

CAID

May 24, 2019

Contents

1 CAID	2
1.1 Dataset	2
1.2 Evaluation metrics	2
1.2.1 Balanced accuracy	3
1.2.2 F1-score	4
1.2.3 MCC	4
1.2.4 Per target accuracy	5
1.2.5 Target correlation matrix	6
1.2.6 ROC curve	7
1.2.7 PR curve	7
1.2.8 pROC/pPR scatter plot	8
1.2.9 Average overall ranking	8
1.2.10 Accuracy correlation between datasets	9
1.3 Consensus	10
1.3.1 Confusion matrix per threshold	10
1.3.2 Accuracy per threshold	11
1.3.3 Percentage of correct/incorrect classifications	12
1.3.4 clustermap of binary predictions correlation	13
1.4 Fully disordered targets	14
1.4.1 Correctly and incorrectly classified full IDPs	14
1.4.2 Full IDPs ROC	15

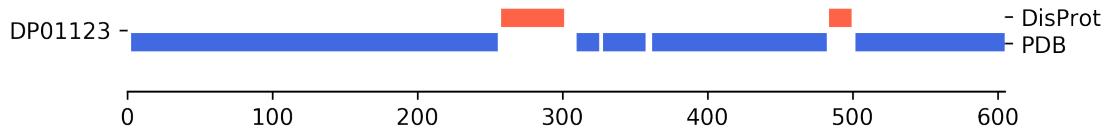
Chapter 1

CAID

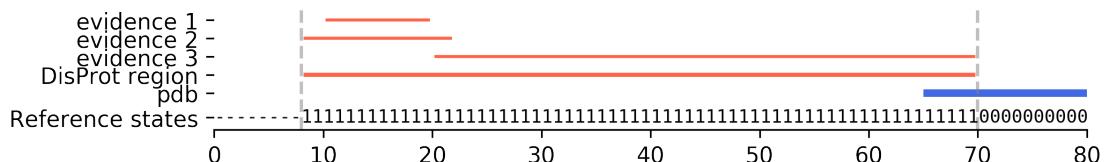
1.1 Dataset

Critical Assessment of Intrinsic Disorder (CAID) is a continuous experiment where prediction methods for intrinsic disorder (ID) are blind tested on unpublished DisProt data.

Current analysis is performed on the **new-disprot-all** dataset with **pdb** negative definition. This means that DisProt defines order for *all its new entries* and *pdb defines order*.



DisProt entries can have annotation covering the same sequence space. In these cases different evidences are merged in a unique continuous region. DisProt merged regions define positive cases (labeled as 1) in reference states. Remaining states are considered undefined.



1.2 Evaluation metrics

Metrics evaluating prediction scores are calculated applying **Default** thresholds to prediction scores. Table is sorted by descending value of BAc column

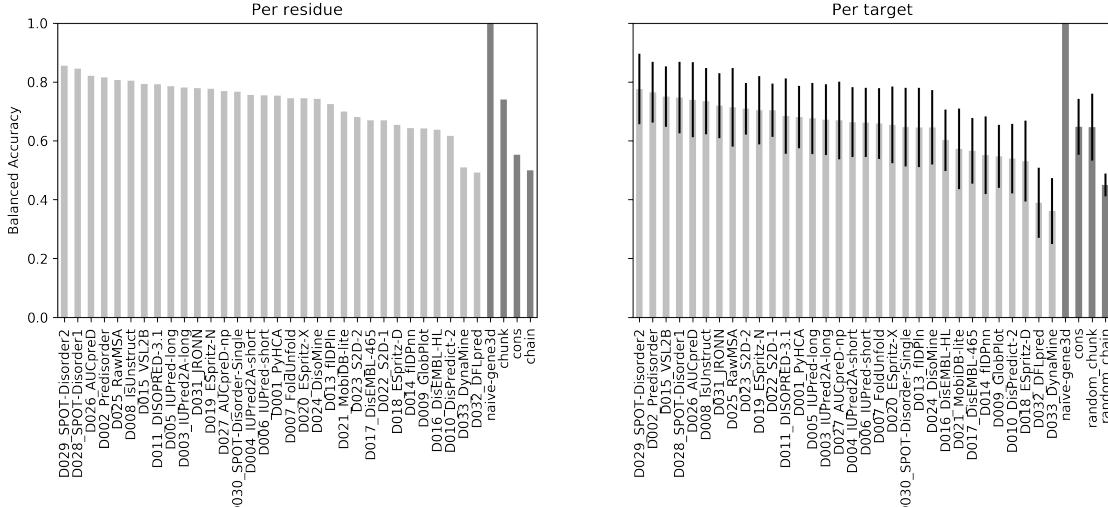
	FN	FP	TN	TP	BAc	F1s	MCC	Pre	Rec	Rec_n	AUC_PRC	AUC_ROC	npred	nref
naive-gene3d	0	0	110553	54604	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	646	646
D029_SPOT-Disorder2	11513	4938	95177	36238	0.855	0.815	0.740	0.880	0.759	0.951	0.895	0.931	610	646
D028_SPOT-Disorder1	13693	6426	104055	40600	0.845	0.801	0.718	0.863	0.748	0.942	0.872	0.927	644	646
D026_AUCpreD	17070	4934	103921	37485	0.821	0.773	0.690	0.884	0.687	0.955	0.878	0.919	644	646
D002_Predisorder	10526	19163	89659	43971	0.815	0.748	0.611	0.696	0.807	0.824	0.845	0.891	642	646
D025_RawMSA	17363	7658	102895	37241	0.806	0.749	0.648	0.829	0.682	0.931	0.841	0.902	646	646
D008_IsUnstruct	13771	15471	95082	40833	0.804	0.736	0.603	0.725	0.748	0.860	0.815	0.879	646	646
D015_VSL2B	10059	25206	85275	44234	0.793	0.715	0.558	0.637	0.815	0.772	0.816	0.874	644	646
D011_DISOPRED-3.1	19395	6698	103855	35209	0.792	0.730	0.632	0.840	0.645	0.939	0.831	0.885	646	646
D005_JUPred-long	19362	8312	102169	35198	0.785	0.718	0.608	0.809	0.645	0.925	0.816	0.871	645	646
D003_JUPred2A-long	19974	7945	102608	34630	0.781	0.713	0.605	0.813	0.634	0.928	0.813	0.869	646	646
D031_JRONN	14112	20200	90281	40448	0.779	0.702	0.545	0.667	0.741	0.817	0.789	0.852	645	646
D019_ESpritz-N	17054	14761	95720	37506	0.777	0.702	0.560	0.718	0.687	0.866	0.787	0.847	645	646
D027_AUCpreD-np	23312	3855	106698	31292	0.769	0.697	0.619	0.890	0.573	0.965	0.846	0.894	646	646
D030_SPOT-Disorder-Single	24220	2498	108055	30384	0.767	0.695	0.629	0.924	0.556	0.977	0.864	0.907	646	646
D004_JUPred2A-short	23880	5614	104939	30724	0.756	0.676	0.581	0.846	0.563	0.949	0.797	0.868	646	646
D006_JUPred-short	23891	5850	104631	30669	0.755	0.673	0.577	0.840	0.562	0.947	0.795	0.866	645	646
D001_PyHCA	18648	16809	93744	35956	0.753	0.670	0.511	0.681	0.658	0.848	0.737	0.828	646	646
D007_FoldUnfold	17940	18853	88908	35715	0.745	0.660	0.489	0.655	0.666	0.825	NaN	NaN	621	646
D020_ESpritz-X	25551	4825	105656	29009	0.744	0.656	0.569	0.857	0.532	0.956	0.802	0.871	645	646
D024_DisoMine	23260	9961	100592	31344	0.742	0.654	0.526	0.759	0.574	0.910	0.737	0.864	646	646
chunk	18991	18991	91561	35612	0.740	0.652	0.480	0.652	0.652	0.828	NaN	NaN	100	100
D013_fIDPln	27013	5937	104616	27505	0.725	0.625	0.528	0.822	0.505	0.946	0.776	0.869	645	646
D021_MobiDB-lite	32125	1517	108964	22435	0.699	0.572	0.531	0.937	0.411	0.986	0.805	0.852	645	646
D023_S2D-2	7710	54763	55718	46583	0.681	0.599	0.350	0.460	0.858	0.504	0.568	0.768	644	646
D022_S2D-1	7237	58208	52273	47056	0.670	0.590	0.333	0.447	0.867	0.473	0.731	0.812	644	646
D017_DisEMBL-465	33266	5291	105190	21027	0.670	0.522	0.435	0.799	0.387	0.952	0.689	0.789	644	646
D018_ESpritz-D	35346	4870	105611	19214	0.654	0.489	0.411	0.798	0.352	0.956	0.722	0.861	645	646
D014_fIDPnn	38167	1460	109093	16351	0.643	0.452	0.435	0.918	0.300	0.987	0.806	0.883	645	646
D009_GlobPlot	33642	11050	99431	20918	0.642	0.483	0.337	0.654	0.383	0.900	0.594	0.714	645	646
D016_DisEMBL-HL	25527	28263	82218	28766	0.637	0.517	0.271	0.504	0.530	0.744	0.560	0.707	644	646
D010_DisPredict-2	31912	20318	90235	22692	0.616	0.465	0.248	0.528	0.416	0.816	0.519	0.673	646	646
cons	4676	89282	21271	49928	0.553	0.515	0.138	0.359	0.914	0.192	0.477	0.666	646	646
D033_DynaMine	53609	25	110456	951	0.509	0.034	0.106	0.974	0.017	1.000	0.759	0.837	645	646
chain	36548	36548	74004	18055	0.500	0.331	0.000	0.331	0.331	0.669	NaN	NaN	100	100
D032_DFLpred	49496	12145	98408	5108	0.492	0.142	-0.025	0.296	0.094	0.890	0.290	0.379	646	646

