

## WEBLEM 7

### Introduction to Structural Blast – VAST and DALI

The protein structures that populate the PDB have provided crucial insights at the atomic level as to the molecular mechanisms that underlie protein function. Indeed, structural studies have had, and continue to have, a significant and sometimes revolutionary impact in all areas of biology. However, structural biology has tended to focus on single proteins or biological systems and, despite significant advances in the general area of structural bioinformatics, the horizontal integration of the vast quantity of structural information available in the PDB has had little or no impact in the larger biological community. This is in contrast to protein sequence information, which is more routinely, automatically and broadly used. Given that structure is more conserved than sequence, structural similarity has the potential to yield a great deal of functional information that sequence relationships cannot provide and to identify relationships between many more pairs of proteins. In this article we argue that the exploitation of statistical and machine learning techniques combined with the vast amount of high-throughput experimental data constantly being generated enable a significant expansion in the scale and diversity of application of structural information to biological problems.

The ultimate potential impact of both global and local structural relationships in inferring function is highlighted by the observation that, given a suitably “loose” definition of structural similarity, the repertoire of structures currently in the structural databases is nearly complete at the domain level. Thus, it can be expected that most newly solved protein structures will have both near and remote structural neighbors which can provide clues as to their function. Programs such as BLAST use local sequence relationships to quickly scan sequence databases. Since structure-based scans of protein structural databases can be carried out very quickly with current technology (typically minutes for a database of tens of thousands of structures), a similar strategy can be used for structural relationships as well, essentially defining a “structural BLAST”

Comparative analyses of protein sequences and structures play a fundamental role in understanding proteins and their functions. Assuming an evolutionary continuity of structure and function, describing the structural similarity relationships between protein structures allows scientists to infer the functions of newly discovered proteins. The most widespread purpose of structural alignment has been to identify homologous residues (encoded by the same codon in the genome of a common ancestor). Mutations manifest in plastic deformations, shifts and rotations of the secondary structure elements (SSEs). A wide spectrum of structural alignment methods exist, which differ in their treatment of structural variations, scoring functions and optimization algorithms.

There are aware of half a dozen web servers that provide structure comparisons against the current weekly updated Protein Data Bank (PDB). Each server is unique because they employ different structure comparison methods.

#### VAST:

The VAST search database and database of precomputed structure alignments have been maintained as complete and redundant collections since their launch, with automated updates occurring on a weekly basis. This was made possible by implementing a fast heuristic that uses a model for the statistical significance of initial alignments of secondary structure vectors (which can be computed quickly), so that the database searches can avoid costly alignment refinements for the large majority of insignificant and uninteresting similarities. The drawbacks are that a heuristic will miss some potentially interesting similarities. The VAST algorithm will not, for example, report similarities between structures deemed to have secondary structure elements. Searches for structural similarity can and should be complemented with searches for sequence similarity, as flexibility of molecular structure and limitations of the structure comparison method may preclude the detection of matches between structures of homologous polypeptides. In general, though, structure comparison methods will pick up many subtle similarities that evade detection by sequence

comparison strategies, and there is no natural cutoff point for a ranked list of similar structures, unlike in the sequence comparison scenario, where matches to non-homologous gene products are considered accidental and uninformative, for the most part.

Results computed by the VAST algorithm have been compared against other approaches a number of times. Although there are subtle differences in retrieval sensitivity and alignment accuracy, it appears fair to state that the large majority of extensive structural similarities, which are indicative of common evolutionary descent and could be used to infer functional similarities, are reported by VAST (and by most if not all of the alternative approaches to detect common substructures).

As structure similarity search strategies have been developed to also detect distant relationships that might not be evident from sequence analysis, most if not all of the current approaches have been implemented so that they use a single protein molecule or rather a single domain as the unit of comparison. This has been true for VAST, in particular. However, the Protein Data Bank is continuing to accumulate structures of larger macromolecular complexes and has started to provide data on what constitutes functionally or biologically relevant macromolecular complexes or biological assemblies. Such assemblies range from simple homooligomers to intricate arrangements of many different components, revealing details on specific molecular interactions and on how these might constrain sequence variation. A small number of approaches have been published in the past few years that examine structural similarity of macromolecular complexes. Here we present a simple strategy that builds on the existing database of pairwise structure alignments computed by VAST and supports the first (to our knowledge) comprehensive and regularly updated collection of macromolecular complex similarities.

## **VAST+ as an extension to existing protein structure comparison**

As information characterizing biological assemblies in macromolecular structure data has become available, it seemed that the biological assembly would be a convenient and informative unit of comparison between individual entries in the structure database. If the goal is to list structures most similar to any particular query, one would have to consider that the query itself may contain a macromolecular complex with a given stoichiometry, and that matching complexes with matching stoichiometry might be more informative ‘structure neighbors’ than, for example, the structures that happen to contain molecules with the strongest local similarity to the query, irrespective of the context.

VAST+ builds on the existing VAST database to generate such a report of structure neighbors. Its goal is to find the largest set of pairs of matching macromolecules between two biological assemblies and to characterize that match and compute instructions for a global superimposition that can be used to visualize the structural similarity. For each pair of structures in MMDDB, VAST+ examines pre-computed structure alignments stored in the VAST database that were computed for the full-length protein molecule components of the default biological assemblies. If such pairwise alignments are found, the alignments between individual protein components of the biological assemblies are compared with each other for compatibility, and compatible/matching alignments are clustered into sets of alignments that together constitute a biological assembly match. Pairwise alignments are compatible (i) if they do not share the same macromolecules, i.e. a protein molecule from one assembly cannot be aligned to two molecules from the other assembly at the same time and (ii) if they generate similar instructions (spatial transformation matrices) for the superpositions of coordinate sets. A simple distance metric can be used to compare transformation matrices and it lends itself to cluster alignment sets efficiently.

Each set of compatible pairwise alignments can be characterized by (i) the number of pairwise matches, i.e. the total number of pairs of protein molecules from the query and subject biological assemblies, that are simultaneously aligned with each other; (ii) the RMSD of the superposition obtained from considering all alignments in the set; (iii) the total length of all pairwise alignments, i.e. the total number of amino acids that are aligned in 3D space; and (iv) percentage of identical residues in the alignments. For each pairwise comparison of two biological assemblies, only the match with the highest number of aligned molecules and the highest number of aligned residues is recorded and reported.

Currently, 53% of polypeptide-containing structures in MMDB have >1 polypeptide chain. The histogram plotted in Figure 1 breaks down the numbers by oligomer size and indicates that large fractions of the oligomeric assemblies have, in general, structure neighbors that match the entire assemblies. It should be noted that the fractions might be somewhat exaggerated, as exact duplicates of a structure would be counted as biological assembly matches, and no attempt was made to remove redundant structures or classify biological assembly matches as informative versus uninformative.

## DALI:

The Dali server is a network service for comparing protein structures in 3D. You submit the coordinates of a query protein structure and Dali compares them against those in the Protein Data Bank (PDB). In favourable cases, comparing 3D structures may reveal biologically interesting similarities that are not detectable by comparing sequences.

User can perform three types of database searches:

- Heuristic PDB search - compares one query structure against those in the Protein Data Bank
- Exhaustive PDB25 search - compares one query structure against a representative subset of the Protein Data Bank
- Hierarchical AF-DB search - compares one query structure against a species subset of the AlphaFold Database

There are two types of structure comparisons of user selected structures:

- Pairwise structure comparison - compares one query structure against those specified by the user
- All against all structure comparison - returns a structural similarity dendrogram for a set of structures specified by the user

## DESCRIPTION OF THE SERVER

### Inputs

The input to the server is one or two protein structures in PDB format. The query structure can be specified as a PDB identifier plus chain identifier, or a PDB file uploaded by the user. There are three cross-linked query forms for the Dali server, Dali Database and pairwise comparison, respectively. For example, the entry point to the Dali server is [http://ekhidna.biocenter.helsinki.fi/dali\\_server](http://ekhidna.biocenter.helsinki.fi/dali_server).

All backbone atoms (N, CA, C, O) are required and the minimum chain length is 30 amino acids. Backbone atoms may be reconstructed from a CA trace using the MaxSprout server at

External links to the Dali database should use `1nnn`, where `1nnn` represents a PDB identifier and `chainid` is optional. Meta-servers may link to, which directly returns the match list and alignment data as plain text.

### Processing

Queries to the Dali Database and pairwise comparison are processed interactively; the result is usually returned within a minute. The Dali server processes up to eight PDB searches in parallel, others are queued. Most PDB-search queries are processed in less than an hour. Results are stored on the server for two weeks. The results of identical queries are retrieved instantly from cache.

The Dali server and Dali database return only the best match of the query to each PDB structure. The pairwise comparison returns also suboptimal matches. The pairwise comparison is based on a systematic branch-and-bound search that returns non-overlapping solutions in decreasing order of alignment score. Suboptimal matches can be of interest in cases of internal symmetries or repeated domains.

Dali Database is updated twice a year and contains precomputed structural alignments of PDB90 against the full PDB. The query structure is mapped to the closest representative in PDB90 and the structure comparison scores are recomputed using the transitive alignment via the representative.

