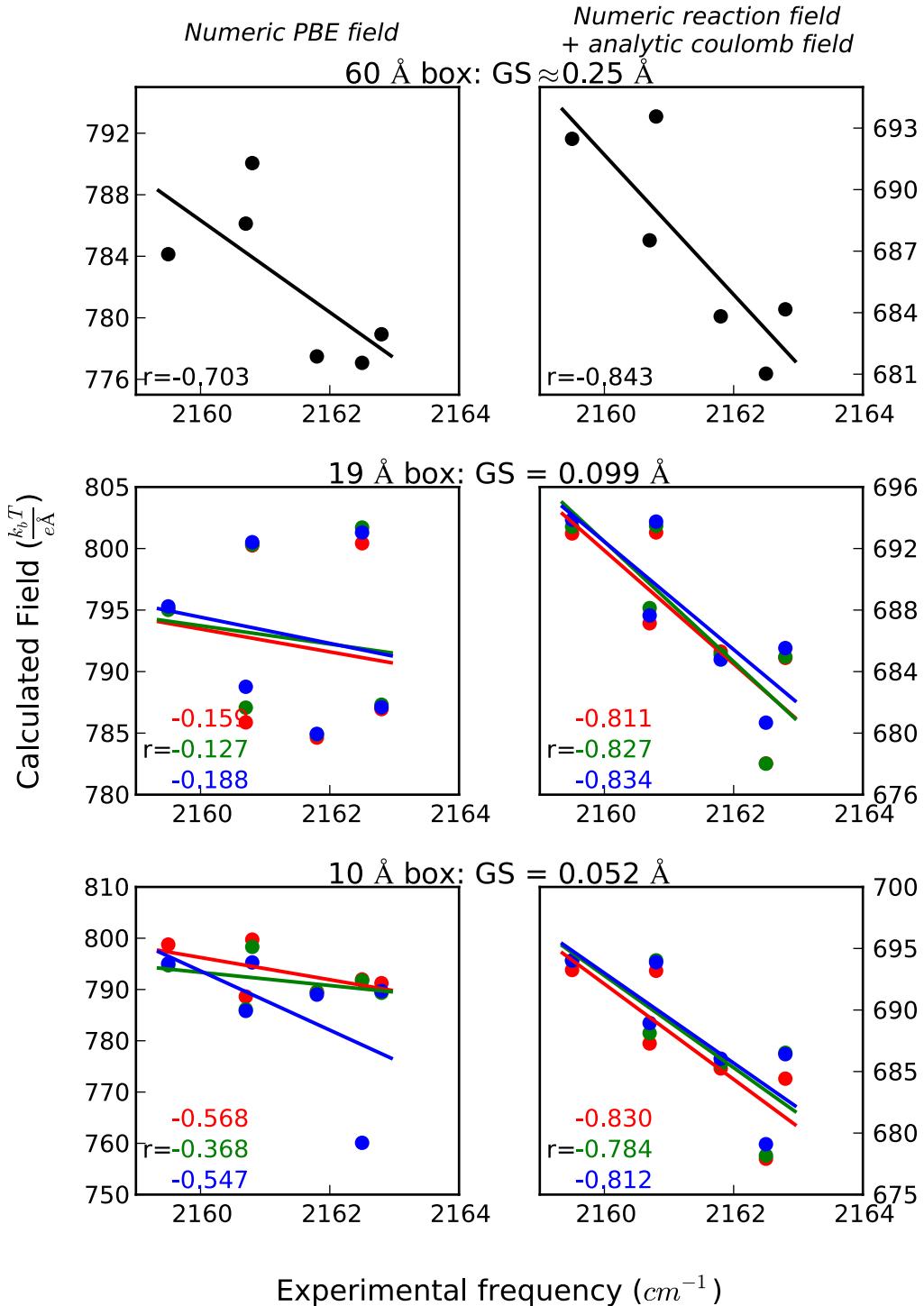


Figure 5-8: Absolute Field Plots for Various Box Locations – Rap E30D/K31

The absolute field calculations from solving numerically solving one LPBE (left) and from summing the numeric reaction field obtained from two LPBE calculations with the analytic Coulomb field (right) compared to the experimental frequencies for various nitrile probe locations on RalGDS docked to Rap E30D/K31. Black: system center of mass (60 Å boxes); red: NC $\delta$ ; green: system center of mass; blue: RalGDS center of mass.

### SCN-labeled RaIGDS mutants docked with Rap K31E

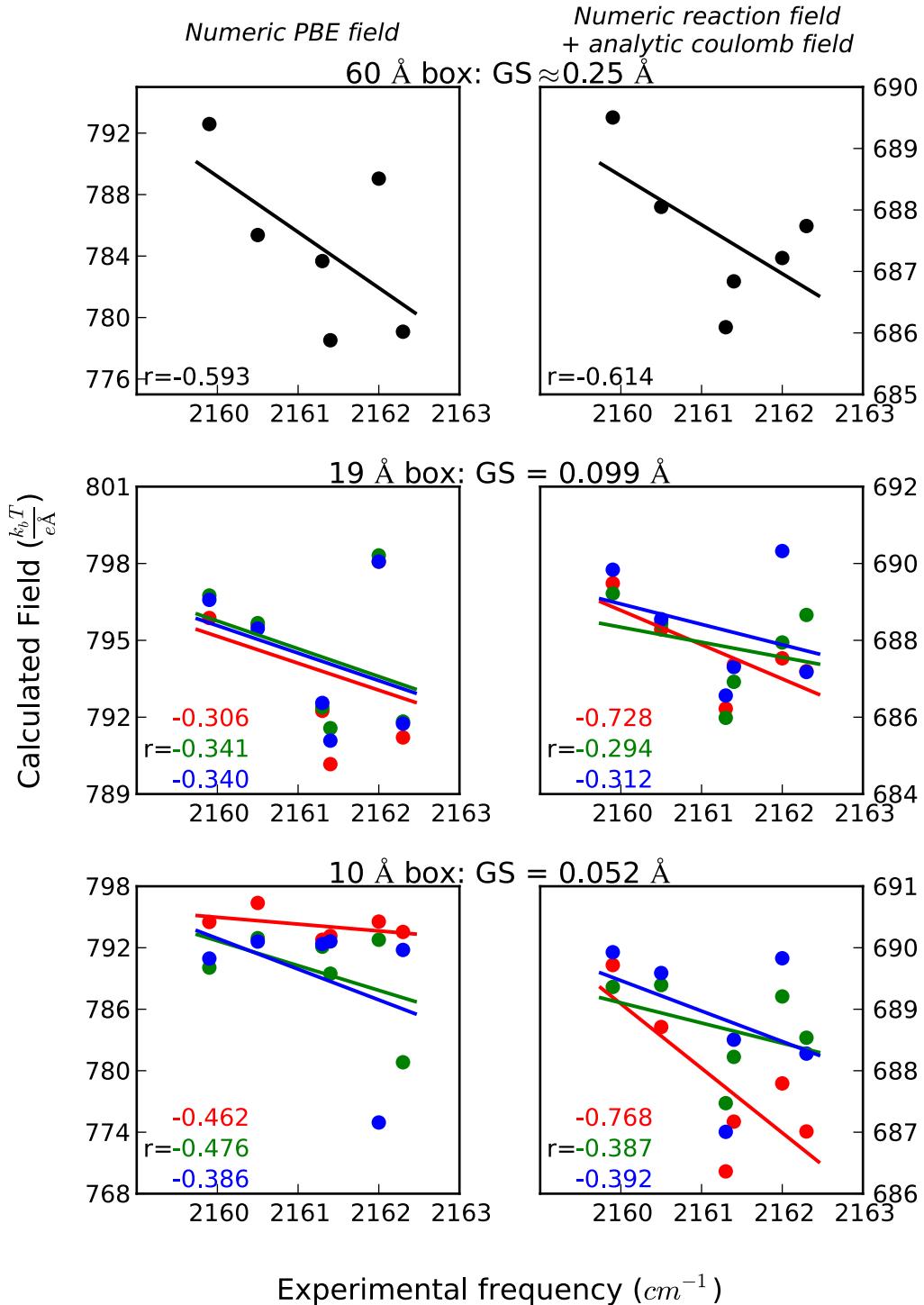


Figure 5-9: Absolute Field Plots for Various Box Locations – Rap E30/K31E

The absolute field from solving numerically solving one LPBE (left) and from summing the numeric reaction field obtained from two LPBE calculations with the analytic Coulomb field (right) calculations compared to the experimental frequencies for various nitrile probe locations on RalGDS docked to Rap E30/K31E. Black: system center of mass (60 Å boxes); red: NC $\delta$ ; green: system center of mass; blue: RalGDS center of mass.

## SCN-labeled RaIGDS mutants docked with Rap E30D/K31E

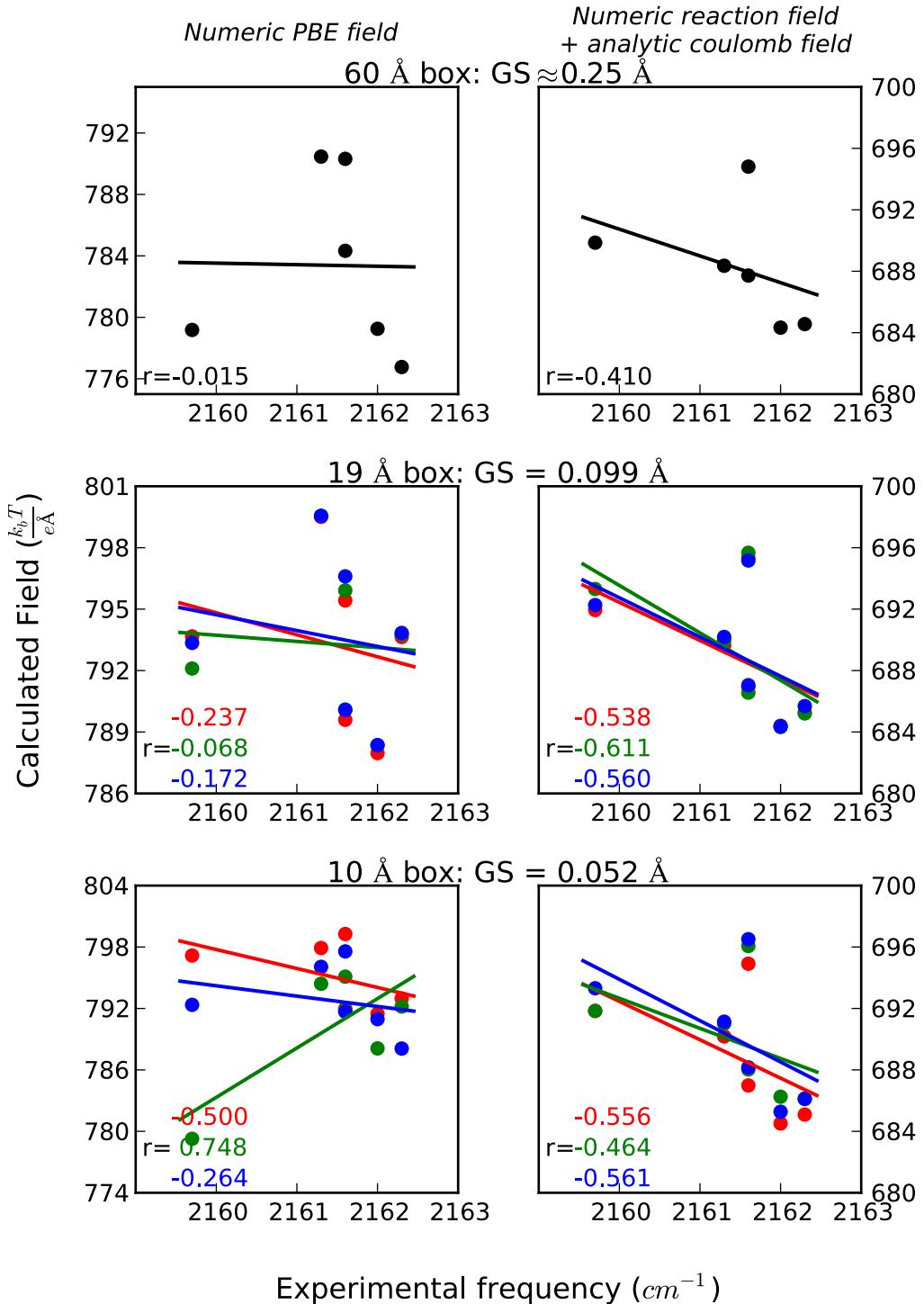


Figure 5-10: Absolute Field Plots for Various Box Locations – Rap E30D/K31E

The absolute field calculations from solving numerically solving one LPBE (left) and from summing the numeric reaction field obtained from two LPBE calculations with the analytic Coulomb field (right) compared to the experimental frequencies for various nitrile probe locations on RalGDS docked to Rap E30D/K31E. Black: system center of mass (60 Å boxes); red: NC $\delta$ ; green: system center of mass; blue: RalGDS center of mass.

### SCN-labeled RaIGDS mutants docked with WT RaI

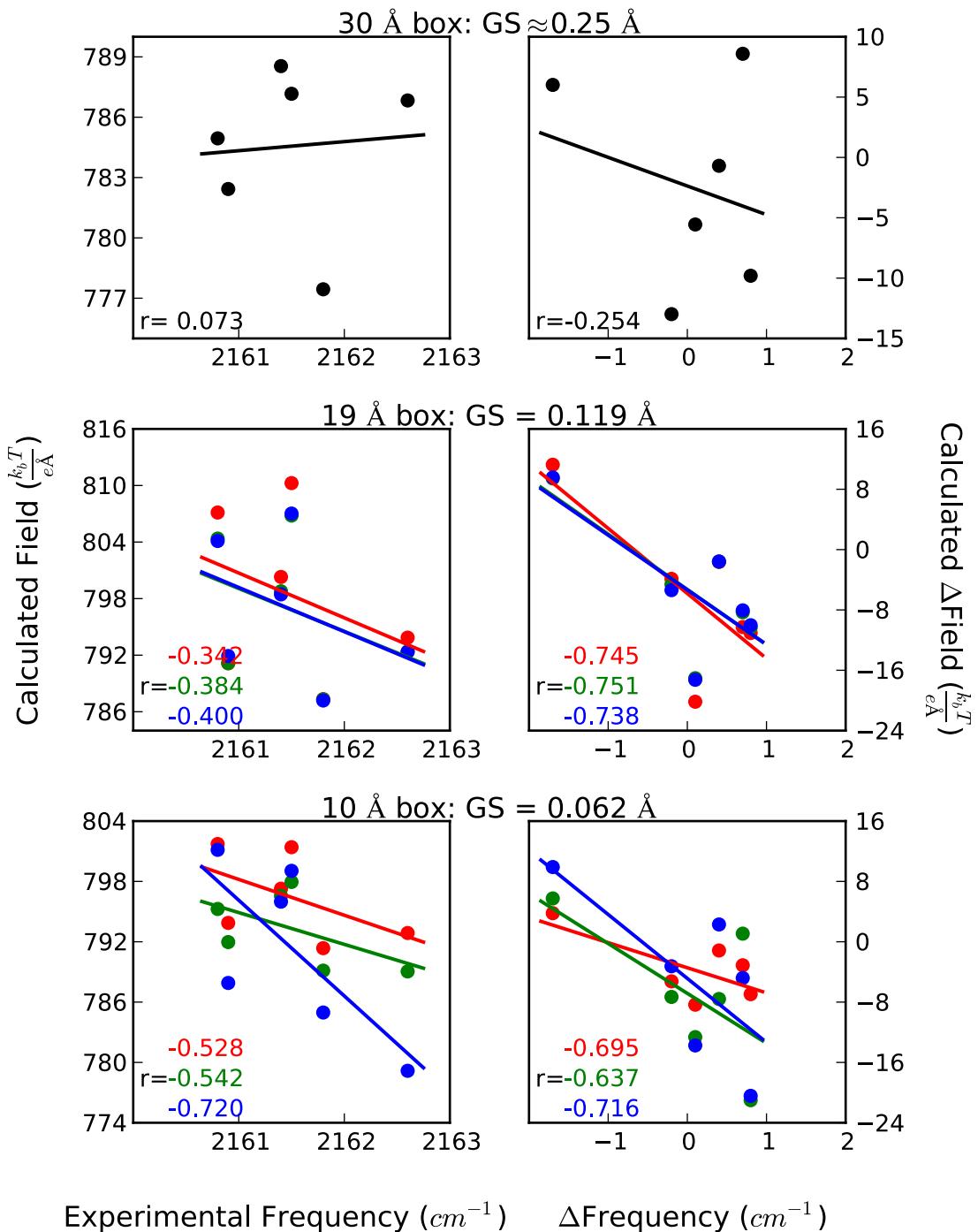


Figure 5-11: Comparing Absolute and Relative Fields - Rap E30/K31

The absolute field (left column) and relative field (right column) calculated using coarse grid spacing compared to the experimental frequencies for various nitrile probe locations on RalGDS docked to Rap E30/K31. Black: system center of mass (30 Å boxes); red: NC $\delta$ ; green: system center of mass; blue: RalGDS center of mass.

### SCN-labeled RaIGDS mutants docked with Rap E30D

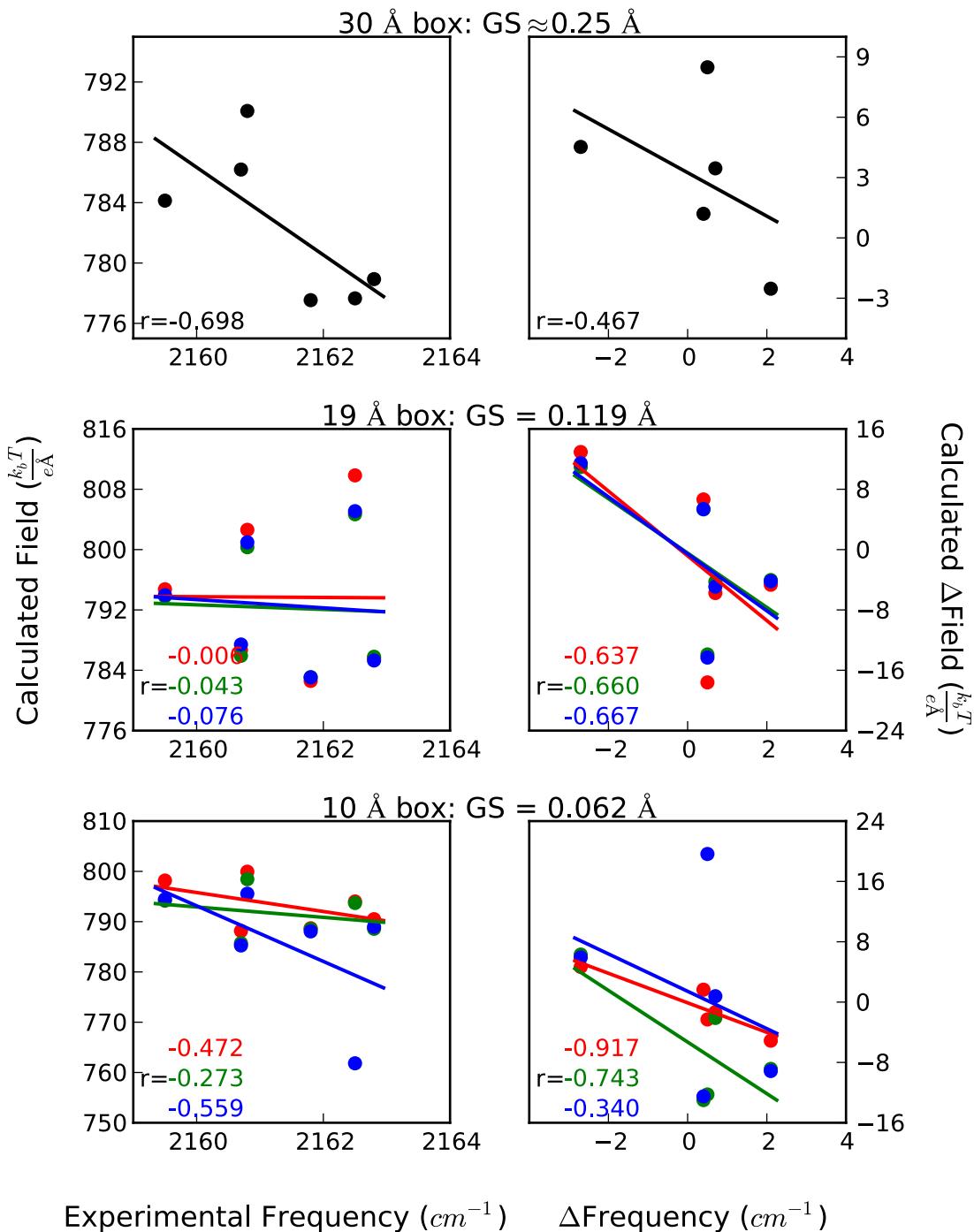


Figure 5-12: Comparing Absolute and Relative Fields - Rap E30D/K31

The absolute field (left column) and relative field (right column) calculated using coarse grid spacing compared to the experimental frequencies for various nitrile probe locations on RalGDS docked to Rap E30D/K31. Black: system center of mass (30 Å boxes); red: NC $\delta$ ; green: system center of mass; blue: RalGDS center of mass.

### SCN-labeled RaIGDS mutants docked with Rap K31E

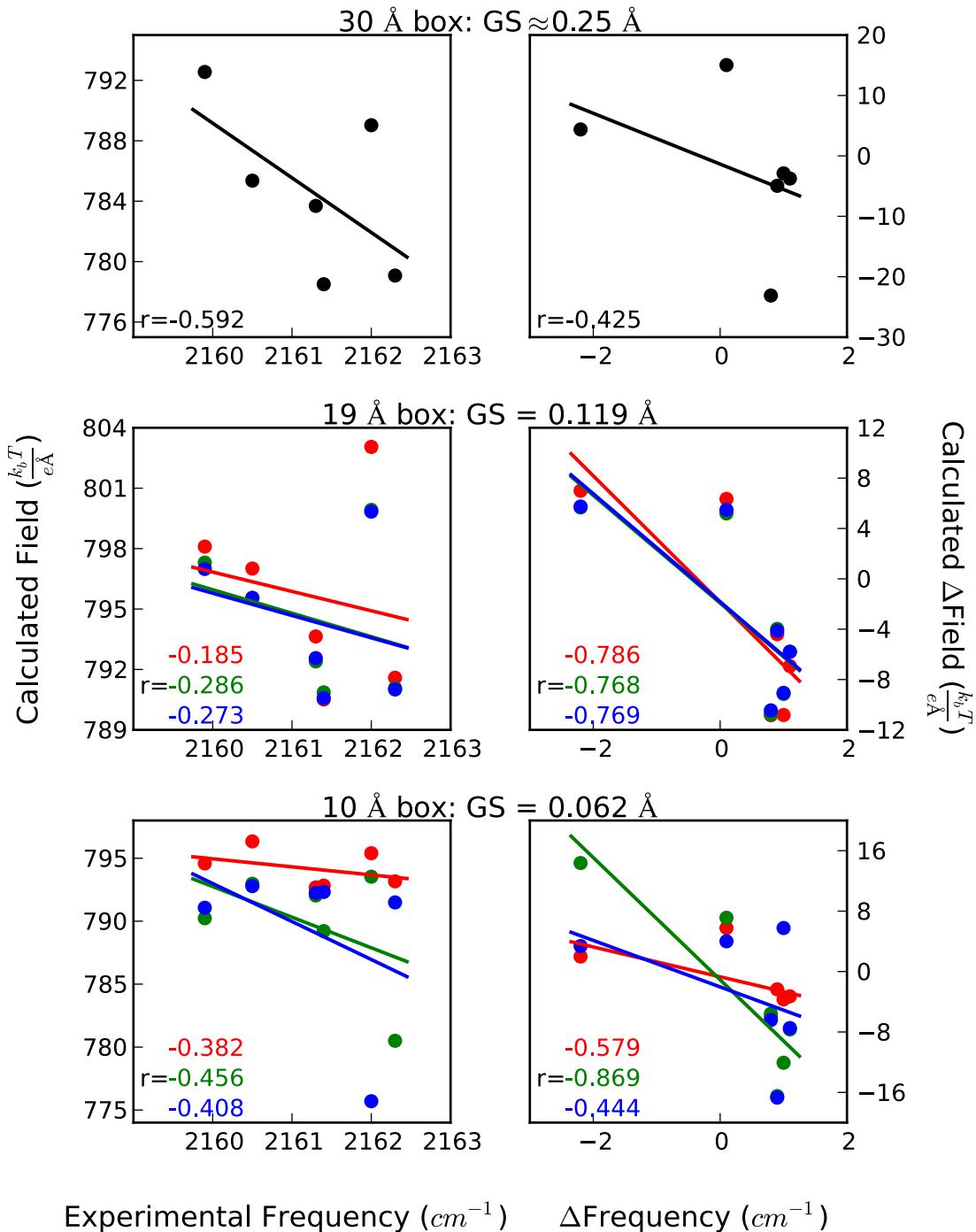


Figure 5-13: Comparing Absolute and Relative Fields - Rap E30/K31E

The absolute field (left column) and relative field (right column) calculated using coarse grid spacing compared to the experimental frequencies for various nitrile probe locations on RalGDS docked to Rap E30/K31E. Black: system center of mass (30 Å boxes); red: NC $\delta$ ; green: system center of mass; blue: RalGDS center of mass.

## SCN-labeled RaIGDS mutants docked with Rap E30D/K31E

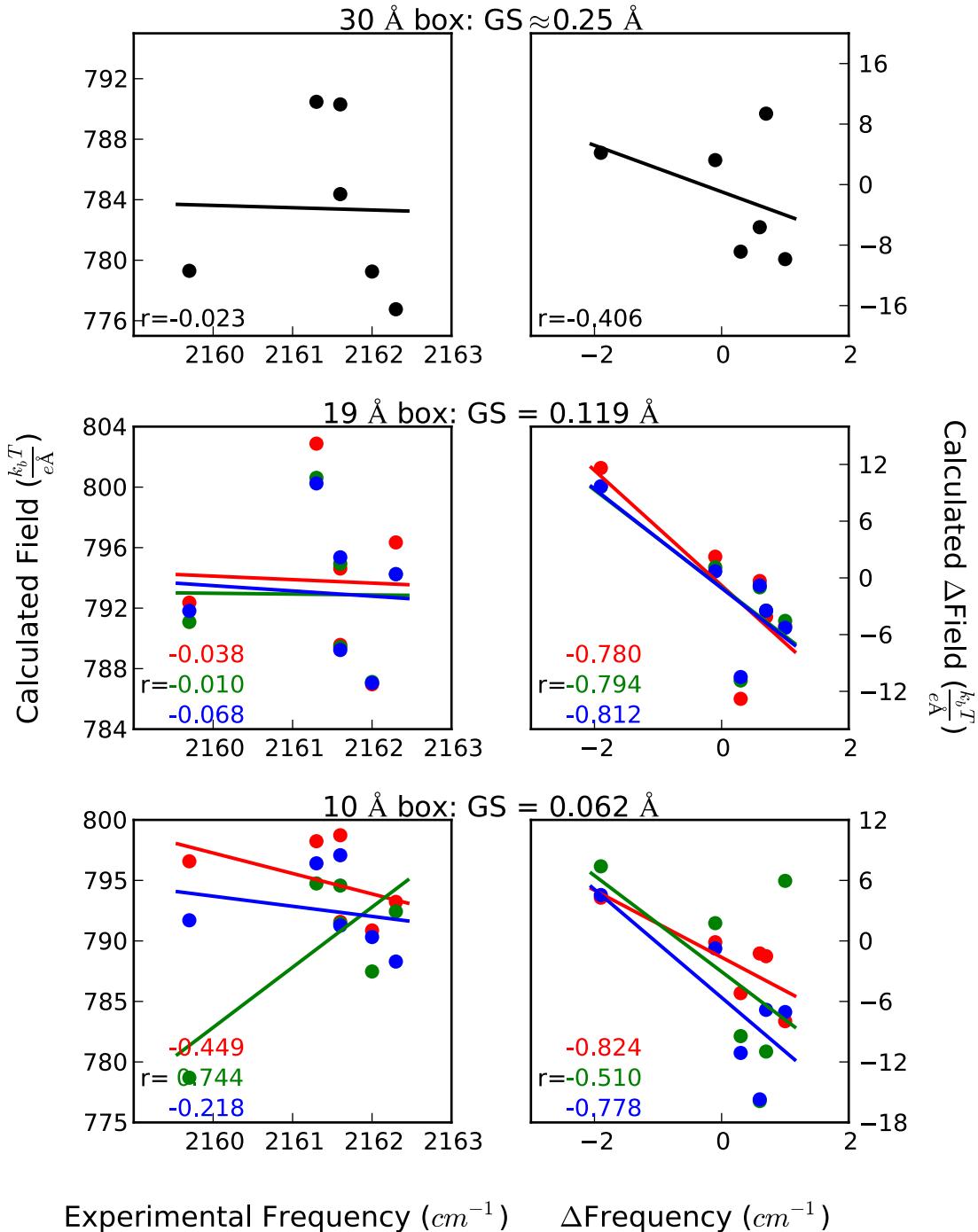


Figure 5-14: Comparing Absolute and Relative Fields - Rap E30D/K31E

The absolute field (left column) and relative field (right column) calculated using coarse grid spacing compared to the experimental frequencies for various nitrile probe locations on RalGDS docked to Rap E30D/K31E. Black: system center of mass (30 Å boxes); red: NC $\delta$ ; green: system center of mass; blue: RalGDS center of mass.

### SCN-labeled RalGDS mutants docked with WT Ral

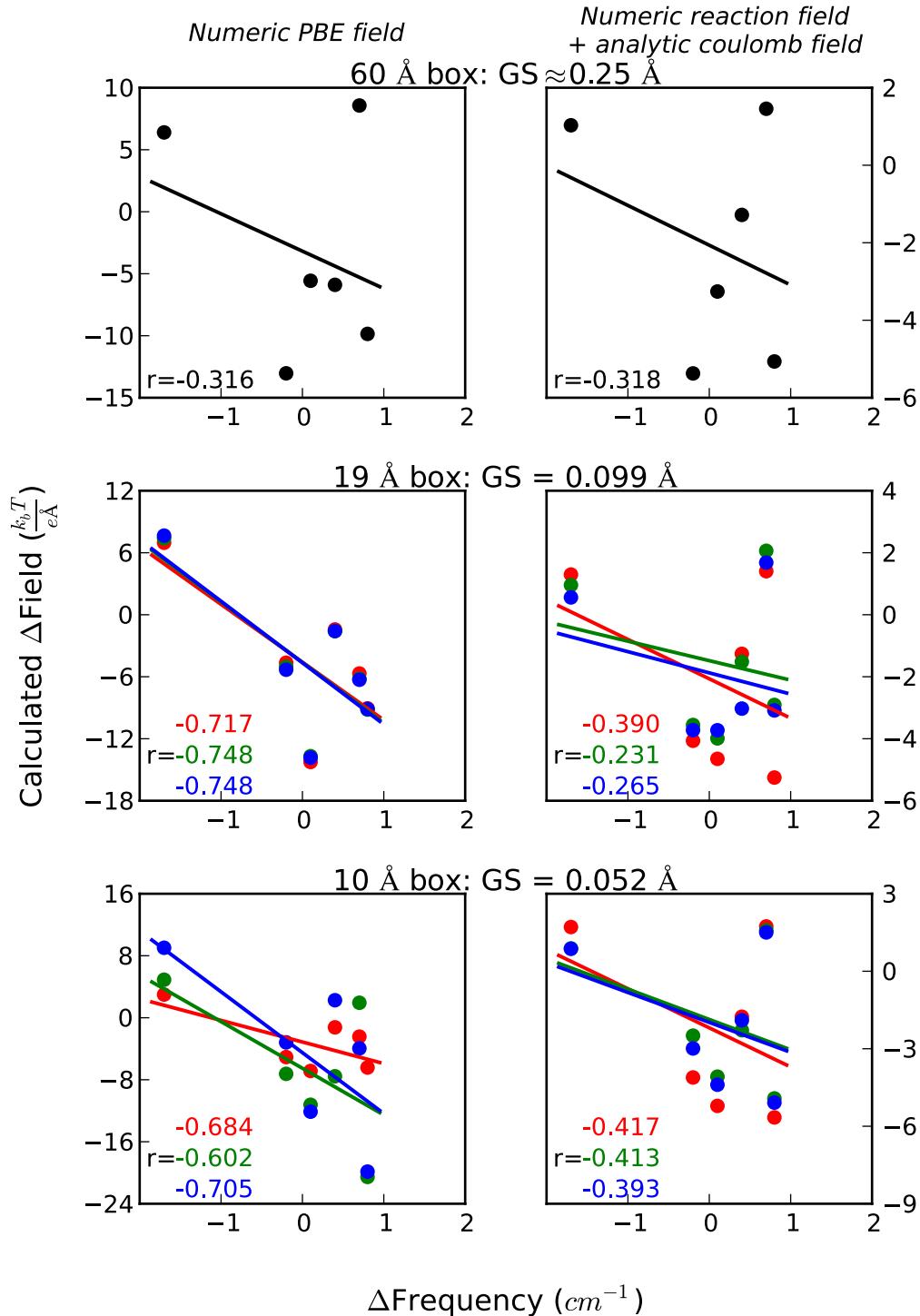


Figure 5-15: Comparing Numeric PBE and RFM Relative Fields - Rap E30/K31

The calculated change in field from subtracting two numerically solved LPBE (left) and from subtracting the numeric reaction fields obtained from four LPBE calculations with the analytic Coulomb fields (right) at various nitrile probe locations on RalGDS due to docking with Rap E30/K31 compared to the experimental change in frequency due to docking. Black: system center of mass (60 Å boxes); red: NC $\delta$ ; green: system center of mass; blue: RalGDS center of mass. Correlation coefficients are labeled in the corresponding color on each plot.

