In silico characterization of the venom-derived LW-9 peptide using PreADMET-based predictions

Ingrid Mayara Cavalcante Trevisan [1], Clailson da Silva Pinheiro [2], Juliana Mozer Sciani [3] and Catarina Rapôso [1]*

[1] Faculty of Pharmaceutical Sciences, Universidade Estadual de Campinas (UNICAMP), 13083-865, Campinas, Brazil;
[2] Departament of Biophysics and Physiology, Federal University of Piauí (UFPI), 65631-430, Piauí, Brazil; clailson@ufpi.edu.br
[3] Natural Products Laboratory, São Francisco University (USF), 12916900, Bragança Paulista, Brazil; juliana.sciani@usf.edu.br, Bragança Paulista, Brazil.

Supplementary material

Table S1. The raw of the PreADMET predictions

[Home] [Server] LQueue] [About] [Remove] [Statistics]

1-TASSER results for job id S743208

(Click on S743208_results.tar.bz2 to download the tarball file including ali modeling results listed on this page. Click on Annotation ofl-TASSER Output to read the instructions for how to interpret the results on this page. Model results are kept on the server for 60 days, there is no way to retrieve the modeling data older than 2 months)

Submitted Sequence in FASTA format

```
>protein
QKKDRFLGLM
```

Predicted Secondary Structure

Sequence QKKDRFLGLM
Prediction CCCCCHHCCC
Conf. Score 9523011149
H:Helix; S:Strand; C:Coil

Predicted Solvent Accessibility

Sequence QKKDRFLGLM
Prediction 8655434447
Values range from **0** (buried residue) to **9** (highly exposed residue)

Predicted normalized B-factor

(B-factor is a value to indicate the extent ofthe inherent thermal mobility ofresidues/atoms in proteins. **In** 1-TASSER, this value is deduced from threading templale proteins from the PDB in combination with the sequence profiles derived from sequence databases. The reported B-factor profile in the figure below corresponds to the normalized B-factor ofthe target protein, defined by B□(B'-u)/s, where B' is the raw B-factor value, u and sare respectively the mean and standard deviation ofthe raw B-factors along the sequence. Click here to read more about predicted normalized B-factor)

Top 10 threading templates used by I-TASSER

(1-TASSER modeling starts from the structnre templales identified by LOMETS from lhe PDB library. LOMETS is a meta-server threading approach containing multiple threading programs, where each threading program can generate tens ofthousands oftemplate aligmnents. 1-TASSER only uses the templates ofthe highest significance in the threading aligmnents, the significance ofwhich are measured by the Z-score, i.e. the difference between the raw and average scores in the unit of standard deviation. Toe templates in this section are the 1O best templates selected from the LOMETS threading programs. Usually, one template of the highest Z-score is selected from each threading progrann, where the threading programs are sorted by lhe average performance in the large-scale benchmark test experiments.)

Rank	PDB Hit	Iden1	Iden2	Cov	Norm. Z-score	Download Align.	
							Sec.Str CCCCCHHCCC **Seq** QKKDRFLGLM
	2bl9A	0.60	0.60	1.00	1.82	Download	HKTDSFVGLM
2	2bl9A	0.67	0.60	0.90	0.07	Download	-KTDSFVGLM
3	7mnkA	0.10	0.10	1.00	1.17	Download	EDQNSLLKMI
4	3k9hA	0.60	0.60	1.00	0.68	Download	KTKDRFLGTI
5	2bl9A	0.60	0.60	1.00	1.80	Download	HKTDSFVGL□

6	2h19	0.57	0.60	0.70	1.15	Download	--TDSFVGL-
7	6bltW	0.38	0.30	O.BD	1.41	Download	--GLRFPSK'I
8	7zbnE	0.20	0.30	1.00	0.37	Download	SSEERYMGAL
9	§ qxfK	0.30	0.30	1.00	3.10	Download	EERLVFLGDN
10	2bl9A	0.60	0.60	1.00	0.14	Download	HKTDSFVGLM

(a) AII lhe residues are colored in black; however, those residues in template which are identical to lhe residue in lhe query sequence are highlighted in color. Coloring scheme is based on lhe properiy oi amino acids, where polar are brighlly coloured while non-polar residues are colored in dark shade. (more about lhe colora used)

(b) Rank of templates represents lhe top ten threading templates used by 1-TASSER.

