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.

In this discussion, all references to APBS are using the RFM.

1.2 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 1-1.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. Changed in electrostatic field due to the monomer docking to each GTPase system is tabulated in Table 1-2, where the data is presented in the same manner as Table 1-1. 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 the monomeric probe.

1.2.1 Grouping by 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 are cases in which 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), cases in which AMOEBA with explicit solvent (with and without charge-penetration corrections) have the highest magnitude in correlation (Ras D30/E31K), APBS has the highest correlation magnitude (Rap E30D/K31, Rap E30D/K31E), and the hybrid TIP3P reaction field has the highest correlation magnitude (Rap E30/K31E, Ras D30E/E31K, Ras D30E/E31) and there are 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 1.99 $cm^{-1}/(k_bT/eÅ)$, and none of the consistently positive. 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.

1.2.2 Grouping by Probe

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 probe and its immediate surroundings are relatively constant and the only changes being observed are on the binding partner.

Once again, there is not a consistently "best" method for calculating the electrostatic field.

1.2.3 General Remarks

With better solvent sampling, the 5 Å water sphere is no better than the purely implicit solvent. The most remarkable difference in the previous study was how well it

improved the fields for the Ral monomers when grouped by probes, but no such improvement is observed.

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 for 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 approximately 1:1 relationship between the two solvent reaction fields, as previously discussed and shown in Figure 1-2.

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) that the local structural environment was, on average, the same for a probe docked to each GTPase, 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 treat add them after the self-consistent induced dipole 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 interaction to dipole-dipole and quadrupole-quadrupole, may merit reassessment of their usefulness.

It's also highly likely that AMOEBA underperforms due to a non-transferability among ensembles. In the next study, we will examine a couple significantly smaller systems which have been sampled in AMOEBA.

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

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	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$_{SCN}$	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$_{SCN}$	-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$_{SCN}$	-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$_{SCN}$	-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$_{SCN}$	-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$_{SCN}$	-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				
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$_{SCN}$	0.405	0.230	0.399	0.300	0.283	0.306	0.422	0.277				
G28C$_{SCN}$	-0.816	-0.349	-0.694	-0.472	-0.376	-0.987	-0.758	-0.443				
N29C$_{SCN}$	-0.359	-0.216	-0.220	-0.379	-0.059	-1.383	-0.328	-0.239				
Y31C$_{SCN}$	-0.523	-0.438	-0.338	-0.805	-0.332	-1.085	-0.368	-0.666				
K32C$_{SCN}$	-0.320	-1.694	0.052	8.800	0.339	0.776	0.053	6.286				
N54C$_{SCN}$	-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				

[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

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	-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 ÅSphere		GROMACS TIP3P Reaction Field		Hybrid TI3P Reaction Field	
	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