## SCN-labeled RaIGDS mutants docked with Rap E30D

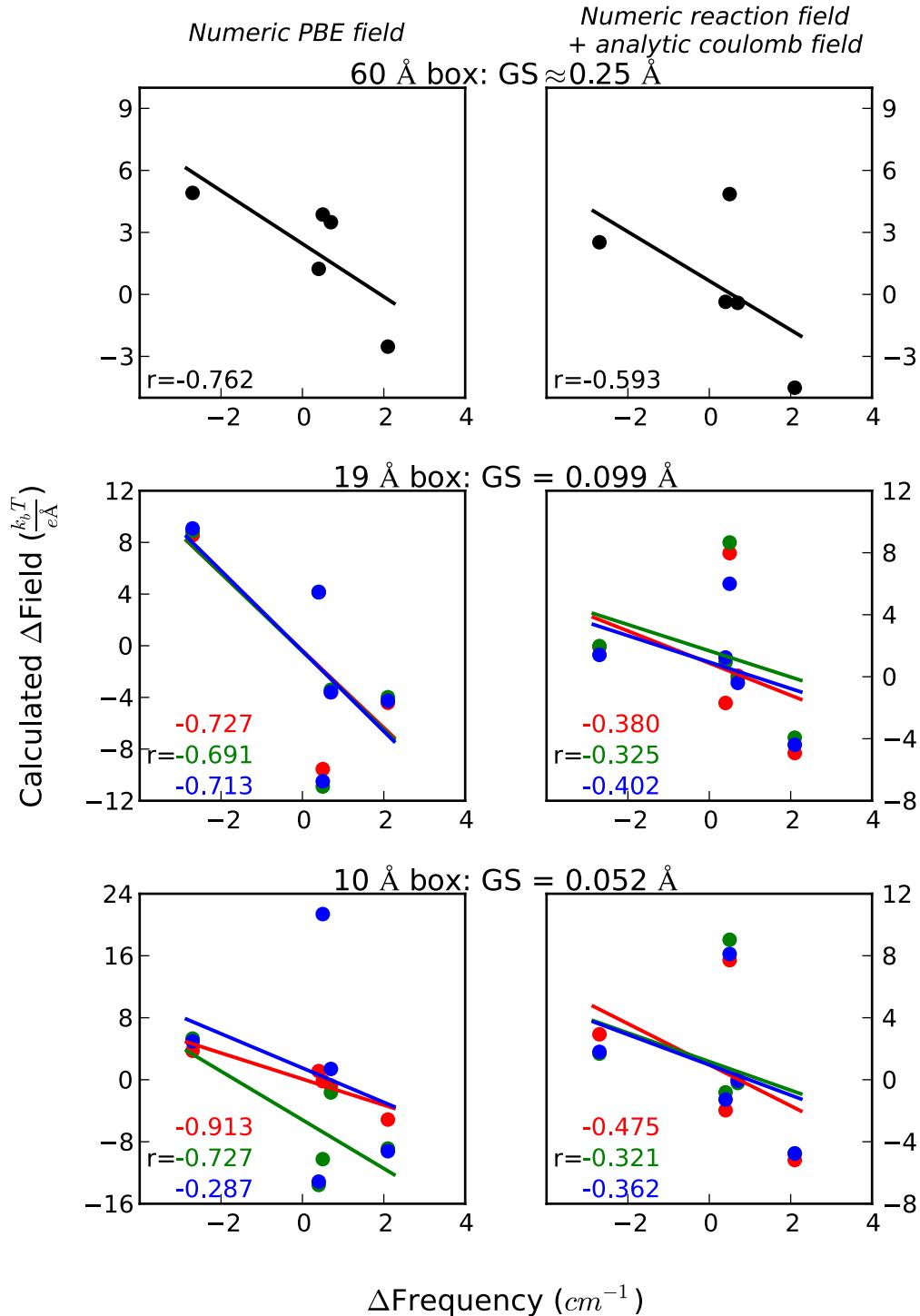


Figure 5-16: Comparing Numeric PBE and RFM Relative Fields - Rap E30D/K31

The calculated change in field from subtracting two numerically solved LPBE (left) and from subtracting the numeric reaction fields obtained from four LPBE calculations with the analytic Coulomb fields (right) at various nitrile probe locations on RalGDS due to docking with Rap E30D/K31 compared to the experimental change in absorption frequency due to docking. Black: system center of mass (60 Å boxes); red: NC $\delta$ ; green: system center of mass; blue: RalGDS center of mass.

### SCN-labeled RalGDS mutants docked with Rap K31E

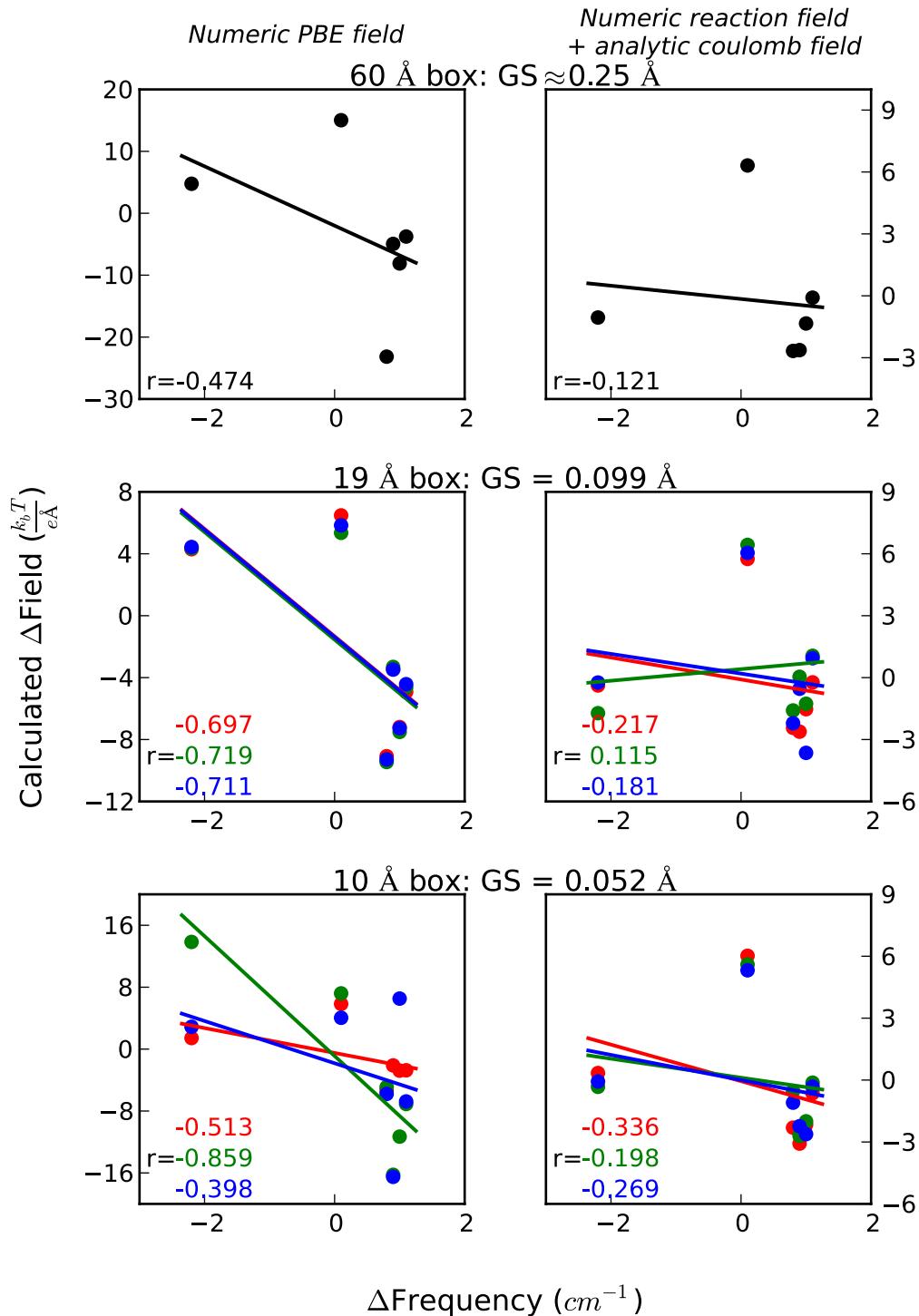


Figure 5-17: Comparing Numeric PBE and RFM Relative Fields - Rap E30/K31E

The calculated change in field from subtracting two numerically solved LPBE (left) and from subtracting the numeric reaction fields obtained from four LPBE calculations with the analytic Coulomb fields (right) at various nitrile probe locations on RalGDS due to docking with Rap E30/K31E compared to the experimental change in absorption frequency due to docking. Black: system center of mass (60 Å boxes); red: NC $\delta$ ; green: system center of mass; blue: RalGDS center of mass.

## SCN-labeled RaIGDS mutants docked with Rap E30D/K31E

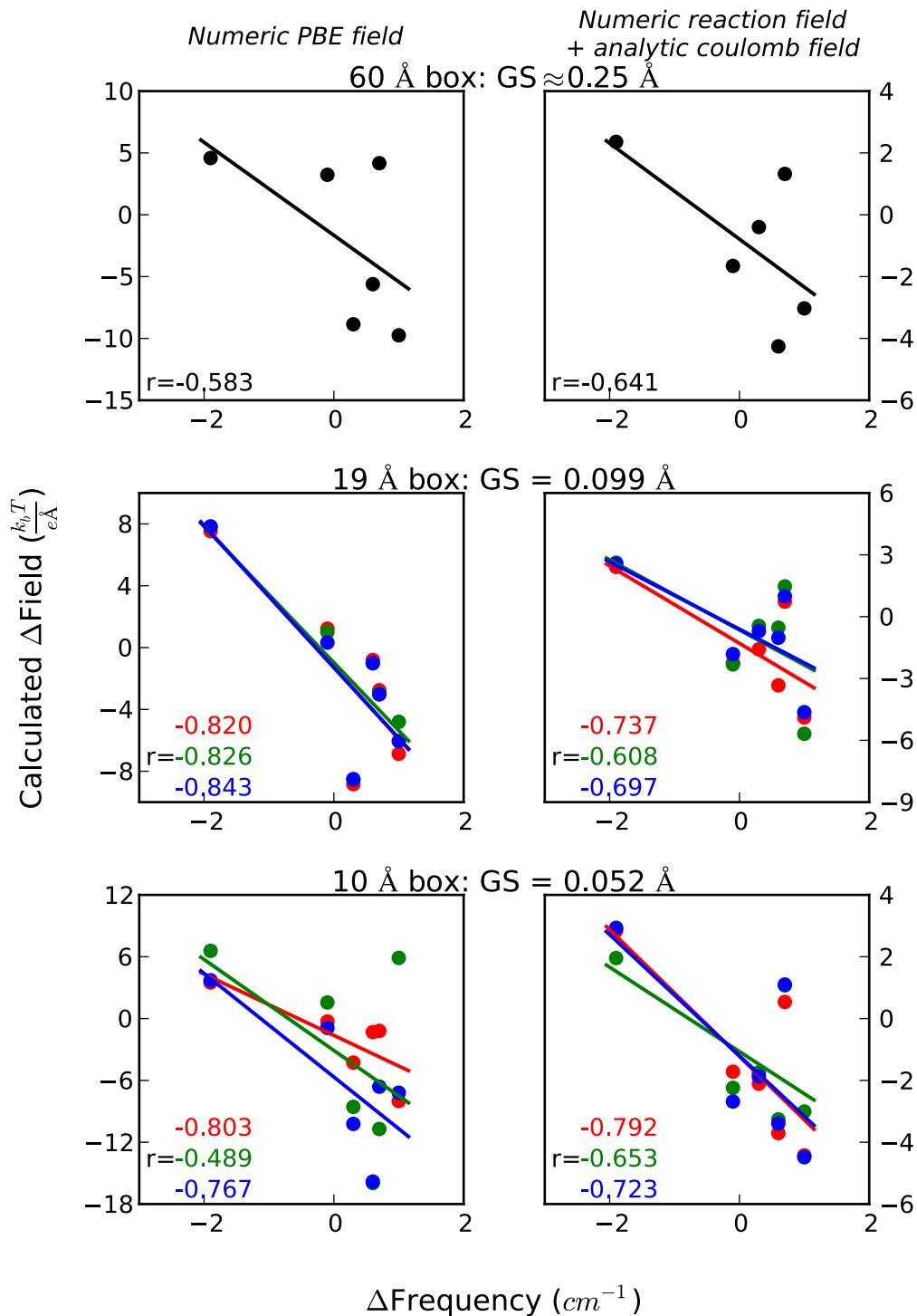


Figure 5-18: Comparing Numeric PBE and RFM Relative Fields - Rap E30D/K31E

The calculated change in field from subtracting two numerically solved LPBE (left) and from subtracting the numeric reaction fields obtained from four LPBE calculations with the analytic Coulomb fields (right) at various nitrile probe locations on RalGDS due to docking with Rap E30D/K31E compared to the experimental change in absorption frequency due to docking. Black: system center of mass (60 Å boxes); red: NC $\delta$ ; green: system center of mass; blue: RalGDS center of mass.

## **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**

Quantitative calculations of electrostatic fields in biological molecules are ubiquitous throughout the biophysical literature in aiding many types of investigations, but their accuracy has proven difficult to confirm from experimental measurements.<sup>152-161</sup>

\*\*\*The only way I knew how to fix this was to not reference a lot of people. I only removed repeat-references though and the total reference-count remains the same We have previously reported the accuracy of calculated protein electrostatic fields determined by docking various functional mutants of the protein Rap1a to six different constructs of the downstream effector protein Ral guanosine dissociation inhibitor carrying a cyanocysteine vibrational probe using the linear Poisson-Boltzmann equation (LPBE), equation (3-4).<sup>88</sup>

The LPBE is an implicit solvent, continuum electrostatic model subject to approximations about the dielectric environment that have already been discussed at length.<sup>6, 7, 72, 74, 88, 162</sup> A second approximation, however, is the difference between bulk water, modeled implicitly as a dielectric of 78, and water that is structurally or chemically important in the system. In our previous work, all solvent was treated as bulk solvent and all explicit references to solvent atoms were removed from structures prior to solving the LPBE. However, it is known that bulk water is distinct from water that is interacting with the surface of the protein, and is identified as so-called hydration water.<sup>163-166</sup> Hydration water is less mobile than bulk water, has a lower dielectric constant than bulk

water,<sup>32</sup> and can play a role in stabilizing protein-protein or protein-ligand interactions.<sup>167</sup>  
<sup>168</sup> Dismissing these persistent waters as part of the bulk solution may be detrimental to the quality of calculations of various properties that depend on molecular structure and charge such as electrostatic potentials and fields. To address this, Layfield *et al.*<sup>117</sup> recently demonstrated the importance of explicit inclusion of the water molecule nearest the nitrile probe in a quantum mechanical calculation of the vibrational frequency of nitrile probes. The included water molecule helped account for polarization effects of the solvent and improved the accuracy their calculated vibrational shifts. Additionally, Fennell *et al.*<sup>169</sup> showed that employing an implicit solvent model that distinguishes between bulk water and hydration water can better account for solvation shell properties than a purely implicit solvent model and yields results consistent with purely explicit models.

Our previous work compared calculated electrostatic fields to experimentally measured vibrational absorption frequencies for six nitrile probes placed along the Rap/RalGDS interface, at positions N27C<sub>SCN</sub>, G28C<sub>SCN</sub>, N29C<sub>SCN</sub>, Y31C<sub>SCN</sub>, K32C<sub>SCN</sub>, and N54C<sub>SCN</sub>, for monomeric RalGDS<sup>37</sup> as well as RalGDS docked to Rap E30/K31, Rap E30D/K31, Rap E30/K31E, and Rap E30D/K31E.<sup>44</sup> This constitutes a set of experiments in which the exact quantity we are trying to calculate, the electrostatic field at the vibrational chromophore, is being directly measured via the VSE. In the work reported here, we test whether including explicit water molecules improves computational correlation to experiment compared to the original, purely implicit solvent, calculations in a purely classical continuum electrostatics calculation.

## 6.2 RESULTS AND DISCUSSION

All structures examined had a total of 56.7 ns of simulation with frames saved every 5 ps, resulting in 11664 frames in which we solved the LPBE. Since the goal of our investigation is to find an easy, generalizable way to calculate electrostatic fields representative of the VSE, one of our criteria is to limit the number of arbitrary user-choices, such as frame picking. Therefore, we used all of the frames saved when calculating our Boltzmann-weighted distributions and averages.

### 6.2.1 Calculations of Absolute Fields

Figure 6-1, Figure 6-2, Figure 6-3, Figure 6-4, and Figure 6-5 compare the calculated electrostatic fields at the nitrile bond midpoint projected along the nitrile bond vector using the various partially-explicit solvent methods (blue) to the experimentally observed nitrile vibrational absorption frequencies for the RalGDS monomer, RalGDS docked to Rap E30/K31, and RalGDS docked to E30D, K31E, and E30D/K31E, respectively, and compared to previous results obtained with no explicit water molecules (red). Results using both the numeric LBPE solutions (left column) as well as the reaction field method (right column) are presented. Error bars have been omitted for clarity, but we consistently observed standard deviations of the electrostatic field calculations ranging from about  $\pm 6.5 \text{ k}_\text{b}T/\text{e}\text{\AA}$  and for vibrational frequencies ranging from about  $\pm 1 \text{ cm}^{-1}$ . The reaction field method yielded higher correlation to experiment than the numeric solutions, consistent with our previously reported results. Further discussion of the results of the absolute calculations will be limited to the reaction field method.

#### 5 Å Sphere

The top rows of Figure 6-1, Figure 6-2, Figure 6-3, Figure 6-4, and Figure 6-5 show the electrostatic fields calculated when all water molecules within 5 Å of the nitrile Nε atom

are explicitly included. In all cases the reaction field method yields better correlation than the numeric method. In Figure 6-1, the monomeric RalGDS field correlations are significantly improved by the inclusion of the solvent sphere and yet in every docked system, the solvent sphere detracts from the overall correlation to experiment.

To understand this observation, we looked at each water molecule within 5 Å of the nitrile Nε and determined the probability of a water molecule staying within 5 Å of the nitrile for  $n$  frames ( $n$  ranging from 1 to 81, the number of frames in each trajectory), shown for the RalGDS monomer in Figure 6-6, as well for RalGDS docked to the various Rap constructs (Figure 6-7, Figure 6-8, Figure 6-9, and Figure 6-10), with the probe location listed as subtitles to each histogram and the trajectory average number of frames (not a Boltzmann-weighted average) listed in the upper-right corner. The error bar on each bin is the standard deviation among the 144 trajectories for that bin. These averages are also tabulated in Table 6-1. We see that the RalGDS monomers have the lowest average number of frames that each water molecule is within the solvent sphere (15.6), while the docked complexes average approximately 18-20, indicating that we have more water moving in-and-out of the simulated system for the monomers than for the dimers.

The improvement in experimental correlation seen (Figure 6-1) using the 5 Å explicit solvent spheres (blue) compared to the entirely implicit solvent model (red) for *only* the RalGDS monomers and not the docked complex may be an artifact of insufficient water sampling in the docked simulations. Not only do the RalGDS monomers have fewer atoms, and thus fewer degrees of freedom to sample, but the docked complexes also have the Rap constructs in close proximity to the nitrile probes. One of the advantages of an implicit solvent model is the effects of the solvent are treated as an average, bulk property, and are not limited by poor sampling. When using explicit solvent, care must be taken to ensure the solvent is fully relaxed and sampled. The

significant translational and rotational mobility of each water molecule requires a significant amount of simulation time to fully sample. Explicitly including some water molecules, we are seeing more water molecules trapped in local minima created by interactions with Rap than in the monomer calculations. This can be seen in Figure 6-6, Figure 6-7, Figure 6-8, Figure 6-9, and Figure 6-10 as the spike in the probability of water molecules staying within the solvent sphere for all 81 snapshots of each 400 ps simulation, such as in the K32C<sub>SCN</sub> RalGDS mutant docked to Rap E30/K31 (Figure 6-7, lower-left corner).

The primary criteria for convergence of our simulations were 1) good sampling of the thiocyanate  $\chi_1$  and  $\chi_2$  dihedral rotamers and 2) the numeric reaction field with entirely implicit solvent cease to significantly change with additional sampling time.<sup>88</sup> Neither of these criteria are significantly dependent on the water being well sampled. In a similar study by Fried *et al.*<sup>170</sup> comparing vibrational frequencies of a carbonyl oscillator to the solvent electrostatics, they performed 2 ns of simulation on solvated small molecule and 20 ns of simulation on a solvated ribonuclease S in order to obtain well-sampled solvent. In contrast, we performed 144 x 400 ps simulations using an enhanced molecular dynamics strategy targeted at the probe only. For the solvent, backbone, and all non-cyanocysteine side chains, a relatively small amount of time was allowed for exploring the free energy landscape; this is a known limitation of the enhanced sampling method used which we accepted based on the hypothesis that the most significant degree of freedom in our system is the nitrile orientation.

To test our system for convergence, we calculated the Boltzmann-weighted average field at each time step  $n\Delta t$ , where  $\Delta t$  is 0.72 ns and  $n$  ranged from 1 to 81 (the number of frames in each trajectory), for all 30 systems examined. Briefly, we used WHAM at each time step to calculate new weights after each additional time step of each

trajectory (0.005 ns per trajectory, 144 trajectories for a total of 0.72 ns per step). Using the new weights and the frames present in calculating them, a Boltzmann-weighted average field was calculated for the 5 Å solvent sphere using both the numeric field method as well as the reaction field method. The average field as a function of time, relative to the average field at time  $n\Delta t = 0$ , is shown in Figure 6-11. The field relative to the initial field was chosen to allow easy visualization using a single y-axis of the average field calculated using the numeric field method (blue) and the reaction field method (green). Relative to the magnitude of the standard deviation in field, the average field is not changing significantly after the first half of the simulation in nearly all systems. To further show this feature, we have plotted the derivative with respect to time of the average field at each time step in Figure 6-12. This was obtained from using the finite difference between steps. In the first one-third to one-half of the simulation, the field gradient is changing rapidly as  $t$  increases, but by the last half of the simulation, the field gradient shows only small fluctuations centered about zero, indicating that, on average, each additional time step is not changing the calculated average significantly. Given this information, it is reasonable to conclude that, even with explicit water near the probe, each additional time step is not significantly altering the average calculated field and the simulation has converged with respect to the electrostatic field experienced by the nitrile probe.

Using APBS and treating all water molecules implicitly yielded better correlations to experimental data than treating water molecules within 5 Å of the nitrile probe explicitly in all the docked simulations given identical sampling. The typical, entirely implicit solvent treatment outperforms this selectively explicit solvent treatment at similar computational costs. Allowing adequate simulation time for the explicit solvent to be fully sampled, the 5 Å solvent sphere calculations and the entirely implicit solvent

calculations should give identical results. If the entirely implicit solvent model returns better results with less computational overhead than trying to ensure the water is fully sampled then, within the framework of this study, the entirely implicit solvent model is better. For a system in which the solvent may not be fully sampled, the 5 Å solvent sphere is likely to be more detrimental than the traditional, entirely implicit solvent model, which is insensitive to solvent under-sampling.

### ***Nearest Water***

The second rows of Figure 6-1, Figure 6-2, Figure 6-3, Figure 6-4, and Figure 6-5 show the electrostatic fields calculated when the water molecule closest to the nitrile Nε is explicitly included. In all cases, the correlation with experiment is not significantly improved with respect to the entirely implicit calculations. It was hypothesized that because the nearest water molecule in the QM/MM calculations done by Layfield *et al.* helped account for polarization effects by including water reorientation due to protein electrostatic fields, the same may be true for a purely classical set of calculations. This hypothesis can be tested against a second experimental measurement, the value of the Stark tuning rate,  $\Delta\bar{\mu}$ . Using the VSE and the slope of our best-fit line, we can back-calculate the solute dielectric, or “ideal” protein dielectric ( $\epsilon'$ ), required for the best-fit line to reproduce the known experimental value of the Stark tuning rate, which are tabulated in Table 6-2 along with the virtual Stark tuning rates (*VSTR*), correlation coefficients  $r$ , and p-values. Errors on the *VSTR* and  $\epsilon'$  were obtained using equation (6-1) (where  $m$  is the slope,  $r$  is the correlation coefficient, and  $n$  is the number of data points in the fit) as the error of the fitted slopes.

$$\sigma_m = \frac{m}{r} \frac{\sqrt{1-r^2}}{\sqrt{n-2}} \quad (6-1)$$

“Ideal” dielectric values reported in Table 6-2 are typically negative, indicating that the calculated field is not following the expected VSE. Although this is a non-physical observation, discussing the magnitudes of the “ideal” dielectrics still give a description of the amount of charge screening required to produce field changes in the same magnitude as the observed shift. Further discussion of these non-VSE fields will be addressed in Chapters 7 and 8.

The magnitudes of the “ideal” protein dielectric<sup>88</sup> are typically slightly larger for the nearest water calculations compared to the entirely implicit solvent calculations. Since dielectric is the macroscopic analog for the molecular property of polarizability, in a system that perfectly includes all polarizability effects, the appropriate solute dielectric is 1.0 and higher dielectric values imply worse modeling of polarizability effects and a need to effectively increase charge screening. The strategy including waters resulted in higher values of the “ideal” protein dielectric, indicating that explicit inclusion of the nearest water molecules, in addition to neither typically nor significantly improving experimental correlation, have failed to capture previously excluded polarizability effects in the continuum calculation.

### ***Hydrogen-bonding Water***

The third rows of Figure 6-1, Figure 6-2, Figure 6-3, Figure 6-4, and Figure 6-5 show the electrostatic fields calculated when the water molecule hydrogen-bonding ( $\text{N--H}$  distance  $\leq 2.25 \text{ \AA}$ ,  $\text{N--H--O}$  smallest bond angle  $\geq 138^\circ$ ) to the nitrile  $\text{Ne}$  is explicitly included. In all cases, the correlation with experiment is not significantly different from the entirely implicit calculations. This is due to the relatively few number of frames in which hydrogen bonding is actually observed. The RalGDS monomers had the highest occurrence, with 10.69% of all frames having a hydrogen-bonding water. The

docked complexes showed fewer frames with hydrogen bonding atoms; Rap E30/K31, Rap E30D/K31E, Rap E30/K31E, and Rap E30D/K31 with 9.92%, 9.39%, 9.36%, and 8.63% respectively. The frequency of frames containing hydrogen-bonding water for each system/probe combination is shown in Table 6-3. Because such a small number of frames even have a hydrogen-bonding water, the field calculations are dominated by frames with zero explicit water molecules—identical to those in the original calculations—and therefore the Boltzmann-weighted average is nearly identical to the original calculations. There are not enough frames containing a water molecule to significantly shift the electrostatic field experienced by the nitrile probe.

Although the 5 Å water sphere calculation significantly altered the calculated electrostatic field, a targeted, enhanced molecular dynamics strategy which may leave solvent under-sampled is detrimental to the overall quality of the calculations in a way not observed when using an entirely implicit solvent model. Additionally, including only the nearest water molecules or water molecules hydrogen-bonding to the nitrile Nε also does not appear to significantly improve experimental correlation, and for absolute field calculations using a LPBE model, no consideration for interfacial or near-site-of-interest water is necessary.

### 6.2.2 Calculations of Field Differences

Figure 6-13, Figure 6-14, Figure 6-15, and Figure 6-16 compare the calculated relative electrostatic fields due to docking at the nitrile bond midpoint projected along the nitrile bond vector using the various explicit solvent methods (blue) to the experimentally observed nitrile vibrational absorption frequencies for RalGDS docked to Rap E30/K31, and RalGDS docked to E30D, K31E, and E30D/K31E, respectively. As in the absolute calculations, error bars have been omitted for clarity and are approximately  $\pm 1 \text{ cm}^{-1}$  in

the x-axis and  $\pm 6.5 k_b T/e\text{\AA}$  in the y-axis. Previously published results using entirely implicit solvent are also shown in red for comparison. Results using both the numeric LBPE solutions (left column) as well as the reaction field method (right column) are presented. Consistent with our previous study, the numeric method yields higher correlation to experiment than the reaction field method for the relative calculations and therefore further discussion of the relative field calculations is limited to those obtained by the numeric method.

### **5 Å Sphere**

The top rows of Figure 6-13, Figure 6-14, Figure 6-15, and Figure 6-16 show the relative electrostatic fields due to RalGDS docking to the Rap E30/K31, Rap E30D/K31, Rap E30/K31E, and Rap E30D/K31E respectively, calculated when all water molecules within 5 Å of the nitrile Nε atom are explicitly included. Although the explicit solvent sphere was largely a failure for the absolute field calculations, the relative field calculations show the same recovery of correlation we previously reported.<sup>88</sup> The numeric relative fields were significantly more correlated to the experimental vibrational energies of the nitrile probe than the reaction field method absolute fields for the Rap E30/K31, Rap E30/K31E, and Rap E30D/K31E. The numeric relative field for RalGDS docked to Rap E30D/K31 was approximately the same as the reaction field correlation. The increased correlations are unlikely to be purely an artifact of the better-correlated RalGDS monomer calculation because of the poor correlation of the absolute field calculations for all of the docked complexes. As before, when taking the *difference* in electrostatic fields obtained from the LPBE, cancellation of errors in the continuum calculations was observed. These observations further support the conclusion of our previous investigation; some non-physical characteristic or important, excluded physical

characteristics not adequately addressed in the absolute calculations can be cancelled out by the subtraction of fields in a comparison of two similar but slightly different states. Additionally, we see slightly greater experimental correlation using the explicit solvent sphere for the relative field calculations than when using the entirely implicit model. However, if trying to efficiently maximizing both absolute and relative field correlations, the additional computational cost of performing both entirely implicit calculations and the 5 Å sphere calculations favors the use of entirely implicit solvent.

### ***Nearest Water and Hydrogen-bonding Water***

The second rows Figure 6-13, Figure 6-14, Figure 6-15, and Figure 6-16 show the relative electrostatic fields due to RalGDS docking to the Rap E30/K31, Rap E30D/K31, Rap E30/K31E, and Rap E30D/K31E respectively, calculated with the water molecule nearest the nitrile N $\epsilon$  atom explicitly included while the third rows show the relative fields calculated explicitly including only hydrogen-bonding waters. In both methods, the relative fields and the “ideal” dielectric constants estimated from the calculation are nearly identical to the original calculations with no explicit water molecules included. Although the nearest water method is guaranteed to have an additional source of charge close to the nitrile, the values for the relative fields were approximately the same as in the implicit model. This suggests that any effect that the nearest water molecule may have on the electrostatic environment of the nitrile probe is just as well modeled explicitly as implicit bulk water. This in turn implies that treating bulk water explicitly as part of the solute dielectric does not detract from the overall quality of the calculation and there is some leeway in assigning bulk versus hydration water. For relative field calculations, no consideration for nearest or hydrogen-bonding water is necessary.

