50TH ANNIVERSARY

OXFORD

Proscan: a structure-based proline design web server

Nathaniel Felbinger[1,2], Helder V. Ribeiro-Filho[1,3] and Brian G. Pierce [1,2,*]

[1]University of Maryland Institute for Bioscience and Biotechnology Research, Rockville, MD 20850, USA
[2]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA
[3]Brazilian Biosciences National Laboratory, Brazilian Center for Research in Energy and Materials, Campinas 13083-100, Brazil

*To whom correspondence should be addressed. Tel: +1 240 314 6271; Email: pierce@umd.edu

Abstract

The ability to control protein conformations and dynamics through structure-based design has been useful in various scenarios, including engineering of viral antigens for vaccines. One effective design strategy is the substitution of residues to proline amino acids, which due to its unique cyclic side chain can favor and rigidify key backbone conformations. To provide the community with a means to readily identify and explore proline designs for target proteins of interest, we developed the Proscan web server. Proscan provides assessment of backbone angles, energetic and deep learning-based favorability scores, and other parameters for proline substitutions at each position of an input structure, along with interactive visualization of backbone angles and candidate substitution sites on structures. It identifies known favorable proline substitutions for viral antigens, and was benchmarked against datasets of proline substitution stability effects from deep mutational scanning and thermodynamic measurements. This tool can enable researchers to identify and prioritize designs for prospective vaccine antigen targets, or other designs to favor stability of key protein conformations. Proscan is available at: https://proscan.ibbr.umd.edu.

Graphical abstract
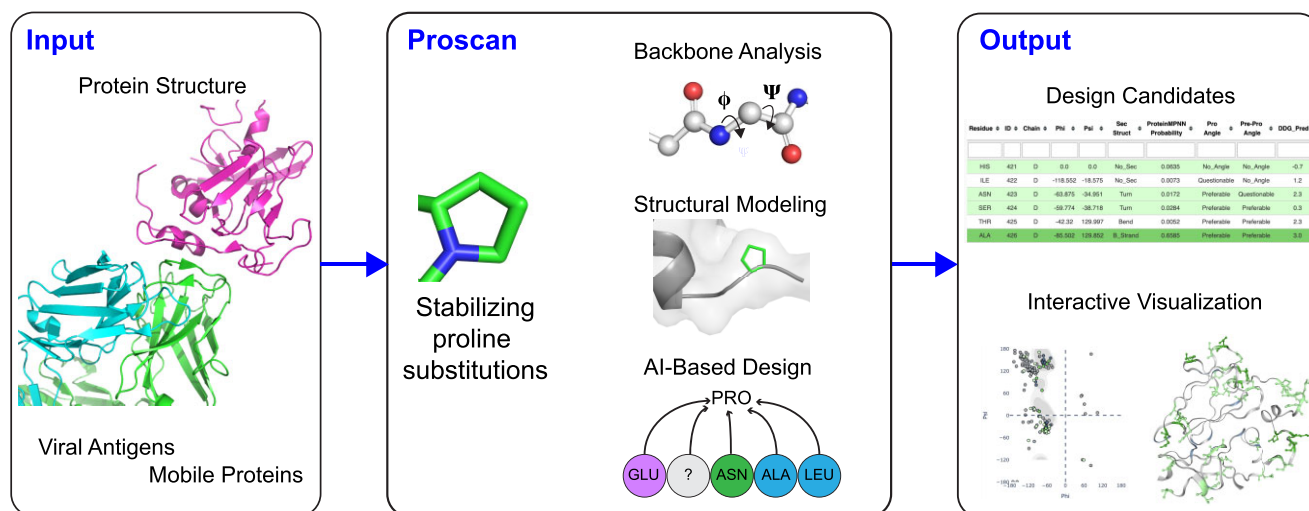
Introduction

Proline substitutions represent a useful and effective protein design strategy, allowing for favoring and rigidification of key preferred local conformations and tertiary structures due to the proline pyrrolidine ring and consequently constrained main chain dihedral angle. This has been successfully utilized for multiple vaccine antigens to help stabilize prefusion conformations of viral proteins, as discussed in a recent review (1). One notable example is the '2P' coronavirus spike design consisting of two consecutive proline substitutions, which was initially found to stabilize the prefusion state of the MERS-CoV spike glycoprotein (as well as spike glycoproteins from other coronaviruses) in 2017 (2). The 2P design was later utilized to stabilize the SARS-CoV-2 prefusion spike, which enabled determination of the first reported SARS-CoV-2 spike structure (3), and it is used in most approved SARS-CoV-2 vaccines (1). Based on the SARS-CoV-2 spike structure, Hsieh et al. identified a set of four additional proline substitutions to further stabilize the prefusion SARS-CoV-2 spike (4), and this 'HexaPro' design (2P plus four more prolines) has been utilized in many subsequent structural and immunogenicity studies (5–8). Examples of proline designs for other viral antigens include the widely used HIV envelope SOSIP design, which contains a key proline residue substitution (I559P)(9), as well as a structure-based proline substitution in the hepatitis C virus (HCV) E2 glycoprotein that led to improved neutralizing antibody

binding and immunogenicity (10). Other studies have explored structure-based proline designs to confer improved stability and rigidification in other systems, for instance to stabilize T cell receptor and antibody loop structures to improve antigen binding (11,12).

