Swiss Institute of Bioinformatics

BIOZENTRUM
Universität Basel

Assessment of Disorder Prediction

CASP7
Asilomar, November 27, 2006

Lorenza Bordoli

Florian Kiefer

Torsten Schwede

Prediction Format and Data

```
The symbols for the 2 state order/disorder prediction are 'O'=order,
'D'=disorder. […] Last column should indicate a probability of a residue
being in the disordered region. The value of this confidence level is in
the range of 0.0 - 1.0.
```

- Disorder: D / O
- Disorder probability: $0.0 \leq p \leq 1.0$
- A residue has been defined as disordered:
  - X-Ray: no coordinates present for crystallized residues (SEQRES)
  - NMR: Residues whose conformation is not sufficiently defined by NMR restraints (ensemble variation or definition by experimentalist in the REMARK section)

CASP7 Disorder Dataset

- Target residue classification:
  - 6 % disordered residues (CASP6: 5.25 %)
  - 94 % ordered residues (CASP6: 94.75%)
- Prediction for 96 targets assessed
  - On a per residue level
  - Only segments > 3 residues were considered.
  - 19’816 residues as prediction targets

Submissions by 19 Predictor Groups

Group	Targets predicted	Residues predicted	P-values	D(cutoff)	O(cutoff)
132	96	19816	cont.	>=0.50	<0.50
140	96	19816	cont.	>=0.50	<0.50
153	95	19723	cont.	>=0.50	<0.50
168	96	19816	cont.	>=0.594843	<=0.553115
188	10	2221	cont.	<0.50	>=0.50 ★
253	96	19816	cont.	>=0.50	<0.50
271	94	19368	cont.	>=0.50	<0.50
272	96	19816	cont.	>=0.50	<0.50
284	95	19494	0.0,1.0	1	0 ★
393	96	19816	cont.	>=0.594843	<=0.553115
443	96	19816	cont.	>=0.50	<0.50
470	96	19816	cont.	>=0.50	<0.50
538	96	19816	cont.	>=0.50	<0.50
572	95	19816	cont.	>=0.50	<0.50
590	95	19541	cont.	>=0.50	<0.50
594	82	16687	cont.*	>=0.85	<=0.85
609	93	19231	cont.**	>0.5	<=0.50
681	75	15180	cont.	>=0.50	<=0.50
686	96	19816	cont.	>0.5	<=0.5

In blue: servers

```
**0,      0.25,     0.33,     0.5,      0.66,     0.75,     1
*In       steps     of        0.01
```

Naïve predictors

- Frequencies of disordered residues at the N- and C- termini is higher than the average
- Several naives predictors were included, predicting 2,3,4,5,6,7,8,9,10 disordered residues at the N- and C- terminus

Group	Targets	Residues predicted	P-values predicted	D(cutoff)	O(cutoff)
Naiv[2-10]	96	19694	0.2,0.8	>=0.50	<0.50

Evaluation Criteria

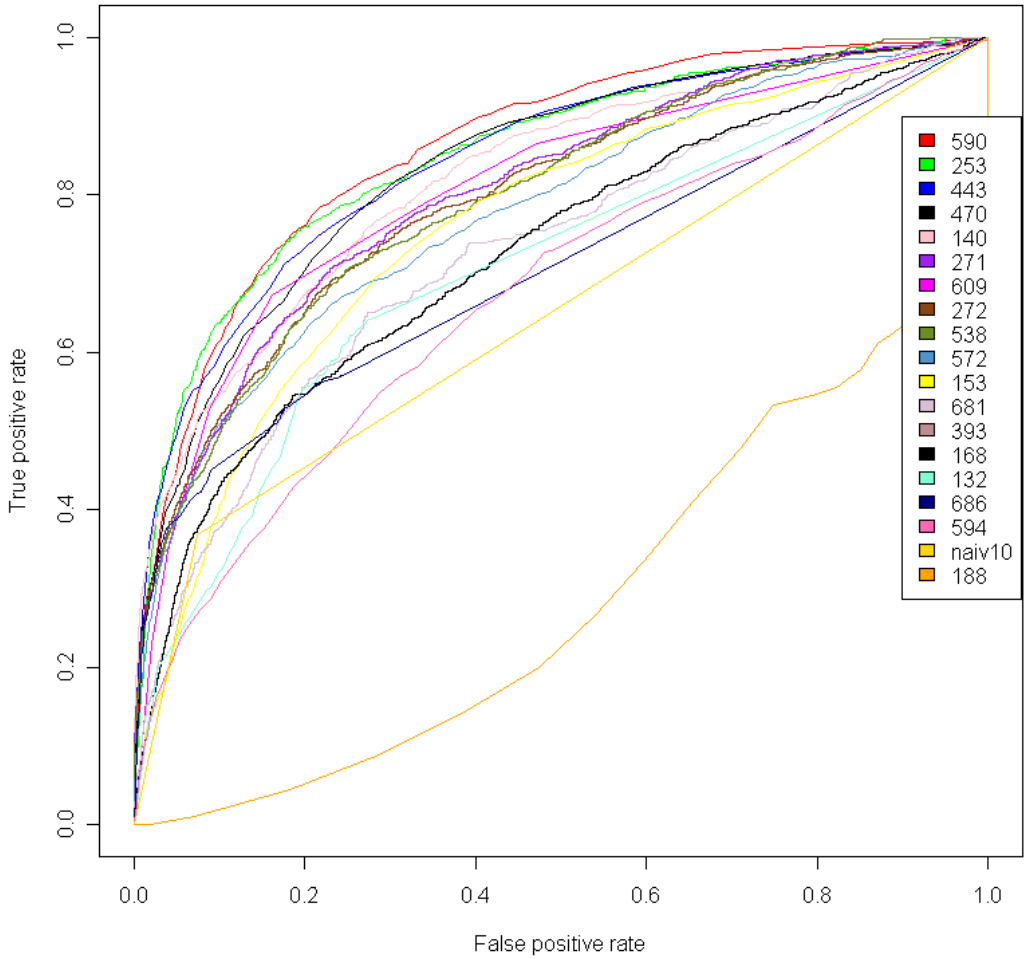
- Continuous prediction (P-value):
  - ROC curves
- Binary Prediction (O / D)
  1. Sensitivity, Specificity, Product and Average
  2. Sw score (CASP6, Dunbrack [1])