For quantifying *changes* in electrostatic fields with *changes* in vibrational absorption frequency due to docking to another protein, explicitly including a 5 Å solvent sphere near the location-of-interest was beneficial to the overall correlation, but because of the poor absolute correlation, comes at an additional cost that may not justify the small increase in correlation over using entirely implicit solvent.

### 6.3 CONCLUSION

When calculating the absolute electrostatic environment of a nitrile covalently bound to the surface of a protein using the LPBE, unlike the results reported for QM/MM vibrational spectra calculations,<sup>117</sup> explicit solvent near the vibrational chromophore should not be included in lieu of an implicit model when solvent sampling may be an issue. We do not see any consistent, significant increase in correlation by explicitly including all water near the nitrile probe, the closest water molecule to the nitrile probe, or water hydrogen bonding to the probe. However, for comparing vibrational absorption energy shifts due to docking, the inclusion of explicit solvent yielded slightly improved results for the 5 Å solvent spheres and nearly identical results for the nearest water and hydrogen-bonding waters compared to the wholly implicit solvent calculation. In computing changes in protein electrostatics due to some perturbation using the LPBE model, there appears to be neither significant, consistent advantage nor disadvantage to explicitly modeling the hydration water when there is little to no persistent hydrogen bonding and the desired information is the difference between two calculations, not an absolute value of the electrostatic field. However, explicitly including water requires an additional preparation step, selecting the waters, and thus is slightly more computationally expensive for very little additional information and significantly more expensive if the absolute fields are also important. We find that, although QM/MM

calculations of the nitrile absorption frequency are largely benefitted by the inclusion of water near the nitrile probe,<sup>117</sup> explicit definition of water molecules near the vibrational Stark probe typically yields neither significant improvement nor deterioration of the predictive capability of vibrational Stark shifts using the LPBE electric field difference calculations and is likely not necessary for future, similar calculations. Additionally, because the use of explicit water does not address the lack of atomic polarizability and the choice of the protein dielectric is arbitrary, additional studies will be aimed at investigating the use of the polarizable model AMOEBA.<sup>13, 15, 171</sup> To address the possibly poor solvent sampling, our next study presented significantly increased the sampling time for each umbrella window.

Table 6-1: Average number of Frames Water Molecules Remained within 5 Å of the Thiocyanate Nε Atom

	<b>Monomer</b>	<b>WT</b>	<b>E30D</b>	<b>K31E</b>	<b>E30D/K31E</b>
<b>N27C<sub>SCN</sub></b>	14.8	16.2	13.2	16.8	15.1
<b>G28C<sub>SCN</sub></b>	14.6	17.0	16.9	16.9	18.2
<b>N29C<sub>SCN</sub></b>	16.5	21.5	20.6	23.5	23.4
<b>Y31C<sub>SCN</sub></b>	17.2	19.5	19.1	17.2	18.3
<b>K32C<sub>SCN</sub></b>	15.6	27.8	22.0	24.7	22.7
<b>N54C<sub>SCN</sub></b>	14.9	18.6	18.8	16.9	18.4
<b>Average<sup>a</sup></b>	15.6 ± 1.0	20.1 ± 4.2	18.4 ± 3.1	19.3 ± 3.7	19.3 ± 3.1

<sup>a</sup>Not a Boltzmann-weighted average.

Table 6-2: Virtual Stark Tuning Rate ( $VSTR$ ), "ideal" Dielectric Constants ( $\epsilon'$ ), Correlation Coefficients (r), and p-values for APBS Calculations of Electrostatic Fields

Method	Implicit Solvent <sup>b</sup>							
	Numeric LPBE Solutions				Reaction Field Method			
	$VSTR^a$	$\epsilon'$	r	p-value	$VSTR^a$	$\epsilon'$	r	p-value
<b>Monomer</b>	-1.7 ± 4.9	-2.3 ± 6.7	-0.172	0.74	-0.99 ± 1.47	-4.0 ± 5.9	-0.319	0.53
<b>WT</b>	-0.29 ± 0.21	-13 ± 10	-0.565	0.22	-0.36 ± 0.22	-11 ± 7	-0.633	0.15
<b>WT Docked</b>	-0.36 ± 0.19	-11 ± 6	-0.684	0.17	-0.66 ± 0.72	-6.0 ± 6.5	-0.417	0.46
<b>E30D</b>	-0.46 ± 0.33	-8.6 ± 6.2	-0.568	0.22	-0.26 ± 0.09	-15 ± 5	-0.83	0.02
<b>E30D Docked</b>	-0.60 ± 0.13	-6.6 ± 1.5	-0.913	0.01	-0.76 ± 0.70	-5.2 ± 4.8	-0.475	0.39
<b>K31E</b>	-1.5 ± 1.4	-2.6 ± 2.5	-0.462	0.34	-0.95 ± 0.40	-4.2 ± 1.7	-0.768	0.05
<b>K31E Docked</b>	-0.62 ± 0.52	-6.4 ± 5.3	-0.513	0.35	-1.1 ± 1.5	-3.5 ± 5.1	-0.336	0.56
<b>E30D/K31E</b>	-0.54 ± 0.47	-7.3 ± 6.4	-0.5	0.29	-0.40 ± 0.30	-9.9 ± 7.4	-0.556	0.23
<b>E30D/K31E Docked</b>	-0.34 ± 0.13	-12 ± 4	-0.803	0.07	-0.49 ± 0.19	-8.1 ± 3.1	-0.792	0.07
<b>5 Å Explicit Solvent Sphere</b>								
Method	Numeric LPBE Solutions				Reaction Field Method			
	$VSTR^a$	$\epsilon'$	r	p-value	$VSTR^a$	$\epsilon'$	r	p-value
	-0.77 ± 0.79	-5.1 ± 5.3	-0.438	0.37	-0.55 ± 0.27	-7.1 ± 3.5	-0.714	0.09
<b>Monomer</b>	-0.80 ± 1.47	-4.9 ± 9.1	-0.263	0.61	-0.89 ± 1.16	-4.4 ± 5.8	-0.359	0.47
<b>WT</b>	-0.32 ± 0.12	-12 ± 5	-0.805	0.07	-0.57 ± 0.45	-7.0 ± 5.5	-0.536	0.32
<b>WT Docked</b>	-0.37 ± 0.21	-11 ± 6	-0.656	0.13	-0.23 ± 0.10	-17 ± 7	-0.769	0.05
<b>E30D</b>	-0.50 ± 0.21	-8.0 ± 3.3	-0.772	0.09	-0.63 ± 0.61	-6.3 ± 6.1	-0.457	0.41
<b>E30D Docked</b>	2.0 ± 3.6	2.0 ± 3.6	0.266	0.60	-11 ± 40	-0.37 ± 1.32	-0.135	0.79
<b>K31E</b>	-0.46 ± 0.16	-8.6 ± 3.1	-0.816	0.06	-0.77 ± 0.55	-5.2 ± 3.7	-0.571	0.28
<b>K31E Docked</b>	-4.7 ± 40.6	-0.85 ± 7.25	-0.058	0.91	-0.91 ± 1.74	-4.3 ± 8.3	-0.253	0.62
<b>E30D/K31E</b>	-0.30 ± 0.10	-13 ± 4	-0.839	0.04	-0.43 ± 0.23	-9.2 ± 4.9	-0.685	0.16
<b>Nearest Water Molecule</b>								
Method	Numeric LPBE Solutions				Reaction Field Method			
	$VSTR^a$	$\epsilon'$	r	p-value	$VSTR^a$	$\epsilon'$	r	p-value
	-1.6 ± 4.5	-2.5 ± 6.9	-0.176	0.73	-0.92 ± 1.38	-4.3 ± 6.4	-0.317	0.53
<b>Monomer</b>	-0.28 ± 0.20	-14 ± 10	-0.572	0.21	-0.33 ± 0.18	-12 ± 7	-0.676	0.12
<b>WT</b>	-0.35 ± 0.18	-11 ± 6	-0.688	0.16	-0.63 ± 0.67	-6.3 ± 6.7	-0.426	0.45
<b>WT Docked</b>	-0.42 ± 0.27	-9.4 ± 6.0	-0.617	0.17	-0.25 ± 0.08	-16 ± 4.9	-0.852	0.02
<b>E30D</b>	-0.55 ± 0.13	-7.2 ± 1.7	-0.905	0.01	-0.69 ± 0.58	-5.8 ± 4.8	-0.511	0.35
<b>E30D Docked</b>	-1.1 ± 0.8	-3.8 ± 2.6	-0.573	0.21	-0.72 ± 0.28	-5.5 ± 2.1	-0.79	0.04
<b>K31E</b>	-0.57 ± 0.48	-6.9 ± 5.8	-0.511	0.35	-1.0 ± 1.4	-3.9 ± 5.5	-0.341	0.56
<b>K31E Docked</b>	-0.59 ± 0.65	-6.7 ± 7.4	-0.412	0.40	-0.41 ± 0.37	-9.6 ± 8.7	-0.487	0.31
<b>E30D/K31E</b>	-0.33 ± 0.11	-12 ± 4	-0.824	0.05	-0.48 ± 0.19	-8.2 ± 3.2	-0.789	0.08
<b>Hydrogen-Bonding Water Molecule</b>								
Method	Numeric LPBE Solutions				Reaction Field Method			
	$VSTR^a$	$\epsilon'$	r	p-value	$VSTR^a$	$\epsilon'$	r	p-value
	-1.7 ± 4.9	-2.3 ± 6.7	-0.171	0.74	-0.97 ± 1.46	-4.1 ± 6.2	-0.315	0.53
<b>Monomer</b>	-0.29 ± 0.21	-14 ± 10	-0.57	0.21	-0.35 ± 0.20	-11 ± 7	-0.652	0.14
<b>WT</b>	-0.36 ± 0.19	-11 ± 5.8	-0.689	0.16	-0.65 ± 0.70	-6.1 ± 6.5	-0.423	0.46
<b>WT Docked</b>	-0.46 ± 0.33	-8.6 ± 6.3	-0.567	0.22	-0.26 ± 0.09	-15 ± 5	-0.836	0.02
<b>E30D</b>	-0.59 ± 0.13	-6.7 ± 1.5	-0.912	0.01	-0.74 ± 0.66	-5.3 ± 4.8	-0.489	0.38

<b>K31E</b>	-1.4 ± 1.3	-2.8 ± 2.6	-0.481	0.31	-0.90 ± 0.37	-4.4 ± 1.8	-0.772	0.05
<b>K31E Docked</b>	-0.61 ± 0.52	-6.5 ± 5.5	-0.508	0.35	-1.1 ± 1.6	-3.6 ± 5.1	-0.334	0.57
<b>E30D/K31E</b>	-0.51 ± 0.43	-7.7 ± 6.5	-0.513	0.28	-0.38 ± 0.28	-10 ± 8	-0.565	0.22
<b>E30D/K31E Docked</b>	-0.33 ± 0.12	-12 ± 5	-0.801	0.07	-0.47 ± 0.18	-8.3 ± 3.2	-0.797	0.07

Table 6-3: Percentage of Total Frames with a Nitrile-Water Hydrogen Bond

	<b>Monomer</b>	<b>WT</b>	<b>E30D</b>	<b>K31E</b>	<b>E30D/K31E</b>
<b>N27C<sub>SCN</sub></b>	10.25 %	11.79 %	7.81 %	10.85 %	9.45 %
<b>G28C<sub>SCN</sub></b>	13.45 %	13.35 %	12.60 %	11.86 %	12.57 %
<b>N29C<sub>SCN</sub></b>	6.51 %	6.69 %	7.33 %	7.79 %	5.82 %
<b>Y31C<sub>SCN</sub></b>	11.31 %	9.98 %	9.08 %	8.95 %	11.40 %
<b>K32C<sub>SCN</sub></b>	9.52 %	6.00 %	5.42 %	2.50 %	4.47 %
<b>N54C<sub>SCN</sub></b>	13.13 %	11.71 %	9.57 %	14.20 %	12.62 %
<b>Average</b>	10.69 %	9.92 %	8.63 %	9.36 %	9.39 %

<sup>a</sup>Not a Boltzmann-weighted average.

### SCN-labeled RaGDS monomer absolute fields

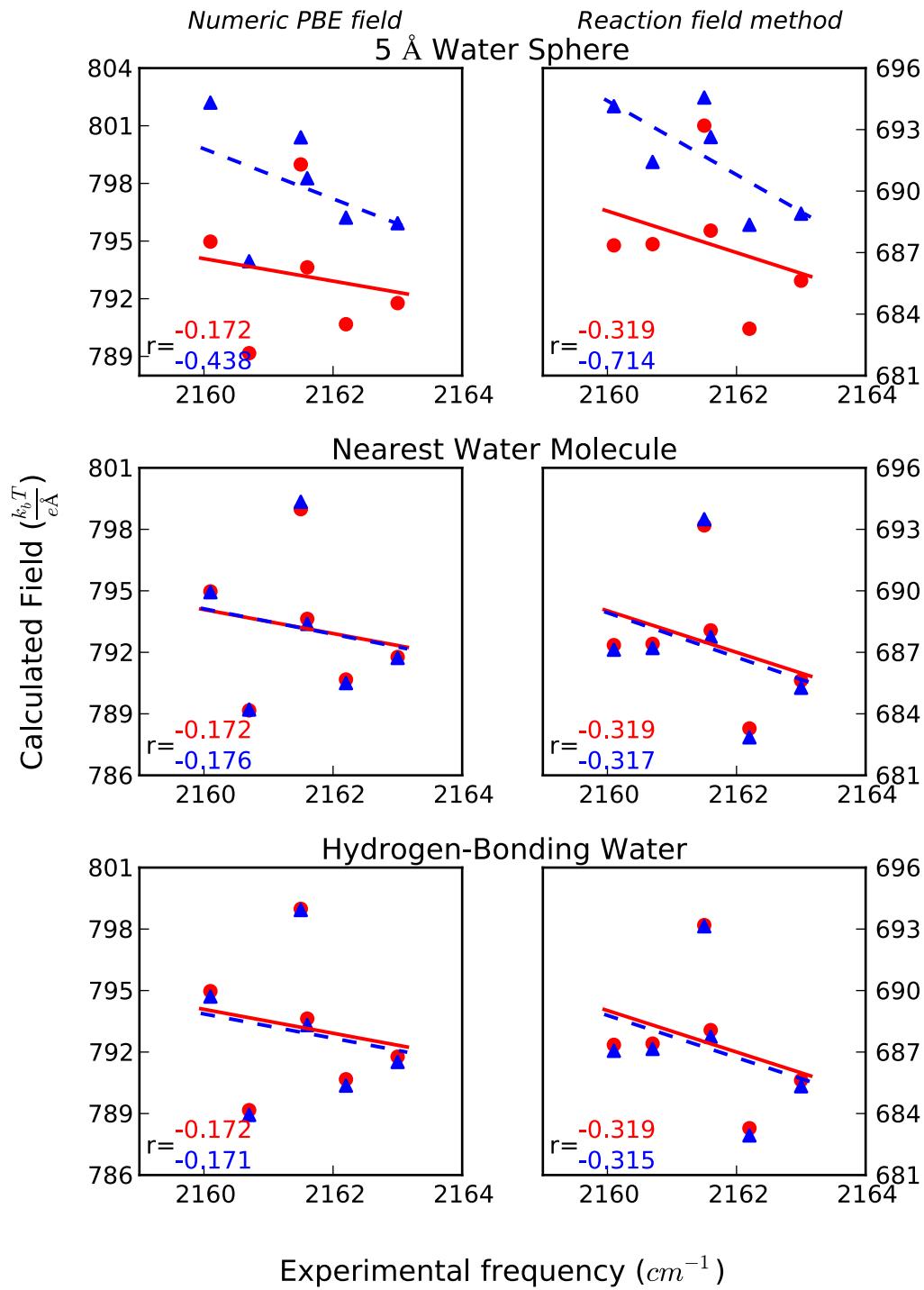


Figure 6-1: PB Absolute Field Plots - Monomer

The absolute field calculations using the numeric LPBE solutions (left column) and from the numeric reaction field method (right column) compared to the experimental frequencies for various nitrile probe locations on the RalGDS monomer. Red: originally reported, entirely implicit solvent fields; blue: fields using the solvent selection model described by the row subheading. Correlation coefficients are listed in the bottom left corner in the corresponding color. Standard deviations are approximately  $\pm 1 \text{ cm}^{-1}$  and  $\pm 6.5 \text{ k}_\text{b} \text{T/e}\text{\AA}$ .

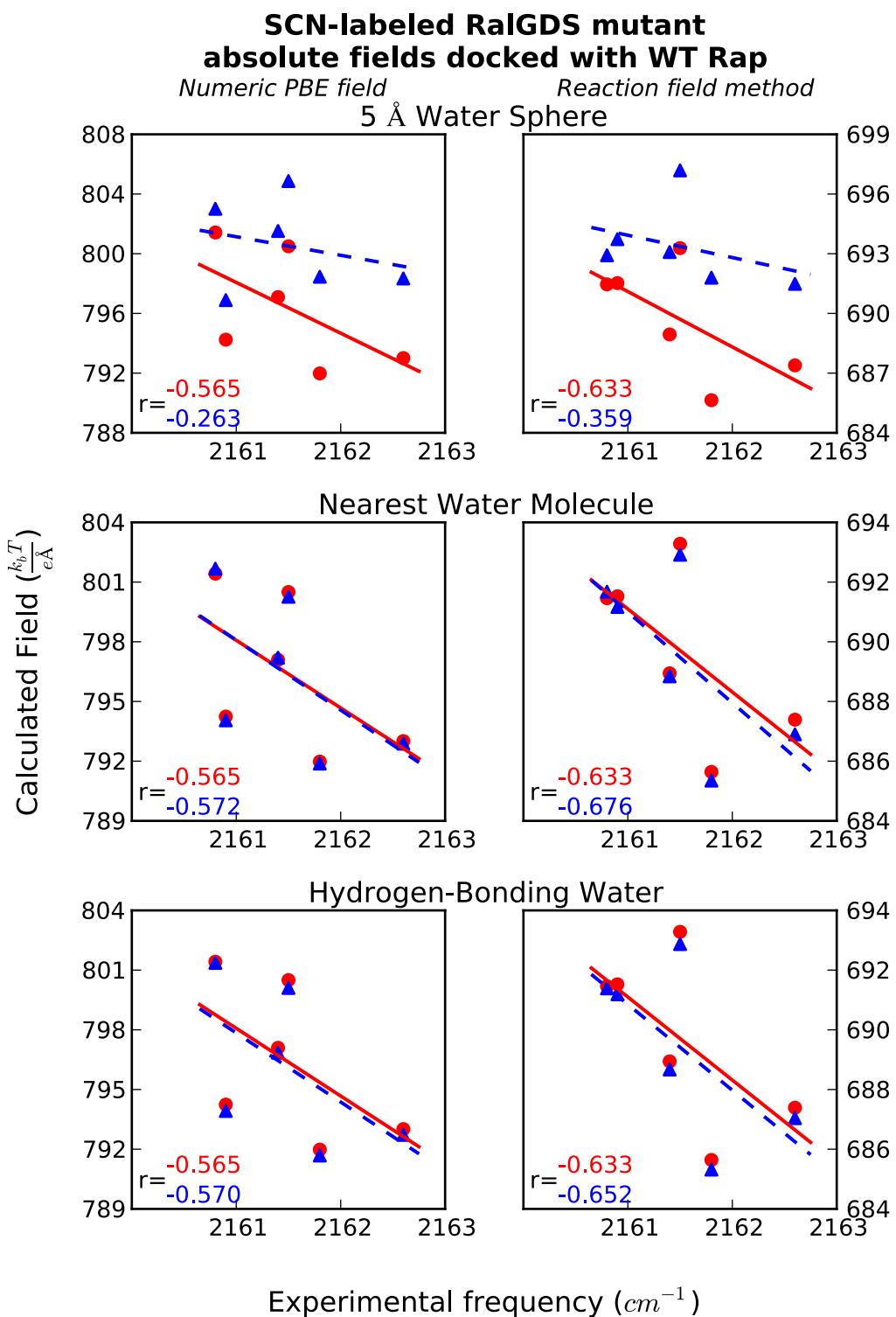


Figure 6-2: PB Absolute Field Plots – Rap E30/K31

The absolute field calculations using the numeric LPBE solutions (left column) and from the numeric reaction field method (right column) compared to the experimental frequencies for various nitrile probe locations on RalGDS docked to Rap E30/K31. Red: originally reported, entirely implicit solvent fields; blue: fields using the solvent selection model described by the row subheading. Correlation coefficients are listed in the bottom left corner in the corresponding color. Standard deviations are approximately  $\pm 1 \text{ cm}^{-1}$  and  $\pm 6.5 \text{ k}_b\text{T/e}\text{\AA}$ .

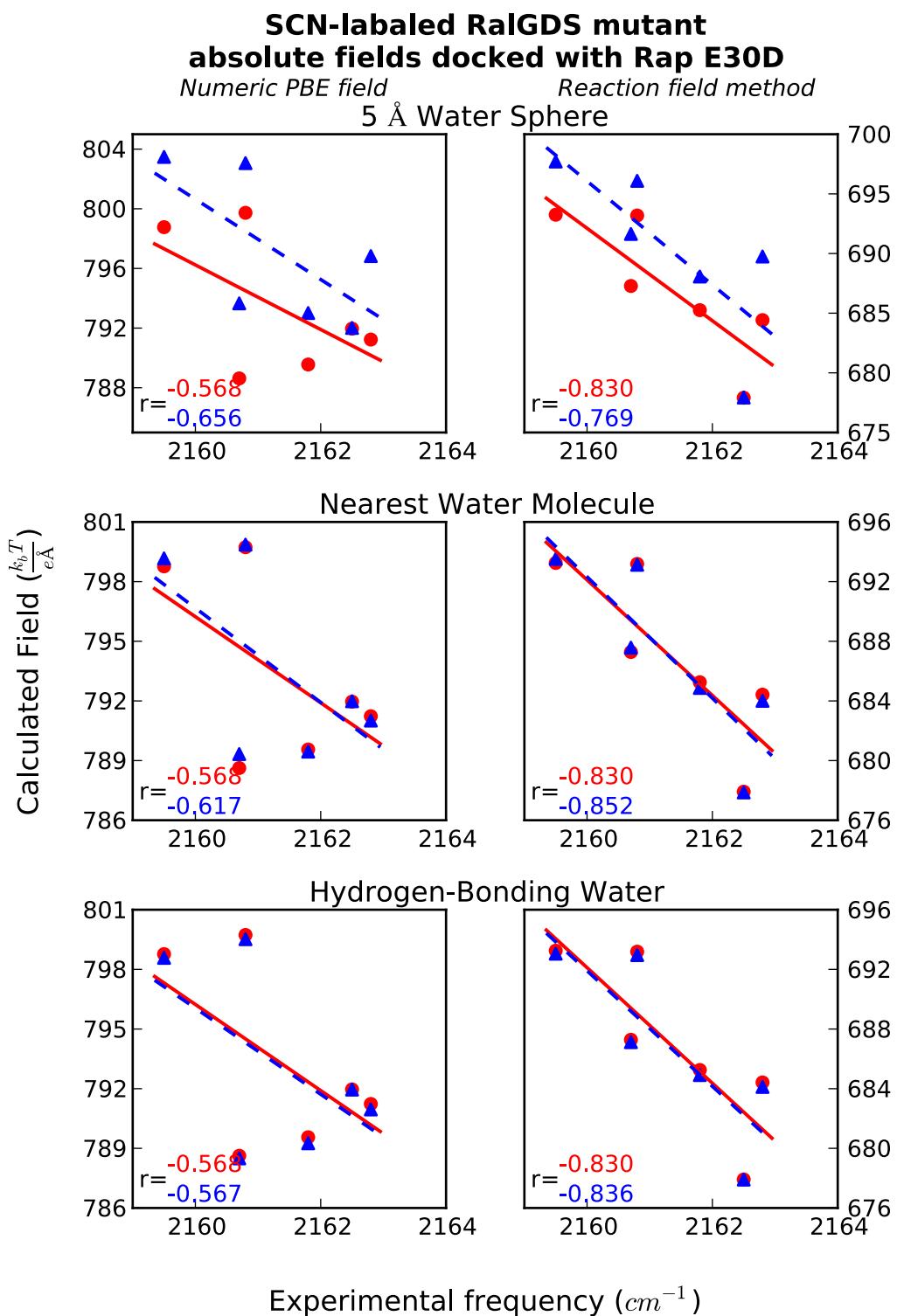


Figure 6-3: PB Absolute Field Plots – Rap E30D/K31

The absolute field calculations using the numeric LPBE solutions (left column) and from the numeric reaction field method (right column) compared to the experimental frequencies for various nitrile probe locations on RalGDS docked to Rap E30D/K31. Red: originally reported, entirely implicit solvent fields; blue: fields using the solvent selection model described by the row subheading. Correlation coefficients are listed in the bottom left corner in the corresponding color. Standard deviations are approximately  $\pm 1 \text{ cm}^{-1}$  and  $\pm 6.5 \text{ k}_\text{b}\text{T}/\text{e}\text{\AA}$ .

**SCN-labeled RaIGDS mutant  
absolute fields docked with Rap K31E**

Numeric PBE field

Reaction field method

5 Å Water Sphere

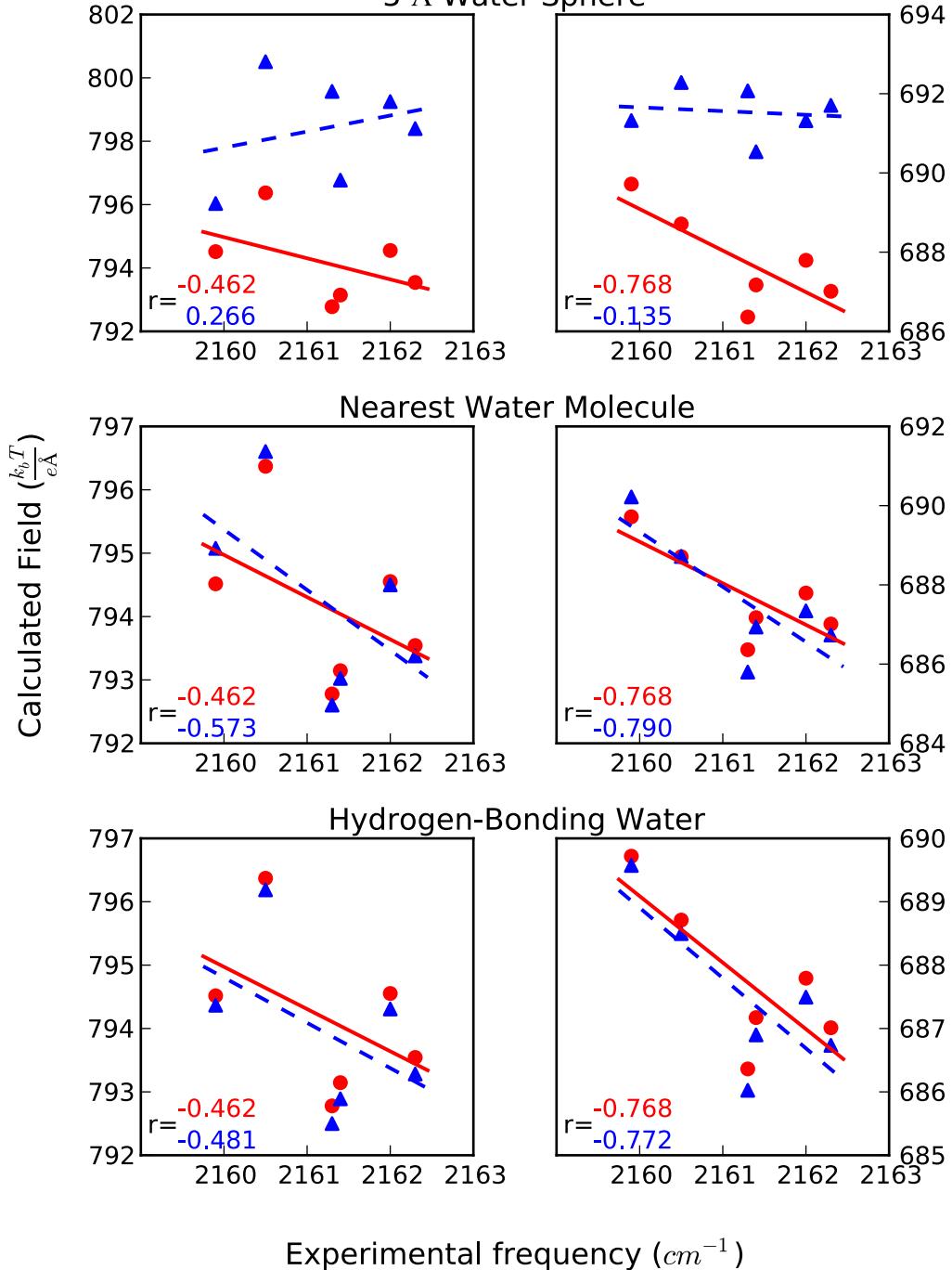


Figure 6-4: PB Absolute Field Plots – Rap E30/K31E

The absolute field calculations using the numeric LPBE solutions (left column) and from the numeric reaction field method (right column) compared to the experimental frequencies for various nitrile probe locations on RalGDS docked to Rap E30/K31E. Red: originally reported, entirely implicit solvent fields; blue: fields using the solvent selection model described by the row subheading. Correlation coefficients are listed in the bottom left corner in the corresponding color. Standard deviations are approximately  $\pm 1 \text{ cm}^{-1}$  and  $\pm 6.5 \text{ k}_\text{b}\text{T}/\text{e}\text{\AA}$ .

**SCN-labeled RaIGDS mutant  
absolute fields docked with Rap E30D/K31E**

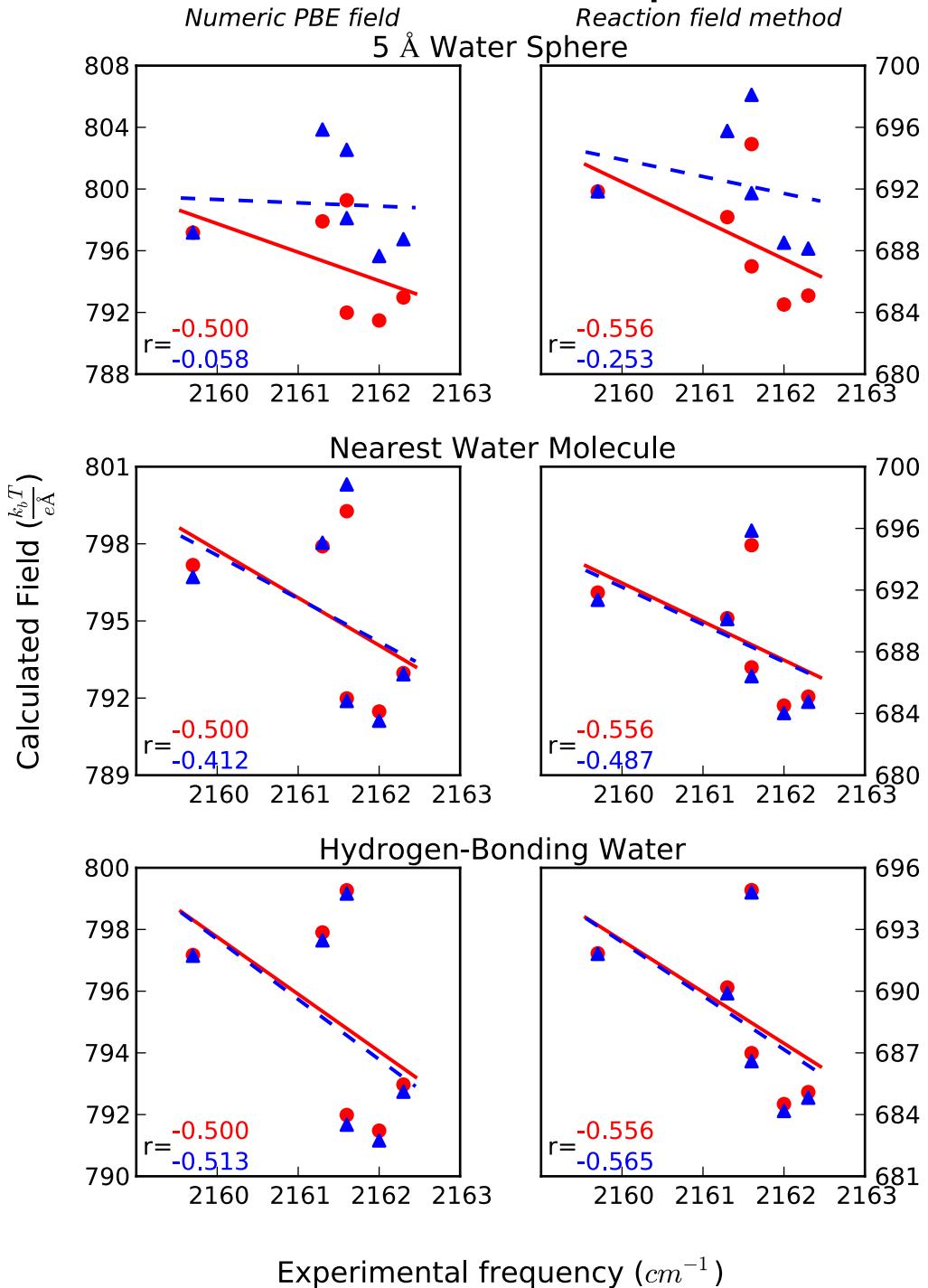


Figure 6-5: PB Absolute Field Plots – Rap E30D/K31E

The absolute field calculations using the numeric LPBE solutions (left column) and from the numeric reaction field method (right column) compared to the experimental frequencies for various nitrile probe locations on RalGDS docked to Rap E30D/K31E. Red: originally reported, entirely implicit solvent fields; blue: fields using the solvent selection model described by the row subheading. Correlation coefficients are listed in the bottom left corner in the corresponding color. Standard deviations are approximately  $\pm 1 \text{ cm}^{-1}$  and  $\pm 6.5 \text{ k}_\text{b}\text{T}/\text{e}\text{\AA}$ .

