Total:8

EXPERIMENT NO: 01 DATE:18.07.2024

Comments: Can Improve

BASICS OF UNIX

AIM:

To understand the basics of linux operating system and to perform various unix commands operation

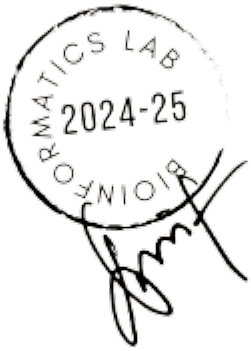
INTRODUCTION

Unix is a powerful, multiuser, multitasking operating system originally developed in the 1960s at Bell Labs by Ken Thompson, Dennis Ritchie, and others. It is known for its simplicity, flexibility, and portability, which have made it a foundational system in the world of computing. Unix provides a command-line interface that allows users to interact directly with the system through commands, enabling precise control over system resources and processes. Its architecture follows a modular design, where small, single-purpose tools can be combined to perform complex tasks. The Unix philosophy emphasizes building simple, reusable programs that can work together, making it highly efficient for developers and system administrators. Over the years, Unix has influenced many modern operating systems, including Linux and macOS, and remains a critical platform for servers, workstations, and embedded systems worldwide.

TOOLS USED

Some key tools used in Unix:

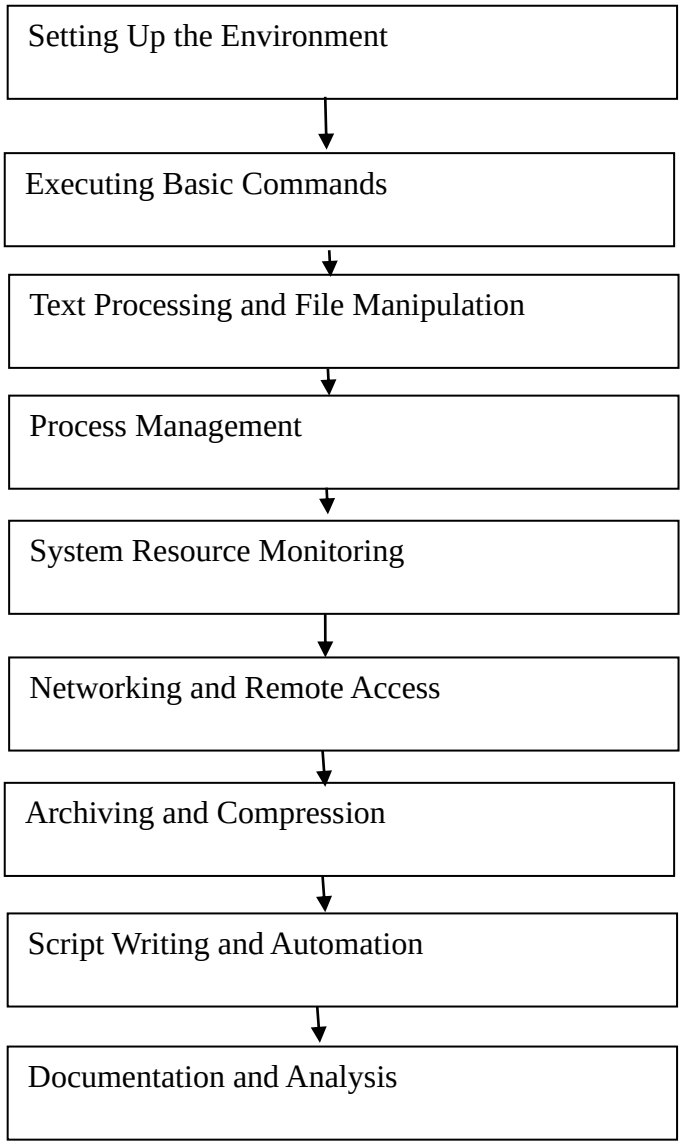
1. Shell
2. File Management Tools
3. Text Processing Tools
4. Process Management Tools
5. Networking Tools
6. Archiving and Compression Tools
7. System Monitoring Tools
8. Programming and Scripting Tools



Marks: 2-2-2-1-1 / Date :04:10:2024 [14:59:35]

METHODS

Comments: Can Improve



RESULTS AND DISCUSSION

Thus, the given activity was completed using the below-mentioned commands and been listed in the table

FIGURES & TABLES

Comments: Can Improve

Command	Description
ls	ls [options] [directory] - Lists files and directories within a directory.
cd	cd [directory] - Changes the current directory to the specified directory.
pwd	pwd - Prints the current working directory.
cp	cp [source] [destination] - Copies files or directories from the source to the destination.
mv	mv [source] [destination] - Moves or renames files or directories.
rm	rm [file] - Removes (deletes) files or directories.
chmod	chmod [mode] [file] - Changes the file permissions.
chown	chown [owner]:[group] [file] - Changes the ownership of a file or directory.
cat	cat [file] - Concatenates and displays the contents of a file.
grep	grep [pattern] [file] - Searches for a specific pattern within files.
sed	sed [options] 'script' [file] - Stream editor for filtering and transforming text in a file.
awk	awk 'pattern {action}' [file] - Pattern scanning and processing language for text processing.
find	find [directory] [criteria] - Searches for files and directories based on given criteria.
ps	ps [options] - Displays information about currently running processes.
kill	kill [PID] - Terminates a process by its Process ID (PID).
tar	tar [options] [archive] [files] - Archives files into a single file or extracts them.
gzip	gzip [file] - Compresses a file using the gzip algorithm.
ssh	ssh [user@hostname] - Securely connects to a remote system via SSH.
scp	scp [source] [destination] - Securely copies files between systems over SSH.
df	df [options] - Displays disk space usage for file systems.
du	du [options] [directory] - Estimates file space usage within a directory.
uptime	uptime - Shows how long the system has been running and the load averages.
top	top - Displays real-time system processes and resource usage.
echo	echo [text] - Prints text to the terminal.
whoami	whoami - Displays the current logged-in user's username.
man	man [command] - Displays the manual page for a specified command.

	MAX MARK	MARKS OBTAINED
INTRODUCTION/ PROBLEM STATEMENT	2	
EXPERIMENTAL PROCEDURE	2	
RESULTS, DATA, FIGURES TABLES ETC	2	
DISCUSSION	2	
VIVA VOCE	2	
APPREANCE & FORMATTING	10	

EXPERIMENT NO: 02 DATE:19.07.2024

Comments: Can Improve

BLAST

AIM:

- To analyze the genome sequences from NCBI database and to learn about the structure, genomic context, genomic regions, conserved domains of a specific enzyme
- Identifying sequences with evolutionary relationships to the query sequence using BLAST search.

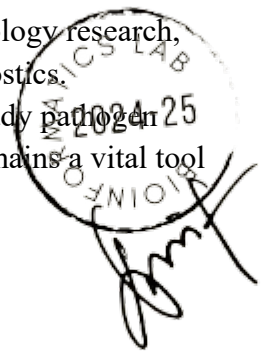
INTRODUCTION:

BLAST (Basic Local Alignment Search Tool) is a powerful algorithm used in bioinformatics to compare a query sequence (DNA, RNA, or protein) against a database of sequences to identify regions of similarity. Developed by Stephen Altschul and colleagues in 1990, BLAST has become an essential tool for researchers working with genetic data, enabling the rapid identification of homologous sequences, gene prediction, and functional annotation.

BLAST operates by breaking down the query sequence into short segments called "words" and searching for matches in the database. It then extends these matches to form high-scoring segment pairs (HSPs), which represent regions of significant similarity between the query and database sequences. This method allows BLAST to quickly and efficiently find regions of interest, making it ideal for large-scale analyses.

There are several versions of BLAST, each designed for different types of sequences and analyses. For example, BLASTn compares nucleotide sequences, BLASTp compares protein sequences, and BLASTx translates a nucleotide sequence into all six possible reading frames and compares it against a protein database. Additionally, the BLAST suite includes tools like PSI-BLAST, which identifies distant protein relatives by iteratively searching for conserved motifs, and MegaBLAST, optimized for comparing highly similar sequences.

The versatility and speed of BLAST have made it a cornerstone of molecular biology research, used in a wide range of applications, from evolutionary studies to clinical diagnostics. Researchers use BLAST to annotate genomes, identify potential drug targets, study pathogen evolution, and much more. As genetic databases continue to expand, BLAST remains a vital tool for navigating and interpreting the vast amount of sequence data available today.



Marks: 2-2-2-1-1 / Date :04:10:2024 [14:59:35]

TOOLS USED

Comments: Can Improve

1. Word Matching
2. Scoring Matrices
3. Heuristic Search
4. High-Scoring Segment Pairs (HSPs)
5. Gap Penalties
6. E-value Calculation
7. Iterative Searches (PSI-BLAST)
8. MegaBLAST

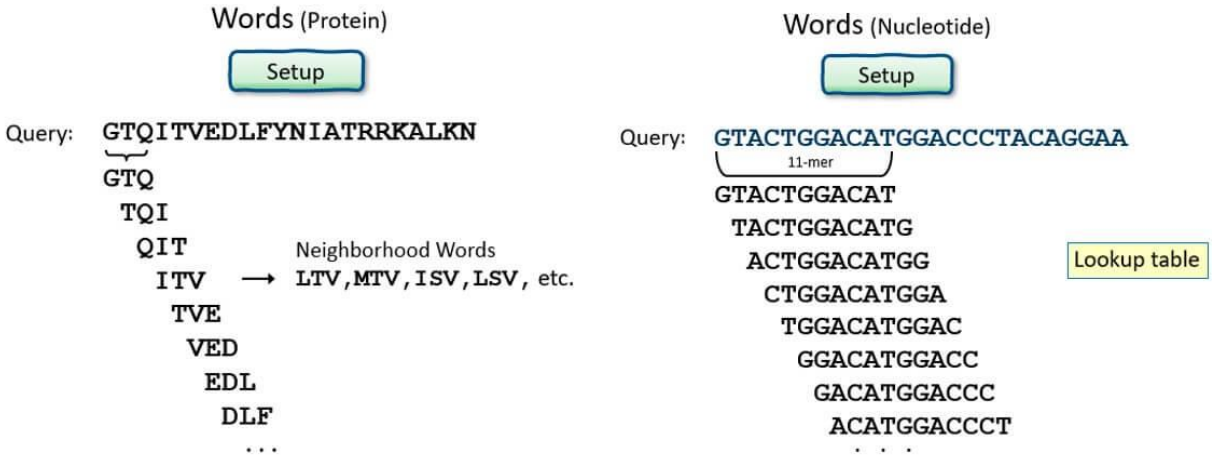


Fig.A Representing the word search in BLAST

METHODS

Comments: Can Improve

1. Start
↓
2. Input Query Sequence
↓
3. Word Generation: Break down the query sequence into smaller "words" of a fixed length.
↓
4. Word Matching: Compare these words against a database to find matches.
↓
5. Initial Alignment: Identify regions where the query sequence matches the database sequence.
↓
6. Scoring Using Matrices: Use scoring matrices like PAM/BLOSUM to evaluate similarity, assigning scores to matches, mismatches, and gaps.
↓
7. High-Scoring Segment Pairs (HSPs): Extend initial matches to form HSPs, identifying segments of significant similarity.
↓
8. Gap Penalties Application: Introduce and penalize gaps (insertions/deletions) to maintain biologically meaningful alignments.
↓
9. E-value Calculation: Calculate the E-value to estimate the significance of the match (lower E-value = more significant match).
↓
10. Iterative Search (PSI-BLAST): For protein sequences, generate a position-specific scoring matrix (PSSM) and perform iterative searches to find distant homologs (if applicable).
↓
11. MegaBLAST (Optional): Use MegaBLAST for highly similar sequences (e.g., in large genomic databases).
↓
12. Output Results: Display alignment results, including matched sequences, scores, and E-values.

