# Supporting Information for "Consistent Refinement of Submitted Models at CASP using a Knowledge-based Potential"

**Chopra et al.**

**FIGURE S1.** Showing the contribution of magnitude of average structural improvements in CαRMSD, GDT_TS and GDT_HA using our KB01 refinement protocol compared to energy minimization with the KB01 potential and MESHI stereochemical correction on all models of CASP7. The KB01 refinement protocol (blue line) and energy minimization by KB01 potential (magenta line) are very close to each other for CαRMSD, GDT_TS and GDT_HA over the complete range of modeling difficulties indicating that most of the structural improvements come from using energy minimization on a native-favoring potential surface like KB01. The MESHI stereochemical correction (green line) has negligible impact on structural improvement.

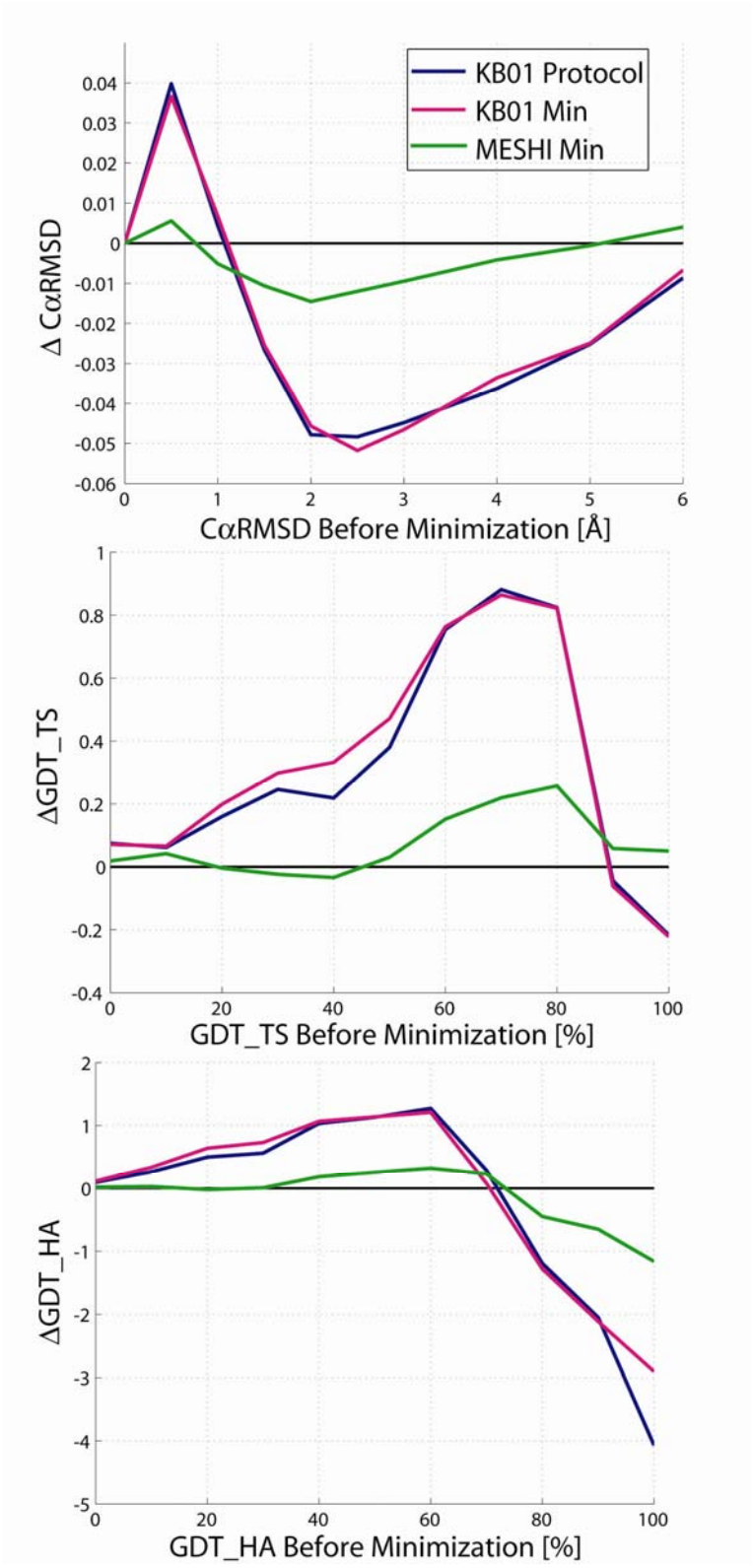
KB01 Protocol
KB01 Min
MESHI Min
Δ CαRMSD
CαRMSD Before Minimization [Å]
ΔGDT_TS
GDT_TS Before Minimization [%]
ΔGDT_HA
GDT_HA Before Minimization [%]