[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

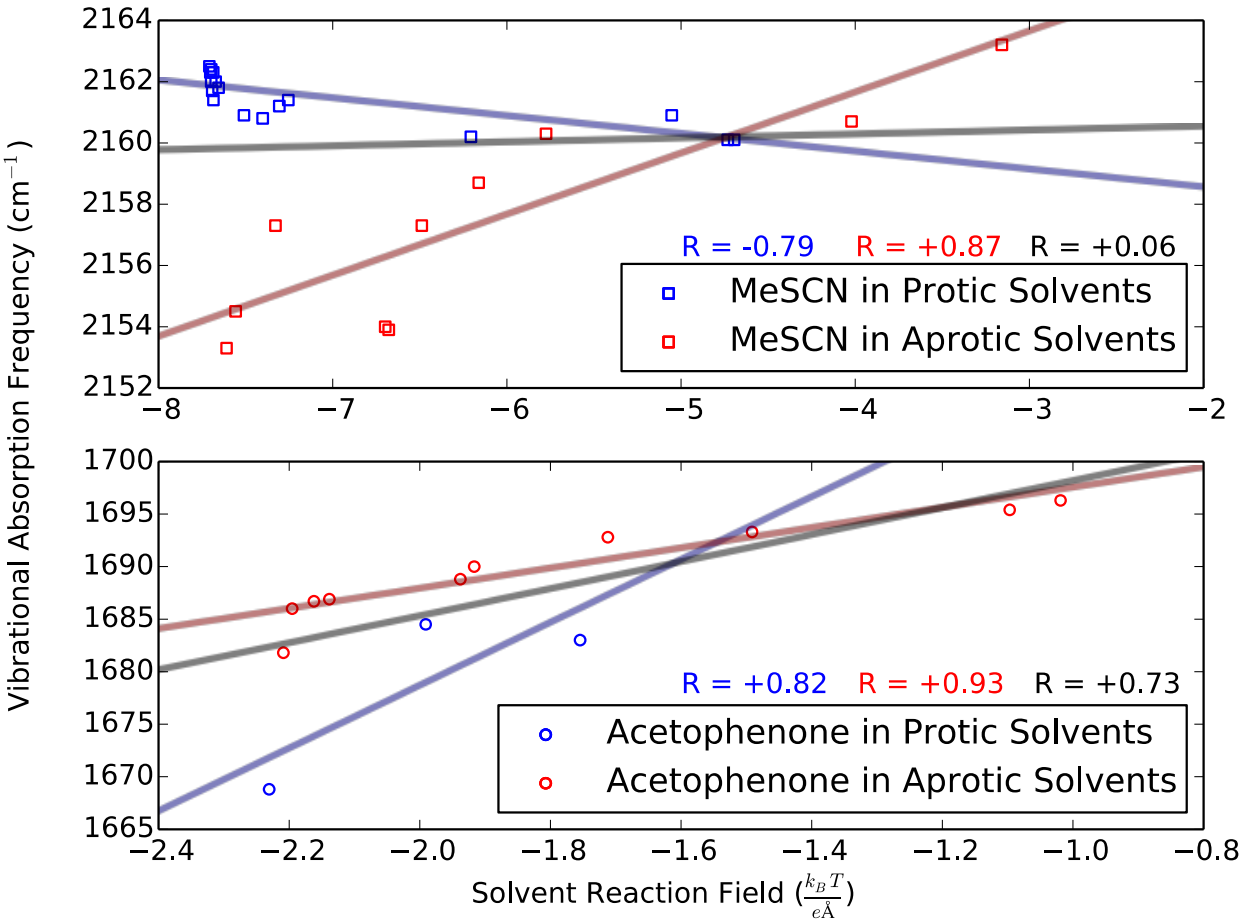

Figure 1-1: Poisson-Boltzmann Solvent Reaction Fields for Methylthiocyanate and Acetophenone in Various Solvents

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. Experimental measurements for methylthiocyanate are unpublished were performed by Christina Ragain, Ph.D., Josh Slocum, and Kelsey Eklund. Experimental measurements for acetophenone were previously reported by Fried *et al.*[1]

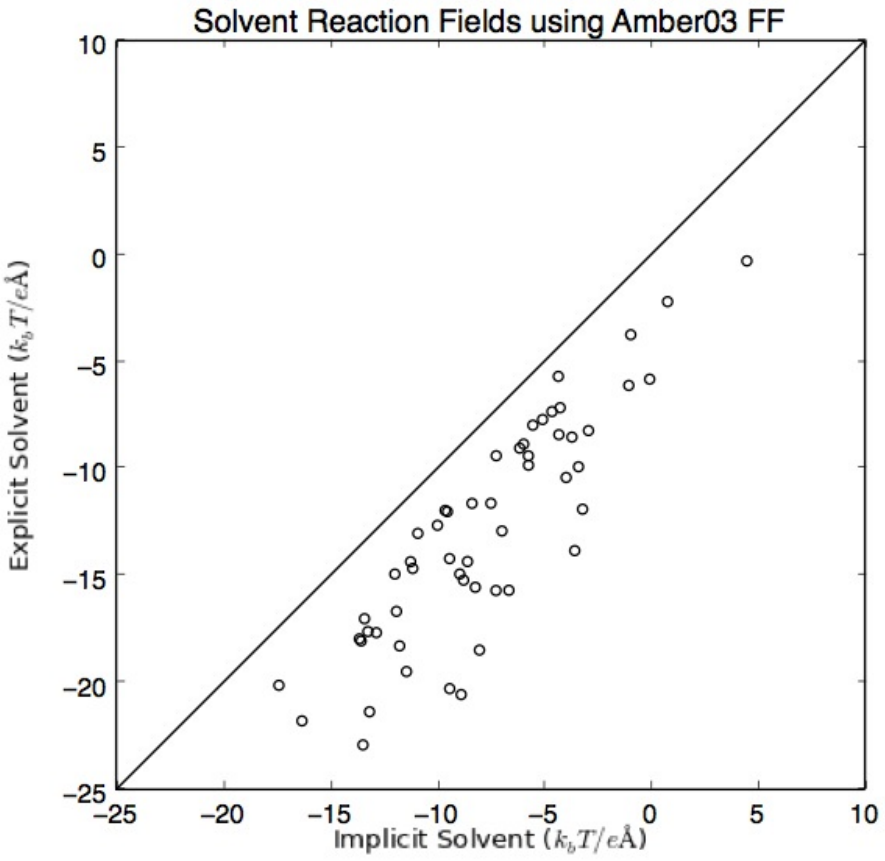

Figure 1-2: Comparison Between Solvent Reaction Fields Calculated using Explicit TI3P Water and Implicit PB Water for All 54 GTPase/Ral Probe Combinations

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.

(1) Fried, S. D.; Bagchi, S.; Boxer, S. G. *J Am Chem Soc* **2013**, *135*, 11181.