Chapter 1 Electrostatic Fields at Protein-Protein Interfaces: Increased Sampling Time and Various Electrostatic Methods: A Case for Simulating in Polarizable Force Fields

1.1 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 *ex post facto* to force the calculated field values to yield experimental Stark shifts consistent with the known Stark tuning rate. 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 which do not contain, and therefore are not perturbed (no matter how slightly), by a VSE probe. This would allow for targeted drug design on biologically active biomolecules which are not dependent on assumptions about a probe's degree of perturbation. Unfortunately, for the model to work thusly we need to be significantly more confident in physical veracity of them. To do this, we need to remove assumptions about protein dielectrics.

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 rearrange is orientation in response to the local electrostatic field. This rearrangement aligns its dipole moment parallel to the electrostatic field, resulting in an electrostatic field produced by the water molecule which is antiparallel to the local electrostatic field, reducing the sum electrostatic field, and therefore screening it for any atom further from the field source than that water

molecule. In contrast, a protein interior is significantly limited in the rotational degrees of freedom of sidechains, 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 sidechains *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, which has led us to the polarizable AMOEBA force field.

In this work we examine a variety of classical field calculation methods: RFM PB in Amber03 with a 10^3 Å^3 second-stage box and 193 grid points in each dimension, 5 Å explicit water sphere also with a 10^3 Å^3 second-stage box and 193 grid points in each dimension, explicit TIP3P using GROMACS reaction field electrostatics, hybrid solvent reaction field electrostatics and solute coulomb field, AMOEBA with PB solvent, and AMOEBA with explicit solvent. In both AMOEBA field methods, we also look at adding in charge-penetration via the fitted charge-penetration and the intuitive charge-penetration parameters previously described. In total, we performed 10 different electrostatic field 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. Furthermore, in addition to the Rap GTPases previously studied, 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.

1.2 RESULTS

<Body text to begin here.>

1.3 DISCUSSION

<Body text to begin here.>

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

Fits/Mutant F vs. $\tilde{\nu}$	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	0.581	8.954	0.578	9.378	0.581	8.967	-0.165	-1.293	-0.188	-1.564	-0.168	-1.331
Rap E30/K31E	-0.316	-1.630	-0.267	-1.357	-0.314	-1.616	-0.192	-0.214	-0.223	-0.264	-0.196	-0.219
Rap E30D/K31	-0.846	-3.272	-0.838	-3.296	-0.845	-3.275	-0.279	-0.266	-0.354	-0.365	-0.284	-0.272
Rap E30D/K31E	-0.819	-4.528	-0.812	-4.542	-0.819	-4.530	-0.558	-0.858	-0.568	-0.970	-0.560	-0.871
Ras D30E/E31K	0.072	0.310	-0.013	-0.061	0.069	0.297	0.313	1.066	0.278	0.970	0.311	1.063
Ras D30E/E31	-0.614	-1.948	-0.572	-2.069	-0.612	-1.957	0.162	0.710	0.126	0.594	0.160	0.705
Ras D30/E31K	-0.308	-3.503	-0.265	-2.947	-0.307	-3.493	-0.926	-9.308	-0.927	-9.951	-0.925	-9.379
Ras D30/E31	0.604	4.142	0.626	4.387	0.605	4.150	-0.302	-0.999	-0.304	-1.088	-0.305	-1.016
Ral	0.380	1.920	0.442	2.212	0.384	1.937	-0.496	-0.035	-0.395	-0.044	-0.478	
N27C$_{SCN}$	0.016	0.355	0.001	0.014	0.015	0.350	0.141	2.194	0.133	2.213	0.140	2.201
G28C$_{SCN}$	-0.422	-1.783	-0.403	-1.862	-0.421	-1.787	-0.260	-1.055	-0.294	-1.306	-0.263	-1.073
N29C$_{SCN}$	-0.295	-5.228	-0.272	-5.159	-0.294	-5.229	0.091	1.317	0.085	1.333	0.090	1.315
Y31C$_{SCN}$	-0.288	-1.849	-0.290	-1.907	-0.288	-1.853	-0.033	-0.073	-0.005	-0.013	-0.030	-0.066
K32C$_{SCN}$	-0.165	-1.341	-0.163	-1.399	-0.164	-1.341	0.385	1.512	0.370	1.542	0.383	1.516
N54C$_{SCN}$	-0.204	-2.078	-0.200	-2.126	-0.204	-2.079	0.085	0.350	0.096	0.418	0.086	0.358
All Points	-0.106	-0.725	-0.096	-0.683	-0.106	-0.723	-0.032	-0.137	-0.050	-0.229	-0.034	-0.146