Where table column names mean:

label	meaning
BAc	balanced accuracy
F1s	F1-score
MCC	Matthew's Correlation Coefficient
Pre	Precision>Selectivity
Rec	Recall/Sensitivity
Rec_n	Specificity
AUC_ROC	Area under the ROC curve
AUC_PRC	Area under the PR curve
npred	number of predicted targets
nref	number of targets in reference

1.2.1 Balanced accuracy

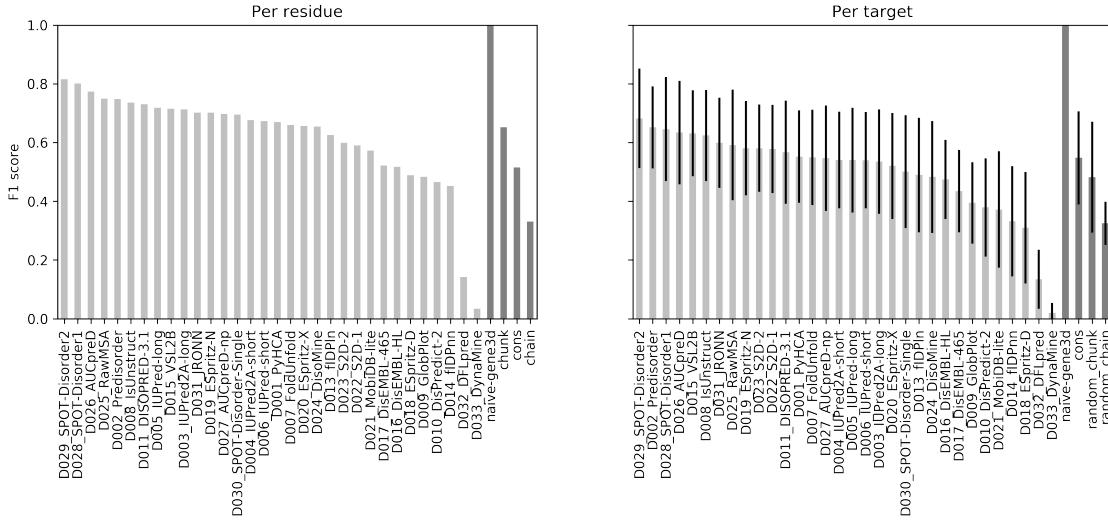
Comparison of predictors and baselines performance by balanced accuracy.



Overall (left panel) and average per-target (right panel) balanced accuracy. Light gray bars represent published prediction methods; dark gray bars represent baseline prediction methods. On the right panel standard deviation is shown as an error bar.

1.2.2 F1-score

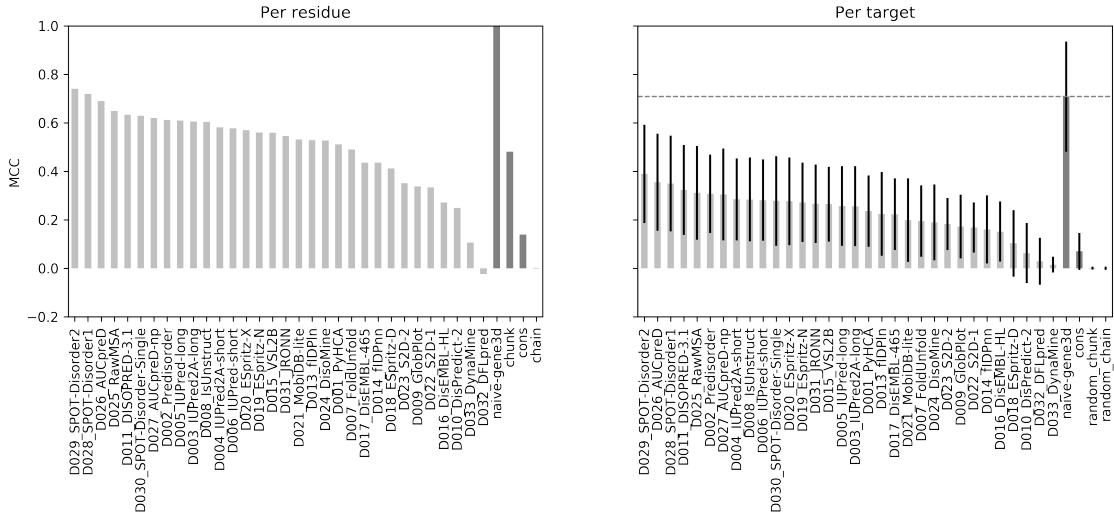
Comparison of predictors and baselines performance by F1-score.



