TopSuite Webserver: A Meta-Suite for Deep Learning-based Protein Structure and Quality Prediction

Daniel Mulnaes[1], Filip Koenig[1], Holger Gohlke[1,2,3]*

[1]Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

[2]John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre (JSC), and Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

Author ORCID
Daniel Mulnaes: 0000-0003-2162-5918
Filip König: 0000-0003-0852-440X
Holger Gohlke: 0000-0001-8613-1447

Running title: TopSuite for protein structure and quality prediction

Key words: structure prediction, comparative modeling, model quality assessment, webserver, machine learning

*Address: Universitätsstr. 1, 40225 Düsseldorf, Germany.
Phone: (+49) 211 81 13662; Fax: (+49) 211 81 13847
Email: gohlke@uni-duesseldorf.de

Abstract

Proteins carry out the most fundamental processes of life such as cellular metabolism, regulation, and communication. Understanding these processes at a molecular level requires knowledge of their 3-D structure. Experimental techniques as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (Cryo-EM) can resolve protein structures but are costly, time-consuming, and do not work for all proteins. Computational protein structure prediction tries to overcome these problems by predicting the structure of a new protein using existing protein structures as a resource. Here we present TopSuite, webservers for protein model quality assessment (TopScore) and template-based protein structure prediction (TopModel). TopScore provides meta-predictions for global and residue-wise model quality estimation using Deep Neural Networks. TopModel predicts protein structures using a top-down consensus approach to aid the template selection and subsequently uses TopScore to refine and assess predicted structures. The TopSuite webserver is freely available at https://cpclab.uni-duesseldorf.de/topsuite/.

Introduction

Proteins carry out their function by virtue of their complex 3D structures. Knowing the 3D structure of a protein is key to predict and understand important properties, particularly its function and potential interactions with other proteins [1] or ligands [2, 3]. Experimental methods for structure determination include X-ray crystallography [4], nuclear magnetic resonance (NMR) spectroscopy [5] and cryo-electron microscopy [6], but these are both costly and time-consuming. Consequently, computational protein structure prediction is an essential part of knowledge-based protein engineering [7], drug-design and -discovery [8], and function assignment [9, 10]. In the last decades, many approaches have been developed for computational protein structure prediction, and the results of CASP competitions [11, 12] have shown the benefit of using consensus methods, which employ complementary algorithms to improve prediction quality. Consensus methods generally employ majority voting for template selection and model averaging during refinement, which can improve quality if primary methods converge on the native fold, but can drive the model away from the native fold if they diverge or converge on the wrong fold.

Recently, we presented TopScore [13] and TopModel [14]. TopScore is a DNN meta-predictor of whole-protein and residue-wise model quality, which has been trained on a massive model ensemble and produces more accurate and consistent superposition-independent quality predictions than any of its constituent primary predictors. Building on the development of TopScore, TopModel is a fully automated meta-method for protein structure prediction. TopModel first submits the target sequence to 12 different primary threading methods. TopModel then uses deep neural networks (DNNs) trained on sequence features, threading scores, and single-template model quality predicted by TopScore to predict the template modeling score (TM-Score, bounded between 0 and 1 with smaller values indicating lower structural similarity between a template and the native structure) [15] of each template. Based on the predicted TM-Score, TopModel then uses top-down consensus to remove false-positive templates. Top-down consensus is advantageous over traditional majority-voting consensus used in other meta-servers like the I-TASSER server [16], particularly for difficult systems where the majority of templates are false positives caused by primary predictors finding similar wrong templates and/or alignments. In such cases, top-down consensus can discard all false-positive templates provided the identified reference-template has the correct fold. After template identification, TopModel uses 18 different primary alignment programs to generate an ensemble of multiple sequence alignments (MSAs) between the templates and the target sequence. These alignments are then modeled, and the models are scored with TopScore. Finally, the highest-ranked single-template and multi-template models are selected for refinement. During refinement, model regions predicted by TopScore to contain errors are iteratively removed, and the remaining model pieces are put together into a final refined model. Unlike traditional model averaging and/or energy minimization refinement protocols used in other meta-servers such as I-TASSER [16] or MULTICOM [17], which also uses majority-voting consensus in the form of clustering, this strategy optimizes the TopScore (A DNN-based score rather than an energy function) to overcome the problems of majority voting mentioned earlier. TopScore and TopModel have been used prospectively and successfully in several applications [2, 18-27].

Until now, applying TopModel and TopScore has required the user to download and install the software suite, which requires ~3 TB of hard drive space due to the many primary predictors and required databases for their use. To avoid this hurdle and improve user-friendliness, we present the TopSuite webserver, which enables users to easily access TopScore and TopModel and inspect the results of a model quality prediction or structure prediction task.

Methods

TopScore. The TopScore server submission page is shown in Figure 1A. The server accepts either a single model or multiple models of the same protein in the form of different PDB files (Figure 1A.1). The files must be submitted as an archive in the .zip, .tgz or tar.gz formats. The maximum sequence length of a single model is 500 residues, and all models must contain the same residues. The maximum number of models submitted in a single evaluation run is 50, and the archive must not be larger than 10 MB.

