Supplementary Material for: Efficient calculation of molecular configurational entropies using an information theoretic approximation

Bracken M. King, Nathaniel W. Silver, and Bruce Tidor
Massachusetts Institute of Technology

September 24, 2011

Demonstration that additional terms cannot decrease MI

We aim to demonstrate that additional terms included to the mutual information cannot decrease its value. That is

$$I(\mathbf{r_i}; \mathbf{r_j}) \leq I(\mathbf{r_i}; \mathbf{r_j}, \mathbf{r_k}). \tag{S1}$$

We start with the fact that the conditional entropy is less than or equal to the unconditioned entropy, and further that additional conditioning terms can only decrease the entropy further [1]

$$S(\mathbf{r_i}) \geq S(\mathbf{r_i}|\mathbf{r_j}) \geq S(\mathbf{r_i}|\mathbf{r_j}, \mathbf{r_k}). \tag{S2}$$

We next negate both sides of the inequality, and add $S(\mathbf{r_i})$ to both sides,

$$S(\mathbf{r_i}) - S(\mathbf{r_i}|\mathbf{r_j}) \leq S(\mathbf{r_i}) - S(\mathbf{r_i}|\mathbf{r_j}, \mathbf{r_k}). \tag{S3}$$

By the definition of mutual information, we can rewrite this expression as

$$I(\mathbf{r_i}; \mathbf{r_j}) \leq I(\mathbf{r_i}; \mathbf{r_j}, \mathbf{r_k}), \tag{S4}$$

which is the relationship we aimed to develop.