Overall (left panel) and average per-target (right panel) F1-score. Light gray bars represent published prediction methods; dark gray bars represent baseline prediction methods. On the right panel standard deviation is shown as an error bar.

1.2.3 MCC

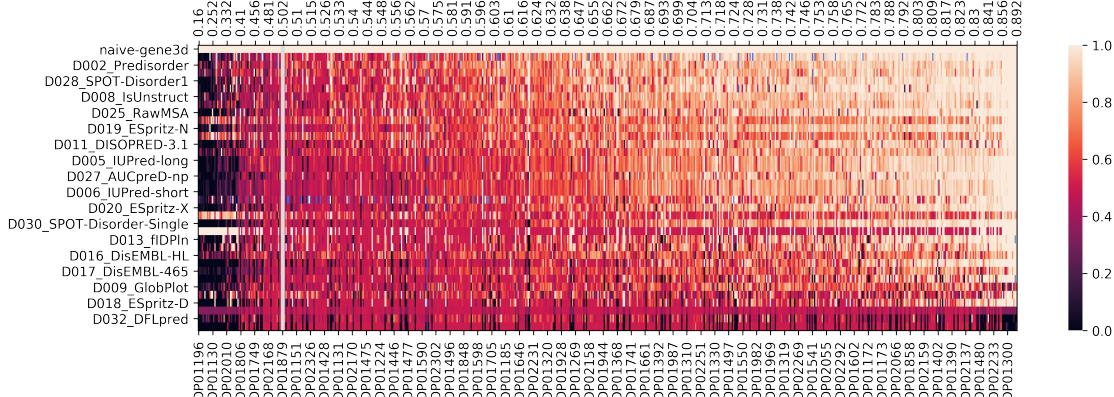
Comparison of predictors and baselines performance by Matthew's Correlation Coefficient.



Overall (left panel) and average per-target (right panel) MCC. Light gray bars represent published prediction methods; dark gray bars represent baseline prediction methods. On the right panel standard deviation is shown as an error bar.

1.2.4 Per target accuracy

Balanced accuracy score for each target for each prediction methods (including baselines)



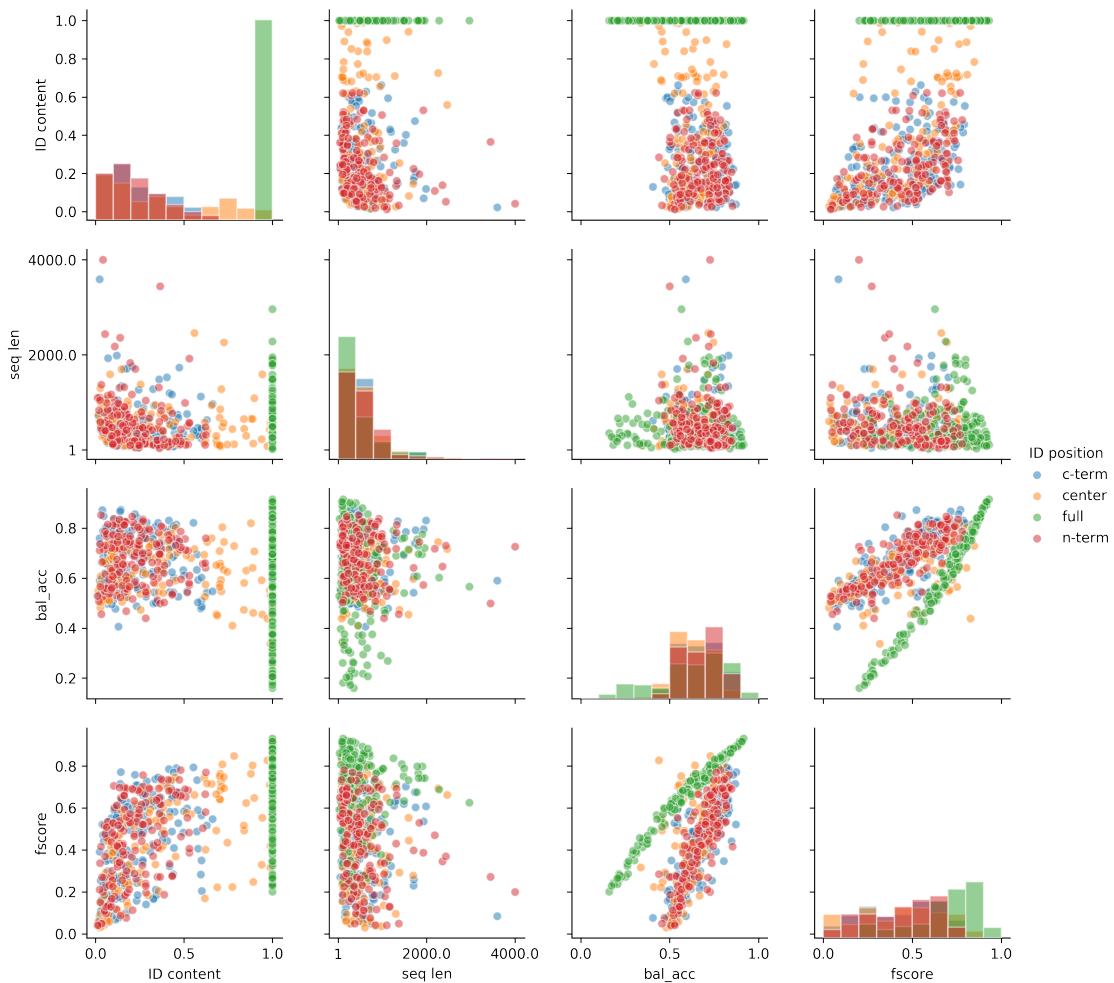
Heatmap of the predictors accuracy for each target. The higher the accuracy the lighter the color. Non-predicted targets are shown in blue. x and y axes are sorted by average accuracy over rows and columns respectively. A white semi-transparent vertical line marks the point where the average accuracy scores for a target is below (left) or above (right) 0.5. Accuracy score approaches 0.5 for a random classifier. Accuracy < 0.5 indicates anti-correlation between predicted and reference classes. Targets with an average accuracy score < 0.5 are:

DP01196	DP01501	DP01432	DP01456	DP01971	DP01500
DP02162	DP01248	DP01949	DP01366	DP01437	DP01130
DP01339	DP01281	DP01190	DP01979	DP01436	DP01555
DP01168	DP01486	DP01177	DP01556	DP02010	DP01870