The success and utility of proline designs in previous studies, in addition to the need for multiple computational tools (10) or large-scale screening strategies required (13), highlights the need for a tool to systematically analyze protein structures to identify candidate proline substitutions. One computational proline design pipeline was reported over 5 years ago (14), but due to its availability as downloadable code, it may not be accessible to researchers with limited computational expertise, while others have proposed a decision tree proline design pipeline (15). Another approach utilized computational mutagenesis in the program Rosetta (16) in conjunction with Ramachandran plot analysis of backbone structure (10), but due to the lack of current availability of the Ramachandran plot web server used in that study (17), and the need to run multiple separate tools in that algorithm, it is not possible to readily utilize it for prospective design studies.

To provide a proline design resource and tool for the community, we have developed the Proscan web server, which includes the Ramachandran plot analysis and Rosetta mutagenesis capabilities previously utilized by our team for antigen design (10), as well as analysis from a recently developed deep learning structure-based design tool, ProteinMPNN (18), and other annotations of interest, including secondary structure and wild-type residue interactions. This easy-to-use server takes protein structures as input, returning results in seconds to minutes (depending on structure size), and its results page includes interactive visualization of results and a formatted table with scores and annotations. Proscan can enable studies to engineer and optimize antigens for current and future vaccine targets, and can be utilized in other design scenarios to stabilize key conformations and modulate protein dynamics.

Materials and methods

Web server implementation

Interface and visualization

The ProScan web server was developed using the Python3 Flask framework and deployed with Apache. Input Protein Data Bank (PDB)(19) and CIF format files are parsed using a combination of the Biopython PDB package (20) and custom scripts, while PDB file retrieval from PDB code is performed using the Biopython PDB package. Ramachandran plots are rendered using plotly (https://plot.ly), using Ramachandran plot angle distributions from from the MolProbity (21) Top8000 structural database. Protein structures are visualized on the results page with NGL Viewer (22).

Structural analysis

Φ and Ψ backbone dihedral angles are calculated for input structures by the Biopython PDB package, and Φ/Ψ angle probability classifications are based on previously established criteria for structural validation (23), with 'Preferred', 'Acceptable' and 'Questionable' referring to within 98%, within 99.5%, and outside of observed proline or pre-proline angle distributions, respectively. Calculations of energetic stability

effects for proline substitutions are performed using Rosetta v2.3 (rosettacommons.org) and the 'interface' mutagenesis protocol, which structurally models mutant side chains and computes stability changes ($\Delta\Delta$Gs) using an energy-based scoring function (16). Rosetta 2.3 is used in this context due to its speed and its utility in previous design and analysis studies (10,11,24), including strong performance in comparative benchmarking of $\Delta\Delta$G for alanine point substitutions (25). ProteinMPNN (18) was downloaded from its Github repository (https://github.com/dauparas/ProteinMPNN) in May 2023 and is run on input structures with default parameters and model weights (v_48_020), and positional proline favorability values (0–1, as a proportion to other possible amino acids at each position) are parsed from ProteinMPNN output. Secondary structure information is obtained from the DSSP program (26), hydrogen bonds are determined by the hbplus program (27), and N-glycans are detected by parsing input structures with a script that extracts glycan information from the LINK and struct_connection records of the input PDB/mmCIF file.

Benchmarking datasets

MegaScale dataset

The $\Delta\Delta$G dataset from the MegaScale deep mutational scanning study (28) was filtered to identify measured values for substitutions to prolines. Sequences of the wild-type proteins were obtained from corresponding PDB structures and clustered with a 30% identity cutoff using MMseqs2 (29), and a PDB was selected from each cluster, excluding NMR structures from selections. This resulted in 16 structures and a total of 652 proline substitutions with measured stability effects.

Protherm dataset

An additional dataset of experimentally measured proline mutation $\Delta\Delta$G values was downloaded from ProthermDB (30). Measured $\Delta\Delta$G values for proline point substitutions were checked against the corresponding original reference to confirm $\Delta\Delta$G value and measured polarity. The final curated dataset contains 27 unique proline mutants from nine protein structures.

