K-Pro: Kinetics Data on Proteins and Mutants

Paola Turina [1], Piero Fariselli [2] and Emidio Capriotti [1]*

*1 - **Department of Pharmacy and Biotechnology (FaBiT),** University of Bologna, Via F. Selmi 3, 40126 Bologna, Italy*
*2 - **Department of Medical Sciences,** University of Torino, Via Santena 19, 10126 Torino, Italy*

***Correspondence to Emidio Capriotti:** emidio.capriotti@unibo.it (E. Capriotti) @EmidioCapriotti (E. Capriotti)*
https://doi.org/10.1016/j.jmb.2023.168245
Edited by Michael Sternberg

Abstract

The study of protein folding plays a crucial role in improving our understanding of protein function and of the relationship between genetics and phenotypes. In particular, understanding the thermodynamics and kinetics of the folding process is important for uncovering the mechanisms behind human disorders caused by protein misfolding. To address this issue, it is essential to collect and curate experimental kinetic and thermodynamic data on protein folding. K-Pro is a new database designed for collecting and storing experimental kinetic data on monomeric proteins, with a two-state folding mechanism. With 1,529 records from 62 proteins corresponding to 65 structures, K-Pro contains various kinetic parameters such as the logarithm of the folding and unfolding rates, Tanford's β and the ϕ values. When available, the database also includes thermodynamic parameters associated with the kinetic data. K-Pro features a user-friendly interface that allows browsing and downloading kinetic data of interest. The graphical interface provides a visual representation of the protein and mutants, and it is cross-linked to key databases such as PDB, UniProt, and PubMed. K-Pro is open and freely accessible through https://folding.biofold.org/k-pro and supports the latest versions of popular browsers.
© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

The protein folding problem encompasses three key aspects: understanding the thermodynamics of the process, determining the protein folding speed to address the Levinthal's puzzle, and developing a computational method for structure prediction.[1,2] While significant progress has been made in the prediction of protein structure with the advent of AlphaFold[3] and the Levinthal's Paradox has been solved for single domain globular proteins,[4,5] the accurate prediction of the folding rate of proteins and their mutants remains a challenging task.[6] Accurate theoretical models that can predict folded protein stability and the kinetics of the folding process,[7–12] are crucial for enhancing our understanding of disease mechanisms that often involve protein unfolding or aggregation.[13,14] The basic hypothesis is that variations in the thermodynamics and kinetics of the folding process may induce protein loss or excess, and/or promote the formation of potentially damaging aggregates.[15–17] The connection between variations in protein stability and human disorders[18] has enhanced the collection of experimental data on protein folding thermodynamics and the implementation of methods for predicting the impact of mutants on protein stability.[19,20]

In contrast, databases for experimental data on folding kinetics are not frequently updated or properly maintained.[21–24] Although a few methods have been developed for predicting folding/unfolding rates[25–28] and their variation upon mutation,[29,30] the lack of curated databases collecting kinetic data hampers the development and validation of such

0022-2836/© 2023 The Authors. Published by Elsevier Ltd.This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

algorithms. To address this issue, we introduce K-Pro, a new database that collects experimental data on the folding kinetics of proteins and mutants.

K-Pro database focuses on proteins which show convincing experimental evidence of undergoing a two-state folding process, i.e. absence of kinetically detectable intermediates, as outlined e.g. in 31. In the current version, K-Pro contains 1,529 records of kinetic data from 62 proteins corresponding to 65 structures, extracted from previously developed databases[21–24] and datasets[10,32,33] and from recently scanned literature using *ThermoScan*.[34] In the present work, this semi-automatic method was only used to facilitate the search for PubMed articles containing both thermodynamic and kinetic experiemntal data on protein folding.

The absence of some kinetic data relative to previous datasets can be attributed to slightly more stringent criteria applied for their evaluation and selection, which exclude repeat proteins, as well as cases with inconsistent or insufficient experimental evidence for the two-state nature of the protein.

The database features a user-friendly web interface for browsing and downloading, with information from UniProt,[35] Protein Data Bank (PDB),[36] and PubMed, cross-linked for each record. Additionally, when the link to the mutation identifier is clicked, a graphical interface allows for the visualization of the three-dimensional structure of the reported proteins. K-Pro is freely accessible without any registration requirements.