DP01769	DP01181	DP01498	DP01124	DP01774	DP01767
DP02149	DP01512	DP01278	DP01806	DP01427	DP01304
DP01765	DP02296	DP01163	DP01600	DP01287	DP01134
DP01898	DP01307	DP01749	DP01285	DP01476	DP01502
DP01195	DP01140	DP01139	DP01701	DP01584	DP01355
DP01494	DP02168	DP01612	DP01288	DP01430	DP02234
DP02169	DP01503	DP01869			

1.2.5 Target correlation matrix

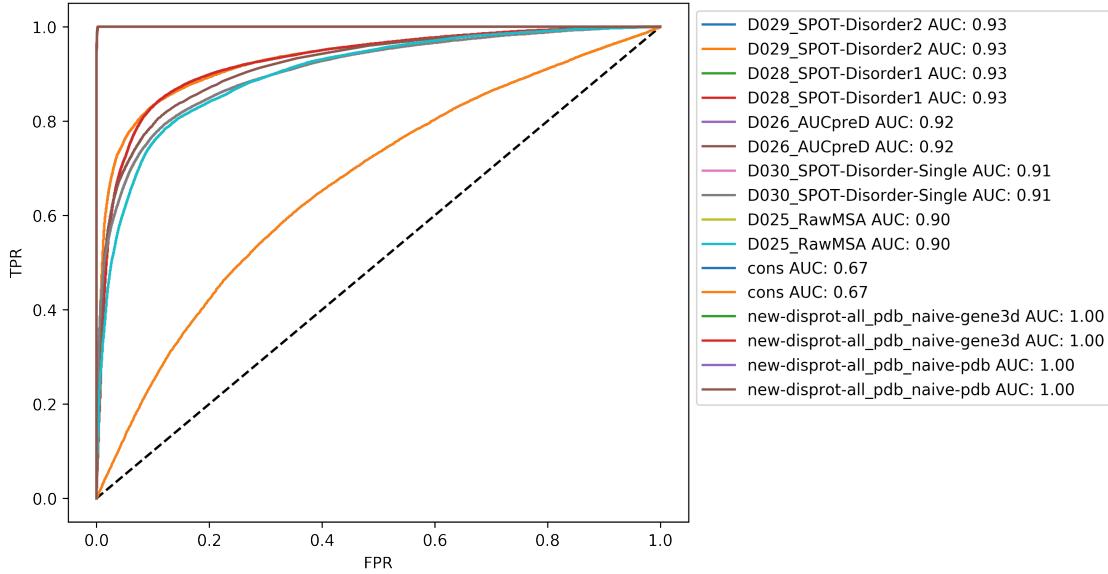
Commonly, experimental data has a bias for low disorder-content. DisProt targets have high disorder-content. Classifiers have been trained/engineered on low disorder-content. I expect difficult targets to have high disorder content. To verify if there is any correlation between target features I'm plotting 4 selected features against each other (Balanced accuracy, F1-Score, Sequence length and ID content). A fifth feature (ID position) divides the datasets in subsets. ID position is calculated as the average of the indexes of disordered residues along the sequence. A correlation is observed in a subset if its points gather around a diagonal.



Correlation matrix of Balanced accuracy, F1-Score, Sequence length and ID content. Average position of disorder is color-coded. Figure matrix is symmetrical. Plots along diagonal axis display single variables distributions. No meaningful correlation is observed.

1.2.6 ROC curve

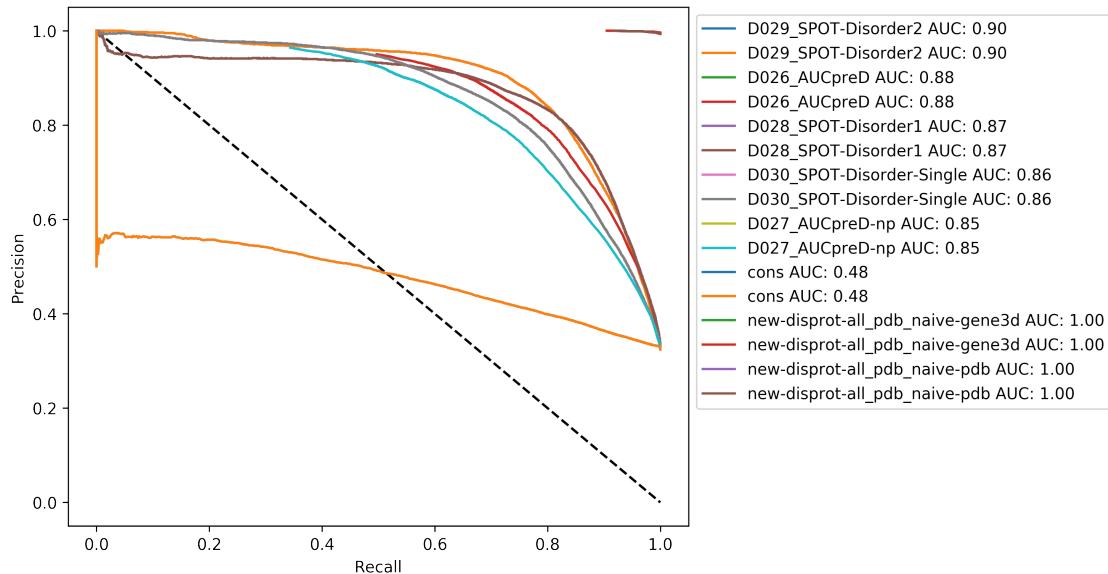
Receiver Operator Characteristic plot for predictors and baselines



False Positive Rate (FPR) on x axis, True Positive Rate (TPR) on y axis. Methods are sorted by their Area Under the Curve (AUC). Only first ten methods plus baselines are shown.

1.2.7 PR curve

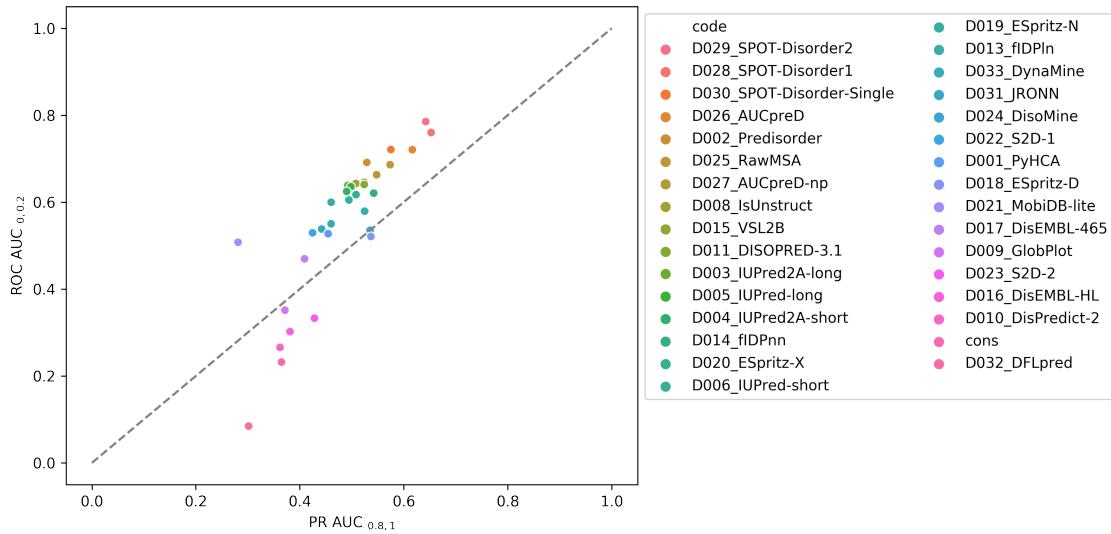
Precision Recall curve plot for predictors and baselines



Recall/Sensitivity on x axis, Precision/Selectivity on y axis. Methods are sorted by their Area Under the Curve (AUC). Only first ten methods plus baselines are shown.

