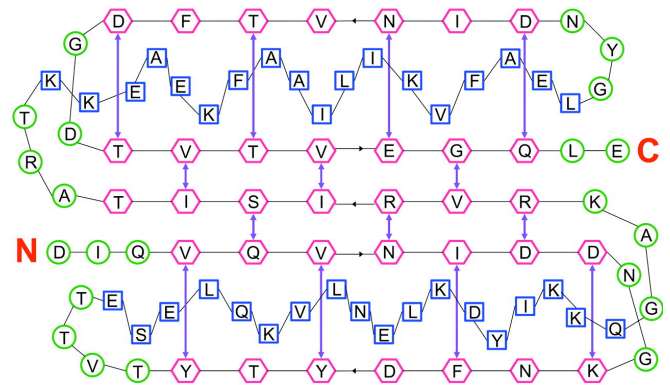
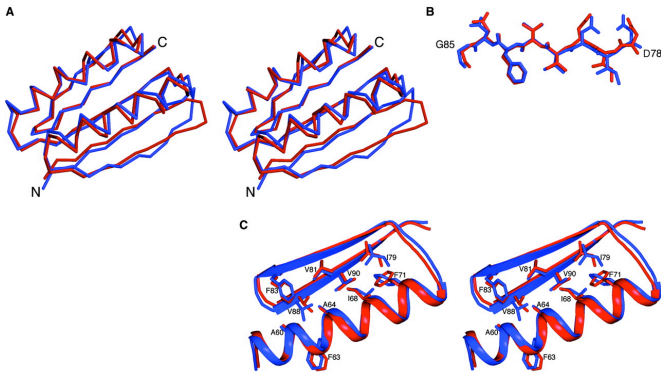


## Protein Design



## 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

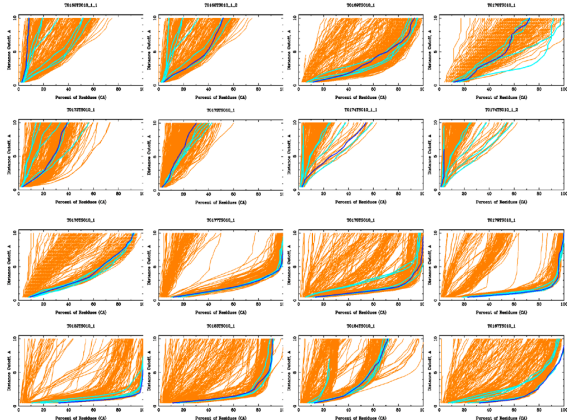
## CASP

CM [28]	CM/FR(H) [22]	FR(H) [8]	FR(A) [8]	FR(A)/NF [8]	NF [5]	T0129	T0130	T0131	T0132
T0133	T0134_1	T0134_2	T0135	T0136_1	T0136_2	T0137	T0138	T0139	T0140
T0141	T0142	T0143_1	T0143_2	T0144	T0145	T0146_1	T0146_2	T0146_3	T0147
T0148_1	T0148_2	T0149_1	T0149_2	T0150	T0151	T0152	T0153	T0154_1	T0154_2
T0155	T0156	T0157	T0158	T0159_1	T0159_2	T0160	T0161		
T0162_1	T0162_2	T0162_3	T0163	T0164	T0165	T0166	T0167	T0168_1	T0168_2
T0169	T0170	T0171	T0172_1	T0172_2	T0173	T0174_1	T0174_2	T0175	T0176
T0177_1	T0177_2	T0177_3	T0178	T0179_1	T0179_2	T0180	T0181	T0182	T0183
T0184_1	T0184_2	T0184_3	T0185_1	T0185_2	T0186_1	T0186_2	T0187_1	T0187_2	T0187_3
T0188	T0189	T0190	T0191_1	T0191_2	T0192	T0193_1	T0193_2	T0194	T0195

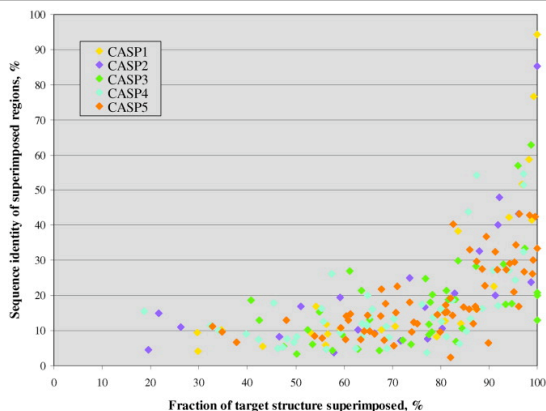
## CASP Questions

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?

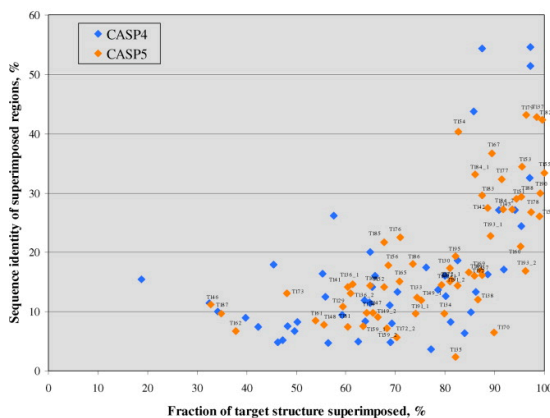
## Hubbard Plots



## Distribution of Target Difficulty



## Distribution of Target Difficulty

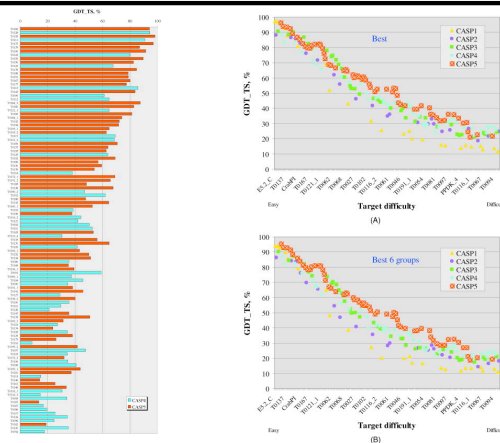


## Overall Model Quality Assessment

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

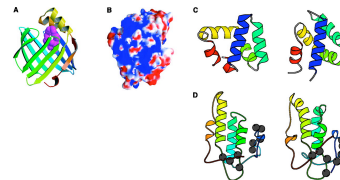
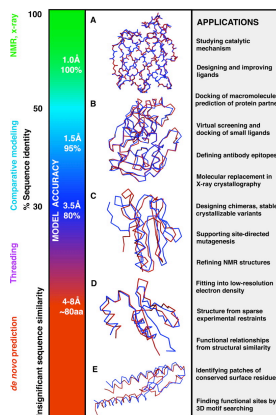


## CASP Problem Areas and Bottlenecks

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