Table of Contents

What is LoopGrafter?

LoopGrafter is a web-based tool designed to provide assistance and guidance in the process of transferring one or more loops in between homologous proteins. In a protein-engineering context, novel properties can be introduced to the designed product by means of loop replacement (Schenkmayerova *et al.* (2021), Heinemann *et al.* (2021)). This process is also known as loop transplantation or **loop grafting**, and is also crucial in humanizing antibodies (Riechmann *et al.* (1988)). Here we refer to the protein that will accept the loop(s) as the **scaffold** protein and the one that serves as donor as the **insert** protein. The result of the interchange is a chimeric protein consisting mainly of the *scaffold* core with some *insert*ed loops that we refer to as **grafted design**. The server makes special emphasis on assessing the **geometrical compatibility** of the loops being interchanged and the **flexibility** properties of the regions selected for loop grafting.

Schenkmayerova A. *et al.* Engineering the protein dynamics of an ancestral luciferase. (2021) *Nat Commun.* **12**, 3616.

Heinemann P.M. *et al.* Active-site loop variations adjust activity and selectivity of the cumene dioxygenase. (2021) *Nat Commun.* **12**, 1095.

Riechmann L. *et al.* Reshaping human antibodies for therapy. (1988) *Nature*. **332**, 323-7.

You have highlighted few keywords on your previous general definition… why did you do so?

The keywords highlighted in bold-type font correspond to key concepts in this LoopGrafter web server. Some of the concepts might be slightly divergent from common perception or intuition, so we will define them in detail:

> ***Loop****:* A loop in LoopGrafter is understood as the set of contiguous amino-acids consisting of a coiled coil segment in a protein and its two flanking regular secondary structures. In structural biology related fields, often only the coiled coil parts are considered as *loops* but here we follow the definitions from a previous classification of protein loops in Bonet *et al*.

(2014). The inclusion of the flanking secondary structures on the definition of the loop enables calculating its geometry and provides further space for performing a good structural interchange in between the scaffold and the insert proteins.

Bonet J. *et al.* ArchDB 2014: structural classification of loops in proteins. (2014) *Nucl Acids Res.* **42,** D315-9.

Homologous proteins: The server helps transferring loops in between two homologous proteins. This is, proteins that are evolutionary and structurally related. Our recommendations based on Kufareva & Abaigan (2012) are that they differ no more than 8 Angstroms RMSD after global superimposition and no more than 20% in sequence length after global alignment. However, the example that can be found below employs proteins that are even more related: scaffold and insert only differ 1.27 Angstrom RSMD and 2% in sequence length.

Kufareva I. & Abaigan R. (2012) Methods of protein structure comparison. *Methods Mol. Biol.* **857**, 231-57.

Loop grafting: Is the process of transferring one or more loops from one protein to another. LoopGrafter is designed to assist in this task.

Scaffold protein*:* Is the protein on which the new loops will be inserted. Meaning that, this protein will provide most of the sequence for the final product of the grafting process.

Insert protein*:* Is the protein from which the selected loops will be extracted and then transferred to the scaffold protein. This protein only provides the selected loops to the final product of the grafting process.

Grafted design: Is the result of the loop grafting process. A chimeric protein consisting mainly of the scaffold protein sequence in which the regions selected for loop grafting have been replaced by their insert protein counter parts.

Loop geometry: The aforementioned classification of protein loops (Bonet et al. (2014)) exploits three angles (δor hoist, θ or packing, and ρ or meridian) and the distance in between the extreme positions of the coiled coil segment of the loop to define its geometry. Here, we present the same metrics for comparison in between different loops. In the classification of loops, cohesive loop class allows up to 15° divergence in the first two angles (δand θ), up to 30° divergence in the third one (ρ), and up to 1 Angstrom difference in the distance parameter D. We strongly recommend grafting loops in between regions from the scaffold and the insert protein that comply with these geometrical criteria.

Protein flexibility*:* It is often the case that flexible regions of a protein are especially interesting targets for engineering. Coiled coil segments have less structural restraints than periodic regular structure elements, and thus are usually more flexible than other parts of the protein. Still, some loops might be more flexible than others, and based on their flexibility properties and

the goals of the engineering experiment certain loops might be better interesting targets for grafting. LoopGrafter provides a quick assessment of the flexibility of the protein based on Normal Mode Analysis (NMA), and a convenient way to visually inspect the differences in flexibility properties in different parts of the input scaffold protein.

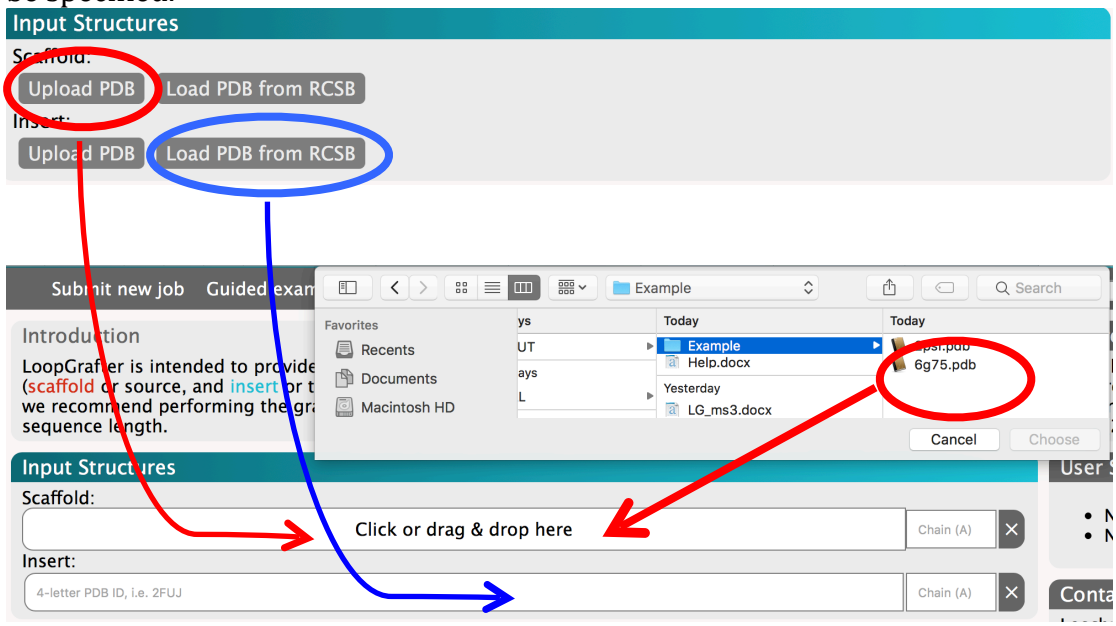
OK. LoopGrafter will generate chimeric proteins in which some loops are transplanted. But… where are those loops exactly inserted in? Which precise positions in my input scaffold protein can accept an insert loop? And how long from the insert protein will be placed in that position?

