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

### 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 did, in fact, yield 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), unchanged (Rap E30D/K31, Rap E30/K31E), and decreased (Ral monomer, Rap E30/K31) correlation were all observed. The first two cases are not unexpected—either sampling was inadequate and improved sampling improved the simulations 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 though—what about increasing the simulation duration would actual decrease how well the *in vitro* system is represented? The answer is proteins are very complex. It’s possible that 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’s also possible that it’s purely coincidence that the local structure sampled happens to also well-correlate to the experimental measurement. Using napkin math, the Ral monomer has approximately 170 alkane-like sidechain dihedrals and Ral docked to Rap E30/K31 has approximately 460 sidechain dihedrals. Assuming all of them have a non-zero probability of being at approximately 60˚, 180˚, and -60˚ and counting each of those states as a single state (for three total probable states), then there are 1080 possible combinations of sidechain dihedrals for Ral and 10220 combinations for Ral docked to Rap E30/K31. Obviously, a large number of these are so energetically unflavored that they can be neglected—there are certain combinations that are physically impossible due to steric overlap. Assuming that only residues with sidechains within 10 Å of the probe are relevant for electrostatic field calculations (1/r2), which for Y31CSCN docked to Rap E30/K31 is 13, and using an averaging number of dihedrals per residue as 1.7, there are still 103 possible dihedral combinations—significantly more tractable. Yet because of energy barriers, more than 103 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 which may only be energetically favorable given specific orientation of residues further than 10 Å from the probe. It’s very easy to fall go down the rabbit hole and get lost in a sea of dihedral permutations.

The number of sidechain degrees of freedom in a protein is massive and it’s impossible to be 100% certain that the entire ensemble is represented in the appropriate proportions at this stage in computational efficiency. In this regard, I have stepped back to a smaller subset of systems: methylthiocyanate, ethylthiocyanate, hexylthiocyanate, and acetyl-cyanocysteine-N-methylamine peptide-like small molecule. Furthermore, since the principles behind using the AMOEBA force field are still a concern (solute dielectric) 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 currently too large for AMOEBA, but these smaller systems can easily be simulated in a reasonable amount of time (approximately 0.8-1.0 ns/day).

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

### Results and Discussion

#### Sampling CN Orientations

#### Electrostatic Fields



Figure ‑: 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) χ1 analogue C4-C5-C6-S dihedral; (cyan) χ2 analogue C5-C6-S-C dihedral. Capped cyanocysteine (bottom): (blue) χ1 N-Cα-Cβ-S dihedral; (green) χ2 Cα-Cβ-S-C dihedral.



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

Hexylthiocyanate (left) χ1 and χ2 analogous two-dimensional dihedral distribution after 5 ns of simulation. Capped cyanocysteine (right) χ1 and χ2 two-dimensional dihedral distribution after 7 ns of simulation.



Figure ‑: 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 (field midpoint) and the SCN atoms.



Figure ‑: 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 ‑: 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. 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.



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. 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