[1] Jin Y. and Dunbrack RL. (2005) Proteins, Suppl 7:167-175

ROC Curves

The area under the curve (AUC) is used as measure of prediction accuracy (*Sroc*)

- For each value of *P*:
  - *TP(P) rate* : fraction of true positives in prediction with *p>P*
  - *FP(P) rate :* fraction of false positives in prediction with *p>P*



Sensitivity /Specificity

- Sensitivity and Specificity are commonly used measures to evaluate predictive accuracy:

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{TP}{N_{disorder}} = \frac{TP(P)}{N_{disorder}}$$

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{TN}{N_{order}} = \frac{TN(P)}{N_{order}}$$

Binary Scores

- Combined scores of Sensitivity and Selectivity:
  - Good methods would ideally have both a high sensitivity and a high specificity
  - The impact of Specificity (Sp) can be seen on the rate of false positive predictions FP = (1 – Sp)
  - Combined to make a single measurement criteria:

$$\sqrt{S_{product}} = \sqrt{S_{sens} \times S_{spec}} = \sqrt{\frac{TP \times TN}{N_{disorder} \times N_{order}}}$$

$$ACC = \frac{S_{sens} + S_{spec}}{2}$$

Sw score (CASP6, Dunbrack)

- Most residues (in CASP7: 94%) are ordered and a simple Q2 measure would not work well.

⇒ weighted score to reward for correctly predicting a disordered residue rather than an ordered one (Sw) introduced by Dunbrack:

$$S_w = \frac{S}{S_{\max}} = \frac{W_{disorder} TP - W_{order} FP + W_{order} TN - W_{disorder} FN}{W_{disorder} N_{disorder} + W_{order} N_{order}}$$

- Weights: $W_{disorder} = 94$, $W_{order} = 6$
- Sw varies between -1 and +1, with a random prediction resulting in Sw = 0

[1] Jin Y. and Dunbrack RL. (2005) Proteins, Suppl 7:167-175

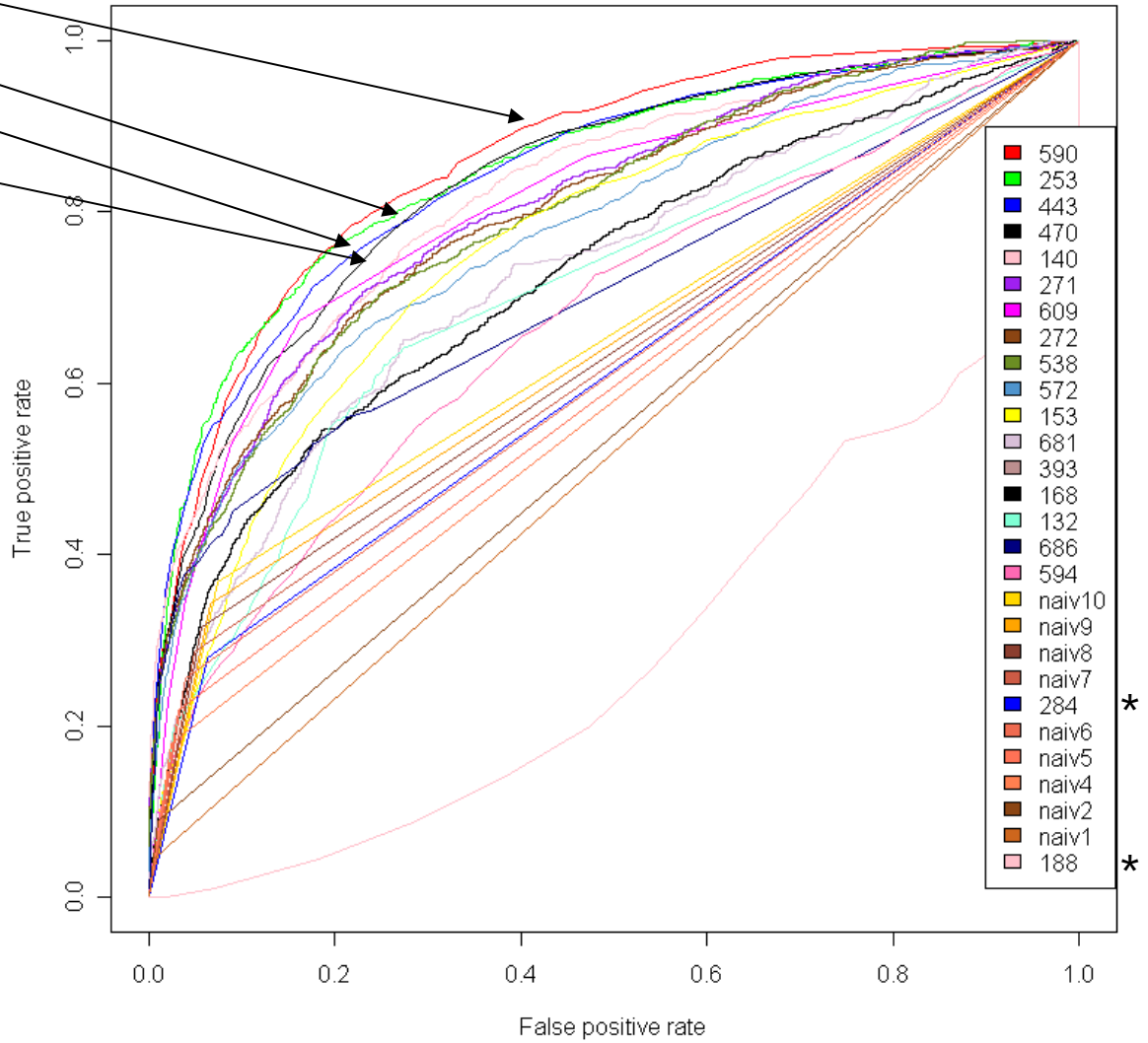
Testing the significance of the assessment

- We used a bootstrapping procedure to test the statistical significance of the assessment:
  - 80% of randomly chosen ***target structures*** were assessed repeatedly 1000 times to derive standard error of each of the binary scores.
  - 80% of randomly chosen ***target residues*** were assessed repeatedly 1000 times to derive standard error of each of the binary scores.

