

Supplementary Table 1. The assessment of disorder prediction methods using standard benchmarking scores, which are dependent on P-value thresholds.

Method	TP	FP	TN	FN	FPR	Sens	Spec	MCC	Acc	Prod	S _w	D cut-off
DISOclust+human	811	2478	15878	262	0.135	0.756	0.865	0.378	0.810	0.654	0.621	0.508
ISTZORAN	778	2991	15365	295	0.163	0.725	0.837	0.325	0.781	0.607	0.562	0.500
<u>DISOclust+DISOPRED</u>	713	2119	16237	360	0.115	0.664	0.885	0.355	0.775	0.588	0.549	0.549
<u>DISOclust</u>	881	5607	12749	192	0.305	0.821	0.695	0.250	0.758	0.570	0.516	0.842
Fais	599	1409	16947	474	0.077	0.558	0.923	0.361	0.741	0.515	0.481	0.500
<u>DISpro</u>	643	2689	15667	430	0.146	0.599	0.854	0.274	0.726	0.511	0.453	0.500
CBRCDR	489	637	17719	584	0.035	0.456	0.965	0.412	0.711	0.440	0.421	0.500
<u>DISOPRED</u>	458	877	17479	615	0.048	0.427	0.952	0.342	0.690	0.406	0.379	0.500

DISOclust+DISOPRED, a consensus method which takes the average of the P-values from DISOclust and DISOPRED; DISOclust+human, a consensus method which takes the average of the P-values from DISOclust, ISTZORAN, CBRCDR and Fais. Server/automated methods are underlined; all other methods were registered as human expert groups at CASP7. TP, number of true positives; FP number of false positives; TN, number of true negatives; FN, number of false negatives; FPR, false positive rate; Sens, sensitivity; Spec, specificity; MCC, Matthews Correlation Coefficient (Matthews, 1975), Acc, accuracy (the mean of sensitivity and specificity scores)(Bordoli, et al., 2007; Peng, et al., 2006), Prod, product of sensitivity and specificity (Jin and Dunbrack, 2005); S_w, weighted score (Jin and Dunbrack, 2005) where W_{dis}=94.477 and W_{ord}=5.523; D cut-off, P-value cut-off used for assigning residues as disordered. Methods were benchmarked using a common subset of 95 CASP7 targets, with a total of 19429 residues (1073 disordered and 18356 ordered). The data are sorted by the S_w score.

Supplementary Table 2. The comparison of methods according to maximal scores and corresponding P-value cut-offs.

Method	pMCC _{max}	MCC _{max}	pAcc _{max}	Acc _{max}	pProd _{max}	Prod _{max}	pS _{wmax}	S _{wmax}	pMean _{max}
DISOclust+human	0.606	0.441	0.475	0.813	0.475	0.661	0.475	0.626	0.508
DISOclust+DISOPRED	0.693	0.393	0.501	0.797	0.501	0.636	0.501	0.595	0.549
ISTZORAN	0.750	0.370	0.440	0.785	0.440	0.616	0.440	0.570	0.518
CBRCDR	0.502	0.414	0.240	0.782	0.240	0.611	0.240	0.564	0.306
Fais	0.680	0.416	0.420	0.771	0.400	0.593	0.420	0.542	0.480
DISOPRED	0.520	0.371	0.120	0.769	0.120	0.589	0.120	0.538	0.220
DISOclust	0.853	0.252	0.838	0.761	0.838	0.573	0.838	0.521	0.842
DISpro	0.750	0.420	0.271	0.744	0.271	0.553	0.271	0.488	0.391

pMCC_{max}, P-value cut-off for maximising the Mathews Correlation Coefficient; MCC_{max}, maximal Matthews Correlation Coefficient; pAcc_{max}, P-value cut-off for maximizing the accuracy score; ACC_{max}, maximal accuracy score; pProd_{max}, P-value cut-off for maximising the product of sensitivity and specificity; Prod_{max}, maximal product of sensitivity and specificity; pS_{wmax}, P-value cut-off for maximising the weighted score; S_{wmax}, maximal weighted score; pMean_{max}, mean P-value for maximising scores. Methods were benchmarked using a common subset of 95 CASP7 targets, with a total of 19429 residues (1073 disordered and 18356 ordered). The data are sorted by the S_{wmax} column.

Supplementary Table 3. The assessment of disorder prediction methods using the DisProt data set and standard benchmarking scores, which are dependent on P-value thresholds.