To be notified when calculations have finished, the user has the option to provide an email address (Figure 1A.2). The email will not be used for any other purpose than notification. Jobs will be kept on the server for up to 7 days before being removed. To guide the user, a documentation page is available, and an example run is provided with different models of the human hemoglobin β-subunit evaluated with TopScore.

TopModel. The TopModel server submission page is shown in Figure 1B. The user must enter a protein sequence (Figure 1B.1), a fasta file content, or upload a fasta file. The server will not accept multiple sequences or sequences longer than 1000 residues. If the inputted sequence contains non-standard amino acids, these will be mutated to alanine in the final model.

To help identify the job and to be notified when calculations have finished, the user has the option to provide a job name and an email address (Figure 1B.2). The email will not be used for any other purpose than notification. Jobs will be kept on the server for up to 7 days before being removed. Four different job options are available (Figure 1B.3):

1) The "Normal Run" job type. This job type requires only a target sequence and starts the default TopModel workflow consisting of the following steps: I) Template identification and single-template modeling. II) Generation of alignment ensembles for multiple templates and multi-template modeling. III) Model quality assessment with TopScore and TopScoreSingle. IV) Model combination and refinement by removing parts of the models predicted by TopScore or TopScoreSingle to contain errors, and combining the remaining parts into a final model. A detailed description of the TopModel workflow can be found in ref. [14].

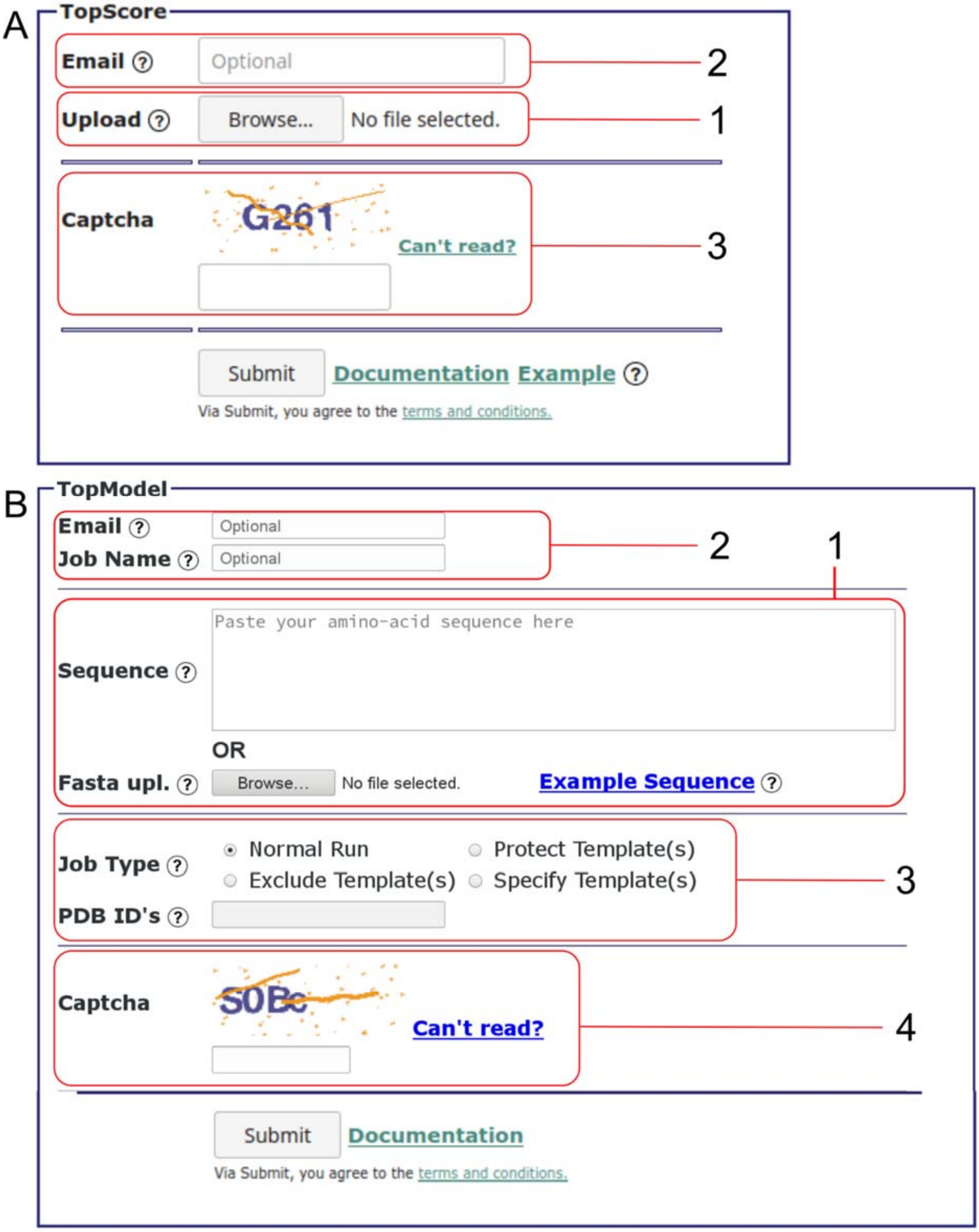
2) The "Protect Templates" job type allows the user to provide PDB IDs that will be protected from being removed by the redundancy and scoring filters in TopThreader. This does not guarantee that these templates are used, if any primary threader does not identify them. However, if the user knows that a specific template has the desired conformation or functional state, this template can be protected.

