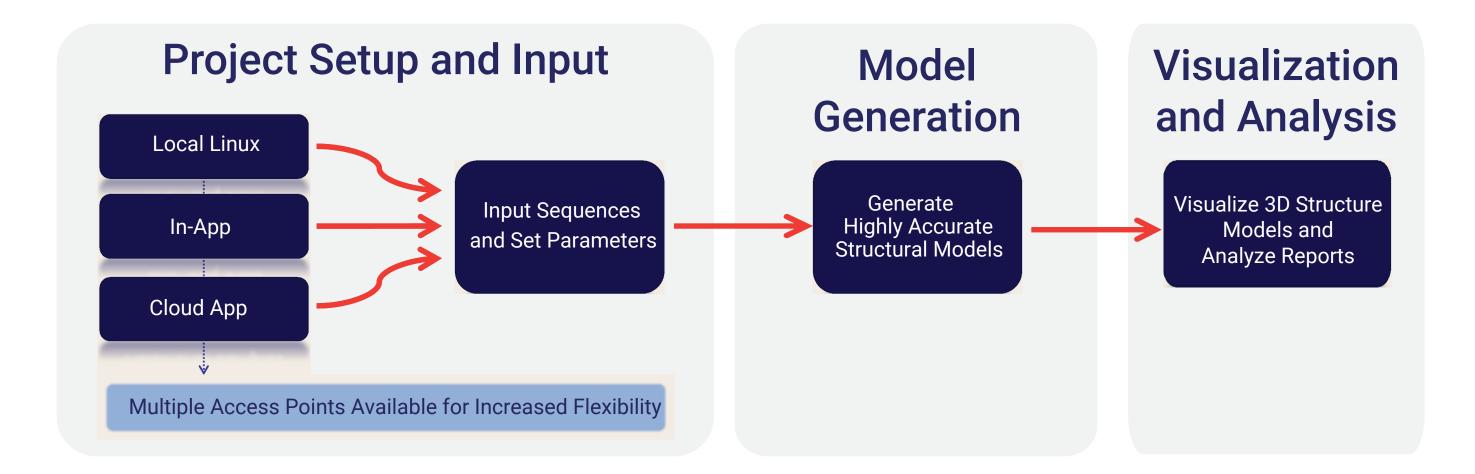


High-Resolution *in silico* Protein Structure Prediction and Docking

Thomas Lynch, PhD, Steven Darnell, PhD, Amanda Mitchell, PhD, Matthew Larsen, PhD, Pavel Pinkas, PhD, Adam Briska, Greg Jakubczak, Frederick Blattner, PhD

<u>Affiliations</u> DNASTAR, Inc., Madison, Wisconsin, USA

Abstract



DNASTAR's new Nova applications allow researchers to generate highly accurate structural models of both proteins and protein-protein complexes that are unattainable through standard modeling methodologies. (1) NovaFold is a protein structure prediction program that utilizes the award-winning I-TASSER algorithm, which combines threading and ab initio techniques to build accurate, full 3D atomic models of proteins with previously unknown structures. (2) NovaFold Antibody is specifically designed to generate models of antibodies and antibody fragments by combining homology modeling for frameworks and ab initio loop prediction. (3) NovaDock is a high-resolution protein-protein docking application that explores flexibility during the modeling process through a particle swarm optimization process encoded by the SwarmDock algorithm. The Nova applications are integrated into the DNASTAR Lasergene Suite, which provides access to a 3D viewer the resulting models and reports analyze in Protean 3D. to

Accurate Protein Structure Prediction

The foundation of NovaFold is the I-TASSER hybrid prediction algorithm, which yields the most accurate protein structural models (up to 2,000 residues). Predictive methods are also applied during NovaFold's automated process that identify potential ligand binding sites and protein functions. From the final report, users can assess model quality through global and local confidence scores, structural and sequence alignments, and apply analysis methods to explore the biophysical properties of the structure.