Content of the database

K-Pro provides experimental kinetic data on both wild-type and mutant proteins, sourced from the literature. These kinetic parameters have been experimentally determined through the use of three main methods: stopped-flow, continuous-flow, and T-jump. Stopped-flow is favored for capturing the rapid dynamics of protein folding, as it provides high temporal resolution, while continuous-flow is better suited for mimicking the conditions of the reaction and monitoring it over a longer period of time. The T-jump method, on the other hand, measures the kinetics of protein folding by rapidly changing the temperature of the reaction and monitoring it over time. The above methods typically monitor the folding reaction by measuring changes in the fluorescence or absorbance of a specific chromophore. The main kinetic parameters stored in the database are ln(k_f) and ln(k_u), *i.e.* the logarithms of the unfolding and folding rate constants, respectively. When available, thermodynamic parameters obtained from the associated equilibrium experiments are also collected.

In detail, each K-Pro entry is identified by a unique accession number, which includes the following information:

1. Protein Sequence: protein name, source, UniProt ID, protein length, wild-type or mutated UniProt sequence.
2. Protein Structure: PDB codes for the wild-type structure (if available), chain name, and mutation details (wild-type and mutant residues along with mutated position). In general, for each protein, the PDB code indicated by the authors of the kinetic measurements was maintained, except in a few cases, in which a crystal structure with a better resolution, and higher coverage, has been published in the meantime. For records with mutation data, the relative solvent accessibility (RSA) and secondary structure (Sec. Str.) of the wild-type residue, calculated with the *dssp* program,[37] are also included.
3. Experimental conditions: temperature, pH, buffer, added ions name and concentration, additives, and measurement method.
4. Literature information: name of the author(s), title of the article and journal, year of publication and PubMed identifier.
5. Kinetic and thermodynamic data: Logarithm of the folding and unfolding rates in water ($\ln(k_f^{H2O})$, $\ln(k_u^{H2O})$) and/or in the presence of denaturant ($\ln(k_f^{DEN})$, $\ln(k_u^{DEN})$), the equilibrium free energy change at zero denaturant (ΔG^{H2O}) (derived from the kinetic data, ΔG_KIN^{H2O}, or from the associated equilibrium measurements, ΔG_EQ^{H2O}), the folding and unfolding slopes of the Chevron plot (m_f, m_u), the concentration of denaturant at 50% of unfolded protein (C_m), the Tanford's β[38] and the folding ϕ value.[39] For mutants, the variation of the logarithm of the folding and unfolding rates in water ($\Delta\ln(k_f^{H2O})$, $\Delta\ln(k_u^{H2O})$) and the variation of the free energy change of unfolding in water ($\Delta\Delta G^{H2O}$), relative to wild-type, are also reported.

The UniProt and PDB identifiers are used to aggregate additional information relative to the enzymatic activity, the protein family and structural class. When available, links to the Enzyme Classification[40], PFAM[41] and CATH[42] identifiers are provided.

If the experimental study was carried out taking as reference a mutated version of the original wild-type protein, such information is reported in the *background_mut* field.

Database implementation and webserver interface

The K-Pro interface, implemented with HTML/JavaScript, processes the user requests, and queries a backend NoSQL database, implemented with MongoDB, which stores data as documents similar to JSON objects. Once the information is retrieved from the database, the server formats a webpage which uses the DataTable plug-in[43] for jQuery library for displaying the output table. The default view of the output table shows, for each record, 12 columns (UniProt, Struc-

ture, Chain, Mutation, Sec. Str., RSA, T, pH, $\ln(k_f^{H2O})$, $\Delta\ln(k_f^{H2O})$, $\ln(k_u^{H2O})$, $\Delta\ln(k_u^{H2O})$). The UniProt and PDB identifiers are linked to the relative webpages of the corresponding databases. When the three-dimensional structure of a protein is not available, a link to the structure predicted with AlphaFold is provided. Furthermore, the link of the mutation field allows the visualization of the three-dimensional structure of the protein and the mutation environment through the JSmol interface.[44] A plus sign button (+) shows all pieces of information associated with each measure. For direct access to the referenced publication, a link to the PubMed database is included.

The K-Pro webserver also implements a REST API backend, accessible through the https://folding.biofold.org/k-pro/api URL. The API interface is implemented by combining flask Welcome to Flask – Flask Documentation,[45] a python framework for web applications, with swagger,[46] a tool for designing and documenting REST API. Queries through the API are organized in 4 groups which are designed to retrieve data from specific proteins, mutation types, experiments, and publications. The API documentation is linked to the main webpage.