The Dali server aims to retrieve a list of 500 structural neighbors of the query structure with the highest Z-scores. Most query structures have strong similarity to a structure already in the PDB. We use fast filters to identify a shortlist of about 100 promising candidates. If these produce strong matches, the search proceeds by walking. Otherwise, the query structure is compared with PDB90 in one versus all fashion, followed by a walk to collect matches to redundant PDB structures (which are over 90% sequence identical to PDB90 representatives).

Walking selects targets for structural comparison from the neighbours of neighbours found so far. The second shell of neighbors is known because all structures in the PDB are stored in a precomputed network of similarities. The pairwise alignments (Q,P) and (P,R) induce a transitive alignment (Q,R), which is used as the starting point of refinement rather than optimizing the alignment from scratch. There are many possible choices of intermediate structure P en route from Q to R. We select the ‘high road’, in other words, the minimum of the Z-scores  $Z(Q,P)$  and  $Z(P,R)$  should be as high as possible. The ‘high road’ may change as more structures are added to the first neighbour shell. To avoid redundant comparisons, we only test induced alignments which are longer than previously obtained ones. When the alignment (Q,R) has been refined, R is added to the first neighbour shell. The walk ends when either there are no new neighbours in the second shell, a specified number of hits (1000) have been reported, or a maximum number of comparisons (1000) have been performed.

## Outputs

The Dali server, Dali Database and pairwise comparison use a common output format and share interactive analysis tools.

The result consists of (i) a list of structural neighbours, ranked by Z-score, and (ii) the alignment data. The results are presented as plain text for downloading by downstream application, and as hypertext for interactive analysis. The default results page reports the top 500 matches to all chains in the PDB. A subset of matches to PDB90, filtered at 90% sequence identity, is provided for convenience.

Selected subsets of matches can be visualized (i) as multiple sequence alignments, or (ii) in multiple 3D superimposition. While sophisticated tools with integrated sequence alignment and structure superimposition views are available, we have chosen Jmol, an open source Java viewer for molecular graphics, because it was most easily accessible to the casual user. Each neighbour is aligned (superimposed) against the query structure in a star-like tree topology. Active sites can be recognized by clusters of conserved residues and ligands. Sequence and structure conservation are calculated within the selected subset of matches.

VAST and DALI are thus very useful structure similarity BLAST tool. VAST provides user with similar structures to their query along with its molecular components and chemicals and non-standard biopolymers, aligned sequences and 3D structure superimposition information which includes information regarding H-bonds, interactions, buried surface area, 2D interaction network and much more. DALI provides user with similar structures to their query along with its pairwise alignment, coordinates information, 3Dsuperimposition results. Describing the structural similarity relationships between protein structures allows scientists to infer the functions of newly discovered proteins.

## REFERENCES:

1. Dey, Fabian; Cliff Zhang, Qiangfeng; Petrey, Donald; Honig, Barry (2013). Toward a “Structural BLAST”: Using structural relationships to infer function. *Protein Science*, 22(4), 359– 366. doi:10.1002/pro.2225
2. Madej, T.; Lanczycki, C. J.; Zhang, D.; Thiessen, P. A.; Geer, R. C.; Marchler-Bauer, A.; Bryant, S. H. (2014). MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Research*, 42(D1), D297–D303. doi:10.1093/nar/gkt1208
3. Dali server. (n.d.). Ekhidna2.Biocenter.helsinki.fi. Retrieved March 14, 2022, from <http://ekhidna2.biocenter.helsinki.fi/dali/>

4. Holm, L.; Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Research*, 38(Web Server), W545–W549. doi:10.1093/nar/gkq366

**WEBLEM 7a****VAST**

(URL: <https://www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi>)

**AIM:**

To perform structure BLAST for tubulin using VAST tool.

**INTRODUCTION:**

Tubulin is the protein that polymerizes into long chains or filaments that form microtubules, hollow fibers which serve as a skeletal system for living cells. Microtubules have the ability to shift through various formations which is what enables a cell to undergo mitosis or to regulate intracellular transport.

The computational detection of similarities between protein 3D structures has become an indispensable tool for the detection of homologous relationships, the classification of protein families and functional inference. Consequently, numerous algorithms have been developed that facilitate structure comparison, including rapid searches against a steadily growing collection of protein structures. To this end, NCBI's Molecular Modeling Database (MMDB), which is based on the Protein Data Bank (PDB), maintains a comprehensive and up-to-date archive of protein structure similarities computed with the Vector Alignment Search Tool (VAST). These similarities have been recorded on the level of single proteins and protein domains, comprising in excess of 1.5 billion pairwise alignments. VAST+, an extension to the existing VAST service, which summarizes and presents structural similarity on the level of biological assemblies or macromolecular complexes. VAST+ simplifies structure neighboring results and shows, for macromolecular complexes tracked in MMDB, lists of similar complexes ranked by the extent of similarity. VAST+ replaces the previous VAST service as the default presentation of structure neighboring data in NCBI's Entrez query and retrieval system.

**METHODOLOGY:**

1. Open homepage for VAST. (URL: <https://www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi>)
2. Retrieve PDB ID for Albumin.
3. Search for similar structures on VAST for the PDB ID.
4. Observe and interpret the results.

## OBSERVATION:

**NCBI** National Center for Biotechnology Information

**VAST+ Similar Structures** 3D structural similarities among biological assemblies

**COVID-19 Information**

Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

**VAST+** is a tool designed to identify macromolecules that have similar 3-dimensional structures, with an emphasis on finding those with similar biological assemblies ("biological units" or "biounits"). The similarities are calculated using purely geometric criteria, and therefore can identify distant homologs that cannot be recognized by sequence comparison.

Input a valid PDB ID or MMDB ID:  PDB ID or MMDB ID  ?

To use VAST+, enter the PDB ID or MMDB ID of any structure that is currently in the Molecular Modeling Database (MMDB). VAST+ will display a list of similar structures, ranking them by the extent of their similarity to the query structure's biological unit. [more...](#)

**Citing VAST**

Gibrat JF, Madej T, Bryant SH. "Surprising similarities in structure comparison." *Curr Opin Struct Biol*. 1996 Jun;6(3): 377-85.  
 Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, Marchler-Bauer A, Bryant SH. "MMDB and VAST+: tracking structural similarities between macromolecular complexes." *Nucl. Acids Res.* 2014 Jan;42(Database issue):D297-303.  
 Thomas Madej, Aron Marchler-Bauer, Christopher Lanczycki, Dachuan Zhang, Stephen H Bryant "Biological Assembly Comparison With VAST" *Methods Mol. Biol.* 2020(2112):175-186

You are here: NCBI > Computational Biology Branch > Structure Group > VAST+ Write to the Help Desk

**GETTING STARTED**

- NCBI Education
- NCBI Help Manual
- NCBI Handbook
- Training & Tutorials

**RESOURCES**

- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

**POPULAR**

- PubMed
- Nucleotide
- BLAST
- PubMed Central
- Gene
- Bookshelf
- Protein
- PubChem
- OMIM
- Genome
- SNP
- Structure

**FEATURED**

- GenBank
- Reference Sequences
- Map Viewer
- Genome Projects
- Human Genome
- Mouse Genome
- Influenza Virus
- Primer-BLAST
- Sequence Read Archive

**NCBI INFORMATION**

- About NCBI
- Research at NCBI
- NCBI Newsletter
- NCBI FTP Site
- NCBI on Facebook
- NCBI on Twitter
- NCBI on YouTube

Copyright | Disclaimer | Privacy | Accessibility | Contact  
 National Center for Biotechnology Information, U.S. National Library of Medicine  
 3500 Rockville Pike, Bethesda MD, 20894 USA



Fig1. Homepage for VAST

RCSB PDB Deposit Search Visualize Analyze Download Learn More Documentation Careers MyPDB

**RCSB PDB** PROTEIN DATA BANK 188431 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

PDB-101 Worldwide Protein Data Bank EMD Data Resource Worldwide Protein Data Bank Foundation

Advanced Search | Browse Annotations Help

Structure Summary 3D View Annotations Experiment Sequence Genome Ligands Versions

**1Z2B**

Biological Assembly 1

Tubulin-colchicine-vinblastine: stathmin-like domain complex

PDB DOI: 10.2210/pdb1Z2B/pdb

Classification: CELL CYCLE  
 Organism(s): Bos taurus, Rattus norvegicus  
 Expression System: Escherichia coli BL21(DE3)  
 Mutation(s): No

Deposited: 2005-03-08 Released: 2005-05-31  
 Deposition Author(s): Gigant, B., Wang, C., Ravelli, R.B.G., Roussi, F., Steinmetz, M.O., Curmi, P.A., Sobel, A., Knissow, M.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION  
 Resolution: 4.10 Å  
 R-Value Free: 0.269  
 R-Value Work: 0.209  
 R-Value Observed: 0.212

wwPDB Validation

3D Report Full Report

Metric Percentile Ranks Value

3D View: Structure | 1D-3D View | Electron Density | Validation Report | Ligand Interaction

Pseudo Symmetry: Asymmetric - C1  
 Pseudo Stoichiometry: Hetero 5-mer - A4B1

Fig2. Tubulin PDB structure

**NCBI** National Center for Biotechnology Information

**VAST+ Similar Structures** 3D structural similarities among biological assemblies

**COVID-19 Information**

[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

**VAST+** is a tool designed to identify macromolecules that have similar 3-dimensional structures, with an emphasis on finding those with similar biological assemblies ("biological units" or "biounits"). The similarities are calculated using purely geometric criteria, and therefore can identify distant homologs that cannot be recognized by sequence comparison.