	Protean 3D - I7IGS3_BABMI-1	1	NovaFold was used to
📄 · 📄 😍 S 💠 🔍 🎄 O			
אין 17IGS3_BABMI-aldolase.novafold 🔂 •17IGS3_BABMI-1 🕅		□ □ ⊗ Ex 🖺 Fe fx Me 🕱 □	generate accurate structural
		Biophysics	models of a putative drug

Lasergene Structural Biology Suite allows users to access the Nova applications locally on a Linux workstation, on the Amazon Cloud as an add-on to Protean 3D, or on the Cloud through a web application. Following model generation, the protein structures and reports can be viewed and analyzed in the Protean 3D molecular viewer.

Antibody Models in Minutes



Using a curated library of antibody frameworks or your own custom templates, NovaFold Antibody builds structural models of antibodies and antibody fragments. Ab initio modeling is used for H3 hypervariable loops, and CDR regions are auto-annotated for user's convenience.

NovaFold Antibody was used to generate four distinct models of a human anti-VEEV antibody. The ensemble of hypervariable H3 loop conformations is highlighted.

		 Veliophysics Amphilicity (Eisenberg) Charge density (Lehninger) Disorder (JRONN) Flexibility (Karplus-Schulz) Hydropathy Secondary structure Stability Surface probability (Emini) Transmembrane Vimmunogenicity Antigenicity Antigenicity Antigenicity T-cell epitopes (DNASTAR) MHC II epitopes (Sette) T-cell epitopes (Sette) T-cell epitopes Voatabases Cleavage sites (Proteases) Domains & sites (PROSITE) Modifications (SUMOylation) Features (Author annotations) Parameters De St & Mo Rendering Backbone Ribbon (cartoon) Side chains Invisible Ligands and water Spheres Surfaces Molecular surface Quality: Fast Normal 1
Structure Analysis		O Transparency
Untitled (17IGS3_BABMI-1:A) KSD 2* Sequence MSDGKGILAADESTGTIKKRFDMINLENTEENRVKYRNLLFTTPLYNQYISGAILFEETLFQNDPNGKPFVEILKENKIIIGIKVDTGLE		Axes Show Euclidean axes
Ruler 10 20 30 40 50 60 70 80 90		∓ Color
		Features
Sequence LIPNTDEYFTKGLDTLSTRCKKYYEAGARFAKWRSVISIDLKTSKPSMLSINTVCDGLAKYASVCQANGLVPIVEPEILADGDHSIETCA Ruler 100 110 120 140 150 160 170 180	U	Sequence Lighting
KSD 2* WWWWW W		- Lighting
Sequence IVTEKVLSVLFKSLYDHGVVLEGTILKINMVTPGFDSKNKSSSQEIAYFTTKALLRTVPPALGGIVFLSGGQSEADATVNLNSINRLGSF		
RCG Sequence		

Model Overview

target from Babesia microti, Aldolase. The example here demonstrates that predictions of high confidence are achievable using templates of low to moderate sequence identity. Top: Highly accurate model 🗐 Мо... 🗖 🗖 (TM-score 0.92, RMSD: 3.4Å) of Aldolase displayed in Protean 3D. Lower Left: Sequence alignment of Aldolase to templates used for structure prediction and threader details. Lower Middle: NovaFold predicted ligand binding sites and functions and connectivity to AmiGO site. Lower Right: Synchronized view of several biophysical methods applied to the sequence.