**FIGURE S1**

**FIGURE S2.** Showing improvement in the high accuracy GDT_HA scores for (A) all CASP7 groups (B) the Baker group (C) the Zhang Server group and (D) the TASSER group. The maximum average GDT_HA improvement is around 1% for all CASP7 as well as the Baker models and around 2% for the Zhang Server and TASSER models for the region of interest. The improvement in GDT_HA by MODELLER is negligible for CASP7 targets of all modeling difficulty, with a change in GDT_HA close to zero. For the KB01 refinement protocol, there are seven models with more than 4% improvement in GDT_HA for Baker, six models with more than 6% improvement in GDT_HA for Zhang Server and three models with more than 10% improvement in GDT_HA for TASSER.

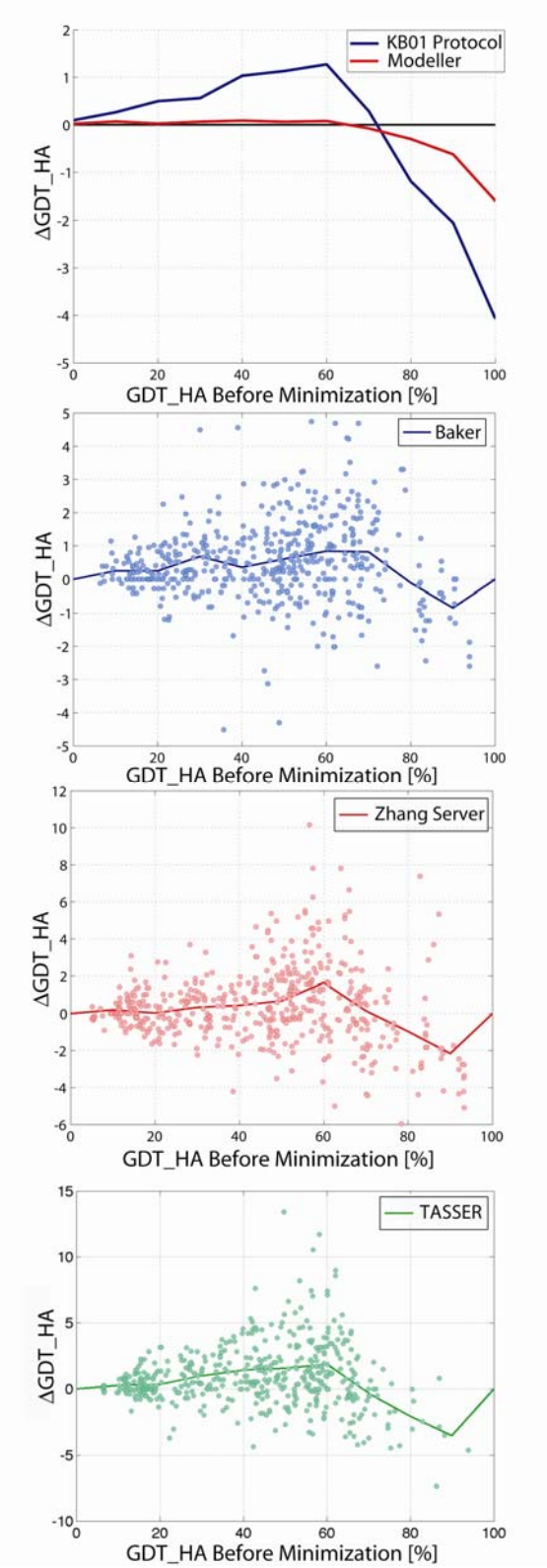
KB01 Protocol
Modeller
ΔGDT_HA
GDT_HA Before Minimization [%]
Baker
ΔGDT_HA
GDT_HA Before Minimization [%]
Zhang Server
ΔGDT_HA
GDT_HA Before Minimization [%]
TASSER
ΔGDT_HA
GDT_HA Before Minimization [%]

**FIGURE S2**

**FIGURE S3.** Mean weighted $C^{\alpha}$ root mean square deviation (wRMS) improvement for near-native decoys (Chopra et al., Proc Natl Acad Sci USA 2008;105: 20239-20244) as a function of the weight of the KB01 term in ENCAD (green curve). Optimal scale values are between 0.1 and 2.0 (magenta dotted lines). A test on all of CASP7 models (blue curve) based on ΔGDT_TS give the same optimal range for the scale of the KB01 non-bonded energy term in ENCAD.