- Standard error of ROC curves were estimated according to Hanley J.A. & McNeil B.J. (1983), Radiology, 148,839-843

ROC curves including naïve predictors

590
253
443
470



Relationship between AUC and Wilcoxon test statistics [1]:

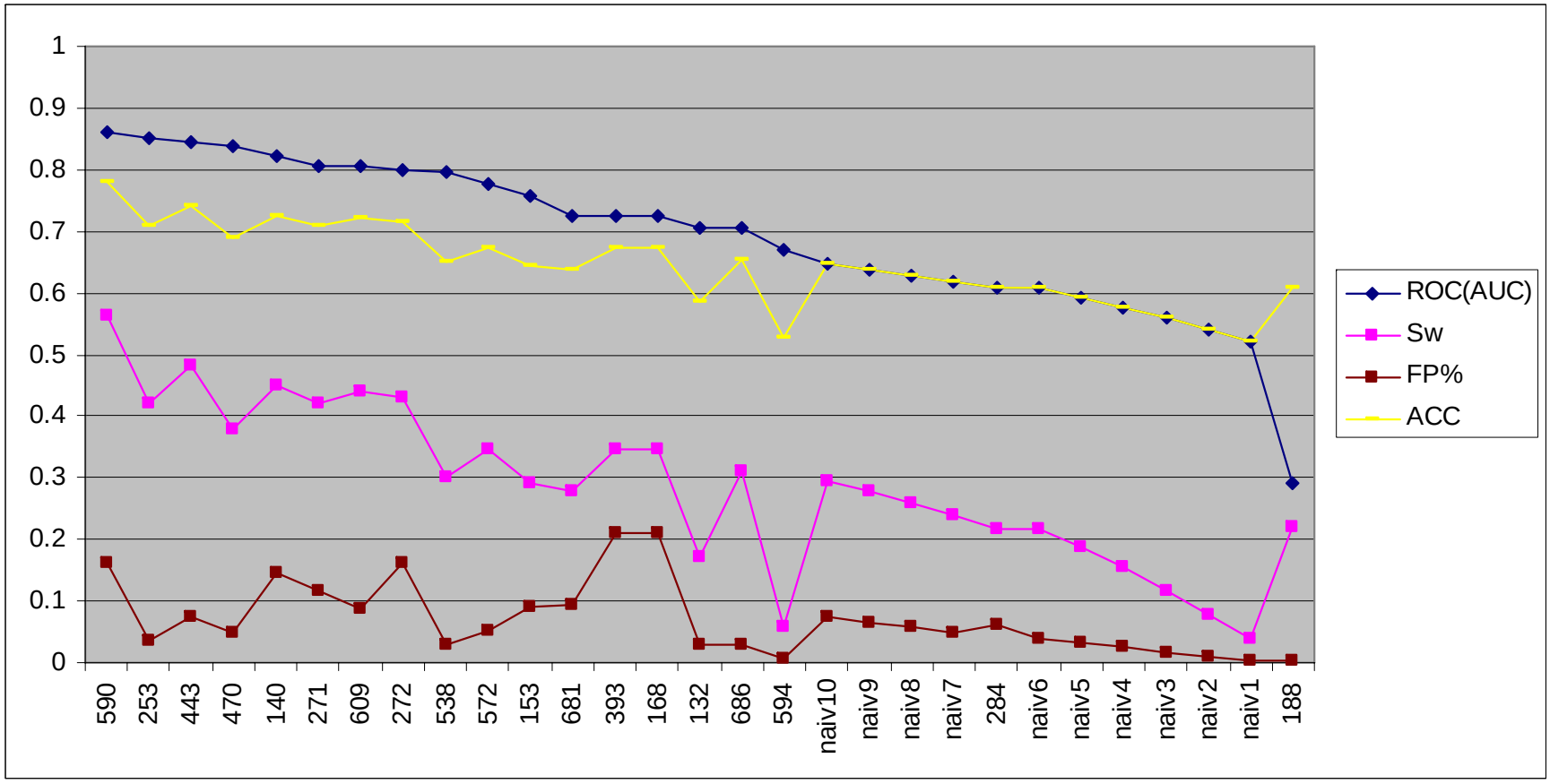
- Performance of a method is different from random

- Statistical significance between two methods.