LoopGrafter structurally superimposes the input scaffold and insert proteis using Combinatorial Extension (CE) algorithm (Shindyalov & Bourne (1998)). A sequence pairing is derived from such structural superimposition using an *ad hoc* designed greedy algorithm, which couples residues from each of the input proteins that are in close distance proximity in the structural superimpostion. Once the user selects the loop or loops to be grafted, all the combinations of insertion points along the flanking regular secondary structures are explored, and those that generate sequence variability are chosen for three-dimensional modelling and evaluation.

Shindyalov I.N. & Bourne P.E. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. (1998) *Protein Eng.* **11**, 739-47.

Let's use the server… How do I input my proteins? What is that license agreement I'm asked about?

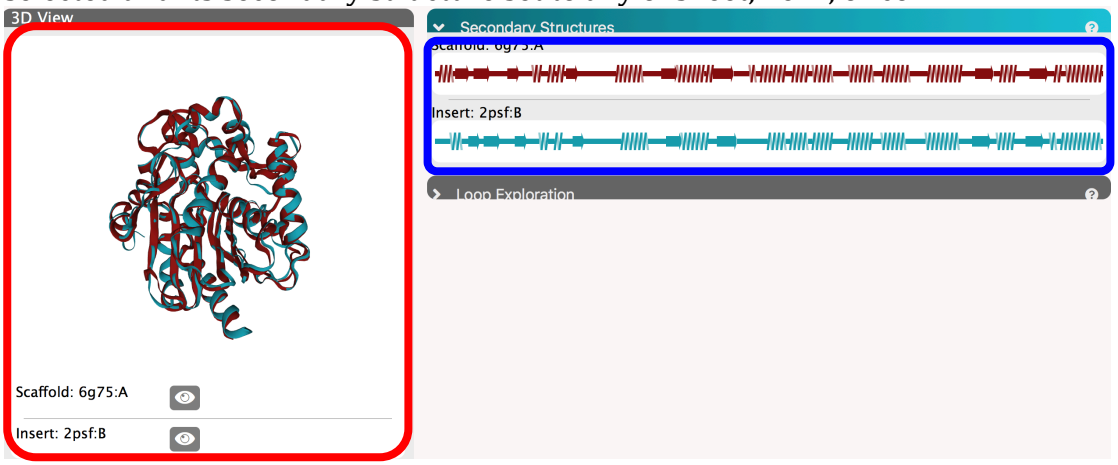
The input proteins can be automatically fetched from the RCSB Protein Data Bank (PDB) or alternatively can be uploaded as PDB-formated files from the user local system. The polypeptidic chain to be considered from the PDB file needs to be specified.



Unfortunately, some pieces of software that LoopGrafter makes use of for three-dimensional modelling and energy evaluation are subject to licensing agreement. These pieces of software are required to generate 3D structures of the grafted designs and to predict their stability. Without the use of such software, LoopGrafter can provide with the sequences of the grafted designs which the user can freely take for further experimentation. In order to take full advantave of these pieces of software (namely MODELLER and Rosetta), academic users are waived from any fee but still required to state they comply with such academic license agreement. Commercial users **must** reach previous agreement with the owners of the third party software rights.

I've loaded my proteins, I see them superimposed and I am presented with a 2D representation of the secondary structure elements of the input scaffold and insert proteins. What can I do here?

LoopGrafter superimposes the input structures using CE (Shindyalov & Bourne (1998)) and determines their secondary structure elements using DSSP (Kabsch & Sander (1983)) . The superimposition is visualised in the 3D view at the left of the screen (red square) and the secondary structure assignment is visualised in the 2D view at the right of the screen (blue square). Here, the automatic secondary structure assignment can be modified. A region on the 2D view can be selected and its secondary structure set to any of sheet, helix, or coil.



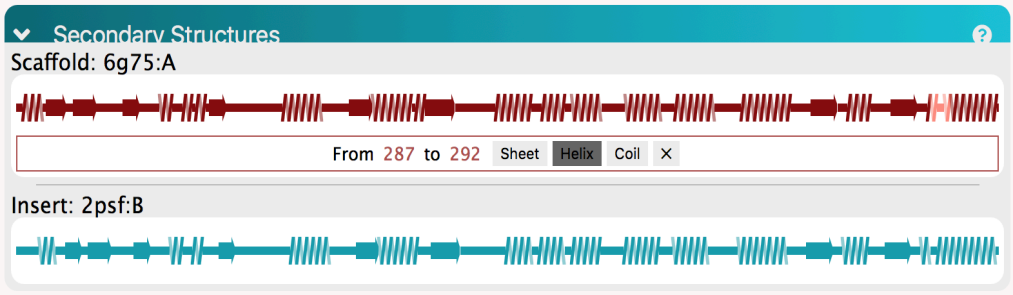
There are several purposes for changing the automatically assigned secondary structure elements:

- First, **the regions to be grafted should be the most similar possible in between the scaffold and the insert**. Often, there are small structural elements inside the loop that would disrupt the definition of the loop itself (a coiled-coil region and its flanking regular secondary structure elements). Here, such small structural elements can be removed or reverted to coil. After the manipulation, the topology of the scaffold and the insert is similar.

Changing a small helix to coil from the scaffold protein to match the topology of the insert.

