

Resource Summary Report

Generated by [ASWG](#) on Aug 3, 2026

Nh3D: A Reference Dataset of Structures of Non-homologous Proteins

RRID:SCR_008212

Type: Tool

Proper Citation

Nh3D: A Reference Dataset of Structures of Non-homologous Proteins
(RRID:SCR_008212)

Resource Information

URL: <http://www.schematikon.org/Nh3D.html>

Proper Citation: Nh3D: A Reference Dataset of Structures of Non-homologous Proteins
(RRID:SCR_008212)

Description: THIS RESOURCE IS NO LONGER IN SERVICE, documented on July 17, 2013. It is freely available as a reference dataset for the statistical analysis of sequence and structure features of proteins in the PDB. It is a dataset of structurally dissimilar proteins. This dataset has been compiled by selecting well resolved representatives from the Topology level of the CATH database which hierarchically classifies all protein structures. These have been pruned to remove: i) domains that may contain homologous elements (by pairwise sequence comparison and structural superposition of aligned residues) ii) internal duplications (by repeat detection) iii) regions with high B-Factor The statistical analysis of protein structures requires datasets in which structural features can be considered independently distributed, i.e. not related through common ancestry, and that fulfill minimal requirements regarding the experimental quality of the structures it contains. However, non-redundant datasets based on sequence similarity invariably contain distantly related homologues. Here a reference dataset of non-homologous protein domains is provided, assuming that structural dissimilarity at the topology level is incompatible with recognizable common ancestry. It contains the best refined representatives of each Topology level, validates structural dissimilarity and removes internally duplicated fragments. The compilation of Nh3D is fully scripted. The current Nh3D list contains 570 domains with a total of 90780 residues. It covers more than 70% of folds at the Topology level of the CATH database and represents more than 90% of the structures in the PDB that have been classified by CATH. Even though all protein pairs are structurally dissimilar, some pairwise sequence identities after global alignment are greater than 30%. Nh3D is freely available as a reference dataset for the statistical analysis of sequence and structure features of proteins in the PDB.

Abbreviations: Nh3D

Resource Type: database, data or information resource

Keywords: duplication, element, feature, fragment, align, alignment, analysis, b-factor, dissimilar, homologous, protein, protein structure databases, residue, sequence, statistical, structurally, structure, topology

Availability: THIS RESOURCE IS NO LONGER IN SERVICE

Resource Name: Nh3D: A Reference Dataset of Structures of Non-homologous Proteins

Resource ID: SCR_008212

Alternate IDs: nif-0000-21286

Record Creation Time: 20220129T080246+0000

Record Last Update: 20260803T040520+0000

Ratings and Alerts

No rating or validation information has been found for Nh3D: A Reference Dataset of Structures of Non-homologous Proteins.

No alerts have been found for Nh3D: A Reference Dataset of Structures of Non-homologous Proteins.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We have not found any literature mentions for this resource.