**Input a valid PDB ID or MMDB ID:**  **Search** 

To use VAST+, enter the PDB ID or MMDB ID of any structure that is currently in the Molecular Modeling Database (MMDB). VAST+ will display a list of similar structures, ranking them by the extent of their similarity to the query structure's biological unit. [more...](#)

**Citing VAST**

Gibrat JF, Madej T, Bryant SH. "Surprising similarities in structure comparison." *Curr Opin Struct Biol*. 1996 Jun;6(3): 377-85.  
 Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, Marchler-Bauer A, Bryant SH. "MMDB and VAST+: tracking structural similarities between macromolecular complexes." *Nucl. Acids Res.* 2014 Jan;42(Database issue):D297-303.  
 Thomas Madej, Aron Marchler-Bauer, Christopher Lanczycki, Dachuan Zhang, Stephen H Bryant. "Biological Assembly Comparison With VAST" *Methods Mol. Biol.* 2020(2112):175-186

You are here: NCBI > Computational Biology Branch > Structure Group > VAST+

Write to the Help Desk

**GETTING STARTED**

- NCBI Education
- NCBI Help Manual
- NCBI Handbook
- Training & Tutorials

**RESOURCES**

- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

**POPULAR**

- PubMed
- Nucleotide
- BLAST
- PubMed Central
- Gene
- Bookshelf
- Protein
- PubChem
- OMIM
- Genome
- SNP
- Structure

**FEATURED**

- GenBank
- Reference Sequences
- Map Viewer
- Genome Projects
- Human Genome
- Mouse Genome
- Influenza Virus
- Primer-BLAST
- Sequence Read Archive

**NCBI INFORMATION**

- About NCBI
- Research at NCBI
- NCBI Newsletter
- NCBI FTP Site
- NCBI on Facebook
- NCBI on Twitter
- NCBI on YouTube

Copyright | Disclaimer | Privacy | Accessibility | Contact  
 National Center for Biotechnology Information, U.S. National Library of Medicine  
 6600 Rockville Pike, Bethesda MD, 20894 USA

**Fig3. Search for tubulin PDB structure**

**NCBI** National Center for Biotechnology Information

**VAST+ Similar Structures** 3D structural similarities among biological assemblies

**1Z2B : Tubulin-Colchicine-Vinblastine: Stathmin-Like Domain Complex**

PDB ID or MMDB ID:  **New Search** 

Biological unit 1: pentameric  
 Source organism: *Rattus norvegicus*, ▼  
 Number of proteins: 5 (RB3 STATHMIN-LIKE DOMAIN 4, TUBULIN ALPHA CHAIN... ▼)  
 Number of chemicals: 9 (Magnesium Ion (2),2-Mercapto-N-[1,2,3,10-Tetram... ▼)

**Similar Structures (3038)**  

**All matching molecules superposed** **Invariant substructure superposed** 

**▲ Hide filters** 

**Filter by number of matching molecules** 

- Complete match, 5 proteins (114)** 
- Partial match, 4 proteins (44)** 
- Partial match, 3 proteins (2)** 
- Partial match, 2 proteins (95)** 
- Partial match, 1 protein (2783)** 

**Filter by taxonomy** 

- Eukaryota (1154)** 
- Bacteria (1686)** 
- Archaea (137)** 
- Viruses (10)** 
- Others (51)** 

**Apply Filter Selection**

Showing 1 to 10 out of 3038 selected structures 

Search within results:  **Go** **Reset** 

PDB ID	Description	Taxonomy	Aligned Protein	RMSD	Aligned Residues	Sequence Identity
1   3N2G	Tubulin-Nsc 613863: Rb3 Stathmin-Like Domain Complex	Ovis aries/Rattus norvegicus	5	0.89 Å	1814	99%

**Fig4. Hit page for Tubulin**

**Filter by number of matching molecules**

Complete match, 5 proteins (114) [x]

Partial match, 4 proteins (44) [x]

Partial match, 3 proteins (2) [x]

Partial match, 2 proteins (95) [x]

Partial match, 1 protein (2783) [x]

**Filter by taxonomy**

Eukarya (1154) [x]

Bacteria (1686) [x]

Archaea (137) [x]

Viruses (10) [x]

Others (51) [x]

Showing 1 to 10 out of 112 selected structures [x]
Search within results: PDB ID or search word

PDB ID	Description	Taxonomy	Aligned Protein	RMSD	Aligned Residues	Sequence Identity
1 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 3N2G	Tubulin-Nsc 613863: Rb3 Stathmin-Like Domain Complex	Ovis aries/Rattus norvegicus	5	0.89 Å	1814	100%
2 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 3HCK	Tubulin-Abt751: Rb3 Stathmin-Like Domain Complex	Ovis aries/Rattus norvegicus	5	0.92 Å	1814	100%
3 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 3N2K	Tubulin-Nsc 613862: Rb3 Stathmin-Like Domain Complex	Ovis aries/Rattus norvegicus	5	0.77 Å	1813	100%
4 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 3HKE	Tubulin-T138067: Rb3 Stathmin-Like Domain Complex	Ovis aries/Rattus norvegicus	5	0.77 Å	1813	100%
5 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 3HKD	Tubulin-Tn16 : Rb3 Stathmin-Like Domain Complex	Ovis aries/Rattus norvegicus	5	0.81 Å	1812	100%
6 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 3E22	Tubulin-Colchicine-Soblidotin: Stathmin-Like Domain Complex	Bos taurus/Rattus norvegicus	5	0.79 Å	1810	99%
7 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 3HKB	Tubulin: Rb3 Stathmin-Like Domain Complex	Ovis aries/Rattus norvegicus	5	0.92 Å	1808	100%
8 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 5KX5	Crystal Structure Of Tubulin-stathmin-ttl-compound 11 Complex	Gallus gallus/Ovis aries/Rattus norvegicus	5	1.66 Å	1808	100%
9 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 1SA0	Tubulin-Colchicine: Stathmin-Like Domain Complex	Bos taurus/Rattus norvegicus	5	0.73 Å	1807	99%
10 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 1SA1	Tubulin-Podophyllotoxin: Stathmin-Like Domain Complex	Bos taurus/Rattus	5	0.91 Å	1806	99%

**Fig5. Hit page after applying filter**

**Showing 1 to 10 out of 112 selected structures [x]**

**Search within results:** PDB ID or search word

Showing 1 to 10 out of 112 selected structures [x]
Search within results: PDB ID or search word

PDB ID	Description	Taxonomy	Aligned Protein	RMSD	Aligned Residues	Sequence Identity
1 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 3N2G	Tubulin-Nsc 613863: Rb3 Stathmin-Like Domain Complex	Ovis aries/Rattus norvegicus	5	0.89 Å	1814	100%

**Aligned Molecules** [x]

**Query structure 1Z2B**

\*Select schematic circles ● or highlighted molecule names to view matches

**Matched structure 3N2G**

2 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 3HCK	Tubulin-Abt751: Rb3 Stathmin-Like Domain Complex	Ovis aries/Rattus norvegicus	5	0.92 Å	1814	100%
3 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 3N2K	Tubulin-Nsc 613862: Rb3 Stathmin-Like Domain Complex	Ovis aries/Rattus norvegicus	5	0.77 Å	1813	100%
4 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 3HKE	Tubulin-T138067: Rb3 Stathmin-Like Domain Complex	Ovis aries/Rattus norvegicus	5	0.77 Å	1813	100%
5 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 3HKD	Tubulin-Tn16 : Rb3 Stathmin-Like Domain Complex	Ovis aries/Rattus norvegicus	5	0.81 Å	1812	100%
6 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 3E22	Tubulin-Colchicine-Soblidotin: Stathmin-Like Domain Complex	Bos taurus/Rattus norvegicus	5	0.79 Å	1810	99%
7 <span style="border: 1px solid #ccc; padding: 2px 5px;">+/-</span> <span style="color: red;">●</span> 3HKB	Tubulin: Rb3 Stathmin-Like Domain Complex	Ovis aries/Rattus	5	0.92 Å	1808	100%

**Fig6. Result for aligned molecules**

## 3N2G: Tubulin-Nsc 613863: Rb3 Stathmin-Like Domain Complex

Citation: [\[2\]](#)

**Stathmin and interfacial microtubule inhibitors recognize a naturally curved conformation of tubulin dimers**

Barbier P, Dorl  ans A, Devred F, Sanz L, Allegro D, Alfonso C, Knossow M, Peyrot V, Andreu JM

