## Force field comparison and analysis

### Overall agreement with experiment.

Overall, the force field combinations we tested (Table 2) produced moderate agreement with the experimental reference data from ITC studies37; see Table 3, Table 4, Figure 3, and Figure 4. Thus, the binding free energy and enthalpy predictions had an RMSE range of 0.9 - 1.8 kcal/mol and 0.9 - 4.0 kcal/mol, respectively. No single force field combination emerged as superior to the others at predicting both binding free energy and enthalpy. The Q4RG-TIP3P force field, which we expect would be the initial, "default" choice for an AMBER user, shows poor correlation (R2 = 0.44) and significant deviation (RMSE = 1.8 kcal/mol) from experimental binding free energies, making it one of the worst force fields tested. For binding enthalpies, the Q4RG-TIP3P force field showed improved correlation (R2 = 0.67), but even greater deviation from experiment (RMSE = 2.0 kcal/mol). Surprisingly, the best force field for predicting the binding free energy was the BGBG-TIP3P combination, which produced the highest correlation (R2 = 0.56) and lowest deviation (RMSE = 0.9 kcal/mol) to experimental values of all force fields tested. The BGBG-TIP3P combination was expected to perform poorly because we crudely parameterized the CD host molecule and made no effort to tune its experimentally known conformational properties, as was done for the Q4MD-CD force field.45 On the other hand, the BGBG-TIP3P combination was among the worst at predicting binding enthalpy, showing low correlation (R2=0.39) and high deviation (RMSE = 2.6 kcal/mol) relative to experiment. The best force field for binding enthalpies, Q4RG-TIP4Pew (R2 = 0.76, RMSE = 0.9 kcal/mol), produced mediocre binding free energy results, similar to those observed for Q4RG-TIP3P.