FIGURES & TABLES

Comments: Can Improve

```
>tr|A4HSF7|A4HSF7_LEIIN Trypanothione reductase OS=Leishmania infantum OX=5671 GN=TRYR PE=1 SV=1
MSRAYDLVVLGAGSGGLEAGWNAAVTHKKKVAVVDVQATHGPPLFAALGGTCVNVGCVPK
KLMVTGAQYMDLIRESGGFGWEMDRESLCPNWKTLIAAKNKVVNSINESYKSMFADTEGL
SFHMGFGALQDAHTVVVRKSEDPHSDVLETLDTEYILIATGSWPTRLGVPGDEFCITSNE
AFYLEDAPKRMLCVGGGYIAVEFAGIFNGYKPCGGYVDLCYRGDLILRGFDTEVRKSLTK
QLGANGIRVRTNLNPTKITKNEDGSNHVHFNDGTEEDYDQVMLAIGRVPRSQALQLDKAG
VRTGKNGAVQVDAYSKTSVDNIYAIGDVTNRVMLTPVAINEGAAFVETVFGGKPRATDHT
KVACAVFSIPPIGTCGMTEEEAAKNYETVAVYASSFTPLMHNISGSKHKEFMIRIITNES
NGEVLGVHMLGDSAPEIIQSVGICMKMGAKISDFHSTIGVHPTSAEELCSMRTPAYFYES
GKRVEKLSSNL
```

Figure 1. Fasta sequence of the desired protein- ***Trypanothione reductase*** **(PDB:2JK6)**

Figure 2. 3D structure of the desired protein- ***Trypanothione reductase***

Total:8

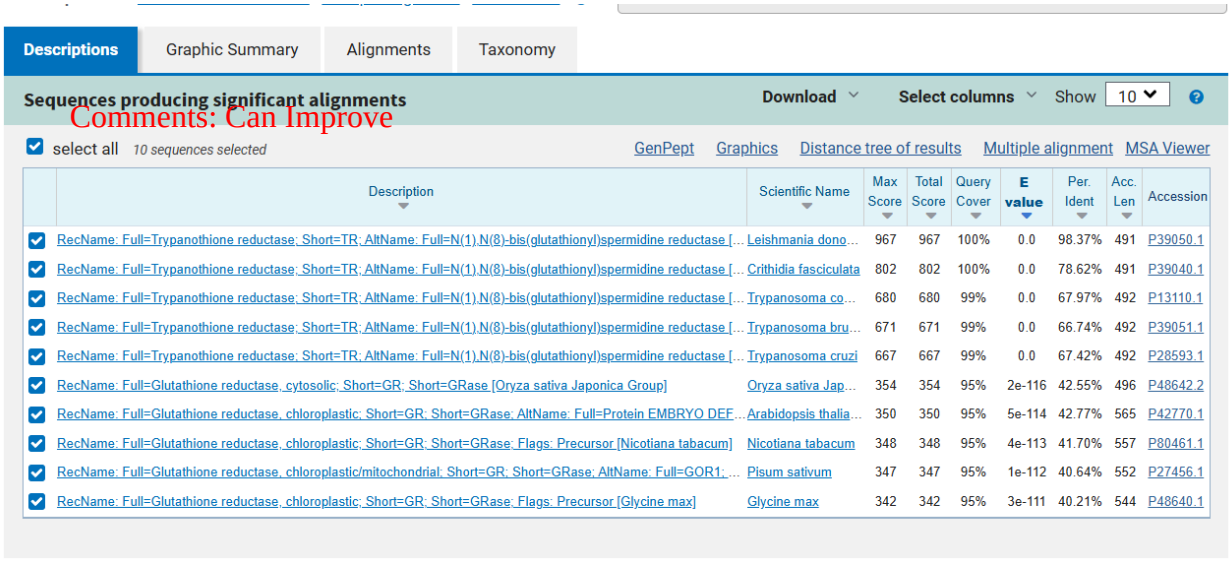


Fig 3. Description summary of the BLAST results of the protein – **CASE A**

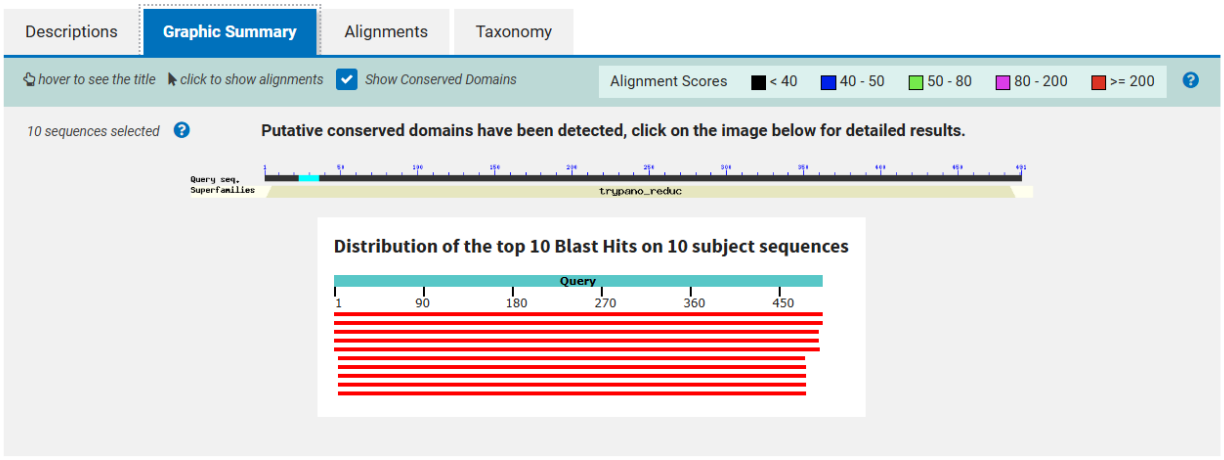


Fig 4. Graphical summary of the BLAST results of the protein

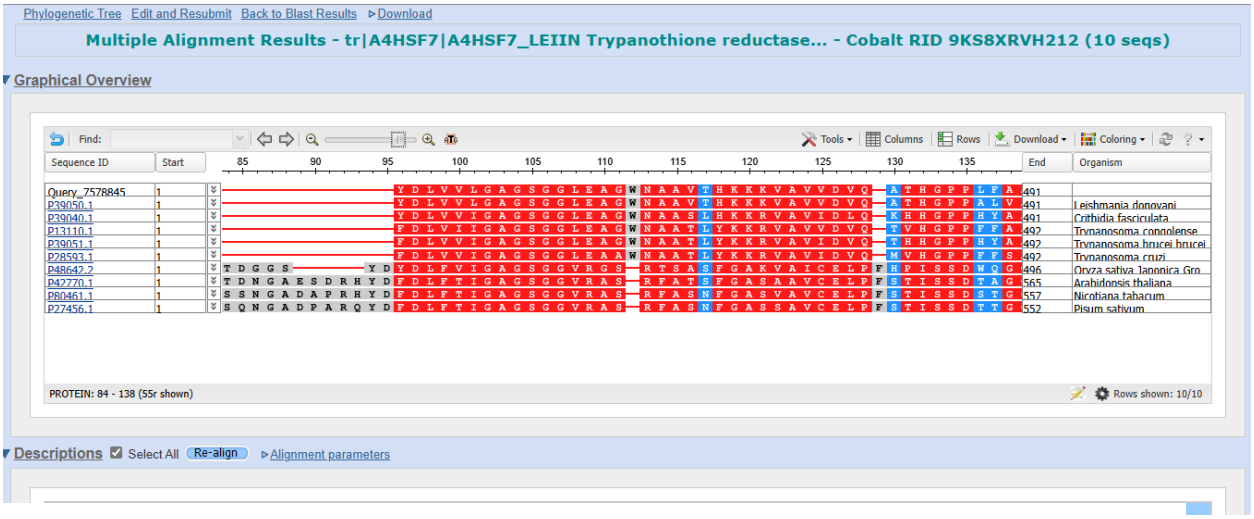
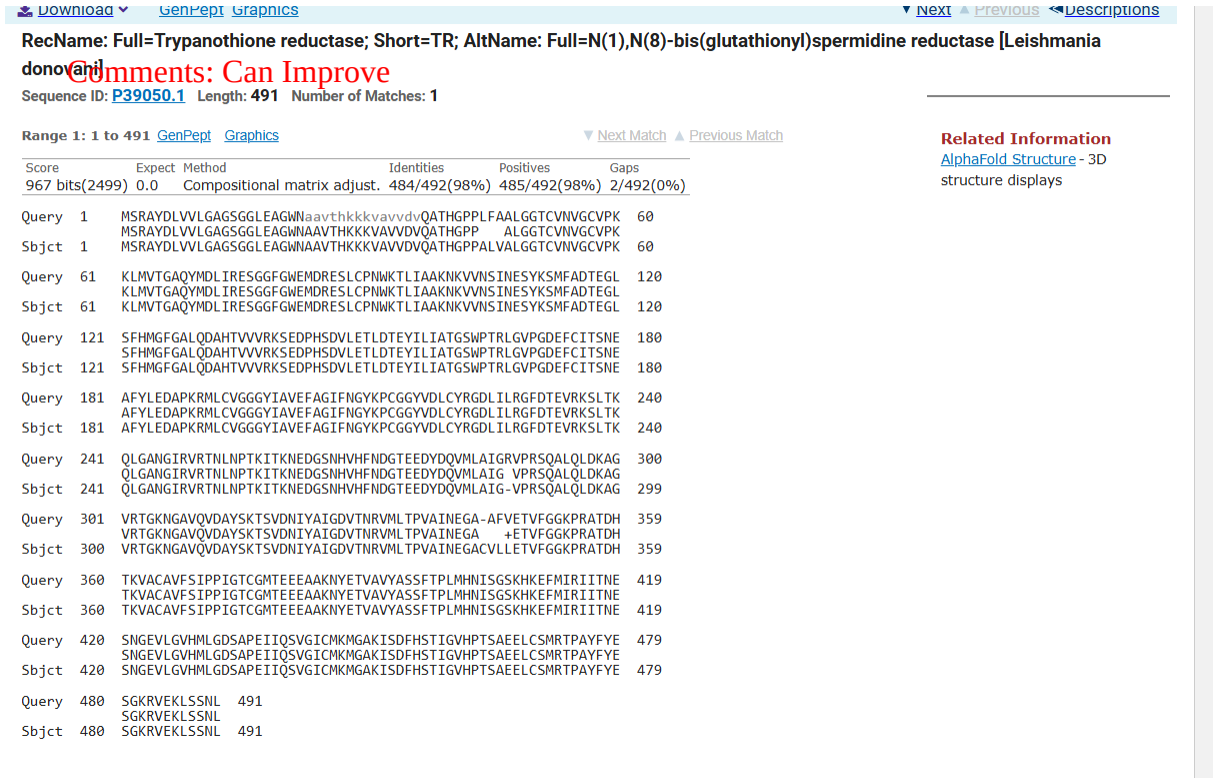