Viral antigens

Viral antigen proline design examples were identified from the literature.

Results

Overview of server

The Proscan web server takes a protein structure as input, which can be provided through the input page (Figure 1A) as a user-input PDB file, or a PDB code to specify structures. Options to control behavior include the capability to only use certain chains, or exclude certain chains, from analysis, and to skip the Rosetta $\Delta\Delta$G calculations, which can speed up running time, particularly for larger structures. The Proscan results page (Figure 1B,C) contains a Ramachandran plot showing the distribution of backbone Φ and Ψ angles, a structural viewer with potential proline substitution sites highlighted, and a table with scores and annotations related to proline substitution analysis. The primary scores and data provided
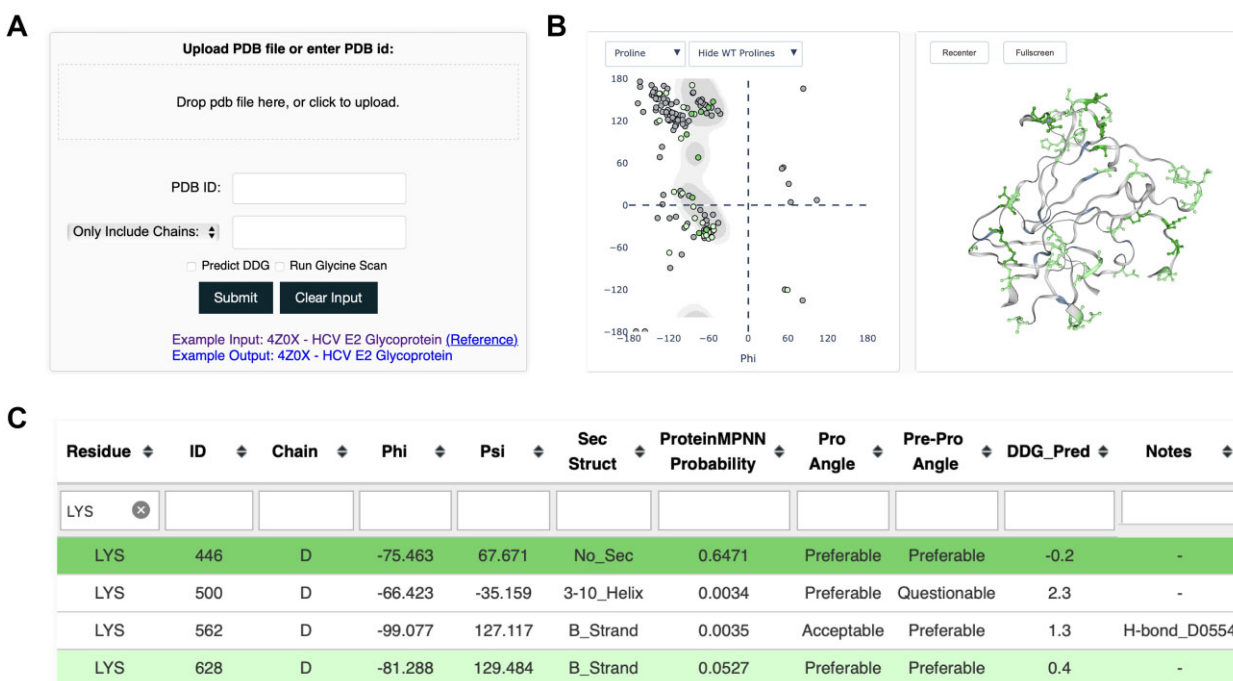
Figure 1. Proscan submission and results pages. (**A**) Proscan submission page, allowing user-input PDB structure or PDB code input, along with user-specified options. (**B**) Results page visualization, with proline Ramachandran plot, showing input structure residue Φ/Ψ backbone angles plotted as points (left), and interactive structure viewer (right). Ramachandran plot points and residues on structure are colored based on predicted proline favorability and score criteria. (**C**) Results page table, showing scores and annotations for each residue and proline substitution favorability. Cells are colored based on score favorability. Example structure shown for (B) and (C) is the HCV E2 glycoprotein core (PDB code 4MWF, chain D) (36). Residues in results table rows and figures are colored by predicted proline favorability (light green = possibly favorable, green = likely favorable).