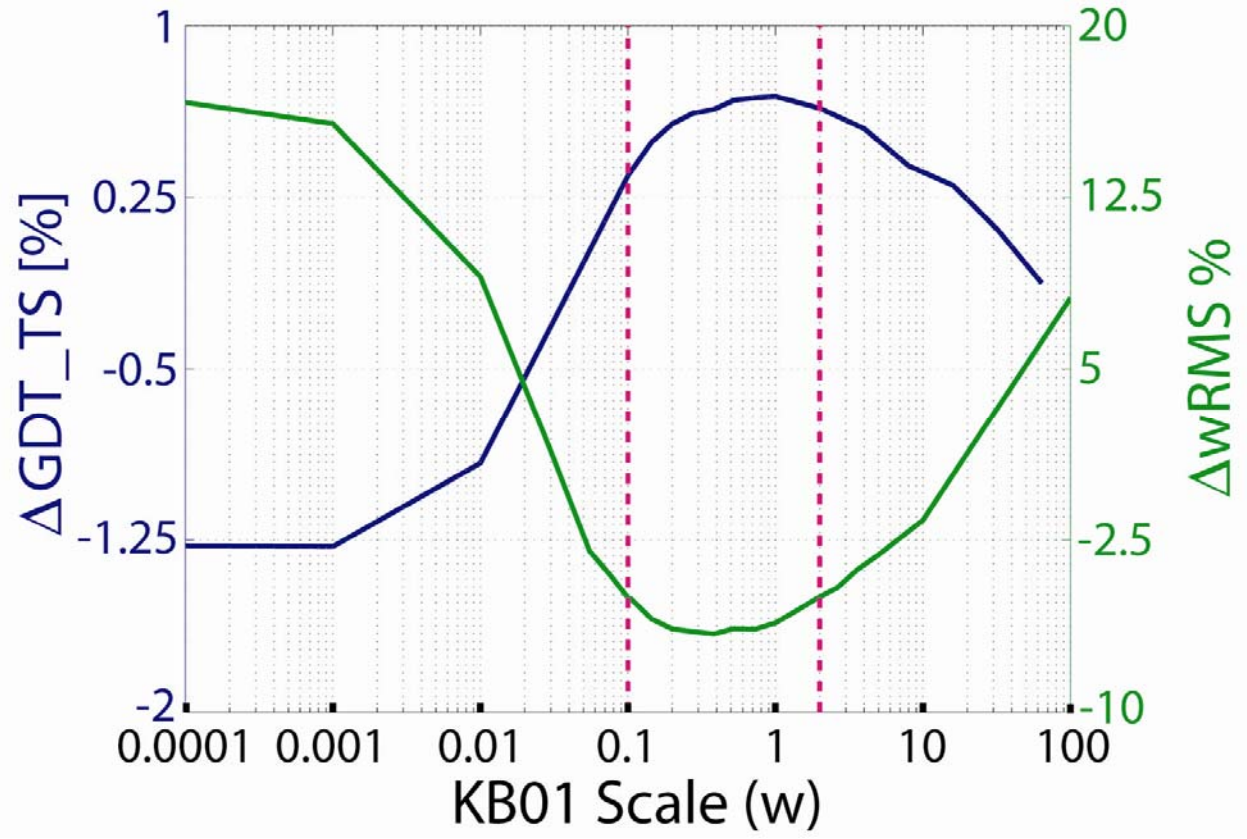


**FIGURE S3**

**FIGURE S4.** Showing the CPU time taken by the KB01 protocol as a function of target chain length. This test used Model 1 from the MetaTASSER group for each of the 114 targets of CASP7 as that group benefits most from the KB01 refinement protocol (Fig. 5). The average time taken to run a model was 106.1 seconds (1.8 minutes). The minimum time is 14 seconds for T0363_D1 (56 residues) and the maximum time is 817 seconds (13.6 minutes) for T0326 (304 residues). For a given chain length, the CPU time shows wide variation but refinement of a protein with up to 300 residues always takes less than 5 minutes.