Fig 5. Multiple alignment summary of the BLAST results of the protein

Marks: 2-2-2-1-1 / Date :04:10:2024 [14:59:35]



Download ▾ GenPept Graphics ▼ Next ▲ Previous Descriptions

RecName: Full=Trypanothione reductase; Short=TR; AltName: Full=N(1),N(8)-bis(glutathionyl)spermidine reductase [Leishmania donovani]

Comments: Can Improve

Sequence ID: P39050.1 Length: 491 Number of Matches: 1

Range 1: 1 to 491 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
967 bits(2499)	0.0	Compositional matrix adjust.	484/492(98%)	485/492(98%)	2/492(0%)

```
Query  1    MSRAYDLVVLGAGSGGLEAGWNaavthkkkvavvdvQATHGPPLFAALGGTCVNVGCVPK  60
            MSRAYDLVVLGAGSGGLEAGWNAAVTHKKKVAVVDVQATHGPP   ALGGTCVNVGCVPK
Sbjct  1    MSRAYDLVVLGAGSGGLEAGWNAAVTHKKKVAVVDVQATHGPPALVALGGTCVNVGCVPK  60

Query  61   KLMVTGAQYMDLIRESGGFGWEMDRESLCPNWKTLIAAKNKVVNSINESYKSMFADTEGL  120
            KLMVTGAQYMDLIRESGGFGWEMDRESLCPNWKTLIAAKNKVVNSINESYKSMFADTEGL
Sbjct  61   KLMVTGAQYMDLIRESGGFGWEMDRESLCPNWKTLIAAKNKVVNSINESYKSMFADTEGL  120

Query  121  SFHMGFGALQDAHTVVVRKSEDPHSDVLETLDTEYILIATGSWPTRLGVPGDEFCITSNE  180
            SFHMGFGALQDAHTVVVRKSEDPHSDVLETLDTEYILIATGSWPTRLGVPGDEFCITSNE
Sbjct  121  SFHMGFGALQDAHTVVVRKSEDPHSDVLETLDTEYILIATGSWPTRLGVPGDEFCITSNE  180

Query  181  AFYLEDAPKRMLCVGGGYIAVEFAGIFNGYKPCGGYVDLCYRGDLILRGFDTEVRKSLTK  240
            AFYLEDAPKRMLCVGGGYIAVEFAGIFNGYKPCGGYVDLCYRGDLILRGFDTEVRKSLTK
Sbjct  181  AFYLEDAPKRMLCVGGGYIAVEFAGIFNGYKPCGGYVDLCYRGDLILRGFDTEVRKSLTK  240

Query  241  QLGANGIRVRTNLNPTKITKNEDGSNHVHFNDGTEEDYDQVMLAIGRVPRSQALQLDKAG  300
            QLGANGIRVRTNLNPTKITKNEDGSNHVHFNDGTEEDYDQVMLAIG VPRSQALQLDKAG
Sbjct  241  QLGANGIRVRTNLNPTKITKNEDGSNHVHFNDGTEEDYDQVMLAIG-VPRSQALQLDKAG  299

Query  301  VRTGKNGAVQVDAYSKTSVDNIYAIGDVTNRVMLTPVAINEGA-AFVETVFGGKPRATDH  359
            VRTGKNGAVQVDAYSKTSVDNIYAIGDVTNRVMLTPVAINEGA   +ETVFGGKPRATDH
Sbjct  300  VRTGKNGAVQVDAYSKTSVDNIYAIGDVTNRVMLTPVAINEGACVLLETVFGGKPRATDH  359

Query  360  TKVACAVFSIPPIGTCGMTEEEAAKNYETVAVYASSFTPLMHNISGSKHKEFMIRIITNE  419
            TKVACAVFSIPPIGTCGMTEEEAAKNYETVAVYASSFTPLMHNISGSKHKEFMIRIITNE
Sbjct  360  TKVACAVFSIPPIGTCGMTEEEAAKNYETVAVYASSFTPLMHNISGSKHKEFMIRIITNE  419

Query  420  SNGEVLGVHMLGDSAPEIIQSVGICMKMGAKISDFHSTIGVHPTSAEELCSMRTPAYFYE  479
            SNGEVLGVHMLGDSAPEIIQSVGICMKMGAKISDFHSTIGVHPTSAEELCSMRTPAYFYE
Sbjct  420  SNGEVLGVHMLGDSAPEIIQSVGICMKMGAKISDFHSTIGVHPTSAEELCSMRTPAYFYE  479

Query  480  SGKRVEKLSSNL  491
            SGKRVEKLSSNL
Sbjct  480  SGKRVEKLSSNL  491
```

Related Information
AlphaFold Structure - 3D structure displays

Fig 6. Alignment summary of the BLAST results of the protein – The query seq and the subject seq is compared

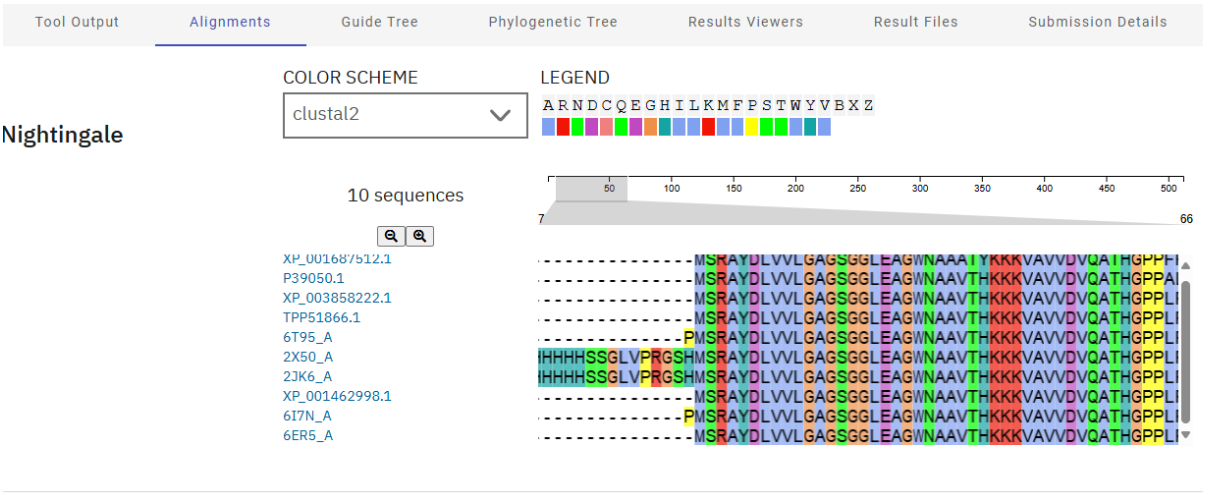


Fig 7. The fasta seq is subjected to conservative domain search (Image from *CLUSTAL OMEGA)*