1.2.8 pROC/pPR scatter plot

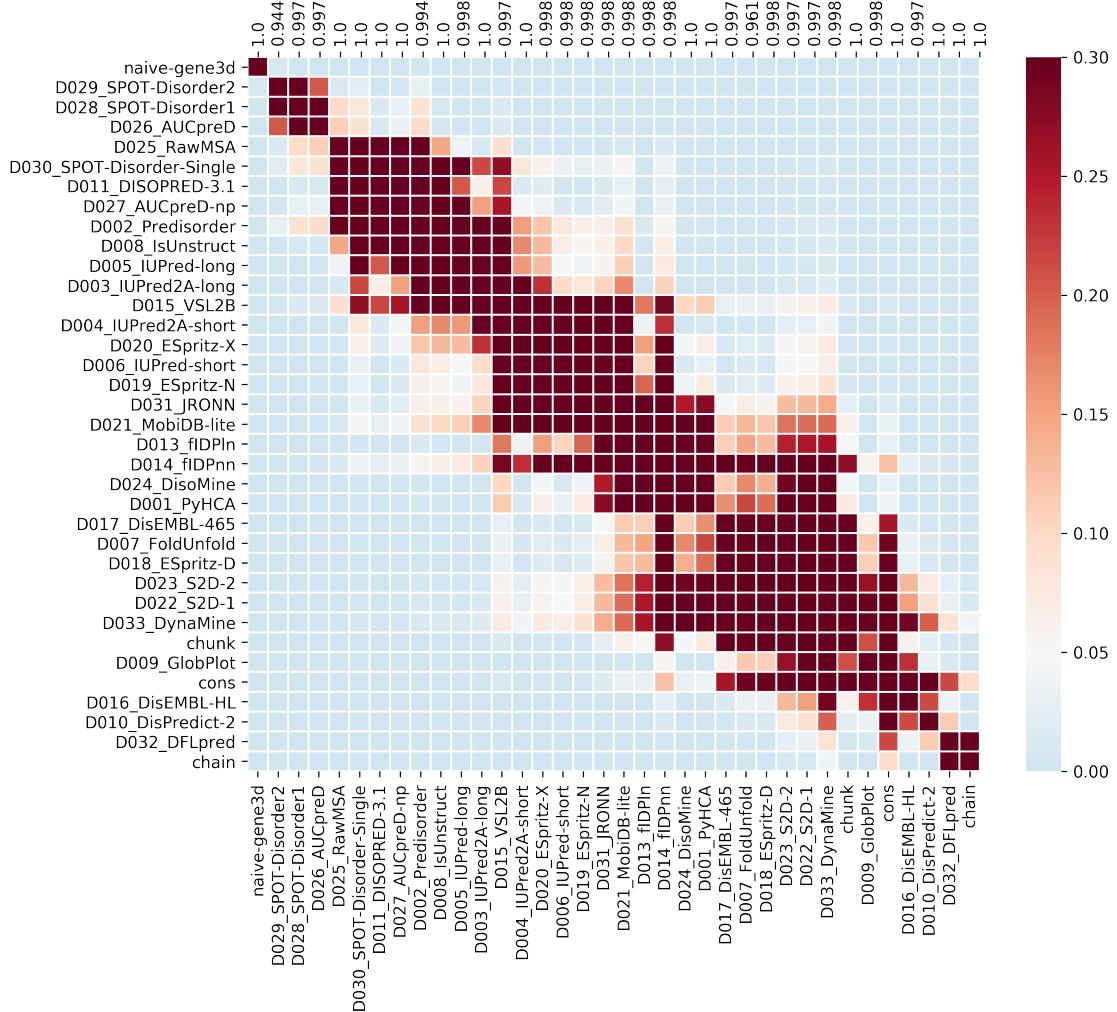
Plot of the AUCs from the ROC curve and PR curve



ROC AUC on the x axis, Precision-Recall (PR) AUC on the y axis. ROC AUC is calculated including ROC curve points with x values (FPR) between 0 and 0.2. PR AUC is calculated including PR curve points with x values (Recall) between 0.8 and 1. Both AUCs are then rescaled to the [0, 1] range.

