Supplementary Material:
Homolog Detection Using Global Sequence Properties:
An Alternate View of Structural Encoding
in Protein Sequences
H.A. Scheraga and S. Rackovsky

1. The Definition of $\Delta_k(P,Q)$

      The unnormalized sine and cosine Fourier coefficients for wave number k and property factor l are denoted by $a_k^{(l)}$ and $b_k^{(l)}$ respectively. Standard scores are defined in terms of the averages of the Fourier coefficients over an ensemble of all possible permutations of the N-member protein sequence, denoted by angle brackets, and the corresponding standard deviation, as follows:

$$Z(a_k^{(l)}) = \frac{a_k^{(l)} - \left\langle a_k^{(l)} \right\rangle_N}{\sigma(a_k^{(l)})}$$

with an analogous expression for $Z(b_k^{(l)})$. Based on results obtained in Reference 13, we define the k-dependent distance function as follows:

$$\Delta_k(P,Q) = [\sum_{l=1}^{10} \{Z(b_k^{(l)}[P]) - Z(b_k^{(l)}[Q])\}^2]^{1/2} \quad (k=1,3,4,6)$$

$$= [\sum_{l=1}^{10} (\{Z(a_k^{(l)}[P]) - Z(a_k^{(l)}[Q])\}^2 + \{Z(b_k^{(l)}[P]) - Z(b_k^{(l)}[Q])\}^2)]^{1/2} \quad (k=2)$$

$$= [\sum_{l=1}^{10} \{Z(a_k^{(l)}[P]) - Z(a_k^{(l)}[Q])\}^2]^{1/2} \quad (k=5)$$

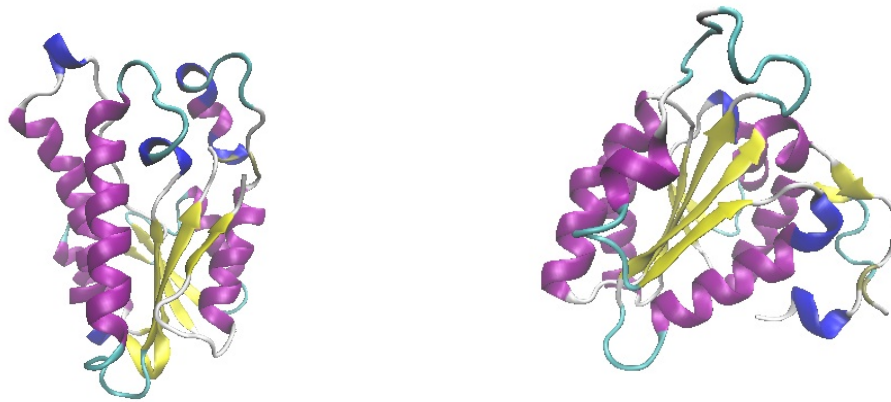$$= [\sum_{l=1}^{10} \{Z(b_0^{(l)}[P]) - Z(b_0^{(l)}[Q])\}^2]^{1/2} \quad (k=0)$$

Figure S1. CASP8 Targets and Closest Database Structures
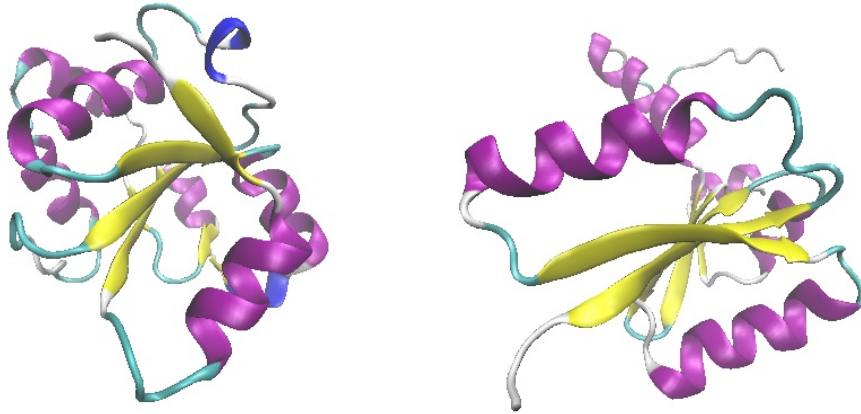
a. CASP8 target T0389 (left) and the minimum-distance structure in the database (right).
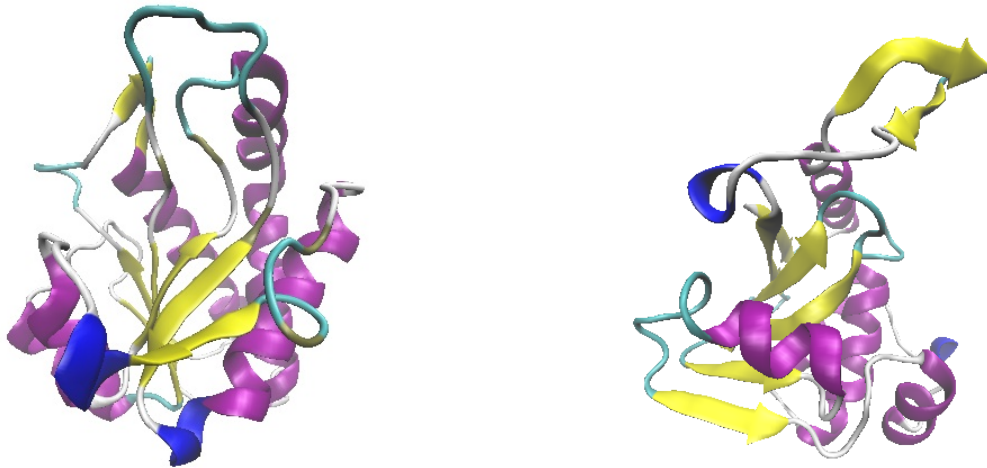
b. CASP8 target T0425 (left) and the minimum-distance structure in the database (right).

c. CASP8 target T0433 (left) and the minimum-distance structure in the database (right).

d. CASP8 target T0457 (left) and the minimum-distance structure in the database (right).

e. CASP8 target T0507 (left) and the minimum-distance structure in the database (right).