User Guide to

Protean 3D

DNASTAR, Inc. 2022

Table of Contents

Welcome to Protean 3D **6**
Separately-licensed services that integrate with Protean 3D 7

Protean 3D Tutorials **8**

Protean 3D Interface **9**
Welcome screen 11
Project tabs 13
Protean 3D toolbar 14
Views 16
Structure view 17
Try it! – Use the Structure view toolbar 19
Default colors and styles in the Structure view 21
Analysis view 23
Try it! – Apply tracks to the Analysis view 27
Part A: Apply and reorder tracks 28
Part B: Compare results from two tracks 32
Part C: Edit track options 35
Part D: Recolor tracks 37
Colors in the Analysis view 39
Sequence view 40
Report view 42
Composition report 44
Try it! – Study surface-exposed cysteines 45
Biophysical Properties report 46
Protein Design reports (Variants and Hot Spot Scan) 48
Filter hot spot scan results 50
NovaFold report 52
NovaFold Report: Molecule 54
NovaFold Report: Templates 55
NovaFold Report: Model Overview 57
NovaFold Report: Model ‘x’ 59
Similar Experimental Structures 60
Predicted Binding Sites 61
Predicted Enzyme Function 63
Predicted Protein Function 64
Open a structure from the NovaFold Report 65
NovaDock report 66
NovaDock Report: Molecules 67
NovaDock Report: Model Overview 69
NovaDock Report: Model ‘x’ 71

Models in Cluster ... 73
Open a structure from the NovaDock Report ... 74
NovaCloud Services view ... 75
Predictions view ... 79
Text view ... 82
Table view ... 84
Pairwise view ... 87
Panels ... 91
Explorer panel ... 92
Molecules section ... 93
Try it! – Use the Molecules section to select and hide a structural element ... 95
Surfaces section ... 98
Examples of surface types ... 101
Features panel ... 103
Try it! – Select and recolor a feature, and change feature options ... 106
(Feature) Options section ... 110
(Feature) Style section ... 113
Tracks panel ... 115
Track Options section ... 117
Biophysics tracks ... 119
Amphiphilicity (Eisenberg) ... 120
Charge Density (Lehninger) ... 121
Disorder (JRONN) ... 122
Flexibility (Karplus-Schulz) ... 123
Hydropathy – Hopp-Woods ... 124
Hydropathy – Kyte-Doolittle ... 125
Hydropathy – Parker ... 126
Secondary Structure – Chou-Fasman ... 127
Secondary Structure – Deleage-Roux ... 129
Secondary Structure – Garnier-Robson ... 130
Secondary Structure – Coiled Coil ... 132
Stability – Aliphatic Index ... 133
Stability – Instability Index ... 134
Stability – Isoelectric Precipitate ... 135
Surface Probability (Emini) ... 136
Transmembrane – Goldman-Engleman-Steitz ... 137
Transmembrane – TopPred ... 138
Transmembrane – von Heijne ... 139
Immunogenicity tracks ... 140
Antigenicity – Jameson-Wolf ... 141
Antigenicity – Welling ... 142
B-Cell Epitopes (DNASTAR) ... 143

MHC II Epitopes (Sette) 145
T-Cell Epitopes – AMPHI 146
T-Cell Epitopes – Rothbard-Taylor 147
Database tracks 148
Cleavage Sites (Proteases) 149
List of proteases 152
Domains & Sites (PROSITE) 155
Update the PROSITE database between releases 159
Modifications (SUMOylation) track 160
NovaFold-related tracks 161
Features (Annotated) tracks 163
Details panel 165
When no document or a blank document is open 167
A single residue or ligand 168
A molecule, feature or region 171
An entire structure 174
Try it! – View details about a structure 175
A selected region or strand 177
A group of structures 179
A track 180
A protease pattern or profile 181
An individual protease match 182
A PROSITE pattern or profile 183
An individual PROSITE match 185
A Protein Design result 186
The NovaFold report 189
The NovaDock report 191
Style panel 192
Rendering section 194
Try it! – Change the rendering of a structure 197
Examples of rendering styles 199
Color section 202
Sequence section 207
Lighting section 208
Models panel 210
Jobs panel 212
Motion Library 214
Try it! – Open a multi-model file from the Motion Library 217

Customize the Look and Layout 220
Customize the appearance and layout of views 221
Customize the layout of other window components 223
Save the layout and apply it to other projects 225

Open a File or Project **226**
Open one or more local files 227
Change the default application for opening files 228
Type or paste a protein sequence 229
Open a PDB file when you know the ID 230
Open a PDB file when you don't know the ID 231
Open a structure from the Motion Library 233
Open an existing Protean 3D project 235

Search for Sequences Online **236**
Search for sequence matches (BLAST) 237
Search wizard: Query 238
Search wizard: Search 241
Search wizard: Options 243
Search wizard: Job 246
Search for text or accession number matches (Entrez) 247
Filter sequence search results 250
Filter text search results 252
Access ID dialog 253

Align Multiple Structures **254**
Try it! – Align multiple structures using jFATCAT 256

Make a Selection **261**
Try it! – Make a selection from the Sequence view 263

Work with Features **267**

Overlay One Structure on Another **269**

Create or Find Variants using Protein Design **270**
Protein Design wizard: Workflow 272
Protein Design wizard: Structure 274
Protein Design wizard: Substitutes 275
Protein Design wizard: Options 277
Protein Design wizard: Job 279
View progress and see variant results 280
Use Protein Design for site-directed mutagenesis 281
Select Neighboring Residues 289

Evaluate Variants Produced in a SeqMan NGen Assembly **290**
Part A: Create a templated assembly in SeqMan NGen 291
Part B (optional): Filter for variants of interest in ArrayStar 292
Part C: View variants in SeqMan Pro 295
Part D: View the protein structure in Protean 3D 296

Predict Structures, Antibody Interactions and Docking **298**
Types of predictions 299
NovaFold: protein structure prediction 300
NovaFold AI: protein structure prediction 301
NovaFold Antibody: antibody interaction prediction 302
NovaDock: protein-protein docking 304
Run Nova applications within the Protean 3D interface 305
Run NovaFold through Protean 3D 306
NovaFold wizard: Sequences 307
NovaFold wizard: Options 309
NovaFold wizard: Submit 314
Run NovaFold AI through Protean 3D 315
NovaFold AI wizard: Sequences 316
NovaFold AI wizard: Options 318
NovaFold AI wizard: Submit 320
Run NovaFold Antibody through Protean 3D 321
NovaFold Antibody wizard: Sequences 323
NovaFold Antibody wizard: Options 325
NovaFold Antibody wizard: Submit 328
Run NovaDock through Protean 3D 329
NovaDock wizard: Structures 330
NovaDock wizard: Options 332
NovaDock wizard: Submit 334
Try it! – Phage display assembly, translation and prediction 335
Part A. Assembly and trimming in SeqMan Ultra 336
Part B: Translation in SeqNinja 341
Part C: Alignment in MegAlign Pro 343
Part D: Structure prediction and docking in Protean 3D 347
Run Nova applications through the DNASTAR website 352
Run Nova applications locally 354
Install and set up the software 355
Uninstall the software 358
Reinstall the software 359
Prepare the input file(s) 360
Set up and run a prediction 361
Required commands 364
Required arguments 365
Optional arguments 366
Supplemental arguments 369
Stop a prediction in progress 377

Save and Export **378**
Save a document 379

Export a still image 381
Export a structure file 382
Export an animation or video 383
Export data from the Text, Table or Pairwise views 384
Export a table of results from the Analysis view 385
Save or export a report 386

Close and Exit 387

Miscellaneous Tasks 388
Use basic editing commands 389
Set preferences 391
Set data preferences 392
Set Internet preferences 393
Set proteins preferences 394
Set temporary file preferences 395
Synchronize sequence updates between two Lasergene applications 396
Turn off usage logging 397
Use a mouse or trackpad with Protean 3D 398
Change the default print size 399
Prevent firewall issues when downloading from the Motion Library 400
Troubleshoot the error “A different version is running” 401
Troubleshoot blank Protean 3D window on Parallels Desktop 402

Appendix 403
Glossary 404
Supported file types 407
IUPAC codes 408
Installed Lasergene file locations 410
Useful links 412
Research references 413
Acknowledgements 417

Welcome to Protean 3D

Protean 3D is Lasergene's application for exploring macromolecular structure, motion, and function. Rich, synchronized graphical views allow you to see the 3D molecular structure, the annotated sequence, and the analysis of applied prediction methods simultaneously, enabling easy identification and analysis of secondary structure elements.

Protean 3D also provides access to the Motion Library, where you can browse and search over 300 animated and annotated macromolecular conformational changes, and as the user interface for NovaFold and NovaFold AI (protein structure prediction), NovaFold Antibody (antibody structure prediction), NovaDock (protein-protein docking) and Protein Design (variant exploration).

The following video is a brief walk-through of the Protean 3D application.

These support pages contain written help and tutorials, as well as video tutorials.

- For help **INSTALLING** Lasergene, see our separate Installation Guide.
- For written **TUTORIALS** with free data, see Protean 3D Tutorials.
- Click here to list all topics with **VIDEOS**. Topics discussing "videos" in relation to Protean 3D will also show up in this search.
- To **PRINT** the current page of the User Guide, click the printer icon in the top right corner ( ).
- To download a **PDF** of the entire User Guide, scroll to the bottom of the table of contents on the left, and press **Download as PDF**.

Copyright DNASTAR, Inc. 2022

Separately-licensed services that integrate with Protean 3D

DNASTAR offers four separately-licensed services that integrate with Protean 3D:

- Predict a protein structure with NovaFold (powered by I-TASSER) or NovaFold AI (powered by AlphaFold 2)
- Predict antibody interactions with NovaFold Antibody
- Explore protein docking with NovaDock
- Create or find variants using Protein Design

Protein Design is always run through the Protean 3D interface. Depending on the type of license, the other services can be run:

- On the cloud from within the Protean 3D interface
- On the NovaCloud through the DNASTAR website
- On a local computer through the command-line.

Regardless of where the prediction is initiated, the results are always viewed through the Protean 3D interface.

For information on licensing these services, please contact support@dnastar.com.

Copyright DNASTAR, Inc. 2022

Protean 3D Tutorials

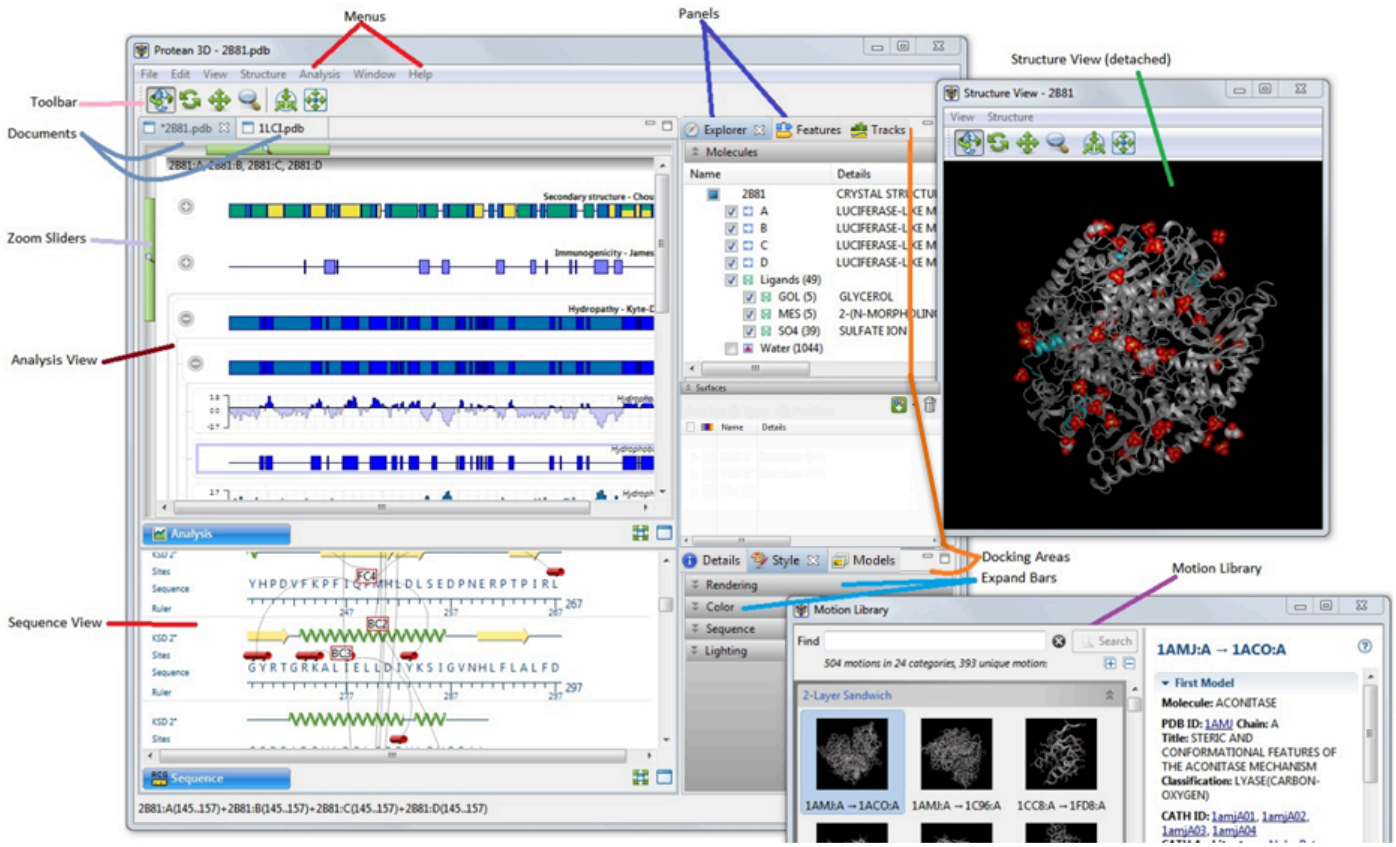
In this User Guide, tutorial topics are located in the table of contents directly below the corresponding help topic. To help you find the tutorials, they are also listed below.

- Use the Structure view toolbar (Structure view)
- Apply tracks to the Analysis view (Analysis view, Tracks panel)
- Make a selection from the Sequence view (Sequence view)
- View details about a structure (Details panel)
- Change the rendering of a structure (Style panel)
- Select and recolor a feature, and change feature options (Features panel)
- Use the Molecules section to select and hide a structural element (Molecules section)
- Study surface-exposed cysteines (Report view)
- Open a multi-model file from the Motion Library (Motion Library window)
- Align multiple structures using jFATCAT
- Phage display assembly, translation and prediction – This is an integrated workflow that begins in SeqMan Ultra, continues through SeqNinja and MegAlign Pro, and ends in Protean 3D. Both NovaFold Antibody and NovaDock play a role in this workflow.

Copyright DNASTAR, Inc. 2022

Protean 3D Interface

Protean 3D's views, panels and tabs (pictured and described below) can be resized and/or undocked and moved anywhere inside or outside the Protean 3D application window. Changes and selections are synchronized between all views.



Item	Description
Projects	Multiple Protean 3D projects (documents) can be open at the same time. Click any project tab to make that document active.
Menus	**File, Edit, View, Modeling, Features, Search, Structure, Analysis, Report, Window,** and **Help**.
Views	Data display windows.
Panels	Dialogs for selecting data and changing options.
Motion Library	A searchable list of previews representing ~500 animated molecular structures from the RCSB Protein Data Bank (PDB).
Tool icons	Different tools are available depending on the active view.

To learn how to optimize the layout and contents of the Protean 3D window, see Customize the look and layout.

Copyright DNASTAR, Inc. 2022

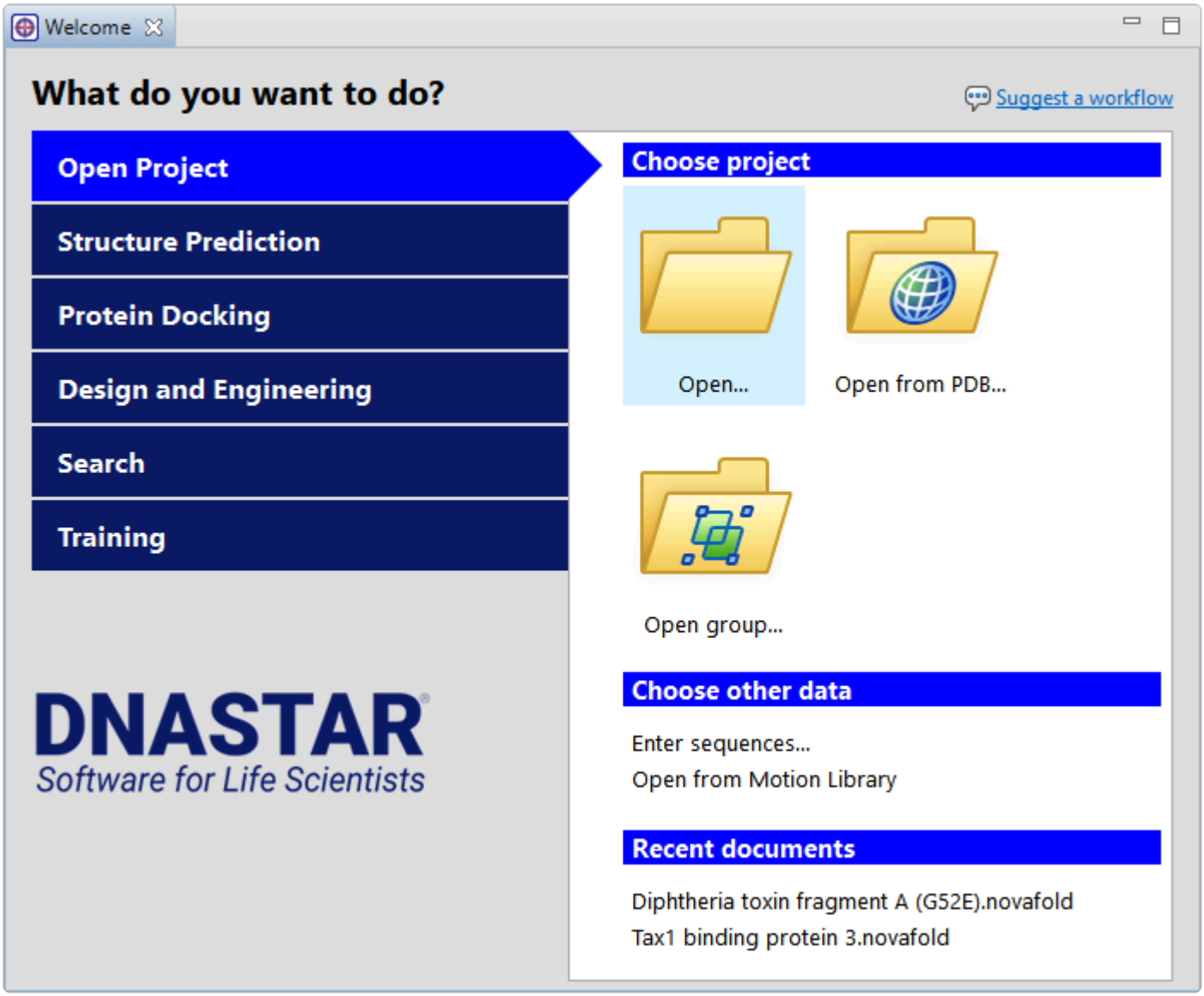
Note that the appearance of the Protean 3D interface changes depending on the file type being viewed:

File type	Appearance of Protean 3D
PDB or other structure file, including Protean 3D (*.structure*) format	All views and panels are enabled.
Sequence file	The Structure view is absent. The Molecules section displays the file name and a list of sequence or locus names, if available.
NovaFold (*.novafold*) structure prediction	When viewing a *.novafold* file, the only available views and panels are the NovaFold Report, Analysis view, Features panel and Tracks panel. To enable all Protean 3D functionality for a predicted structure, go to the Report view and click the **Open model in new document link**.
NovaFold Antibody (*.antibody*) structure prediction	All views and panels are enabled. Light and heavy chain annotations are displayed in the Style section in the **S-S Bond** group.
NovaDock (*.novadock*) protein docking prediction	When viewing a *.novadock* file, the only available view is the NovaDock Report. No panels are enabled.

Copyright DNASTAR, Inc. 2022

Welcome screen

The Welcome “project” opens when you launch Protean 3D and is a central location for opening protein projects, searching NCBI databases, getting help, and much more. As you add projects, the Welcome “project” will remain as the leftmost tab unless you manually move or close it. We recommend leaving the Welcome tab in place, since it has many time-saving shortcuts to useful commands.



Open Project	**Open**	Shortcut to **File > Open**. See Open one or more local files.
	Open from PDB	Shortcut to **File > Open from PDB**. See Open a PDB file when you know the ID.
	Open group	Shortcut to **File > Open Group**. See Open one or more local files.
	Enter sequences	Shortcut to **File > Enter Sequences** See .

Copyright DNASTAR, Inc. 2022

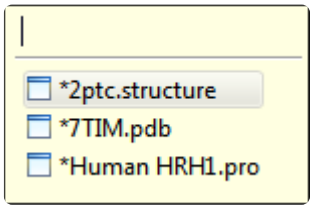
	Open from Motion Library	Shortcut to **File > Open from Motion Library**. See Open a structure from the Motion Library.
	Recent documents	Shortcut to **File > Recent Documents**.
Structure Prediction	**New protein structure with NovaFold**	Shortcut to **Modeling > Protein Structure Prediction > New NovaFold Prediction**. See Run NovaFold through Protean 3D.
	New protein structure with NovaFold AI.	Shortcut to **Modeling > Protein Structure Prediction > New NovaFold AI Prediction**. See Run NovaFold AI through Protean 3D.
	New antibody structure with NovaFold Antibody	Shortcut to **Modeling > Antibody Structure Prediction > New NovaFold Antibody Prediction**. See Run NovaFold Antibody through Protean 3D.
Protein Docking	**New protein-protein complex with NovaDock**	Shortcut to **Modeling > Protein-Protein Docking > New NovaDock Prediction**. See Run NovaDock through Protean 3D.
Design and Engineering	**Create a specific variant**	Shortcut to **Modeling > Protein Design > Create a specific variant**. See Create or Find Variants Using Protein Design.
	Scan for hot spots that provide folding stability	Shortcut to **Modeling > Protein Design > Scan for hot spots that provide folding stability**. See Create or Find Variants Using Protein Design.
Search	**Search for sequences**	Shortcut to **Search > Search for Sequences**. See Search for sequence matches.
	Search for text in NCBI Nucleotide	Shortcut to **Search > Search for Text In > NCBI Nucleotide**. See Search for text or accession number matches.
	Search for text in NCBI Protein	Shortcut to **Search > Search for Text In > NCBI Protein**. See Search for text or accession number matches.
Training	**Protean 3D help**	Launches this User Guide at the first topic, Welcome to Protean 3D.
	Lasergene tutorials	Launches the DNASTAR website at the Training page.
	(List of workflows)	Launches this User Guide at the selected workflow topic.

Copyright DNASTAR, Inc. 2022

Project tabs

You can open multiple projects within a single instance of Protean 3D. Each project is displayed in a separate tab and is initially assigned a unique name (*Untitled[#].structure*).

Tasks pertaining to Protean 3D project tabs:

Task	How To
To make a project the active project	Click on its project tab.
To open an additional Protean 3D window	Select the **Window > New Window** command.
To open a project chooser	Press **Ctrl+;** (Win) or **Cmd+;** (Mac). Alternatively, press the **>>** icon that appears to the right of the project tabs. This icon is displayed when multiple projects are open and the Protean 3D window is too small to display each one in an individual tab. The project chooser appears as a yellow popup with a line for typing in text at the top, and a list of open projects below.  Use your mouse to click on the project you want to make active. If you do not have a mouse, make a selection using the keyboard arrow keys, then press **Enter**. To instead search for a project using a filter, type text into the text field and press **Enter**. The text field supports the “?” and “*” wildcards.

Copyright DNASTAR, Inc. 2022

Protean 3D toolbar

The toolbar is located below the Protean 3D menus, and can be shown or hidden using the **Controls > Show/Hide Toolbar** command. The first two tools on the left of the toolbar affect the entire Protean 3D project, while the remaining tools affect only the Structure view. See that topic for a video demonstrating those tools.

The Structure view tools can either be “sticky” or “transient,” as color-coded in the table below.

- **Sticky** – Protean 3D stays in the selected mode until you choose a different function.
- **Transient** – The selected mode only stays active as long as you are actually using the controls.

Task	How To
To open a file	Press the **Open a local file** tool ( ). For information on the various options, see Open a file or project.
To save the document	Press the **Save to a file** tool ( ), use **File > Save**, or press **Ctrl/Cmd+S**. The document is saved in *.structure* file format.
To rotate the molecule in any direction	Use the **Rotate** tool ( ), press **Structure > Rotate**, or hold down the left mouse button and move the mouse in any direction.
To rotate the structure around the z-axis	Use the **Rotate Around Z-Axis** tool ( ), use **Structure > Rotate Around Z-axis** press **Shift+click**, or hold down the left mouse button and move the mouse right to rotate in a clockwise direction; or left to rotate in a counter-clockwise (anti-clockwise) direction.
To move the molecule relative to the background	Use the **Move** tool , use **Structure > Move**, press **Alt+click** (Win) or **+click** (Mac), or hold down the left mouse button and drag the mouse in any direction to move the structure the same way.
Note for the three rows above: If the molecule has 20,000 or more atoms, the side chains will temporarily disappear in the Structure view when you rotate or move the molecule. This circumvents slow-downs that would occur if the completely-rendered molecule was refreshed throughout the rotation or move. The side chains will appear again as soon as you stop rotating or moving the molecule.	
To zoom in or out	Use the **Zoom** tool ( ), use **Structure > Zoom**, use the mouse’s scroll wheel, or hold down the left mouse button and rag the mouse up to zoom in and down to zoom out.
To select neighboring residues	Use the **Select Neighbors** tool ( ), use **Structure > Select Neighbors**, or

Copyright DNASTAR, Inc. 2022

	press **Shift+double-click**.
To move a selection to the center of the screen and make it the center of rotation	Use the **Center on Selection** tool ( ), use **Structure > Center on Selection**, or press **Alt+double-click**. **Alt + double-click** both selects an item and centers the selection on it.
To re-center and resize the model to fit in the view	Use the **Reset Structure View** tool ( ), use **Structure > Reset Structure View**, or press **Ctrl/Cmd+Shift+R**.

Copyright DNASTAR, Inc. 2022

Views

The majority of the Protean 3D window is comprised of one or more views. To learn what can be observed and done within each view, click the links below.

General views:

- Structure – Displays a physical representation of the molecule.
- Analysis – Shows a graphical representation of all prediction tracks that have been applied to the active sequence using the Tracks panel .
- Sequence – Shows the amino acid or nucleotide sequence, secondary structures, binding sites and other features.
- Report

NovaCloud-related views:

- NovaCloud Services view – After launching NovaCloud Services, the Services view is used to log in and launch a setup wizard for NovaFold, NovaFold Antibody, or NovaDock predictions.
- Predictions view – Once a prediction has been started on NovaCloud Services, you can monitor its progress and download its eventual output from the Predictions view.
- Report (Composition, Biophysical Properties, Protein Design, NovaFold and NovaDock results)

Online search views:

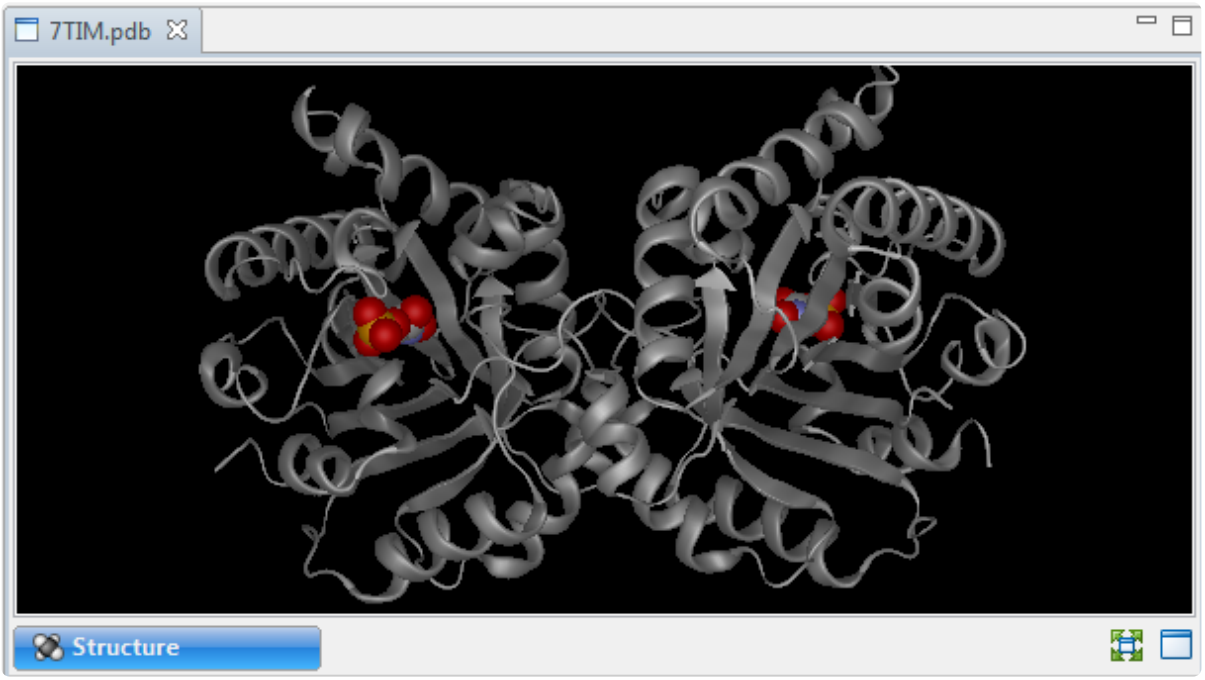
- Text
- Table
- Pairwise

To learn how to change the layout or contents of a view, see Customize the look and layout.

Copyright DNASTAR, Inc. 2022

Structure view

The Structure view displays a physical representation of the molecule in your structure file. By default, the Structure view is shown in the upper left of the Protean 3D window. If the Structure view is hidden or absent, you can display it using **View > Analysis > Show**. Alternatively, you can reset the entire Protean 3D window to the default layout using **View > Document Layout > Restore DNASTAR Default Layout**.



The Structure view can be rotated, zoomed, moved relative to the background, or seen in full-screen mode. Selections within this view are synchronized with other views in Protean 3D. Similarly, selections made elsewhere are highlighted simultaneously in the Structure view.

> **Note:** If the active document is a sequence file, rather than a structure, the Structure view is inaccessible.

Tasks that pertain to the Structure view:

Task	How To
To move, rotate or zoom in on the view	Use the Protean 3D toolbar in the upper left. The following video demonstrates some of the tools available from the Structure view:

Copyright DNASTAR, Inc. 2022

To make selections in the view	See Make a selection.
To change the layout of the view or what is displayed in it	See Customize the look and layout.
To change the rendering of the structure	Use the Rendering section of the Style panel.
To change the colors	Use the Color section of the Style panel.
To change the lighting	Use the Lighting section of the Style panel.
To add a molecular or solvent-accessible surface to the structure	Use the Surfaces section of the Explorer panel. Once a surface has been created, it can be controlled through the Rendering section of the Style panel.
To create a duplicate of a selected portion of a chain	Use **Structure > Duplicate Selection**. The duplicate chain appears in the Explorer and Features panels with the original chain name followed by a number in parentheses, e.g., *Chain B (1)*.

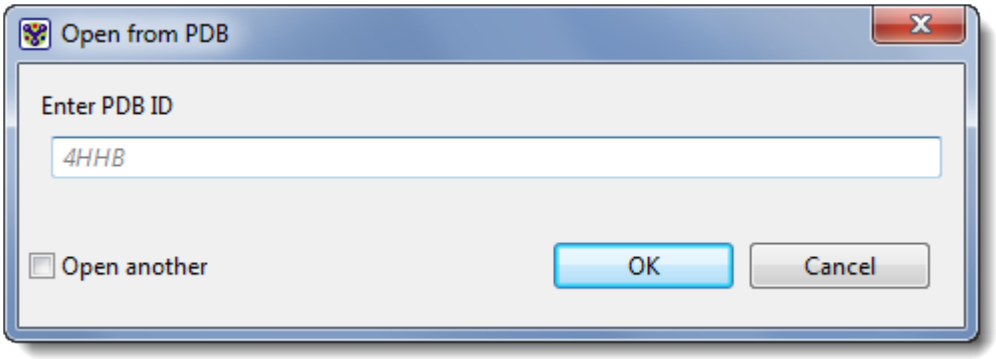
Copyright DNASTAR, Inc. 2022

Try it! – Use the Structure view toolbar

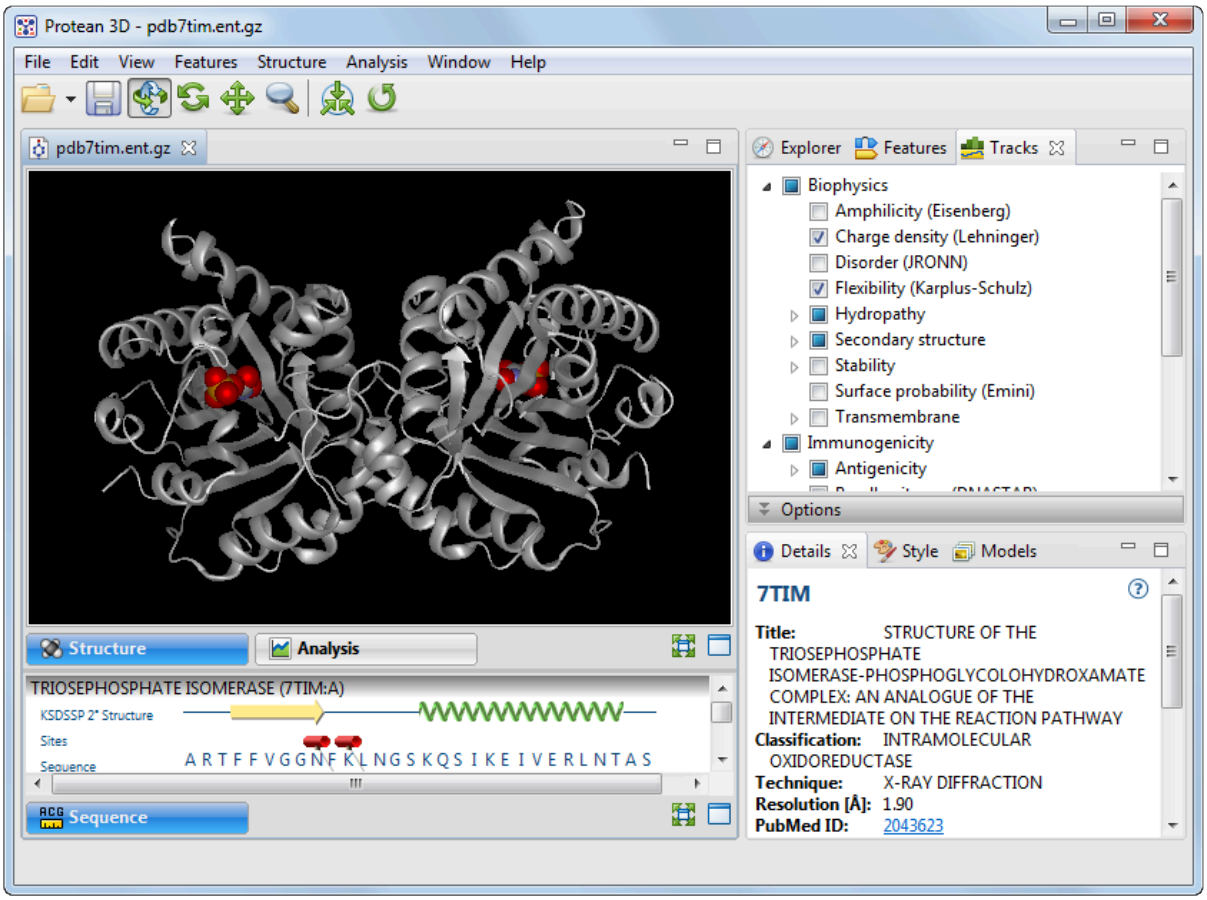
This tutorial provides practice using the toolbar from within the Structure view.

1. Within Protean 3D, choose **File > Open from PDB**. You may initially see a yellow warning icon ( ) prompting you to type in a valid ID.

2. Type **7TIM** into the text box.



3. Press **OK** to open the structure in Protean 3D.



4. In the Structure view, zoom in and out on the structure by clicking the **Zoom** tool ( ) and dragging the mouse up and down. (If you don't see the zoom icon, use the **View > Show Toolbar** command to restore the toolbar. Another alternative is to use the commands in the **Structure** menu instead of the toolbar icons.)

5. Slide the molecule partially out of view and back again by clicking the **Move** tool ( ) and dragging the mouse in any direction.

6. Rotate the molecule by clicking the **Rotate** tool ( ) and dragging the mouse in any direction.

7. To return the structure to its original orientation, click the **Reset Structure View** tool ( ).

Copyright DNASTAR, Inc. 2022

Default colors and styles in the Structure view

The tables below shows the default colors and styles of various entities in the Structure view.

- To select a different color for a selection or a different set of colors for the entire molecule, use the Color section of the Style panel.
- To select a different rendering style, use the Rendering section of the Style panel.

Element	Abbrev.	Default Color
Hydrogen	H	White
Carbon	C	Light Gray
Nitrogen	N	Lilac
Oxygen	O	Red
Phosphorus	P	Orange
Sulfur	S	Yellow
Zinc	Zn	Dark Red
Bromine	Br	Dark Red
Iron	Fe	Orange
Boron	B	Tan
Beryllium	Be	Dark Green
Magnesium	Mg	Dark Green
Strontium	Sr	Dark Green
Barium	Ba	Dark Green
Radon	Ra	Dark Green
Fluorine	F	Green
Chlorine	Cl	Light Green
Helium	He	Aqua
Neon	Ne	Aqua
Argon	Ar	Aqua
Krypton	Kr	Aqua
Xenon	Xe	Aqua
Sodium	Na	Blue
Lithium	Li	Purple
Rubidium	Rb	Purple
Cesium	Cs	Purple
Iodine	I	Violet
Titanium	Ti	Gray
Calcium	Ca	Dark Gray
Potassium	K	Dark Gray
Others		Pink

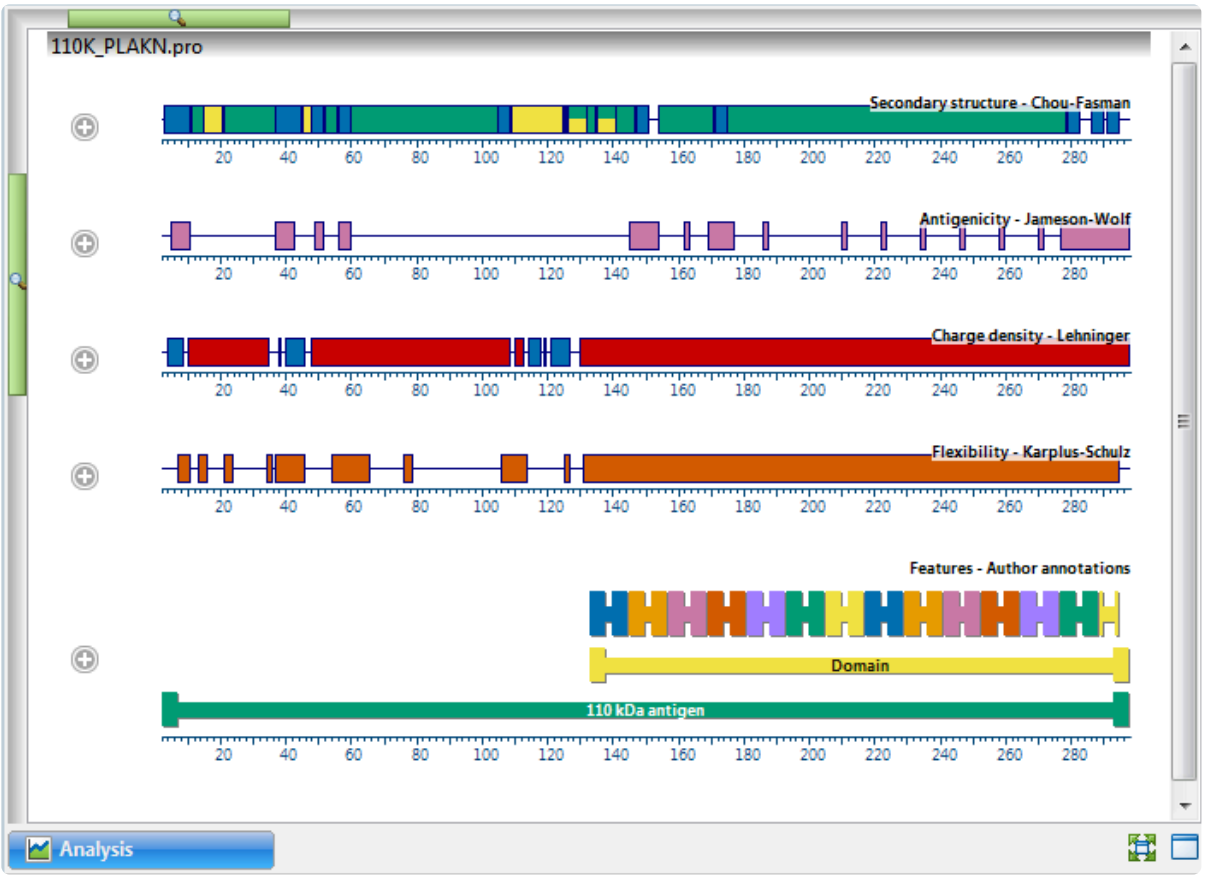
Copyright DNASTAR, Inc. 2022

Component	Default Rendering Style
Backbone	Ribbon (cartoon)
Side chains	Invisible
Ligands and water	Spheres

Copyright DNASTAR, Inc. 2022

Analysis view

The Analysis view lets you quickly find regions of interest on a protein sequence. By default, the Analysis view is shown in the bottom left of the Protean 3D window. If the Analysis view is hidden or absent, you can display it using **View > Analysis > Show**. Alternatively, you can reset the entire Protean 3D window to the default layout using **View > Document Layout > Restore DNASTAR Default Layout**.



Once a sequence has been opened, the Analysis view shows a default selection of “prediction tracks” that can be used to evaluate secondary structural characteristics of a sequence. You can add more tracks to the Analysis view or hide unwanted tracks using the Tracks panel.

Selected features can be displayed along the same axis, if desired. Initially, a synopsis is displayed for each of the applied prediction tracks, while a consensus is displayed for the Feature/Annotation method. The synopsis or consensus is in the form of a colorful cartoon showing regional conclusions for the method or features.

Tracks are displayed as either line graphs or region plots, and all results are displayed on a common horizontal scale. Line graphs and region plots are two ways of visualizing the results of the same method. Line graphs are composed of a series of connected points, calculated at each residue and averaged over the window you specify. Region plots use the calculated data to make a yes/no decision, drawing a bar to represent a “yes” answer. The threshold for calling a region a distinct type may be adjusted when editing track options.

Copyright DNASTAR, Inc. 2022

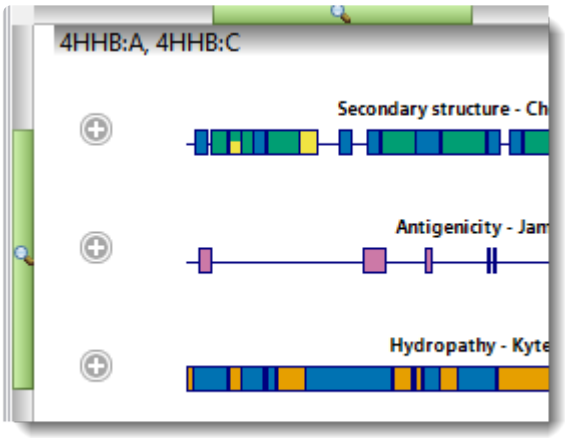
Columns represent clusters of identical sequences, while rows represent method calculations.

The following video is a brief introduction to the Analysis view:

Tasks that pertain to the Analysis view:

Task	How To
To open individual displays showing the “channels” or components that contributed to the synopsis	Click on the “plus sign” to the left of each synopsis or consensus.
To zoom in/out	Grab the green horizontal or vertical slider with your mouse and drag left/right or up/down to zoom in or out in the view.

Copyright DNASTAR, Inc. 2022

	 Alternatively, you can use any of the four **Analysis > Zoom** commands or use *Ctrl/Cmd+Shift+(arrow direction).
To expand or collapse a synopsis	Click the tree control on the left of the synopsis.
To change the amplitude of line graphs	Use **Analysis > Plot in Peak Scale** to set the amplitude of each line graph in the Analysis view to its own maximum range. Or use **Analysis > Plot in Constant Scale** to set the amplitude of each line graph in the Analysis view to the maximum range for the parent method. The latter is useful for comparing method results for two different assays.
To edit method options	See Track Options section.
To rearrange tracks	Select a synopsis. Drag the row and drop it in the desired position. All channels within the synopsis will move as a single group.
To collapse all channels on the view	Use **Analysis > Collapse All Rows**. This causes only the parent level synopses for each applied method to be displayed.
To copy a selection from the view	Right-click on the selection and choose **Copy**.
To create a new feature	Select the range for the feature, then right-click on the selection and choose **New Feature**.
To delete a feature	Right-click in the feature and choose **Delete Feature**.
To show/hide individual features	Check or uncheck them in the Features panel.
To create a virtual mutation within the current selection	Right-click on the selection and choose **Create Variants**. The Protein Design wizard opens with the Substitutes screen already populated for the selected range.
To make selections in the view	See Make a selection. To select everything except the current range, right-click on the selection and choose **Invert Selection**.

Copyright DNASTAR, Inc. 2022

To remove chains/ sequences from all synopses	Uncheck the corresponding box(es) in the Molecules section.
To change the color for a channel display	Select a channel (not a synopsis) and change the **Fill** color in the Color section. If you make a change to one item (e.g. an alpha helix), the color will change for all items in the view that share that class.
To toggle between displaying or hiding a ruler under each synopsis	Use the **Analysis > Show Rulers** or **Analysis > Hide Rulers** command or the **Ctrl/ Cmd+** keyboard shortcut. Rulers display major ticks every ‘n’ residues, where ‘n’ is a multiple of 10 and varies depending upon the zoom level. Ruler numbering for sequences or synopses always starts at “1.” Numbering for select chains uses the position labels defined in the structure file.
To change the layout of the view or what is displayed in it	See Customize the look and layout.
To view tooltips (information balloons) about a feature or other item	Hover over it.
To restore the default display	Use **Analysis > Reset Analysis View**. This reverts to the default horizontal and vertical zoom level for the current window size, and collapses all rows.

Copyright DNASTAR, Inc. 2022

Try it! – Apply tracks to the Analysis view

Protean 3D's Analysis View allows you to apply a variety of prediction tracks to your protein sequence. In this four-part tutorial, you will locate potential calcium-binding regions for a human calmodulin sequence. We know from the literature that these regions are correlated with predicted turn regions and with flexible regions.

- Part A: Apply and reorder tracks
- Part B: Compare results from two tracks
- Part C: Edit track options
- Part D: Recolor tracks

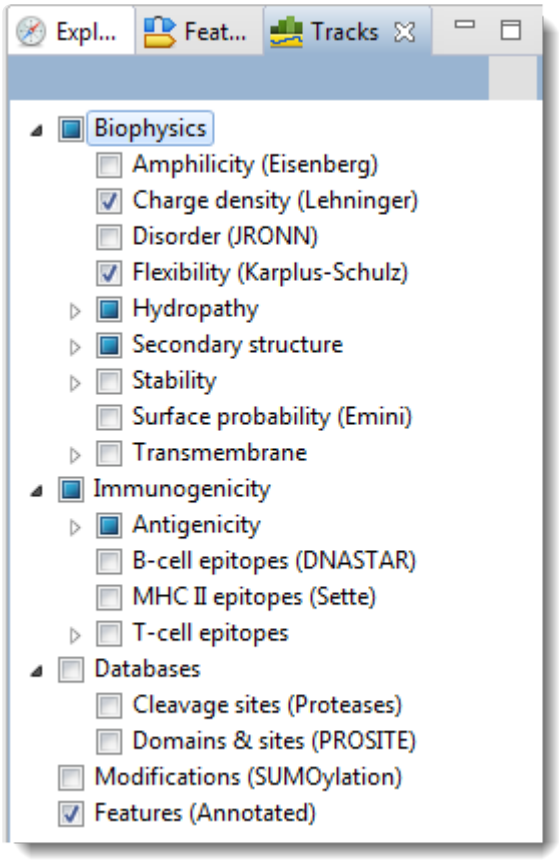
Copyright DNASTAR, Inc. 2022

Part A: Apply and reorder tracks

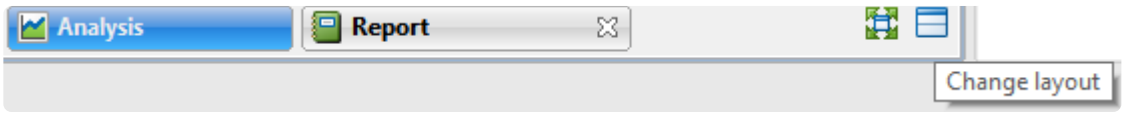
1. If you have not yet downloaded and extracted the tutorial data, click here to download it. Then decompress (unzip) the file archive using the method of your choice.

2. In Protean 3D, use **File > Open** to open the protein sequence file *Protean 3D Data/Human Calmodulin.pro*. The Analysis view opens with some tracks already applied by default.

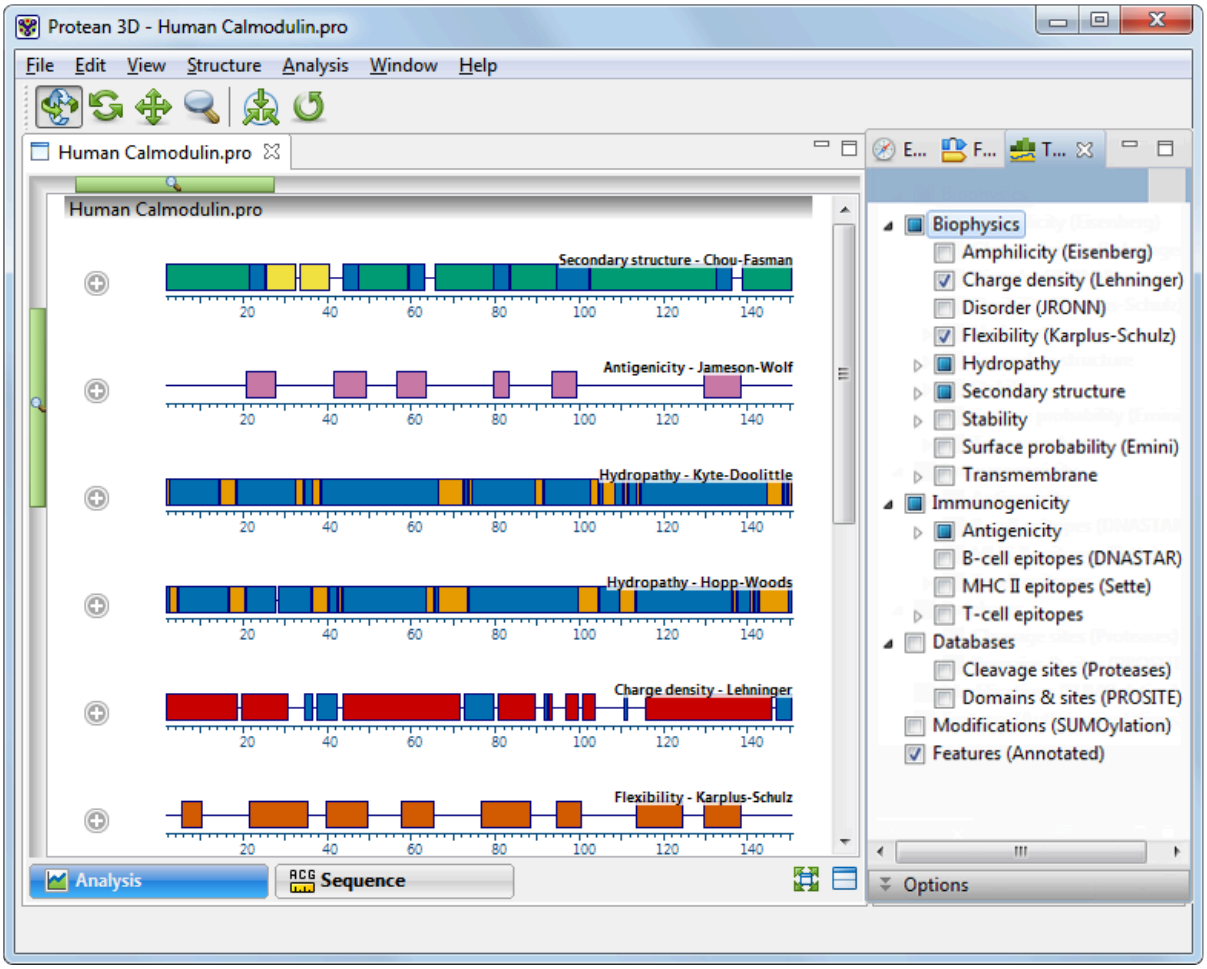
3. To see a list of the applied tracks, open the Tracks panel using **View > Tracks > Tracks**.



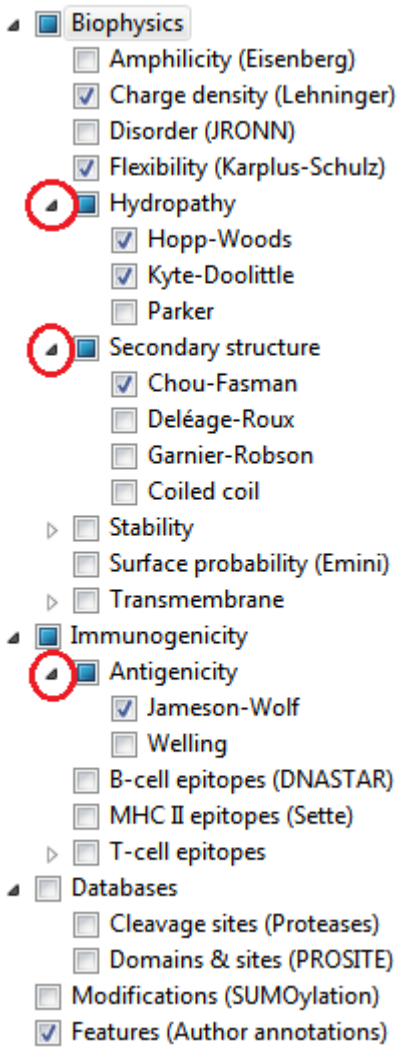
4. Maximize the Analysis view by clicking its **Change layout** tool and choosing **One** ( ).



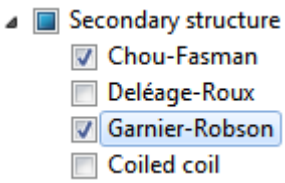
The Protean 3D document now appears as follows:



5. In the Tracks panel, click the expand arrows to the left of some tracks to reveal their sub-tracks:

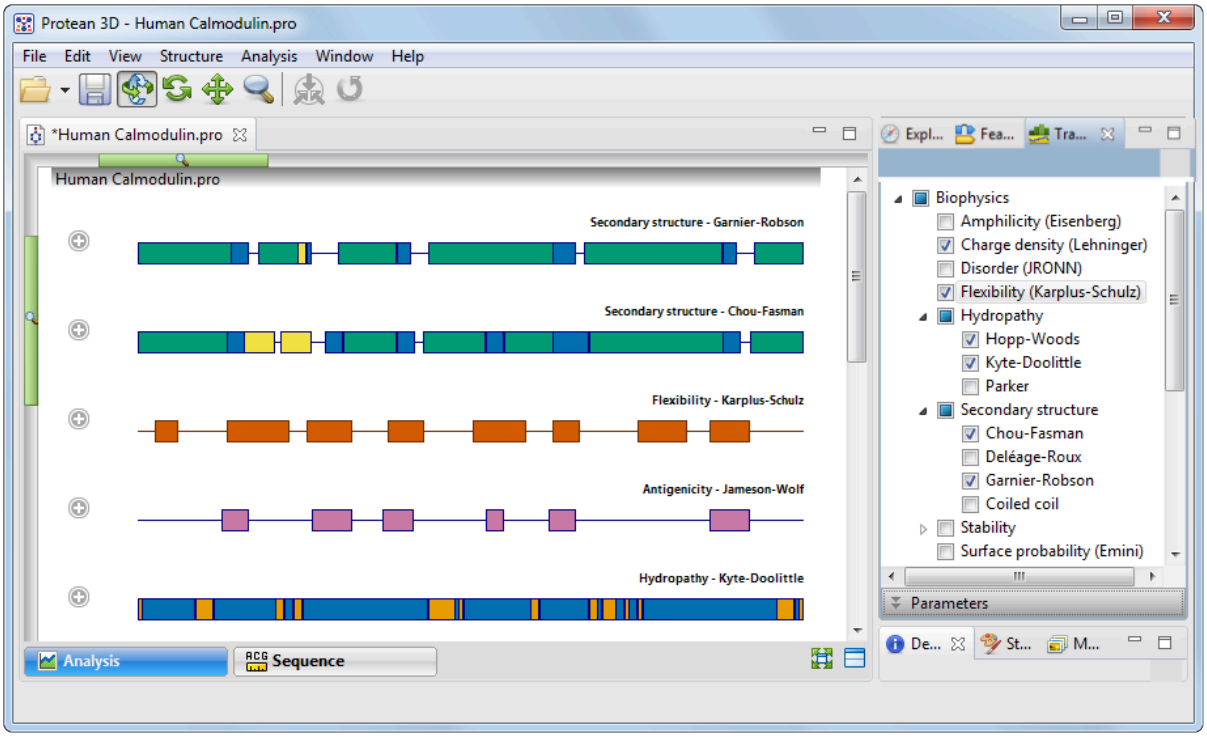


6. Add the Garnier-Robson track to the Analysis view by checking the box to its left.



7. Reorder the tracks so the first three tracks are Garnier-Robson, Chou-Fasman and Flexibility-Karplus-Schulz. This is done within the Analysis view by double-clicking on each track synopsis, and then dragging & dropping it at the desired location.

Copyright DNASTAR, Inc. 2022

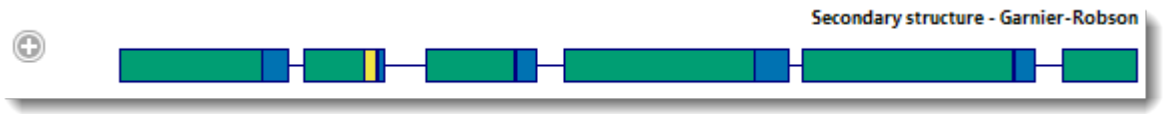


Proceed to [Part B: Compare results from two tracks](#).

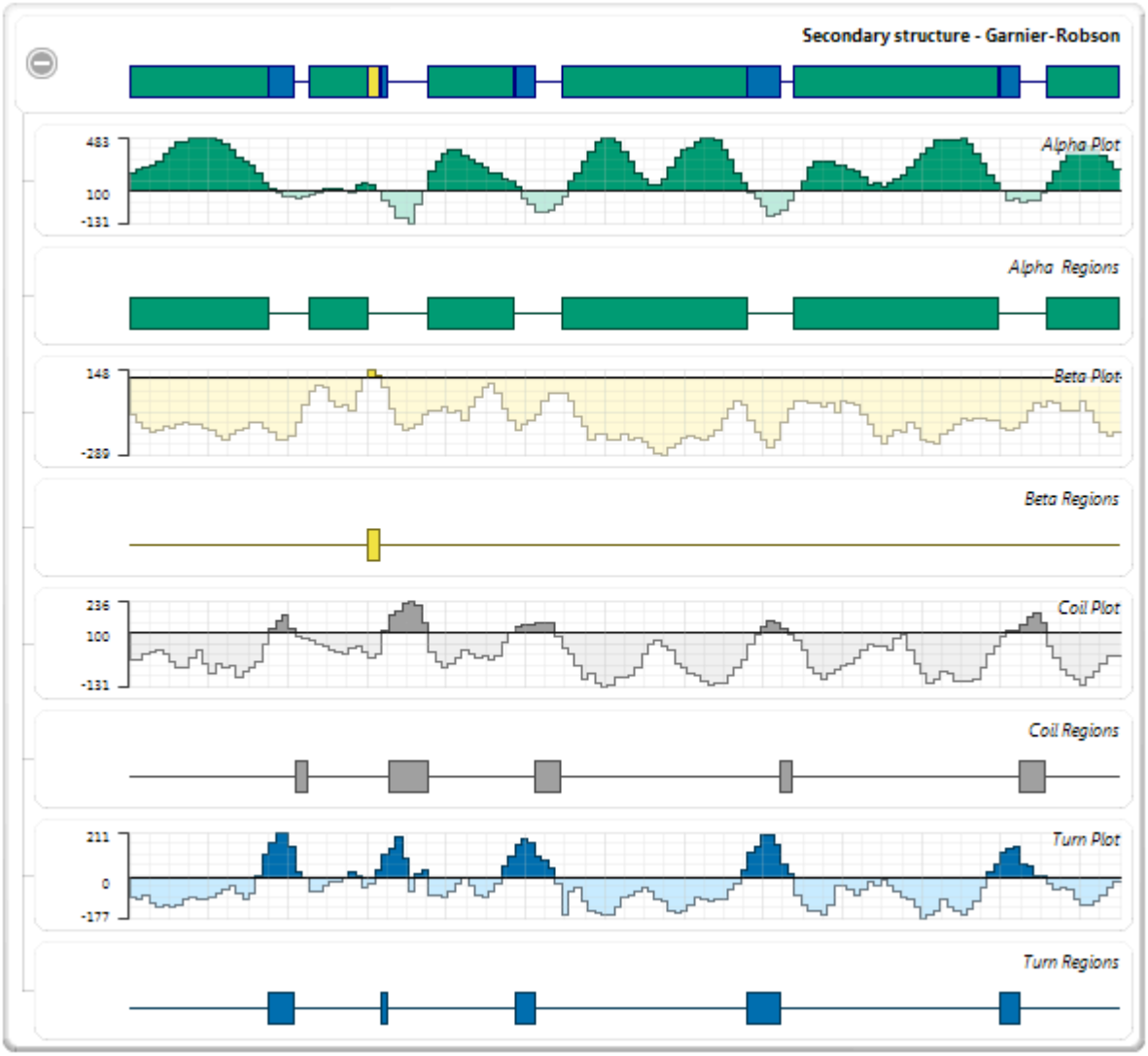
Part B: Compare results from two tracks

This part of the tutorial follows Part A: Apply and reorder tracks. Part B demonstrates how the results of one track can be corroborated by looking at a similar graph from another track.

So far, only the synopsis row for each track is visible in the Analysis view.

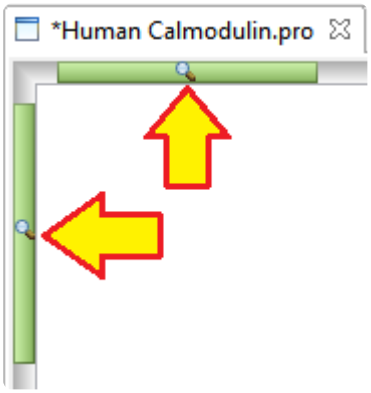


1. To view the individual graphs, click the plus signs ( ) to the left of the first three synopses. The individual track graphs are now displayed below each synopsis. The name of each graph appears on the right of the Analysis View.

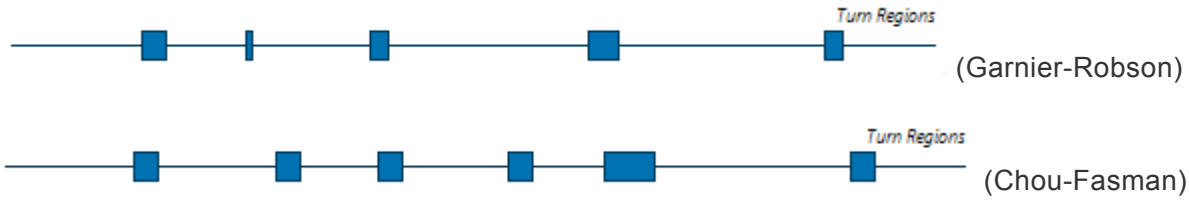


2. To rescale the display so that you can view it more comfortably, use the zoom sliders located on the

top and left sides of the Analysis view.



3. Examine the Garnier-Robson and Chou-Fasman “Turn Region” plots. These plots show the predicted turn regions, and thus, the probable calcium binding regions for the protein sequence. (The plots are shown as adjacent in the graphic below, but they are not actually adjacent in the Analysis view).

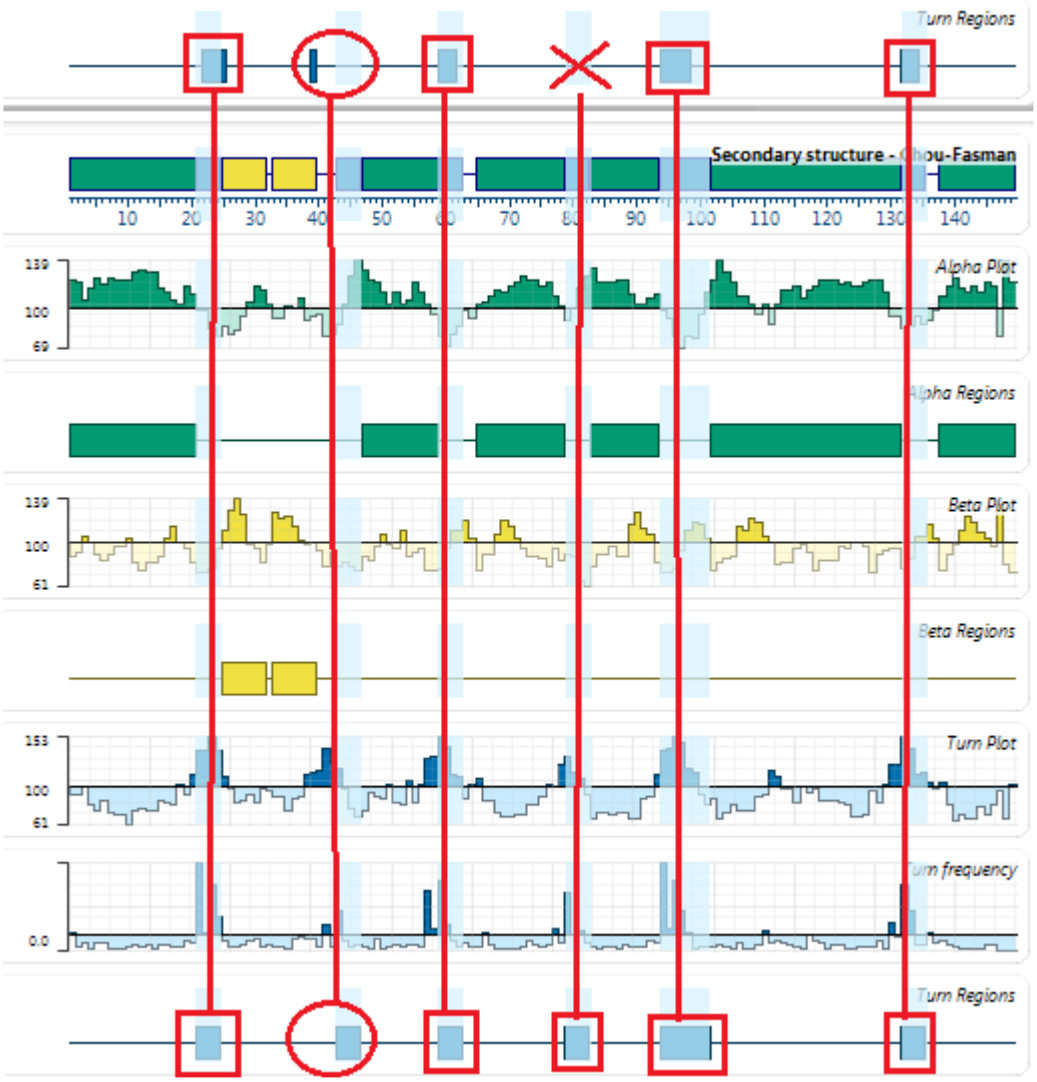


4. To verify which turn regions are the same in both tracks, select the Chou Fasman “Turn Regions” by holding down the **Ctrl** or **Cmd** key while clicking once on each of the six blue boxes.

5. Visually follow the highlighting upwards to see corresponding turn regions highlighted in the Garnier-Robson track.

   Of the six turn regions found in Chou-Fasman, four are nearly exact matches (shown by red boxes, below) with the turn regions in Garnier-Robson, and another is a close match (shown by red circles, below). Thus, the Garnier-Robson track provides partial corroboration of the turn regions from the Chou-Fasman track.

Copyright DNASTAR, Inc. 2022



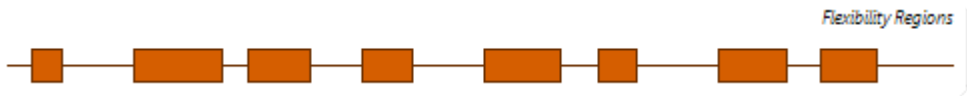
Proceed to Part C: Edit Track Options.

Part C: Edit track options

This tutorial follows Part B: Compare results from two tracks.

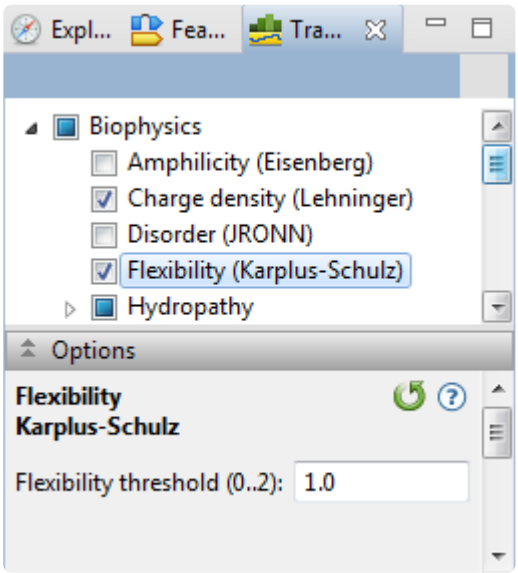
1. Scroll down the Analysis view to the Flexibility-Karplus-Schulz track. Click on the plus sign to reveal two sub-plots called **Flexibility Regions** and **Flexibility Plot**.

   Note that under the current (default) options, there are 8 boxes in **Flexibility Regions** compared to 6 for Chou-Fasman "Turn Regions":

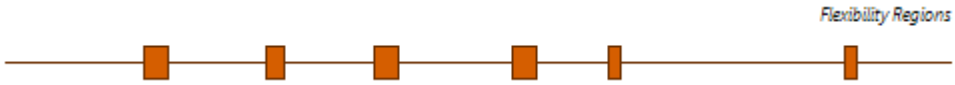
   By making the options more stringent, we can find which of these areas are most likely to be flexibility regions.

2. Open the options for the Flexibility (Karplus-Schulz) track by going to the Tracks panel and clicking on the words "Flexibility (Karplus-Schulz)." Any editable options are displayed in the Track Options section:

3. Change the **Flexibility threshold** to "1.07."

   Note that the Flexibility Regions graph now corresponds almost exactly with Chou-Fasman Turn Regions selected in Part B. This provides additional confirmation that these regions may correspond to calcium binding regions.

Copyright DNASTAR, Inc. 2022

4. Return to the default options by clicking the **Restore Default Values** tool ( ).

5. Remove the Flexibility (Karplus-Schulz) track from the Analysis view by removing its checkmark in the Tracks panel.

Proceed to Part D: Recolor tracks.

Copyright DNASTAR, Inc. 2022

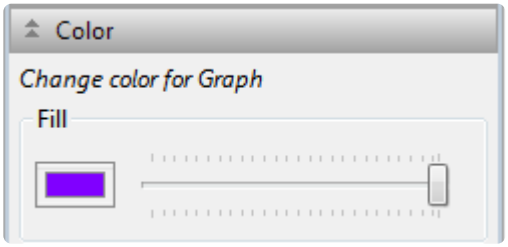
Part D: Recolor tracks

This part of the tutorial follows Part C: Edit track options. In Part D, you will learn how to recolor a graph track in the Analysis view.

1. Double-click on any green **Alpha Regions** graph to select it.

2. If it is not already open, open the Color section by clicking on it. Alternatively, use **View > Style > Color**.

3. Click on the green box under **Fill**, and choose bright purple from the color picker.



Since the selected graph is of the “alpha helix” type, all graphs of the same type will be simultaneously re-colored bright purple, including graphs in other tracks.

Copyright DNASTAR, Inc. 2022

If many individual track graphs are displayed at the same time, the Analysis view can become crowded. To instantly close all of the individual graphs, and display only the synopsis rows, use the **Analysis > Collapse All Rows** command.

This marks the end of the tutorial Apply tracks to the Analysis view.

Colors in the Analysis view

Default colors in the Analysis view are assigned by track classification and channel type:

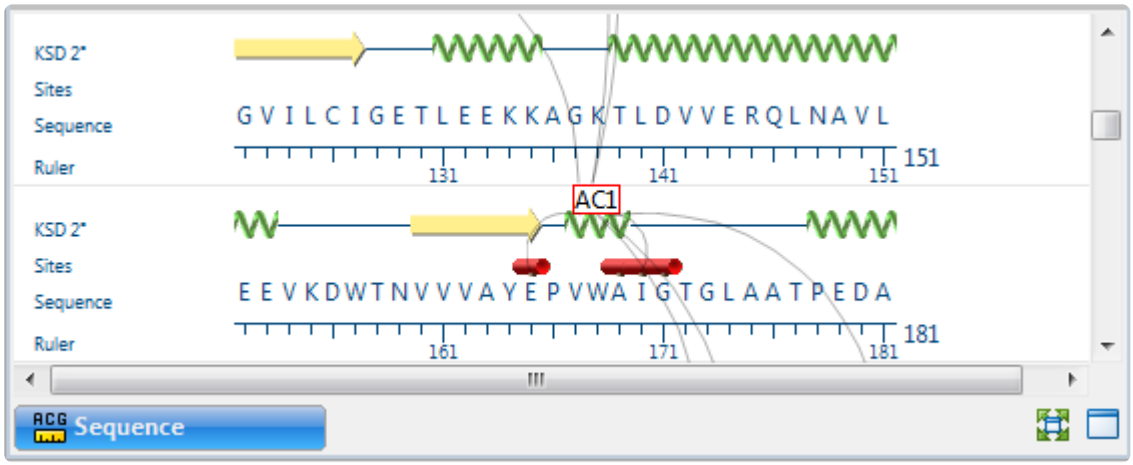
Method Class	Channel	Color
Secondary Structure	Helix	Aquamarine
Secondary Structure	Sheet	Yellow
Secondary Structure	Beta turn	Blue
Secondary Structure	Coil	None
Secondary Structure	Coiled Coil	Orange
Charge Density	Positive (basic/amine)	Blue
Charge Density	Negative (acidic/carboxylic)	Red
Hydrophobicity	Hydrophobic (oily)	Orange
Hydrophobicity	Hydrophilic (water)	Blue
Surface Probability	Index	Purple
Flexibility	Index	Vermilion
Disorder	Index	Vermilion
Antigenicity	Index	Red-purple
Stability	Stable	Gray
Stability	Unstable	Vermilion

Copyright DNASTAR, Inc. 2022

Sequence view

The Sequence view displays the amino acid or nucleotide sequence, secondary structures, binding sites, disulfide bonds, and a ruler. Colorful graphics denote features derived from annotations in the structure file.

By default, the Sequence view is shown vertically in the middle of the Protean 3D window. If the Sequence view is hidden or absent, you can display it using **View > Sequence > Show**. Alternatively, you can reset the entire Protean 3D window to the default layout using **View > Document Layout > Restore DNASTAR Default Layout**.



Selections within this view are synchronized with other views in Protean 3D. Similarly, selections made elsewhere will be highlighted simultaneously in the Sequence view.

Descriptions of graphics in the Sequence view:

	Protein sequence residues and the ruler.
	Structural gaps are displayed as grayed-out residues. If a structural gap comprises an entire row, no ruler will appear under it, since PDB files do not contain numbering information for gaps. A ruler is shown, however, if the gap occurs in the same row as regular sequence. In this case, the number to the right of the ruler for that row shows the position of the last residue before the gap.
	Helical structure.
	Site.
	Beta sheet secondary structure.

	Site hubs use the site identifier listed in the PDB file. This 3-letter code is the same one that is used in the Style. The lines between hubs indicate the locations where bonds occur.
	Disulfide bond residues.

Tasks that pertain to the Sequence view:

Task	How To
To change the font style or size	Use the **Font** menu in the Sequence section.
To change the order of rows	Click+drag* the row label (e.g. Sequence, KDS 2°, Sites, Ruler) to the desired position.
To create a new feature	Select the range for the feature, then right-click on the selection and choose **New Feature**.
To control the display of features	Use the checkboxes in the feature Style sectionto toggle between showing and hiding features. Selections may be made for individual features or for entire feature classes.
To delete a feature	Right-click in the feature and choose **Delete Feature**.
To make selections in the view	See Make a Selection. To select everything except the current range, right-click on the selection and choose **Invert Selection**.
To copy a selection from the view	Right-click on the selection and choose **Copy**.
To create a virtual mutation within the current selection	Right-click on the selection and choose **Create Variants**. The Protein Design wizard opens with the Substitutes screen already populated for the selected range.
To change the layout of the view or what is displayed in it	See Customize the look and layout.

Copyright DNASTAR, Inc. 2022

Report view

Protean 3D offers a variety of different Report views, depending on the structure or result (variant search, BLAST search) being viewed.

If the entire sequence or “nothing” is selected, the report applies to the entire protein. If an entire multi-chain molecule is selected, all chains are represented in the report. If a subset of the sequence is selected (a chain, feature, etc.), the report shows analysis information for only the selected portion.

Selecting a row in any Report view selects the same element in other views. Similarly, selections made elsewhere will be highlighted simultaneously in the Report view.

To learn more about specific report types and how to access them, see the topics below.

Document type	Report name
Protean 3D structure file (*.structure*) NovaFold Antibody result file (*.antibody*)	Composition, Biophysical Properties, Variants
Protein Design result file (*.structure*)	Composition, Biophysical Properties, Variants, Hot Spot Scan
NovaFold result file (*.novafold*)	NovaFold
NovaDock result file (*.novadock*)	NovaDock

Tool-related tasks in the Composition, Variants and Biophysical Properties views:

Task	How To
To open a new file	Choose an option from the the **Open a local file** ( ) drop-down menu.
To save the document in *.structure* file format	Press the **Save to a file** tool ( ).
To switch to another type of report	Use the drop-down menu in the view header.
To add and view a report not yet in the drop-down list	Use the **Add reports to this view** tool ( ) in the view header.
To delete the current report	Use the **Remove the current report** tool ( ) in the view header.
To export the report contents	Use the **Export** tool ( ) in the view header. Report contents can be saved in

Copyright DNASTAR, Inc. 2022

	Comma-separated values (*.csv*), Tab-separated values (*.tsv*) or Extensible table (*.xtf*) format.
To open a copy of the current report in a new Report tab	Use the **Clone this view** tool ( ) in the view header. After using this command, you may wish to move one of the views to a different location so you can compare multiple reports at the same time.
To show the active view in full-screen mode	Use the **View > Full Screen** command or the the **Full screen** tool at the bottom right corner of the view.
To exit full screen mode	Press the **Esc** key or use the **Exit full screen** tool) at the bottom of the full-screen view.
To configure the number and location of visible views.	Use the **View > Change layout** command or the **Change layout** tool (image varies) in the bottom right corner of the view. The following options are available: • **One** – One – To show only the current view. • **Two Horizontal** (Two Horizontal) – To show the current view on top and one additional view below it. • **Two Vertical** (Two Vertical) – To show the current view on the left, and one additional view to the right. • **Three Horizontal** (Three Horizontal) – To show the current view on top and two additional views below it. • **Three – Left** (Three - Left) – To show the current view on the left and two additional views stacked one on top of the other on the right. • **Three – Right** (Three - Right) – To show the current view on the right and two additional views stacked one on top of the other on the left. • **Three – Top** (Three - Top) – To show the current view on the top and two additional views side-by-side on the bottom. • **Three – Bottom** (Three - Bottom) – To show the current view on the bottom and two additional views side-by-side on the top.

Copyright DNASTAR, Inc. 2022

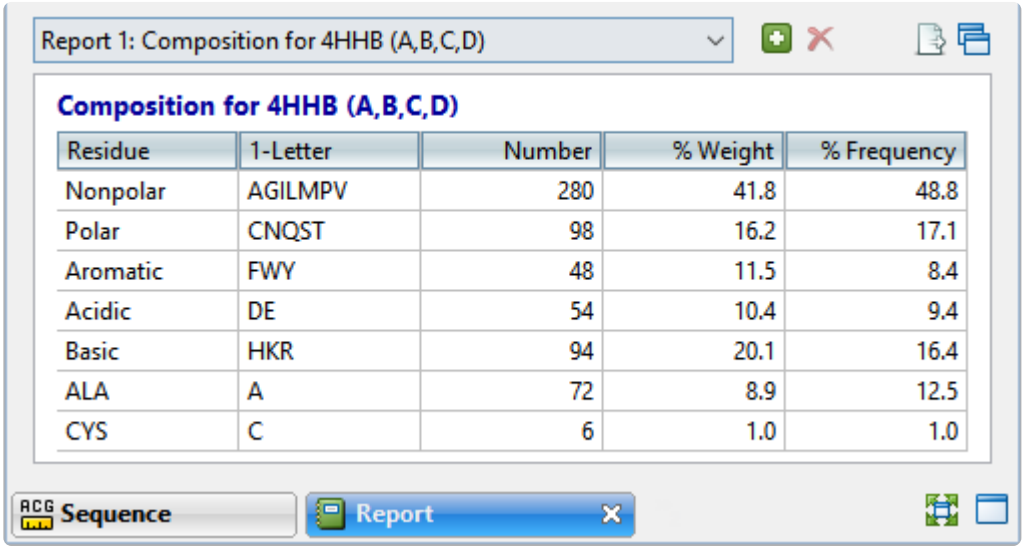
Composition report

The Composition report is the default type of Report view for most situations. The Composition report table provides information about each individual amino acid, as well as amino acid categories.

To open the Composition report, do either of the following:

- Use the **Report > Composition from Document/Selection/Structure (chain names)** command.
- Use the **View > Report > Show** command or press the **Report** tab (Report). If the Composition report is not current displayed, select it from the drop-down menu in the upper left corner of the Report view. If a Composition report is not in the menu list, add it using the **Add reports to this view** tool ( ).

An example Composition report is shown below:



Composition for 4HHB (A,B,C,D)

Residue	1-Letter	Number	% Weight	% Frequency
Nonpolar	AGILMPV	280	41.8	48.8
Polar	CNQST	98	16.2	17.1
Aromatic	FWY	48	11.5	8.4
Acidic	DE	54	10.4	9.4
Basic	HKR	94	20.1	16.4
ALA	A	72	8.9	12.5
CYS	C	6	1.0	1.0

Each column is described below:

Column name	Description
Residue	The amino acid code or category.
1-Letter	The 1-letter IUPAC amino acid code.
Number	The residue-level count.
% Weight	the percent weight of the category vs. the entire selection.
% Frequency	The percent frequency of observing a category.

For information about the tools available in this view, see Report view.

Copyright DNASTAR, Inc. 2022

Try it! – Study surface-exposed cysteines

Surface-exposed cysteine residues are of interest in industrial applications because they tend to form disulfide bridges with other molecules, disrupting the stability of the original protein.

The following example shows one way to study cysteines in a protein structure:

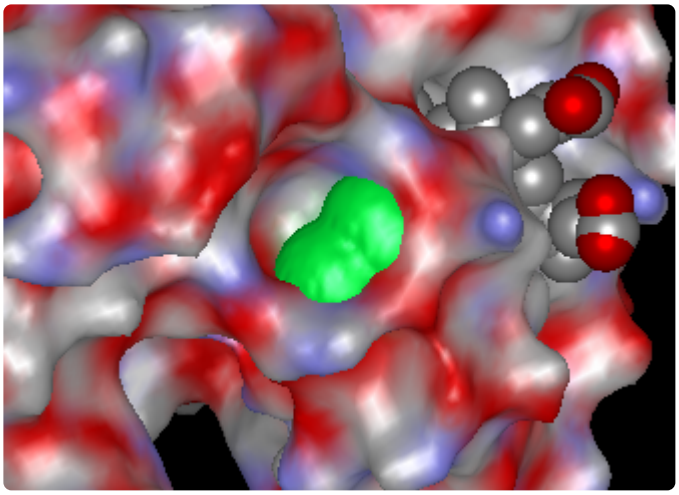
1. Launch Protean 3D and open any desired structure file.

2. Open the Composition report and select all cysteines by clicking on the CYS row.

3. Choose a **Fill** color for the cysteines (**View > Style > Color**), that will differentiate them from the colors used to show surfaces. Bright green, blue or yellow are good options.

4. Apply a surface (**View > Explorer > Surfaces**).

5. Rotate and visually examine the structure in the Structure view to detect the surface-exposed cysteines. In the image below, the cysteine residue is shown in green.



6. To confirm the visual analysis, select each cysteine residue in the Structure view and refer to the Details panel (**View > Details**) for relevant statistics.

Copyright DNASTAR, Inc. 2022

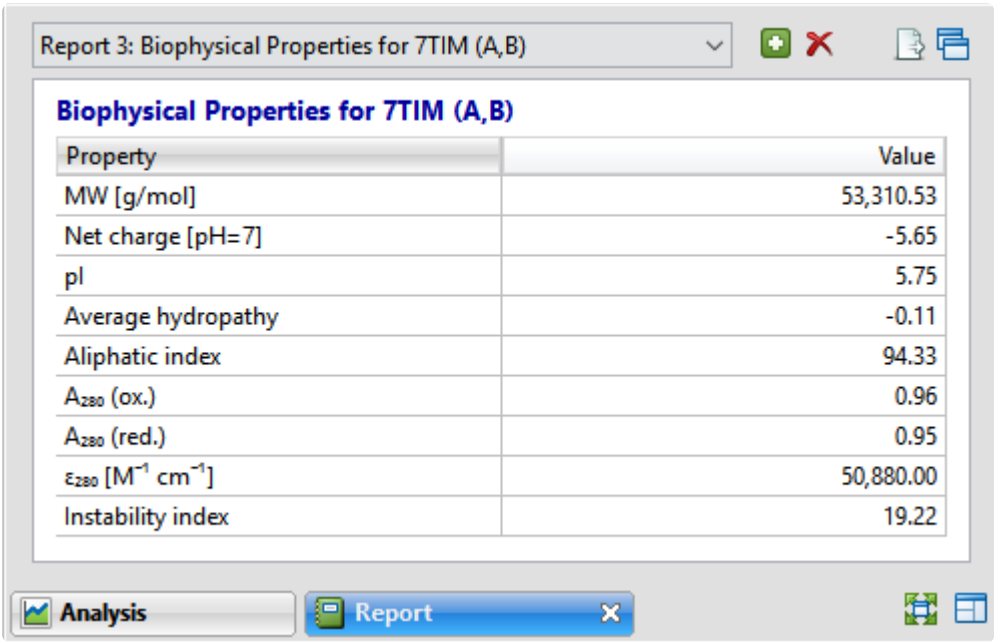
Biophysical Properties report

The Biophysical Properties report is an optional type of Report view that can be generated on demand.

To open the Biophysical Properties report, do either of the following:

- Use the **Report > Biophysical Properties from Document/Selection** command.
- Use the **View > Report > Show** command or press the **Report** tab (Report). If the Biophysical Properties report is not current displayed, select it from the drop-down menu in the upper left corner of the Report view. If a Biophysical Properties report is not in the menu list, add it using the **Add reports to this view** tool ( ).

An example Biophysical Properties report is shown below:



Property	Value
MW [g/mol]	53,310.53
Net charge [pH=7]	-5.65
pI	5.75
Average hydropathy	-0.11
Aliphatic index	94.33
A_{280} (ox.)	0.96
A_{280} (red.)	0.95
ε_{280} [M^{-1} cm^{-1}]	50,880.00
Instability index	19.22

Each Property in the table is described below:

Property	Description
MW [g/mol]	The molecular weight of the selected region.
Net charge [pH=7]	The pH-dependent sum of charges in a population of molecules. Protean 3D calculates the net charge for protein regions and amino acid residues, and assumes a pH of 7.
pI	The isoelectric point (the pH at which the residue carries no net electrical charge), calculated from pKa tables from Lehninger et al. (2005) . This calculation is available only for amino acid selections.

Copyright DNASTAR, Inc. 2022

Average hydropathy	The Kyte-Doolittle hydropathy value, calculated using the method of Kyte, J. and Doolittle, R.F. (1982) . This calculation is available only for amino acid selections.
Aliphatic index	The relative volume occupied by aliphatic side chains (alanine, valine, isoleucine and leucine). See note below for reference. This calculation is available only for amino acid selections.
A_{280} (ox.)	The absorbance (optical density) for the residue in an oxidized environment. This value is affected by presence of disulfide bonds. See note below for reference. This calculation is available only for amino acid selections.
A_{280} (red.)	The absorbance (optical density) for the residue in a reduced environment. This value is affected by presence of disulfide bonds. See note below for reference. This calculation is available only for amino acid selections.
ε_{280} [$M^{-1}cm^{-1}$]	The extinction coefficient (amount of light a residue absorbs). Protean 3D assumes all Cys residues are reduced. See note below for reference. This calculation is available only for amino acid selections.
Instability index	An estimate of the stability of the protein in a test tube. A protein whose instability index is smaller than 40 is predicted as stable, while a value above 40 predicts that the protein may be unstable. This value is calculated using the method of Guruprasad et al. (1990) .

Note: For aliphatic index, absorbance and extinction coefficient, see Gasteiger et al. (2005). In all three calculations, the sequence is treated as independent residues. There is no algorithmic dependence on length.

For information about the tools available in this view, see Report view.

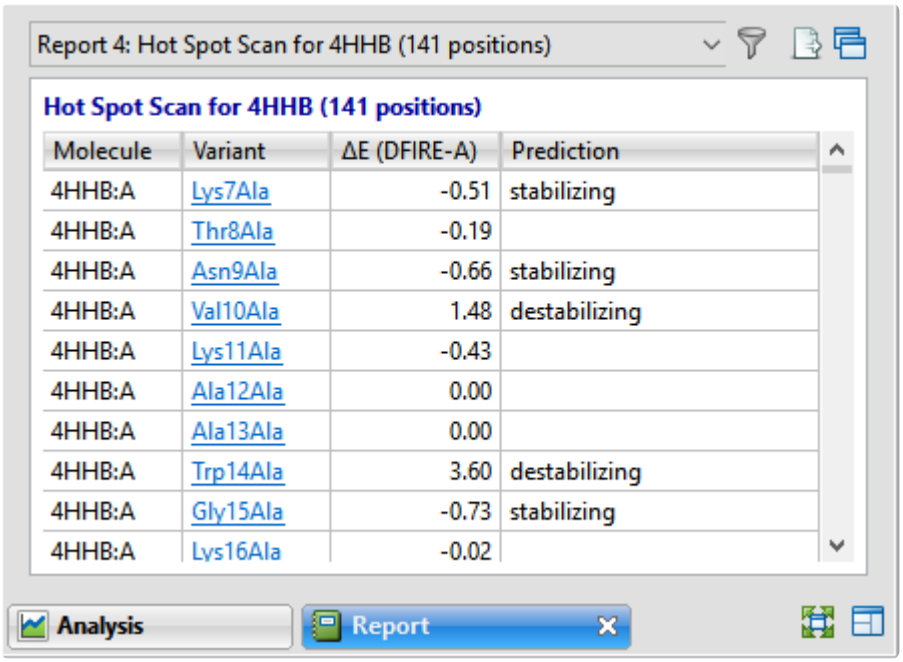
Copyright DNASTAR, Inc. 2022

Protein Design reports (Variants and Hot Spot Scan)

The Protein Design reports, “Variants” and “Hot Spot Scan,” are two similar versions of the Report view. Both versions contain data that can be used to predict whether a SNP might be stabilizing or destabilizing to a protein structure. These reports are available in the following circumstances:

- A Protein Design result is open. By default, hot spot scan results cause the Hot Spot Scan report to open; variants results cause the Variants report to open.
- A variant from the the “SNP to Structure” workflow is open. In this case, instructions from SeqMan Pro cause Protean 3D to open two structures. One is the original structure from the Protein Data Bank, while the other is the variant version calculated by Protean 3D.

An example Hot Spot Scan report is shown below:



Molecule	Variant	ΔE (DFIRE-A)	Prediction
4HHB:A	Lys7Ala	-0.51	stabilizing
4HHB:A	Thr8Ala	-0.19	
4HHB:A	Asn9Ala	-0.66	stabilizing
4HHB:A	Val10Ala	1.48	destabilizing
4HHB:A	Lys11Ala	-0.43	
4HHB:A	Ala12Ala	0.00	
4HHB:A	Ala13Ala	0.00	
4HHB:A	Trp14Ala	3.60	destabilizing
4HHB:A	Gly15Ala	-0.73	stabilizing
4HHB:A	Lys16Ala	-0.02	

To open a Protein Design report, click on the **Done** link for a Protein Design job in the Jobs panel, use the **View > Report > Show** command or press the **Report** tab (Report). If the desired report is not currently displayed in the Report view, select it from the drop-down menu in the upper left corner of the Report view. If a Variants or Hot Spot Scan report is not in the menu list, add it using the **Add reports to this view** tool ( ).

To open a variant structure from the report table, double-click its row.

To see experimental notes below the table, select a table row (molecule). To hide the notes, click the “Hide Notes* link above the table.

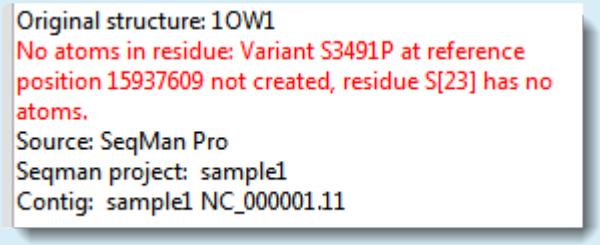
To learn about the tools available in this view, see Report view.

Interpreting Protein Design results:

The report table lists information about the variant structure(s). If desired, you can filter hot spot scan results.