Table 1. Proscan assessments of previously described viral antigen proline designs

PDB	Amino acid	Residue number	Chain	Phi	Psi	ProteinMPNN	Ramachandran Proline	Rosetta ΔΔG
Hepatitis C virus E2 H445P (10)								
4Z0X	HIS	445	C	− 66.4	147.0	0.712	Preferable	− 0.6
SARS-CoV-2 Spike Hexapro (4)								
6VSB	PHE	817	A	− 54.8	− 50.7	0.316	Preferable	− 0.8
6VSB	ALA	892	A	− 63.5	147.8	0.113	Preferable	− 1.7
6VSB	ALA	899	A	− 58.6	− 36.3	0.056	Preferable	− 1.2
6VSB	ALA	942	A	− 68.3	166.8	0.667	Preferable	− 1.3
RSV F S215P (37)								
7UJA	SER	215	A	− 77.2	− 61.34	0.742	Questionable	− 0.2
Ebola gp T577P (38)								
5JQ3	THR	577	B	− 102.3	− 7.0	0.015	Preferable	0.6
Dengue E T280P (39)								
1TG8	THR	280	A	− 59.45	− 45.7	0.872	Preferable	− 0.9

by Proscan are Ramachandran plot favorability annotations for proline backbone conformation (current residue) and pre-proline backbone conformation (previous residue), Rosetta stability change for the proline substitution (ΔΔG) in approximate units of kcal/mol, and proline positional probability score from the recently developed deep learning protein design tool ProteinMPNN (18). Rosetta ΔΔG and Ramachandran plot analysis, which primarily represent side chain substitution favorability and backbone compatibility, respectively, have previously been used to select favorable proline designs for viral antigen and other design targets (10–12). The deep learning protein design tool ProteinMPNN uses a graph-based neural network to predict amino acid favorability at each position of a structure, and was shown to be effective at design of structures and protein interfaces (18), as well as protein stability improvement (31). The interactive Proscan output page allows users to download results tables for further analysis, filter the table and viewers based on specific residues, and to filter the table by score criteria and annotations (e.g. secondary structure, hydrogen bond for wild-type residue). For convenience, positions are highlighted based on favorability defined by ProteinMPNN score cutoffs (as noted below), but can be sorted or filtered by users as needed based on additional or other criteria. Proscan running time depends on input protein size, and typically takes seconds to several minutes for most standard sized proteins (<400 residues), while it only takes seconds with Rosetta ΔΔG calculations omitted.
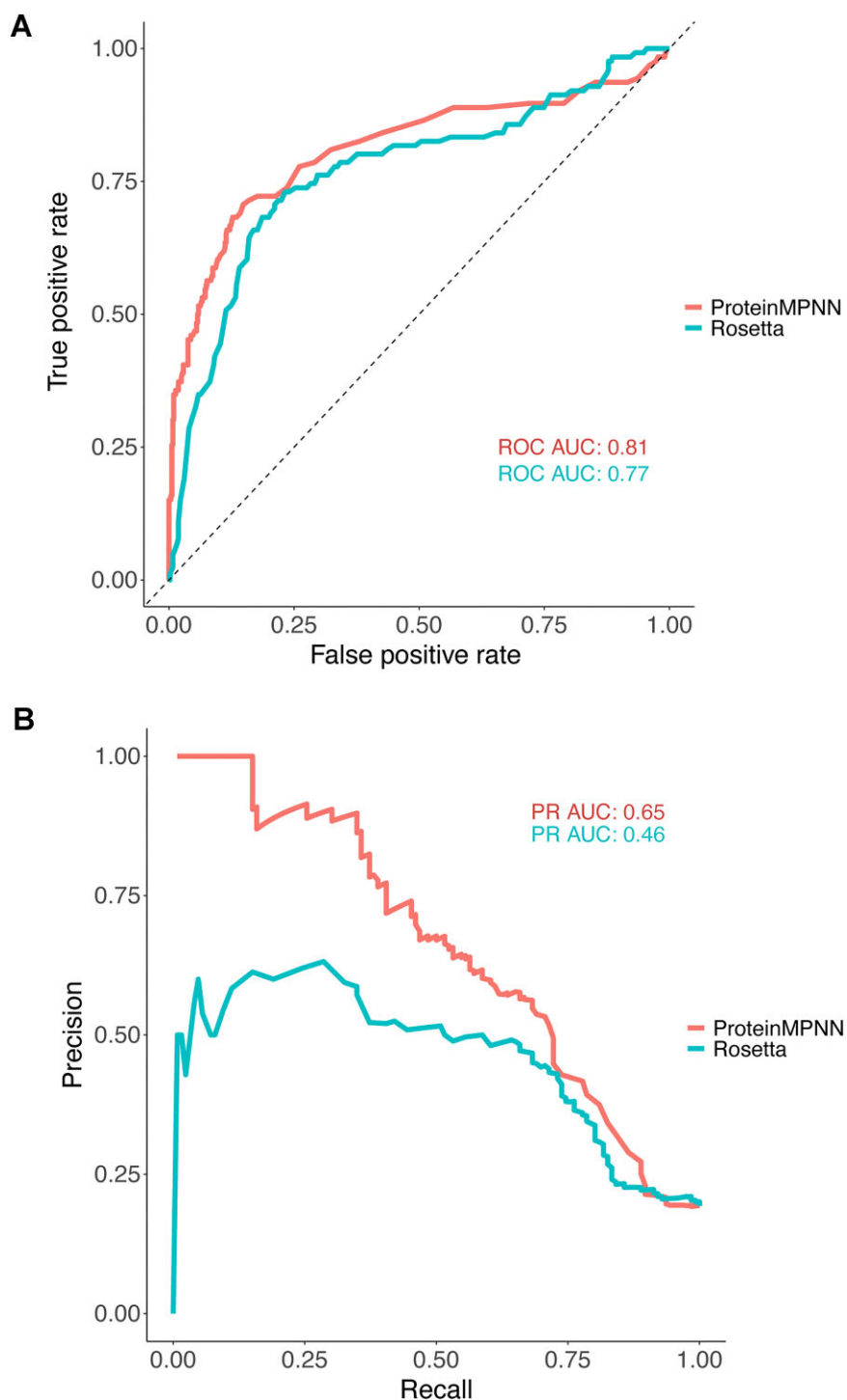
Figure 2. Classification accuracy of ProteinMPNN and Rosetta for proline substitution stability effects from the MegaScale protein stability dataset (28). (**A**) Receiver operating characteristic (ROC) curves for ProteinMPNN and Rosetta scoring of stabilizing versus destabilizing proline substitutions (0.5 kcal/mol $\Delta\Delta$G cutoff). Numbers of points are 126 (stabilizing) and 526 (destabilizing), and area under curve (AUC) values are shown. (**B**) Precision-recall curves for ProteinMPNN and Rosetta scoring of the same proline substitutions from (A), with AUC values shown.