*J Biol Chem* (2010) **285** p.31672-81

### Abstract

Tubulin is able to switch between a straight microtubule-like structure and a curved structure in complex with the stathmin-like domain of the RB3 protein (T(2)RB3). GTP hydrolysis following microtubule assembly induces protofilament curvature and disassembly. The conformation of the labile tubulin heterodimers is unknown. One important question is whether free GDP-tubulin dimers... [read more](#)

PDB ID:

3N2G [Download](#) [\[?\]](#)

MMDB ID:

83668 [\[?\]](#)

PDB Deposition Date: 2010/5/18 [\[?\]](#)

Updated in MMDB: 2012/11 [\[?\]](#)

Experimental Method: x-ray diffraction [\[?\]](#)

Resolution: 4    [\[?\]](#)

Source Organism: *Rattus norvegicus* [\[?\]](#)

Similar Structures:

[VAST+](#) [\[?\]](#)

[Download sequence data](#) [\[?\]](#)

[Biological Unit](#)

[Asymmetric Unit](#) [\[?\]](#)

Biological Unit for 3N2G: pentameric; determined by author and by software (PISA) [\[?\]](#)

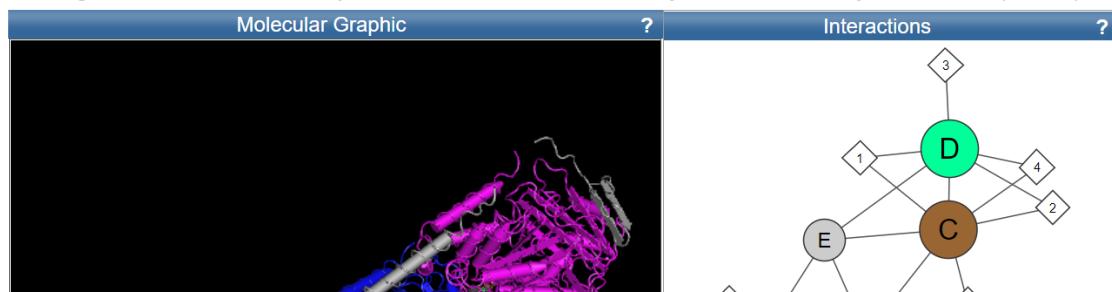


Fig7. Result page for structure (3N2G) similar to Tubulin

### Molecular Components in 3N2G [\[?\]](#)

Label	Count	Molecule
<b>Proteins (5 molecules)</b>		
<b>A</b> <b>C</b>	2	<b>Tubulin Alpha Chain</b> 
<b>B</b> <b>D</b>	2	<b>Tubulin Beta Chain</b> 
<b>E</b>	1	<b>Stathmin-4</b> 
<b>Chemicals and Non-standard biopolymers (9 molecules)</b>		
<b>1</b>	2	<b>Guanosine-5'-Triphosphate</b> 
<b>2</b>	3	<b>Magnesium Ion</b> 

Fig8. Molecular components and chemical and non-standard biopolymers for 1N5U

Aligned Sequences [?](#)

[Close](#)

1Z2B\_A: TUBULIN ALPHA CHAIN  
3N2G\_A: TUBULIN ALPHA CHAIN

[Visualize 3D structure superposition](#) [?](#)

```

1Z2B_A 2 RECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPsdktigggdDSFNTFFSETGAGKH 61
3N2G_A 2 RECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPsdktigggdDSFNTFFSETGAGKH 61

1Z2B_A 62 VPRAVFVDELEPTVIDEVRTGTYRQLFHPQLITGKEDAANNYARGHYTIGKEIIDLVLDR 121
3N2G_A 62 VPRAVFVDELEPTVIDEVRTGTYRQLFHPQLITGKEDAANNYARGHYTIGKEIIDLVLDR 121

1Z2B_A 122 IRKLADQCTGLQGFLVFHSFGGGTGSGFTSLLMERLSVDYGKKSLEFSIYPAPQVSTAV 181
3N2G_A 122 IRKLADQCTGLQGFLVFHSFGGGTGSGFTSLLMERLSVDYGKKSLEFSIYPAPQVSTAV 181

1Z2B_A 182 VEPYNSILTTHTTLEHSDCAFMDNEAIYDICRRLNDIERPTYTNLNRLLIQIVSSITAS 241
3N2G_A 182 VEPYNSILTTHTTLEHSDCAFMDNEAIYDICRRLNDIERPTYTNLNRLLIQIVSSITAS 241

1Z2B_A 242 LRFDGALNVDLTFQTNLVPYPRHPLATYAPVISAEKAYHEQLSVAEITNACFEPANQ 301
3N2G_A 242 LRFDGALNVDLTFQTNLVPYPRHPLATYAPVISAEKAYHEQLSVAEITNACFEPANQ 301

1Z2B_A 302 MVKCDPRHGKYMACCLLYRGDVPKDVNAAIATIKTKRSIQFVDCPTGFKVGINYQPPT 361
3N2G_A 302 MVKCDPRHGKYMACCLLYRGDVPKDVNAAIATIKTKRTIQFVDCPTGFKVGINYQPPT 361

1Z2B_A 362 VVPGGDLAKVQRAVCMLSNTTAAIAEAWARLDHKFDLMYAKRAFVHVVGEGMEEGFSEA 421
3N2G_A 362 VVPGGDLAKVQRAVCMLSNTTAAIAEAWARLDHKFDLMYAKRAFVHVVGEGMEEGFSEA 421

1Z2B_A 422 REDMAALEKDYEVG 437
3N2G_A 422 REDMAALEKDYEVG 437

```

[Visualize 3D structure superposition](#) [?](#)

1Z2B\_B: TUBULIN BETA CHAIN  
3N2G\_B: TUBULIN BETA CHAIN

[Visualize 3D structure superposition](#) [?](#)

1Z2B\_B 2 REIVHIQAGQCGNQIGAKFWEVISDEHGDPTGSYHGDSDLQLERINVYYNEATGNKVP 61

Fig9. Result page for aligned sequences

File Select View Style Color Analysis Help [Toolbar](#) [-](#) [one-letter seq](#) [Search](#) [?](#)

Alternate (Key "a") Save iCn3D PNG Image H-Bonds & Interactions View Selection Toggle Highlight Remove Labels

Structure Alignment of [1Z2B](#) and [3N2G](#) from VAST+

[Select residues in aligned sequences](#)

+ Selection: Name:  Save Clear

		10			20			30			
1Z2B_A	TUBULIN ALPHA CHAIN	*	*	*	*	*	*	*	*	*	*
3N2G_A	TUBULIN ALPHA CHAIN	*	*	*	*	*	*	*	*	*	*
1Z2B_A	2 RECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPs										
3N2G_A	2 RECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPs										
1Z2B_B	TUBULIN BETA CHAIN	*	*	*	*	*	*	*	*	*	*
3N2G_B	TUBULIN BETA CHAIN	*	*	*	*	*	*	*	*	*	
1Z2B_B	4 REIVHIQAGQCGNQIGAKFWEVISDEHGDPTGSYHGD										
3N2G_B	4 REIVHIQAGQCGNQIGAKFWEVISDEHGDPTGSYHGD										
1Z2B_C	TUBULIN ALPHA CHAIN	*	*	*	*	*	*	*	*	*	*
3N2G_C	TUBULIN ALPHA CHAIN	*	*	*	*	*	*	*	*	*	
1Z2B_C	2 RECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPs										
3N2G_C	2 RECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPs										
1Z2B_D	TUBULIN BETA CHAIN	*	*	*	*	*	*	*	*	*	*
3N2G_D	TUBULIN BETA CHAIN	*	*	*	*	*	*	*	*	*	
1Z2B_D	2 RECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPs										
3N2G_D	2 RECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPs										

> load alignment 52185,1,83668,1 | parameters &showalignseq=1&alignh=52185,1,83668,1&atype=0  
>

Fig10. Result page for 3D structure superimposition

File Select View Style Color Analysis Help Toolbar - one-letter seq Search ? All atoms

Alternate (Key "a") Save iCn3D PNG Image H-Bonds & Interactions View Selection Toggle Highlight Remove Labels Select residues in aligned sequences

Structure Alignment of 1Z2B and 3N2G from VAST+

**Hydrogen bonds/interactions between two sets of atoms**

1. Choose interaction types and their thresholds:

Hydrogen Bonds 3.8 Å  Salt Bridge/Ionic 6 Å  Contacts/Interact

Halogen Bonds 3.8 Å  π-Cation 6 Å  π-Stacking

2. Select the first set: selected 1Z2B 1Z2B\_A 1Z2B\_B 1Z2B\_C

3. Select the second set: non-selected 1Z2B 1Z2B\_A 1Z2B\_B

4. Cross Structure Interactions: No

3D Display Interactions

Highlight Interactions in Table Sort Interactions on: Set 1 Set 2

2D Interaction Network to show interactions between two lines of residue nodes

2D Interaction Map to show interactions as map

2D Graph(Force-Directed) to show interactions with strength parameters in 0-200:

Helix or Sheet: 100 Coil or Nucleotide: 50 Disulfide Bond: 50

Hydrogen Bond: 50 Salt Bridge/Ionic: 50 Contacts: 25

Halogen Bonds: 50 π-Cation: 50 π-Stacking: 50

(Note: you can also adjust thresholds at #1 to add/remove interactions.)