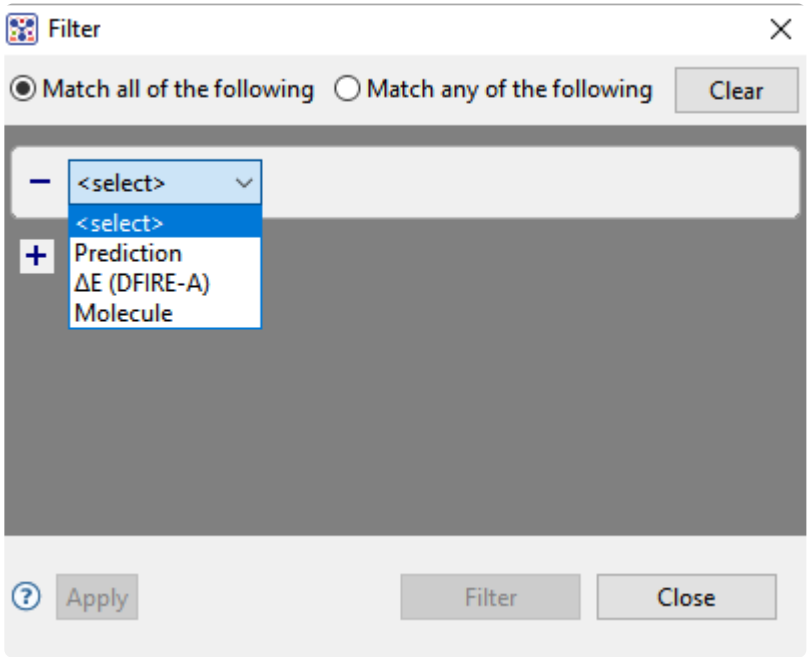
Table column	Description
Molecule	The structure and chain where the variant is located.
Variant	Each variant is shown in the form: [original amino acid] [position] [substitute amino acid]. Amino acids are shown using 3-letter IUPAC codes. In the Hot Spot Scan version of the report, each item in this column is hyperlinked. Click the link to open the Variants version of the report for the specified variant.
ΔE (DFIRE-A)	Displays the change in energy value based on the DFIRE calculation (reference). This number can be used to predict whether the mutation is stabilizing or destabilizing to the protein structure. A positive number > 0.5 is considered destabilizing to the structure when compared to the original amino acid; a negative number < -0.5 is considered stabilizing. Numbers between -0.5 and +0.5 are considered neutral.
Prediction	Stabilizing or destabilizing. Blank cells indicate a neutral prediction, neither stabilizing nor destabilizing.

Note: If the view pertains to a SNP-to-Structure result and the residue has no atoms in the PDB structure, you will see an error message similar to the one below. In this case, return to SeqMan Pro and choose a different row in the Variants table.



Filter hot spot scan results

When you press the **Filter results** tool ( ) in the header above the Hot Spot Scan report, there are two options: **Filter** and **Clear Filter**. Clicking on the **Filter** option opens the following dialog:



- Use the **<select>** drop-down menu to choose a filter type, then enter the desired options.

Filter type	Description	Options
Prediction	Stabilizing or destabilizing.	
ΔE (DFIRE-A	Displays the change in energy value based on the DFIRE calculation (reference). This number can be used to predict whether the mutation is stabilizing or destabilizing to the protein structure. A positive number > 0.5 is considered destabilizing to the structure when compared to the original amino acid; a negative number < -0.5 is considered stabilizing. Numbers between -0.5 and +0.5 are considered neutral.	
Molecule	The structure and chain where the variant is located.	

- To add another filter, press the plus icon (+). To delete a filter row, press the minus icon (−).
- Once you have set up all desired queries, go to the top of the dialog and choose whether to **Match all of the following** or **Match any of the following** queries.

Once you are finished setting up filters, do any of the following:

Copyright DNASTAR, Inc. 2022

- To apply the filters, press **Apply** or **Filter**.
- To clear all filters and start over, press **Clear**.
- To exit the dialog without filtering, press **Close**.

Copyright DNASTAR, Inc. 2022

NovaFold report

When you open a completed NovaFold prediction (*.novafold*) in Protean 3D, a specialized Report view contains detailed results about the prediction.

The sections of the NovaFold Report are:

- Molecule
- Templates
- Model Overview
- Model 'x' and its subsections:
    - Similar Experimental Structures
    - Predicted Binding Sites
    - Predicted Enzyme Function
    - Predicted Protein Function

The following video is a brief introduction to the NovaFold Report:

When viewing prediction results in the Report view, it may seem natural to compare the similarities between each section of a model and a template. However, this is most appropriate for molecules where a homolog (both similar in protein fold and sequence) is known. NovaFold uses a "template ensemble" approach, where different pieces of multiple templates are assembled in different combinations. This improves the structural diversity of the sampled structural conformations for the query molecule, which leads to more accurate models for molecules without clear homologous templates.

In addition, NovaFold also uses a model refinement technique when building the final all-atom models. By design, it modifies and repacks local regions in the model looking for lower-energy conformational intermediates.

Copyright DNASTAR, Inc. 2022

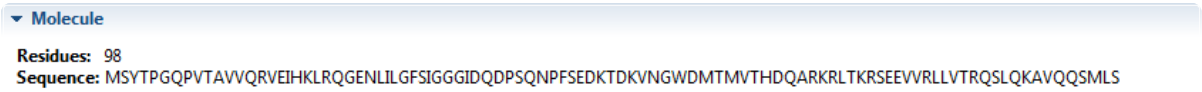
In summary, there are multiple algorithms in play which are designed to distort the final models away from the starting templates in search for the lowest-energy conformations.

To copy text from the NovaFold Report, see Use basic editing commands.

To open a template or structure from the report, see Open a structure from the NovaFold Report.

NovaFold Report: Molecule

The Molecule section of the NovaFold Report contains the number of residues in the sequence and the query sequence that was submitted for folding.

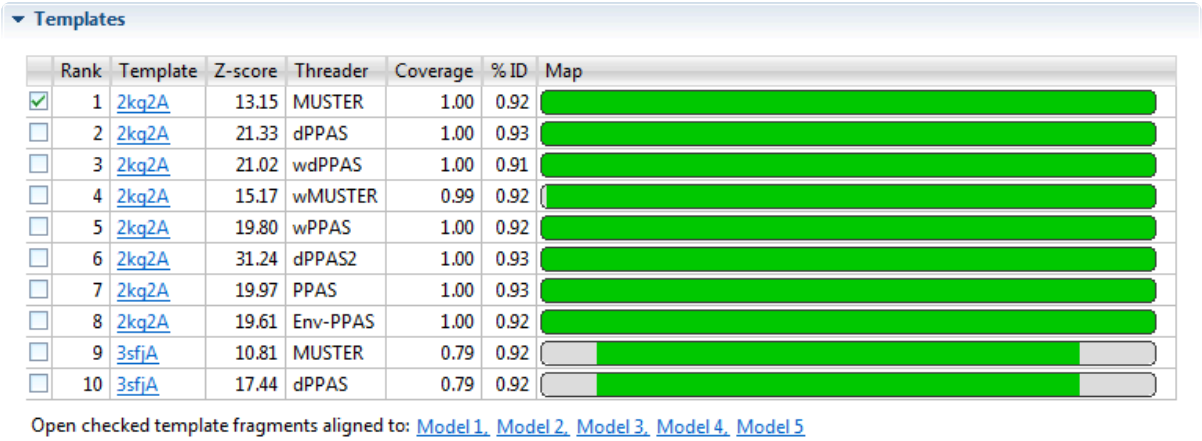


The next section of the NovaFold Report is the Templates section.

Copyright DNASTAR, Inc. 2022

NovaFold Report: Templates

In NovaFold, protein fragments from existing PDB structures are used as templates for building a prediction model. The Templates section of the NovaFold Report lets you compare the sequence similarity of a query sequence to the top ten template sequences in MegAlign Pro.



Templates

	Rank	Template	Z-score	Threader	Coverage	% ID	Map
☑	1	2kq2A	13.15	MUSTER	1.00	0.92	
☐	2	2kq2A	21.33	dPPAS	1.00	0.93	
☐	3	2kq2A	21.02	wdPPAS	1.00	0.91	
☐	4	2kq2A	15.17	wMUSTER	0.99	0.92	
☐	5	2kq2A	19.80	wPPAS	1.00	0.92	
☐	6	2kq2A	31.24	dPPAS2	1.00	0.93	
☐	7	2kq2A	19.97	PPAS	1.00	0.93	
☐	8	2kq2A	19.61	Env-PPAS	1.00	0.92	
☐	9	3sfjA	10.81	MUSTER	0.79	0.92	
☐	10	3sfjA	17.44	dPPAS	0.79	0.92	

Open checked template fragments aligned to: Model 1, Model 2, Model 3, Model 4, Model 5

- **Rank** - Displays a list of the top ten threading templates, with “1” representing the best template. Use the checkboxes at the left of the **Rank** column to specify which template(s) to align against a selected model (see bottom of this row).
- **Template** - The name of the template with a link to the entry on the PDB website.
- **Z-score** - A normalized score describing the similarity between the query and the template. Higher is better. A score ≥ 1.0 is considered to be significant.
- **Threader** - Threading algorithm that selected the template. NovaFold uses eight different algorithms.
- **%Coverage** - Fraction that the template overlaps the query sequence **(0,1]**.
- **%ID** - The percentage of the query sequence that is an exact match to the template sequence. Templates with low sequence similarity may still be structurally similar to the query sequence, but selecting these relevant templates is challenging.
- **Map** - A minimap showing the alignment of the query sequence to the template.
    - **Green** - ID ≥ 70
    - **Amber** - 30% ≤ ID <70
    - **Red** - ID <30
    - **Gray** - gaps

Copyright DNASTAR, Inc. 2022

Check one or more templates, and then click any of the links entitled “Open checked template fragments aligned to Model ‘n’” to open a new Protean 3D project showing each of the checked templates aligned with the selected model. In the Protean 3D document containing the requested alignment, each protein fragment will appear as an entry in the Features panel, and as a region feature in the Analysis view.

The next section of the NovaFold Report is the Model Overview section.

Copyright DNASTAR, Inc. 2022

NovaFold Report: Model Overview

The Model Overview section of the NovaFold Report allows you compare the structural similarity of a NovaFold prediction to experimental structures in the RCSB Protein Data Bank (PDB).

Model Overview

Model	TM-score	TM error	RMSD	RMSD error	C-score	Cluster size	Density score
1	0.79	±0.09	2.83	±2.05	0.57	20070	0.69
2	0.23	±0.06	14.97	±3.54	-4.70	102	0.00
3	0.21	±0.05	15.80	±3.23	-5.00	31	0.00
4	0.21	±0.05	15.80	±3.23	-5.00	10	0.00
5	0.21	±0.05	15.80	±3.23	-5.00	2	0.00

- **Model** – The model ID number.
- **TM-score** - Template modeling score [0, 1], best = 1. See below for information on color coding.
- **TM error** – The error (± value) for the TM-score calculation.
- **RMSD** - Root mean square deviation [0, ~30], lower is better. See below for information on color coding.
- **RMSD error** – The error (± value) for the RMSD calculation.
- **C-score** - Confidence score [-5,2], with higher numbers denoting increased confidence.
- **Cluster size** – During the “ab initio” simulations performed by NovaFold, thousands of plausible structure conformations are recorded. These conformations are clustered and one representative from each cluster is selected as the scaffold that is used to create the final prediction(s). **Cluster size** is the number of conformations in the cluster. A larger **Cluster size** is indicative of a lower-energy conformation. The lowest-energy conformations are associated with biologically-relevant structures. Therefore, with regard to determining ranking, a higher **Cluster size** is better.
- **Density score** - This score is related to the tightness of the cluster and similarity of the cluster to the threading templates. It is calculated as:

  Density score = $N_{cluster}$ / (N_{total} x RMSD)

  …where

- ◦ $N_{cluster}$ is the number of recorded conformations assigned to the cluster that the final model represents.

- N_{total} is the total number of recorded conformations from the simulation process.
- RMSD is the average root-mean-squared deviation of the cluster's conformations to the "centroid" cluster representative. The minimum value is 0.8 to prevent a division-by-zero singularity.

Density scores range from 0 to 1.25. A higher score is better with regard to determining ranking.

****

TM-score, **TM error**, **RMSD** and **RMSD error** are color-coded to indicate the predicted quality of the resulting models.

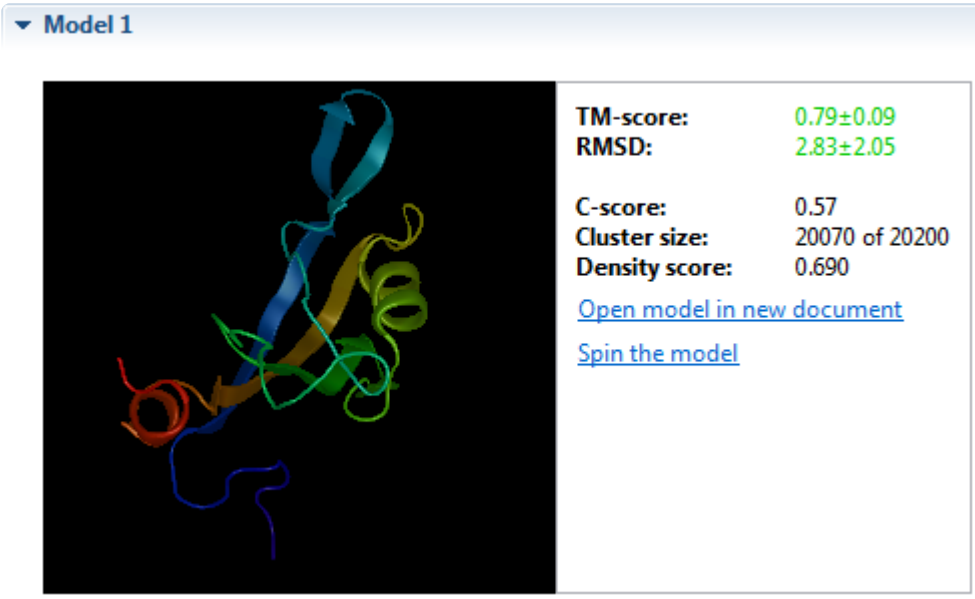
- **Green** – **TM-score** ≥ 0.75. Models are appropriate for drug design, drug screening, ligand docking, and molecular replacement.
- **Orange** - 0.50 ≤ **TM-score** < 0.75. Models are appropriate for mutagenesis design and predicting active sites, disease substitutions, alternative splicing patterns, and domain boundaries.
- **Red** - 0.17 ≤ **TM-score** < 0.50. Models might be appropriate for predicting domain boundaries.
- **Black** – **TM-score** < 0.17. Similarity between models can occur at random.

Below this section of the NovaFold Report are sections for each individual model.

Copyright DNASTAR, Inc. 2022

NovaFold Report: Model ‘x’

For each model number in the NovaFold Report, there is an expandable report section featuring a mini Structure view with a rendering of the predicted structure.



The scoring information at the top right reiterates data displayed in the Model Overview (see row above).

There are two links at the bottom right:

- Click **Open model in new document** to open the selected model as a Protean 3D *.structure* project in a new tab.

- Click **Spin the model** to spin the model left. When the model is spinning, you can stop it at any point by clicking **Stop spinning the model**.

After this, the following subsections may appear. These are discussed in separate topics:

- Similar Experimental Structures
- Predicted Binding Sites
- Predicted Enzyme Function
- Predicted Protein Function

Copyright DNASTAR, Inc. 2022

Similar Experimental Structures

The Similar Experimental Structures section of the NovaFold Report is visible only if you checked **Predict ligand binding sites and protein function** in the NovaFold Options screen.

Similar Experimental Structures

	Rank	PDB ID	TM-score	RMSD	Coverage	% ID
☑	1	2kg2A	0.90	1.35	0.89	0.89
☐	2	3sfjA	0.71	1.32	0.92	0.92
☐	3	2h2bA	0.62	2.77	0.22	0.22
☐	4	1v6bA	0.62	3.17	0.21	0.21
☐	5	2dazA	0.62	3.12	0.20	0.20
☐	6	3qdoA	0.61	3.05	0.21	0.21
☐	7	2fneC	0.61	2.72	0.22	0.22
☐	8	2xkxA	0.61	2.84	0.26	0.26
☐	9	2xkxA1	0.61	0.00	0.00	0.26
☐	10	1x5qA	0.61	2.68	0.24	0.24

Open model aligned with checked experimental structures

- **Rank** - Displays a list of the ten most similar templates, with “1” representing the best match. Use the check boxes at the left of the **Rank** column to specify which template(s) to align against a selected model (see bottom of this row).
- **PDB ID** - The name of the template with a link to the entry on the PDB website.
- **TM-score** - Template modeling score [0, 1], best = 1.
- **RMSD** - Root mean square deviation [0, ~30], lower is better.
- **%Coverage** - Fraction that the template overlaps the query sequence **(0,1]**.
- **%ID** - The percentage of the query sequence that is an exact match to the template sequence. Templates with low sequence similarity may still be structurally similar to the query sequence, but selecting these relevant templates is challenging.

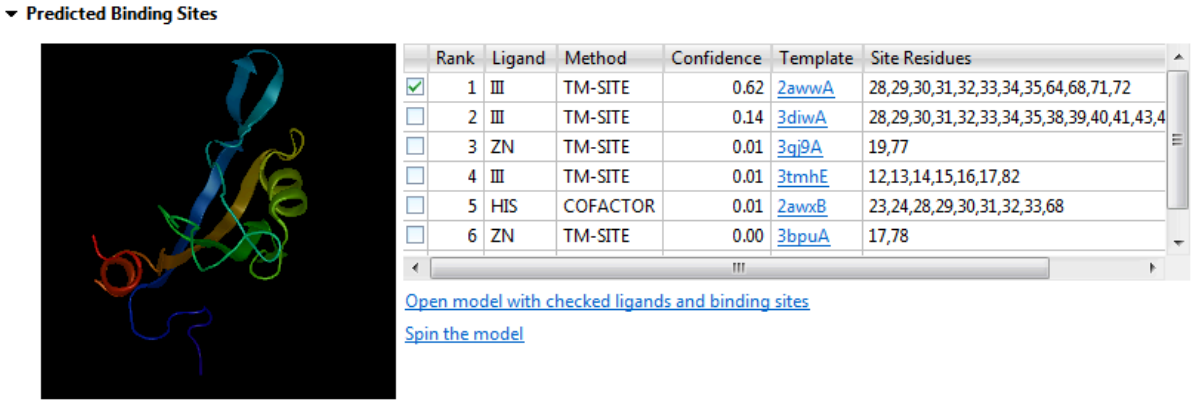
Click the link **Open model aligned with checked experimental structures** to open a new Protean 3D project showing the query sequence aligned to each of the checked templates.

The next section of the NovaFold Report is Predicted Binding Sites.

Copyright DNASTAR, Inc. 2022

Predicted Binding Sites

If you checked **Predict ligand binding sites and protein function** in the NovaFold Options screen, the Predicted Binding Sites section of the NovaFold Report is available for Model 1, and for any model with a predicted **TM-score** ≥ 0.5. This section lists predicted ligand binding sites for a NovaFold prediction in a model structure.



The thumbnail image is a mini Structure view that shows a rendering of the predicted structure and the ligands checked in the **Rank** column of the table.

- **Rank** - Orders the ligands that the query sequence is likely to bind to from one to ten, with one being the most likely. Use the checkboxes at the left of the **Rank** column to specify which ligands to include as features when opening the model structure (see bottom of this row).

- **Ligand** - Abbreviation for the ligand name.

- **Method** - The proprietary algorithm used in the calculation: **TM-SITE**, **COFACTOR** or **S-SITE**.

> **Note:** Since S-SITE predictions are sequence based, checking a row with this method does not cause the ligand to be displayed in the thumbnail view.

- **Confidence** - Confidence score [0,1], with higher numbers denoting increased confidence.

- **Template** - The name of the template with a link to the entry on the PDB website.

- **Site Residues** - The numbers of the residues involved in the predicted binding site.

There are two links at the right:

- Click **Open model with checked ligands and binding sites** to open the model structure with each of the checked ligands as Features in the Features panel. This allows you to view similarities between a predicted binding site and known binding sites.

Copyright DNASTAR, Inc. 2022

- Click **Spin the model** to spin the model left. When the model is spinning, you can stop it at any point by clicking **Stop spinning the model**.

In certain cases, the next section of the NovaFold Report is Predicted Enzyme Function.

Copyright DNASTAR, Inc. 2022

Predicted Enzyme Function

If you checked **Predict ligand binding sites and protein function** in the NovaFold Options screen, the Predicted Enzyme Function section of the NovaFold Report is available for Model 1, and for any model with a predicted **TM-score** ≥ 0.5. During calculation, NovaFold uses a database of known enzymes to predict whether the query sequence has an enzymatic function.

Predicted Enzyme Function

	Rank	EC Number	Confidence	Template	Site Residues
☑	1	1.14.13.39	0.18	1qavA	30,32,64,71
☐	2	3.1.4.11	0.11	1ihjB	36
☐	3	3.1.3.48	0.07	2vphB	30,47
☐	4	3.3.2.6	0.06	1qw6A	N/A
☐	5	1.7.99.3	0.06	1qs6X	N/A
☐	6	1.14.17.4	0.06	1w9yA	N/A
☐	7	1.-.-.-	0.06	1zpuA	N/A
☐	8	1.14.11.-	0.06	2og6A	N/A
☐	9	3.1.3.48	0.06	1vj6A	29,82
☐	10	3.4.11.2	0.06	2dqmA	N/A

Open model with checked active sites

- **Rank** - Orders the predicted ligands from one to ten, with one being the most likely. Use the checkboxes at the left of the **Rank** column to specify which enzymes to show when opening the model structure (see bottom of this row).
- **EC Number** - The enzyme number and a link to the entry on the BRENDA database.
- **Confidence** - Confidence score [0,1], with higher numbers denoting increased confidence.
- **Template** - The name of the template with a link to the entry on the PDB website.
- **Site Residues** - The numbers of the residues involved in the predicted binding site.

Click the link **Open model with checked active sites** to open the model structure with each of the checked active sites listed as features in the Features panel.

In certain cases, the next section of the NovaFold Report is Predicted Protein Function.

Copyright DNASTAR, Inc. 2022

Predicted Protein Function

If you checked **Predict ligand binding sites and protein function** in the NovaFold Options screen, the Predicted Protein Function section of the NovaFold Report is available for Model 1, and for any model with a predicted **TM-score** ≥ 0.5.

Predicted Protein Function

GO Molecular Function (GO-score): beta-catenin binding (0.67), protein C-terminus binding (0.67)

GO Biological Process (GO-score): negative regulation of cell proliferation (0.67), activation of Cdc42 GTPase activity (0.67), negative regulation of Wnt receptor signaling pathway (0.67), negative regulation of protein localization at cell surface (0.67)

GO Cellular Component (GO-score): cytoplasm (0.67), nucleus (0.67)

This section shows possible protein functions of the query sequence using terms from the Gene Ontology database's three major GO classifications:

- GO Molecular Function
- GO Biological Process
- GO Cellular Component

Each GO term is followed by a Confidence score [0,1], with higher numbers denoting increased confidence.

This section is the last one in the NovaFold Report. If you would like to open a template or structure from within the report, see Open a structure from the NovaFold Report.

Copyright DNASTAR, Inc. 2022

Open a structure from the NovaFold Report

There are several links in the NovaFold Report that allow you to open a model as a new Protean 3D document. When opened, each selected model will appear as a .pdb-formatted structure in a separate tab of the Protean 3D window.

The following table describes the locations of links in the NovaFold Report that allow you to open documents.

To open this…	…do this in the NovaFold Report:
One or more selected template fragments aligned to the experimental structure	In the Templates section of the report, check rows for the templates you wish to open, then click the link below the table (e.g., “Model 1”).
A specific model	In the Model section of interest (e.g., Model 3), to the right of the thumbnail graphic, click the **Open model in new document** link.
One or more selected templates aligned to a specific model.	In the Model section of interest, under Similar Experimental Structures, check rows for the templates you wish to align with the model, then click the **Open model aligned with checked experimental structures** link.
A specific model, along with predicted ligands and binding sites	In the Model section of interest, under Predicted Binding Sites, click the **Open model with checked ligands and binding sites** link.
A specific model, along with predicted active sites	In the Model section of interest, under Predicted Enzyme Function, click the **Open model with checked active sites** link.

Copyright DNASTAR, Inc. 2022

NovaDock report

When you open a completed NovaDock prediction (*.novadock*) in Protean 3D, a specialized Report view contains detailed results about the prediction.

The sections of the NovaDock Report are:

- Molecules
- Model Overview
- Model and its subsection:
    - Models in Cluster

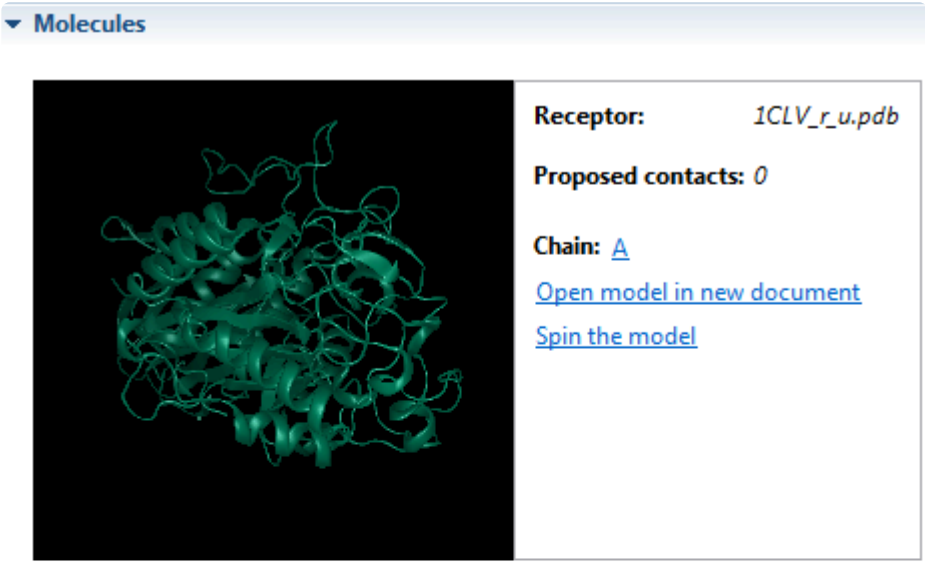
To copy text from the NovaDock Report, see Use basic editing commands.

To open a template or structure from the report, see Open a structure from the NovaDock Report.

Copyright DNASTAR, Inc. 2022

NovaDock Report: Molecules

The Molecules section of the NovaDock Report displays information about the receptor and ligand structures that were used as inputs in the NovaDock prediction. To view or hide this section of the report, click on the expand arrow (small blue triangle) in its header.



In the Molecules section, thumbnail preview images show the input molecules rendered in the “ribbon” motif, with the receptor shown in green, and the ligand shown in yellow.

Because the original structures were user-supplied, they may have had missing atoms or residues, or non-standard residues. The structures shown in the Molecules section are the “repaired” versions of those files (if any repairs were made). If any repairs were made, they will be listed in the **Repairs** section of the Details panel.

Thumbnail previews are interactive and can be controlled with the mouse or with the links provided:

Task	How to…
To rotate the model manually	Slide the mouse laterally while holding down the left mouse button.
To spin the model around its axis automatically	Click the link **Spin the model** to spin the model slowly clockwise, as viewed from above. Once the model is spinning, this link changes to **Stop spinning the model**.
To open the model as a new Protean 3D document	Click the link **Open model in new document**.
To zoom in/out	Use the mouse's scroll wheel.

Information about each molecule appears to the right of its thumbnail preview.

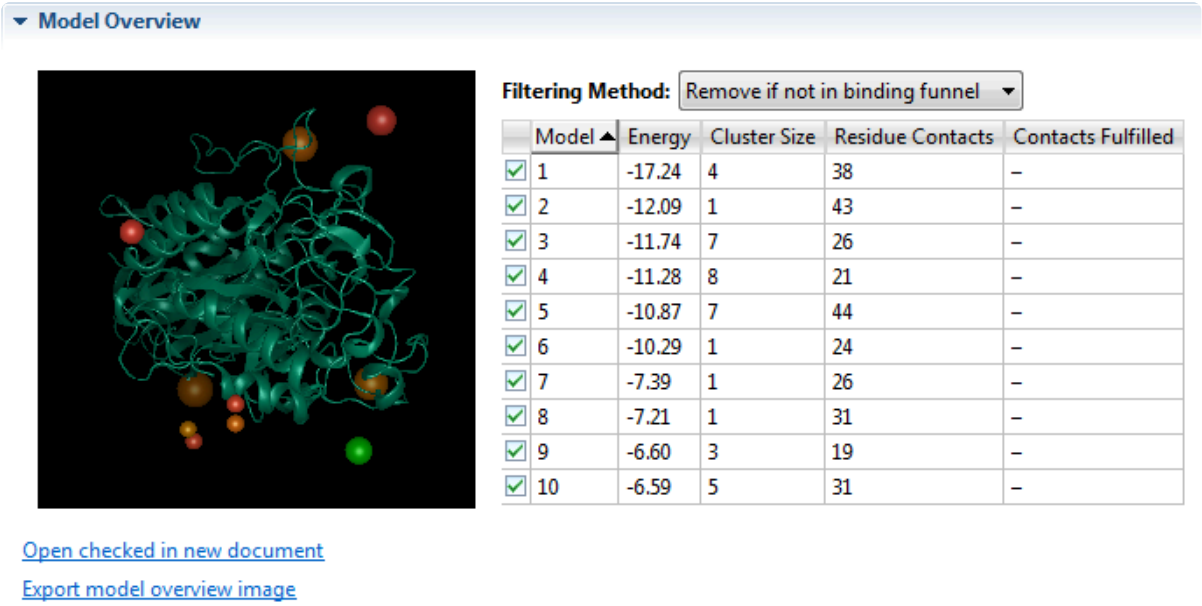
Item	Description
Receptor/ Ligand	The name of the input file specified as the receptor or ligand when setting up a NovaDock prediction.
Proposed contacts	Number of residue contacts, if any, proposed by the user when initiating the NovaDock prediction.
Chain	The letter representing the chain name. Clicking the chain name link selects the receptor and/or ligand in the two thumbnail images. The selected chain(s) will be highlighted in light blue.

The next section of the NovaDock Report is the Model Overview section.

Copyright DNASTAR, Inc. 2022

NovaDock Report: Model Overview

This section of the NovaDock Report represents the top ten ligand-receptor docking models predicted by NovaDock. To view or hide this section of the report, click on the expand arrow (small blue triangle) in its header.



Model	Energy	Cluster Size	Residue Contacts	Contacts Fulfilled
1	-17.24	4	38	–
2	-12.09	1	43	–
3	-11.74	7	26	–
4	-11.28	8	21	–
5	-10.87	7	44	–
6	-10.29	1	24	–
7	-7.39	1	26	–
8	-7.21	1	31	–
9	-6.60	3	19	–
10	-6.59	5	31	–

The thumbnail preview image acts as an overview, with colored “heat bubble” spheres representing the centers of mass for ligand contacts. Each sphere represents the position of the ligand for one of the models listed in the corresponding table. Green is used to denote the lowest energy position (best) and red, the highest. Sphere size is proportional to cluster size, with larger being better.

The thumbnail preview is interactive and can be controlled with the mouse. To spin the model around its axis manually, slide the mouse laterally while holding down the left mouse button. To zoom in/out, use the mouse’s scroll wheel.

On the right, use the **Filtering Method** drop-down menu to specify whether or not to use funnel filtering, an energy based analysis method:

- **Remove if not in binding funnel** – Applies funnel filtering to models in the table, thereby eliminating models that do not land in an “energy funnel.”
- **No filtering** – Does not apply funnel filtering.

The Model Overview table contains sortable columns. Click once on a column header to make it active. Then click on the black triangle in the column header to sort the column numerically. Click on the triangle a second time to sort in the opposite direction.

The table below describes the columns in the Model Overview table.

Copyright DNASTAR, Inc. 2022

Column	Description
Model	The Model ID number. Models are ranked on a scale of 1 to 10 based on their **Energy** scores (see below). Use the checkboxes to the left of the **Model ID** column to specify which model(s) you would like to open when using the **Open checked in new document** link.
Energy	The energy score for each model. Lower (i.e., a larger negative number) is better.
Cluster size	During the simulations performed by NovaDock, thousands of plausible docking models are recorded. These models are clustered and one representative from each cluster is selected as the scaffold that is used to create the final prediction(s). **Cluster size** is the number of conformations in the cluster. A larger **Cluster size** is indicative of a lower-energy conformation. The lowest-energy conformations are associated with biologically-relevant structures. Therefore, with regard to determining ranking, a higher **Cluster size** is better.
Residue Contacts	The number of ligand residues in contact with receptor residues in a particular model.
Contacts Fulfilled	If a particular ligand or receptor residue was proposed when the NovaDock prediction was initiated, this column shows if it was included in the interface.

To open one or more models as new Protean 3D documents – Leave checkmarks only next to the models you wish to open. Then click the link **Open checked in new document**.

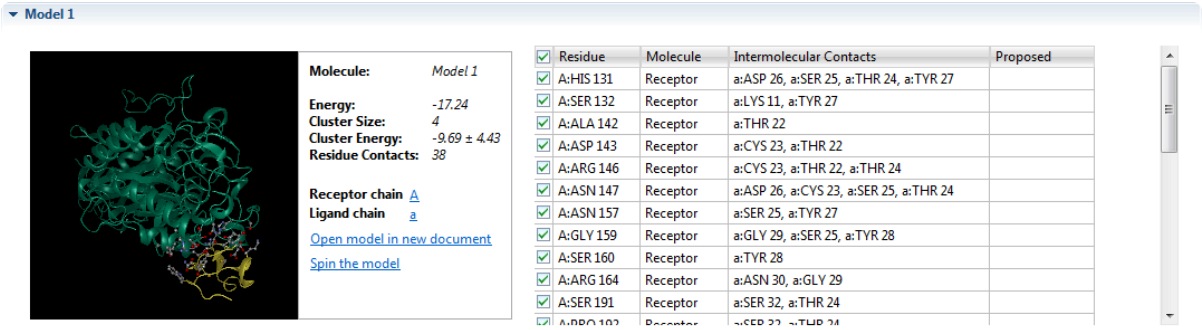
To export the thumbnail preview image – Click the link **Export model overview image**. You may choose to save the image in *.png*, *.gif*, *.jpg* or *.tiff* formats.

Below this section of the NovaDock Report are sections for each individual model.

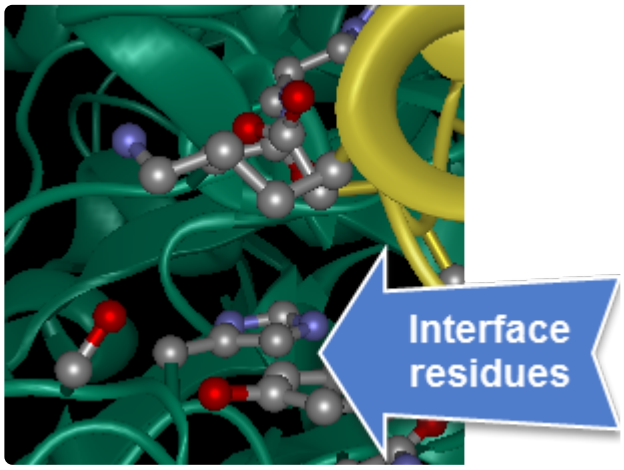
Copyright DNASTAR, Inc. 2022

NovaDock Report: Model ‘x’

Below the Model Overview section of the NovaDock Report are expandable sections for each of the individual models (e.g., Model 1, Model 2, etc.). To view or hide this section of the report, click on the expand arrow (small blue triangle) in its header.



The image in each Model shows the predicted ligand-receptor docking model. The receptor and ligand chains are represented by green and yellow ribbons, respectively. Interface residues are rendered in “ball and stick” format, and can be seen more clearly by zooming in on the image, as described in the table below.



The thumbnail preview is interactive and can be controlled with the mouse:

Task	How to…
To spin the model manually	Slide the mouse laterally while holding down the left mouse button.
To zoom in/out	Use the mouse’s scroll wheel.
To view the location of a particular residue	Select the residue in the table on the right by clicking on its row. The residue will be highlighted in light blue in the image.
To view the location of the receptor and ligand chains	To the right of the image, click the links next to **Receptor chain** or **Ligand chain**. The associated chain will be highlighted in light blue.

Copyright DNASTAR, Inc. 2022

Information about the model appears to the right of the thumbnail preview.

Item	Description
Molecule	Model ID number.
Energy	The energy score for each model. Lower is better.
Cluster Size	During the simulations performed by NovaDock, thousands of plausible docking models are recorded. These models are clustered and one representative from each cluster is selected as the scaffold that is used to create the final prediction(s). **Cluster size** is the number of conformations in the cluster. A larger **Cluster size** is indicative of a lower-energy conformation. The lowest-energy conformations are associated with biologically-relevant structures. Therefore, with regard to determining ranking, a higher **Cluster size** is better.
Cluster Energy	The average energy for the clusters in the model.
Residue Contacts	The number of ligand residues in contact with receptor residues in a particular model.
Receptor chain	A link that can be pressed to select the receptor chain (green ribbon) in the thumbnail preview image. The associated chain will be highlighted in light blue.
Ligand chain	A link that can be pressed to select the ligand chain (yellow ribbon) in the thumbnail preview image. The associated chain will be highlighted in light blue.

Additional information about the model appears in the Model table. Selections made in this table are simultaneously selected in the thumbnail preview.

Column	Description
Residue	A label for each interface residue. It consists of the chain where the residue is located, its amino acid abbreviation and the residue number (e.g., A:HIS 131).
Molecule	The molecule to which the interface residue belongs (i.e., Receptor or Ligand).
Intermolecular Contacts	A list of contacts between the two binding partners.
Proposed	If a particular ligand residue was proposed when the NovaDock prediction was initiated, this column shows if that residue was included in the interface.

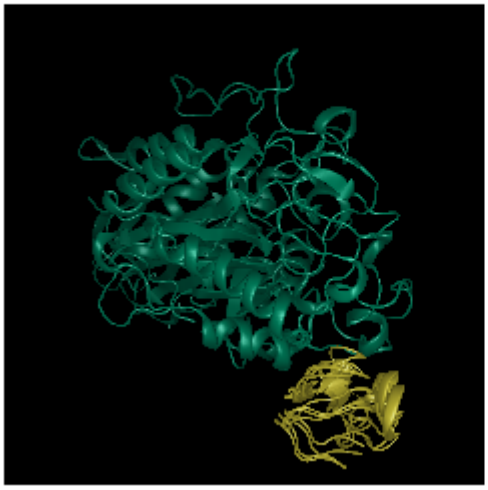
Each Model section also includes a separate area called Models in Cluster.

Copyright DNASTAR, Inc. 2022

Models in Cluster

In the NovaDock Report, Models in Cluster is a subsection of the Model section. To view or hide this section of the report, click on the expand arrow (small blue triangle) in its header.

Models in Cluster 1



☑	Model ID	Energy	RMSD
☑	1a	-17.24	0.00
☑	1b	-8.29	2.25
☑	1c	-7.17	2.33
☑	1d	-6.05	1.89

Open checked in new document

The thumbnail preview shows alternative conformations for that model clustered together in a single image. For example, if the model has a Cluster Size of ‘4’ and all boxes in the table are checked (as they are by default), then all four clusters will be overlaid, causing the image to appear “blurry.” Check or uncheck boxes in the table to show or hide alternative conformations, as desired. Your selections will be reflected in the thumbnail preview.

The table below describes the columns in the Models in Cluster table.

Column	Description
Model ID	Displays a list of alternative versions of the model, ranked alphabetically by energy score. Use the checkboxes at the left of the **Model ID** column to specify which alternative model(s) to display in the thumbnail preview, and also to specify which model(s) to open as structure documents when using the **Open checked in new document** link.
Energy	The energy score for each alternative model. Lower is better.
RMSD	Root mean square deviation [0, ~30], lower is better.

To open one or more alternative models as new Protean 3D documents, leave checkmarks only next to the models you wish to open. Then click the link **Open checked in new document**.

Copyright DNASTAR, Inc. 2022

Open a structure from the NovaDock Report

There are several links in the NovaDock Report that allow you to open a model as a new Protean 3D document. When opened, each selected model will appear as a .pdb-formatted structure in a separate tab of the Protean 3D window.

The following table describes the locations of links in the NovaDock Report that allow you to open documents.

To open this…	…do this in the NovaDock Report:
Receptor and/or ligand structures used as NovaDock inputs	In the Molecules section, click on either of the **Open model in a new document** links. The structures will include any repairs made in the course of making the NovaDock prediction.
One or more docking models	In the Model Overview section, put checkmarks only next to the models of interest. Remove the other checkmarks. Then click **Open checked in new document**.
A specific docking model	In the Model section of interest (e.g., Model 3), click on **Open model in new document**.
One or more versions of a specific docking model	In the Models in Cluster area, put checkmarks only next to the models of interest. Remove the other checkmarks. Then click **Open checked in new document**.

Copyright DNASTAR, Inc. 2022

NovaCloud Services view

NovaCloud Services allows you to make a prediction using NovaFold, NovaFold AI, NovaFold Antibody, or NovaDock.

Open the Services view using any of the following menu commands or keyboard shortcuts. Note that some commands are only available if a protein sequence document is active.

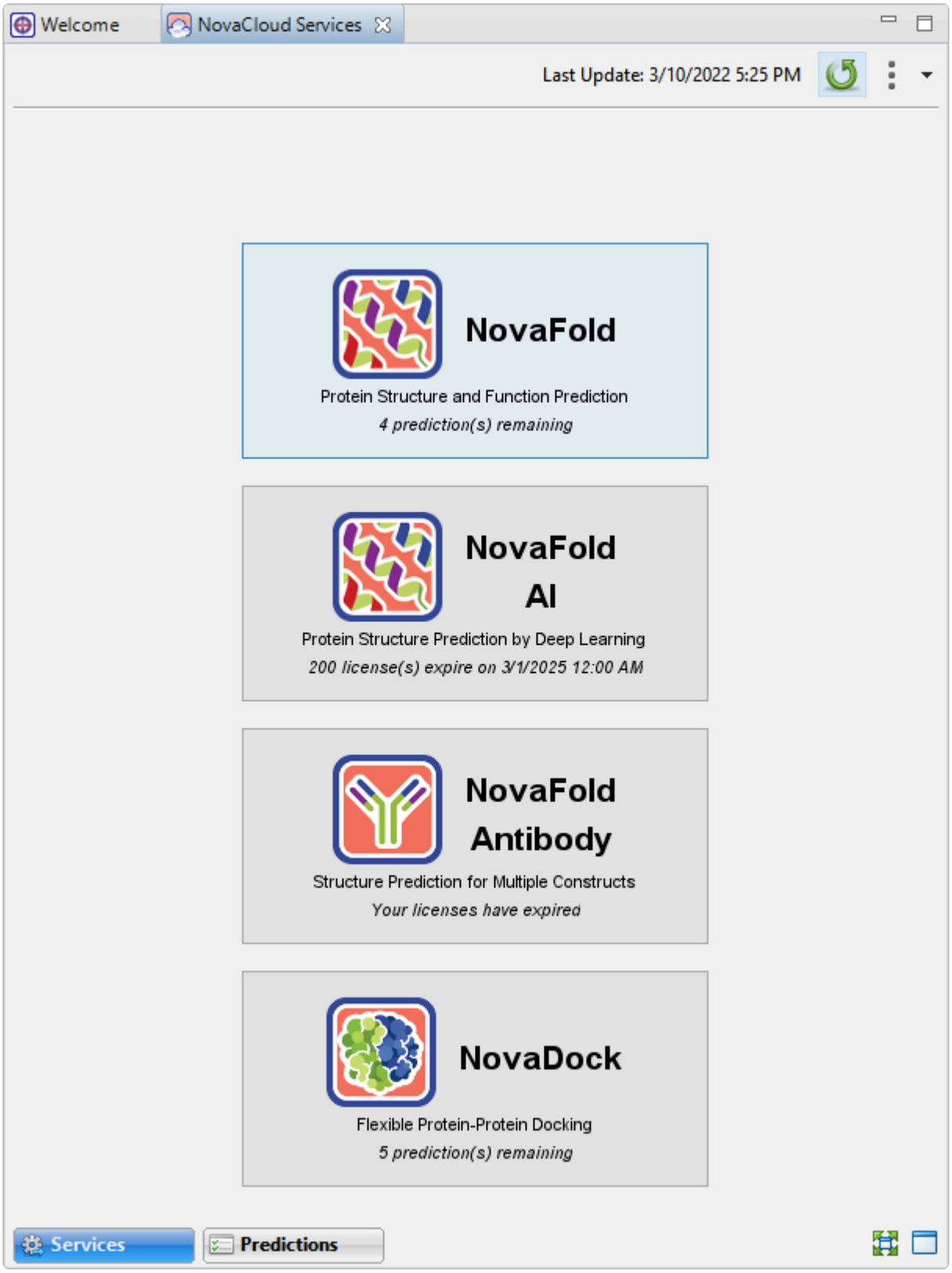
- **Modeling > Protein Structure Prediction > Connect to NovaCloud Services**
- **Modeling > Antibody Structure Prediction > Connect to NovaCloud Services**
- **Modeling > Protein-Protein Docking > Connect to NovaCloud Services**
- **File > New NovaCloud Prediction**
- **View > Services > Show**
- **View > NovaCloud Services > Connect to NovaCloud Services**
- **Ctrl+Alt+N** (Windows) or **Option+Cmd+N** (Macintosh)

If you have not yet logged in, enter the same email and password you use to log in to the DNASTAR website. To prevent having to enter this information in the future, check the **Remember me** box. Then press the **Log in** button. Links are provided if you have forgotten your password or do not yet have a DNASTAR account, as is often the case for network licensees.

Copyright DNASTAR, Inc. 2022

Once you have successfully logged in, the Services view displays a large button for each of the three NovaCloud applications. Each button describes the associated workflow and the number of predictions remaining on your license.

Copyright DNASTAR, Inc. 2022



Tasks that pertain to the NovaCloud Services view:

Task	Description

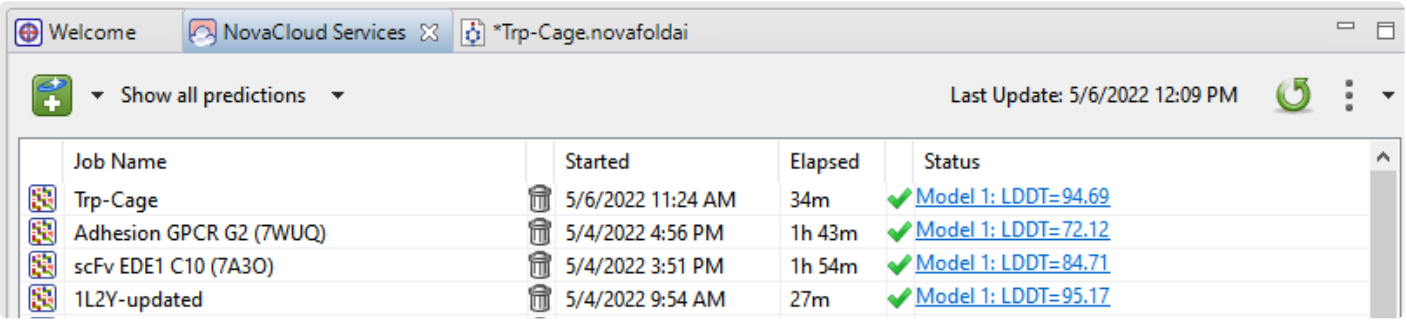
To launch a NovaCloud Services application	Press the button with the name of the service. This launches one of the prediction wizards: NovaFold, NovaFold AI, NovaFold Antibody, or NovaDock. After you follow the steps in the wizard and begin a prediction, you can check its status and results in the Predictions view.
To manually sync to the NovaCloud Server	Press the **Refresh** button ( ).The right side of the header displays the date and time the server was last contacted.
To access help for this view	Click the **Additional actions** tool ( ) on the right of the header, and choose **Help Contents**.
To go to your account on the DNASTAR website	Click the **Additional actions** tool ( ) on the right of the header, and choose **My Account**.
To log out from the NovaCloud Server	Click the **Additional actions** tool ( ) on the right of the header, and choose **Log Out**.
To change the layout of the view or what is displayed in it	See Customize the look and layout.

Copyright DNASTAR, Inc. 2022

Predictions view

After launching NovaCloud Services, two views become available. The first of these is NovaCloud Services, which is used to log in and launch a setup wizard for NovaFold, NovaFold Antibody, or NovaDock predictions. Once a prediction has been started, you can monitor its progress and download its eventual output from the Predictions view.

Note that some licenses allow predictions to be run in batches of five predictions at a time, with additional predictions placed in a queue. If you have a limited number of licenses, predictions will only be run until you run out of licenses. For example, if you have three licenses remaining and attempt to run five predictions, only the first three predictions will actually be run.



Description of columns in the Predictions view table:

Column	Description
Job Name	The name specified in the prediction wizard. Note that the same Job Name can be used for more than one prediction.
(Delete Job)	To delete a job, press the trash can icon in its row.
Started	The date and time the prediction began.
Elapsed	The elapsed time of an in-progress prediction; or the time to completion, for finished jobs.
Status	An icon showing the status of the run, followed by a link to open the finished prediction, if the prediction was successful. The status column may also include the calculated TM-Score and RMSD. The following table describes each status icon. • **Queued** – – Submitted and placed into the queue • **Running** – – Running on the NovaCloud • **Done** – – Prediction complete; results available • **Failed** – – Prediction failed

	• **Stopped** – – Prediction stopped by user

Tasks that pertain to the Predictions view:

Task	How to…
To open the results of one or more predictions	Click on the link in the **Status** column.
To download the results of one or more predictions	Right-click on the link in the **Status** column and choose **Download** from the context menu.
To delete a prediction	In the **Job Name** column, select the row you want to delete and then press the trash can icon ( ).
To stop one or more predictions that are in progress	Select one or more rows. Then right-click on any part of the selection and choose **Stop**. Once stopped, a prediction cannot be restarted from the Predictions view. You must instead initiate a new prediction.
To begin a new folding or docking prediction	Click on the **Start new prediction** tool ( ), on the left of the header, and make a choice from the menu. NovaFold NovaFold AI NovaFold Antibody NovaDock The corresponding wizard is launched, allowing you to set up the prediction.
To limit the contents of the table to a specific set of predictions	Click **Show all predictions**, on the left of the header, and make a choice from the menu. Show all predictions Show predictions from the past week Show active predictions Show completed predictions Show failed predictions
To manually sync to the NovaCloud Server	Press the **Refresh** button ( ).The right side of the header displays the date and time the server was last contacted.
To access help for this	Click the **Additional actions** tool ( ) on the right of the header, and choose

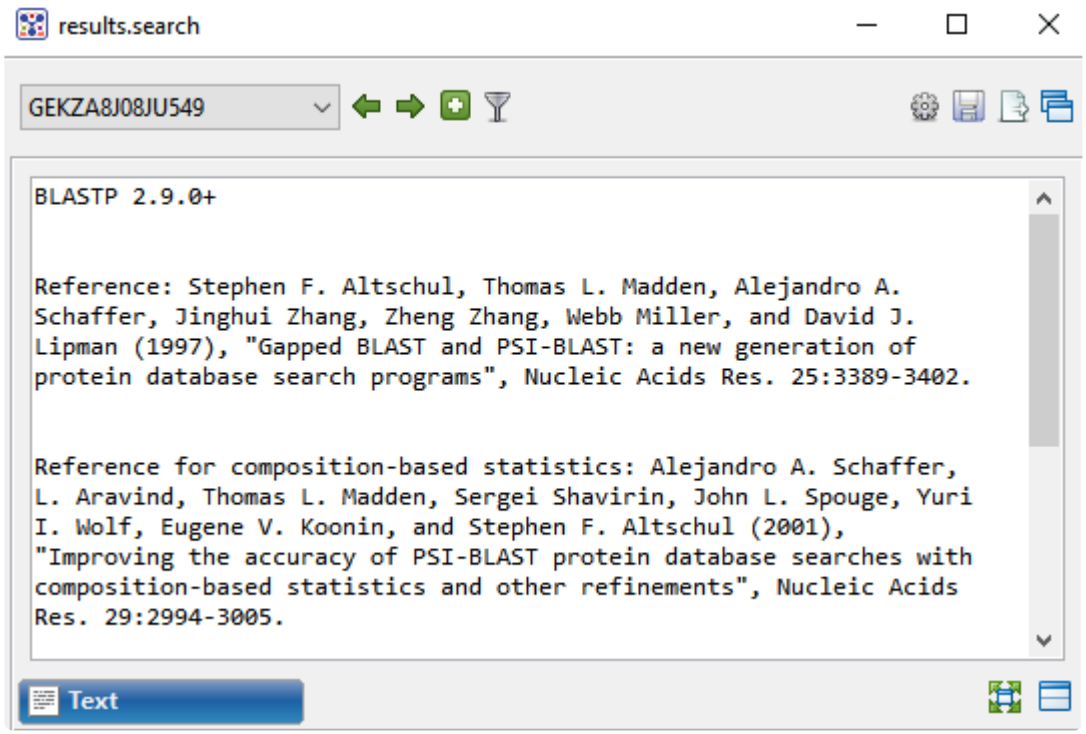
view	**Help Contents**.
To go to your account on the DNASTAR website	Click the **Additional actions** tool ( ) on the right of the header, and choose **My Account**.
To log out from the NovaCloud Server	Click the **Additional actions** tool ( ) on the right of the header, and choose **Log Out**.
To change the layout of the view or what is displayed in it	See Customize the look and layout.

Copyright DNASTAR, Inc. 2022

Text view

After the Jobs panel shows the **Status** of a BLAST search or Entrez search is **Done**, pressing the “Done” hyperlink opens the search results simultaneously in the Text, Table, and Pairwise views. All three views are contained in a large pop up window that is separate from the main window.

An example Text view is shown below:



In order, the view shows the following:

- The version of BLAST used in the search (e.g., “BLASTP 2.9.0+”).
- References for the BLAST search engine.
- A description of the BLAST database used in the search, including the number of sequences.
- Information about the Query sequence, including its length and the Request ID (RID) number.
- A table of sequences producing significant alignments.
- Pairwise alignments of each significant match with summary information on the match.

Tasks that pertain to the Text view:

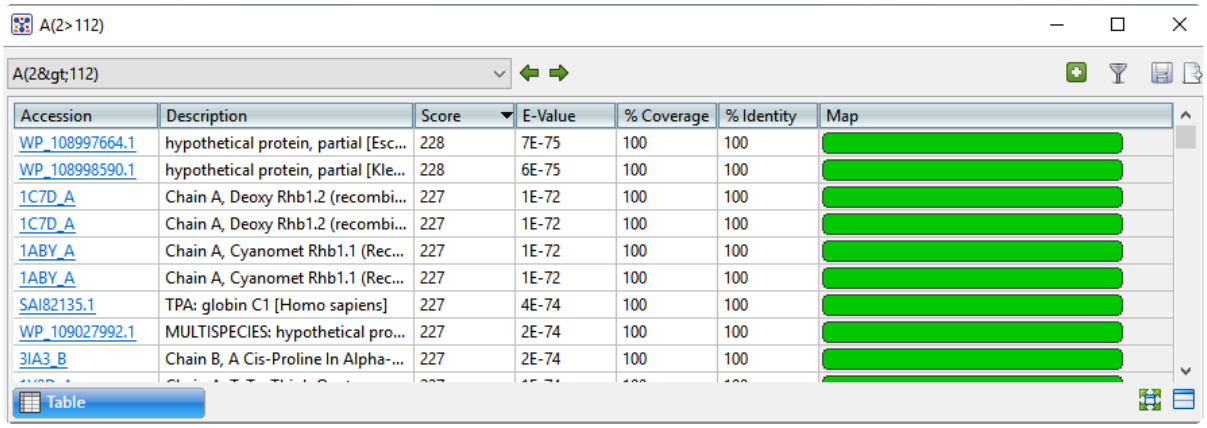
Task	How to…
To make selections in the view	See Make a Selection.
To copy all text	Right-click anywhere in the view and choose **Copy**.
To select all text	Right-click anywhere in the view and choose **Select All**.
To change the layout of the view or what is displayed in it	See Customize the look and layout.

Copyright DNASTAR, Inc. 2022

Table view

After the Jobs panel shows the **Status** of a BLAST search or Entrez search is **Done**, pressing the “Done” hyperlink opens the search results simultaneously in the Text, Table, and Pairwise views. All three views are contained in a large pop up window that is separate from the main window.

If no significant matches were identified, the Table view will be empty. An example of a populated Table view is shown below:



Accession	Description	Score	E-Value	% Coverage	% Identity	Map
WP_108997664.1	hypothetical protein, partial [Esc...	228	7E-75	100	100	
WP_108998590.1	hypothetical protein, partial [Kle...	228	6E-75	100	100	
1C7D_A	Chain A, Deoxy Rhb1.2 (recombi...	227	1E-72	100	100	
1C7D_A	Chain A, Deoxy Rhb1.2 (recombi...	227	1E-72	100	100	
1ABY_A	Chain A, Cyanomet Rhb1.1 (Rec...	227	1E-72	100	100	
1ABY_A	Chain A, Cyanomet Rhb1.1 (Rec...	227	1E-72	100	100	
SAI82135.1	TPA: globin C1 [Homo sapiens]	227	4E-74	100	100	
WP_109027992.1	MULTISPECIES: hypothetical pro...	227	2E-74	100	100	
3IA3_B	Chain B, A Cis-Proline In Alpha-...	227	2E-74	100	100	

Columns in the view’s table are described below:

Column	Description
Accession	The NCBI “Accession” number for the match. Press the link to open the corresponding page on the NCBI website.
Description	The NCBI “Definition” field text.
Score	The similarity score for the match. (See the NCBI definition for **raw score**). In general, a higher score denotes a better match.
E-Value	The chance that the match is expected to occur by chance. (See the NCBI definition for **E value**). In general, a lower score denotes a better match.
%Coverage	The percentage of the query sequence that aligns to a match in the BLAST database. Example: Query has 596 bases, of which 565 align to a match in the NCBI database. **%Coverage** = 565/596 = 95%.
%Identity	The percentage of bases that match when the query and result sequence are aligned. (NCBI definition). Example: Query has 565 bases that align to the match and 3 bases that don’t match. **%Identity** = 562/565 = 99.5%.

Copyright DNASTAR, Inc. 2022

Map	Graphic showing areas of agreement between the query and the match.

Tasks that pertain to the Table view:

Task	How To…
To launch the NCBI page for matching database entry	Click the hyperlink in the table's Accession column.
To copy all information from a search result row	Right-click on the row and choose **Copy**.
To save the sequence match in *.gbk* format	Right-click on the row and choose **Download sequences**. In the ensuing Save As dialog, choose a location in which to save the sequence. By default, the sequence is named using its Accession number.
To display a different search job	Select it from the drop-down menu in the upper left of the window (GEKZA8J08JU549).
To display a pairwise alignment using a particular search result	If both the Table and Pairwise views are open, click on the desired row in the Table view to see the corresponding pairwise alignment in the Pairwise view. As an alternative, open the Pairwise view manually and choose the desired target from the **with** drop-down menu.
To display search results for previous query (in a single job with multiple queries)	Use the **Previous query** tool ( ).
To display search results for the next query (in a single job with multiple queries)	Use the **Next query** tool ( ).
To filter results from a sequence search	Use the **Filter results** tool ( ). Filtering for this situation is described here.
To filter results from a text search	Use the **Filter results** tool ( ). Filtering for this situation is described here.
To add sequences or matches	Use the **Add sequences or subsequences** tool ( ). Sequences are added back into the host application, not the job's document.
To launch the Search Online dialog	Use the **Change search strategy** tool ( ). See Search for Sequences Online for details.

Copyright DNASTAR, Inc. 2022

To save the search job to disk	Use the **Save as** tool ( ).
To export data from the view	Use the **Export data** tool ( ), described in the topic Export data from the Text, Table or Pairwise views.
To change the layout of the view or what is displayed in it	See Customize the look and layout.

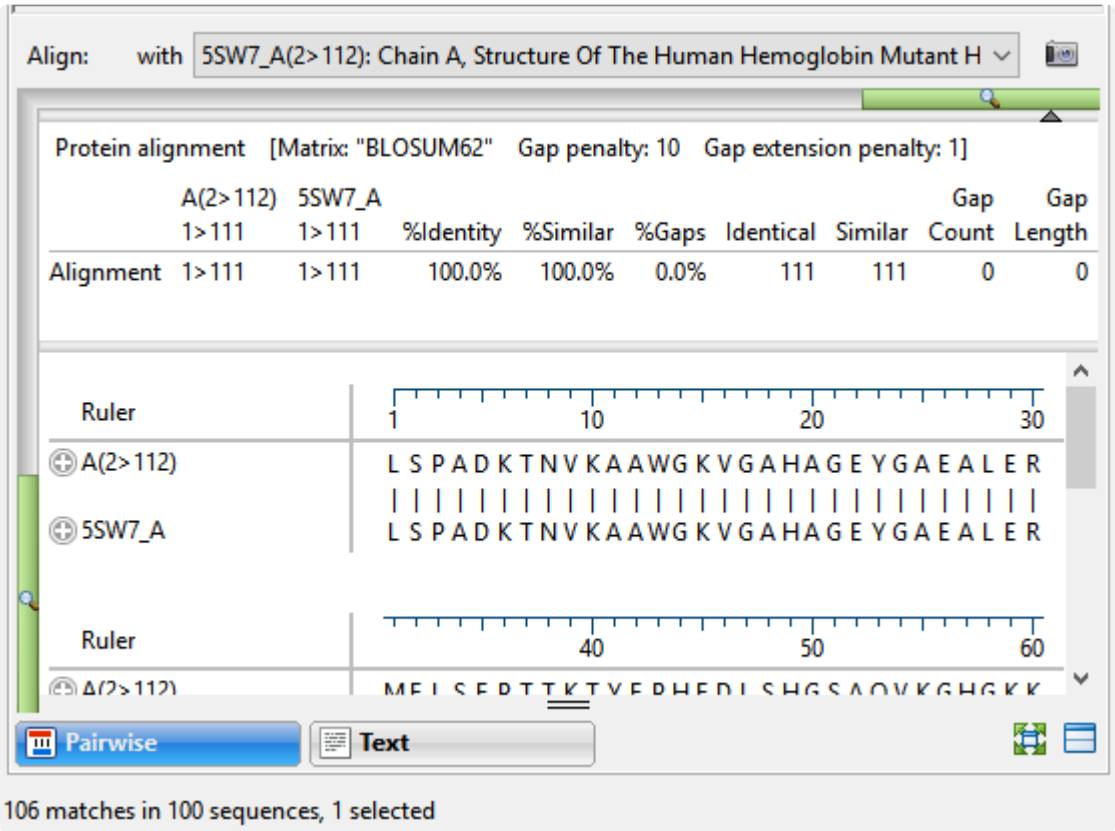
Pairwise view

After the Jobs panel shows the **Status** of a BLAST search or Entrez search is **Done**, pressing the “Done” hyperlink opens the search results simultaneously in the Text, Table, and Pairwise views. All three views are contained in a large pop up window that is separate from the main Protean 3D window.

The Pairwise view displays the query sequence aligned pairwise to a BLAST search match. This view is nearly identical to the same view in the Lasergene application MegAlign Pro.

To open a pairwise alignment for any completed BLAST results, do any of the following:

- Go to the Jobs panel (**View > Jobs**) and click on the **Done** link for a completed BLAST search job. A new window will open with the Table view on top and the Pairwise view on the bottom.
- If the view has been closed, return to it by clicking on the Pairwise tab (Pairwise).
- If the Pairwise view is open but you want to view results for a different BLAST search job, use the menu in the top left corner of the view to select the job.



The summary area near the top of the view describes the alignment engine and settings that were used to create the pairwise alignment. Protein alignments always use the BLOSUM62 matrix, while nucleotide alignments use NUC44. This section features a one-row table whose columns are described below.

Value	Description
(Query name)	The query ID with the range of nucleotides or amino acids that were included in the query shown below the name.
(Target name)	The accession number of the matching database entry with the range of nucleotides or amino acids included in the match shown below the accession number.
% Identity	Number of matching residues divided by the sum of all residues and gap characters. The quotient is multiplied by 100 to give the similarity as a percent.
% Gaps	The gap length divided by the consensus length. The quotient is multiplied by 100 to give the similarity as a percent.
Identical	Number of residues that are identical in both sequences.
Gap Count	Number of gaps (of any length) in either sequence of an alignment.
Gap Length	Number of gap characters in an alignment.
Score	The cumulative score for an alignment derived by adding the matrix-derived match scores and subtracting gap penalties.
Length	Number of residues in the consensus, including gaps that match in both sequences.

Below the summary area is the colorful sequences area. For each sequence, the amino acid codes are displayed on the right, and their labels on the left. The order is always “query sequence” first, and “target sequence” second. The blue consensus ruler numbers and ticks any gaps found in the consensus.

Click the plus sign at the left of a row to reveal two tracks: **Gap Fraction** and **Ruler**. These default tracks cannot be edited, and additional tracks cannot be displayed.

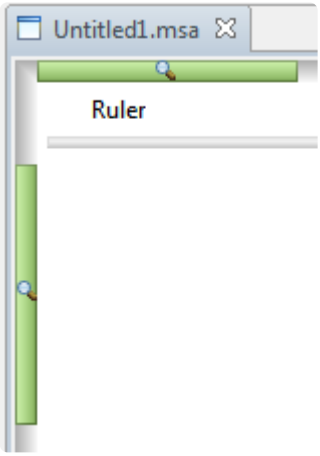
- **Gap Fraction** – If data is available, displays a graph track with a scale of 0-1 that shows the proportion of gapped positions in a sliding window centered on each position.

- **Ruler** – The black gap-aware sequence ruler uses pre-alignment numbering for the sequences.

<u>Tasks that pertain to the Pairwise view:</u>

Task	How To…
To create a new pairwise alignment using a different target sequence	Make a selection from the drop-down menu at the top of the view. The target sequence should be the same length or longer than the query. (Note that the **Query** drop-down menu has an unchangeable default sequence).
To zoom in or out	The Pairwise view features independent vertical and horizontal zoom in/out

Copyright DNASTAR, Inc. 2022

	sliders. Grab a green slider with your mouse and drag left/right or up/down to zoom in or out in the view. 
To display data tracks	See the topic Tracks panel.
To display search results for previous query (in a single job with multiple queries)	Use the **Previous query** tool ( ).
To display search results for the next query (in a single job with multiple queries)	Use the **Next query** tool ( ).
To filter results from a sequence search	Use the **Filter results** tool ( ). Filtering for this situation is described here.
To filter results from a text search	Use the **Filter results** tool ( ). Filtering for this situation is described here.
To add sequences or matches	Use the **Add sequences or matches** tool ( ). Sequences are added back into the host application, not the job's document.
To launch the Search Online dialog	Use the **Change search strategy** tool ( ). See Search for Sequences Online for details.
To save the search job to disk	Use the **Save as** tool ( ).
To export data from the view	Use the **Export data** tool ( ), described in the topic Export data from the Text, Table or Pairwise views.
To export an image of the alignment	Use the **Export alignment as image** tool ( ) in the upper right corner of the view. Images can be saved in PDF (*.pdf*), PowerPoint (*.pptx*), PNG (*.png*) or JPEG (*.jpg*) formats.
To change the layout of the	See Customize the look and layout.

Copyright DNASTAR, Inc. 2022

view or what is displayed in it	

Panels

Settings panels allow you to optimize the look, contents and layout of the Protean 3D view. To learn what can be done within each panel and their expandable sections, click the links below.

Panel name	Sections within the panel
Explorer	Molecules Surfaces
Features	Options Style
Tracks	Track options
Details	(none)
Style	Rendering Color Sequence Lighting
Models	(none)
Jobs	(none

To learn how to change the layout of panels, see Customize the look and layout.

Copyright DNASTAR, Inc. 2022

Explorer panel

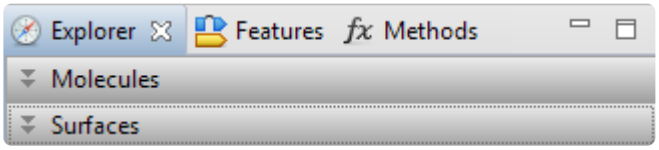
The Explorer panel, represented by a compass icon ( ), includes two expandable sections that allow you to view, select or hide structural elements and to apply surfaces to the structure.

Opening the Explorer panel:

Do any of the following:

- Click on the **Explorer** tab
- Choose **View > Explorer**
- Press **Ctrl+Alt+X** (Windows) or **Option+Cmd+X** (Macintosh)

Using the Explorer panel:



Each section of this panel is described in a separate topic:

- Molecules
- Surfaces

Copyright DNASTAR, Inc. 2022

Molecules section

The Molecules section is located in the Explorer panel lists the structures and components (chains, ligands and water) for each structure in the current document. If you display/hide/select items in the Molecules section, these choices are simultaneously displayed/hidden/selected in the other Protean 3D views.

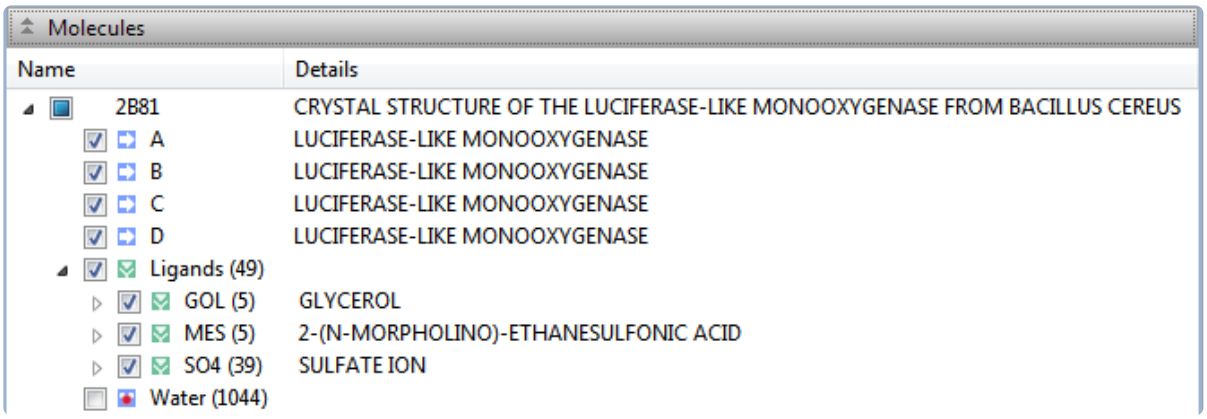
Opening the Molecules section:

Do any of the following:+

- Click the expand bar entitled “Molecules”
- Use **View > Explorer > Molecules**
- Press **Ctrl+Alt+M** (Windows) or **Option+Cmd+M** (Macintosh)

> **Note**: The Molecules section can only be opened if a document is open in the Protean 3D application window.

Using the Molecules section:



All independent protein chains, DNAs, RNAs, small molecules and metal ions bound to the complex are listed separately in tree format. Expand the arrows to reveal the elements of each structure. Each molecule is named by its common chemical name and by the chemical formula, where possible.

Chain names and details are taken directly from the PDB file. Protean 3D automatically generates chain names when none are available in the file, as is often the case in ATOM-only files.

- When viewing a sequence file, such as a GenBank file, the Molecules section displays only the file name and a list of sequence or locus names.

- When viewing a multiple-structure document, including prediction model/template alignments launched from the NovaFold Report, each structure or template protein fragment appears as a

separate section.

Icons in the Molecules section:

Icon	Function
	Denotes a protein chain.
	Denotes a DNA chain.
	Denotes an RNA chain.
	Denotes a ligand.
	Denotes water.
	A checkmark denotes that the element is currently displayed in the Protean 3D views . All checkboxes are initially checked by default, but can be manually unchecked to hide various structural elements in the Structure view and to hide sequence in the Sequence view. Expand arrows to the left of the checkboxes can be used to reveal/hide additional structural elements. Expand arrow icons vary by operating system and may not be visible unless the mouse is hovering above them.

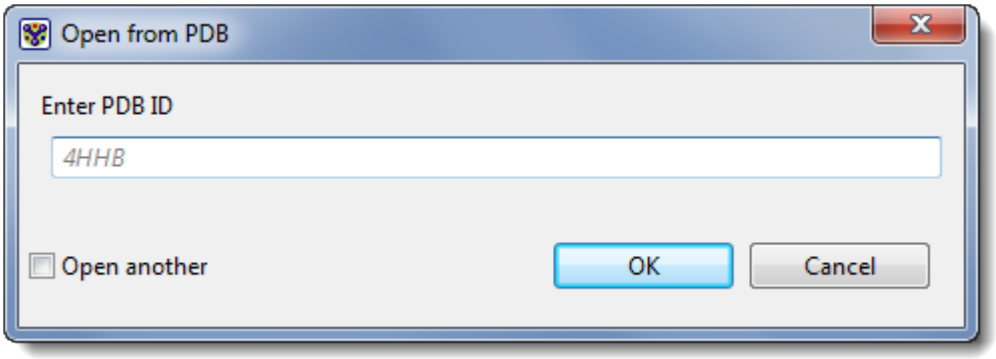
Copyright DNASTAR, Inc. 2022

Try it! – Use the Molecules section to select and hide a structural element

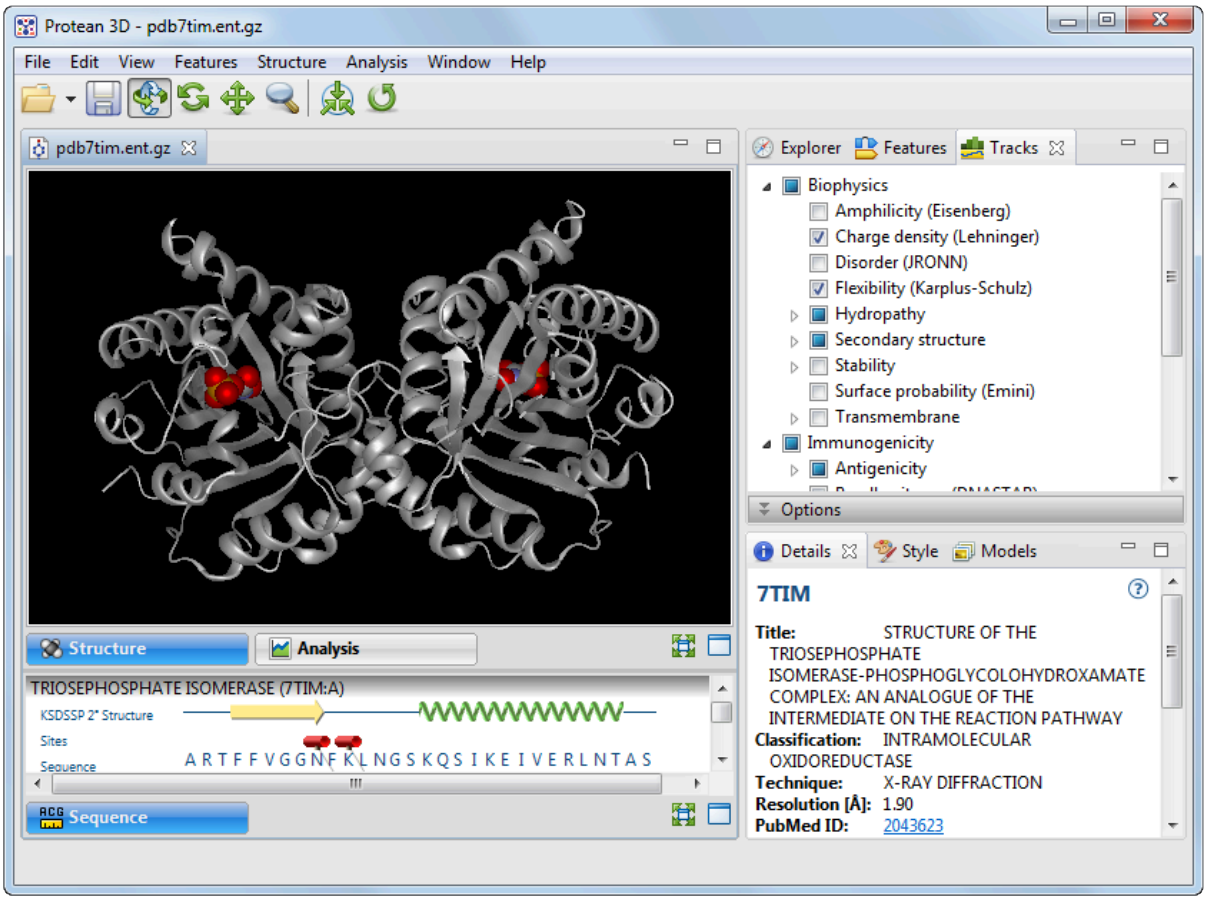
This tutorial shows how to select a particular structural element from the Molecules section.

1. Within Protean 3D, choose **File > Open from PDB**. You may initially see a yellow warning icon ( ) prompting you to type in a valid ID.

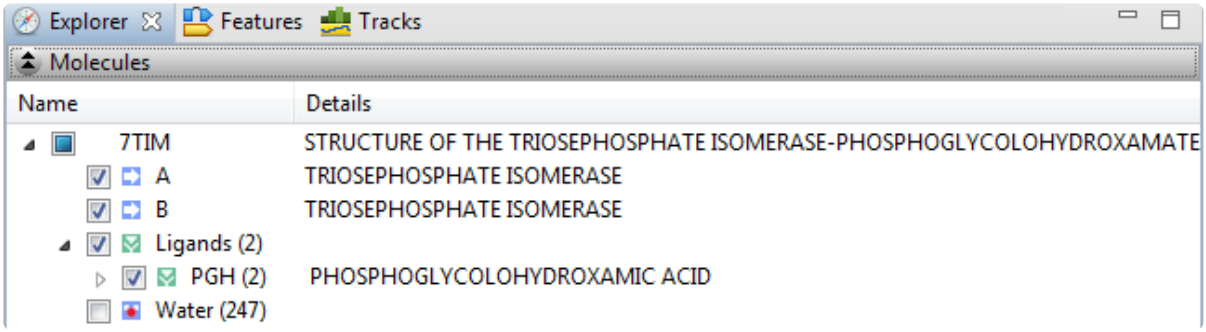
2. Type **7TIM** into the text box.



3. Press **OK** to open the structure in Protean 3D.

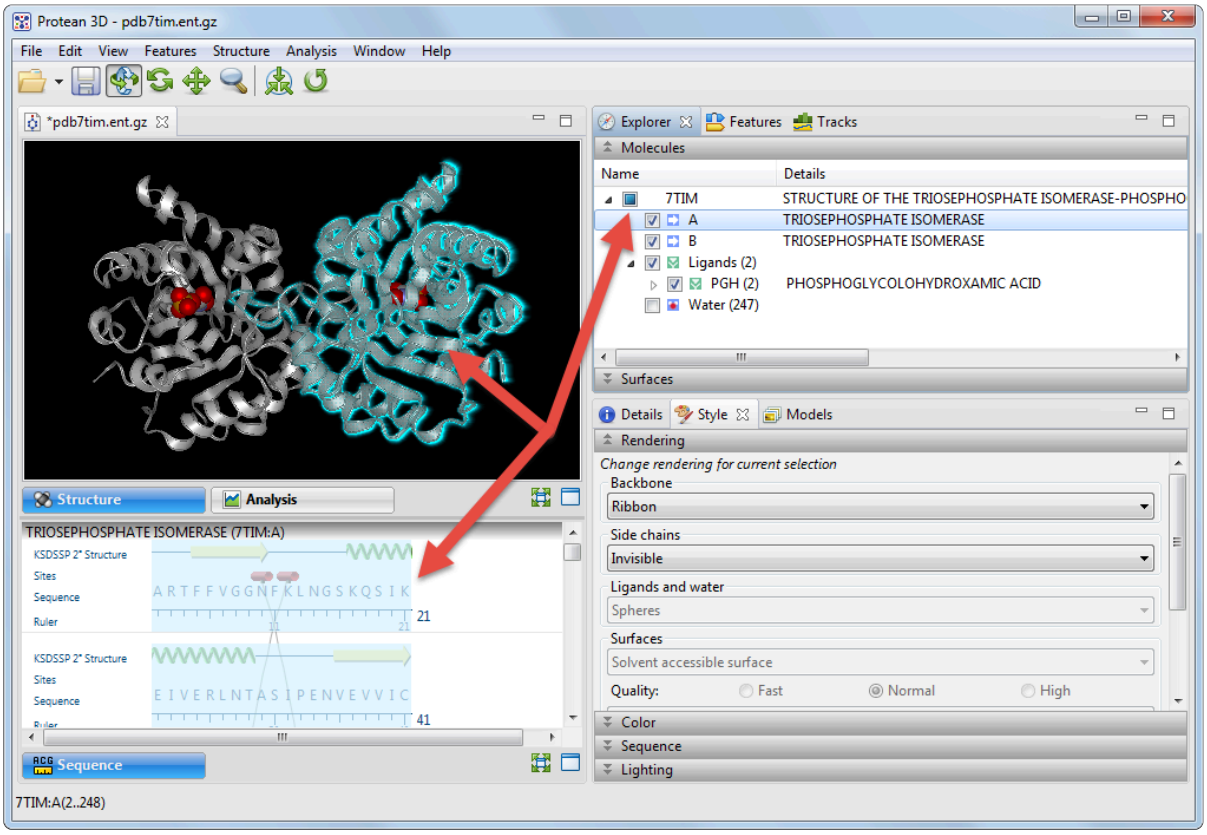


4. In the Explorer panel, open the Molecules section. Alternatively, open this section using the command **View > Explorer > Molecules**.

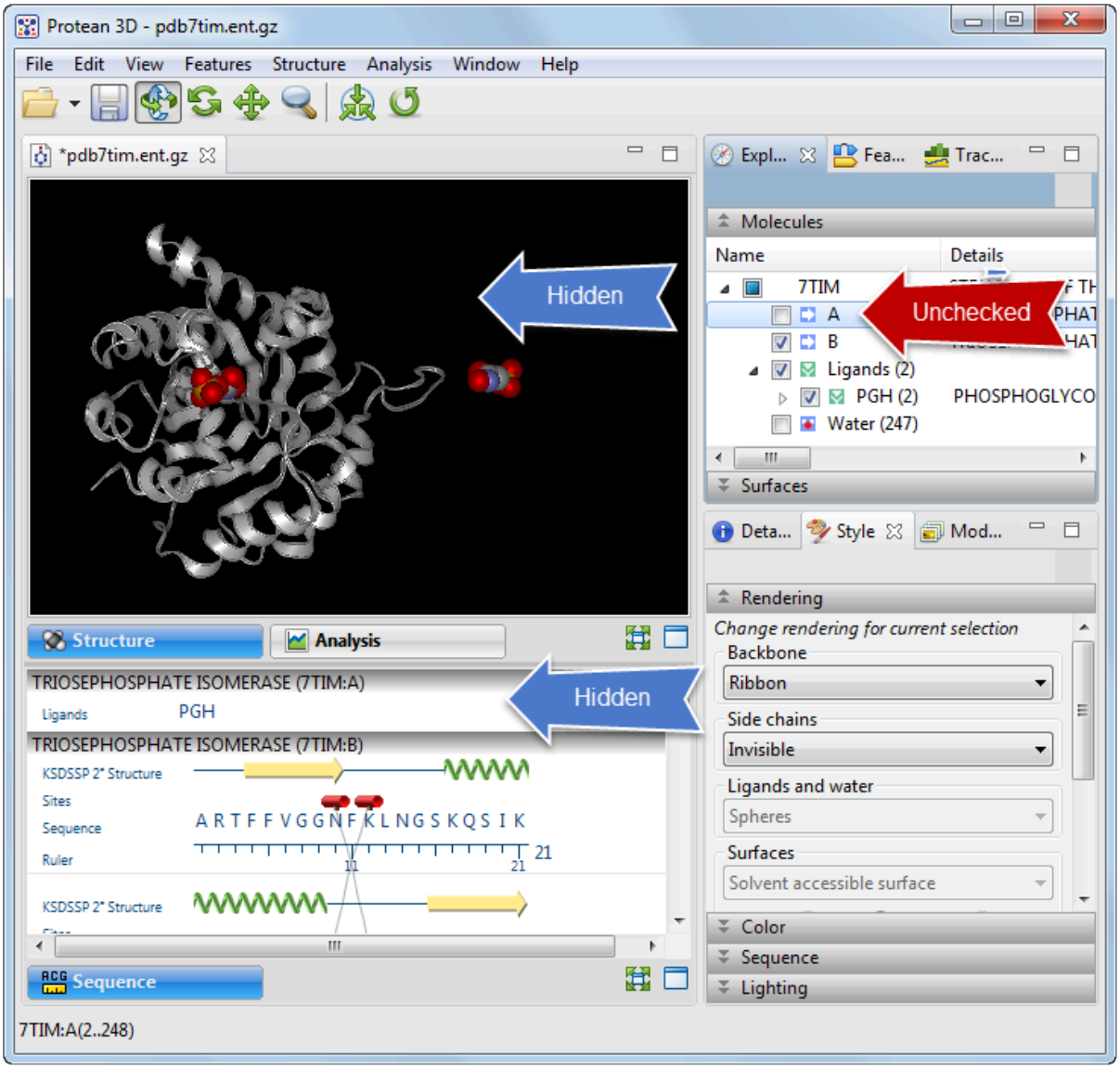


Notice that, by default, all boxes are checked except for **Water**.

5. Select the first *Triosephosphate Isomerase* subunit by clicking on the letter "*A*" (not on its checkbox). Observe the corresponding selections in the Structure and Sequence views.



6. Uncheck the box to the left of "*A*." Observe that the analogous *Triosephosphate Isomerase* is now hidden in both views.



7. To re-display the hidden elements, add check marks next to all boxes except **Water**.

Surfaces section

The Surfaces section is located in the Explorer panel and is used to create molecular or solvent-accessible surfaces for selections in the Structure view.

The following video shows the Surfaces section in action:

Opening the Surfaces section:

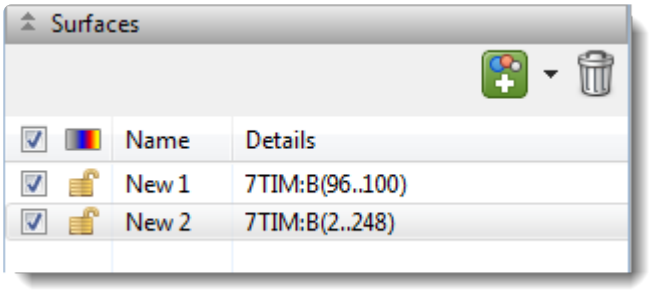
Do any of the following:

- Click the expand bar entitled **Surfaces**
- Use **View > Explorer > Surfaces**
- Press **Ctrl+Alt+S** (Windows) or **Option+Cmd+S**

Note: The Surfaces section can only be opened if a document is open in the Protean 3D application window.

Using the Surfaces section:

Copyright DNASTAR, Inc. 2022



Conceptually, molecular surfaces are mapped by "rolling" a spherical probe (typically the size of a water molecule) over the molecule's surface. When a surface is applied in Protean 3D, a grid-based approach is used to calculate molecular surface coordinates that approximate the molecular surface, following the method of Xu and Zhang.

- **Molecular surface** (Solvent-excluded surface) – The surface coordinates display the predicted point at which water/solvent would be excluded from surface interactions.
- **Solvent-accessible** surface – The surface coordinates display the predicted point at which water/solvent would just be able to interact with the molecule's surface. The volume of the **Solvent-accessible surface** is always greater that of the **Molecular surface**.

Tools in the Surfaces section:

Tool	Name	Function
	Create a surface	Creates a surface for the currently selected item(s); or for all chains, if nothing is selected. Use the associated drop-down menu to select the surface style and speed. The current default surface appears at the top of the list. Clicking the icon without using the menu will create a surface using the current default surface style and speed. To view samples of each surface style and speed, see Examples of surface types.
	Delete a surface	Deletes the currently selected row and removes the surface from the Structure view.

Tasks that can be performed within the Surfaces section:

Task	How To
To wrap surfaces around all chains only	Double-click on the Structure view background to de-select any elements. Then click on the **Create a surface** tool ( ) to apply the same surface type that you last used. To choose a different surface type, instead click on the arrow to the right of the icon and choose a surface type from the resulting menu.

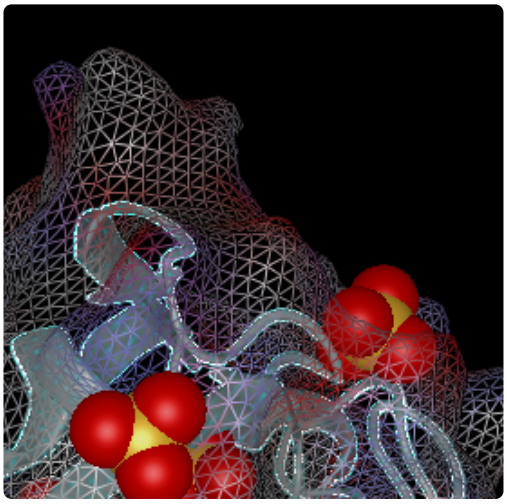
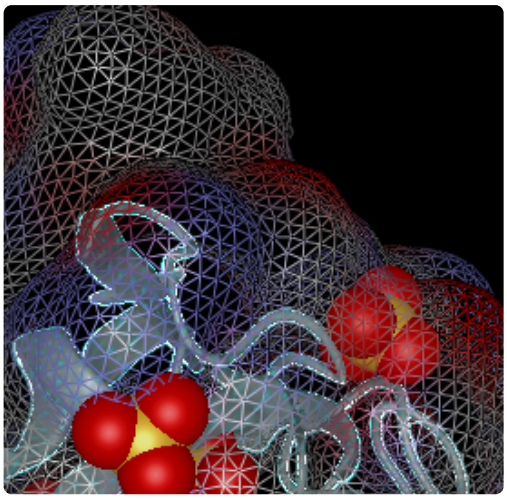
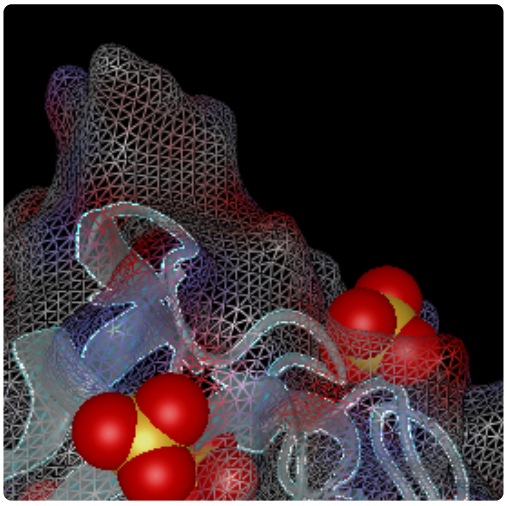
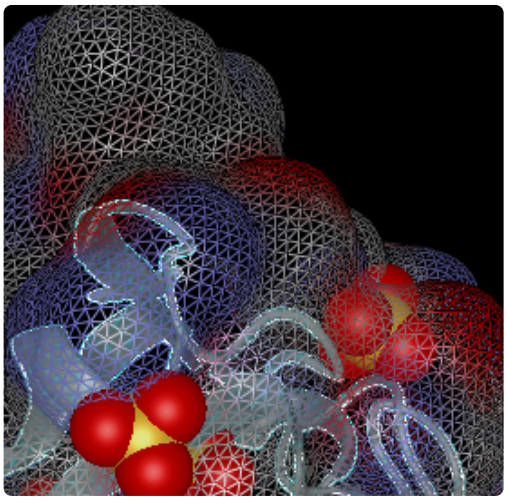
Copyright DNASTAR, Inc. 2022

To wrap surfaces around a selection	Select a residue, ligand, chain, or whole structure. Then click the **Create a surface** tool ( ) to apply the same surface type that you last used. To choose a different surface type, instead click on the arrow to the right of the icon and choose a surface type from the resulting menu.
To toggle between displaying/ hiding an existing surface	Check or uncheck the box at the left of the row.
To toggle between displaying/ hiding all surfaces	Check or uncheck the box on the left of the header row.
To change the color of an existing surface	By default, the color of the surface matches the color of the underlying structural component, including colors applied by the use of **Color Styles** in the Color section. To change the color of the surface (and its underlying structure), click on the surface to select it, then change its **Fill** color in the Color section.
To lock the color of an existing surface	When you change **Color Styles** in the Color section, the color of the surface will also change unless you "lock" it first in the Surfaces section by clicking the associated padlock icon.
To change the name or details for an existing surface	Double-click in the **Name** or **Details** column and type in the desired text.
To view details about an existing surface	Click on the surface in the **Name** column, then open the Details panel and look in the Structural Properties section. "Surface section in current selection" displays the name of the selection and the surface area in square angstroms.
To change the style of an existing surface	Click on the surface to select it, then change the Rendering section settings.