Benchmarking and score cutoffs

To examine whether Proscan can correctly assess known beneficial proline substitutions, we used Proscan to assess a set of previously published proline substitutions in structures of viral antigen proteins that have high resolution structures of the unmutated proteins available. As shown in Table 1, Proscan favorably assessed the HCV E2, SARS-CoV-2 spike Hexapro, Ebola virus gp, RSV F and Dengue E proline designs based on most metrics. It should be noted that structures containing the designed proline substitutions (except for HCV E2 and Dengue E designs) are within the possible ProteinMPNN training set, however, it is not clear whether the ProteinMPNN model and its output would be biased by those potential instances.

To more systematically assess the predictive performance of Proscan and its output scores, we used the recently published MegaScale deep mutational scanning dataset (28), from which we obtained a set of over 600 measured proline substitution stability effects (Supplementary Table S1). We examined the performance of scores and criteria to classify substitutions that approximately stabilize or maintain stability ($\Delta\Delta G < 0.5$) versus those that disrupt stability ($\Delta\Delta G > 0.5$). Note that the $\Delta\Delta G$ values were negated from original dataset values (28) to align their polarity with most $\Delta\Delta G$ measurements and predictions (positive meaning less favorable), and units are approximately in kcal/mol. As seen in Figure 2A, both Rosetta and ProteinMPNN scores classify stabilizing vs. non-stabilizing substitutions with similar performance based on received operating characteristic (ROC) curve and area under the curve (ROC AUC). Given the somewhat imbalanced dataset (approximately 4:1 destabilizing vs. stabilizing substitutions), we also compared ProteinMPNN and Rosetta with precision-recall curves (Figure 2B), showing that ProteinMPNN is better overall at prediction of proline substitution effects, which is not unexpected given that modeling backbone and other effects from proline substitutions in physics-based modeling programs such as Rosetta can be challenging.

We performed a more detailed analysis of classification accuracies for Ramachandran plot criteria, Rosetta ($\Delta\Delta G$ value cutoff 0.1), and ProteinMPNN (probability score cutoff values 0.2 and 0.01) with the MegaScale dataset (Supplementary Table S2). While backbone angle-based classifications alone (current residue Preferable, or Preferable/Acceptable/No_angle for proline) were effective at identifying some negative (destabilizing) mutations, many false positive (destabilizing) mutations passed the angle filters. Rosetta $\Delta\Delta G$ at the tested cutoff showed some specificity (precision = 0.59) but did not correctly identify the majority of the favorable/neutral substitutions (recall = 0.35). ProteinMPNN performed well in classifying substitutions, with the stricter cutoff (0.2) showing a 1.0 precision value (although detecting only 14% of favorable/neutral substitutions), and the more permissive cutoff (0.01) showing a better balance, with 0.64 precision and 0.56 recall. Combination of the backbone angle-based filters did not markedly improve performance of either Rosetta and ProteinMPNN at the tested score cutoffs. Based on the superior performance of ProteinMPNN in this context, Proscan uses ProteinMPNN score and the two defined score cutoffs (0.2, 0.01) to highlight highly favorable and favorable predicted substitutions, respectively, in the results page table and figures.

