CASP6
Disorder Prediction Assessment

Yumi Jin
Roland Dunbrack

Prediction Format

```
PFRMAT DR
TARGET T0205
MODEL  1
S D   0.679
E D   0.552
E D   0.510
I D   0.500
T O   0.254
D O   0.160
G O   0.097
V O   0.472
N O   0.366
…
```

ROC Curves

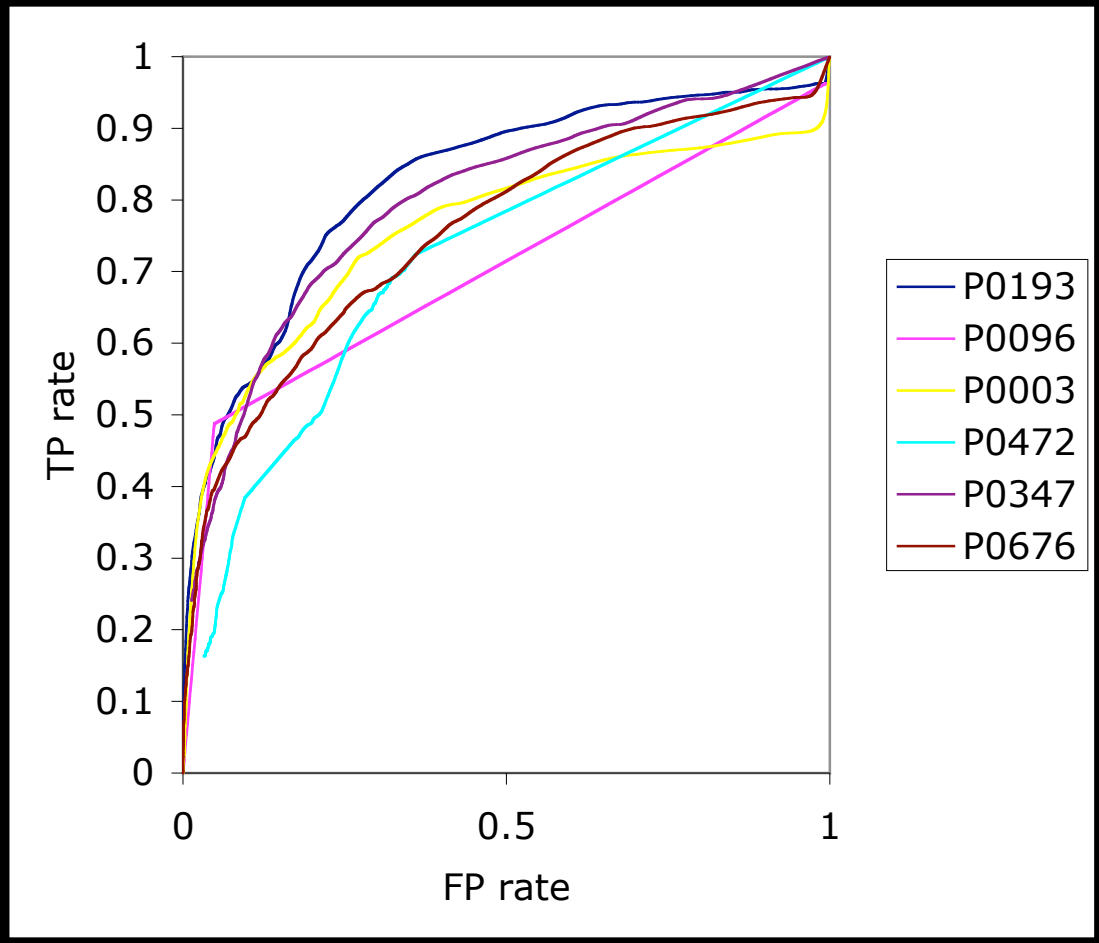
For each value of P in increments of 0.01:

$TP(P)$= Number of true positives in predictions with $p>P$
$FP(P)$= Number of false positives in predictions with $p>P$

$$r_{TP}(P) = \frac{TP(P)}{n_{disordered}}$$

$$r_{FP}(P) = \frac{FP(P)}{n_{ordered}}$$

ROC Curves



$$ROCScore = \int_0^1 r_{TP}\left(r_{FP}\right) dr_{FP}$$

ROC Scores

Group	N	ROC	ROC66	Scores used
193	69	0.775	0.823	
347	69	0.743	0.798	
003	69	0.733	0.758	
676	61	0.660	0.755	
472	64	0.648	0.717	
096	68	0.648	0.703	0.2, 0.7
633	67	0.645	0.681	
686	60	0.634	0.693	
461	68	0.626	0.661	0.2,0.4,0.6,0.8,1.0
675	62	0.622	0.636	0.0, 1.0
536	69	0.610	0.652	
673	62	0.591	0.593	0.0, 1.0
019	45	0.577	0.587	0.3, 0.7
667	69	0.572	0.632	
018	25	0.560	0.657	0.3, 0,7
674	62	0.553	0.571	0.0, 1.0
245	63	0.545	0.557	
545	67	0.528	0.538	
060	69	0.526	0.476	
679	58	0.489	0.484	all ordered

ROC66 = ROC without targets T0218, T0219, T0245

Group	Np	Nres	TP	FP	TN	FN	Sens	Spec	Prod	S
193	66	14490	764	2303	11119	304	0.715	0.828	0.593	7.43
096	65	14392	542	593	12731	526	0.507	0.955	0.485	6.32
003	66	14490	530	686	12736	538	0.496	0.949	0.471	6.08
347	66	14490	544	1137	12285	524	0.509	0.915	0.466	5.80
018	23	4814	83	44	4538	149	0.358	0.990	0.354	5.61
676	58	12592	385	569	11167	471	0.450	0.952	0.428	5.57
060	66	14490	425	474	12948	643	0.398	0.965	0.384	4.95
536	66	14490	367	234	13188	701	0.344	0.983	0.338	4.45
461	65	14311	445	1523	11733	610	0.422	0.885	0.373	4.19
686	57	12466	276	414	11198	578	0.323	0.964	0.312	4.13
675	59	12736	505	3388	8484	359	0.584	0.715	0.418	4.01
472	61	13319	407	1339	10937	636	0.390	0.891	0.348	3.72
633	64	14050	575	3731	9271	473	0.549	0.713	0.391	3.58
667	59	12736	282	1152	10720	582	0.326	0.903	0.295	3.36
019	44	9685	171	120	8863	531	0.244	0.987	0.240	3.21
673	59	12736	397	3052	8820	467	0.459	0.743	0.341	2.83
674	59	12736	154	243	11629	710	0.178	0.980	0.175	2.55
679	54	12120	139	60	11211	710	0.164	0.995	0.163	2.42
545	64	14079	415	4032	9024	608	0.406	0.691	0.280	1.34
245	60	13239	52	720	11654	813	0.060	0.942	0.057	0.76

Total #residues = 14490
Total number D's = 1068
Total number O's = 13422

Bad targets: 217, 218, 245
Sens=TP/(TP+FN) Spec=TN/(TN+FP)
S = (TP*Wt_tp + FP*Wt_fp + TN*Wt_fn + FN*Wt_fn)/Nres
where WT_tp=92.63 WT_fp=-7.37 WT_tn=7.37 WT_fn=-92.63