### SCN-labeled RaIGDS Monomers

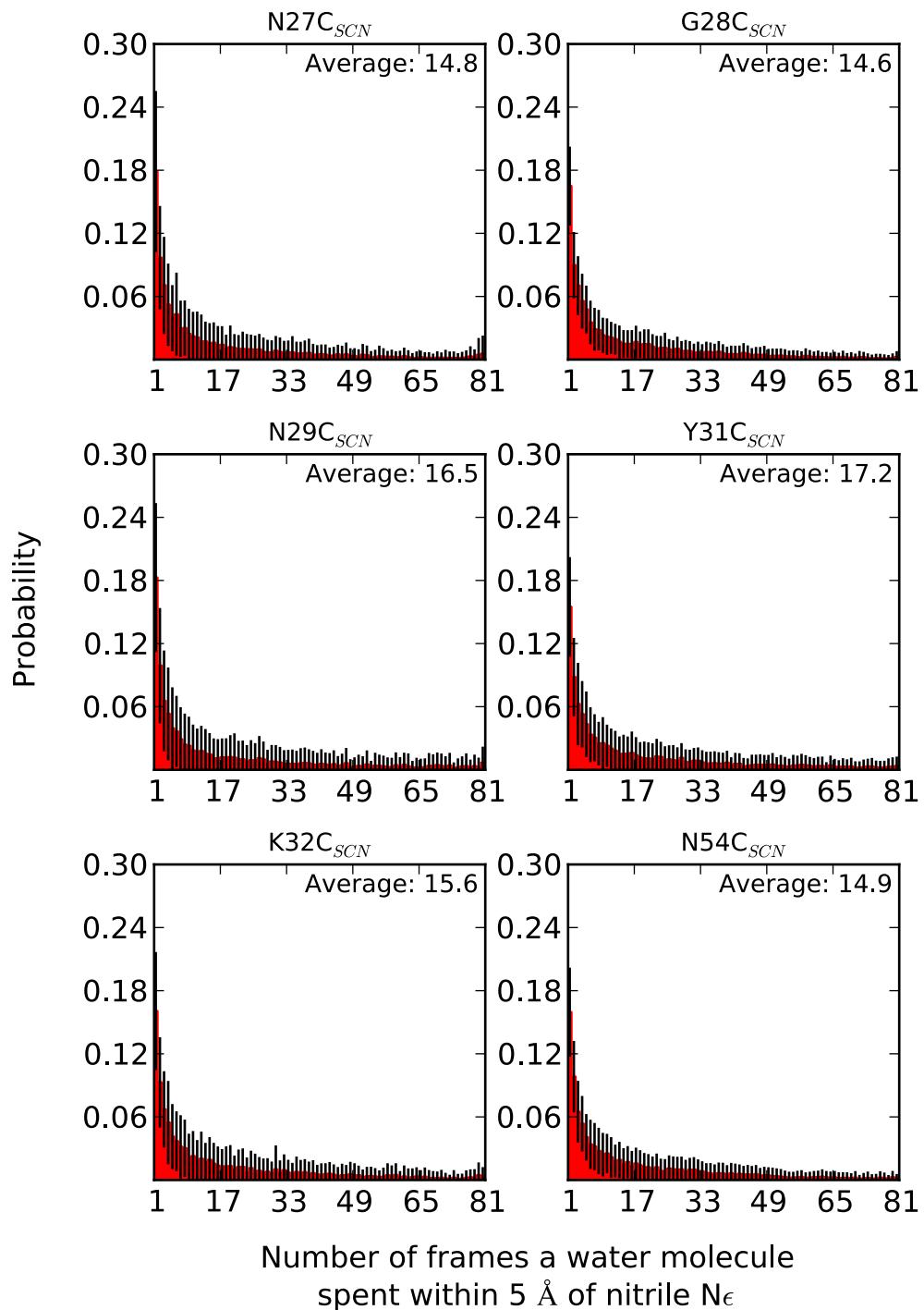


Figure 6-6: Probability of Water Lifetimes - Monomer

Observed probability of a water molecule being within 5 Å of the nitrile Nε for some number of frames in the simulation for various probe locations on the RalGDS monomer. The error bar on each bin is the standard deviation among the 144 trajectories for that bin.

SCN-labeled RalGDS mutants docked with Rap WT

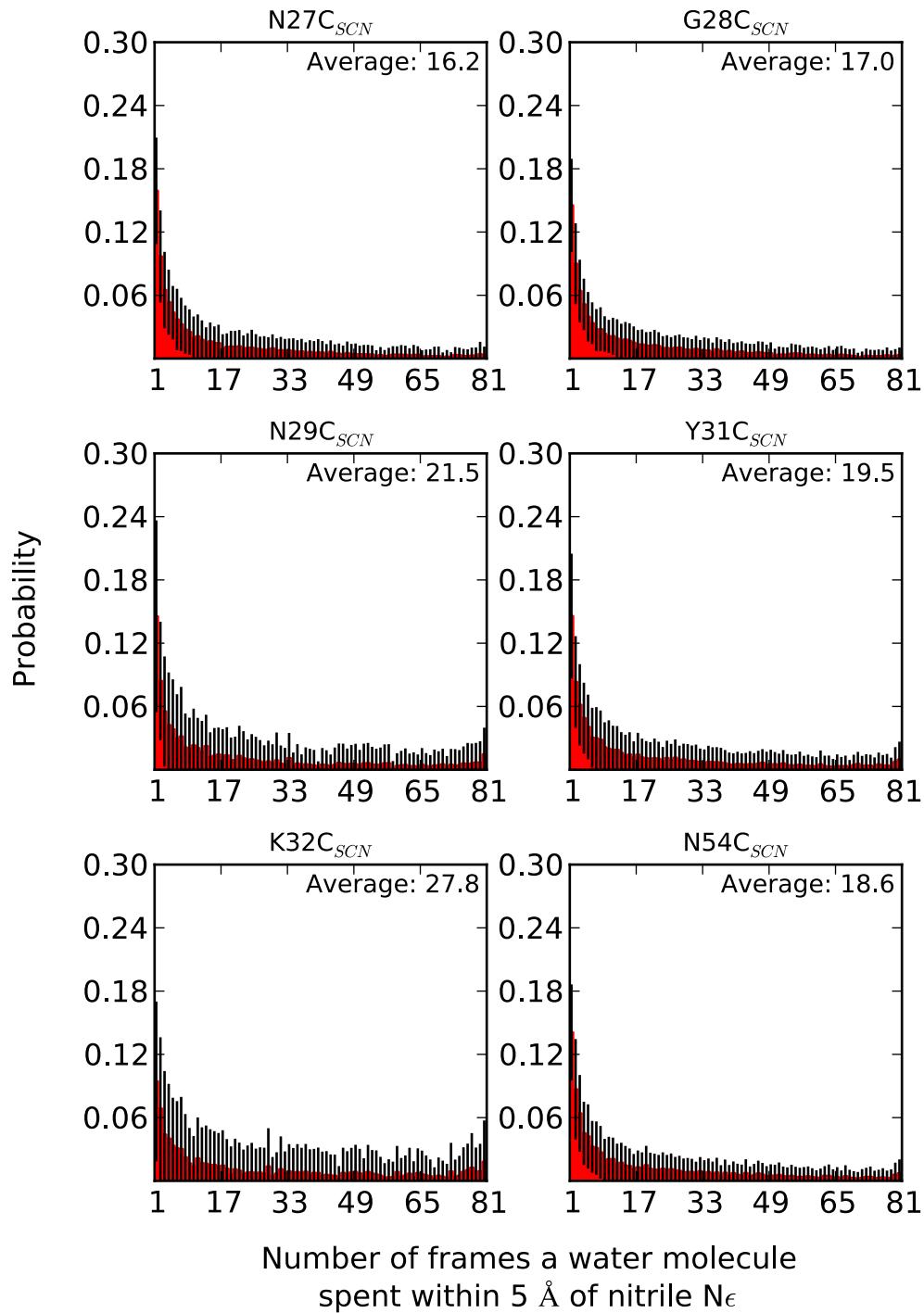


Figure 6-7: Probability of Water Lifetimes – Rap E30/K31

Observed probability of a water molecule being within 5 Å of the nitrile Nε for some number of frames in the simulation for various probe locations on the RalGDS monomer docked to Rap E30/K31. Probabilities are not Boltzmann-weighted. The error bar on each bin is the standard deviation among the 144 trajectories for that bin.

SCN-labeled RalGDS mutants docked with Rap E30D

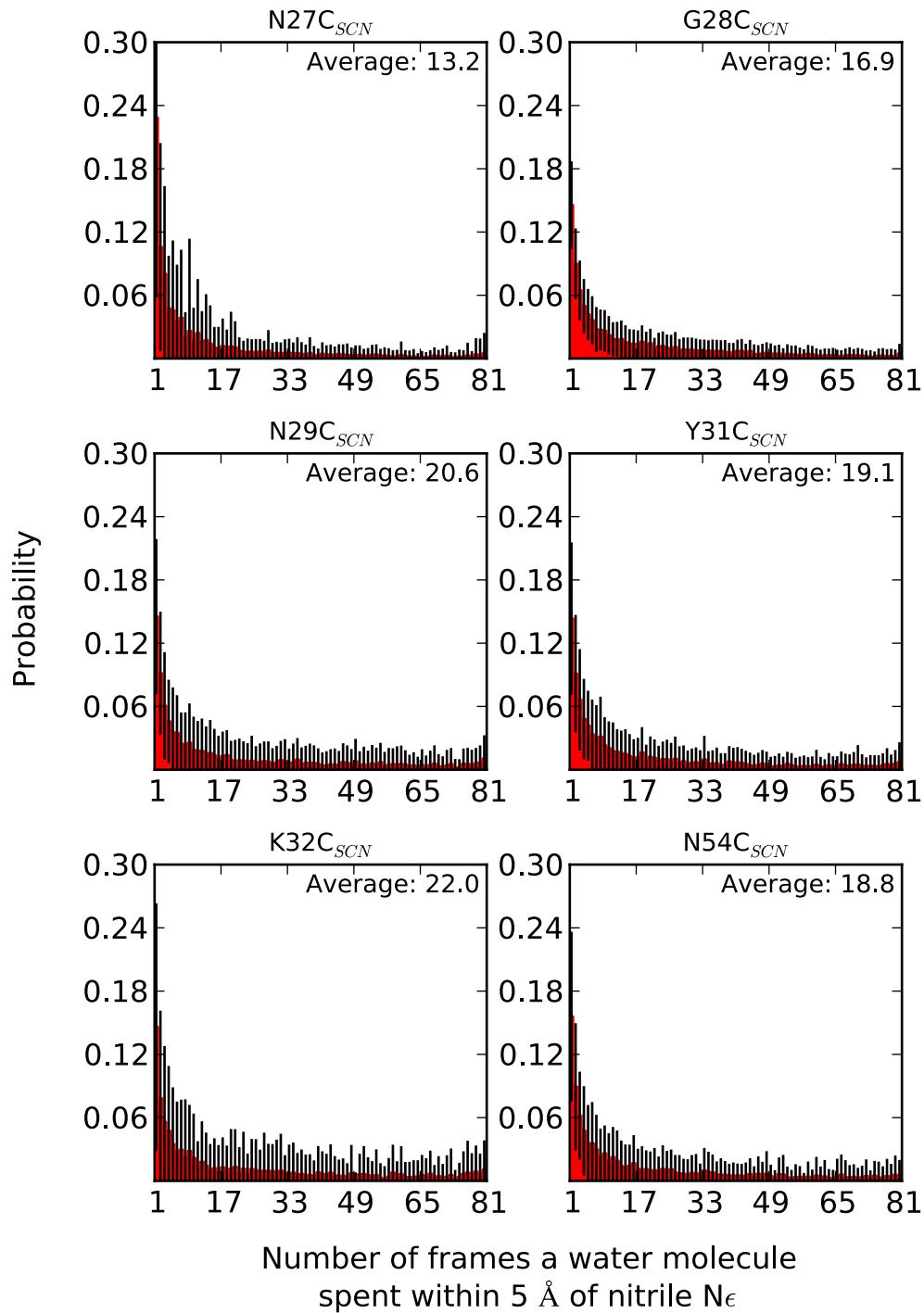


Figure 6-8: Probability of Water Lifetimes – Rap E30D/K31

Observed probability of a water molecule being within 5 Å of the nitrile Nε for some number of frames in the simulation for various probe locations on the RalGDS monomer docked to Rap E30D/K31. The error bar on each bin is the standard deviation among the 144 trajectories for that bin.

SCN-labeled RaIGDS mutants docked with Rap K31E

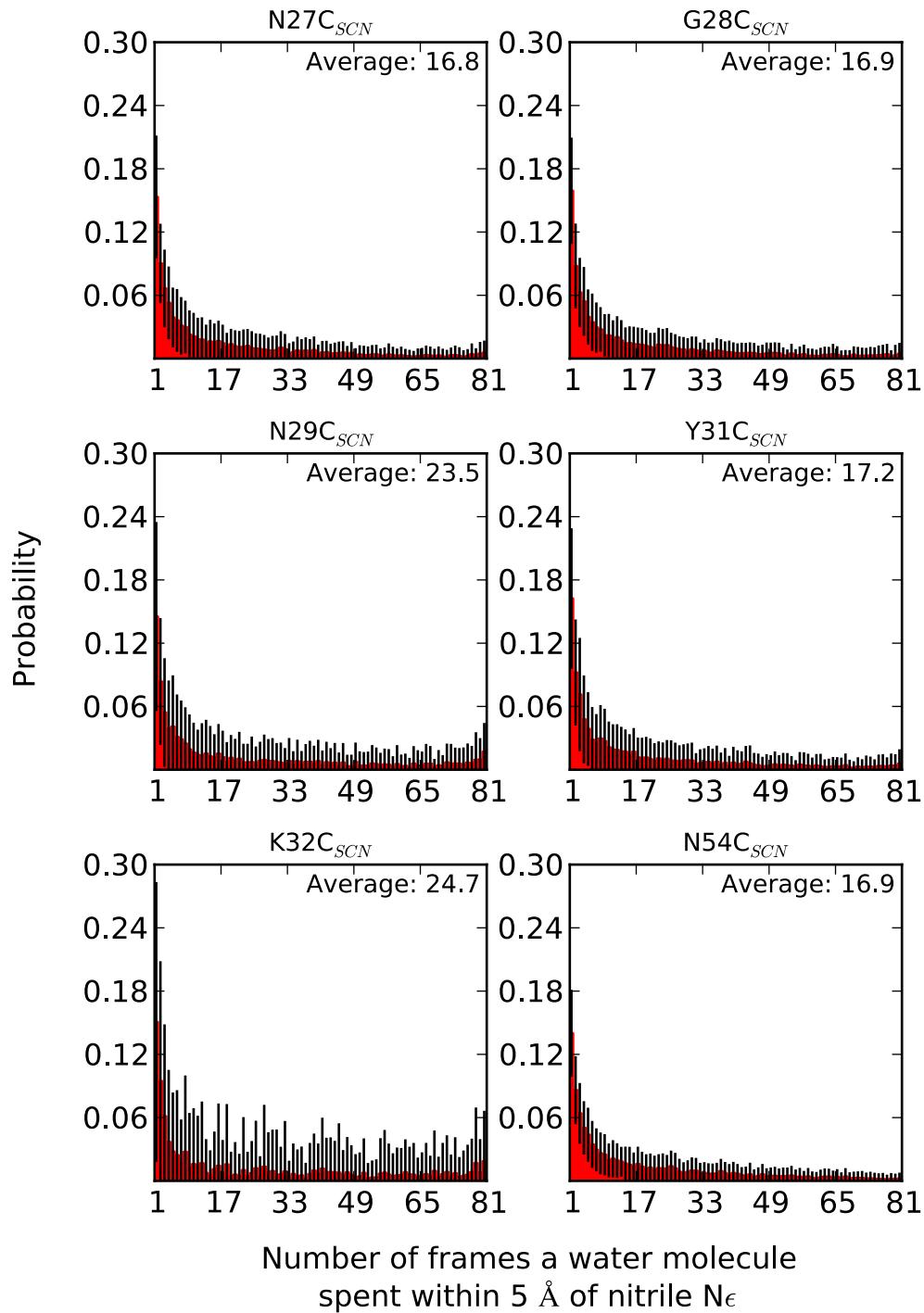


Figure 6-9: Probability of Water Lifetimes – Rap E30/K31E

Observed probability of a water molecule being within 5 Å of the nitrile Nε for some number of frames in the simulation for various probe locations on the RalGDS monomer docked to Rap E30/K31E. The error bar on each bin is the standard deviation among the 144 trajectories for that bin.

SCN-labeled RaIGDS mutants docked with Rap E30D/K31E

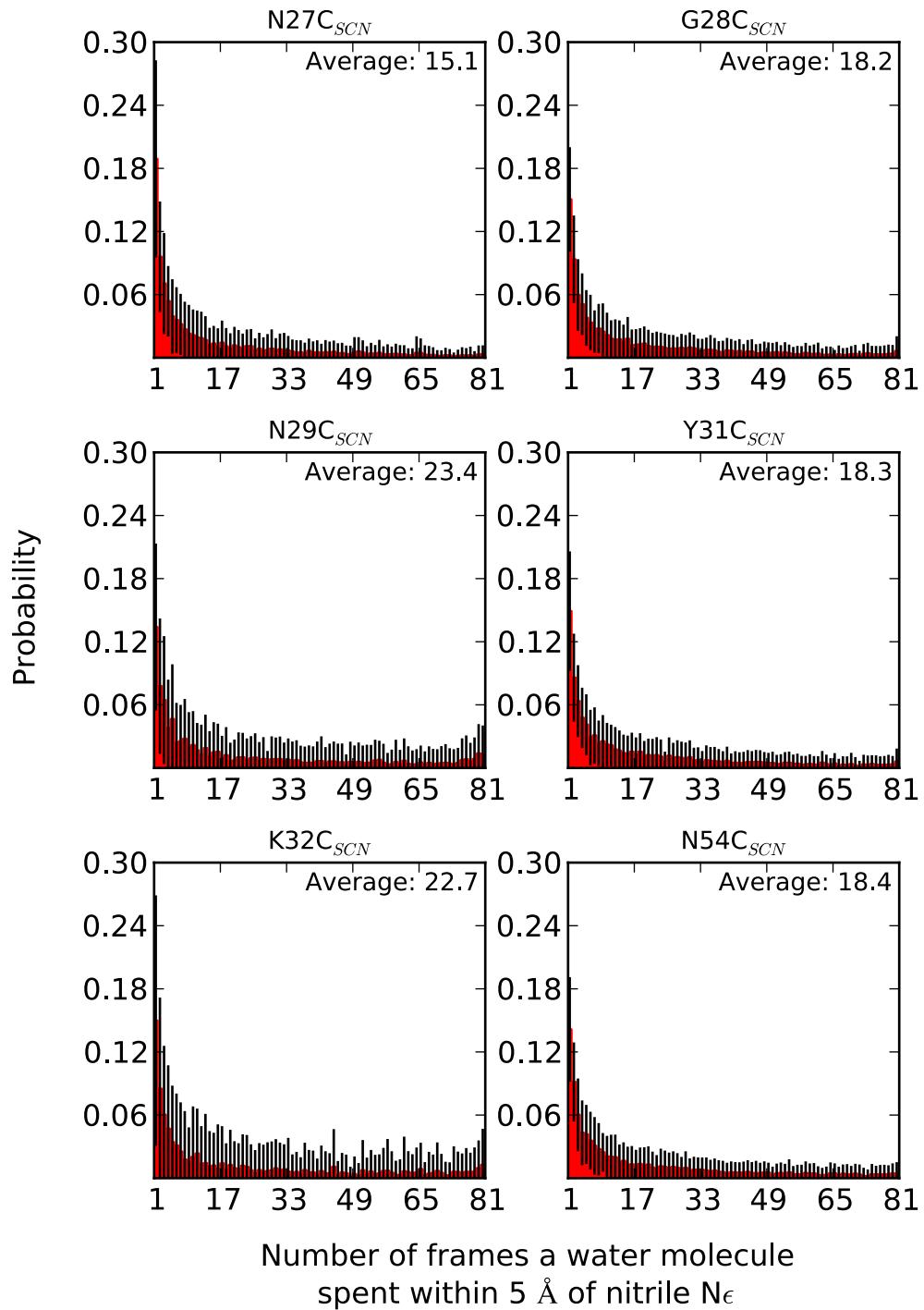


Figure 6-10: Probability of Water Lifetimes – Rap E30D/K31E

Observed probability of a water molecule being within 5 Å of the nitrile Nε for some number of frames in the simulation for various probe locations on the RalGDS monomer docked to Rap E30D/K31E. The error bar on each bin is the standard deviation among the 144 trajectories for that bin.

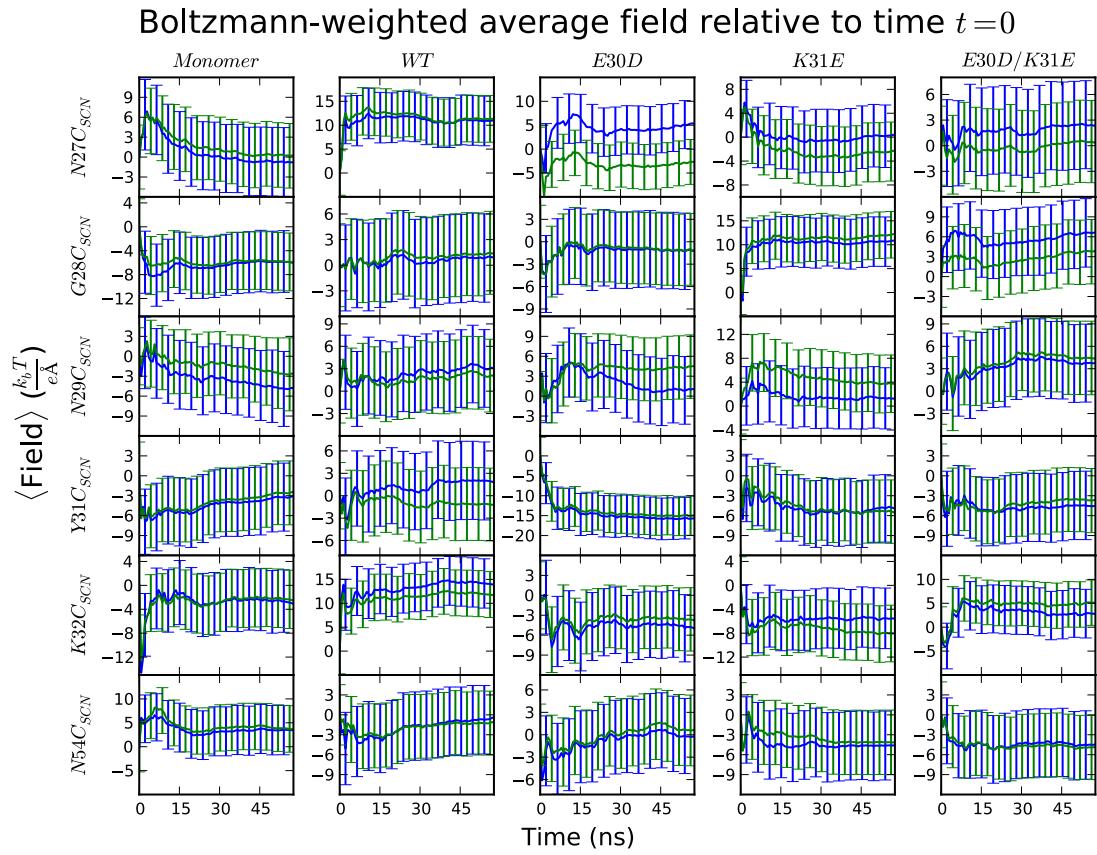


Figure 6-11: Average PB Field as a Function of Sampling Time

Boltzmann-weighted average field explicitly including water molecules within 5 Å of the nitrile Nε as a function of time relative to the field at time  $t = 0$ . Green: field calculated using the numeric field method; Blue: field calculated using the reaction field method. Error bars are standard deviations of the average field calculated from time  $t = 0$  and  $t = n\Delta t$ .

## Time gradient of the Boltzmann-weighted average field at time $t$

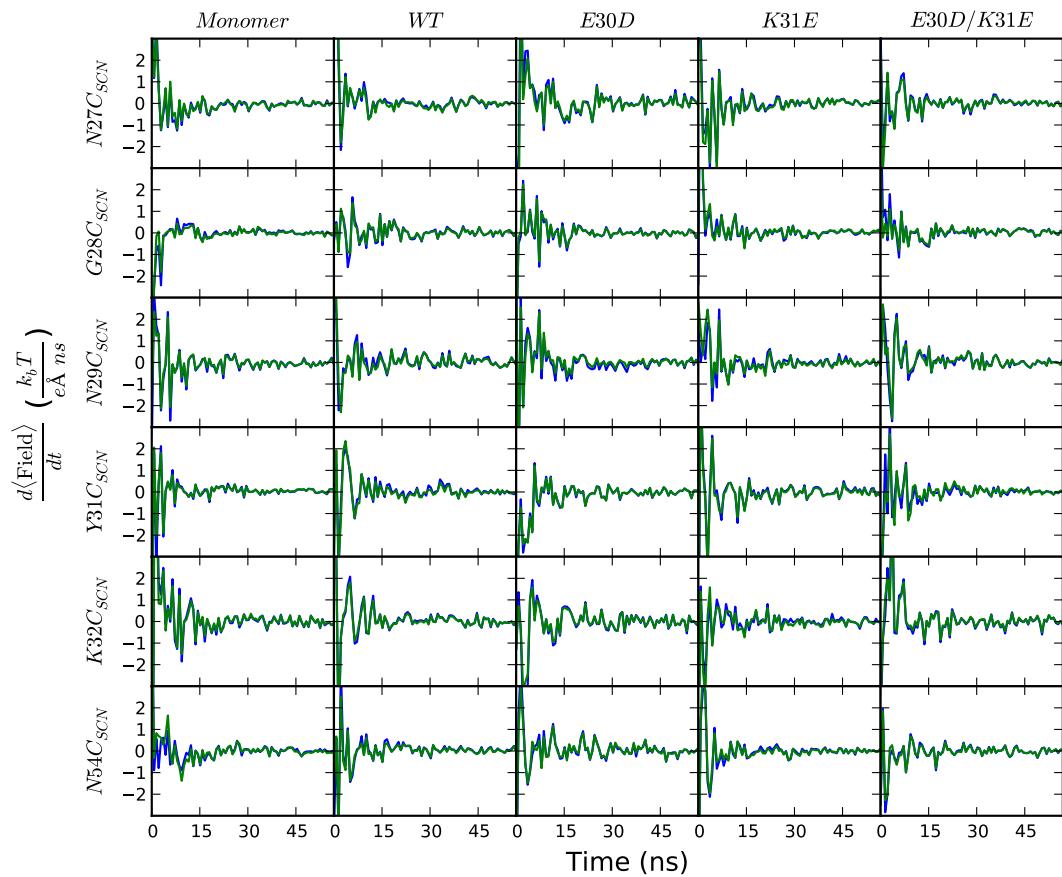


Figure 6-12: Average Field Gradient as a Function of Sampling Time

The first derivative with respect to time of the Boltzmann-weighted average field at each time step for the various Rap and RalGDS constructs. Green: numeric field method; Blue: reaction field method.

**SCN-labeled RaIGDS mutant  
field differences docked with WT Rap**  
 Numeric PBE field      Reaction field method  
 5 Å Water Sphere

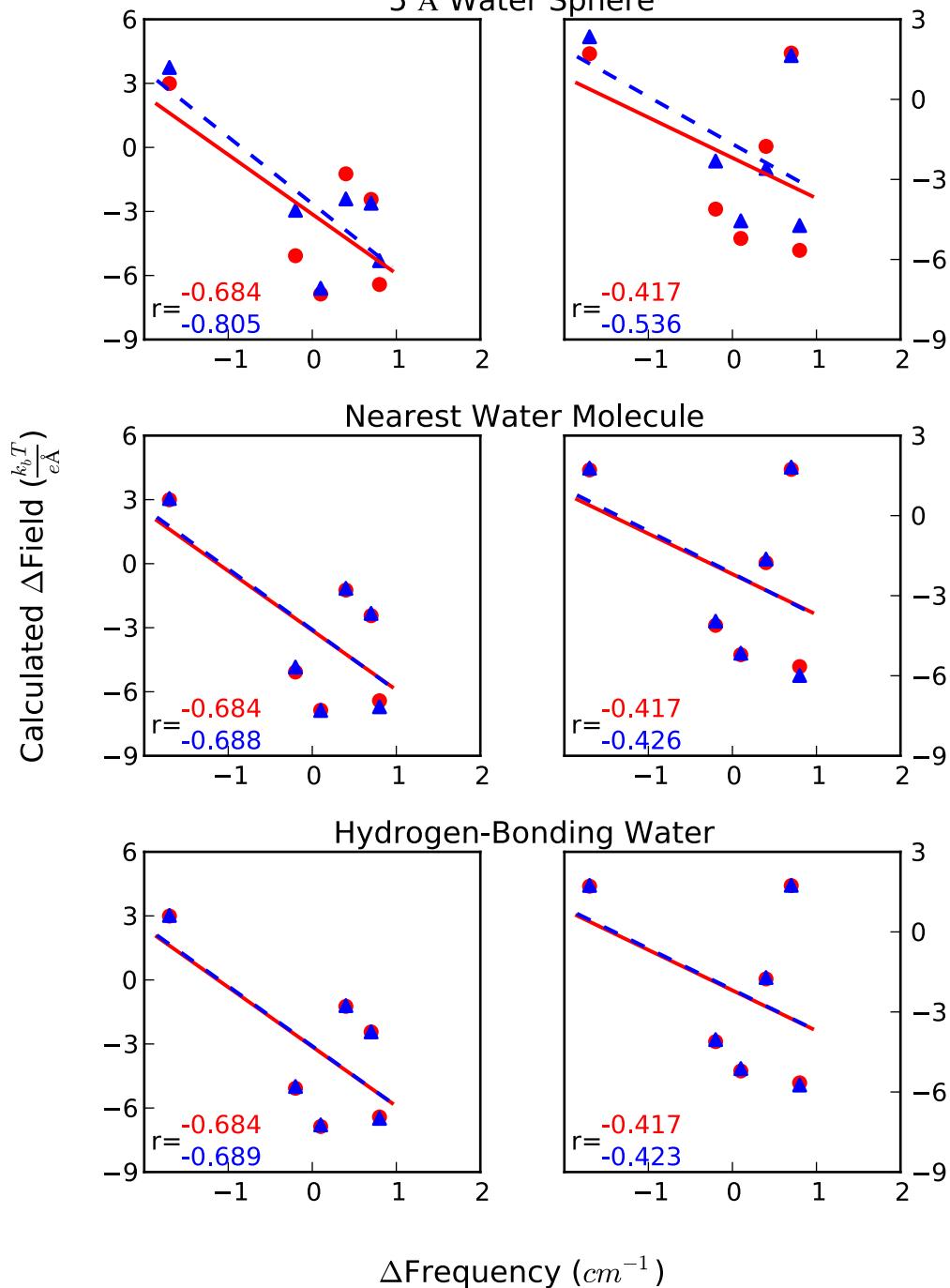


Figure 6-13: PB Relative Field Plots – Rap E30/K31

The relative field calculations using the numeric LPBE solutions (left column) and from the numeric reaction field method (right column) compared to the experimental changes in frequencies for various nitrile probe locations on RalGDS upon docking to Rap E30/K31. Red: originally reported, entirely implicit solvent fields; blue: fields using the solvent selection model described by the row subheading. Correlation coefficients are listed in the bottom left corner in the corresponding color. Standard deviations are approximately  $\pm 1 \text{ cm}^{-1}$  and  $\pm 6.5 \text{ k}_\text{b}\text{T/e\AA}$ .