To provide additional information regarding the predictive performance of Proscan and its metrics, we identified a set of 27 experimentally measured proline substitution stability changes ($\Delta\Delta G$s) from the Protherm database (30) (Supplementary Table S3). Computed ProteinMPNN proline probability (log-transformed) and Rosetta $\Delta\Delta G$ values calculated by Proscan with the wild-type structures were found to be highly correlated with the experimentally measured $\Delta\Delta G$s, with Pearson correlations of 0.69 and 0.66 for ProteinMPNN and Rosetta, respectively (Supplementary Figure S1).

Example use case

As an example use case for Proscan, we analyzed the antigen chain alone from the HC84.26.5D-E2$_{434-446}$ antibody-antigen complex structure (PDB code 4Z0X) (32), to deter-mine whether the unbound antigen chain without the context of the bound antibody would show favorable scores for the H445P substitution, as we observed for that position in the antibody-antigen complex structure in Table 1. Inputting PDB code 4Z0X, and specifying chain C for 'Only Include Chains' to scan the antigen chain alone gives the output results table shown in Supplementary Figure S2. As can be seen, residue 445 is the most favorable position for a proline substitution for the unbound antigen based on ProteinMPNN and Rosetta scores, with ProteinMPNN score of 0.69 (out of 1) and Rosetta $\Delta\Delta G$ of -1.0.

Discussion

We developed the Proscan server to enable researchers to readily identify and prioritize favorable proline substitutions to stabilize viral antigens and other proteins of interest. Given the range of antigens already successfully engineered with prolines (1), it is quite possible that additional viral and pathogen antigens can be optimized with proline substitutions to favor preferred conformations. Due to current and future emerging pathogens, such design approaches are useful tools in pandemic preparedness (33), as exemplified by the S-2P and Hexapro coronavirus spike designs (4,34), and it is possible that future structure-based antigen designs can be performed from sequence using structural models such as accurate deep learning-based modeling methods such as AlphaFold (35).

There are several possible limitations and considerations for Proscan usage. While Proscan is unable to perform multistate design directly to disfavor certain conformations (e.g. post-fusion antigen conformations) while favoring others (pre-fusion conformation), by downloading Proscan results tables for two or more analyzed structures, the results can be analyzed and compared to effectively address such scenarios. Additionally, while limited structure quality (e.g. resolution > 3.5 Å) may affect Proscan's performance and output utility due to the sensitivity of some of the structure analysis programs, structure refinement can be performed by users to resolve questionable or erroneous geometries prior to Proscan input. While Proscan is expected to prioritize favorable proline substitutions, more detailed simulations of potential dynamic and energetic effects of substitutions can be performed on the top-ranked set of candidates from Proscan for further prioritization.

Data availability

The data underlying this article are available in the article and in its online supplementary material.

Supplementary data

Supplementary Data are available at NAR Online.