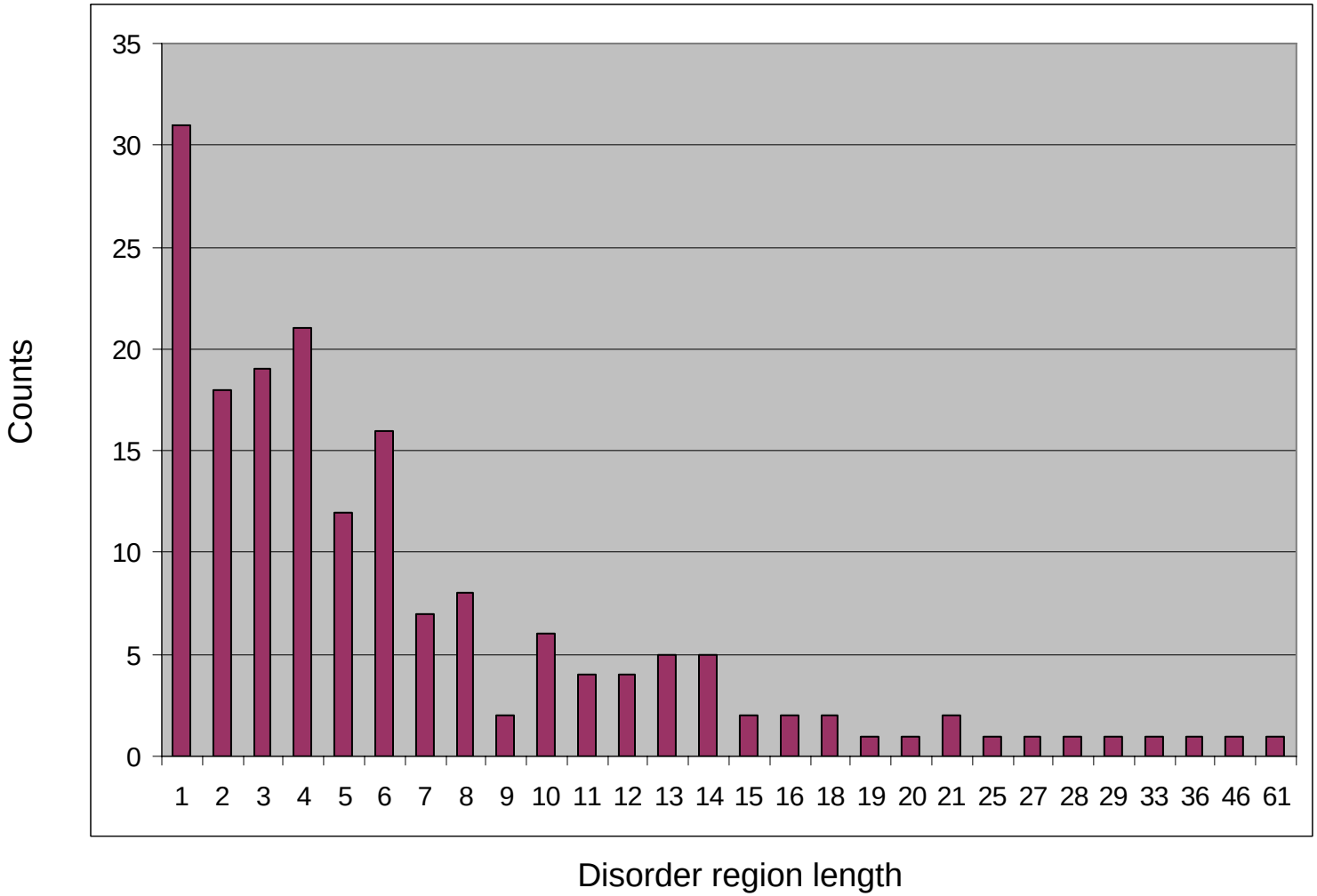
[1] Hanley J.A. & McNeil B.J. (1983), Radiology, 148,839-843