3) Similar to the "Protect Templates" option, the "Exclude Templates" job type forcibly removes the provided PDB IDs at the threading step. This is useful for benchmarking or for the removal of templates with undesirable conformations or known false positives.

4) The "Specify Template" job type is an additional feature of the TopModel server. This option skips the template identification step and the refinement step (steps I and IV) of a normal TopModel run and produces alignments and models using only the provided PDB IDs as templates. This, therefore, requires the additional information of chains for every PDB ID. The model generation that way is much faster than a regular TopModel job because template identification and refinement are the most time-consuming steps of the workflow. However, this job type should be used only when it is known that the templates provided are strongly homologous to the target sequence because alignments are made without using primary threaders to guide the match between the target sequence and the template.

A documentation page is accessible to explain the necessary inputs and the output results to guide the user. Additionally, a pre-calculated example run with the sequence of human hemoglobin beta subunit (HBB) is available. The runtime depends on the given sequence length and the chosen job type. A webpage reports on the progress of a given job and provides a runtime estimate after the job has started. Since TopModel does not include domain-parsing or *ab-initio* folding tools, it may be less suitable for modeling large multi-domain proteins or for modeling proteins without any available template with the correct fold. Currently, we are in the process of extending TopSuite with these functionalities.

Implementation. The TopSuite webserver has been implemented in PHP, HTML, and Python, as have the underlying routines of TopScore and TopModel. All molecular structures are visualized with NGL Viewer. Given the computational demand of our approaches, jobs are queued on the server, and it may take some time before calculations start depending on the server load.



Via Submit, you agree to the terms and conditions.

Paste your amino-acid sequence here

● Normal Run ○ Protect Template(s)
○ Exclude Template(s) ○ Specify Template(s)

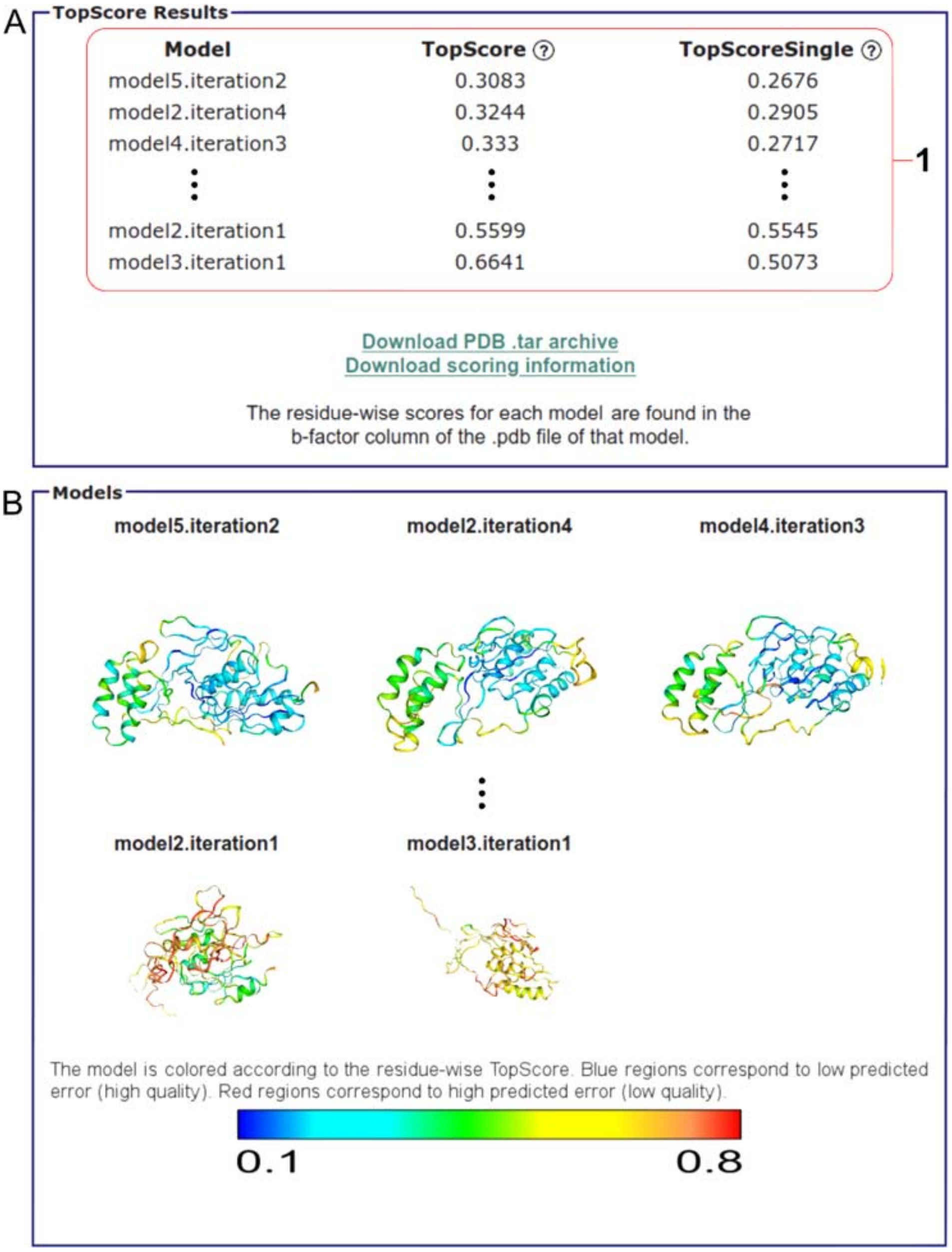
Via Submit, you agree to the terms and conditions.