**SCN-labeled RaIGDS mutant  
field differences docked with Rap E30D**  
 Numeric PBE field      Reaction field method  
 5 Å Water Sphere

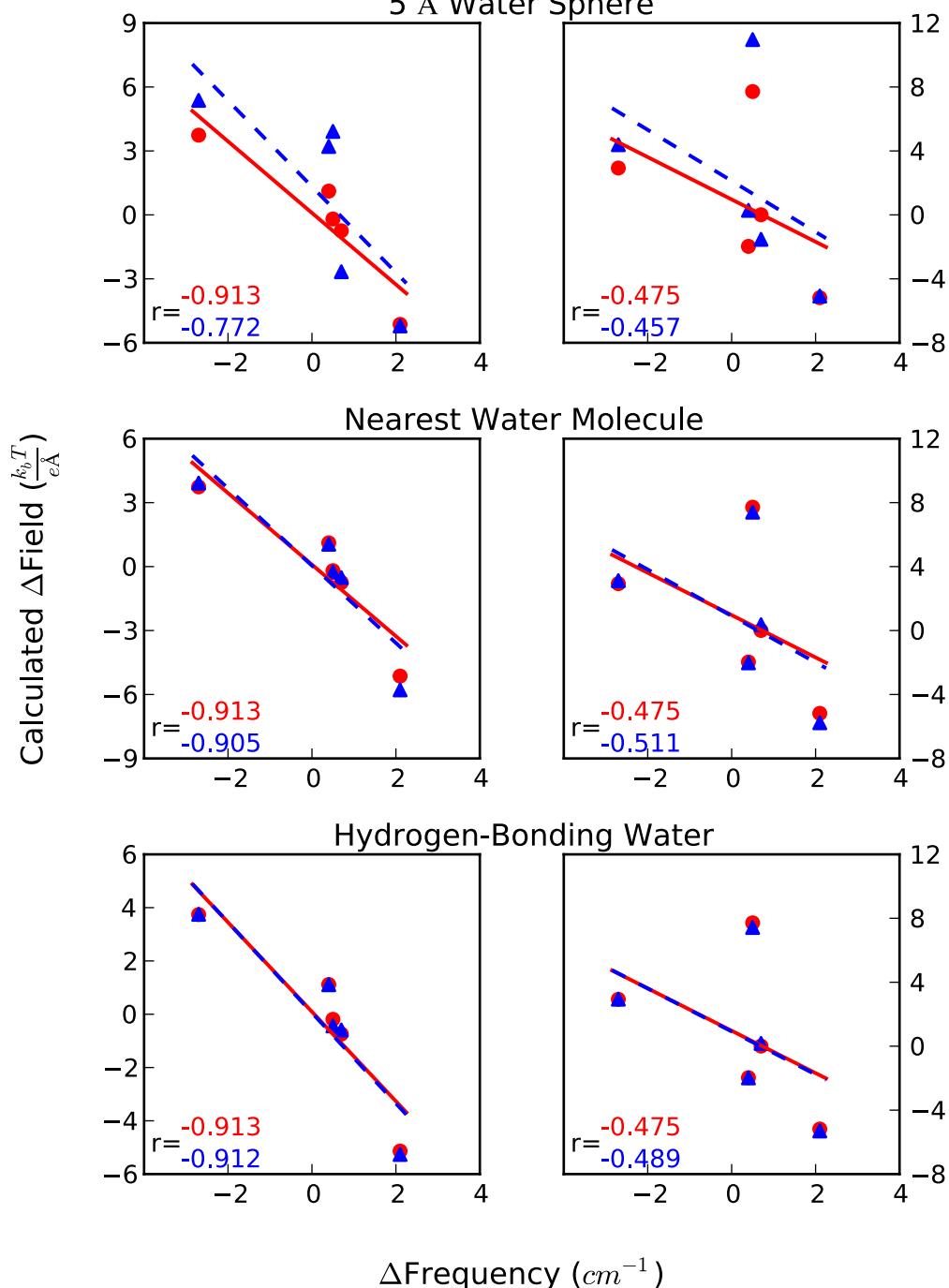


Figure 6-14: PB Relative Field Plots – Rap E30D/K31

The relative field calculations using the numeric LPBE solutions (left column) and from the numeric reaction field method (right column) compared to the experimental changes in frequencies for various nitrile probe locations on RalGDS upon docking to Rap E30D/K31. Red: originally reported, entirely implicit solvent fields; blue: fields using the solvent selection model described by the row subheading. Correlation coefficients are listed in the bottom left corner in the corresponding color. Standard deviations are approximately  $\pm 1 \text{ cm}^{-1}$  and  $\pm 6.5 \text{ k}_\text{b}\text{T/e}\text{\AA}$ .

**SCN-labeled RalGDS mutant  
field differences docked with Rap K31E**  
 Numeric PBE field      Reaction field method

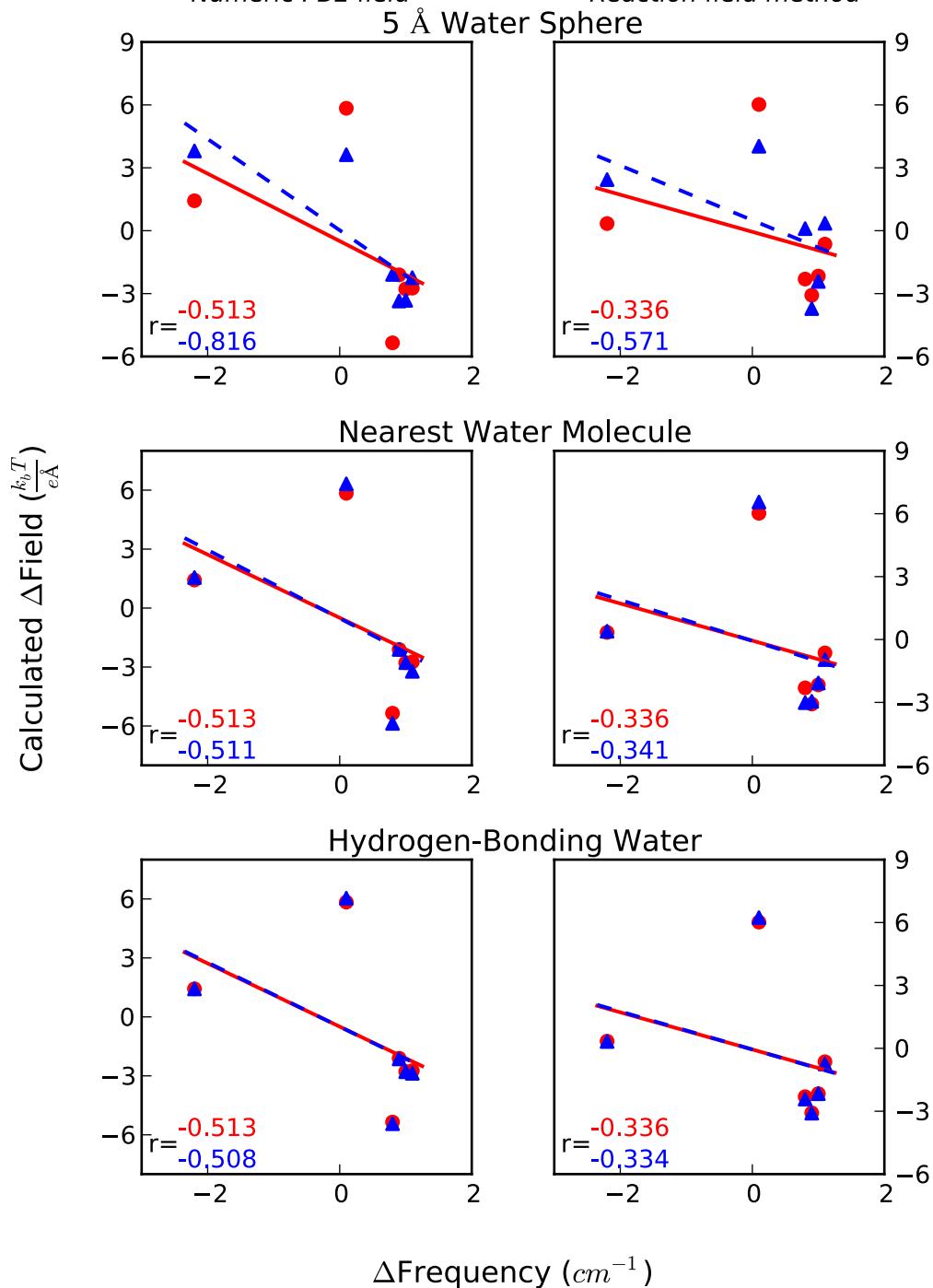


Figure 6-15: PB Relative Field Plots – Rap E30/K31E

The relative field calculations using the numeric LPBE solutions (left column) and from the numeric reaction field method (right column) compared to the experimental changes in frequencies for various nitrile probe locations on RalGDS upon docking to Rap E30/K31E. Red: originally reported, entirely implicit solvent fields; blue: fields using the solvent selection model described by the row subheading. Correlation coefficients are listed in the bottom left corner in the corresponding color. Standard deviations are approximately  $\pm 1 \text{ cm}^{-1}$  and  $\pm 6.5 \text{ k}_\text{b}\text{T/e\AA}$ .

**SCN-labeled RaIGDS mutant  
field differences docked with Rap E30D/K31E**

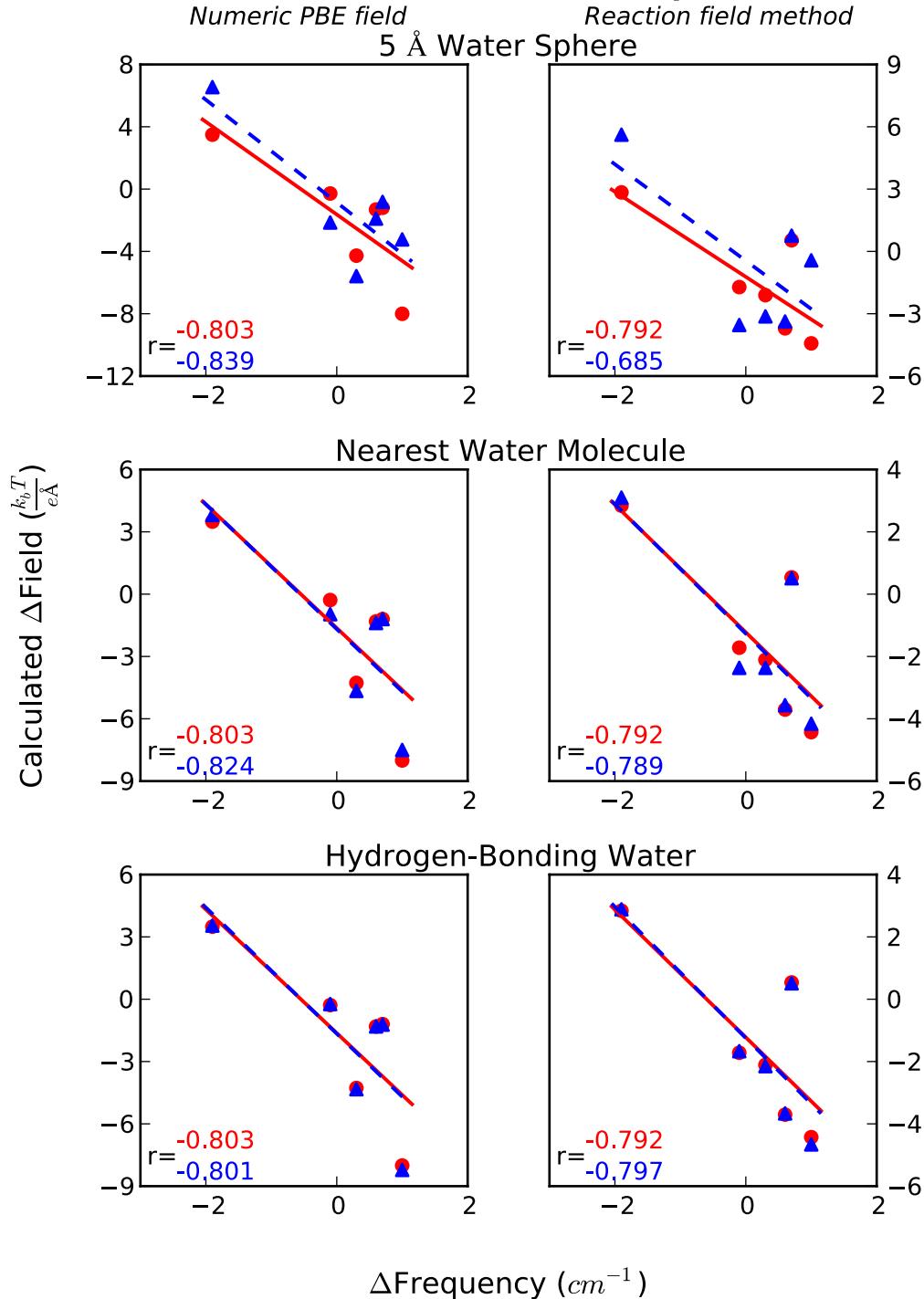


Figure 6-16: PB Relative Field Plots – Rap E30D/K31E

The relative field calculations using the numeric LPBE solutions (left column) and from the numeric reaction field method (right column) compared to the experimental changes in frequencies for various nitrile probe locations on RalGDS upon docking to Rap E30D/K31E. Red: originally reported, entirely implicit solvent fields; blue: fields using the solvent selection model described by the row subheading. Correlation coefficients are listed in the bottom left corner in the corresponding color. Standard deviations are approximately  $\pm 1 \text{ cm}^{-1}$  and  $\pm 6.5 \text{ k}_\text{b}\text{T/e}\text{\AA}$ .

## **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**

One of the principle grievances with PB electrostatics is the arbitrary choice of solute dielectric. In fact, it can be trivially shown that the solute dielectric is just a scaling factor and can be adjusted after results are known to force the calculated field values to yield experimental Stark shifts consistent with known Stark tuning rates. While this may be beneficial from the stance of a machine learning algorithm where the relationship is the most important factor and the physics are just an afterthought, it is unsatisfying for trying to predict fields of new or interesting molecules. In fact, it is my opinion that the ideal use of these calculations would be to calculate the field in regions of biological molecules that do not contain, and therefore are not perturbed (no matter how slightly), by a VSE probe. This would allow for targeted drug design that is not dependent on a (even minimally) perturbative probe. Unfortunately, for any electrostatic model to work thusly, a higher degree of confidence in the physical veracity is required, which includes removing the dependence on a dielectric constant, which is nonphysical and inappropriate for an atomistic, biological system.

A dielectric constant is a macroscopic bulk property describing the atomic polarizability of a material. At the atomic level, however, a dielectric constant is a relatively meaningless quantity that acts to indiscriminately screen electric charge. Water is known to have a relatively high dielectric of 78-80 at 298 K. This high dielectric is a result of each water molecule's ability to change its orientation in response to an external electrostatic field. The total energy of the system is minimized when the molecular

dipole moment is aligned parallel to the electrostatic field. This creates a local electrostatic field, generated by the water molecule, which is antiparallel to the external field. The total electrostatic field is the sum of the local molecular field and the external field. For any point that is separated from the field source by a rearranged dipole, the effective field is reduced or screened by that dipole. In contrast, a protein interior is significantly limited in the rotational degrees of freedom of side chains, and therefore has less capable of reorienting in response to a local electrostatic field. This results in a lower effective dielectric constant and less charge screening. Protein side chains can, however, respond to a local field via an induced dipole moment, which has the same effect as rotating a permanent dipole moment and reducing the effect field further from the source. Conventional point charge force fields cannot account for the induced dipole moments directly. For these reasons, we have recently considered the polarizable AMOEBA force field.

In this work we examine a variety of classical field calculation methods. As a reference, we reinvestigated using the RFM and entirely implicit solvent in Ambe03 with a  $10^3$  Å<sup>3</sup> second-stage box and 193 grid points in each dimension centered on the nitrile Cδ, which has been labeled PB in this study. Because these parameters yielded the most consistent results, they were deemed an appropriate reference to test each method against, where the goal was to find a better electrostatic method. We also reexamined using 5 Å explicit water sphere also with a  $10^3$  Å<sup>3</sup> second-stage box and 193 grid points in each dimension centered on the nitrile Cδ to reevaluate included select explicit solvent molecules, given a larger ensemble. Also in Amber03, we tested using explicit TIP3P water with GROMACS reaction field electrostatics, as well as the hybrid TIP3P solvent reaction field electrostatics with solute coulomb field.

Using the same ensembles, we investigated using AMOEBA self-consistent multipoles with implicit, PB solvent as well as with explicit AMOEBA water. In both AMOEBA field methods we also considered the charge penetration model previously described, which separates short-range electrostatic interactions into a positive effective core contribution and a damped electron distribution interaction. We looked at using both the intuitive charge penetration  $\alpha$  parameters (CP) as well as the fitted charge penetration  $\alpha$  parameters (CPf). In total, we compared our experimental absorption energies to fields calculated using 10 different classical electrostatic methods.

In addition to examining a variety of classical electrostatic field models, we significantly increased the simulation time for each 2D umbrella window from 0.4 ns to 2.0 ns each, for a total of 288 ns for each system. Although we previously demonstrated that our simulations had converged to a single average field value (Figure 6-11, Figure 6-12), the question of simulation convergence can never been entirely satisfied. The simulations were increased by a factor of five to allow further side chain and solvent sampling while still using umbrella sampling to fully explore all nitrile orientations and reduce potential criticism regarding insufficient sampling. **\*\*\*Can I say that, or is it too informal?** Furthermore, because our lab has collected data on both Rap as well as Ras, in addition to the Rap GTPases previously simulated, we have also included simulations on Ras D30/E31, Ras D30E/E31, Ras D30/E31K, and Ras D30E/E31K, each bound to each of the six previously discussed nitrile probes. In total 54 different systems were each simulated for 288 ns, resulting in 15,552 total ns of simulation.

## 7.2 RESULTS AND DISCUSSION

Each electrostatic field method was plotted against the experimental vibrational absorption frequencies for the appropriate systems and the resulting correlation

coefficients and virtual Stark tuning rates have been tabulated in Table 7-1. The table has been broken up into three sections: a single GTPase mutation, indicated in the leftmost column, and all of the probes it could be docked to; a single probe, indicated in the leftmost column, and all of the GTPase systems it could be docked to (including the undocked, monomeric stat); all 54 systems, indicated by "All Points". Each major column shows a different electrostatic method. Change in electrostatic field due to the monomer docking to each GTPase system is tabulated in Table 7-2, where the data is presented in the same manner as Table 7-1. The case where a single probe location is measured when docked to each GTPase is excluded because all the results would be shifted by a constant amount, whatever the particular shift and calculated field is for that monomeric probe.

### 7.2.1 Electrostatic Fields Along the Interface of Each GTPase

First we examine how well a single mutation is seen at different points along the Ral surface by comparing the field calculated at each probe site when docked to the same GTPase. Moving the probe location while keeping the mutation constant allows us to look at the field at multiple locations in the protein while keeping the cause of the field constant. By then making a mutation and scanning across the surface of the protein again, we can see how well changes in the field due to a mutation are also calculated. This is the way data has previously been presented and is included for consistency.

The first important observation is that no single method stands out as the "best" field calculation method. There is a case in where AMOEBA, AMOEBA with intuitive charge penetration parameters (AMOEBA CP), and AMOEBA with fitted charge penetration parameters (AMOEBA CPf) all have the largest magnitude in correlation (Rap E30/K31, Ral). There is a case where AMOEBA with explicit solvent (with and

without charge penetration corrections) has the highest magnitude in correlation (Ras D30/E31K). Sometimes APBS has the highest correlation magnitude (Rap E30D/K31, Rap E30D/K31E) and other times the hybrid TIP3P reaction field has the highest correlation magnitude (Rap E30/K31E, Ras D30E/E31K, Ras D30E/E31). There are also cases where each respective model has the lowest correlation magnitude. Furthermore, there is no consistent trend regarding the sign of the *VSTR*. In general, the *VSTR* is negative, although there are cases in each model where a positive value is calculated. The known Stark tuning rate is positive,  $1.99 \text{ cm}^{-1}/(k_b T/e\text{\AA})$ , and none of the methods yield a consistently positive *VSTR*. The direction of the correlation, and therefore sign of the *VSTR*, will be addressed further, but at this point it is sufficient to say that none of the models are consistently able to model the changes in fields due to changing locations of the probe with any significant degree of consistency.

One of the reasons for using AMOEBA was the inclusion of atomic polarizability, eliminating the need to select for a protein dielectric based on the observed *VSTR*. Put another way, we hypothesized that AMOEBA would do a better job at predicting a *VSTR* which matched the experiment Stark tuning rate. However, we observed that the average order of correct magnitudes was AMOEBA (CP, CPf) > APBS > hybrid TIP3P reaction field > 5 Å water sphere > experimental Stark tuning rate > AMOEBA explicit water (CP, CPf) > explicit TIP3P reaction field electrostatics. Implicit solvent AMOEBA was further from the experimental Stark tuning rate than implicit solvent Amber03 while explicit solvent AMOEBA was also further from the experimental Stark tuning rate than explicit TIP3P solvent in Amber03.

## 7.2.2 Electrostatic Fields at a Single Point on Ral

The advantage of looking at a single probe and making changes elsewhere in the system, while leaving the probe location constant, is we are able to see how the probe's local field changes as a function of relatively distant residues without having to be concerned with changes due to the local environment, such as different degrees of solvent accessibility or hydrogen bonding. The local structure of the probe and its immediate surroundings are relatively constant (which is easily confirmed with MD) and the only changes being observed are on the binding partner.

Contrary to looking at a single GTPase and different probes, a single electrostatic method did consistently yield the highest (or close to highest) correlation to experimental measurement—APBS. The only exceptions are K32C and N54C, where the magnitude of the correlation is similar for PB as for the most correlated method, but the sign was opposite.

Once again, we look at how the magnitudes of the *VSTR* compare to the experimental Stark tuning rate: APBS > hybrid TIP3P reaction field > AMOEBA (CP, CPf) > experimental Stark tuning rate > 5 Å water sphere > explicit TIP3P reaction field electrostatics > AMOEBA explicit water (CP, CPf). In this, implicit solvent AMOEBA best predicted the magnitude of the Stark tuning rate. Because we are not looking at a single point in space and making changes elsewhere (rather than making changes at a single point in space and observing that one change from multiple perspectives), it was hypothesized that a single, uniform solute dielectric (and therefore *VSTR*) would be more appropriate. If a protein interior is not a uniform dielectric, then observing the same change at two different points can be complicated by varying amounts of charge screening between the changed potential and the observed field location. Observing changes from a single point, where the electrostatic field propagates along the same path

to get from the mutation site to the measurement site, means the dielectric "path" should be constant. However, because AMOEBA explicit models polarizability, the dielectric "path" is irrelevant; only models that require information about a solute dielectric should significantly benefit from a single dielectric "path" which can be treated as an average dielectric constant. AMOEBA should be equally capable of looking at a many field perturbations from one location as looking at a single field perturbation at many locations. Yet, surprisingly, AMOEBA typically did a significantly better job of predicting electrostatic fields at various points in space due to a single perturbation.

### 7.2.3 Changes in Field upon Docking

Previously we observed that looking at the difference field calculations typically yielded an improved correlation to experiment, possibly by cancelling nonphysical aspects of the continuum calculation when taking the difference between two systems. That observation was not observed for all of the various field calculation strategies. Table 7-2 shows the relative field correlations and *VSTR*. Looking at fields calculated in APBS, correlation only increased significantly for Rap E30/K31 and Ras D30E/E31K (it is interesting to note that these two GTPases share amino acid identity at positions 30 and 31). In these two cases, looking at the change in field upon docking to the monomer, which had essentially zero correlation to experiment, significantly improved the correlation to experiment. For the other GTPases, the difference calculation either decreased correlation or had no significant effect. Interestingly, the two GTPases that were improved by looking at the difference calculation were, like the monomer, poorly correlated to experiment in the absolute calculations, while the other six were already moderately-to-well correlated. It is likely that the absolute fields in the monomer, Rap E30/K31, and Ras D30E/E31K are all poorly correlated to the experimental absorption

energies for the same reason(s). Taking the difference had the previously observed effect of canceling out some error, while the other six GTPases either did not have that error, or had it to a significantly less degree, and subtracting the monomer field reintroduced or exasperated that error. For every other electrostatic method, the difference calculations did not have any systematic change in the correlation to experiment.

#### 7.2.4 Overall Electrostatic Field Predictions Along Protein Surfaces

We want to know how well each model, overall, is capable of predicting the electrostatic field via the VSE. The last row on Table 7-1 and Table 7-2 show the absolute and relative correlation coefficient and *VSTR*, respectively, for each electrostatic method when all 54 systems are examined together. None of the AMOEBA methods are significantly correlated to experiment. APBS has the highest correlation (*p*-value = 0.0042), followed by the 5 Å water sphere (0.0217). When taking the difference, the correlations increase for APBS (*p*-value < 0.0001) and the hybrid TIP3P reaction field electrostatics (*p*-value = 0.0025). The explicit solvent AMOEBA (CP, CPf) calculations also saw an improved correlation, but the correlation is still too small to have likely statistical significance (*p*-values = 0.1430, 0.1878, 0.1471).

### 7.3 PB IMPLICIT WATER AND EXPLICIT TIP3P WATER ARE ELECTROSTATICALLY EQUIVALENT

We performed PB calculates on fixed orientations of two small molecules, methylthiocyanate and acetophenone, in various solvents (represented only as a dielectric constant in the PB model). We then compared them to experimental measurements of the vibrational absorption frequencies for methylthiocyanate in those solvents, obtained by Christina Ragain, Ph.D, Josh Slocum, and Kelsey Eklund, and vibrational absorption frequencies for acetophenone previously reported by Fried *et al.*,<sup>170</sup> shown in Figure 7-1.

Electrostatic fields calculated using PB implicit solvent can qualitatively reproduce the VSE only in solvents that do not hydrogen bond to the vibrational chromophore. This is not entirely surprising; PB treats solvent as a continuum of dielectric and if there is a specific interaction between the solvent and the solute that cannot be well represented by a continuum, it is neglected.

In the upper-right corner of Figure 7-1, a subpopulation of data points can be seen. Figure 7-2 focuses on the subpopulation of methylthiocyanate whose dielectric has been modulated by the addition of varying amount of glycerol (blue) and varying amounts of ethanol (red). What is interesting about this data is the strong, negative correlation between the solvent reaction field and the measured absorption energy. Although the relationship between the electrostatic field and the absorption energy does not obey the VSE, a relationship does clearly exist. Again, because PB treats solvent as a dielectric continuum, there is no regard given for specific solvent-solute interactions. In spite of that, this disregarded interaction is constant given similar-enough solvents. In the data shown, as long as the solvent consists only of either water+glycerol or water+ethanol, the missing information in PB is the same and a relationship between the PB electrostatic field and experimental absorption energy is still seen. However, trying to compare the water+glycerol data together with the water+ethanol data does not correlate to experiment because the two PB continuum solvent environments have different sources of error, which are not constant between the two solvents.

From this observation, two hypothesizes can be drawn. 1) The variation in correlations using PB electrostatics is due to the inability of PB to capture specific solute-solvent interactions, which are known to be important.<sup>17, 18, 29</sup> 2) The reason we observe negative correlation between electrostatic fields and vibrational absorption energies is because, although lacking a good description of water-nitrile interactions, those

interactions *in vitro* are relatively constant among the different systems and the missing physics are therefore constant, resulting in data which is correlated but also missing an important part of the model.

Furthermore, we have compared the solvent reaction field using PB implicit solvent to TIP3P explicit solvent in the GROMACS reaction field electrostatics model, shown in Figure 7-3. From this figure, it is clear that the explicit TIP3P solvent reaction field is linearly correlated to the implicit PB solvent reaction field with a slope of approximately 1 and an offset of approximately -5. Given that solvent reaction fields from PB do not adequately model solute-solvent interactions (Figure 7-1, Figure 7-2) and explicit TIP3P solvent reaction fields are offset PB solvent reaction fields, it is reasonable to conclude that TIP3P also does not adequately model solute-solvent interactions to the degree required for VSE calculations. Further investigation is currently being pursued in our lab using TIP3P, TIP4P, TIP5P, and AMOEBA water.

#### 7.4 GENERAL REMARKS

The GROMACS explicit TIP3P reaction field electrostatic method was typically no better than APBS, and in some cases significantly worse (Ras D30E/E31). The hybrid TIP3P reaction field method typically looked very much like the APBS field (due to both using the same solute fields), with the difference being the implicit PB SRF is used in the APBS results while the explicit TIP3P solvent reaction field is used for the hybrid method. The general agreement between the two is due to the linear relationship between the two solvent reaction fields, as previously discussed and shown in Figure 7-3.

The charge penetration corrections were not significantly different from the AMOEBA fields without charge penetration. Due to 1) the short-range nature of the correction and 2) the local structural environment was, on average, the same for a probe

docked to each GTPase construct, the average of the charge penetration correction field was approximately constant, resulting in a uniform shift for all field calculations and no overall change in correlation. It should be noted that the current state of the charge penetration corrections add them after the self-consistent induced dipole moment calculations, and it may be that further development within the procedure, including using the charge penetration corrected fields when calculating self-consistent induced dipoles and expanding the corrections from just monopole-monopole interactions to dipole and quadrupole interactions, may merit reassessment of their usefulness.

Correlations between the calculated standard deviation of the electrostatic field using the various field calculation methods and the experimental full width at half peak maximum are shown in Table 7-3. In general, there is little-to-no correlation between the two. When a statistically significant correlation is observed ( $p\text{-value} \leq 0.10$ :  $R > 0.72$  for  $N=6$ ;  $R>0.58$  for  $N=9$ ;  $R> 0.23$  for  $N=54$ ), the correlation is negative, indicating that more electrostatic field states result in a narrower vibrational absorption peak and fewer electrostatic field states result in a broader vibrational absorption peak. However, we would expect the opposite trend; broader absorption peaks should be indicative of a larger number of states which would manifest itself in simulation as a broader distribution of electrostatic field environment, and thus it is likely that the significantly correlated cases may actually be random noise.

It is also highly likely that AMOEBA field calculations underperform due to non-transferability among ensembles. In the next study, we will examine a few significantly smaller systems that have been sampled in AMOEBA.