GS3_BABMI						▼ Predi	icted Bi	inding Sites							
Molecule											- Rank Ligan	d Method	Confidence	Template	Site Residues
esidues: 339											✓ 1 5MM	TM-SITE		<u>3tu9C</u>	9,11,12,13,16,84,122,124,165,167,20
equence: ISDGKGILAADESTGT	IKKREDMI	NLENTEENRY	KYRNLLETTP	LYNOY	YISGAILFEETLFQNDPNGKPFVEILKENKIIIGIKVDT						2 13P	TM-SITE		1adoA	207,248,249,250,251,277,278,279,28
LELIPNTDEYFTKGLD	TLSTRCK	YYEAGARFA	KWRSVISIDLK	TSKP	SMLSINTVCDGLAKYASVCQANGLVPIVEPEILADG		17			GZ	3 111	TM-SITE		2ot0A	12,16,19,20,122,124,250,280,283
					SSQEIAYFTTKALLRTVPPALGGIVFLSGGQSEADAT QRAKENSMACLGEFNDSEVEDASKDTLFEKRYVY		4				4 III	TM-SITE	0.11	<u>3lgeB</u>	12,15,16,20,124,171,215,249,250,251
To see between											5 PO4	TM-SITE	0.11	2qdgB	124,165,167,248,250,280
Templates								γ		∇	6 N3P	TM-SITE	0.09	<u>2ot1A</u>	12,20,280,283,287
- Rank Template	Z-Score	Threader	% Coverage	% ID	Мар			Y			7 PO4	TM-SITE	0.04	2qdqB	13,84,86,88,124
2 1 2pc4C		MUSTER	99	60				5			8 PO4	COFACTOR	0.03	<u>1zalA</u>	11,12,13,16,84
2 <u>3mbfA</u>	3.85	dPPAS	94	43					-Par		9 111	S-SITE	0.01	<u>3lgeC</u>	12,15,16,20,124,167,171,215,220,251
3 <u>2pc4C</u>	4.76	wdPPAS	99	60				and the second sec				1	1		
4 <u>2pc4C</u>	3.68	WMUSTER	99	60				-			Open model with	checked ligan	ds and binding	sites	
5 <u>2pc4C</u>	5.32	WPPAS	99	60							Spin the model				
6 <u>3mbfA</u>	6.44	dPPAS2	94	43				_				17IGS3_BA	BMI-aldolase.novafold	×	- 8
7 <u>3mbfA</u>	4.62	PPAS	94	43		* Predi	icted Er	nzyme Func	tion			17IGS3_8	ABMI		
8 3mbfA	7.48	Env-PPAS	94	43			Rank	EC Number	Confidence	Template	Site Residues		000		Features - Predicted annotations
9 <u>1aldA</u>	2.95	MUSTER	99	53			1	4.1.2.13	0.64	1xdmB	11,165,207				
10 <u>2pc4C</u>	3.84	dPPAS	99	60			2	4.1.2.13	0.54	1aldA	5,50)-N-M-)1-1	<u>,</u> , ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	Model 1 KSD 2* Structure
Open checked temp	ate fraom	ents aligned t	o: Model 1	Model	2. Model 3. Model 4. Model 5		3	4.1.2.13	0.53	<u>1xfbA</u>	11,165,207	0			-2pc4C
			incut in the				4	4.1.2.13	0.52	2pc4C	11,165,207		-		3mbfA

5 <u>4.1.2.13</u>

6 4.1.2.13



High-Resolution Protein Docking

NovaDock accurately predicts the flexible binding interface of two binding parts and generates an efficient report displa potential binding poses. In the Noval report, users are able to evaluate er scores, cluster sizes, observable lie alternate contacts, and doc conformations for each model. From interactive view, users can easily ide interfacial residues and create documents to interrogate the model molecular vie 3D Protean the 🗖 ' 🗟 😵 🖏 🖉 🔁 '