Figure 1. TopSuite submission webpages. (A) The TopScore submission webpage with an upload function for

the target structures **(1)**, an optional email submission field **(2),** and the required Captcha query **(3)**. **(B)** The TopModel submission webpage with a text-input window and file upload for sequence submission **(1),** an optional email submission and job name field **(2)**, a selection field for the different job types **(3),** and the required Captcha query **(4)**.

Results

TopScore. The TopScore results page is partitioned into two sections, as seen in Figure 2. The first section contains a table of the model names and their corresponding TopScore and TopScoreSingle (Figure 2A.1). TopScore is a measure of the error in the protein that uses clustering information from an ensemble of models; TopScoreSingle is a measure of the error that uses only information from a single model. Therefore, the latter measure is independent of the model ensemble and estimated solely by the model 3D structure itself. Both scores are defined as 1-lDDT [28] (lDDT: local Distance Difference Test), are bounded by [0, 1], and indicate how uncertain the inter-atomic distances in the model are. Lower scores thus indicate models of better quality. In turn, highly flexible structures or mobile regions of a structure can be expected to have a higher uncertainty and a higher score. For an in-depth explanation of TopScore and TopScoreSingle, please see ref. [13]. The models are ranked according to their TopScore. The table and an archive of the model files with the residue-wise TopScore written in the b-factor column are offered for download (Figure 2A). The second section (Figure 2B) shows every model structure, ordered according to their respective TopScore, with the NGL Viewer [29, 30]. The structures are colored according to the residue-wise TopScore, with blue indicating a low error and red a high error.



Model	TopScore ⓘ	TopScoreSingle ⓘ
model5.iteration2	0.3083	0.2676
model2.iteration4	0.3244	0.2905
model4.iteration3	0.333	0.2717
⋮	⋮	⋮
model2.iteration1	0.5599	0.5545
model3.iteration1	0.6641	0.5073

Download PDB .tar archive
Download scoring information

The residue-wise scores for each model are found in the b-factor column of the .pdb file of that model.

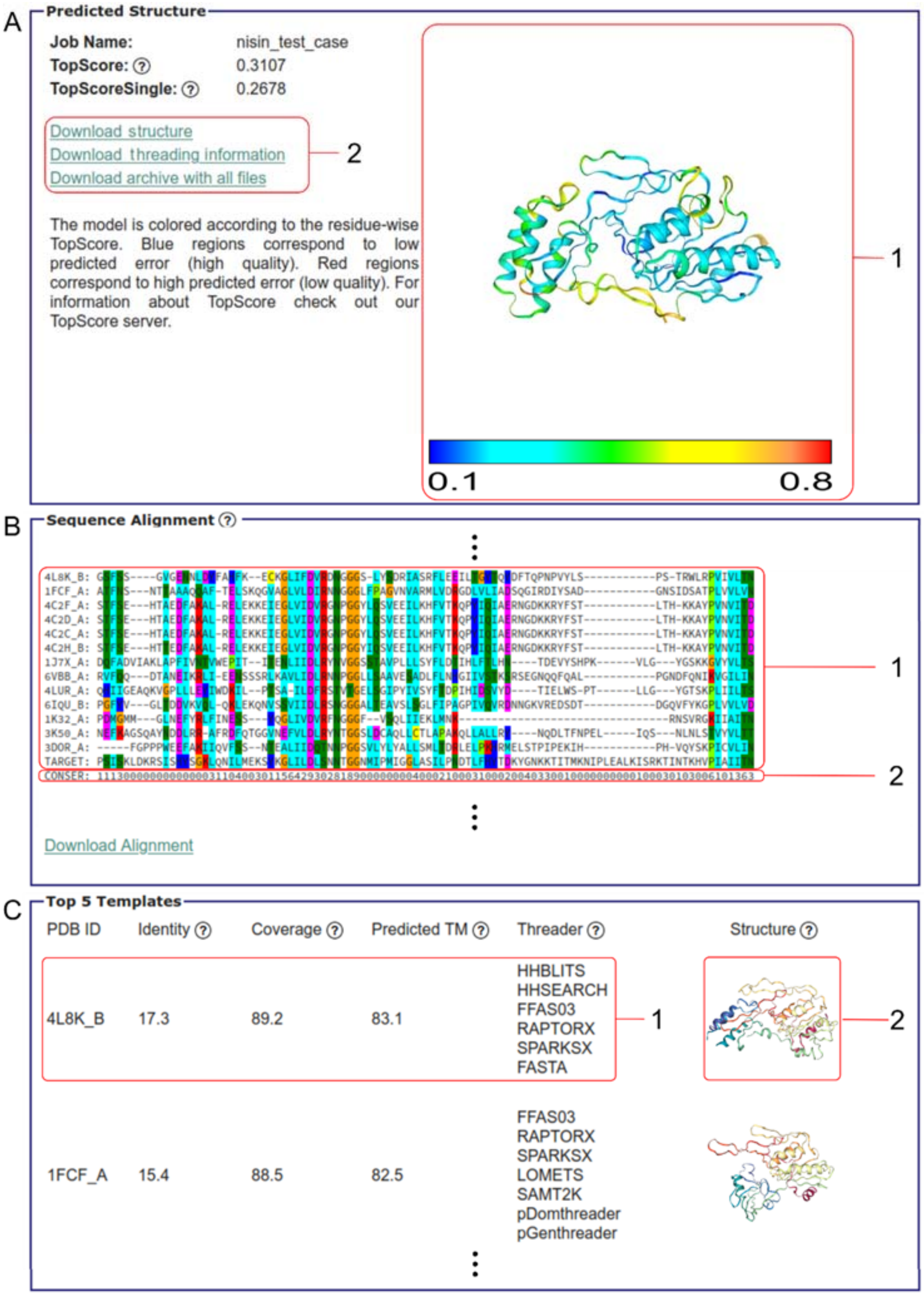
The model is colored according to the residue-wise TopScore. Blue regions correspond to low predicted error (high quality). Red regions correspond to high predicted error (low quality).