Total:8

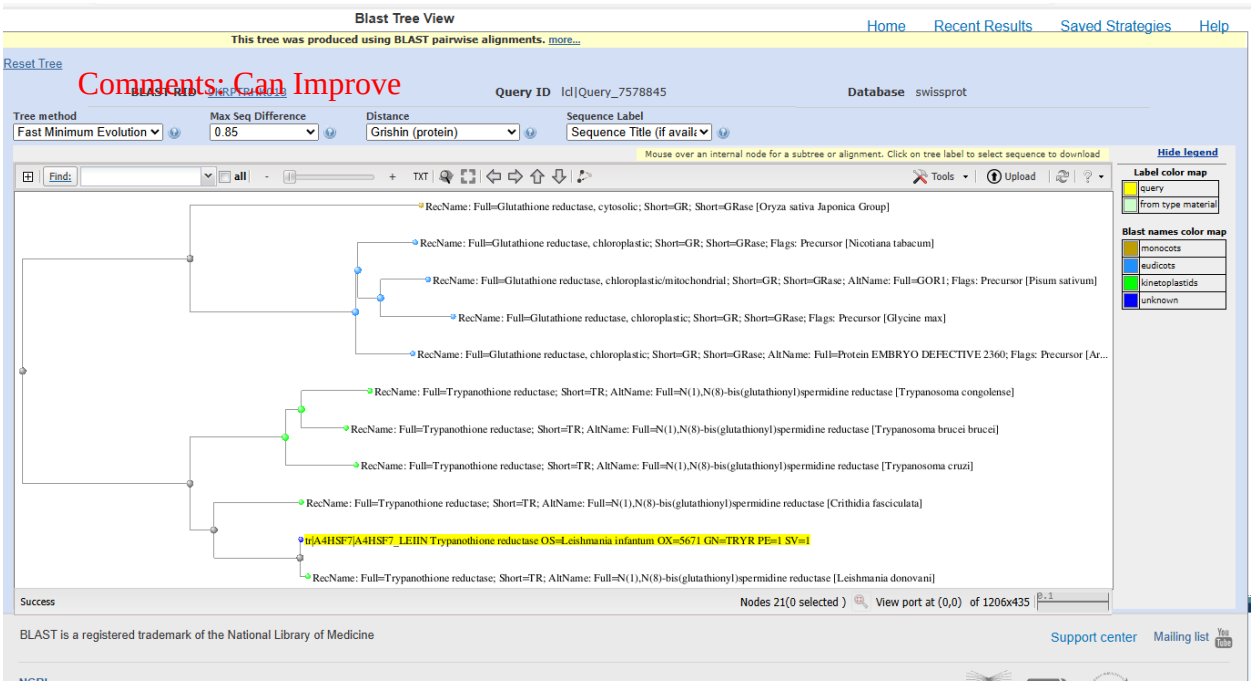


Fig 8. Distance tree alignment of the BLAST results of the protein

CASE B

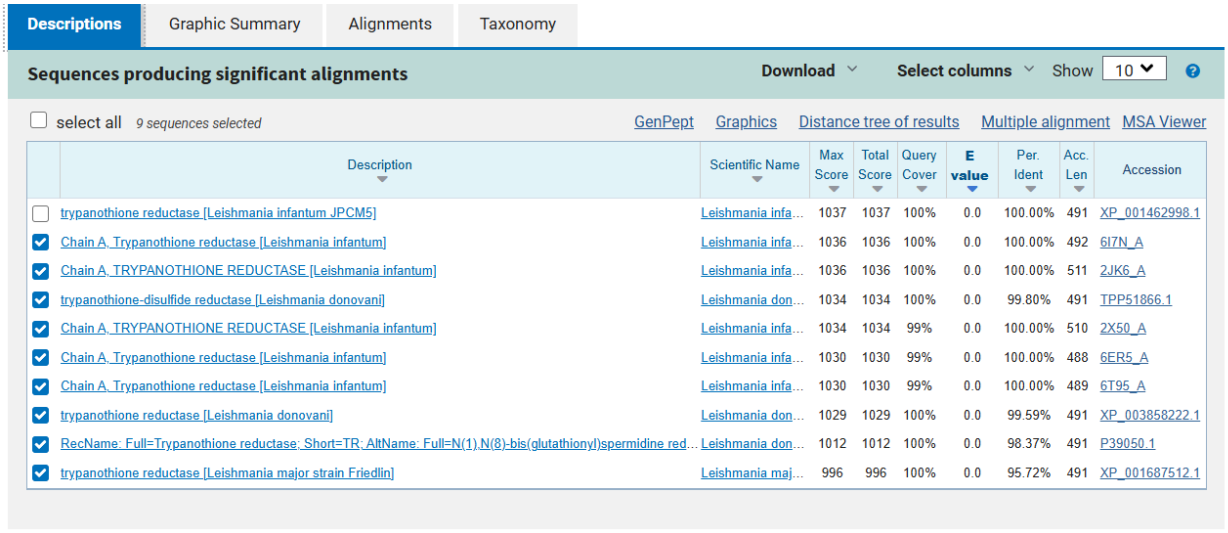


Fig 9. The threshold and the matrices changed and the sequence is subjected to BLAST

Marks: 2-2-2-1-1 / Date :04:10:2024 [14:59:35]

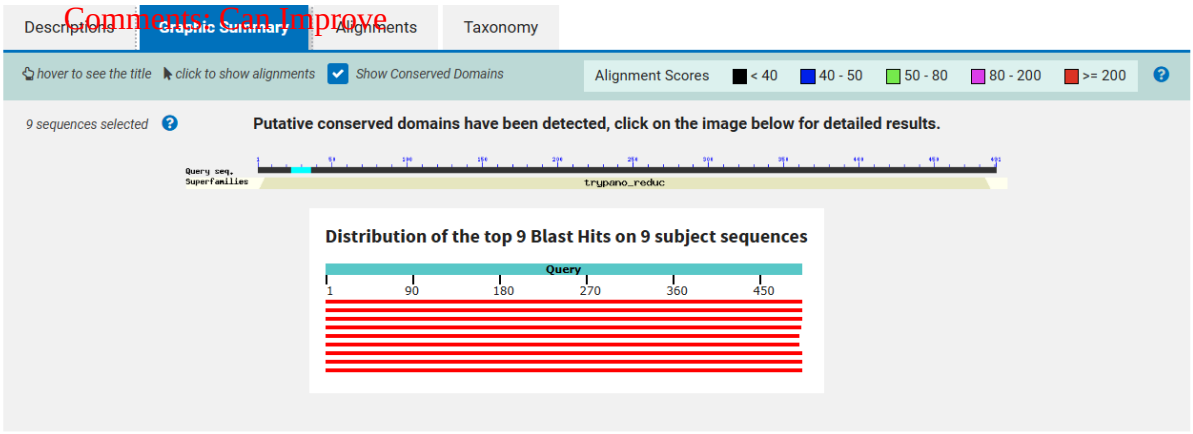


Fig 10. Graphical summary of the BLAST results of the protein

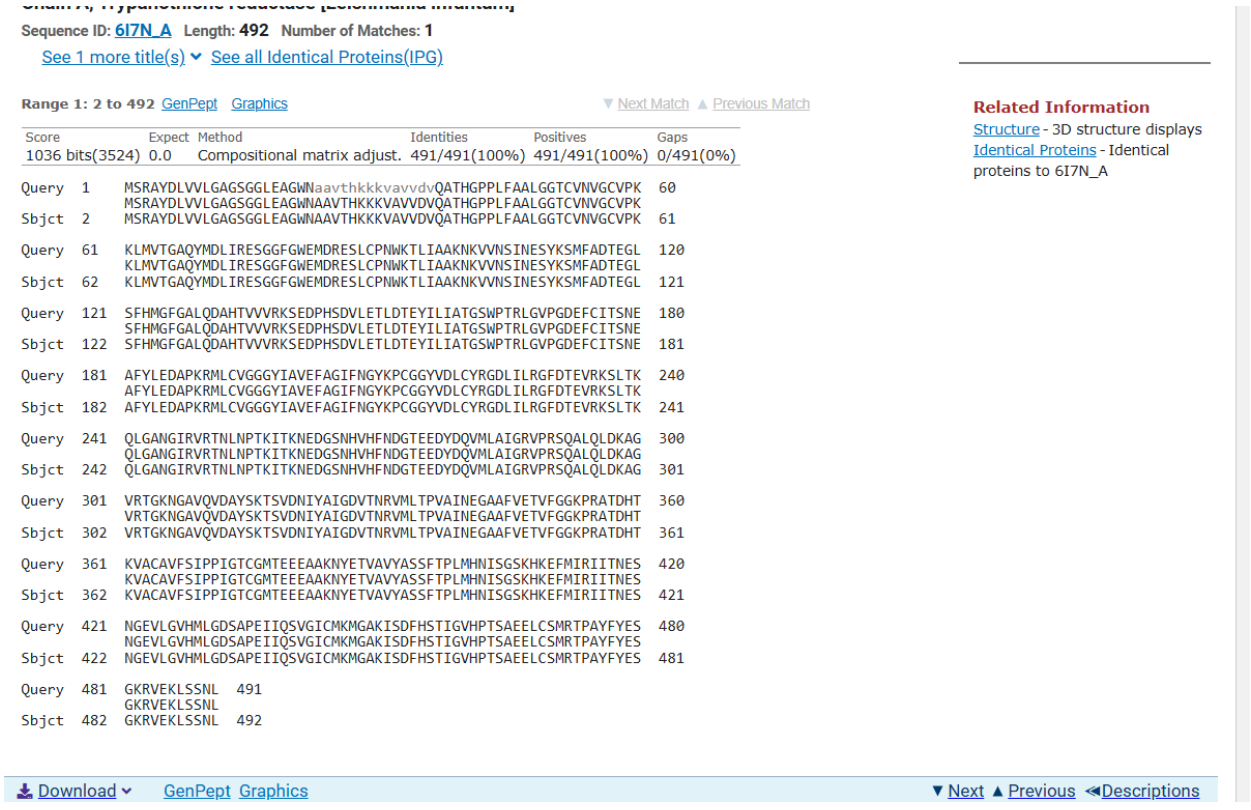


Fig 3. Alignment summary of the BLAST results of the protein

RESULTS AND DISCUSSION

Comments: Can Improve

In the study, the protein sequence of Trypanothione reductase was obtained from UniProt and its 3D structure was analyzed. The sequence was subjected to a BLAST search, and the resulting description, graphical representation, and alignment summary were thoroughly examined.

In **CASE A**, the BLAST search was performed using the default settings, which included the default threshold, 10 sequences, and the inclusion of low-complexity regions. This approach successfully identified conserved domains within the protein sequence, providing a foundational understanding of its functional regions and potential homologs.

In **CASE B**, adjustments were made to the threshold and scoring matrices. These modifications led to a more refined and efficient search, offering deeper insights into the protein's homologs and functional domains. By optimizing the search parameters, CASE B demonstrated an enhanced ability to identify relevant sequences and conserved regions, thereby improving the overall accuracy and relevance of the BLAST results.

This comparative analysis between the default settings and optimized parameters highlights the importance of fine-tuning BLAST search conditions to achieve more precise and informative outcomes in protein sequence analysis.

	MAX MARK	MARKS OBTAINED
INTRODUCTION/ PROBLEM STATEMENT	2	
EXPERIMENTAL PROCEDURE	2	
RESULTS, DATA, FIGURES TABLES ETC	2	
DISCUSSION	2	
VIVA VOCE	2	
APPREANCE & FORMATTING	10	

Marks: 2-2-2-1-1 / Date :04:10:2024 [14:59:35]

EXPERIMENT NO: 03 DATE:26.07.2024

Comments: Can Improve

MOLECULAR VISUALIZATION USING PYMOL

AIM

To learn Molecular Visualization and Structural Analysis

INTRODUCTION

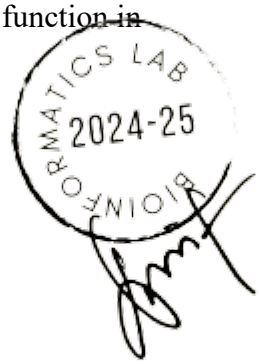
PyMOL is a widely-used molecular visualization software that allows scientists to create detailed 3D representations of molecular structures. Developed by Warren Lyford DeLano in 2000, PyMOL is now an essential tool in structural biology, bioinformatics, and computational chemistry. It is particularly valued for its ability to produce high-quality images and animations that aid in the analysis and presentation of complex molecular data.

One of PyMOL's key features is its versatility in visualizing a wide range of molecular structures, including proteins, nucleic acids, small molecules, and macromolecular complexes. The software supports various file formats, such as PDB (Protein Data Bank) files, which contain atomic coordinates for molecular structures, allowing users to import and manipulate structural data with ease.

PyMOL's interface, though powerful, is user-friendly, catering to both beginners and advanced users. It offers a command-line interface for precise control and scripting, as well as a graphical user interface (GUI) for more intuitive operations. Users can perform tasks such as molecular modeling, structure refinement, and surface analysis, as well as explore molecular interactions, such as hydrogen bonds, hydrophobic contacts, and electrostatic potential.

In addition to static images, PyMOL can generate animations to illustrate dynamic processes like protein folding, ligand binding, or conformational changes. These visualizations are invaluable in research and education, as they help to convey complex molecular mechanisms in a more accessible and understandable way.

PyMOL is a crucial tool for anyone working in the fields of molecular biology, chemistry, or related disciplines. Its ability to visualize and analyze molecular structures in detail makes it indispensable for understanding the intricate relationships between structure and function in biological systems



Marks: 2-2-2-1-1 / Date :04:10:2024 [14:59:35]

Comments: Can Improve

TOOLS USED

pyMOL

RESULTS AND DISCUSSION

1. Molecular Visualization: Using PyMOL, we achieved clear and detailed 3D visualizations of the target macromolecules. The software allowed us to generate high-resolution images and animations of protein structures, including their secondary and tertiary folds. Visualization of complex formations and binding sites was facilitated, revealing the spatial arrangement and interaction between different molecular components.