Beta-Lactamase-1JTG.novadock
MovaDock-1JTG-comparison.structure X

Beta-lactamase_Inhibitor		
Malagulas		
Model Overview		
Filtering method: Remove if not in binding funne	el 🗘	
Model Energy Cluster Size Residue Contact	ts Contacts Fulfilled	
	-	
✓ 2 -32.42 6 59 ✓ 3 -30.14 6 55	-	
• • • • • • • • • • • • • • • • • • • •	-	
5 -22.67 4 28	-	
✓ 6 -22.14 2 43 ✓ 7 -21.37 2 40	-	
$\boxed{\begin{array}{c} \hline \\ \hline $	-	
9 -19.21 1 39	-	
6 ∎ 1 0 -18.84 5 34	-	
Cluster energy: -42.3 -18.8 Export model overview image		
* Model 1		
		Intermolecular Contacts
Molecule: Model 1	Molecule	
Molecule: Model 1	Receptor	B:ASP 49
Molecule: Model 1 Energy: -42.28 Cluster size: 24 ✓ A:SER 70 A:SER 70	Receptor Receptor	B:ASP 49 B:ASP 49
Molecule: Model 1 Energy: -42.28 ALYS 73	Receptor	B:ASP 49
Molecule: Model 1 ✓ A:SER 70 Energy: -42.28 ✓ A:LYS 73 Cluster size: 24 ✓ A:LYS 73 Cluster energy: -34.97 ± 4.44 ✓ A:GLN 99 Residue contacts: 62 ✓ A:ASN 100 Ligand chain: B ✓ A:SP 101	Receptor Receptor Receptor	B:ASP 49 B:ASP 49 B:TRP 150 B:ARG 160, B:TRP 150 B:ARG 160
Molecule: Model 1 Image: Active state sta	Receptor Receptor Receptor Receptor Receptor Receptor	B:ASP 49 B:ASP 49 B:TRP 150 B:ARG 160, B:TRP 150 B:ARG 160 B:TRP 112, B:TRP 162
Molecule: Model 1 Image: Algorithm of the system Energy: -42.28 Image: Algorithm of the system Cluster size: 24 Image: Algorithm of the system Ligand chain: B Image: Algorithm of the system Deen model in new document Image: Algorithm of the system Image: Algorithm of the system Open model in new document Image: Algorithm of the system Image: Algorithm of the system	Receptor Receptor Receptor Receptor Receptor	B:ASP 49 B:ASP 49 B:TRP 150 B:ARG 160, B:TRP 150 B:ARG 160 B:TRP 112, B:TRP 162 B:TRP 162
Molecule: Modecl 1 ✓ A:SER 70 Energy: -42.28 ✓ A:LYS 73 Cluster size: 24 ✓ A:GLN 99 Residue contacts: 62 ✓ A:ASN 100 Ligand chain: B ✓ A:ASP 101 Receptor chain: A ✓ A:LEU 102 Open model in new document ✓ A:VAL 103	Receptor Receptor Receptor Receptor Receptor Receptor Receptor	B:ASP 49 B:ASP 49 B:TRP 150 B:ARG 160, B:TRP 150 B:ARG 160 B:TRP 112, B:TRP 162 B:TRP 162 B:GLU 73, B:PHE 142, B:TYR 143
Molecule: Mode/1 ✓ A:SER 70 Energy: -42.28 ✓ A:LYS 73 Cluster size: 24 ✓ A:GLN 99 Residue contacts: 62 ✓ A:ASN 100 Ligand chain: B ✓ A:ASN 100 Open model in new document ✓ A:AULU 102 Open model in new document ✓ A:VAL 103 Spin the model ✓ A:GLU 104	Receptor Receptor	B:ASP 49 B:ASP 49 B:TRP 150 B:ARG 160, B:TRP 150 B:ARG 160 B:TRP 112, B:TRP 162 B:TRP 162

Color Features

Sequence

∓ Lighting① Details ☆

NovaDock-1JTG

near-perfect structural model (RMSD: 1.27Å) when compared to the co-crystal x-ray structure of the E. coli TEMP-1 beta lactamase and beta lactamase inhibitor protein II complex (PDB ID: 1JTG). No

