elements and global chain contour similarity is retained for structures with a TM-score to native of 0.4. Such structures are a rich source of information that needs to be better exploited, both in the design of improved structure prediction algorithms and approaches to model refinement. Moreover, while the notion of discrete folds is convenient as a classification tool, it may perhaps be more productive to focus on the key structural features that a pair proteins have in common without resorting to arbitrary assignments of fold similarity. If such similarity is detected, then as demonstrated for the QS structures, quite high quality models can be built, regardless of whether the templates have any local secondary structure in common, nor whether they are evolutionarily related or not. The key challenge is to develop methods that can routinely identify this similarity in the limit when their evolutionary relationship, if any, cannot be detected.

Acknowledgements/Dedication

This paper is dedicated with deep affection and admiration to Professor Harold Scheraga on the occasion of his 90th birthday. This research was supported in part by grant No. GM-48835 of the Division of General Medical Sciences of the National Institutes of Health. We thank the Taylor, Trovato and Zhou groups for generously providing us with their structural libraries.

Supporting information available:

Figure S.1 Fraction of unmatched targets in the Taylor, Zhou and QS sets to the PDB300 library as a function of the number of target protein residues. Red (black) indicates a TM-score threshold of 0.45 (0.50). Dashed (black) lines are for the original (smoothed) structures. This material is available free of charge via the Internet at http://pubs.acs.org.