(e) Ident1 is lhe percentage sequence idenlity of lhe templates in the threading aligned region with lhe query sequence.

(d) Ident2 is lhe percentage sequence identity of lhe whole template chains with query sequence.

(e) Cov represents the coverage of lhe threading alignment and is equal to lhe number oi aligned residues divided by lhe length of query protein.

(f) Norm. Z-score is lhe normalized Z-score of lhe threading alignments. Alignment with a Normalized Z-score >1 mean a good alignment and vice versa.

(g) Download Align. providas lhe 3D structure of the aligned regions of lhe threading templates.

(h) The top 1O alignments reporied above (in order of their ranking) are from lhe following threading programs:

1: MUSTER 2: FFAS-3D 3: SPARKS-X 4: HHSEARCH2 5: Neff-PPAS 6: HHSEARCH 7: wdPPAS 8: PROSPECT2 9: SP3 10: FFAS03

Top 5 final models predicted by I-TASSER

(For each target, 1-TASSER simulations generate a large ensemble of structnral conformations, called decoys. To select the final models, 1-TASSER uses the SPICKER prograrn to cluster ali the decoys based on the pair-wise structnre similarity, and reports up to five models which corresponds to the five largest structnre clusters. The confidence of each model is quantitatively measured by C-score that is calculated based on the significance ofthreading template aligmnents and the convergence pararneters ofthe structnre assembly simulations. C-score is typically in the range of [-5, 2], where a C-score **of a higher value signifies a model with a higher confidence and vice-versa. TM-score and RMSD are estimated based on C-score and protein length following the correlation observed between these** qualities. Since the top *5* models are ranked by the cluster size, it is possible that the lower-rank models have a higher C-score in rare cases. Although the first model has a better quality in most cases, it is also possible that the lower-rank models have a better quality than the higher-rank models as seen in our benchmark tests. Ifthe 1-TASSER simulations converge, it is possible to have less than *5*

clusters generated; this is usually an indication that the models have a good quality because of the converged simulations.)

- More about C-score
- Local structure accuracy_Jlrofile of the top five models

(By right-click on the images, you can expor! image file or change the configurations, e.g. modifying the background color or stopping the spin ofyour models)

- Download Model 1
- C-score=-0.47 (Read more about C-score)
- Estimated TM-score = 0.65±0.13
- Estimated RMSD = 0.7±0.7Á

- Download Model 2
- C-score = -0.78

Proteins structurally close to the target in the PDB (as identified by IM.:ª1.ign)

(After lhe structure assembly simulation, I-TASSER uses the TM-align structural alignment program to match the first I-TASSER model to ali structures in the PDB library. This section reports the top 10 proteins from the PDB lha!have the closest structural similarity, i.e. the highest TM-score, to the predicted 1-TASSER model. Dueto the structural similarity, these proteins often have similar function to the target. However, users are encouraged to use the data in the next section 'Predicted function using COACH' to infer the function ofthe target protein, since COACH has been extensively trained to derive biological functions from multi-source of sequence and structure features which has on average a higher accuracy than lhe function annotations derived only from the global structure comparison.)

Top 10 Identified stuctural analogs in PDB

t;□□e□ Rank	PDB Hit	TM-score	RMSDª	IDENª	Cov	Alignment
	5lzbz	0.722	0.27	0.200	0.900	Download
2	fui!Y.8	0.706	0.49	0.200	1.000	Download
3	6ut5G2	0.700	0.38	0.100	1.000	Download
4	6emkA	0.700	0.82	0.100	1.000	Download
5	;)id5o	0.698	0.39	0.000	1.000	Download
6	7o41A2	0.696	1.05	0.000	1.000	Download
7	2lfwB	0.696	0.99	0.100	1.000	Download
8	3s2cA	0.693	0.51	0.100	1.000	Download
9	To□	0.693	0.73	O.DOO	1.000	Download
10	2kOgA	0.692	0.20	0.000	0.800	Download

(a) Query structure is shown in cartoon, while lhe structural analog is displayed using backbone trace.