## 7.5 CONCLUSION

From this extensive set of calculations, there are three major conclusions. 1) None of the methods are particularly reliable or consistent than our benchmark, APBS. APBS has the highest correlation with experiment ( $R=-0.384$ ,  $N=54$ ,  $p\text{-value}=0.0042$ ), but the correlation is still weak and negative instead of positive. 2) Solvent reaction fields obtained using TIP3P explicit water and PB continuum water are linearly correlated given adequate sampling. In electrostatic field calculations, using TIP3P water is equivalent to using a PB continuum of water. This is a likely reason both explicit TIP3P solvent as well as implicit PB solvent fail to adequately reproduce electrostatic fields which qualitatively obey the vibrational Stark effect. Instead, we often see correlations in the opposite direction what VSE predicts. 3) Sampling in Amber03 and performing electrostatics in AMOEBA is typically worse than performing the monopole electrostatics in Amber03. It is likely that the ensembles are not transferable. For these reasons, the next study presented uses the AMOEBA force field to simulated and calculate electrostatic fields using explicit AMOEBA water for four small molecules containing a thiocyanate probe.

Table 7-1: Correlation Coefficients (R) and Virtual Stark Tuning Rates ( $VSTR^a$ ) for Absolute Field Calculations using Various Electrostatic Models

F vs. $\bar{\nu}$	AMOEBA		AMOEBA CP		AMOEBA CPF		AMOEBA Explicit Water		AMOEBA Explicit Water CP		AMOEBA Explicit Water CPF	
	R	VSTR	R	VSTR	R	VSTR	R	VSTR	R	VSTR	R	VSTR
Rap E30/K31	0.581	0.112	0.578	0.107	0.581	0.112	-0.165	-0.773	-0.188	-0.639	-0.168	-0.751
Rap E30/K31E	-0.316	-0.613	-0.267	-0.737	-0.314	-0.619	-0.192	-4.670	-0.223	-3.793	-0.196	-4.573
Rap E30D/K31	-0.846	-0.306	-0.838	-0.303	-0.845	-0.305	-0.279	-3.764	-0.354	-2.742	-0.284	-3.680
Rap E30D/K31E	-0.819	-0.221	-0.812	-0.220	-0.819	-0.221	-0.558	-1.166	-0.568	-1.031	-0.560	-1.148
Ras D30E/E31K	0.072	3.230	-0.013	-16.500	0.069	3.367	0.313	0.938	0.278	1.031	0.311	0.941
Ras D30E/E31	-0.614	-0.513	-0.572	-0.483	-0.612	-0.511	0.162	1.408	0.126	1.682	0.160	1.419
Ras D30/E31K	-0.308	-0.286	-0.265	-0.339	-0.307	-0.286	-0.926	-0.107	-0.927	-0.100	-0.925	-0.107
Ras D30/E31	0.604	0.241	0.626	0.228	0.605	0.241	-0.302	-1.001	-0.304	-0.919	-0.305	-0.984
Ral	0.380	0.521	0.442	0.452	0.384	0.516	0.090	3.992	0.070	5.013	0.087	4.142
N27C <sub>SCN</sub>	0.016	2.817	0.001	73.333	0.015	2.857	0.141	0.456	0.133	0.452	0.140	0.454
G28C <sub>SCN</sub>	-0.422	-0.561	-0.403	-0.537	-0.421	-0.560	-0.260	-0.948	-0.294	-0.766	-0.263	-0.932
N29C <sub>SCN</sub>	-0.295	-0.191	-0.272	-0.194	-0.294	-0.191	0.091	0.759	0.085	0.750	0.090	0.760
Y31C <sub>SCN</sub>	-0.288	-0.541	-0.290	-0.524	-0.288	-0.540	-0.033	-13.750	-0.005	-76.154	-0.030	-15.231
K32C <sub>SCN</sub>	-0.165	-0.746	-0.163	-0.715	-0.164	-0.746	0.385	0.661	0.370	0.649	0.383	0.660
N54C <sub>SCN</sub>	-0.204	-0.481	-0.200	-0.470	-0.204	-0.481	0.085	2.857	0.096	2.391	0.086	2.793
All Points	-0.106	-1.380	-0.096	-1.463	-0.106	-1.383	-0.032	-7.279	-0.050	-4.361	-0.034	-6.851
	APBS		APBS 5 ÅSphere		GROMACS TIP3P Reaction Field		Hybrid TI3P Reaction Field					
	R	VSTR	R	VSTR	R	VSTR	R	VSTR	R	VSTR	R	VSTR
Rap E30/K31	-0.128	-1.252	-0.206	-0.970	-0.427	-0.416	-0.075	-1.800				
Rap E30/K31E	-0.659	-0.280	-0.646	-0.380	-0.293	-1.870	-0.739	-0.301				
Rap E30D/K31	-0.892	-0.301	-0.862	-0.453	-0.884	-0.813	-0.815	-0.503				
Rap E30D/K31E	-0.875	-0.205	-0.841	-0.280	-0.366	-1.338	-0.740	-0.326				
Ras D30E/E31K	0.203	2.340	0.112	3.542	0.174	1.840	0.494	0.602				
Ras D30E/E31	-0.599	-0.650	-0.527	-0.886	-0.073	-5.252	-0.718	-0.461				
Ras D30/E31K	-0.709	-0.135	-0.792	-0.119	-0.866	-0.129	-0.494	-0.157				
Ras D30/E31	0.620	0.332	0.471	0.544	-0.136	-4.000	0.781	0.207				
Ral	-0.191	-1.776	-0.031	-13.113	0.053	12.073	-0.345	-0.902				
N27C <sub>SCN</sub>	0.405	0.230	0.399	0.300	0.283	0.306	0.422	0.277				
G28C <sub>SCN</sub>	-0.816	-0.349	-0.694	-0.472	-0.376	-0.987	-0.758	-0.443				
N29C <sub>SCN</sub>	-0.359	-0.216	-0.220	-0.379	-0.059	-1.383	-0.328	-0.239				
Y31C <sub>SCN</sub>	-0.523	-0.438	-0.338	-0.805	-0.332	-1.085	-0.368	-0.666				
K32C <sub>SCN</sub>	-0.320	-1.694	0.052	8.800	0.339	0.776	0.053	6.286				
N54C <sub>SCN</sub>	-0.178	-0.668	-0.142	-1.010	-0.051	-6.326	0.270	0.739				
All Points	-0.384	-0.568	-0.312	-0.808	-0.159	-1.795	-0.211	-1.008				

<sup>a</sup> $VSTR$  has units of  $\text{cm}^{-1}/(\text{k}_\text{b}\text{T}/\text{e}\text{\AA})$ . The known  $VSTR$  is  $1.99 \text{ cm}^{-1}/(\text{k}_\text{b}\text{T}/\text{e}\text{\AA})$ .

Table 7-2: Correlation Coefficients (R) and Virtual Stark Tuning Rates ( $VSTR^a$ ) for Relative Field Calculations using Various Electrostatic Models

$\Delta F$ vs. $\Delta \tilde{\nu}$	AMOEBA		AMOEBA CP		AMOEBA CPf		AMOEBA Explicit Water		AMOEBA Explicit Water CP		AMOEBA Explicit Water CPf	
	R	VSTR	R	VSTR	R	VSTR	R	VSTR	R	VSTR	R	VSTR
Rap E30/K31	-0.469	-0.200	-0.467	-0.198	-0.469	-0.200	0.840	0.256	0.834	0.249	0.838	0.255
Rap E30/K31E	-0.019	-14.776	0.028	10.312	-0.016	-17.522	0.283	1.359	0.265	1.396	0.280	1.366
Rap E30D/K31	-0.052	-8.722	-0.016	-27.887	-0.050	-9.000	-0.046	-10.421	-0.100	-4.648	-0.051	-9.384
Rap E30D/K31E	0.218	1.002	0.225	0.965	0.220	0.993	0.897	0.456	0.876	0.449	0.896	0.456
Ras D30E/E31K	0.126	0.923	0.132	0.823	0.128	0.906	-0.036	-6.246	-0.062	-3.449	-0.037	-6.000
Ras D30E/E31	0.612	0.072	0.574	0.072	0.609	0.073	0.432	0.132	0.426	0.126	0.431	0.132
Ras D30/E31K	0.311	0.743	0.382	0.646	0.316	0.732	0.171	1.187	0.129	1.472	0.169	1.197
Ras D30/E31	0.651	0.477	0.721	0.410	0.655	0.473	0.021	12.000	0.028	8.426	0.020	12.375
All Points	0.016	0.204	0.034	0.454	0.018	0.223	0.202	1.777	0.182	1.710	0.200	1.769
	APBS		APBS 5 ASphere		GROMACS TIP3P Reaction Field		Hybrid TI3P Reaction Field					
	R	VSTR	R	VSTR	R	VSTR	R	VSTR	R	VSTR		
Rap E30/K31	-0.973	-0.196	-0.481	-0.680	0.299	1.171	-0.699	-0.284				
Rap E30/K31E	-0.688	-0.272	-0.529	-0.511	-0.010	-41.250	-0.663	-0.348				
Rap E30D/K31	-0.728	-0.351	-0.690	-0.634	-0.434	-1.464	-0.797	-0.543				
Rap E30D/K31E	-0.387	-0.513	-0.177	-1.930	0.789	0.621	-0.328	-1.427				
Ras D30E/E31K	-0.604	-0.589	-0.175	-1.909	0.092	3.183	-0.134	-2.000				
Ras D30E/E31	0.264	0.366	0.445	0.445	0.210	0.556	-0.093	-2.134				
Ras D30/E31K	-0.393	-0.515	0.015	14.887	0.064	4.231	-0.359	-0.546				
Ras D30/E31	-0.265	-1.128	0.136	3.328	-0.288	-1.674	0.098	2.773				
All Points	-0.549	-5.224	-0.278	-1.945	0.018	0.115	-0.403	-3.204				

<sup>a</sup> $VSTR$  has units of  $\text{cm}^{-1}/(k_b T/e\text{\AA})$ . The known  $VSTR$  is  $1.99 \text{ cm}^{-1}/(k_b T/e\text{\AA})$ .

Table 7-3: Correlation Coefficients for Field Standard Deviations Compared to Experimental Full Width at Half Peak Maximum (FWHM) using Various Electrostatic Methods

Fits/Mutant	AMOEBA	AMOEBA CP	AMOEBA CPf	AMOEBA Explicit Water	AMOEBA Explicit Water CP	AMOEBA Explicit Water CPf
Rap E30/K31	-0.571	-0.571	-0.571	-0.754	-0.753	-0.754
Rap E30/K31E	0.159	0.158	0.159	-0.456	-0.446	-0.456
Rap E30D/K31	0.073	0.073	0.073	-0.796	-0.804	-0.796
Rap E30D/K31E	-0.294	-0.297	-0.294	-0.328	-0.330	-0.328
Ras D30E/E31K	-0.058	-0.058	-0.058	0.400	0.402	0.400
Ras D30E/E31	-0.324	-0.322	-0.324	-0.075	-0.057	-0.075
Ras D30/E31K	0.634	0.633	0.634	-0.062	-0.062	-0.062
Ras D30/E31	-0.348	-0.348	-0.348	0.006	0.002	0.006
N27C <sub>SCN</sub>	0.110	0.109	0.110	0.251	0.253	0.251
G28C <sub>SCN</sub>	0.348	0.348	0.348	0.573	0.574	0.573
N29C <sub>SCN</sub>	-0.605	-0.605	-0.605	-0.815	-0.816	-0.815
Y31C <sub>SCN</sub>	0.111	0.111	0.111	0.025	0.030	0.025
K32C <sub>SCN</sub>	-0.130	-0.129	-0.130	-0.277	-0.276	-0.277
N54C <sub>SCN</sub>	0.249	0.247	0.249	0.064	0.062	0.064
All Points	-0.125	-0.125	-0.125	-0.139	-0.137	-0.139
				GROMACS		
	APBS	APBS 5 ÅSphere	TIP3P Reaction Field		Hybrid TIP3P Reaction Field	
Rap E30/K31	-0.666	-0.675	-0.641	-0.741		
Rap E30/K31E	-0.077	0.371	0.167	-0.296		
Rap E30D/K31	-0.103	-0.352	-0.537	-0.209		
Rap E30D/K31E	-0.485	-0.586	-0.274	-0.525		
Ras D30E/E31K	0.057	0.281	0.297	0.157		
Ras D30E/E31	-0.414	-0.408	-0.047	-0.413		
Ras D30/E31K	0.550	0.718	0.164	0.441		
Ras D30/E31	-0.594	-0.384	-0.127	-0.746		
N27C <sub>SCN</sub>	0.107	0.221	0.191	0.173		
G28C <sub>SCN</sub>	0.383	0.298	0.262	0.434		
N29C <sub>SCN</sub>	-0.536	-0.748	-0.880	-0.547		
Y31C <sub>SCN</sub>	0.024	0.296	0.527	0.031		
K32C <sub>SCN</sub>	-0.287	-0.264	-0.453	-0.199		
N54C <sub>SCN</sub>	0.261	0.321	0.049	0.304		
All Points	-0.189	-0.136	-0.118	-0.235		

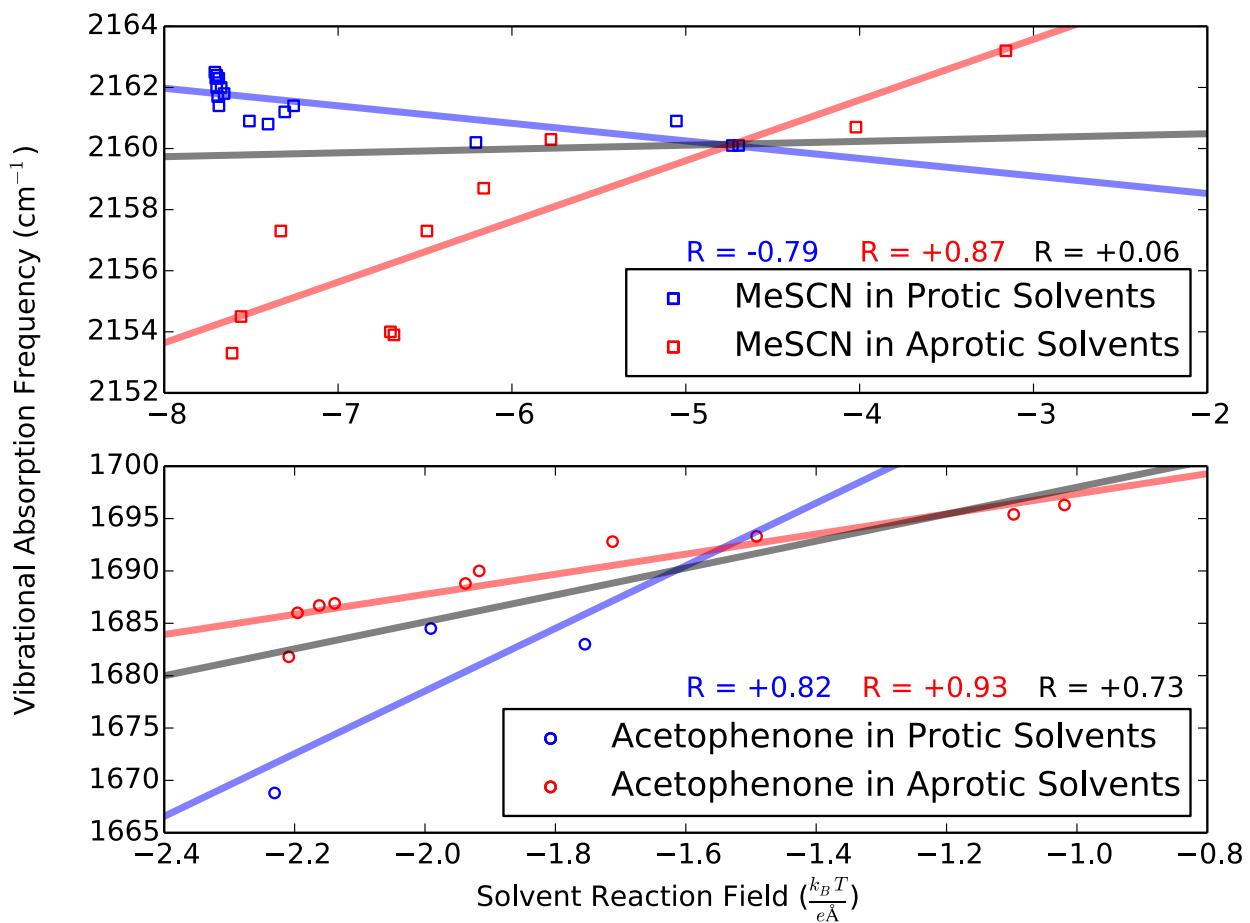


Figure 7-1: Poisson-Boltzmann Solvent Reaction Fields for Methylthiocyanate and Acetophenone in Various Solvents

Solvent Reaction Fields on (top) methylthiocyanate and (bottom) acetophenone calculated using APBS where each solvent is described as a dielectric continuum. Blue: solvents that can donate a hydrogen bond to the vibrational chromophore; red: solvents that cannot hydrogen bond to the vibrational chromophore. Best-fit lines and correlations coefficients are included, with black being the aggregate of all data points. Experimental measurements for methylthiocyanate are unpublished and taken by Christina Ragain, Ph.D., Josh Slocum, and Kelsey Eklund. Experimental measurements for acetophenone were previously reported by Fried *et al.*<sup>170</sup>

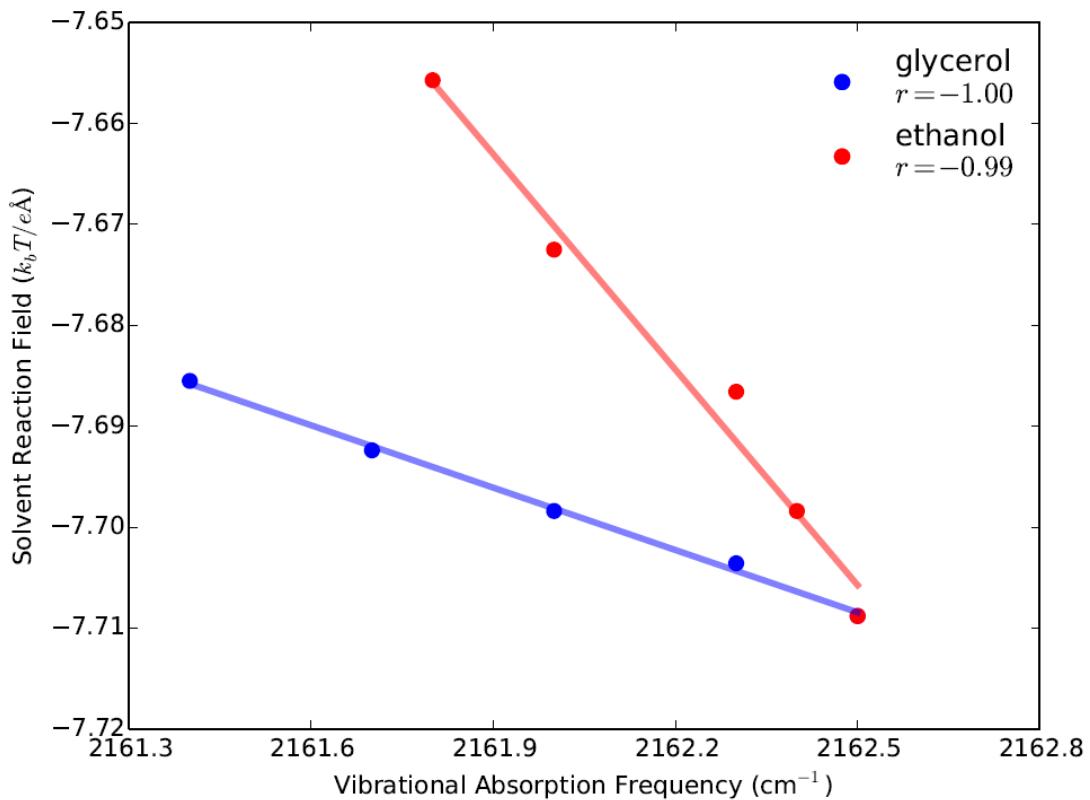


Figure 7-2: Methylthiocyanate Solvent Reaction Fields at Various Dielectrics, Modulated by Glycerol and Ethanol

The solvent reaction field calculated in APBS for methylthiocyanate at various dielectrics. (Red) Dielectric constant was modulated by addition of glycerol. (Blue) Dielectric constant was modulated by addition of ethanol. The right-most data point is 100% water and shared by both fits.

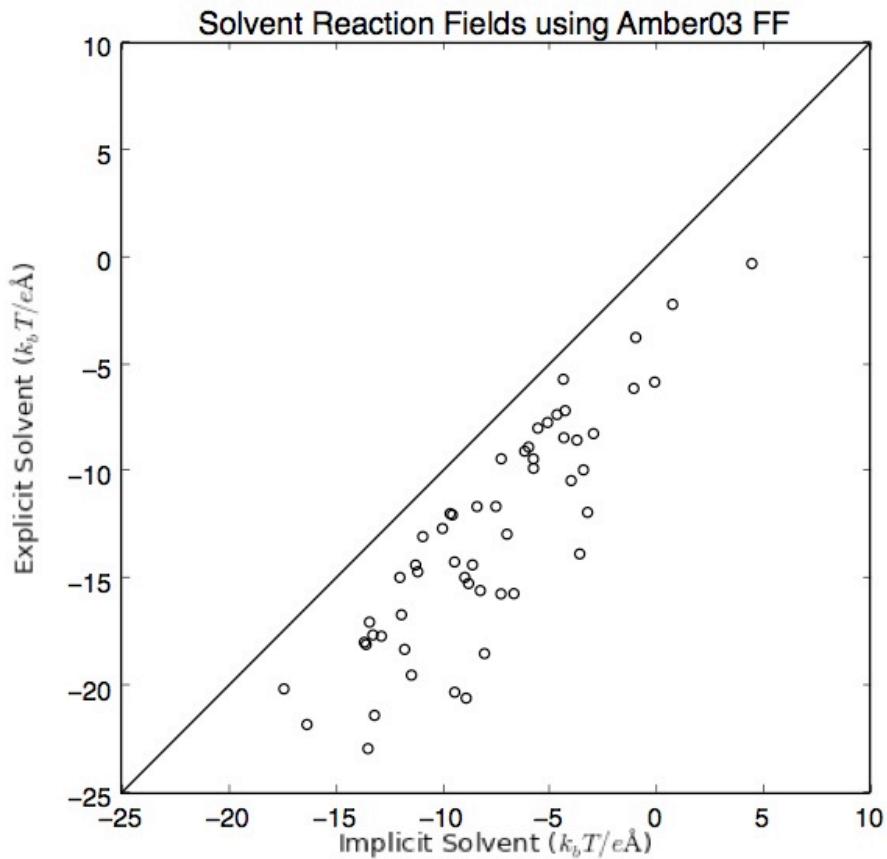


Figure 7-3: Comparison Between Solvent Reaction Fields Calculated using Explicit TIP3P Water and Implicit PB Water for All 54 GTPase/Ral Probe Combinations

The solvent reaction fields using explicit TIP3P water plotted against the solvent reaction fields using implicit PB water with Amber03 point charges for all explicitly defined atoms. The line along  $y=x$  is not a best fit line and is meant to show that the two models are linearly correlated with a slope of approximately 1 and the implicit solvent being consistently less negative

## **Chapter 8 Comparing Experimental Absorption Energies to Electrostatic Fields Sampled in AMOEBA for Small Thiocyanate-labeled Molecules**

### **8.1 INTRODUCTION**

One of the chief difficulties in protein electrostatics is obtaining adequate sampling. Throughout my research, I have operated under the assumption that the most important degree of freedom is the nitrile dipole orientation. Expanding from one-dimensional umbrella sampling to two-dimensional umbrella sampling yielded improved correlation between calculated PB fields and experimentally measured vibrational absorption energies. However, increasing the umbrella window time from 400 ps to 2000 ps did not yield the same degree of improvement. Looking strictly at the RFM, instances of improved (Rap E30D/K31E), relatively unchanged (Rap E30D/K31, Rap E30/K31E), and decreased (Ral monomer, Rap E30/K31) correlation were all observed (comparing Table 6-2 to Table 7-2). The first two cases are not unexpected; either sampling was inadequate and improved sampling improved the simulation's physical veracity or the sampling was adequate enough and improved sampling does not have a significant effect.

The case where increasing the sampling actually decreases the correlation to experiment is perplexing; what about increasing the simulation duration would actually decrease how well the *in vitro* system is represented? The answer is proteins are very complex. One possibility is increasing simulation time could allow a rare local minimum to be visited but not escaped, resulting in an over-representation of that structure. Likewise, it is also possible that the local structure sampled coincidentally happens to also be well correlated to the experimental measurement. The Ral monomer has approximately 170 alkane-like side chain dihedrals and Ral docked to Rap E30/K31 has

approximately 460 side chain dihedrals. Assuming all of them have a non-zero probability of being at approximately  $60^\circ$ ,  $180^\circ$ , and  $-60^\circ$  and counting each of those states as a single state (for three total probable states), then there are about  $10^{80}$  possible combinations of side chain dihedrals for Ral and  $10^{220}$  combinations for Ral docked to Rap E30/K31. Obviously, a large number of these are so energetically unfavored that they can be neglected; there are certain combinations that are physically impossible due to steric overlap (which is one of the reasons we previously used 2D umbrella sampling about our cyanocysteine probe side chain dihedrals: to ensure that all probably states were visited).

The relevant number of states can be further reduced by remembering that we are looking at protein electrostatic fields at specific points in space, which is determined using equation (3-2). Assuming that only residues with side chains within 10 Å of the probe are relevant for electrostatic field calculations (which scale as  $1/r^2$ , where  $r$  is distance), which for Y31C<sub>SCN</sub> docked to Rap E30/K31 is 13, and using an average number of dihedrals per residue as 1.7 (460 side chain dihedrals per 270 residues), there are still  $10^3$  possible dihedral combinations—significantly more tractable. Yet because of energy barriers, significantly more than  $10^3$  frames are needed to see all possible states; the scope of how many more frames is dependent on the size of the barriers. But again, there are orientations of these 13 residues that may only be energetically favorable given specific orientation(s) of residues further than 10 Å from the probe. To summarize, sampling all possible, or even reasonable, permutations of even a small protein is never going to be viable.

For these reasons, we have moved away from small-moderately sized proteins and will be examining a set of smaller systems: methylthiocyanate (mescn/MeSCN), ethylthiocyanate (etscn/EtSCN), hexylthiocyanate (hxscn/HxSCN), and acetyl-

cyanocysteine-N-methylamine (ace-cnc-nme/ACE-CNC-NME) peptide-like small molecule, shown in Figure 8-1. Furthermore, since the principles behind using the AMOEBA force field are still a concern (solute dielectric versus electronic polarizability) and it was clear that sampling in Amber03 and performing field calculations in AMOEBA was unsuccessful, these molecules have also been simulated in the AMOEBA force field. The solvated protein system is too large for the current implementation of AMOEBA in Tinker 7, but these smaller systems can easily be simulated in a reasonable amount of time (approximately 0.8-1.0 ns/day).

It has been shown that both AMOEBA as well as GAFF<sup>36</sup> can reproduce experimental Stark shifts for a given probe in a variety of solvents.<sup>172</sup> However, interesting biology involves water interacting with many different solutes. The central focus of my work has been to understand how to quantify electrostatic fields in biologically-relevant systems. In that regard, a good reproduction of Stark shifts of different protein systems in water strictly using MD has not been reported. Here we investigate how well Stark shifts can be reproduced for various thiocyanate-containing solutes that have fewer relevant degrees of freedom and therefore will be easier to obtain a complete ensemble. The goal here is a proof-of-concept and may be used as a springboard for future works on increasingly larger probe-containing systems.

Simulations on these same small molecules in Amber03 have also been started by an undergraduate in our lab for future comparison.

## 8.2 RESULTS AND DISCUSSION

Parameterization (page 33) and simulation protocols (page 59) have previously been described. To recap, snapshots were collected every 0.1 ps. MeSCN and EtSCN were each sampled for 4.0 ns. HxSCN was sampled for 5.0 ns. ACE-CNC-NME was

sampled for 7.0 ns. All simulations used the RESPA integrator<sup>61</sup> and Bussi thermostat<sup>62</sup> with a 2.0 fs time step at 298 K at NVT.

### 8.2.1 Sampling CN Orientations

Figure 8-2 shows the one-dimensional dihedral distributions for all non-hydrogen dihedral angles. It is clear that ethylthiocyanate is able to sample all of the expected alkane-like dihedral space in 4 ns. Hexylthiocyanate and the ACE-CNC-NME are more difficult to assess from Figure 8-2 due to being unable to distinguish between dihedral permutations. Figure 8-3 therefore shows the two-dimensional dihedral distribution for these two molecules for the dihedrals involving SCN. Looking at Figure 8-3, hexylthiocyanate appears to behave alkane-like, with all alkane-like windows visited at least briefly. However, either the  $(\chi_1, \chi_2)$  analogous-pairs  $(180^\circ, 60^\circ), (180^\circ, -60^\circ), (-60^\circ, -60^\circ)$ , and  $(60^\circ, 60^\circ)$  are particularly favorable or the system has not yet had enough time to fully sample the other alkane-like regions. Likewise, ACE-CNC-NME appears to very much favor the  $(180^\circ, -60^\circ)$  alkane region, with a small-but-significant probability at  $(60^\circ, 60^\circ)$ . ACE-CNC-NME may actually favor those regions or, more likely, it has not yet had adequate time to sample other dihedral pairs.

Because of it is capped and therefore few possible solute-thiocyanate interactions should be present, it is hypothesized that the ACE-CNC-NME simply has not yet had adequate time to sample all states and an enhanced method, such as umbrella sampling, should be used to generate a complete ensemble. The same may be said for hexylthiocyanate also, although to a lesser degree. Looking at Figure 8-4, it can clearly be seen that ethylthiocyanate and hexylthiocyanate can rotate dihedrals much faster than the capped cyanocysteine, which spends nearly 6 ns in a single  $(\chi_1, \chi_2)$  conformation, only to escape it near the end. It is possible that more sampling could significantly alter

the dihedral probability distribution as well as see the probe revert back to the prior state and remain there; more sampling or an enhanced MD technique is needed in either case.

### 8.2.2 Electrostatic Fields

Figure 8-5 shows the average calculated field as a function of time for fields calculated using Figure 8-5A) the induced method (IM, equation (3-9)), Figure 8-5B) the midpoint method (MPM, equation (3-10)), Figure 8-5C) MPM where the monopole, permanent dipole, induced dipole, and quadrupole contributions of the SCN atoms have been removed, and Figure 8-5D) MPM where the monopole, permanent dipole, and quadrupole contributions of the SCN atoms have been removed. The IM is trivial to calculate using built-in output from Tinker during the simulation run. The MPM method required additional work-up after the simulation was completed; both methods have previously been described.

For all four methods, the average field value stops significantly changing within the first 2 ns all small molecules, indicating either convergence with respect to electrostatic field or oversampling within some number of local minima and inadequate sampling within other structures. It's likely the former for MeSCN and EtSCN and the latter for hexylthiocyanate and ACE-CNC-NME, based on MeSCN having very few structural degrees of freedom, Figure 8-2, and Figure 8-3. Once again, it is likely that enhanced MD (or simply more simulation time) are necessary for the larger two molecules, HxSCN and ACE-CNC-NME.

Due to the close proximity between the SCN atoms and the nitrile midpoint ( $\sim 0.57 \text{ \AA}$ ), SCN will be the dominating contributor to the electrostatic field. In the IM, this is significantly reduced by the damping factors in the SCF subroutines of Tinker. The MPM does not use such damping terms, however, which is why the magnitudes in

Figure 8-5B, Figure 8-5C, Figure 8-6, Figure 8-7B, and Figure 8-7D are so large. Since we are not interested in the self-field of the probe, but rather the external field felt by the probe, we have looked at removing the field contributions of the SCN atoms, which is identical to what was done in the previous chapter. The SCN field is due to the charges and orientations of SCN, which should, on average, be approximately constant for any given system; the force field parameters defining bond lengths and angles should define the average local SCN structure. Because AMOEBA also has an induced dipole term, however, the field due to the induced dipoles on the S, C, and N atoms is *not* constant. Therefore, we have looked at both removing and keeping the SCN induced dipole field contribution. Figure 8-6 shows the contribution of each multipole part on SCN to the electrostatic field at the bond midpoint. As expected, all of the permanent terms (monopole, permanent dipole, quadrupole, permanent total) are approximately constant for all probes. Again, this is unsurprising given all the SCN atoms share multipole parameters and have very similar bonded and angle parameters (Code and Parameters 2-1, Code and Parameters 2-2, Code and Parameters 2-3, Code and Parameters 2-4). The induce dipole field contributions (and total including induced dipole contributions) vary because it is a response to the local field experienced by the probe—the exact value we are trying to vary by modulating the atoms SCN is attached to.