1.2.9 Average overall ranking

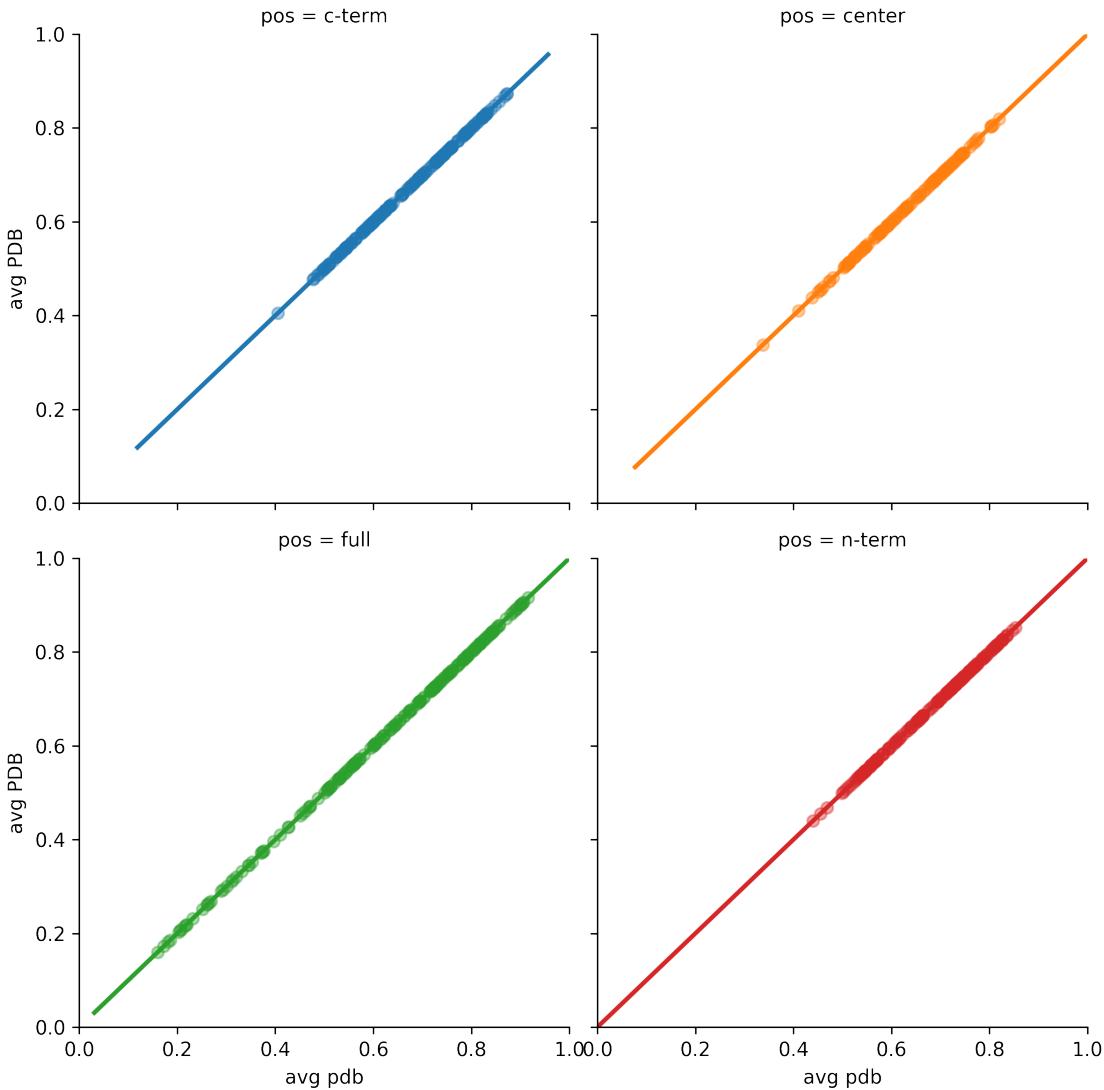
Predictor ranking and ranking statistical significance. Predictors are ranked on average rank from metrics scores: Balanced Accuracy, MCC, Precision, Recall, F1-score, F1-scores on negatives, Precision on negatives, Specificity, ROC AUC, PR AUC.



Heatmap of the p-value associated to the statistical significance of the difference between ranking distributions. Colorormap is centered on 0.05 so that any pvalue above 0.05 is red-ish. Red color indicates that the ranking difference between two predictors is not statistically significant. Top tick labels of x axis display prediction coverage for each predictor.

1.2.10 Accuracy correlation between datasets

Per target average balanced accuracy correlation between *simple* and *pdb* negative definition. Datasets is divided by average disorder position in targets.



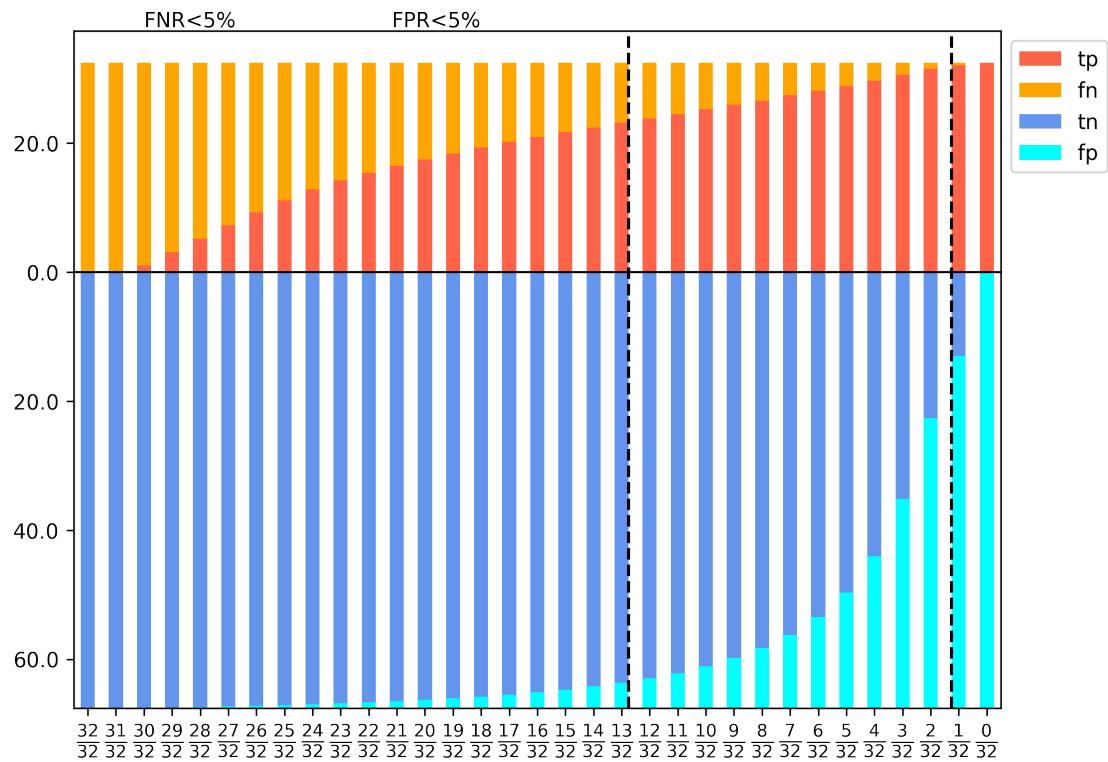
Average balanced accuracy for targets with reference negative defined by the *simple* rule on x axis. Average balanced accuracy for targets with reference negative defined by the *pdb* rule on y axis. Each panel includes only targets with a specific average disorder position (C-terminal, N-Terminal, central, full-disorder)

1.3 Consensus

Consensus among all prediction methods was calculated as the fraction of positive predictions per residue.

