DNASTAR®
Software for life scientists

Accuracy of NovaFold Protein Folding Predictions

Introduction

Revolutions in biology and computer science over the past half-century have created a vast potential for improving human and environmental health, and are poised to usher in a new era of engineered drug development.

A drug often achieves a therapeutic objective by binding to a target protein in a patient. Such binding may result, for example, in the disruption of normal substrate binding and/or modulation of the shape or motion of the protein. Clearly, knowledge of the protein's 3D structure and dynamics is a significant aid to engineering new drugs and therapies. But until recently, the drug discovery process and related research studies have relied on high-throughput screening experiments to identify biologically active compounds from a library of hundreds of thousands of molecules.

In 2013, DNASTAR released NovaFold®, a revolutionary 3D protein structure prediction application based on the award-winning I-TASSER algorithm. NovaFold uses iterative assembly simulations to predict 3D protein structure models using protein sequence as input.

NovaFold is available as a separately licensed service within Protean 3D™, which is part of the Lasergene Structural Biology Suite. NovaFold predictions are visualized and analyzed using DNASTAR's graphic-rich Protean 3D application. Results for each prediction include up to five ranked models, with full details reported for all models. NovaFold predictions take place on a high performance cloud computing cluster on the Amazon EC2 web service. (See our white paper on DNASTAR Cloud Security). Alternatively, NovaFold is also available as a local installation for the Linux operating system.

DNASTAR's goals for NovaFold are:

1) To positively impact human health by accelerating drug discovery efforts with the construction of high-confidence structural models of drug targets and biopharmaceuticals.

2) To support a wide range of research objectives for enhanced understanding of how molecular changes impact protein structure and function.

NovaFold's I-TASSER Prediction Algorithm

The engine behind NovaFold is the I-TASSER algorithm, developed by Professor Yang Zhang of the Department of Computational Medicine and Bioinformatics at the University of Michigan. This algorithm utilizes a combination of "threading" and "*ab initio* folding" in predicting protein structure.

- Threading attempts to match portions of the query sequence to template sequences. The template sequences, and their experimentally solved structures, are part of the RCSB Protein Data Bank (PDB).
- *Ab initio* folding uses biophysical properties of the query sequence and simulations to determine the likely structure(s) of the protein.

I-TASSER's Performance in Blind Studies

For twenty years, the biennial Critical Assessment of Protein Structure Prediction (CASP) competition has sponsored blind studies where approximately 100 participating laboratories worldwide test their prediction tools against unpublished protein structures. I-TASSER has won the Protein Structure Prediction Center's three most recent Critical Assessment of Protein Structure Prediction (CASP) experiments, spanning the years from 2012 to present.

Table 1 lists the top ten automated servers using the Z-score for each group. Z-score measures the significance of the predictions compared against the average results. I-TASSER, referred to in CASP as "Zhang-Server," consistently outperformed all other methods. This unswerving success in the CASP competition clearly demonstrates that I-TASSER is the leading structure prediction method available.

Note: References related to the I-TASSER algorithm are provided at the end of this paper.

Table 1. Z-scores of the ten best automated prediction groups in CASP 10-12

For detailed results of each competition, click the header links.

Place	CASP10 (127 targets)		CASP11 (126 targets)		CASP12 (146 targets)	
	Group	Z-score	Group	Z-score	Group	Z-score
1st	**Zhang-Server**	**104.0**	**Zhang-Server**	**132.4**	**Zhang-Server**	**98.7**
2nd	QUARK	97.5	QUARK	125.5	QUARK	93.5
3rd	RaptorX-ZY	79.2	nns	77.7	Baker-Rosetta	92.4
4th	HHpredA	68.1	Myprotein-me	68.7	GOAL	79.6
5th	PMS	74.2	Baker-Rosetta	68.1	RaptorX	67.3
6th	Baker-Rosetta	79.3	MulticonConst	60.8	ToyPred-email	63.8
7th	Tasser-VMT	68.2	Tasser-VMT	43.6	Multicom-Construct	36.4
8th	PconsM	66.9	RaptorX	31.3	Multicom-Cluster	34.3
9th	MulticonNovel	57.7	HHPredA	22.0	Multicom-Novel	33.2
10th	MUfold-Servr	39.1	Falcon_topo	17.2	intFOLD4	26.4

Resources and Free Trial Software

Click the links to watch brief videos introducing NovaFold and showcasing NovaFold prediction results.

Would you like some hands-on experience using NovaFold? Our free trial software lets you view sample NovaFold results and work with their model structures in a fully-functional version of Protean 3D. Simply:

1) Download and install a free trial of Lasergene. When the Navigator opens, click on its Protean 3D bar to launch the application.

2) Follow the step-by-step instructions in this written tutorial to learn how to view sample prediction results from NovaFold. For more information, refer to the help topics describing the result report and "details panel."

3) Open one of the sample predictions as a Protean 3D structure. Refer to the online Protean 3D help or this tutorial for information on viewing the model structure, formatting, making selections, exporting, and much more.

To start making 3D structure predictions using your own protein sequences, please contact us at support@dnastar.com to obtain software and licenses for either local or cloud-based versions of NovaFold.

References

1) Kabsch W and Sander C (1983). "Dictionary of protein secondary structure - pattern recognition of hydrogen bonded and geometrical features." Biopolymers 22 - 2577-2637.

2) Moult J (2005). "A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction." Curr Opin Struct Biol 15: 285-289.

3) Roy A, Kucukural A, Zhang Y (2010). "I-TASSER - a unified platform for automated protein structure and function prediction." Nature Protocols Vol 5, 725-738.

4) Roy A, Yang J, Zhang Y (2012). "COFACTOR - An accurate comparative algorithm for structure-based protein function annotation." Nucleic Acids Research 40 - W471-W477.

5) Ye Y and Godzik A. (2003) "Flexible structure alignment by chaining aligned fragment pairs allowing twists." Bioinformatics19 Suppl 2:ii246-55.

6) Zhang Y (2008). "I-TASSER server for protein 3D structure prediction." BMC Bioinformatics Vol 9, 40.