**** Not assessed well by ROC curves***

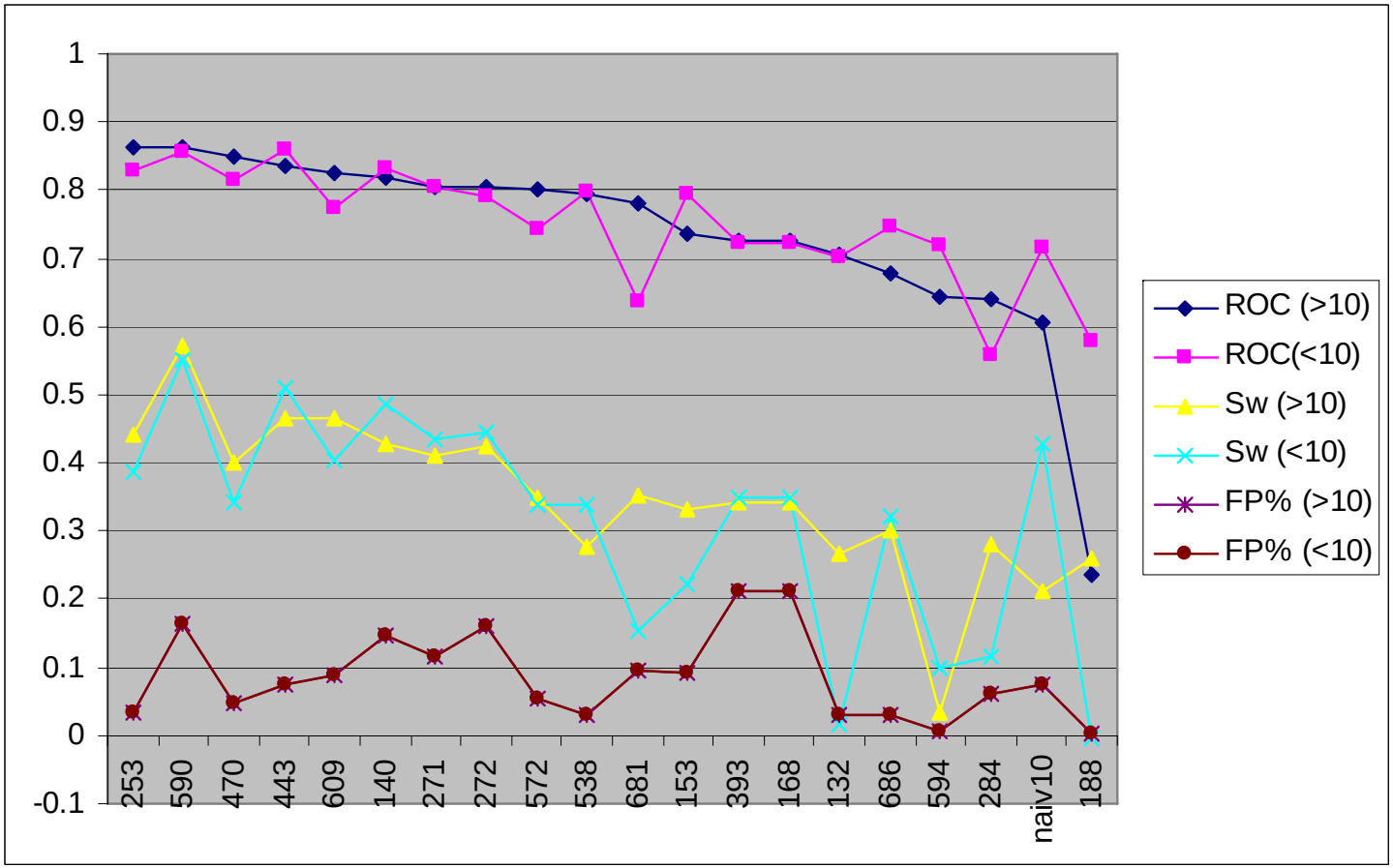
Disorder Prediction ranked by ROC-AUC



Distribution of the length of disordered regions in CASP7

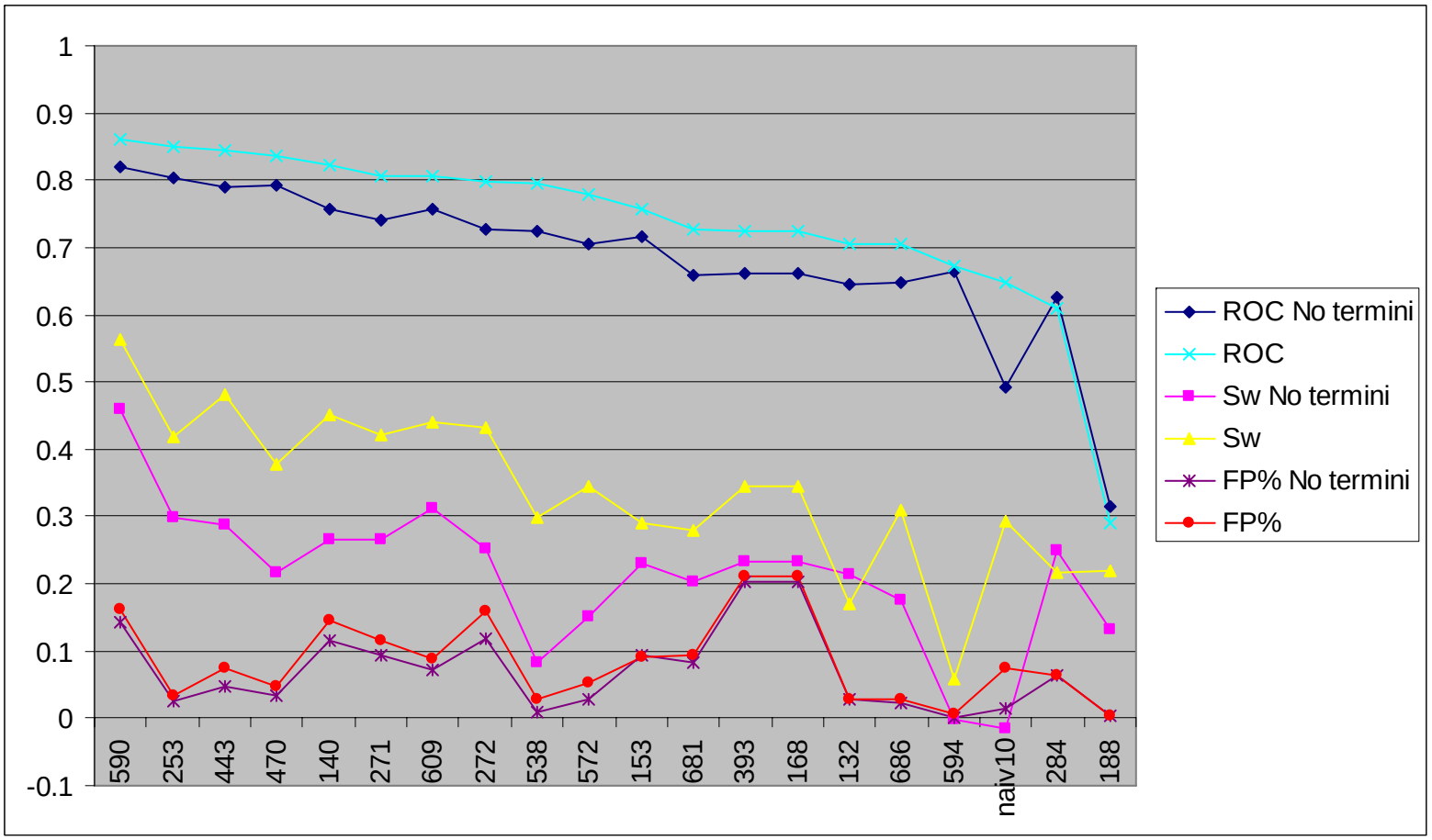


Short (<= 10) vs. Long (> 10) disordered regions



- Groups 253 and 470 are more accurate on long disordered regions
- Group 443 more accurate on short disordered regions
- Group 590 comparable accuracy on both long and short regions

Disorder Prediction ranked by ROC-AUC



- Internal disordered regions are more difficult to be detected than terminal disordered regions

Results

- In summary, groups 590, 253 followed by 443 & 470 give better performance than the other groups.
- 590 has highest sensitivity but a relatively high FP rate.
- At a lower FP rate groups 253 & 470 have comparable Sw and ACC scores to the rest of the top groups.
- Groups 253 and 470 more accurate on long disordered regions.
- Groups 443 more accurate on short disordered regions.
- Group 590 comparable accuracy on both long and short regions.