F vs. $\tilde{\nu}$	APBS		APBS 5 ÅSphere		GROMACS TIP3P Reaction Field		Hybrid TI3P Reaction Field	
	R	VSTR	R	VSTR	R	VSTR	R	VSTR
Rap E30/K31	-0.128	-0.798	-0.206	-1.031	-0.427	-2.403	-0.075	-0.556
Rap E30/K31E	-0.659	-3.572	-0.646	-2.629	-0.293	-0.535	-0.739	-3.318
Rap E30D/K31	-0.892	-3.319	-0.862	-2.207	-0.884	-1.230	-0.815	-1.988
Rap E30D/K31E	-0.875	-4.888	-0.841	-3.573	-0.366	-0.747	-0.740	-3.072
Ras D30E/E31K	0.203	0.427	0.112	0.282	0.174	0.543	0.494	1.661
Ras D30E/E31	-0.599	-1.539	-0.527	-1.128	-0.073	-0.190	-0.718	-2.168
Ras D30/E31K	-0.709	-7.424	-0.792	-8.383	-0.866	-7.771	-0.494	-6.381
Ras D30/E31	0.620	3.009	0.471	1.838	-0.136	-0.250	0.781	4.825
Ral	-0.191	-0.563	-0.031	-0.076	0.053	0.083	-0.345	-1.109
N27C$_{SCN}$	0.405	4.339	0.399	3.330	0.283	3.265	0.422	3.609
G28C$_{SCN}$	-0.816	-2.862	-0.694	-2.120	-0.376	-1.013	-0.758	-2.256
N29C$_{SCN}$	-0.359	-4.624	-0.220	-2.636	-0.059	-0.723	-0.328	-4.182
Y31C$_{SCN}$	-0.523	-2.281	-0.338	-1.242	-0.332	-0.922	-0.368	-1.502
K32C$_{SCN}$	-0.320	-0.590	0.052	0.114	0.339	1.288	0.053	0.159
N54C$_{SCN}$	-0.178	-1.497	-0.142	-0.990	-0.051	-0.158	0.270	1.353
All Points	-0.384	-1.761	-0.312	-1.237	-0.159	-0.557	-0.211	-0.992

[a]VSTR has units of $cm^{-1}/(k_bT/eÅ)$. The known VSTR is 1.99 $cm^{-1}/(k_bT/eÅ)$.

Table 1-2: Correlation Coefficients (R) and Virtual Stark Tuning Rates (VSTR[a]) for Relative Field Calculations using Various Electrostatic Models

Fits/Mutant ΔF 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	-5.002	-0.467	-5.063	-0.469	-5.002	0.840	3.913	0.834	4.021	0.838	3.918
Rap E30/K31E	-0.019	-0.068	0.028	0.097	-0.016	-0.057	0.283	0.736	0.265	0.716	0.280	0.732
Rap E30D/K31	-0.052	-0.115	-0.016	-0.036	-0.050	-0.111	-0.046	-0.096	-0.100	-0.215	-0.051	-0.107
Rap E30D/K31E	0.218	0.998	0.225	1.036	0.220	1.007	0.897	2.194	0.876	2.227	0.896	2.194
Ras D30E/E31K	0.126	1.083	0.132	1.215	0.128	1.104	-0.036	-0.160	-0.062	-0.290	-0.037	-0.167
Ras D30E/E31	0.612	13.794	0.574	13.904	0.609	13.779	0.432	7.564	0.426	7.953	0.431	7.576
Ras D30/E31K	0.311	1.345	0.382	1.547	0.316	1.367	0.171	0.842	0.129	0.679	0.169	0.835
Ras D30/E31	0.651	2.095	0.721	2.439	0.655	2.113	0.021	0.083	0.028	0.119	0.020	0.081
All Points	0.016	-4.906	0.034	-2.205	0.018	-4.477	0.202	-0.563	0.182	-0.585	0.200	-0.565