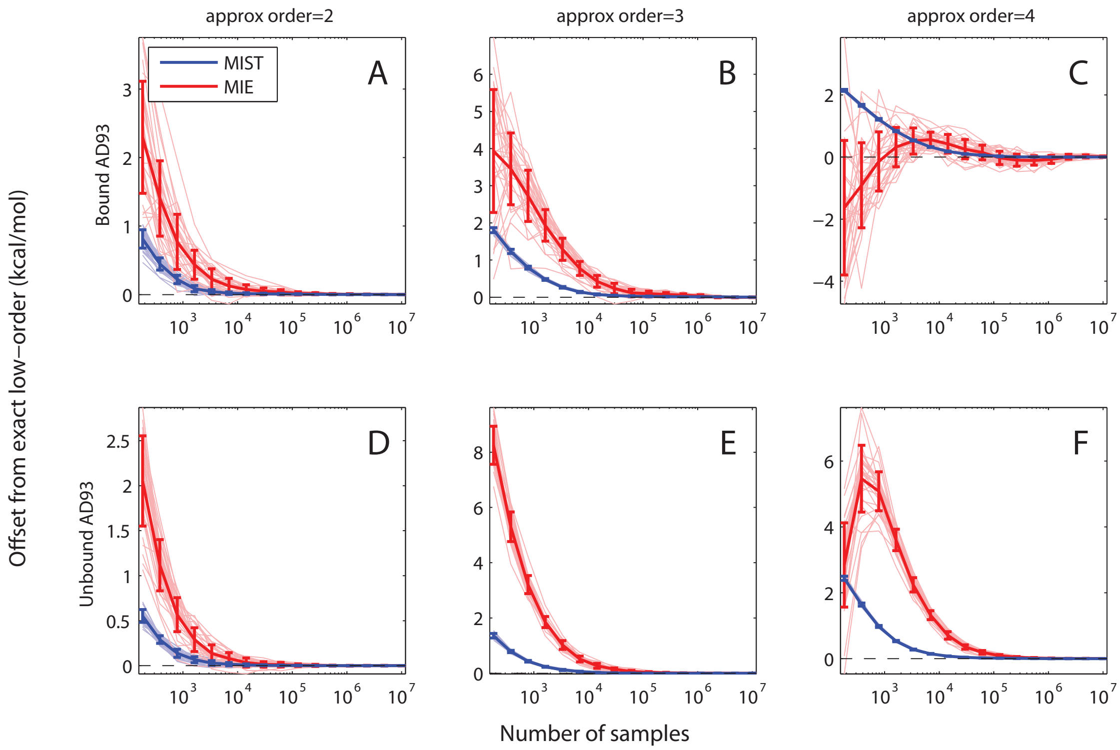

Figure S1: **Convergence in AD93 rotameric systems**: For each of the eight idealized rotameric systems, we sampled with replacement from the 5×10^4 configurations representing the system, according to the Boltzmann distribution determined by the relative energies of each configuration. These samples were then used to estimate the marginal entropies of all combinations of 1–4 torsions prior to application of MIST (blue lines) or MIE (red lines) to compute $-TS^{\circ}$. This procedure was repeated 50 times for each system, and the deviation of each run from the exact result to the same order approximation are shown (pale lines), as well as the mean and standard deviation across the 50 runs (thick lines). Results for bound (top row) and unbound (bottom row) AD93 are shown here.

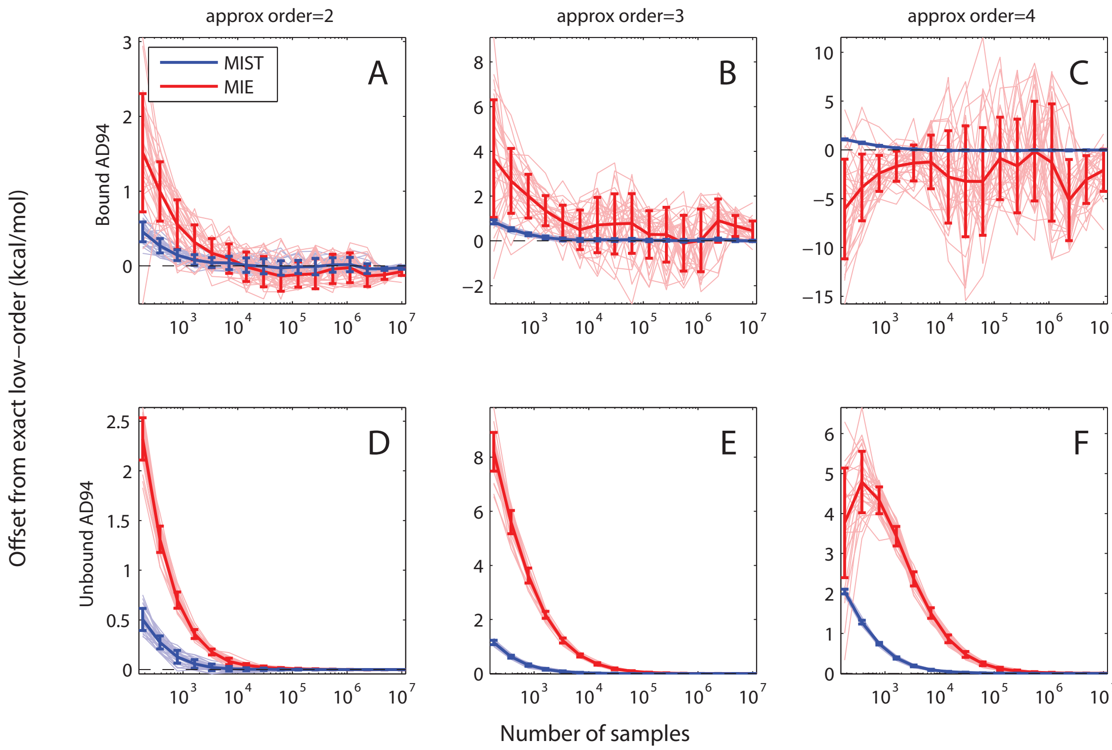

Figure S2: **Convergence in AD94 rotameric systems**: For each of the eight idealized rotameric systems, we sampled with replacement from the 5×10^4 configurations representing the system, according to the Boltzmann distribution determined by the relative energies of each configuration. These samples were then used to estimate the marginal entropies of all combinations of 1–4 torsions prior to application of MIST (blue lines) or MIE (red lines) to compute $-TS^{\circ}$. This procedure was repeated 50 times for each system, and the deviation of each run from the exact result to the same order approximation are shown (pale lines), as well as the mean and standard deviation across the 50 runs (thick lines). Results for bound (top row) and unbound (bottom row) AD94 are shown here.