Conflict of interest statement

None declared

References

1. Sanders,R.W. and Moore,J.P. (2021) Virus vaccines: proteins prefer prolines. *Cell Host Microbe*, **29**, 327–333.
2. Pallesen,J., Wang,N., Corbett,K.S., Wrapp,D., Kirchdoerfer,R.N., Turner,H.L., Cottrell,C.A., Becker,M.M., Wang,L., Shi,W., *et al.* (2017) Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E7348–E7357.
3. Wrapp,D., Wang,N., Corbett,K.S., Goldsmith,J.A., Hsieh,C.L., Abiona,O., Graham,B.S. and McLellan,J.S. (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, **367**, 1260–1263.
4. Hsieh,C.L., Goldsmith,J.A., Schaub,J.M., DiVenere,A.M., Kuo,H.C., Javanmardi,K., Le,K.C., Wrapp,D., Lee,A.G., Liu,Y., *et al.* (2020) Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science*, **369**, 1501-1505.
5. Barnes,C.O., Jette,C.A., Abernathy,M.E., Dam,K.A., Esswein,S.R., Gristick,H.B., Malyutin,A.G., Sharaf,N.G., Huey-Tubman,K.E., Lee,Y.E., *et al.* (2020) SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature*, **588**, 682–687.
6. Wang,Z., Schmidt,F., Weisblum,Y., Muecksch,F., Barnes,C.O., Finkin,S., Schaefer-Babajew,D., Cipolla,M., Gaebler,C., Lieberman,J.A., *et al.* (2021) mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature*, **592**, 616–622.
7. Jones,B.E., Brown-Augsburger,P.L., Corbett,K.S., Westendorf,K., Davies,J., Cujec,T.P., Wiethoff,C.M., Blackbourne,J.L., Heinz,B.A., Foster,D., *et al.* (2021) The neutralizing antibody, LY-CoV555, protects against SARS-CoV-2 infection in nonhuman primates. *Sci. Transl. Med.*, **13**, eabf1906.
8. Lu,M., Chamblee,M., Zhang,Y., Ye,C., Dravid,P., Park,J.G., Mahesh,K.C., Trivedi,S., Murthy,S., Sharma,H., *et al.* (2022) SARS-CoV-2 prefusion spike protein stabilized by six rather than two prolines is more potent for inducing antibodies that neutralize viral variants of concern. *Proc. Natl. Acad. Sci. U.S.A.*, **119**, e2110105119.
9. Sanders,R.W., Derking,R., Cupo,A., Julien,J.P., Yasmeen,A., de Val,N., Kim,H.J., Blattner,C., de la Pena,A.T., Korzun,J., *et al.* (2013) A next-generation cleaved, soluble HIV-1 Env trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. *PLoS Pathog.*, **9**, e1003618.
10. Pierce,B.G., Keck,Z.Y., Wang,R., Lau,P., Garagusi,K., Elkholy,K., Toth,E.A., Urbanowicz,R.A., Guest,J.D., Agnihotri,P., *et al.* (2020) Structure-based design of Hepatitis C virus E2 glycoprotein improves serum binding and cross-neutralization. *J. Virol.*, **94**, e00704-20.
11. Pierce,B.G., Hellman,L.M., Hossain,M., Singh,N.K., Vander Kooi,C.W., Weng,Z. and Baker,B.M. (2014) Computational design

12. Haidar,J.N., Zhu,W., Lypowy,J., Pierce,B.G., Bari,A., Persaud,K., Luna,X., Snavely,M., Ludwig,D. and Weng,Z. (2013) Backbone flexibility of CDR3 and the kinetics of immune recognition of antigens: a computational and experimental study. *J. Mol. Biol.*, **426**, 1583–1599.
13. Sullivan,J.T., Sulli,C., Nilo,A., Yasmeen,A., Ozorowski,G., Sanders,R.W., Ward,A.B., Klasse,P.J., Moore,J.P. and Doranz,B.J. (2017) High-throughput protein engineering improves the antigenicity and stability of soluble HIV-1 envelope glycoprotein SOSIP trimers. *J. Virol.*, **91**, e00862-17.
14. Rawi,R., Shen,C.H., Kwong,P.D. and Chuang,G.Y. (2018) CRISPro: an automated pipeline for protein conformation stabilization by Proline. *J. Chem. Inf. Model.*, **58**, 2189–2192.
15. Bajaj,K., Madhusudhan,M.S., Adkar,B.V., Chakrabarti,P., Ramakrishnan,C., Sali,A. and Varadarajan,R. (2007) Stereochemical criteria for prediction of the effects of proline mutations on protein stability. *PLoS Comput. Biol.*, **3**, e241.
16. Kortemme,T. and Baker,D. (2002) A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 14116–14121.
17. Anderson,R.J., Weng,Z., Campbell,R.K. and Jiang,X. (2005) Main-chain conformational tendencies of amino acids. *Proteins*, **60**, 679–689.
18. Dauparas,J., Anishchenko,I., Bennett,N., Bai,H., Ragotte,R.J., Milles,L.F., Wicky,B.I.M., Courbet,A., de Haas,R.J., Bethel,N., *et al.* (2022) Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, **378**, 49–56.
19. Rose,P.W., Beran,B., Bi,C., Bluhm,W.F., Dimitropoulos,D., Goodsell,D.S., Prlic,A., Quesada,M., Quinn,G.B., Westbrook,J.D., *et al.* (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res.*, **39**, D392–D401.
20. Cock,P.J., Antao,T., Chang,J.T., Chapman,B.A., Cox,C.J., Dalke,A., Friedberg,I., Hamelryck,T., Kauff,F., Wilczynski,B., *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.
21. Williams,C.J., Headd,J.J., Moriarty,N.W., Prisant,M.G., Videau,L.L., Deis,L.N., Verma,V., Keedy,D.A., Hintze,B.J., Chen,V.B., *et al.* (2018) MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.*, **27**, 293–315.
22. Rose,A.S., Bradley,A.R., Valasatava,Y., Duarte,J.M., Prlic,A. and Rose,P.W. (2018) NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics*, **34**, 3755–3758.
23. Lovell,S.C., Davis,I.W., Arendall,W.B. 3rd, de Bakker,P.I., Word,J.M., Prisant,M.G., Richardson,J.S. and Richardson,D.C. (2003) Structure validation by Calpha geometry: phi,psi and cbeta deviation. *Proteins*, **50**, 437–450.
24. Pierce,B.G., Keck,Z.Y., Lau,P., Fauvelle,C., Gowthaman,R., Baumert,T.F., Fuerst,T.R., Mariuzza,R.A. and Foung,S.K.H. (2016) Global mapping of antibody recognition of the hepatitis C virus E2 glycoprotein: implications for vaccine design. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E6946–E6954.
25. Yin,R., Guest,J.D., Taherzadeh,G., Gowthaman,R., Mittra,I., Quackenbush,J. and Pierce,B.G. (2021) Structural and energetic profiling of SARS-CoV-2 receptor binding domain antibody recognition and the impact of circulating variants. *PLoS Comput. Biol.*, **17**, e1009380.
26. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
27. McDonald,I.K. and Thornton,J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, **238**, 777–793.
28. Tsuboyama,K., Dauparas,J., Chen,J., Laine,E., Mohseni Behbahani,Y., Weinstein,J.J., Mangan,N.M., Ovchinnikov,S. and Rocklin,G.J. (2023) Mega-scale experimental analysis of protein folding stability in biology and design. *Nature*, **620**, 434–444.