1.3.1 Confusion matrix per threshold

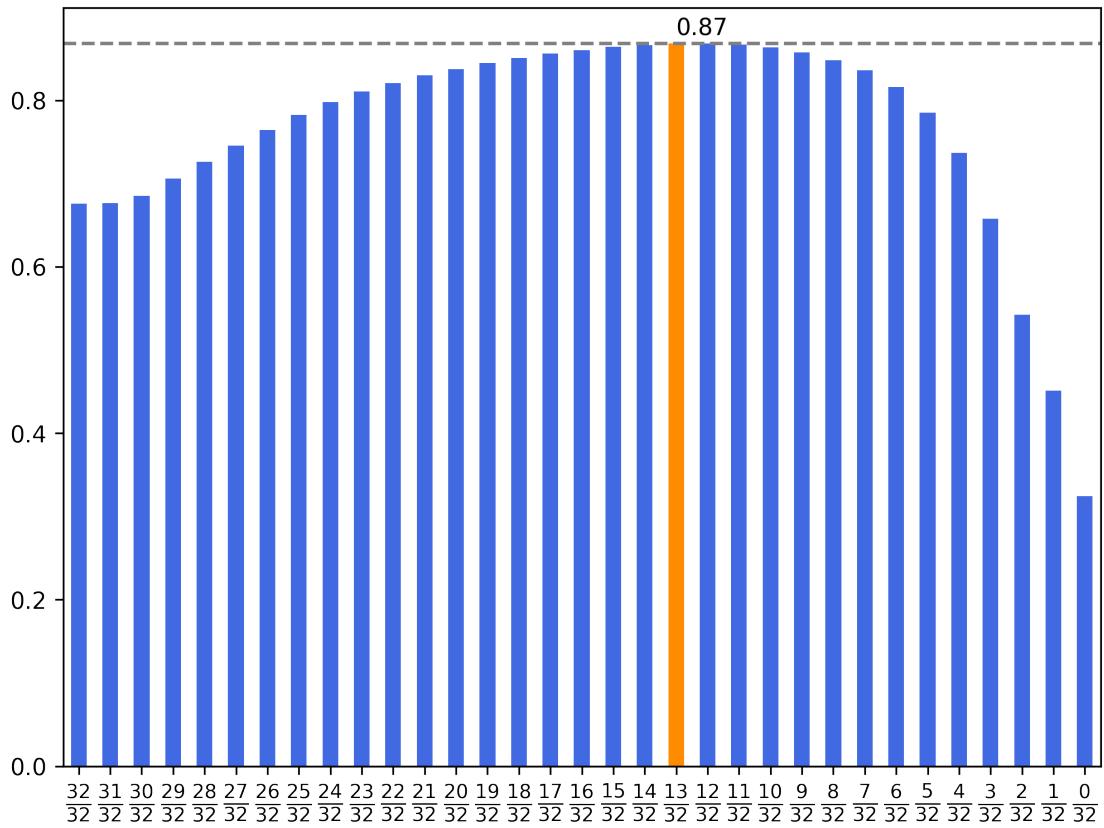
Predicted and actual positive and negatives for each threshold on the consensus score.



Percentage of correct and wrong assignment of positives (above 0) and negatives (below 0) for each threshold of the consensus.

1.3.2 Accuracy per threshold

Balanced accuracy score for each threshold of the consensus.



Accuracy distribution for each consensus threshold. Bar of max threshold is highlighted in orange.

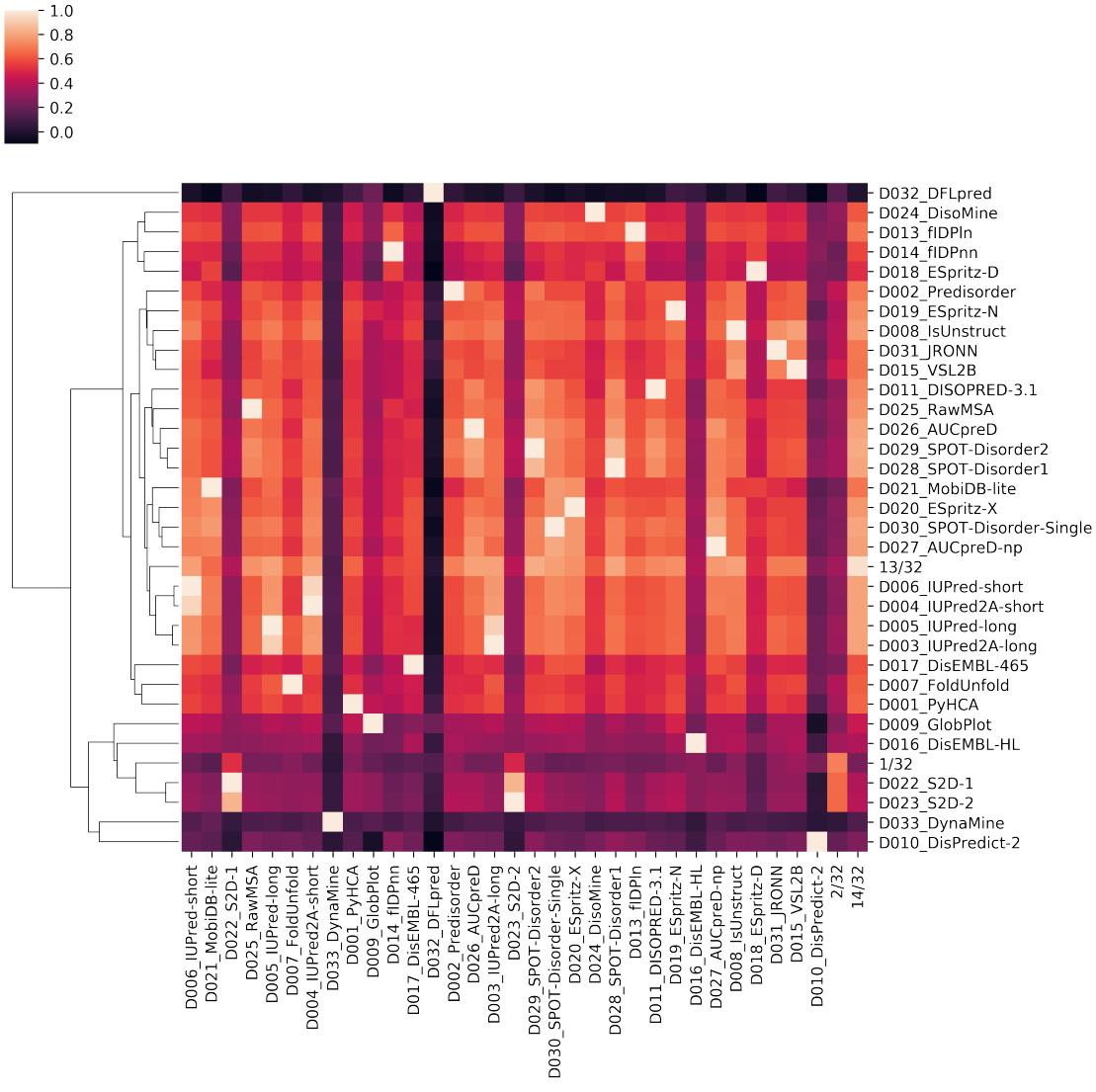
1.3.3 Percentage of correct/incorrect classifications

Percentage of correct and incorrect classifications for positives (defined by DisProt), negatives (defined by PDB) and undefined residues for each predictor.

