GeneSilico MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins

Łukasz P. Kozłowski[1,*], Janusz M. Bujnicki[1,2]

1 Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, ul. Trojdena 4, 02-109 Warsaw, Poland,
2 Laboratory of Bioinformatics, Institute of Molecular Biology and Biotechnology, Faculty of Biology, ul. Umultowska 89, 61-614 Poznan, Poland
* lukaskoz@genesilico.pl

Intrinsically unstructured proteins (IUPs) or simply disordered proteins lack a well-defined three-dimensional structure, but still remains functional. The discovery of IUPs challenged the traditional protein structure paradigm, which stated that a specific well-defined structure defines the function of the protein. To date, approximately 60 methods for computational prediction of protein disorder from sequence have been made publicly available.

Given the diversity of existing intrinsic disorder prediction methods, we developed **meta-prediction method** which is based on arbitrarily chosen 13 **disorder predictors**. The final consensus was weighted by the accuracy of the methods. The MetaDisorder was tested in CASP8 being the best method then.

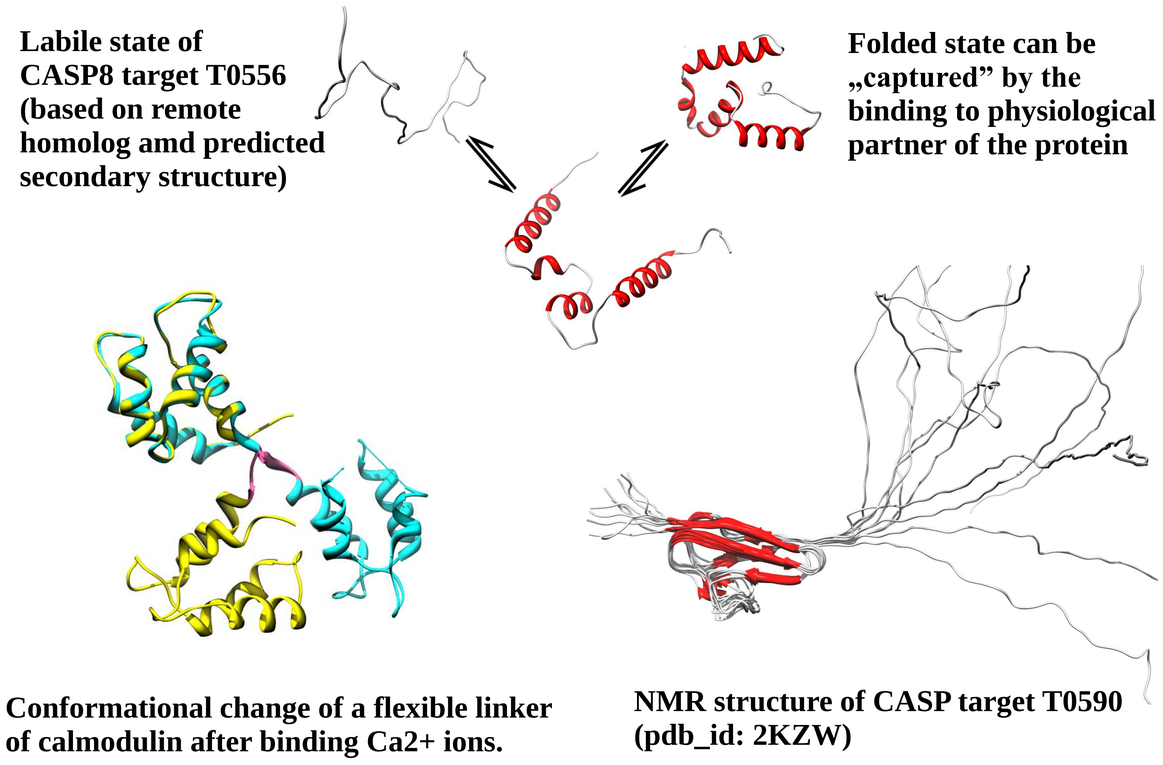
Next, we have developed a disorder predictor **GSmetaDisorder3D that uses alignments** to known protein structures, reported by the protein fold-recognition methods, to infer the potentially structured and unstructured regions.

Finally, we combined this two components into a meta-meta predictor called **GSmetaDisorderMD**, which was one of the top scoring method in the CASP9.