29. Steinegger,M. and Soding,J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.*, **35**, 1026–1028.

30. Nikam,R., Kulandaisamy,A., Harini,K., Sharma,D. and Gromiha,M.M. (2021) ProThermDB: thermodynamic database for proteins and mutants revisited after 15 years. *Nucleic Acids Res.*, **49**, D420–D424.

31. Sumida,K.H., Nunez-Franco,R., Kalvet,I., Pellock,S.J., Wicky,B.I.M., Milles,L.F., Dauparas,J., Wang,J., Kipnis,Y., Jameson,N., *et al.* (2024) Improving protein expression, stability, and function with ProteinMPNN. *J. Am. Chem. Soc.*, **146**, 2054–2061.

32. Keck,Z.Y., Wang,Y., Lau,P., Lund,G., Rangarajan,S., Fauvelle,C., Liao,G.C., Holtsberg,F.W., Warfield,K.L., Aman,M.J., *et al.* (2016) Affinity maturation of a broadly neutralizing human monoclonal antibody that prevents acute hepatitis C virus infection in mice. *Hepatology*, **64**, 1922–1933.

33. Corbett,K.S., Edwards,D.K., Leist,S.R., Abiona,O.M., Boyoglu-Barnum,S., Gillespie,R.A., Himansu,S., Schafer,A., Ziwawo,C.T., DiPiazza,A.T., *et al.* (2020) SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature*, **586**, 567–571.

34. Kirchdoerfer,R.N., Wang,N., Pallesen,J., Wrapp,D., Turner,H.L., Cottrell,C.A., Corbett,K.S., Graham,B.S., McLellan,J.S. and Ward,A.B. (2018) Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Sci. Rep.*, **8**, 15701.

35. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Zidek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.

36. Kong,L., Giang,E., Nieusma,T., Kadam,R.U., Cogburn,K.E., Hua,Y., Dai,X., Stanfield,R.L., Burton,D.R., Ward,A.B., *et al.* (2013) Hepatitis C virus E2 envelope glycoprotein core structure. *Science*, **342**, 1090–1094.

37. Krarup,A., Truan,D., Furmanova-Hollenstein,P., Bogaert,L., Bouchier,P., Bisschop,I.J.M., Widjojoatmodjo,M.N., Zahn,R., Schuitemaker,H., McLellan,J.S., *et al.* (2015) A highly stable prefusion RSV F vaccine derived from structural analysis of the fusion mechanism. *Nat. Commun.*, **6**, 8143.

38. Rutten,L., Gilman,M.S.A., Blokland,S., Juraszek,J., McLellan,J.S. and Langedijk,J.P.M. (2020) Structure-based design of prefusion-stabilized filovirus glycoprotein trimers. *Cell Rep.*, **30**, 4540–4550.

39. Kudlacek,S.T., Metz,S., Thiono,D., Payne,A.M., Phan,T.T.N., Tian,S., Forsberg,L.J., Maguire,J., Seim,I., Zhang,S., *et al.* (2021) Designed, highly expressing, thermostable dengue virus 2 envelope protein dimers elicit quaternary epitope antibodies. *Sci. Adv.*, **7**, eabg4084.