Database statistics

The K-Pro database contains 1,529 entries from 62 two-state folding proteins, corresponding to 65 structures. The mapping between protein sequences and structures always exhibits some exceptions. The discrepancy in the numbers arises from the presence of multiple domains belonging to the same protein and mutants linked to different protein structures. The majority of the kinetic data records are from single-point mutants, accounting for approximately 87.7% of the database, with the remaining 12% split between wild-type (6%) and multiple-point (6.3%) mutations. For the 1,341 single-point mutants, a structural analysis shows that 43% of the mutated residues were in a helical conformation, 29% were in beta regions, and 27% were in a coil conformation. Based on the fraction of the surface of the wild-type residue exposed to solvent (RSA), the mutated sites were then divided into three groups: buried (RSA $\leq$ 10), intermediate (10 < RSA < 40), and exposed (RSA $\geq$ 40). According to this classification, buried, exposed and intermediate sites each makes up ~33% of the total. To assess the impact of single amino acid substitutions on protein folding kinetics, the distribution of the variations in the logarithms of folding and unfolding rates was analyzed. The data collected in K-Pro suggest that the majority of mutations slow down the folding process, decreasing the average logarithm of k_f^{H2O} by approximately 0.6, and accelerate the unfolding process, increasing the logarithm of k_u^{H2O} by approximately 1.9. In both cases, the standard deviation of the distribution is similar to the mean, reaching 0.9 for $\Delta\ln(k_f^{H2O})$ and 1.8 for $\Delta\ln(k_u^{H2O})$. The difference between $\Delta\ln(k_f^{H2O})$ and $\Delta\ln(k_u^{H2O})$, which is proportional to the variation of Gibbs free energy change of the unfolding process upon mutation, mostly results in a negative value, in agreement with the observation that mutations tend on average to have a destabilizing effect on the protein structures.[47] The results of such analysis are summarized in Figure 1.

The distributions of K-Pro entries based on protein length and their UniProt and PDB identifiers are summarized in Tables S4–S6. The composition of the databases in terms of structural and functional classification described by the PFAM, CATH and the enzymatic activity (EC number) are reported in Tables S7–S9.

At the current stage, K-Pro collects additional ~500 data on mutants relative to the previously selected datasets[32,33] and a similar number of data on wild-type proteins as does PFDB.[24] Furthermore, K-Pro includes data on ~100 multiple site mutants. From the user standpoint, K-Pro features an intuitive interface that allows querying the database by command line and the direct visualization of both the protein structures and the structural environment of the mutated sites.

Data retrieval

For querying the database, the K-Pro web interface implements both basic and advanced searches. The basic mode allows users to submit simple queries using UniProt and PDB identifiers, and to select subsets of different types of mutations (wild-type, single, double, and multiple). The search for kinetic data linked with alternative PDB structures can be performed by selecting the option "*Match UniProt*". This option allows the retrieval of data from all the PDB structures associated with the same UniProt ID. The advanced interface enables more refined queries, based on the experimental techniques used to measure the kinetic constants, on the experimental conditions, and on publication information. In the output webpage, the selected data are displayed on a table with expandable records, which includes the kinetics data. On the top of the page, an *output.tsv* link allows downloading the result of your query in tab separated format. Queries can also be performed through the command line, by running curl and accessing specific API endpoints. An example of how to run a search via the command line is provided on the API documentation webpage. The