CASP5 vs CASP6

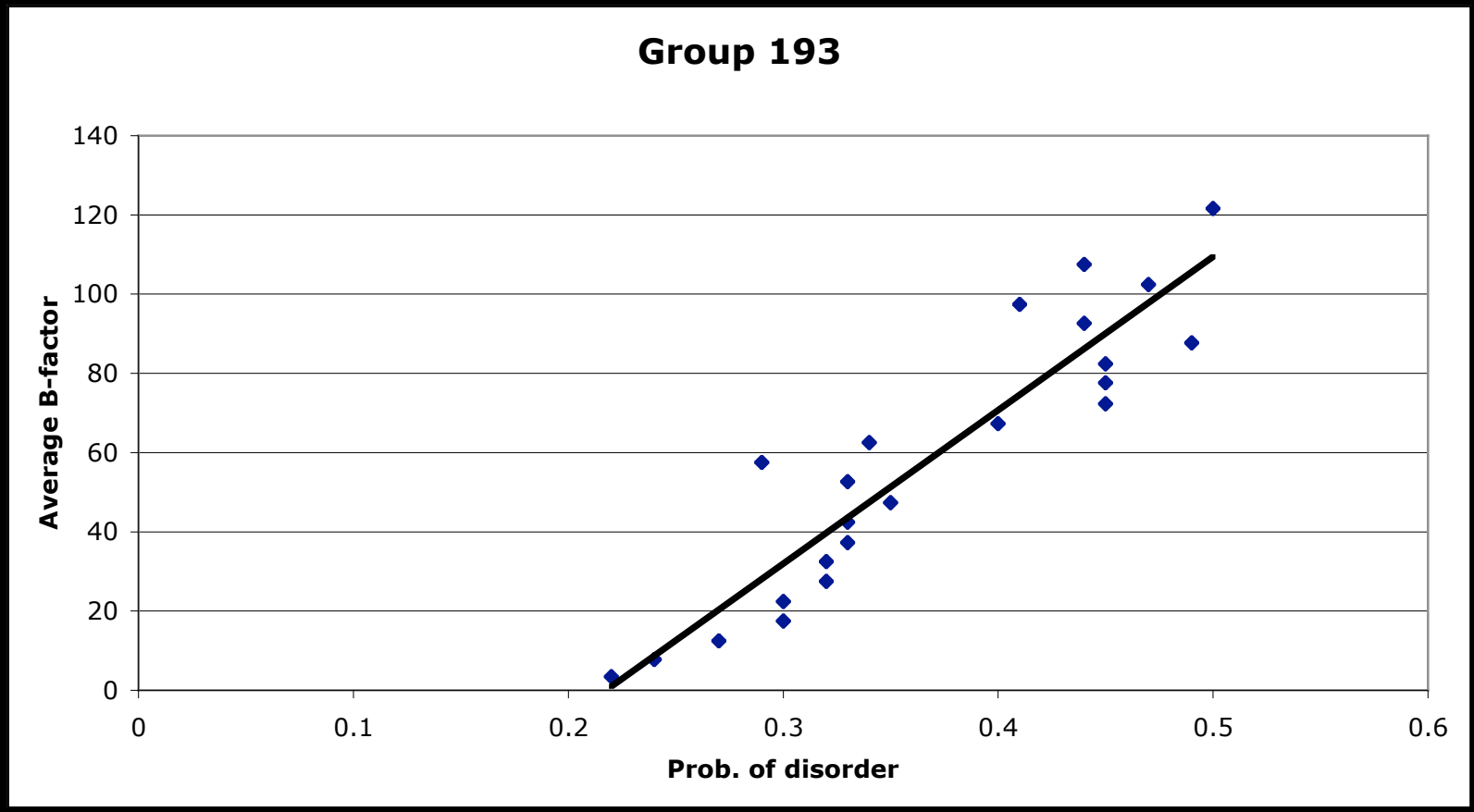
CASP5

Group	N	Spec.	Sens.	Prod.	Score
20	36	0.564	0.973	0.549	5.84
68	56	0.618	0.916	0.566	5.62
355	56	0.694	0.770	0.535	4.57
454	56	0.420	0.901	0.379	3.90