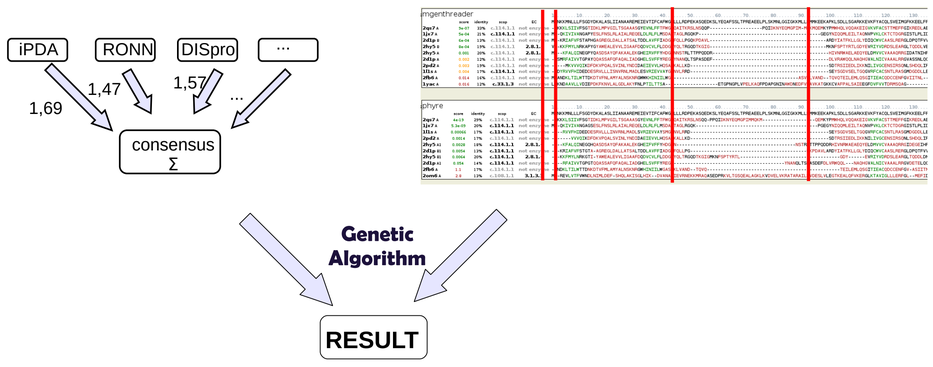
In CASP10 edition, we have been testing different approach. The method called **sDisPred**, which is simple disorder prediction method using **trivial features**, was developed in order to assess how well disorder can be predicted using obvious information e.g. disorder content in homologous sequences, statistics derived from DisProt database, and predicted secondary structure. If no reliable information can be obtained for some parts of the sequence, the consensus from locally installed disorder methods is used.

A series of disorder predictors described here are available as a MetaDisorder web server at **http://iimcb.genesilico.pl/metadisorder/**

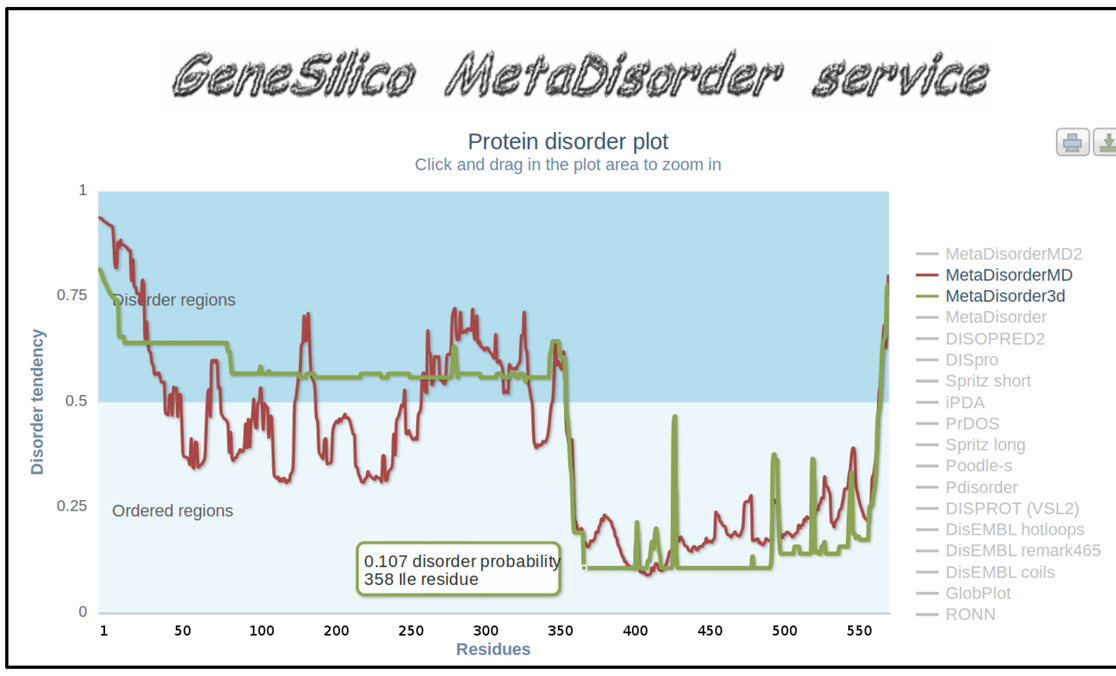
Intrinsically disordered proteins are highly flexible