(b) Ranking of proteins is based on TM-score of the structural ali9nment between the query structure and known structures in the PDB library.

(e) RMSDª is the RMSD between residues that are structurally aligned by TM-align.

(d) IDENª is the percentage sequence identity in the structurally aligned region.

(e) Cov represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by length of the query protein.

Predicted function using COFACTOR and COACH

(This section reports biological annotations ofthe target protein by COFACTOR and COACH based on the 1-TASSER structure prediction. While COFACTOR deduces protein functions (ligand-binding sites, EC and GO) using structure comparison and protein-protein networks, COACH is a meta-server approach that combines multiple function annotation results (on ligand-binding sites) from the COFACTOR, TM-SITE and S-SITE progrars.)

Ligand binding sites

t;□□e□ Rank	C-score	C□□=□er	p□r	N□□e	□ºoV:□□:,:t	Ligand Binding Site Residues
1	0.32	208	1lghA	LYC	ReP-. Mui!	2,3,5,6
2	0.22	142	2ol1C	UMP	Bfill, Mui!	1,2,3
3	0.11	66	□l!!!'.l!::I	Nuc.Acid	ReP-. M.Y11	3,5,6,8,9

Download the residue-specific ligand binding probability, which is estimated by SVM.

Download the all possible binding ligands and detailed prediction summary.

Download the templates clustering results.

(a) **C-score** is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a more reliable prediction.

(b) **Cluster size** is the total number of templates in a cluster.

(e) **Lig Name** is name of possible binding ligand. Click lhe name to view its information in the BioliP database.

(d) **Repisa** single complex structure with lhe mos! representativa ligand in the cluster, i.e., lhe one listed in the **Lig Name** column.
Mult is lhe complex structures wilh all polenlial binding ligands in lhe cluster.

Enzyme Commission (EC) numbers and active sites

t;□□□ Rank	CscoreEC	p□r	TM-score	RMSDª	IDENª	Cov	EC Number	Aclive Site Residues
1	0.317	3caiA	0.658	0.95	0.000	0.900	2.6.1.-	**NA**
2	0.287	11Qgt,	0.651	0.90	0.100	1.000	*l/1 1,1*	NA
3	0.270	3f9zD	0.587	0.35	0.000	0.800	□	**NA**
4	0.268	2ztgA	0.661	0.80	0.100	1.000	*lUJ..,I*	NA
5	0.266	1zkkB	0.573	0.37	0.000	0.800	2.1.1.43	NA

Click on lhe radio buttons to visualize predicted aclive site residues.

(a) CscoreEC is lhe confidence score for the EC number prediction. CscoreEC values range in between [0-1]; where a higher score indicates a more reliable EC number prediction.

(b) TM-score is a measure of global structural similarity between query and templale protein.

(c) RMSDª is the RMSD between residues that are structurally aligned by TM-align.

(d) IDEN[a] is lhe percenlage sequence idenlily in lhe slruclurally aligned region.

(e) Cov represents lhe coverage oi global structural alignment and is equal to lhe number oi structurally aligned residues divided by length oi lhe query prolein.