Confusion matrix

Gr.590		Prediction	
		Order	Disorder
Actual	Order	15365	2991
	Disorder	295	781

Gr.253		Prediction	
		Order	Disorder
Actual	Order	17990	637
	Disorder	590	490

Gr.443		Prediction	
		Order	Disorder
Actual	Order	17218	1409
	Disorder	478	602

Gr.470		Prediction	
		Order	Disorder
Actual	Order	17750	877
	Disorder	621	459

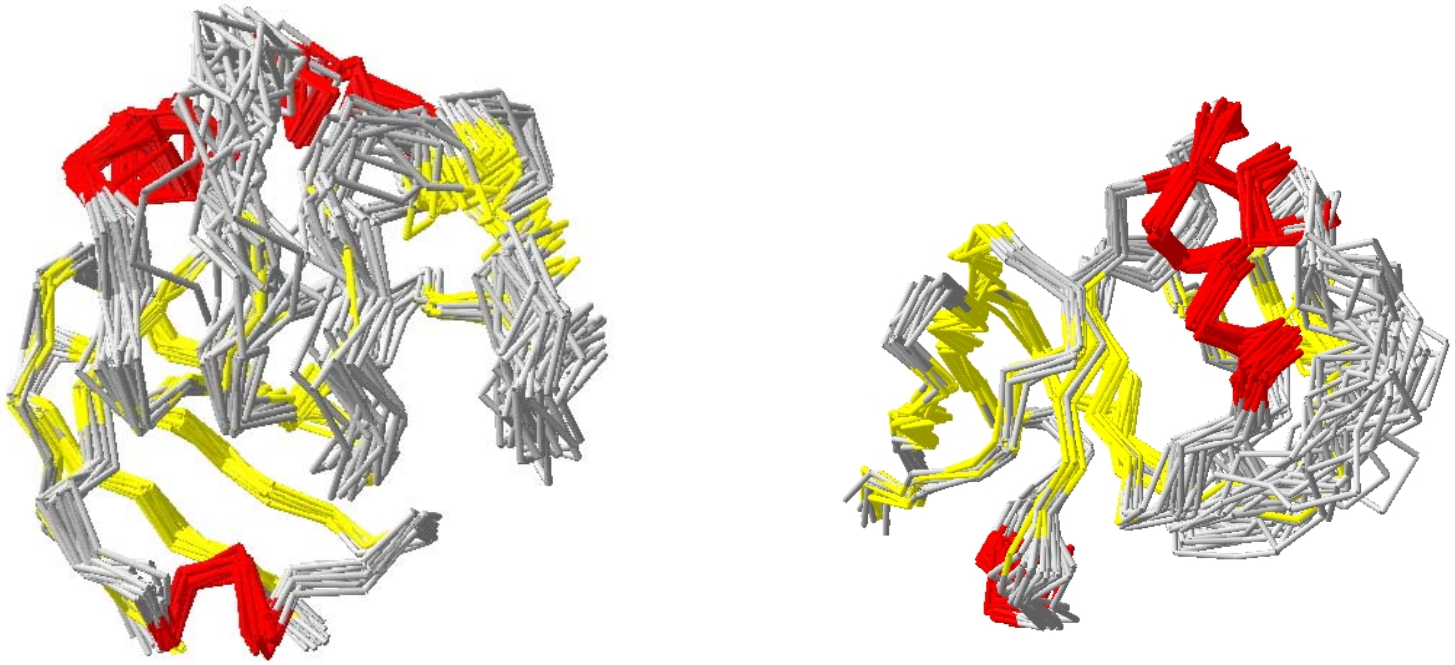
Gr.140		Prediction	
		Order	Disorder
Actual	Order	15910	2717
	Disorder	435	645

Gr.609		Prediction	
		Order	Disorder
Actual	Order	16492	1589
	Disorder	491	551

TN	FP
FN	TP

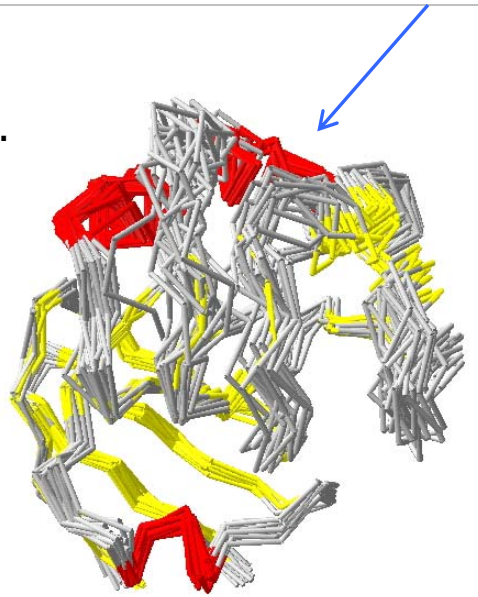
Examples

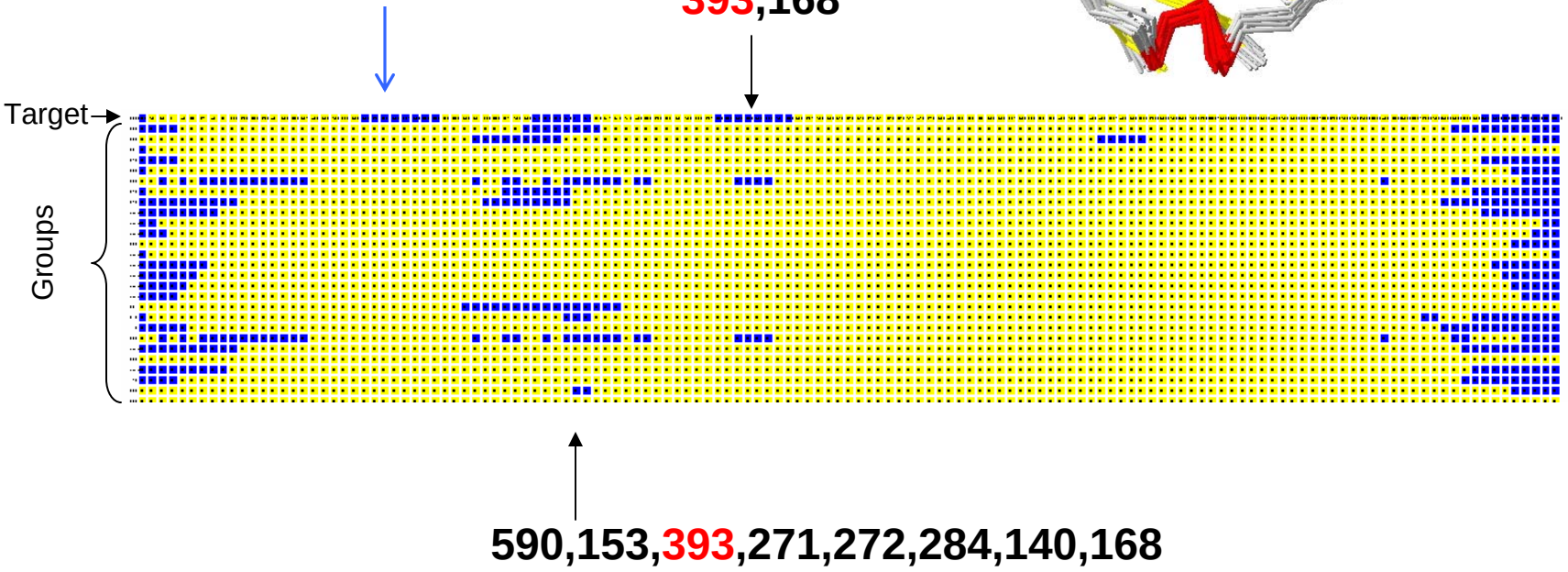
- NMR-structure (T0357, 2HFQ):