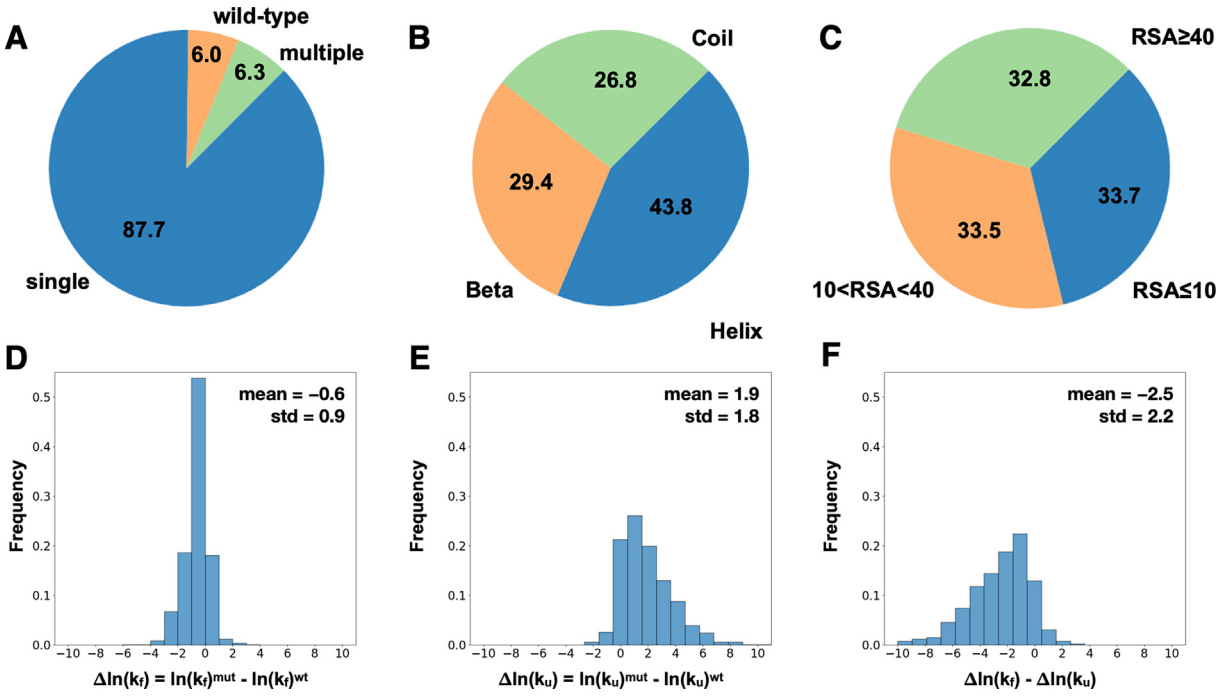

Fig. 1. K-Pro statistics: (A) Frequencies of wild-type, single and multiple mutants in the 1,529 records of K-Pro. (B) Frequencies of the secondary structure of the wild-type residue in 1,327 single mutants. (C) Frequencies of the buried (RSA $\leq$ 10), intermediate (10 < RSA < 40) and exposed (RSA $\geq$ 40) sites for the single mutants. (D)–(F) Distributions of the $\Delta\ln(k_f^{H2O})$, $\Delta\ln(k_u^{H2O})$ and their difference ($\Delta\ln(k_f^{H2O}) - \Delta\ln(k_u^{H2O})$) for single-point mutants. The complete statistics on data in panels (A)–(C) are reported in Tables S1–S3.