### 8.2.3 Correlating Small Molecule Fields to Experiment

Figure 8-7 shows the calculated fields plotted against the experimental absorption energies. It is important to keep in mind that, no matter how well correlated any of the data may be, there are only three data points in each fit (there are no experimental absorption frequency measurements for ACE-CNC-NME at this time and it is simply placed along the best-fit line based on its calculated field) and all interpretations of the

data need to weary of this. In fact, a squared correlation coefficient of 0.97 (Figure 8-7D), still only has a p-value of 0.11 for 3 data points—there is an 11% likelihood that the most correlated data are correlated by chance.

In the induced method (Figure 8-7A), MPM total field (Figure 8-7B), and MPM field less permanent SCN field (Figure 8-7D) we see a positive correlation between calculated fields and experimental absorption energies. However, in the MPM field less the total SCN field (Figure 8-7C), we see a strong negative correlation. The only difference between Figure 8-7(A, B, D) and Figure 8-7C is that Figure 8-7C does *not* include the probe's response to its external field environment via the SCN induced dipole field. As previously described, when the positive vector direction is defined as pointing from the nitrile carbon to the nitrile nitrogen ( $\overrightarrow{\text{CN}}$ ), then the *VSTR* is the inverse of the slope, equation (3-15), and the sign of the slope is the sign of the *VSTR*. In Figure 8-7(A, B, D) the *VSTR* is positive, as expected, while in Figure 8-7C it is negative. This is the first time we have isolated a specific interaction that influences the sign of the *VSTR*. It appears as though including the part of field due to the vibrational chromophore's response to an external electrostatic field (via induced dipole moments on the SCN atoms) is important for accurately predicting qualitative vibrational Stark shifts.

The observation that removing all of the SCN contribution to the field, including the induced dipole contribution, results in negative correlations merited re-visiting the previously reported results of Ral docked to the GTPase constructs. Due to the way fields were calculated and stored, I could not examine adding back in only the induced dipole contribution without months of repeated calculations. Furthermore, we looked at removing the entire side chain,  $\text{CH}_2\text{SCN}$ , for the same reason that it should be approximately constant. We can, however, look at total field without any contributions removed, which would reintroduce the seemingly vital probe induced dipole field. The

correlations with and without removing any atom contributions are plotted in Figure 8-8. For all the monopole methods, there is essentially zero change—the CH<sub>2</sub>SCN is essentially constant among all simulations. For the explicit solvent AMOEBA, there is also essentially no change. For AMOEBA with implicit solvent, the magnitude of the correlation increases, although it becomes more negative rather than changing signs. It is likely that allowing for only electronic polarization and not orientation rotation in AMOEBA is insufficient and contributes to the lack of transferability among ensembles. For future calculations, keeping the field due to the induced dipole on probe may be important, but for our previous results it's likely that the non-transferability of ensembles from Amber03 to AMOEBA is more significant and convoluting.

We also looked at field standard deviations compared to experimental FWHM, shown in Figure 8-9. Aside from the MPM total field (Figure 8-9B), the correlations are in the positive direction, although, again, with only three data points, their significance is questionable. While removing the permanent SCN field contributions (Figure 8-9D) appears to increase the correlation between field standard deviations and experimental FWHM, it is again necessary to remember that only three data points exist, with the ACE-CNC-NME data point plotted along the best-fit equation, and the statistical significance is too low to make a definitive statement.

### 8.3 CONCLUSION

Sampling SCN-labeled small molecules in AMOEBA appears to be promising and merits further investigation. For the first time we have been able to identify the term responsible for the sign of the *VSTR*. It appears that the electronic polarization of the probe is important for correctly quantifying the direction of vibrational shifts due to the

VSE. It's also likely necessary to perform some sort of enhanced MD on the cyanocysteine and hexylthiocyanate to ensure correct ensembles.

AMOEBA has already been shown to reproduce qualitative shifts due to the VSE at a single probe in different solvents.<sup>172</sup> Here we have shown that AMOEBA can reproduce qualitative shifts due to the VSE for multiple solutes in water. Combining the results of the previous report with this preliminary study indicates that AMOEBA may be able to adequately reproduce VSE shifts in more complicated solute/solvent environments, such as proteins. Addition experiments and simulations should be examined to further investigate.

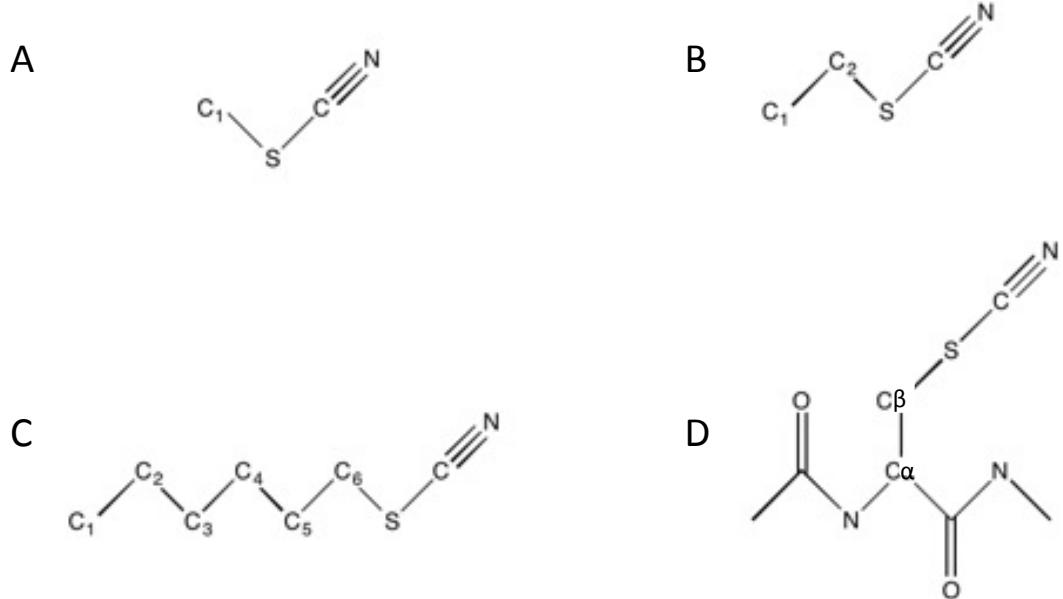


Figure 8-1: Thiocyanante-Labeled Small Molecules

Thiocyanate-labeled small molecules (A) methylthiocyanate, (B) ethylthiocyanate, (C) hexylthiocyanate, and (D) acetyl-cyanocysteine-N-methylamine. Hydrogen atoms are unlabeled.

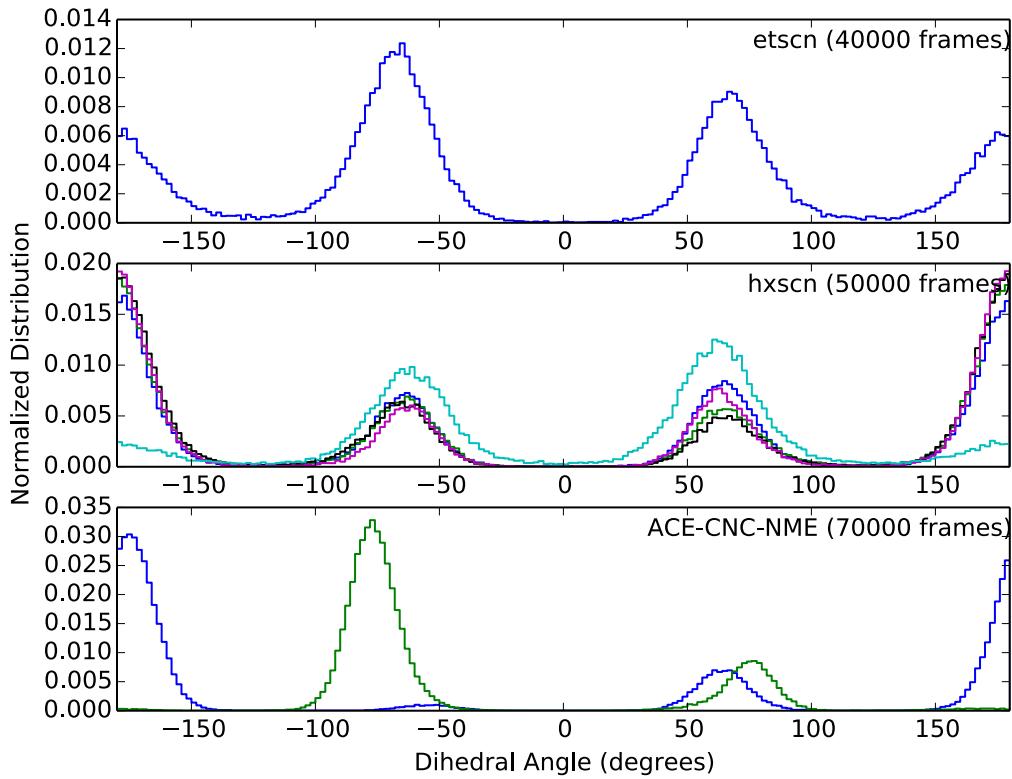


Figure 8-2: One-Dimensional Dihedral Probability Distributions

Non-Hydrogen dihedral probability distributions for ethylthiocyanate (top), hexylthiocyanate (middle), and capped cyanocysteine (bottom). Ethylthiocyanate (top): (blue) C1-C2-S-C dihedral. Hexylthiocyanate (middle): (blue) C1-C2-C3-C4 dihedral; (green) C2-C3-C4-C5 dihedral; (black) C3-C4-C5-C6 dihedral; (magenta)  $\chi_1$  analogue C4-C5-C6-S dihedral; (cyan)  $\chi_2$  analogue C5-C6-S-C dihedral. Capped cyanocysteine (bottom): (blue)  $\chi_1$  N-C $\alpha$ -C $\beta$ -S dihedral; (green)  $\chi_2$  C $\alpha$ -C $\beta$ -S-C dihedral.

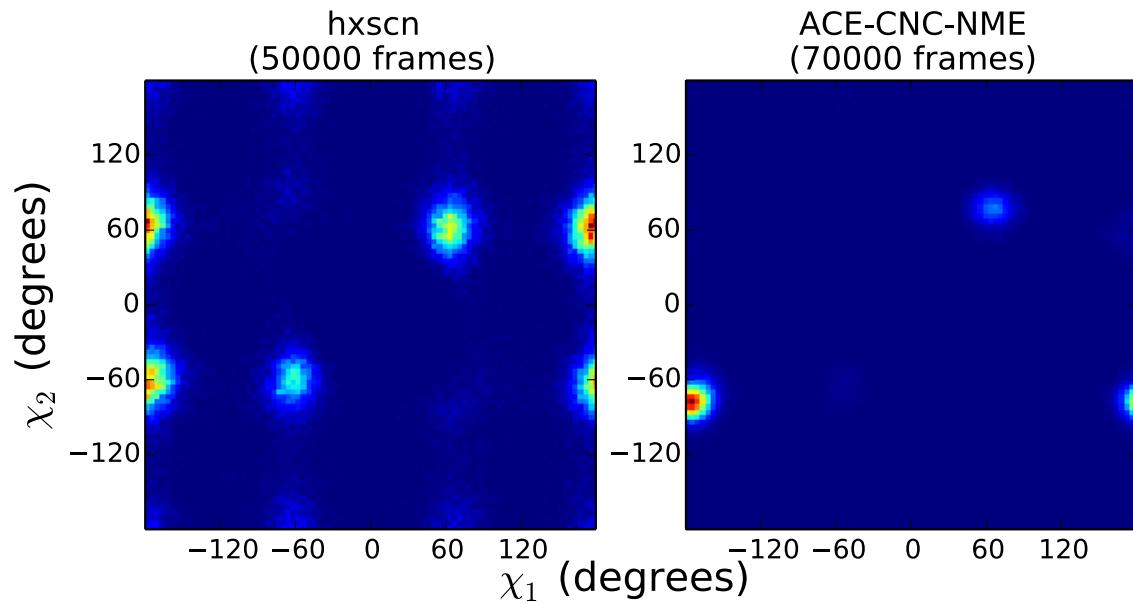


Figure 8-3: Two-Dimensional Dihedral Probability Distributions for Hexylthiocyanate and Capped Cyanocysteine

Hexylthiocyanate (left)  $\chi_1$  (C4-C5-C6-S) and  $\chi_2$  (C5-C6-S-C) analogous two-dimensional dihedral distribution after 5 ns of simulation. Capped cyanocysteine (right)  $\chi_1$  (N-C $\alpha$ -C $\beta$ -S) and  $\chi_2$  (C $\alpha$ -C $\beta$ -S-C) two-dimensional dihedral distribution after 7 ns of simulation.

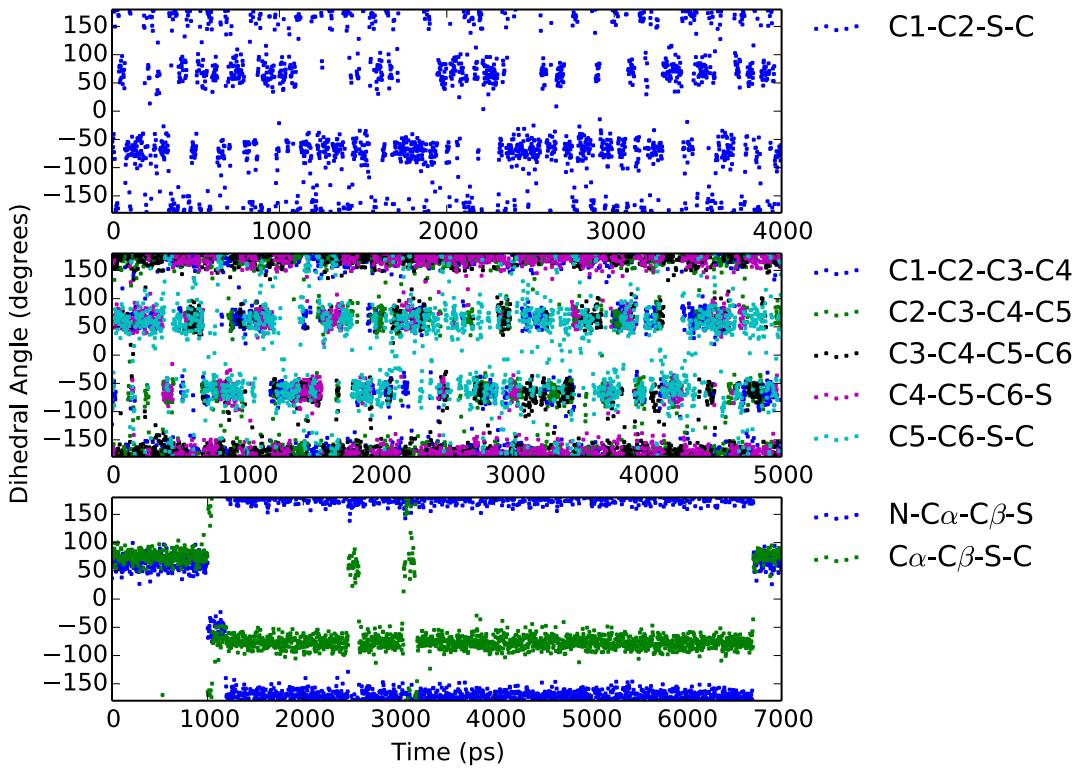


Figure 8-4: Dihedral Angles at Each Time Step

Non-Hydrogen dihedral angles as a function of time for ethylthiocyanate (top), hexylthiocyanate (middle), and capped cyanocysteine (bottom). Ethylthiocyanate (top): (blue) C1-C2-S-C dihedral. Hexylthiocyanate (middle): (blue) C1-C2-C3-C4 dihedral; (green) C2-C3-C4-C5 dihedral; (black) C3-C4-C5-C6 dihedral; (magenta)  $\chi_1$  analogue C4-C5-C6-S dihedral; (cyan)  $\chi_2$  analogue C5-C6-S-C dihedral. Capped cyanocysteine (bottom): (blue)  $\chi_1$  N-C $\alpha$ -C $\beta$ -S dihedral; (green)  $\chi_2$  C $\alpha$ -C $\beta$ -S-C dihedral.

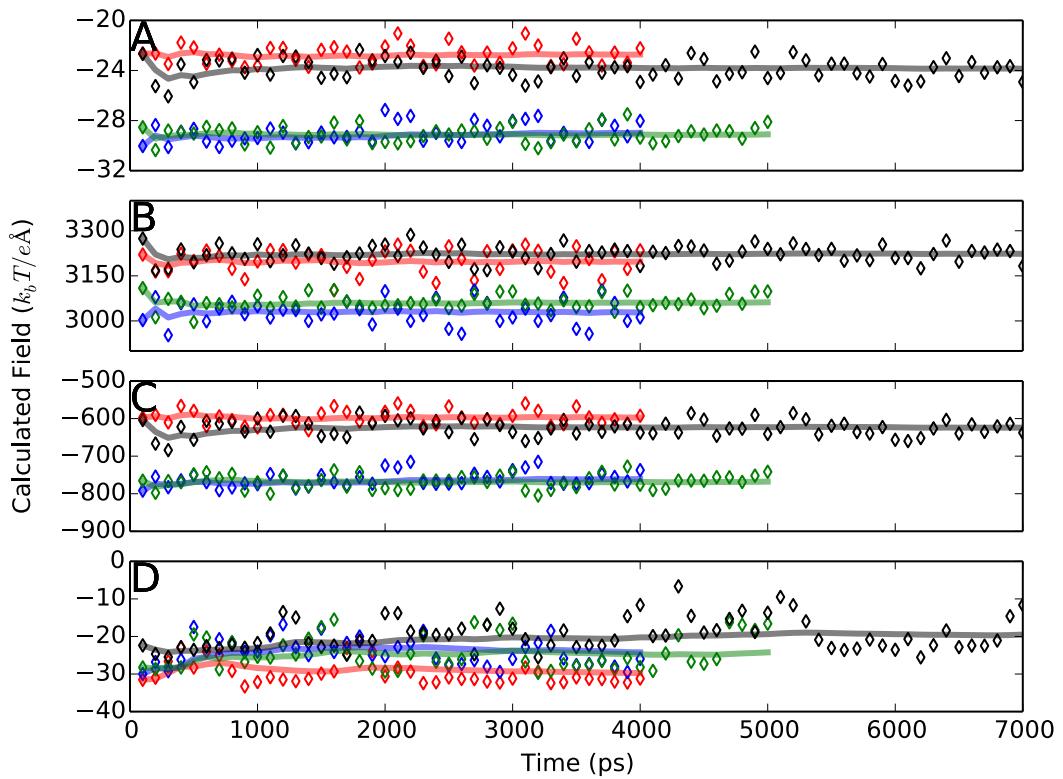


Figure 8-5: Average Electrostatic Field as a Function of Simulation Time

Average electrostatic field for methylthiocyanate (red), ethylthiocyanate (blue), hexylthiocyanate (green), and capped cyanocysteine (black) after some amount of frames, indicated on the x-axis. Diamonds are the average for the previous 1000 frames (100 ps).

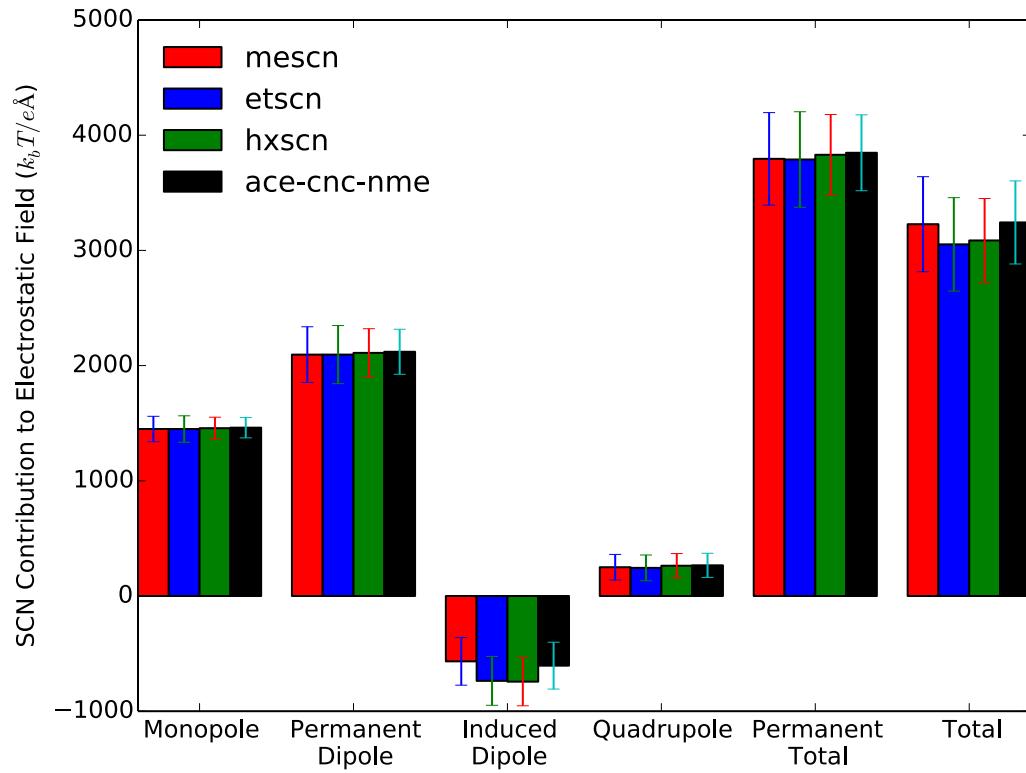


Figure 8-6: SCN Field Contributions are Constant

Contributions to the electrostatic field at the nitrile bond midpoint due to different multipole contributions from the SCN atoms. SCN is the dominating contributor to the electrostatic field due to the close proximity between the location of interest (nitrile bond midpoint) and the SCN atoms.

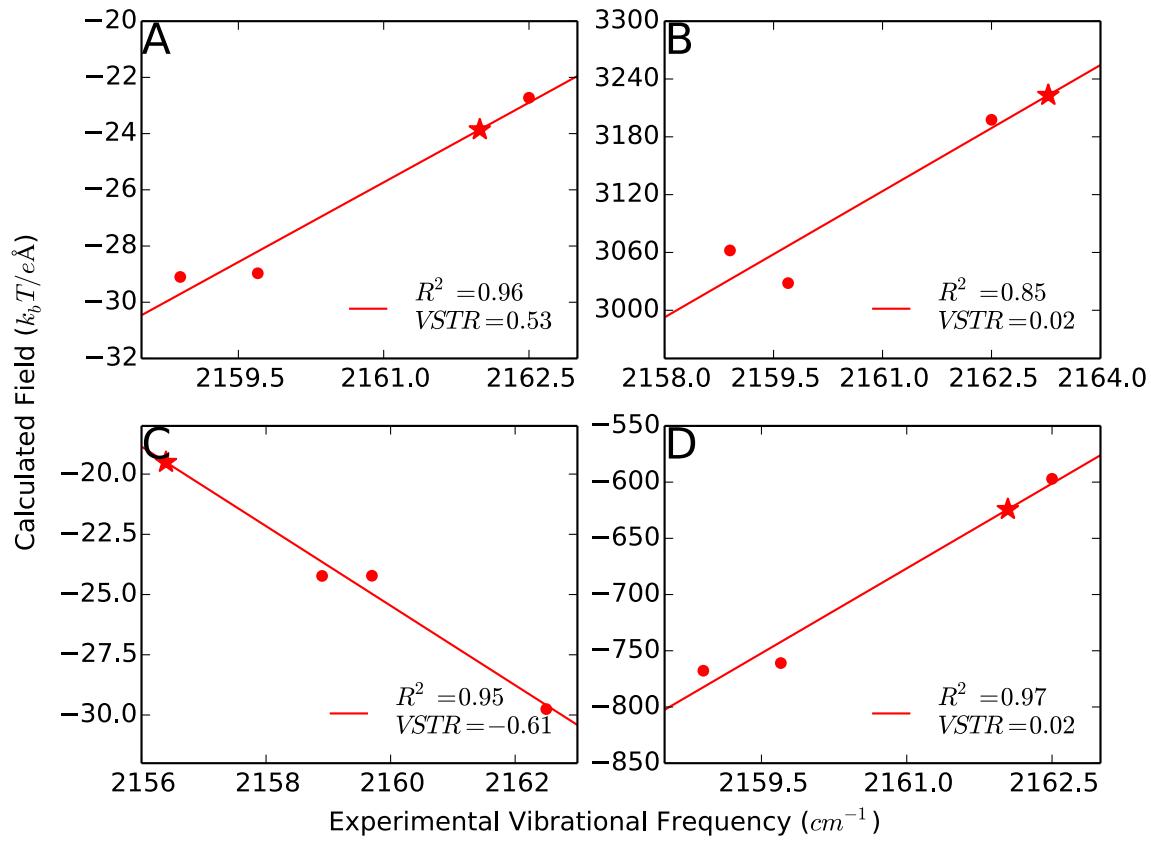


Figure 8-7: Small Molecule AMOEBA Fields Plotted Against Experimental Vibrational Absorption Energies

Calculated fields plotted against experimentally measured vibrational absorption energies. Squared correlation coefficients and  $VSTR$  are indicated in the bottom right corner of each plot. The capped cyanocysteine does not have an experimental absorption energy at this time and has therefore been speculated based on the best-fit equation obtain from the other three molecules and indicated with a star data point. A) Fields calculated with the IM. B) Total fields calculated using MPM. C) Fields calculated using MPM where the monopole, permanent dipole, *induced* dipole, and quadrupole fields due to SCN atoms have been removed. D) Fields calculated using MPM where the monopole, *permanent* dipole, and quadrupole fields due to SCN atoms have been removed.

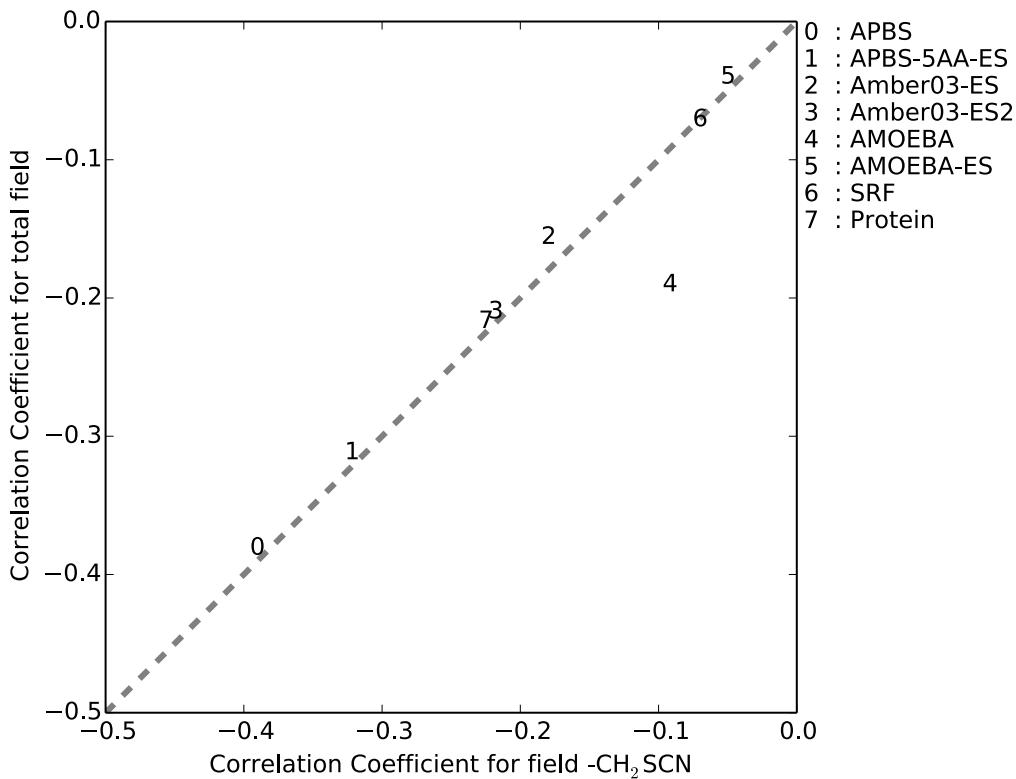


Figure 8-8: Comparing Total Field Correlations to Field Less CH<sub>2</sub>SCN Correlations

The correlation between the total field and experimental vibrational absorption energies was compared to the field less CH<sub>2</sub>SCN and experimental vibrational absorption energies. The dashed line is the line y=x, not a best-fit line. 0) APBS; 1) APBS 5 Å water sphere; 2) GROMACS explicit TIP3P reaction field electrostatics; 3) hybrid TIP3P reaction field electrostatics; 4) AMOEBA (CP and CPf have been excluded due to being nearly identical to without); 5) AMOEBA with explicit solvent (CP and CPf have been excluded due to being nearly identical to without); 6) the PB solvent reaction field; 5) the analytic Coulomb solute field.

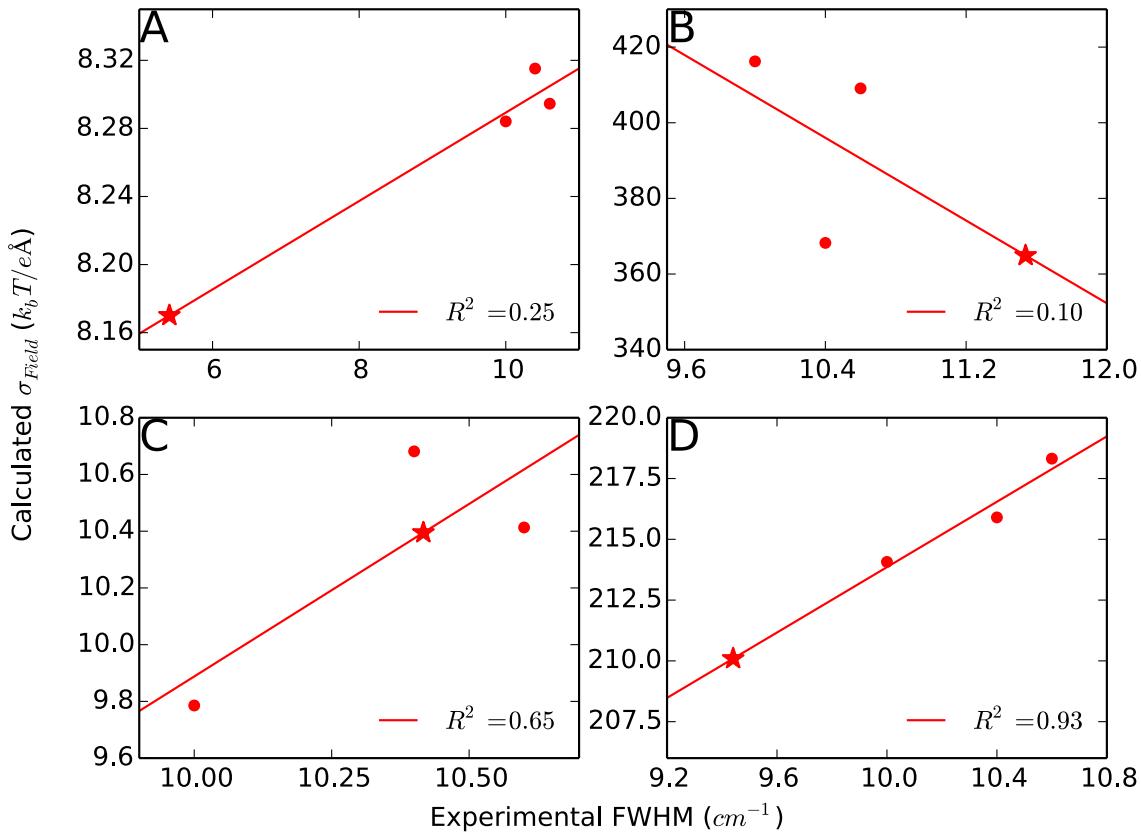


Figure 8-9: Small Molecule AMOEBA Field Standard Deviations Plotted Against Experimental FWHM

Calculated field standard deviations plotted against experimentally measured full width at half peak maximum. Squared correlation coefficients are indicated in the bottom right corner of each plot. The capped cyanocysteine does not have an experimental absorption energy at this time and has therefore been speculated based on the best-fit equation obtained from the other three molecules and indicated with a star data point. A) Fields calculated with the IM. B) Total fields calculated using MPM. C) Fields calculated using MPM where the monopole, permanent dipole, *induced* dipole, and quadrupole fields due to SCN atoms have been removed. D) Fields calculated using MPM where the monopole, *permanent* dipole, and quadrupole fields due to SCN atoms have been removed.