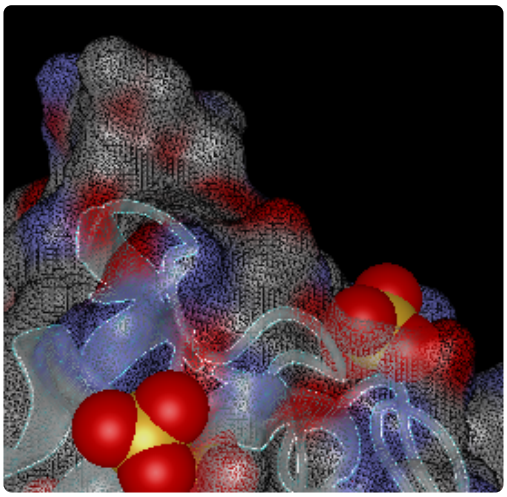
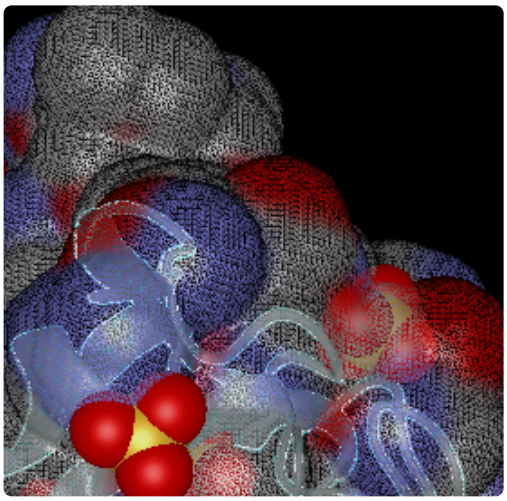
Copyright DNASTAR, Inc. 2022

Examples of surface types

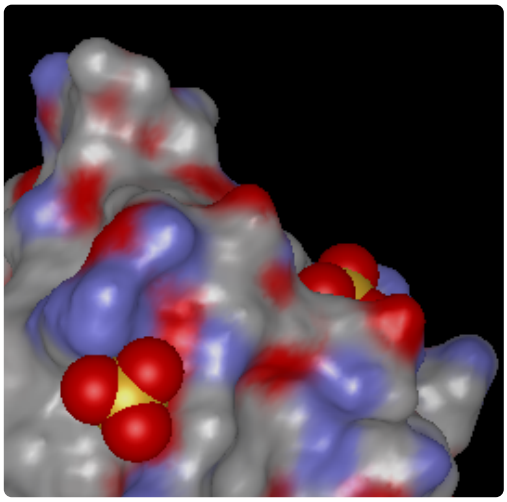
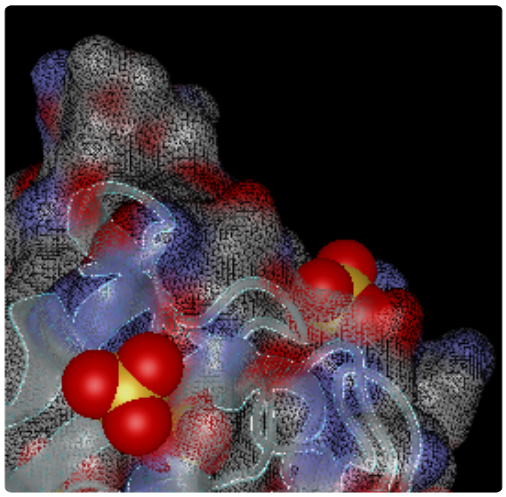
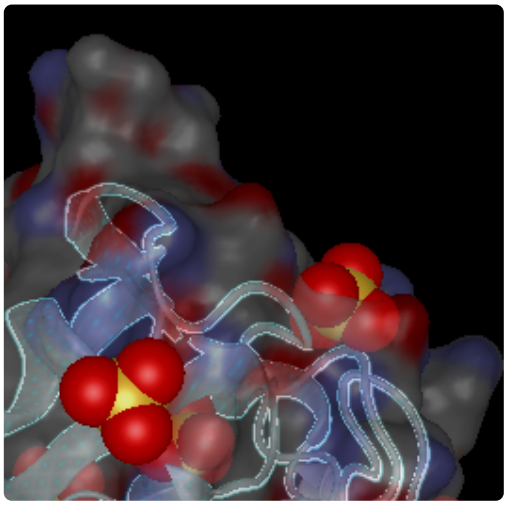
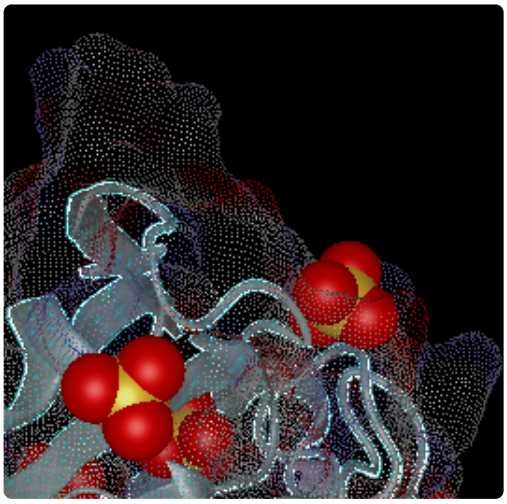
Two types of surfaces and three levels of surface resolution are available, and are controlled via options in the Surfaces and Rendering sections. The following images show the identical portion of a structure using the six different rendering choices. All images are shown with the "Mesh" surface type (selected in the Rendering section, as described below this table).

Molecular surface - Fast		Solvent accessible surface - Fast	
Molecular surface – Normal quality		Solvent accessible surface – Normal quality	

Copyright DNASTAR, Inc. 2022

Molecular surface – High quality		Solvent accessible surface – High quality	

The next set of images show the identical portion of a structure using four different combinations of rendering styles and transparency. Any style may be shown as 0-100% transparent. These options are set by opening the Rendering section and using the lower-most drop-down menu and the Transparency slider.

Solid		Mesh	
Solid (50% Transparency selected)		Points	

Copyright DNASTAR, Inc. 2022

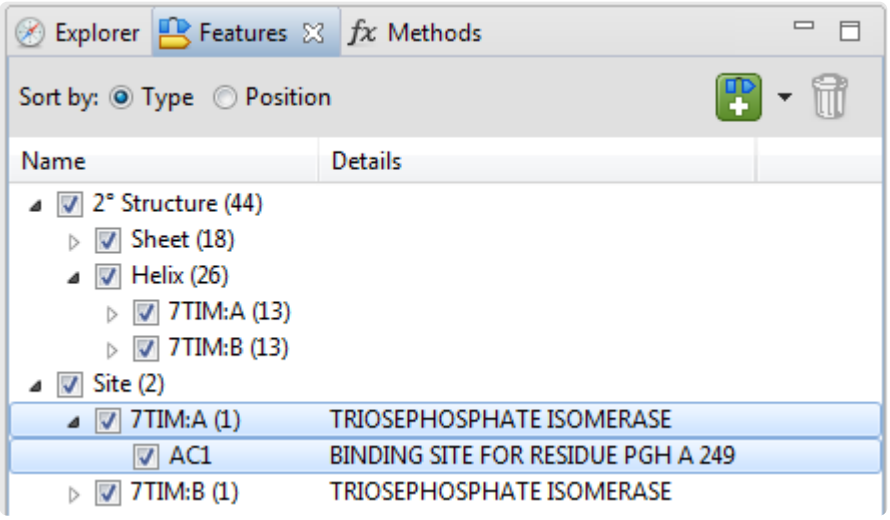
Features panel

The Features panel, represented by a “feature” icon ( ), shows the incidence of sequence features by type. This panel consists of an expandable tree listing all sites, disulfides (S-S bonds), helices, and sheets from the original PDB file, as well as Protean 3D’s calculation of secondary structure. In addition, it includes two expandable areas that allow you to specify options and styles for features in non-PDB files.

Opening the Features panel:

Do any of the following:

- Click on the **Features** tab
- Use **View > Features > Features** .
- Press **Ctrl+Alt+F** (Windows) or **Option+Cmd+F** (Macintosh)

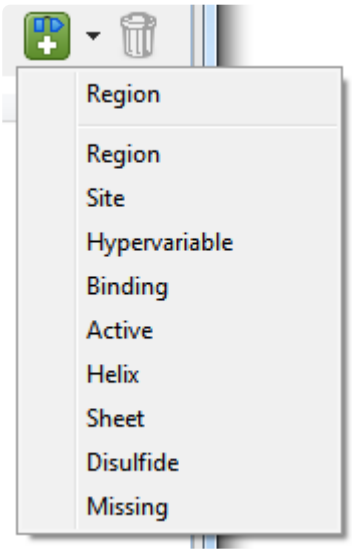


Note: The Features panel is blank unless a document is open in the Protean 3D application window.

Using the Features panel:

Task	How To
To create a new feature	1) In any Protean 3D view, select the desired range for the feature. The selection can be discontinuous, but must be on a single molecule; it cannot span molecules. If you inadvertently create a feature from selections on multiple chains, multiple features with the same type and name will be added to the table.

Copyright DNASTAR, Inc. 2022

	2) Use the **Create feature** tool ( ). Clicking directly on the tool image adds a new “Region” feature. To create a feature of a different type, click on the arrow to the right of the tool and make a selection from the drop-down menu.  A feature named **New feature 1** is created and is displayed graphically in the Analysis view. (User-created features are never visible in the Sequence view.) Subsequent features created in the same way are numbered incrementally. 3) If desired, edit the feature in the (feature) Options area.
To delete a feature	To delete a feature that was created using the above method, select it from the list in the Features panel and press the **Delete feature** tool ( ) or use the **Features > Delete Feature** menu command. Features that were imported with the sequence or structure cannot be deleted.
To order the tree alphabetically by feature type	Choose the **Type** radio button.Upon expanding a feature type, the features are sorted alphabetically by chain or sequence, and subsequently, by feature name.
To order features into their respective chains or sequences	Choose the **Position** radio button. Those features that traverse multiple chains are categorized as "Intermolecular." Upon expanding a chain or sequence, the features are organized by the order in which they appear in the chain or sequence.
To expand the feature list	Click the arrows to the left of items in the **Names** column. Expand arrows to the left of the check boxes can be used to reveal/hide additional elements. Expand arrow icons vary by operating system and may not be visible unless the mouse is hovering above them.

Copyright DNASTAR, Inc. 2022

	Note that all structure files have two secondary structure assignments. The PDB secondary structure is the one annotated in the original PDB file. The KSD secondary structure is an automatic secondary structure assignment calculated by Protean 3D using the Kabsch and Sander DSSP method. See Work with Features for additional information.
To show/hide feature annotations in Sequence & Analysis views	Check/uncheck the box next to the feature ( ). Note that changing the visibility of any helix or sheet affects *all* helices and sheets in both views.

The two areas related to the Features panel are described in separate topics:

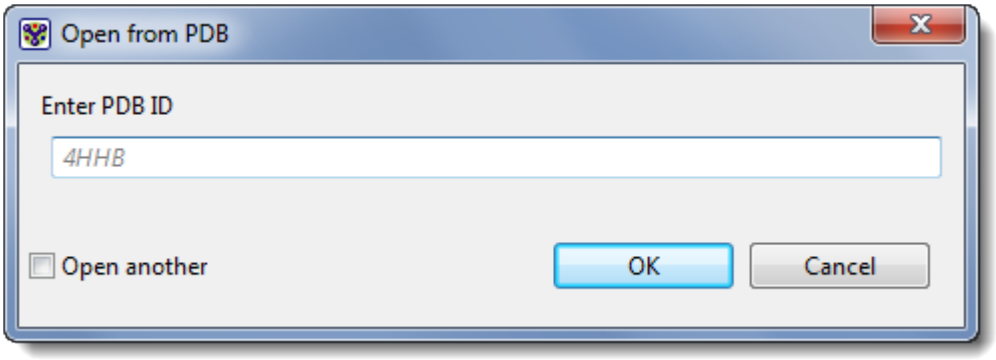
- Options
- Style

Try it! – Select and recolor a feature, and change feature options

In this tutorial, you will use the Features panel to select a feature to recolor. Next, you will edit some of the feature options.

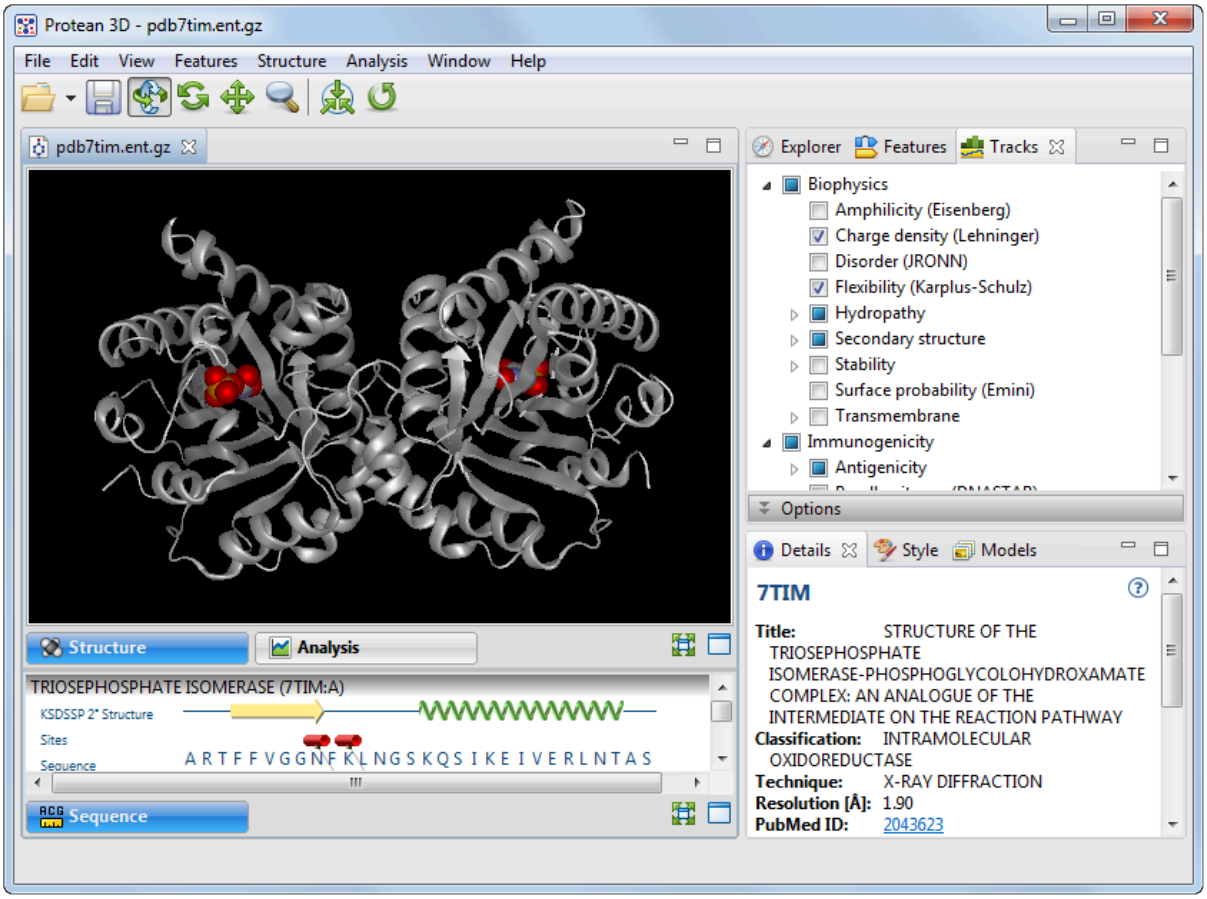
1. Within Protean 3D, choose **File > Open from PDB**. You may initially see a yellow warning icon ( ) prompting you to type in a valid ID.

2. Type **7TIM** into the text box.



3. Press **OK** to open the structure in Protean 3D.

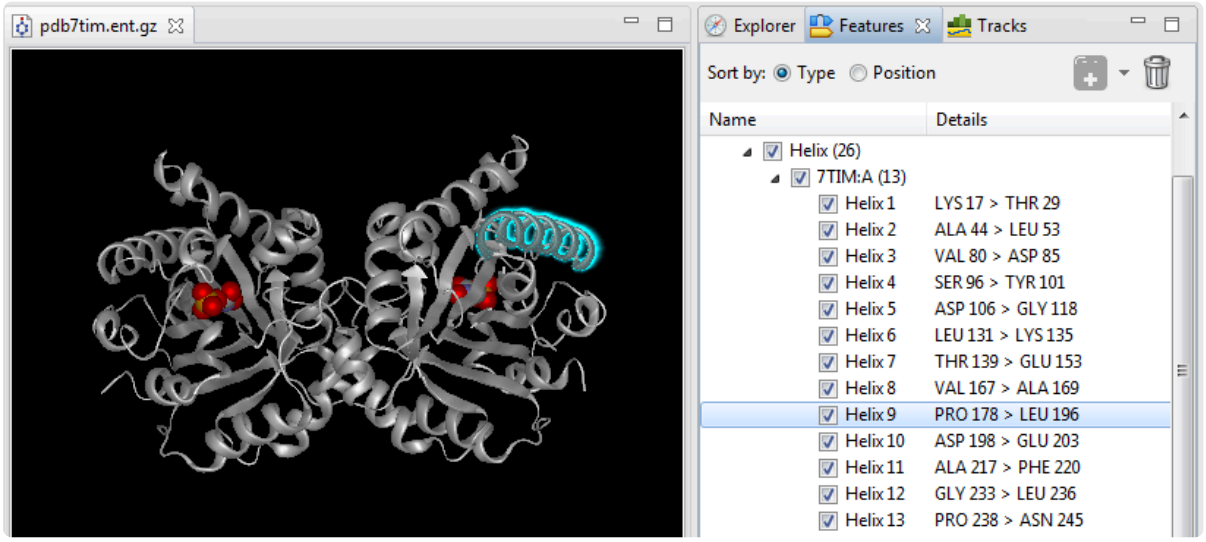
Copyright DNASTAR, Inc. 2022



4. Open the Features panel using **View > Features > Features**, or click on the **Features** tab to make it the active panel.

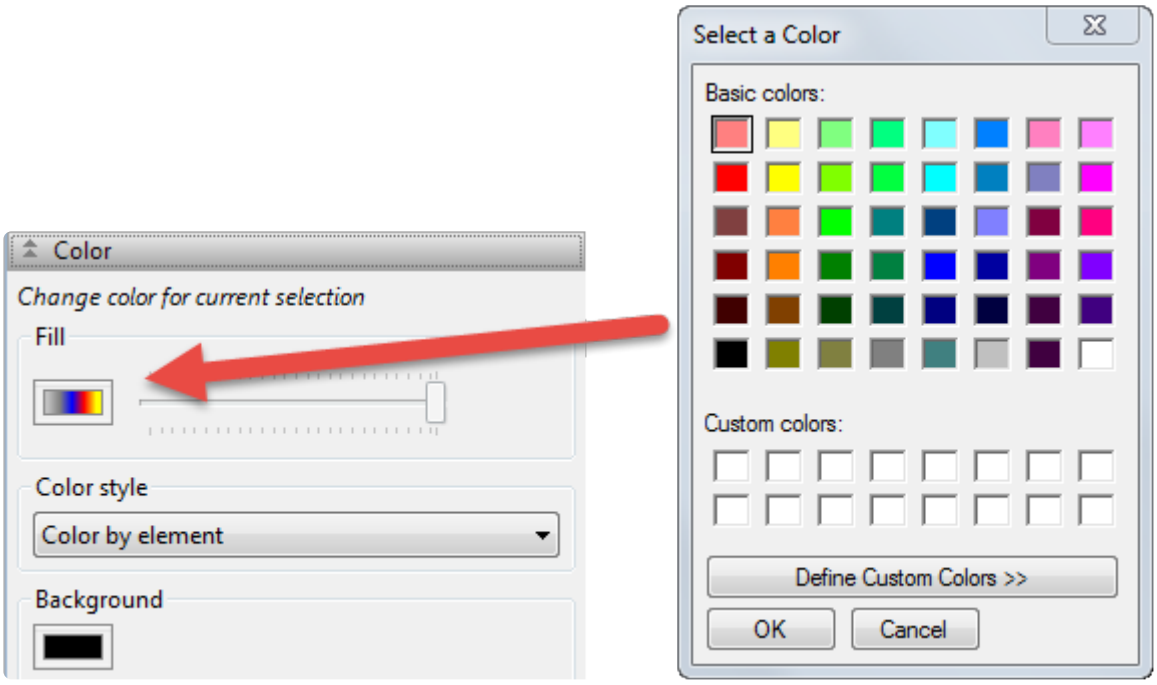
5. Use the expand arrows to expand **KSDSSP 2° Structure** > **Helix (26)** > **7TIM:A (13)**. If the expand arrows are hidden, hover over them with the mouse.

6. Select the longest helix, **Helix 9**, by clicking on its name.

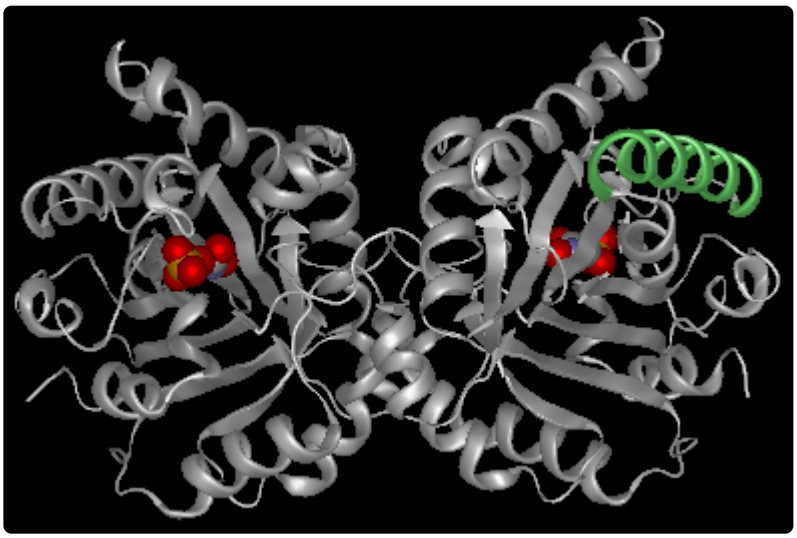


Copyright DNASTAR, Inc. 2022

7. Select **View > Style > Color** from the menu. In the Color section, click on the box under the word **Fill** to open the color picker for your operating system.

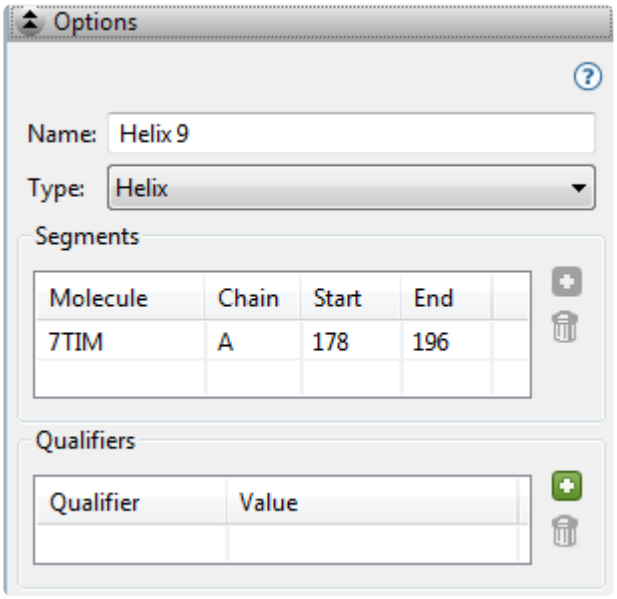


8. On Windows, choose any shade of light green and click **OK** to close the window and return to Protean 3D. On Macintosh, you must close the color picker after selecting the light green color. Double-click on the black background to unselect Helix 9. Observe that Helix 9 now appears in green.



9. To edit the feature further, select Helix 9 again in the Features panel, then choose **View > Features > Options**. Information about the selected feature is shown.

Copyright DNASTAR, Inc. 2022



10. Within the Options section:

    a. Change the **Name** of the feature to “Longest helix.”

    b. Insert your cursor in the **Start** column and change the starting point of the feature from **178** to **177**.

    c. In the Qualifier section, click on the **Add a Qualifier** tool ( ) to add a blank /note. In the **Value** column, type “Extended left coordinate from 178 to 177.”

    d. Click on the arrow to the right of **/note** to reveal a drop-down menu. Choose any qualifier type from the list.

Copyright DNASTAR, Inc. 2022

(Feature) Options section

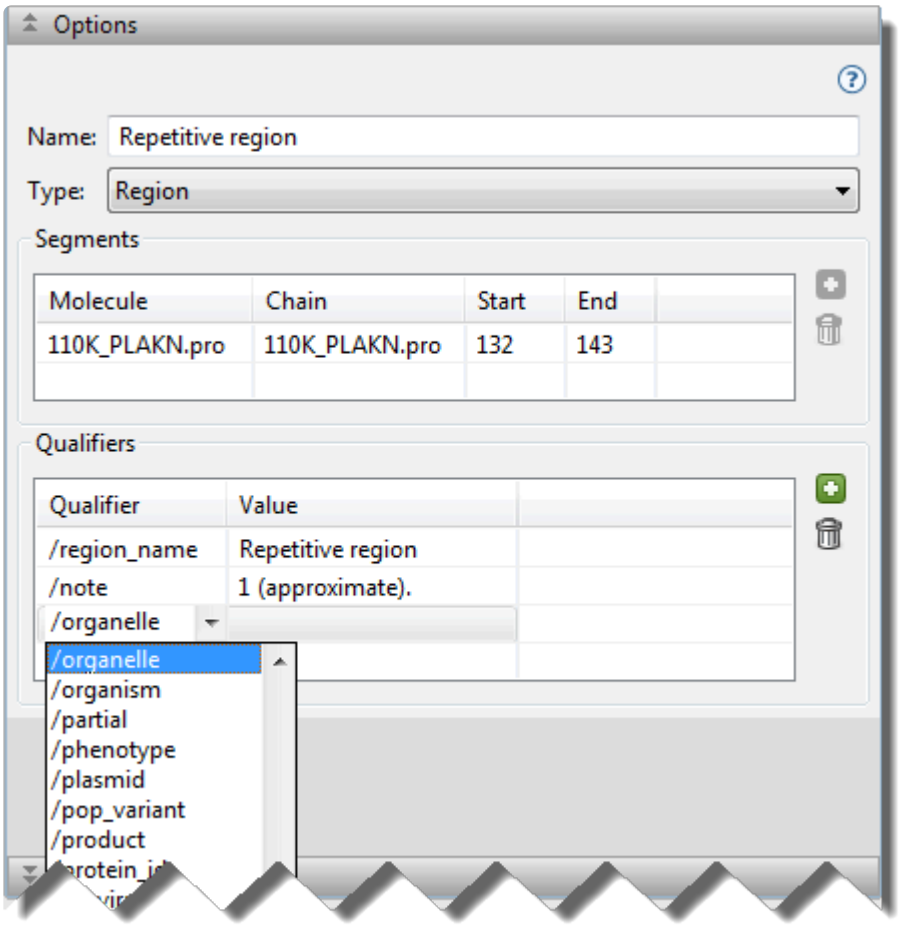
The Options section is located in the Features panel and, in some circumstances, lets you edit the bounds of secondary structure elements and/or change their descriptions.

Opening the Options section:

Do either of the following:

- Click the expand bar entitled “Options”
- Use **View > Features > Options**

The Feature Options section appears similar to the image below:



Tools in the Options section:

Tool	Name	Description
	Create a segment	To create a new feature segment from a selected range of sequence.

	Delete a segment	To delete the selected segment.
	Add a qualifier	To create a new qualifier for a selected feature.
	Delete a qualifier	To delete the selected qualifier.

Tasks that can be done within the (feature) Options section:

Before performing any of the tasks below, you must first select one or more items from the Features panel or in the (feature) Options section itself. If you select a sequence file feature, user-created feature, or a structural element in a PDB file, the (feature) Options dialog is enabled. If you select a PDB site, however, all options in the dialog are disabled.

Any changes made in this dialog are reflected in the **Name** and/or **Details** columns of the Features panel and on the Features track, if it is displayed in the Analysis view.

Task	How To
To change the feature name	Overwrite the default name in the **Feature name** box. **Note:** If multiple features are selected in the Features panel, you can change the name and type for all of them simultaneously. The effect is similar to what will happen in a true multiple chain feature, except that the individual features are displayed in the Features panel.
To change feature type	Make a selection from the **Feature type** drop-down menu. Available options include **Region**, **Site**, **Hypervariable**, **Binding**, **Active**, **Helix**, **Sheet**, **Disulfide**, and **Missing**. When you make a new selection, the feature automatically moves to the corresponding category in the Feature panel. **Note:** If multiple features are selected in the Features panel, you can change the name and type for all of them simultaneously. The effect is similar to what will happen in a true multiple chain feature, except that the individual features are displayed in the Features panel.
To change the start and end points for the feature or one of its segments	Insert the cursor into a **Start** or **End** cell and overwrite the original entry with a new number.
To add a new segment to	Select the target feature in the Features panel. Then, from any Protean 3D view, select the desired range for the new segment. The selection must be on the same chain as the original feature. In the Options section's Segment portion, click the **Create a segment** tool ( ) to add

an existing feature	a new segment row to the table. This tool is only enabled when the current selection is different from the existing segments. Note that after creating a new segment, you may need to manually reselect the feature in the Feature panel before adding another segment.
To delete a segment	To delete a segment from a multi-segment feature (i.e., two or more segments on a single chain), select the corresponding row in the Options section's Segment portion and press the **Delete a segment** tool ( ). If the feature only has one segment, this tool is disabled.
To add a new qualifier to an existing feature, or to edit an existing qualifier	Select the target feature in the Features panel. In the Qualifier section, click the **Add a qualifier** tool ( ) to add a new row to the table. In the **Qualifier** column of the new row, use the drop-down menu to choose the type of qualifier from a list. In the **Value** column, type in the desired text. Edits can be made to existing qualifiers in the same way.
To delete a qualifier	To delete a qualifier, select the corresponding row in the Qualifiers section and press the **Delete a qualifier** tool ( ).

Copyright DNASTAR, Inc. 2022

(Feature) Style section

The Style section is located in the Features panel and provides options for customizing the look of features in the Analysis view.

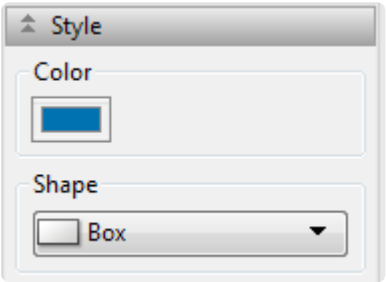
Opening the Style section:

Do either of the following:

- Click the expand bar entitled “Features”
- Use **View > Style > Style**

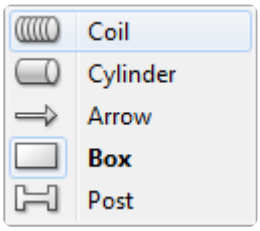
The Style sectionis only enabled when all of the following are true: 1) a document is open in the Protean 3D application window, 2) a feature is selected, 3) and the selected feature is not a strand, helix or PDB-derived feature. If the section is disabled, try selecting a different type of feature. Alternatively, you can create a new feature by selecting a portion of the sequence and choosing **Features > New Feature**.

Using the Style section:



- The current color used for the feature is shown in the **Color** box. To choose a different color, click on the box and make a new selection using the color chooser for your operating system.

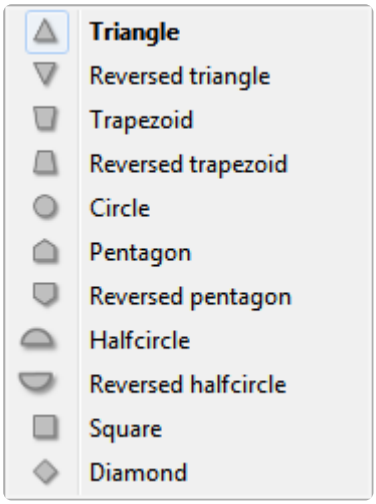
- To change the shape for a region feature, first select an individual region feature in the Features panel, then make a selection from the **Shape** drop-down menu. When a site feature is selected, this menu is disabled.



- To change the shape for a site feature, first select an individual site feature in the Features panel, then make a selection from the **Pin** drop-down menu. When a region feature is selected, this menu is

Copyright DNASTAR, Inc. 2022

disabled.


Triangle
Reversed triangle
Trapezoid
Reversed trapezoid
Circle
Pentagon
Reversed pentagon
Halfcircle
Reversed halfcircle
Square
Diamond

Tracks panel

The Tracks panel, represented by a graph icon ( ), provides a list of prediction tracks sorted by classification. Prediction tracks are used to evaluate the sequence and structural characteristics of a protein. Results of tracks chosen in this panel are displayed in the Analysis or Pairwise views on a common horizontal scale. The Tracks panel includes one expandable section for setting options for a selected track type.

Opening the Tracks panel:

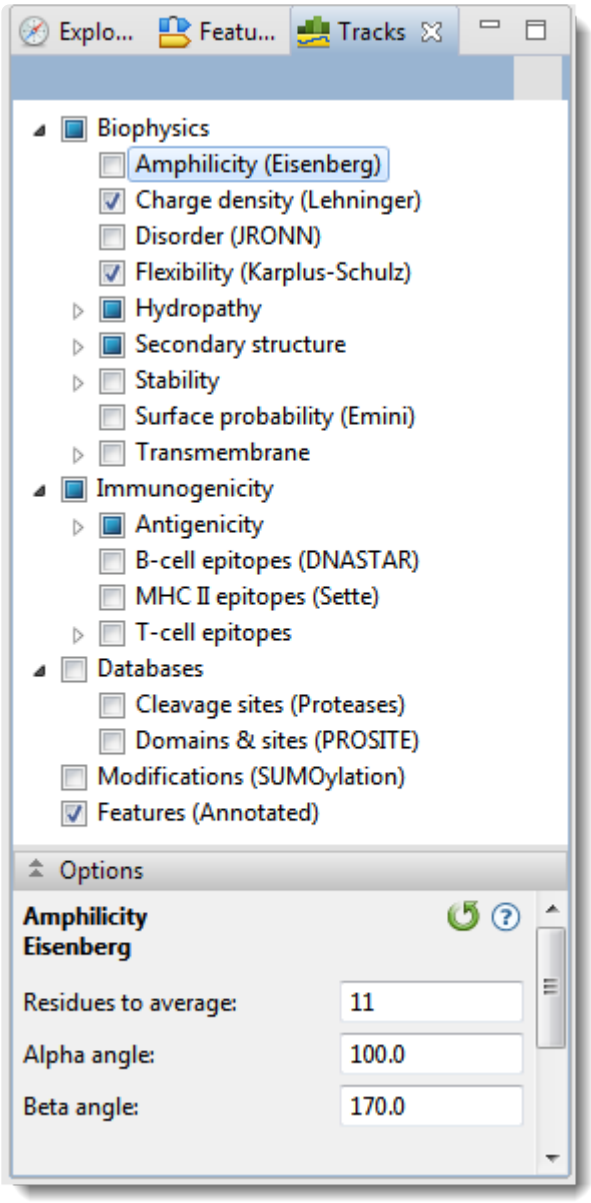
Do any of the following:

- Click on the **Tracks** tab
- Use **View > Tracks > Tracks**
- Press **Ctrl+Alt+T** (Windows) or **Option+Cmd+T** (Macintosh)
- Double-click on a synopsis or channel in the Analysis view; this simultaneously selects the method and opens the corresponding parameter dialog.

Note: The Tracks panel is blank unless a document is open in the Protean 3D application window.

Using the Tracks panel:

Copyright DNASTAR, Inc. 2022



Task	How To
To reveal/hide sub-tracks	Click on the expand arrows to the left of some tracks (e.g., *Hydropathy*).
To learn about the available tracks and to modify options for the selected track	Use the Track Options section. Some options allow you to type in values. Note that if you type in a disallowed value, a small red/white “X” will appear, and a message balloon will notify you of the allowable range.
To apply a prediction track to the active sequence	Check the box to the left of the method name. Checking or unchecking a top-level box will cause its sub-tracks to also be checked/unchecked. Prediction results are displayed in the Analysis view.
To view details about a track	Select the track by clicking on its name and then open the Details panel.

Copyright DNASTAR, Inc. 2022

Track Options section

The Track Options section is located in the Tracks panel and provides an editable display of options for a selected Protean 3D track.

Opening the Track Options section:

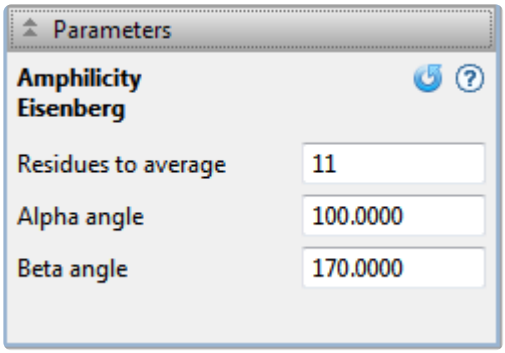
Do either of the following:

- Click the expand bar entitled “Options”
- Use **View > Tracks > Options**

The Track Options sectionis only active when a document is open in the Protean 3D application window and when a method has been selected in the Tracks panel.

Using the Track Options section:

This section contains very different options, depending on the selected track. An example is shown below:



To learn about options available for specific tracks, see:

Biophysics tracks	Immunogenicity tracks	Database tracks	Other track types
Amphiphilicity Charge Density Disorder (JRONN) Flexibility **Hydropathy:** Hopp-Woods Kyte-Doolittle Parker	B-cell epitopes MHC II epitopes **Antigencity:** Jameson-Wolf Welling **T-cell epitopes:** AMPHI Rothbard-Taylor	Cleavage sites Domains & sites	Modifications Features NovaFold-related

Copyright DNASTAR, Inc. 2022

Secondary structure: Chou-Fasman Deleage-Roux Garnier-Robson Coiled coil **Stability:** Aliphatic index Instability index Isoelectric precipitate Surface probability **Transmembrane:** Goldman-Engleman-Steitz TopPred von Heijne			

To return to the default settings for any Track Options dialog, click .

Copyright DNASTAR, Inc. 2022

Biophysics tracks

The Tracks panel lets you choose which biophysics tracks to apply to the Analysis view. The following links describe the methods and explain how to change their Track Options settings, if you desire to do so:

- Amphiphilicity
- Charge Density
- Disorder (JRONN)
- Flexibility
- **Hydropathy:**
  - Hopp-Woods
  - Kyte-Doolittle
  - Parker
- **Secondary structure:**
  - Chou-Fasman
  - Deleage-Roux
  - Garnier-Robson
  - Coiled coil
- **Stability:**
  - Aliphatic index
  - Instability index
  - Isoelectric precipitate
  - Surface probability
- **Transmembrane:**
  - Goldman-Engleman-Steitz
  - TopPred
  - von Heijne

Copyright DNASTAR, Inc. 2022

Amphiphilicity (Eisenberg)

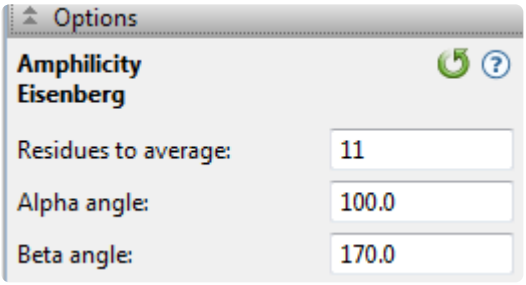
The Amphiphilicity (Eisenberg) track predicts the Eisenberg Moment using the approach of Eisenberg et al., 1984. Hydrophobic moments are semi-empirical quantities based on computation and experimental measurements, which describe the distribution of hydrophilic and hydrophobic residue groups in a protein. They describe asymmetry of hydrophobicity, or amphiphilicity.

To apply this track to the sequence:

In the Tracks panel, check the box next to *Amphiphilicity (Eisenberg)*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- **Residues to Average** - the number of residues to be averaged for the hydrophobic moment and hydrophobicity. The default is 11.
- **Alpha Angle** and **Beta Angle** - the angles used in moment and amphipathic region plots to assign emergence angles for successive side chains. Only change these options if you are looking for a structure with atypical α-helix and ß-sheet angles, such as a 3-10 helix. The defaults are 100 and 170, respectively.

Click if you wish to return to the default values.

Copyright DNASTAR, Inc. 2022

Charge Density (Lehninger)

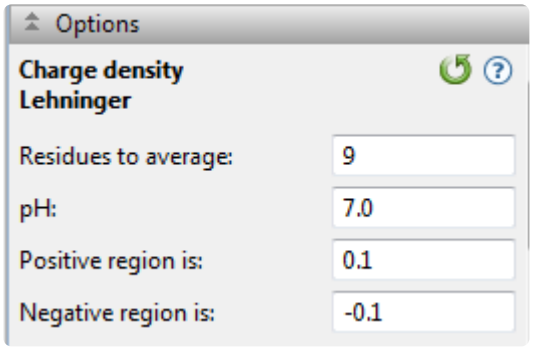
The Charge Density (Lehninger) track predicts regions of positive and negative charge by summing charge over a specific range of residues. DNASTAR developed this method using the pK tables of Lehninger, 2005. Since charged residues tend to lie on the surfaces of proteins, this method aids in predicting surface characteristics. A set of data is generated that can be displayed three ways: a line plot graph of average charge, a region graph of negative charge and another region graph of positive charge. The method is sensitive to small changes in parameter settings, so you may wish to set up multiple plots to compare how parameter changes affect each plot.

To apply this track to the sequence:

In the Tracks panel, check the box next to *Charge Density (Lehninger)*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- **Residues to Average** – the size of the window used to average the charge on consecutive residues. Inputting a high value will smooth the curve, while a low value will make the method more sensitive to charge differences between adjacent amino acids. The default is 9.
- **pH** – desired pH for calculating pK values for each residue. The default is 7.0.
- **Positive Region is** - the threshold value for a region to be considered positive. The default is 0.1.
- **Negative Region is** – the threshold value for a region to be considered negative. The default is -0.1.

Click if you wish to return to the default values.

Copyright DNASTAR, Inc. 2022

Disorder (JRONN)

The Disorder (JRONN) track predicts disordered regions of proteins, which are unlikely to form well-defined three-dimensional structures. These regions may affect both solubility and the ability to crystallize. DNASTAR developed this track based on the regional order neural network (RONN) analysis method of Troshin et al., 2011, and Yang et al., 2005.

To apply this track to the sequence:

In the Tracks panel, check the box next to *Disorder (JRONN)*. The track will now be visible in the Analysis view

To edit track options:

There are no editable options for this track.

Copyright DNASTAR, Inc. 2022

Flexibility (Karplus-Schulz)

The Flexibility (Karplus-Schulz) track predicts backbone chain flexibility using the approach of Karplus and Schulz, 1985. The method is useful for resolving antigenic sites, as these regions tend to be among the most flexible in a polypeptide sequence.

To apply this track to the sequence:

In the Tracks panel, check the box next to *Flexibility (Karplus-Schulz)*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:

- **Flexibility threshold (0..2)** – the value (between 0 and 2) where flexible residues are separated from rigid residues. In the default of 1, residues with Bnorm values below 1 are rigid and those above 1 are flexible. The most flexible residue is assigned a value of 2 and the least flexible is assigned 0.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

Hydropathy – Hopp-Woods

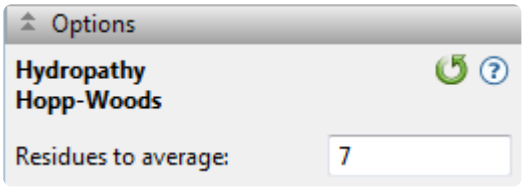
The Hydropathy – Hopp-Woods track predicts protein antigenic determinants by searching protein sequences for the area of greatest local hydrophilicity. Hopp and Woods, 1981, make two assumptions in using hydrophilicity to find antigenic determinants: 1) antigenic determinants are usually found on regions with a high degree of exposure to solvents and 2) antigenic determinants commonly possess charged hydrophilic side chains. Each residue is assigned a hydrophilicity value and these values are averaged over a hexamer window.

To apply this track to the sequence:

In the Tracks panel, expand *Hydropathy* and check the box next to *Hopp-Woods*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- **Residues to Average** – the size of the window over which hydropathy values are averaged. The Protean 3D default is 7, while the original authors use 6. There is very little difference in antigenic site prediction with this change, except that the larger default value requires a slightly larger antigenic site. The authors recommend setting this value between five and eight. If you set the value below five, too many regions may be falsely predicted. Conversely, values set above eight can create a smooth curve, making it difficult to locate areas of potential antigenic sites.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

Hydropathy – Kyte-Doolittle

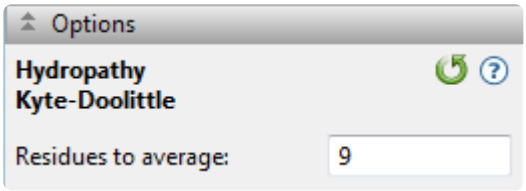
The Hydropathy – Kyte-Doolittle track predicts regional hydropathy of proteins from their amino acid sequences, using the approach of Kyte and Doolittle, 1982. Hydropathy values are assigned for all amino acids and are then averaged over a user-defined window. The average is plotted at the midpoint of the window. Residue hydropathy assignments are derived from water-vapor transfer free energies and the interior-exterior distribution of residue side-chains.

To apply this track to the sequence:

In the Tracks panel, expand *Hydropathy* and check the box next to *Kyte-Doolittle*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- **Residues to Average** – the size of the window over which hydropathy values are averaged. It is recommended you select 7 to 11 residues. The default is 9.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

Hydropathy – Parker

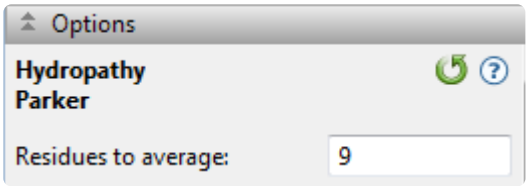
The Parker track uses the method of Parker et al., 1986, which is a hydrophilicity scale based on high-performance liquid chromatography retention times of model synthetic peptides. Hydrophilicity measures have been used extensively in the prediction of antigenic amino acid residues. The Parker method uses a modified Hopp-Woods algorithm together with a new set of hydrophilicity options.

To apply this track to the sequence:

In the Tracks panel, expand *Hydropathy* and check the box next to *Parker*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- **Residues to average** – number of residues to be averaged when constructing each point in the plot. The default is 9.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

Secondary Structure – Chou-Fasman

The Secondary Structure – Chou-Fasman track predicts secondary structure of proteins from the crystallographic structures of their amino acid sequences.

Secondary structure prediction methods from the1970s and 1980s rely on the propensities of individual amino acids observed in either alpha helices or beta strands/sheets. For a local region in the protein sequence, these propensity values are summed for each individual class. The secondary structure for central residue of the region is predicted by comparing whether propensity sum is greater than the threshold.

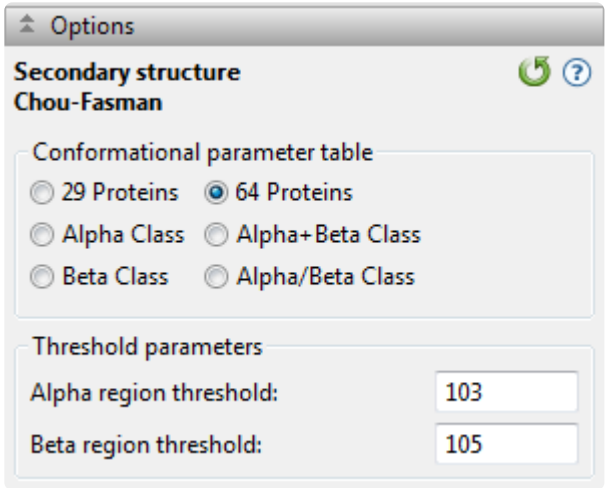
Chou and Fasman, 1978, identified several protein classes, with different data sets used to train the model for each class. The class for 29 Proteins, later augmented to 64 Proteins (Chou, 1990), is the more general model. The other four classes are more appropriate for specific types of protein folds: Alpha, Beta, Alpha plus Beta and Alpha/Beta (i.e., some combination of the two structural elements).

To apply this track to the sequence:

In the Tracks panel, expand *Secondary Structure* and check the box next to *Chou-Fasman*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



Choose which conformation parameter table to use:

- **29 Proteins** – to use parameters from only 29 protein structures.

- **64 Proteins** – to use parameters from all 64 protein structures.
- Alpha Class, Beta Class, Alpha plus Beta Class or Alpha/Beta Class* – to use the Chou classification method parameter tables.

If you select **29 Proteins** or **64 Proteins**, enter numbers in the Threshold parameters section:

- **Alpha region threshold** – P_α has a default of 103.
- **Beta region threshold** – $P_ß$ has a default of 105.

These values are constants which limit whether a region is defined as α-helix or ß-sheet. If $P_\alpha > 103$ and $P_\alpha > P_ß$ a region is predicted as helical. If $P_ß > 105$ and $P_ß > P_\alpha$ a region is predicted as a beta sheet. The default values are from Chou and Fasman (1978).

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

Secondary Structure – Deleage-Roux

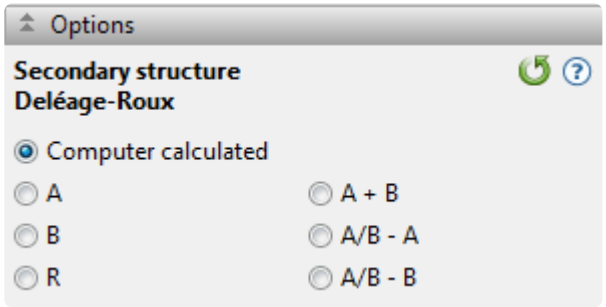
The Secondary Structure – Deléage-Roux track predicts secondary structure using the approach of Deléage and Roux, 1987. The method uses an independent prediction of the protein's structural class to bias the prediction of its secondary structure. Protean 3D begins its calculations by predicting the class of the protein from its amino acid composition. Class designations are identical to those of Nakashima et al., 1986, except for α/ß class proteins. Protean 3D divides α/ß class proteins into two subclasses, α/ß-α and α/ß-ß, according to their distance from the α and ß classes. The letter R is used by Deléage and Roux to designate the "irregular" class of proteins. Next, secondary structure is predicted from the frequency of occurrence of each residue in the four states (helix, sheet, turn and coil) using the secondary structure database of Kabsch and Sander.

To apply this track to the sequence:

In the Tracks panel, expand *Secondary Structure* and check the box next to *Deleage-Roux*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- If you know the structural class for your protein (**A**, **B**, **A+B**, **A/B – A**, **A/B – B**, **R**), choose it by clicking on a button.
- Use the default, **Computer Calculated**, if you do not know to which class your protein belongs.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

Secondary Structure – Garnier-Robson

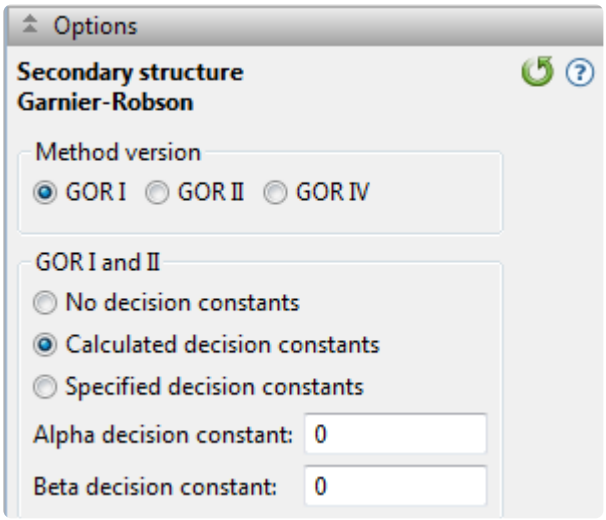
The Secondary Structure – Garnier-Robson track , from Garnier et al. (1978), Garnier and Robson (1990), and Garnier, Gibrat and Robson (1996), predicts protein structure from the amino acid sequence. The method examines the propensity of a given residue to exist in a certain structure. It is a statistical approach which uses examination of known crystallographic structures to define regions as α-helical (H), ß-pleated sheet (E - extended chains), ß-turns (T) and coil (C).

To apply this track to the sequence:

In the Tracks panel, expand *Secondary Structure* and check the box next to *Garnier-Robson*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



Specify which version of the GOR method you wish to use: **GOR I**, **GOR II** or **GOR IV**. GOR II uses the same algorithm as GOR I, but is based on updated propensity tables. GOR IV does not predict turns, but only predicts the probability of areas being helices, sheets or coils. GOR IV is the most modern secondary structure prediction method in Protean 3D.

If you choose GOR I or GOR II, you may select from among three options related to the decision constant:

- **No Decision Constants** - if you do not want Protean 3D to make assumptions about the global α-helix and ß-sheet content.
- **Calculated Decision Constants** - to use computer-derived constants from global α-helix and ß-sheet

Copyright DNASTAR, Inc. 2022

probabilities. Constants are based on three protein classes: proteins with less than 20%, 20-50%, and greater than 50% α-helix or ß-sheet.

- **Specified Decision Constants** - if you have prior knowledge of circular dichroism data. If you select this button, you need to input values for **Alpha Decision Constant** and **Beta Decision Constant**.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

Secondary Structure – Coiled Coil

The Secondary Structure – Coiled Coil track predicts structures according to the formula of Lupas et al., 1991. You may use this track with the Transmembrane – Goldman-Engleman-Steitz track to predict transmembrane alpha helices.

To apply this track to the sequence:

In the Tracks panel, expand *Secondary Structure* and check the box next to *Coiled Coil*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:

- **Threshold** – threshold needed to determine a positive score in the region graph. 1.3 is the default value because it is the minimum score for a known coiled-coil. If you need less stringent coiled-coil predictions (e.g., for globular proteins), you may want to lower this to 1.1. If you want to predict very stringent coiled-coils, (e.g., for fibrous proteins), raise it to 1.5.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

Stability – Aliphatic Index

The Stability – Aliphatic Index track (Gasteiger et al. 2005) predicts regional stability by calculating the relative volume occupied by aliphatic side chains. This is a positive indicator of globular protein thermostability. This method is only available for amino acid selections. Detailed information about this method is available on the ExPASY website.

To apply this track to the sequence:

In the Tracks panel, expand *Stability* and check the box next to *Aliphatic Index*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:

- **Residues to average** – number of residues to be averaged when constructing each point in the plot. The default is 9.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

Stability – Instability Index

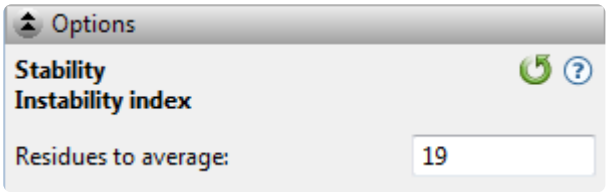
The Stability – Instability Index track estimates the stability of the protein in a test tube. The Instability Index is calculated using the approach of Guruprasad et al., 1990., which predicts regional instability by calculating the weighted sum of dipeptides that occur more frequently in unstable proteins when compared to stable proteins. Detailed information about this method is available on the ExPASY website. An instability index less than 40 predicts a stable protein, whereas values higher than 40 denote a potentially unstable protein.

To apply this track to the sequence:

In the Tracks panel, expand *Stability* and check the box next to *Instability Index*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- **Residues to Average** – number of residues to be averaged when constructing each point in the plot. The default is 19. The closer a calculation is to this value, the more likely the chance of aggregation. Over small windows, this method gives an idea of the potential for local instability.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

Stability – Isoelectric Precipitate

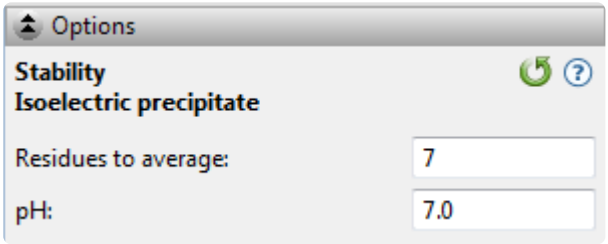
The Stability – Isoelectric Precipitate track predicts regional instability by identifying uncharged regions at a given pH, which have an increased potential to aggregate if surface exposed. This method uses the pKa tables from Lehninger et al., 2005.

To apply this track to the sequence:

In the Tracks panel, expand *Stability* and check the box next to *Stability – Isoelectric Precipitate*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- **Residues to average** – number of residues to be averaged when constructing each point in the plot. The default is 7.
- **pH** – Protean 3D highlights regions with a pI ± 0.5 pH units from the pH specified here. The default is 7.0.

Click if you wish to return to the default settings.

Copyright DNASTAR, Inc. 2022

Surface Probability (Emini)

The Surface Probability (Emini) track predicts the probability that a given region lies on the surface of a protein, using the approach of Emini et al., 1985.

To apply this track to the sequence:

In the Tracks panel, check the box next to *Surface Probability (Emini)*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:

- **Surface decision threshold (0..6)** – the value (between 0 and 6) that determines whether or not a given hexapeptide is considered to be on the surface or not. The default used by Emini is 1. Higher values will predict fewer surface residues and lower values will predict more.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

Transmembrane – Goldman-Engleman-Steitz

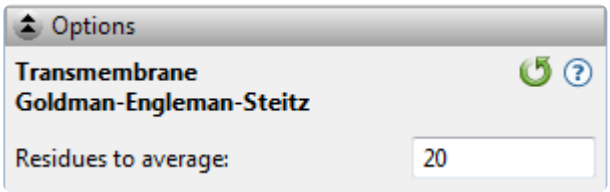
The Transmembrane – Goldman-Engleman-Steitz track predicts non-polar alpha helices which may span a cell membrane, using the approach of Engleman et al., 1986. To search for non-polar helices, scales of residue polarity are applied across a protein sequence. This algorithm is much more accurate at predicting membrane bound helices than other hydropathy methods, especially if the helix contains polar residues.

To apply this track to the sequence:

In the Tracks panel, expand *Transmembrane* and check the box next to *Goldman-Engleman-Steitz*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- **Residues to Average** – number of residues to be averaged when constructing each point in the plot. The default is 20, which reflects the size of a common transmembrane alpha helix. 21 residues are required to span a lipid bilayer 30Å long, as the interval between residues in an alpha helix is 1.5Å.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

Transmembrane – TopPred

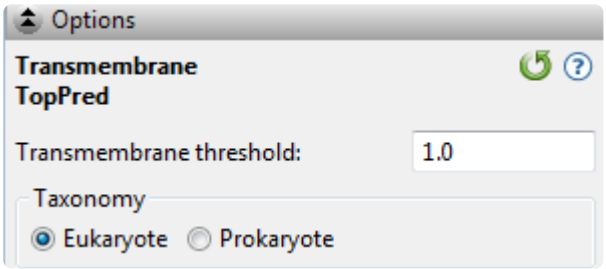
The Transmembrane – TopPred track utilizes hydropathy analysis to propose probable transmembrane regions in a protein sequence, using the methods of von Heijne, 1992, and Claros and von Heijne, 1994. Possible membrane-protein topologies are considered and ranked according to the “positive-inside rule,” an observation that the cytoplasmic parts of membrane proteins more frequently contain positively charged residues compared to non-cytoplasmic parts. Additional considerations are made for eukaryote proteins, including the net charge difference of residues flanking the N-terminal transmembrane segment.

To apply this track to the sequence:

In the Tracks panel, expand *Transmembrane* and check the box next to *TopPred*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- **Transmembrane threshold** – Regions with a maximum hydrophobicity score equal to or greater than this threshold will be reported as transmembrane regions, as long as they are not adjacent to, nor overlap with, other transmembrane regions. A setting of 0.6 will cause Protean 3D to report all putative regions, while a threshold of 1.0 will cause the application to be more selective and only report regions with a higher probability of being transmembrane regions. The default is 1.0.

- **Taxonomy** – Select the taxonomy of the organism from which the sequence is derived. The selection influences Protean 3D’s prediction of whether a region is cytoplasmic or extracellular.

Click if you wish to return to the default settings.

Copyright DNASTAR, Inc. 2022

Transmembrane – von Heijne

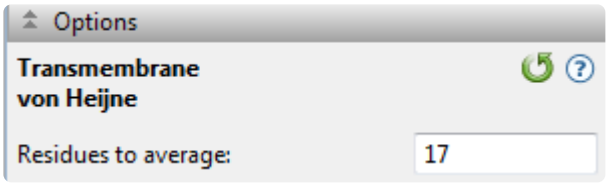
The Transmembrane – von Heijne track predicts hydropathy based on the interaction between the signal sequence and membrane, using the approach of von Heijne et al., 1981. The Transmembrane – von Heijne method is similar to Goldman-Engleman-Steitz, another Protean 3D method that uses the Hopp-Woods algorithm. The von Heijne method uses estimates of the standard free energy transfer of residues that transition from a random coil conformation in water to a helical conformation in a non-polar environment.

To apply this track to the sequence:

In the Tracks panel, expand *Transmembrane* and check the box next to *von Heijne*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- **Residues to Average** – number of residues to be averaged when constructing each point in the plot. The default is 17.

Click if you wish to return to the default settings.

Copyright DNASTAR, Inc. 2022

Immunogenicity tracks

The Tracks panel lets you choose which immunogenicity tracks to apply to the Analysis view. The following links describe the methods and explain how to change their Track Options settings, if you desire to do so:

- **Antigencity:**
  - Jameson-Wolf
  - Welling
- B-cell epitopes
- MHC II epitopes
- **T-cell epitopes:**
  - AMPHI
  - Rothbard-Taylor

Copyright DNASTAR, Inc. 2022

Antigenicity – Jameson-Wolf

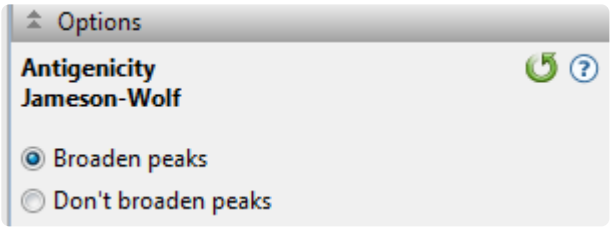
The Jameson-Wolf track predicts potential antigenic determinants by combining existing methods for protein structural predictions, using the approach of Jameson and Wolf , 1988. Method results appear as multiple peaks in the Antigenic Index plot, with each peak signifying a potential antigenic determinant. The Antigenic Region graph summarizes the antigenic regions using a threshold value of 1.0. The Jameson-Wolf output is affected by parameter changes made to the Hopp-Woods, Chou-Fasman and Garnier-Robson methods, but not to changes made to the Surface Probability or Karplus-Schulz methods.

To apply this track to the sequence:

In the Tracks panel, expand *Immunogenicity*, then *Antigenicity*, and check the box next to *Jameson-Wolf*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- **Broaden Peaks** - causes the following smoothing procedure to be applied: For values A[i] that exceed 1.0, where i is a sequence position, find the position p of the peak value within the following 9 residues. If neither structure prediction method predicts helix at position p, apply A[p+n] = A[p+n] + 0.2 * (5 - abs(n)) * A[p] for n = -4 to 4.

- **Don't Broaden Peaks** - causes the method to produce the default index of antigenicity by combining values for hydrophilicity (Hopp-Woods, H[<2 > 0.5 <1 > 0 <-1 > -0.5 <-2 >]), surface probability (Emini, S[<1 > 1.0 <0 >]), flexibility (Karplus-Schulz, F[<1 > 1.0 <0 >]), and the secondary structure predictions of Chou-Fasman (CF[T=2, t=1, 0]) and Garnier-Robson (GR[T=2, t=1, 0]) using the following formula: A = 0.3H + 0.15S + 0.15F +0.2CF + 0.2GR.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

Antigenicity – Welling

The Welling track predicts antigenicity using the method of Welling et al., 1985. Previous methods assumed that antigenic regions were mainly hydrophilic regions on the surface of a protein. By contrast, the Welling method is based on the percentage of each amino acid present in known antigenic determinants compared to the percentage of the amino acids in the average composition of a protein.

To apply this track to the sequence:

In the Tracks panel, expand *Immunogenicity*, then *Antigenicity*, and check the box next to *Welling*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:

- **Residues to average** – number of residues to be averaged when constructing each point in the plot. The default is 9.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

B-Cell Epitopes (DNASTAR)

The B-Cell Epitopes (DNASTAR) prediction track uses a bioinformatics approach that uses experimental antibody recognition data. For a set of peptides provided by NeoClone® Biotechnology International, LLC (Madison, WI USA), the antigenicity of each peptide was determined by measuring the relative binding affinity of the peptide against hybridoma-secreted antibodies raised in mice inoculated with the peptide.

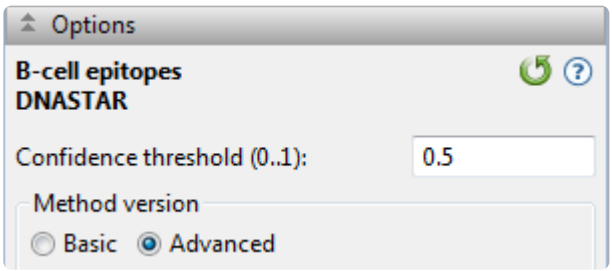
The following video is a brief introduction to the B-Cell Epitopes (DNASTAR) track.

To apply this track to the sequence:

First open a file or project. Then, in the Tracks panel, expand *Immunogenicity* and check the box next to *B-Cell Epitopes (DNASTAR)*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- **Confidence score threshold** – A confidence score with a value from 0 to 1 is assigned to each residue in a protein sequence. By default, residues with a confidence score greater than 0.5 are classified as putative antigenic residues. Changing this threshold value will update the Analysis View accordingly. The default is 0.5.

- **Method version** – Choose whether to use the **Basic** or **Advanced** version of the method. **Basic** uses the legacy "NeoClone" algorithm from Protean 3D version 10.1, while **Advanced** uses an updated algorithm.

Click if you wish to return to the default settings.

Antigenic classifications:

A cross-validated machine learning approach was used to train two models that predict antigenic and highly antigenic regions. The antigenic models consider sequence-based helix, sheet, beta turn, and flexibility predictions as well as hydropathy, antigenicity, and stability indices when predicting antigenic regions; the highly antigenic models consider a modified combination of secondary structure predictions, hydropathy, antigenicity, and stability.

- Highly antigenic – At least 5% of hybridoma cultures create antibodies that tightly bind the peptide. The highly antigenic model may predict narrow regions of a protein sequence to be antigenic. Use these predictions to highlight regions from the antigenic model with an even greater likelihood of generating an immunological response.

- Antigenic – At least 2.5% of hybridoma cultures create antibodies that tightly bind the peptide. The antigenic model may predict wide-spanning regions of a protein sequence to be antigenic. Increase the threshold value (e.g. 0.75) to focus your analysis on regions with a higher antigenic confidence level.

- Non-antigenic – Less than 2.5% of hybridoma cultures create antibodies that tightly bind the peptide.

Copyright DNASTAR, Inc. 2022

MHC II Epitopes (Sette)

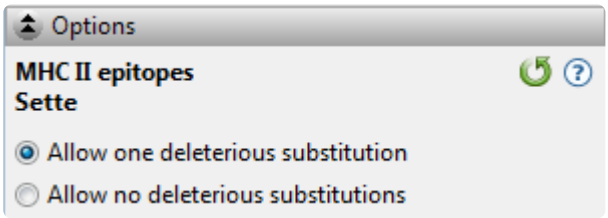
The MHC II Epitopes (Sette) track predicts peptide antigenic sites which interact with mouse MHC II haplotype d proteins. There are separate motifs for IAd and IEd haplotypes, and Sette et al., 1989, reports that this method finds more than 75% of known epitope sites for each. However, the method is only effective for x and y motifs for the d haplotype of mouse MHC II proteins. It can be used to design peptide epitopes when used in conjunction with the T-Cell Epitopes – AMPHI and the T-Cell Epitopes – Rothbard-Taylor tracks.

To apply this track to the sequence:

In the Tracks panel, expand *Immunogenicity* and check the box next to *MHC II Epitopes (Sette)*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



Choose how to set the threshold for a positive IAd region:

- **Allow one deleterious substitution** – to set it to 400. The is the default choice.
- **Allow no deleterious substitutions** – to set it to 729.

There are no options available for the IEd method.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

T-Cell Epitopes – AMPHI

The T-Cell Epitopes – AMPHI track uses the model of Margalit and Berzofsky, 1987 to predict immunodominant helper T-lymphocyte antigenic sites from primary sequence data. The underlying assumption of the AMPHI method is that T-cell antigenic sites are composed of amphipathic helices. The method is useful only for T-cell epitopes and not B-cell epitopes (the latter normally requires greater knowledge of the tertiary structure of the binding site).

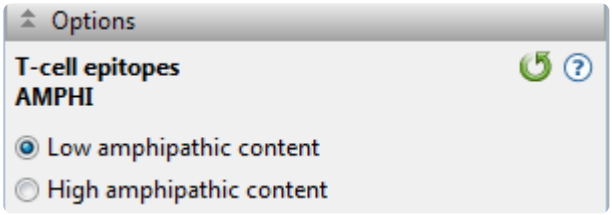
Track results can be displayed as a hydrophobicity graph, an AMPHI region graph, an alpha-helix graph, a 3-10 helix graph (scaled for the amphipathic index) and/or an AMPHI intensity graph. Use the intensity graph to determine the relative intensity of predicted amphipathic regions. For best results, use this method in combination with the MHC II Epitopes and T-Cell Epitopes – Rothbard-Taylor tracks.

To apply this track to the sequence:

In the Tracks panel, expand *Immunogenicity*, then *T-cell epitopes*, and then check the box next to *AMPHI*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- **Low amphipathic content** – uses a minimum amphipathic index of 4 to create the region graph from three (or more) consecutive regions of length 7. This is the default choice.
- **High amphiphatic content** – uses an amphipathic index of 8 to create the region plot for three (or more) consecutive regions of length 11. Choose this option only for those proteins known to have high amphipathic content.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

T-Cell Epitopes – Rothbard-Taylor

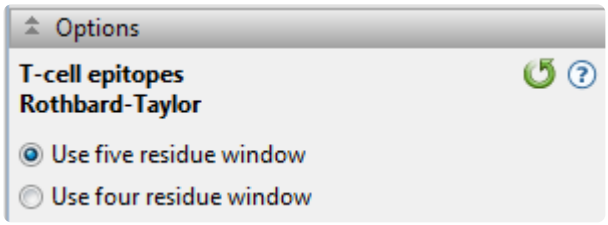
The T-Cell Epitopes – Rothbard-Taylor track locates potential T-lymphocyte antigenic determinants which contain a common sequence motif. Although Rothbard and Taylor, 1988, found that the method predicted 80% of the antigenic determinants for their peptide database, the presence or absence of the motif is not sufficient to guarantee antigenicity. Furthermore, the method is useful only for T-cell epitopes and not B-cell epitopes, which often require greater knowledge of the tertiary structure of the binding site. For best results, use this method in combination with the T-Cell Epitopes – AMPHI and the MHC II Epitopes tracks.

To apply this track to the sequence:

In the Tracks panel, expand *Immunogenicity*, then *T-cell epitopes*, and then check the box next to *Rothbard-Taylor*. The track will now be visible in the Analysis view

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- Choose between **Use five residue window** and **Use four residue window**. A window of five will locate both 4-mers and 5-mers, which fit motif criteria. The default is **Use five residue window**.

Click if you wish to return to the default value.

Copyright DNASTAR, Inc. 2022

Database tracks

The Tracks panel lets you choose which database tracks to apply to the Analysis view. The following links describe the methods and explain how to change their Track Options settings, if you desire to do so:

- Cleavage sites
- Domains & sites

Copyright DNASTAR, Inc. 2022

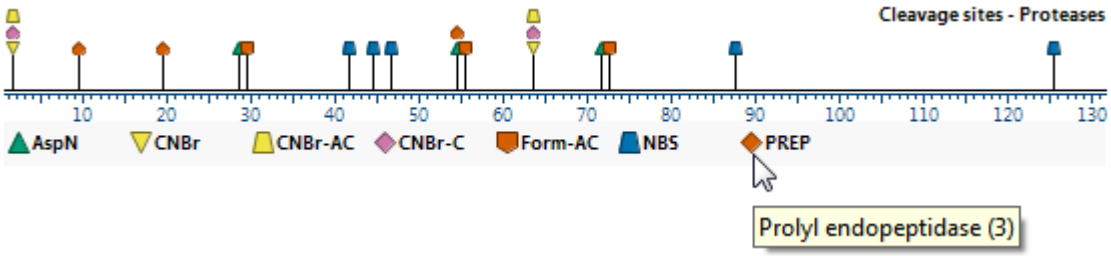
Cleavage Sites (Proteases)

The Cleavage sites (Proteases) track predicts cleavage sites for proteases (also termed peptidases or proteinases) and chemicals in a protein sequence using known proteolytic signatures. When you apply Cleavage sites (Proteases) to the Analysis view, Protean 3D searches the protein sequence against the protease library. and displays matches to protease and chemical cleavage sites as markers with different shapes and colors. For a table showing which proteases and chemicals are included in this method, see List of proteases.

The following video demonstrates the use of the Cleavage sites (Proteases) track:

To apply this track to the sequence:

In the Tracks panel, expand *Databases* and check the box next to *Cleavage Sites (Proteases)*. The track will now be visible in the Analysis view, with a legend matching the marker symbol to a specific protease.

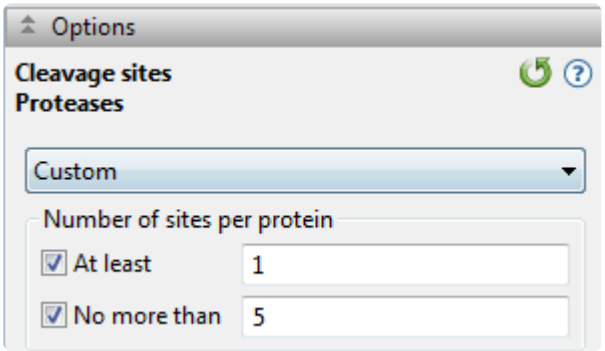


In addition, a list of cleavage sites will appear in the Tracks panel under "Cleavage Sites (Proteases).

Copyright DNASTAR, Inc. 2022

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- The drop-down menu lets you select from the following filters based on the protease type or the number of cleavage sites:

Filter type	Filter description
All	All enzymes and chemicals
Chemicals	Only chemicals
Cuts infrequently	Cleaves a sequence at least once but no more than five times
Cuts once	Cleaves a sequence once
Enzymes	Only enzymes
Fusion tag cleavage	Enterokinase, Factor Xa, HRV3C protease, TEV protease, Thrombin
Mass spectrometry	Arg-C, Asp-N, Chymotrypsin, Glu-C, Lys-C, Pepsin, Pancreatic elastase, Thermolysin, Trypsin
Nonspecific	Arg-C, Asp-N, Asp-N (N-terminal Glu), BNPS or NCS/urea, Chymotrypsin, Chymotrypsin (low specificity), Clostripain, CNBr, CNBr (methyl-Cys), CNBr (with acids), Formic acid, Glu-C (AmAc buffer), Glu-C (Phos buffer), Hydroxylamine, Iodosobenzoic acid, Lys-C, Lys-N, Lys-N (Cys modified), Mild acid hydrolysis, NBS (long exposure), NBS (short exposure), NTCB, Pancreatic elastase, Pepsin A, Pepsin A (low specificity), Prolyl endopeptidase, Proteinase K, Thermolysin, Trypsin, Trypsin (Arg blocked), Trypsin (Cys modified), Trypsin (Lys blocked)
Specific	Caspase-1, Caspase-2, Caspase-3, Caspase-4, Caspase-5, Caspase-6, Caspase-7, Caspase-8, Caspase-9, Caspase-10, Enterokinase, Factor Xa, Granzyme B, HRV3C protease, TEV protease, Thrombin
Custom	Allows you to manually define the number of cleavage sites per sequence by using

Copyright DNASTAR, Inc. 2022

	the text boxes below the drop-down menu.

> **Note:** Choosing any filter except for **Cuts infrequently** or **Cuts once** will cause the **Number of Sites per Protein** section to be cleared.

- In the **Number of Sites per Protein** area, check one or both boxes, if desired, then enter a number in the corresponding text box.

Click if you wish to return to the default value.

Working with proteases in the Analysis view:

Task	How To…
To select an individual cleavage site	Click on its marker above the sequence ruler.
To select all visible cleavage sites for a protease or chemical	Click on its name in the legend.
To view the protease or chemical name and cleavage site	Hover over the marker above the sequence ruler.
To view the full name of the protease or chemical and the number of times it cuts the sequence	Hover over its name in the legend.
To view detailed information about a protease	Select its name from the list. The Details panel will then display in-depth information about the protease, including its classification.
To create a tab-delimited text file with information about the cleavage frequency of all compounds	Click the "Export Proteases" link in the **Actions** section of the Details panel.

Copyright DNASTAR, Inc. 2022

List of proteases

The following is a list of proteases available in Protean 3D via the Cleavage Sites (Proteases) analytical method:

Name	Synonyms	Recognition Sequences
Arg-C	Arginyl peptidase, Endoproteinase Arg-C, Tissue kallikrein	R-x
Asp-N	Endoproteinase Asp-N, Peptidyl-Asp metalloendopeptidase	x-D
Asp-N (N-terminal Glu)	Endoproteinase Asp-N, Peptidyl-Asp metalloendopeptidase	x-[DE]
BNPS or NCS/urea	3-Bromo-3-methyl-2-(2-nitrophenylthio)-3H-indole, BNPS-skatol, N-chlorosuccinimide/urea	W-x
Caspase-1	ICE, Interleukin-1β-Converting Enzyme	[FLWY]-x-[AHT]-D-{DEKPQR}
Caspase-10	Flice2, Mch4	I-E-A-D-x
Caspase-2	Ich-1, Nedd2	D-V-A-D-{DEKPQR} or D-E-H-D-{DEKPQR}
Caspase-3	Apopain, CPP32, Yama	D-M-Q-D-{DEKPQR} or D-E-V-D-{DEKPQR}
Caspase-4	ICEII, Ich-2, TX	L-E-V-D-{DEKPQR} or [LW]-E-H-D-{DEKPQR}
Caspase-5	ICEIII, TY	[LW]-E-H-D-x
Caspase-6	Mch2	V-E-[HI]-D-{DEKPQR}
Caspase-7	CMH-1, ICE-LAP3, Mch-3	D-E-V-D-{DEKPQR}
Caspase-8	FLICE, MASH, Mch5	[IL]-E-T-D-{DEKPQR}
Caspase-9	ICE-Lap6, Mch6	L-E-H-D-x
Chymotrypsin		[FY]-{P} or W-{MP}
Chymotrypsin (low specificity)		[FLY]-{P} or W-{MP} or M-{PY} or H-{DMPW}
Clostripain	Clostridiopeptidase B	R-x
CNBr	Cyanogen bromide	M-x
CNBr (methyl-	Cyanogen bromide	M-x or x-C

Copyright DNASTAR, Inc. 2022

Cys)		
CNBr (with acids)	Cyanogen bromide	[MW]-x
Enterokinase	Enteropeptidase	[DE](4)-K-x
Factor Xa	Coagulation factor Xa	[AFGILTVM]-[DE]-G-R-x
Formic acid		D-x
Glu-C (AmAc buffer)	Endoproteinase Glu-C, V8 protease, Glutamyl endopeptidase	E-x
Glu-C (Phos buffer)	Endoproteinase Glu-C, V8 protease, Glutamyl endopeptidase	[DE]-x
Granzyme B	Cytotoxic T-lymphocyte proteinase 2, Granzyme-2, GranzymeB, Lymphocyte protease, SECT, T-cell serine protease 1-3E	I-E-P-D-x
HRV3C protease	Human rhinovirus 3C protease, Picornain 3C, Protease 3C	L-E-V-L-F-Q-G-P
Hydroxylamine	Hydroxylammonium	N-G
Iodosobenzoic acid	2-Iodosobenzoic acid	W-x
Lys-C	Endoproteinase Lys-C, Lysyl endopeptidase	K-x
Lys-N	Endoproteinase Lys-N, Peptidyl-Lys metalloendopeptidase, Armillaria mellea neutral proteinase	x-K
Lys-N (Cys modified)	Endoproteinase Lys-N, Peptidyl-Lys metalloendopeptidase, Armillaria mellea neutral proteinase	x-[CK]
Mild acid hydrolysis		D-P
NBS (long exposure)	N-Bromosuccinimide	[HWY]-x
NBS (short exposure)	N-Bromosuccinimide	[WY]-x
NTCB	2-Nitro-5-thiocyanatobenzoic acid, 2-Nitro-5-thiocyanobenzoic acid	x-C
Pancreatic elastase	Pancreatopeptidase E, Elastase-1	[AGSV]-x
Pepsin A	Pepsin	{HKR}-{P}-{R}-[FLWY]-{P} or

Copyright DNASTAR, Inc. 2022

		{HKR}-{P}-[FLWY]-x-{P}@
Pepsin A (low specificity)	Pepsin	{HKR}-{P}&#45@;{R}@-[FL]-{P} or {HKR}-{P}-[FL]-x-{P}
Prolyl endopeptidase	Prolyl oligopeptidase, Post-proline cleaving enzyme	[HKR]-P-{P}
Proteinase K	Endopeptidase K, Peptidase K	[AEFILTVWY]-x
TEV protease	Tobacco etch virus protease, Nuclear-inclusion-a endopeptidase	E-x-x-Y-x-Q-[GS]
Thermolysin	Thermophilic-bacterial protease	{DE}-[AFILMV]-{P}
Thrombin	Factor IIa	x-x-G-R-G-x or [AFGILTVW]-[AFGILTVW]-P-R- {DE}-{DE}
Trypsin	Trypsin-1	x-[KR]-{P} or W-K-P or M-R-P *But not:* [CD]-K-D or C-K-[HY] or C-R-K or R-R-[HR]
Trypsin (Arg blocked)		K-{P}
Trypsin (Cys modified)		[RKC]-{P}
Trypsin (Lys blocked)		R-{P}

Note: Recognition sequences are formatted using PROSITE pattern syntax.

Copyright DNASTAR, Inc. 2022

Domains & Sites (PROSITE)

The Databases – Domains & sites (PROSITE) track searches chains against common patterns found in the PROSITE database (installed with your software). When you apply the method, Protean 3D will search the protein sequence against the pattern library and display found patterns and their labels. Protean 3D will attempt to locate the database automatically. If it is not found, you will be prompted to locate it.

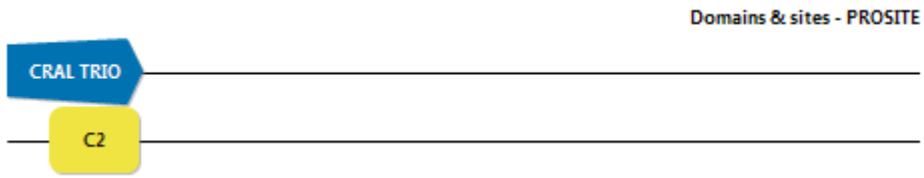
By default, the PROSITE database (*prosite.dat*) is installed to the *Prosite* folder, found in the following directory:

- **Windows 7 & Windows 8:** C:\Users\Public\Documents\DNASTAR\Lasergene 10 Data\Prosite
- **Macintosh:** Hard Drive:Applications:DNASTAR:Lasergene 10 Data:Prosite

The following video is a brief walkthrough that shows how to use the Domains & Sites (PROSITE) track:

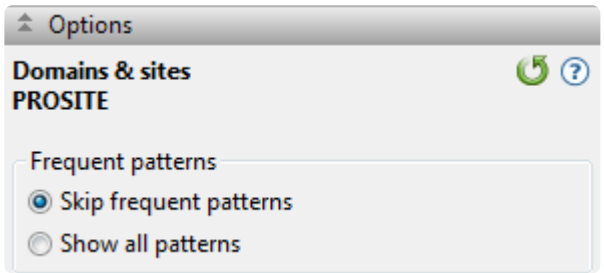
To apply this track to the sequence:

In the Tracks panel, expand *Databases* and check the box next to *Domains & Sites (PROSITE)*. The track will now be visible in the Analysis view. In addition, a list of findings will appear in the Tracks panel under "Domains & Sites (PROSITE)."



To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



Choose whether you want to **Skip Frequent Patterns** or **Show All Patterns**. If you choose **Skip Frequent Patterns** (the default), Protean 3D will not display common sites, such as asparagine glycosylation and tyrosine kinase phosphorylation. Thus, fewer sites will be displayed.

Click if you wish to return to the default settings.

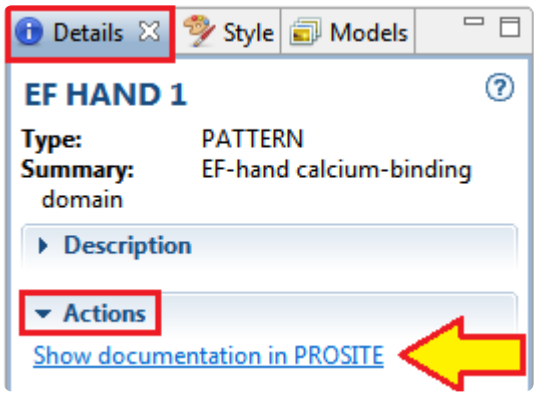
Working with PROSITE results:

In the Analysis view, the PROSITE synopsis shows all pattern hits found in a given cluster of chains. The patterns found in the same location in all chains are shown in solid color and in front of the sequence line. The patterns found only in some of the chains are shown as semitransparent and behind the sequence line.

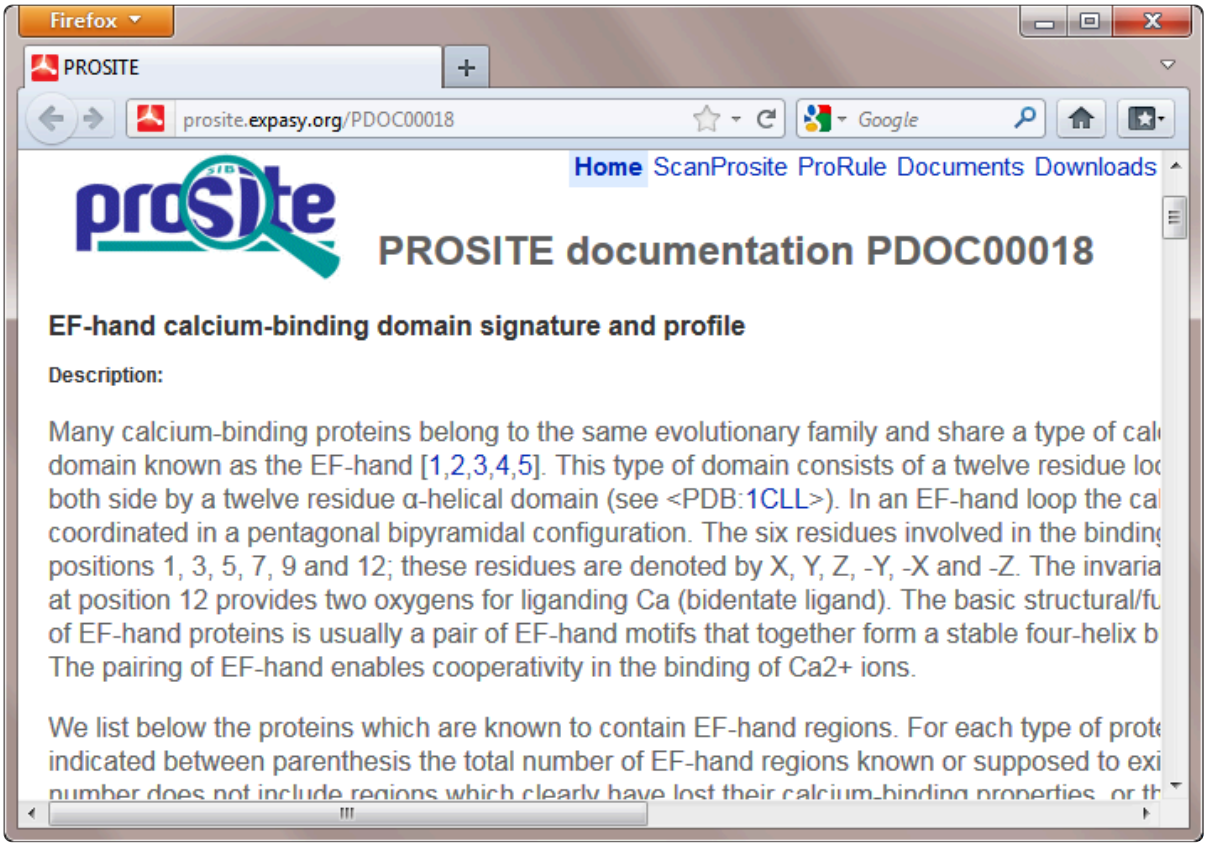
Expanding the synopsis reveals separate rows containing patterns found in all chains and patterns found only in some of the chains. You can further expand the latter row to reveal two additional rows: "by pattern," and "by chain."

Task	How To
To view the residue range that falls within the pattern	Hover over any PROSITE pattern match shape in the Analysis View. By selecting a shape or marker, this information and additional information will be displayed in the Details panel.
To view the PROSITE entry online	Open the Details panel and click on the hyperlink in the “Actions” section.

Copyright DNASTAR, Inc. 2022



A browser window will open at the relevant PROSITE entry.



To search the PROSITE database online	Click the “Scan PROSITE” link at the top of the page.

Note: DNASTAR is authorized by Geneva Bioinformatics (GeneBio) S.A. to incorporate PROSITE® data into Lasergene. PROSITE® is a registered trademark of the Swiss Institute of Bioinformatics (SIB) and GeneBio has been granted exclusive rights by the SIB for the

Copyright DNASTAR, Inc. 2022

licensing of PROSITE. Note that access to PROSITE via Lasergene requires a user license which can be obtained by contacting GeneBio at sib-db@genebio.com or info@genebio.com.

Update the PROSITE database between releases

You may update the PROSITE database at any time following the steps below:

1. Go to this ExPASy web page, then select the **Downloads** tab from the top of the page.
2. From the ftp directory, right-click on the file *prosite.dat* and select **Save Link As**.
3. In the Save dialog, navigate to the appropriate folder for your operating system:
   - Windows 7 & Windows 8: C:\Users\Public\Documents\DNASTAR\Lasergene ‘x’ Data\Prosite
   - Macintosh: Applications:DNASTAR:Lasergene ‘x’ Data:Prosite

   …where ‘x’ represents the version number.
4. Click **Save**. When prompted, confirm that you wish to overwrite the existing *prosite.dat* file.

The updated PROSITE database is now ready to use in Protean and Protean 3D. Before using the PROSITE database, please read the note near the top of the *prosite.dat* file regarding licensing for commercial entities.

Copyright DNASTAR, Inc. 2022

Modifications (SUMOylation) track

The Modifications (SUMOylation) track predicts sites where lysine undergoes the addition of small, ubiquitin-like modifiers after translation using the approach of Ren et al., 2009. The covalent attachment of this small, ubiquitin-like protein to other proteins can modify and regulate cell function.

Modifications (SUMOylation) provides individual region plots for each of four sumoylation site patterns: **KxD**, **KxE**, **IKxEP** and **KxE**. The weights for the amino acids in each window are calculated and used for a system of linear equations which result in a correlation coefficient between the training data and the known false pattern positives. This correlation value is used as a threshold for predicting a sumoylation pattern.

To apply this track to the sequence:

In the Tracks panel, check the box next to *Modifications (SUMOylation)*.

The track will now be visible in the Analysis view. The SUMOylation method displays results as "markers" on the synopsis. Expand the groups below the synopsis, if present, to see channels showing different groups and propensities.

To edit method options:

Select the method in the Tracks panel. Open the Track Options section, which appears as follows:

- **Threshold (0.0..1.0)** – The default value of 0.3 maximizes the correlation coefficient between the training data set and the known false pattern positives. Decreasing this value increases both true positives and false positives; increasing this value decreases false positives at the expense of decreasing true positives as well.

Click if you wish to return to the default settings.

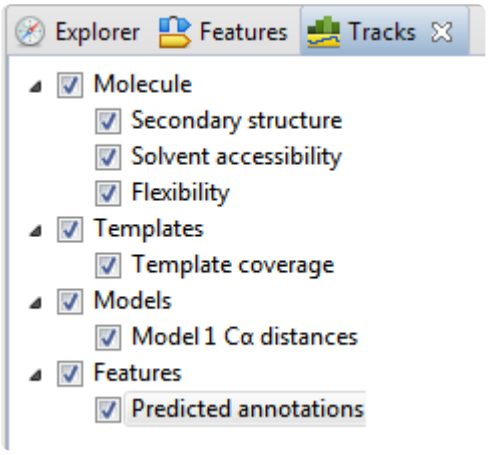
Copyright DNASTAR, Inc. 2022

NovaFold-related tracks

After running a NovaFold prediction and opening the result file (*.novafold*), the Tracks panel displays tracks specific to this type of prediction.

To apply this track to the sequence:

Tracks related to NovaFold predictions are listed below. In the Tracks panel, check the box next to the track name to make it visible in the Analysis view. Not all method classes are available for a given prediction.



- **Molecule > Secondary structure** – NovaFold predicts the location of secondary structure elements using a collection of artificial neural networks trained to recognize patterns in a molecule's evolutionary profile. Models tend to be more accurate in regions with helices and sheets.

- **Molecule > Solvent accessibility** – NovaFold predicts regions that are accessible to the surrounding solvent using a collection of artificial neural networks trained to recognize patterns in a molecule's evolutionary profile. Models tend to be more accurate in regions with lower solvent accessibility.

- **Molecule > Flexibility** – NovaFold predicts flexible regions by identifying ranges with an increased flexibility profile, which is derived from the combination of sequence-based predictions and crystallographic temperature factors from the selected templates.

- **Templates > Template coverage** – The template coverage is proportional to the fraction of threading templates mapped to a given residue in the molecule. Models tend to be more accurate in regions with higher coverage values. Regions that appear as gaps in the NovaFold Report minimap, and after exporting to MegAlign Pro, may appear to simply have "relatively low coverage" in this method display. This is because the first two displays consider only the top 10 template results, while this method considers all template results.

- **Models > Model 'n' Cα distances** – NovaFold predicts the pairwise distances between Cα atoms in a predicted model and its theoretical ideal structure using a support vector machine trained to recognize

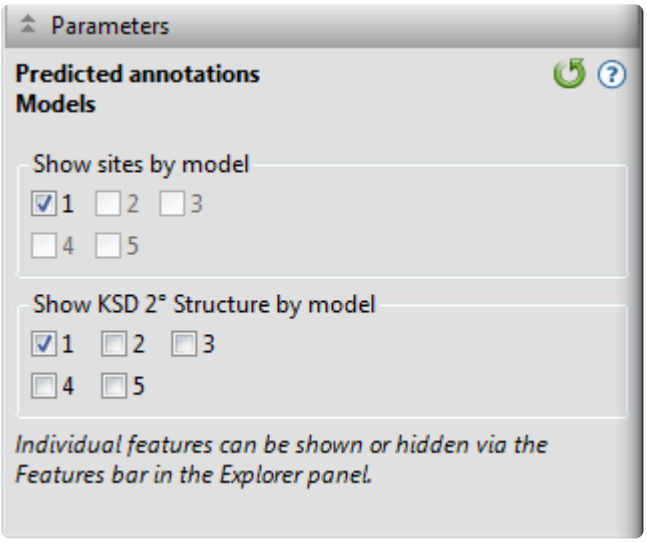
Copyright DNASTAR, Inc. 2022

patterns in template coverage, secondary structure, solvent accessibility, and simulated structural diversity. Models tend to be more accurate in regions with lower distance values.

- **Features > Predicted annotations** – Features include matching regions from threading templates and predicted annotations, such as secondary structure and functional sites.

