

Resource Summary Report

Generated by [kravitz2](#) on Jul 30, 2026

Protein Classification Benchmark Collection

RRID:SCR_007561

Type: Tool

Proper Citation

Protein Classification Benchmark Collection (RRID:SCR_007561)

Resource Information

URL: <http://net.icgeb.org/benchmark/>

Proper Citation: Protein Classification Benchmark Collection (RRID:SCR_007561)

Description: It was created in order to create standard datasets on which the performance of machine learning methods can be compared. The collection contains datasets of sequences and structures, each subdivided into positive/negative training/test sets. Such a subdivision is called a classification task. Typical tasks include the classification of structural domains in the SCOP and CATH databases based on their sequences, as well as various functional and taxonomic classification tasks. Running a performance evaluation test on an entire database can include many different classification tasks. These ensembles of classification tasks are encoded in a simple matrix format - called the cast matrix or membership table - that specifies the role of each sequence (or structure) in the different calculations. Each column of this matrix is a subdivision of the objects (rows) into positive/negative training/test sets. Typically, a database record contains such an ensemble of classification tasks, encoded in a single cast matrix. In addition, there is a collection of distance matrices that contain an all vs. all comparison of the datasets using methods as BLAST, Smith-Waterman, 3D-comparisons etc. Evaluation of a method on a given database consists of calculating a performance measure such as a receiver operating curve (ROC) AUC value. Results of evaluation are deposited along with the data, each dataset is evaluated at least by one classification method, such as 1NN (nearest neighbour) or SVM (support vector machines), ANN (artificial neural networks), RF (random forests) etc.. There are small datasets meant for program developers, as well as downloadable programs for various classification algorithms.

Synonyms: Benchmark

Resource Type: data or information resource, database

Keywords: classification, machine learning, sequence, standard, standard dataset, structure, technology

Resource Name: Protein Classification Benchmark Collection

Resource ID: SCR_007561

Alternate IDs: nif-0000-02600

Record Creation Time: 20220129T080242+0000

Record Last Update: 20260729T044205+0000

Ratings and Alerts

No rating or validation information has been found for Protein Classification Benchmark Collection.

No alerts have been found for Protein Classification Benchmark Collection.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We found 24 mentions in open access literature.

Listed below are recent publications. The full list is available at [kravitz2](#) .

Shi C, et al. (2025) CellBinDB: a large-scale multimodal annotated dataset for cell segmentation with benchmarking of universal models. GigaScience, 14.

Xu M, et al. (2025) A study on the classification and prediction of firefighter's operational fatigue level. PloS one, 20(5), e0323911.

Lei Z, et al. (2022) Coordination modulation of iridium single-atom catalyst maximizing water oxidation activity. Nature communications, 13(1), 24.

Yang Y, et al. (2020) E-jet 3D printed drug delivery implants to inhibit growth and metastasis of orthotopic breast cancer. Biomaterials, 230, 119618.

Weng G, et al. (2019) HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. Nucleic acids research, 47(W1), W322.

Dos Santos Vasconcelos CR, et al. (2018) Building protein-protein interaction networks for Leishmania species through protein structural information. BMC bioinformatics, 19(1), 85.

Feng T, et al. (2017) HawkRank: a new scoring function for protein-protein docking based on weighted energy terms. Journal of cheminformatics, 9(1), 66.

Adang EM, et al. (2016) Efficiency of the implementation of cardiovascular risk management in primary care practices: an observational study. Implementation science :

IS, 11, 67.

Zhang C, et al. (2016) The Relationship between B-cell Epitope and Mimotope Sequences. Protein and peptide letters, 23(2), 132.

Basu S, et al. (2016) Finding correct protein-protein docking models using ProQDock. Bioinformatics (Oxford, England), 32(12), i262.

Basu S, et al. (2016) DockQ: A Quality Measure for Protein-Protein Docking Models. PloS one, 11(8), e0161879.

Sun P, et al. (2015) Conformational B-cell epitope prediction method based on antigen preprocessing and mimotopes analysis. BioMed research international, 2015, 257030.

Segura J, et al. (2015) VORFFIP-driven dock: V-D2OCK, a fast and accurate protein docking strategy. PloS one, 10(3), e0118107.

Liu H, et al. (2015) Applying Side-chain Flexibility in Motifs for Protein Docking. Genomics insights, 8, 1.

Maheshwari S, et al. (2015) Predicted binding site information improves model ranking in protein docking using experimental and computer-generated target structures. BMC structural biology, 15, 23.

Ohue M, et al. (2014) MEGADOCK: an all-to-all protein-protein interaction prediction system using tertiary structure data. Protein and peptide letters, 21(8), 766.

Vreven T, et al. (2013) Exploring angular distance in protein-protein docking algorithms. PloS one, 8(2), e56645.

Moal IH, et al. (2013) The scoring of poses in protein-protein docking: current capabilities and future directions. BMC bioinformatics, 14, 286.

Swapna LS, et al. (2012) Comparison of tertiary structures of proteins in protein-protein complexes with unbound forms suggests prevalence of allostery in signalling proteins. BMC structural biology, 12, 6.

Zhang W, et al. (2011) Prediction of conformational B-cell epitopes from 3D structures by random forests with a distance-based feature. BMC bioinformatics, 12, 341.