Examples: T0357

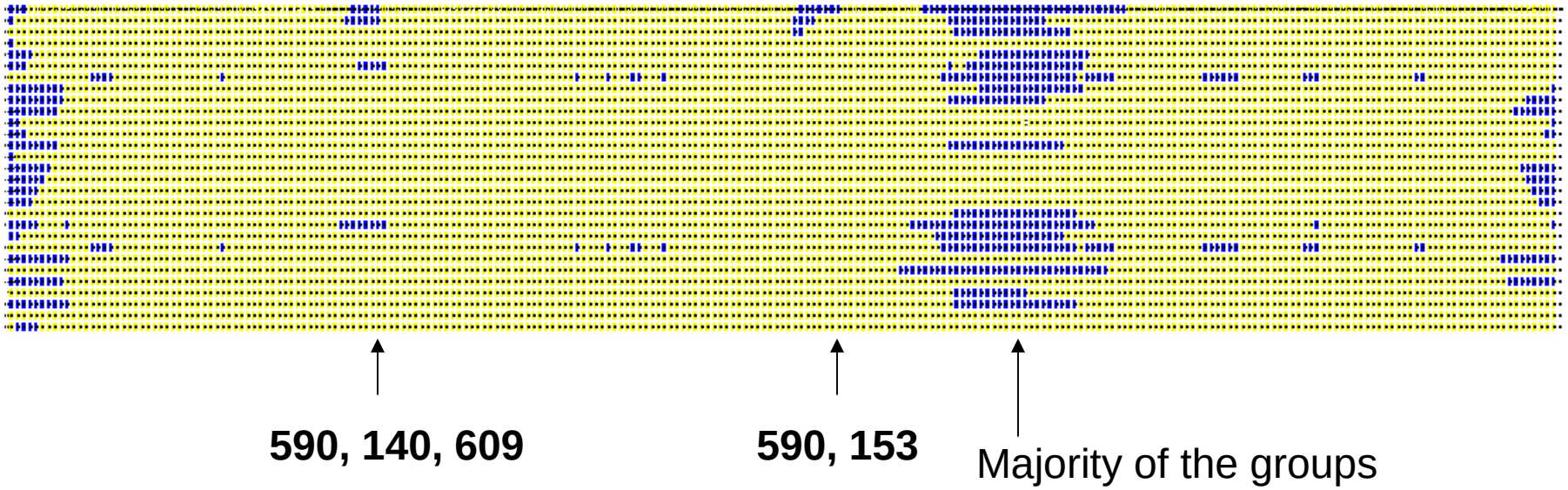
- Nobody identified the disorder segment 24-30.