## References

1. Gilson, M. K.; Honig, B. H., Calculation of Electrostatic Potentials in an Enzyme Active-Site. *Nature* **1987**, *330* (6143), 84-86.
2. 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.
3. Honig, B.; Nicholls, A., Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268*, 1144-1149.
4. Nielsen, J. E.; Andersen, K. V.; Honig, B.; Hooft, R. W. W.; Klebe, G.; Vriend, G.; Wade, R. C., Improving Macromolecular Electrostatics Calculations. *Protein Eng.* **1999**, *12*, 657-662.
5. Warshel, A., What about protein polarity? *Nature* **1987**, *330*, 15-16.
6. Warshel, A.; Levitt, M., Theoretcial Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227-249.
7. Warshel, A.; Papazyan, A., Electrostatic effects in macromolecules: fundamental concepts and practical modeling. *Curr. Opin. Struct. Biol.* **1998**, *8*, 211-217.
8. Ensign, D. L.; Webb, L. J., Statistical, Geometric, and Physical Factors Determining Electrostatic Fields at the Ras/Effecter Interface. **2010**, *in preparation*.
9. Jiao, D.; King, C.; Grossfield, A.; Darden, T. A.; Ren, P. Y., Simulation of Ca<sup>2+</sup> and Mg<sup>2+</sup> solvation using polarizable atomic multipole potential. *J Phys Chem B* **2006**, *110* (37), 18553-18559.
10. Petersen, A.; Aarestrup, F. M.; Hofshagen, M.; Sipila, H.; Franklin, A.; Gunnarsson, E., Harmonization of antimicrobial susceptibility testing among veterinary diagnostic laboratories in the five Nordic countries. *Microb Drug Resist* **2003**, *9* (4), 381-388.
11. Jiao, D.; Golubkov, P. A.; Darden, T. A.; Ren, P., Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc. Natl. Acad. Sci.* **2008**, *105*, 6290-6295.
12. Burrell, K. H.; Allen, S. L.; Anderson, P. M.; Austin, M. E.; Baity, W. F.; Bakalarski, J. P.; Baker, D. R.; Baldwin, D. E.; Barber, D. E. G.; Bastasz, R., et al., Overview of recent experimental results from the DIII-D advanced tokamak program. *Nucl Fusion* **2003**, *43* (12), 1555-1569.
13. Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A., et al., Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549-2564.
14. Petersen, R. K.; Jorgensen, C.; Rustan, A. C.; Froyland, L.; Muller-Decker, K.; Furstenberger, G.; Berge, R. K.; Kristiansen, K.; Madsen, L., Arachidonic acid-dependent inhibition of adipocyte differentiation requires PKA activity and is associated with sustained expression of cyclooxygenases. *J Lipid Res* **2003**, *44* (12), 2320-2330.

15. Schnieders, M. J.; Baker, N. A.; Ren, P.; Ponder, J. W., Polarizable Atomic Multipole Solutes in a Poisson-Boltzmann Continuum. *J. Chem. Phys.* **2007**, *126*, 124114-1-47.
16. Petersen, T. W.; van den Engh, G., Stability of the breakoff point in a high-speed cell sorter. *Cytom Part A* **2003**, *56A* (2), 63-70.
17. Lindquist, B. A.; Corcelli, S. A., Nitrile Groups as Vibrational Probes: Calculations of the CN Infrared Absorption Line Shape of Acetonitrile in Water and Tetrahydrofuran. *J. Phys. Chem. B* **2008**, *112*, 6301-6303.
18. Lindquist, B. A.; Haws, R. T.; Corcelli, S. A., Optimized Quantum Mechanics/Molecular Mechanics Strategies for Nitrile Vibrational Probes: Acetonitrile and para-Tolunitrile in Water and Tetrahydrofuran. *J. Phys. Chem. B* **2008**, *112*, 13991-14001.
19. Goforth, H. W.; Laurent, D.; Prusaczyk, W. K.; Schneider, K. E.; Petersen, K. F.; Shulman, G. I., Effects of depletion exercise and light training on muscle glycogen supercompensation in men. *Am J Physiol-Endoc M* **2003**, *285* (6), E1304-E1311.
20. Beghetto, E.; Buffolano, W.; Spadoni, A.; Del Pezzo, M.; Di Cristina, M.; Minenkova, O.; Petersen, E.; Felici, F.; Gargano, N., Use of an immunoglobulin G avidity assay based on recombinant antigens for diagnosis of primary Toxoplasma gondii infection during pregnancy. *J Clin Microbiol* **2003**, *41* (12), 5414-5418.
21. Wolkoff, P.; Skov, P.; Franck, C.; Petersen, L. N., Eye irritation and environmental factors in the office environment - hypotheses, causes and a physiological model. *Scand J Work Env Hea* **2003**, *29* (6), 411-430.
22. Allen, R.; Petersen, S.; Moore, P.; Allen, L.; Heller, Y.; Gurewitsch, E., Do antepartum and intrapartum risk factors differ between mild and severe shoulder dystocia? *Am J Obstet Gynecol* **2003**, *189* (6), S208-S208.
23. Mayo, A. L.; Morris, T. H.; Peltier, S.; Petersen, E. C.; Payne, K.; Holman, L. S.; Tingey, D.; Fogel, T.; Black, B. J.; Gibbs, T. D., Active and inactive groundwater flow systems: Evidence from a stratified, mountainous terrain. *Geol Soc Am Bull* **2003**, *115* (12), 1456-1472.
24. Petersen, J. E.; Kemp, W. M.; Bartleson, R.; Boynton, W. R.; Chen, C. C.; Cornwell, J. C.; Gardner, R. H.; Hinkle, D. C.; Houde, E. D.; Malone, T. C., et al., Multiscale experiments in coastal ecology: Improving realism and advancing theory. *Bioscience* **2003**, *53* (12), 1181-1197.
25. Perseghin, G.; Petersen, K.; Shulman, G. I., Cellular mechanism of insulin resistance: potential links with inflammation. *Int J Obesity* **2003**, *27*, S6-S11.
26. Petersen, K. K.; Hansen, C. W., Compact Campanula carpatica 'Uniform' without chemical growth regulators. *Eur J Hortic Sci* **2003**, *68* (6), 266-+.
27. Cerutti, D. S.; Baker, N. A.; McCammon, J. A., Solvent Reaction Field Potential inside an Uncharged Globular Protein: A Bridge between Implicit and Explicit Solvent Models. *J. Chem. Phys.* **2007**, *127*, 10.1063/1.2771171-1-28.
28. Petersen, G.; Martin, U.; Singhal, A.; Criner, G. J., Intralobar sequestration in the middle-aged and elderly adult: Recognition and radiographic evaluation. *J Thorac Cardiov Sur* **2003**, *126* (6), 2086-2090.

29. Choi, J. H.; Oh, K. I.; Lee, H.; Lee, C.; Cho, M., Nitrile and thiocyanate IR probes: Quantum chemistry calculation studies and multivariate least-square fitting analysis. *J. Chem. Phys.* **2008**, *128*, 134506.
30. Oh, K. I.; Choi, J. H.; Lee, J. H.; Han, J. B.; Lee, H.; Cho, M., Nitrile and thiocyanate IR probes: Molecular dynamics simulation studies. *J. Chem. Phys.* **2008**, *128*, 154504.
31. Eves, N. D.; Petersen, S. R.; Jones, R. L., Effects of helium and 40% O<sub>2</sub> on graded exercise with self-contained breathing apparatus. *Can J Appl Physiol* **2003**, *28* (6), 910-926.
32. Sheu, S. Y.; Yang, D. Y., Determination of Protein Surface Hydration Shell Free Energy of Water Motion: Theoretical Study and Molecular Dynamics Simulation. *J Phys Chem B* **2010**, *114* (49), 16558-16566.
33. Thorbjorn, L.; Petersen, G. H., The epifauna on the carbonate reefs in the Arctic Ikka Fjord, SW Greenland. *Ophelia* **2003**, *57* (3), 177-201.
34. Petersen, P. E., The World Oral Health Report 2003: continuous improvement of oral health in the 21st century - the approach of the WHO Global Oral Health Programme. *Community Dent Oral* **2003**, *31*, 3-23.
35. 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.
36. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
37. 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.
38. 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.
39. 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.
40. 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.
41. Nassar, N.; Horn, G.; Herrmann, C.; Scherer, A.; McCormick, F.; Wittinghofer, A., The 2.2 Å 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.
42. 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)).
43. Ensign, D. L.; Webb, L. J., Factors Determining Electrostatic Fields at the Ras/Effecter Interface. *Proteins* **2011**, *79*, 3511-3524.

44. 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.
45. 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.
46. 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.
47. 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.
48. 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.
49. Humphrey, W.; Dalke, A.; Schulten, K., VMD - Visual Molecular Dynamics. *J. Molec. Graphics* **1996**, *14*, 33-38.
50. 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.
51. Smith, T. F.; Waterman, M. S., Identification of Common Molecular Subsequences. *J Mol Biol* **1981**, *147* (1), 195-197.
52. Russell, R. B.; Barton, G. J., Multiple protein sequence alignment from tertiary structure comparison. *PROTEINS: Struct. Funct. Genet.* **1992**, *14*, 309-323.
53. 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.
54. 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.
55. 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.
56. 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.
57. Roux, B., The calculation of the potential of mean force using computer simulations. *Comp. Phys. Commun.* **1995**, *91*, 275-282.
58. 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.
59. Mardia, K. V., Statistics of Directional Data. *J. Roy. Statist. Soc. Ser. B (Methodol.)* **1975**, *37* (3), 349-393.

60. Wu, J. C.; Chattree, G.; Ren, P. Y., Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor Chem Acc* **2012**, *131* (3).
61. Tuckerman, M. E.; Berne, B. J.; Rossi, A., Molecular-Dynamics Algorithm for Multiple Time Scales - Systems with Disparate Masses. *J Chem Phys* **1991**, *94* (2), 1465-1469.
62. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J Chem Phys* **2007**, *126* (1).
63. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular-Dynamics with Coupling to an External Bath. *J Chem Phys* **1984**, *81* (8), 3684-3690.
64. Berendsen, H. J. C.; Spoel, D. v. d.; Drunen, R. v., GROMACS: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Commun.* **1995**, *91*, 43-56.
65. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* **2008**, *4* (3), 435-447.
66. Lindahl, E.; Hess, B.; Spoel, D. v. d., GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Mod.* **2001**, *7*, 306-317.
67. Honig, B.; Nicholls, A., Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268* (5214), 1144-1149.
68. Fogolari, F.; Brigo, A.; Molinari, H., The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J Mol Recognit* **2002**, *15* (6), 377-392.
69. Fogolari, F.; Zuccato, P.; Esposito, G.; Viglino, P., Biomolecular electrostatics with the linearized Poisson-Boltzmann equation. *Biophys J* **1999**, *76* (1), 1-16.
70. Holst, M.; Saied, F., Multigrid Solution of the Poisson-Boltzmann Equation. *J Comput Chem* **1993**, *14* (1), 105-113.
71. Holst, M. J.; Saied, F., Numerical-Solution of the Nonlinear Poisson-Boltzmann Equation - Developing More Robust and Efficient Methods. *J Comput Chem* **1995**, *16* (3), 337-364.
72. Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W., Modeling electrostatic effects in proteins. *Biochim. Biophys. Acta* **2006**, *1764*, 1647-1676.
73. Schnieders, M. J.; Baker, N. A.; Ren, P. Y.; Ponder, J. W., Polarizable atomic multipole solutes in a Poisson-Boltzmann continuum. *J Chem Phys* **2007**, *126* (12).
74. Schutz, C. N.; Warshel, A., What are Dielectric "Constants" of Proteins and How to Validate Electrostatic Models? *Proteins* **2001**, *44*, 400-417.
75. Ensign, D. L.; Webb, L. J., *Submitted 2013*.
76. Le Questel, J. Y.; Berthelot, M.; Laurence, C., Hydrogen-bond acceptor properties of nitriles: a combined crystallographic and ab initio theoretical investigation. *J Phys Org Chem* **2000**, *13* (6), 347-358.
77. Freitag, M. A.; Gordon, M. S.; Jensen, J. H.; Stevens, W. J., Evaluation of charge penetration between distributed multipolar expansions. *J Chem Phys* **2000**, *112* (17), 7300-7306.

78. Piquemal, J. P.; Gresh, N.; Giessner-Prettner, C., Improved formulas for the calculation of the electrostatic contribution to the intermolecular interaction energy from multipolar expansion of the electronic distribution. *J Phys Chem A* **2003**, *107* (48), 10353-10359.
79. Wang, Q.; He, C.; Qi, R.; Ponder, J. W.; Piquemal, J.; Ren, P., A general model for treating short-range electrostatic penetration in classical mechanics force field. *In Preparation* **2015**.
80. Jeziorski, B.; Moszynski, R.; Szalewicz, K., Perturbation-Theory Approach to Intermolecular Potential-Energy Surfaces of Van-Der-Waals Complexes. *Chem Rev* **1994**, *94* (7), 1887-1930.
81. 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.
82. 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.
83. 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.
84. Andrews, S. S.; Boxer, S. G., Vibrational Stark Effects of Nitriles I. Methods and Experimental Results. *J. Phys. Chem. A* **2000**, *104*, 11853-11863.
85. 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.
86. 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.
87. Fafarman, A. T.; Boxer, S. G., Nitrile Bonds as Infrared Probes of Electrostatics in Ribonuclease S. *J Phys Chem B* **2010**, *114* (42), 13536-13544.
88. Ritchie, A. W.; Webb, L. J., Optimizing Electrostatic Field Calculations with the Adaptive Poisson-Boltzmann Solver to Predict Electric Fields at Protein-Protein Interfaces I: Sampling and Focusing. *J Phys Chem B* **2013**, *117* (39), 11473-11489.
89. Krauss, G., *Biochemistry of Signal Transduction and Regulation*. 3 ed.; WILEY-VCH Verlag: Weinheim, Germany, 2003.
90. Cox, A. D.; Der, C. J., The dark side of Ras: regulation of apoptosis. *Oncogene* **2003**, *22*, 8999-9006.
91. Downward, J., Targeting Ras Signalling Pathways in Cancer Therapy. *Nature Canc. Rev.* **2002**, *3*, 11-22.
92. 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.
93. 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.

94. 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.
95. 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.
96. Herrmann, C., Ras-effector interactions: after one decade. *Curr. Opinion Struct. Biol.* **2003**, *13*, 122-129.
97. 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.
98. 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.
99. 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.
100. 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.
101. Lee, L. P.; Tidor, B., Optimization of Binding Electrostatics: Charge Complementarity in the Barnase-Barstar Protein Complex. *Protein Science* **2001**, *10*, 362-377.
102. Simonson, T., Macromolecular electrostatics: continuum models and their growing pains. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243-252.
103. Villa, J.; Warshel, A., Energetics and Dynamics of Enzymatic Reactions. *J. Phys. Chem. B* **2001**, *105*, 7887-7907.
104. 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.
105. Jiao, D.; Zhang, J.; Duke, R. E.; Li, G.; Schnieders, M. J.; Ren, P., Trypsin-Ligand Binding Free Energies from Explicit and Implicit Solvent Simulations with Polarizable Potential. *J. Comput. Chem.* **2009**, *30*, 1701-1711.
106. Jiao, D. A.; Ren, P. Y., Calculation of protein-ligand binding free energy by a polarizable force field. *Abstr Pap Am Chem S* **2009**, 237.
107. Sigala, P. A.; Fafarman, A. T.; Bogard, P. E.; Boxer, S. G.; Herschlag, D., Do Ligand Binding and Solvent Exclusion Alter the Electrostatic Character within the Oxyanion Hole of an Enzymatic Active Site. *J. Am. Chem. Soc.* **2007**, *129*, 12104-12105.
108. Zhong, S.; MacKerell, A. D., Binding Response: A Descriptor for Selecting Ligand Binding Site on Protein Surfaces. *J. Chem. Inf. Model.* **2007**, *47*, 2303-2315.

109. Alakoskela, J. M. I.; Kinnunen, P. K. J., Control of a Redox Reaction on Lipid Bilayer Surfaces by Membrane Dipole Potential. *Biophys. J.* **2001**, *80*, 294-304.
110. Alakoskela, J. M. I.; Soderlund, T.; Holopainen, J. M.; Kinnunen, P. K. J., Dipole Potential and Head-Group Spacing are Determinants for the Membrane Partitioning of Pregnanolone. *Mol. Pharma.* **2004**, *66*, 161-168.
111. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A., Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.* **2001**, *98*, 10037-10041.
112. Baran, K. L.; Chimenti, M. S.; Schlessman, J. L.; Fitch, C. A.; Herbst, K. J.; Garcia-Moreno, B. E., Electrostatic Effects in a Network of Polar and Ionizable Groups in Staphylococcal Nuclease. *J. Mol. Biol.* **2008**, *379*, 1045-1062.
113. Bredenberg, J. H.; Russo, C.; Fenley, M. O., Salt-Mediated Electrostatics in the Association of TATA Binding Proteins to DNA: A Combined Molecular Mechanics/Poisson-Boltzman Study. *Biophys. J.* **2008**, *94*, 4634-4645.
114. Cui, Q.; Sulea, T.; Schrag, J. D.; Munger, C.; Hung, M. N.; Naim, M.; Cygler, M.; Purisima, E. O., Molecular-Dynamics-Solvated Interaction Energy Studies of Protein-Protein Interactions: The MP1-p14 Scaffolding Complex. *J. Mol. Biol.* **2008**, *379*, 787-802.
115. Ji, C. G.; Zhang, J. Z. H., Electronic Polarization is Important in Stabilizing the Native Structures of Proteins. *J. Phys. Chem. B* **2009**, *113*, 16059-16064.
116. Kiel, C.; Selzer, T.; Shaul, Y.; Schreiber, G.; Herrmann, C., Electrostatically optimized Ras-binding Ral guanine dissociation stimulator mutants increase the rate of association by stabilizing the encounter complex. *Proc. Natl. Acad. Sci.* **2004**, *101*, 9223-9228.
117. Layfield, J. P.; Hammes-Schiffer, S., Calculation of Vibrational Shifts of Nitrile Probes in the Active Site of Ketosteroid Isomerase upon Ligand Binding. *J Am Chem Soc* **2013**, *135* (2), 717-725.
118. Lindquist, B. A.; Furse, K. E.; Corcelli, S. A., Nitrile groups as vibrational probes of biomolecular structure and dynamics: an overview. *Phys. Chem. Chem. Phys.* **2009**, *11*, 8119-8132.
119. Oh, K. I.; Choi, J. H.; Lee, J. H.; Han, J. B.; Lee, H.; Cho, M., Nitrile and thiocyanate IR probes: Molecular dynamics simulation studies. *J Chem Phys* **2008**, *128* (15).
120. Baker, N. A., Poisson-Boltzmann methods for biomolecular electrostatics. *Method Enzymol* **2004**, *383*, 94-+.
121. Rocchia, W.; Alexov, E.; Honig, B., Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *J. Phys. Chem. B* **2001**, *105*, 6507-6514.
122. Simonson, T.; Archontis, G.; Karplus, M., A Poisson-Boltzmann Study of Charge Insertion in an Enzyme Active Site: The Effect of Dielectric Relaxation. *J. Phys. Chem. B* **1999**, *103*, 6142-6156.
123. Teixeira, V. H.; Cunha, C. A.; Machuqueiro, M.; Oliveira, A. S. F.; Victor, B. L.; Soares, C. M.; Baptista, A. M., On the use of different dielectric constants for computing

- individual and pairwise terms in Poisson-Boltzmann studies of protein ionization equilibrium. *J Phys Chem B* **2005**, *109* (30), 14691-14706.
124. Swanson, J. M. J.; Adcock, S. A.; McCammon, J. A., Optimized radii for Poisson-Boltzmann calculations with the AMBER force field. *J Chem Theory Comput* **2005**, *1* (3), 484-493.
125. Jiao, D.; Zhang, J. J.; Duke, R. E.; Li, G. H.; Schnieders, M. J.; Ren, P. Y., Trypsin-Ligand Binding Free Energies from Explicit and Implicit Solvent Simulations with Polarizable Potential. *J Comput Chem* **2009**, *30* (11), 1701-1711.
126. Muegge, I.; Qi, P. X.; Wand, A. J.; Chu, Z. T.; Warshel, A., The reorganization energy of cytochrome c revisited. *J Phys Chem B* **1997**, *101* (5), 825-836.
127. Rasmussen, T. D.; Ren, P. Y.; Ponder, J. W.; Jensen, F., Force field modeling of conformational energies: Importance of multipole moments and intramolecular polarization. *Int J Quantum Chem* **2007**, *107* (6), 1390-1395.
128. Ren, P. Y.; Chun, J. H.; Thomas, D. G.; Schnieders, M. J.; Marucho, M.; Zhang, J. J.; Baker, N. A., Biomolecular electrostatics and solvation: a computational perspective. *Q Rev Biophys* **2012**, *45* (4), 427-491.
129. Shi, Y.; Wu, C. J.; Ponder, J. W.; Ren, P. Y., Multipole Electrostatics in Hydration Free Energy Calculations. *J Comput Chem* **2011**, *32* (5), 967-977.
130. Carlsson, J.; Ander, M.; Nervall, M.; Aqvist, J., Continuum solvation models in the linear interaction energy method. *J Phys Chem B* **2006**, *110* (24), 12034-12041.
131. de Carvalho, S. J.; Fenley, M. O.; da Silva, F. L. B., Protein-Ion Binding Process on Finite Macromolecular Concentration. A Poisson-Boltzmann and Monte Carlo Study. *J Phys Chem B* **2008**, *112* (51), 16766-16776.
132. Freedman, H.; Huynh, L. P.; Le, L.; Cheatham, T. E.; Tuszyński, J. A.; Truong, T. N., Explicitly Solvated Ligand Contribution to Continuum Solvation Models for Binding Free Energies: Selectivity of Theophylline Binding to an RNA Aptamer. *J Phys Chem B* **2010**, *114* (6), 2227-2237.
133. Lei, B. L.; Hameed, M. D. M. A.; Hamza, A.; Wehenkel, M.; Muzyka, J. L.; Yao, X. J.; Kim, K. B.; Zhan, C. G., Molecular Basis of the Selectivity of the Immunoproteasome Catalytic Subunit LMP2-Specific Inhibitor Revealed by Molecular Modeling and Dynamics Simulations. *J Phys Chem B* **2010**, *114* (38), 12333-12339.
134. Suarez, D.; Diaz, N.; Merz, K. M., Molecular dynamics simulations of the dinuclear zinc-beta-lactamase from bacteroides fragilis complexed with imipenem. *J Comput Chem* **2002**, *23* (16), 1587-1600.
135. Zhou, Z. G.; Madura, J. D., Relative free energy of binding and binding mode calculations of HIV-1 RT inhibitors based on dock-MM-PB/GS. *Proteins-Structure Function and Bioinformatics* **2004**, *57* (3), 493-503.
136. Baran, K. L.; Chimenti, M. S.; Schlessman, J. L.; Fitch, C. A.; Herbst, K. J.; Garcia-Moreno, B. E., Electrostatic effects in a network of polar and ionizable groups in staphylococcal nuclease. *J Mol Biol* **2008**, *379* (5), 1045-1062.
137. Forsyth, W. R.; Antosiewicz, J. M.; Robertson, A. D., Empirical Relationships Between Protein Structure and Carboxyl pKa Values in Proteins. *Proteins* **2002**, *48*, 388-403.

138. Harms, M. J.; Castaneda, C. A.; Schlessman, J. L.; Sue, G. R.; Isom, D. G.; Cannon, B. R.; Garcia-Moreno, B. E., The pKa Values of Acidic and Basic Residues Buried at the Same Internal Location in a Protein Are Governed by Different Factors. *J. Mol. Biol.* **2009**, 389, 34-47.
139. Matousek, W. M.; Ciani, B.; Fitch, C. A.; Garcia-Moreno, B.; Kammerer, R. A.; Alexandrescu, A. T., Electrostatic contributions to the stability of the GCN4 leucine zipper structure. *J Mol Biol* **2007**, 374 (1), 206-219.
140. Meng, Y. L.; Roitberg, A. E., Constant pH Replica Exchange Molecular Dynamics in Biomolecules Using a Discrete Protonation Model. *J Chem Theory Comput* **2010**, 6 (4), 1401-1412.
141. Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. N., Hydrogen Bonding Markedly Reduces the pKa of Buried Carboxyl Groups in Proteins. *J. Mol. Biol.* **2006**, 362, 594-604.
142. Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A., Predicting absolute ligand binding free energies to a simple model site. *J Mol Biol* **2007**, 371 (4), 1118-1134.
143. Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S., Predicting small-molecule solvation free energies: An informal blind test for computational chemistry. *J Med Chem* **2008**, 51 (4), 769-779.
144. Nicholls, A.; Wlodek, S.; Grant, J. A., SAMPL2 and continuum modeling. *J Comput Aid Mol Des* **2010**, 24 (4), 293-306.
145. Falconer, R. J.; Penkova, A.; Jelesarov, I.; Collins, B. M., Survey of the year 2008: applications of isothermal titration calorimetry. *J Mol Recognit* **2010**, 23 (5), 395-413.
146. Wennerberg, K.; Rossman, K. L.; Der, C. J., The Ras superfamily at a glance. *J Cell Sci* **2005**, 118 (5), 843-846.
147. Holst, M., Adaptive numerical treatment of elliptic systems on manifolds. *Adv Comput Math* **2001**, 15 (1-4), 139-191.
148. Bank, R. E.; Holst, M., A new paradigm for parallel adaptive meshing algorithms. *Siam J Sci Comput* **2000**, 22 (4), 1411-1443.
149. Baker, N. *Adaptive Poisson-Boltzmann Solver*, 1.3; 2010.
150. Gunner, M. R.; Zhu, X. Y.; Klein, M. C., MCCE analysis of the pK(a)s of introduced buried acids and bases in staphylococcal nuclease. *Proteins-Structure Function and Bioinformatics* **2011**, 79 (12), 3306-3319.
151. Wallace, J. A.; Wang, Y. H.; Shi, C. Y.; Pastoor, K. J.; Nguyen, B. L.; Xia, K.; Shen, J. K., Toward accurate prediction of pK(a) values for internal protein residues: The importance of conformational relaxation and desolvation energy. *Proteins-Structure Function and Bioinformatics* **2011**, 79 (12), 3364-3373.
152. Pearson, J. G.; Oldfield, E.; Lee, F. S.; Warshel, A., Chemical-Shifts in Proteins - a Shielding Trajectory Analysis of the Fluorine Nuclear-Magnetic-Resonance Spectrum of the Escherichia-Coli Galactose Binding-Protein Using a Multipole Shielding Polarizability Local Reaction Field Molecular-Dynamics Approach. *J Am Chem Soc* **1993**, 115 (15), 6851-6862.

153. Danielson, M. A.; Falke, J. J., Use of <sup>19</sup>F NMR To Probe Protein Structure and Conformational Changes. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 163-195.
154. Forsyth, W. R.; Antosiewicz, J. M.; Robertson, A. D., Empirical relationships between protein structure and carboxyl pK(a) values in proteins. *Proteins-Structure Function and Bioinformatics* **2002**, *48* (2), 388-403.
155. Castaneda, C. A.; Fitch, C. A.; Majumdar, A.; Khangulov, V.; Schlessman, J. L.; Garcia-Moreno, B. E., Molecular determinants of the pK(a) values of Asp and Glu residues in staphylococcal nuclease. *Proteins-Structure Function and Bioinformatics* **2009**, *77* (3), 570-588.
156. Harms, M. J.; Castaneda, C. A.; Schlessman, J. L.; Sue, G. R.; Isom, D. G.; Cannon, B. R.; Garcia-Moreno, B., The pK(a) Values of Acidic and Basic Residues Buried at the Same Internal Location in a Protein Are Governed by Different Factors. *J Mol Biol* **2009**, *389* (1), 34-47.
157. Nielsen, J. P.; Petersen, H. H., Objective measurement of health pigs - Application of acute phase proteins. *Acta Vet Scand* **2003**, 125-125.
158. Urbanek, D. C.; Vorobyev, D. Y.; Serrano, A. L.; Gai, F., The two-dimensional vibrational echo of a nitrile probe of the villin HP35 protein. *J. Phys. Chem. Lett.* **2010**, *1*, 3311-3315.
159. Waegele, M. M.; Gai, F., Computational modeling of the nitrile stretching vibration of 5-cyanoindole in water. *J. Phys. Chem. Lett.* **2010**, *1*, 781-786.
160. Waegele, M. M.; Gai, F.; Culik, R. M., Site-Specific Spectroscopic Reporters of the Local Electric Field, Hydration, Structure, and Dynamics of Biomolecules. *J. Phys. Chem. Lett.* **2011**, *2*, 2598-2609.
161. Waegele, M. M.; Tucker, M. J.; Gai, F., 5-Cyanotryptophan as an infrared probe of local hydration status of proteins. *Chem. Phys. Lett.* **2009**, *478*, 249-253.
162. Neves-Petersen, M. T.; Petersen, S. B., **2003**, - Volume 9, - 395.
163. Chen, X. F.; Weber, I.; Harrison, R. W., Hydration water and bulk water in proteins have distinct properties in radial distributions calculated from 105 atomic resolution crystal structures. *J Phys Chem B* **2008**, *112* (38), 12073-12080.
164. Higo, J.; Nakasako, M., Hydration structure of human lysozyme investigated by molecular dynamics simulation and cryogenic X-ray crystal structure analyses: On the correlation between crystal water sites, solvent density, and solvent dipole. *J Comput Chem* **2002**, *23* (14), 1323-1336.
165. Nakasako, M., Large-scale networks of hydration water molecules around bovine beta-trypsin revealed by cryogenic X-ray crystal structure analysis. *J Mol Biol* **1999**, *289* (3), 547-564.
166. Yokomizo, T.; Higo, J.; Nakasako, M., Patterns and networks of hydrogen-bonds in the hydration structure of human lysozyme. *Chem Phys Lett* **2005**, *410* (1-3), 31-35.
167. Chung, E.; Henriques, D.; Renzoni, D.; Zvelebil, M.; Bradshaw, J. M.; Waksman, G.; Robinson, C. V.; Ladbury, J. E., Mass spectrometric and thermodynamic studies reveal the role of water molecules in complexes formed between SH2 domains and tyrosyl phosphopeptides. *Struct Fold Des* **1998**, *6* (9), 1141-1151.

168. Funahashi, J.; Takano, K.; Yamagata, Y.; Yutani, K., Positive contribution of hydration structure on the surface of human lysozyme to the conformational stability. *J Biol Chem* **2002**, *277* (24), 21792-21800.
169. Fennell, C. J.; Kehoe, C. W.; Dill, K. A., Modeling aqueous solvation with semi-explicit assembly. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108* (8), 3234-3239.
170. Fried, S. D.; Bagchi, S.; Boxer, S. G., Measuring Electrostatic Fields in Both Hydrogen-Bonding and Non-Hydrogen-Bonding Environments Using Carbonyl Vibrational Probes. *J Am Chem Soc* **2013**, *135* (30), 11181-11192.
171. Ren, P. Y.; Ponder, J. W., Polarizable atomic multipole water model for molecular mechanics simulation. *J Phys Chem B* **2003**, *107* (24), 5933-5947.
172. Fried, S. D.; Wang, L. P.; Boxer, S. G.; Ren, P. Y.; Pande, V. S., Calculations of the Electric Fields in Liquid Solutions. *J Phys Chem B* **2013**, *117* (50), 16236-16248.