Figure 2. TopScore results webpage. (A) The global TopScore and TopScoreSingle for each model are presented **(1)** along with links for downloading the scoring information and PDB files of the models with b-factors colored by the residue-wise TopScore. **(B)** The models are provided as interactive NGL Viewer applets and colored according to residue-wise TopScore (blue: low error, red: high error). Some data is omitted for clarity as indicated by three consecutive dots.

TopModel. The TopModel results page is divided into three sections, as seen in Figure 3. The first section (Fig. 3A) presents modeling accuracy metrics in form of TopScore and TopScoreSingle [13]. Furthermore, it contains an NGL Viewer window with the predicted structure colored according to the residue-wise TopScore (Fig. 3A.1). The predicted structure, the template information, and an archive containing all available information, including the multiple sequence alignment, are also provided for download (Fig. 3A.2).

The second section (Fig. 3B) shows a multiple sequence alignment. This alignment is produced by aligning all identified compatible templates (TM-Score > 0.5) to the final model using TopAligner [14]. The templates are ordered according to the predicted TM-Score to the native structure. Colors are used to group the residues according to their physicochemical properties (Fig. 3B.1). A row with conservation score (Fig. 3B.2) indicates the normalized fraction of amino acids equal to the one in the target sequence at a given position (9 indicating 100% conservation and 0 indicating 0% conservation). A download link for the multiple sequence alignment is provided.

The third section (Fig. 3C) gives an overview of the best five templates used for model generation. These templates are ranked and ordered according to their predicted TM-Score to the native structure. The PDB ID, chain identifier, sequence identity, sequence coverage, predicted TM-Score, and a list of primary threaders that identified the template are provided (Fig. 3C.1). The structure of each template is presented with the NGL Viewer colored according to residue number (Fig. 3C.2).



Predicted Structure

Job Name: nisin_test_case
TopScore: ? 0.3107
TopScoreSingle: ? 0.2678

Download structure
Download threading information
Download archive with all files

The model is colored according to the residue-wise TopScore. Blue regions correspond to low predicted error (high quality). Red regions correspond to high predicted error (low quality). For information about TopScore check out our TopScore server.