2. Structural Analysis: PyMOL's analysis tools enabled us to perform detailed examination of molecular surfaces, electrostatic potentials, and ligand interactions. We observed specific binding sites and interactions, which were crucial for understanding the functional mechanisms of the proteins. The ability to color-code different structural features and apply various representations (e.g., sticks, spheres, ribbons) provided deeper insights into the molecular structure and function.

FIGURES & TABLES

Fig1. Fetch the protein ***Trypanothione reductase*** **(PDB:2JK6)**

Comments: Can Improve

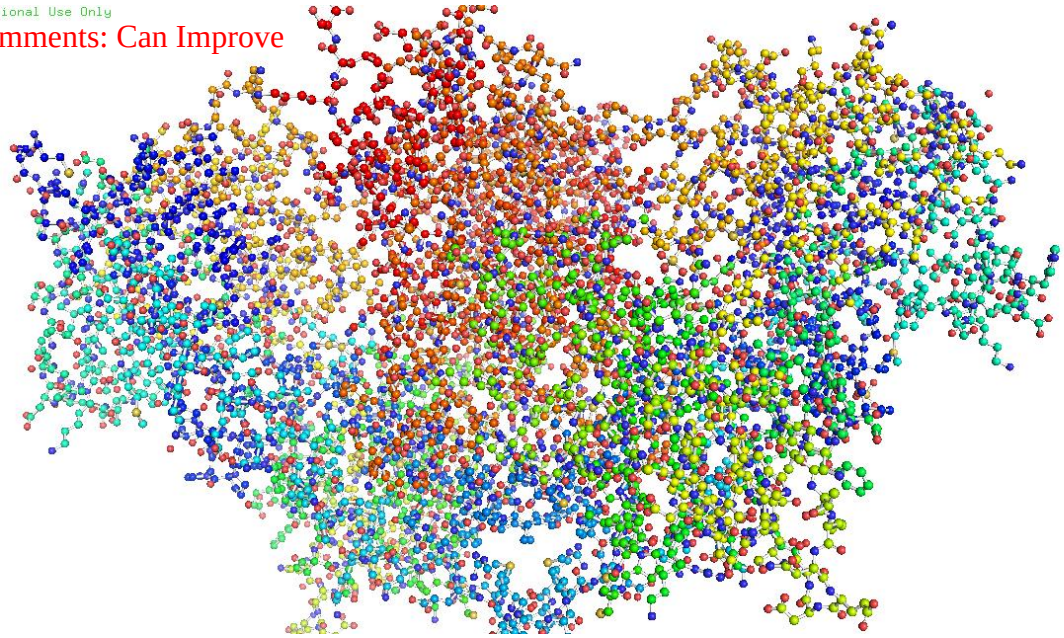


Fig2. Change the preset to ball and stick model

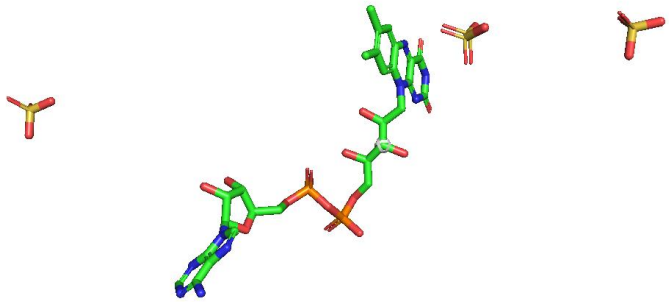


Fig3. Select the ligand

Comments: Can Improve

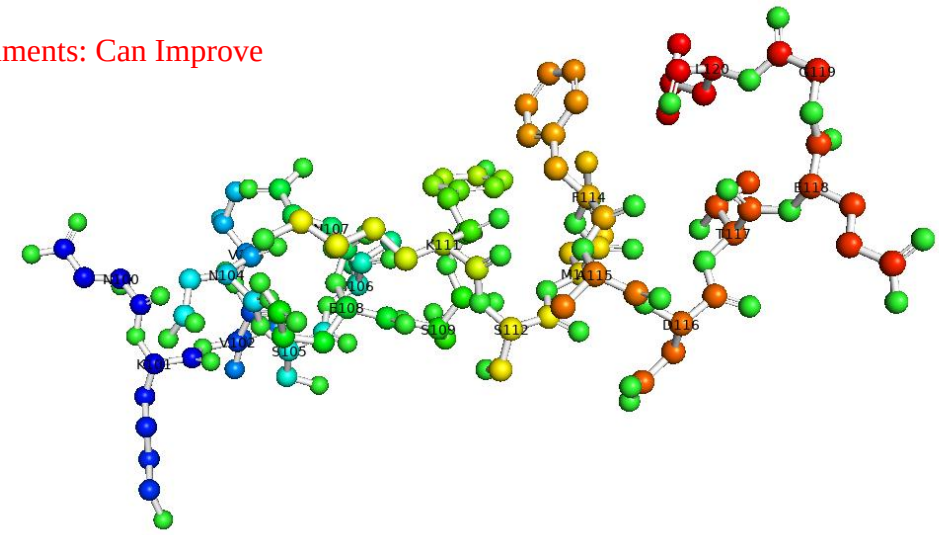


Fig4. Labelling the ligand

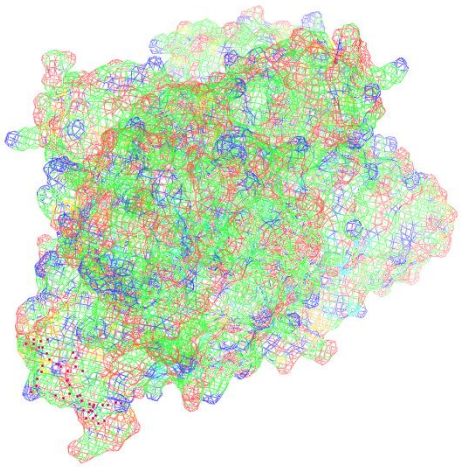


Fig5. Change the preset to mesh

Comments: Can Improve

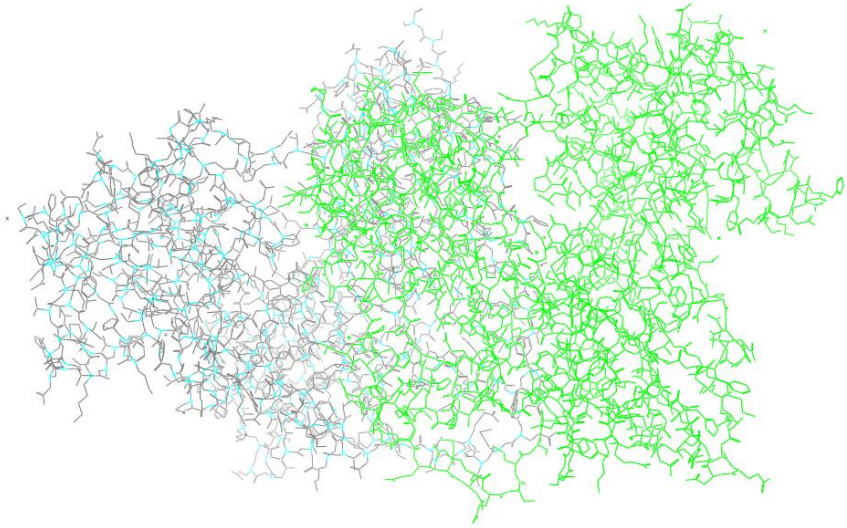


Fig6. Change the preset to lines

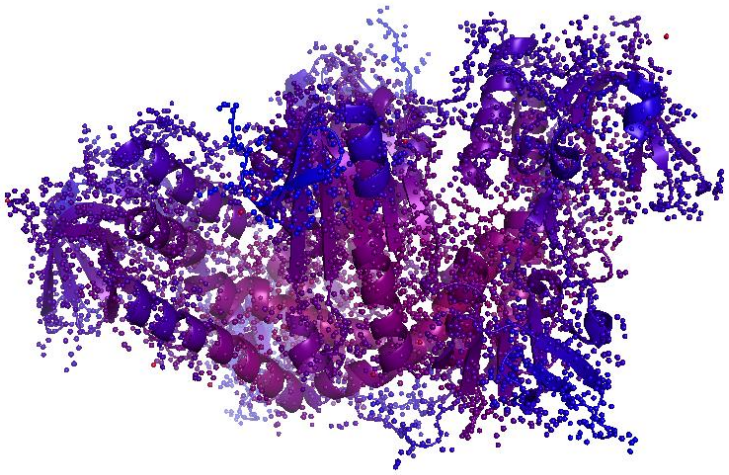


Fig7. Change the preset to ribbon and colour to spectrum

Comments: Can Improve

Fig8. Select chain A

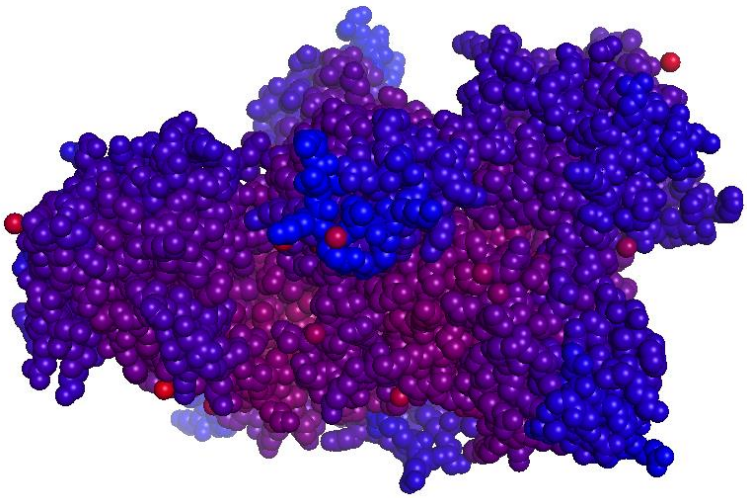


Fig9. Change the present to spheres

Comments: Can Improve

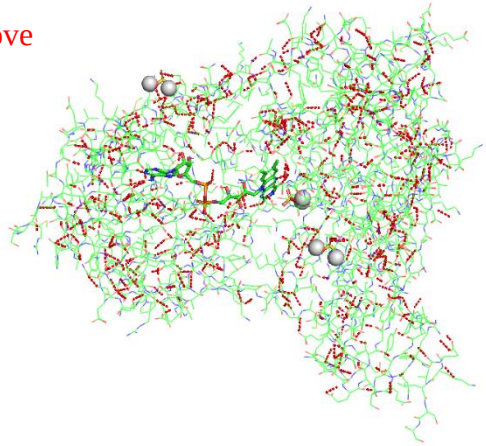


Fig9. Polar bonds between the chain A

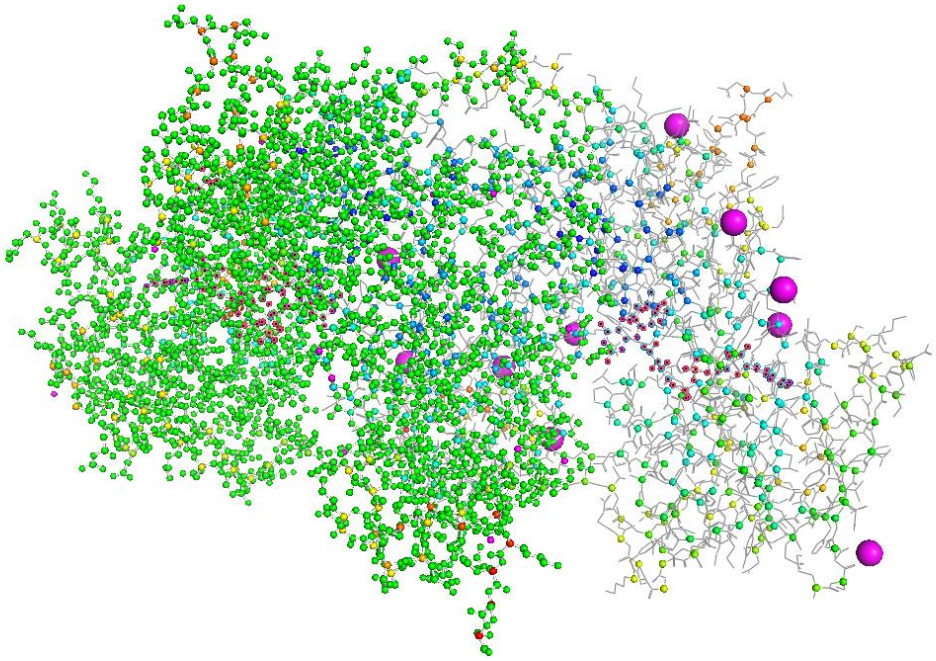


Fig10. Distinguishing between chain A and Chain B

Total:8

METHODS

Comments: Can Improve

1. Start
↓
2. Obtain Molecular Data
↓
3. Load Data into PyMOL
↓
4. Prepare the Structure
↓
5. Visualize Molecular Structure
↓
6. Analyze Interactions
↓
7. Generate Visual Outputs
↓
8. Save Results
↓
9. End

	MAX MARK	MARKS OBTAINED
INTRODUCTION/ PROBLEM STATEMENT	2	
EXPERIMENTAL PROCEDURE	2	
RESULTS, DATA, FIGURES TABLES ETC	2	
DISCUSSION	2	
VIVA VOCE	2	
APPREANCE & FORMATTING	10	

Marks: 2-2-2-1-1 / Date :04:10:2024 [14:59:35]

Total:8

EXPERIMENT NO: 04 DATE: 02.08.2024

Comments: Can Improve

HOMOLOGY MODELING OF PROTEIN

AIM

To predict Molecular Structures and Understanding Molecular Interactions

INTRODUCTION

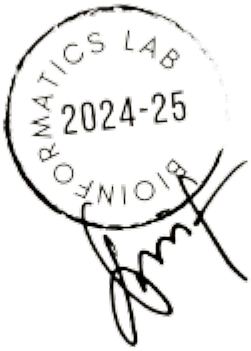
Modeller is a powerful computational tool used in molecular modeling to build three-dimensional structures of biological macromolecules. Developed by André P. M. P. G. D. S. G. G. R. R. M. W. (Wang et al., 2004), Modeller is designed to predict the structure of proteins, nucleic acids, and complex biomolecular assemblies based on known homologous structures.

At its core, Modeller utilizes comparative or homology modeling, which involves creating models by aligning a target sequence with homologous sequences of known structures. This process allows researchers to predict the 3D conformation of the target molecule with a high degree of accuracy. Modeller incorporates sophisticated algorithms for building and refining these models, accounting for the geometry and energetics of the molecular system.

The tool is widely used in structural biology, drug design, and bioinformatics. By providing insights into molecular structure and function, Modeller aids in understanding biological processes and facilitating the development of new therapeutic agents. Its versatility and robust performance make it an essential component in the computational toolkit of researchers in various fields of molecular science.

TOOLS USED

1. **Homology Modeling:** Modeller
2. **Structure Validation Tools:** PROCHECK and VERIFY3D
3. **Visualization Software:** PyMOL, Chimera
4. **Sequence Alignment Tools:** CLUSTALW or MUSCLE



Marks: 2-2-2-1-1 / Date :04:10:2024 [14:59:35]

METHODS

Comments: Can Improve

1. Input Data Collection
(Target Sequence Identification, Template Selection)
↓
2. Sequence Alignment
(Alignment Tools, Generate Alignment File)
↓
3. Model Building
(Initial Model Construction, Model Refinement)
↓
4. Model Validation
(Quality Assessment, Adjustments)
↓
5. Molecular Dynamics