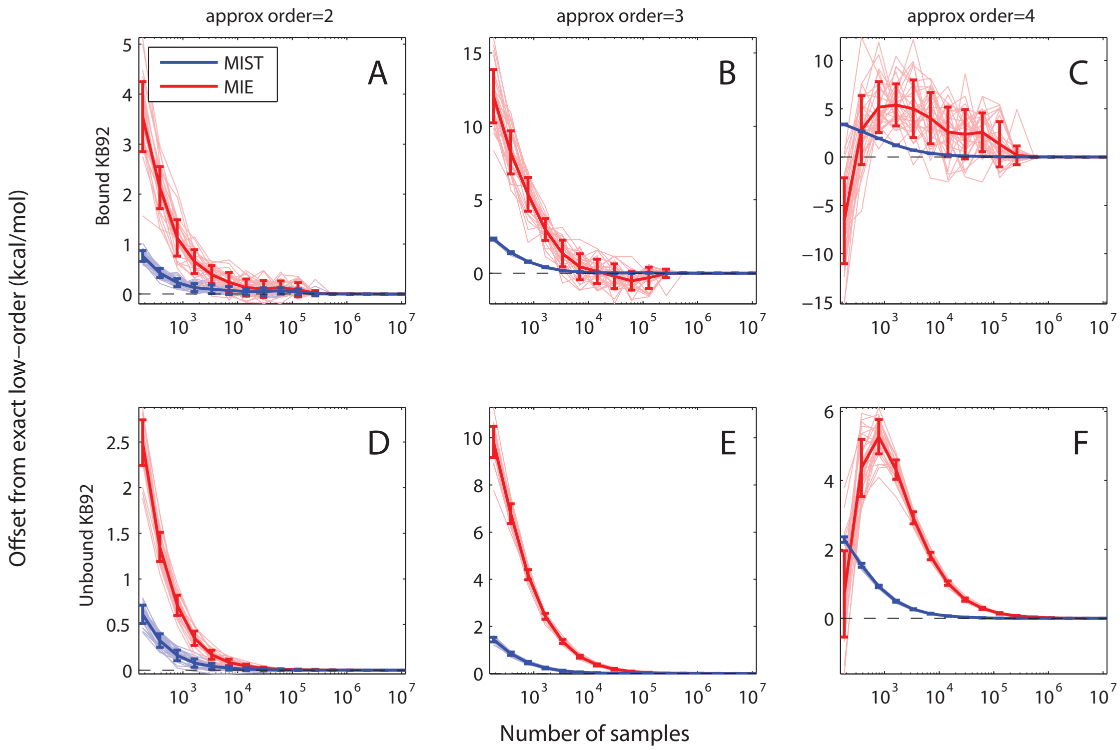

Figure S3: **Convergence in KB92 rotameric systems**: For each of the eight idealized rotameric systems, we sampled with replacement from the 5×10^4 configurations representing the system, according to the Boltzmann distribution determined by the relative energies of each configuration. These samples were then used to estimate the marginal entropies of all combinations of 1–4 torsions prior to application of MIST (blue lines) or MIE (red lines) to compute $-TS°$. This procedure was repeated 50 times for each system, and the deviation of each run from the exact result to the same order approximation are shown (pale lines), as well as the mean and standard deviation across the 50 runs (thick lines). Results for bound (top row) and unbound (bottom row) KB92 are shown here.

References

[1] T. M. Cover and Joy A Thomas. *Elements of Information Theory*. Wiley-Interscience, Hoboken, N.J., 2nd edition, 2006.