	DisProt		PDB		Undefined		
	TP	FN	TN	FP	TN	FP	
D006_IUPred-short	56.2	43.8	94.7	5.3	66.2	33.8	
D021_MobiDB-lite	41.1	58.9	98.6	1.4	77.2	22.8	
D022_S2D-1	86.7	13.3	47.3	52.7	22.3	77.7	
D025_RawMSA	68.2	31.8	93.1	6.9	60.6	39.4	
D005_IUPred-long	64.5	35.5	92.5	7.5	58.9	41.1	
D007_FoldUnfold	66.6	33.4	82.5	17.5	48.8	51.2	
D004_IUPred2A-short	56.3	43.7	94.9	5.1	66.3	33.7	
D033_DynaMine	1.7	98.3	100.0	0.0	97.9	2.1	
D001_PyHCA	65.8	34.2	84.8	15.2	53.7	46.3	
D009_GlobPlot	38.3	61.7	90.0	10.0	70.0	30.0	
D014_fIDPnn	30.0	70.0	98.7	1.3	93.0	7.0	
D017_DisEMBL-465	38.7	61.3	95.2	4.8	76.1	23.9	
D032_DFLpred	9.4	90.6	89.0	11.0	89.4	10.6	
D002_Predisorder	80.7	19.3	82.4	17.6	41.8	58.2	
D026_AUCpreD	68.7	31.3	95.5	4.5	54.0	46.0	
D003_IUPred2A-long	63.4	36.6	92.8	7.2	59.6	40.4	
D023_S2D-2	85.8	14.2	50.4	49.6	24.7	75.3	
D029_SPOT-Disorder2	75.9	24.1	95.1	4.9	45.8	54.2	
D030_SPOT-Disorder-Single	55.6	44.4	97.7	2.3	65.3	34.7	
D020_ESpritz-X	53.2	46.8	95.6	4.4	67.5	32.5	
D024_DisoMine	57.4	42.6	91.0	9.0	67.3	32.7	
D028_SPOT-Disorder1	74.8	25.2	94.2	5.8	48.8	51.2	
D013_fIDPln	50.5	49.5	94.6	5.4	79.3	20.7	
D011_DISOPRED-3.1	64.5	35.5	93.9	6.1	49.9	50.1	
D019_ESpritz-N	68.7	31.3	86.6	13.4	51.6	48.4	
D016_DisEMBL-HL	53.0	47.0	74.4	25.6	63.8	36.2	
D027_AUCpreD-np	57.3	42.7	96.5	3.5	65.1	34.9	
D008_IsUnstruct	74.8	25.2	86.0	14.0	49.3	50.7	
D018_ESpritz-D	35.2	64.8	95.6	4.4	88.4	11.6	
D031_JRONN	74.1	25.9	81.7	18.3	47.3	52.7	
D015_VSL2B	81.5	18.5	77.2	22.8	40.0	60.0	
D010_DisPredict-2	41.6	58.4	81.6	18.4	72.5	27.5	

1.3.4 clustermap of binary predictions correlation

Correlation of binary states between predictors.



Heatmap of the correlation of binary prediction states for each couple of predictors. Pearson R is calculated between all predictions. Clustering is based on Euclidean distance calculated over an array (column) of R correlation coefficients.

1.4 Fully disordered targets

Statistics calculated for the subset of targets that are reported as completely disordered in DisProt.

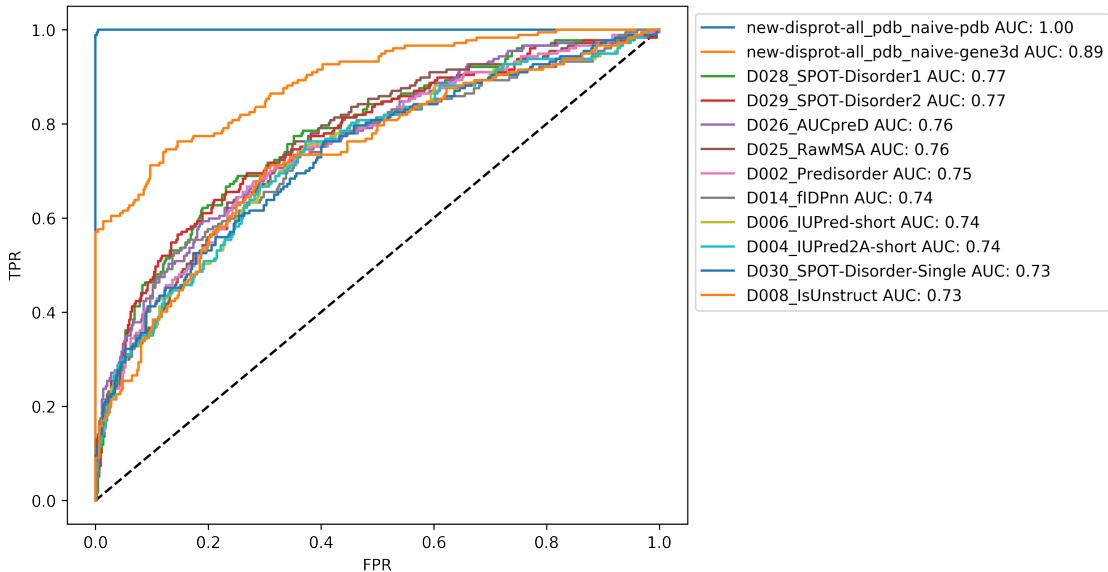
1.4.1 Correctly and incorrectly classified full IDPs

Number of correctly and incorrectly classified full IDPs with a prediction tolerance of 5%.

Actual Predicted	Positives		Negatives	
	TP	FN	FP	TN
random_chain	189	0	0	457
naive-gene3d	189	0	4	453
random_chunk	189	0	4	453
cons	189	0	116	341
D023_S2D-2	188	1	20	437
D022_S2D-1	188	1	32	425
D015_VSL2B	187	2	25	432
D002_Predisorder	186	3	9	448
D001_PyHCA	185	4	5	452
D008_IsUnstruct	185	4	13	444
D019_ESpritz-N	183	6	7	450
D031_JRONN	182	7	8	449
D026_AUCpreD	181	8	12	445
D009_GlobPlot	179	10	0	457
D006_IUPred-short	179	10	3	454
D004_IUPred2A-short	179	10	4	453
D028_SPOT-Disorder1	179	10	24	433
D016_DisEMBL-HL	178	11	2	455
D007_FoldUnfold	176	13	97	360
D017_DisEMBL-465	175	14	0	457
D011_DISOPRED-3.1	174	15	10	447
D027_AUCpreD-np	173	16	7	450
D003_IUPred2A-long	173	16	8	449
D020_ESpritz-X	172	17	5	452
D005_IUPred-long	172	17	11	446
D029_SPOT-Disorder2	168	21	19	438
D030_SPOT-Disorder-Single	167	22	7	450
D024_DisoMine	165	24	43	414
D025_RawMSA	164	25	23	434
D010_DisPredict-2	160	29	16	441
D013_fIDPnn	159	30	33	424
D014_fIDPnn	138	51	10	447
D021_MobiDB-lite	137	52	3	454
D018_ESpritz-D	114	75	43	414
D032_DFLpred	81	108	0	457
D033_DynaMine	49	140	0	457

1.4.2 Full IDPs ROC

ROC for the classification power of Full IDPs. Average disorder scores for each target is compared to full IDPs (positives) and partial IDPs (negatives). 5% prediction tolerance is applied.



FPR on the x axis, TPR on the y axis. Methods are sorted by their AUC. Only first 12 methods are shown.