Gene Ontology (GO) terms

Top 1O homologous GO templates in PDB

Rank	CscoreGO	TM-score	RMSD[a]	IDEN[a]	Cov	□□B	Associated GO Terms
	0.27	0.6922	0.20	0.00	0.80	2k0gà	GO:0000166 GO:0016020 GO:0005249 GO:0006811 GO:0030552 GO:0005267 GO:0008076 GO:0006810 GO:0071805 GO:0016021 GO:0006813 GO:0005216 GO:0001932 GO:0005952 GO:0008603
2	0.27	0.6737	0.50	0.10	1.00	J.fmili	GO:0008565 GO:0009306 GO:0016020
3	0.27	0.5732	0.37	0.00	0.80	lzkkB	GO:0005515 GO:0018024
4	0.26	0.6768	0.80	0.10	1.00	2hm2Q	GO:0005634 GO:0006469 GO:0032088 GO:0033209 GO:0045087 GO:0050718 GO:0005515 GO:0005737 GO:0008385 GO:0005829 GO:0004197
5	0.24	0.6739	0.44	0.00	1.00	füdF	GO:0003677 GO:0006351 GO:0006355 GO:0009408 GO:0016987 GO:0005515 GO:0003700 GO:0006352
6	0.23	0.6719	1.43	0.00	0.90	□g2ê,	GO:0051287 GO:0016628 GO:0005737 GO:0055072 GO:0042168 GO:0006635 GO:0055114 GO:0016491 GO:0019290 GO:0003858 GO:0005488 GO:0008152
7	0.22	0.5398	0.59	0.20	1.00	lv33A	GO:0016779 GO:0006269 GO:0006260 GO:0046872 GO:0005658 GO:0006351 GO:0016740 GO:0003899 GO:0003896
8	0.20	0.4871	0.90	0.00	1.00	2avkA	GO:0046872
9	0.20	0.4428	1.71	0.00	0.80	la34A	GO:0019028 GO:0019012 GO:0005198
10	0.20	0.5074	0.67	0.10	0.90	2zt9C	GO:0046872 GO:0015979 GO:0042651 GO:0016021 GO:0016020 GO:0006810 GO:0009579 GO:0022900 GO:0005506 GO:0020037 GO:0031361

Consensus prediction of GO terms

Molecular Function	GO:0005515 GO:0016208 GO:0022843 GO:0030551				GO:0015079 GO:0019887 GO:0042054 GO:0016279 GO:0004175 GO:0000996					
GO-Score	0.59	0.55	0.55	0.55	0.55	0.55	0.53	0.53	0.52	0.48
Biological Process	GO:0071804 GO:0034220 GO:0050706 GO:0050716 GO:0033673 GO:0002376 GO:0045859 GO:0019221								GO:0032731 GO:0071356	
GO-Score	0.55	0.55	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52

Cellular Component	GO:0034705 GO:0044445 GO:0043231			GO:0005887
GO-Score	0.55	0.52	0.52	0.49

(a) Cscore[a0] is a combined measure for evaluating global and local similarity between query and template protein. It's range is [0-1] and higher values indicate more confident predictions.

(b) TM-score is a measure of global structural similarity between query and template protein.

(e) RMSD[a] is the RMSD between residues that are structurally aligned by TM-align.

(d) IDEN[a] is the percentage sequence identity in the structurally aligned region.

(e) Cov represents the coverage of global structural alignment and is equal to the number of structurally aligned residues divided by length of the query protein.

(f) Toe second table shows a consensus GO terms amongst the top scoring templates. Toe GO-Score associated with each prediction is defined as the average weight of the GO tenn, where the weights are as□ template.

[Click on S743208_results.tar.bz2 to download the tarball file including ali modeling results listed on this page]

Please cite the following articles when you use the I-TASSER server:

- Wei Zheng, Chengxin Zhang, Yang Li, Robin Pearce, Eric W. Bell, Yang Zhang. Folding non-homology proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. Cell Reports Methods, 1: 100014 (2021).
- Chengxin Zhang, Peter L. Freddolino, and Yang Zhang. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. Nucleic Acids Research, 45: W291-299 (2017).
- Jianyi Yang, Yang Zhang. I-TASSER server: new development for protein structure and function predictions, Nucleic Acids Research, 43: Wl 74-W181, 2015.

[Back to the Results Pag□]

Top five models with local structure error profiles

The local accuracy is defined as the distance deviation (in Angstrom) between residue positions in the model and the native structure. Since the native structure is unknown, the distance errors in the following plots are estimated by ResQ using support vector regression that makes use of the coverage of threading alignment, divergence ofI-TASSER simulation decoys, and sequence-based secondary structure and solvent accessibility predictions. The numerical data ofthe ResQ prediction are listed in this file: lscore.txt.

More details of ResQ and the local structure error prediction can be found at
J Yang, Y Wang, Y Zhang. ResQ: An approach to unified estimation of B-factor and residue-specific error in protein stmcture prediction, Journal of Molecular Biology, 428: 693-701 (2016).

Generated 3D models

Download the estimated local accuracy of models

- Download Model 1
- C-score=-0.47 (Read more about C-score)
- Estimated TM-score = 0.65±0.13
- EstimatedRMSD=0.7±0.7Á

- Download Model 2
- C-score = -0.78

- Download Model 3
- C-score = -1.79