	APBS		APBS 5 ÅSphere		GROMACS TIP3P Reaction Field		Hybrid TI3P Reaction Field	
	R	VSTR	R	VSTR	R	VSTR	R	VSTR
Rap E30/K31	-0.973	-5.091	-0.481	-1.470	0.299	0.854	-0.699	-3.522
Rap E30/K31E	-0.688	-3.674	-0.529	-1.959	-0.010	-0.024	-0.663	-2.873
Rap E30D/K31	-0.728	-2.848	-0.690	-1.578	-0.434	-0.683	-0.797	-1.842
Rap E30D/K31E	-0.387	-1.951	-0.177	-0.518	0.789	1.611	-0.328	-0.701
Ras D30E/E31K	-0.604	-1.698	-0.175	-0.524	0.092	0.314	-0.134	-0.500
Ras D30E/E31	0.264	2.729	0.445	2.248	0.210	1.797	-0.093	-0.469
Ras D30/E31K	-0.393	-1.942	0.015	0.067	0.064	0.236	-0.359	-1.832
Ras D30/E31	-0.265	-0.886	0.136	0.301	-0.288	-0.597	0.098	0.361
All Points	-0.549	0.191	-0.278	0.514	0.018	-8.706	-0.403	0.312

[a]VSTR has units of $cm^{-1}/(k_bT/eÅ)$. The known VSTR is 1.99 $cm^{-1}/(k_bT/eÅ)$.

Table 1-3: 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
$N27C_{SCN}$	0.110	0.109	0.110	0.251	0.253	0.251
$G28C_{SCN}$	0.348	0.348	0.348	0.573	0.574	0.573
$N29C_{SCN}$	-0.605	-0.605	-0.605	-0.815	-0.816	-0.815
$Y31C_{SCN}$	0.111	0.111	0.111	0.025	0.030	0.025
$K32C_{SCN}$	-0.130	-0.129	-0.130	-0.277	-0.276	-0.277
$N54C_{SCN}$	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

	APBS	APBS 5 ÅSphere	GROMACS TIP3P Reaction Field	Hybrid TI3P 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
$N27C_{SCN}$	0.107	0.221	0.191	0.173
$G28C_{SCN}$	0.383	0.298	0.262	0.434
$N29C_{SCN}$	-0.536	-0.748	-0.880	-0.547
$Y31C_{SCN}$	0.024	0.296	0.527	0.031
$K32C_{SCN}$	-0.287	-0.264	-0.453	-0.199
$N54C_{SCN}$	0.261	0.321	0.049	0.304
All Points	-0.189	-0.136	-0.118	-0.235

