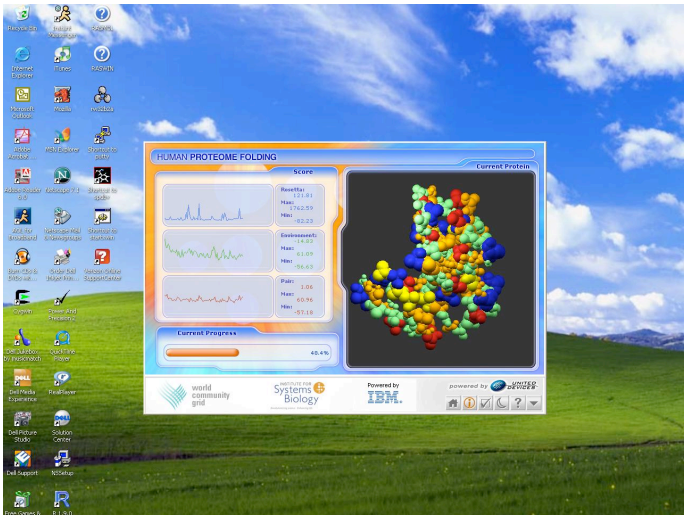
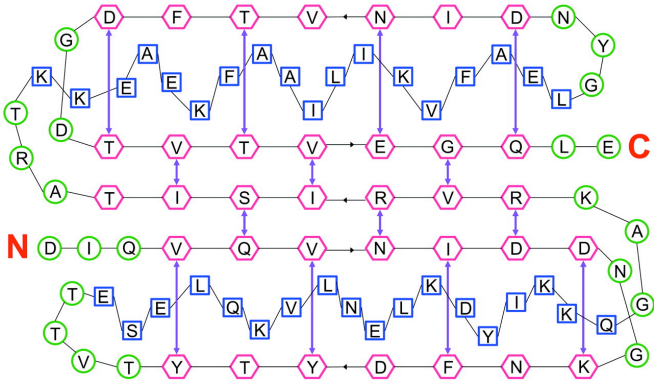
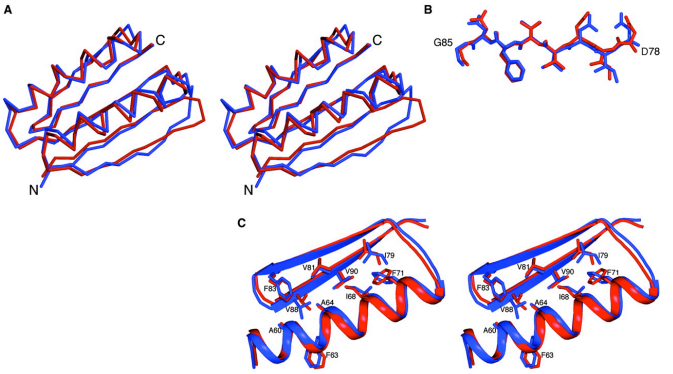



## Protein Design



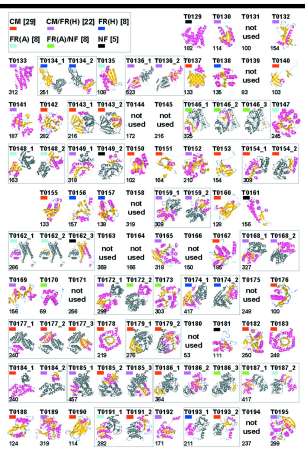

## Protein Design




## The Five Categories of CASP Targets

1. CM/E (Comparative Modeling / Easy) ←
   Structural homolog found by BLAST.
2. CM/H (Comparative Modeling / Hard) ←
   structural homolog found by 5 rounds of PSI-BLAST.
3. FR/H (Fold Recognition / Homology) ←
   Structural comparison to PDB finds a structure found by PSI-BLAST.
4. FR/A (Fold Recognition / Analogy) ←
   Finds a similar structure, no evidence of sequence homology.
5. NF (New Fold) ← nothing “similar” in the PDB