To edit track options for “Predicted annotations” :

Select the **Predicted annotations** track in the Tracks panel. Open the Track Options section:

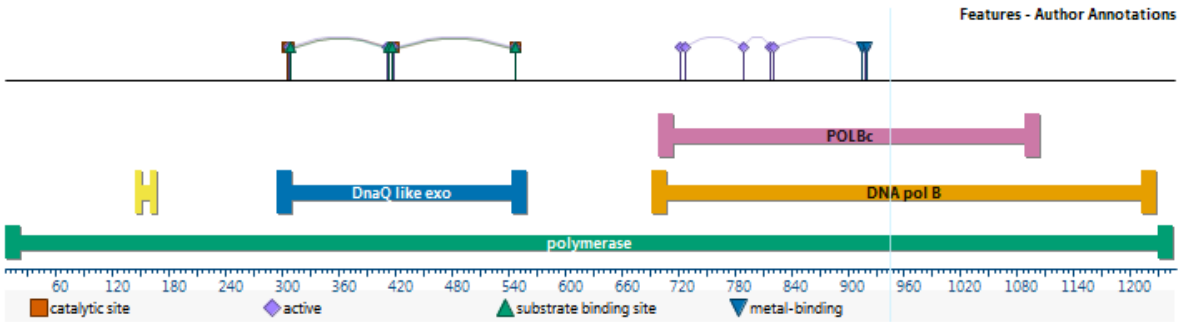


- **Show sites by model** – Check the boxes for each model whose sites you wish to display in the Analysis view.

- **Show KSD 2 Structure by model** – (Only present in certain circumstances) Check the boxes for each model whose KSD secondary structures you wish to display.

Features (Annotated) tracks

In the Analysis view, you can display selected features from the protein sequence file on the same axis with the prediction tracks. PDB files contain sites, disulfide bonds, helices, and sheets/strands, while Lasergene and Genbank files support many more feature types, including user-created feature types.



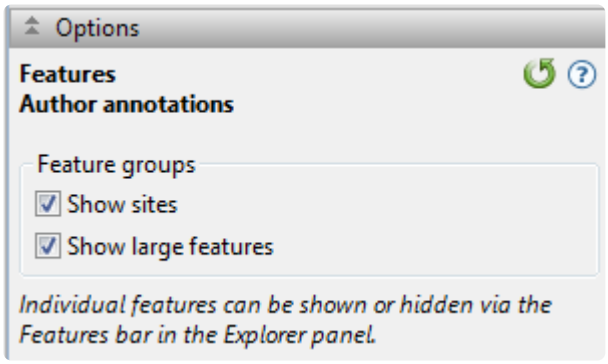
To apply this track to the sequence:

In the Tracks panel, check the box next to *Features (Annotated)*. The track will now be visible in the Analysis view. Initially, the display shows a consensus of the selected features. Feature types are differentiated via colorful icons with legends and/or by text labels.

> **Note:** When viewing prediction model/template alignments from NovaFold, each region feature represents a protein fragment from a template.

To edit track options:

Select the track in the Tracks panel. Open the Track Options section, which appears as follows:



- Uncheck **Show Sites** if you wish to hide the upper row showing disulfide bonds and sites.

Copyright DNASTAR, Inc. 2022

- Uncheck **Show Large Features** to hide large, usually unimportant, features spanning most or all of the sequence.

Click if you wish to return to the default values.

Copyright DNASTAR, Inc. 2022

Details panel

The Details panel, represented by a blue “information” icon ( ), lets you view details about the structure, sequence or selection. Information in this section is updated continuously to reflect the currently selected data.

Opening the Details panel:

Do any of the following:

- Click on the **Details** tab
- Use **View > Details**
- Press **Ctrl+Alt+I** (Windows) or **Option+Cmd+I** (Macintosh)

Using the Details panel:

The contents of the panel depends on the context of the selection in the document:

- (no document open)
- A single residue or ligand
- A molecule, feature or region
- An entire structure
- A selected region or strand
- A group of structures
- A track
- A protease pattern or profile
- An individual protease match
- A PROSITE pattern or profile
- An individual PROSITE match
- A NovaFold prediction
- A NovaDock prediction

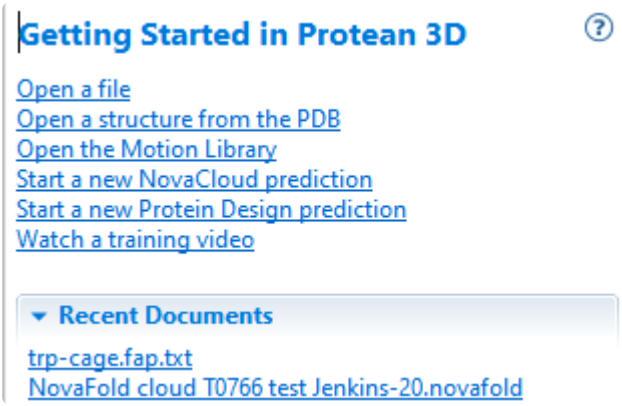
Copyright DNASTAR, Inc. 2022

Copying text from the Details panel:

Text and graphics in the Details panel can be copied to your computer's clipboard using basic editing commands.

When no document or a blank document is open

When no document or a blank document is open, the Details panel displays information for getting started in Protean 3D.



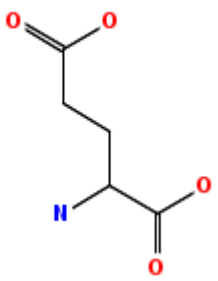
Click on the links to:

- Open a file
- Open a structure from the PDB
- Open the Motion Library
- Start a new NovaCloud prediction
- Start a new Protein Design prediction
- Watch a training video.

Copyright DNASTAR, Inc. 2022

A single residue or ligand

If the active selection is a single residue or ligand, the Details panel contains the following information:

Title **GLUTAMIC ACID (GLU)**	The component name and 3-letter code.
Summary **Structure:** 7TIM **Chain:** A **Position:** 152 **Type:** L-PEPTIDE LINKING	• **Structure** – The name of the parent structure. • **Chain** – The name of the parent chain. • **Position** – The sequence position. • **Type** – The component type.
2D Structure 2D Structure 	The chemical structure for the component.
Biophysical Properties Biophysical Properties **MW [g/mol]:** 105.093 **Net charge [pH=7]:** -0.01 **pI:** 5.24 **Average hydropathy:** -0.80 **Aliphatic index:** 0.00 **A_{280}:** 0.00 **ε_{280} [M^{-1} cm^{-1}]:** 0.00	Displays the physical properties of the current selection, and may include: • **MW [g/mol]** – The molecular weight of the selected component. • **Formal charge** – The sum of formal charges in a molecule for a given electronic state. This is shown only for non-amino acid selections. • **Net charge** – The pH-dependent sum of charges in a population of molecules. Protean 3D calculates the net charge for protein regions and amino acid residues, and assumes a pH of 7. • **pI** – The isoelectric point (the pH at which the residue carries no net electrical charge), calculated from pKa tables from Lehninger et al. (2005). This is available only for amino acid selections.

Copyright DNASTAR, Inc. 2022

	• **Average hydropathy** – The Kyte-Doolittle hydropathy value, calculated using the method of Kyte, J. and Doolittle, R.F. (1982). This is shown only for amino acid selections. • **Aliphatic index** – The relative volume occupied by aliphatic side chains (alanine, valine, isoleucine and leucine). See note below for reference. This is shown only for amino acid selections. • **A_{280}** – The absorbance (optical density) for the residue, which is affected by presence of disulfide bonds. See note below for reference. This is shown only for amino acid selections. • **ε_{280}; [$M^{-1}cm^{-1}$]** – The extinction coefficient (amount of light a residue absorbs). Protean 3D assumes all Cys residues are reduced. See note below for reference. This is shown only for amino acid selections. • **Instability index** – An estimate of the stability of the protein in a test tube. A protein whose instability index is smaller than 40 is predicted as stable, while a value above 40 predicts that the protein may be unstable. This value is calculated using the method of Guruprasad et al.. **Note**: For aliphatic index, absorbance and extinction coefficient, see Gasteiger et al.. In all three calculations, the sequence is treated as independent residues. There is no algorithmic dependence on length.
Structural Properties ▾ Structural Properties *Dihedral angles* **φ:** -69.7° **ψ:** -30.1° **ω:** 178.3° **Median B-factor:** 22.97 *Surface area in current selection* **New 1 [Å²]:** 40.47	Displays angular and other structural properties of the current selection. A dihedral angle is the angle between two geometric planes defined by four atoms and is restricted by chemical and steric restraints. **Note**: If the active document is a multi-model document, values are reported for the first model only. • **Dihedral angles φ** (phi), **ψ** (psi), and **ω** (omega) are displayed only for proteins. **φ** and **ψ** define the structure of a protein's backbone since **ω** is a rigid angle describing the peptide bond. • **Dihedral angles α** (alpha) through **ζ** (zeta) and **χ** (chi) are displayed only for nucleotides. **α** through **ζ** are defined by the position of a residue's backbone atoms while **χ** describes the orientation of the base.

Copyright DNASTAR, Inc. 2022

	• Median B-factor – The median temperature factor in a residue or ligand. For NMR files, this is typically 0.0. • Occupancy – A list of alternate location codes and occupancies. • **Surface area in current selection** is displayed only when a surface has been applied to all or part of a selection. The display shows the area, in square Angstroms, for the part of the surface that is contained in the selection.
Actions Actions Export selection details to file Copy 2D structure to clipboard Show GLU in Ligand Expo	Click the relevant link to export data to a file, copy data to the clipboard, or open a browser window to the relevant page of the PDB or Ligand Expo databases.

A molecule, feature or region

If the active selection consists of one or more molecules, features or regions, the Details panel contains the following information:

Title **Selected Region**	The name or type of molecule, feature or region. If a region is selected, the text will read "Selected Region."
Summary **Residues:** 162 **Water:** 4 *Selection* **Structure:** 7TIM **Chain:** A **Position:** 10, 12, 17..29, 44..53, 80..85, 95..101, 106..118, 131..135, 139..153, 165, 167..171, 178..196, 198..203, 211, 217..220, 230, 232..236, 238..245, 620..623 **Chain:** B **Position:** 5..10, 36..41, 59..63, 90..93, 122..127, 160..165, 206..209, 228..231	• **Residues**, **Ligands**, **Water** – The number of residues, ligands, and water molecules in the selection. • **Structure** – The name of the parent structure. • **Chain** and **Position** – The names of the parent chains and residue ranges (for structures). • **Sequence** – Sequence name and positions of selected residues (for sequences).
Biophysical Properties Biophysical Properties **MW [g/mol]:** 1248.40 **Net charge [pH=7]:** 2.00 **pI:** 10.00 **Average hydropathy:** -0.79 **Aliphatic index:** 56.67 **A_{280}:** 0.00 **ε_{280} [M^{-1} cm^{-1}]:** 0.00 **Instability index:** -19.62	Displays the physical properties of a continuous selection, including: • **MW [g/mol]** – The molecular weight of the selected region. • **Net charge** – The pH-dependent sum of charges in a population of molecules. Protean 3D calculates the net charge for protein regions and amino acid residues, and assumes a pH of 7. • **pI** – The isoelectric point (the pH at which the residue carries no net electrical charge), calculated from pKa tables from Lehninger et al. (2005). This calculation is available only for amino acid selections. • **Average hydropathy** – The Kyte-Doolittle hydropathy value, calculated using the method of Kyte, J. and Doolittle, R.F. (1982). This calculation is available only for amino acid selections. • **Aliphatic index** – The relative volume occupied by aliphatic side chains (alanine, valine, isoleucine and leucine). See note below for reference. This calculation is available only for amino

	acid selections. • **A_{280}** – The absorbance (optical density) for the residue, which is affected by presence of disulfide bonds. See note below for reference. This calculation is available only for amino acid selections. • **ε_{280}; [$M^{-1}cm^{-1}$]** – The extinction coefficient (amount of light a residue absorbs). Protean 3D assumes all Cys residues are reduced. See note below for reference. This calculation is available only for amino acid selections. • **Instability index** – An estimate of the stability of the protein in a test tube. A protein whose instability index is smaller than 40 is predicted as stable, while a value above 40 predicts that the protein may be unstable. This value is calculated using the method of Guruprasad et al. (1990)|topic=Research References. **Note**: For aliphatic index, absorbance and extinction coefficient, see Gasteiger et al. (2005)|topic=Research References. In all three calculations, the sequence is treated as independent residues. There is no algorithmic dependence on length.
Structural Properties ▾ Structural Properties **Distance [Å]:** 5.77 *Between C (A:CYS 11) and H (A:TYR 14)* *Surface area in current selection* **INSULIN (A) [Å²]:** 225.73	For structures, displays the structural properties of a continuous selection, including: • **Distance** - displays the minimum distance in angstroms (Å) between a pair of residues or ligands. The interface displays the two atoms selected for the calculation. Distance is calculated for Model 1 only. • **Surface area in current selection** – is displayed only when a surface|topic=The Surfaces Area has been applied to all or part of a selection. The display shows the area, in square Angstroms, for the part of the surface that is contained in the selection.
Features in Selection ▾ Features in Selection **Site:** AC1 **Helix:** 13 helices **Sheet:** 9 strands	When available, displays a list of features located within the range of the current selection. Available information may include a list of **Site** names or **Disulfide Bond** names, and the numbers of **Helices** or **Sheets** in the selection. Click on any hyperlink to open details for that feature.

Copyright DNASTAR, Inc. 2022

Actions ▾ Actions Export selection details to file	Click the link to export the data to a file.

Copyright DNASTAR, Inc. 2022

An entire structure

If a document contains just one structure and nothing is selected, or if the structure is directly selected in the Molecules section, the Details panel contains the following information:

Title **1AON**	The structure name or file name.
Summary **Title:** CRYSTAL STRUCTURE OF THE ASYMMETRIC CHAPERONIN COMPLEX GROEL/GROES/(ADP)7 **Classification:** COMPLEX (GROEL/GROES) **Technique:** X-RAY DIFFRACTION **Resolution [Å]:** 3.00 **PubMed ID:** 9285585	If available, the summary may display the structure **Title**, **Classification**, experimental **Technique** (e.g. x-ray, NMR, EM), experimental **Resolution** (for x-ray only) and/or a link to **PubMed**.
Molecules ▾ **Molecules** *21 biopolymers with 8337 residues* GROEL **Chains:** A, B, C, D, E, F, G, H, I, J, K, L, M, N **Type:** Protein **Length:** 547 **UniProt ID:** P0A6F5 GROEL/GROES COMPLEX **Chains:** O, P, Q, R, S, T, U **Type:** Protein **Length:** 97 **UniProt ID:** P0A6F9	Displays a count of the molecules and residues in a structure document, as well as basic information about the molecule. For each chain, the following information may be provided: • The **Molecule** name, with a link allowing you to select all chains for this molecule. • A list of **Chains** for the molecule, with links allowing you to select any individual chain. • The **Type** and **Length** of biopolymer. • A link to the **UniProt** database.
Actions ▾ **Actions** Show 1AON in PDB	Click the relevant link to export data to open a browser window to the relevant page of the **PDB** website.

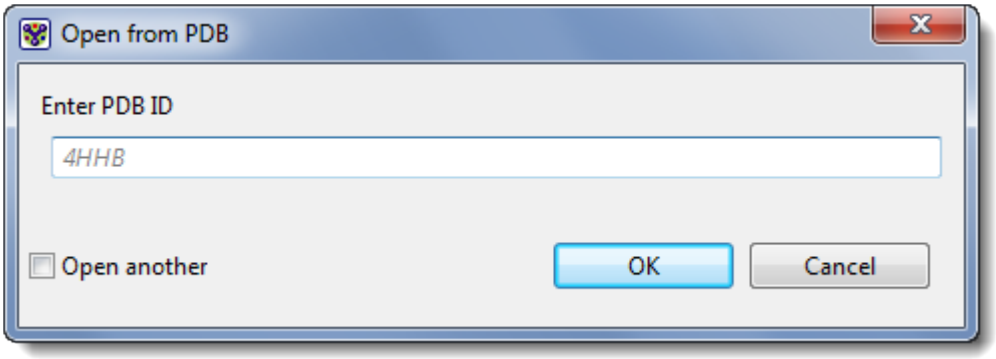
Copyright DNASTAR, Inc. 2022

Try it! – View details about a structure

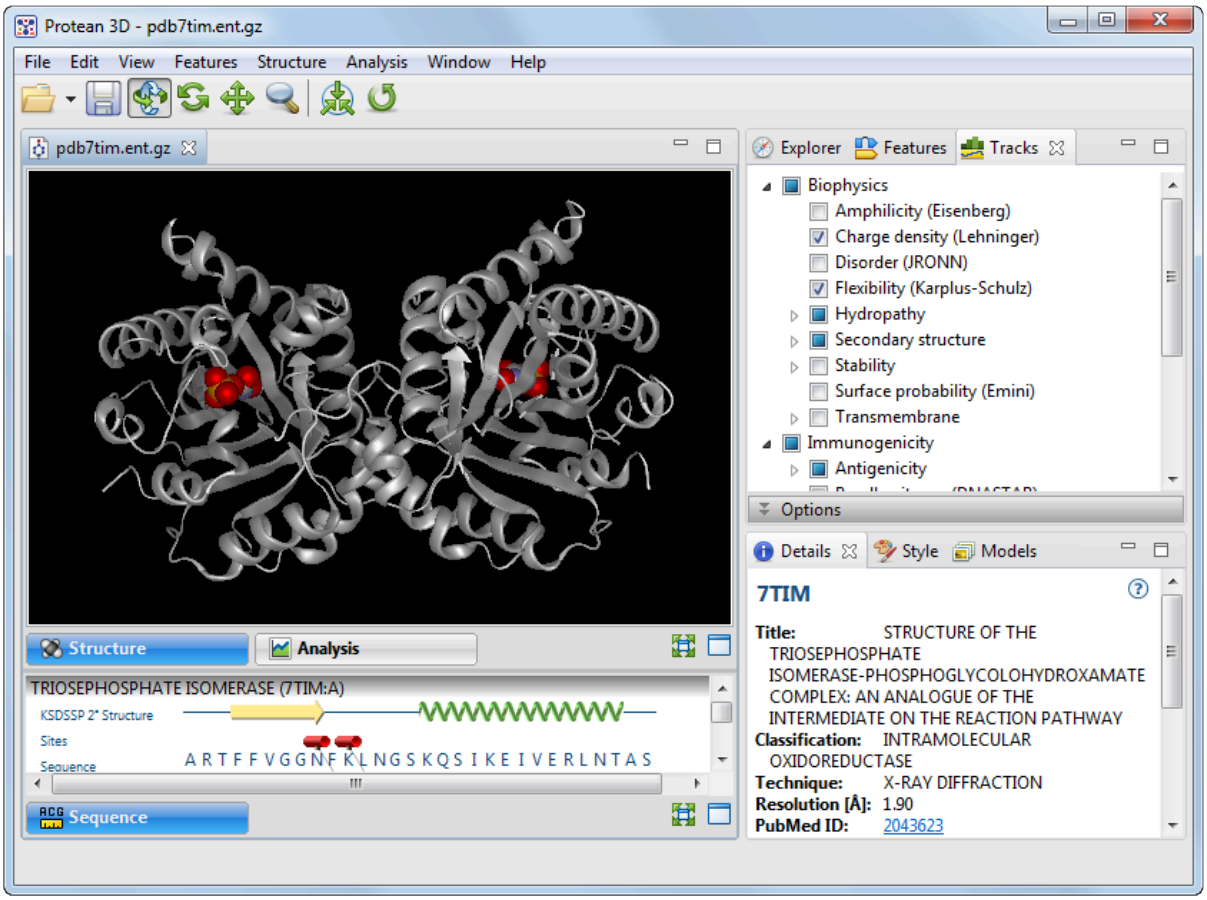
This tutorial demonstrates how to use the Details panel to view information about a structure or sequence.

1. Within Protean 3D, choose **File > Open from PDB**. You may initially see a yellow warning icon ( ) prompting you to type in a valid ID.

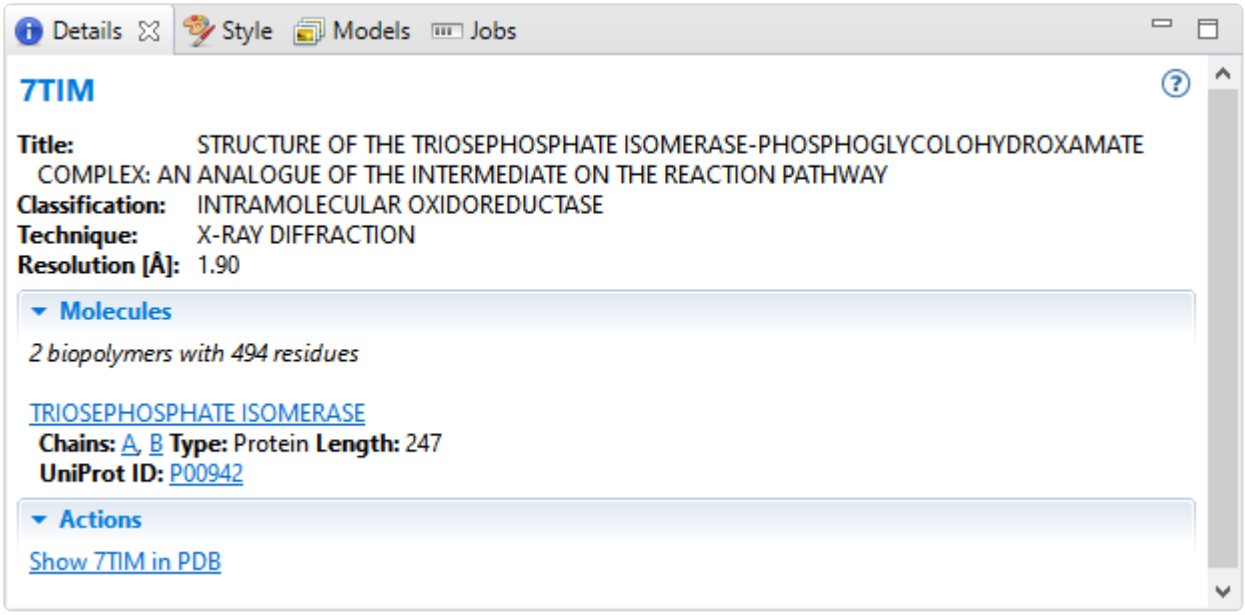
2. Type **7TIM** into the text box.



3. Press **OK** to open the structure in Protean 3D.



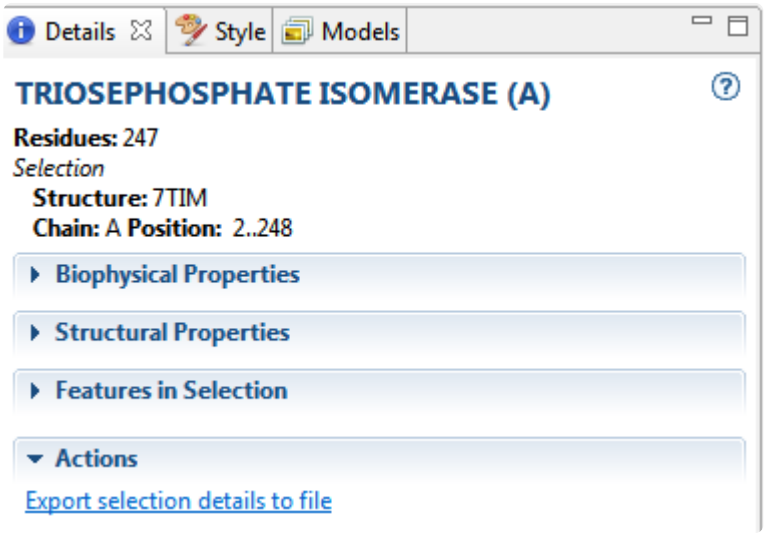
4. If the Details panel is not visible, click on its tab or use **View > Details**.



The Details panel displays a different subset of information depending on your selection. Because no specific selection has been made here, the panel displays information about the structure as a whole.

5. Click on the **Uniprot ID** and/or **Show 7TIM in PDB** hyperlinks to open the relevant web pages. After viewing the web page(s), return to Protean 3D.

6. In the Molecules section of the Explorer panel, click on A TRIOSEPHOSPHATE ISOMERASE to select that molecule in all views. Note that the Details panel now contains different expand bars from before, including sections for **Biophysical Properties**, **Structural Properties**, and **Features in Selection**. The hyperlinked task in the Actions bar has also changed, and now allows you to export the selection details to a file.



7. Click on the triangles to the left of each bar to reveal more details.

Copyright DNASTAR, Inc. 2022

A selected region or strand

If a region or single strand is selected in any of the views, the Details panel contains the following information:

Titles **1AON**	The strand identifier or the words “Selected Region.”
Summary **Residues:** 324 *Selection* **Structure:** structure **Chain:** B **Position:** 1..81 **Chain:** D **Position:** 1..81 **Structure:** structure_Variant1 **Chain:** B **Position:** 1..81 **Chain:** D **Position:** 1..81	• **Residues**, **Ligands**, **Water** – The number of residues, ligands, and water molecules in the selection. • **Structure** – The name of the parent structure. • **Chain** – The name of the parent chain. • **Position** – The sequence position.
Biophysical Properties ▾ **Molecules** *21 biopolymers with 8337 residues* GROEL **Chains:** A, B, C, D, E, F, G, H, I, J, K, L, M, N **Type:** Protein **Length:** 547 **UniProt ID:** P0A6F5 GROEL/GROES COMPLEX **Chains:** O, P, Q, R, S, T, U **Type:** Protein **Length:** 97 **UniProt ID:** P0A6F9	Displays the physical properties of the current selection, and may include: • **MW [g/mol]** – The molecular weight of the selected component. • **Formal charge** – The sum of formal charges in a molecule for a given electronic state. This is shown only for non-amino acid selections. • **Net charge** – The pH-dependent sum of charges in a population of molecules. Protean 3D calculates the net charge for protein regions and amino acid residues, and assumes a pH of 7. • **pI** – The isoelectric point (the pH at which the residue carries no net electrical charge), calculated from pKa tables from Lehninger et al. (2005). This is available only for amino acid selections. • **Average hydropathy** – The Kyte-Doolittle hydropathy value, calculated using the method of Kyte, J. and Doolittle, R.F. (1982). This is shown only for amino acid selections.

Copyright DNASTAR, Inc. 2022

	• **Aliphatic index** – The relative volume occupied by aliphatic side chains (alanine, valine, isoleucine and leucine). See note below for reference. This is shown only for amino acid selections. • **A_{280}** – The absorbance (optical density) for the residue, which is affected by presence of disulfide bonds. See note below for reference. This is shown only for amino acid selections. • **ε_{280}; [$M^{-1}cm^{-1}$]** – The extinction coefficient (amount of light a residue absorbs). Protean 3D assumes all Cys residues are reduced. See note below for reference. This is shown only for amino acid selections. • **Instability index** – An estimate of the stability of the protein in a test tube. A protein whose instability index is smaller than 40 is predicted as stable, while a value above 40 predicts that the protein may be unstable. This value is calculated using the method of Guruprasad et al.. **Note**: For aliphatic index, absorbance and extinction coefficient, see Gasteiger et al.. In all three calculations, the sequence is treated as independent residues. There is no algorithmic dependence on length.
Structural Properties ▾ Actions Show 1AON in PDB	• **Distance** - displays the minimum distance in angstroms (Å) between a pair of residues or ligands. The interface displays the two atoms selected for the calculation. Distance is calculated for Model 1 only.
Features in Selection **Selected Region**	• **Sheet** – The number of strands in the selections. • **Hypervariable** – Click on the provided link to view information about the hypervariable region (HVR).
Actions **Residues:** 7 *Selection* **Structure:** Anti-IgE Fab_Model1 **Chain:** L **Position:** 41..47	Click the link to export the data to a file.

A group of structures

If a document contains multiple structures and nothing is selected, or if multiple structures are selected in the Molecules section, the Details panel contains the following information:

Titles 1AON	The structure names or file names appear in separate sections for each structure.
Summary* **Title: CRYSTAL STRUCTURE OF THE ASYMMETRIC CHAPERONIN COMPLEX GROEL/GROES/(ADP)7 **Classification:** COMPLEX (GROEL/GROES) **Technique:** X-RAY DIFFRACTION **Resolution [Å]:** 3.00 **PubMed ID:** 9285585 **PDB ID:** 1AON *21 biopolymers with 8337 residues* GROEL **Chains:** A, B, C, D, E, F, G, H, I, J, K, L, M, N **Type:** Protein **Length:** 547 **UniProt ID:** P0A6F5 GROEL/GROES COMPLEX **Chains:** O, P, Q, R, S, T, U **Type:** Protein **Length:** 97 **UniProt ID:** P0A6F9	A summary appears below each title. If available, the summary may display the structure **Title**, **Classification**, experimental **Technique** (e.g. x-ray, NMR, EM), experimental **Resolution** (for x-ray only) and/or a link to **PubMed**. Below that is a count of the molecules and residues in a structure document. For each chain, the following information may be provided: • The **Molecule** name, with a link allowing you to select all chains for this molecule. • A list of **Chains** for the molecule, with links allowing you to select any individual chain. • The **Type** and **Length** of biopolymer. • A link to the **UniProt** database.
Actions	(No actions are available for multiple structures.)

Copyright DNASTAR, Inc. 2022

A track

If the active selection is a track, the Details panel contains the following information:

Title **Chou-Fasman**	The track method or pattern name.
Summary **Type:** Secondary structure **Summary:** The Chou-Fasman method predicts the secondary structure of Alpha, Beta, and Turn conformations.	The track method or pattern classification and a brief description.
Current Settings Current Settings **Alpha region threshold:** 103 **Beta region threshold:** 105 **Conformational parameter table:** 64 Proteins	The parameters and current values.

A protease pattern or profile

If the active selection is a protease class listed in the Tracks panel, the Details panel contains the following information:

Title **Asp-N**	The name of the protease.
Summary **Type:** Protease **Summary:** A metallo-endoproteinase with a zinc ion coordinated by three His and a catalytic Glu	• **Type:** Pattern classification ("protease" or "chemical") • **Summary:** The MEROPS protease family
Description Description **Synonyms:** Endoproteinase Asp-N, Peptidyl-Asp metalloendopeptidase **Pattern:** x†D **EC Number:** 3.4.24.33 **MEROPS ID:** M72.001 **Matches:** 3 matches on 1 sequence	• **Synonyms:** List of synonyms, if any • **Pattern:** The pattern defining a site. • **EC Number:** A link to the BRENDA Comprehensive Enzyme Information System entry for the protease (for enzymes only) • **MEROPS ID:** A link to the MEROPS database entry (enzymes only) • **PubChem ID:** A link to the PubChem entry (for chemicals only) • **Matches:** Number of matches found. • **Comments:** Any comments associated with the record.
Actions Actions Export full protease report to file	Click the link to launch a Save dialog that allows you to save a tab-delimited text file with information about the cleavage frequency of all compounds.

Copyright DNASTAR, Inc. 2022

An individual protease match

If the active selection is an individual protease in the Analysis view, the Details panel contains the following information:

Title **Site: Asp-N**	The name of the protease.
Summary **Type:** Protease **Summary:** A metallo-endoproteinase with a zinc ion coordinated by three His and a catalytic Glu **Group:** ysth2b1.pro **Sequence:** ysth2b1.pro **Position:** 54	• **Type:** Pattern classification ("protease" or "chemical") • **Summary:** The MEROPS protease family • **Structure, Chain** and **Position** (for structures) • **Group**, **Sequence** and **Position(s)** (for sequences).
Description Description **Synonyms:** Endoproteinase Asp-N, Peptidyl-Asp metalloendopeptidase **Pattern:** x+D **EC Number:** 3.4.24.33 **MEROPS ID:** M72.001 **Matching sequence:** ysth2b1.pro	• **Synonyms:** List of synonyms, if any • **Pattern:** The pattern defining a site. • **EC Number:** A link to the BRENDA Comprehensive Enzyme Information System entry for the protease (for enzymes only) • **MEROPS ID:** A link to the MEROPS database entry (enzymes only) • **PubChem ID:** A link to the PubChem entry (for chemicals only) • **Comments:** Any comments associated with the record. • **Matching sequence:** The matching sequence for this cleavage site.
Actions	(No actions are available for this selection.)

Copyright DNASTAR, Inc. 2022

A PROSITE pattern or profile

If the active selection is a PROSITE pattern or profile in the Tracks panel, the Details panel contains the following information:

Title **NAC AB**	The name of the PROSITE domain or site.
Summary **Type:** MATRIX **Summary:** NAC A/B domain profile	• **Type** – Pattern classification • **Summary** – Short description
Description Description **PROSITE ID:** PS51151 **Matrix type:** Protein domain **Version:** 20.79 **Taxon range:** Archaea; Eukaryotes **Max repeats:** 1 **Matches:** 2 on 2 sequences **Best score:** 295 **Worst score:** 295 **Threshold:** 294	• **PROSITE ID** A link to the PROSITE database. • **Pattern** The PROSITE representation of the pattern defining a site. For information about conventions, see this PROSITE page Depending on availability, some or all of the following information may be displayed: • **Profile type -** A matrix can represent a protein domain, repeat region, localization signal, or composition. • **Version -** Database version. • **Taxon range -** Kingdoms where pattern is expected: Archaea, Bacteriophages, Eukaryotes, Prokaryotes, or Viruses. • **Max repeats -** Maximum number of repeats observed in at least one protein sequence. • **Matches -** Number of matches. • **Best score -** Profile score of best match in class with link to the match (profile type only). • **Worst score -** Profile score of worst match in class with link to the match (profile type only, shown for two or more matches). • **Threshold -** the threshold similarity score used to identify

Copyright DNASTAR, Inc. 2022

	sequences that match a profile (profiles only).
Actions 	Click the link to open a browser window to the relevant documentation page of the PROSITE database.

Copyright DNASTAR, Inc. 2022

An individual PROSITE match

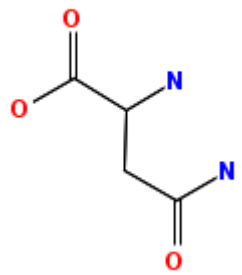
If the active selection is an individual PROSITE pattern match in the Analysis view, the Details panel contains the following information:

Title **Match: EF HAND 2**	The name of the PROSITE domain or site.
Summary **Summary:** EF-hand calcium-binding domain profile **Type:** MATRIX **Sequence:** ysth2b1.pro **Position:** 88..108	Pattern classification and short description, as well as structure or chain name and range (for structures), and chain name and positions of selected residues (for sequences).
Description Description **PROSITE ID:** PS50222 **Profile type:** Protein domain **Profile score:** -16.00 **Threshold:** -16 **Matching sequence:** nkkstisareiqtavrlilpg	• **PROSITE ID -** A link to the PROSITE database. • **Pattern -** The PROSITE representation of the pattern defining a site. For information about conventions, see this PROSITE page. • **Profile type -** A matrix can represent a protein domain, repeat region, localization signal, or composition. • **Profile score -** A similarity score for an alignment between a sequence and a profile (profiles only). • **Threshold -** the threshold similarity score used to identify sequences that match a profile (profiles only). • **Sites -** List of predicted functional sites and their corresponding residues. • **Matching sequence -** The sequence that matches the pattern.
Actions Actions Show documentation in PROSITE	Click the link to open a browser window to the relevant documentation page of the PROSITE database.

Copyright DNASTAR, Inc. 2022

A Protein Design result

If a Protein Design result is open, the Details panel contains the following information:

Title Pro4Asn	The name of the Protein Design job.
Summary **Structure:** structure_Variant1 **Chain:** A **Position:** 4 **Type:** L-PEPTIDE LINKING	• **Structure** – The name of the parent structure. • **Chain** and **Position** – The names of the parent chains and residue ranges (for structures). • **Type** – The component type.
2D Structure 2D Structure 	The chemical structure of the new amino acid residue. If more than one residue has been changed, this section shows the number of changed residues, the specific changes and their positions.
Biophysical Properties Biophysical Properties **MW [g/mol]:** 132.118 **Net charge [pH=7]:** -0.02 **pI:** 5.52 **Average hydropathy:** -3.50 **Aliphatic index:** 0.00 **A_{280} (ox.):** 0.00 **A_{280} (red.):** 0.00 **ε_{280} [M^{-1} cm^{-1}]:** 0.00	Displays the physical properties of a continuous selection, including: • **MW [g/mol]** – The molecular weight of the selected region. • **Net charge** – The pH-dependent sum of charges in a population of molecules. Protean 3D calculates the net charge for the selected amino acid residues, and assumes a pH of 7. • **pI** – The isoelectric point (the pH at which the residue carries no net electrical charge), calculated from pKa tables from Lehninger et al. (2005). This calculation is available only for amino acid selections. • **Average hydropathy** – The Kyte-Doolittle hydropathy value, calculated using the method of Kyte, J. and Doolittle, R.F. (1982). This calculation is available only for amino acid selections.

Copyright DNASTAR, Inc. 2022

	• **Aliphatic index** – The relative volume occupied by aliphatic side chains (alanine, valine, isoleucine and leucine). See note below for reference. This calculation is available only for amino acid selections. • **A_{280}** – The absorbance (optical density) for the residue, which is affected by presence of disulfide bonds. See note below for reference. This calculation is available only for amino acid selections. • **ε_{280}; [$M^{-1}cm^{-1}$]** – The extinction coefficient (amount of light a residue absorbs). Protean 3D assumes all Cys residues are reduced. See note below for reference. This calculation is available only for amino acid selections. **Note**: For aliphatic index, absorbance and extinction coefficient, see Gasteiger et al. (2005)|topic=Research References. In all three calculations, the sequence is treated as independent residues. There is no algorithmic dependence on length.
Structural Properties ▾ Structural Properties *Dihedral angles* **φ:** -100.2° **ψ:** -1.8° **ω:** 174.5° **Median B-factor:** 25.31	For structures, displays the structural properties of a continuous selection, including: • **φ** and **ψ** – The "phi" and "psi" dihedral angles in the backbones of the protein. • **Median B-factor** – Reflects the fluctuation of atoms about their average positions, which is related to protein dynamics.
Feature Details ▾ Feature Details **experiment:** predicted by NovaDesign (16.0.0.294) **delta_energy:** -0.06089417128464447 **potential:** DFIRE_A	• **experiment** – Reports the tool and version, "predicted by Protein Design ($VERSION). • **delta_energy** – The change in energy calculated between a variant and wild-type. • **potential** – The energy function used to evaluate sequence perturbations, currently DFIRE-A (reference).
Actions ▾ Actions Export selection details to file Copy 2D structure to clipboard Show ASN in Ligand Expo	Click the links to: • **Export selection details to file** – To save details related to the current selection. • **Copy 2D structure to clipboard** – To copy the image in the 2D Structure section to the clipboard.

	• **Show [residue name] in Ligand Expo** – To view the entry for the substitute residue on the Protein Data Bank (PDB) Ligand Expo website.

The NovaFold report

For most selections made in the NovaFold report, the Details panel provides the information shown in the table below.

Title **Diphtheria toxin fragment A**	The molecule name.
Summary **Residues:** 193 **Started:** 9/17/2014 3:26 PM	• **Residues** - Number of residues in the sequence. • **Started** - Date and time the prediction was submitted to the NovaFold server.
Options Options **Function:** Predicted **Models:** 5 **Templates:** 20 **Simulation:** Extended **Homologs:** Included	Displays the settings from the NovaFold Option screen. • **Function** - Displays whether binding sites were **Predicted** or **Not Predicted**, depending on whether or not **Predict ligand binding sites and protein function** was checked in the NovaFold Options screen. • **Models** – The number of prediction models created. • **Templates** – The number of model templates used. • **Simulation** - The simulation effort, which describes maximum simulation runtimes [**Extended**, **Quick**]. • **Homologs** - Whether templates are filtered based on sequence identity [**Included**, **Excluded**].
Status Status **Prediction ID:** d4852633a7f48020bb55597b52394d2d **Version:** 4.0.0.1 **Library:** 2014-W33-4 **Computation:** 19h 40m	• The first three items in this section (**Prediction ID, Version, Library**) are for diagnostic use only. • **Computation** – The time in hours and minutes needed to create the prediction.
Actions	Click **Send template alignment to MegAlign Pro** to launch the MegAlign Pro application already populated with the

▾ **Actions** Send template alignment to MegAlign Pro	alignment of the query sequence against the top ten threading templates.

Copyright DNASTAR, Inc. 2022

The NovaDock report

The Details panel is the only panel enabled when viewing the NovaDock report. This panel contains expandable areas that can be displayed or hidden by clicking the small blue triangle in its header.

Title **1CLV results.novadock**	The name of the NovaDock file.
Summary **Receptor:** 1 biopolymer with 470 residues **Ligand:** 1 biopolymer with 32 residues **Started:** 2016-08-30 07:57:59	• **Receptor/Ligand** – A description of the receptor and ligand molecules, including their lengths in residues. • **Started** – Date and time when the NovaDock prediction job began.
Options Options **Normal modes:** 5 (ligand), 5 (receptor) **Proposed contacts:** 0	• **Normal modes** – The number of normal modes of motion used to sample molecular flexibility during the docking process. • **Proposed contacts** – Number of residue contacts, if any, proposed by the user when initiating the NovaDock prediction. • **Repairs** – Repairs, if any, that were automatically made to the **Ligand** or **Receptor** in the course of making the NovaDock prediction.
Status Repairs **Ligand:** No repairs needed **Receptor:** No repairs needed	• **Prediction ID** – If available, an internal identification value assigned to the prediction. • **Version** – Version number of the NovaDock application used to make the docking prediction. • **Computation** – Time (in hours/minutes/seconds) needed to complete the docking prediction.

Copyright DNASTAR, Inc. 2022

Style panel

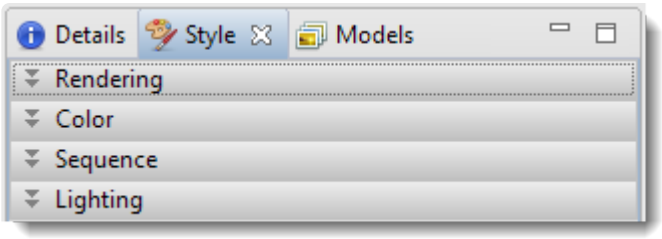
The Style panel, represented by a palette icon ( ), includes four expandable sections that affect the appearance of the Structure and/or Sequence views.

Opening the Style panel:

Do any of the following:

- Click on the **Style** tab
- Choose **View > Style**
- Press **Ctrl+Alt+Y** (Windows) or **Option+Cmd+Y** (Macintosh)

Using the Style panel:



Each section of this panel is described in a separate topic:

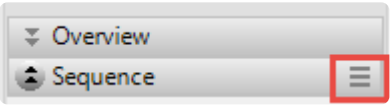
- Rendering
- Color
- Sequence
- Lighting

To save styles and apply them to future projects:

Once you have set up styles, you can save them and then apply them to future projects, either by default or only when you specify that they be applied. You can choose whether to save and apply styles for just one particular view or for all views.

Note that some of the instructions below refer to the “hamburger” menu. To access commands on these menus, click on the expand bar of interest in the Style panel. Then click on the tool consisting of three horizontal lines.

Copyright DNASTAR, Inc. 2022



Task	How To
Save all current styles to a file that can be applied to future projects	Use **View > Document Style > Save Style**.
Save the styles from one particular view to a file that can be applied to that view in future projects	Click the “hamburger” menu on the right of a Style panel expand bar and choose **Save (Expand bar name) Style**.
Save all current styles to a file that will be applied as the default for future projects	Use **View > Document Style > Save Style as Default**.
Save the current styles from one particular view to a file that will be applied as the default for that view in future projects	Click the “hamburger” menu on the right of a Style panel expand bar and choose **Save (Expand bar name) Style as Default**.
Apply styles that were saved earlier to the current project	Use **View > Document Style > Apply Style**
Apply styles that were saved for this view previously and apply them to the same view in the current project	Click the “hamburger” menuon the right of a Style panel expand bar and choose **Apply (Expand bar name) Style**.
Restore all DNASTAR default styles to the current project	Use **View > Document Style > Restore DNASTAR Default Style**.
Restore the DNASTAR default style for one particular view to that view in the current project	Click the “hamburger” menu on the right of a Style panel expand bar and choose **Restore DNASTAR Default (Expand bar name) Style**.

Copyright DNASTAR, Inc. 2022

Rendering section

The Rendering section is located in the Style panel, and provides options for customizing the look of the Structure view.

The following video is a brief walkthrough showing how to apply a different rendering style to highlight features of interest:

Opening the Rendering section:

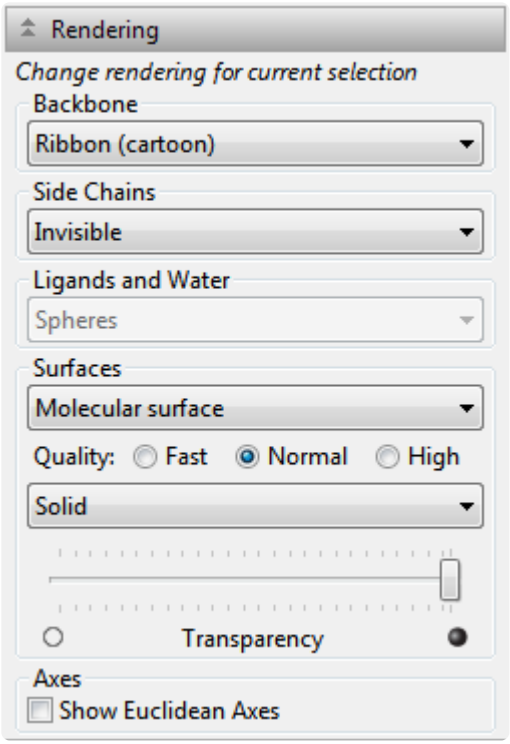
Do either of the following:

- Click the expand bar entitled **Rendering**
- Use **View > Style > Rendering**

> **Note:** The Rendering section is only enabled when a document is open in the Protean 3D application window.

Using the Rendering section:

Copyright DNASTAR, Inc. 2022

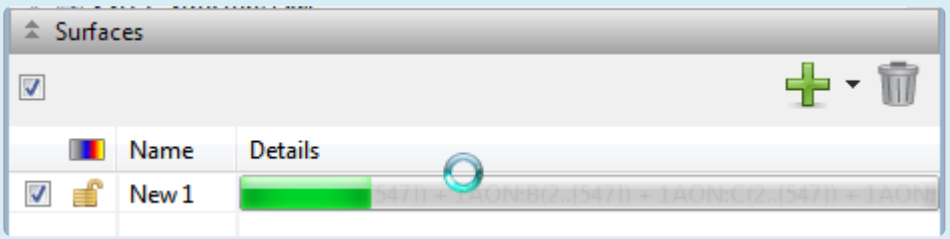


Task	How To
To globally change the appearance of a structural category (backbone, side chains, or ligands and water) in the Structure view	Make sure nothing is selected by double-clicking on the background of the Structure view. Select the desired rendering styles for each category (**Backbone**, **Side Chains**, **Ligands and Water**) from the corresponding drop-down menus. To view samples of each rendering style, see Examples of rendering styles. If the molecule is particularly large, note that some styles may take noticeably longer to update than others.
To change the appearance of a selected region	Make a selection in the Structure or Sequence view or in the Explorer panel. Select a new rendering style from one of the available menus. Rendering options will vary depending on what is included in your selection.
To display axes for the structure	Check the **Show Euclidean Axes** box. If you are rotating a structure, checking this box can allow you to compare the current position of the molecule to the original orientation.
To change the type of surface in the Structure view	Select the surface in the Surfaces section, then use the **Surfaces** drop-down menu in the Rendering section to choose between **Molecular surface** and **Solvent accessible surface**. See Examples of surface types for a visual comparison of these types.
To change the quality of a surface in the Structure view	Select the surface in the Surfaces section. Then, in the Rendering section to the right of **Quality**, click **Fast** for fastest surface rending with the least detail; **Normal** for average quality; or **High** for the slowest but most detailed rendering. See Examples of surface types for a visual comparison of the three options.

Copyright DNASTAR, Inc. 2022

To change the rendering of a surface in the Structure view	Select the surface in the Surfaces section. Then, in the Rendering section, make a selection using the drop-down menu above the **Transparency** slider. Select **Solid** for solid rendering, **Mesh** to display only the structure framework, or **Points** to show only the outline of the structure using dots. See Examples of surface types for a visual comparison of these surface rendering styles.
To change the transparency of a surface in the Structure view	Select the surface in the Surfaces section. Then, in the Rendering section, use the **Transparency** slider to make the surface more or less transparent. See See Examples of surface types for a visual comparison of an opaque and a semi-transparent surface.

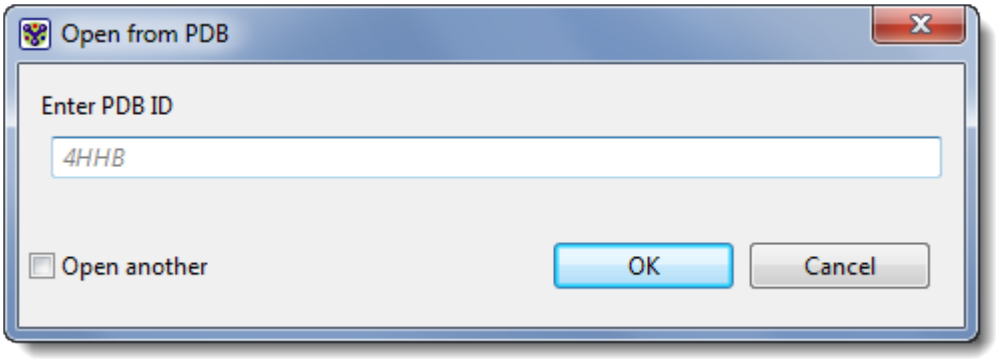
Note: Depending on the size of the structure, changes that affect the surface type and quality may take time for Protean 3D to calculate. In this case, a progress bar will appear in the Details column on the Surfaces section.



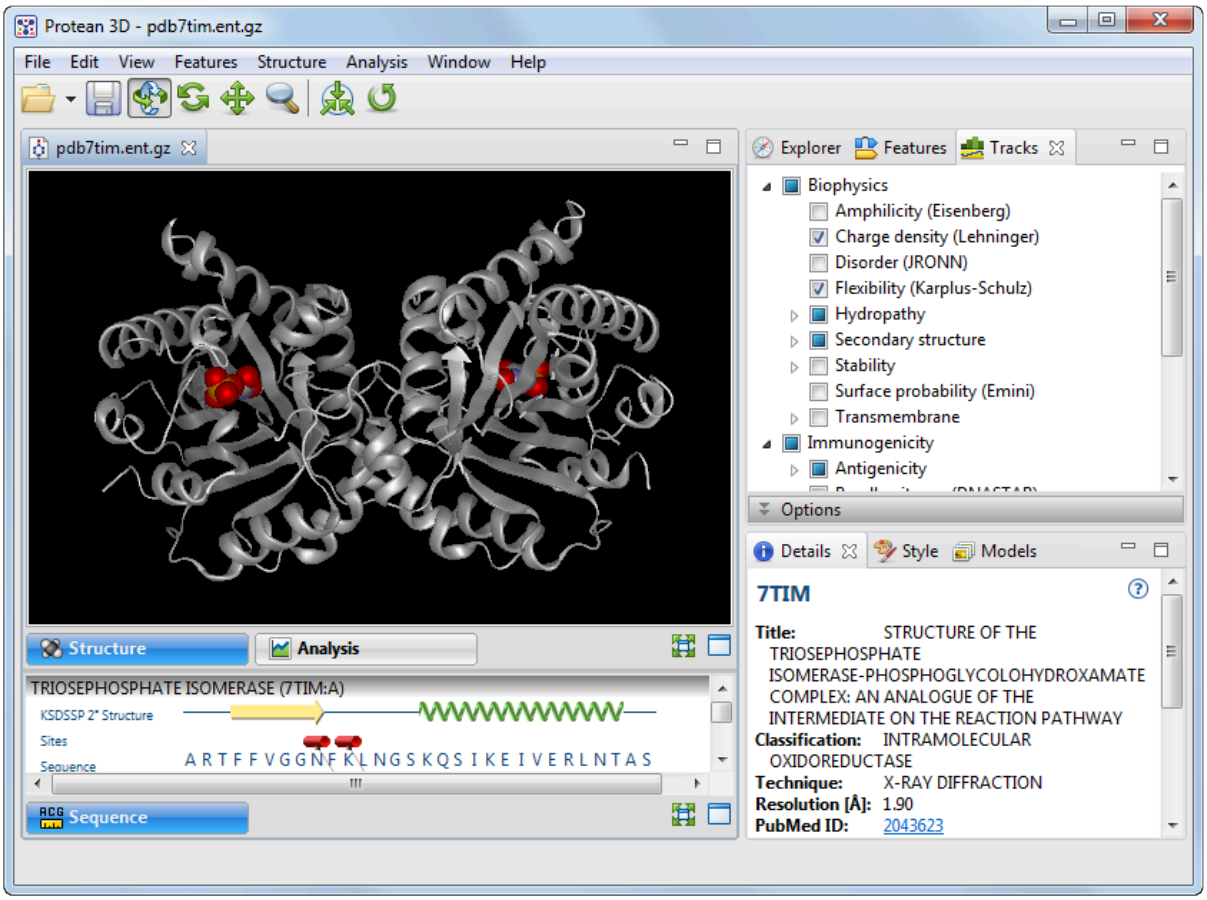
Try it! – Change the rendering of a structure

1. Within Protean 3D, choose **File > Open from PDB**. You may initially see a yellow warning icon ( ) prompting you to type in a valid ID.

2. Type **7TIM** into the text box.



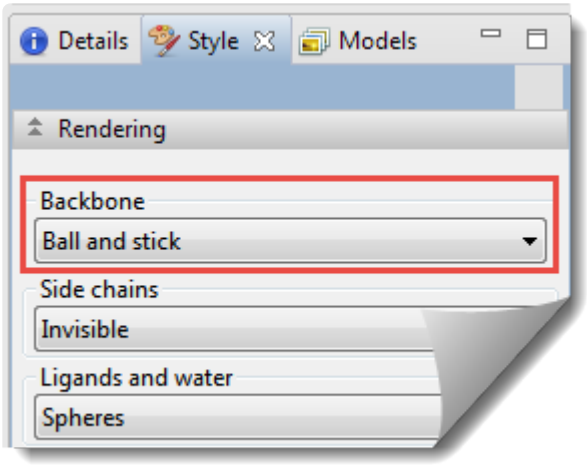
3. Press **OK** to open the structure in Protean 3D.



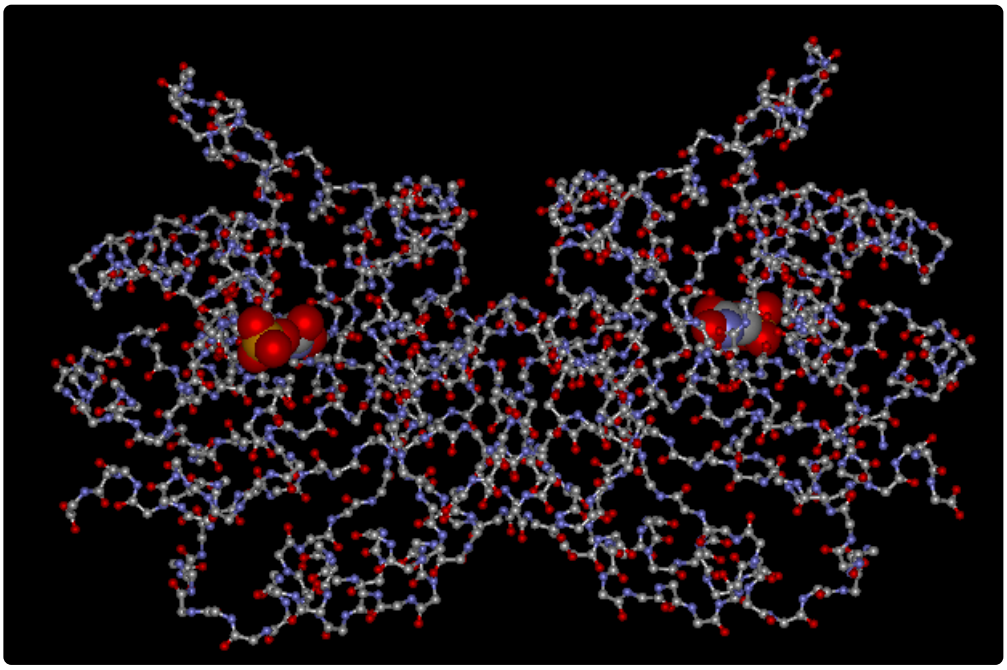
4. Go to the **Style** panel bar and expand the **Rendering** section. If the Rendering bar isn't visible, use the **View > Style > Rendering** command to reveal it.

Copyright DNASTAR, Inc. 2022

5. Select **Ball and stick** from the **Backbone** drop-down menu.

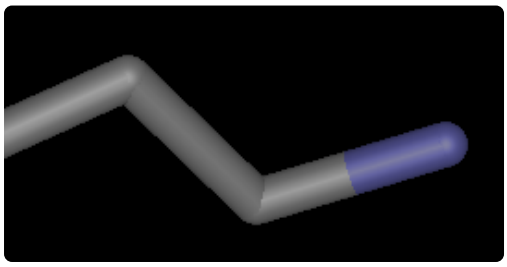
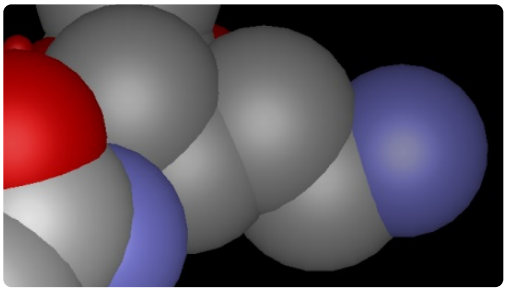
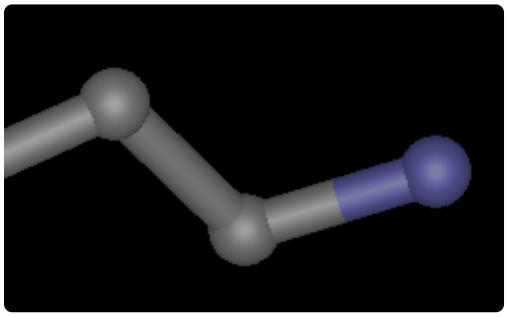
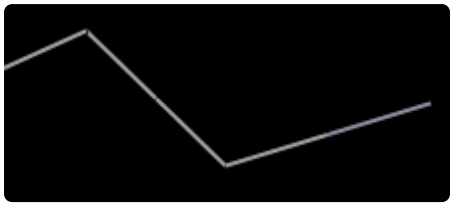


The structure now appears as follows:

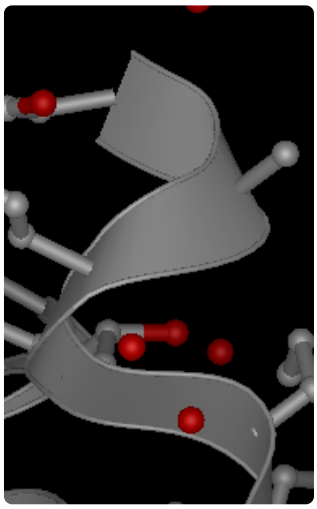
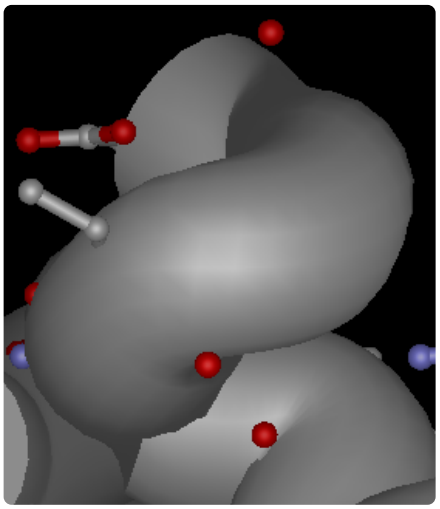


Examples of rendering styles

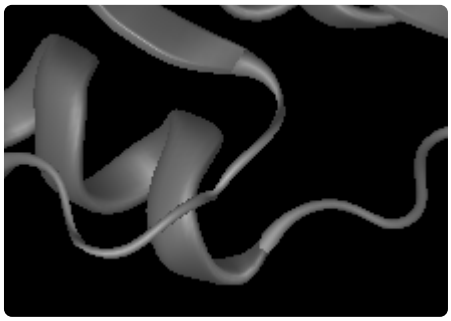
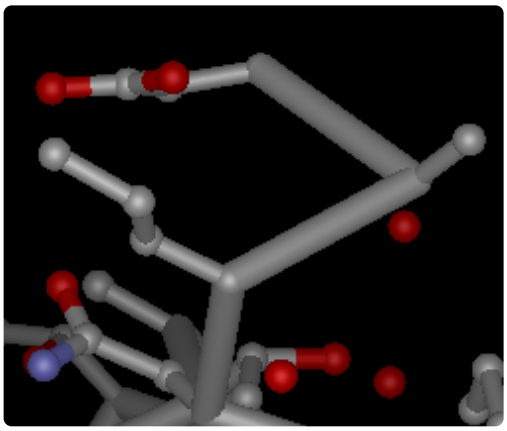
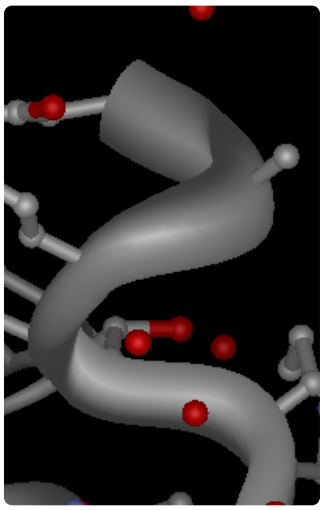
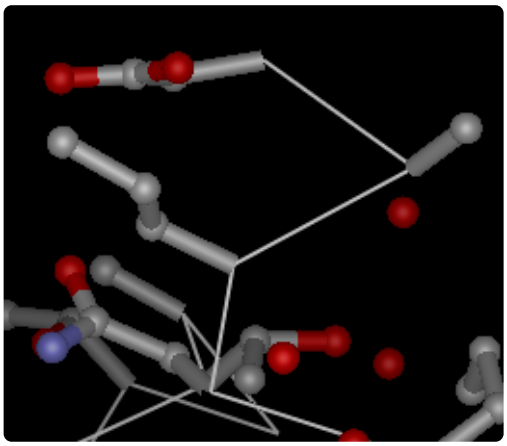
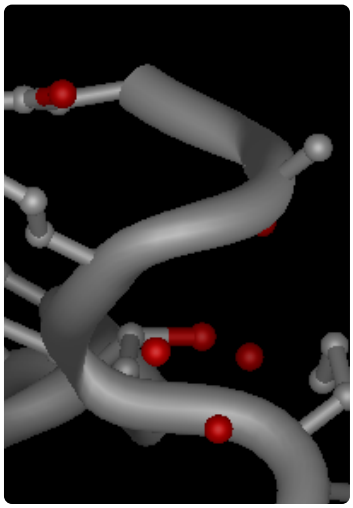
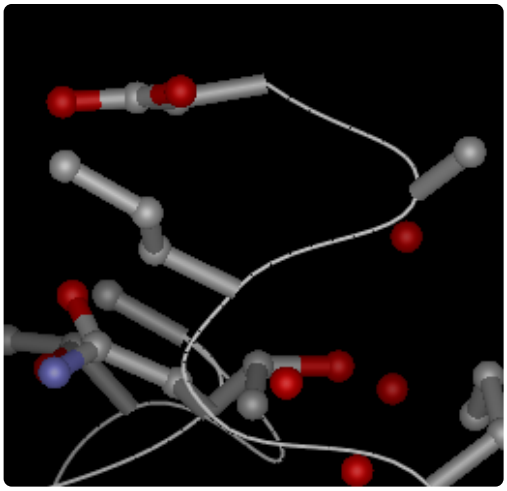
In the Rendering section, the following choices are available in all of the first three drop-down menus:

Sticks		Spheres	
Ball and Stick		Points	
Wireframe		Invisible	

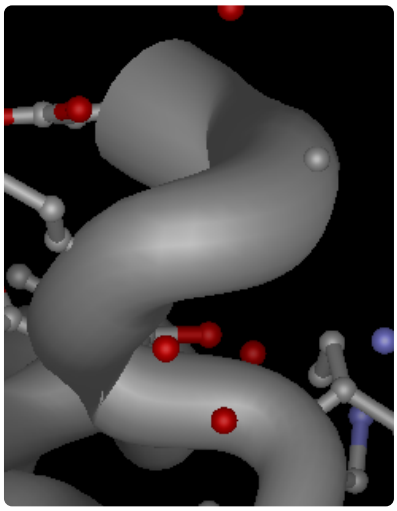
These choices are available in the **Backbone** menu only:

Ribbon		Tube (occupancy)	

Copyright DNASTAR, Inc. 2022

Ribbon (cartoon)		Trace	
Ribbon (solid)		Trace (wireframe)	
Tube		Trace (spline)	

Copyright DNASTAR, Inc. 2022

Tube (temperature)			

For all three “Tube” options, a thicker tube is indicative of more uncertainty with respect to the position of the atom. Temperature factors and occupancies are relative within the PDB file and should not be directly compared to those in other PDB files.

Tube (occupancy) – Draws the tube in proportion to the occupancy measured over time during x-ray crystallography experiments. Occupancy shows the percentage of time that the atom is in the given location, averaged over a time scale. The occupancy value is always between 0 and 1.

Tube (temperature) – Draws the tube in proportion to the temperature factor (B-factor) measured over time during x-ray crystallography experiments. The temperature value is always between 0 and 1. Low B-factors in the hydrophobic center of a molecule usually result in thinner tubes towards the center. Conversely, higher B-factors near the surface, among the outside chains, produce thicker tubes.

Copyright DNASTAR, Inc. 2022

Color section

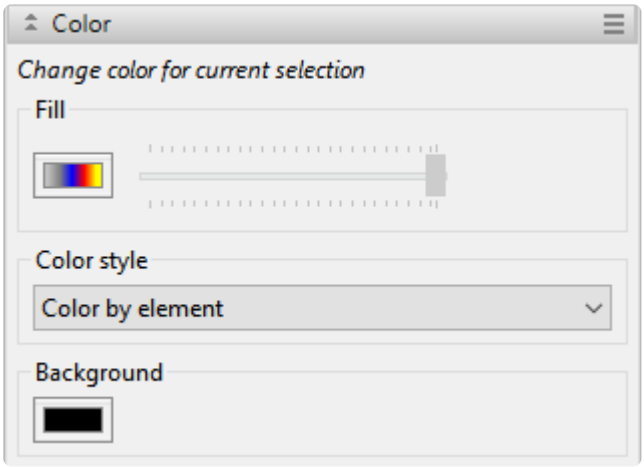
The Color section, located in the Style panel, lets you adjust colors of selected components of the Structure view or of that view's background.

Opening the Color section:

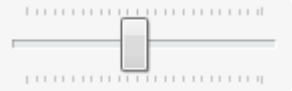
Do either of the following:

- Click the expand bar entitled **Color**
- Use **View > Style > Color**.

> **Note:** The Color area is only enabled when a document is open in the Protean 3D application window.



The following icons are used in this section:

Icon	Meaning
	Indicates a "mixed state" of colors for the current selection. (Or for the entire structure, if nothing is selected.)
	One specific color has already been assigned to the current selection.
	Opacity slider. This can be used to make a manually-chosen color more opaque or more translucent.

Tasks that can be performed in the Colors section:

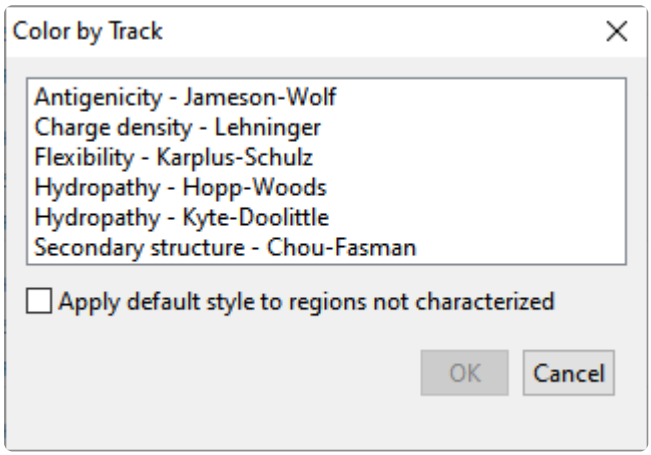
Copyright DNASTAR, Inc. 2022

Task	How To
Change the color of a selected structural component or feature in the Structure View	First make a selection in the Structure or Sequence views, or the Molecules or feature Style sections, then click the color box under **Fill**. If you don't make a selection, Protean 3D assumes you would like modify the color for all elements.
Change the color style for the entire structure	Like **Fill** color (see row above), color styles can be applied to the entire structure, or just to the current selection. The **Color Style** drop-down menu offers the following choices: • **Color by element** – Displays each element in the default CPK color. Choose this option to return to the default color scheme. • **Color by element (except carbon)** – Displays non-carbon atoms in the CPK color scheme without changing the colors of the carbon atoms. Before using this option, first color the structure using the **Color Style** or **Fill** color that you want to use for the carbon atoms. • **Color by structure** – Automatically colors each structure (i.e., all molecules in each structure) in a different color from any of the other structures. • **Color by chain** – Displays each chain with a different color. • **Color by position** – Displays the entire structure as a color gradient using a rainbow color scheme. The N terminus is colored in blue. The color gradient gradually moves through the entire spectrum of a rainbow (omitting indigo and violet), until reaching the C terminus, which is colored red. • **Color by amino acid chemistry** – Displays acidic, basic, polar-neutral and hydrophobic components in different colors. Note that if the style is applied globally, or if DNA, RNA, ligands or solvent are included in the selection, these items will be exempted from the color change.

Copyright DNASTAR, Inc. 2022

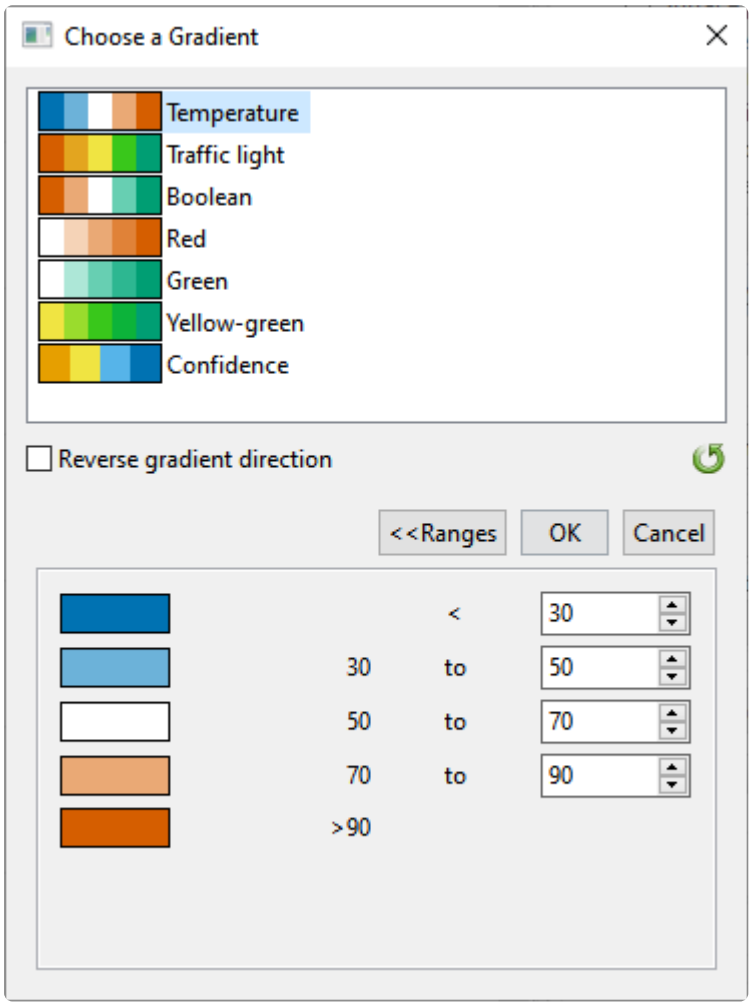
Amino Acid	Side Chain Chemistry	Color
Arg (R), His (H), Lys (K)	Basic (positive charge)	blue
Asp (D), Glu (E)	Acidic (negative charge)	red
Cys (C), Asn (N), Gln (Q), Ser (S), Thr (T)	Polar (neutral)	green
Ala (A), Gly (G), Ile (I), Leu (L), Met (M), Pro (P), Val (V)	Nonpolar (aliphatic)	gold
Phe (F), Trp (W), Tyr (Y)	Aromatic	yellow

- **Color by track** – Colors the molecule in the Structure view to correspond to colors in an Analysis view track of your choice. A popup appears with a list of all tracks currently displayed in the Analysis view.



Choose the desired track from the list. (Note: If the analysis track of interest isn't on the list, click **Cancel**, check the desired track(s) in the Tracks panel, then come back to the Color section and choose **Color by track** again). If you want to apply the default style, **Color by element** to regions of the molecule that are not colored in the Analysis view track, you can check **Apply default style to regions not characterized**.

- **Color by temperature** – When you select this option, the "Choose a Gradient" popup appears, with only the upper half displayed initially.



Choose from a variety of monochrome and multi-colored heat maps reflecting the temperature factor (B-factor) for each atom. The temperature represents the uncertainty for a residue's position in an experimental crystal structure, where higher values indicate greater uncertainty. To see a legend for the selected color scheme, click on **Ranges** to reveal the lower half of the popup. If desired, you can specify custom range thresholds for each color. You can also click the checkbox in this dialog to reverse the direction of the gradient.

When used with a NovaFold AI-predicted structure, the **Confidence** option shows residue-level confidence in the prediction. By default, this color scheme uses the same colors and thresholds used by EMBL-EBI in their AlphaFold 2 model database. Residues with the highest level of confidence are shown in dark blue, while those with the lowest level are displayed in orange.

- **Fill with solid color** – Colors the entire structure a single color.

Override a color style for one	Apply a color style using the **Color Style** menu. First select the desired component. Then choose a separate color for that component by clicking the box under **Fill**.

particular component	
Change the background color of the Structure View	Click the color box under **Background**.
Undo the last color change	Use the **Edit > Undo Fill Color** command.
Return all colors to their default values	Clear the selection, then choose **Color by element (default)** from the Color Style drop-down menu.
Change the transparency of the active color	After making a selection and choosing a color, move the slider right to make the color more opaque, or left to make it more transparent.

Copyright DNASTAR, Inc. 2022

Sequence section

The Sequence section, located in the Style panel, allows you to change display characteristics of the Sequence view.

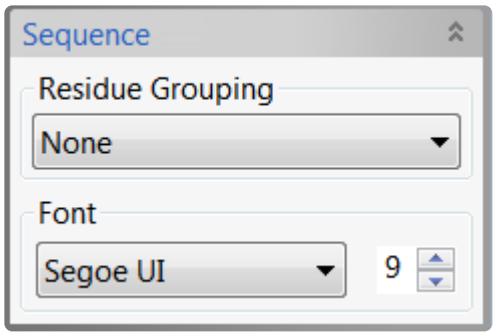
Opening the Sequence section:

Do either of the following:

- Click the expand bar entitled **Sequence**
- Use **View > Style > Sequence**

> **Note:** The Sequence section is only enabled when a document is open in the Protean 3D application window.

Using the Sequence section:



Task	How To
To change the grouping of residues in the Sequence view	Use the **Residue Grouping** menu to choose between showing uninterrupted sequence or grouping it in sets of ten.
To change the Sequence view font style	Use the **Font** menu to choose between seven available fonts. Recently selected fonts appear at the top of the menu, and other available fonts appear below the row of dashes.
To change the Sequence view font size	Use the up and down arrows ( ) next to the number to select a font size of **8** or larger.

Copyright DNASTAR, Inc. 2022

Lighting section

The Lighting section, located in the Style panel, allows you to change lighting characteristics of the Structure view. Up to three “directional” lights may be deployed, in addition to displaying “ambient” light.

The following video shows the Lighting section in action:

To open the Lighting section:

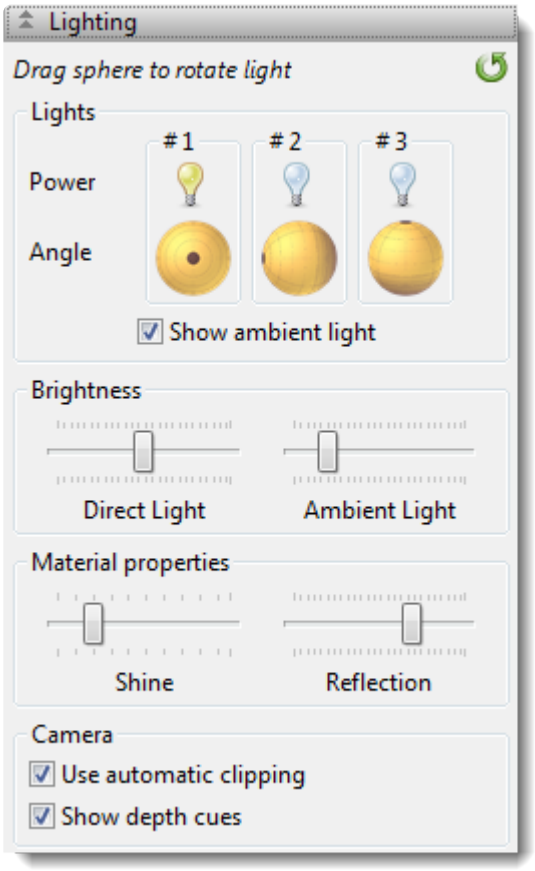
Do either of the following:

- Click the expand bar entitled **Lighting**
- Use **View > Style > Lighting**

> **Note:** The Lighting section is only enabled when a document is open in the Protean 3D application window.

Using the Lighting section:

Copyright DNASTAR, Inc. 2022



Task	How To
To turn a light source on or off	Toggle between light on/off by clicking the light bulbs in the **Power** section. To turn ambient light on or off, check or uncheck the box next to **Show Ambient Light**.
To change the lighting angle	Use your mouse to rotate any of the three spheres in the **Angle** section. Changes will only be apparent when the power for the corresponding light is turned on.
To change the amount of lighting	Use the **Direct Light** and **Ambient Light** sliders in the **Brightness** section.
To change the reflectivity of the structure	Use the **Shine** and **Reflection** sliders in the **Material Properties** section.
To hide distracting elements in the foreground as you zoom in on the Structure view	Check **Use automatic clipping**.
To add a “3-D” look	Check the **Show depth cues** box in the **Depth Cues** section. This causes the back of the structure appear to fade out, an effect which is most noticeable on a white background.
To return to the default settings	Click the **Revert** icon ( ).

Copyright DNASTAR, Inc. 2022

Models panel

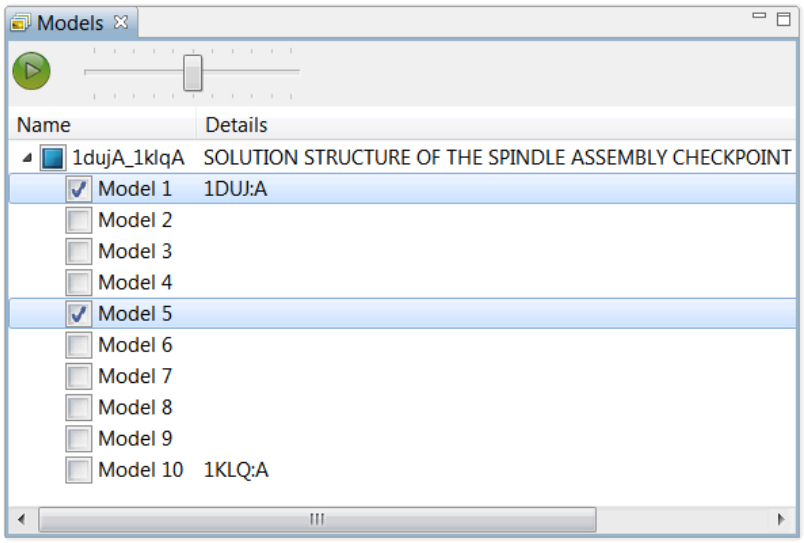
The Models panel, represented by a picture icon ( ), lets you control the visibility of individual models in multi-model PDB files.

Opening the Models panel:

Do any of the following:

- Click on the **Models** tab
- Use **View > Models**
- Press **Ctrl+Alt+O** (Windows) or **Option+Cmd+O** (Macintosh)

Using the Models panel:



The PDB names (if available) of the first and last model are shown in the Details column, with intermediate models having only a Model number. The Models panel requires that at least one Model be selected at all times. In the illustration above, **Model 1** and **Model 5** are checked, meaning that both interim folding models are overlaid in the Structure view. The checkbox to the left of the file name is filled in, indicating that some models are checked and others are unchecked.

If you are viewing a single-model file, the Models window is non-functional and contains only **Model 1**, which cannot be unchecked.

Task	How To
To show or hide the model	Click on the corresponding checkbox.
To select a single model and unselect	Click on model name.

Copyright DNASTAR, Inc. 2022

all others	
To select all models	Check the box to the left of the file name.
To return to the state before selecting all models	Uncheck the box to the left of the file name.
To play the animation	Click the play button ( ).
To stop the animation	Click he pause button ( ).
To control the speed of the animation	Move the slider ( ) left to go slower, or right to go faster.

Copyright DNASTAR, Inc. 2022

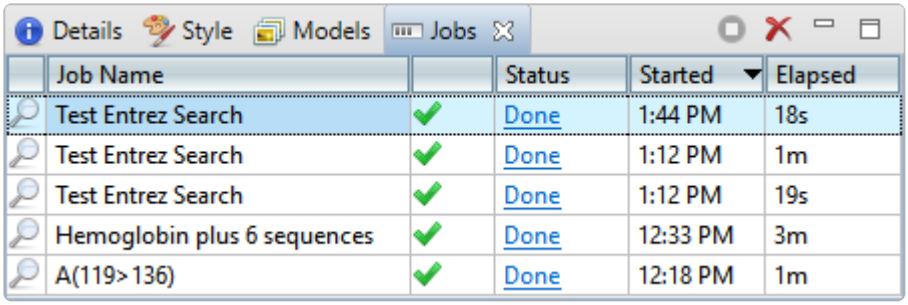
Jobs panel

After initiating an NCBI BLAST or Entrez search, you can monitor its progress and download its eventual output from the Jobs panel. This panel is not visible in a project until you have started running at least one search.

To open the Jobs panel:

Do any of the following:

- Click on the **Jobs** tab.
- Press **Ctrl+Alt+J** (Win) or **Option+Cmd+J** (Mac).
- Use **View > Jobs** or **Search > Show Searches**.



The Jobs panel consists of a header and a table with the following columns:

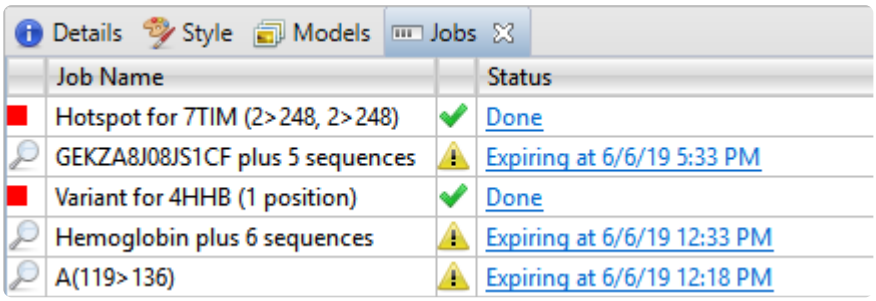
Column	Description
Job Name	Text searches are all named with the field and term of the first search criterion followed by “plus # terms,” where # represents the number of additional criteria. Sequence searches use the name specified in the Job screen of the Search wizard. The same **Job Name** can be used for more than one prediction.
Status	A word and icon indicating the status of the search. • **Started** – **Run** or **Run Job** were pressed in the Search wizard • **Queued** – – Submitted and placed into the queue • **Running** – – Running • **Done** – – Job complete; results available • **Failed** – – Job failed • **Stopped** – – Job stopped by user
Started	The date and time the job began.
Elapsed	The total elapsed time between starting and completing the job.

Copyright DNASTAR, Inc. 2022

Tasks that pertain to the Jobs panel:

Task	How To
To open the results of one or more jobs	Click on the **Done** (Done) or **Expiring** links for that job. See below this table for more information on job expiration. You can also open a saved *.search* file using **Search > Open Search File**.
To stop one or more jobs that are in progress	Select one or more rows with a Status of **Started**, **Queued** or **Running**. Then press the **Stop** tool ( ) in the header above the table. Once stopped, a job cannot be restarted from the Jobs panel. You must instead initiate a new search.
To delete a finished job	Select one or more rows with a Status of **Done**, **Failed** or **Stopped**. Then press the **Delete** tool ( ) in the header above the table.
To open/reopen the progress dialog for a job	Select one or more rows with a Status of **Started**, **Queued** or **Running**. Then right-click and choose **Open Progress dialog**.

Sequence (BLAST) and text (Entrez) search jobs are intended to be ephemeral and expire after 7 days. The Jobs panel notifies you of upcoming expiration dates.



All jobs are recorded in the location specified in the Temporary File preferences dialog. If you are a Macintosh user and elect to keep the default save location, the operating system may remove jobs after a reboot. To avoid this, we recommend choosing a different location in which to save temporary files.

Copyright DNASTAR, Inc. 2022

Motion Library

The Motion Library consists of over 500 animated and annotated macromolecular motions illustrating structural conformational changes. The intermediate structures in the animation are an interpolation between two experimentally-determined protein conformations calculated by the Morph Server.

The structures that define each Motion Library motion are chemically plausible, meaning that bond lengths, bond angles, and steric interactions within the structures are confined insofar as possible to reasonable geometries. The models do not include kinetic or other dynamic considerations. Nevertheless, the interpolated intermediates are likely similar to a subpopulation of the biologically-relevant conformational change intermediates (Krebs and Gerstein, 2000).

Annotations in the Motion Library are derived from the Protein Data Bank (PDB), the CATH protein structure classification database, and the Gene Ontology (GO) database. Annotations are based on the accurate assignment of PDB identifiers for the first and last conformations in the motion.

The following video is a brief introduction to the Motion Library:

To open the Motion Library:

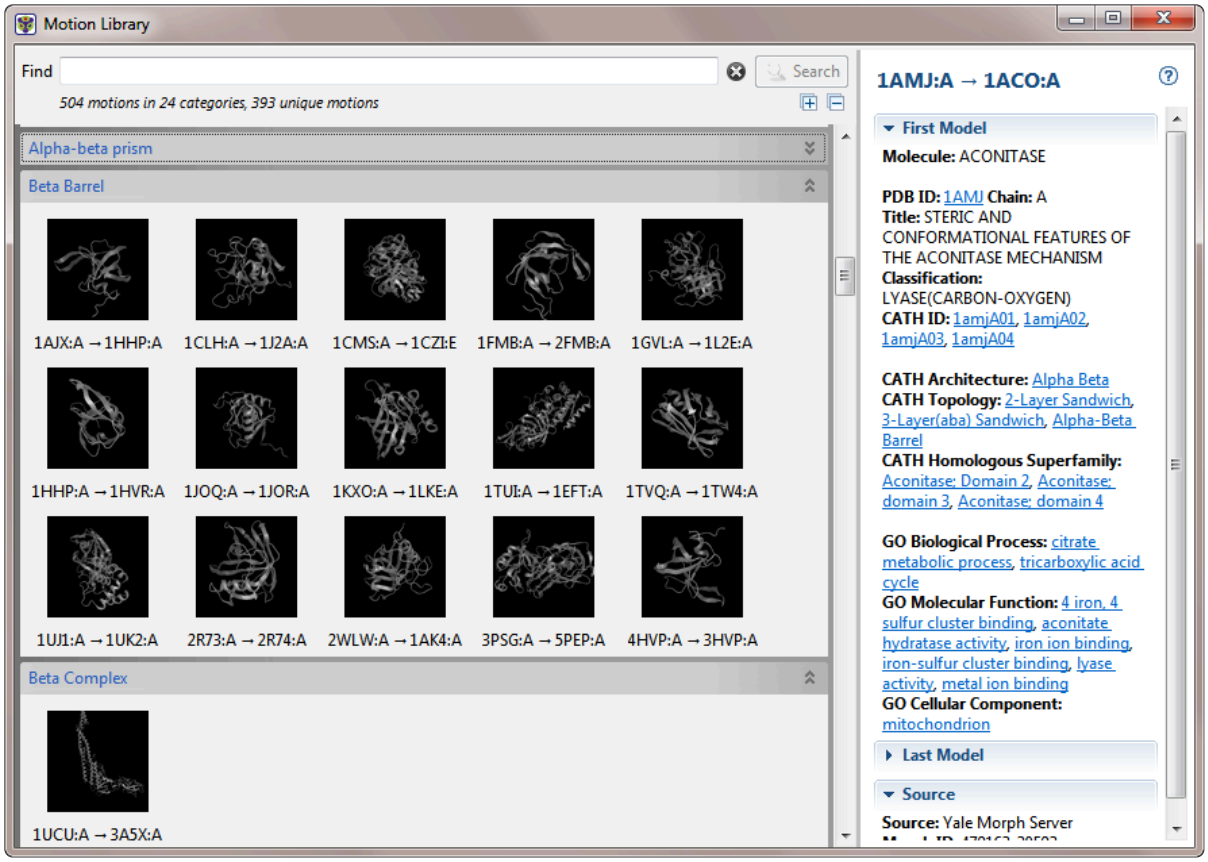
Do either of the following:

- Use **View > Motion Library**
- Press **Ctrl+Alt+L** (Windows) or **Option+Cmd+L** (Macintosh).

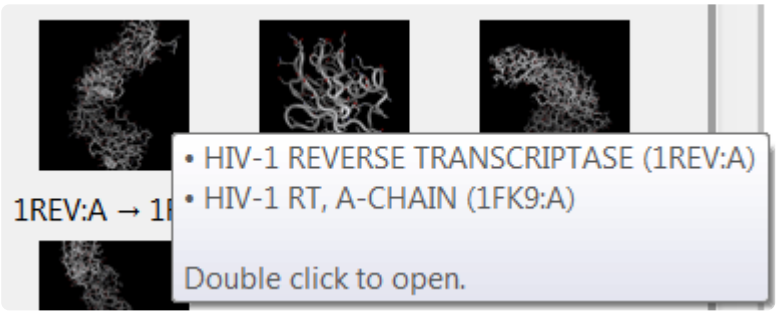
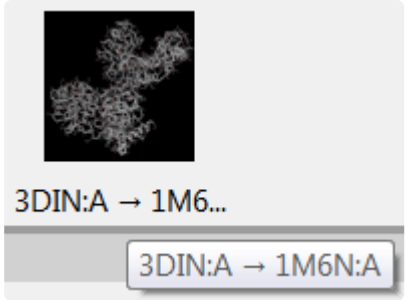
To use the Motion Library:

In the Motion Library Window, motion previews appear as image thumbnails. These represent all of the protein motions in the Library as animated previews grouped by CATH Architecture (i.e. the “A” term of the CATH hierarchy: “Alpha-Beta Complex,” “Propeller,” etc.). Motions without a CATH ID are grouped into an “Uncategorized” section at the bottom.

Underneath each preview are partial PDB IDs of the first and last structure sequences in the animation. Each preview group is part of an expand bar that can be collapsed or expanded as desired.



Task	How To
To open a motion in the main viewer	Double-click on its preview.
To expand or collapse all categories (expand bars) simultaneously	Click the **Expand/Collapse All Categories** tools ( ) under the **Search** button.
To view the names and PDB IDs of the first/last	Hover over the preview.

structures	
To view the full PDB IDs of the first/last structure	Hover over the partial ID listed below the preview. 
To search for a particular motion	Use the **Find** field and associated **Search** button. These let you perform a simple (non-Boolean) “exact term” search of the Motion Library. After executing a search, only the motions that match the search term will remain visible in the preview area. If your search term contains a misspelling, you will be redirected to a help dialog. After performing a search, clicking the “x” icon ( ) to the left of the **Search** button clears the results and makes all motions visible again in the preview area.
To view detailed information about a motion	Click once on the motion’s preview to display its information in the Information Panel.The information appears in three expand bars: **First Model**, **Last Model**, and **Source**. Some information is available for all motions in the Motion Library. CATH and GO information is shown if available, and links to the corresponding web pages. The PDB ID links to the entry in the RSCB Protein Data Bank.

Copyright DNASTAR, Inc. 2022

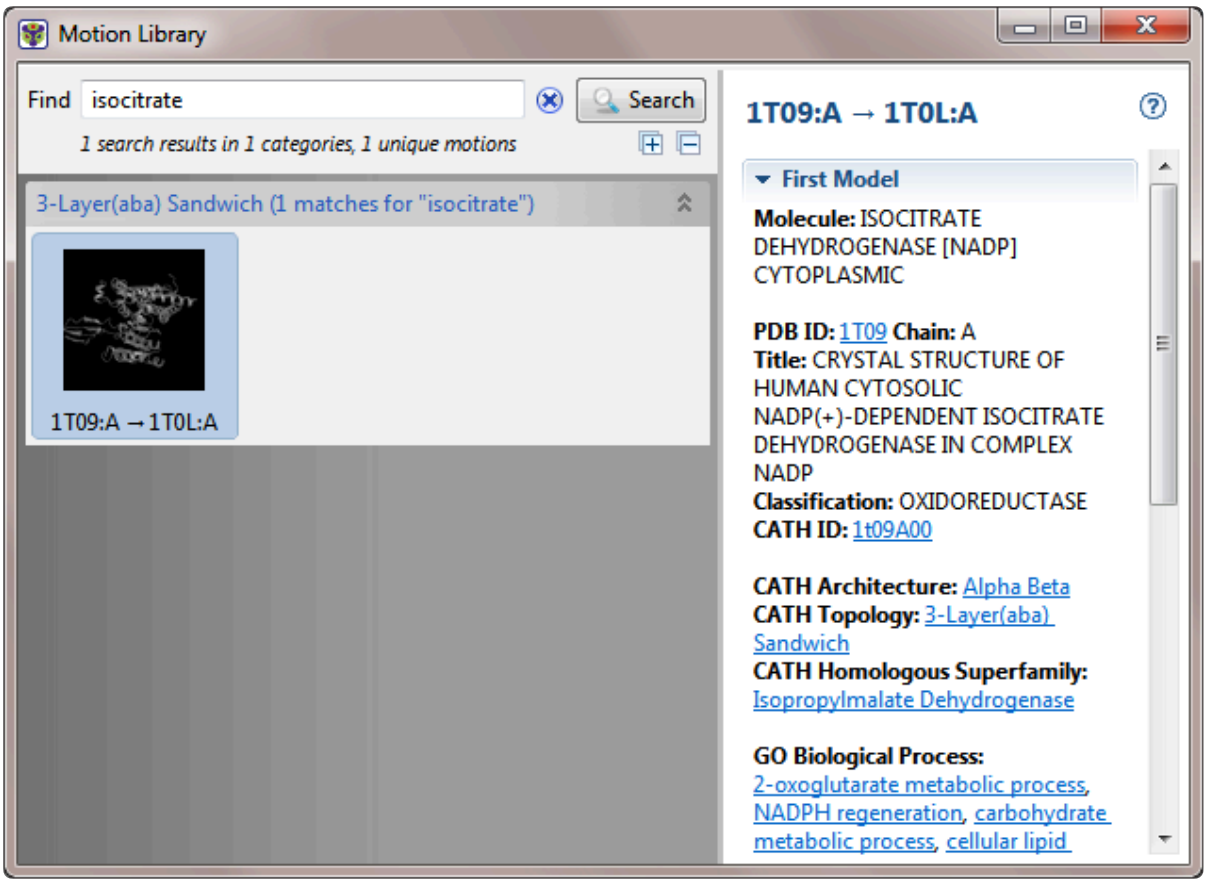
Try it! – Open a multi-model file from the Motion Library

This tutorial shows how to open and work with a multi-model file from the Motion Library.