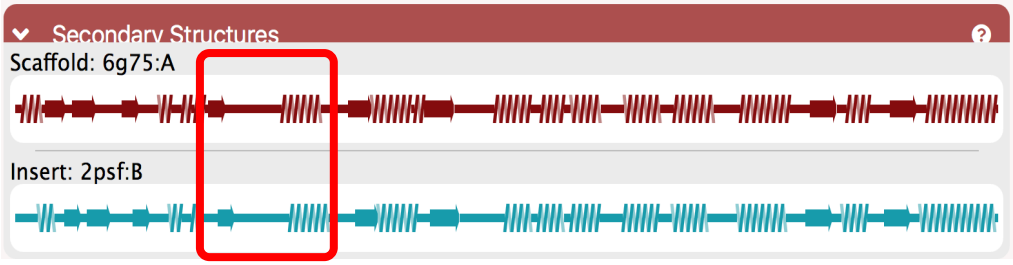
The topology of the scaffold and insert protein is similar after the manipulation.

- Second, certain structures may contain **apparently broken secondary structure elements**, or with discontinuities inside them. Short flanking regular secondary structure elements make poor insertion points, and if by means of repairing such discontinuities the regular secondary structure element can be extended, LoopGrafter can explore better options for inserting loops in the selected region.

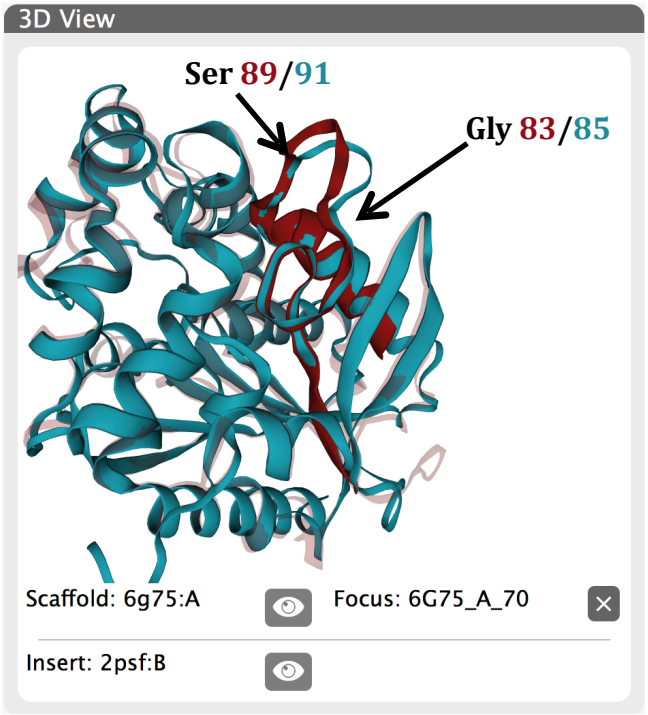


Extending a broken helix.

- Third, **grafting smaller regions**. LoopGrafter tries to find insertion points in the flanking secondary structures of the grafted loop. It is not uncommon that design efforts are targeted only to the coiled-coil part of the loop (i.e. grafting complementarity determining regions or CDRs in antibodies). If the superimposition of the initially assigned coil region is good in between the two proteins, the flanking regular secondary structures can be extended “inwards” to allow LoopGrafter to produce grafted designs that contain smaller portions of the insert protein.



The region in the box is considered for grafting, but the flanking N-terminal β-sheet is very short and the coiled region very large. The researchers in charge of this design would like to change only residues inside the coil part.



Fortunately, the superimposition of the coil part is good for the two proteins, and the flanking secondary structures can be extended, the N-terminal β-sheet up to residue glycine 83 and the C-terminal α-helix up to residue serine 89. The same extension needs to be done for the insert protein counterpart.

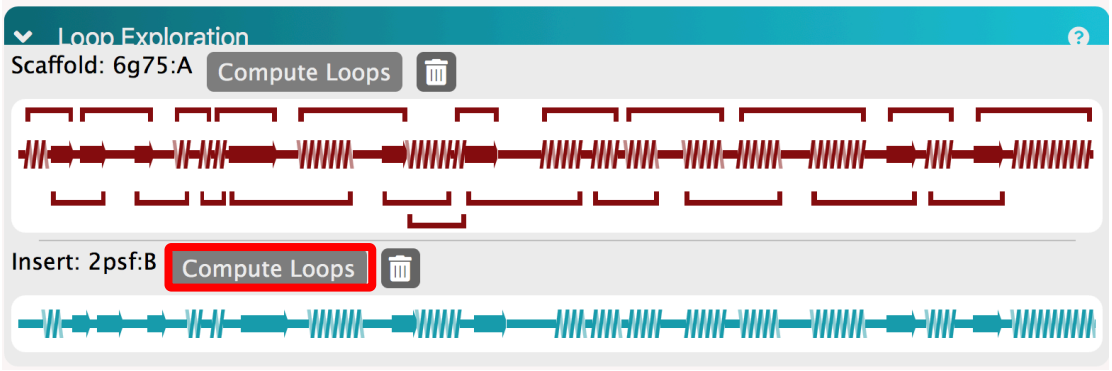
Result after extending the flanking secondary structures.

Shindyalov I.N. & Bourne P.E. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. (1998) *Protein Eng.* **11**, 739-47.
Kabsch W. & Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. (1983) *Biopolymers.* **22**, 2577-37.

Great! I have defined the secondary structures of my input proteins as I want them to be. What is next?