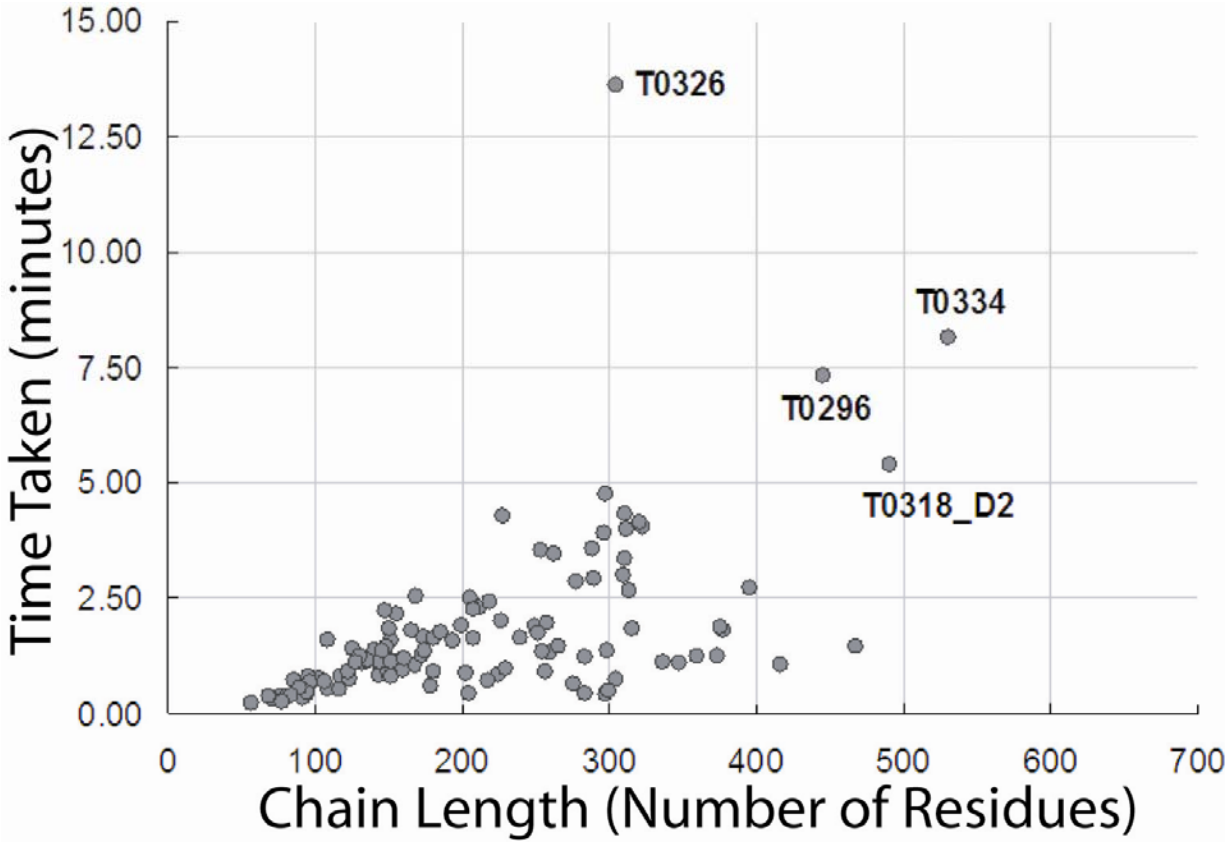


**FIGURE S4**

**FIGURE S5.** Showing how KB01 refinement improves GDT_TS and GDT_HA values for residues classified by DSSP as “Helix” (red line), “Sheet” (blue line) and “Coil” (green line). These results are averaged over all models of CASP7. Most improvement in GDT_TS and GDT_HA values is for helical residues, with less improvement for coil residues and little improvement for sheet residues. Both helical and coil residues are refined for almost the entire range of modeling difficulty, whereas sheet residues are not refined when the initial GDT_TS ranges from 20 to 60%. The “All” GDT_TS and GDT_HA (grey lines) data are comparable to data in Figures 1(B) and S2. Note that the “All” GDT_TS and GDT_HA values are calculated as the weighted mean of “Helix”, “Sheet” and “Coil” GDT_TS and GDT_HA values using the overall secondary structure composition for all CASP7 models (36.85% helix, 19.97% sheet and 43.18% coil).