1. Choose **File > Open from Motion Library** to open the floating Motion Library window.

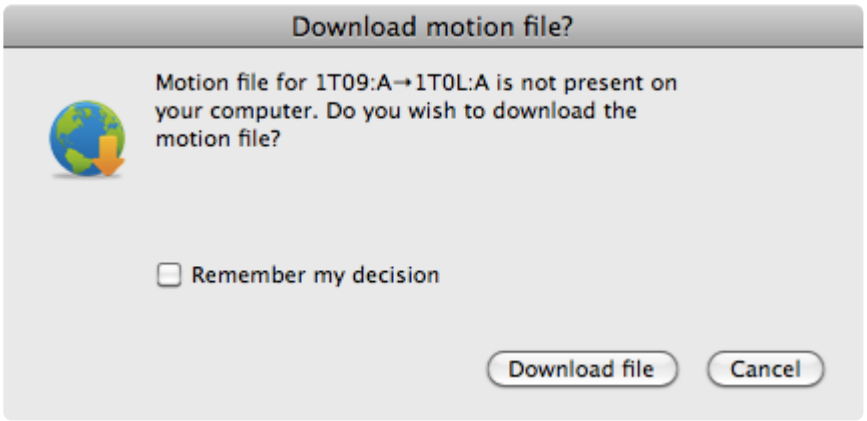
2. In the Find area, enter *isocitrate* and click **Search**. One match will be shown. Notice that the animated thumbnails that did not match the search term are now hidden.

   The panel on the right displays detailed information about the first and last models in the animation. Hyperlinks lead to additional data about the molecule in the PDB, CATH, and GO databases.



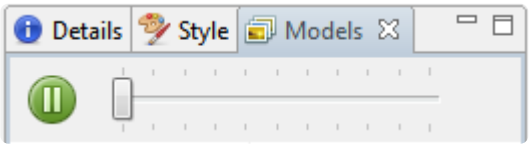
3. Double-click on the thumbnail to open the animation as a new document. If you have not previously downloaded the Motion Library, you will be prompted to download the file you selected.

Copyright DNASTAR, Inc. 2022



4. Check the **Remember my decision** box if you would like to download individual Motion Library files in the future without being prompted. Then click **Download file**.

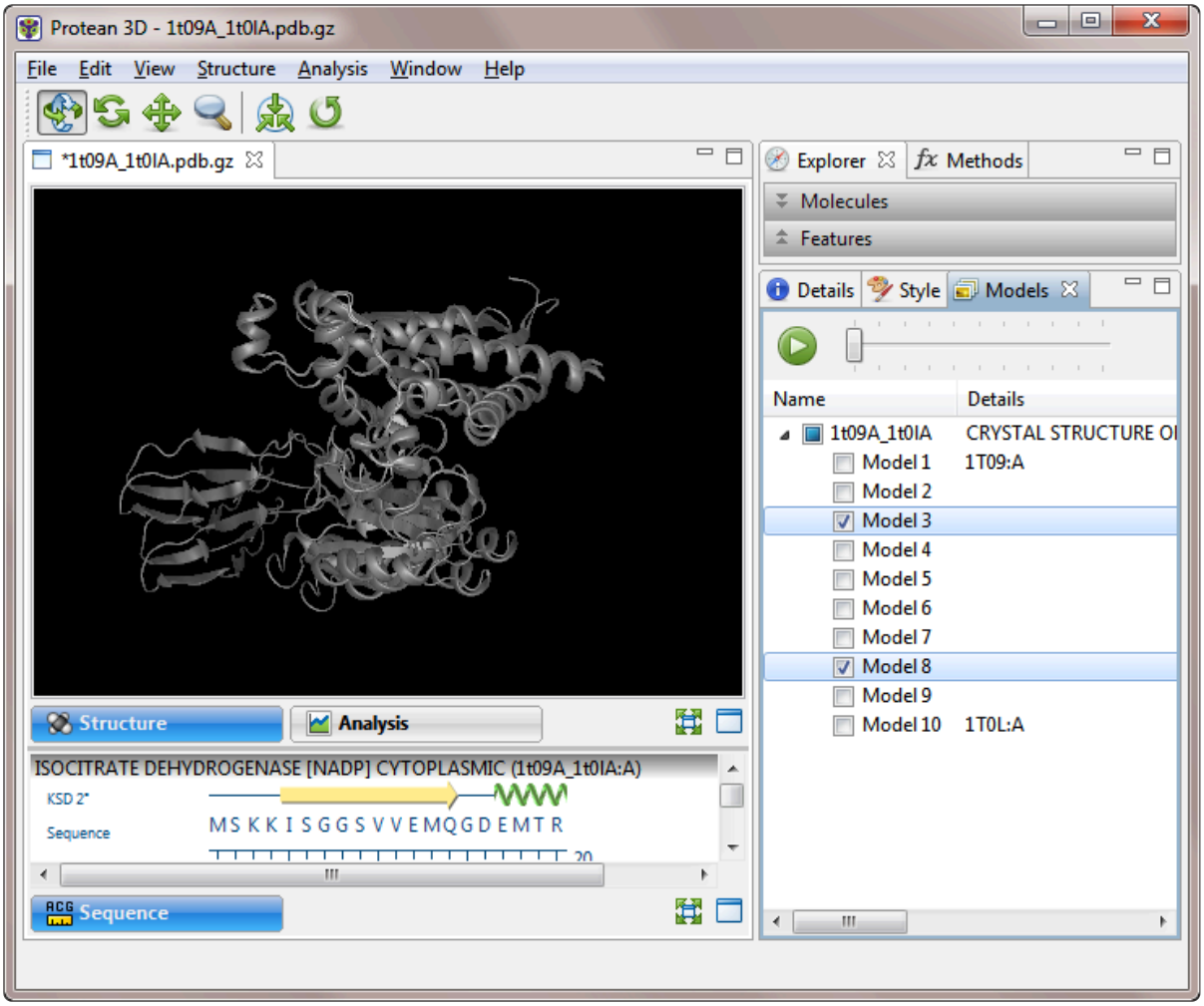
5. In the Models panel, move the slider to the left to slow the animation.



6. Click the **Pause** button ( ) to stop the animation. The **Model** last displayed in the Structure view as a “movie frame” remains checked and in view.

7. Check a second **Model** box to overlay the two models in the Structure view.

Copyright DNASTAR, Inc. 2022



Overlays are helpful for checking where multiple models in NMR files agree and disagree. They can also be used as presentation graphics to show the path of motion in a movie file.

You can also perform the same tasks (e.g., selecting, recoloring or hiding various structural elements and features, changing the structure rendering) with multi-model files as you can with single-model files . For ideas, see the Try Its! that use the PDB file *7TIM*.

Customize the Look and Layout

The following topics describe the many ways you can customize the look and layout of Protean 3D.

- Use the Style panel to change the look of a view (font styles and sizes, colors, number of decimal places, etc.)
- Customize the appearance and layout of views
- Customize the layout of other window components
- Apply information tracks to a view

Once you have customized the look and layout as desired, you can:

- Save Styles for individual views and apply them to other projects
- Save the entire layout and apply it to other projects

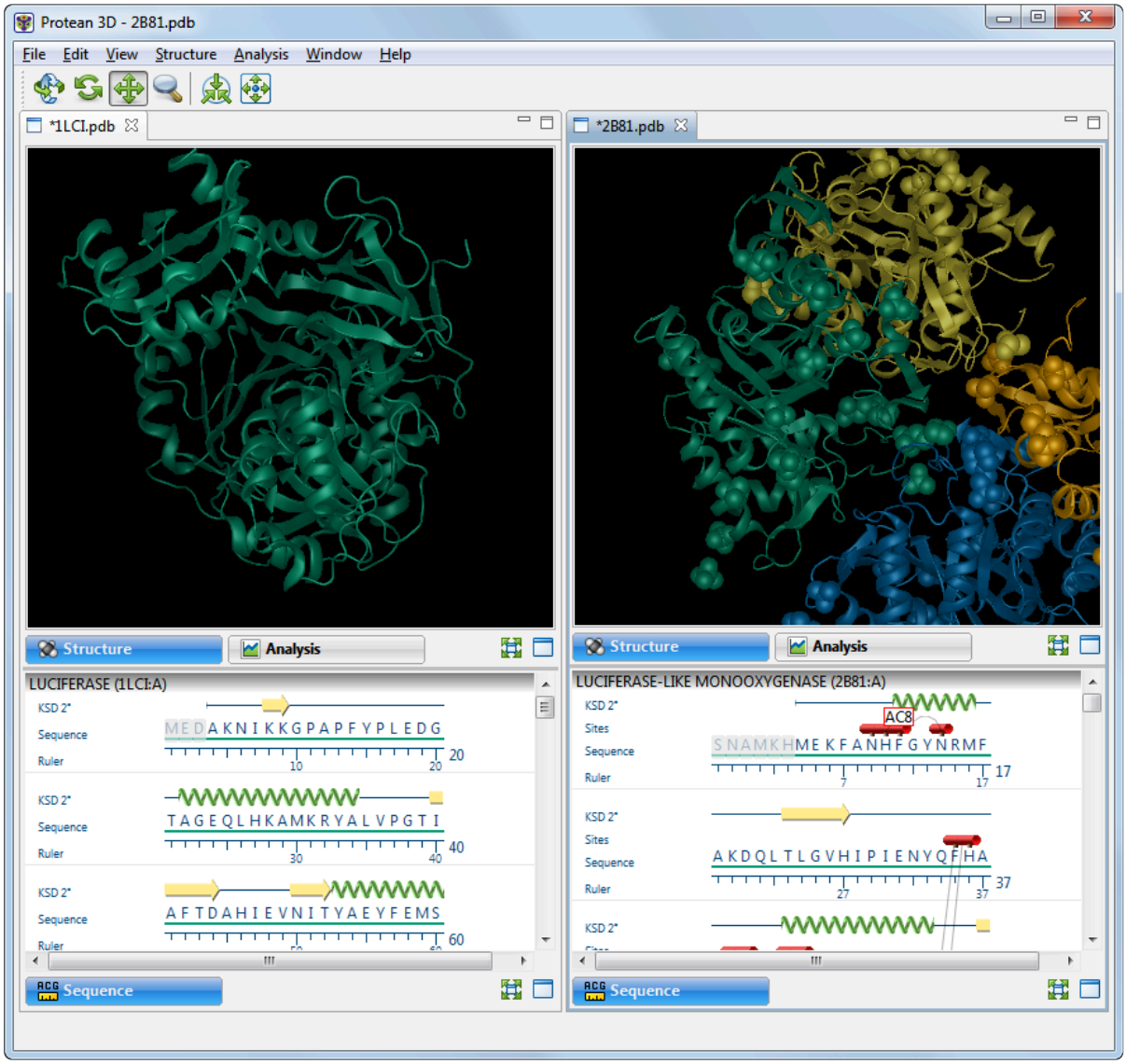
Copyright DNASTAR, Inc. 2022

Customize the appearance and layout of views

The following table lists ways in which you can customize the appearance and layout of views.

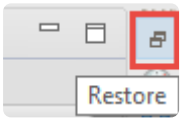
Task	How to…
To show a view	Use **View > (View Name) > Show**.
To detach the view	Use **View > (View Name) > Detach** or right-click on the view tab and choose **Detach (View Name)**. Detaching a view turns it into a free-floating window.
To reattach all detached views to their original locations	Use **View > Reattach All Views**.
To configure the number and location of visible views.	Use the **View > Document Layout** command or the **Change layout** tool (image varies) in the bottom right corner of the view to access possible layouts. The following options are available: • **One** – One – To show only the current view. • **Two Horizontal** (Two Horizontal) – To show the current view on top and one additional view below it. • **Two Vertical** (Two Vertical) – To show the current view on the left, and one additional view to the right. • **Three Horizontal** (Three Horizontal) – To show the current view on top and two additional views below it. • **Three – Left** (Three - Left) – To show the current view on the left and two additional views stacked one on top of the other on the right. • **Three – Right** (Three - Right) – To show the current view on the right and two additional views stacked one on top of the other on the left. • **Three – Top** (Three - Top) – To show the current view on the top and two additional views side-by-side on the bottom. • **Three – Bottom** (Three - Bottom) – To show the current view on the bottom and two additional views side-by-side on the top.
To show the active view in full-screen mode	Use **View > Full Screen**, click the the **Full screen** tool at the bottom right corner of the view, or press **Ctrl/Cmd+Shift+F**.

Copyright DNASTAR, Inc. 2022

To exit full-screen mode and return to normal mode	Press the **Esc** key or **Ctrl/Cmd+Shift+F**, or click on the **Exit Full Screen** tool ( ) at the bottom of the full-screen view.
To view two or more documents simultaneously on a split screen	By default, only one document at a time is displayed, even if several documents are open. To view two or more documents simultaneously on a split screen, drag a project tab in a downward, right or left direction. When you see a black arrow, release the mouse button.  Documents can be re-docked in their original positions by reversing this process.

Customize the layout of other window components

Controls for this application's view, panels and expand bars are described in the following table. For information that pertains only to views, also see Customize the appearance and layout of views.

Task	"How To"
Show or open a hidden item	Use the associated **View** menu command, click on a closed expand bar, click on a plus sign, or click the **Restore** icon ().
Show/Hide the side panels	Double-click on the project name to toggle between showing and hiding the side panels on the right of the window.
Display panel sub-sections	Click on a closed expand bar (medium gray bar) or use the associated **View** menu command. If you can't see all the information in an expanded section, close other expanded sections and/or increase the size of the whole panel as described elsewhere in this table.
Hide panel sub-sections	Click on an open expand bar.
Maximize an item	use the **Maximize** icon (.
Change the relative sizes of panes	Drag the divider bar between them to the left, right, up or down.
Hide or close a visible item	Use the **Close** icon ( ) or **Minimize** icon ( ), click on an open expand bar, or click on a minus sign. To close one or more tabs, right-click on a project, view or panel tab and choose **Close**, **Close Others**, **Close Tabs to Right**, or **Close All**. Some of these options are only available for certain tabs.
Move a view, panel or project to another location	Right-click on the view tab and choose **Move (View Name) to (Location)**. In some cases, you can use your mouse to grab a project, view or panel tab, then drag and drop it in the desired location.
Scroll vertically	Use the vertical scrollbar or the mouse wheel.
Scroll horizontally	Use the horizontal scrollbar, or hold down the **Alt** key (Win) or **Option** key (Mac) and drag the mouse.

Copyright DNASTAR, Inc. 2022

Return to the default side panel layout	Choose **View > Reset Side Panel Layout**. This command restores the default layout of panels, including opening any default panels that have been closed; and restoring panels that have been minimized or maximized. It will not reattach a detached view. In that case, use **View > Reattach All Views** instead.
Return to the default layout for the entire window	Choose **View > Document Layout > Restore DNASTAR Default Layout**.

Copyright DNASTAR, Inc. 2022

Save the layout and apply it to other projects

You can change the layout of Protean 3D's views and panels as described in Customize the look and layout.

Once you have arrived at a layout you like, you can save it and then apply it to future projects, either by default or only when you specify that it be applied.

Task	How To
Save the current layout to a file that can be applied to future projects	Use **View > Document Layout > Save Layout**.
Save the current layout to a file that will be applied as the default layout for future projects	Use **View > Document Layout > Save Layout as Default**.
Apply a layout that was saved earlier to the current project	Use **View > Document Layout > Apply Layout**.
Restore the DNASTAR default layout to the current project	Use **View > Document Layout > Restore DNASTAR Default Layout**.

Copyright DNASTAR, Inc. 2022

Open a File or Project

The following topics discuss different ways of opening structure files, protein sequence files, or saved Protean 3D projects;

- Open one or more local files (protein sequences, structures, or NovaCloud results files)
    - (Change the default application for opening files)
- Open a Protein Data Bank file when you know the PDB ID
- Open a Protein Data Bank file when you don't know the PDB ID
- Open a structure from Protean 3D's Motion Library
- Open an existing Protean 3D project

Copyright DNASTAR, Inc. 2022

Open one or more local files

Protean 3D allows you to open a wide variety of supported file types, including protein sequences, structure files, and prediction results from the Nova applications.

To open a single file within a Protean 3D project:

Use the **File > Open** command, or press **Ctrl+O** (Win) or **Cmd+O** (Mac). For an additional shortcut to this command, see The Welcome Screen.

Once a document is open in Protean 3D, you can open additional files using the same command. If no document is currently open, you can also use the "Open a file" link in the Details panel.

In the rare case that you have a protein file whose type is not supported in Protean 3D, try opening it with SeqBuilder Pro. If you are able to view the file there, you can then save it in Lasergene protein file format (*.pro*) and open the saved file in Protean 3D.

To open multiple files within a single Protean 3D project:

Choose **File > Open Group**. For an additional shortcut to this command, see The Welcome Screen.

Protean 3D allows you to open multiple files (*.structure*, *.pdb*, *.pro*, etc.) within a single document. This is particularly useful if you are working with PDB "split entry" structure files, in which one structure has been split between two or more files due to PDB format constraints. This functionality can also be used to overlay related structures in order to investigate their similarities and differences. All panels and views in Protean 3D support the viewing of multiple-structure documents. This feature is not intended for use in investigating protein-protein docking. Animation in the Models panel is disabled for these documents.

Copyright DNASTAR, Inc. 2022

Change the default application for opening files

Windows 7 and Windows 8:

1. Open the **Control Panel**.
2. Go to **Programs > Default Programs > Associate a file type or protocol with a specific program.**
3. Select extension you wish to change in the **Extensions** list and then click the **Change program** button.
4. Click the **Browse** button in the Open With dialog box.
5. Navigate to and select the application you want the files to open in. (Lasergene applications are located by default in **C:\Program Files\DNASTAR\Lasergene 'x'** or **C:\Program Files (x86)\DNASTAR\Lasergene 'x'**, where 'x' represents the version number).
6. Click **Open**.
7. Click **OK** to close the Open With dialog box, and then click **Close**.

Macintosh:

1. Select a file of the type you wish to change in Finder.
2. Go to **File > Get Info**.
3. Click the dropdown arrow in the **Open With** field, and navigate to the application you want the files to open in. (Lasergene applications are located by default in **Applications:DNASTAR:Lasergene 'x'**, where 'x' represents the version number).
4. Click the **Change All** button to make the application you selected the default for all files of this type.

Copyright DNASTAR, Inc. 2022

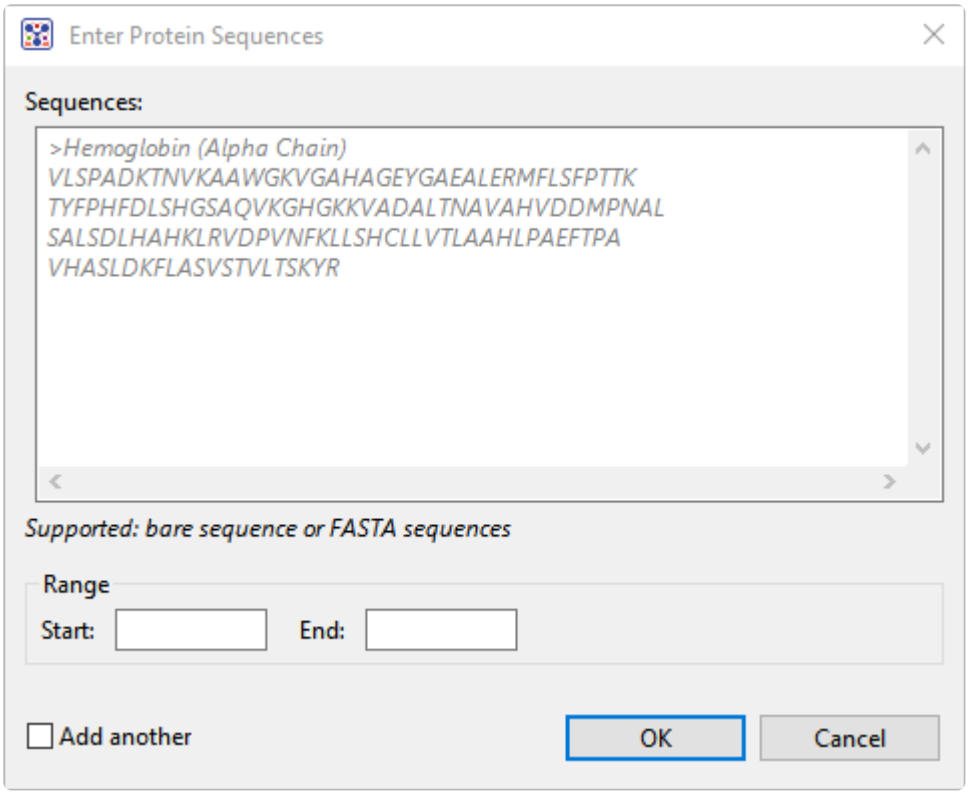
Type or paste a protein sequence

To enter a protein sequence by pasting or typing text into a box:

1. Do either of the following:

    - Choose **File > Enter Sequences**.

    - From the Welcome screen, click the **Open Project** tab and then the link **Enter sequences**.



2. Following the model provided, type or paste a valid amino acid sequence into the **Sequences** field. You can enter or copy in FASTA-formatted sequences or IUPAC protein sequence characters. If a yellow or red error icon appears, hover over it to see a tooltip describing the issue (e.g., “sequence must be in FASTA format,” etc.).

3. (optional) If you want to enter only a portion of the sequence in the **Sequences** box, enter **Start** and **End** coordinates in the boxes provided.

4. (optional) If you want to enter additional sequences, check the **Add another** box.

5. Press **OK**.

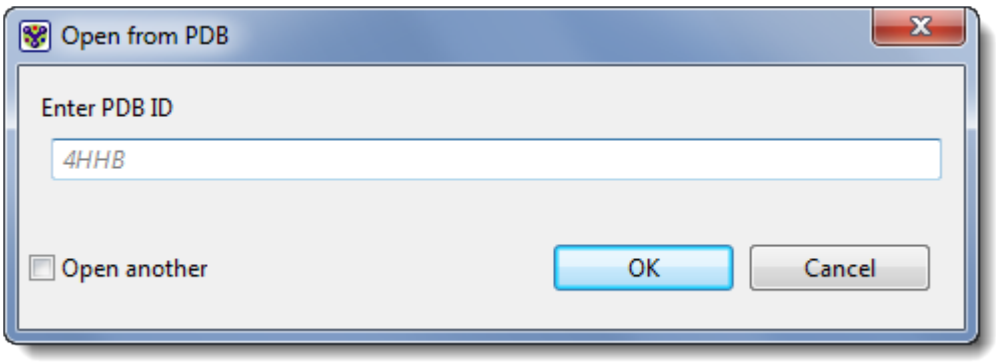
Copyright DNASTAR, Inc. 2022

Open a PDB file when you know the ID

To open a PDB file when you know the ID:

1. Choose **File > Open from PDB** or press **Ctrl/Cmd+D**. For an additional shortcut to this command, see The Welcome Screen.

2. Type a valid 4-digit ID in the text box. All PDB IDs consist of a single-digit number followed by three additional characters. These three characters may consist of numbers and/or letters of the English alphabet.



3. (optional) If you plan to enter additional IDs, check **Open another**.

4. Press **OK**. The structure will open in Protean 3D. If you checked **Open another** in the previous step, the Open from PDB dialog will re-open automatically.

Note: Protean 3D always uses the sequence from the original PDB file. When a PDB file is saved as a *.structure* file, two copies of the original sequence are stored in that file. The first copy, used whenever the file is opened in Protean 3D, cannot be edited. The second copy can be edited with DNASTAR's SeqBuilder Pro, and the edits can be viewed in some additional DNASTAR Lasergene applications.

Copyright DNASTAR, Inc. 2022

Open a PDB file when you don't know the ID

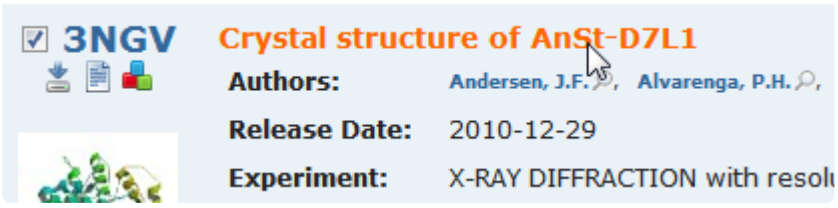
Protein structure files are available for free download via the RCSB Protein Data Bank. If you don't know the 4-digit PDB ID for a structure, you can locate the structure on the PDB website, and then download it to use later in Protean 3D. Alternatively, you can write down its number and import it using Protean 3D's **File > Open from PDB** command, as described here.

1. Go to the Protein Data Bank home page. Locate the search bar near the top of the window.

2. Enter a term (e.g. 'luciferase,' 'feline leukemia,' etc.).

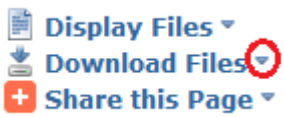
3. In the ensuing results page, refine the query, if desired.

4. From the list of results, click on a hyperlinked structure name or PDB ID to open the page for that structure.



5. To open the structure in Protean 3D, do either of the following:

- Make a notation of the PDB ID, then follow these instructions.
- Download the file to your hard drive and then open it in Protean 3D. To download the file, first locate the **Download Files** menu in the top right corner of the PDB structure page. Click on the triangle to open the menu, and then choose either **PDB File (Text)** or **PDB File (.gz)**.



Copyright DNASTAR, Inc. 2022

For more information and guides to using PDB, see the topic Useful Links.

Note: Protean 3D always uses the sequence from the original PDB file. When a PDB file is saved as a *.structure* file, two copies of the original sequence are stored in that file. The first copy, used whenever the file is opened in Protean 3D, cannot be edited. The second copy can be edited with DNASTAR's SeqBuilder Pro, and the edits can be viewed in some additional DNASTAR Lasergene applications.

Copyright DNASTAR, Inc. 2022

Open a structure from the Motion Library

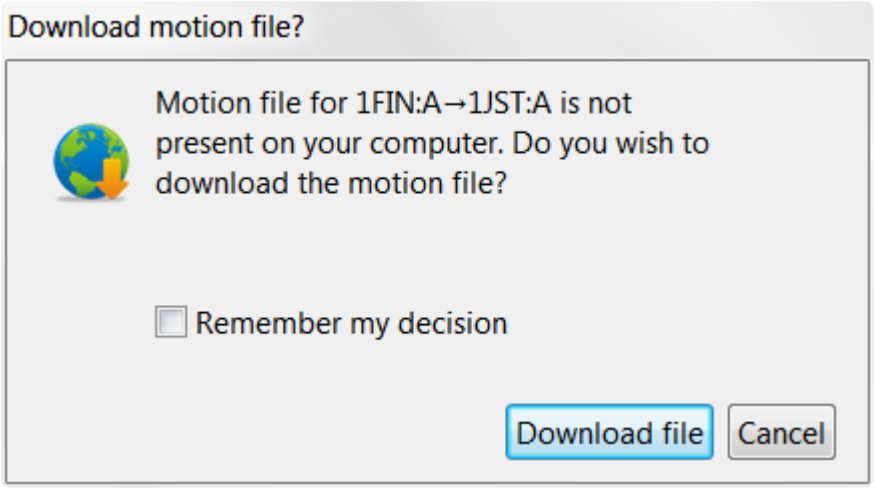
To open a structure from Protean 3D's Motion Library:

1. Open the Motion Library using any of these methods:

- ◦ Using the **File > Open From Motion Library** command
  - ◦ Using the **View > Motion Library** command
  - ◦ Pressing **Ctrl+Alt+L** (Windows) or **Option+Cmd+L** (Macintosh).
  - ◦ For an additional method, see The Welcome Screen.

A progress bar may be visible as the library opens.

2. Find the motion you wish to open, then double-click on its preview image.
3. If you have not downloaded the Motion Library to your hard disc, you will be prompted to download the file you selected:



   a. (optional) If you would like to download individual Motion Library files in the future without being prompted, check the **Remember my decision** box. (Also, see the note below).

   b. Press the **Download file** button.

The motion you selected opens in a new Protean 3D window as a *.pdb.gz* structure. This structure can be optimized as desired and saved as a Protean 3D *.structure* file.

Note: To download the Motion Library to your hard disk, thereby eliminating all future

download prompts, use the **Help > Download Motion Library** command. Protean 3D functions are not available during the Motion Library download. You will see a progress bar during the procedure, and will be notified when the download has concluded.

Copyright DNASTAR, Inc. 2022

Open an existing Protean 3D project

To open an existing Protean 3D project (*.msa*):

Use the **File > Open** command, select the **Open a file** tool ( ), or press **Ctrl/Cmd+O**. For an additional shortcut to this command, see The Welcome Screen.

Once a project is open in Protean 3D, you can open additional files using the same command.

If you moved or renamed any of the sequences prior to reopening a project, you may be prompted to supply the new name(s) or navigate to the location(s) of those sequences.

Note about viewing a project in an earlier version of Protean 3D:

If one user creates a Protean 3D project using Lasergene version 10.1 (or higher), then sends the results to a second user who has Lasergene version 10.0, the second user will be able to open the multiple-file project, and the information for all the structures will appear in the Explorer panel. However, only the first structure will be displayed in the Structure view, and the Analysis view will be disabled. If the second user saves the project in Lasergene version 10.0, all information regarding everything other than the first structure will be discarded. The only way the second user can access the full functionality of a multiple-file document is to upgrade to Lasergene version 10.1 (or higher).

Copyright DNASTAR, Inc. 2022

Search for Sequences Online

This application lets you search National Center for Biotechnology Information (NCBI) databases for matches to text or sequence that you specify.

- Query part or all of a nucleotide or protein sequence against one of NCBI's sequence databases using the Basic Alignment Search Tool (BLAST).
- Query text or accession numbers against entries NCBI's Entrez database.

Once a search has been initiated, you can use the Jobs panel to monitor the search, stop an in-progress search, or open search results in the Text and Table views.

Tools available in the Search wizards:

Some or all of the following optional tools may be available in the bottom left corner of the Search wizards:

1. To display the help for this topic, press the question mark icon ( ).
2. To enter an NCBI Access ID, press the **Set Access ID** ( ) tool. This causes the Access ID dialog to pop up. See Access ID dialog for information.
3. To submit an idea for a new feature, press the **Suggest a feature** ( ) tool.
4. To view information about your service plan, press the **Show Service Plan** ( ) tool. For sequence or text search services, the ensuing dialog describes the state of your service plan and any warning or error text messages. The license control has three states: **Active**, **Expiring**, and **Expired**.

Copyright DNASTAR, Inc. 2022

Search for sequence matches (BLAST)

The BLAST Search wizard is used to search a query sequence against one of NCBI's databases. The sequence can be part or all of an active sequence, another sequence file on your computer, or typed/pasted-in text. The wizard consists of four consecutive screens, each described in a separate topic:

- Query screen
- Search screen
- Options screen
- Job screen

To launch the BLAST Search wizard, use one of the following commands or tools. Some commands let you launch the wizard with the workflow pre-selected and the Search list already pre-populated.

Command, shortcut or tool	Description
Search > Sequence Search	Launches the wizard at the Query screen.
Use the "Change search strategy" tool ( ) from the Text, Table or Pairwise views	
In the Welcome screen, choose **Search** then **Search for sequences**.	
Search > Sequence Search with Selection	Launches the wizard at the Query screen, whose table is pre-populated with the currently-selected portion of sequence.
Press **Ctrl/Cmd+B**	
Search > Sequence Search with All Sequences	Launches the wizard at the Query screen, whose table is pre-populated with all sequences in the document.

Copyright DNASTAR, Inc. 2022

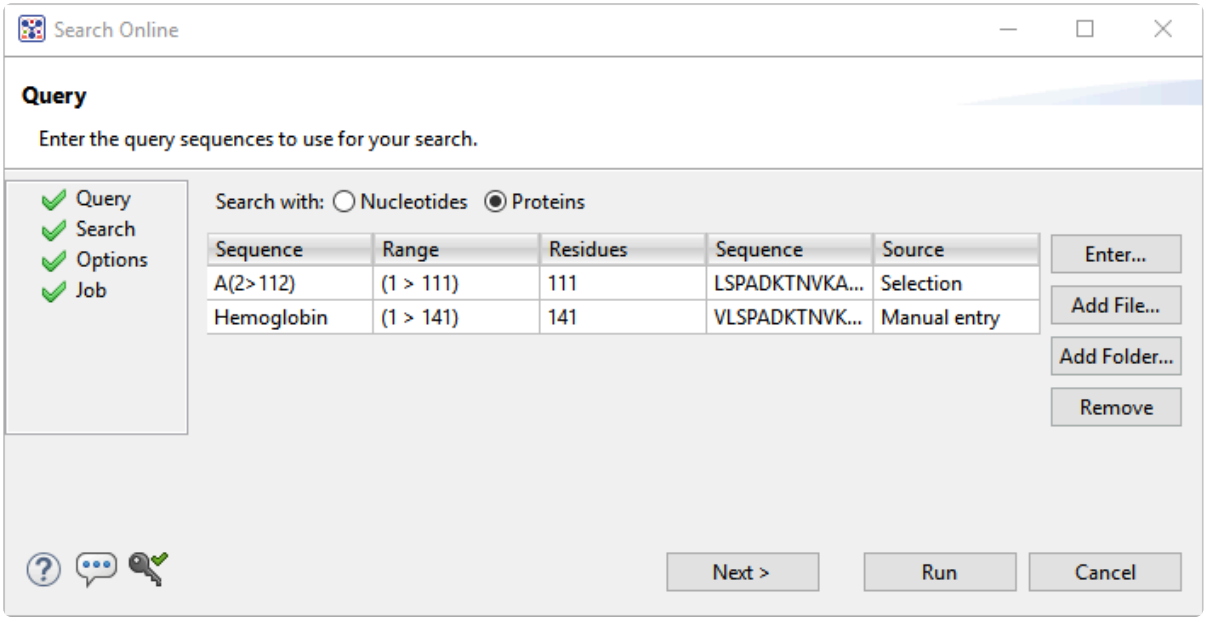
Search wizard: Query

The Query screen is the first of four screens in the BLAST Search wizard. The Query screen is where you add the query sequence, i.e., the sequence for which you wish to locate matches in the NCBI BLAST database.

At the top of the screen, next to “Search with,” select the **Nucleotides** or **Proteins** button.

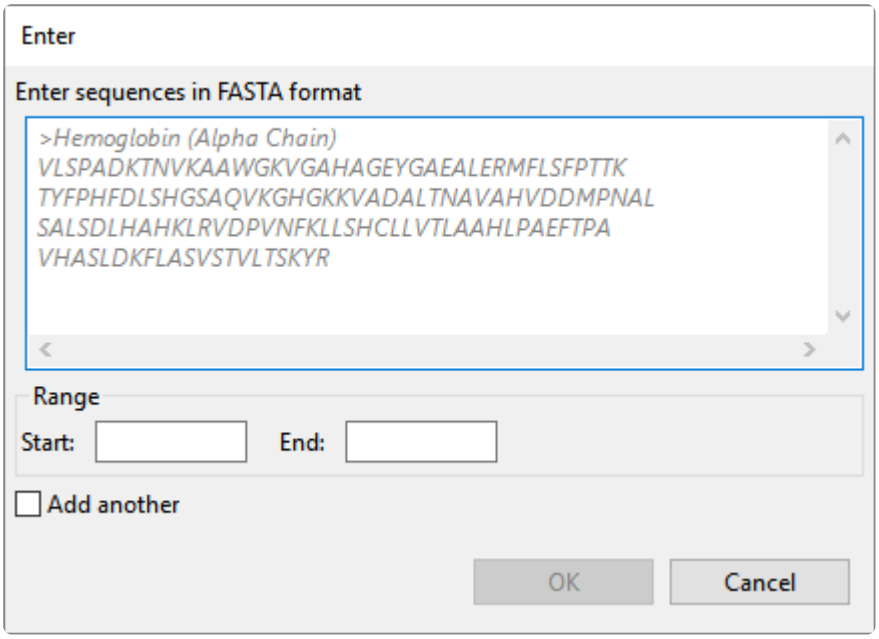
To add one or more query sequence files or a folder of sequences:

Press the **Add** or **Add Folder** buttons. In the file explorer, navigate to and select the desired file(s)/folder(s), and then click **Open**. If you add a multi-file sequence or multiple sequences, each sequence will appear on a separate row in the table.



To paste or type in the query sequence:

Press **Enter**.



In the Enter dialog:

1. Type or paste in the desired sequence following the “dimmed” example in the dialog. A header line is optional.

2. (optional) If you wish to use only a range of this sequence, specify the **Start** and **End** positions.

3. (optional) If you want to add another query sequence after this one, check the **Add another** box. After clicking **OK** to the first sequence, a new Enter dialog will open to add the next query. Otherwise, leave this box unchecked.

4. Press **OK**. If you add additional sequences, each sequence will appear on a separate row in the table.

To remove a file that has already been added:

Select the unwanted file from the list and click **Remove**.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

Once you have set up the query sequence(s) as desired, do either of the following:

- To choose the BLAST program to use and database to search, press **Next** to continue to the Search screen.

- To run the search directly, press **Run**.

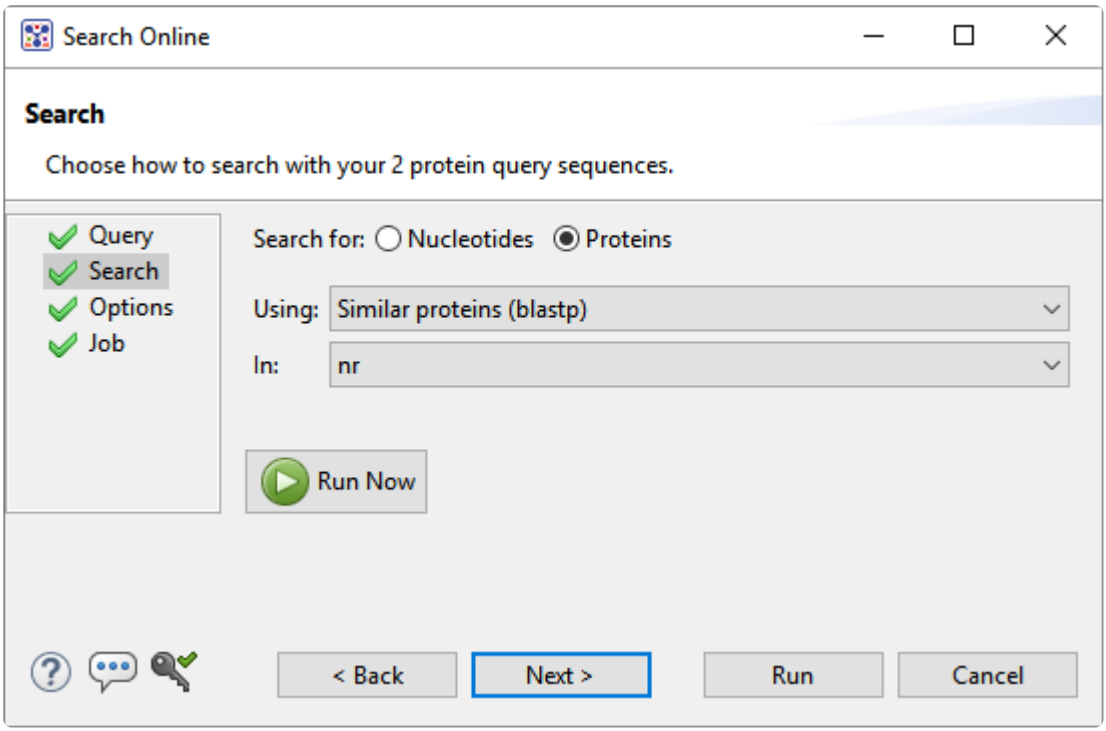
Copyright DNASTAR, Inc. 2022

- To close the Search wizard without starting a search, press **Cancel**.

Copyright DNASTAR, Inc. 2022

Search wizard: Search

The Search screen is the second of four screens in the BLAST Search wizard. The Search screen is used to choose the BLAST program and database.

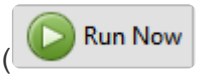


- Specify whether you wish to search for **Nucleotides** or **Proteins**.
- Select the desired BLAST program from the **Using** drop-down menu. For information on the different BLAST programs, see the BLAST Program Selection Guide PDF. Available options depend on the combination of the selections you made in **Search with** (Query screen) and **Search for** (this screen).
    - If you selected **Proteins** on both screens, the **Using** menu contains the unchangeable default of **Similar proteins (blastp)**.
    - If you selected **Nucleotides** on both screens, Choices are **Highly similar sequences (megablast)**, **Similar nucleotides (blastn)**, and **Similar translated nucleotides (tblastx)**.
    - If you selected **Nucleotides** in the Query screen and Proteins in the Search screen, the **Using** menu contains the unchangeable default of **Similar proteins (blastx)**.
    - If you selected **Proteins** in the Query screen and **Nucleotides** in the Search screen, the **Using** menu contains the unchangeable default of **Similar translated nucleotides (tblastn).**
- Select the desired BLAST database from the **In** drop-down menu. Available choices depend on whether you selected **Nucleotides** or **Proteins** in this screen.

Copyright DNASTAR, Inc. 2022

To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

Once you have made the desired selections, do any of the following:

- To go back to the Query screen, press **Back**.
- To add additional parameters to the search, press **Next** to continue to the Options screen.
- To start the search immediately, press the **Run Now** button (

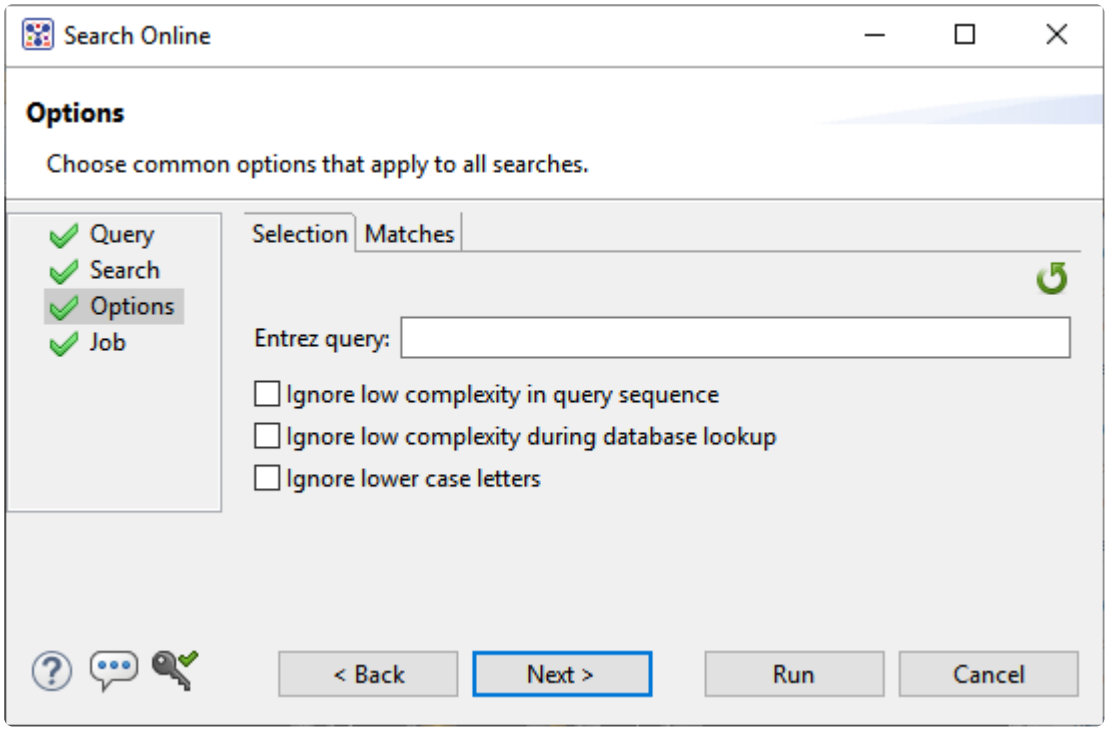
) or **Run** button. The Jobs panel in the application will open and contain information on the status of the search. During the search, a progress dialog is displayed where you can cancel the search using the **Stop** button, hide the dialog using the **Hide** button, or set up another query using the **Settings...** button. Use the Jobs panel to monitor an online search, stop an in-progress search, or open search results in the Text, Table and Pairwise views.
- To close the Search wizard without starting a search, press **Cancel**.

Copyright DNASTAR, Inc. 2022

Search wizard: Options

The Options screen is the third of four screens in the BLAST Search wizard, and has two tabs: **Selection** and **Matches**. These tabs allow you to customize miscellaneous parameters relating to the BLAST search.

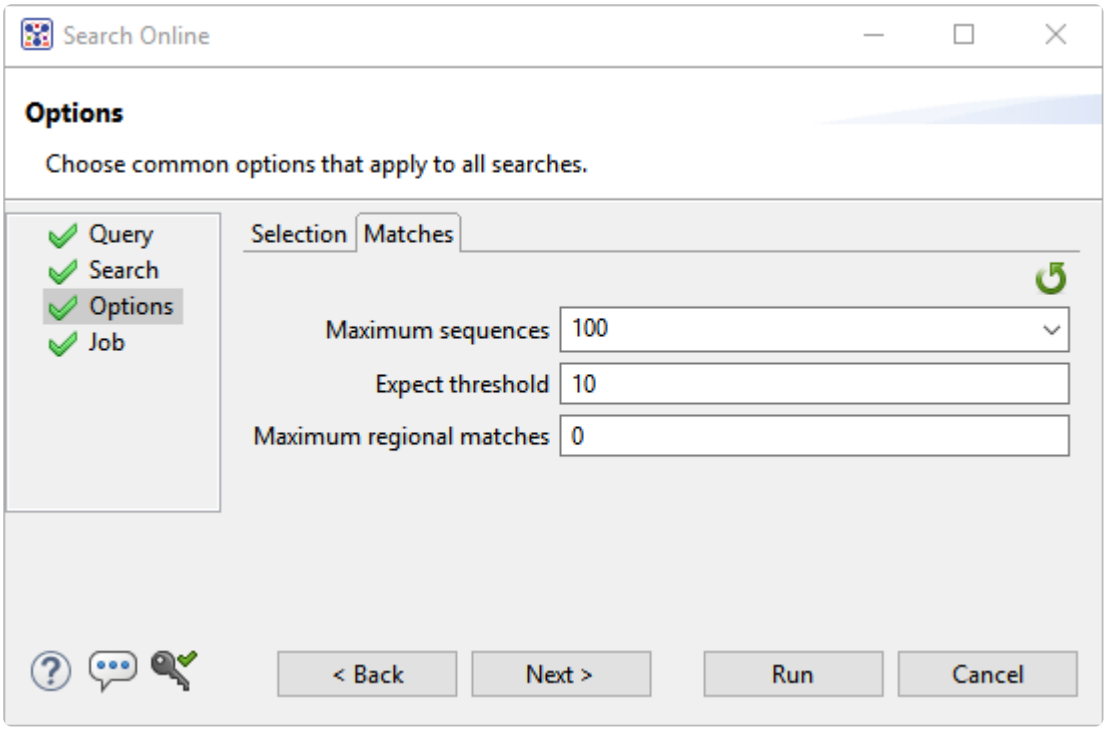
- Click the **Selection** tab to make it active.



In the **Selection** tab:

  - To filter the database you're searching using a text term, enter a text string in the **Entrez** query field.
  - To filter low-complexity regions in the query, check the box next to **Ignore low complexity in query sequence**. The **blastn** search program will use **DUST** for filtering; all other programs will use **SEG**.
  - To mask the query while producing seeds used to scan database (but not for extensions), check the box by **Ignore low complexity during database lookup**.
  - To mask any letters that were lower-case in the FASTA input, check the box next to **Ignore lower case letters**.

- Click the **Matches** tab to make it active.

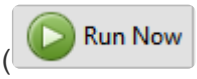


In the **Matches** tab:

- Enter a number next to **Maximum sequences**, the maximum number of aligned sequences to display. The actual number of alignments may be greater than this number. The default is **100**.
- Enter a number beside **Expect threshold**, the expected number of chance matches in a random model. The default is **10**.
- Enter a number next to **Maximum regional matches** to limit the number of matches to a query range. The default is **0**.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

Once you have made the desired selections, do any of the following:

- To continue filtering, press **Next** to continue to the Job screen.
- To start the search immediately, press the **Run Now** button (

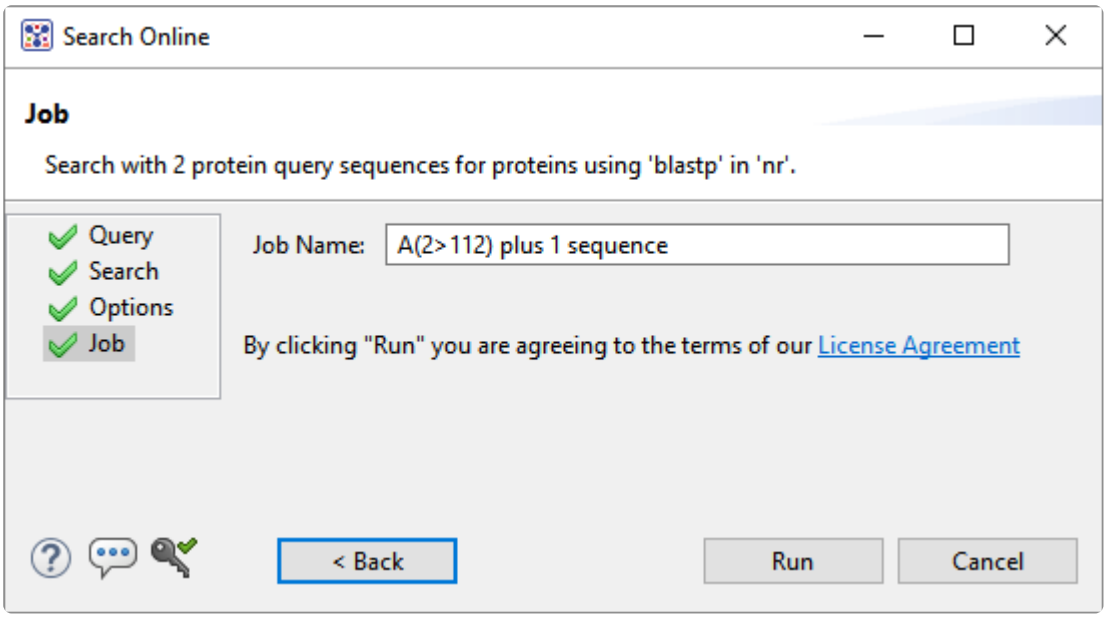
) or **Run** button. The Jobs panel in the application will open and contain information on the status of the search. During the search, a progress dialog is displayed where you can cancel the search using the **Stop** button, hide the dialog using the **Hide** button, or set up another query using the **Settings...** button. Use the Jobs panel to monitor an online search, stop an in-progress search, or open search results in the Text, Table and Pairwise views.

- To close the Search wizard without starting a search, press **Cancel**.

Copyright DNASTAR, Inc. 2022

Search wizard: Job

The Job screen is the fourth of four screens in the BLAST Search wizard, This screen allows you to overwrite the default search name with a custom name and initiate the BLAST search.



The text box contains an automatically-generated name for the search job. If you wish, you can overwrite this default name with a custom name.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

After entering a custom job name (or keeping the default name), do either of the following:

- To start the search , press the **Run** button. The Jobs panel in the application will open and contain information on the status of the search. During the search, a progress dialog is displayed where you can cancel the search using the **Stop** button, hide the dialog using the **Hide** button, or set up another query using the **Settings…** button. Use the Jobs panel to monitor an online search, stop an in-progress search, or open search results in the Text, Table and Pairwise views.

- To close the Search wizard without starting a search, press **Cancel**.

Copyright DNASTAR, Inc. 2022

Search for text or accession number matches (Entrez)

The Entrez Search wizard is used to search a text query or accession number against NCBI's Entrez databases for nucleotides or proteins.

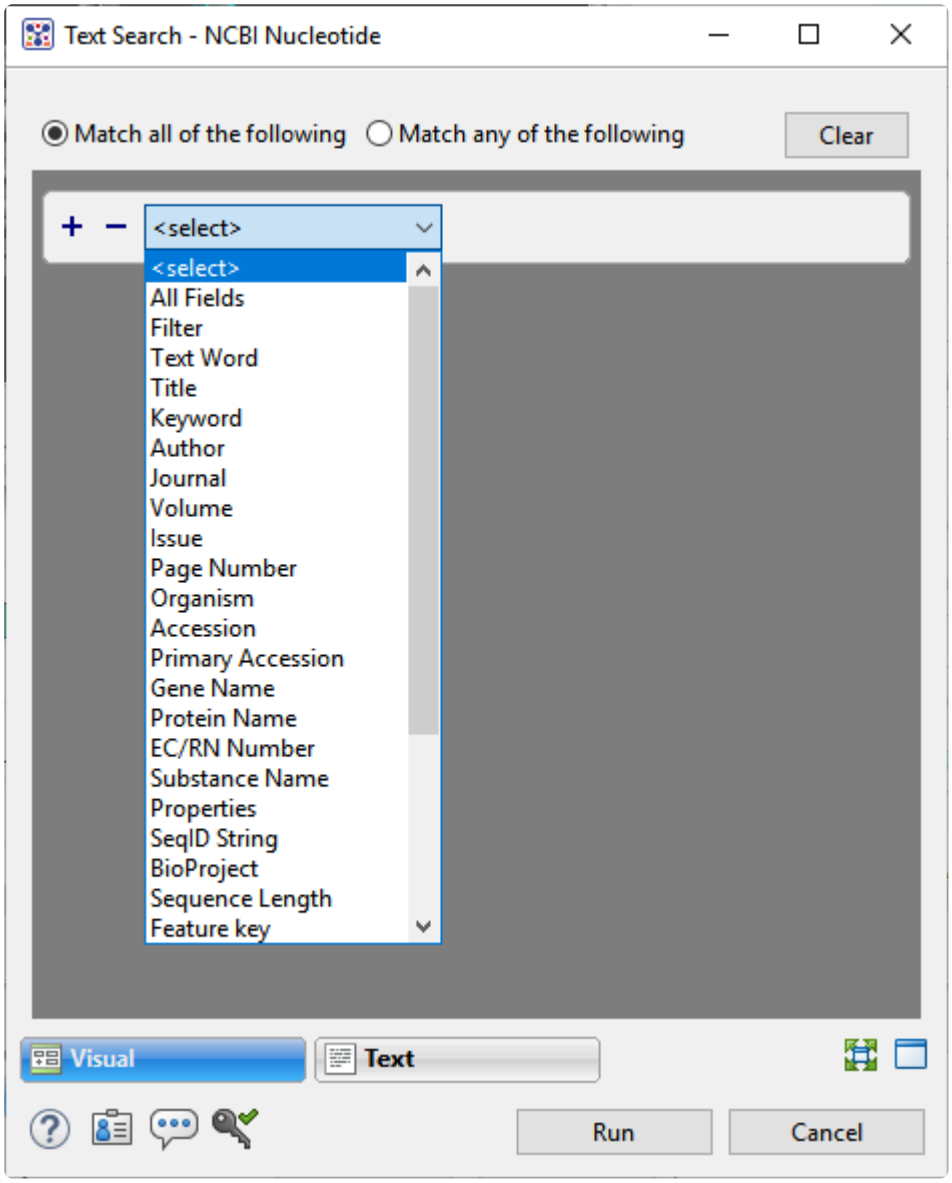
To launch the Entrez search dialog, choose **Search > Text Search in > NCBI Nucleotide** or **Search > Text Search in > NCBI Protein**. For shortcut to these commands, see The Welcome screen.

The **Visual** view is displayed by default and is described below.

> **Note:** The Visual view is only for non-nested queries using the same type of operator (AND, OR). To perform a more complicated search, press the **Text** tab near the bottom left corner and create the search in the Text view.

1. Use the **<select>** drop-down menu to choose a criterion. The field headings vary depending upon the online database that you are using. Choosing **All fields** will cause the application to search every word of every listing for the query term in the database.

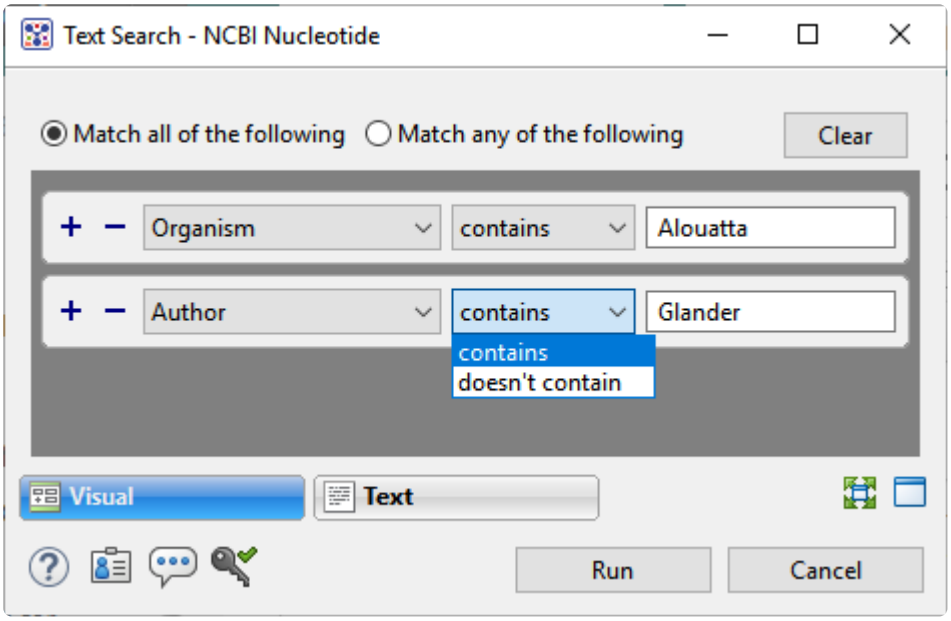
Copyright DNASTAR, Inc. 2022



Once you make a selection, a second drop-down menu and an empty text box appear to the right.

- In the second drop-down menu on the row, choose **contains** or **doesn't contain** (see image below).
- In the text box, type the text that should be found in any results (in the case of **contains**) or should not be found in any results (in the case of **doesn't contain**).

2. To add another criterion, press the plus icon (+). To delete a criterion, press the minus icon (−).



3. Once you have set up all desired criteria, go to the top of the dialog and choose whether to **Match all of the following** or **Match any of the following** criteria.

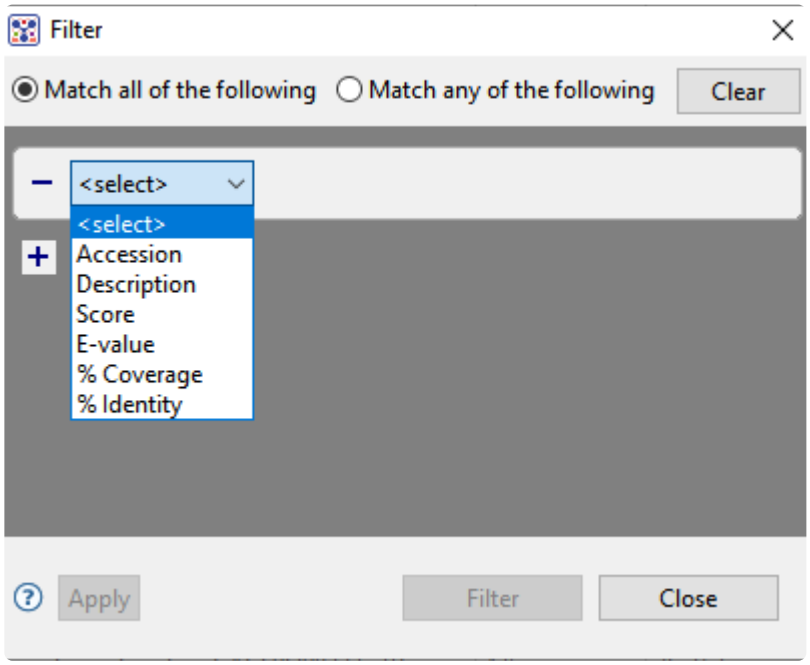
4. To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

5. Click **Run** to send the query to the database. To instead exit the dialog without searching, press **Cancel**.

During the search, a progress dialog is displayed. Once the search is completed, any matches are displayed in the Jobs panel.

Filter sequence search results

When you press the **Filter results** tool ( ) in the header above the Table or Pairwise views, there are two options: **Filter** and **Clear Filter**. Clicking on the **Filter** option opens the following dialog:



- Use the **<select>** drop-down menu to choose a filter type, then enter the desired options.

Filter type	Description	Options
Accession	The Accession number.	
Description	The NCBI “Definition” field text.	
Score	The similarity score for the match. (See the NCBI definition for **raw score**). In general, a higher score denotes a better match.	
E-value	The chance that the match is expected to occur coincidentally. (See the NCBI definition for **E value**). In general, a lower score denotes a better match.	
% Coverage	The percentage of the query sequence that aligns to a match in the BLAST database. Example: Query has 596 bases, of which 565 align to a match in the NCBI database. **%Coverage** = 565/596 = 95%.	
% Identity	The percentage of bases that match when the query and result sequence are aligned. (NCBI definition). Example: Query has 565 bases that align to the match and 3 bases that don’t	

Copyright DNASTAR, Inc. 2022

	match. **%Identity** = 562/565 = 99.5%.	

- To add another filter, press the plus icon (+). To delete a filter row, press the minus icon (−).
- Once you have set up all desired queries, go to the top of the dialog and choose whether to **Match all of the following** or **Match any of the following** queries.

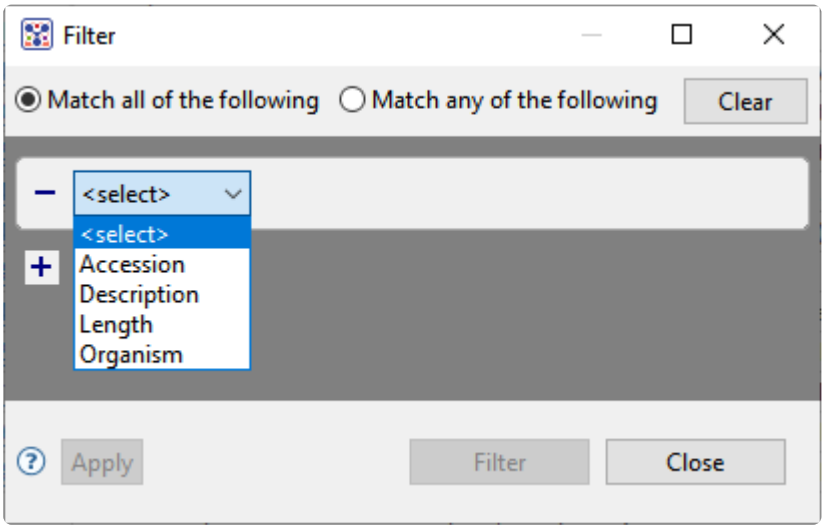
Once you are finished setting up filters, do any of the following:

- To apply the filters, press **Apply** or **Filter**.
- To clear all filters and start over, press **Clear**.
- To exit the dialog without filtering, press **Close**.

Copyright DNASTAR, Inc. 2022

Filter text search results

When you press the **Filter results** tool ( ) in the header above the Table or Pairwise views, there are two options: **Filter** and **Clear Filter**. Clicking on the **Filter** option opens the following dialog:



- Use the **<select>** drop-down menu to choose a filter type, then enter the desired options.

Filter type	Description	Options
Accession	The Accession number.	
Description	The NCBI “Definition” field text.	
Length	The length of the match sequence in bases.	
Organism	The species name.	

- To add another filter, press the plus icon ( ). To delete a filter row, press the minus icon ( ).
- Once you have set up all desired queries, go to the top of the dialog and choose whether to **Match all of the following** or **Match any of the following** queries.

Once you are finished setting up filters, do any of the following:

- To apply the filters, press **Apply** or **Filter**.
- To clear all filters and start over, press **Clear**.
- To exit the dialog without filtering, press **Close**.

Copyright DNASTAR, Inc. 2022

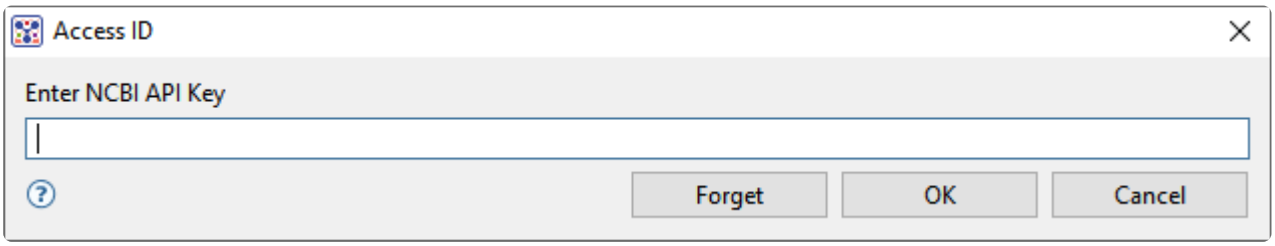
Access ID dialog

As of December 2018, NCBI requires an API key for faster access to NCBI services, or to run more searches concurrently. To get a key, start by setting up an NCBI account. After registering your account, click on your user name and request your key using **API Key Management**. Once you have your key, you can enter it in Lasergene applications to enable NCBI's premium search access.

To add your key to a Lasergene application:

1. Begin an Entrez database search using **Search > Text Search in > NCBI Nucleotide** or **Search > Text Search in > NCBI Protein**.

2. Press the **Set Access ID** ( ) tool in the bottom left corner of the dialog; or, from the Table view, press the **Export data** tool and choose **Set Access ID**. In either case, the Access ID pop up appears.



- If an NCBI API key is not already present in the text box, enter your key and press **OK**.

- If an NCBI API key is present but you no longer use it, press **Forget** to remove it from the application's memory. Then enter the desired key and press **OK**.

- To exit without saving your changes, press **Cancel**.

Copyright DNASTAR, Inc. 2022

Align Multiple Structures

Homologous proteins often have similar structures. A structure alignment attempts to identify these similarities by comparing the shape and 3D conformation between two structures. Unlike simple superposition methods, structure alignments do not require prior knowledge of equivalent residues between structures.

The results of a structure alignment are viewed in the Structure view as a set of superimposed structures, where each alignment minimizes the root mean squared deviation (RMSD), or the average distance between atoms in superimposed structures. The RMSD values are also reported in the Details panel for a group of structures. Rigid-body alignment is used to determines the best alignment without allowing the structures to bend or twist.

The following video shows a brief walkthrough of the structural alignment workflow:

To align multiple structures:

1. Open a document containing two or more structures.

   Note: Since Protean 3D treats all units in biological assemblies as individual structures, you may also align such units. However, this workflow is not generally useful.

2. (optional) If you wish to align specific regions, select the desired regions. Each selection must contain more than 8 residues with atomic coordinates.

> **Note:** If aligning units of a biological assembly, you must specify the chains or regions that you wish to align. If nothing is selected, all components of the biological assembly will automatically be aligned to the first unit.

3. For whole structures, choose between **Structure > Align Structures > Rigid-body Alignment (TM-Align)** and **Structure > Align Structures > Rigid-body Alignment (jFATCAT)**.

   If you have selected specific regions to align, choose between **Structure > Align Structures > Rigid-body Alignment for Selected Regions (TM-Align)** and **Structure > Align Structures > Rigid-body Alignment for Selected Regions (jFATCAT)**.

   Protean 3D will then calculate a series of pairwise structure alignments using the first structure in the document as the alignment base.

If the **Structure > Align Structures** command is disabled, the usual sub-command is replaced by one of the following:

Sub-command	Correction procedure:
No Document Open	Follow step 1, above.
Requires Two or More Protein Structures	Add one or more structures to the existing structure using **File > Add to Document**.
Requires Selections in Two or More Protein Structures (Each > 8 Residues)	You cannot align multiple regions from the same structure. You need to select regions in two or more structures, as per step 2, above.

References for Protean 3D's structure alignment algorithms include Prlic et al. (2010), Ye et al. (2003) and Zhang and Skolnick (2005). The last of these pertains specifically to the TM-Align method.

Copyright DNASTAR, Inc. 2022

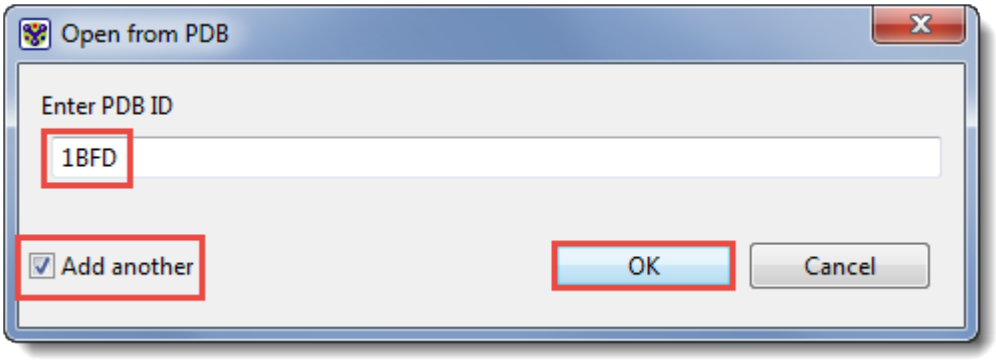
Try it! – Align multiple structures using jFATCAT

This tutorial demonstrates how to align two structurally homologous protein files within a single document. This technique is useful for discovering similarities and differences between related structures. The files used in this tutorial will be opened from the RCSB Protein Data Bank (PDB) using the Protean 3D interface:

- *1BFD* – The crystal structure of benzoylformate decarboxylase
- *1PVD* – The crystal structure of pyruvate decarboxylase

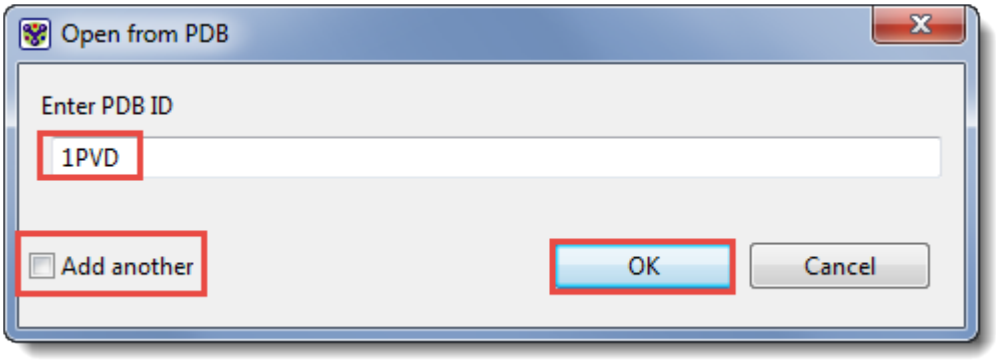
These two proteins exhibit structural homology, although they have only a 19% sequence identity. Their structural homology may indicate either distant sequence homology or convergent evolution.

1. Choose **File > Open Group** to launch the Open Group dialog.
2. Keep the default setting of **Center on-screen to align similar structures**.
3. Click the **Add PDB ID** button.
4. Type **1BFD** into the text box. You may initially see a yellow warning icon (⚠) prompting you to type in a valid ID. Check the box next to **Add another** and press **OK**.

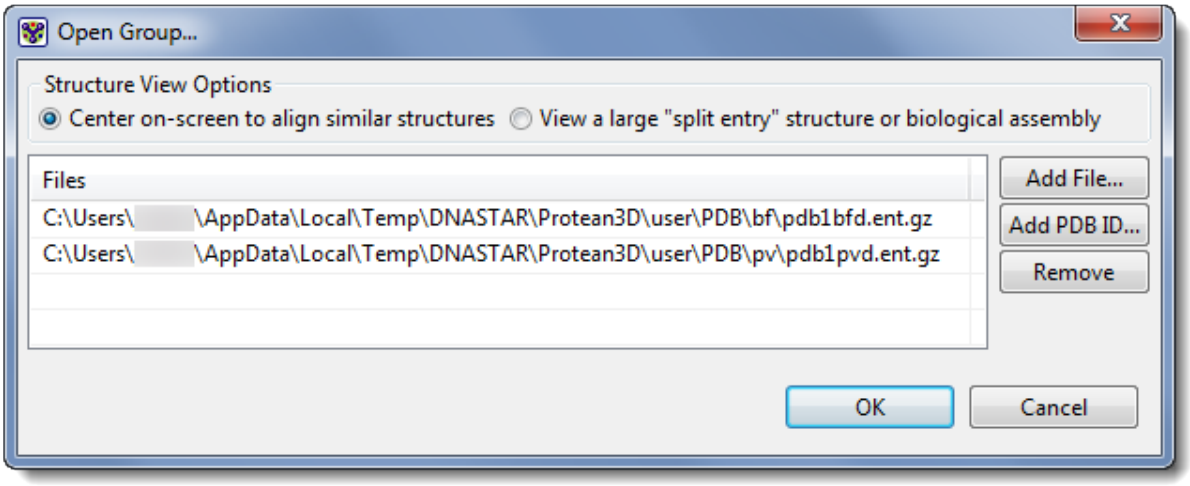


The first structure is added to the Open Group file list.

5. When the Open from PDB dialog re-opens automatically, type **1PVD** into the text box. Uncheck the box next to **Add another** and press **OK**.



The second structure is added to the Open Group file list. (Note: Blurred areas below denote where the user name appears).

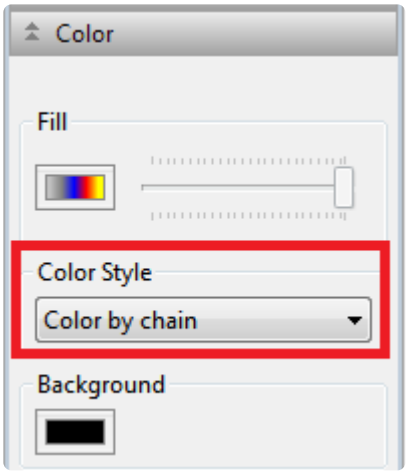


6. Click the **OK** button.

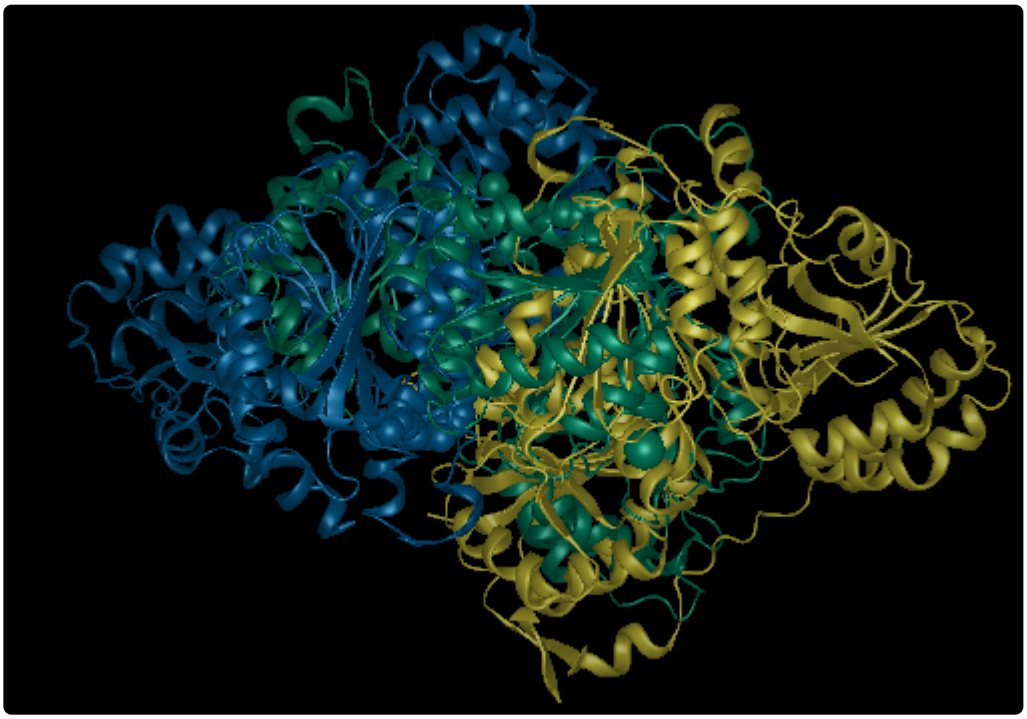
   Now that both structures are open in a single document, we will color them by chain to facilitate their comparison.

7. Double-click on the black background of the Structure view to remove any existing selection.

8. Choose **View > Style > Color** to open the Color bar. From the **Color Style** drop-down menu, select **Color by chain**.



The *1BFD* structure now consists of a single green chain, while the *1PVD* structure consists of two chains: one blue and the other yellow. Since we have not yet aligned the structures, the three chains overlap randomly.

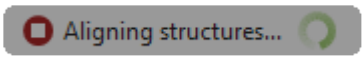


9. To select the chains you wish to align, open the Molecules section using **View > Explorer > Molecules**. Hold down the **Ctrl** (Win) or **Cmd** (Mac) button and click once on the name of the "A" chain from *1BFD* and again on the name of the "B" chain from *1PVD*.

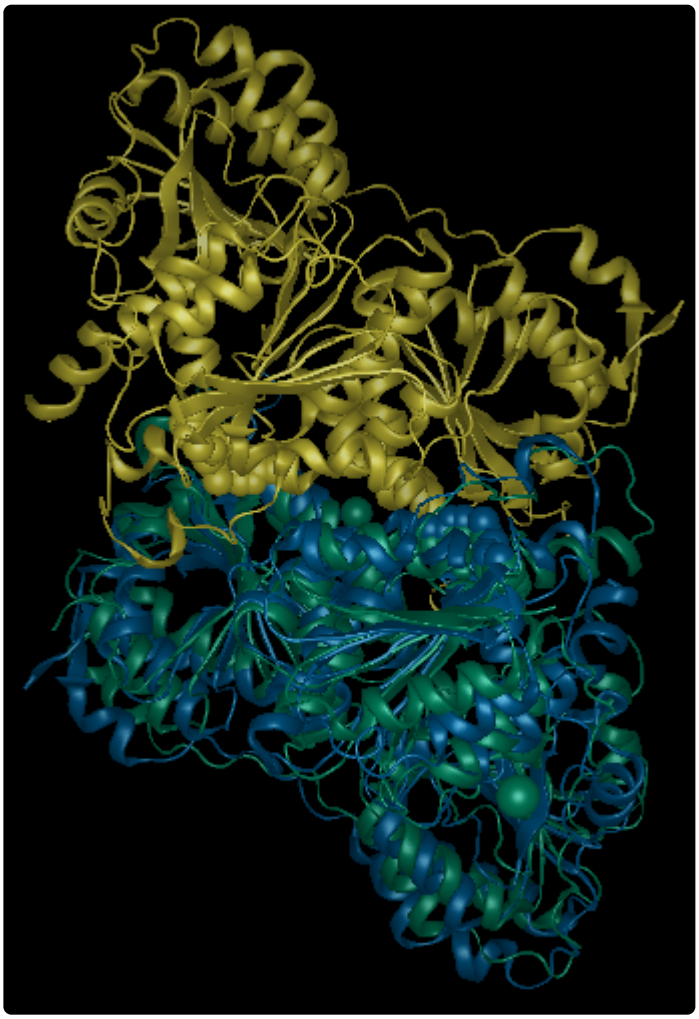
> p(banner tip). **Note:** Checking the boxes in the Molecules section only toggles between displaying or hiding the chains in the views. To select an item, you must click on its name.

10. To align the selected chains, choose **Structure > Align Structures** > **Rigid-body Alignment for**

Selected Regions (jFATCAT). During alignment, a status bar appears in the lower left corner of the Structure View. The status bar includes a red button that could be used to stop the alignment, if desired in the future.

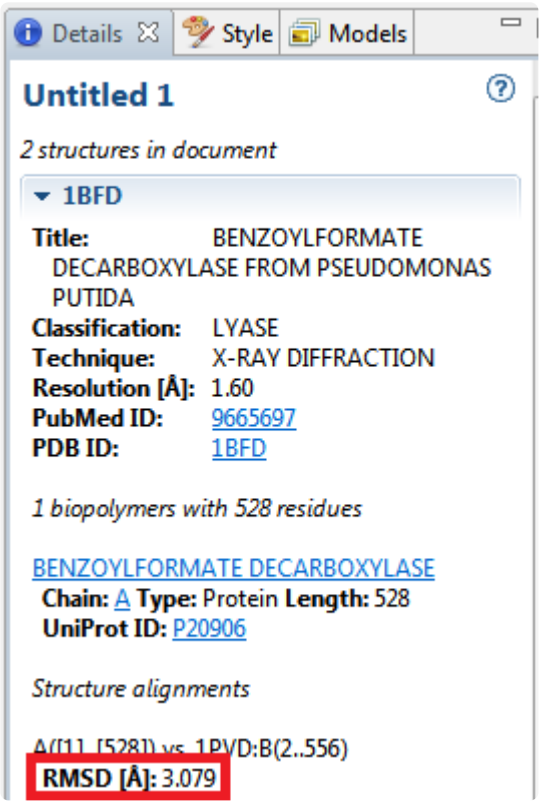


11. When alignment is complete, double-click on the black background to clear the selection. Observe that chain A of protein *1BFD* (green) is now aligned to chain B of protein *1PVD* (blue).



12. Open the Details panel using the **View > Details** command, and look for the Root Mean Square Deviation (RMSD). Note that the RMSD value for the two proteins is 3.079 Å.

Copyright DNASTAR, Inc. 2022



RMSD quantifies the degree of similarity by describing the average distance, in Angstroms (Å), between the atoms of two superimposed structures. The RMSD value is inversely proportional to the degree of similarity between the structures.

As a very general rule of thumb:

- RMSD ≤ 0.5 Å indicates near-identical structural similarity.
- RMSD between 0.5-2.0 Å corresponds to a high degree of similarity.
- RMSD between 2.0-5.0 Å indicates a lower, but still noticeable, degree of similarity.

The RMSD in this case is 3.079, which is compatible with our knowledge that the two sequences have a 19% sequence identity.

Copyright DNASTAR, Inc. 2022

Make a Selection

Selections can be made from various views, panels and panel sections. Selections are synchronized between all views.

To Select…	Location	Method
A residue	Structure View	Double-click on one of the atoms within the desired residue. Double-click on the background to de-select.
	Sequence view	Click on the residue. To move between residues, use the keyboard arrow keys.
	Analysis view	Click on the vertical space that aligns with the residue. (Do not click directly on a decision region.)
Contiguous residues	Sequence view	Hold down the mouse button and drag the mouse or use **Shift+click**.
	Analysis view	Click on the vertical space that aligns with a residue. Use **Shift+click** to select contiguous residues.
Non-contiguous residues	Structure view	Use **Ctrl+double-click** (Win) or **Cmd+double-click** (Mac).
	Sequence view	Use **Ctrl+click** (Win) or **Cmd+click** (Mac).
	Analysis view	Click on the vertical space that aligns with a residue. Use **Ctrl+click** (Win) or **Cmd+click** (Mac) to select additional, non-contiguous residues.
All residues in a binding site	Sequence view	Click on a site label or hub. To group or ungroup residues, use the **Residue Grouping** menu in the Sequence section.
Both members of a disulfide bond	Sequence view	Click on a yellow sphere associated with a disulfide bond.
A chain	Sequence view	Double-click on the chain.
	Molecules section	Click on the name of the chain. It is possible to select hidden (unchecked) chains.
A ligand	Sequence view	Click on the ligand.
	Structure view	Double-click on an atom of the ligand.

Copyright DNASTAR, Inc. 2022

Contiguous structural elements	Molecules section	**Shift+click** on the names of the structural elements.
Non-contiguous chains	Molecules section	**Ctrl+click** (Win) or **Cmd+click** (Mac) on the chain name.
Non-contiguous ligands	Structure view	**Crtl+double-click** (Win) or **Cmd+double-click** (Mac) on each part of the ligand.
A feature	Sequence view	Click on the corresponding feature icon to select a helix, sheet, site or disulfide.
	Features section	Click on the name of the feature. It is possible to select features that are hidden (unchecked).
Contiguous features	Features section	**Shift+click** on the names of the features.
Non-contiguous features	Sequence view	**Ctrl+click** (Win) or **Cmd+click** (Mac) on the corresponding feature icons to select helices, sheets, sites or disulfides.
	Features section	**Ctrl+click** (Win) or **Cmd+click** (Mac) on the names of the features.
A region of a decision plot	Analysis view	Click on a region to select it. **Shift+click**, **Ctrl+click** (Win) or **Cmd+click** (Mac) to select multiple regions.
A single synopsis or an individual channel (row)	Analysis view	Double-click on the synopsis or channel.

Note: When a selection is made in the Molecules section, the Sequence view will jump to that selection.

- If a single item (chain, ligand or solvent) is selected, the Sequence view will scroll to the beginning of the chain or the ligand row where the ligand is displayed.

- If multiple items are selected, the Sequence view will jump to the beginning of the first item selected. This is true even if the selected item is hidden in the Sequence View. In that case, the sequence will scroll to the position where the selected item would be displayed if not hidden.

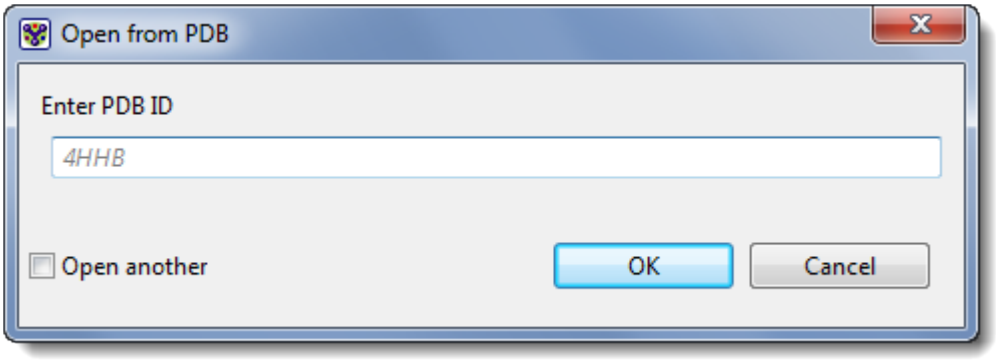
Copyright DNASTAR, Inc. 2022

Try it! – Make a selection from the Sequence view

In this tutorial, you will learn how to make a selection from the Structure and Sequence views. You will also confirm that the same selection appears in other views.

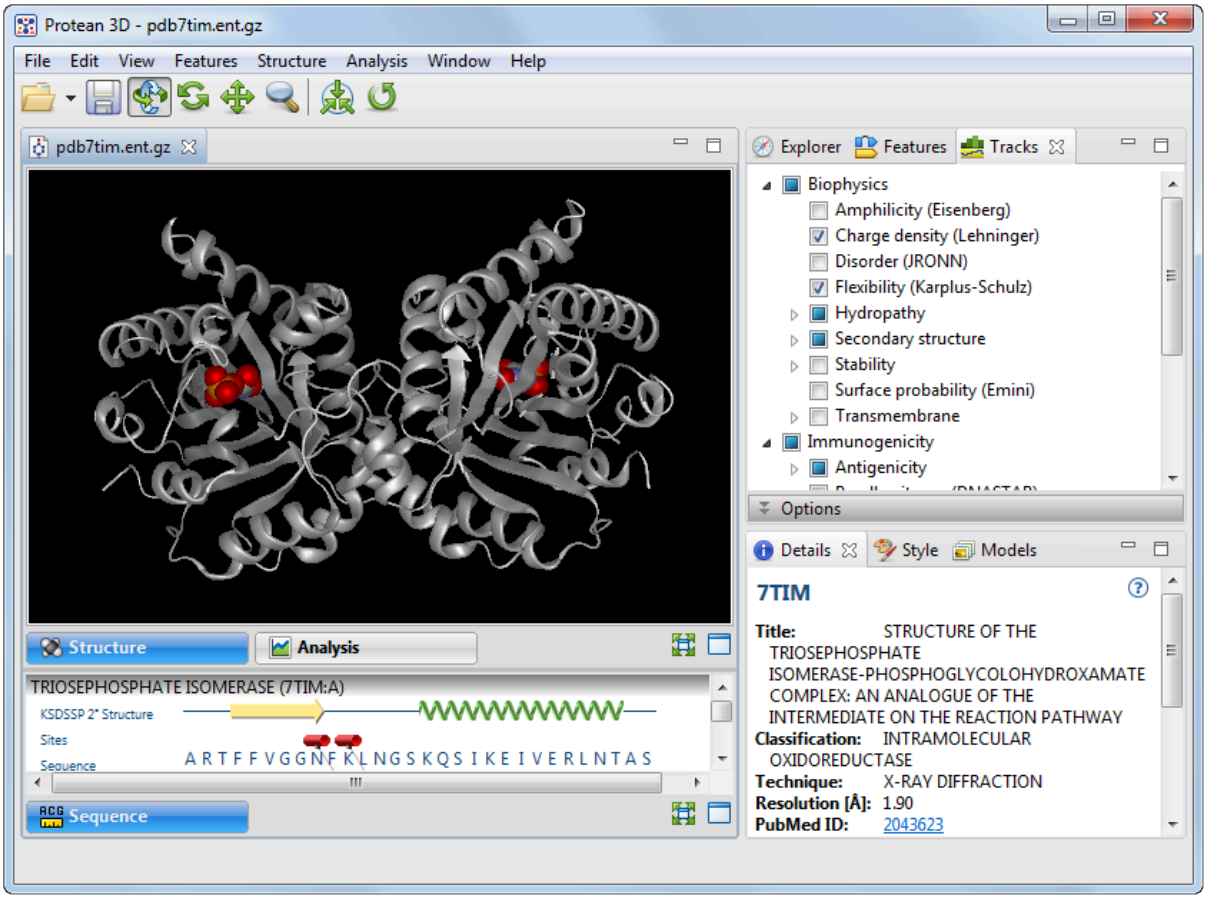
1. Within Protean 3D, choose **File > Open from PDB**. You may initially see a yellow warning icon ( ) prompting you to type in a valid ID.

2. Type **7TIM** into the text box.

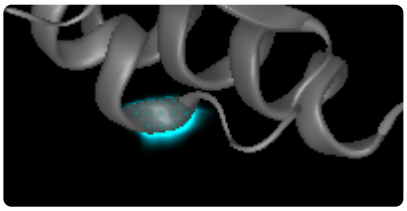


3. Press **OK** to open the structure in Protean 3D.

Copyright DNASTAR, Inc. 2022

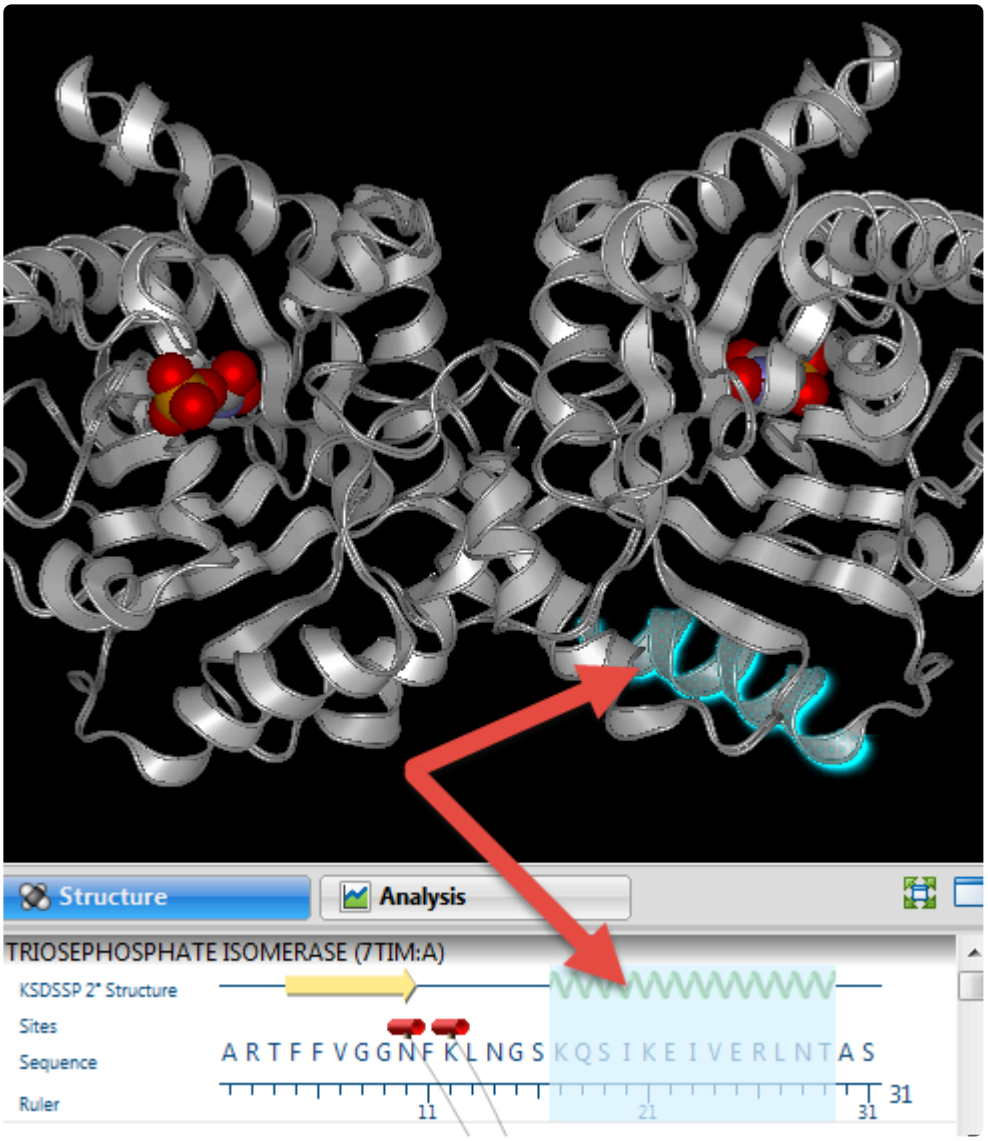


4. In the Structure view, double-click on any residue in the molecule to select it and give it a blue highlight in all views. Deselect by double-clicking on the black background.