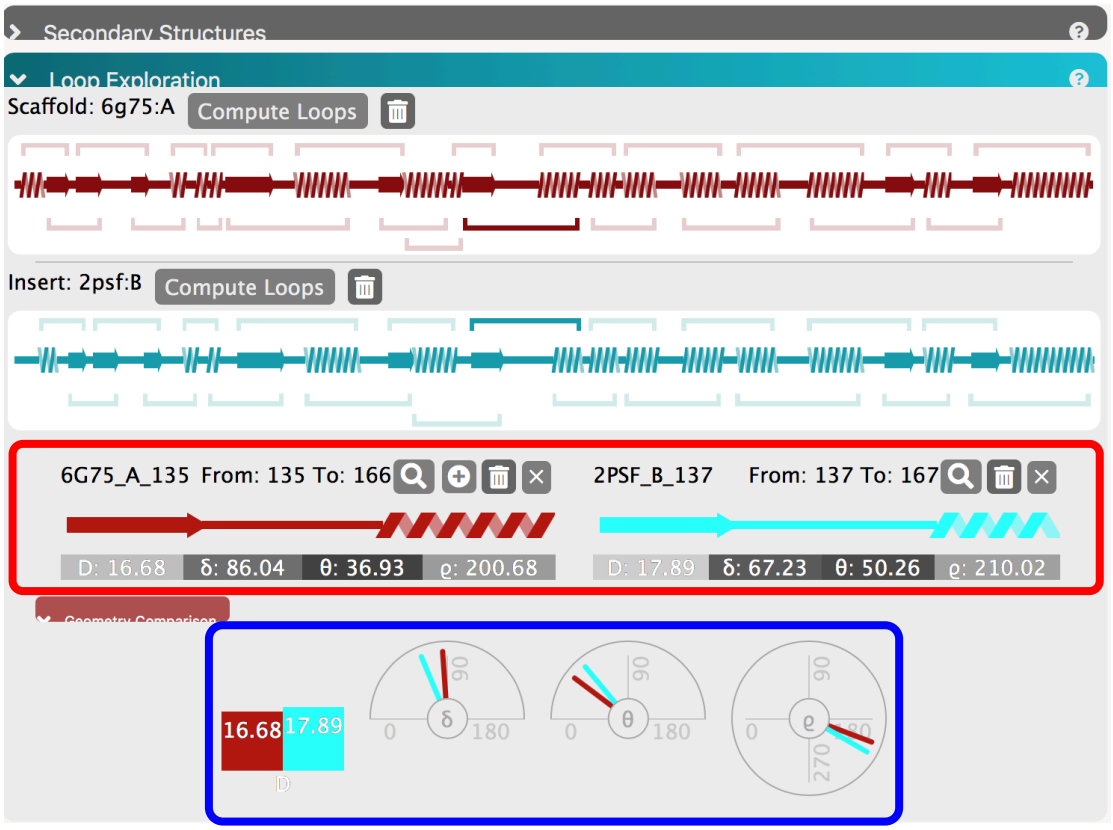
Based on the current definitions of secondary structures, LoopGrafter calculates loops according to Bonet et al. (2014) on both input proteins. To get the

calculations done, the user only needs to click on the “Compute Loops” button for each of the input proteins. Once the loops are calculated, they are shown as square brackets above or below the 2D representation of the protein.



Loops have been calculated for the Scaffold protein. By clicking on the “Compute Loops button” (red rectangle) loops will be calculated for the Insert protein.

Clicking on any square bracket representing a loop brings additional information about it. Details about its geometry (distance D and hoistδ, packing θ , and meridian ρ angles) are shown upon hovering the mouse over and are in a separate view below after clicking on the loop. If two or more loops are selected, their geometries can be directly compared in a dedicated view.

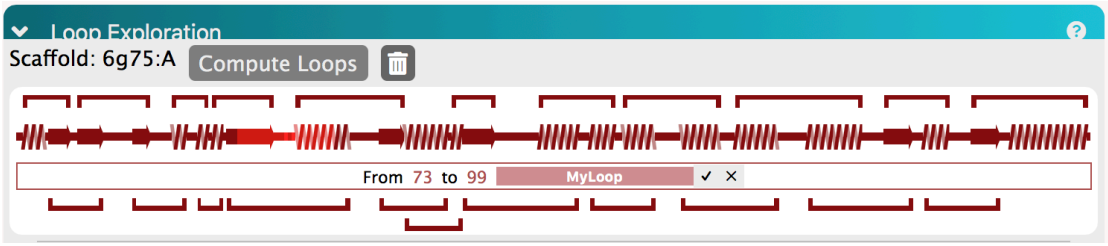


Loop details (red box) and geometrical comparison (blue box) after selection (highlighted loops in the upper view)

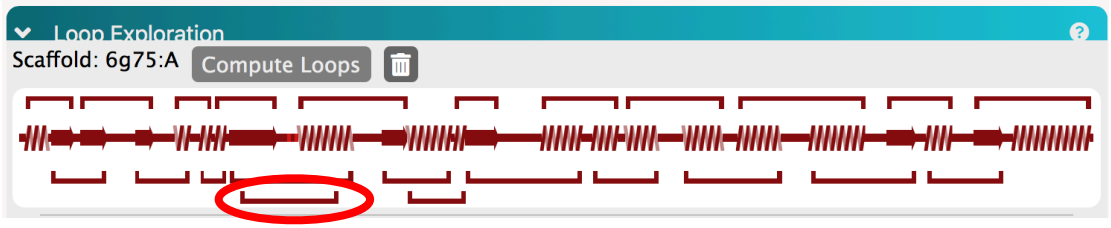
Selected loops can be highlighted in the 3D view (magnifying glass icon) and can be removed from the protein (trash bin icon) or from the selection (cross icon).

Also, loops from the scaffold protein can be selected for replacement (plus sign icon).

In this step, the user can define their own loops if so desires. This is an alternative to manipulating the assignment of secondary structure elements. A new loop can be defined by clicking and dragging on the region where the loop is to be defined.



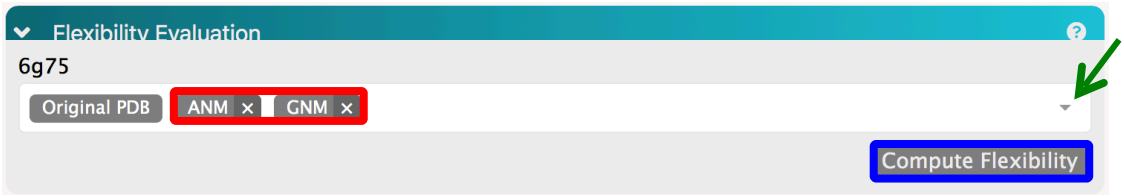
Definition of a new loop in the Scaffold protein. The effect of this action will be similar to shortening the flanking secondary structures of the loop.



Newly defined loop.
Bonet J. *et al.* ArchDB 2014: structural classification of loops in proteins. (2014) *Nucl Acids Res.* **42,** D315-9.

Up so far I have secondary structures and loops defined on my proteins. The next step reads Flexibility evaluation. What is this about?