Buried solvent accessible surface area in the interface

Calculate solvent accessible surface area in the interface:

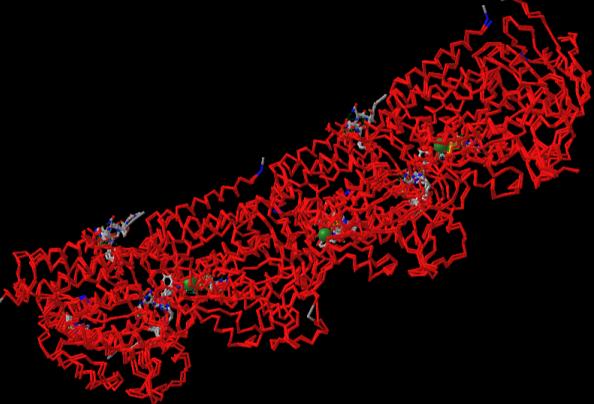
Set 1: 1Z2B, Surface: 56097.10 Å<sup>2</sup>  
 Set 2: non-selected, Surface: 10143.75 Å<sup>2</sup>  
 Total Surface: 56170.29 Å<sup>2</sup>  
 Buried Surface for Set 1: 526.39 Å<sup>2</sup>  
 Buried Surface for Set 2: 8433.20 Å<sup>2</sup>

Fig11. Buried surface information

File Select View Style Color Analysis Help Toolbar - one-letter seq Search ? Selection

Alternate (Key "a") Save iCn3D PNG Image H-Bonds & Interactions View Selection Toggle Highlight Remove Labels Select residues in aligned sequences

Structure Alignment of



+ Selection: Name: alseq\_1 Save Clear

1Z2B\_A 3N2G\_A TUBULIN ALPHA CHAIN TUBULIN ALPHA CHAIN 10 20 30  
 1Z2B\_A 3N2G\_A 2 REC I S I H V G Q A G V Q I G N A C W E L Y C L E H G I Q P D G Q M P S C  
 1Z2B\_A 3N2G\_A 2 REC I S I H V G Q A G V Q I G N A C W E L Y C L E H G I Q P D G Q M P S C

1Z2B\_B 3N2G\_B TUBULIN BETA CHAIN TUBULIN BETA CHAIN 10 20 30  
 1Z2B\_B 3N2G\_B 4 R E I V H I Q A G Q C G N Q I G A K F W E V I S D E H G I D P T G S Y H G D  
 1Z2B\_B 3N2G\_B 4 R E I V H I Q A G Q C G N Q I G A K F W E V I S D E H G I D P T G S Y H G D

1Z2B\_C 3N2G\_C TUBULIN ALPHA CHAIN TUBULIN ALPHA CHAIN 10 20 30  
 1Z2B\_C 3N2G\_C 2 R E C I S I H V G Q A G V Q I G N A C W E L Y C L E H G I Q P D G Q M P S C  
 1Z2B\_C 3N2G\_C 2 R E C I S I H V G Q A G V Q I G N A C W E L Y C L E H G I Q P D G Q M P S C

1Z2B\_D 3N2G\_D TUBULIN BETA CHAIN TUBULIN BETA CHAIN 10 20 30  
 1Z2B\_D 3N2G\_D 2 R E C I S I H V G Q A G V Q I G N A C W E L Y C L E H G I Q P D G Q M P S C

> display interaction 3d | 1Z2B non-selected | hbonds,salt bridge,interactions,halogen,pi-cation,pi-stacking | false | threshold 3.8 6 4 3.8 6 5.5

Fig12. Result for itneractions

File Select View Style Color Analysis Help Toolbar | one-letter seq. Search ?

Alternate (Key "a") Save iCn3D PNG Image H-Bonds & Interactions View Selection of Toggle Highlight Remove Labels Selection

Select residues in aligned sequences

1. Choose interaction types and their thresholds:

<input checked="" type="checkbox"/> Hydrogen Bonds	3.8	A	<input checked="" type="checkbox"/> Salt Bridge/Ionic	6	A	<input checked="" type="checkbox"/> Contacts/Interactions
<input checked="" type="checkbox"/> Halogen Bonds	3.8	A	<input checked="" type="checkbox"/> π-Cation	6	A	<input checked="" type="checkbox"/> π-Stacking

2. Select the first set: selected 1Z2B, 1Z2B\_A, 1Z2B\_B, 1Z2B\_C

3. Select the second set: non-selected selected 1Z2B, 1Z2B\_A, 1Z2B\_B

4. Cross Structure Interactions: No

3D Display Interactions

Highlight Interactions in Table Sort Interactions on: Set 1 Set 2

2D Interaction Network to show interactions between two lines of residue nodes

2D Interaction Map to show interactions as map

2D Graph(Force-Directed) to show interactions with strength parameters in 0-200:

Helix or Sheet: 100 Coll or Nucleotid: 50 Disulfide Bond: 50

Hydrogen Bond: 50 Salt Bridge/Ionic: 50 Contacts: 25

Halogen Bonds: 50 π-Cation: 50 π-Stacking: 50

(Note: you can also adjust thresholds at #1 to add/remove interactions.)

Buried Surface Area

5. **Reset** and select new sets

hbonds, salt bridge, interactions, halogen, π-cation, π-stacking between Two Sets:

Set 1: selected **Highlight in 3D**  
Set 2: non-selected **Highlight in 3D**

The interfaces are:

interface\_1 **Highlight in 3D**  
interface\_2 **Highlight in 3D**

Note: Each checkbox below selects the corresponding residue. You can click "Save Selection" in the "Select" menu to save the selection and click on "Highlight" button to clear the checkboxes.

3872 hydrogen bond pairs:

Atom 1	Atom 2	Distance(Å)	Highlight in 3D
ARG \$1Z2B.A:2@NE	GLU \$1Z2B.A:3@OE2	3.8	<b>Highlight</b>
GLU \$1Z2B.A:3@O	GLN \$1Z2B.A:133@N	2.9	<b>Highlight</b>
GLU \$1Z2B.A:3@OE2	ARG \$1Z2B.A:2@NE	3.8	<b>Highlight</b>
GLU \$1Z2B.A:3@OE2	THR \$1Z2B.A:130@OG1	3.7	<b>Highlight</b>
CYS \$1Z2B.A:1@N	THR \$1Z2B.A:51@O	3.2	<b>Highlight</b>

**Fig13. Result for interactions**

File Select View Style Color Analysis Help Toolbar | one-letter seq. Search ?

Alternate (Key "a") Save iCn3D PNG Image H-Bonds & Interactions View Selection Toggle Highlight Remove Labels Selection

Structure Alignment of 4L8U and 1N5U from VAST+

Show interactions between two lines of residue nodes

Hold Ctrl key to select multiple nodes/lines.  
Green: H-Bonds; Cyan: Salt Bridge/Ionic; Grey: contacts  
Magenta: Halogen Bonds; Red: π-Cation; Blue: π-Stacking

SVG PNG JSON Scale: 1

> line graph interaction pairs | 1NSU selected | hbonds,salt bridge,interactions,halogen,pi-cation,pi-stacking | true | threshold 3.8 6 4 3.8 6 5.5

**Fig14. Result for 2D interaction network**

## RESULT:

PDB ID for tubulin structure was searched in structure similarity BLAST tool, VAST and 152 similar structures were retrieved.

## CONCLUSION:

VAST is a useful structure similarity BLAST tool which provides user with similar structures to their query along with its molecular components and chemicals and non-standard biopolymers, aligned sequences and 3D structure superimposition information which includes information regarding H-bonds, interactions, buried surface area, 2D interaction network and much more. Describing the structural similarity relationships between protein structures allows scientists to infer the functions of newly discovered proteins.