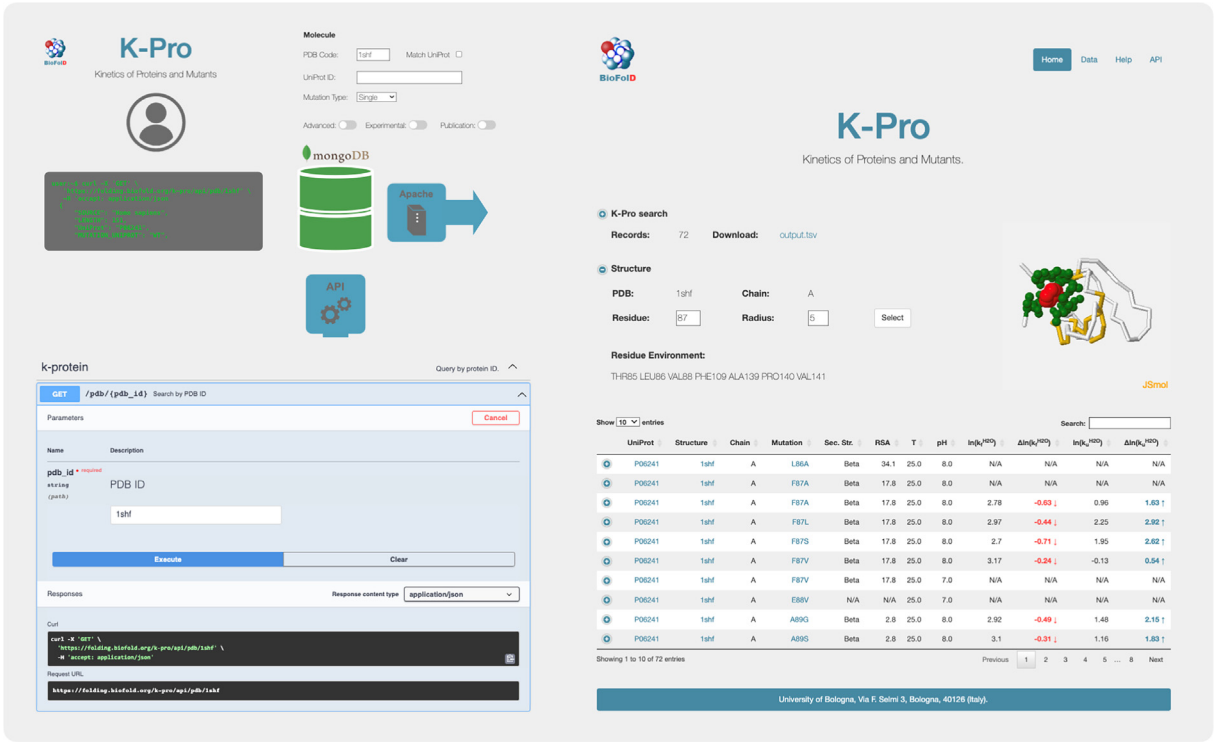

Fig. 2. Web interface of the K-Pro database. Top-left corner: K-Pro searches through the standard web interface or by command line. Bottom-left: an example of querying the database through the REST API endpoint. On the right, the K-Pro output is shown, including the visualization of the 3D protein structure.

server returns its output in JSON format, which is a commonly used data exchange format, known for its readability and ease of parsing by both humans and machines. Figure 2 highlights the main features of the K-Pro web interface and displays an example of output.

Funding

This work was supported by the PRIN project, "Integrative tools for defining the molecular basis of the diseases: Computational and Experimental methods for Protein Variant Interpretation" of the Ministero Istruzione, Università e Ricerca (PRIN201744NR8S) and by the initiatives of the European Network for Life Science ELIXIR.

CRediT authorship contribution statement

Paola Turina: Writing – review & editing, Data curation, Validation. **Piero Fariselli:** Writing – review & editing, Funding acquisition, Conceptualization. **Emidio Capriotti:** Writing – original draft, Writing – review & editing, Software, Supervision.

DATA AVAILABILITY

Data are made available on the website of the database

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.jmb.2023.168245.

Received 14 February 2023;
Accepted 17 August 2023;
Available online 23 August 2023

Keywords:
protein folding;
folding kinetics;
protein variant;
transition state;
folding rate constant

References

1. Dill, K.A., Ozkan, S.B., Weikl, T.R., Chodera, J.D., Voelz, V.A., (2007). The protein folding problem: when will it be solved? *Curr Opin Struct Biol* **17**, 342–346. https://doi.org/10.1016/j.sbi.2007.06.001.
2. Compiani, M., Capriotti, E., (2013). Computational and theoretical methods for protein folding. *Biochemistry* **52**, 8601–8624. https://doi.org/10.1021/bi4001529.
3. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., et al., (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
4. Ivankov, D.N., Finkelstein, A.V., (2020). Solution of Levinthal's paradox and a physical theory of protein folding times. *Biomolecules* **10**, 250. https://doi.org/10.3390/biom10020250.