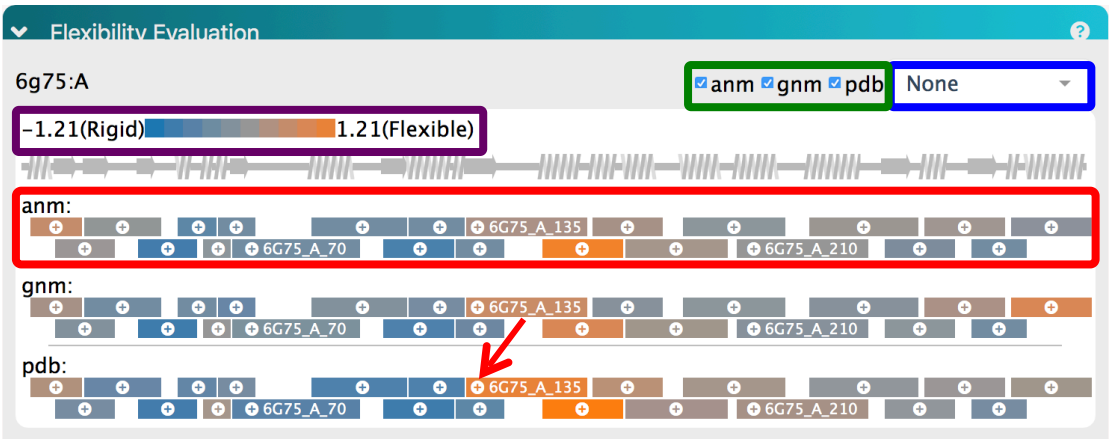
Coiled-coil parts of proteins tend to be flexible, some more than others. This often is a source of problems in protein production, and might be one of the causes inspiring loop grafting. In this step LoopGrafter reads flexibility values in the form of B-factors from the input PDBs when available, offers the possibility to predict flexibility using two different Elastic Network Models (Gaussian or GNM and Anisotropic or ANM) and also offers the user to input their own flexibility values for the scaffold protein.



ANM and GNM calculations can be deactivated by clicking on the cross icon (red box). Custom flexibility values can be inputted from a local file by unfolding the menu pointed by the green arrow. Pressing the button framed in the blue box starts the flexibility calculation.

All these flexibility assessments are visualized in an abstract linear representation of the scaffold protein, and are given as an average value per loop.

Additionally, flexibility values per residue or per secondary structure element can be visualized in a 2D representation of the protein. The flexibility methods visualized can be selected on demand.



The average flexibility value for each loop is represented in the red box as calculated by ANM. The flexibility is colour-coded according to the scale highlighted in the purple box. Individual methods providing flexibility information can be displayed on demand by selecting or de-selecting them in the green box. The menu depicted in the blue box allows the user to display the flexibility values per residue or per secondary structure in a 2D representation of the scaffold protein. Clicking on the "plus" sign adds the loop the list of loops to be replaced (red arrow).

Scaffold Loop List

Name	From	To	anm	gnm	pdb
Candidate					
6G75_A_135	135	166	0.21	0.53	0.96
Preserved					
6G75_A_14	14	26	0.47	0.18	0.29
6G75_A_21	21	35	0.04	-0.29	-0.4
6G75_A_29	29	48	-0.06	-0.48	-0.56
6G75_A_44	44	58	-0.91	-0.96	-0.93
6G75_A_55	55	64	-0.68	-0.46	-0.6
6G75_A_62	62	68	-0.11	-0.03	0.14
6G75_A_66	66	74	-0.44	-0.59	-0.39
6G75_A_70	70	103	-0.48	-0.31	-0.73
6G75_A_92	92	118	-0.46	-0.25	-0.83
6G75_A_112	112	130	-0.86	-0.84	-0.95
6G75_A_119	119	134	-0.43	-0.38	-0.85
6G75_A_132	132	143	-0.44	-0.53	-0.7
6G75_A_156	156	176	1.04	0.72	1.21
6G75_A_170	170	187	0.18	-0.19	0.39
6G75_A_179	179	205	0.16	0.12	0.17
6G75_A_195	195	221	-0.11	-0.13	0.15
6G75_A_210	210	244	0.01	-0.18	0.18
6G75_A_230	230	258	0.05	-0.22	-0.1
6G75_A_251	251	268	-0.32	-0.11	-0.1
6G75_A_262	262	282	0.09	0.2	0.03
6G75_A_275	275	288	-0.48	-0.48	-0.38
6G75_A_286	286	306	-0.15	0.73	0.12

Each loop representation has a plus sign icon on it. Clicking on that icon adds the loop to the list of loops to be replaced (this is, the loop is selected for grafting). Below the 3D view, there is a panel that keeps the user informed at all times of the selection for replacement status of the loops from the scaffold protein.
Snapshot of the Scaffold Loop list, showing that one loop has been selected as a candidate for replacement (grafting). Clicking on the magnifying glass icon in the list highlights the loop in the 3D view panel just above.

Now that I selected a loop for replacement, two new tabs appear below: "Correlation Evaluation" and "Loop Pairing". What should I do?