CASP6

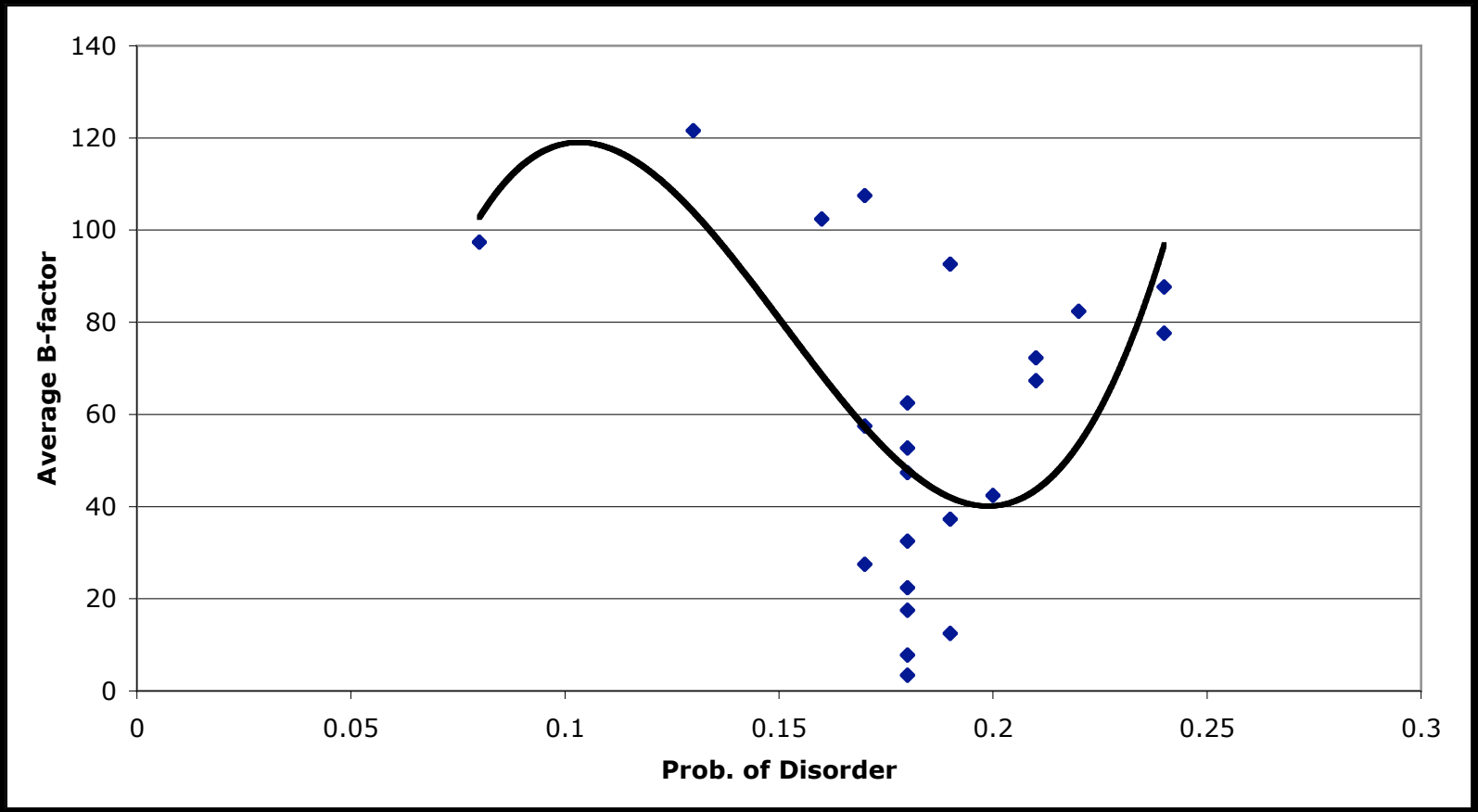
Group	N	Spec.	Sens.	Prod.	Score
193	66	0.715	0.828	0.593	6.57
96	65	0.507	0.955	0.485	5.07
3	66	0.496	0.949	0.471	4.84
347	66	0.509	0.915	0.466	4.66
676	58	0.450	0.952	0.428	4.31
18	23	0.358	0.990	0.354	4.20
60	66	0.398	0.965	0.384	3.65
675	59	0.584	0.715	0.418	3.43
461	65	0.422	0.885	0.373	3.11
536	66	0.344	0.983	0.338	3.09
633	64	0.549	0.713	0.391	3.00
686	57	0.323	0.964	0.312	2.81
472	61	0.390	0.891	0.348	2.62
667	59	0.326	0.903	0.295	2.20
673	59	0.459	0.743	0.341	2.15
19	44	0.244	0.987	0.240	1.81
674	59	0.178	0.980	0.175	1.15
679	55	0.163	0.995	0.162	1.00
545	64	0.406	0.691	0.280	0.80
245	60	0.060	0.942	0.057	-0.55

Correlation of P with the B-factors



r=0.92

Correlation of P with the B-factors



r=0.25 for a straight line (not shown)

Correlations

CASP6

Group	N	r
193	61	0.92
3	61	0.89
472	56	0.85
347	61	0.78
633	59	0.74
461	60	0.63
676	53	0.62
96	60	0.61
245	55	0.49
19	43	0.34
18	18	0.30
686	52	0.25
675	54	0.23
667	54	0.06
679	51	0.00
673	54	-0.09
674	54	-0.11
545	59	-0.12
536	61	-0.15
60	61	-0.64

CASP5

Group	N	r
68	50	0.89
355	50	0.76
454	50	0.64
20	31	-0.09

How?

193 ISTZORAN (Zoran Obradovic, Temple University)
neural network

096 CaspIta (Tosatto et al., Univ. of Padova)
support vector machines

003 Jones UCL (David Jones, University College London)
support vector machines

347 DRIP PRED (server from Bob MacCallum, Stockholm)
Kohonen self-organizing maps

472 Softberry (good at B-factor correlation)
A combination of “a neural network, linear discriminant function, and a smoothing procedure”

Conclusion

Group 193 is best on all measures, on both no-density segments and B-factors, and is significantly better than next 3 groups, 096, 003, 347 on no-density segments, who are about the same as each other. Groups 3, 347, and 472 are good at B-factors.