## REFERENCES:

1. Madej, T.; Lanczycki, C. J.; Zhang, D.; Thiessen, P. A.; Geer, R. C.; Marchler-Bauer, A.; Bryant, S. H. (2014). MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Research*, 42(D1), D297–D303. doi:10.1093/nar/gkt1208
2. Tubulin: Mystery of vital cell protein solved after 30 years. (n.d.). Retrieved March 19, 2022, from <https://www2.lbl.gov/Science-Articles/Archive/3D-tubulin.html#:~:text=Tubulin%20is%20the%20protein%20that,or%20to%20regulate%20intracellular%20transport>.
3. Similar Protein Structure Assemblies. (2014). Nih.gov. Retrieved March 19, 2022, from <https://www.ncbi.nlm.nih.gov/Structure/vastplus/vastplus.cgi>
4. Bank, R. P. D. (n.d.-b). RCSB PDB - 4L8U: X-ray study of human serum albumin complexed with 9 amino camptothecin. [Www.rcsb.org](https://www.rcsb.org/structure/4L8U). Retrieved March 19, 2022, from <https://www.rcsb.org/structure/4L8U>
5. Bank, R. C. S. B. P. D. (n.d.). *1Z2B: Tubulin-colchicine-vinblastine: Stathmin-like domain complex*. RCSB PDB. Retrieved March 19, 2022, from <https://www.rcsb.org/structure/1Z2B>
6. U.S. National Library of Medicine. (n.d.). *3N2G: Tubulin-NSC 613863: RB3 stathmin-like domain complex*. National Center for Biotechnology Information. Retrieved March 19, 2022, from <https://www.ncbi.nlm.nih.gov/Structure/pdb/3N2G>
7. iCn3D: Web-based 3D Structure Viewer. (n.d.). [Www.ncbi.nlm.nih.gov](https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html?showalignseq=1&align=112122). Retrieved March 19, 2022, from <https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html?showalignseq=1&align=112122>

## WEBLEM 7b

### DALI

(URL: <http://ekhidna2.biocenter.helsinki.fi/dali/>)

#### AIM:

To perform structural blast for tubulin using DALI tool

#### Introduction:

Tubulin is the protein that polymerizes into long chains or filaments that form microtubules, hollow fibers which serve as a skeletal system for living cells. Microtubules have the ability to shift through various formations which is what enables a cell to undergo mitosis or to regulate intracellular transport.

The DALI server is a network service for comparing protein structures in 3D. You submit the coordinates of a query protein structure and DALI compares them against those in the Protein Data Bank (PDB). In favourable cases, comparing 3D structures may reveal biologically interesting similarities that are not detectable by comparing sequences.

User can perform three types of database searches:

- **Heuristic PDB search** - compares one query structure against those in the Protein Data Bank
- **Exhaustive PDB25 search** - compares one query structure against a representative subset of the Protein Data Bank
- **Hierarchical AF-DB search** - compares one query structure against a species subset of the AlphaFold Database

#### METHODOLOGY:

1. Open homepage for DALI. (URL: <http://ekhidna2.biocenter.helsinki.fi/dali/>)
2. Enter Albumin PDB ID.
3. Observe similar structures matches against PDB25, PDB50, PDB90 and all PDB structures.
4. Interpret the results.

# DALI

## PROTEIN STRUCTURE COMPARISON SERVER

About PDB search PDB25 AF-DB search Pairwise All against all Gallery References Statistics Download

The DALI server is a network service for comparing protein structures in 3D. You submit the coordinates of a query protein structure and DALI compares them against those in the Protein Data Bank (PDB). In favourable cases, comparing 3D structures may reveal biologically interesting similarities that are not detectable by comparing sequences.

Check queue status [here](#). Megauers please consider downloading the standalone program.

You can perform three types of database searches:

- Heuristic **PDB search** - compares one query structure against those in the Protein Data Bank
- Exhaustive **PDB25** search - compares one query structure against a representative subset of the Protein Data Bank
- Hierarchical **AF-DB** search - compares one query structure against a species subset of the AlphaFold Database

and two types of structure comparisons of user selected structures:

- **Pairwise** structure comparison - compares one query structure against those specified by the user
- **All against all** structure comparison - returns a structural similarity dendrogram for a set of structures specified by the user

Citation:

1. Holm L (2020) Using DALI for protein structure comparison. Methods Mol. Biol.

**Fig1. Homepage for DALI**

The screenshot shows the RCSB PDB website. At the top, there is a navigation bar with links for RCSB PDB, Deposit, Search, Visualize, Analyze, Download, Learn, More, Documentation, and Careers. A "MyPDB" button is also present. The main content area features the PDB logo and a search bar with a "PDB Archive" dropdown and a magnifying glass icon. Below the search bar, there are links for "Advanced Search" and "Browse Annotations", along with a world map and social media links (Facebook, Twitter, YouTube). The main content area displays the structure of PDB ID 1Z2B, which is a "Biological Assembly 1". The structure is shown as a 3D ribbon model with various colored domains. To the right of the structure, the ID "1Z2B" is displayed, along with the title "Tubulin-colchicine-vinblastine: stathmin-like domain complex". Below the title, the PDB DOI is listed as 10.2210/pdb1Z2B/pdb. The classification is "CELL CYCLE", and the organism is "Bos taurus, Rattus norvegicus". The expression system is "Escherichia coli BL21(DE3)" and there are no mutations. The deposition date is 2005-03-08 and the release date is 2005-05-31. The deposition author is Gigant, B., Wang, C., Ravelli, R.B.G., Roussi, F., Steinmetz, M.O., Curni, P.A., Sobel, A., Knossow, M. The experimental data snapshot shows the method as X-RAY DIFFRACTION, resolution as 4.10 Å, and R-value Free as 0.269. The wwPDB Validation table provides percentile ranks and values for various metrics:

Metric	Percentile Ranks	Value
Rfree	40	0.276
Clashscore	9.8%	40
Ramachandran outliers	36.1%	9.8%
Sidechain outliers	1.7%	36.1%
RSRZ outliers	Worse	Better

**Fig2. Tubulin PDB structure**

# DALI

## PROTEIN STRUCTURE COMPARISON SERVER

About PDB search PDB25 AF-DB search Pairwise All against all Gallery References Statistics Download

### PDB search

Compare query structure against Protein Data Bank.

STEP 1 - Enter your query protein structure

Structures may be specified by concatenating the PDB identifier (4 characters) and a chain identifier (1 character) or, alternatively, you may upload a PDB file.

OR upload file  No file chosen

1z2bA TUBULIN ALPHA CHAIN  
1z2bB TUBULIN ALPHA CHAIN  
1z2bC TUBULIN ALPHA CHAIN  
1z2bD TUBULIN ALPHA CHAIN  
1z2bE TUBULIN ALPHA CHAIN

Job title

E-mail

STEP 3 - Submit your job

If the same structure has been submitted recently, you will be redirected to the result page of the previous instance.

**Fig3. PDB search for tubulin**

## Results: lmao

### Chain: 1z2bA

- [Matches against PDB25](#) [Correlation matrix](#)
- [Matches against PDB50](#)
- [Matches against PDB90](#)
- [Matches against full PDB](#)
- [Download matches against PDB25](#)
- [Download matches against PDB50](#)
- [Download matches against PDB90](#)
- [Download matches against full PDB](#)

Results will be deleted after one week.

**Fig4. Result page of Tubulin**

## Matches against PDB25:

## Results: lmao

## Query: 1z2bA

## MOLECULE: TUBULIN ALPHA CHAIN;

Select neighbours (check boxes) for viewing as multiple structural alignment or 3D superimposition. The list of neighbours is sorted by Z-score. Similarities with a Z-score lower than 2 are spurious. Each neighbour has links to pairwise structural alignment with the query structure, and to the PDB format coordinate file where the neighbour is superimposed onto the query structure.

Structural Alignment  Expand gaps 3D Superimposition (PV) SANS PANZ Pfam Reset Selection

## Summary

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
□ 1:	7m2w-A	47.6	1.8	408	453	30	<a href="#">PDB</a>	MOLECULE: TUBULIN GAMMA CHAIN;
□ 2:	3zid-A	32.4	2.7	317	360	16	<a href="#">PDB</a>	MOLECULE: TUBULIN/FTSZ, GTPASE;
□ 3:	1w59-A	24.7	3.1	291	350	12	<a href="#">PDB</a>	MOLECULE: CELL DIVISION PROTEIN FTSZ HOMOLOG 1;
□ 4:	4xcq-A	23.9	3.5	280	303	11	<a href="#">PDB</a>	MOLECULE: TUBZ;
□ 5:	4e17-A	23.6	4.0	313	382	11	<a href="#">PDB</a>	MOLECULE: PLASMID REPLICATION PROTEIN REPX;
□ 6:	2xka-F	23.2	3.8	313	414	12	<a href="#">PDB</a>	MOLECULE: FTSZ/TUBULIN-RELATED PROTEIN;
□ 7:	3zbq-A	19.8	3.5	263	315	13	<a href="#">PDB</a>	MOLECULE: PHIKZ039;
□ 8:	4fkz-A	7.8	3.9	174	384	10	<a href="#">PDB</a>	MOLECULE: UDP-N-ACETYLGLUCOSAMINE 2-EPIMERASE;
□ 9:	4zht-A	7.3	3.7	167	384	9	<a href="#">PDB</a>	MOLECULE: BIFUNCTIONAL UDP-N-ACETYLGLUCOSAMINE 2-EPIMERASE/

### Fig5. Result for similar structures

## Pairwise Structural Alignments

Notation: three-state secondary structure definitions by DSSP (reduced to H=helix, E=sheet, L=coil) are shown above the amino acid sequence. Structurally equivalent residues are in uppercase, structurally non-equivalent residues (e.g. in loops) are in lowercase. Amino acid identities are marked by vertical bars.