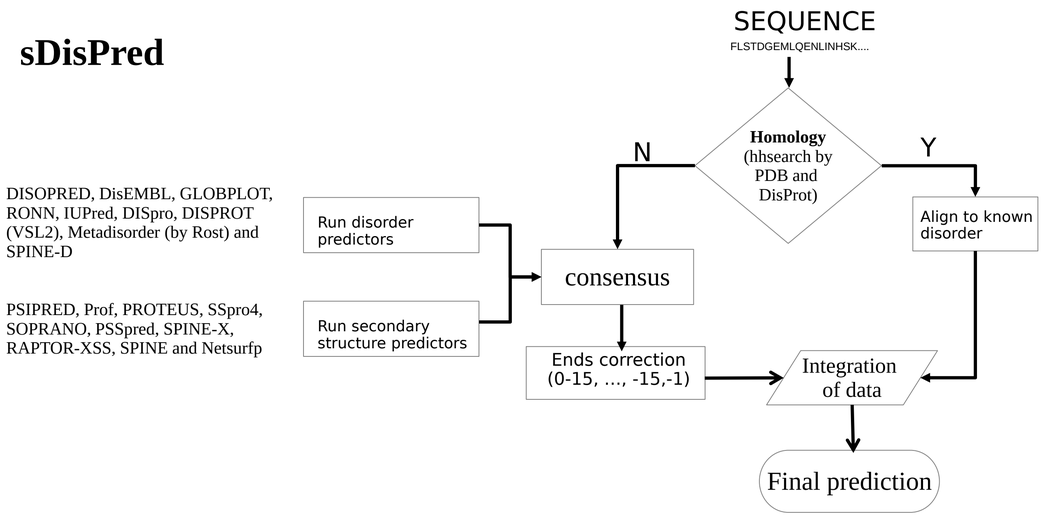
GeneSilico MetaDisorder



MetaDisorder uses consensus from primary disorder methods (13 different programs). Individual methods accuracy is used as a weight for consensus. Additionally, the smoothing filter and 7 amino acids window was employed. Next component uses information from the coverage of the target sequence by known structures, as reported by protein-fold recognition methods To combine all components of the algorithm we used genetic algorithm.



sDisPred



As the 90 percent of CASP targets are from TMB category it is expected that most of targets can be easily predicted based only on the information from homologs. This should be also true for disorder prediction. sDisPred pipeline is negative and positive scan through databases, PDB and DISPROT respectively, in order to find obvious disorder regions with high homology to those which are already known. The regions which cannot be reliably aligned to PDB or DISROT templates are considered “new”and they are validated by second part of pipeline which predict for them disorder and secondary structure using a bunch of programs.

Method performance

Comparison to build-in predictors

	Sw	*AUC*
FloatCons	0.470	0.811
BinCons	0.460	0.806
iPDA	0.453	0.779
DisPSSMP	0.446	0.785
POODLE-L	0.439	0.792
DISPROT	0.419	-
Pdisorder	0.394	-
RONN	0.376	0.740
IUPred (long)	0.373	0.759
POODLE-S	0.372	0.763
DISOPRED	0.368	-
IUPred (short)	0.348	0.747
PrDOS	0.330	0.745
Spritz	0.327	-
Dispro	0.276	0.760
DisEMBL	0.209	-
GlobPlot	0.164	-

Independent, blind test

CASP8

	ACC	*Sw*	*AUC*	*Sens*	*Spec*
FloatCons	**0.831**	**0.662**	0.908	**0.758**	0.904
BinCons	0.830	0.661	0.897	0.741	0.920
DisoClust	0.822	0.644	0.908	0.727	0.917
MULTICOM	0.830	0.660	0.896	0.796	0.864
Mahmood-Torda	0.809	0.619	**0.918**	0.641	**0.978**
POODLE-L	0.794	0.588	0.895	0.646	0.942

CASP9

	ACC	*Sw*	*AUC*	*Sens*	*Spec*
FloatCons	0.713	0.427	0.813	0.572	0.855
GSmetadisorder3D	0.653	0.401	0.798	0.553	0.811
GSmetadisorderMD	0.738	0.476	0.816	**0.654**	0.822
GSmetadisorderMD2	**0.758**	**0.516**	0.839	**0.654**	**0.862**
PrDOS2	0.754	0.509	**0.854**	0.608	0.855
MULTICOM-REFINE	0.750	0.500	0.821	0.650	0.850

* different versions or components of MetaDisorder are marked in bold

REFERENCE

Kozlowski LP, Bujnicki JM. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. BMC Bioinformatics. 2012 May 24;13:111. doi: 10.1186/1471-2105-13-111.

ACKNOWLEDGEMENTS

The consensus approach could not be done without the availability of third-party methods and servers. We would like to thank all developers for kindly making their programs freely available. LK was supported by Polish Ministry of Science grant N N301 190139. JMB was supported by the NIH grant 1R01GM081680-01. Additionally, LPK was supported by EMBO in the form of travel fellowship in CASP10.