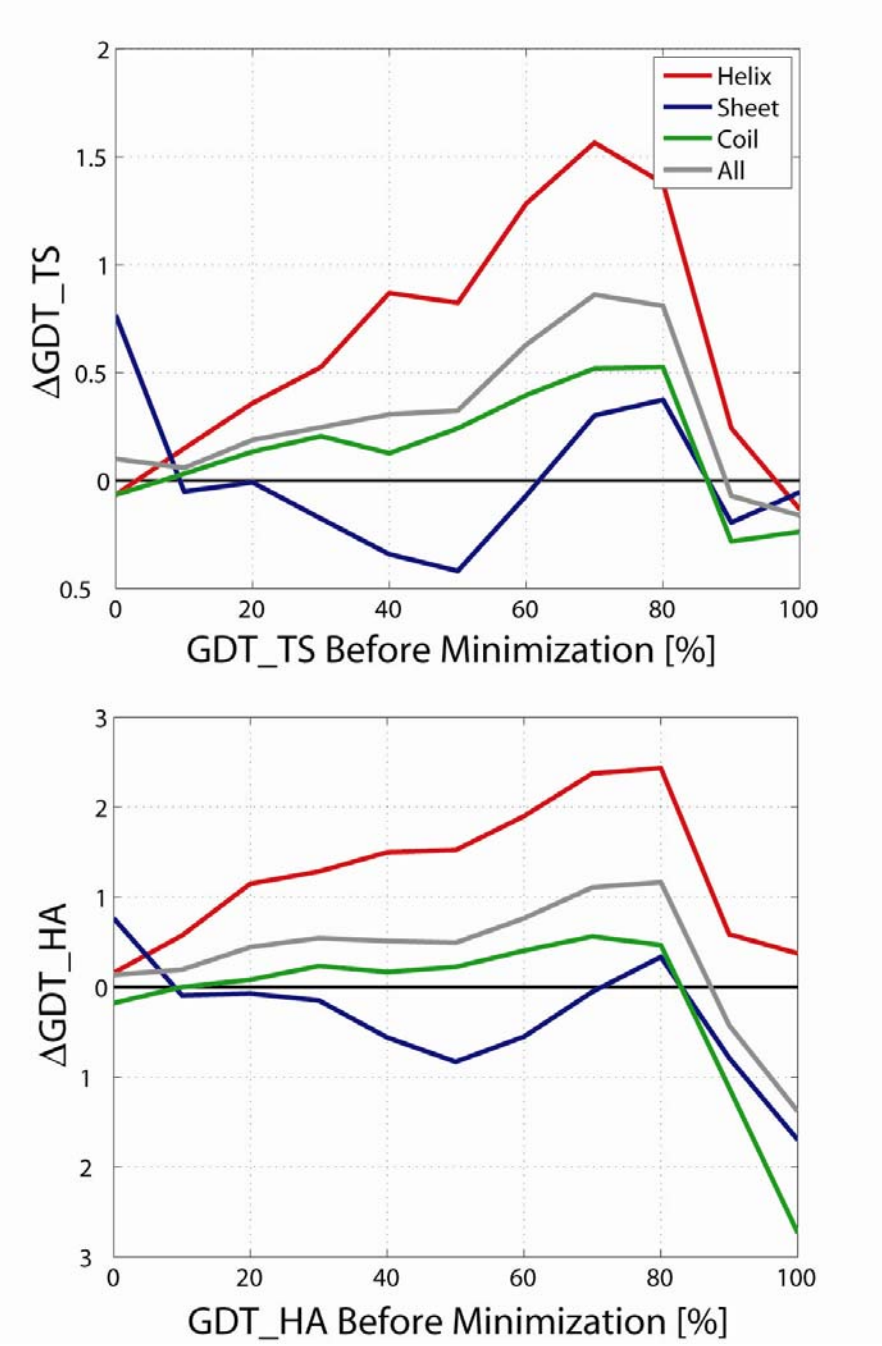
Helix
Sheet
Coil
All
ΔGDT_TS
GDT_TS Before Minimization [%]
ΔGDT_HA
GDT_HA Before Minimization [%]

**FIGURE S5**