**No 1: Query=1z2bA Sbjct=7m2wA Z-score=47.6**

[back to top](#)

**Fig6. Results for pairwise structural alignment**

```

REMARK Coordinates of 7m2w rotated and translated as follows:
REMARK | 0.40520 0.53216 -0.74338 | | x | | 173.000 |
REMARK | -0.07305 -0.79168 -0.60655 | * | y | + | 510.000 |
REMARK | -0.91131 0.30008 -0.28191 | | z | | 313.000 |
REMARK
REMARK HOH and TIP residues excluded. Only first MODEL passed through.
REMARK
HEADER CELL CYCLE 17-MAR-21 7M2W
TITLE ENGINEERED DISULFIDE CROSS-LINKED CLOSED CONFORMATION OF THE YEAST
TITLE 2 GAMMA-TURC(SS)
COMPND MOL_ID: 1;
COMPND 2 MOLECULE: TUBULIN GAMMA CHAIN;
COMPND 3 CHAIN: B, A, C, D;
COMPND 4 SYNONYM: GAMMA-TUBULIN;
COMPND 5 ENGINEERED: YES;
COMPND 6 MUTATION: YES;
COMPND 7 MOL_ID: 2;
COMPND 8 MOLECULE: SPINDLE POLE BODY COMPONENT SPC97;
COMPND 9 CHAIN: E, G;
COMPND 10 ENGINEERED: YES;
COMPND 11 MOL_ID: 3;
COMPND 12 MOLECULE: SPINDLE POLE BODY COMPONENT SPC98;
COMPND 13 CHAIN: F, H;
COMPND 14 ENGINEERED: YES;
COMPND 15 MOL_ID: 4;
COMPND 16 MOLECULE: SPINDLE POLE BODY COMPONENT 110;
COMPND 17 CHAIN: U, K, X, Y;
COMPND 18 SYNONYM: EXTRAGENIC SUPPRESSOR OF CMD1-1 MUTANT PROTEIN 1, NUCLEAR
COMPND 19 FILAMENT-RELATED PROTEIN 1, SPINDLE POLE BODY SPACER PROTEIN SPC110;
COMPND 20 ENGINEERED: YES
AUTHOR A.F.BRILOT,A.S.LYON,A.ZELTER,S.VISWANATH,A.MAXWELL,M.J.MACCOSS,
AUTHOR 2 E.G.MULLER,A.SALI,T.N.DAVIS,D.A.AGARD

```

**Fig7. Result for coordinates of similar structure**

## Matches against PDB50:

### Results: lmao

#### Query: 1z2bA

MOLECULE: TUBULIN ALPHA CHAIN;

Select neighbours (check boxes) for viewing as multiple structural alignment or 3D superimposition. The list of neighbours is sorted by Z-score. Similarities with a Z-score lower than 2 are spurious. Each neighbour has links to pairwise structural alignment with the query structure, and to the PDB format coordinate file where the neighbour is superimposed onto the query structure.

Structural Alignment  Expand gaps  3D Superimposition (PV)  SANS  PANZ  Pfam  Reset Selection

#### Summary

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
<input type="checkbox"/> 1:	2btq-A	54.0	1.5	410	436	38	<a href="#">PDB</a>	MOLECULE: TUBULIN BTUBA;
<input type="checkbox"/> 2:	7pqc-B	54.0	1.6	426	451	99	<a href="#">PDB</a>	MOLECULE: TUBULIN BETA CHAIN;
<input type="checkbox"/> 3:	2btq-B	53.5	1.8	388	391	41	<a href="#">PDB</a>	MOLECULE: TUBULIN BTUBA;
<input type="checkbox"/> 4:	7pqc-A	51.5	1.8	418	445	41	<a href="#">PDB</a>	MOLECULE: TUBULIN BETA CHAIN;
<input type="checkbox"/> 5:	3cb2-A	51.4	1.8	407	432	32	<a href="#">PDB</a>	MOLECULE: TUBULIN GAMMA-1 CHAIN;
<input type="checkbox"/> 6:	7m2w-A	47.6	1.8	408	453	30	<a href="#">PDB</a>	MOLECULE: TUBULIN GAMMA CHAIN;
<input type="checkbox"/> 7:	7anz-B	44.6	1.8	375	410	28	<a href="#">PDB</a>	MOLECULE: TUBULIN GAMMA CHAIN;
<input type="checkbox"/> 8:	3zid-A	32.4	2.7	317	360	16	<a href="#">PDB</a>	MOLECULE: TUBULIN/FTSZ, GTPASE;
<input type="checkbox"/> 9:	4b46-A	30.6	2.7	307	330	18	<a href="#">PDB</a>	MOLECULE: CELL DIVISION PROTEIN FTSZ;

**Fig8. Result for similar structures**

## Pairwise Structural Alignments

Notation: three-state secondary structure definitions by DSSP (reduced to H=helix, E=sheet, L=coil) are shown above the amino acid sequence. Structurally equivalent residues are in uppercase, structurally non-equivalent residues (e.g. in loops) are in lowercase. Amino acid identities are marked by vertical bars.

No 1: Query=1z2bA Sbjct=2btqA Z-score=54.0

[back to top](#)

### Fig9. Result for similar structures

```

REMARK Coordinates of 2btq rotated and translated as follows:
REMARK | 0.28475 -0.95786 0.03764 | | x | | 92.000 |
REMARK | -0.37807 -0.07614 0.92264 | * | y | + | 137.000 |
REMARK | -0.88090 -0.27695 -0.38382 | | z | | 57.000 |
REMARK
REMARK HOH and TIP residues excluded. Only first MODEL passed through.
REMARK
HEADER STRUCTURAL PROTEIN 06-JUN-05 2BTQ
TITLE STRUCTURE OF BTUBAB HETERODIMER FROM PROSTHECOBACTER DEJONGEII
COMPND MOL_ID: 1;
COMPND 2 MOLECULE: TUBULIN BTUBA;
COMPND 3 CHAIN: A;
COMPND 4 ENGINEERED: YES;
COMPND 5 MOL_ID: 2;
COMPND 6 MOLECULE: TUBULIN BTUBB;
COMPND 7 CHAIN: B;
COMPND 8 ENGINEERED: YES
AUTHOR D.SCHLIEPER,J.LOWE
HELIX 1 1 GLY A 12 GLY A 31 1
HELIX 2 2 GLU A 74 SER A 85 1
HELIX 3 3 ASN A 90 ALA A 92 5
HELIX 4 4 ASN A 104 LEU A 110 1
HELIX 5 5 GLY A 111 LYS A 130 1
HELIX 6 6 GLY A 146 TYR A 163 1
HELIX 7 7 SER A 176 SER A 180 5
HELIX 8 8 THR A 184 ALA A 200 1
HELIX 9 9 ASN A 208 ARG A 217 1
HELIX 10 10 THR A 225 PHE A 246 1
HELIX 11 11 SER A 255 VAL A 264 1
HELIX 12 12 GLY A 291 PHE A 300 1
HELIX 13 13 SER A 310 GLY A 314 5
HELIX 14 14 ASP A 329 LEU A 344 1

```

**Fig10. Result for coordinates of similar structure**

## Query: 1z2bA

MOLECULE: TUBULIN ALPHA CHAIN;

Select neighbours (check boxes) for viewing as multiple structural alignment or 3D superimposition. The list of neighbours is sorted by Z-score. Similarities with a Z-score lower than 2 are spurious. Each neighbour has links to pairwise structural alignment with the query structure, and to the PDB format coordinate file where the neighbour is superimposed onto the query structure.

Structural Alignment  Expand gaps 3D Superimposition (PV) SANS PANZ Pfam Reset Selection

## Summary

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
□ 1:	5w3f-A	55.4	1.5	424	440	74	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA-1 CHAIN;
□ 2:	2btq-A	54.0	1.5	410	436	38	<a href="#">PDB</a>	MOLECULE: TUBULIN BTUBA;
□ 3:	7pqc-B	54.0	1.6	426	451	99	<a href="#">PDB</a>	MOLECULE: TUBULIN BETA CHAIN;
□ 4:	2btq-B	53.5	1.8	388	391	41	<a href="#">PDB</a>	MOLECULE: TUBULIN BTUBA;
□ 5:	5ubq-A	52.2	1.7	425	441	87	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA CHAIN;
□ 6:	5mlv-C	52.0	1.7	418	430	40	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA-1 CHAIN;
□ 7:	7pqc-A	51.5	1.8	418	445	41	<a href="#">PDB</a>	MOLECULE: TUBULIN BETA CHAIN;
□ 8:	3cb2-A	51.4	1.8	407	432	32	<a href="#">PDB</a>	MOLECULE: TUBULIN GAMMA-1 CHAIN;
□ 9:	5w3f-B	51.3	1.7	416	427	40	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA-1 CHAIN;
□ 10:	7m2w-A	47.6	1.8	408	453	30	<a href="#">PDB</a>	MOLECULE: TUBULIN GAMMA CHAIN;
□ 11:	6v5v-g	45.8	1.6	356	369	33	<a href="#">PDB</a>	MOLECULE: TUBULIN GAMMA-1 CHAIN;
□ 12:	7zpq-B	44.6	1.8	375	410	28	<a href="#">PDB</a>	MOLECULE: TUBULIN GAMMA CHAIN;

Fig11. Result for similar structures

## Pairwise Structural Alignments

Notation: three-state secondary structure definitions by DSSP (reduced to H=helix, E=sheet, L=coil) are shown above the amino acid sequence. Structurally equivalent residues are in uppercase, structurally non-equivalent residues (e.g. in loops) are in lowercase. Amino acid identities are marked by vertical bars.