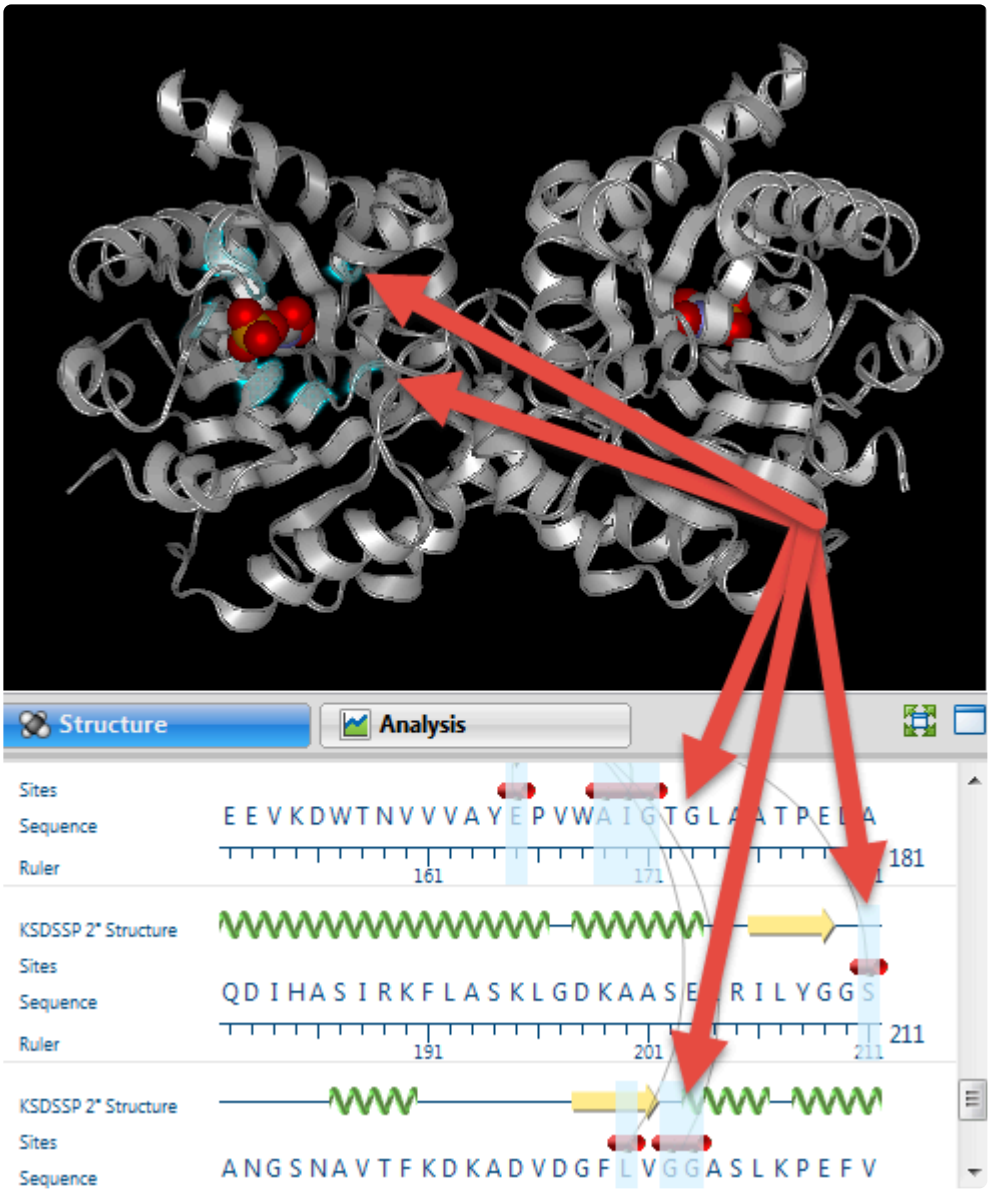


5. In the Sequence view (bottom left of the Protean 3D window), select any helix (represented by a green spiral) by clicking on it. The selected helix will appear highlighted (light blue) in both the Sequence and Structure views. If you don't see a highlighted element in the Structure view, try zooming in on the molecule or rotating it for a better view.

Copyright DNASTAR, Inc. 2022



6. Click on any red tube icon in the Sequence view to simultaneously select all residues involved in a binding site. Note that the corresponding binding site also appears highlighted in the Structure view.


Structure
Analysis
Sites
Sequence
Ruler
161
171
181
KSDSSP 2* Structure
Sites
Sequence
Ruler
191
201
211
KSDSSP 2* Structure
Sites
Sequence

Work with Features

The following video is a brief introduction to working with features in Protean 3D:

For detailed information, see the table below:

Task	How To
To control which features are displayed in the Sequence and Analysis views	See Features panel.
To control the appearance of features in the views	See Options section and Style section.
To create a new feature	Select a portion of the sequence or a structural element and do one of the following: • Select the Features panel's **Create a feature** tool • Use **Features > New Feature** • Press **Ctrl/Cmd+K** The Options section of the Features panel automatically becomes populated with a default name, as well as information about the molecule and endpoints of the feature.
To delete a feature	Select the feature in one of the views or panels and do one of the following: • Select the Features panel's **Delete a feature** tool

Copyright DNASTAR, Inc. 2022

	• Use **Features > Delete Feature**
To choose which annotations to display for a structure, or to reset annotations that were previously edited in the Options section	Use **Features > Reset Secondary Structure** and choose one of the following sub-commands: • **Annotated Assignments** – To display author annotations from a PDB or other structure file. • **Calculated Assignments (KSDSSP)** – To display features calculated automatically from the structure file. The calculated annotations will be saved when you save the structure file.

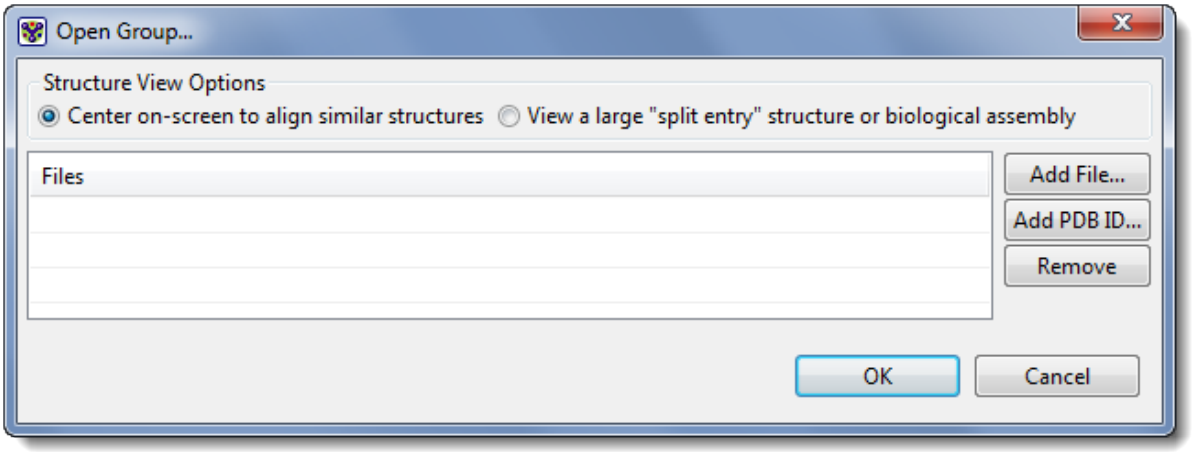
Copyright DNASTAR, Inc. 2022

Overlay One Structure on Another

To overlay a structure on an already-open structure:

1. Open the first structure file.

2. Choose **File > Add to Document** or press **Ctrl/Cmd+E**.



3. In the ensuing dialog, choose the placement of the structure(s):

- ◦ To place the center of one structure directly on top of the center of the other structure, choose **Center on-screen to align similar structures**.

  ◦ To place the structures using the internal coordinates of their structure files, choose **View a large "split entry" structure or biological assembly**.

4. Within the same dialog, add or remove files from the Files area:

- ◦ To open a file chooser and add one or more files to the list, click **Add File**.

  ◦ To choose a file with a known ID from the PDB database, select **Add PDB ID**. The ensuing dialog is the same one used for the menu command **File > Open from PDB**. See Opening a PDB File When You Know the ID for more information.

  ◦ To remove a file from the list, select it and choose **Remove**.

Once the desired files appear in the Files area, click **OK** to open them in Protean 3D, or **Cancel** to exit from the dialog without opening any files.

Copyright DNASTAR, Inc. 2022

Create or Find Variants using Protein Design

Protein Design, accessed through Protean 3D, lets you make one or more changes at the residue-level on a protein structure, and calculates whether these changes are likely to be stabilizing or destabilizing compared to the original structure. Two workflows are available: **Create specific variant** and **Scan for hotspots that provide stability**. Variant calculations are performed for one residue at a time, not for groups of residues as a single unit. Note that Protein Design is not suitable for use in modeling changes involving indels. For indels, we recommend adding or removing amino acids from the sequence and submitting the edited sequence as a query for NovaFold.

The Protein Design wizard consists of five consecutive wizard screens, each described in a separate topic:

- Workflow screen
- Structure screen
- Substitutes screen
- Options screen
- Job screen

To launch the Protein Design wizard, choose one of the following commands. Some commands let you launch the wizard with the workflow pre-selected and the Substitutes table already pre-populated.

Command	Opens Wizard at	Workflow screen selection	Substitutes screen pre-populated with:
File > New Protein Design Prediction	Workflow screen	None	Nothing
Modeling > Protein Design > New Protein Design Prediction	Workflow screen	None	Nothing
Modeling > Protein Design > Create Variant With > Selection	Substitutes screen	**Create specific variant** workflow	Selected sequence range
Structure > Create Variants or right-click on a selected range in the Analysis or Sequence view and choose **Create Variants**			
Modeling > Protein Design > Create Variant With > Document	Substitutes screen	**Create specific variant** workflow	Entire sequence range
Modeling > Protein Design > Scan for Hot Spots With > Selection	Substitutes screen	**Scan for hotspots that provide stability** workflow	Selected sequence range
Modeling > Protein Design > Scan for Hot Spots	Substitutes	**Scan for hotspots**	Entire sequence

Copyright DNASTAR, Inc. 2022

With > Document	screen	**that provide stability** workflow	range

For additional shortcuts, see The Welcome Screen.

Tools available in the Protein Design wizard:

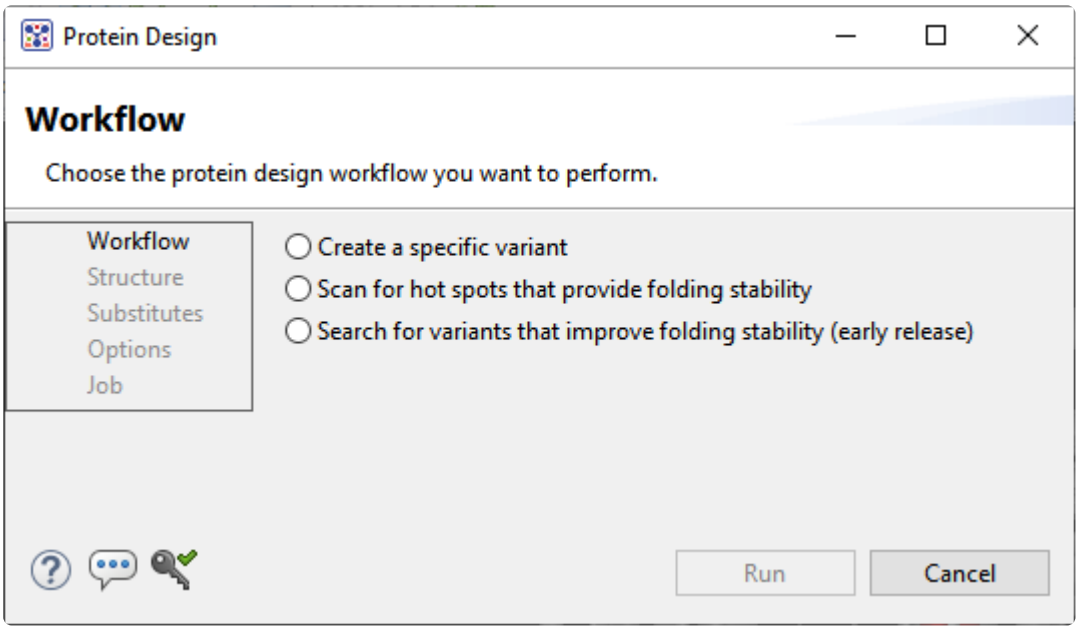
The following tools are available in the bottom left corner of each Protein Design wizard screen:

- To display the help for this topic, press the question mark icon ( ).
- To enter an NCBI Access ID, press the **Set Access ID** ( ) tool. This causes the Access ID dialog to pop up. See Access ID dialog for information.
- To submit an idea for a new feature, press the **Suggest a feature** ( ) tool.
- To view information about your service plan, press the **Show Service Plan** ( ) tool. For sequence or text search services, the ensuing dialog describes the state of your service plan and any warning or error text messages. The license control has three states: **Active**, **Expiring**, and **Expired**.

Copyright DNASTAR, Inc. 2022

Protein Design wizard: Workflow

When using the commands **File > New Protein Design Prediction** or **Modeling > Protein Design > New Protein Design Prediction**, the Protein Design wizard opens at the Workflow screen.



Choose one of the following workflows:

- **Create specific variant** – Allows the user to specify one or more amino acid substitutions in a known structure after which Protein 3D calculate the side chain rotamer conformation at each position with the lowest predicted change in energy. This option is similar to the **Structure > Create Variants** functionality in Protean 3D version 15.3 and only requires that you have a current Protean 3D license. Note that substitution variations are not naturally-occuring SNPs, but rather user defined variants.

- **Scan for hotspots that provide stability** – Search for amino acid positions that are important for protein stability by performing computational alanine (or serine) scanning over a selected region of the protein structure. The technique predicts the contribution of a specific residue to the stability of the protein fold by independently mutating every position to alanine (or serine) and calculating the energetic difference. Hot spots are the subset of residues that contribute largely to the folding energy, thus hot spots are residues whose variants destabilize the structure. Positions where one the substitution is predicted to increase stability are also identified. The hotspot workflow is designed to support PCR site-directed mutagenesis, a workflow of special interest to molecular biologists. Site-directed mutagenesis utilizes both Protean 3D/Protein Design and SeqBuilder Pro. First, use Protean 3D/Protein Design to try different mutations at different positions and predict which combination will produce the most stable protein. Results are returned for each scanned sequence. The Protein Design reports shows the predicted stability or instability in numerical format. Second, use the primer design workflow in SeqBuilder Pro to create a DNA fragment with the mutation. Once you create the primers, use virtual cloning wizard to introduce the fragment into your expression clone using the cloning method of your choice. The hotspot option requires a separate Protein Design license.

Copyright DNASTAR, Inc. 2022

Contact support@dnastar.com for more information.

- **Search for variants that improve folding stability (early release)** – This workflow is similar to the **Create a specific variant** workflow, but allows you to specify the optimal combination of substitutions at different positions. When you select this workflow, the Substitutes and Options wizard screens provide additional options that are not available for the other two workflows. See those user guide topics for details. Note that this workflow is designated as being in the “early release” stage because the energy function currently being used (DFIRE-A) tends to bias results to preferentially select aromatic or other large residues. Future versions of this workflow will use an energy function that could improve this behavior by accounting for other factors such as the energy related to solvation.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

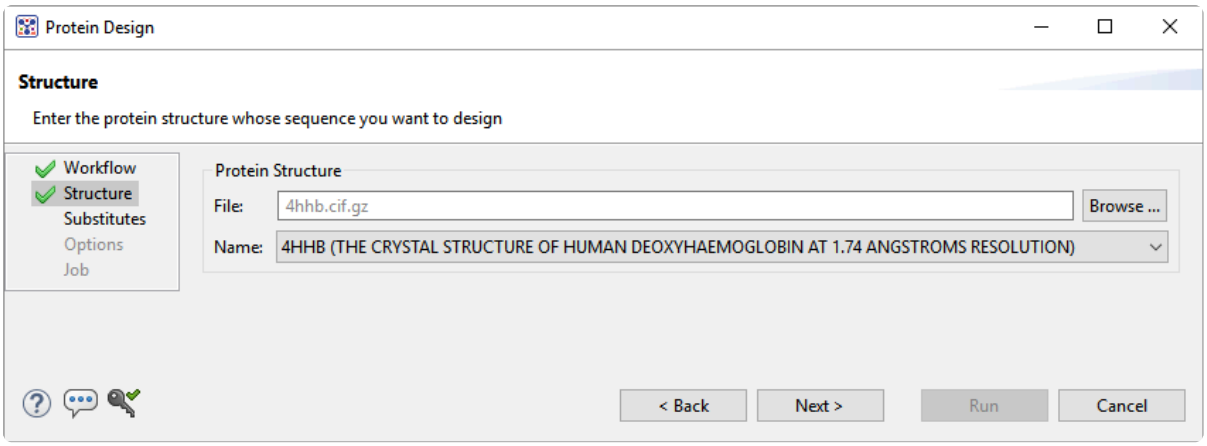
Once you have made the desired selections, do any of the following:

- To specify additional options, press **Next** to continue to the Structure screen.
- To begin the variant creation procedure immediately, press the **Run** button. This will create an alanine substitution at every position in every chain of the current structure.
- To close the Protein Design wizard without starting a search, press **Cancel**.

Copyright DNASTAR, Inc. 2022

Protein Design wizard: Structure

The Structure screen is the second screen in the Protein Design wizard.



Press the **Browse** button next to the **File** field to select the structure file. The following file types are supported: PDB (*.pdb*, *.ent*), Gzip PDB (*.pdb.gz*, *.ent.gz*), mmCIF (*.cif*), Gzip mmCIF (*.cif.gz*), Protean 3D structure (*.structure*). Once you make a selection, the file name will appear in the **File** field and a summary description of the structure will appear in the **Name** field.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

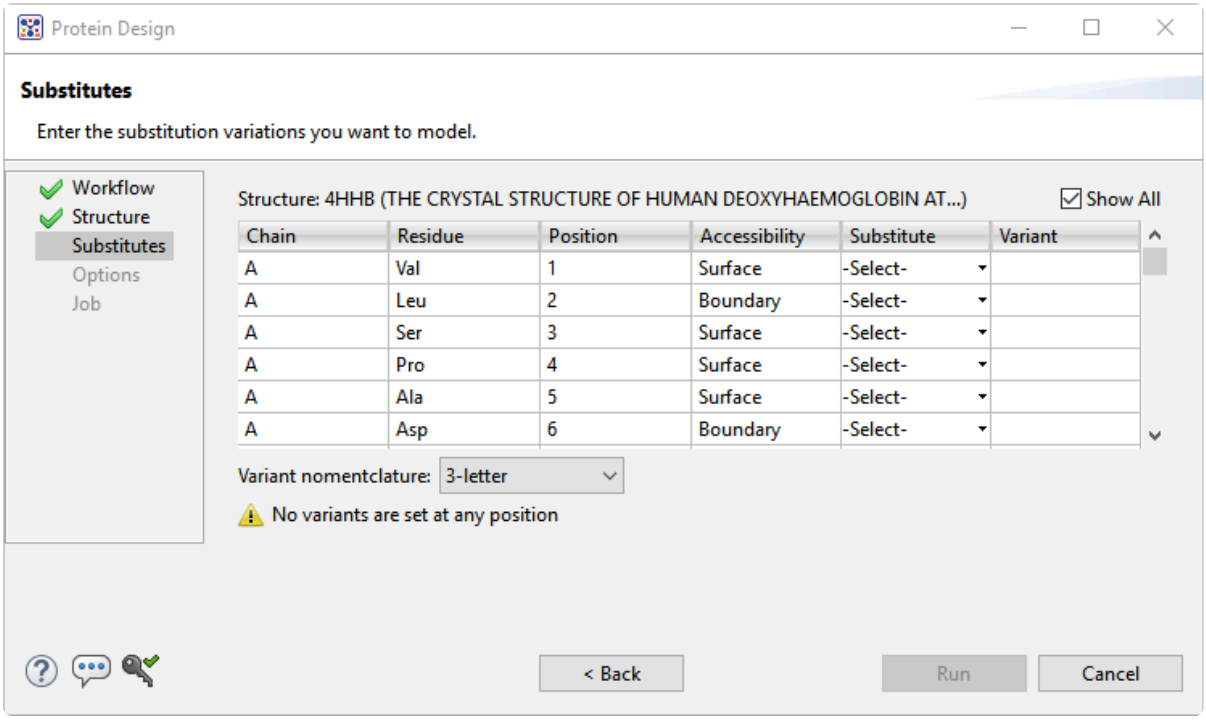
Continue by doing one of the following:

- To specify additional options, press **Next** to continue to the Substitutes screen.
- To return to the Workflow screen, press **Back**.
- To begin the variant creation procedure immediately, press the **Run** button. For the next steps, see View progress and see variant results.
- To close the Protein Design wizard without starting a search, press **Cancel**.

Copyright DNASTAR, Inc. 2022

Protein Design wizard: Substitutes

Substitutes is the third screen of the Protein Design wizard. When using certain menu commands, the wizard opens to this screen, already prepopulated with the currently-selected residues.



Each row of the table represents a single residue.

- **Chain** – The chain where the residue is located.
- **Residue** – The amino acid code for the residue.
- **Position** – The numerical position of the residue on the sequence.
- **Accessibility** – The solvent accessibility of a residue. Possible values are **Surface**, **Core**, or **Boundary**.
- **Substitute** – A drop-down menu from which you can choose a different amino acid residue or set of residues for the position. The selection made in this menu will replace the original residue for that row. If you selected **Create a specific variant** on the Workflow screen, menu options will include only individual amino acid codes.

  If you selected **Search for variants that improve folding stability (early release)** on that screen, this menu will also include the following options:

  - **All** – All 20 amino acids
  - **Core** – AVLIFYW (those likely to be found in protein cores)

- **Surface** – ANQSTHDEKR (those likely to be found on protein surfaces)
- **Boundary** – Those likely to be found at the union of the **Core** and **Surface** groups
- **Polar** – CNQST
- **Nonpolar** – AGILMPV
- **Aromatic** – FWY
- **Acidic** – DE
- **Basic** – HKR
- **Charged** – DEKRH

Core, **Surface**, and **Boundary** are defined in Pierce et al. **Nonpolar**, **Aromatic**, **Acidic**, **Basic**, **Charged** are defined the same as in the Composition report and other locations in Protean 3D.

- **Variant** – After choosing a **Substitute** residue, this column will display an automatically-generated name for the variant. The name consists of the original residue, its position, and the substitute residue, in that order.

Use the **Variant nomenclature** drop-down menu to choose from the following options. The selected option will be reflected in how residues are listed in the **Residue**, **Substitute**, and **Variant** columns:

- **1-letter** – To use 1-letter amino acid codes.
- **3-letter** – To use 3-letter amino acid codes.
- **Genome (HGVS)** – To use the Human Genome Variation Study's naming convention.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

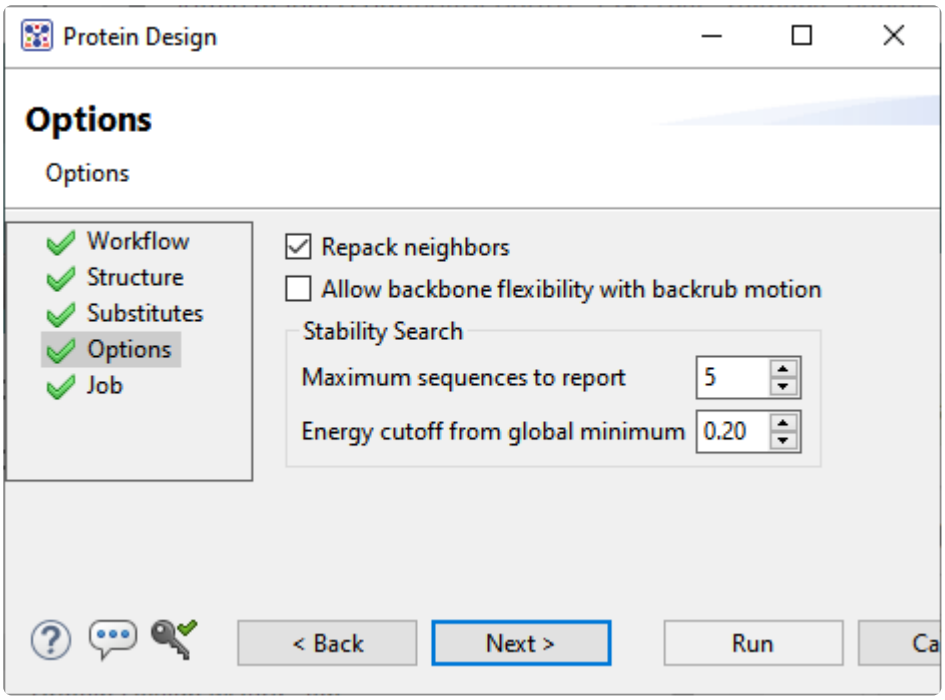
Continue by doing one of the following:

- To specify additional options, press **Next** to continue to the Options screen.
- To return to a previous screen, press **Back**.
- To begin the variant creation procedure immediately, press the **Run** button. For the next steps, see View progress and see variant results.
- To close the Protein Design wizard without starting a search, press **Cancel**.

Copyright DNASTAR, Inc. 2022

Protein Design wizard: Options

Options is the fourth screen of the Protein Design wizard.



Enter the substitution variations you wish to model.

- **Repack neighbors** – When this box is checked, as it is by default, the residues in contact with the variant positions are also repacked during modeling. This can help better accommodate changes occurring at selected positions.

- **Allow backbone flexibility with backrub motion** – When this box is checked (it is unchecked by default), the backbone atoms at the selected positions are allowed to move using the “backrub” algorithm. This allows sampling local backbone plasticity in a realistic manner (reference).

The following two options are only visible if you selected **Search for variants that improve folding stability (early release)** in the Workflow screen:

- **Maximum sequences to report** – In the box, specify the maximum number of new sequences to report from the search. Sequences are ranked by lowest energy.

- **Energy cutoff from global minimum** – In the box, specify an energy threshold value. Sequences whose energy is higher than that of the lowest-energy sequence PLUS this threshold value will be excluded from the list of results.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

Copyright DNASTAR, Inc. 2022

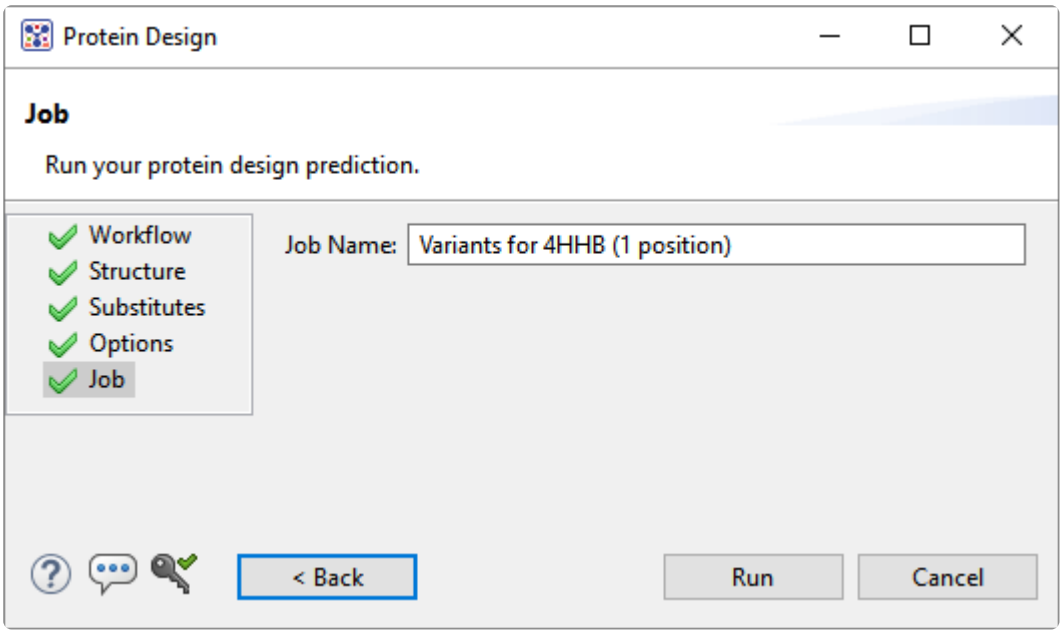
Once you have made the desired selections, continue by doing one of the following:

- To specify additional options, press **Next** to continue to the Job screen.
- To return to a previous screen, press **Back**.
- To begin the variant creation procedure immediately, press the **Run** button. For the next steps, see View progress and see variant results.
- To close the Protein Design wizard without starting a search, press **Cancel**.

Copyright DNASTAR, Inc. 2022

Protein Design wizard: Job

Job is the fifth and final screen in the Protein Design wizard.



By default, the **Job Name** for the Create Variants workflow consists of the following text: **Variants for {PDB ID}** followed by the number of variant positions in parentheses. By default, the **Job Name** for the Hot Spots workflow consists of the following text: **Hot Spot Scan for {PDB ID}** followed by the number of variant positions in parentheses. If you wish, you may overwrite part or all of the default name by typing custom text into the box.

To learn about the optional tools in the bottom left corner of the screen, see the bottom of this topic.

Continue by doing one of the following:

- To return to a previous screen, press **Back**.
- To begin the variant creation procedure, press the **Run** button. For the next steps, see View progress and see variant results.
- To close the Protein Design wizard without starting a search, press **Cancel**.

Copyright DNASTAR, Inc. 2022

View progress and see variant results

Once you have pressed the **Run** button in the Protein Design wizard, you can use the Jobs panel to monitor progress, stop the run, or open results.

After the variant has been created, the following views and panels can be used for analysis:

- The Protein Design reports, Variants, Hot Spot Scan – These two reports show results for the variants and hot spot scan workflows. Once the report is open, you can open a specific variant structure by double-clicking its row in the report table.

- Structure view – By default, variant amino acids are displayed with side chains. For non-mutants, only the backbone is shown. To easily visualize the environment around the variant, select neighboring residues and change their colors.

- Details panel – Shows tabular results, including the change in energy score and prediction of the effect (stabilizing, destabilizing, no effect).

- Molecules section – Shows the structure and chain. The variant name is formed using standard 1-letter, 3-letter or HGVS nomenclature, as specified in the Substitutes screen.

Copyright DNASTAR, Inc. 2022

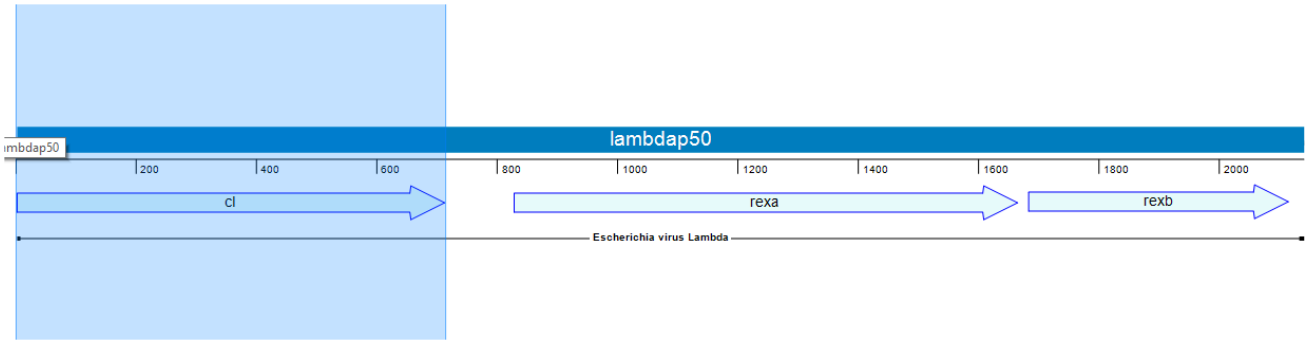
Use Protein Design for site-directed mutagenesis

Site-directed mutagenesis is a molecular biology method that is used to make specific changes to the DNA sequence of a gene. As of Lasergene 16.0, you can perform site-directed mutagenesis using Protean 3D's Protein Design service together with SeqBuilder Pro. The following case study is based on a poster that DNASTAR presented at the 2019 PEGS Summit in Boston.

Introduction:

Nature commonly evolves proteins to provide sufficient fold stability to optimize its activity—not its energetic stability. Lasergene's site-directed mutagenesis workflow is designed to help improve the developability of a protein by predicting which mutations may lead to increased protein stability and developability.

In the following example, we'll use Lasergene's site-directed mutagenesis workflow to replicate the findings of Lim, et al., 1994 (*Proc. Natl. Acad. Sci. USA, 91:423-7*). Lim made three specific mutations to the repressor protein *cI*. This protein is part of the *lambda P50* operon (PDB ID: 1LMB), which also includes the genes *rexa* and *rexb*.



Lim's study found that mutating the repressor protein *cI* at three specific points conferred 0.5 kcal/mol more stability compared to the wild type.

As we describe below, Lasergene's site-directed mutagenesis workflow not only replicates Lim's findings, but early on—in Step #2—reveals other mutations that could be even more stabilizing. Our results were presented by us in a poster at the 2019 PEGS conference. **Note that we do not provide data for the workflow described and it should be considered a "use case" rather than a tutorial. As of November 2021, we are planning to write this as a tutorial with downloadable data in the near future.**

Note: The coordinates in this blog post match the current PDB structure, not the coordinates reported from the 1994 paper or the poster cited above. UniProt indicates position 1 is a removed initiator methionine and the chain starts at position 2 (serine). By contrast, RCSB numbers the start of the chain after post-translational modification as position 1. This blog post uses a sequence that follows the UniProt convention rather than the RCSB convention.

Copyright DNASTAR, Inc. 2022

Part I: Scan for hotspots and create variants on protein structure

Since we have a protein structure as our starting point, we will begin with the following steps. If a protein structure was not available, we could instead start by creating variant primers as shown in Part II.

1. Launch Protean 3D and use **File > Open from PDB** to open the protein sequence **1LMB**.

   We will begin by performing “hot spot scanning” using alanine. This hot spot workflow is used to predict folding hot spots; residues that contribute to fold stability more significantly than others. The more destabilizing the alanine variant, the greater the probability the position and residue identity is important to the fold.

2. Choose **Modeling > Protein Design > Scan for Hot Spots With > Document**. Keep all defaults and choose **Run**.

   In approximately a minute or less, the results of the alanine scan appear in the Report view. The Prediction column indicates that the positions referenced by Lim 36, 40 and 47 (in addition to other positions) are all predicted to be “destabilizing.”

Report 2: Hot Spot Scan for 1LMB (179 positions)

Hot Spot Scan for 1LMB (179 positions)

Molecule	Variant	ΔE (DFIRE-A)	Prediction
1LMB:3	E34A	-0.20	
1LMB:3	S35A	-0.44	
1LMB:3	V36A	1.34	destabilizing
1LMB:3	A37A	-0.00	
1LMB:3	D38A	-0.41	
1LMB:3	K39A	-0.40	
1LMB:3	M40A	0.94	destabilizing
1LMB:3	G41A	-0.25	
1LMB:3	M42A	0.47	
1LMB:3	G43A	-0.37	
1LMB:3	Q44A	-0.17	
1LMB:3	S45A	-0.34	
1LMB:3	G46A	-0.73	stabilizing
1LMB:3	V47A	1.00	destabilizing
1LMB:3	G48A	-0.86	stabilizing
1LMB:3	A49A	-0.01	

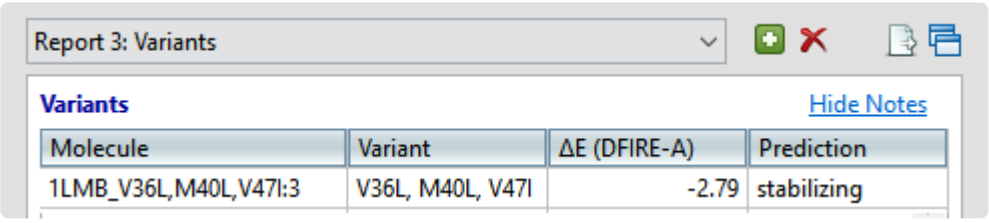
The Report view contains several residues that are even more destabilizing (i.e., have higher DFIRE-A potentials) than the ones in Lim’s 1994 paper. Ideally, these would be our first choice for increasing the developability of the protein. However, for purposes of replicating Lim’s findings, we will proceed with the three positions referenced above.

Copyright DNASTAR, Inc. 2022

3. To manually insert variants into these positions of interest, choose **Modeling > Protein Design > Create Variant With > Document**. Then use the **Substitute** drop-down menus to change position 36 from **V(al)** to **Leu**; position 40 from **M(et)** to **Leu**; and position 47 from **V(al)** to **Ile**.

   If we weren't replicating Lim's study, which amino acids would we substitute for the originals? Unfortunately, there is no "magical" way to determine which variants will work best. We can simply run different combinations and see which ones lead to increased stability.

   Press **Run**. The new Report shows that this combination of variants is predicted to be "stabilizing."



Molecule	Variant	ΔE (DFIRE-A)	Prediction
1LMB_V36L,M40L,V47I:3	V36L, M40L, V47I	-2.79	stabilizing

4. Export the structure file using **File > Export Data > Export All Structures**. In the **Save to** section, type in **1LMB** and use the **Browse** button to save the file to any convenient location, such as your desktop. Keep the default **Format** of **Protein Data Bank (.pdb)**. Press **Export**. This will cause two files to be saved, the original file (*1LMB.pdb*) and the mutated version (*1LMB_V36L, M40L, V47I.pdb*).

<u>Part II: Introduce variants in sequence and create mutated primer</u>

1. Launch SeqBuilder Pro and use **File > Open** to open an annotated version of the unmutated file for the *1LMB.pdb* protein. (This file was made using manual sequence annotation).

2. Reverse translate the sequence to DNA using **Sequence > New Sequence > Back Translate**. Choose **File > Save**. (Note: You can see the locations of the three variants in the image in Step 5).

3. Select all nucleotides from the beginning of the cI gene to the right (downstream) of the sequence and delete them. Then use **File > Save As** to save the remaining fragment as *Upstream fragment.sbd*.

4. Open the sequence saved in Step 2, above. Select all nucleotides to the left (upstream) of the *cI* gene and delete them. Then use **File > Save As** to save the remainder of the sequence as *cI mutated fragment.sbd*.

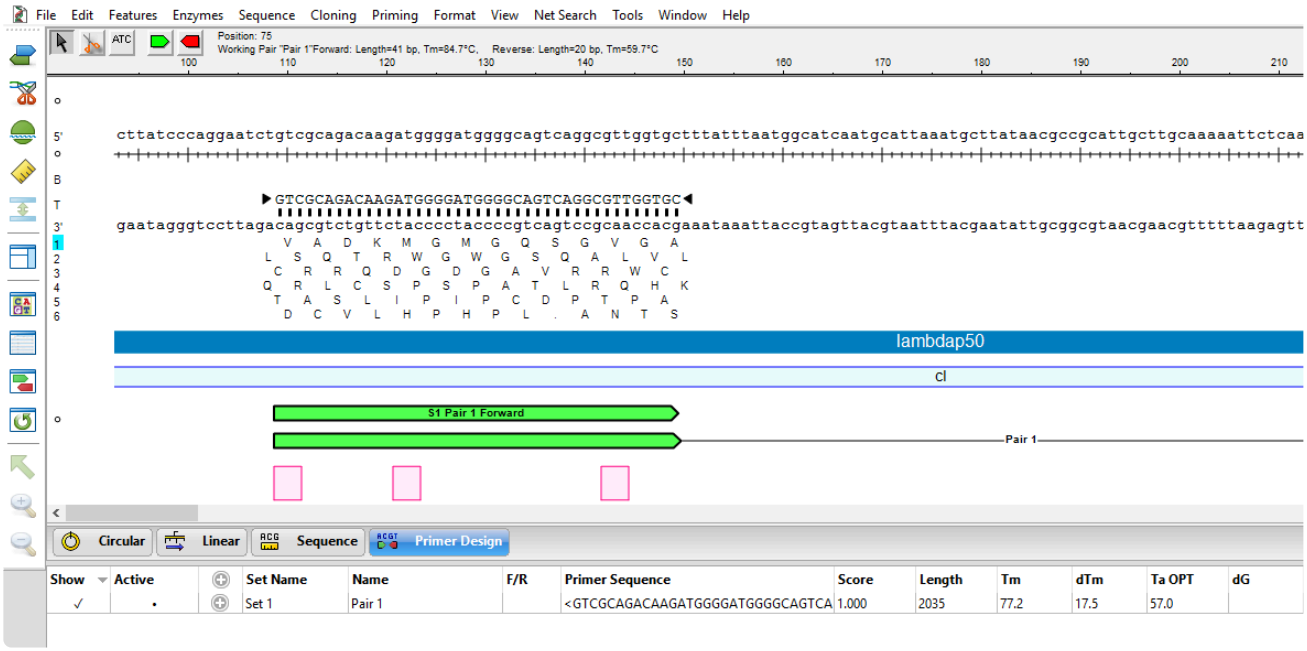
   We now have two sequence files that—laid end to end—would be identical to the original nucleotide sequence.

5. In this step, we will use the SeqBuilder Pro primer workbench to create PCR primer pairs at the exact desired location with respect to features and translational frames. Choose **Edit > Select All**, then **Priming > Create Primer Pairs**. Keep the default setting Primers end exactly at selection and press **OK**. Once the primers are created, the forward primer is extended to 41 bp in length to fully cover all

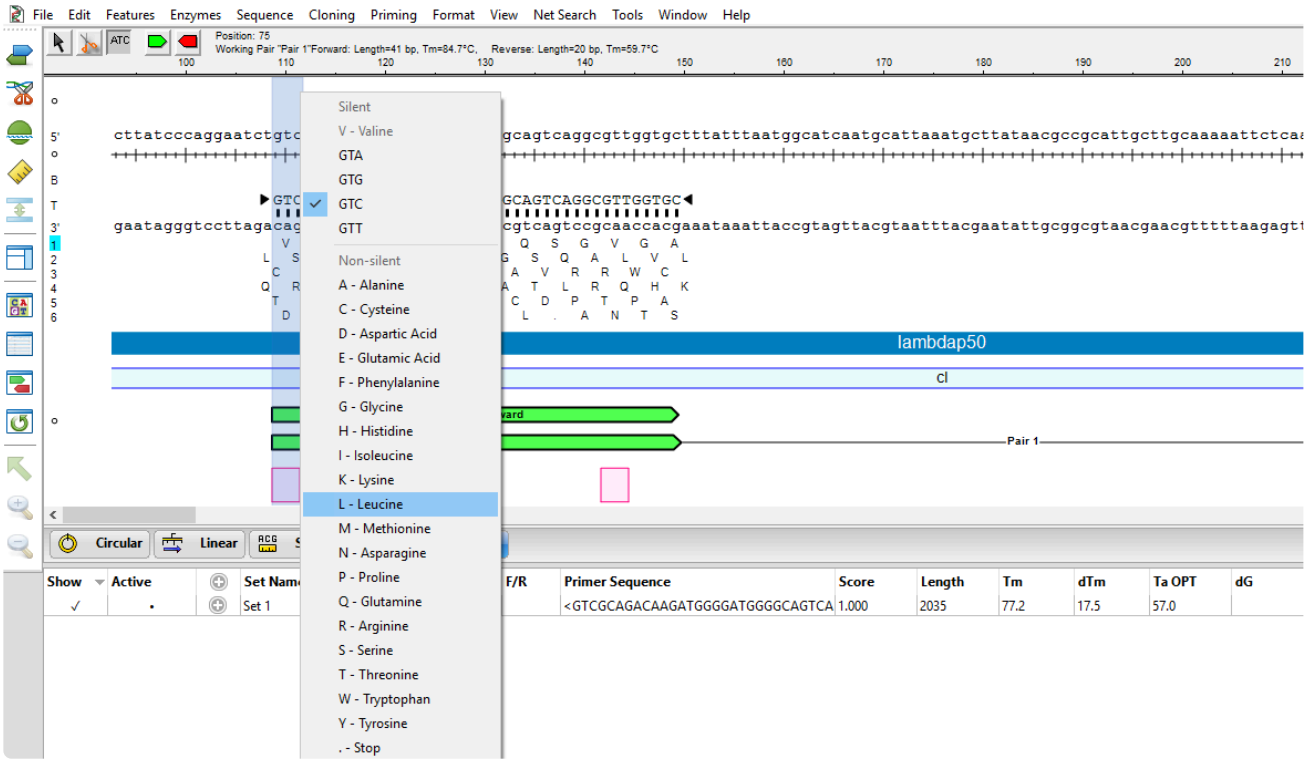
Copyright DNASTAR, Inc. 2022

three mutation areas.

The image below shows that the three codons where the variants will be introduced (pink) all occur on the forward primer (green).



6. Next, we need to recreate the variants of interest using the Introduce **Codon Change/Mutation** tool (ATC). This tool allows us to easily modify triplet codons within primers to introduce the desired mutations. Select the tool, then hover over the first codon (**V** in Reading Frame 1) and choose the desired substitute: **L**.



7. In the same way, introduce the planned variants at the second and third codons. The mutated bases are represented by a pink lower case font, as shown below.

At this point, either or both primers could be further adjusted to improve T~m~ differences and primer-dimer formation.

8. Save the *cI mutated fragment.sbd* project using **File > Save**.

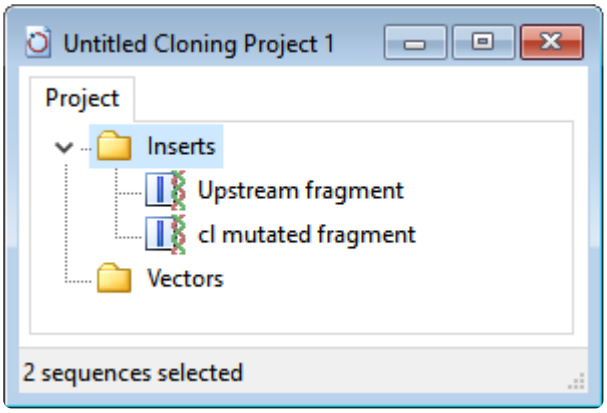
9. Reopen the sequence fragment *Upstream fragment.sbd* and create primer pairs using the commands in Step 4, above. Save the project using **File > Save**.

SeqBuilder Pro provides multi-fragment PCR based cloning options so that the mutated gene of interest can be cloned into an appropriate expression vector. In the next steps, we will clone the mutated cI gene (and the rest of the *lambdap50 operon*) and the upstream (non-mutated) fragment of the *cI* gene into an *E.coli* expression vector system (pBAM/Thio) using PCR-based Gibson assembly.

10. Open a new, empty cloning project using **File > New Cloning Project**.

11. Click on the **Inserts** folder to select it. Then right-click on the **Inserts** folder and choose **Import**. Use **Ctrl+click** (Win) or **Cmd+click** (Mac) to select the files *cI mutated fragment.sbd* and *Upstream fragment.sbd*, then click **Open**.

Copyright DNASTAR, Inc. 2022

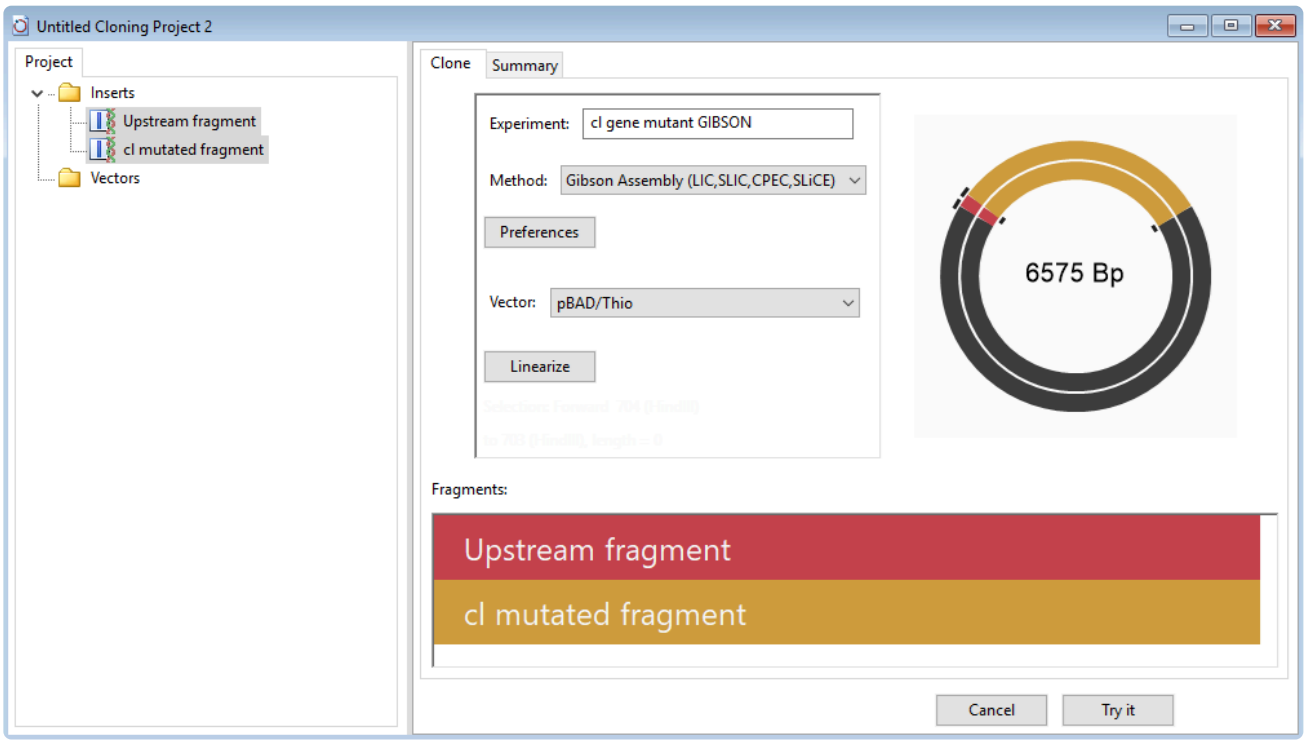


12. Choose **Cloning > Gibson, LIC Cloning**.

13. Next to **Experiment**, type in **cl gene mutant**.

14. Use the **Vector** drop-down menu and click the plus sign next to **Restriction Vectors** to expand the folder. Then click on the first vector in that folder, **pBAD/THIO**.

15. Drag the two sequence fragments from the **Inserts** folder and drop them in the **Fragments** area on the bottom right. Make sure the upstream fragment is on top. In the cloning dialog, name it **cl gene mutant**.



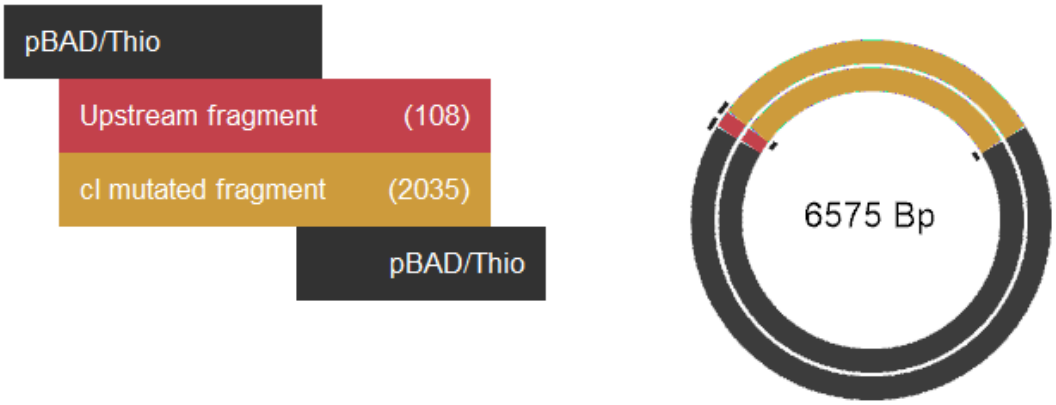
16. Press the **Linearize** button and specify a **Left Cut** at **NaeI (676)** and **Right Cut** at **HindIII (703)**. Press **Select**.

17. Press **Try it**, then **Make it**. The Cloning Summary shows a rendering of the fragments and clone, all

primer information, and the list of steps needed to create the clone in the lab. The red text in the Primers area is the junction that puts the two sequence fragments back together.

Cloning Summary

Fragments



Primers

Overlap	Primer	Anneal	T/B	Tm	Ta
pBAD/Thio	ATCCGGTGATGACGATGACAATGAGCACAAAAAAGAAACC	Upstream fragment	T	60.45	52.85
cl mutated fragment	CAACTTGTCTGCAAGAGATTCCTGGGATAAGCCAAGTT	Upstream fragment	B	60.48	52.88
Upstream fragment	CTTGGCTTATCCCAGGAATCTCTTGCAGACAAGTTG	cl mutated fragment	T	60.51	52.91
pBAD/Thio	ATAGGCTTACCTTCACAAAGAAAACCCGGCGCTGA	cl mutated fragment	B	60.53	52.93

Steps

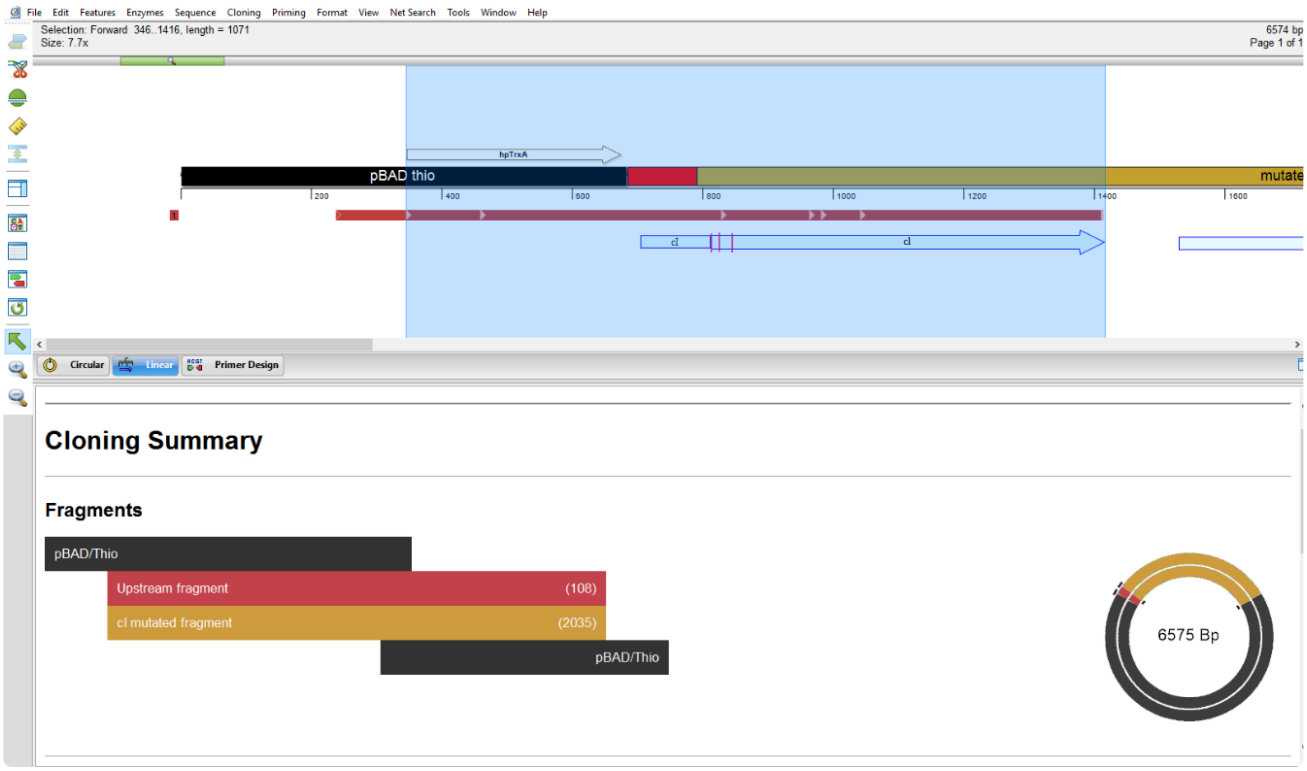
```
1) Cut pBAD/Thio with HindIII

2) Amplify Upstream fragment using primers:
        Top:    5'-ATCCGGTGATGACGATGACAATGAGCACAAAAAAGAAACC-3'
        Bottom: 5'-CAACTTGTCTGCAAGAGATTCCTGGGATAAGCCAAGTT-3'

3) Amplify cl mutated fragment using primers:
        Top:    5'-CTTGGCTTATCCCAGGAATCTCTTGCAGACAAGTTG-3'
        Bottom: 5'-ATAGGCTTACCTTCACAAAGAAAACCCGGCGCTGA-3'
```

18. Open expression clone *cl gene mutant GIBSON Clone 1* in SeqBuilder Pro by double-clicking on it.

    The Linear view shows all the of the vector and insert feature annotations and lets us apply ORF detection and translation to confirm that the gene of interest remains in-frame with the fusion protein.

19. Apply the ORF for “Top Frame 1,” shown as a red bar in the image below. The small white triangles along the ORF denote the *atg* start sites.

20. Click on the start site corresponding to *hpTrxA*, a mutated version of the *E. coli* protein thioredoxin (trxA) used to create a metal-binding domain that allows purification of thioredoxin fusion on metal-chelating resins.



The selection contains the entire *cI* gene feature; verification that the transcript is in-frame.

Select Neighboring Residues

After creating or finding variants using Protein Design, a common situation is the desire to visualize the environment around the variant. You may want to know how the new sidechain fits in its environment, and whether or not it changes the rotamer of any other residues.

Protean 3D allows you to easily locate neighboring residues using the “select neighbors” function. This function changes an existing selection to one consisting of those residues adjacent to the original selection.

Note that the command can be used with any selection, so it is not necessary to create variants before using this functionality.

To locate “neighbors” of a current selection, do any of the following:

- Choose the **Structure > Select Neighbors** command
- Use the **Select Neighbors** tool ( ).
- Hold down the **Shift** key and double-click in the Structure view.

When a single residue is selected, neighbors include all adjacent residues. When multiple residues, a ligand or molecule are selected, neighbors might encompass interfacing groups of residues on an adjacent chain.

Once neighbors are selected, you may wish to change their display color for ease of recognition.

Copyright DNASTAR, Inc. 2022

Evaluate Variants Produced in a SeqMan NGen Assembly

If you are working with templated human assemblies, Lasergene's "SNP to Structure" workflow lets you combine genomic sequencing and variant level data with structure files from the RCSB Protein Data Bank (PDB) to model point mutations on the protein structure and assess the effect on protein stability. By combining structural bioinformatics with sequencing technologies, this integrated workflow can guide genomic and molecular biology researchers to create structure-based hypotheses and to investigate possibilities not evident by sequences alone.

This workflow requires that you be licensed to use several Lasergene applications: SeqMan NGen, SeqMan Pro and Protean 3D (all required), and ArrayStar (optional).

Only Part D of the workflow involves Protean 3D. However, all parts of the workflow are described below.

- Part A: Create a templated assembly in SeqMan NGen
- Part B (optional): Filter for variants of interest in ArrayStar
- Part C: View variants in SeqMan Pro
- Part D: View the protein structure in Protean 3D

Copyright DNASTAR, Inc. 2022

Part A: Create a templated assembly in SeqMan NGen

To evaluate the effect of a variant on protein structure, the first step is to create a templated assembly in SeqMan NGen.

Create an assembly using the genome template package *Homo_sapiens-GRCh38-Ensembl-dbSNP150.zip*. This template contains the mapping of the sequences of PDB structures to the human genomic coordinates, and will later allow SeqMan Pro to communicate with Protean 3D. SeqMan NGen outputs an *.astar* and an *.assembly* package to use in later steps.

Continue to Part B (optional): Filter for variants of interest in ArrayStar or Part C: View variants in SeqMan Pro.

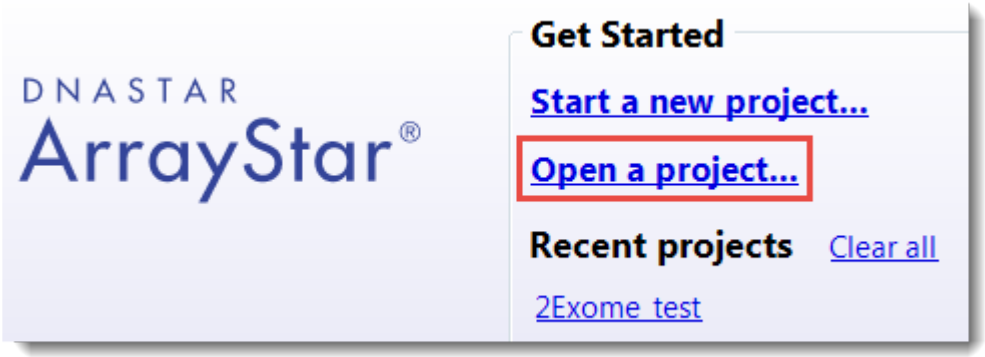
Copyright DNASTAR, Inc. 2022

Part B (optional): Filter for variants of interest in ArrayStar

This is continued from Part A: Create a templated assembly in SeqMan NGen.

If your assembly has a large number of variants, you can use ArrayStar to filter them down to a smaller group of interest before sending them to SeqMan Pro for viewing. This optional step is highly recommended for all but very small assemblies.

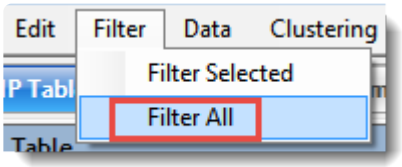
1. Launch ArrayStar and choose **Open a project**.



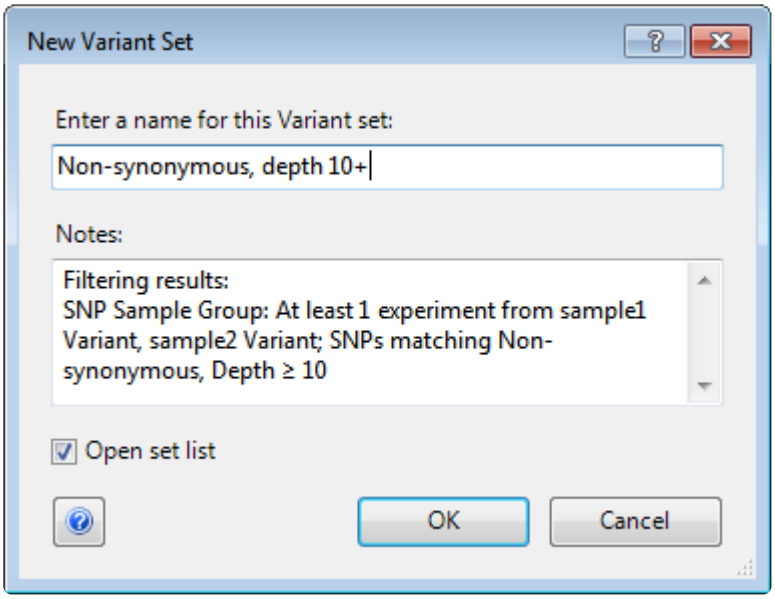
2. Navigate to and select the *.astar* file output by SeqMan NGen.

3. When the file has loaded, click on the **SNP Table** tab.

4. Use **Filter > Filter All** to perform any desired filtering.

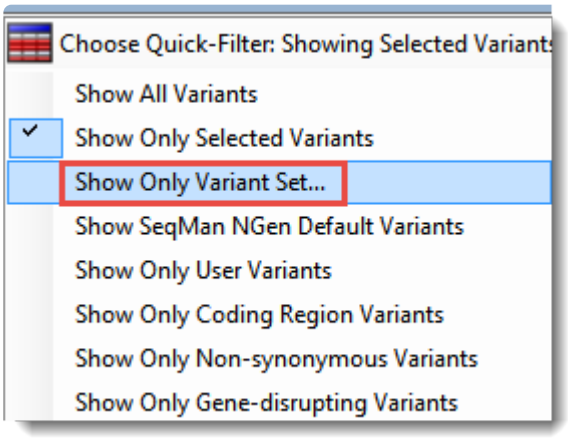


5. At the conclusion of filtering, click the **Remember Results as a Variant Set** tool () above the Search Results table. Type in any desired name and press **OK**.



6. In the Action section in the center-right of the window, click the link **Select and show the table of this set's Variants**.

7. Use the **Choose Quick-Filter** drop-down to select **Show Only Variant Set**.



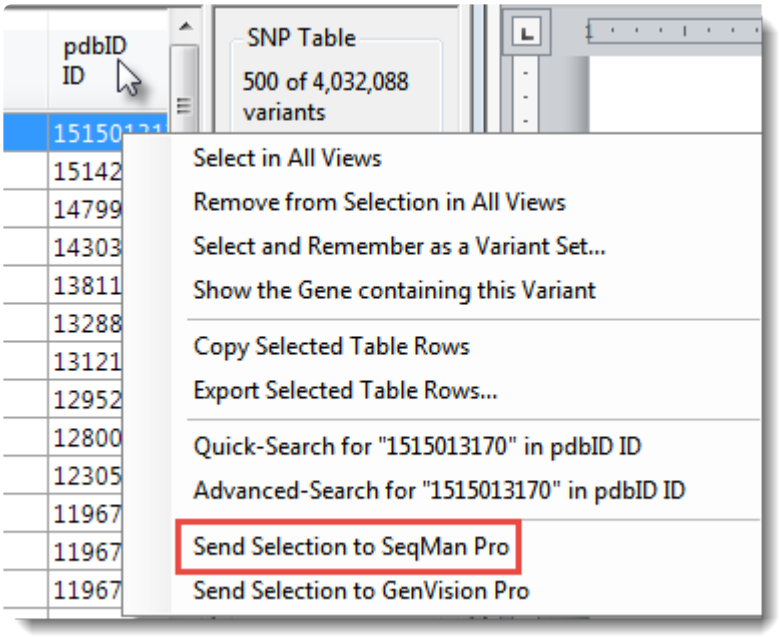
In the ensuing dialog box, select the named set and click **OK**.

8. Click on **Add/Manage Columns**. Select **SeqMan NGen Assembly Variants**, then **pdbID**. Click **Add Column** and **OK**.

9. In the SNP Table, click on the **pdbID** column header to sort the column and locate rows with PDB entries.

10. Within the subset of rows with PDB entries, select any row of interest. Then right-click on it and choose **Send Selection to SeqMan Pro**.

Copyright DNASTAR, Inc. 2022



11. If prompted, select an individual sample of interest and press **OK**.

Continue to Part C: View variants in SeqMan Pro.

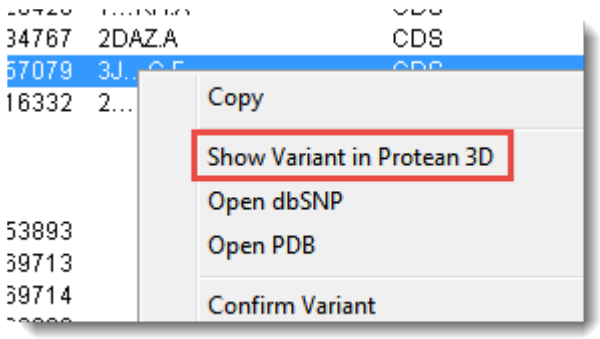
Part C: View variants in SeqMan Pro

Whether you are coming from Part A: Create a templated assembly in SeqMan NGen or the optional Part B: Filter for variants of interest in ArrayStar, the next step is to view the variants in SeqMan Pro.

1. If you are coming from Part B, the sample is automatically selected and its Alignment view is opened. Continue directly to Step 2. If you are coming from Part A, launch SeqMan Pro and use **File > Open** to open the *.assembly* file.

2. Choose **Variant > Variant Report**. (If desired variants are being filtered out, you may need to click on the **Show All** button.)

3. Click on the **PDB ID** column header to sort items with PDB IDs to the top. Select one or more rows, then right-click within the selection and choose **Show Variant in Protean 3D** (or use **Variant > Show Variant in Protean 3D**).

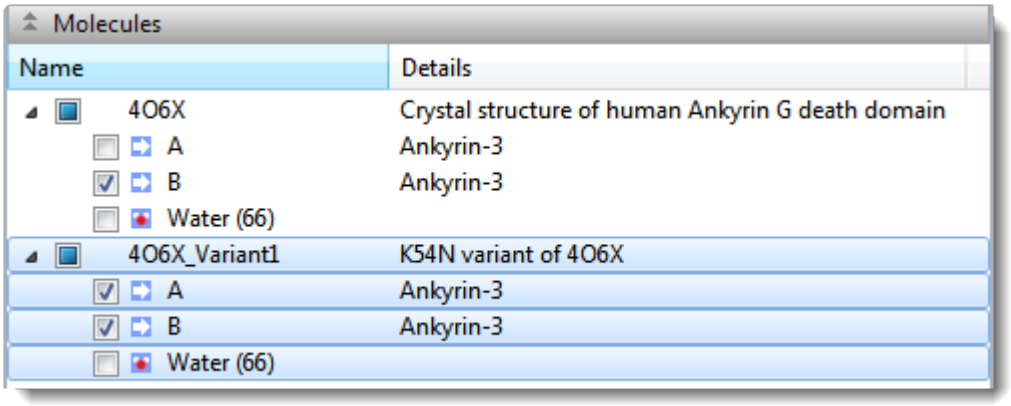


Continue to Part D: View the protein structure in Protean 3D.

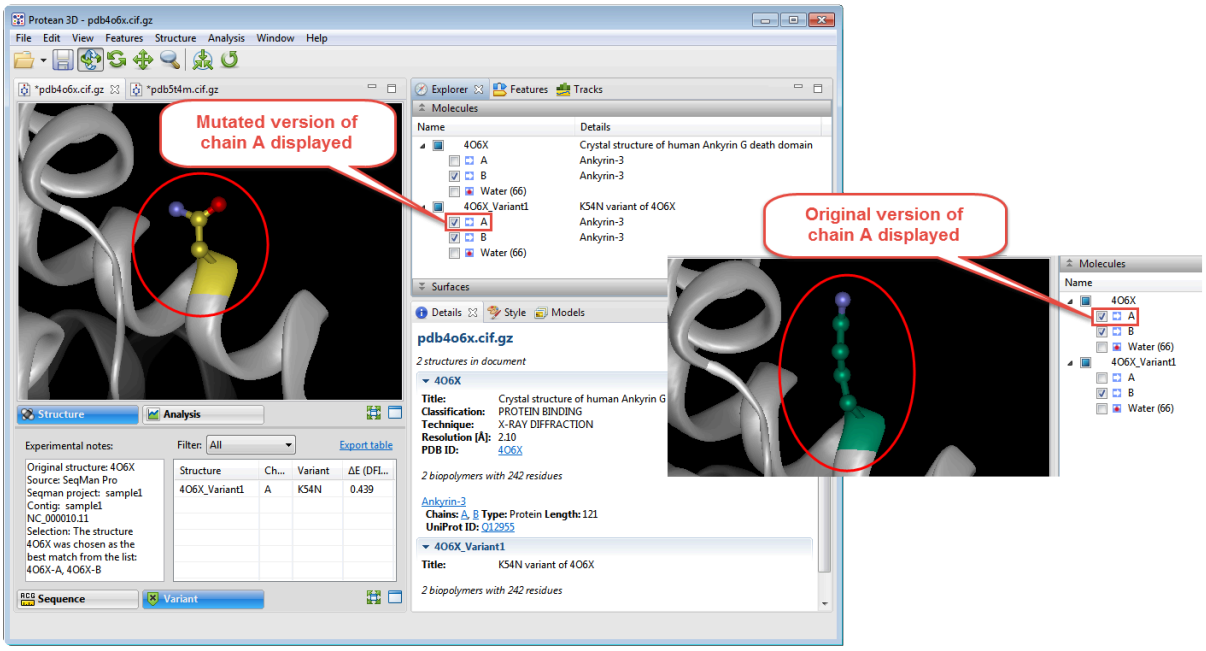
Part D: View the protein structure in Protean 3D

After finishing Part C, the protein structure with the variant of interest opens in Protean 3D.

In the Molecules section, two near-identical copies of the structure appear. The upper structure is the original structure from the Protein Data Bank (PDB). The lower structure is the variant version calculated by Protean 3D.



- Use the Structure view to observe the mutated side chain along the backbone. To show/hide the different versions of each chain, check/uncheck boxes in the Molecules section.



- To see notes about the structure chosen as the best match, look in the Notes box in the Variants report.

```
Original structure: 4WKA
Source: SeqMan Pro
Seqman project:  sample1
Contig:  sample1 NC_000001.11
Selection: The structure 4WKA was chosen
as the best match from the list:
1GUV-A, 1HKI-A, 1HKJ-A, 1HKK-A, 1HKM-A,
1LG1-A, 1LG2-A, 1LQ0-A, 1WAW-A,
1WB0-A, 4WJX-A, 4WK9-A, 4WKA-A,
4WKF-A, 4WKH-A
```

> **Note:** If the residue has no atoms, you will see an error message in this box. In this case, return to SeqMan Pro and choose a different row in the Variants table.

Protean 3D uses several metrics to determine the “best” PDB file to display when a variant is located in a CDS that is associated with more than one PDB file. Quality is the first consideration, with high-resolution crystal structures > NMR structures > low-resolution crystal structures > other techniques. This ordering is refined by alignment to the corresponding Uniprot sequences. This refinement considers the percent match and the number of gaps before the variant position. If two structures are still tied as the “best,” the largest structure is chosen.

- To predict whether the mutation is stabilizing or destabilizing to the protein structure, use the table on the right of the Variants view. The **delta-E (DFIRE-A)** column displays the change in energy value based on the DFIRE calculation (reference). This number can be used to predict whether the mutation is stabilizing or destabilizing to the protein structure. A positive number is considered destabilizing to the structure when compared to the original amino acid; a negative number is considered non-destabilizing.

> **Note:** If multiple entries appear in the table, you can sort them by clicking on their column headers, or filter them using the **Filter** drop-down menu. If desired, you can save the table by clicking the **Export table** link.

- To further explore potential impact of mutation on protein stability and function, apply a solvent-accessible surface.

- In the Analysis view, analyze secondary structure characteristics to interrogate the mutation’s effect on protein flexibility, amphiphilicity, charge density, hydropathy, and more.

Copyright DNASTAR, Inc. 2022

Predict Structures, Antibody Interactions and Docking

Prediction of protein structures, antibody interactions and protein-protein docking take place in the Nova applications, which are licensed separately from Protean 3D. To learn more about each application, see Types of predictions.

All four Nova applications can be run through the Protean 3D interface. In addition, some can be run locally and/or from the NovaCloud, which is accessed through the DNASTAR website.

- Run the Nova applications through the Protean 3D interface. (This is the only option for NovaFold AI).
- Run the Nova applications using NovaCloud, via the DNASTAR website
- Run the Nova applications locally

Copyright DNASTAR, Inc. 2022

Types of predictions

Click the links below to learn about the four Nova applications.

- NovaFold and NovaFold AI use two different algorithms for protein structure prediction
- NovaFold Antibody for antibody interaction prediction
- NovaDock for protein-protein docking exploration

Copyright DNASTAR, Inc. 2022

NovaFold: protein structure prediction

NovaFold, DNASTAR's 3D protein structure prediction program, can use a sequence file to predict ligand binding sites and protein function and to build 3-D protein structure models based on iterative assembly simulations. DNASTAR provides NovaFold as a local or cloud-based, separately licensed service within Protean 3D. The NovaFold client provides an easy to use interface for submitting structure prediction tasks to a high performance computing cluster on the Amazon EC2 web service.

For step-by-step instructions, see Run NovaFold through Protean 3D, Run Nova applications through the DNASTAR website, or Run Nova applications locally.

NovaFold uses 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.

Professor Zhang's algorithms have won the Protein Structure Prediction Center's five most recent Critical Assessment of Protein Structure Prediction (CASP) experiments. Each CASP experiment is a biennial, world-wide evaluation of structure prediction methods, with approximately 100 participating laboratories.

> **Note -** In CASP, I-TASSER is referred to as "Zhang-Server."

- CASP11 Results - #1 Zhang-Server
- CASP10 Results – #1 Zhang-Server
- CASP9 Results - #1 Zhang-Server and QUARK (another Zhang service)
- CASP8 Results - #1 Zhang-Server
- CASP7 Results - #1 Zhang-Server

For references related to the I-TASSER algorithm, see Research References.

Copyright DNASTAR, Inc. 2022

NovaFold AI: protein structure prediction

NovaFold AI is a protein structure prediction application that incorporates the award-winning AlphaFold 2 AI system developed by DeepMind. AlphaFold 2 was the top-ranked protein structure prediction method in the CASP14 challenge in 2020, significantly outperforming the other participating teams by determining the structure of proteins with accuracy comparable to laboratory experiments.

DNASTAR provides NovaFold AI as a cloud-based, separately licensed service within Protean 3D. The NovaFold AI client provides an easy to use interface for submitting structure prediction tasks to a high performance computing cluster on the Amazon EC2 web service.

Like NovaFold, NovaFold AI can use a sequence file to predict ligand binding sites and protein function and to build 3-D protein structure models based on iterative assembly simulations. NovaFold AI enables you to seamlessly utilize the AlphaFold 2 method to predict the protein structure of an amino acid sequence, and then visualize and analyze your results. Both the prediction and analysis take place within the existing NovaCloud Services interface of Protean 3D.

For step-by-step instructions, see Run NovaFold AI through Protean 3D, Run Nova applications through the DNASTAR website, or Run Nova applications locally.

Copyright DNASTAR, Inc. 2022

NovaFold Antibody: antibody interaction prediction

NovaFold Antibody is specifically designed to generate models of antibodies and antibody fragments (Fv, Fab, VH, sdAb) and, with the use of NovaDock, predict the structure of antibody-antigen complexes. DNASTAR provides NovaFold Antibody as a local or cloud-based, separately licensed service within Protean 3D™. Antibodies frequently have two chains of interest: a light and a heavy chain. When you submit a prediction using NovaFold Antibody, you can provide a light chain, a heavy chain, or both. If you submit two chains, the software will model the complex; if you submit one, it will instead model the chain.

For step-by-step instructions, see Run NovaFold Antibody through Protean 3D, Run Nova applications through the DNASTAR website, or Run Nova applications locally.

The following video is a brief introduction to NovaFold Antibody:

The NovaFold Antibody algorithm utilizes a combination of homology modeling and *ab initio* loop prediction, resulting in highly accurate predictions. During modeling, NovaFold Antibody searches the input sequence(s) against thousands of non-redundant protein antibody structures from PDB and finds the best template matches for the chain or complex.

During modeling, NovaFold Antibody gives particular consideration to the complementary determining region (CDR) loops, hypervariable regions of an antibody that react dominantly with an antigen. The three CDR loops on the heavy chain are known as H1, H2 and H3, while the three on the light chain are L1, L2 and L3. H3 is commonly the most important region in antigen binding. Due to its increased length and flexibility, it is also the most difficult to model. As such, NovaFold Antibody limits H3 loop modeling to fifteen or fewer residues, a length commonly seen in antibody modeling problems. Finally, NovaFold Antibody performs energy minimization calculations to construct the final predicted structure model for the antibody

Copyright DNASTAR, Inc. 2022

chain or complex. This entire process takes approximately 5-15 minutes on a standard workstation computer.

After running a NovaFold Antibody prediction, the resulting *.antibody* structure file can be viewed and analyzed in DNASTAR's Protean 3D application. Protean 3D's Structure view and Features panel facilitate examination of antibody features, including the six annotated CDR loops for light and heavy chains.

NovaFold Antibody can also be used to create an input structure for NovaDock, DNASTAR's protein-protein docking application that can also be used to predict atomic-level antibody/antigen interactions. When setting up a NovaDock prediction, simply use Protean 3D to export the predicted structure in PDB format and specify the exported file as one of the two binding partners.

Copyright DNASTAR, Inc. 2022

NovaDock: protein-protein docking

NovaDock is DNASTAR's protein docking application. DNASTAR provides NovaDock as a local or cloud-based, separately licensed service within Protean 3D. NovaDock's algorithm is based on SwarmDock, developed in Dr. Paul Bates' laboratory at the Cancer Research UK's London Research Institute, and ongoing at the Francis Crick Institute. The SwarmDock algorithm is provided to DNASTAR under license from Cancer Research Technology Limited.

For step-by-step instructions, see Run NovaDock through Protean 3D, Run Nova applications through the DNASTAR website, or Run Nova applications locally.

The following video is a brief introduction to NovaDock:

NovaDock is used to predict atomic interactions between two binding partners, both of which must be proteins. One binding partner is the ligand ('L': usually the smaller partner or an antibody); the other is the receptor ('R': usually the larger partner or an antigen). Each binding partner can consist of one or multiple protein chains. NovaDock does not use a library of templates, but instead makes docking predictions based on a type of energy calculation known as "particle swarm optimization."

One of several optional arguments directs NovaDock to use residues which are believed (or known) to be in the protein interface.

NovaDock requires a complete protein structure that uses only standard residues and has no missing residues, atoms or chains. Before the docking simulation, NovaDock automatically replaces any missing atoms or replaces non-standard residues with the closest natural equivalent. If any missing data were replaced, the NovaDock Report (viewed within Protean 3D), will include a notation to that effect in its Details panel. NovaDock then uses "normal modes of motion" to explore protein flexibility during the docking simulation.

Copyright DNASTAR, Inc. 2022

Run Nova applications within the Protean 3D interface

Click the links below to learn how to run the various Nova applications from within the Protean 3D interface:

- NovaFold
- NovaFold AI
- NovaFold Antibody
- NovaDock

Copyright DNASTAR, Inc. 2022

Run NovaFold through Protean 3D

This topic is a general overview on how to set up, run and view a NovaFold prediction **through the Protean 3D interface**. If you are instead running NovaFold locally on a Linux or Macintosh machine, see Run Nova applications locally. If you are running NovaFold online, see Run Nova applications through the DNASTAR website.

1. If you are not currently logged into NovaCloud Services, open the NovaCloud Services view and log in. If you are already logged in, you can skip ahead and launch the NovaFold wizard by doing any of the following:

    - From the Welcome screen, click **Structure Prediction** on the left and then **New protein structure with NovaFold** on the right.
    - Choose **Modeling > Protein Structure Prediction > New NovaFold Prediction**.
    - From the NovaCloud Services view, press the **NovaFold** button.

2. Follow the steps in the three NovaFold wizard screens. Options, Sequences, and Submit.

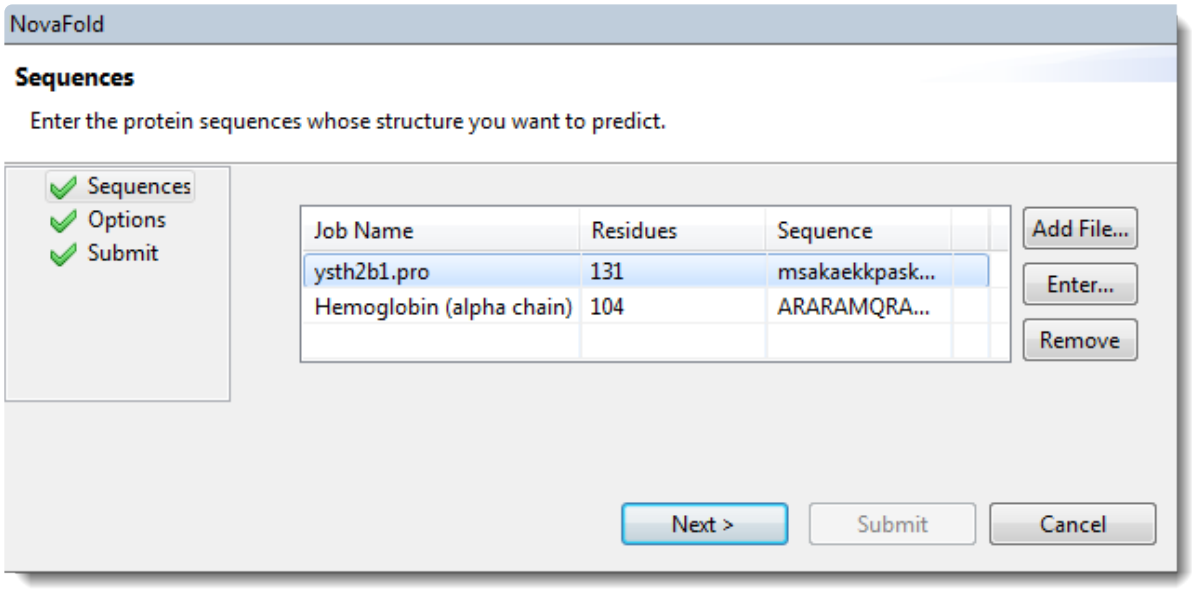
3. Use **View > NovaCloud Services > Predictions** to monitor run status and open or download predictions from within the Predictions view. NovaFold results can be opened on any computer running Protean 3D 12.1 and higher; a NovaFold license is not required to view results.

Open results appear as a new project tab in Protean 3D with the NovaFold Report active. To learn which views and panels are enabled with this type of project, see Protean 3D Interface.

Copyright DNASTAR, Inc. 2022

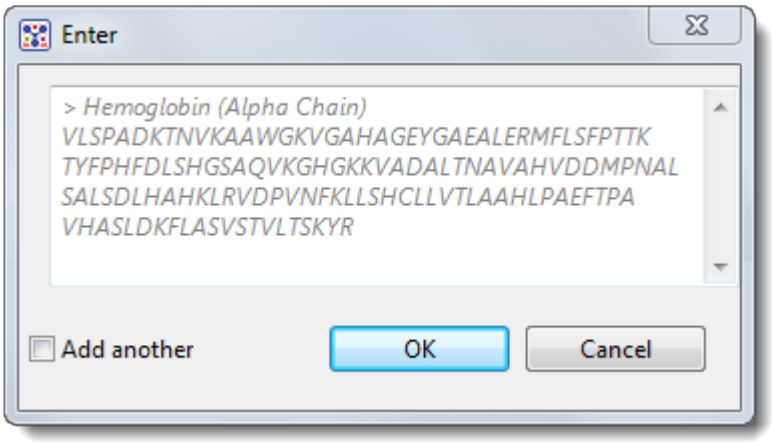
NovaFold wizard: Sequences

When setting up a NovaFold prediction in Protean 3D's NovaFold wizard, the first screen is Sequences. This screen primarily consists of a job table with three columns. When populated, the **Job Name** column displays the name of each query protein sequence or its FASTA title tag. The **Residues** column displays the number of residues in the sequence, and the **Sequence** column displays the first part of the amino acid sequence in IUPAC format.



Task	How To
To upload a file containing one or more sequences in FASTA format	Click **Add File** and navigate to the file. The maximum sequence length is 2,000 residues. Multi-sequence FASTA files are imported as a batch, with each sequence listed on a separate row of the table.
To change the information in any cell of the table	Insert the cursor and typing over the existing text.
To type or paste in a sequence	Click the **Enter** button, then type or paste a valid amino acid sequence into the field provided. If a yellow "error triangle" appears, hover over it to see a tooltip describing the issue (e.g., "sequence must be in FASTA format," etc.). Press **OK** to add the sequence to the job table. If you wish to enter additional sequences, check the **Add another** box before pressing **OK**.

Copyright DNASTAR, Inc. 2022

	
To delete a job from the job table	Select the job and press **Remove**.

Click **Next** to proceed to the Options screen.

NovaFold wizard: Options

When setting up a NovaFold prediction in Protean 3D's NovaFold wizard, the second screen is Options. This screen allows you to edit or specify prediction options. During the prediction step, options specified in this screen are applied to all jobs listed in the Sequences screen.

Copyright DNASTAR, Inc. 2022

NovaFold

Options

Set advanced options that will apply to all predictions.

- Sequences
- Options
- Submit

Enhanced conformational search: ☐

Predict ligand binding sites and protein function: ☑

Number of models: 5

Number of threading templates: 20

Simulation effort: Extended

Add PDB file to template library

Location: Browse...

Add distance and contact restraints or choose a list

```
DIST    58  CE187  NE2    6.4
DIST    42  OH 97  ND2    3.4
CONTACT 43  61
CONTACT 46  61
```

Exclude PDB templates by name or choose a list

```
1MBN:A
1ABW:A    95
1ABW:B    90
```

☐ Remove sequences at 90 % identity

< Back | Next > | Submit | Cancel

Edit or enter information in this wizard screen as desired. Or you can accept the defaults as they are, and simply press **Next** to continue to the Submit screen.

Copyright DNASTAR, Inc. 2022

- **Enhanced conformation search** – If this box is checked, NovaFold will perform DNASTAR's experimental method for enhancing the structural diversity of the normal template set. For templates selected by protein threading, a proprietary process samples alternate structural conformations and replaces a subset of the templates with lower energy conformations. This box is unchecked, by default.

- **Predict ligand binding sites and protein function** – If this box is checked, NovaFold will perform structure/function relationship calculations in addition to the folding prediction. These calculations are designed to elucidate a possible structural mechanism for a putative biological function. Uncheck the box if you do not need this additional prediction information and/or to obtain a faster folding prediction.

- **Number of models** – Select the maximum number of models to create [1, 5]. The default is 5.

- **Number of threading templates** – Select the maximum number of templates to use for simulations [1, 50]. The default is 20.

- **Simulation effort** – Specify whether to run **Quick** (5 hour) or **Extended** (50 hour) simulations. The default is **Extended**.

- The fourth drop-down menu on this screen lets you choose from the following options:

    - **Add PDB template by name** – To nominate a PDB entry to use as a template in the modeling prediction, along with other templates selected by NovaFold. This option requires you to type in a four digit alphanumeric PDB code, followed by a colon and the chain name. For example: *4HHB:*

    - **Add file with your sequence aligned to PDB template** – Select this radio button to provide a text file containing a user-defined template structure and the alignment between that template and the query sequence. Then press **Browse** and select a text file containing either the pairwise FASTA-formatted sequence alignment between query and template, or the standard PDB format 3D structural coordinates of a single protein chain. See an example file further down this topic.

    - **Add PDB file to template library** – Select this radio button to nominate a local 3D structure to use as a template in the modeling prediction, along with other templates selected by NovaFold. Then press **Browse** and select a text file representing a single protein chain in standard PDB format.

  This menu is followed by a **Location** box. Use the **Browse** button to navigate to the desired file location.

- **Add distance or contact restraints** – To restrict predictions based on distance and/or contact restraints (e.g., active sites, zinc fingers, disulfide bonds), type the information into the text box provided. This information will consist of a list of pairwise distances between two atoms ("i" and "j"), or contacts between two residues ("i" and "j"). Alternatively, click on the "choose a list" link to browse to

Copyright DNASTAR, Inc. 2022

a text file containing the same information. Both distances and contacts can be described in different rows of the same restraint file. **IMPORTANT:** The text file may ***not*** contain any lower-case letters. See an example file further down this topic.

- **Exclude PDB templates by name** – To exclude specified PDB structures from being used as templates for the prediction, type their name and chain information into the text box. Alternatively, click on the “choose a list” link to browse to a text file containing the same information.

- **Remove sequences at ___% identity** – Check this box to remove structural templates at or above the specified sequence identity from consideration. This setting was designed primarily for DNASTAR’s internal benchmarking. However, it is also useful for reducing the influence of closely homologous templates. For example, you could select **100%** to exclude perfect matches.

Click **Next** to proceed to the Submit screen, or **Back** to return to the Sequences screen.

Example text file for **Add file with your sequence aligned to PDB template**:

The following is a file for mammoth myoglobin (query) against whale myoglobin (target structure). “ATOM” rows 6-1211 have been omitted for space. The format for ATOM records is described on this PDB web page.

```
>query
MGLSDGEWELVLKTWGKVEADIPGHGLEVFVRLFTGHPETLEKFDKFKHLKTEGEMKASE
DLKKQGVTVLTALGGILKKKGHHQAEIQPLAQSHATKHKIPIKYLEFISDAIIHVLQSKH
PAEFGAD---------------------------
>1MBN:A
-VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASE
DLKKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRH
PGDFGADAQGAMNKALELFRKDIAAKYKELGYQG


ATOM 1 N VAL A 1 -2.900 17.600 15.500 1.00 0.00 N
ATOM 2 CA VAL A 1 -3.600 16.400 15.300 1.00 0.00 C
ATOM 3 C VAL A 1 -3.000 15.300 16.200 1.00 0.00 C
ATOM 4 O VAL A 1 -3.700 14.700 17.000 1.00 0.00 O
ATOM 5 CB VAL A 1 -3.500 16.000 13.800 1.00 0.00 C
…
ATOM 1212 NE2 GLN A 152 -1.600 24.200 -1.500 1.00 0.00 N
ATOM 1213 N GLY A 153 1.500 24.700 -6.400 1.00 0.00 N
ATOM 1214 CA GLY A 153 1.100 24.000 -7.600 1.00 0.00 C
ATOM 1215 C GLY A 153 0.300 22.700 -7.500 1.00 0.00 C
ATOM 1216 O GLY A 153 -0.900 22.800 -7.100 1.00 0.00 O
TER 1217 GLY A 153
```

Copyright DNASTAR, Inc. 2022

Example text file for **Add distance and contact restraints**:

DIST 12 HG21 50 HB1 8.1
DIST 14 HA 57 1HE 6.2
DIST 21 HB2 43 HD11 4.0
DIST 124 CA 84 CA 17.4
DIST 36 UNK 120 CA 17.4
CONTACT 33 6
CONTACT 60 29
CONTACT 37 345
CONTACT 109 42

Column requirements in the text file:

Given two residues that contact one another (‘Residue i’ and ‘Residue j’) or two atoms at a distance from one another (‘Atom i’ and ‘Atom j’:

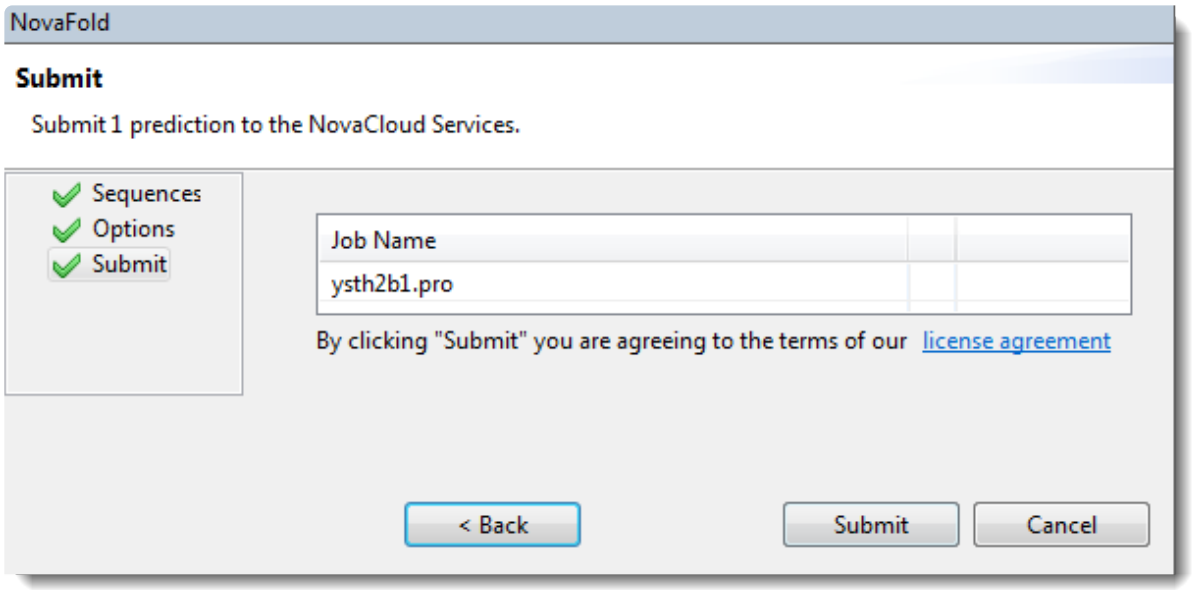
Column name	Description
***Distance* rows contain the following columns from left to right:**	
“DIST” (without quotes)	
Res_No.i	Residue sequence number for Residue i
Atom_type_i	Atom name for contacting atom of Residue i
Res_No_j	Residue sequence number for Residue j
Atom_type_j	Atom name for contacting atom of Residue j
Distance in Angstroms	
***Contact* rows contain the following columns from left to right:**	
“CONTACT” (without quotes)	
Res_No.i	Residue sequence number for Residue i
Res_No.j	Residue sequence number for Residue j

In both cases, UNK can be used in a row to represent an unknown atom.