(Simulation Setup, Run Simulations)
↓
6. Visualization and Analysis
(Visualization Tools, Interaction Analysis)
↓
7. Documentation and Reporting
(Document Findings, Publication or Further Research)

RESULTS AND DISCUSSION

The protein sequence of Trypanothione reductase was obtained from UniProt, and its 3D structure was modeled through homology modeling. The resulting model was evaluated using both the DOPE score and the Ramachandran plot.

The DOPE score revealed some inconsistencies in the loop regions of the model, indicating that these regions might benefit from further refinement. The DOPE profile, which assesses the quality of protein structures based on statistical potentials, highlighted areas where the model could be improved, particularly in loop regions.

The Ramachandran plot provided additional validation of the model's geometry, showing the distribution of dihedral angles and confirming that most residues were in favorable regions. There were no residues were in disallowed regions, which is consistent with the findings from the DOPE score analysis.

To address these issues, it is recommended to perform loop refinement modeling.

FIGURES & TABLES

Comments: Can Improve

```
HITS FOUND IN ITERATION:     1


Dynamically allocated memory at     amaxprofile [B,KiB,MiB]:      1030490    1006.338     0.983
> 1aogA                     1     577   85750     485     491   66.94     0.0          2   484     2   486     1   484
> 1ojt                      1    1173   21550     482     491   29.10     0.0          3   446     5   468     7   463
> 3grs                      1    1468   32400     461     491   37.04     0.0          4   438     4   472     3   461
> 1trb                      1    1815    5750     316     491   27.72    0.81E-05      5   163   156   332   108   291
> 1d7yA                     1    2104    5850     401     491   30.30    0.67E-05      6   182   136   331    76   273
> 1dxlA                     1    2456   21750     467     491   29.40     0.0          7   440     6   468     5   453
> 1ebdA                     1    2656   21850     455     491   29.48     0.0          8   434     6   466     5   445
> 1nhp                      1    3142    6450     447     491   26.69    0.35E-06      9   218    96   329    48   283
> 1fecA                     1    3228  102600     485     491   78.76     0.0         10   485     2   486     1   485
> 1fl2A                     1    3356    4850     310     491   28.09    0.83E-03     11   165   151   329   102   279
> 1gesA                     1    3766   36900     448     491   38.34     0.0         12   442     5   472     3   448
> 1h6vA                     1    4097   30850     490     491   32.69     0.0         13   451     5   466     4   468
> 1hyuA                     1    4415    5050     521     491   29.21    0.59E-03     14   165   151   329   313   490
> 1jehA                     1    5010   18600     478     491   26.11     0.0         15   449     1   476     2   472
> 3ladA                     1    5807   21250     472     491   27.79     0.0         16   447     2   468     1   457
> 1lpfA                     1    5938   21150     472     491   28.01     0.0         17   448     2   468     1   457
> 1lvl                      1    5993   16950     458     491   28.05     0.0         18   427     7   466     8   442
> 1mo9A                     1    6221   11650     522     491   25.79     0.0         19   440     3   472    41   513
> 1onfA                     1    6831   30800     439     491   37.67     0.0         20   428     5   472     2   439
> 1xdiA                     1    9281   15400     459     491   26.02     0.0         21   429     7   463     4   445
> 1q1rA                     1   10119    4700     421     491   25.23    0.27E-02     22   197   151   355   100   321
> 1xhcA                     1   10821    7900     346     491   30.92    0.14E-09     23   198   135   355    85   291
```