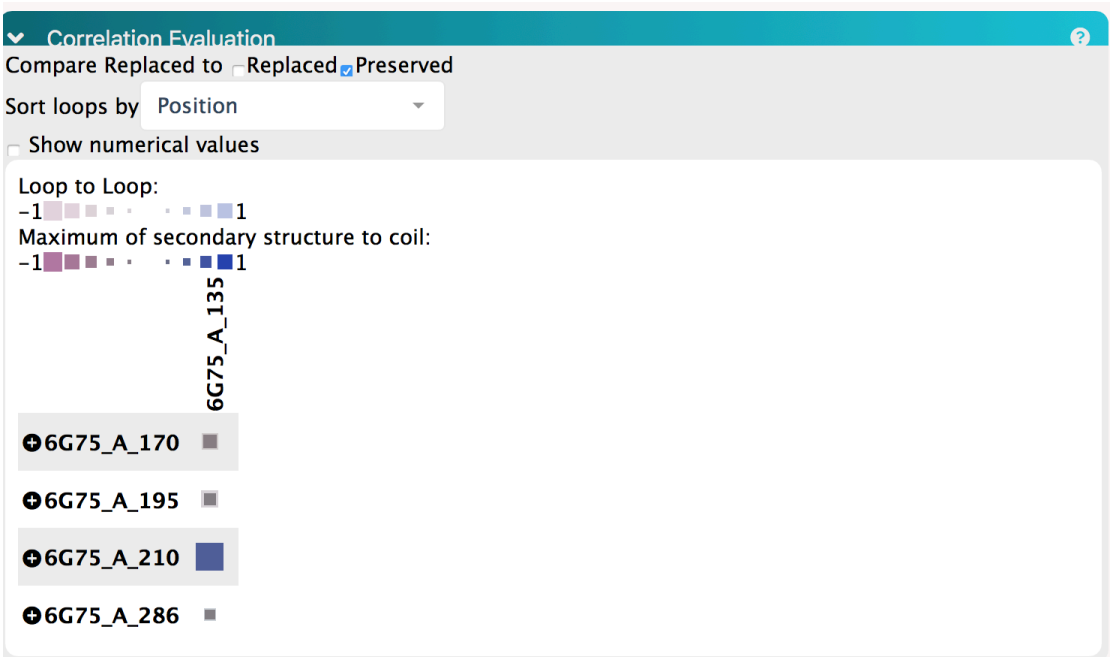
Once at least one loop has been selected for grafting, the server is ready to offer the user the possibility to match the scaffold loop with its insert protein counterpart and proceed with the proper grafting of the loop. However, it might be interesting to stop for a minute and loop how the selected loop moves in the protein in relation to other loops. The "Correlation Evaluation" step is designed

for this task. Specifically, if the user has opted for calculating the Anisotropic Network Model, this option will be available.

Here the dynamic behaviour of the selected loop(s) (vertical tags) is compared to that of others in the scaffold protein (horizontal tags) in the form of motion cross-correlations. This measure indicates if the two regions (loops) move along in the same direction over time. Positive cross-correlations indicate similar direction of motions over time even if there is a time lapse in between the motions of the two elements (like a pair of windscreen wipers do in cars). Negative cross-correlations indicate opposite direction of motions in the same conditions (like double swing saloon doors do in western films).

Two metrics are presented at the same time: “Loop to Loop” is shown in light shade and indicates the overall cross-correlation between the two loops. The second one, “Maximum, of secondary structure to coil”, is shown in dark shade and might result more interesting but requires an extra bit of explanation. Here, the cross-correlation represented is that of the coiled-coil part of the selected loop(s) to the secondary structure elements of any other loop in the protein. This figure is relevant since the coiled-coil part in loops is often responsible for certain biophysical effects such as occluding cavities or access pathways, or modulating the affinity for a particular ligand. This metric gives a more specific picture about how other parts of the protein move along with the relevant section of the selected loop(s).

Loops that are highly cross-correlated (either positively or negatively) might be interesting for grafting altogether. This is because if such motions are coordinated it is likely that they are co-regulated as well. By transplanting all the elements relevant to such regulation is more likely that the behaviour of such loops in the insert protein will be transferred to the scaffold one. Also, if such motions are important to the loops biological functions, transplanting them altogether increases the possibility that such functions are transferred to the grafted design.

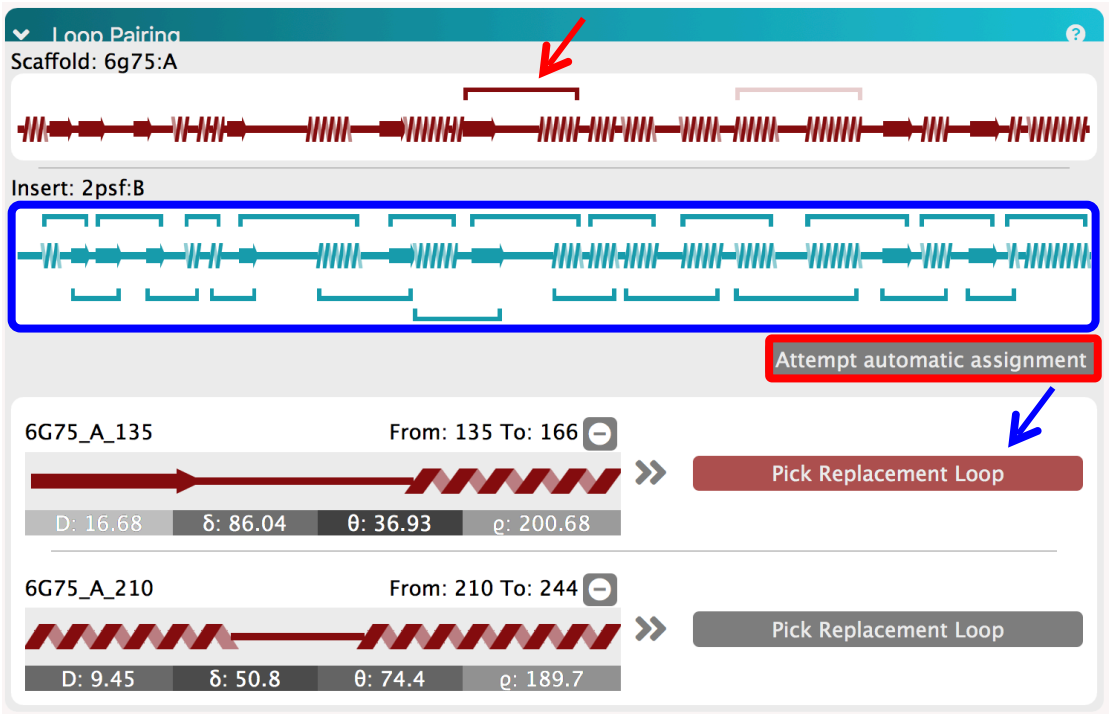


Cross-correlation evaluation view. The loop 6G75_A_210 is highly cross-correlated to the selected loop 6g75_A_135 and it might be interesting to select it for grafting as well.