Sequence Alignment ?

```
⋮
4L8K_B: GSFSS----GVGENNLDYFAHFK--ECKGLIFDVRDNGGGS-LYSDRIASRFLEEILTGYTQYDFTQPNPVYLS-----------PS-TRWLRPVIVLTN
1FCF_A: ATFNS---NTTAAAQQAF-TELSKQGVAGLVLDIRNNGGGLFPAGVNVARMLVDRGDLVLIADSQGIRDIYSAD-----------GNSIDSATPLVVLVN
4C2F_A: STFSE---HTAEDFAKAL-RELEKKEIEGLVIDVRGNPGGYLQSVEEILKHFVTKQPYIQIAERNGDKKRYFST-----------LTH-KKAYPVNVITD
4C2D_A: STFSE---HTAEDFAKAL-RELEKKEIEGLVIDVRGNPGGYLQSVEEILKHFVTKQPYIQIAERNGDKKRYFST-----------LTH-KKAYPVNVITD
4C2C_A: STFSE---HTAEDFAKAL-RELEKKEIEGLVIDVRGNPGGYLQSVEEILKHFVTKQPYIQIAERNGDKKRYFST-----------LTH-KKAYPVNVITD
4C2H_B: STFSE---HTTEDFAKAL-RELEKKEIEGLVIDVRGNPGGYIQSVEEILKHFVTKQPYIQIAERNGDKKRYFST-----------LTH-KKAYPVNVITD
1J7X_A: DQFADVIAKLAPFIVNTVWEPIT--ITENLIIDLRYNVGGSSTAVPLLLSYFLDTIHLFTLHN----TDEVYSHPK------VLG---YGSKKGVYVLTS
6VBB_A: RVFQQ---DTANEIKRLI-EENSSSRLKAVLIDLRNNPGGLLSAAVESADLFLNNGIIVSTKSRSEGNQQFQAL-----------PGNDFQNIKVGILIN
4LUR_A: QHIIGEAQKVGPLLLEYIWDKIL--PTSA-ILDFRSTVTGELSGIPYIVSYFTDPIHIDSVYD----TIELWS-PT------LLG---YGTSKPLIILTS
6IQU_B: PGFYV---GLTDDVKVQL-QKLEKQNVSSVIIDLRSNGGGALTEAVSLSGLFIPAGPIVQVRDNNGKVREDSDT-----------DGQVFYKGPLVVLVD
1K32_A: PDMGMM---GLNEFYRLFINESS---YQGLIVDVRFNGGGF--VSQLIIEKLMNK------------------------------RNSVRGKIIAITN
3K50_A: NEFKAGSQAYNDDLRR-AFRDFQTGGVNEFVLDLRYNTGGSLDCAQLLCTLAPAKQLLALLRY----NQDLTFNPEL-----IQS---NLNLSTVYVLTT
3DOR_A: -----FGPPPWEEFAKIIQVFSS--NTEALIIDQTNNPGGSVLYLYALLSMLTDRLELPKHRMELSTPIPEKIH-----------PH-VQYSKPICVLIN
TARGET: PSISKLDKRSISHVSGKLQNILMEKSYKGLILDLSNNTGGNMIPMIGGLASILPNDTLFHYTDKYGNKKTITMKNIPLEALKISRKTINTKHVPIAIITN
CONSER: 111300000000000000031104003011564293028189000000000040002100031000200403300100000000000100030103006101363
⋮
```

Download Alignment

Top 5 Templates

PDB ID	Identity ?	Coverage ?	Predicted TM ?	Threader ?	Structure ?
4L8K_B	17.3	89.2	83.1	HHBLITS HHSEARCH FFAS03 RAPTORX SPARKSX FASTA	
1FCF_A	15.4	88.5	82.5	FFAS03 RAPTORX SPARKSX LOMETS SAMT2K pDomthreader pGenthreader	

⋮

Figure 3. TopModel results webpage. (A1) The predicted structure is presented as an interactive NGL Viewer applet. The residues are colored according to the residue-wise TopScore. **(A2)** A PDB file of the predicted structure is available for download, along with a file containing information about the templates identified by meta-threading. **(B1)** All template sequences are presented as a multiple-sequence alignment to the target model. Conserved residues are colored according to their physio-chemical properties by the ClustalX coloring scheme. **(B2)** A conservation score indicates the normalized fraction of conserved amino (9 indicating 100% conservation, 0 indicating 0% conservation). **(C1)** An overview of the best templates is provided along with sequence identities, target coverages, predicted TM scores, and a list of primary threaders that identified the template. **(C2)** The templates are provided as interactive NGL Viewer applets, colored according to residue index. Some data is omitted for clarity as indicated by three consecutive dots.

Example Case. To demonstrate the utility of the webserver, we used the Nisin Resistance Protein (NSR) from *S. agalactiae* as a test case. The Nisin Resistance Protein belongs to the S41 Protease family and hydrolyzes the antimicrobial peptide Nisin [20]. As one crystal structure of NSR is available (PDB ID 4Y68), we used the "Exclude Template" job type and excluded this PDB ID to avoid solving a trivial task with a perfect template. The resulting model showed an RMSD of 3.5 Å compared to the crystal structure (Figure 4), demonstrating the strength of TopModel in predicting good models despite having no close homologs (the highest sequence identity is 17%). The region with the highest predicted error is the linker between the two protein domains (Fig. 4 red box), which is not found to be helical in any of the primary threading models and not predicted as helical by secondary structure prediction with PSIPRED4 [14]. As a result, TopModel also fails to predict these residues as helical.