Copyright DNASTAR, Inc. 2022

NovaFold wizard: Submit

When setting up a NovaFold prediction in Protean 3D's NovaFold wizard, the third screen is Submit. This screen displays the prediction jobs that were set up in the Sequences and Options screens, but have not yet been run.



Click **Submit** to accept the terms of the license agreement and initiate all listed jobs on the NovaCloud web service. Note that term license holders can submit as many predictions as desired, but those predictions will be run in batches of five. Alternatively, press **Back** to return to the Options screen. To remove or add predictions to the list before submitting, click **Back** twice to return to the Sequences screen.

Once a prediction has been started, go to the Predictions view to monitor its progress, and to view or save finished prediction results.

Copyright DNASTAR, Inc. 2022

Run NovaFold AI through Protean 3D

This topic is a general overview on how to set up, run and view a NovaFold AI prediction **through the Protean 3D interface**.

1. If you are not currently logged into NovaCloud Services, open the NovaCloud Services view and log in. If you are already logged in, you can skip ahead and launch the NovaFold AI wizard by doing any of the following:

    - From the Welcome screen, click **Structure Prediction** on the left and then **New protein structure with NovaFold AI** on the right.

    - Choose **Modeling > Protein Structure Prediction > New NovaFold AI Prediction**.

    - From the NovaCloud Services view, press the **NovaFold AI** button.

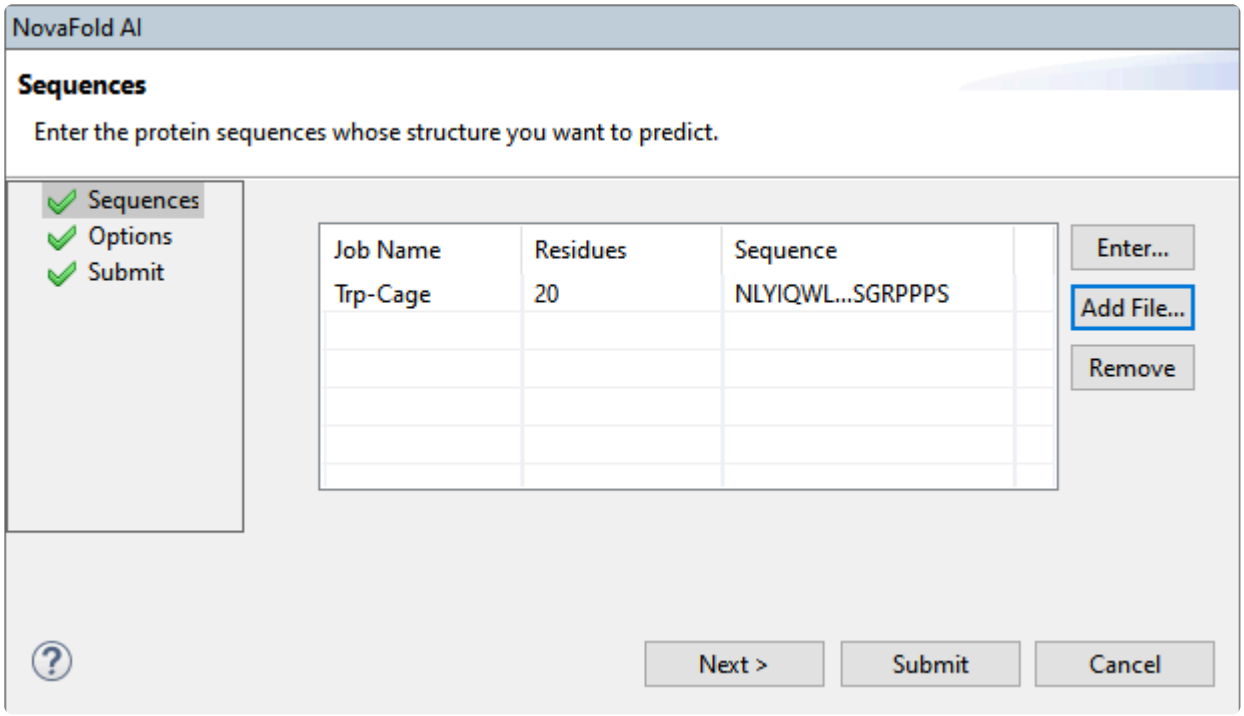
2. Follow the steps in the three NovaFold AI wizard screens: Sequences, Options and Submit.

3. Use **View > NovaCloud Services > Predictions** to monitor run status in the Predictions view. Once the prediction is complete, use the **Status** column link to open and/or download the predicted models.

Copyright DNASTAR, Inc. 2022

NovaFold AI wizard: Sequences

When you predict protein structure with NovaFold AI, the wizard begins with the Sequences screen. This screen primarily consists of a job table with three columns. When populated, the **Job Name** column displays the name of each query protein sequence or its FASTA title tag. The **Residues** column displays the number of residues in the sequence, and the **Sequence** column displays the first part of the amino acid sequence in IUPAC format.



Task	How To
To upload a file containing the protein sequence in FASTA format	Click **Add File** and navigate to the file. The maximum sequence length is 2,000 residues per sequence. Multi-sequence FASTA files are imported as a batch, with each sequence listed on a separate row of the table.
To change the information in any cell of the table	Insert the cursor and type over the existing text.
To type or paste in the sequence	Click the **Enter** button. Type or paste two valid amino acid sequences into the field provided, in FASTA format. If a yellow “error triangle” appears, hover over it to see a tooltip describing the issue (e.g., “sequence must be in FASTA format,” etc.). Press **OK** to add the sequence to the job table. If you wish to enter additional sequences, check the **Add another** box before

Copyright DNASTAR, Inc. 2022

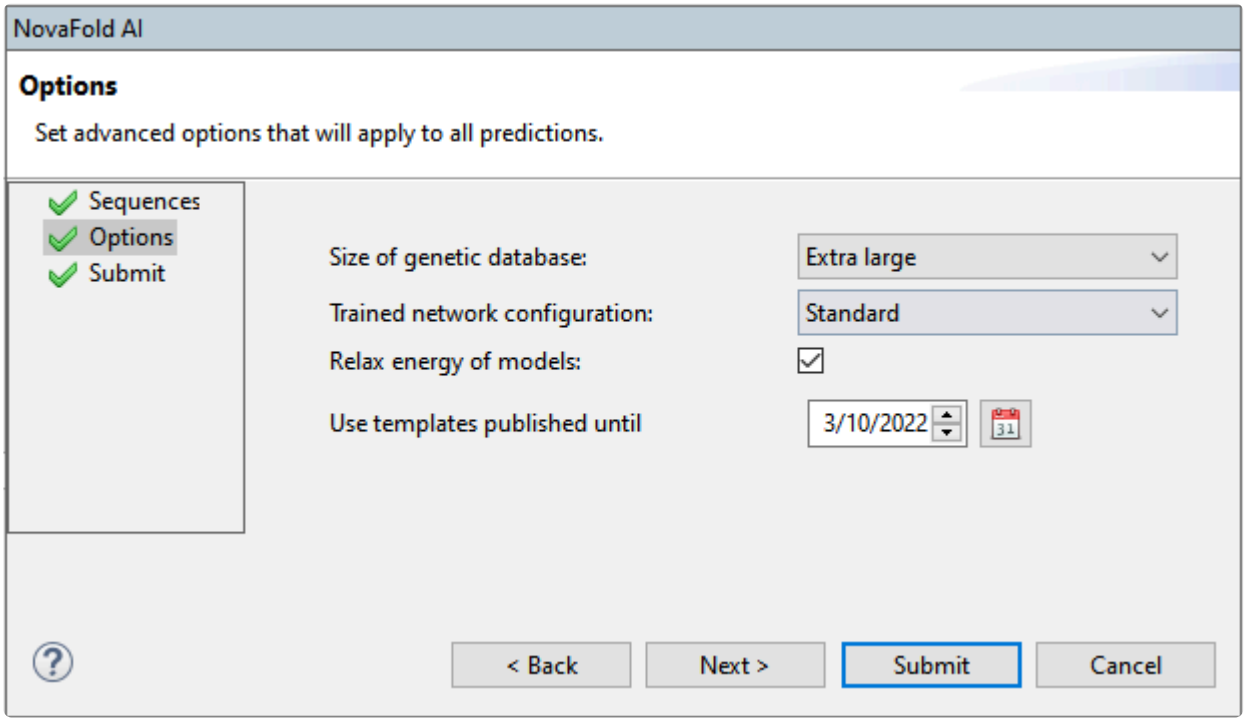
	pressing **OK**.
To delete a job from the job table	Select the job and press **Remove**.

Click **Submit** to submit the job without further customization, or click **Next** to proceed to the Options screen. If you choose **Submit**, you can close this dialog as soon as the **Close** button is enabled.

Copyright DNASTAR, Inc. 2022

NovaFold AI wizard: Options

When you predict a protein structure with NovaFold AI, the second wizard screen is Options. During the prediction step, options specified in this screen are applied to all jobs listed in the Sequences screen.



Edit or enter information in this wizard screen as desired. Or you can accept the defaults as they are, and simply press **Next** to continue to the Submit screen.

- **Size of genetic database** – NovaFold AI constructs a deep multiple sequence alignment as input for the structure prediction process. Use the drop-down menu to choose either:
    - **Extra large** – To search through all 2 billion protein sequences in the full BFD.
    - **Large** – To search a reduced version of the BFD prepared by the AlphaFold group.
- **Trained network configuration** – Use the drop-down menu to choose between:
    - **Standard** – The default model.
    - **Extra ensembling** – The model used during the CASP14 competition. It requires 8x more computation but yields only a minor gain in modeling accuracy.
    - **Extra metrics** – A model tuned to provide a pairwise confidence measure (predicted C-alpha differences; also used by NovaFold/I-TASSER). This model is less accurate than the standard model.

Copyright DNASTAR, Inc. 2022

- **Relax energy of models** – Check the box if you wish to minimize the energy of the prediction. Leave the box unchecked to use the raw coordinates from the AI model.

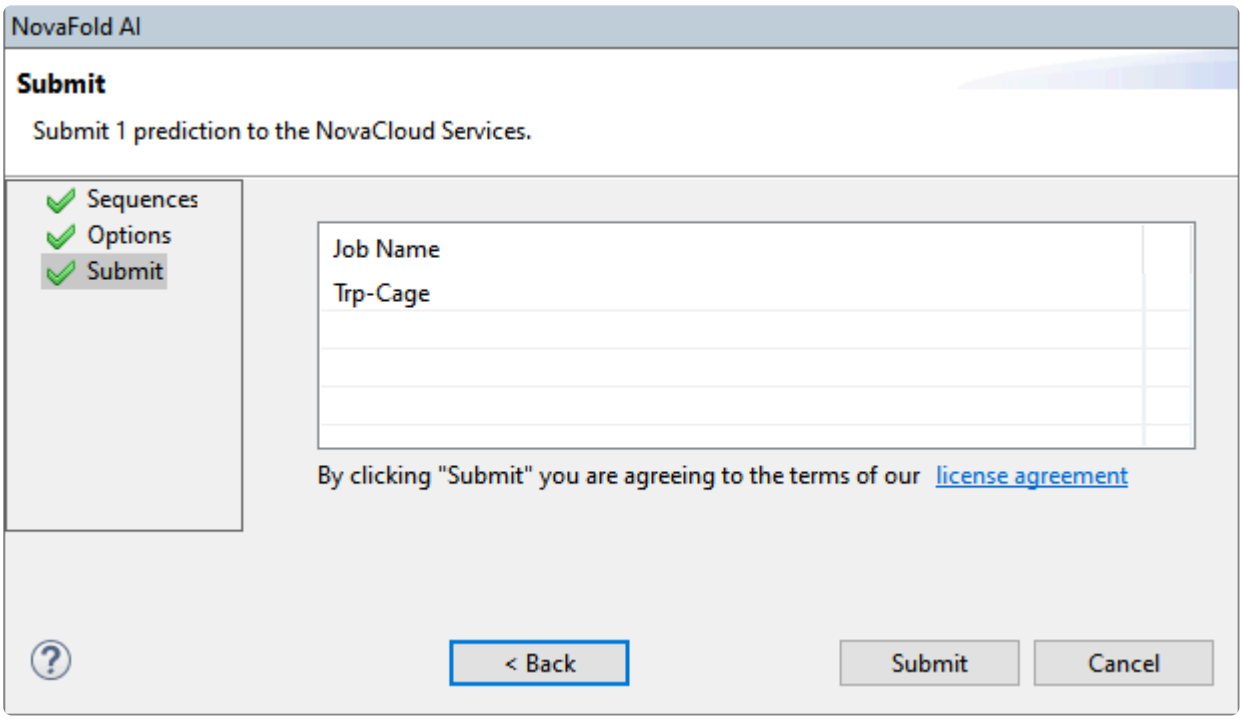
- **Use templates published until** – Today's date is shown by default. To limit templates to those published before an earlier date, specify the date by using the up/down triangles or by clicking on the **Choose a date** tool (). This option was designed for DNASTAR's internal benchmarking purposes. It can be used to reproduce the CASP14 experiment with a contemporary set of the Protein Data Bank.

Click **Next** to proceed to the Submit screen, or **Back** to return to the Sequences screen.

Copyright DNASTAR, Inc. 2022

NovaFold AI wizard: Submit

When you predict protein structures with NovaFold AI, the third wizard screen is Submit. This screen displays the prediction jobs that were set up in the Sequences and Options screens, but have not yet been run.



Click **Submit** to accept the terms of the license agreement and initiate all listed jobs on the NovaCloud web service. Note that term license holders can submit as many predictions as desired, but those predictions will be run in batches of five. Alternatively, press **Back** to return to the Options screen. To remove or add predictions to the list before submitting, click **Back** twice to return to the Sequences screen.

Once you have submitted the run, use **Close** to close the wizard.

Monitoring progress and viewing prediction results:

Once a prediction has been started, go to the Predictions view to monitor its progress, and to save and open finished prediction results.

Copyright DNASTAR, Inc. 2022

Run NovaFold Antibody through Protean 3D

This topic is a general overview on how to set up, run and view a NovaFold Antibody prediction **through the Protean 3D interface**. If you are instead running NovaFold Antibody locally on a Linux or Macintosh machine, see Run Nova applications locally. If you are running NovaFold Antibody online, see Run Nova applications through the DNASTAR website.

1. If you are not currently logged into NovaCloud Services, open the NovaCloud Services view and log in. If you are already logged in, you can skip ahead and launch the NovaFold Antibody wizard by doing any of the following:

    - From the Welcome screen, click **Structure Prediction** on the left and then **New antibody structure with NovaFold Antibody** on the right.

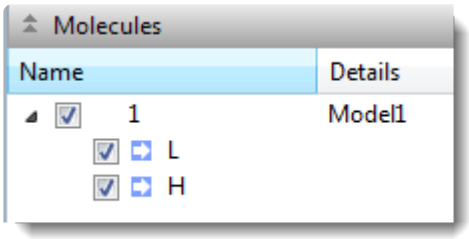
    - Choose **Modeling > Antibody Structure Prediction > New NovaFold Antibody Prediction**.
    - From the NovaCloud Services view, press the **NovaFold Antibody** button.

2. Follow the steps in the three NovaFold Antibody wizard screens: Sequences, Options and Submit.

3. Use **View > NovaCloud Services > Predictions** to monitor run status and open or download predictions from within the Predictions view. NovaFold Antibody results can be opened on any computer running Protean 3D 12.1 and higher; a NovaFold license is not required to view results.

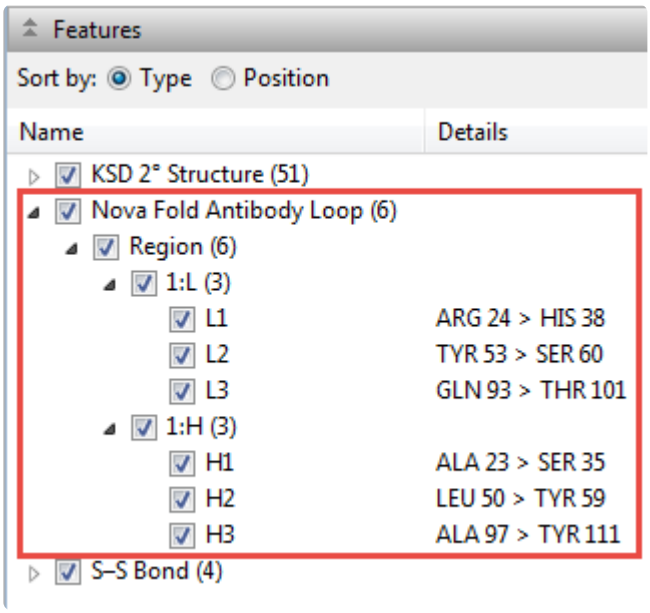
Open results appear as a new project tab in Protean 3D. Unlike NovaFold and NovaDock, NovaFold Antibody has no specialized Report view. Instead, all Protean 3D views and panels are enabled, as for any structure file. However, the NovaFold Antibody project differs from a typical Protean 3D in two ways:

- In the Molecules section, the light and heavy chains are designated as **L** and **H**.



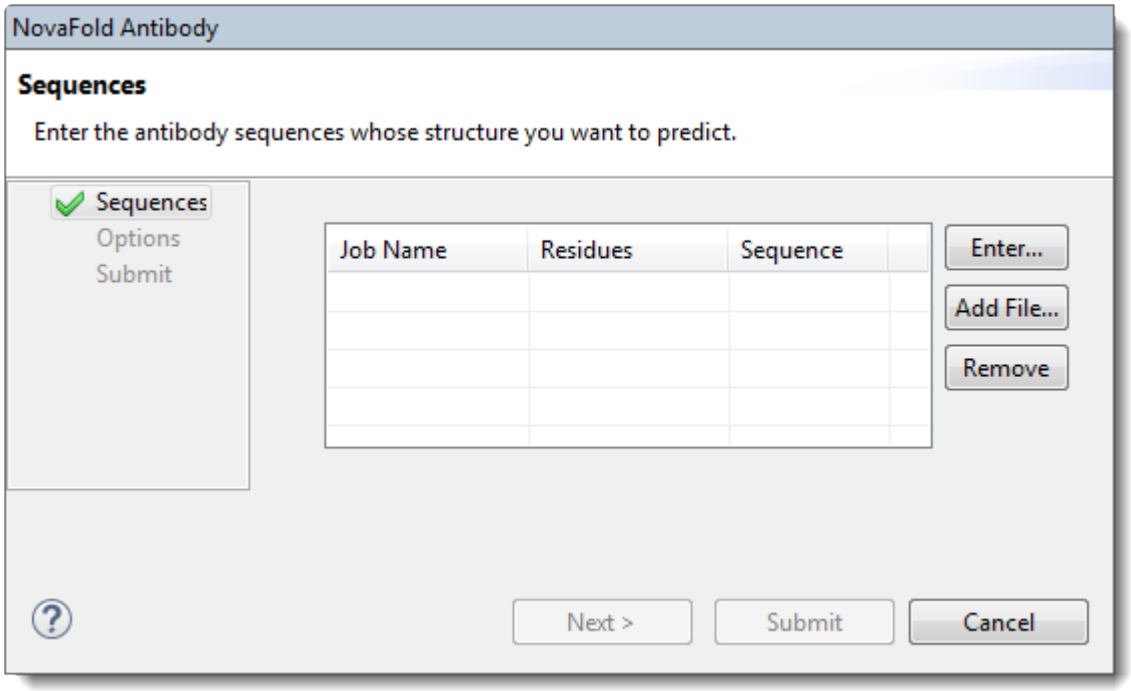
- In the Style section, a new item called **NovaFold Antibody Loop** is available By using the corresponding expand triangles, you can expand this item and its subtopics to reveal the three light (L1, L2, L3) and/or three heavy (H1, H2, H3) chain annotations.

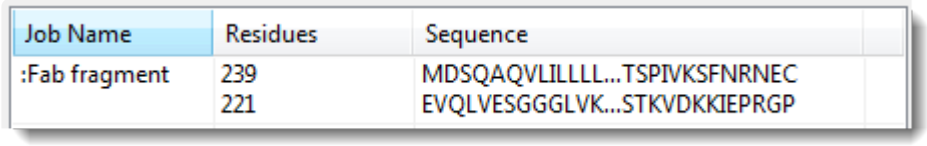
Copyright DNASTAR, Inc. 2022

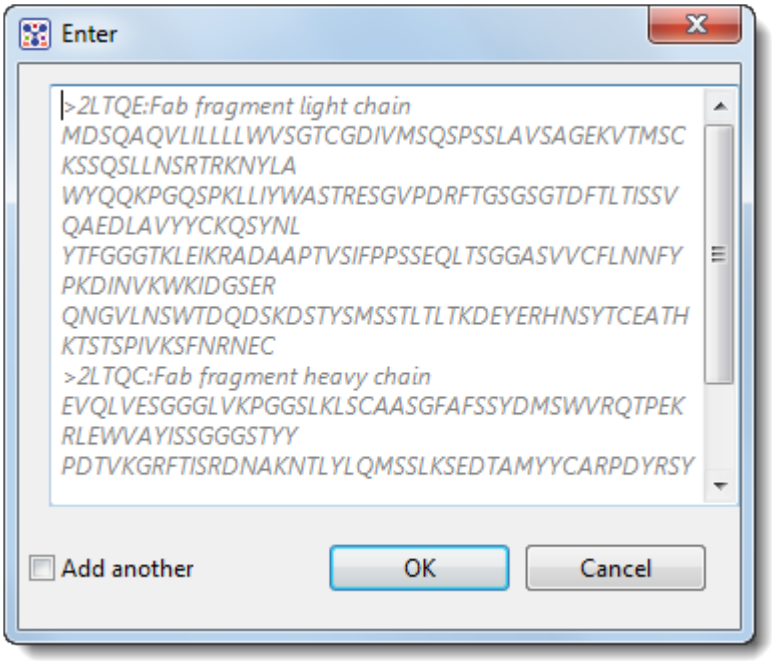

Features
Sort by: Type Position
Name
Details
KSD 2° Structure (51)
Nova Fold Antibody Loop (6)
Region (6)
1:L (3)
L1
ARG 24 > HIS 38
L2
TYR 53 > SER 60
L3
GLN 93 > THR 101
1:H (3)
H1
ALA 23 > SER 35
H2
LEU 50 > TYR 59
H3
ALA 97 > TYR 111
S–S Bond (4)

NovaFold Antibody wizard: Sequences

When you predict antibody interactions with NovaFold Antibody, the wizard begins with the Sequences screen. This screen primarily consists of a job table with three columns. When populated, the **Job Name** column displays the name of each query protein sequence or its FASTA title tag. The **Residues** column displays the number of residues in the sequence, and the **Sequence** column displays the first part of the amino acid sequence in IUPAC format.



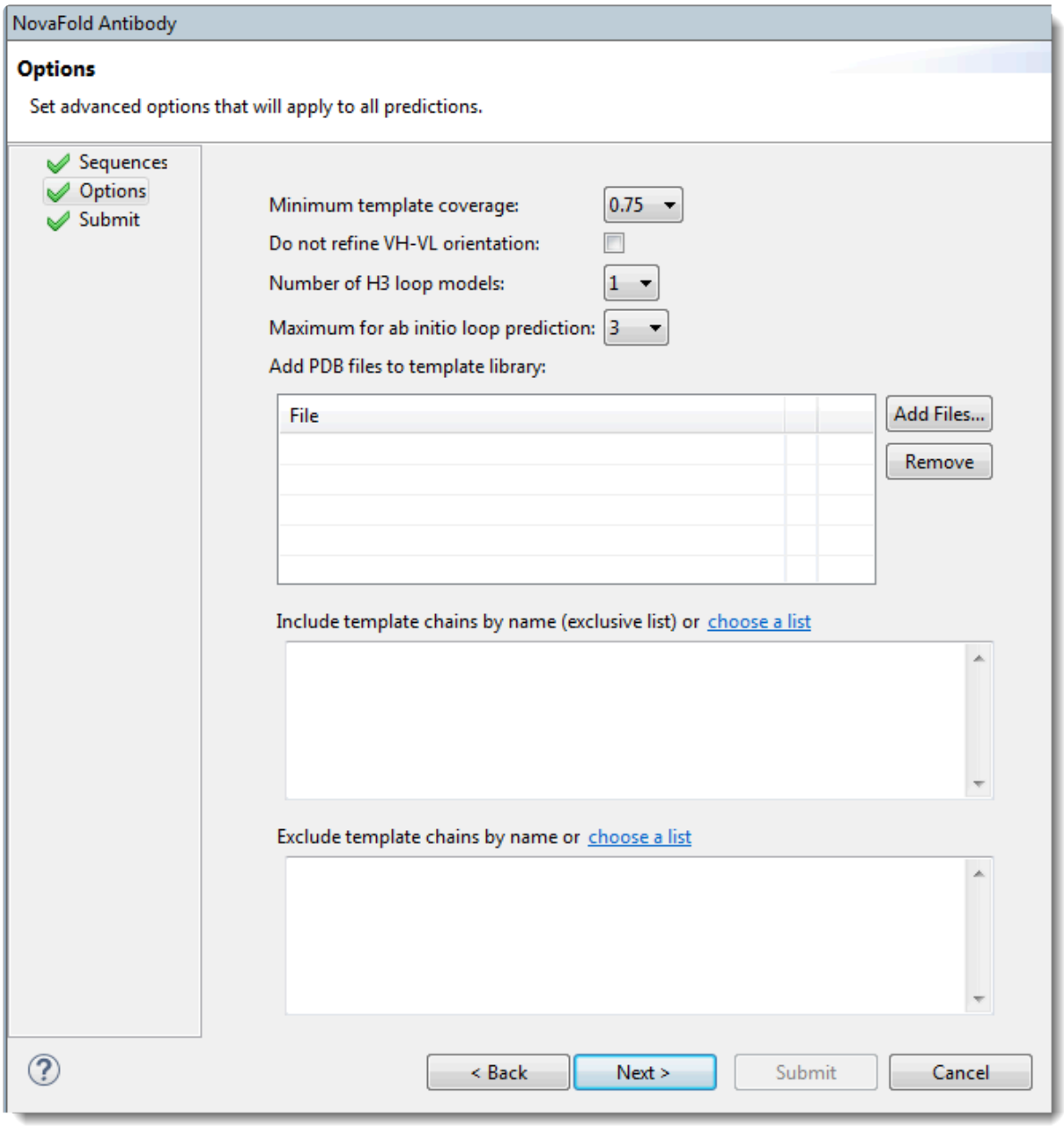
Task	How To
To upload a file containing the light chain, heavy chain, or both in FASTA format	Click **Add File** and navigate to the file. The maximum sequence length is 2,000 residues per sequence. Multi-sequence FASTA files are imported as a batch, with each chain's sequence listed on a separate row of the table. 
To change the information in any cell of the table	Insert the cursor and type over the existing text.
To type or paste in the sequence	Click the **Enter** button. Type or paste two valid amino acid sequences into the field provided, in FASTA format. If a yellow "error triangle" appears, hover over it to see a

chains	tooltip describing the issue (e.g., “sequence must be in FASTA format,” etc.). Press **OK** to add the sequence to the job table. If you wish to enter additional sequences, check the **Add another** box before pressing **OK**. 
To delete a job from the job table	Select the job and press **Remove**.

Click **Submit** to submit the job without further customization, or click **Next** to proceed to the Options screen. If you choose **Submit**, you can close this dialog as soon as the **Close** button is enabled.

NovaFold Antibody wizard: Options

When you predict antibody interactions with NovaFold Antibody, the second wizard screen is Options. During the prediction step, options specified in this screen are applied to all jobs listed in the Sequences screen.



Edit or enter information in this wizard screen as desired. Or you can accept the defaults as they are, and simply press **Next** to continue to the Submit screen.

- **Minimum template coverage** – To ensure the selected template exceeds a particular fractional coverage, you may specify a non-zero value using the drop-down menu. This compels NovaFold Antibody to locate a template with both variable and constant domains.

Copyright DNASTAR, Inc. 2022

> **Note**: If ‘0’ were an option here, NovaFold Antibody would select a template framework based on its statistical significance to the query sequence, without considering coverage criteria. The threader would pick the template with the highest (log-likelihood score)/(background log-likelihood) for matching a sequence. If a single domain matched, only that domain would be proposed as a model. Therefore, the most significant template would likely not cover the entire query.

- **Do not refine VH-VL orientation** – By default, NovaFold Antibody optimizes the rigid-body orientation between the light and heavy antibody chains to remove atomic clashes if they were introduced during the modeling process. To skip the optimization step, check this box.

- **Number of H3 loop models** – The H3 loop is the generally the hardest region of the antibody structure to predict. NovaFold Antibody offers a template-based approach that uses a machine learning model to choose the best templates for the H3 loop. To specify how many results models to output, choose a value from the drop-down menu.

- **Maximum for ab initio loop prediction** – To specify a cutoff for switching from *ab initio* "distance guided" prediction of the H3 loop to a template-based prediction. Example: The default setting of ‘3’ means loops of length 3 or shorter would use the *ab initio* methodology, while loops longer than 3 residues would be built with the template-based approach.

- **Add PDB files to template library** – Press **Browse** to nominate a local 3D structure to use as a template in the modeling prediction, along with other templates selected by NovaFold Antibody. Select a text file representing a single protein chain in standard PDB format. If you need to remove a file from the list, select it and press **Remove**.

- **Include template chains by name (exclusive list)** – To define one or more chains from a PDB structure as the template(s) to use in the modeling prediction, type the information into the text area. Each line in the file should represent one template using the format [PDB ID][CHAIN ID][HEAVY/LIGHT]. The PDB ID is four characters, and starts with a number, followed by three letters or numbers. The CHAIN ID is case sensitive and is typically a single character. Examples:

  1IGT:A Heavy
  1IGT:A Light

  Alternatively, press **choose a list** to open a browser where you can navigate to a local file containing this information.

- **Exclude template chains by name** – To exclude certain templates from the library (i.e., to prevent these templates from being considered) during structure prediction. Templates can be excluded by name, using the following format: [PDB ID][CHAIN ID]. Example: 1wor:A. A percent sequence identity can be specified at the end, if desired. For example, 1wor:A 70 would specify a 70% sequence identity cutoff. If no number is specified, the percent sequence identity is assumed to be 100%. An asterisk (*) may be used to designate any chain in the PDB file. For example, 1wor:*.

Copyright DNASTAR, Inc. 2022

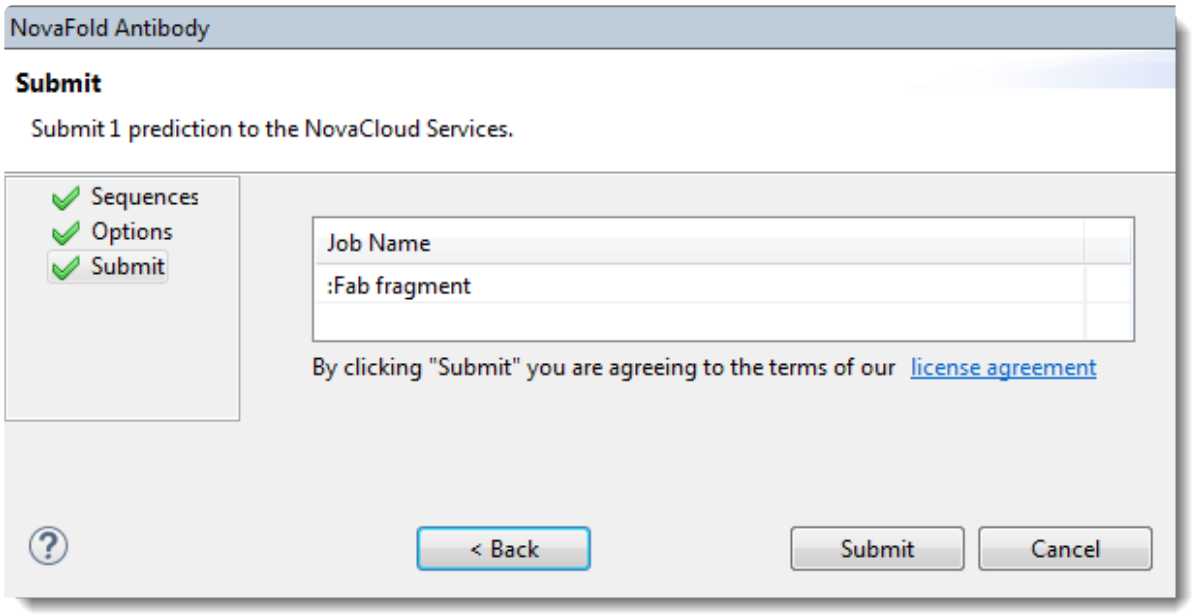
Alternatively, press **choose a list** to open a browser where you can navigate to a local file containing this information.

Click **Next** to proceed to the Submit screen, or **Back** to return to the Sequences screen.

Copyright DNASTAR, Inc. 2022

NovaFold Antibody wizard: Submit

When you predict antibody interactions with NovaFold Antibody, the third wizard screen is Submit. This screen displays the prediction jobs that were set up in the Sequences and Options screens, but have not yet been run.



Click **Submit** to accept the terms of the license agreement and initiate all listed jobs on the NovaCloud web service. Note that term license holders can submit as many predictions as desired, but those predictions will be run in batches of five. Alternatively, press **Back** to return to the Options screen. To remove or add predictions to the list before submitting, click **Back** twice to return to the Sequences screen.

Monitoring progress and viewing prediction results:

Once a prediction has been started, go to the Predictions view to monitor its progress, and to view or save finished prediction results.

Unlike NovaFold and NovaDock, NovaFold Antibody does not result in a specialized version of the Report view. Instead, the NovaFold Antibody prediction results (*.antibody*) file can be viewed and analyzed in Protean 3D like any other structure. The Structure view and Features section facilitate examination of antibody features, including the six annotated CDR loops for light and heavy chains.

Copyright DNASTAR, Inc. 2022

Run NovaDock through Protean 3D

This topic is a general overview on how to set up, run and view a NovaDock prediction **through the Protean 3D interface**. If you are instead running NovaDock locally on a Linux or Macintosh machine, see Run Nova applications locally. If you are running NovaDock online, see Run Nova applications through the DNASTAR website.

1. If you are not currently logged into NovaCloud Services, open the NovaCloud Services view and log in. If you are already logged in, you can skip ahead and launch the NovaDock wizard by doing any of the following:

    - From the Welcome screen, click **Protein Docking** on the left and then **New protein-protein complex with NovaDock** on the right.

    - Choose **Modeling > Protein-Protein Docking > New NovaDock Prediction**.
    - From the NovaCloud Services view, press the **NovaDock** button.

2. Follow the steps in the three NovaDock wizard screens: Structures, Options and Submit.

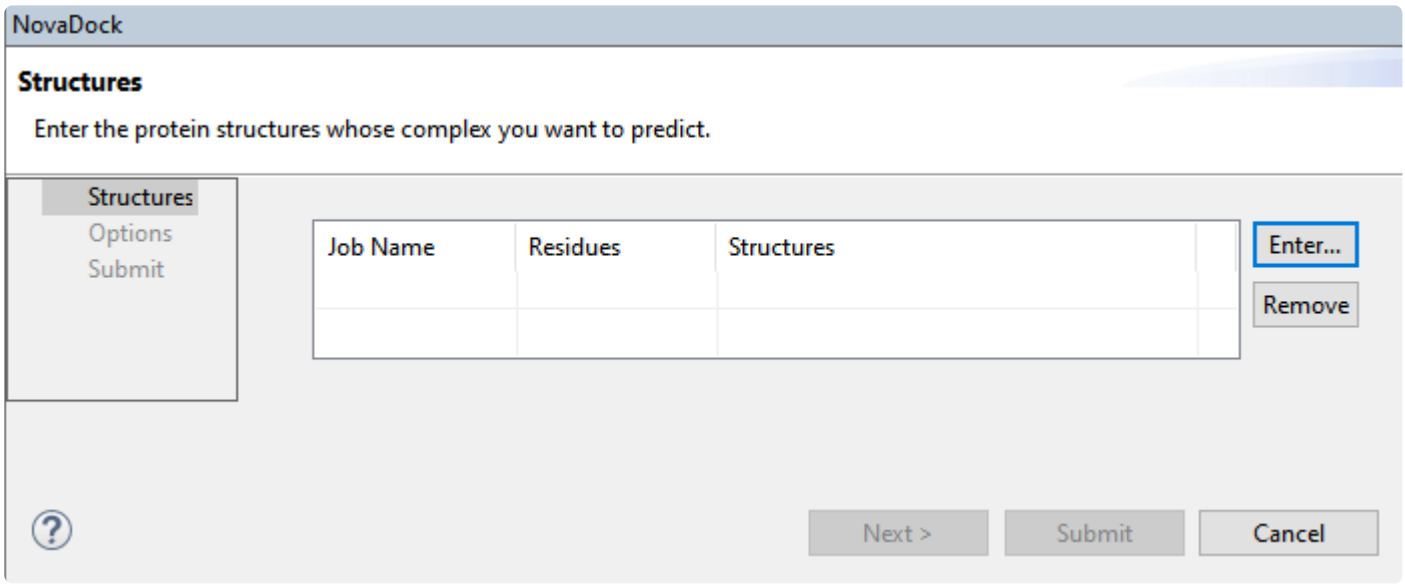
3. Use **View > NovaCloud Services > Predictions** to monitor run status and open or download predictions from within the Predictions view. NovaDock results can be opened on any computer running Protean 3D 14.0 and higher; a NovaDock license is not required to view results.

Open results appear as a new project tab in Protean 3D with the NovaDock Report active. To learn which views and panels are enabled with this type of project, see Protean 3D Interface.

Copyright DNASTAR, Inc. 2022

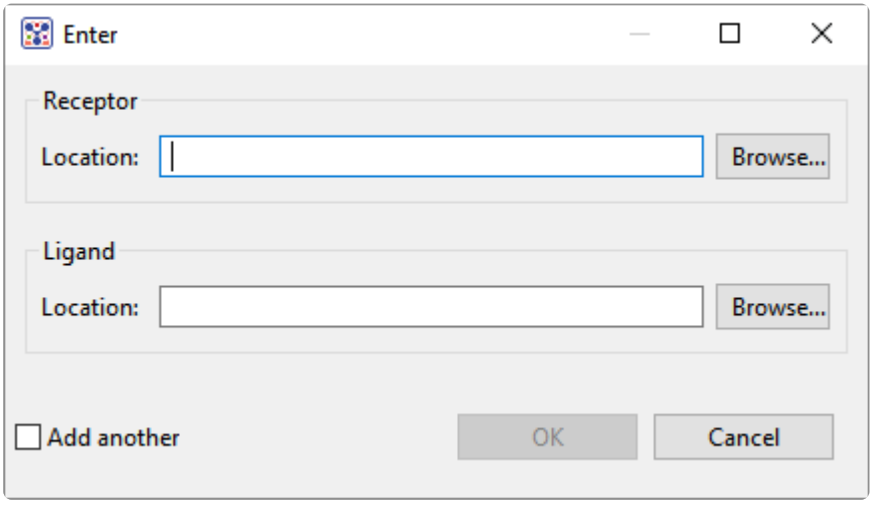
NovaDock wizard: Structures

When you predict protein docking with NovaDock, the first wizard screen is Structures. This screen prompts you to specify structure files to use as the receptor and as the ligand(s) in a docking simulation.



To enter the antibody and antigen files, click **Enter**.

In the **Receptor** section, use the **Browse** button to browse to the antigen file in FASTA format. In the **Ligand** section, use the **Browse** button to browse to the antibody file in FASTA format. The maximum sequence length in both cases is 2,000 residues.



When you are finished, click **OK**. If you wish to open the popup again to add additional files, check the **Add another** box before pressing **OK**. Each imported structure appears on a separate row of the table. Multi-sequence FASTA files are imported as a batch, with each sequence listed on a separate row.

Here are descriptions of each column in the now-populated table:

Copyright DNASTAR, Inc. 2022

- **Job Name** displays the name of each query protein sequence or its FASTA title tag. Information in this column can be changed by inserting the cursor and type over the existing text.
- **Residues** displays the number of residues in the sequence.
- **Structures** displays the path and file name of each structure file.

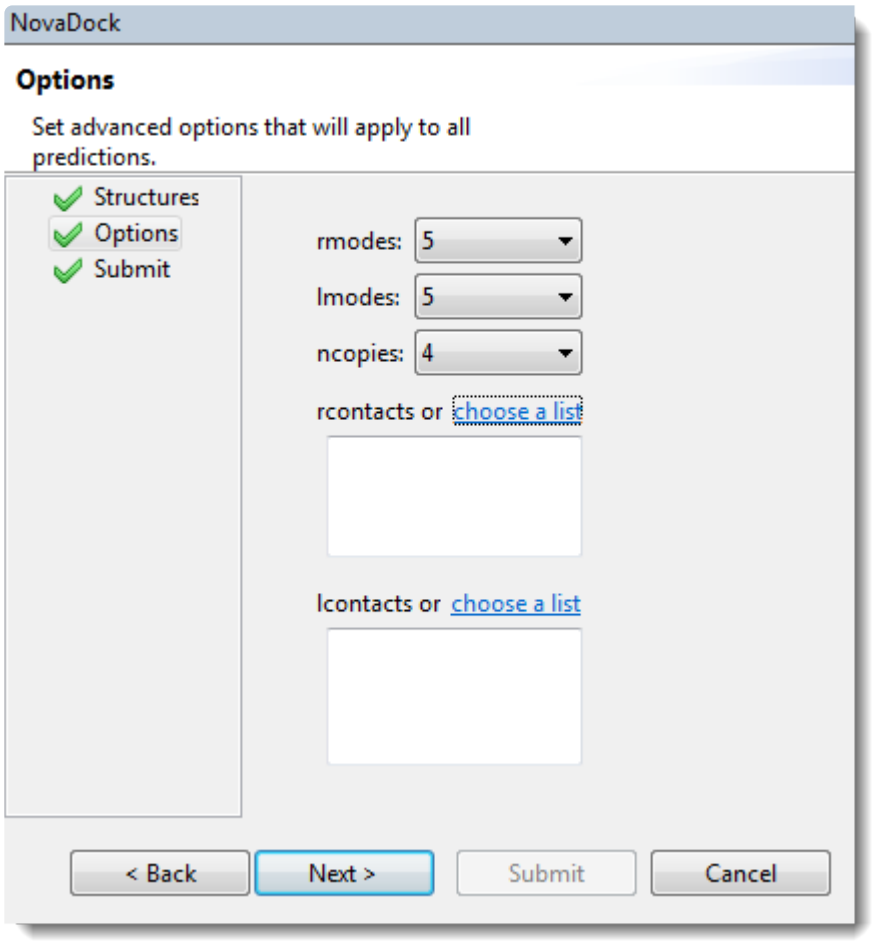
If you want to delete a job from the table, select the job and press **Remove**.

Click **Submit** to submit the job without further customization, or click **Next** to proceed to the Options screen. If you choose **Submit**, you can close this dialog as soon as the **Close** button is enabled.

Copyright DNASTAR, Inc. 2022

NovaDock wizard: Options

When you predict protein docking with NovaDock, the second wizard screen is Options. This screen allows you to edit or specify prediction options. During the prediction step, options specified in this screen are applied to all docking jobs listed in the Structures screen.



Edit or enter information in this wizard screen as desired. Or you can accept the defaults as they are, and simply press **Next** to continue to the Submit screen.

- **rmodes** – Number of normal modes of motion to explore for the receptor. We suggest using the default (5) to model large, concerted conformational changes and using a higher number (e.g. 20) to model more localized fluctuations, such as within antibody interfaces.

- **lmodes** – Number of normal modes of motion to explore for the ligand. We suggest using the default (5) to model large, concerted conformational changes and using a higher number (e.g. 20) to model more localized fluctuations, such as within antibody interfaces.

- **ncopies** – Number of copies of each swarm point to run. The default is 4.

- **rcontacts** – To list proposed contact residues for the receptor. This causes NovaDock to focus

Copyright DNASTAR, Inc. 2022

docking around the specified contacts, reducing computation time. Each line represents one contact residue using the format: [CHAIN ID][RESIDUE NUMBER]

- ◦ The CHAIN ID is case sensitive and is typically a single character.
  - ◦ The RESIDUE NUMBER is an integer and may include an insertion code.

  Examples:

  R 15
  R 15A

  As an alternative to typing in proposed contact residues, you can click the “choose a list” link to navigate to a file containing a list of contacts.

- **Icontacts** – To list proposed contact residues for the ligand. Each line represents one contact residue using the format: [CHAIN ID] [RESIDUE NUMBER]

- ◦ The CHAIN ID is case sensitive and is typically a single character.
  - ◦ The RESIDUE NUMBER is an integer and may include an insertion code.

  Examples:

  L 15
  L 15A

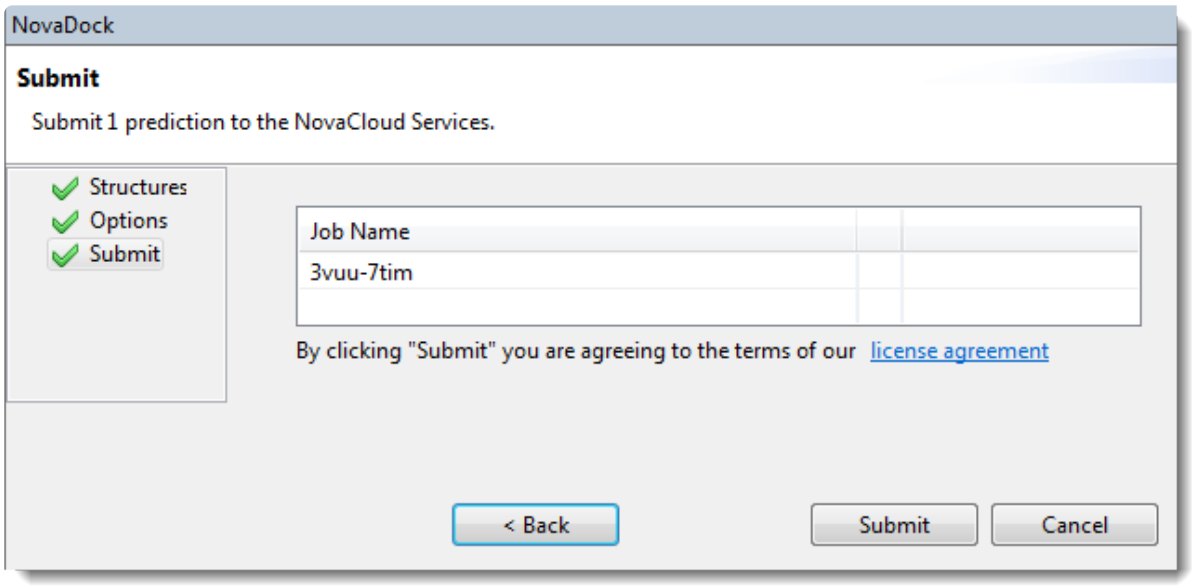
  As an alternative to typing in proposed contact residues, you can click the “choose a list” link to navigate to a file containing a list of contacts.

Click **Next** to proceed to the Submit screen, or **Back** to return to the Structures screen.

Copyright DNASTAR, Inc. 2022

NovaDock wizard: Submit

The Submit screen is the third of three screens in the NovaDock prediction wizard and displays the prediction jobs that were set up in the Structures and Options screens, but have not yet been run.



Click **Submit** to accept the terms of the license agreement and initiate all listed jobs on the NovaCloud web service. Note that term license holders can submit as many predictions as desired, but those predictions will be run in batches of five. Alternatively, press **Back** to return to the Options screen. To remove or add predictions to the list before submitting, click **Back** twice to return to the Structures screen.

Once a prediction has been started, go to the Predictions view to monitor its progress, and to view finished prediction results.

Copyright DNASTAR, Inc. 2022

Try it! – Phage display assembly, translation and prediction

Phage display panning is a tool used for the discovery of novel ligands against various targets of interest. In phage display panning, you will typically start with a batch of 96 _.abi. read files, each corresponding to one well of a 96-well ELISA assay. At the end of the workflow, you will be able to view and analyze predicted 3D antibody structures and see how they dock to the corresponding antigen structure. The phage display workflow utilizes four Lasergene applications, but is fast and easy to learn and do.

Before beginning the tutorial, let's start by reviewing the laboratory steps leading up to the *.abi* sequence reads. The technique (summarized in this NCBI article) begins in the laboratory with these three steps:

1. Create a library of plasmids containing variable CDR loops

2. Make a single-chain variable fragment (scFv) for each plasmid in the library. An scFv is a fusion protein that includes the variable regions of heavy and light chains connected by a short linker peptide about 10-25 residues in length. Leave at least a portion of vector in each sequence.

3. Screen each plasmid against a specific antigen. This screening is typically done using a 96-well ELISA assay. The output is one sequence read file for each well, typically in *.abi* format.

Now that the laboratory portion is complete, which of the tested plasmids have an antibody sequence that is bound to an antigen? The answer can be found quickly and easily using Lasergene's integrated phage display workflow. The tutorial is divided into four parts, with each part taking place in a different Lasergene application.

Part A. Use SeqMan Ultra to *de novo* assemble reads, then trim down to the target antibody translation sequence in the correct reading frame
Part B. Use SeqNinja to batch translate the sequences
Part C. Use MegAlign Pro to align sequences and create a phylogenetic tree
Part D. Use Protean 3D and the NovaApps to create a homology model of the sequence and simulate docking of the predicted antibody structures to the target epitope

Start with Part A: Assembly and trimming in SeqMan Ultra.

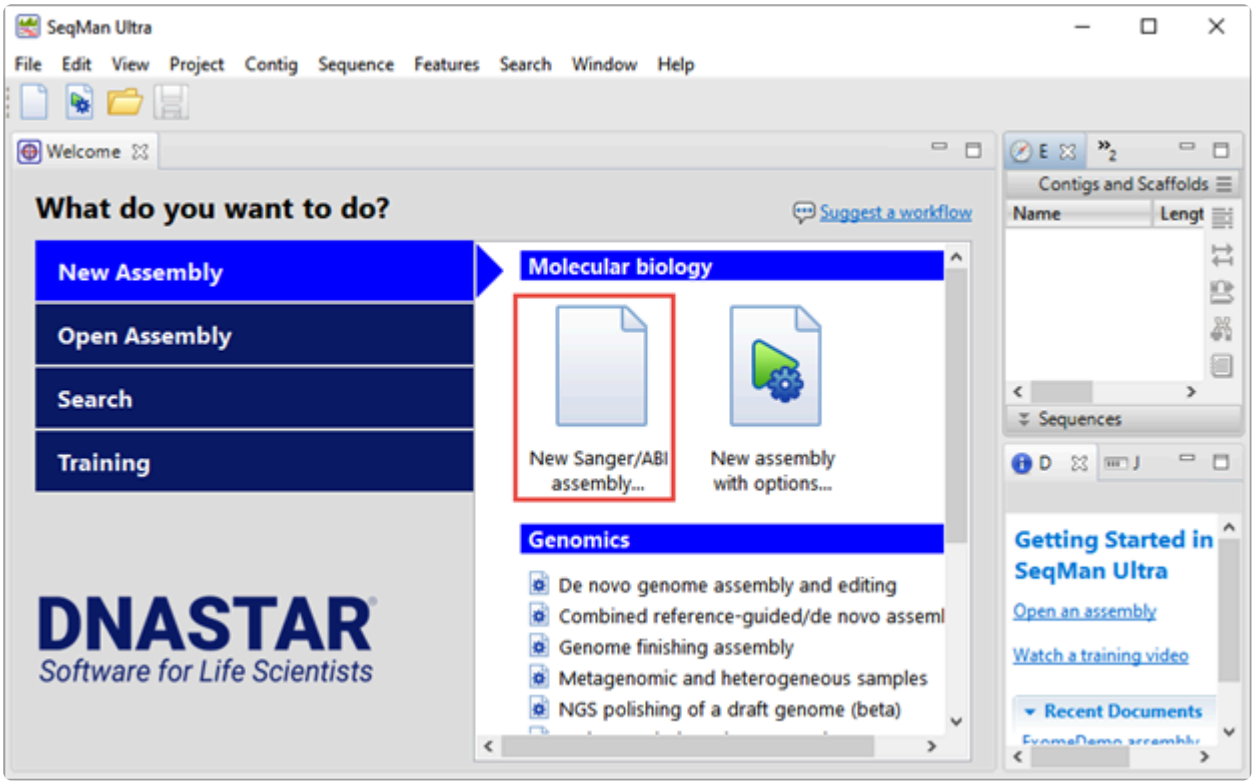
Copyright DNASTAR, Inc. 2022

Part A. Assembly and trimming in SeqMan Ultra

This section of the tutorial describes the procedure for working with *.abi* files, by far the most common file type used in this workflow. (NGS files are handled differently; please contact us for instructions.) In this section, you will use SeqMan Ultra to de novo assemble reads, then trim down to the target antibody translation sequence in the correct reading frame

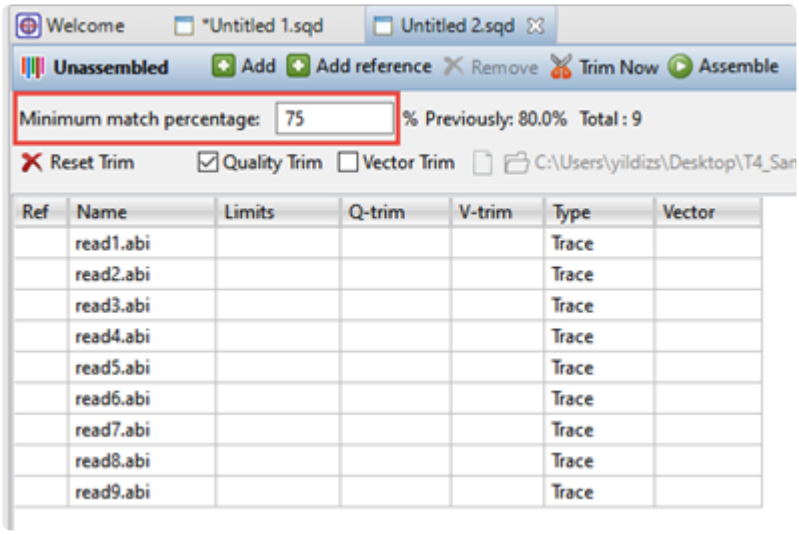
1. Click here to download the tutorial data. Extract it to any convenient location (e.g., your desktop).

2. Launch **SeqMan Ultra** and click on **New Sanger/ABI assembly**.



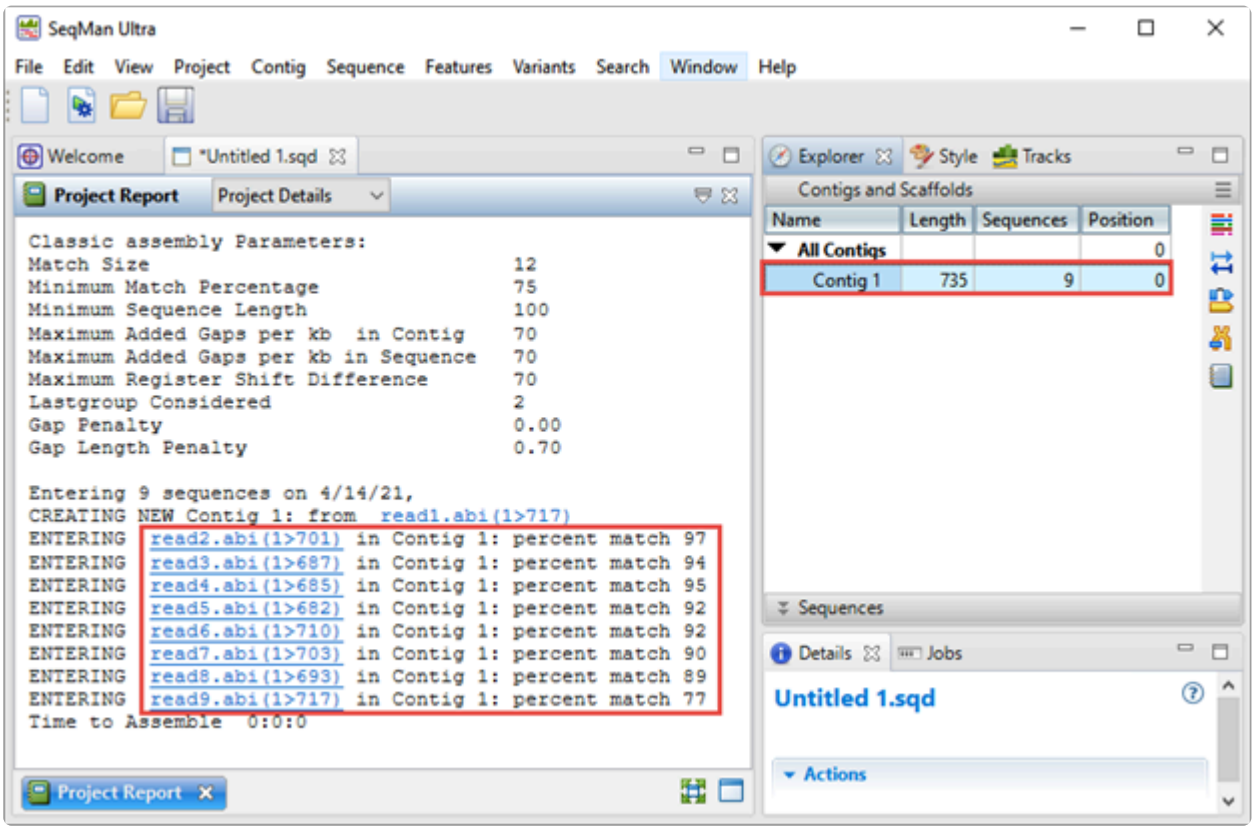
3. Use the **Add** button (image below) to add the *.abi* sequence reads.

4. For this specific data set to assemble as a single contig, you will need to change the **Minimum match percentage** to **75**.



5. Click the **Assemble** button to *de novo* assemble the reads.

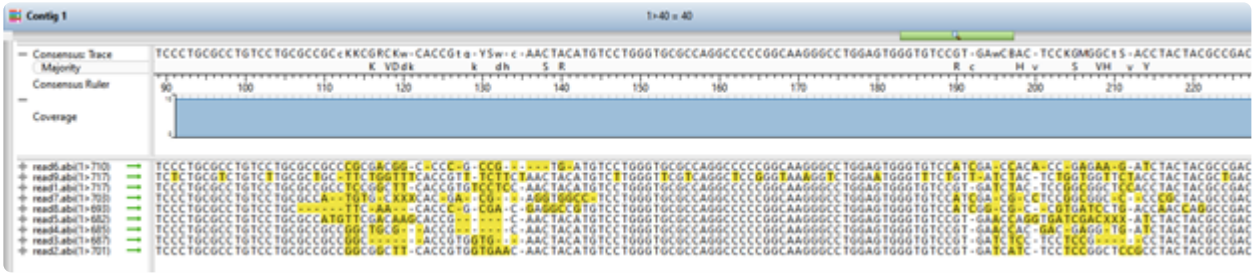
The Explorer panel on the right (image below) shows any resulting contigs; ideally all sequences will be placed in a single contig. The Project Report view on the left shows the “percent match” statistics, which can be used to judge the quality of the assembly. Each “percent match” row corresponds with a specific well from the 96-well plate.



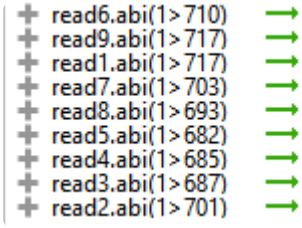
In the next steps, you will use SeqMan Ultra to trim down to the target antibody translation sequence in the correct reading frame.

6. Double-click on the contig in the Explorer panel to open its Alignment view.

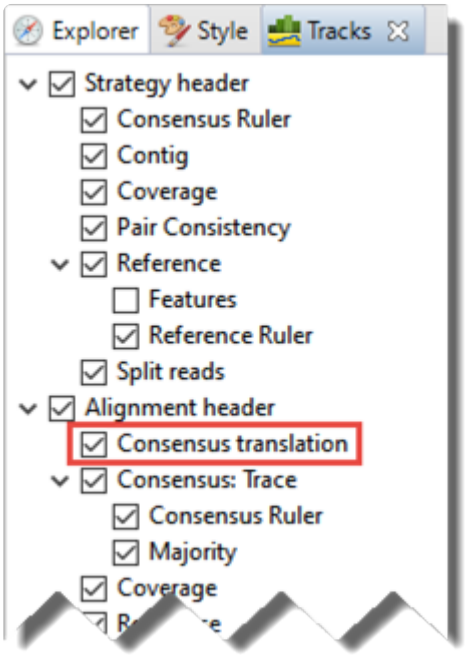
   By default, variances in a germline antibody--usually from a deliberate mutation--will appear highlighted in yellow. The three highlighted areas along the sequence--two of which are shown below--denote the CDR (Complimentary Determining Region) heavy & light chains. These regions allow for different affinities for an antibody targeting a specific antigen.



   In this example, the right-facing green arrows in each row denote that each read is already in the forward orientation. (Later, we'll explain how to automatically correct any reads that happen to be in the reverse orientation).



7. To display the translation track, first click on the **Tracks** tab in the upper right of the SeqMan Ultra window. Check the box next to **Consensus translation**.



Copyright DNASTAR, Inc. 2022

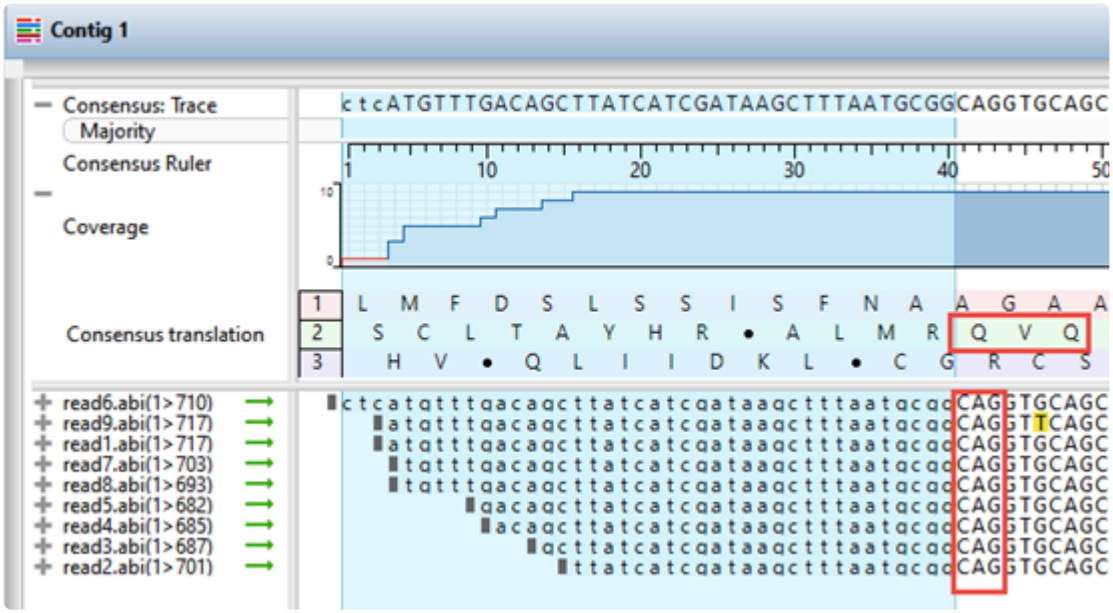
By default, the Alignment view shows the **Consensus translation** for all three reading frames in the top strand (see middle portion of the image below).

In the next steps, you will manually trim bases to remove everything except the area between the start and stop codons of the target antibody translation sequence.

8. In the **Consensus translation** track, look for the sequence “QVQ.” This indicates the “start” location. The nucleotide sequence corresponding to the first “Q” is “CAG.”

> **Note:** If you need help finding the “start” location, press the **Search Alignment** tool ( ) to display the Search bar. Then make the following choices and use the green arrows to locate all instances of “CAG.”
>
> Find: In Consensus | CAG | Allow ambiguity
>
> Find the instance that is associated with “QVQ” in the **Consensus translation** track.

9. In the **Consensus: Trace** track at the top of the view, insert the cursor at the junction just to the left of the “C” in “CAG”. Holding down the left mouse button, drag your mouse all the way to the left and release the button. All sequence to the left of “C” is now selected.



10. Press the **Delete** key (or right-click and choose **Delete**) to delete the unwanted portion of sequence.

11. Use the same technique to delete all sequence to the right of the “stop” codon; this location varies by project.

The sequences are now identical in length in the correct reading frame and are ready for export.

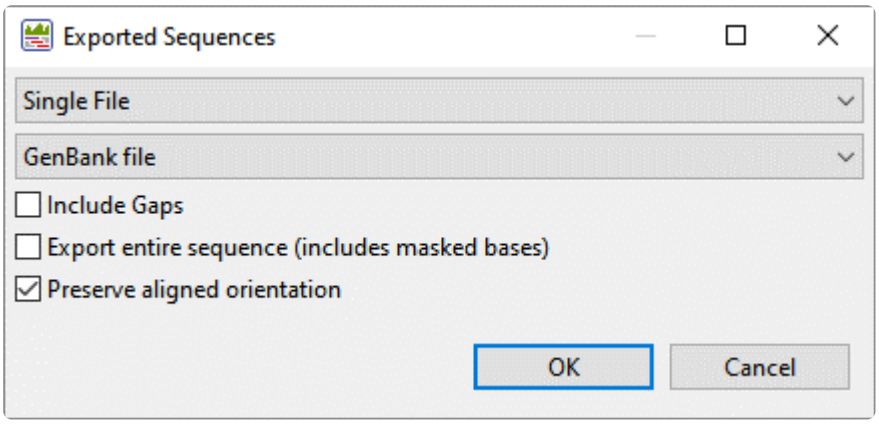
Copyright DNASTAR, Inc. 2022

12. Export the sequences as follows. First, use **Edit > Select All** to select all of the trimmed sequences. Then choose **Contig > Export Subsequences**.

13. In the popup dialog, choose **GenBank file** from the second drop-down menu.

> **Note:** If you had kept the default FASTA format, the sequences would have a semicolon added to their labels. A later step in this workflow does not support labels with semicolons, so you would have had to remove those semicolons manually. Choosing **Genbank file** in this step circumvents this.

14. Check **Preserve aligned orientation** to automatically reverse-complement reads that are in the reverse orientation. Leave everything else at the default settings and press **OK**.



When prompted, choose a file name and location for the exported sequences.

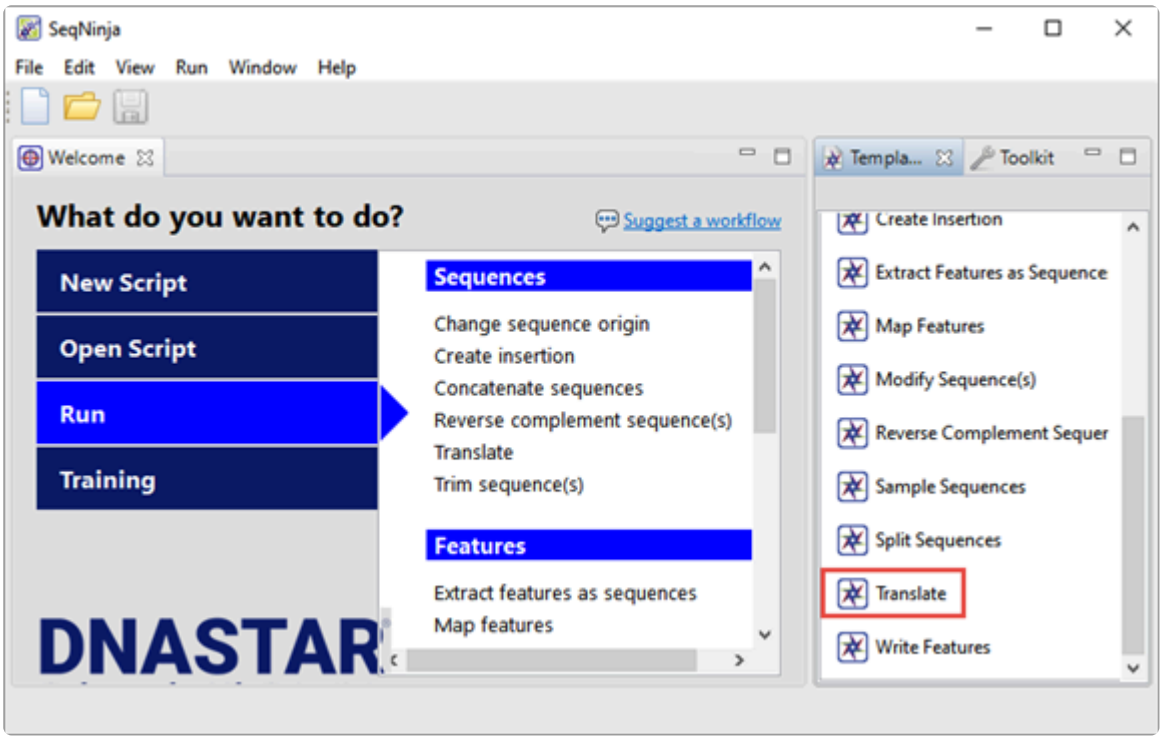
15. Close SeqMan Ultra. There is no need to save the project.

Proceed to Part B: Translation in SeqNinja.

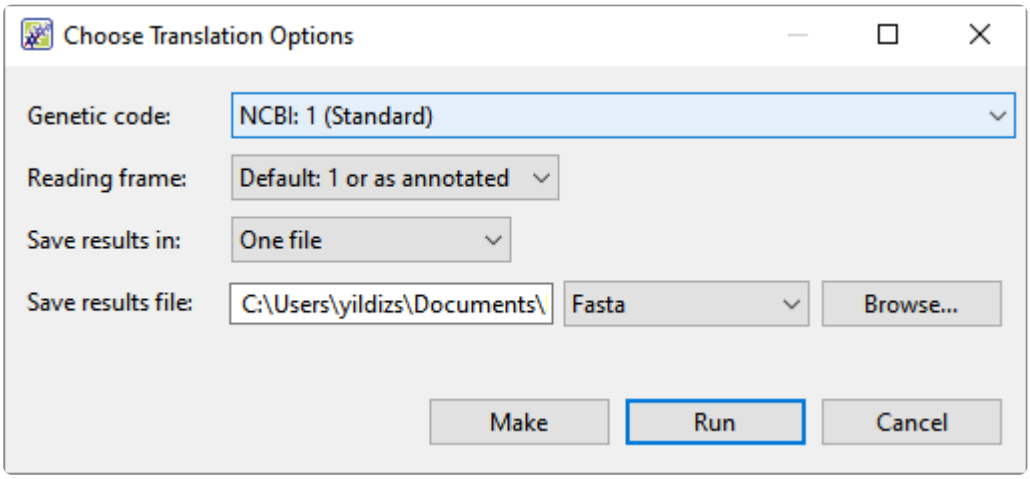
Part B: Translation in SeqNinja

In this part of the tutorial, you will use SeqNinja to batch translate the sequences.

1. Launch **SeqNinja** and choose **Translate** from the menu on the right.

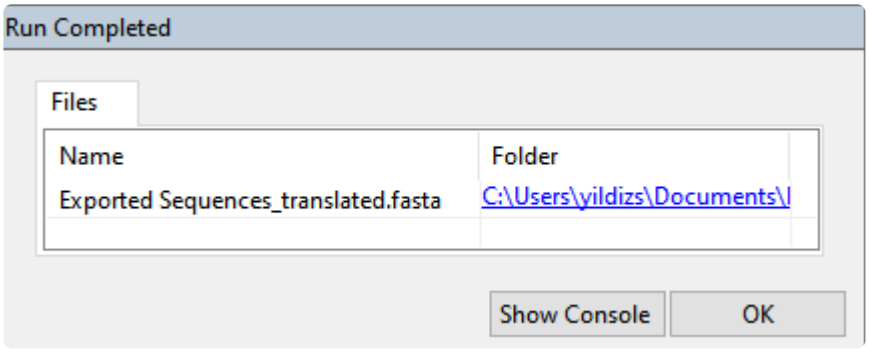


2. Navigate to and open the *Exported Sequences.gbk* file. The Choose Translation Options dialog appears.



3. Leave settings at the default values and press **Run** to perform the batch sequence translation. When prompted to save the script, click **No**. When the Run Completed pop-up appears, click **OK**.

Copyright DNASTAR, Inc. 2022



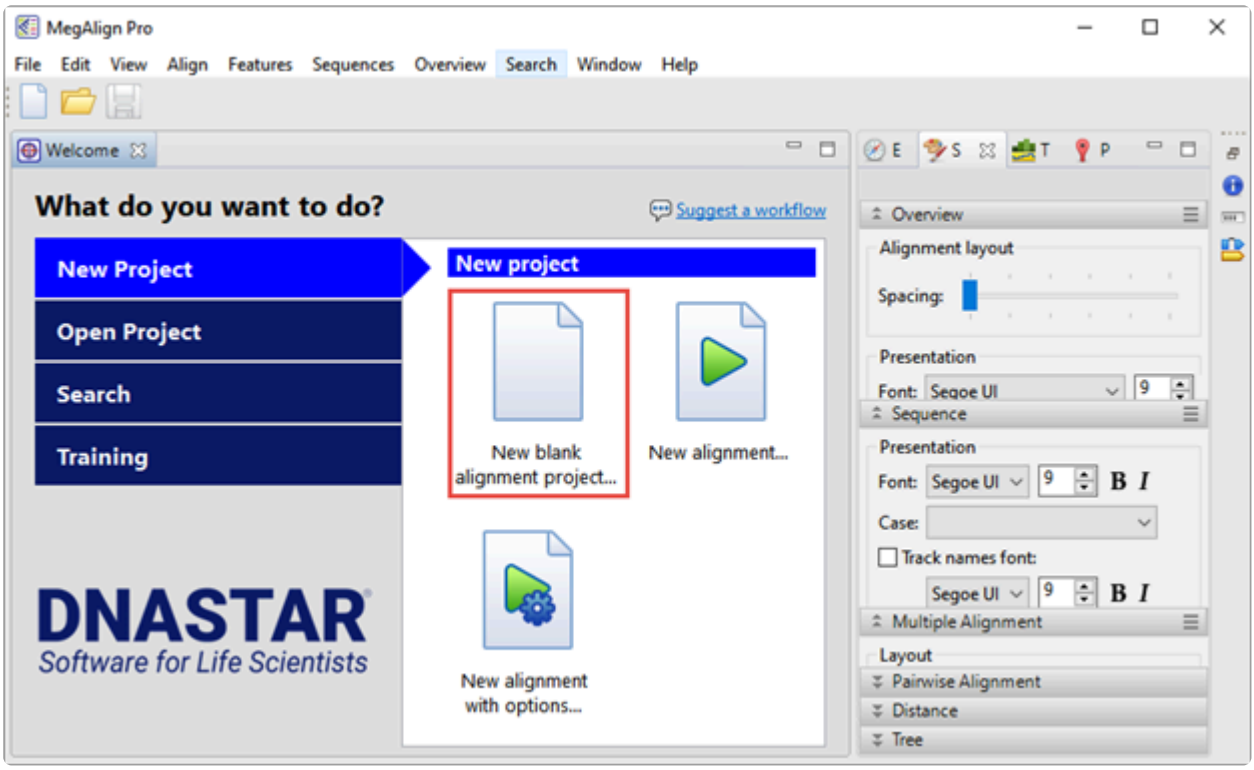
4. Close SeqNinja. There is no need to save the project.

Proceed to Part C: Alignment in MegAlign Pro.

Part C: Alignment in MegAlign Pro

In this section of the tutorial, you will use MegAlign Pro to align sequences and create a phylogenetic tree.

1. Launch **MegAlign Pro** and choose **New blank alignment project**.

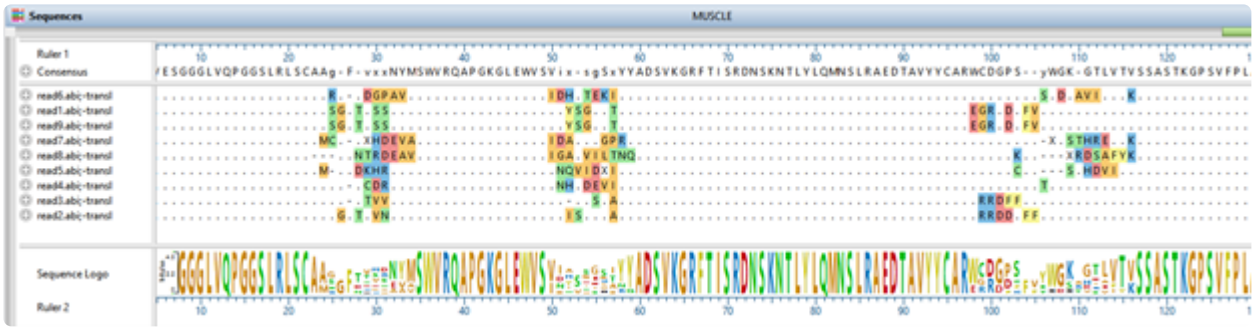


2. Use **File > Add Sequences** and add the *Exported Sequences_translated.fasta* file exported from SeqNinja. When prompted to select the sequence type, click **Protein**.

3. Perform the multiple alignment by choosing **Align > Align Using MUSCLE**. (Any method can be used in this workflow but using MUSCLE will ensure that your project will match the screenshots in this tutorial).

4. To more easily see the locations of the three CDRs, click the **Style** tab on the right. If necessary, expand the **Multiple Alignment** section by clicking on its gray bar. Change the **Comparison** drop-down menu to **Show only differences from reference**. In this case, the “reference” is simply the consensus. The Sequences view now appears as below. The three colorful areas denote the three CDR regions.

Copyright DNASTAR, Inc. 2022

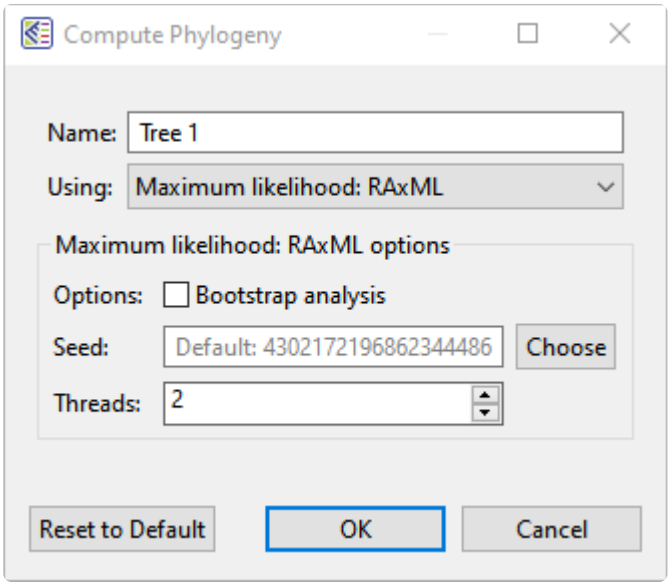


In the next steps, you will build a phylogenetic tree to see how the sequences are related.

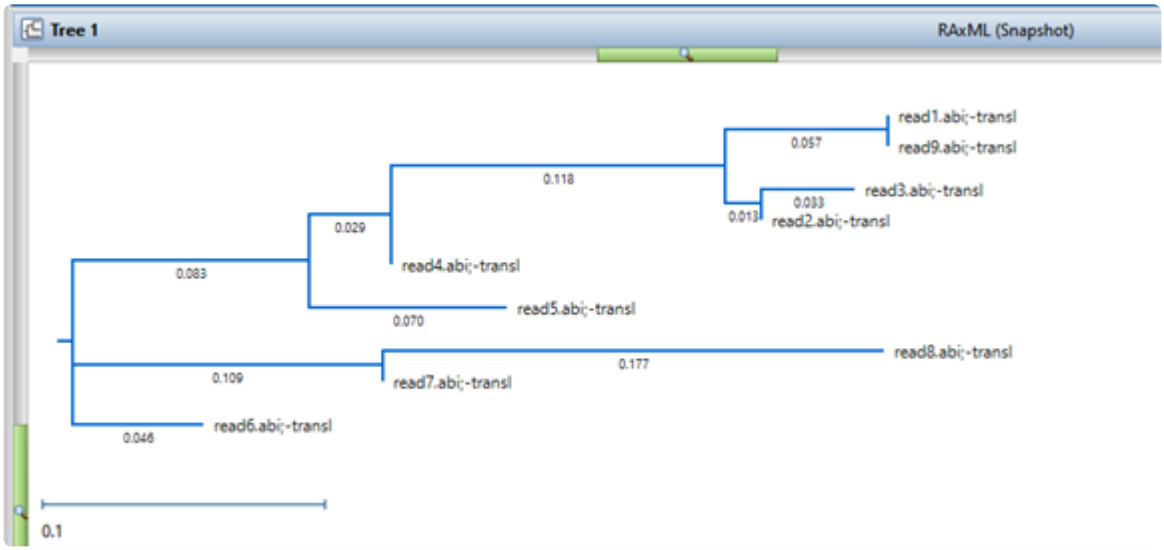
5. Click on the **Tree** tab (

).

6. Choose **Maximum Likelihood: RaxML**. Keep the default settings and click **OK**.

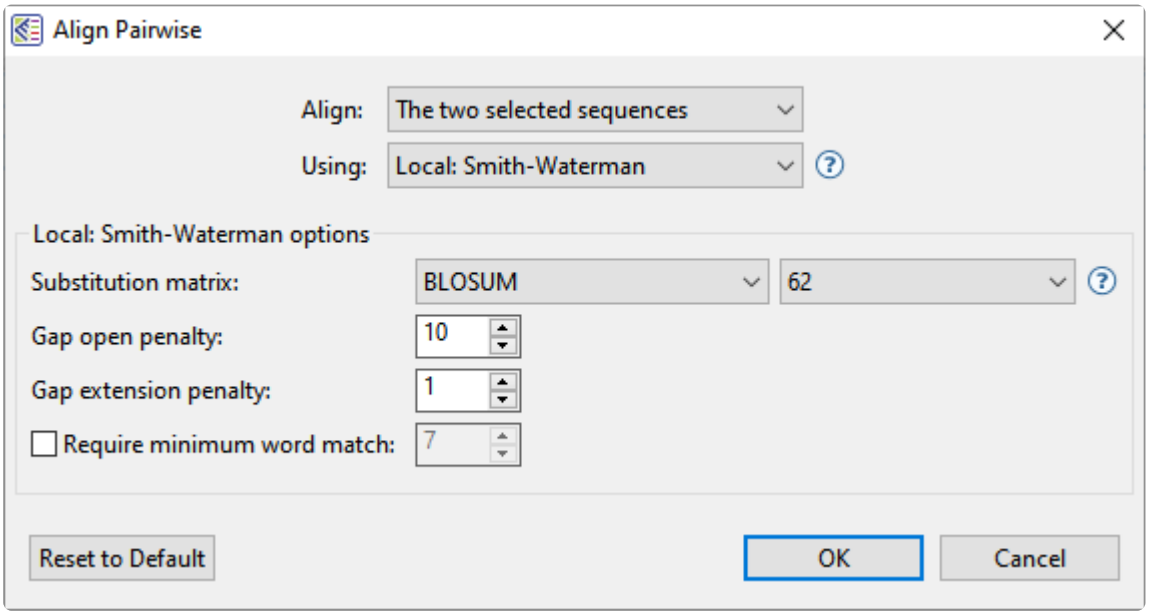


MegAlign Pro displays the calculated tree.

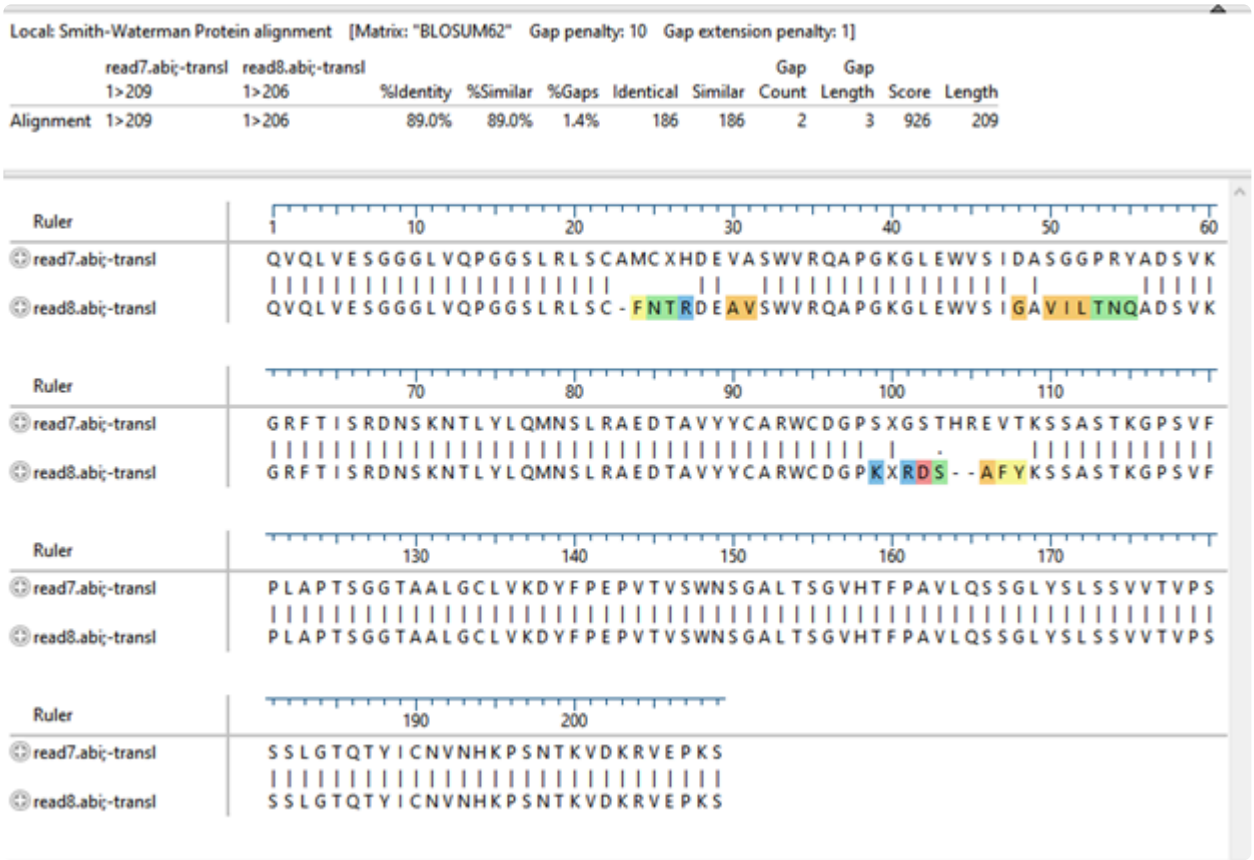


7. (optional) If you'd like to see the sequences in the Overview or Sequences view in the same order in which they appear in the tree, you can use **Tree > Order Sequences Like Tree**.

8. (optional) To get a more accurate comparison of specific sequences, you can perform one or more pairwise alignments. Choose any two sequences from any view, then right-click and choose Align Pairwise. Keep the default **BLOSUM 62** matrix (ideal for studying conservation) and press **OK**.



The default color scheme shows the three CDRs with colored backgrounds. Statistics appear in the header.



You will now export the first read sequence to use in the next section of this tutorial.

9. In any view, single-click on the sequence name *read1.abi-transl* and choose **File > Export Data > Sequences**. Keep the defaults and press **Save**.

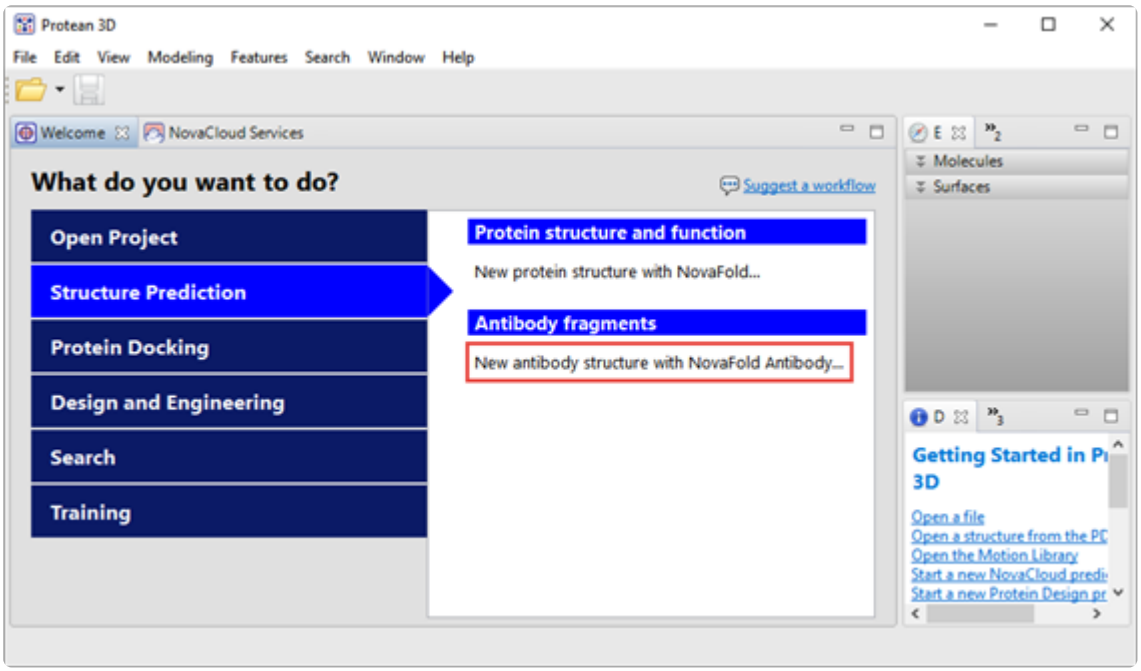
10. Close MegAlign Pro. There is no need to save the project.

Proceed to Part D: Structure prediction and docking in Protean 3D.

Part D: Structure prediction and docking in Protean 3D

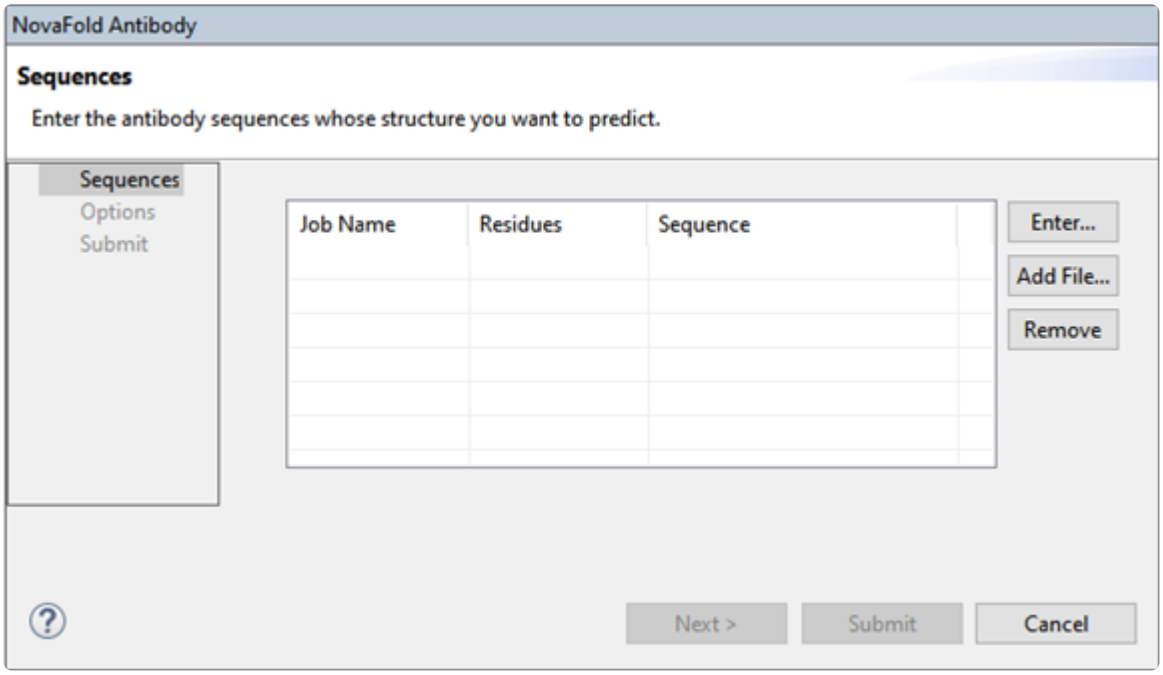
In this portion of the tutorial, you will use Protean 3D and the NovaApps to create a homology model of the sequence and simulate docking of the predicted antibody structures to the target epitope.