Fig1. Highest matched templates

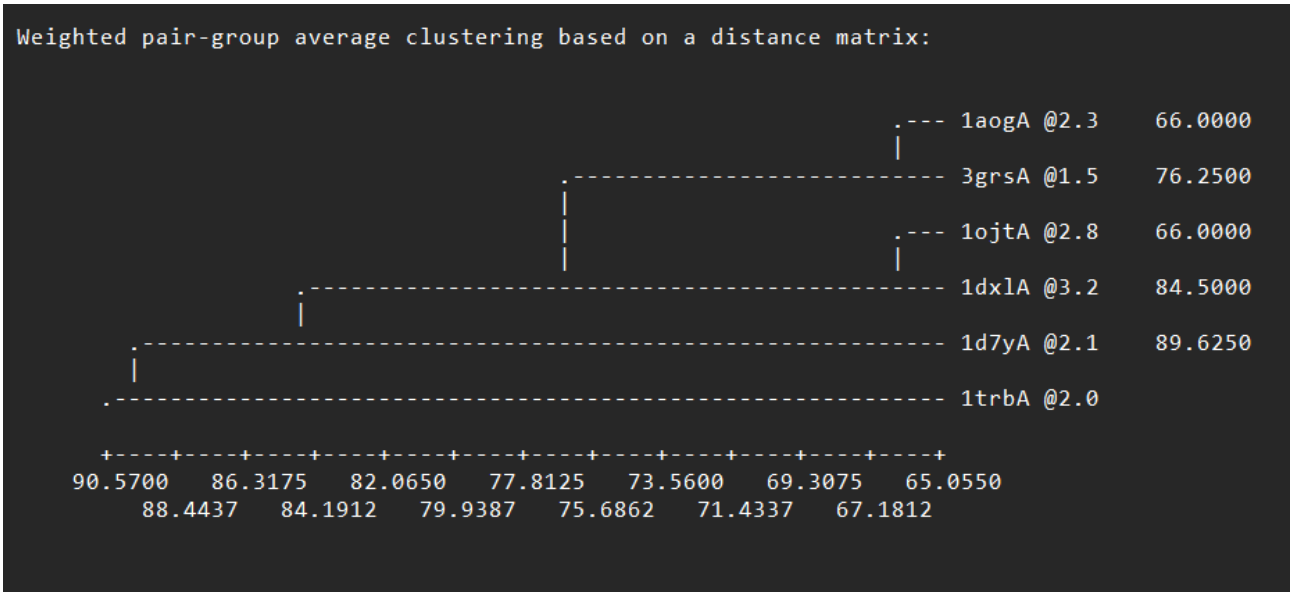


Fig2. Comparison and clustering of templates based on resolution

Marks: 2-2-2-1-1 / Date :04:10:2024 [14:59:35]

```
 _aln.pos         10        20        30        40        50        60
3grsA     -VASYDYLVIGGGSGGLASARRAAELG-ARAAVVES---H------KLGGTCVNVGCVPKKVMWNTAV
A4HSF7    MSRAYDLVVLGAGSGGLEAGWNAAVTHKKKVAVVDVQATHGPPLFAALGGTCVNVGCVPKKLMVTGAQ
 _consrvd     **  * * *****     **       ***     *       ************ *   *

 _aln.p   70        80        90       100       110       120       130
3grsA     HSEFMHDHADYGFP-SCEGK-FNWRVIKEKRDAYVSRLNAIYQNNLTKSH-IEIIRGHAAFTSDPKPT
A4HSF7    YMDLIRESGGFGWEMDRESLCPNWKTLIAAKNKVVNSINESYKSMFADTEGLSFHMGFGALQDAHTVV
 _consrvd           *     *    **          *   *  *              *  *

 _aln.pos  140       150       160       170       180       190       200
3grsA     IEVS-------GKKYTAPHILIATGGMPSTPHESQIPGASLGITSDGFFQLEELPGRSVIVGAGYIAV
A4HSF7    VRKSEDPHSDVLETLDTEYILIATGSWPTRLG---VPGDEFCITSNEAFYLEDAPKRMLCVGGGYIAV
 _consrvd    *               ******  *        **    ***   * **  * *   ** *****

 _aln.pos    210       220       230       240       250       260       270
3grsA     EMAGILSAL---GSKTSLMIRHDKVLRSFDSMISTNCTEELENAGVEVLKFSQVKEVKKTLSGLEVSM
A4HSF7    EFAGIFNGYKPCGGYVDLCYRGDLILRGFDTEVRKSLTKQLGANGIRVRTNLNPTKITKNEDGSNHVH
 _consrvd * ***       *    *  * *  ** **       *  *   *  *          *   *

 _aln.pos      280       290       300       310       320       330       340
3grsA     VTAVPGRLPVMTMIPDVDCLLWAIGRVPNTKDLSLNKLGIQTDDKGHIIVDEFQNTNVKGIYAVGDVC
A4HSF7    FN--------DGTEEDYDQVMLAIGRVPRSQALQLDKAGVRTGKNGAVQVDAYSKTSVDNIYAIGDVT
 _consrvd               * *    ******    * * * *  *   *   **    * *  *** ***

 _aln.pos       350       360       370       380       390       400
3grsA     GKALLTPVAIAAGRKLAHRLFEYKEDSKLDYNNIPTVVFSHPPIGTVGLTEDEAIHKYGIENVKTYST
A4HSF7    NRVMLTPVAINEGAAFVETVFGGK-PRATDHTKVACAVFSIPPIGTCGMTEEEAAKNYE--TVAVYAS
 _consrvd     ******  *       *  *     *       *** ***** * ** **   *    *  *

 _aln.p  410       420       430       440       450       460       470
3grsA     SFTPMYHAVTKRKTKCVMKMVCANK-EEKVVGIHMQGLGCDEMLQGFAVAVKMGATKADFDNTVAIHP
A4HSF7    SFTPLMHNISGSKHKEFMIRIITNESNGEVLGVHMLGDSAPEIIQSVGICMKMGAKISDFHSTIGVHP
 _consrvd ****  *     * *  *     *     * * ** *    *  *      ****   **  *   **

 _aln.pos  480       490       500
```

Comments: Can Improve

Fig3. Alignment of template and the query sequence

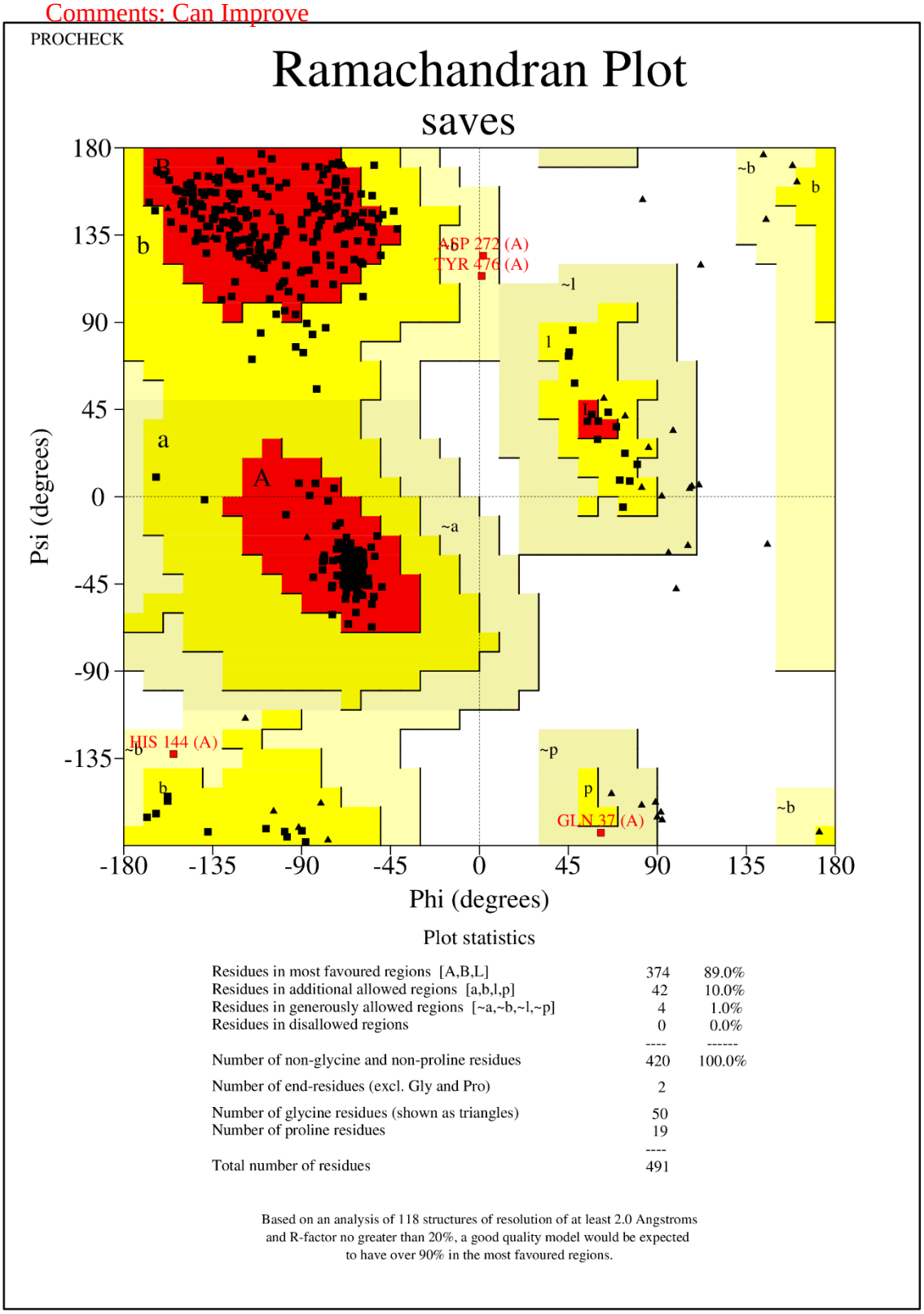
```
<< end of ENERGY.

>> Summary of successfully produced models:
Filename                          molpdf     DOPE score    GA341 score
----------------------------------------------------------------------
A4HSF7.B99990001.pdb          2636.63208   -48288.99609        1.00000
A4HSF7.B99990002.pdb          2820.45093   -48445.58203        1.00000
A4HSF7.B99990003.pdb          2540.28906   -47276.41406        1.00000
A4HSF7.B99990004.pdb          2622.63232   -47723.83594        1.00000
A4HSF7.B99990005.pdb          2752.07617   -47257.96094        1.00000

Total CPU time [seconds]                                :      60.63
```

Fig4. Selecting the most appropriate template

Comments: Can Improve



saves_01.ps

Fig5. Ramachandran plot

Marks: 2-2-2-1-1 / Date :04:10:2024 [14:59:35]