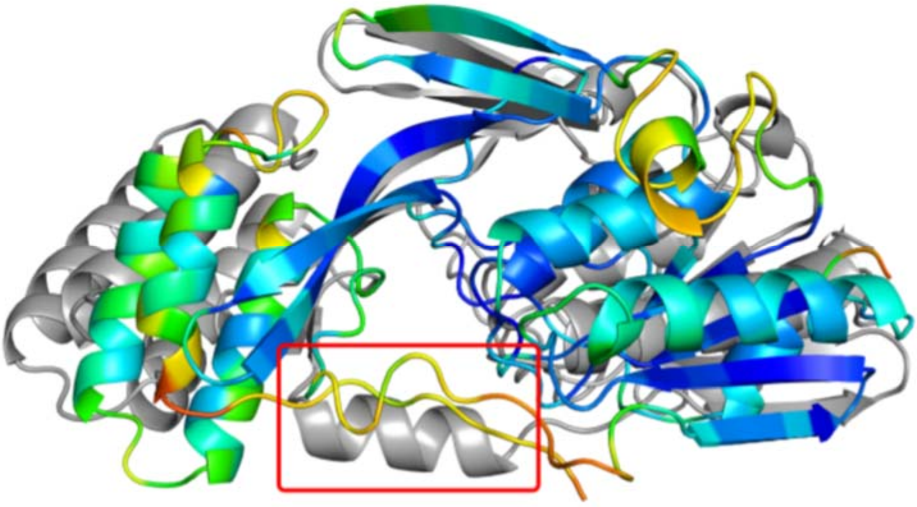


Figure 4. Comparison of the TopModel server prediction of NSR to the crystal structure (PDB ID 4Y68, gray). The model is colored according to TopScore, with blue regions indicating the highest confidence. A red box indicates the misfolded helix linker.

Concluding remarks

In this work, we presented TopSuite, a webserver for convenient access to TopModel and TopScore. TopScore is a deep learning-based meta-method for protein model quality assessment, which has higher and more consistent performance than any of its constituent primary predictors on different datasets and quality criteria. TopModel is a template-based meta-method for protein structure prediction that uses top-down consensus for template selection and optimizes TopScore during model refinement, mitigating the drawbacks of traditional majority-voting consensus, model clustering, and model averaging during template selection and refinement. The TopScore server accepts up to 50 models for a single run. The

TopModel server allows sequence inputs of up to 1000 residues and provides various options for job types to meet the user needs. The servers present the results in an easy-to-overview style, enables visual inspection of structures using the NGL Viewer, and provides scoring information, multiple sequence alignments, and download links. We provided an example case by applying TopModel on the Nisin Resistance Protein from *S. agalactiae* and presented the structural results compared to the crystal structure to illustrate the method's predictive power. We expect that the easy access to the TopScore and TopModel servers will aid researchers from the life sciences who are not bioinformatics experts in predicting protein structures and evaluating the quality of protein structure models.

Author Contributions

DM and HG jointly conceived the study. DM developed and implemented the methods and wrote the manuscript. FK implemented the webservers and wrote the manuscript. HG supervised and managed the project, secured funding and resources for the project, and revised the manuscript. All authors reviewed and approved the manuscript.

Notes

The authors declare no competing interests.

Acknowledgments

This work was supported by the Bundesministerium für Bildung und Forschung (BMBF) − Förderkennzeichen 031L0182 / InCelluloProtStruct and in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) − Projektnummer 267205415 / SFB 1208 (project A03). We are grateful for the computational support and infrastructure provided by the "Zentrum für Informations- und Medientechnologie" (ZIM) at the Heinrich Heine University Düsseldorf.

References

1. Janin, J., Assessing predictions of protein-protein interaction: the CAPRI experiment. *Protein Sci.*, **2005**. 14(2), 278-283.
2. Gohlke, H., Hergert, U., Meyer, T., Mulnaes, D., Grieshaber, M. K., Smits, S. H. J., and Schmitt L., Binding region of alanopine dehydrogenase predicted by unbiased molecular dynamics simulations of ligand diffusion. *J. Chem. Inf. Mod.*, **2013**. 53(10),

2493-2498.

3. Yang, J., Roy, A., and Zhang, Z., Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, **2013**, btt447.

4. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., & Phillips, D. C., A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, **1958**, 181(4610), 662-666.

5. Huang, C., and Kalodimos, C. G., Structures of large protein complexes determined by nuclear magnetic resonance spectroscopy. *Annu. Rev. Biophys.*, **2017**, 46, 317-336.

6. Egelman, E.H., The current revolution in cryo-EM. Biophys.l J., **2016**. 110(5), 1008-1012.

7. Aehle, W., Sobek, H., Amory, A., Vetter, R., Wilke, D., and Schomburg, D., Rational protein engineering and industrial application: Structure prediction by homology and rational design of protein-variants with improved 'washing performance'- the alkaline protease from Bacillus alcalophilus. *J. Biotechnol.*, **1993**, 28(1), 31-40.

8. Cavasotto, C. N., and Phatak, S. S., Homology modeling in drug discovery: current trends and applications. Drug Discov. Today, **2009**. 14(13), 676-683.

9. Roy, A., Yang, J., and Zhang, Y., COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res.*, **2012**, gks372.

10. Roche, D. B., Buenavista, M. T., and McGuffin, L. J., The FunFOLD2 server for the prediction of protein-ligand interactions. *Nucleic Acids Res.*, **2013**. 41(W1), W303-W307.

11. Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., and Tramontano, A., Critical assessment of methods of protein structure prediction (CASP) - Round X. *Proteins: Struct. Funct. Bioinform.*, **2014**, 82(S2), 1-6.

12. Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., and Tramontano, A., Critical assessment of methods of protein structure prediction (CASP) - Round XII. *Proteins: Struct. Funct. Bioinform.*, **2018**. 86(S1), 7-15.