5. Finkelstein, A.V., Bogatyreva, N.S., Ivankov, D.N., Garbuzynskiy, S.O., (2022). Protein folding problem: enigma, paradox, solution. *Biophys Rev* **14**, 1255–1272. https://doi.org/10.1007/s12551-022-01000-1.
6. Chang, C.C.H., Tey, B.T., Song, J., Ramanan, R.N., (2015). Towards more accurate prediction of protein folding rates: a review of the existing Web-based bioinformatics approaches. *Brief Bioinform* **16**, 314–324. https://doi.org/10.1093/bib/bbu007.
7. Dill, K.A., (1990). Dominant forces in protein folding. *Biochemistry* **29**, 7133–7155.
8. Tokuriki, N., Tawfik, D.S., (2009). Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* **19**, 596–604. https://doi.org/10.1016/j.sbi.2009.08.003.
9. Goldenzweig, A., Fleishman, S.J., (2018). Principles of protein stability and their application in computational design. *Annu Rev Biochem* **87**, 105–129. https://doi.org/10.1146/annurev-biochem-062917-012102.
10. Glyakina, A.V., Galzitskaya, O.V., (2020). How quickly do proteins fold and unfold, and what structural parameters correlate with these values? *Biomolecules* **10**, 197. https://doi.org/10.3390/biom10020197.
11. Eaton, W.A., (2021). Modern kinetics and mechanism of protein folding: a retrospective. *J Phys Chem B* **125**, 3452–3467. https://doi.org/10.1021/acs.jpcb.1c00206.
12. Ooka, K., Liu, R., Arai, M., (2022). The Wako-Saitô-Muñoz-Eaton model for predicting protein folding and dynamics. *Mol Basel Switz* **27**, 4460. https://doi.org/10.3390/molecules27144460.
13. Chiti, F., Dobson, C.M., (2017). Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annu Rev Biochem* **86**, 27–68. https://doi.org/10.1146/annurev-biochem-061516-045115.
14. Dobson, C.M., Knowles, T.P.J., Vendruscolo, M., (2020). The amyloid phenomenon and its significance in biology and medicine. *Cold Spring Harb Perspect Biol* **12**, https://doi.org/10.1101/cshperspect.a033878 a033878.
15. Meisl, G., Kirkegaard, J.B., Arosio, P., Michaels, T.C.T., Vendruscolo, M., Dobson, C.M., Linse, S., Knowles, T.P.J., (2016). Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nature Protoc* **11**, 252–272. https://doi.org/10.1038/nprot.2016.010.