No 1: Query=1z2bA Sbjct=5w3fA Z-score=55.4

[back to top](#)

Fig12. Result for pairwise structural alignment

```

REMARK Coordinates of 5w3f rotated and translated as follows:
REMARK | 0.39885 -0.50887 0.76286 | | x | | -11.000 |
REMARK | -0.13412 0.79058 0.59749 | * | y | + | -390.000 |
REMARK | -0.90715 -0.34062 0.24708 | | z | | 366.000 |
REMARK
REMARK HOH and TIP residues excluded. Only first MODEL passed through.
REMARK
HEADER HYDROLASE 07-JUN-17 5W3F
TITLE YEAST TUBULIN POLYMERIZED WITH GTP IN VITRO
COMPND MOL_ID: 1;
COMPND 2 MOLECULE: TUBULIN ALPHA-1 CHAIN;
COMPND 3 CHAIN: A;
COMPND 4 MOL_ID: 2;
COMPND 5 MOLECULE: TUBULIN BETA CHAIN;
COMPND 6 CHAIN: B;
COMPND 7 SYNONYM: BETA-TUBULIN
AUTHOR S.C.HOWES,E.A.GEYER,B.LAFRANCE,R.ZHANG,E.H.KELLOGG,S.WESTERMANN,
AUTHOR 2 L.M.RICE,E.NOGALES
HELIX 1 AA1 GLY A 10 HIS A 28 1 19
HELIX 2 AA2 GLU A 72 ASN A 81 1 10
HELIX 3 AA3 TYR A 84 PHE A 88 5 5
HELIX 4 AA4 HIS A 89 GLU A 91 5 3
HELIX 5 AA5 ASN A 103 HIS A 108 1 6
HELIX 6 AA6 VAL A 111 GLU A 114 5 4
HELIX 7 AA7 ILE A 115 CYS A 130 1 16
HELIX 8 AA8 GLY A 144 TYR A 162 1 19
HELIX 9 AA9 VAL A 183 ALA A 199 1 17
HELIX 10 AB1 ASN A 207 ASN A 217 1 11
HELIX 11 AB2 SER A 224 VAL A 239 1 16
HELIX 12 AB3 THR A 240 ARG A 244 5 5
HELIX 13 AB4 ASN A 254 VAL A 261 1 8
HELIX 14 AB5 SER A 288 GLU A 298 1 11

```

**Fig13. Result for coordinates of similar structure**

## Query: 1z2bA

MOLECULE: TUBULIN ALPHA CHAIN;

Select neighbours (check boxes) for viewing as multiple structural alignment or 3D superimposition. The list of neighbours is sorted by Z-score. Similarities with a Z-score lower than 2 are spurious. Each neighbour has links to pairwise structural alignment with the query structure, and to the PDB format coordinate file where the neighbour is superimposed onto the query structure.

Structural Alignment  Expand gaps  3D Superimposition (PV)  SANS  PANZ  Pfam  Reset Selection

## Summary

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
<input type="checkbox"/> 1:	1z2b-A	75.0	0.0	427	427	100	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA CHAIN;
<input type="checkbox"/> 2:	1sa0-A	69.5	0.3	427	427	99	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA CHAIN;
<input type="checkbox"/> 3:	3hkc-A	69.1	0.5	426	426	100	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA CHAIN;
<input type="checkbox"/> 4:	3hke-A	69.0	0.5	427	427	100	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA CHAIN;
<input type="checkbox"/> 5:	3n2k-A	68.9	0.5	427	428	100	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA CHAIN;
<input type="checkbox"/> 6:	3e22-A	68.9	0.5	427	427	98	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA-1C CHAIN;
<input type="checkbox"/> 7:	3hkb-A	68.6	0.5	427	427	100	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA CHAIN;
<input type="checkbox"/> 8:	1z2b-C	68.5	0.6	420	427	99	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA CHAIN;
<input type="checkbox"/> 9:	1sa0-C	68.5	0.5	421	421	99	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA CHAIN;
<input type="checkbox"/> 10:	3ed1-F	68.4	0.5	421	421	99	<a href="#">PDB</a>	MOLECULE: ALPHA-TUBULIN;
<input type="checkbox"/> 11:	3n2g-A	68.1	0.5	427	428	100	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA CHAIN;
<input type="checkbox"/> 12:	3hkd-A	68.1	0.5	427	428	100	<a href="#">PDB</a>	MOLECULE: TUBULIN ALPHA CHAIN.

**Fig14. Result for similar structures**

## Pairwise Structural Alignments

Notation: three-state secondary structure definitions by DSSP (reduced to H=helix, E=sheet, L=coil) are shown above the amino acid sequence. Structurally equivalent residues are in uppercase, structurally non-equivalent residues (e.g. in loops) are in lowercase. Amino acid identities are marked by vertical bars.

No 1: Query=1z2bA Sbjct=1z2bA Z-score=75.0

[back to top](#)

### Fig15. Result for pairwise structural alignment

```

REMARK Coordinates of 1z2b rotated and translated as follows:
REMARK | 1.00000 0.00000 -0.00000 | | x | | -0.000 |
REMARK | -0.00000 1.00000 0.00000 | * | y | + | 0.000 |
REMARK | 0.00000 -0.00000 1.00000 | | z | | 0.000 |
REMARK
REMARK HOH and TIP residues excluded. Only first MODEL passed through.
REMARK
HEADER CELL CYCLE 08-MAR-05 1Z2B
TITLE TUBULIN-COLCHICINE-VINBLASTINE: STATHMIN-LIKE DOMAIN COMPLEX
COMPND MOL_ID: 1;
COMPND 2 MOLECULE: TUBULIN ALPHA CHAIN;
COMPND 3 CHAIN: A, C;
COMPND 4 MOL_ID: 2;
COMPND 5 MOLECULE: TUBULIN BETA CHAIN;
COMPND 6 CHAIN: B, D;
COMPND 7 MOL_ID: 3;
COMPND 8 MOLECULE: RB3 STATHMIN-LIKE DOMAIN 4;
COMPND 9 CHAIN: E;
COMPND 10 SYNONYM: STATHMIN-LIKE PROTEIN B3, RB3-SLD;
COMPND 11 ENGINEERED: YES
AUTHOR B.GIGANT,C.WANG,R.B.G.RAVELLI,F.ROUSSI,M.O.STEINMETZ,P.A.CURMI,
AUTHOR 2 A.SOBEL,M.KNOSSOW
HELIX 1 1 GLY A 10 GLY A 29 1 20
HELIX 2 2 THR A 73 ARG A 79 1 7
HELIX 3 3 HIS A 88 GLU A 90 5 3
HELIX 4 4 ASN A 102 TYR A 108 1 7
HELIX 5 5 ILE A 110 ALA A 126 1 17
HELIX 6 6 GLY A 143 TYR A 161 1 19
HELIX 7 7 VAL A 182 GLU A 196 1 15
HELIX 8 8 ASP A 205 ASN A 216 1 12
HELIX 9 9 THR A 223 ALA A 240 1 18
HELIX 10 10 ALA A 240 ASP A 245 1 6

```

**Fig16. Result for coordinates of similar structure**

## RESULT:

PDB ID for albumin structure was searched in structure similarity BLAST tool, DALI and similar structures matches against PDB25, PDB50, PDB90 and all PDB structures were retrieved.

## CONCLUSION:

DALI is a useful structure similarity BLAST tool which provides user with similar structures to their query along with its pairwise alignment, coordinates information, 3D superimposition results. Describing the

structural similarity relationships between protein structures allows scientists to infer the functions of newly discovered proteins.

## REFERENCES:

1. Tubulin: Mystery of vital cell protein solved after 30 years. (n.d.). Retrieved March 19, 2022, from <https://www2.lbl.gov/Science-Articles/Archive/3D-tubulin.html#:~:text=Tubulin%20is%20the%20protein%20that,or%20to%20regulate%20intracellular%20transport>.
2. Dali server. (n.d.). Ekhidna2.Biocenter.helsinki.fi. Retrieved March 14, 2022, from <http://ekhidna2.biocenter.helsinki.fi/dali/>
3. Bank, R. P. D. (n.d.-b). RCSB PDB - 4L8U: X-ray study of human serum albumin complexed with 9 amino camptothecin. [Www.rcsb.org](http://www.rcsb.org). Retrieved March 14, 2022, from <https://www.rcsb.org/structure/4L8U>
4. Dali server. (n.d.). Ekhidna2.Biocenter.helsinki.fi. Retrieved March 14, 2022, from <http://ekhidna2.biocenter.helsinki.fi/barcosel/tmp//4l8uA/>