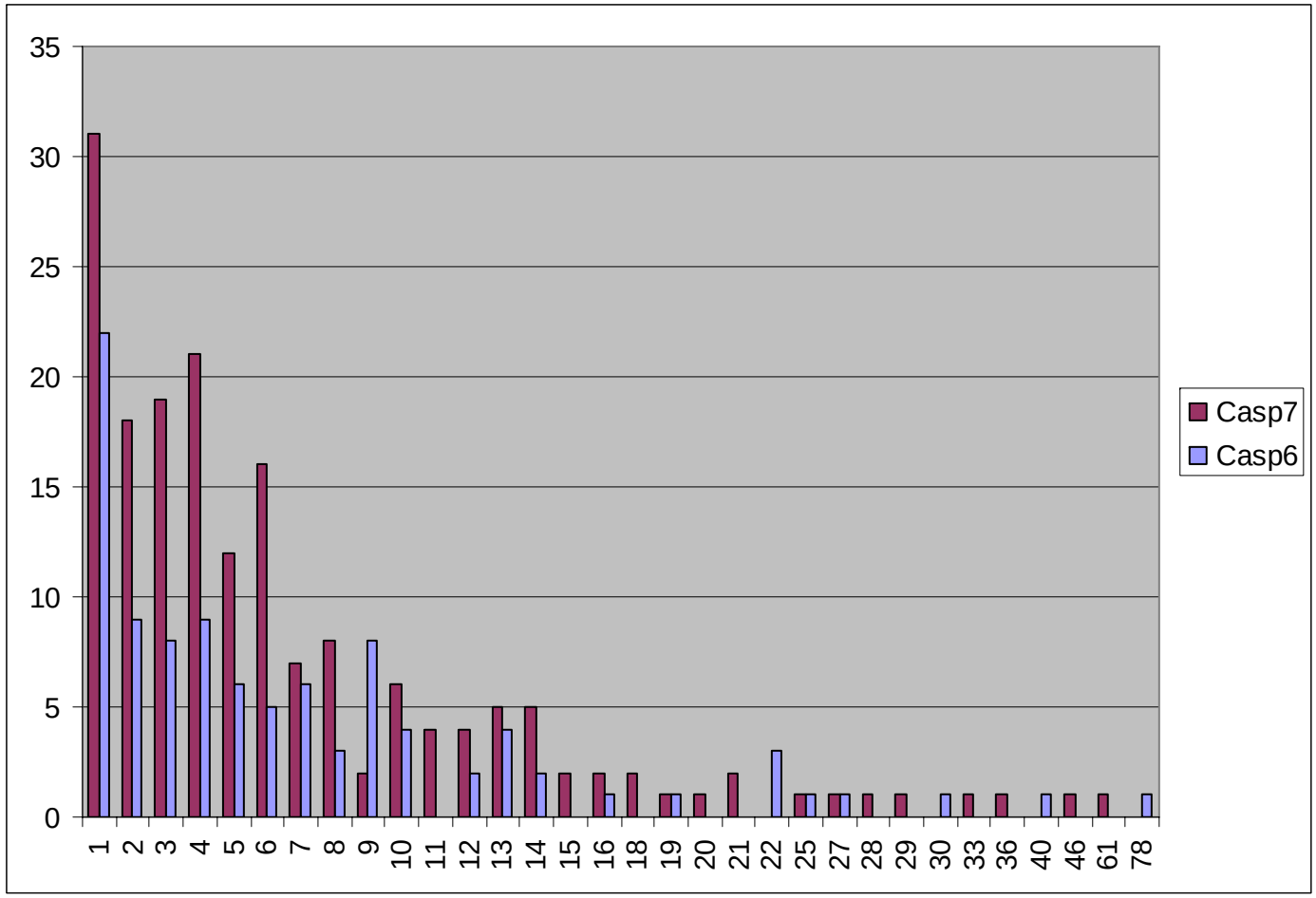


Examples identified correctly

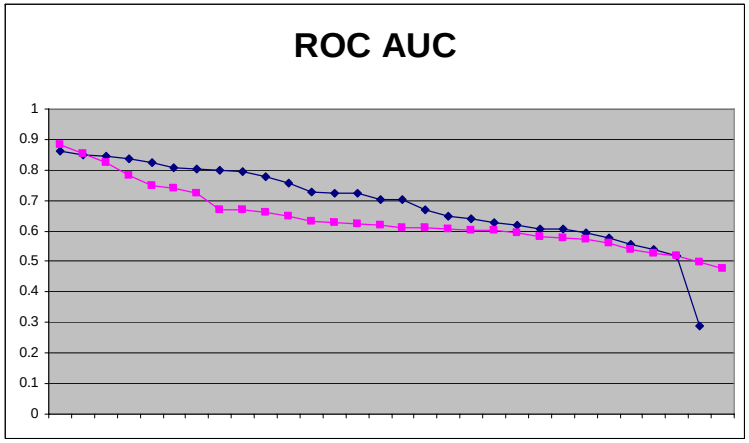
- X-Ray structure (T0293, 2H00):
  - 32, 7 & 5 residue long disordered regions
  - Correctly identified by 590

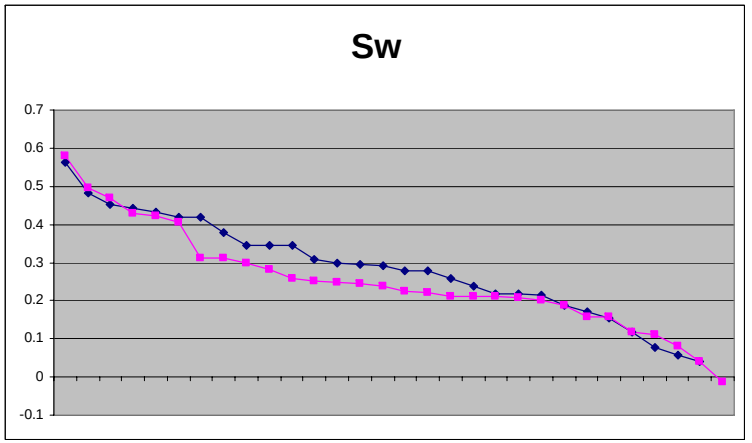


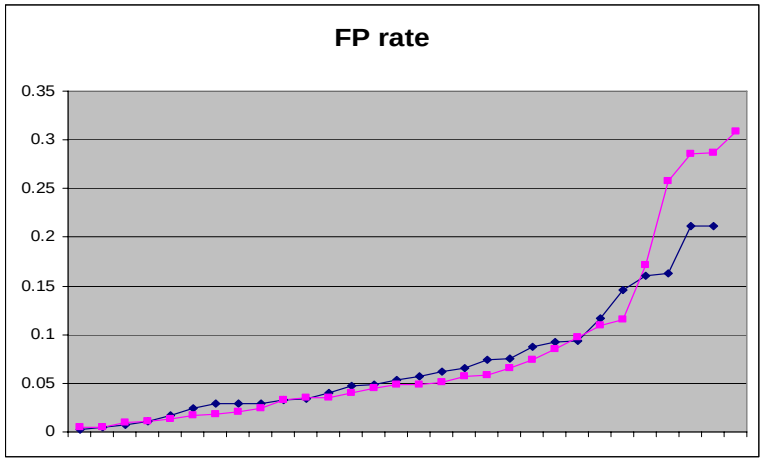
CASP6 vs. CASP7



CASP6 vs. CASP7







- Caveat: Comparison is based on ***same scores*** for a ***different data set (CASP6)***.
- If we assume the comparison as valid, no improvement over CASP6 is observed.

Methods

- Group 590

  ISTZORAN - *K. Peng, P. Radivojac, S. Vucetic, A.K. Dunker, Z. Obradovic* (SVM, two specialized predictors, for long and short disordered regions.

- Group 253

  CBRC-DR – *K. Shimizu, S. Hirose, N. Inoue, S. Kanai and T. Noguchi* (SVM, specialized predictors for long and short disordered regions.)

- Group 443

  fais - *T. Ishida, and K. Kinoshita*

- Group 470

  DISOPRED – *K. Bryson and D. Jones* (SVM trained on PSI-BLAST Profiles)

Acknowledgements

- Neil Clark
- Randy Read
- Michael Tress
- Alfonso Valencia
- Roland Dunbrack
- Yumi Jin
- Anna Tramontano
- Krzysztof Fidelis
- Andriy Kryshtafovych
- Torsten Schwede
- Florian Kiefer
- Jürgen Kopp
- James Battey
- John Moult
- Burkhard Rost

Discussion panel

- Applications: What are the main applications of disorder prediction? What are their requirements? How sensitive are they to over / under prediction?
- Performance: Are current disorder prediction methods performing sufficiently well for these practical applications? Do we deliver?
- Definition of "Disorder": Is there consensus about what we define as a "disordered region"? Based on which experimental evidence?
- Data: What experimental data is required for method development?
- Is the CASP7 data adequate for disorder prediction assessment?
- Human vs. servers: Does human intervention improve disorder prediction?
- Future?