16. Michaels, T.C.T., Šarić, A., Habchi, J., Chia, S., Meisl, G., Vendruscolo, M., Dobson, C.M., Knowles, T.P.J., (2018). Chemical kinetics for bridging molecular mechanisms and macroscopic measurements of amyloid fibril formation. *Annu Rev Phys Chem* **69**, 273–298. https://doi.org/10.1146/annurev-physchem-050317-021322.
17. Dobson, C.M., (2001). The structural basis of protein folding and its links with human disease. *Philos Trans R Soc Lond B Biol Sci* **356**, 133–145. https://doi.org/10.1098/rstb.2000.0758.
18. Martelli, P.L., Fariselli, P., Savojardo, C., Babbi, G., Aggazio, F., Casadio, R., (2016). Large scale analysis of protein stability in OMIM disease related human protein variants. *BMC Genomics* **17** (Suppl 2), 397. https://doi.org/10.1186/s12864-016-2726-y.
19. Marabotti, A., Scafuri, B., Facchiano, A., (2021). Predicting the stability of mutant proteins by computational approaches: an overview. *Brief Bioinform* **22**, https://doi.org/10.1093/bib/bbaa074 bbaa074.
20. Pancotti, C., Benevenuta, S., Birolo, G., Alberini, V., Repetto, V., Sanavia, T., Capriotti, E., Fariselli, P., (2022). Predicting protein stability changes upon single-point mutation: a thorough comparison of the available tools on a new dataset. *Brief Bioinform* **23**, https://doi.org/10.1093/bib/bbab555 bbab555.
21. Fulton, K.F., Bate, M.A., Faux, N.G., Mahmood, K., Betts, C., Buckle, A.M., (2007). Protein Folding Database (PFD 2.0): an online environment for the International Foldeomics Consortium. *Nucleic Acids Res* **35**, D304–D307. https://doi.org/10.1093/nar/gkl1007.
22. Bogatyreva, N.S., Osypov, A.A., Ivankov, D.N., (2009). KineticDB: a database of protein folding kinetics. *Nucleic Acids Res* **37**, D342–D346. https://doi.org/10.1093/nar/gkn696.
23. Wagaman, A.S., Coburn, A., Brand-Thomas, I., Dash, B., Jaswal, S.S., (2014). A comprehensive database of verified experimental data on protein folding kinetics. *Protein Sci Publ Protein Soc* **23**, 1808–1812. https://doi.org/10.1002/pro.2551.
24. Manavalan, B., Kuwajima, K., Lee, J., (2019). PFDB: a standardized protein folding database with temperature correction. *Sci Rep* **9**, 1588. https://doi.org/10.1038/s41598-018-36992-y.
25. Plaxco, K.W., Simons, K.T., Baker, D., (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* **277**, 985–994. https://doi.org/10.1006/jmbi.1998.1645.
26. Gromiha, M.M., Selvaraj, S., (2001). Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction. *J Mol Biol* **310**, 27–32. https://doi.org/10.1006/jmbi.2001.4775.
27. Ivankov, D.N., Finkelstein, A.V., (2004). Prediction of protein folding rates from the amino acid sequence-predicted secondary structure. *PNAS* **101**, 8942–8944. https://doi.org/10.1073/pnas.0402659101.
28. Capriotti, E., Casadio, R., (2007). K-Fold: a tool for the prediction of the protein folding kinetic order and rate. *Bioinforma Oxf Engl* **23**, 385–386. https://doi.org/10.1093/bioinformatics/btl610.
29. Gromiha, M.M., Huang, L.-T., (2011). Machine learning algorithms for predicting protein folding rates and stability of mutant proteins: comparison with statistical methods. *Curr Protein Pept Sci* **12**, 490–502. https://doi.org/10.2174/138920311796957630.
30. Chaudhary, P., Naganathan, A.N., Gromiha, M.M., (1864). Prediction of change in protein unfolding rates upon point mutations in two state proteins. *BBA* **2016**, 1104–1109. https://doi.org/10.1016/j.bbapap.2016.06.001.
31. Gianni, S., Guydosh, N.R., Khan, F., Caldas, T.D., Mayor, U., White, G.W.N., DeMarco, M.L., Daggett, V., et al., (2003). Unifying features in protein-folding mechanisms. *PNAS* **100**, 13286–13291. https://doi.org/10.1073/pnas.1835776100.
32. Naganathan, A.N., Muñoz, V., (2010). Insights into protein folding mechanisms from large scale analysis of mutational effects. *PNAS* **107**, 8611–8616. https://doi.org/10.1073/pnas.1000988107.
33. Chaudhary, P., Naganathan, A.N., Gromiha, M.M., (2015). Folding RaCe: a robust method for predicting changes in protein folding rates upon point mutations. *Bioinforma Oxf Engl* **31**, 2091–2097. https://doi.org/10.1093/bioinformatics/btv091.
34. Turina, P., Fariselli, P., Capriotti, E., (2021). ThermoScan: semi-automatic identification of protein stability data from PubMed. *Front Mol Biosci* **8** https://doi.org/10.3389/fmolb.2021.620475.
35. UniProt Consortium, (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506–D515. https://doi.org/10.1093/nar/gky1049.
36. wwPDB Consortium, (2019). Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res* **47**, D520–D528. https://doi.org/10.1093/nar/gky949.
37. Kabsch, W., Sander, C., (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637. https://doi.org/10.1002/bip.360221211.
38. Tanford, C., (1970). Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* **24**, 1–95.
39. Matouschek, A., Kellis, J.T., Serrano, L., Fersht, A.R., (1989). Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**, 122–126. https://doi.org/10.1038/340122a0.
40. McDonald, A.G., Boyce, S., Tipton, K.F., (2015). Enzyme classification and nomenclature. *In: Encycl. Life Sci.*. John Wiley & Sons Ltd, pp. 1–11. https://doi.org/10.1002/9780470015902.a0000710.pub3.
41. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., et al., (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412–D419. https://doi.org/10.1093/nar/gkaa913.
42. Sillitoe, I., Bordin, N., Dawson, N., Waman, V.P., Ashford, P., Scholes, H.M., Pang, C.S.M., Woodridge, L., et al., (2021). CATH: increased structural coverage of functional space. *Nucleic Acids Res* **49**, D266–D273. https://doi.org/10.1093/nar/gkaa1079.
43. DataTables | Table plug-in for jQuery, (n.d.). https://datatables.net/ (accessed February 9, 2023).
44. Hanson, R.M., Prilusky, J., Renjian, Z., Nakane, T., Sussman, J.L., (2013). JSmol and the next-generation web-based representation of 3D molecular structure as applied to proteopedia. *Isr J Chem* **53**, 207–216. https://doi.org/10.1002/ijch.201300024.

45. Welcome to Flask — Flask Documentation (2.2.x), (n.d.). https://flask.palletsprojects.com/en/2.2.x/ (accessed February 9, 2023).
46. API Documentation & Design Tools for Teams | Swagger, (n.d.). https://swagger.io/ (accessed February 9, 2023).
47. Benevenuta, S., Birolo, G., Sanavia, T., Capriotti, E., Fariselli, P., (2022). Challenges in predicting stabilizing variations: an exploration. *Front Mol Biosci* **9**, https://doi.org/10.3389/fmolb.2022.1075570 1075570.