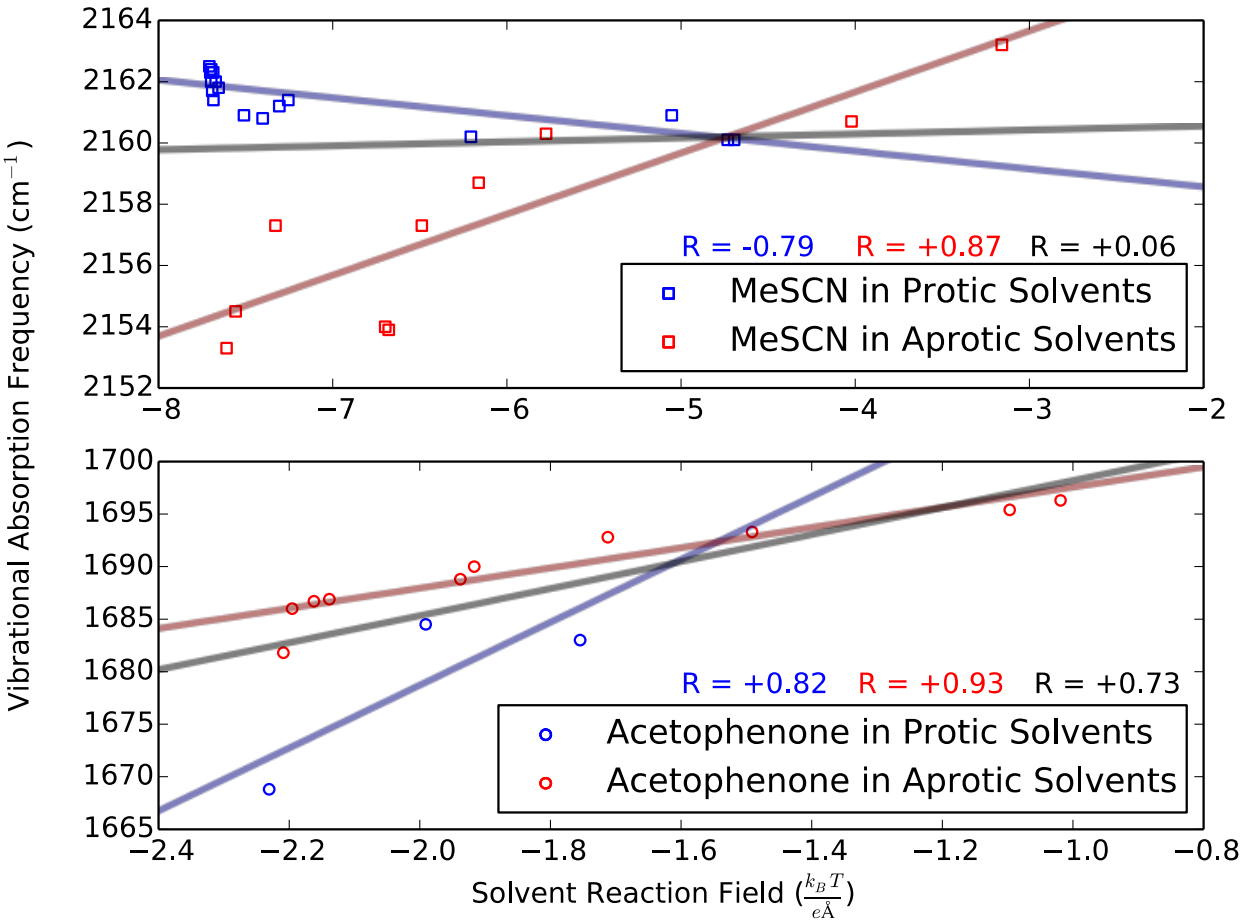

Figure 1-1: Poisson-Boltzmann Solvent Reaction Fields for Methylthiocyanate and Acetophenone in Various Solvents

Solvent Reaction Fields on (top) methylthiocyanate and (bottom) acetophenone calculated using APBS where each solvent is described as a dielectric continuum. Blue: solvents which can donate a hydrogen bond to the vibrational chromophore; red: solvent which cannot hydrogen bond to the vibrational chromophore. Best fit lines and correlations coefficients are included, with black being the aggregate of all data points.

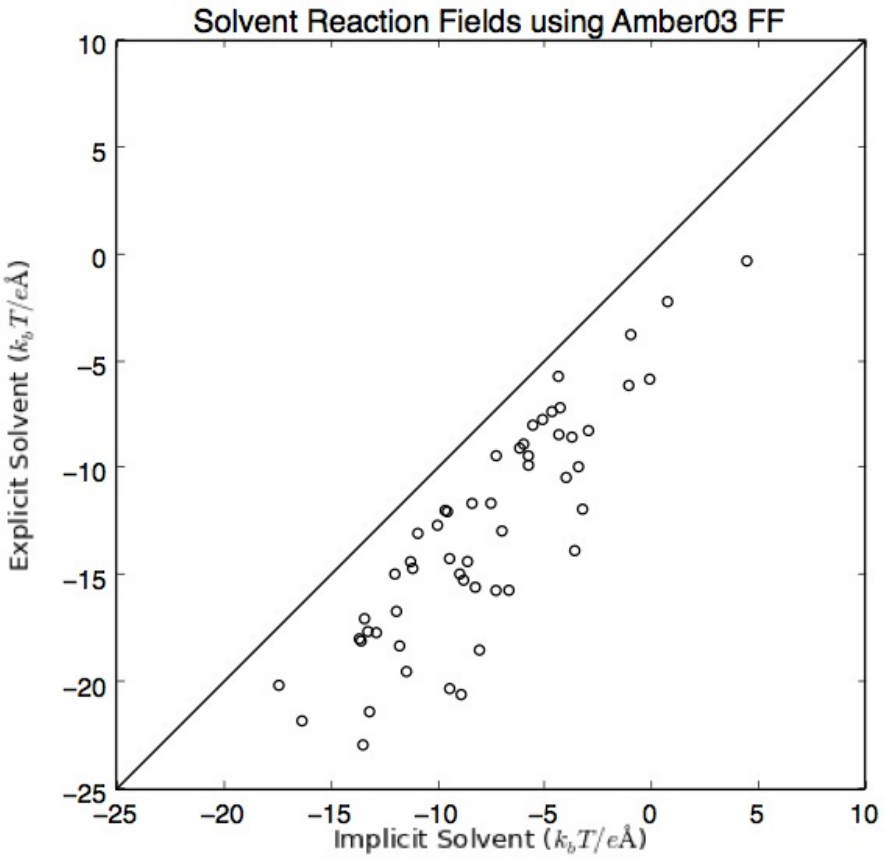

Figure 1-2: Comparison Between Solvent Reaction Fields Calculated using Explicit TI3P Water and Implicit PB Water for All 54 GTPase/Ral Probe Combinations

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 1:1 with the implicit solvent being consistently less negative.