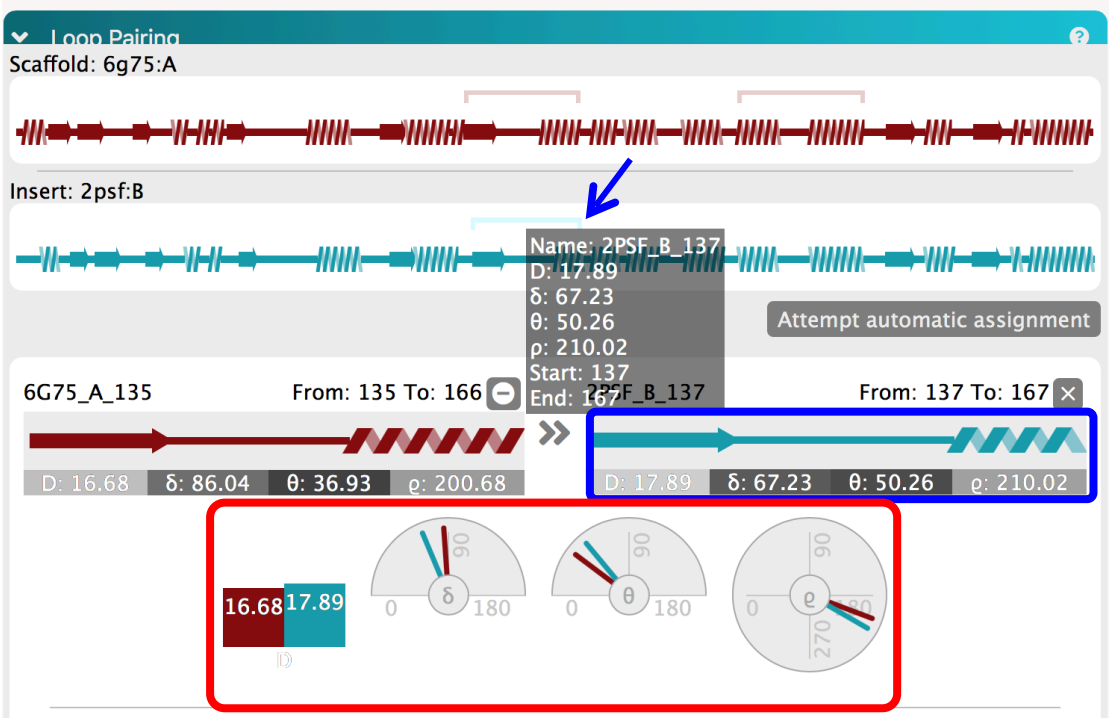
Finally into Loop Paring! Should I click on the "Attempt automatic assignment" button?

The only purpose of this step is to indicate which loops from the insert protein will be used to replace the selected loops in the scaffold one. For the grafting process to be successful, the loops need to be structurally equivalent and well superimposed in the 3D view. Thus, there is not much of an option. The "Attempt automatic assignment button" will try to do this job automatically. However, if the user has been adjusting the definition of secondary structures or loops, optimal solutions might not be found by the automatic pairing. Thus, the option of doing the paring manually is also available.

To manually pair loops, either click on the bracket representation of the scaffold loop in the 2D view of the protein or on the corresponding "Pick Replacement Loop" button. The loops on the insert protein become available for selection and by clicking any of them the pairing is made. A geometrical comparison of the loops to be replaced (in the scaffold) and its replacement (from the insert protein) is conveniently presented at this moment. We remind the recommendations for geometrical compatibility: a maximum divergence of 15°for hoist (δ) and packing (θ) angles, up to 30° for meridian (ρ) angle, and up to 1 Angstrom difference in the distance parameter D.



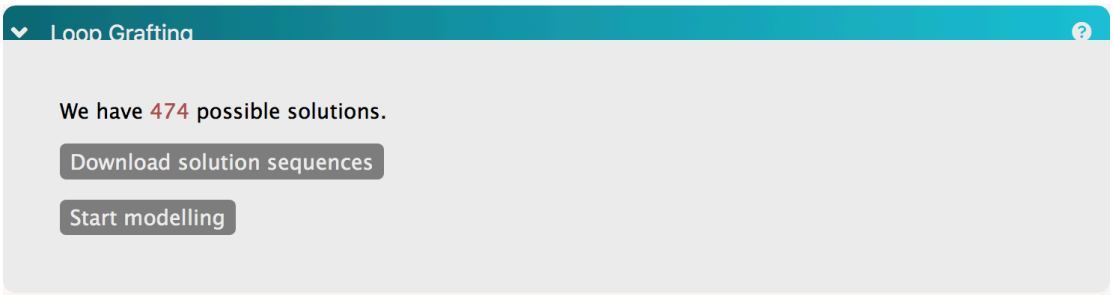
The "Attempt automatic assignment" button (red box) tries to automatically perform the loop paring, Failing this or finding the user a better option, a manual pairing can be achieved by selecting the relevant loop (red arrow) or clicking on the corresponding "Pick Replacement Loop Button". The insert loops become available for selection (blue box).



Once the appropriate loop is selected (blue arrow) , the geometrical details of the loop are displayed (blue box) and a geometrical comparison with the scaffold loop to be replace can be done in the bottom view (red box)

It seems I have almost done it! This step reads "Loop Grafting" and the only thing available is a button named "Calculate grafting sequences". I guess I should click on it…

Before generating the three-dimensional models of the different grafted designs, the server provides the users with the sequences that is going to evaluate. It is worth noting that this is the end of the pipeline for users that do not own a MODELLER license. After calculating all the possible insertion points for the different loops to be grafted (see *Where does the loop(s) replacement happen in the input proteins?* section), the user is presented with the option to download the sequences of the resulting grafting designs and to start the modelling of such grafting designs.