Method	TP	FP	TN	FN	FPR	Sens	Spec	MCC	Acc	Prod	S _w	D cut-off
DISOclust+DISOPRED	3429	11771	38369	1037	0.235	0.768	0.765	0.326	0.767	0.588	0.510	0.549
DISOPRED	2470	3664	46476	1996	0.073	0.553	0.927	0.417	0.740	0.513	0.460	0.500
DISOclust	4166	30291	19849	300	0.604	0.933	0.396	0.187	0.664	0.369	0.315	0.842

DISOclust+DISOPRED, a consensus method which takes the average of the P-values from DISOclust and DISOPRED. TP, number of true positives; FP number of false positives; TN, number of true negatives; FN, number of false negatives; FPR, false positive rate; Sens, sensitivity; Spec, specificity; MCC, Matthews Correlation Coefficient (Matthews, 1975), Acc, accuracy (the mean of sensitivity and specificity scores)(Bordoli, et al., 2007; Peng, et al., 2006), Prod, product of sensitivity and specificity (Jin and Dunbrack, 2005); S_w, weighted score (Jin and Dunbrack, 2005), where W_{dis}=91.821 and W_{ord}=8.179; D cut-off; P-value cut-off used for assigning residues as disordered (values for DISOclust and DISOclust+DISOPRED were derived using the CASP7 data set, see Supplementary Table 2). Methods were benchmarked using a set of 199 proteins obtained from the DisProt resource with <25% sequence identity to other structures within the PDB. A total of 54606 residues were evaluated (4466 disordered and 50140 ordered). The LOMETS server was used to obtain models for DISOclust. Any models built using homologous templates ($\geq 25\%$ sequence identity) were excluded. The data are sorted by the S_w score.

Supplementary Table 4. The change in DISOclust performance using different numbers of fold recognition models obtained from the LOMETS meta-server (Wu and Zhang, 2007).

Max. number of models per target	Method	AUC	SE	AUC _{0-0.1}
81	DISOclust+DISOPRED	0.8557	0.0036	0.0480
27	DISOclust+DISOPRED	0.8526	0.0036	0.0479
9	DISOclust+DISOPRED	0.8442	0.0037	0.0467
-	DISOPRED	0.8223	0.0039	0.0462
81	DISOclust	0.7688	0.0042	0.0243
27	DISOclust	0.7548	0.0043	0.0182
9	DISOclust	0.7191	0.0045	0.0143

AUC, area under ROC curve; SE, standard error of AUC (Hanley and McNeil, 1982); AUC_{0-0.1}, the area under the curve for false positive rates between 0 and 10%. Methods were benchmarked using a set of 199 proteins obtained from the DisProt resource with <25% sequence identity to other structures within the PDB. A total of 54606 residues were evaluated (4466 disordered and 50140 ordered). The LOMETS server was used to obtain models for DISOclust. Any models built using homologous templates ($\geq 25\%$ sequence identity) were excluded. In the cases where 81 models were used, 9 different fold recognition methods were used to build models - HHsearch (Soding, 2005), SPARKS2 & SP3(Zhou and Zhou, 2005), FUGUE (Shi, et al., 2001), PROSPECT2 (Zhang and Skolnick, 2004), SAM-T02 (Karplus, et al., 2003), PAINT, PPA-I & PPA-II (Wu and Zhang, 2007) - and a maximum of 9 models were taken from each method (the top hit to the target was excluded); in cases where 27 models were used, 3 different methods were used to build models - HHsearch, SPARKS2 & SP3; in cases where 9 models were used, all models were taken from 1 method - HHsearch. The DISOPRED results are shown for comparison.

Supplementary References

- Bordoli, L., Kiefer, F. and Schwede, T. (2007) Assessment of disorder predictions in CASP7, *Proteins*, **69 Suppl 8**, 129-136.
- Hanley, J.A. and McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, **143**, 29-36.
- Jin, Y. and Dunbrack, R.L., Jr. (2005) Assessment of disorder predictions in CASP6, *Proteins*, **61 Suppl 7**, 167-175.
- Karplus, K., Karchin, R., Draper, J., Casper, J., Mandel-Gutfreund, Y., Diekhans, M. and Hughey, R. (2003) Combining local-structure, fold-recognition, and new fold methods for protein structure prediction, *Proteins*, **53 Suppl 6**, 491-496.
- Matthews, B.W. (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme, *Biochim Biophys Acta*, **405**, 442-451.

- Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K. and Obradovic, Z. (2006) Length-dependent prediction of protein intrinsic disorder, *BMC Bioinformatics*, **7**, 208.
- Shi, J., Blundell, T.L. and Mizuguchi, K. (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties, *J Mol Biol*, **310**, 243-257.
- Soding, J. (2005) Protein homology detection by HMM-HMM comparison, *Bioinformatics*, **21**, 951-960.
- Wu, S. and Zhang, Y. (2007) LOMETS: a local meta-threading-server for protein structure prediction, *Nucleic Acids Res*, **35**, 3375-3382.
- Zhang, Y. and Skolnick, J. (2004) Scoring function for automated assessment of protein structure template quality, *Proteins*, **57**, 702-710.
- Zhou, H. and Zhou, Y. (2005) SPARKS 2 and SP3 servers in CASP6, *Proteins*, **61 Suppl 7**, 152-156.