					1	1 1			.13 0.33 180111	00,124,102		■ III	
lodel	TM-Score	TM Error	RMSD	RMSD Error	C-Score	Cluster Size	Density Score	7 4.1.2	.21 0.26 <u>2v81A</u>	207			
1	0.92	±0.06	3.39	±2.36	1.51	18578	0.92	8 4.1.2	.14 0.23 <u>2yw3E</u>	207	0	Temputs - Temp	ate coverage
2	0.51	±0.15	10.28	±4.64	-1.66	1011	0.04	9 4.1.3		207	Ŭ	۰	<u> </u>
3	0.33	±0.11	15.05	±3.50	-3.49	135	0.01	10 4.1.2	.4 0.19 <u>1j2wA</u>	122		Molecule - Solven	accessibility
4	0.26	±0.08	17.59	±2.56	-4.38	115	0.00	Open model with	checked active sites		¢		
5	0.26	±0.08	17.59	±2.56	-4.38	102	0.00) -	la - Flavibility
								Predicted Protein	Function		0		
lodel 1								GO Molecular Fu	Inction (GO-score): fructose-b	oisphosphate aldolase ac	ti		
								<u>binding</u> (0.66), <u>a</u>	ctin binding (0.66), phosphatid	ylcholine binding (0.64),		Malecule - Second	ary structure
					M-score:	0.92±0.06		GO Biological Pr	ocess (GO-score): <u>glycolysis</u> (0.99), <u>gluconeogenesis</u>	(
				R	MSD:	3.39±2.36		shape (0.66), pla bomeostasis (0.6	telet activation (0.66), ATP bio 66), actin filament organization	synthetic process (0.66)	,	16 J Models - Model 1	Ca distances
	9				-score: luster size	1.51 18578 of 1	20200	(0.64), positive r	egulation of ATPase activity (0.	64), response to drug (0			1
	$\sim \sim $	× 1	AA			re: 0.920	0200	development (0.6 (0.64), response	64), response to amino acid stir to starvation (0.64), response	<u>mulus</u> (0.64), <u>response t</u> to organic cyclic compo	2		
				<u>o</u>	pen model	in new docume	nt	response to zinc (0.64)	ion (0.64), cellular response to	extracellular stimulus (C		16 Models - Model2	Cadistances
		195		S	pin the mo	del		(0.84)				o	
								GO Cellular Com	iponent (GO-score): <u>cytosol</u> (C 66), <u>plasma membrane</u> (0.64), <u>r</u>	0.88), <u>I band</u> (0.66), <u>plate</u>	<u>el</u>	16 Models - Model 3	Ca distances
		17 B						membrane (0.64)), perinuclear region of cytoplas	sm (0.64), centriolar sate			
			5D									۰	Cadistances
	2	S J						Model 2			0		
								Model 3			Repor	rt 🛛 🔛 Analysis	1
								> Model 4					
eport			Analysis					Report	Analysis				

Research reported in this publication was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Numbers R44GM110814 and 5R44GM100520. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

		2 structures in document Model 1 Title: Model 1 - fr (Beta-Lactamase-1JTC 2 biopolymers with 429 f Structure alignments A(26290) + Model 1 vs. 1JTG:A(26288) + 1JT RMSD [Å]: 1.105
Structure	Analysis	▼ 1JTG Title: CRYSTAL S OF TEM-1 BETA-LACTA BETA-LACTAMASE INH PROTEIN COMPLEX Classification: HYDROLA Technique: X-RAY DIFF Resolution [Å]: 1.73
KSD 2° Sites Sequence	HPETLVKVKDAEDQLGARVGYIELDLNSGKILESFRPEERFPMMSTFKVLLCGAVLSRIDAGQEQLGRRIHYSQNDLVEYSPVTEKHLTD	PubMed ID: 11573088 PDB ID: 1JTG PDB ID: 1JTG
Ruler KSD 2°		 4 biopolymers with 856 BETA-LACTAMASE TEM Chains: <u>A</u> , <u>C</u> Type: Prot
Sites Sequence Ruler	GMTVRELCSAAITMSDNTAANLLLTTIGGPKELTAFLHNMGDHVTRLDRWEPELNEAIPNDERDTTMPVAMATTLRKLLTGELLTLASRQ	263 UniProt ID: <u>P62593</u>
KSD 2*	125 135 145 155 165 175 185 195 205 205	BETA-LACTAMASE INHIB PROTEIN Chains: <u>B</u> , <u>D</u> Type: Prot 165

arison.structurestructuraltemplatesorpriordel1Model 1 - frknowledgewereusedinthedocking simulations.thedocking simulations.thedocking simulations.gyo) + Model 1 vstrop:Theresultsoverviewwithgyo) + Model 1 vstrop:Theresultsoverviewwithgyo) + Model 1 vstrop:Theresultsoverviewwithgyo) + Model 1 vstrop:Theresultsoverviewwithgyo) + Model 1 vstrop:Theresultsoverviewwithinterfacialcontactresiduesidentified.Bottom:Structuralalignmentofthemodel(green and yellow))againstthe1JTGx-rayco-crystalob/mers with 856the1JTGx-rayco-crystalstructure(blue and orange).ot ID:ID:FP2593thethetheco-crystal