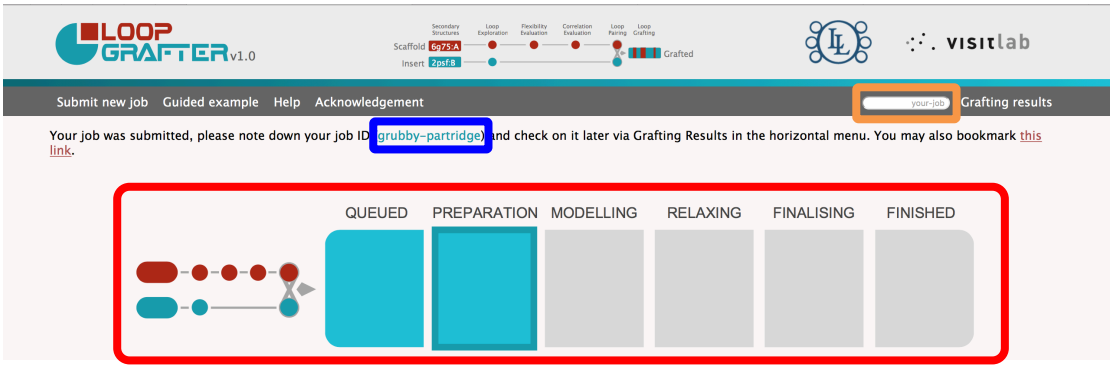


The sequences of the grafting designs are already calculated and ready to be modelled.

How exciting! Let's see those grafted designs 3D models!

By clicking on the "Start modelling" button the pipeline producing such 3D models of the grafted designs is started. The 3D coordinates of each of the fragments from the insert protein to be transplanted to the scaffold is structurally superimposed to it using CE and a sequence-pairing is derived from the same *ad-hoc* designed algorithm used to explore the insertion boundaries. This sequence paring is used as a guide to build a sequence alignment that informs MODELLER how to build the 3D model of each grafted design. After the model is built and evaluated using MODELLER DOPE score (which produces arbitrary units), it is minimised using Rosetta FastRelax. In turn, Rosetta evaluates the minimised model producing a score in Rosetta Energy Units.

This process might take long time and the user is informed about the status of the progress in a visual progress bar. Also, if an e-mail address was provided at the beginning of the process in the input proteins step, the user is noticed when the grafting process is concluded. Each submission has a unique identifier that can be used to retrieve the results at any time from the server using the dedicated search box.



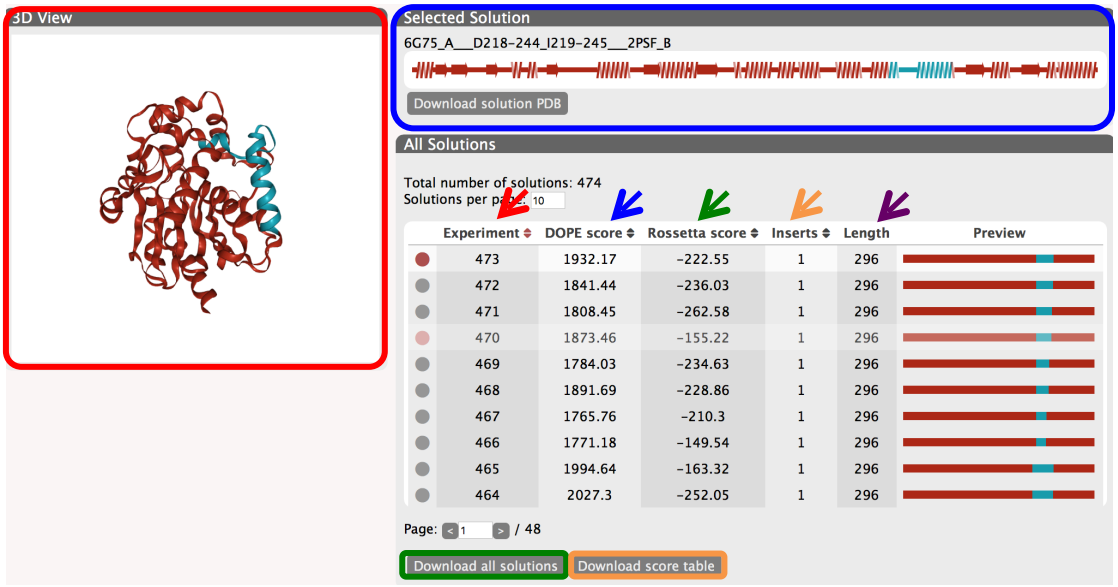
View of the progress-tracking page. The progress status of the submitted job can be assessed by the completion of individual tasks (red box). The name of the job-id is given in a prominent font (blue box, it is worth to note that the job-id is always visible at the bottom left of the web server) and can be entered in the dedicated search box to retrieve the results (orange box).

So… Let's see those results!

When the grafting process is finished, a message appears below the progress bar informing that "Grafting is finished. Click here to retrieve the results.", and by following the link the user is directed to the results page.

All the grafted designs are shown in a paginated table, each grafted design is assigned a unique ID (Experiment column), and is annotated with MODELLER (DOPE, arbitrary units) and Rosetta (FastRelax, Rosetta Energy Units) scores. It is important to note that lower values are better for both scores. Rosetta scores are only produced if the ownership of a Rosetta license was stated in the initial step of the process. Also for each grafted design the number of inserted loops and the total length of the grafted construct is indicated. Additionally a graphical overview of the grafted region in the context of the entire scaffold protein is provided for each solution.

One grafted design can be selected at a time by clicking on the appropriate row on the table, and then it is shown in more detail in the "Selected Solution" 2D view box at the top of the page and in the 3D view at the left. The grafted region is highlighted in a shade of blue and by clicking on the "Download solution PDB" button the 3D structure for that particular grafted design can be downloaded. The complete set of results can be downloaded as a compressed file clicking on the "Download all solutions" button at the bottom of the page. Next to it, the "Download score table" button provides a table containing the calculated scores and annotations for each grafting variant as well as their sequences.



View of the results page, with one grafting design selected. The 2D (red box) and 3D (blue box) representations of the resulting chimeric protein can be seen in the top and the left of the page with the insert region highlighted in a blue hue. ID (red arrow), MODELLER (DOPE, blue arrow) and Rosetta (green arrow) scores, the number of inserts (orange arrow) and the length (purple arrow) of all the produced grafted designs are shown in the paginated table in the main body of the results page. A compressed file with the complete set of results can be downloaded by clicking on the "Download all solutions" button (green box) and a summary table containing all relevant information can be obtained as well by clicking on the "Download score table" button (orange box)

The guided example function from the main page of the server is really cool, but I would like to follow the example from a printable source. Do you have the example written anywhere?

Yes, of course. The remaining of this help section is the written description of the step-by step example provided in the Guided Example functionality. We hope this will be useful to you! You can also download a pdf version here.