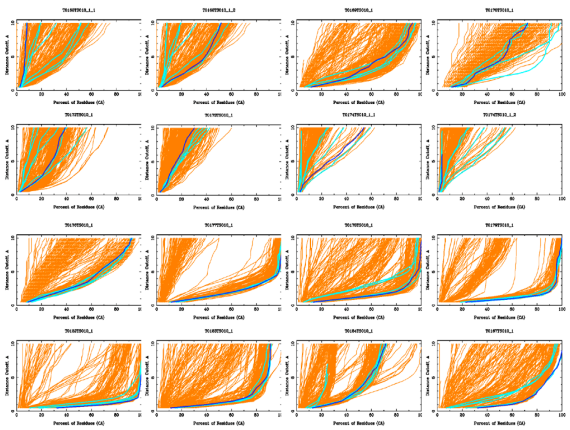
## CASP



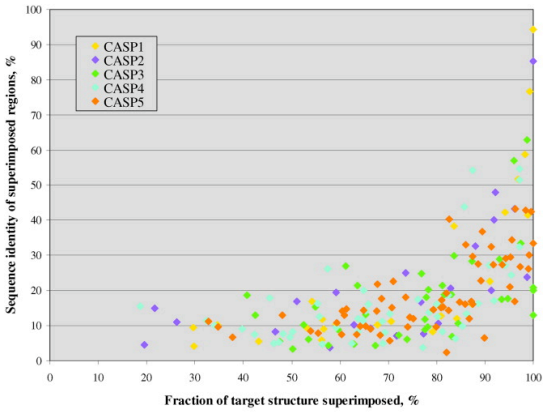

## CASP Questions

1. Are the models produced similar to the corresponding experimental structure?
2. Is the mapping of the target sequence onto the proposed structure (i.e. the alignment) correct?
3. Have similar structures that a model can be based on been identified?
4. Are the details of the models correct?
5. Has there been progress from the earlier CASPs?
6. What methods are most effective?
7. Where can future effort be most productively focused?

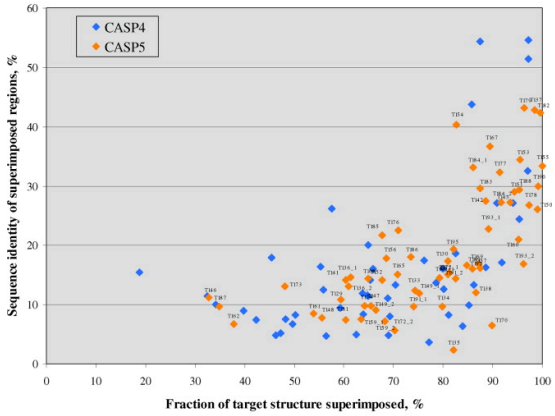
## Hubbard Plots




## Distribution of Target Difficulty




## Distribution of Target Difficulty

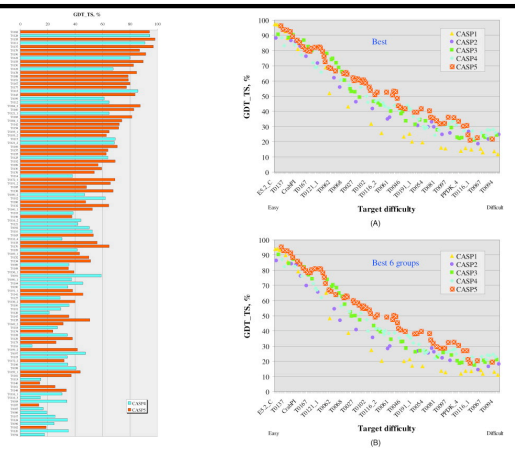



## Overall Model Quality Assessment

Venclovas et al (2003): "A large sample of possible structure superpositions of the model on the corresponding experimental structure is generated by superposing all sets of three, five, and seven consecutive Ca along the backbone (each peptide segment provides one superposition). Each of these initial superpositions is iteratively extended, including all residue pairs under a specified threshold in the next iteration, and continuing until there is no change in included residues. The procedure is conducted by using thresholds of 1, 2, 4, and 8 Å, and the superposition that includes the maximum number of residues, is selected for each threshold ... GDT_TS is then obtained by averaging over the four superposition scores for the different thresholds:

$$GDT\_TS = (N1+N2+N4+N8) / 4$$

## CASP5 Progress

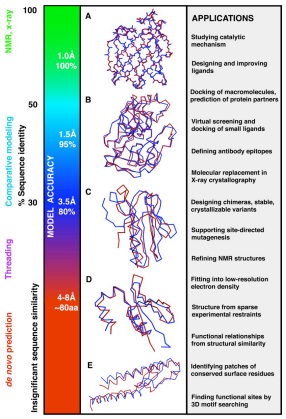


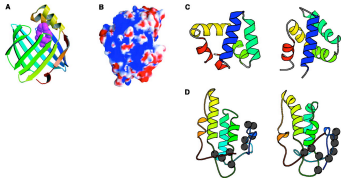
## CASP Problem Areas and Bottlenecks

1. Alignment of a sequence onto a template fold.
2. Model refinement - improving accuracy of initial models.
3. Accurately modeling regions of insertion and deletion relative to a template structure.
4. Improved fold recognition, particularly for analogous, analogous/new fold targets.
5. Improved New Fold methods (for recognizing new folds).

Always the same …

## Applications of Structure Prediction




Baker and Sali, Science 294 (2001)

## Ten Most Wanted


## CASP 6

90 Domains assessed:

| | | |
|---|---|---|
| CM / E | → | 25 |
| CM / H | → | 18 |
| FR / H | → | 21 |
| FR / A | → | 16 |
| NF | → | 9 |