1. Launch **Protean 3D** and click the **Structure Prediction** tab on the left.



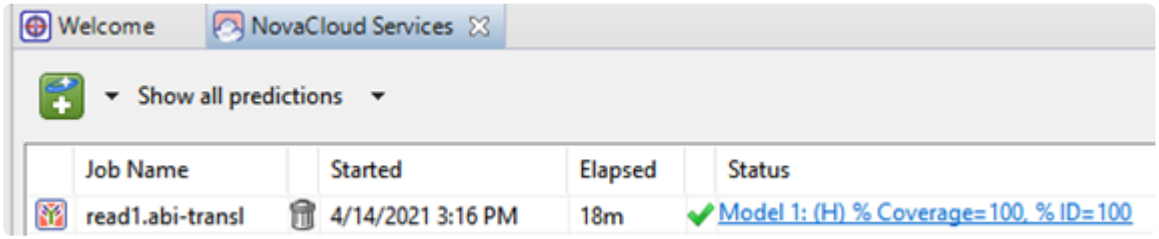
2. To create homology models of sequence you just exported from MegAlign Pro, choose **New antibody structure with NovaFold Antibody**. This launches the NovaFold Antibody wizard at the **Sequences** screen.

Copyright DNASTAR, Inc. 2022



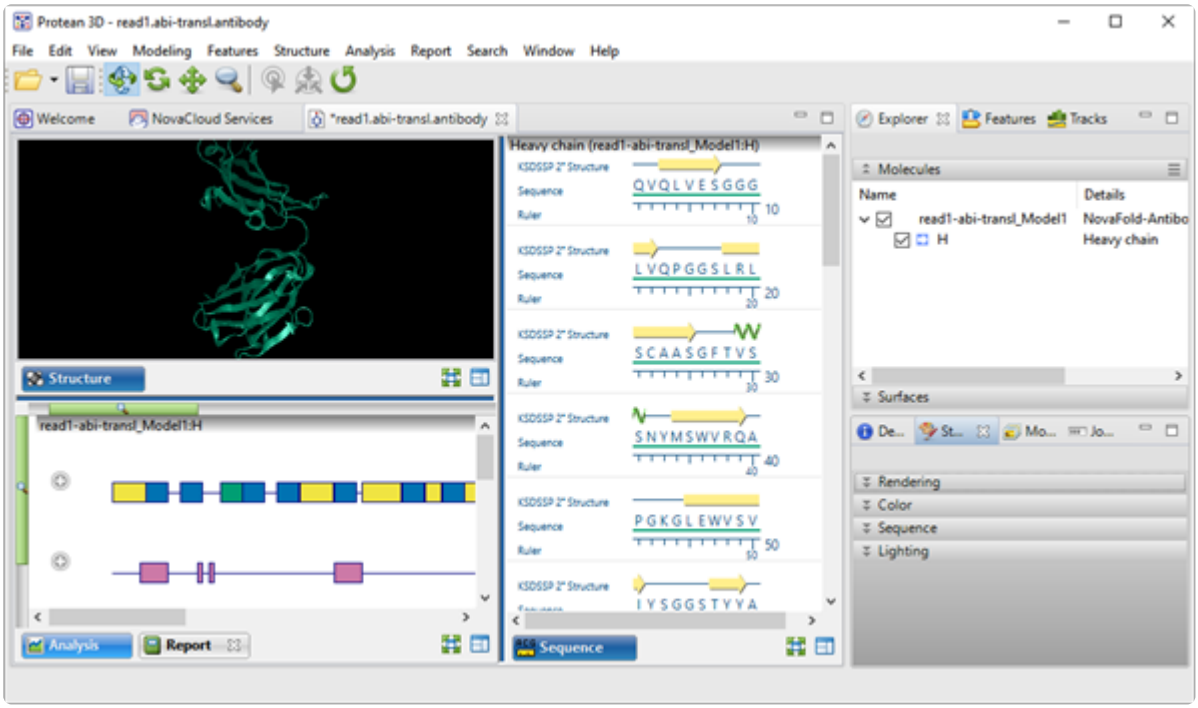
3. Click **Add File** and open *Untitled1-ungapped-sequences.fasta*. Click **Submit** to begin the antibody model prediction. Once you see the **Close** button, press it.

4. To see the status of the prediction, click on the **Predictions** tab. This prediction will take approximately 20 minutes. Once the prediction is complete, a blue link will appear under **Status**.



5. Click the blue link to open the structure in Protean 3D.

Copyright DNASTAR, Inc. 2022

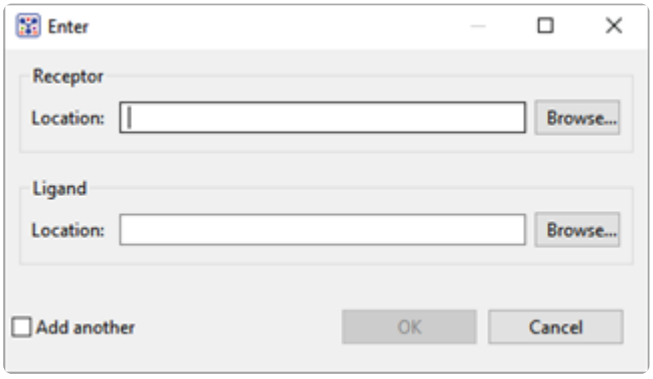


6. Save the structure using **File > Export Data > Export Structure**.

   In the next steps, you'll be using NovaDock to simulate docking of the predicted antibody structure to its target epitope. In real life, you could use the file you just exported. However, this tutorial utilized artificial reads, so that exported antibody does not correspond with a known antigen. Therefore, the rest of this tutorial will use separately provided antibody and antigen structures.

7. Choose **Structure > Predict Structures > New Protein-Protein Complex with NovaDock**.

8. In the Structures screen, click **Enter**.

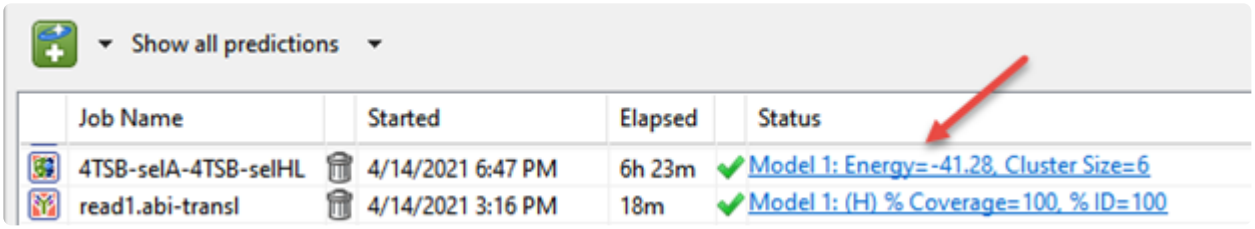


9. In the **Receptor** section above, **Browse** to the file *4TSB-selA.pdb*. In the **Ligand** section below, **Browse** to the file *4TSB-selHL.pdb*. Press **OK**.

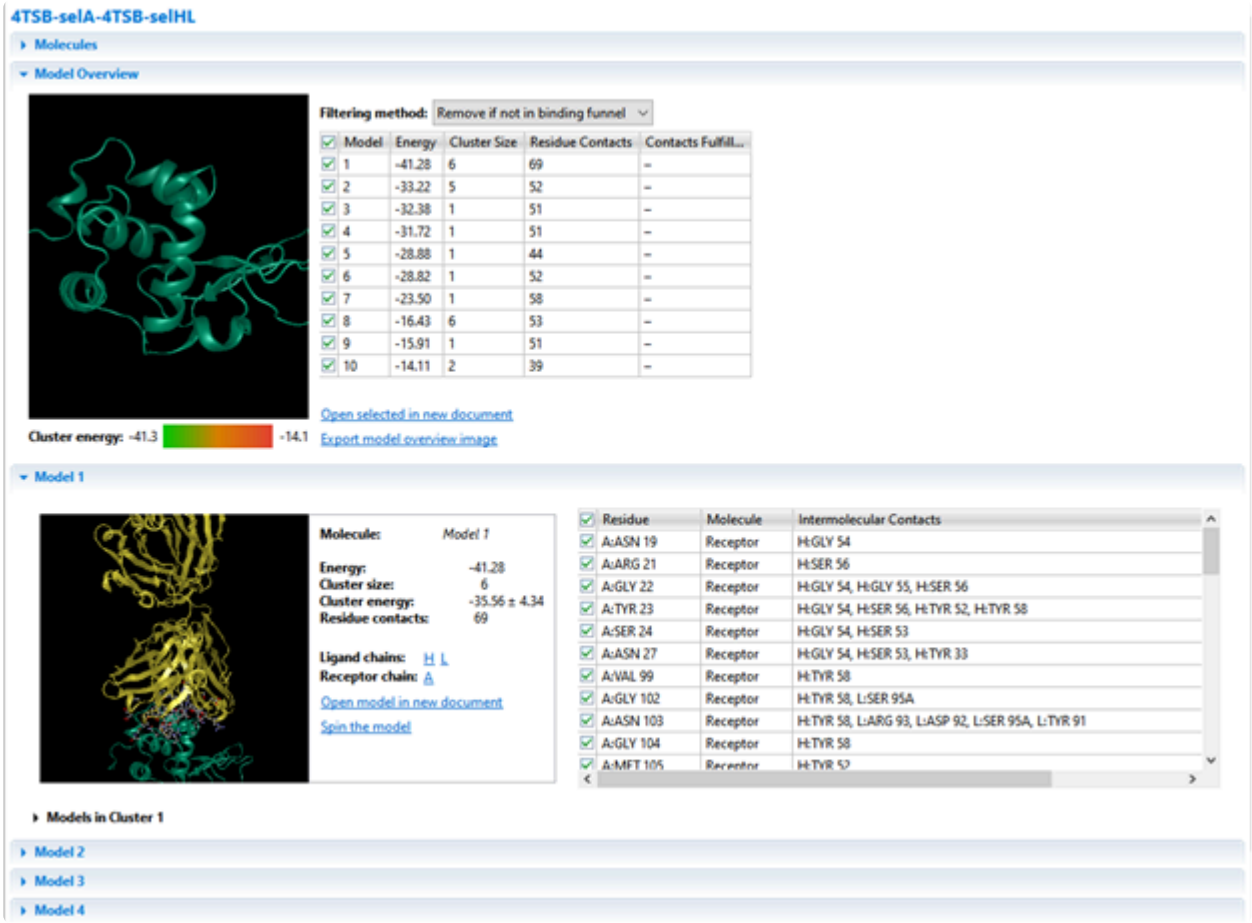
10. Press **Submit** to begin the prediction. When the **Close** button appears, click on it.

11. Use the **Predictions** tab to monitor the prediction status. This prediction will take approximately 7

hours. When the prediction is finished, you will see a blue link in the Status section with the word “Energy” in it.



12. Click on the link to open the docking model predictions in Protean 3D.



13. Open Model 1 by clicking the link Open model in new document.

14. (optional) Explore the antigen/antibody interface by doing one or more of the following:

    - Use **View > Style > Rendering** to apply a Surface to the sequence.

    - Use **View > Style > Sequence**, then choose **Color by Track** from the drop-down menu. By choosing different types of tracks, you could ascertain, for example, that the molecule is binding to the epitope in an area with a polar charge difference.

Copyright DNASTAR, Inc. 2022

- Find residues within a specified number of angstroms of any selected residue. To do this, select a residue in any view, then choose **Structure > Select Neighbors**. To change the angstrom value, choose **Edit > Preferences**, then choose **Proteins** from the left. Change the **Cutoff distance** value as desired. You could then use the Style panel to color the selected neighboring residues differently from other parts of the molecule.

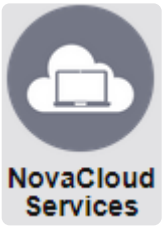
- Export the data in *.csv* format using **File > Export Data > Export Report**.

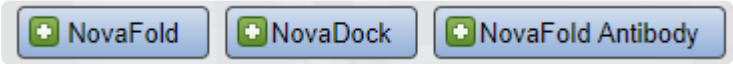
Copyright DNASTAR, Inc. 2022

Run Nova applications through the DNASTAR website

This topic is a general overview on how to set up, run and view NovaFold, NovaFold Antibody and NovaDock through the DNASTAR website. If you are instead running these services through the Protean 3D interface, see “Run NovaFold / NovaFold Antibody / NovaDock through Protean 3D.” If you are running these services locally on a Linux or Macintosh machine, see Run Nova applications locally.

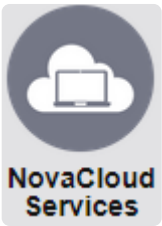
To initiate a NovaCloud prediction:

1. Go to the DNASTAR My Account page. If you are not already logged in, you will be prompted to log in.
2. Click the **NovaCloud Services** button: 

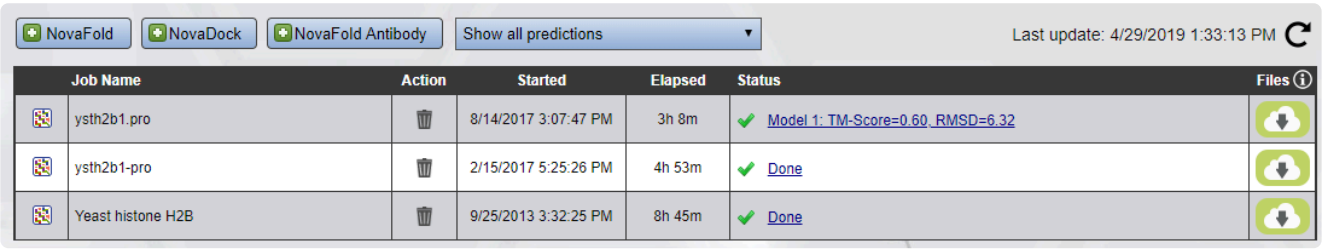
3. Choose the Nova application of interest: 

4. Follow the instructions in the online wizard.

To monitor run status and download results:

1. Go to the DNASTAR My Account page. If you are not already logged in, you will be prompted to log in.
2. Click the **NovaCloud Services** button: 

. The table shows the status of each run.



Copyright DNASTAR, Inc. 2022

3. To download results, click the download link in the **Files** column. When prompted, choose **Download**.

To view results:

Double-click on the downloaded result file to launch Protean 3D.

NovaFold Antibody prediction results (*.antibody*) file can be viewed and analyzed in Protean 3D like any other structure. The Structure view and Features panel facilitate examination of antibody features, including the six annotated CDR loops for light and heavy chains. For information on viewing other types of results, see View NovaFold results or View NovaDock results.

Copyright DNASTAR, Inc. 2022

Run Nova applications locally

Three separately-licensed Protean 3D services — NovaFold, NovaFold Antibody and NovaDock — can be initiated online (see note below) or locally, through the terminal of a Linux or Macintosh computer. This topic and its subtopics discuss how to initiate predictions locally.

To run a local prediction:

1. Install and set up the software on a Linux or Macintosh machine. All three of the “Nova applications” are installed, though each is licensed separately.

2. License the desired applications by contacting support@dnastar.com.

3. Prepare one or more input files.

4. Set up and run a prediction.

5. View the results in Protean 3D on a Windows or Macintosh computer. NovaFold Antibody results open as a typical structure file, but have some additional information available. For information on viewing other results, see View NovaFold results and View NovaDock results.

Note: This topic applies to Nova applications run locally on a Linux or Macintosh machine. If you are running the Nova applications through Protean 3D or the DNASTAR website, instead see:

* Predict a Protein Structure with NovaFold
* Predict Antibody Interactions with NovaFold Antibody
* Explore Protein Docking with NovaDock

Copyright DNASTAR, Inc. 2022

Install and set up the software

Before installing NovaFold, NovaFold Antibody, or NovaDock for local use, you must first install Docker, a computing platform for distributed applications that is needed to run the programs.

On a Windows or Macintosh computer:

Download and install the DNASTAR License Server by following these instructions: Windows / Macintosh.

On the Linux or Macintosh computer where you will be using the Nova applications:

> **Note:** DNASTAR began supporting Macintosh for this use since February 29, 2019.

1. Log in to your DNASTAR account; you will automatically be taken to the My Licenses page. If you are already logged in, click the My Account link at the top of the DNASTAR header to go to the same page.

2. Use the button in the **Installers** column to download the installer archive file. Five downloadable archives are available, but you only need to download the installer package; the other four archives are downloaded automatically during the installation process. However, if desired, you may download and save all five archives to your Linux machine.

3. Unpack the installer archive file on your computer (e.g., using the `tar` command).

4. Install Docker using either of these methods, both of which require `sudo / root` privileges:

- Follow the instructions at Digital Ocean: "How To Install and Use Docker on Ubuntu 18.04."

- Enter the following command in the terminal to run a script distributed in the installer archive file:

  ```
  $ get-docker
  ```

  Unless you add qualified users to the `docker` group, Docker requires `sudo / root` privileges. If you would like to add users, append their names to the `get-docker` command:

  Linux: `$ get-docker $[USERNAME]`

  Macintosh: `$ ./get-docker-mac`

  If you are one of the users to whom access was granted, you must log out and log back in for changes to take effect.

Copyright DNASTAR, Inc. 2022

5. Continue by installing the application.

    - On Macintosh:

    a. Move ‘Docker.app’ to the ‘Applications’ folder.

    b. Double-click ‘Docker.app’ in the ‘Applications’ folder to start Docker.

    c. Configure the Docker virtual machine to run MARKET_NAME.

    d. Open Docker preferences (**Status menu > Docker icon > Preferences**).

    e. On the Disk panel, increase ‘Disk image size’ to more than 200 GB, then apply.

    f. On the Advanced panel, increase ‘CPUs’ to at least 4 and ‘Memory’ to at least 16 GB, then apply.

- ◦ On Linux, run the script located in the unpacked application image folder. Installation usually takes 5-20 minutes.

    - **Simple install:** You will be prompted to accept the DNASTAR License Agreement (“yes”), and to enter your DNASTAR user name and password.

      ```
      $ install
      ```

    - **Custom install:** Depending on the use of optional arguments (i.e., everything following `install`, below), you can run the installation without any further input:

      ```
      $ install [-h] [-D DOWNLOAD_DIR] [-s SUBSCRIBER] [-I INPUT_DIR] [-y]
      ```

      The following table describes each of the available arguments:

Optional argument	Description
-D DOWNLOAD_DIR, --download-dir DOWNLOAD_DIR	Specify a directory in which to save downloaded files.
-s SUBSCRIBER, --subscriber SUBSCRIBER	Authenticate with your DNASTAR subscriber number instead of user name and password.
-I INPUT_DIR, --input-dir INPUT_DIR	Specify the directory containing the local image and volume data files.
-y, --yes	Automatically respond to prompts with ‘yes.’

6. Verify that the repository is showing the correct version of the application:

Copyright DNASTAR, Inc. 2022

```
$ docker images
```

7. (optional) If you need to resolve local servers by hostname, you can define the local DNS server by editing the `DOCKER_OPTS` variable to read `DOCKER_OPTS="--dns $[LOCAL_DNS_IP]"`. In Ubuntu, the variable is defined in `/etc/default/docker`.

> **Note:** If you previously installed the Nova applications and are having issues with reinstallation, see Uninstall the software and Reinstall the software.

Once you have successfully installed Docker and the Nova applications, the next step is to prepare the input file.

Copyright DNASTAR, Inc. 2022

Uninstall the software

To remove a Nova application:

If you want to remove one of the Nova applications, use either of these commands:

```
# /opt/dnastar//uninstall
$ sudo/opt/dnastar//uninstall
```

Note that removal requires root access or elevated (sudo) privileges. If there are multiple components installed in the '/opt/dnastar' directory, you will need to repeat this command for each component.

Copyright DNASTAR, Inc. 2022

Reinstall the software

If you have installed the Nova applications successfully in the past, subsequent reinstallation may result in one or more error messages. Simple steps for resolving these errors are described below:

Error: failed to remove one or more images

This issue occurs when one or more Docker containers reference the image being removed. To address this issue:

1. Find out which containers reference the image by running: `docker ps – a`
2. Remove the Docker containers one by one using this command: `docker rm value`

   …where `value` is the either the “Container ID” or the “Name” in the table resulting from Step 1.

The following table contains a list of the two Docker commands used to resolve the issue above, as well as other useful commands.

Script	Definition
`docker images`	To view a list of installed Docker images.
`docker rmi value`	To remove a particular Docker image, where `value` is the “Image ID.”
`docker ps K/ font>`	To view a list of Docker containers that are currently running.
`docker ps -a`	To view a complete list of Docker containers.
`docker rm value`	To remove a particular Docker container, where `value` is the “Container ID” or the “Name.”

Copyright DNASTAR, Inc. 2022

Prepare the input file(s)

Application	Directory contents
NovaFold	A FASTA-format protein sequence file containing a single sequence or chain. Maximum sequence length is 2,000 residues.
NovaFold Antibody	A FASTA-format protein sequence file containing either one or two sequences, representing the light and/or heavy chains of an antibody. If two sequences/chains are included, NovaFold Antibody models the complex. If one sequence/chain is included, NovaFold Antibody instead models the individual sequence or chain.
NovaDock	Two protein structure PDB files: one representing the ligand (L) and the other representing the receptor ®. You can obtain a protein structure file by downloading it from PDB or by predicting a protein sequence's 3D structure using NovaFold or NovaFold Antibody.

The next step is to set up and run a prediction.

Copyright DNASTAR, Inc. 2022

Set up and run a prediction

Note: Before initiating a prediction, you must install Docker and the Nova applications, and prepare the input file(s).

Predictions are initiated by typing a script consisting of commands and “arguments,” which are parameters that affect those commands. Commands and arguments are described in detail in separate topics: Required commands, Required arguments, Optional arguments, and Supplemental arguments.

The following step-by-step instructions discuss a generic script:

1. Launch the terminal.

2. Type the relevant application command and its parameters.

   --lshost=ipaddress (or =hostname)

   In this step, lshost must be in lower-case letters and preceded by two dashes (-).

3. (optional) Set up a persistent environment variable to define the license server address. Configuration varies by Linux distribution and command shell. Please consult your distribution’s documentation to determine the preferred way to set environment variables.

   For example, if you are using a Bourne-style shell like <code>bash</code>, you can set the environment variable for your user account by adding the line:

   export DNASTAR_LSHOST=<license server IP address>

   to one of the following files:

   $HOME/.profile

   $HOME/.bash_profile

   $HOME/.bashrc

   Note that DNASTAR_LSHOST must be in capital letters and should not be preceded by any dashes. In addition, no spaces should be entered before or after the equal sign.

   To test if the environment variable has been set, open a new shell and run the <code>env</code> command to confirm DNASTAR_LSHOST is listed.

Copyright DNASTAR, Inc. 2022

4. (optional) Append any additional run parameters:

   novafold --datadir value1 --seqname value2 --seqfile value3

   novafold-antibody --datadir value1 - – seqname value2 --seqfile value3

   novadock --output-dir value1 --name value2 --lfile value3 --rfile value4

   …where:

novafold novafold-antibody novadock	Command to run the associated application. A single Linux machine should only run one prediction and one application at a time since all resources will be used.
value1	Path to the workspace or output directory.
value2	Any desired name consisting of alphanumeric characters, hyphens and/or underscores. No spaces are allowed.
value3	Path to an input file, such as a sequence file or a ligand PDB file.
value4	Path to an input file, such as a receptor PDB file.

5. (optional) If you did not set up the environment variable in Step 3, above, append the IP address or hostname specification for your DNASTAR License Server:

   --LSHOST ipaddress (or hostname)

6. (optional) Append any additional run parameters. An example NovaFold script appears below these steps. For a list of available parameters for all three Nova applications, see Required commands, Required arguments, Optional arguments, and Supplemental arguments.

7. Press **Enter** to begin the prediction. For NovaFold and NovaDock, the prediction may require several hours up to several days to finish, depending on the input. NovaFold Antibody predictions usually require only about 5-10 minutes. If you wish to stop a prediction that is in progress, see Stop a prediction in progress.

After a prediction has finished successfully, result files will be added to the workspace folder (*datadir*) or the output directory referenced in the script.

Example script:

novafold --datadir ~/Desktop/BioluminescentSeqs --seqname firefly_luciferase --hours 100 --GO true --EC true --LBS true

The NovaFold script above denotes that the protein sequence ~/Desktop/BioluminescentSeqs/seq.fasta should be used as the query, and each simulation should run for a maximum of 100 hours. Besides the structure prediction, the following optional predictions should be made: protein function GO terms, enzyme active sites, and ligand binding sites. Results are to be saved as ~/Desktop/BioluminescentSeqs/results.novafold.

Copyright DNASTAR, Inc. 2022

Required commands

The following table shows commands that are required when scripting a local Nova application prediction. Only one of the three commands is required, depending on the application. The three rightmost columns show whether the command applies to NF (NovaFold), NFA (NovaFold Antibody) or ND (NovaDock).

Name	Definition	Arguments and Defaults	NF	NFA	ND
novafold	Run a NovaFold structure prediction of the designated query sequence. Requires pre-installation of the ‘dnastar/novafold’ Docker image.		✔		
novafold-antibody	Run a NovaFold antibody structure prediction. Requires pre-installation of the ‘dnastar/novafold-antibody’ Docker image.			✔	
novadock	Run a NovaDock protein-protein complex prediction. Requires pre-installation of the ‘dnastar/novadock’ Docker image.				✔

Copyright DNASTAR, Inc. 2022

Required arguments

The following table shows arguments that are required when scripting a local Nova application prediction. Only a subset of these arguments is required, depending on the application. The three rightmost columns show whether the argument applies to NF (NovaFold), NFA (NovaFold Antibody) or ND (NovaDock).

Name	Definition	Arguments and Defaults	NF	NFA	ND
--datadir DATADIR	The path to the location where output/results will be written.		✔	✔	
--seqname SEQNAME	The unique name of the query sequence.		✔	✔	
--output-dir OUTPUT, -o OUTPUT	The path to the result output directory.				✔
--name NAME	The name for this complex prediction.				✔
--lfile LFILE	Path to a ligand PDB file that includes all chains. The ligand is often, though not always, smaller than the receptor.				✔
--rfile RFILE	Path to a receptor PDB file that includes all chains. The receptor is often, though not always, larger than the ligand.				✔

Copyright DNASTAR, Inc. 2022

Optional arguments

The following table shows optional arguments when scripting a local Nova application prediction. Only a subset of these arguments is available, depending on the application. The three rightmost columns show whether the argument applies to NF (NovaFold), NFA (NovaFold Antibody) or ND (NovaDock).

Name	Definition	Arguments and Defaults	NF	NFA	ND
--force	Performs these two functions: * Deletes any existing Docker container sharing the same --seqname or --name as the current prediction input. * After the current run has finished, deletes the Docker container for that run. The --force command has additional functionality related to reinstallation.		✔	✔	✔
-h, --help	To display a list containing a subset of options and then to exit.		✔	✔	✔
--log [CRITICAL, ERROR, WARNING, INFO, DEBUG, NOTSET]	To enable specific types of logging.	default: [INFO]	✔	✔	✔
--lshost LSHOST	To define the network license server, where value is the IP address (or hostname) of the DNASTAR License Server. For capitalization rules, see Set up and run a prediction. **Note:** This parameter is not required if the environment variable has already been set.		✔	✔	✔
--recover	To recover results from a stopped Docker container.		✔	✔	✔
--version	To output the NovaFold version and exit.		✔	✔	✔
--supplemental-help	To display a list containing additional NovaFold options and exit. An argument shown without brackets is required. The use of brackets signifies that the argument is optional.		✔		

Copyright DNASTAR, Inc. 2022

--library LIBRARY	The version of library volume for prediction.	default: [latest novafold-antibody-library]	✔	✔	
--seqfile SEQFILE	To designate a local FASTA file as the query. SEQFILE represents the local path and filename of the query sequence, which must be in FASTA format. If you provide a path with no filename, NovaFold Local will look for a file named *seq.fasta* located in the specified directory.		✔	✔	
--lcontacts LCONTACTS	To specify a file listing proposed contact residues for the ligand. Each line represents one contact residue using the format: [CHAIN ID] [RESIDUE NUMBER] * The CHAIN ID is case sensitive and is typically a single character. * The RESIDUE NUMBER is an integer and may include an insertion code. Examples: L 15 L 15A				✔
--rcontacts RCONTACTS	To specify a file listing proposed contact residues for the receptor. Using this argument causes NovaDock to focus docking around the specified contacts, reducing computation time. Each line represents one contact residue using the format: [CHAIN ID][RESIDUE NUMBER] * The CHAIN ID is case sensitive and is typically a single character. * The RESIDUE NUMBER is an integer and may include an insertion code. Examples: R 15 R 15A				✔

Copyright DNASTAR, Inc. 2022

--lmodes LMODES	Number of normal modes of motion to explore for the ligand. We suggest using the default to model large, concerted conformational changes and using a higher number (e.g. 20) to model more localized fluctuations, such as within antibody interfaces.	default: 5			✔
--rmodes RMODES	Number of normal modes of motion to explore for the receptor. We suggest using the default to model large, concerted conformational changes and using a higher number (e.g. 20) to model more localized fluctuations, such as within antibody interfaces.	default: 5			✔
--ncopies NCOPIES	Number of copies of each swarm point to run.	range: [1,50] default: 4			✔

Copyright DNASTAR, Inc. 2022

Supplemental arguments

The following table shows supplemental arguments when scripting a local Nova application prediction. Only a subset of these arguments is available, depending on the application. The three rightmost columns show whether the argument applies to NF (NovaFold), NFA (NovaFold Antibody) or ND (NovaDock).

Name	Definition	Defaults	NF	NFA	ND
--verbose – v	To enable verbose output.		✔	✔	✔
--welcome	To show a welcome message.		✔	✔	✔
--EC value	To predict enzyme active sites.	allowed: [true, false] default: false	✔		
--enhancedSearch value	To perform DNASTAR's experimental method for enhancing the structural diversity of the normal template set. For templates selected by protein threading, a proprietary process samples alternate structural conformations and replaces a subset of the templates with lower energy conformations. **Note**: This option typically adds 30-60 minutes to the prediction time but, in some cases, improves the accuracy of the prediction. We recommend running the prediction with and without this search option.	allowed: [true, false] default: false	✔		
--GO value	To predict protein function GO terms.	allowed: [true, false] default: false	✔		
--homoflag value	To use all templates or exclude homologs for benchmarking.	allowed: [real, benchmark] default: real	✔		
--hours value	To set a maximum simulation runtime.	range: [1,200] default: 50	✔		
--idcut value	To set a sequence identity cutoff for benchmarking.	range: [0,1] default: 0.3	✔		

Copyright DNASTAR, Inc. 2022

--LBS value	To predict ligand binding sites.	allowed: [true, false] default: false	✔		
--light value	To enable fast mode (override 'hours' option to 5).	allowed: [true, false] default: false	✔		
--nmodel value	To specify the maximum number of models to create.	range: [1,10] default: 5	✔		
--ntemp value	To specify the maximum number of templates to be used from each threader.	range: [1,50] default: 20	✔		
--restraint1 value	To provide a text file containing a collection of distance and/or contact restraints (*e.g.*, active sites, zinc fingers, disulfide bonds): * Pairwise distances between two atoms (i and j) * Contact between two residues (i and j) If both Distance and Contact are specified, they are described in different rows in the same restraint file. Value represents the path and filename of the text file containing the distance and/or contact information. A file located outside of the *datadir* data directory will be copied into the *datadir*. **IMPORTANT** – The text file may ***not*** contain any lower-case letters. **Note: For detailed information and an example, see the text below this table.**		✔		
--restraint2 value	To provide a text file containing a user-defined template structure and the alignment between that template and the query sequence. Value represents the path and filename of a text file containing the information below: * The pairwise FASTA-formatted sequence alignment between query and template. * The standard PDB format 3D structural coordinates of		✔		

	a single protein chain of the desired template. A file located outside of the *datadir* data directory will be copied into the *datadir*. The alignment file may only include one template sequence. **Note: For detailed information and an example, see the text below this table.**				
--restraint3 value	To nominate a specific single-chained PDB structure as a template in the modeling prediction, along with other templates selected by NovaFold. Value represents the PDB and chain desired for the user template in the format [PDB ID]:[CHAIN ID]. * The CHAIN ID is case sensitive. * An underscore (_) may be used to designate the first listed chain in the PDB. * Downloading the designated file from the Protein Data Bank requires Internet connectivity. Example: 7tim:A Since 'A' is the first chain, 7tim:_ would work as an alternative expression. **For additional information, see the Note below this table.**		✔		
NovaFold: --restraint4 value **NovaFold Antibody:** --add ADD	To nominate a local 3D structure (in PDB format) as a template in the modeling prediction, along with other templates selected by NovaFold or NovaFold Antibody. Value and ADD represents the path and filename of a standard PDB format text file. NovaFold requires a single protein chain; NovaFold Antibody can accept multiple chains. **For additional information, see the Note below this table.**		✔	✔	
NovaFold:	To exclude certain templates from the library (i.e., to	default	✔	✔	

--temp_excl value **NovaFold Antibody:** --exclude EXCLUDE	prevent these templates from being considered) during structure prediction, where value and EXCLUDE represents the name of the file containing the list of structures to exclude. In both NovaFold and NovaFold Antibody, templates can be excluded solely by name. In addition, NovaFold lets you specify a sequence identity cutoff value, such that all templates with an identity at that threshold or higher are excluded. By default, if no sequence identity cutoff is specified, a value of 100% is used. The tab delimited file listing templates to exclude must have the following format: [PDB ID][CHAIN ID]. Example: 1wor:A A percent sequence identity can be specified at the end, if desired. For example, 1wor:A 70 would specify a 70% sequence identity cutoff. If no number is specified, the percent sequence identity is assumed to be 100%. An asterisk (*) may be used to designate any chain in the PDB file. For example, 1wor:*	sequence identity cutoff: 100%			
--include INCLUDE	To define one or more chains from a PDB structure as the template(s) to use in the modeling prediction. INCLUDE represents a file listing the only templates to use. Each line represents one template using the format [PDB ID][CHAIN ID][HEAVY/LIGHT] * The PDB ID is four characters, and starts with a number, followed by three letters or numbers. * The CHAIN ID is case sensitive and is typically a single character. Examples: 1IGT:A Heavy 1IGT:A Light			✔	

Copyright DNASTAR, Inc. 2022

--models MODELS	The H3 loop is the generally the hardest region of the antibody structure to predict. NovaFold Antibody offers a template-based approach that uses a machine learning model to choose the best templates for the H3 loop. The --models option lets you specify how many results models to output, each using a unique H3 loop template.	range: [1-10] default: 1	✔		
--max-abinitio-h3 MAX_ABINITIO_H3	To specify a cutoff for switching from *ab initio* "Distance Guided" prediction of the H3 loop to a template-based prediction. Example: The default setting of '3' means loops of length 3 or shorter would use the *ab initio* methodology, while loops longer than 3 residues would be built with the template-based approach.	range [3-15] default: 3		✔	
--min-coverage	The default behavior of NovaFold Antibody (--min-coverage = 0.0) is to select a template framework based on its statistical significance to the query sequence; coverage criteria is not considered. The threader picks the template with the highest (log-likelihood score)/(background log-likelihood) for matching a sequence. If a single domain matches, only that domain will be proposed as a model. The result is that the most significant template may not always cover the entire query. To ensure the selected template exceeds a particular fractional coverage, specify a higher number for the --min-coverage. This compels NovaFold Antibody to locate a template with both variable and constant domains.	range: [0.0-1.0] default: 0.0		✔	
--no-orient-refine	By default, NovaFold Antibody optimizes the rigid-body orientation between the light and heavy antibody chains to remove atomic clashes if they were introduced during the modeling process. To skip the optimization step, use the argument --no-orient-refine.			✔	

Notes:

* During a folding prediction, the threader and user template alignments are each ranked. Therefore, user provided templates may not necessarily appear in the NovaFold Report's 'top ten' template list.

Regarding the template-based restraint commands (--restraint2, --restraint3/INCLUDE and --restraint4/ADD):

* Only one template-based restraint parameter can be used in the novafold command string.

* The add option can be used multiple times to introduce multiple templates to the modeling process with the novafold-antibody command. The include option can be used in combination with the add option.

****

Column requirements in the --restraint1 text file:

Distance rows contain the following columns from left to right:

DIST
Res_No.i
Atom_type_i
Res_No_j
Atom_type_j
Distance in Angstroms

Contact rows contain the following columns from left to right (see definitions below):

CONTACT
Res_No.i
Res_No.j

In both cases, UNK can be used in a row to represent an unknown atom.

Column definitions for a --restraint1 text file:

Given two residues that contact one another (‘Residue i’ and ‘Residue j’) or two atoms at a distance from one another (‘Atom i’ and ‘Atom j’):

Res_No_i – Residue sequence number for Residue i.
Atom_type_i – Atom name for contacting atom of Residue i.
Res_No_j – Residue sequence number for Residue j.
Atom_type_j – Atom name for contacting atom of Residue j.

Example text file for --restraint1:

Copyright DNASTAR, Inc. 2022

DIST 12 HG21 50 HB1 8.1

DIST 14 HA 57 1HE 6.2
DIST 21 HB2 43 HD11 4.0
DIST 124 CA 84 CA 17.4
DIST 36 UNK 120 CA 17.4
CONTACT 33 6
CONTACT 60 29
CONTACT 37 345
CONTACT 109 42

****

When using --restraint2, note that:

- The length of the aligned template residues must be ≥ 25% of the length of the query sequence.
- In the coordinate section, the ATOM record indices need to be numbered sequentially, beginning at 1.

Example text file for --restraint2:

The following is a -restraint2 file for mammoth myoglobin (query) against whale myoglobin (target structure). “ATOM” rows 6-1211 have been omitted for space. The format for ATOM records is described on this PDB web page.

```
>query
MGLSDGEWELVLKTWGKVEADIPGHGLEVFVRLFTGHPETLEKFDKFKHLKTEGEMKASE
DLKKQGVTVLTALGGILKKKGHHQAEIQPLAQSHATKHKIPIKYLEFISDAIIHVLQSKH
PAEFGAD---------------------------
>1MBN:A
-VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASE
DLKKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRH
PGDFGADAQGAMNKALELFRKDIAAKYKELGYQG
ATOM 1 N VAL A 1 -2.900 17.600 15.500 1.00 0.00 N
ATOM 2 CA VAL A 1 -3.600 16.400 15.300 1.00 0.00 C
ATOM 3 C VAL A 1 -3.000 15.300 16.200 1.00 0.00 C
ATOM 4 O VAL A 1 -3.700 14.700 17.000 1.00 0.00 O
ATOM 5 CB VAL A 1 -3.500 16.000 13.800 1.00 0.00 C
...
ATOM 1212 NE2 GLN A 152 -1.600 24.200 -1.500 1.00 0.00 N
ATOM 1213 N GLY A 153 1.500 24.700 -6.400 1.00 0.00 N
ATOM 1214 CA GLY A 153 1.100 24.000 -7.600 1.00 0.00 C
ATOM 1215 C GLY A 153 0.300 22.700 -7.500 1.00 0.00 C
```

Copyright DNASTAR, Inc. 2022

ATOM 1216 O GLY A 153 -0.900 22.800 -7.100 1.00 0.00 O
TER 1217 GLY A 153

Stop a prediction in progress

Once you have set up and run a prediction, the process will normally stop running only after the prediction is finished. If you need to stop a prediction before it is complete:

- If you are in the terminal with the running process, press **Ctrl+C** to stop the prediction.
- If you are in a different terminal, instead enter the command:

  ```
  docker stop seqname
  ```

  … where `seqname` is the same one used when the run was initiated.

Once you terminate a prediction, you may run a new prediction, if desired.

Copyright DNASTAR, Inc. 2022

Save and Export

Click on one of the links below to learn how to:

- Save a document
- Export a still image
- Export a structure file
- Export an animation or video
- Save or export a report

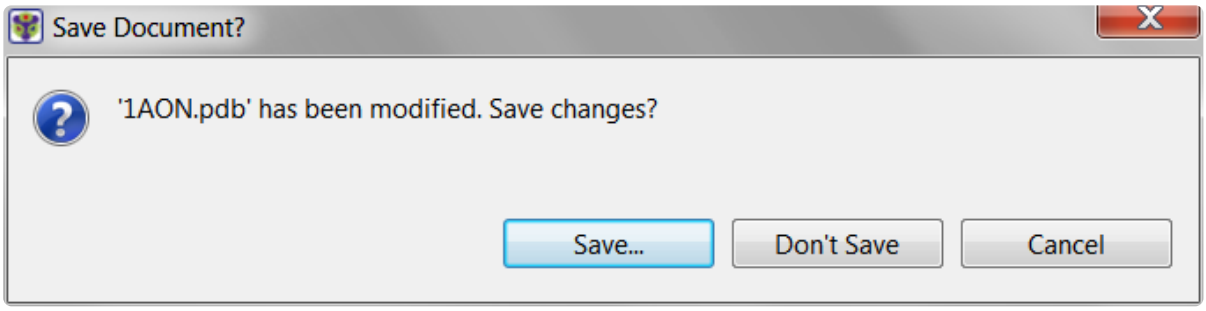
Copyright DNASTAR, Inc. 2022

Save a document

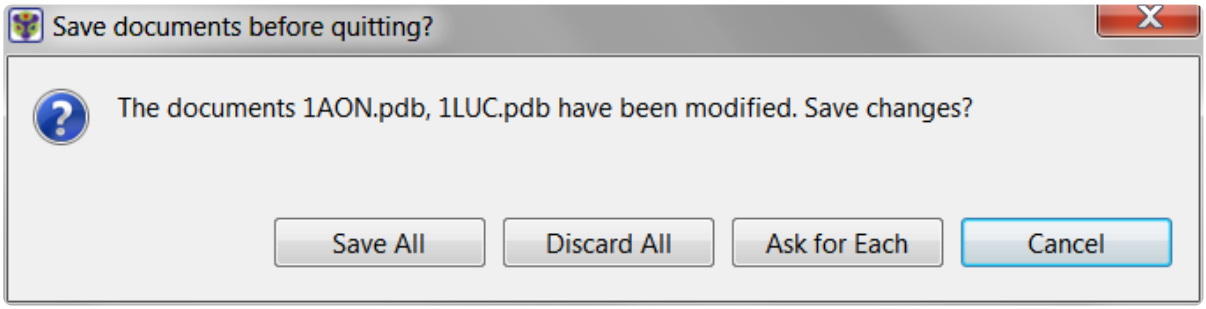
To save a new document in the *.structure* format, or to overwrite a previously saved version of the document, choose **File > Save**, or use **Ctrl+S** (Win) or **Cmd+S** (Mac). If you want to preserve the earlier version of a *.structure* document, or store a copy of the document with a different name or location, use **File > Save As** instead. Both commands save a “snapshot” of the Protean 3D document at the time of the save, including the settings, view layout and placement of panels.

Motion Library files originally have the extension *.pdb.gz*. Saved copies are assigned the *.structure* extension.

If only one document is open during an **Exit** (Win) or **Quit** (Mac), or if you **Save** or **Save As** for the first time, Protean 3D will open the Save Document dialog.



If there are multiple edited and unsaved documents open during an **Exit** (Win) or **Quit** (Mac), Protean 3D will display the following prompt:



- **Save All** saves all documents with their changes.
- **Discard All** exits without saving edits for any of the documents.
- **Ask for Each** opens individual Save dialogs for each open document (see first image, above).
- **Cancel** lets you return to Protean 3D without saving any documents.

A Protean 3D document saved in *.structure* format contains both structure and sequence information. Re-opening a saved *.structure* file in Protean 3D lets you see both types of data simultaneously. In contrast,

Copyright DNASTAR, Inc. 2022

opening the *.structure* file with other Lasergene applications allows you to view only the sequence and features portions of the file.

Copyright DNASTAR, Inc. 2022

Export a still image

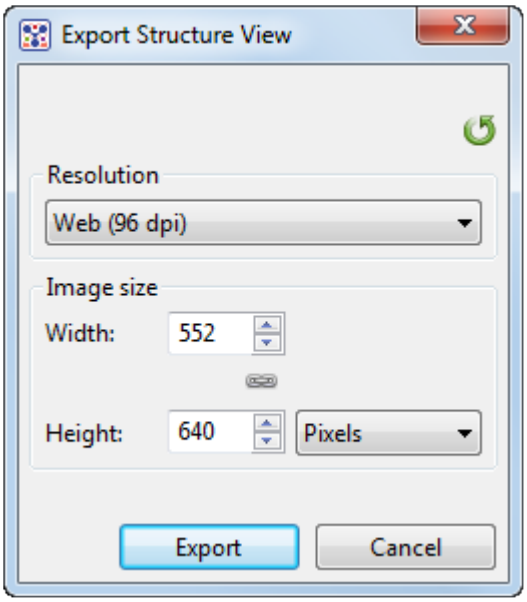
To export a picture of a view as a still image:

The method below can be used to save a “screenshot” from the Structure, Analysis or Sequence views or to export a still image from an animated sequence.

1. Choose **File > Export Image**.

2. In the Save dialog, enter a **File name** and make a selection from the **Save as type** drop-down menu. Depending on the view, you may be able to save the image as a *.pdf*, *.png*, *.jpg*, *.pptx* (Microsoft PowerPoint), or *.gif* file.

   If you are exporting the Structure view, another dialog will open.



3. Select the desired export options:

-
  - Choose from available image resolution using the **Resolution** drop-down menu.
  - In the **Width** and **Height** boxes, type in numbers or use the up/down arrows to specify numbers for the image size. If the chain icon is unbroken, making changes to either of these items will automatically cause the other to change proportionally.
  - The unbroken chain icon ( ) denotes that the aspect ratio (proportions) of the image are constrained by default. If you wish to size the **Width** and **Height** independently, click the icon to “break” the chain ( ).

4. Click the **Export** button.

Copyright DNASTAR, Inc. 2022

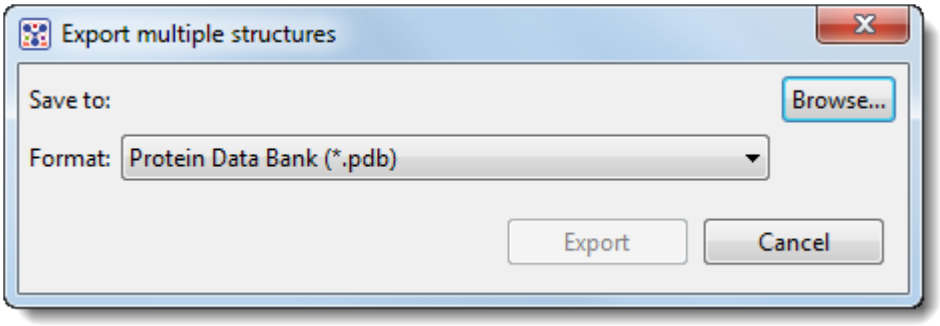
Export a structure file

To export the selected portion of a currently-displayed structure (or entire structure, if nothing is selected), choose one of the following **File > Export Data** options:

- **Export Structure** – Available for documents with one structure. By default, the structure is exported in **Protein Data Bank** (*.pdb*) format. If desired, you may instead specify **Macromolecular Crystallographic Information File** (*.cif*) format. Make any other desired changes in the dialog, and then press **Save** to export the structure.

> **Note:** To be exported in Protein Data Bank (*.pdb*) format, a structure must contain ≤ 99,999 atoms and ≤ 62 chains. If you attempt to export a structure that exceeds these limits, you will receive an error message that the export failed. However, you can still export a portion of the structure that does not exceed the limits. For example, if a three-chained structure is too large to export, try selecting just one or two of the chains before re-attempting the export.

- **Export All Structures** – Available for documents with two or more structures. After selecting this option, a popup dialog opens:



Choose a file format from the dropdown menu (*.pdb* or *.cif*). Click **Browse** to navigate to the location where you wish to save the structure. When you have made the desired selections, press **Export**.

- **Export All Selections** – Available when a selection is made, and applies only to selected regions. By default, the structure is exported in **Protein Data Bank** (*.pdb*) format. If desired, you may instead specify **Macromolecular Crystallographic Information File** (*.cif*) format. Make any other desired changes in the dialog, and then press **Save** to export the structure.

Copyright DNASTAR, Inc. 2022

Export an animation or video

To save an animated sequence in QuickTime Movie (*.mov*) or *.avi* format, choose **File > Export Movie**. Both formats are saved to play back at 6 frames per second (166 ms/frame), which is one-quarter the speed of standard television programs and films. This speed was chosen because it is optimal for viewing Motion Library files, most of which are 10-30 frames in length.

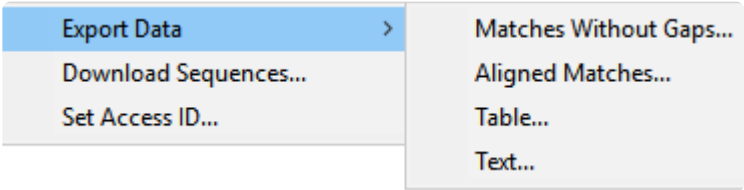
To save as an animated *.gif* choose **File > Export Image**. Animated *.gifs* are exported with a frame rate of 10 milliseconds/frame.

Note: Playback speeds may vary with the size of the structure and the type of program that is reading the file.

Copyright DNASTAR, Inc. 2022

Export data from the Text, Table or Pairwise views

The **Export data** tool ( ), located in the header above the Text, Table or Pairwise views, lets you export data from those views. This tool provides a choice of three options, one of which has four sub-options.



Options include:

- **Export Data >**
  - **Matches without Gaps** – To export an ungapped FASTA (*.fasta*) file consisting of matching regions of displayed matches, but not the query sequence itself.
  - **Aligned Matches** – To export a gapped FASTA (*.fasta*) file with displayed matches aligned against the query.
  - **Table** – To save Table view contents in either Comma- or tab-separated value (*.csv* or *.tab*) formats.
  - **Text** – To save the Text view contents in Text (*.txt*) format.
- **Download Sequences** – To download the selected match sequence in GenBank (*.gbk*) format.

The commands above all launch a Save As dialog. Choose a file name, location and file type (fixed, in some cases) then press **Save**.

- **Set Access ID** – To launch the Access ID dialog, in which you can specify your NCBI Access ID (optional). See Access ID dialog for information.

Copyright DNASTAR, Inc. 2022

Export a table of results from the Analysis view

To export a table showing results from all methods currently applied to the Analysis view, use **File > Export Data > Export Analysis Data**.

Choose the **Save as type**; either tab-separated values (*.tsv*) or comma-separated values (*.csv*). The *.tsv* file type is designed to be opened in a text viewer like Notepad, Wordpad, or Notepad++. Both types can be opened using standard spreadsheet applications, though *.csv* is recommended for spreadsheet use.

The image below shows an example of a saved file that has been opened in Microsoft Excel.

	A	B	C	D	E	F	G
1	Molecule	Residue	Position	Secondary structure|Chou-Fasman|Alpha Plot	Secondary structure|Chou-Fasman|Alpha Regions	Secondary structure|Chou-Fasman|Beta Plot	Secondary structure|Chou-Fasman|Beta Regions
2	4HHB:A	VAL	1	88	FALSE	109	FALSE
3	4HHB:A	LEU	2	99	FALSE	88	FALSE
4	4HHB:A	SER	3	93	FALSE	75	FALSE
5	4HHB:A	PRO	4	105	FALSE	70	FALSE
6	4HHB:A	ALA	5	111	FALSE	87	FALSE
7	4HHB:A	ASP	6	95	FALSE	84	FALSE
8	4HHB:A	LYS	7	93	FALSE	109	FALSE
9	4HHB:A	THR	8	93	FALSE	109	FALSE
10	4HHB:A	ASN	9	108	TRUE	95	FALSE
11	4HHB:A	VAL	10	124	TRUE	98	FALSE
12	4HHB:A	LYS	11	125	TRUE	88	FALSE
13	4HHB:A	ALA	12	111	TRUE	92	FALSE
14	4HHB:A	ALA	13	106	TRUE	90	FALSE
15	4HHB:A	TRP	14	96	TRUE	112	TRUE
16	4HHB:A	GLY	15	86	TRUE	102	TRUE
17	4HHB:A	LYS	16	105	TRUE	100	TRUE

Copyright DNASTAR, Inc. 2022

Save or export a report

To save a Report view :

The Report is automatically saved when you save a document.

To export the most recently-viewed report:

1. Either use the **Export data** tool ( ) in the view header, or choose **File > Export Data > Export Report**.
2. Specify the spreadsheet format in which you wish to save the report (*.csv*, *.tsv*, *.xtf*).
3. Press **Save**.

Copyright DNASTAR, Inc. 2022

Close and Exit

To close an active document:

Choose **File > Close** or press **Ctrl/Cmd+S**.

To exit from Protean 3D:

Choose **File > Exit** (Windows) or **Protean 3D > Quit Protean 3D** (Macintosh).

Copyright DNASTAR, Inc. 2022

Miscellaneous Tasks

The following topics cover tasks and troubleshooting tips that do not fit into any of the main help categories:

General tasks	Troubleshooting
Use basic editing commands Set preferences Synchronize sequence updates between two Lasergene applications Turn off usage logging Use a mouse or trackpad with Protean 3D Change the default print size	Prevent firewall issues when downloading from the Motion Librar Troubleshoot the error Troubleshoot blank Protean 3D window on Parallels Desktop

Copyright DNASTAR, Inc. 2022

Use basic editing commands

Protean 3D's basic editing commands are similar to those found in Microsoft Windows and other text editing programs.

Task	Command	Keyboard shortcut	Notes
To remove a portion of text	**Edit > Cut**	**Ctrl/ Cmd+X**	Text is placed on the clipboard.
To copy an item	**Edit > Copy**	**Ctrl/ Cmd+C**	To copy a portion of text, sequence, or a graphic element from any protein sequence or structure file, select the desired portion of the sequence or structure (e.g., a chain from the Molecules section, a feature, etc.). Then choose **Edit > Copy**. The selected portion of sequence is saved to the clipboard in FASTA format. It can then be pasted into a text editor, spreadsheet utility, etc. **Note:** Copying is not currently enabled in the Report views.
To paste an item	**Edit > Paste**	**Ctrl/ Cmd+V**	This pastes the item in the clipboard at the cursor insert point.
To delete a portion of text	**Edit > Delete**	**Delete** key	This command does not cause the text to be put on the clipboard.
To select everything in the view	**Edit > Select All**	**Ctrl+A**	This selects all sequence, all atoms, and everything in the Molecules and Style sections.
To invert the selection	**Edit > Invert Selection**		After making a selection in any view, using this command causes all other parts of the sequence/structure to be selected instead, and the original selection to be unselected.
To return the document to its state before the most recent operation	**Edit > Undo**	**Ctrl/ Cmd+Z**	**Undo** can be used to remove changes in rendering, color, visibility, selection and animation speed.
To reinstate the most recent change	**Edit > Redo**	**Ctrl/ Cmd+Y**	This cancels use of the **Undo** command.

Copyright DNASTAR, Inc. 2022

Note: In addition to the options above, editing commands are also available through context (right-click) menu options in many of Protean 3D's panels and areas.

Copyright DNASTAR, Inc. 2022

Set preferences

You can set preferences in Protean 3D for data, Internet, proteins and temporary files.

To open the Set Preferences dialog:

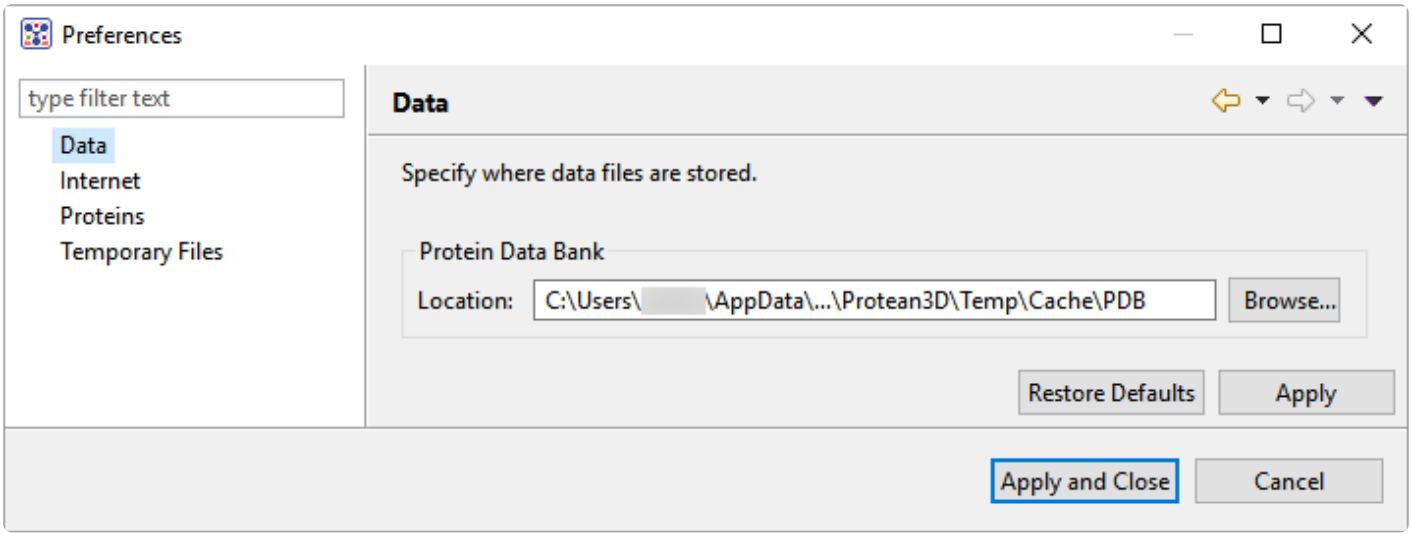
On Macintosh, choose **Protean 3D > Preferences**. On Windows, choose **Edit > Preferences**.

Copyright DNASTAR, Inc. 2022

Set data preferences

The Data Preferences dialog designates the folder in which PDB data files are stored when using the **File > Open from PDB** command. By default, PDB files are downloaded to the *\Protean3D\Temp\Cache\PDB* folder.

The dialog is accessed via the command **Edit > Preferences** (Win) or **Protean 3D > Preferences** (Mac). If the Data section is not already active, choose **Data** from the menu on the left.



To save PDB files in an alternative location, use the **Browse** button to navigate to the desired folder. If you maintain a local copy of the PDB, you may wish to change the path to point to your local PDB folder.

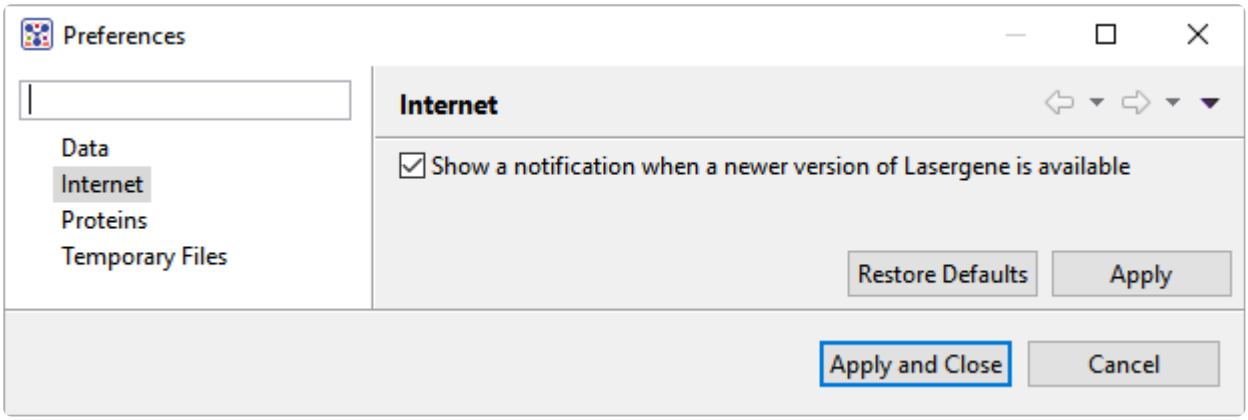
Click **Apply** to save your changes while keeping the preferences dialog open, or **Apply and Close** to save the changes and exit from the dialog. If you instead wish to return to default settings for all four drop-down menus, click **Restore Defaults**. To leave the dialog without applying changes, choose **Cancel**.

Copyright DNASTAR, Inc. 2022

Set Internet preferences

The Internet Preferences dialog lets you choose whether you wish to be notified about Lasergene updates.

To access the dialog, use **Edit > Preferences** (Win) or **Protean 3D > Preferences** (Mac). If the Internet section is not already active, choose **Internet** from the menu on the left.



Check the box if you would like Protean 3D to display a notification when a newer version of Lasergene is available. Uncheck the box if you do not wish to receive this notification.

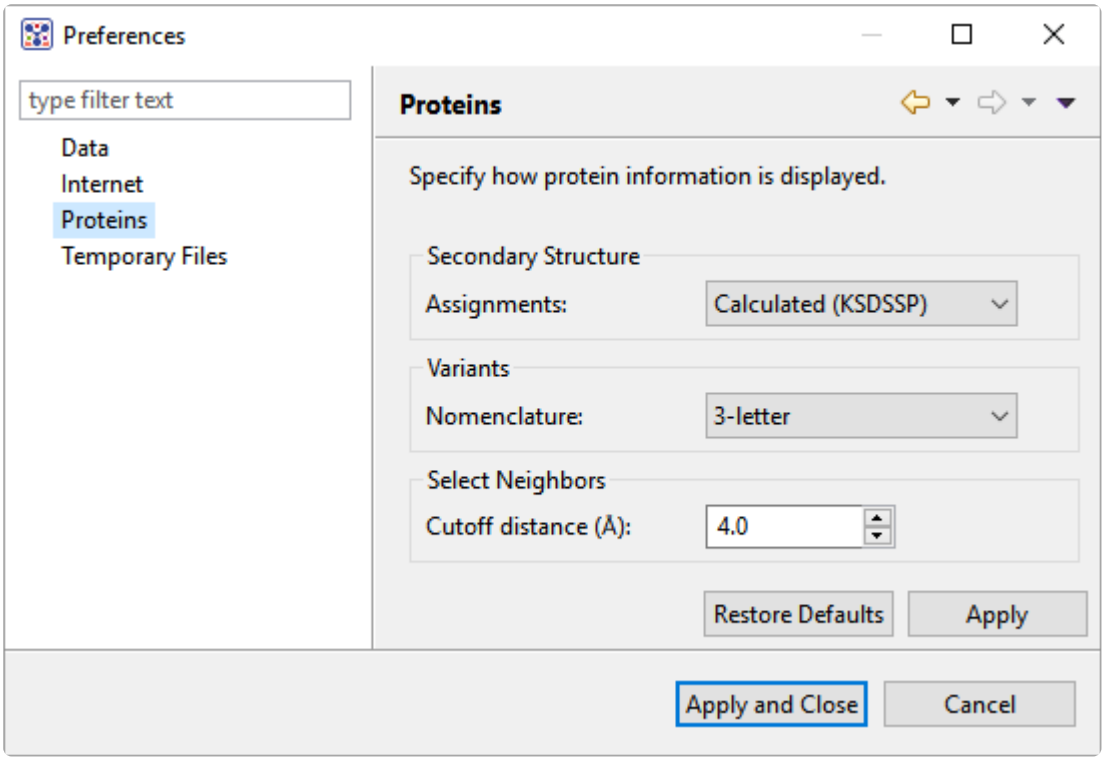
Click **Apply** to save your changes while keeping the preferences dialog open, or **Apply and Close** to save the changes and exit from the dialog. If you instead wish to return to default settings for all four drop-down menus, click **Restore Defaults**. To leave the dialog without applying changes, choose **Cancel**.

Copyright DNASTAR, Inc. 2022

Set proteins preferences

The Proteins Preferences dialog lets you adjust how protein information is displayed.

The dialog is accessed via the command **Edit > Preferences** (Win) or **Protean 3D > Preferences** (Mac). If the Proteins section is not already active, click on the word "Proteins" on the left.



- In the **Assignments** drop-down menu, select whether you want the default rendering for protein structures to be **Calculated (KSDSSP)** (Kinase Sequence Database secondary structure prediction) or **Annotated**.

- In the **Nomenclature** drop-down menu, specify whether mutations should use **1-letter**, **3-letter**, or **Genome (HGVS)** (Human Genome Variation Society) amino acid codes.

- In the **Cutoff distance (Å)** section, specify the likelihood that an item will be considered a “neighbor” of the current selection. The default value of 4.0 Å is optimal for most circumstances.

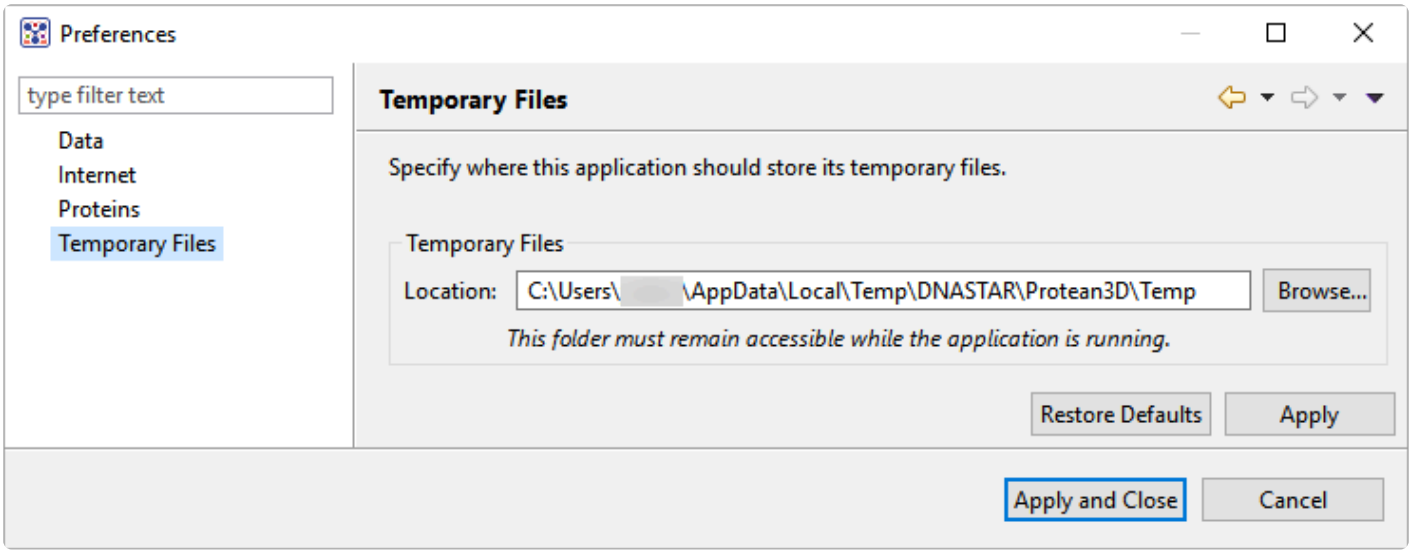
Click **Apply** to save your changes while keeping the preferences dialog open, or **Apply and Close** to save the changes and exit from the dialog. If you instead wish to return to default settings for all four drop-down menus, click **Restore Defaults**. To leave the dialog without applying changes, choose **Cancel**.

Copyright DNASTAR, Inc. 2022

Set temporary file preferences

Protean 3D creates and uses temporary files while a project is open. The need for temporary file space grows with the number and length of the sequences and amount of track data in open alignment projects. If you expect the temporary files will be too large to fit in the default location, you may specify a different location for these files using the Temporary Files Preferences dialog.

To access the dialog, use **Edit > Preferences** (Win) or **Protean 3D > Preferences** (Mac). If the Temporary Files section is not already active, choose **Temporary Files** from the menu on the left.



To change the temporary file location, close all Protean 3D projects and then return to this dialog. Select the **Browse** button, navigate to the new location and click **OK**. As noted in the dialog, the location you choose must remain accessible while Protean 3D is running. The new location will be used for all projects that are subsequently created or opened.

Click **Apply** to save your changes while keeping the preferences dialog open, or **Apply and Close** to save the changes and exit from the dialog. If you instead wish to return to default settings for all four drop-down menus, click **Restore Defaults**. To leave the dialog without applying changes, choose **Cancel**.

Copyright DNASTAR, Inc. 2022

Synchronize sequence updates between two Lasergene applications

Synchronous updating allows you to work on the same file in multiple Lasergene applications at the same time, and is available in SeqBuilder Pro and GeneQuest, and—on a limited basis—in Protean 3D. Changes made to a sequence in one application are automatically applied in the other shared applications.

Protean 3D does not allow the editing of sequences or features, and synchronous updating is not supported for .pdb files. As a result, Protean 3D has limited, unidirectional synchronous updating capacity compared to some other Lasergene applications. Changes made in another application may be synchronized to Protean 3D, but not the reverse, as in these two scenarios:

- A file is open in another Lasergene application and you have made unsaved changes to it (e.g., trimming, adding a feature). When you initially open the same file in Protean 3D, any changes made to the original sequence or its features will also appear there.

- The same file is already open in both Protean 3D and another application. You make changes to the file in the other application. Only feature edits will synchronize to Protean 3D; sequence edits are ignored.

Copyright DNASTAR, Inc. 2022

Turn off usage logging

By default, usage logging is enabled in Lasergene version 11 and later. To opt out of usage logging, launch the DNASTAR Navigator and go to **View > Preferences > Lasergene** (Win) or **DNASTAR Navigator > Preferences > Lasergene** (Mac). Uncheck the box and click **Apply** and then **OK**.

Copyright DNASTAR, Inc. 2022

Use a mouse or trackpad with Protean 3D

Protean 3D's user interface is controlled by a combination of menu commands, toolbar icon selections and mouse clicks. For best results, we recommend using a 3-button mouse with Protean 3D. However, single-button versions and trackpads are also supported.

If you are using a 1-button mouse, Protean 3D offers alternatives such as using a menu command in conjunction with a mouse click.

If you are a Mac user with a Mac Mighty Mouse, or similar, you can follow the 3-button mouse instructions, as long as the second and third mouse buttons are enabled. Please refer to the manufacturer's literature for more information.

If you are a trackpad user, you may need to change the configuration of your trackpad to recognize scrolling motions:

- Windows 8: Select **Control Panel > Mouse > Device Settings > (Select the trackpad) > Settings > Scrolling** and change the settings to allow scrolling. When this Protean 3D user's guide calls for scrolling, do so by drawing one finger up or down along the right edge of the trackpad.

- Windows 7: Select **Control Panel > Hardware and Sound > Devices and Printers > Mouse**, and choose the **Device Settings tab**. Change the settings to allow scrolling. When this Protean 3D user's guide calls for scrolling, do so by drawing one finger up or down along the right edge of the trackpad.

- Macintosh: Select **System Preferences > Trackpad**. Change the settings to allow scrolling. When this Protean 3D user's guide calls for scrolling, do so by drawing two fingers together across the trackpad.

Copyright DNASTAR, Inc. 2022

Change the default print size

Unless otherwise specified on your computer, the default printing size for all DNASTAR applications is “U.S. Letter” size (8.5 × 11 inches = 21.6 × 28 cm). You can temporarily change to another size, such as A4, from within the application. However, the next time you print, the default size will again be U.S. Letter size.

To correct this issue, you need to change your computer’s default printing size:

On Windows 7 and Windows 8:

1. Open the Control Panel.
2. In the Hardware and Sound category, click on **Printers** or **View devices and printers**.
3. Right-click on your printer and choose **Properties**.
4. In the **General** tab, click on the **Printing Properties** button.
5. In the ensuing **Printing Shortcuts** tab, choose the desired default printing size from the **Printing Sizes** drop-down menu.
6. Click the **Apply** button, then the **OK** button.

On Macintosh:

1. Open System Preferences.
2. Click on **Printers and Scanners**.
3. Choose the desired size from the **Default paper size** drop-down menu.

Note: Changes to the default printing size will not be applied until after you restart the DNASTAR application.

Copyright DNASTAR, Inc. 2022

Prevent firewall issues when downloading from the Motion Library

When attempting to download files from the Motion Library on Windows, you may receive the following error message: *I/O error when communicating with DNASTAR Motion Library Server.* This issue may occur if Windows Firewall is **On** and **Outbound Connections** are **Blocked**. Simply adding Protean 3D to the list of allowed programs is not likely to correct the issue. There are several possible work-arounds:

- Download the entire Motion Library from our website.
- Add one of these rules to the Windows Firewall Advanced Settings dialog, which is accessed through the Windows Control Panel's "Security Settings":
- Allow Protean 3D to make outbound HTTP connections.
- Allow Protean 3D to pass through the Firewall on port 80.

For more information on creating rules, consult Microsoft's support page for Windows Firewall Advanced Settings.

Copyright DNASTAR, Inc. 2022

Troubleshoot the error “A different version is running”

If you receive the error “A different version of Lasergene is running” when you try to launch a Lasergene application, but another version of Lasergene actually isn’t running, there may be a lingering .state file from another version that needs to be deleted.

The solution is to locate and delete files named *STARDM*.state*, …where ‘*’ represents the version number.

These files may exist in the following directories:

Windows 7 and Windows 8:

- *C:\Users\username\AppData\Local\DNASTAR\DataManager*
- *C:\ProgramData\DNASTAR\DataManager*

Macintosh:

- *Hard Drive:Library:Preferences:DNASTAR:DataManager*
- *Hard Drive:Users:username:Library:Preferences:DNASTAR:DataManager*

If you find any of the *.state* files listed above, delete them. Once the files are deleted, the error should no longer appear.

Copyright DNASTAR, Inc. 2022

Troubleshoot blank Protean 3D window on Parallels Desktop

The Protean 3D application launches, but all panes remain blank after opening a file. This issue has been observed when using Parallels Desktop 5 in combination with Microsoft Windows 7, but may also occur in other versions.

To resolve this issue, enable 3D acceleration for Parallels Desktop. See this Parallels web page for instructions.

Copyright DNASTAR, Inc. 2022

Appendix

Glossary

Most of the definitions in the table below are courtesy of the PDB Glossary of Technical Terms.

Term	Definition
Active site, binding site, or site	The region of a protein that binds substrates and that is involved in chemical reactions. Generally an active site contains very specific amino acids that participate in the binding.
Alpha carbon	The carbon in an amino acid to which the side chain is attached.
Alpha helix or helix	A secondary structural motif of a protein. It is characterized by hydrogen bonds between the carbonyl group (-C=O) of one amino acid and the amino (N-H) group of the amino acid 4 residues below it along the helix. The helix makes one complete turn every 3.6 amino acids.
Atom and bond	Base units for displaying molecular structure.
B-factor or temperature factor	A measure of how much the position of an atom deviates from that given in the atomic coordinates. This deviation is due to thermal motion and crystal imperfections. Temperature factors are also termed B-factors.
Beta sheet	A secondary structural motif of a protein. It is characterized by 2 or more amino acid chains lying side-by-side where hydrogen bonds are formed between carbonyl groups (C=O) of one chain and amino groups (N-H) of the adjacent chain. This arrangement forms a sheet-like structure.
CATH	A hierarchical classification of protein structural relationships which groups proteins at four major levels: Class©, Architecture(A), Topology(T) and Homologous superfamily (H).
Chain	A single biopolymer, either a polypeptide (protein) or polynucleotide (DNA, RNA).
CPK model	A space filling three-dimensional representation of a molecule. Given the centers of the molecule atoms and the relative van der Waals radii, a spherical CPK-type representation of a molecule can be built.
Data (NMR)	In NMR spectroscopy, the initial data collected include high-resolution multidimensional NMR spectra which reveal correlations between atoms that lie either within a few bonds of each other (J coupling) or within a short distance of each other (NOE coupling). NMR resonances are then assigned to specific atoms in the molecule and couplings are assigned to pairs or groups of atoms. These couplings are then used to produce a list of restraints that specify that particular atoms in the final model be near each other. Finally, a computer program is used to produce a model of the molecule which fits all of the restraints.
Data (X-ray)	In X-ray crystallography, the initial data include the measured positions and intensities of the reflections in the diffraction pattern produced by the macromolecular crystal. The relative phases of the waves that produced the diffraction pattern are also determined. An electron-density map of the molecule is then computed from the positions, intensities and phases of

Copyright DNASTAR, Inc. 2022

	all of the reflections. A model of the molecule contained within the crystal is then built to fit the electron-density map.
Disordered regions (NMR)	In NMR spectroscopy, disordered regions of molecules result in high variability between the models of the ensemble. Atom positions in these regions are very uncertain.
Disordered regions (X-ray)	In X-ray crystallography, disordered regions of molecules appears as weak regions of electron density and as regions with high temperature factors. Atom positions in these regions are very uncertain.
Disulfide bond	Covalent bond formed between the sulfur atoms of two cysteine residues in a protein. Disulfide bonds often stabilize protein structure or link multiple proteins in a complex.
Gene Ontology (GO)	A bioinformatics initiative to standardize the representation of gene and gene product attributes across species and databases. The ontology covers three domains: cellular component, molecular function and biological process.
Insertion code	A non-numerical component of the residue identifier that historically exists to shift the natural numbering of residues so that key residues maintain the same index number across homologous proteins.
Kabsch and Sander DSSP method (KSDSSP, KSD)	Features created by Protean 3D and shown/hidden on the Sequence View via the Features Panel. KSDSSP is an implementation of the method for assigning secondary structure developed by Kabsch and Sander (1983). In brief, the algorithm identifies hydrogen bonds along the backbone of a protein structure, then classifies the secondary structure based on the observed hydrogen bonding patterns. The original DSSP algorithm recognizes eight types of secondary structure (G = 3:10 helix, H = alpha helix, I = pi helix, E = beta strand, B = beta bridge, T = turn, S = high curvature, L = loop/other). KSDSSP reports three types of secondary structures: helix/H (G, H, I), strand/S (E, B), coil/C (T, S, L). Protean 3D uses the HSC classification from KSDSSP and creates features of helices and strands. Reference: Kabsch W. and Sander C. (1983) "Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features" Biopolymers 22:2577.
Ligand	A molecule that binds specifically to another molecule (usually a protein or nucleic acid) to form a complex. A ligand can be another protein, but is often a small molecule.
Model	A conformation of a set of molecules, including positions of atoms.
Nuclear Magnetic Resonance (NMR)	Nuclear magnetic resonance is a property that magnetic active nuclei have in a magnetic field when responding to applied electromagnetic pulses or perturbations. The measurements of such responses provide rich structural, dynamic, and kinetic information about the molecule.
Occupancy (X-	The occupancy of an atom is a measure of the fraction of molecules in the crystal in which

Copyright DNASTAR, Inc. 2022

ray)	the atom is actually present in the position specified by the model. If all of the molecules in the crystal are identical, then occupancies for all of the atoms are 1.00. Otherwise, two or more conformations may be observed for a small region of an atom, such as a side chain. The atoms are then given occupancies equal to the fraction of side chains in each distinct conformation.
Protein Data Bank (PDB)	The single repository of experimentally determined structures of proteins, nucleic acids and complex biomolecular assemblies as managed by the Worldwide Protein Data Bank. PDB can also refer to a particular file format for describing structures.
Residue	An amino or nucleic acid that is part of a biopolymer by means of a condensation reaction.
Secondary structure	The arrangement a protein arising from hydrogen bonding interactions within contiguous segments of the polypeptide backbone. The two main types of secondary structure are the alpha helix and the beta-pleated sheet.
Side chain	The variable group in an amino acid that is attached to the alpha carbon.
Structure	A non-empty set of models of a set of molecules.
Temperature factor	A measure of how much the position of an atom deviates from that given in the atomic coordinates. This deviation is due to thermal motion and crystal imperfections. Temperature factors are also termed B-factors.
Theoretical model	Theoretical models are produced by methods other than X-ray diffraction and NMR spectroscopy. Theoretical models include homology models and models derived from simulations of folding and molecular dynamics.
X-ray crystallography	A method used to determine the detailed three-dimensional structure of molecules present in a crystal in which an X-ray beam is aimed at the crystal and the resulting scattered rays are studied.

Copyright DNASTAR, Inc. 2022

Supported file types

For a list of file formats supported by this DNASTAR product, please see the File Formats page of our website.

Copyright DNASTAR, Inc. 2022

IUPAC codes

IUPAC codes for amino acids:

1-Letter Code	3-Letter Code	Amino Acid
A	Ala	Alanine
B	Asx	Aspartic Acid or Asparagine
C	Cys	Cysteine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
J	Xle	Leucine or Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
O	Pyl	Pyrrolysine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
U	Sec	Selenocysteine
V	Val	Valine
W	Trp	Tryptophan
X	Xaa	Unspecified or unknown
Y	Tyr	Tyrosine

Copyright DNASTAR, Inc. 2022

Z	Glx	Glutamic Acid or Glutamine

IUPAC codes for nucleotides:

Symbol	Description	Bases Represented			
A	adenosine/adenine	A			
C	cytidine/cytosine		C		
G	guanosine/guanine			G	
T	thymidine/thymine				T
U	uridine/uracil				U
W	weak	A			T
S	strong		C	G	
M	amino	A	C		
K	keto			G	T
R	purine	A		G	
Y	pyrimidine		C		T
B	not A		C	G	T
D	not C	A		G	T
H	not G	A	C		T
V	not T	A	C	G	
N or -	any base (not a gap)	A	C	G	T

The non-standard nucleotide codes “I” (Inosine) or “X” (xanthine) are not supported.

Copyright DNASTAR, Inc. 2022

Installed Lasergene file locations

The following file names use ‘x’ to represent the version number.

File Category	Application	Path
Application[2]	ArrayStar	Windows: C:\Program Files (x86)\DNASTAR\Lasergene x\ ArrayStar
	SeqNinja (command line)	Windows: C:\Program Files (x86)\DNASTAR\Lasergene x\ SeqNinjaCL
	All others	Windows: C:\Program Files (x86)\DNASTAR\Lasergene x Macintosh: /Applications/ DNASTAR/Lasergene x
Data Manager[2] (*DMx, DMx.exe)*	SeqBuilder Pro, Protean 3D, GeneQuest, MegAlign Pro	Windows: C:\Program Files (x86) \DNASTAR\Lasergene x Macintosh: /Applications/ DNASTAR/Lasergene x
Data Manager State File	SeqBuilder Pro, Protean 3D, GeneQuest, MegAlign Pro	Windows: C:\Program Data\ DNASTAR\DataManager Macintosh: ~/Library/ Application Support/ DNASTAR/DataManager, ~/Library/Preferences/ DNASTAR/DataManager
License Manager[2]	All	Windows: C:\Program Files (x86)\DNASTAR\License Manager Macintosh: /Applications/ DNASTAR/License Manager
Server License File (*lservrc*), Server License Manager [2], *Server Executables (_lserv, lservnt.exe*)	All	Windows: C:\Program Files (x86)\DNASTAR-LicenseServer\Server

Copyright DNASTAR, Inc. 2022

		Macintosh: ~/Library/DNASTAR-LicenseServer
Standalone & Trial Licenses (*.license), License Server Client License (*.lshost), Key Server Client License (*.keyhost)	All	Windows: C:\Program Data\DNASTAR\Licenses Macintosh: ~/Library/Application Support/DNASTAR/Licenses
Preferences	Protean 3D, Navigator, SeqNinja (DNA*), GenVision Pro	Windows: C:\Users\<user>\DNASTAR Macintosh: ~/Library/DNASTAR
	ArrayStar	Windows: C:\Users\<User>\AppData\Roaming\DNASTAR\ArrayStar
	All others	Windows: C:\Users\<user>\AppData\Local\DNASTAR\ Macintosh: ~/Library/Preferences

AppData is a hidden folder in Windows. To unhide the folder, go to **Organize > Folder** and **Search Options > View > Show Hidden files and folders**.

Copyright DNASTAR, Inc. 2022

Useful links

Name and Link	Description
CATH Protein Structure Classification Database	A curated classification of protein domain structures. Each protein is assigned to a homologous superfamily (group of domains related by evolution). The CATH classification procedure uses a combination of automated and manual techniques which include computational algorithms, empirical and statistical evidence, literature review and expert analysis.
Chemical Component Dictionary (CCD)	An external reference file describing all residue and small molecule components found in PDB entries. Protean 3D's algorithm uses the CCD to look up information about non-standard residues and ligands (e.g. common name, most similar standard residue) when parsing PDB files. The full version of the most recent CCD was included with your Protean 3D installation. DNASTAR will continue to update the file as it encounters new CCD entries.
Gene Ontology	A bioinformatics initiative to standardize the representation of gene and gene product attributes across species and databases.
RCSB Protein Databank and Worldwide Protein Databank	Open-source databases of annotated protein files output by crystallographic or NMR instruments. Files can be downloaded in Protein Databank (PDB) format and other file formats. To learn more: * PDB Getting Started Guide * Narrated PDB tutorial from Open Helix * Narrated tutorials from PDB
Morph Server (part of the Database of Macromolecular Movements)	A database of protein motion video clips. Also provides tools for protein flexibility and geometric analysis.

Copyright DNASTAR, Inc. 2022

Research references

Chou PY (1990). "Prediction of protein structural classes from amino acid composition." In Prediction of Protein Structure and the Principles of Protein Conformation, Plenum Press, New York, NY, pp. 549-586.

Chou PY and Fasman GD (1978). "Prediction of the secondary structure of proteins from their amino acid sequence." Advances in Enzymology 47: 45-148.

Claros MG and von Heijne G (1994). "TopPred II: An Improved Software For Membrane Protein Structure Predictions." Comput. Appl. Biosci. 10(6): 685-686. [Reference pertains to Protean 3D's TopPred algorithm].

Deléage G and Roux B (1987). "An algorithm for protein secondary structure prediction based on class prediction." Protein Engineering 1: 289-294.

Deveaud and Schuerer (Pasteur Institute). "New implementation of the original TopPred program, based on the G. von Heijne algorithm."

Eisenberg D, Weiss RM and Terwilliger TC (1984). "The hydrophobic moment detects periodicity in protein hydrophobicity." Proceedings of the National Academy of Science (USA), 81(1): 140-144.

Emini EA, Hughes J, Perlow D and Boger J (1985). "Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide." Journal of Virology, 55: 836-839.

Engleman DM, Steitz TA and Goldman A (1986). "Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins." Annual Review of Biophysics and Biophysical Chemistry, 15: 321-54.

Garnier J, Gibrat JF, Robson B (1996). "GOR method for predicting protein secondary structure from amino acid sequence." Methods Enzymol. 266:540-53. [Reference pertains to Protean 3D's GOR IV algorithm].

Garnier J, Osguthorpe DJ and Robson B (1978). "Analysis of the accuracy and implications of simple method for predicting the secondary structure of globular proteins." Journal of Molecular Biology, 120: 97-120. [Reference pertains to Protean 3D's Garnier-Robson method].

Garnier J and Robson B (1990). "The GOR Method for Predicting Secondary Structures in Proteins." In Prediction of Protein Structure and the Principles of Protein Conformation, ed. G.D. Fasman (2nd ed). Plenum Press, New York. [Reference pertains to Protean 3D's GOR II algorithm].

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD and Bairoch A (2005). "Protein Identification and Analysis Tools on the ExPASy Server." In The Proteomics Protocols Handbook, ed. John M. Walker, Humana Press, pp. 571-607.

Gill SC and von Hippel PH (1989). "Calculation of protein extinction coefficients from amino acid sequence data." Analytical Biochemistry, 182: 319-326.

Copyright DNASTAR, Inc. 2022

Guruprasad K, Reddy BVB and Pandit MW (1990). "Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence." Protein Eng., 4, 155 – 161.

Hopp TP and Woods KR (1981). "Prediction of protein antigenic determinants from amino acid sequences." Proceedings of the National Academy of Science, 78: 3824-3828.

Jameson BA and Wolf H (1988). "The antigenic index: a novel algorithm for predicting antigenic determinants." CABIOS, 4: 181-186.

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

Karplus PA and Schultz GE (1985). "Prediction of chain flexibility in proteins." Naturwissenschaft, 72: 212-213.

Krebs WG and Gerstein M (2000). “The Morph Server: A standardized system for analyzing and visualizing macromolecular motions in a database framework.” Nucleic Acids Res 28: 1665-75. [Pertains to the Motion Library.]

Kyte J and Doolittle RF (1982). "A simple method for displaying the hydropathic character of a protein." Journal of Molecular Biology, 157: 105-132.

Lathrop RH, Webster TA, and Smith TF (1987). “Ariadne: Pattern-directed inference and hierarchical abstraction in protein structural recognition.” Communications of the ACM 30: 909-921. [Pertains to Ariadne patterns.]

Lehninger A (2005). Principles of Biochemistry, 4th Edition, Chapter 3, page 78, Table 3-1.

Lehninger A, Nelson D, and Cox M (2005). “Principles of biochemistry.” New York: W.H. Freeman.

Lupas A, Van Dyke M and Stock J (1991). "Predicting coiled coils from protein sequences." Science 252: 1162-1164.

Margalit H, Berzofsky J et al. (1987). "Prediction of immunodominant helper T cell antigenic sites from the primary sequence." Journal of Immunology, 138: 2213-2229.

Nakashima H, Nishikawa K and Ooi T (1986). “The folding type of a protein is relevant to the amino acid composition.” Journal of Biochemistry (Tokyo), 99: 153-162.

Parker JMR, Guo D and Hodges RS (1986). “New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and x-ray-derived accessible sites.” Biochemistry, 25: 5425-5432.

Parry DA (1982). “Coiled-coils in alpha-helix-containing proteins: analysis of the residue types within the

Copyright DNASTAR, Inc. 2022

heptad repeat and the use of these data in the prediction of coiled-coils in other proteins." Biosci Rep. 1982 Dec;2(12):1017-24.

Pierce NA, Spriet JA, Desmet J, Mayo SL. (2000) "Conformational splitting: A more powerful criterion for dead-end elimination." J Comput Chem. 21:999-1009.

Prlic A, Bliven S, Rose PW, Bluhm WF, Bizon C, Godzik A and Bourne PE (2010). "Precalculated Protein Structure Alignments at the RCSB PDB website." Bioinformatics. 26(23):2983-5.

Ren J et al. (2009). "Systematic study of protein sumoylation: Development of a site-specific predictor of SUMOsp 2.0.""Proteomics, 9: 3409-3412.

Rothbard JB and Taylor WR (1988). "A sequence pattern common to T cell epitopes." EMBO Journal, 7: 93-100.

Sasagawa T, Okuyama T and Teller D (1982). "Prediction of peptide retention times in reversed-phase high-performance liquid chromatography during linear gradient elution." Journal of Chromatography, 240: 329-340.

Sette A et al. (1989). "Prediction of major histocompatibility complex binding regions of protein antigens by sequence pattern analysis." Proceedings of the National Academy of Science, 86: 3296-3300.

Troshin PV, Procter JB and Barton GJ (2011). "Java Bioinformatics Analysis Web Services for Multiple Sequence Alignment – JABAWS:MSA." Bioinformatics; doi: 10.1093/bioinformatics/btr304.

von Heijne G (1981). "On the hydrophobic nature of signal sequences." Eur. J. Biochem., 116: 419-422. [Reference pertains to Protean 3D's von Heijne method].

von Heijne G (1992). "Membrane Protein Structure Prediction: Hydrophobicity Analysis and the 'Positive Inside' Rule." J. Mol. Biol. 225(2): 487-494. [Reference pertains to Protean 3D's TopPred algorithm].

Webster TA, Lathrop RH., and Smith TF (1987). "Prediction of a common structural domain in aminoacyl-trna synthetases through use of a new pattern-directed inference system." Biochemistry 26: 6950-6957. [Pertains to Ariadne patterns.]

Webster TA, Tsai H, Kula M, Mackie GA and Schimmel PR (1984). "Specific sequence homology and three-dimensional structure of an aminoacyl transfer RNA synthase." Science 226: 1315-1317. [Pertains to Ariadne patterns.]

Welling GW, Weijer WJ, van der Zee R and Welling-Wester S (1985). "Prediction of sequential antigenic regions in proteins." FEBS Letters, Vol. 188, No. 2: 215-218.

White A, Handler P and Smith E (1964). Principles of Biochemistry, 3rd Edition, p. 124.

Xu D and Zhang Y (2009). "Generating triangulated macromolecular surfaces by Euclidean Distance

Copyright DNASTAR, Inc. 2022

Transform." PLoS One 4(12):e8140. [Reference pertains to Protean 3D's EDTSurf algorithm].

Yang ZR, Thomson R, McMeil P and Esnouf RM (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 21: 3369-3376.

Ye Y and Godzik A. (2003) "Flexible structure alignment by chaining aligned fragment pairs allowing twists." Bioinformatics. 19 Suppl 2:ii246-55.

Zhang Y and Skolnick J (2005). "TM-align: a protein structure alignment algorithm based on the TM-score." Nucleic Acids Res. 33(7):2302-9. [Reference pertains to Protean 3D's TM-align algorithm].

Copyright DNASTAR, Inc. 2022

Acknowledgements

DNASTAR thanks the authors for the PDB structure files used in our written tutorials:

Structure of the triosephosphate isomerase-phosphoglycolohydroxamate complex: an analogue of the intermediate on the reaction pathway.

Davenport, R.C., Bash, P.A., Seaton, B.A., Karplus, M., Petsko, G.A., Ringe, D.

Journal: (1991) Biochemistry **30:** 5821-5826

PubMed: 2043623
Search Related Articles in PubMed

PubMed Abstract:

The glycolytic enzyme triosephosphate isomerase (TIM) catalyzes the interconversion of the three-carbon sugars dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde 3-phosphate (GAP) at a rate limited by the diffusion of substrate to the enzyme. We have solved the three-dimensional structure of TIM...

The crystal structure of benzoylformate decarboxylase at 1.6 A resolution: diversity of catalytic residues in thiamin diphosphate-dependent enzymes.

Hasson, M.S., Muscate, A., McLeish, M.J., Polovnikova, L.S., Gerlt, J.A., Kenyon, G.L., Petsko, G.A., Ringe, D.

Journal: (1998) Biochemistry **37:** 9918-9930

PubMed: 9665697
DOI: 10.1021/bi973047e
Search Related Articles in PubMed

PubMed Abstract:

The crystal structure of the thiamin diphosphate (ThDP)-dependent enzyme benzoylformate decarboxylase (BFD), the third enzyme in the mandelate pathway of Pseudomonas putida, has been solved by multiple isomorphous replacement at 1.6 A resolution and refined to an R-factor of

Crystal structure of the thiamin diphosphate-dependent enzyme pyruvate decarboxylase from the yeast Saccharomyces cerevisiae at 2.3 A resolution.

Arjunan, P., Umland, T., Dyda, F., Swaminathan, S., Furey, W., Sax, M., Farrenkopf, B., Gao, Y., Zhang, D., Jordan, F.

Journal: (1996) J.Mol.Biol. **256:** 590-600

PubMed: 8604141
DOI: 10.1006/jmbi.1996.0111
Search Related Articles in PubMed

PubMed Abstract:

The crystal structure of pyruvate decarboxylase (EC 4.1.1.1), a thiamin diphosphate-dependent enzyme isolated from Saccharomyces cerevisiae, has been determined and refined to a resolution of 2.3 A. Pyruvate decarboxylase is a homotetrameric enzyme which crystallizes with two

Copyright DNASTAR, Inc. 2022