13. Mulnaes, D. and Gohlke H., TopScore: Using Deep Neural Networks and Large Diverse Data Sets for Accurate Protein Model Quality Assessment. *J. Chem. Theory Comput.* **2018**, 14(11), 6117-6126.

14. Mulnaes, D., Porta, N., Clemens, R., Apanasenko, I., Reiners, J., Gremer, L., Neudecker, P., Smits, S.H. and Gohlke, H., TopModel: Template-based protein structure prediction at low sequence identity using top-down consensus and deep neural networks. *J. Chem. Theory Comput.*, **2020**. 16(3), 1953-1967.

15. Zhang, Y. and Skolnick J., TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.*, **2005**. 33(7), 2302-2309.

16. Zhang, Y., I-TASSER server for protein 3D structure prediction. *BMC Bioinform.*, **2008**, 9(1), 40.

17. Wang, Z., Eickholt J., and Cheng J., MULTICOM: a multi-level combination approach to protein structure prediction and its assessments in CASP8. *Bioinformatics*, **2010**. 26(7), 882-888.

18. Widderich, N., Pittelkow, M., Höppner, A., Mulnaes, D., Buckel, W., Gohlke, H., Smits, S.H. and Bremer, E., Molecular dynamics simulations and structure-guided mutagenesis provide insight into the architecture of the catalytic core of the ectoine hydroxylase. *J. Mol. Biol.*, **2014**. 426(3), 586-600.

19. Zhang, Z., Gu, Q., Vasudevan, A.A.J., Hain, A., Kloke, B.P., Hasheminasab, S., Mulnaes, D., Sato, K., Cichutek, K., Häussinger, D. and Bravo, I.G., Determinants of FIV and HIV Vif sensitivity of feline APOBEC3 restriction factors. *Retrovirology*, **2016**, 13(1), 46.

20. Khosa, S., Frieg, B., Mulnaes, D., Kleinschrodt, D., Hoeppner, A., Gohlke, H. and

Smits, S.H., Structural basis of lantibiotic recognition by the nisin resistance protein from Streptococcus agalactiae. *Sci. Rep.*, **2016**. 6, 18679.

21. Milić, D., Dick, M., Mulnaes, D., Pfleger, C., Kinnen, A., Gohlke, H. and Groth, G., Recognition motif and mechanism of ripening inhibitory peptides in plant hormone receptor ETR1. *Sci. Rep.*, **2018**. 8(1), 3890.
22. Furtmann, F., Porta, N., Hoang, D.T., Reiners, J., Schumacher, J., Gottstein, J., Gohlke, H. and Smits, S.H., Characterization of the nucleotide-binding domain NsrF from the BceAB-type ABC-transporter NsrFP from the human pathogen Streptococcus agalactiae. *Sci. Rep.*, **2020**. 10(1), 1-16.
23. Khan, I., Gratz, R., Denezhkin, P., Schott-Verdugo, S.N., Angrand, K., Genders, L., Basgaran, R.M., Fink-Straube, C., Brumbarova, T., Gohlke, H. and Bauer, P., Calcium-promoted interaction between the C2-domain protein EHB1 and metal transporter IRT1 inhibits Arabidopsis iron acquisition. *Plant Physiol.*, **2019**, 180(3), 1564-1581.
24. Verma, N., Dollinger, P., Kovacic, F., Jaeger, K.E. and Gohlke, H., The Membrane-Integrated Steric Chaperone Lif Facilitates Active Site Opening of Pseudomonas aeruginosa Lipase A. *J. Comput. Chem.* **2020**. 41(6), 500-512.
25. Liedgens, L., Zimmermann, J., Wäschenbach, L., Geissel, F., Laporte, H., Gohlke, H., Morgan, B. and Deponte, M., Quantitative assessment of the determinant structural differences between redox-active and inactive glutaredoxins. *Nat. Commun.*, **2020**,. 11(1), 1-18.
26. Twizerimana, A.P., Scheck, R., Becker, D., Zhang, Z., Wammers, M., Avelar, L., Pflieger, M., Häussinger, D., Kurz, T., Gohlke, H. and Münk, C., Cell type-dependent escape of capsid inhibitors by simian immunodeficiency virus SIVcpz. *J. Virol.*, **2020** 94(23).
27. Viegas, A., Dollinger, P., Verma, N., Kubiak, J., Viennet, T., Seidel, C.A., Gohlke, H., Etzkorn, M., Kovacic, F. and Jaeger, K.E., Structural and dynamic insights revealing how lipase binding domain MD1 of Pseudomonas aeruginosa foldase affects lipase activation. *Sci. Rep.*, **2020**. 10(1), 1-15.
28. Mariani, V., Biasini, M., Barbato, A. and Schwede, T., lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*, **2013**. 29(21), 2722-2728.
29. Rose, A.S., Bradley, A.R., Valasatava, Y., Duarte, J.M., Prlić, A. and Rose, P.W., NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics*, **2018**. 34(21), 3755-3758.
30. Rose, A.S. and Hildebrand, P.W., NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.*, **2015**. 43(W1), W576-W579.