Comments: Can Improve

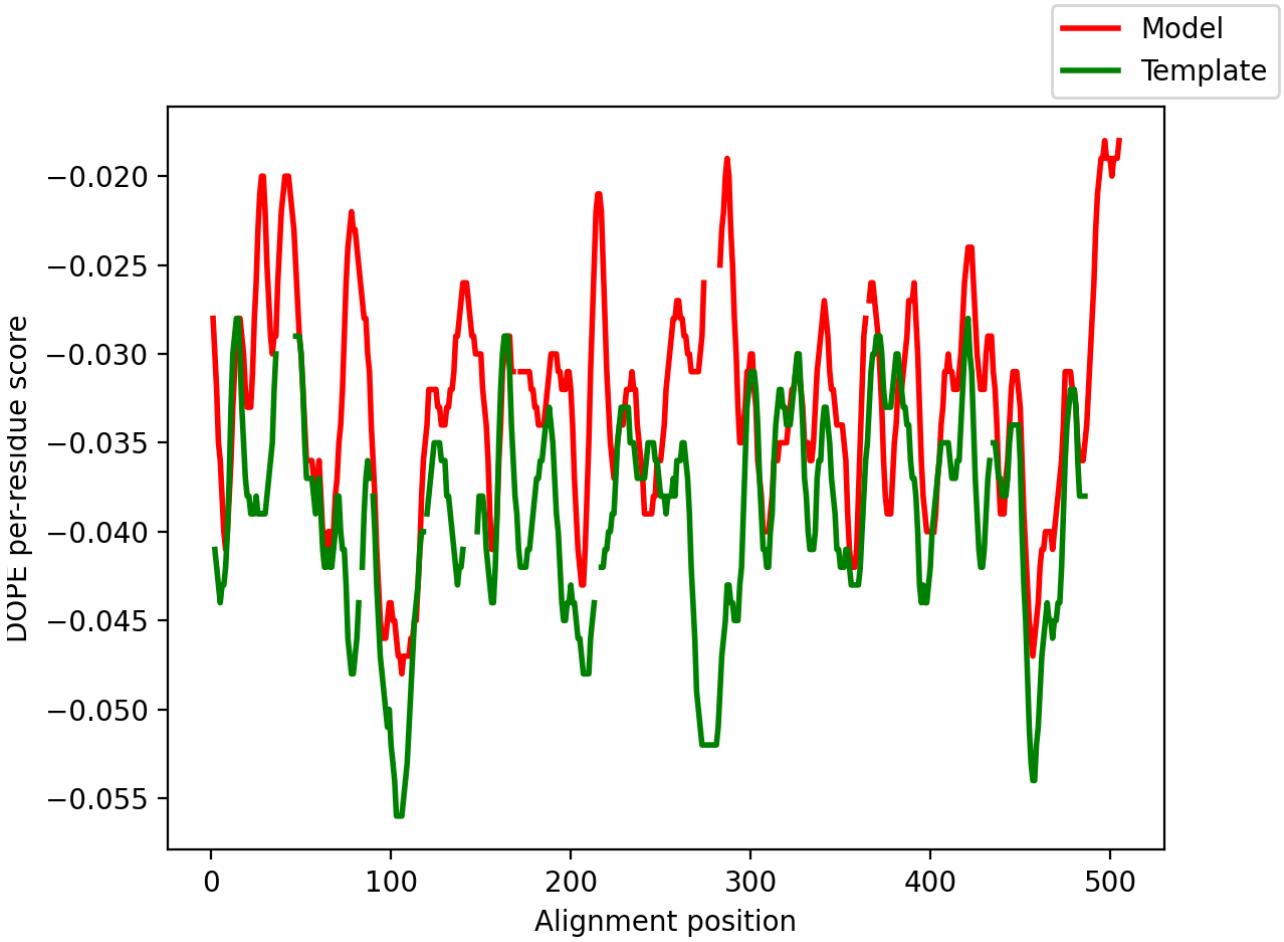


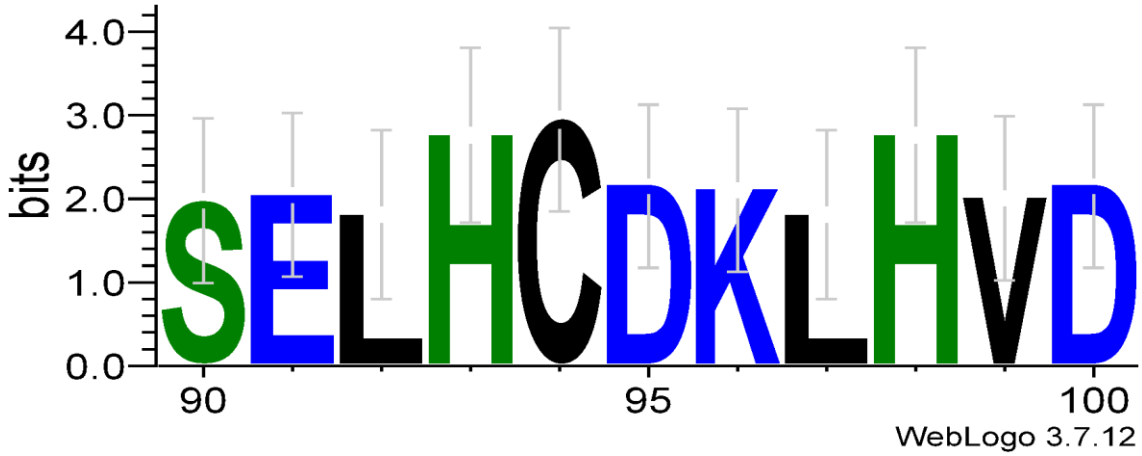
Fig6. DOPE profile

	MAX MARK	MARKS OBTAINED
INTRODUCTION/ PROBLEM STATEMENT	2	
EXPERIMENTAL PROCEDURE	2	
RESULTS, DATA, FIGURES TABLES ETC	2	
DISCUSSION	2	
VIVA VOCE	2	
APPREANCE & FORMATTING	10	

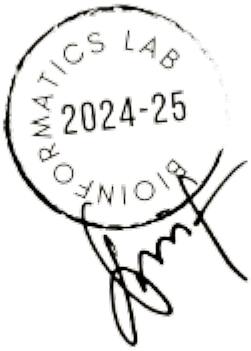
Marks: 2-2-2-1-1 / Date :04:10:2024 [14:59:35]

WB006-MSA BENCHMARKING

Comments: Can Improve



Eventually, the results of ClustalW's inference demonstrate a promising alignment with highly conserved residues across the sequences of globin. It is important to benchmark several tools to guarantee optimal alignment for further analysis. The most suitable tool will be chosen based on the degree to which it can handle structural features, speed, gap management, and evolutionary relationship, all of which are critical given the varied and evolutionary nature of globins.
The degree of amino acid conservation at a given location is indicated by the height of each letter. Smaller letters denote variability, whereas larger ones represent strong conservation. For example, there is major conservation of histidine (H) and serine (S) at positions 91 and 98, respectively.The residues that are highly conserved at these places, especially the histidine (H) between positions 94 and 98, may be important for the structural or functional integrity of the globin protein. It's believed that these residues help bind or preserve the core structure.



Marks: 2-2-2-1-1 / Date :04:10:2024 [14:59:35]

Total:8

Advanced Homology Modelling

Comments: Can Improve

Steps:

1. Generate a multiple structure alignment of the family
2. A pairwise alignment is proceeded with the use of a structure-dependent gap penalty
3. We build the new model for the target sequence based on the alignment against the multiple templates
4. DOPE potential is used to evaluate the new model coordinates

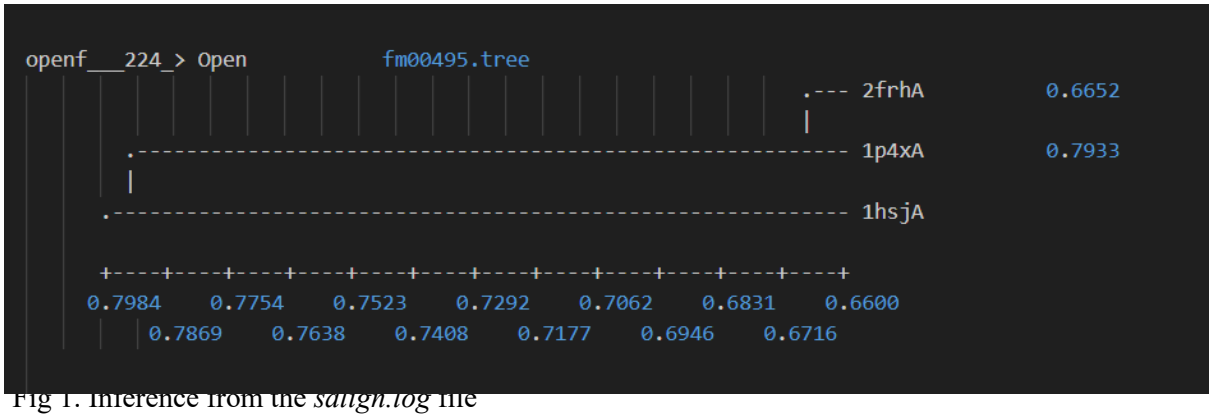


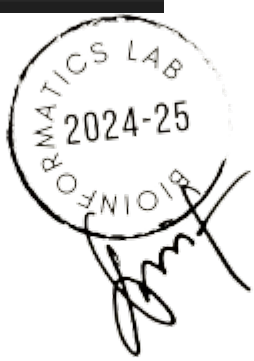
Fig 1. Inference from the *salign.log* file

```
<< end of ENERGY.

>> Summary of successfully produced models:
Filename                          molpdf     DOPE score    GA341 score
----------------------------------------------------------------------
q.B99990001.pdb                601.58698   -12426.70020        1.00000
q.B99990002.pdb                590.99518   -12282.94727        1.00000
q.B99990003.pdb                517.68054   -12210.61133        1.00000
q.B99990004.pdb                567.20978   -12312.70703        1.00000
q.B99990005.pdb                518.30328   -12422.77539        1.00000

Total CPU time [seconds]                                :       9.84
```

Fig 2. Previous result from the basic modelling with appropriate molpdf score



Marks: 2-2-2-1-1 / Date :04:10:2024 [14:59:35]

```
<< end of ENERGY.

Comments: Can Improve
>> Summary of successfully produced models:
Filename                          molpdf
----------------------------------------
q.B99990001.pdb                 3381.60645
q.B99990002.pdb                 3438.99756
q.B99990003.pdb                 3496.30249
q.B99990004.pdb                 3412.14697
q.B99990005.pdb                 3457.91211
```

Fig 3. The modelling score after advance modelling

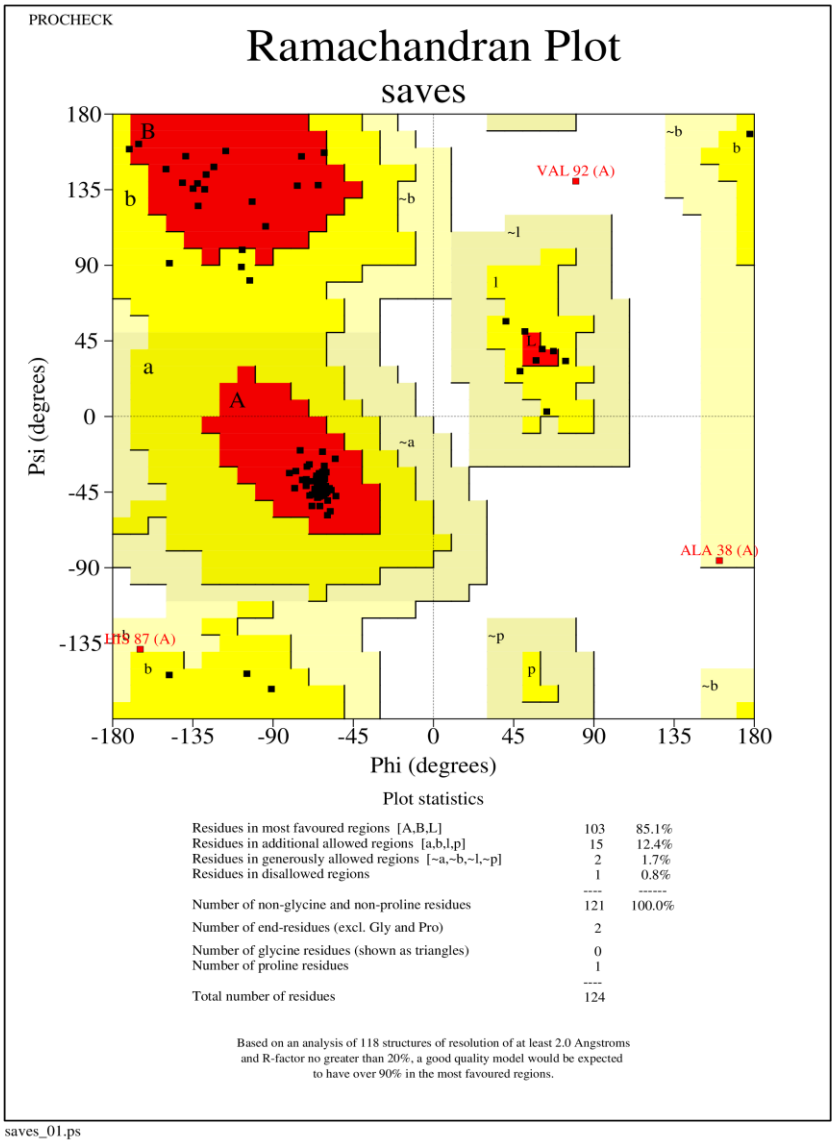


Fig 4. Ramachandran plot for the advance modelled protein – providing the insights about the stereochemical model

Comments: Can Improve

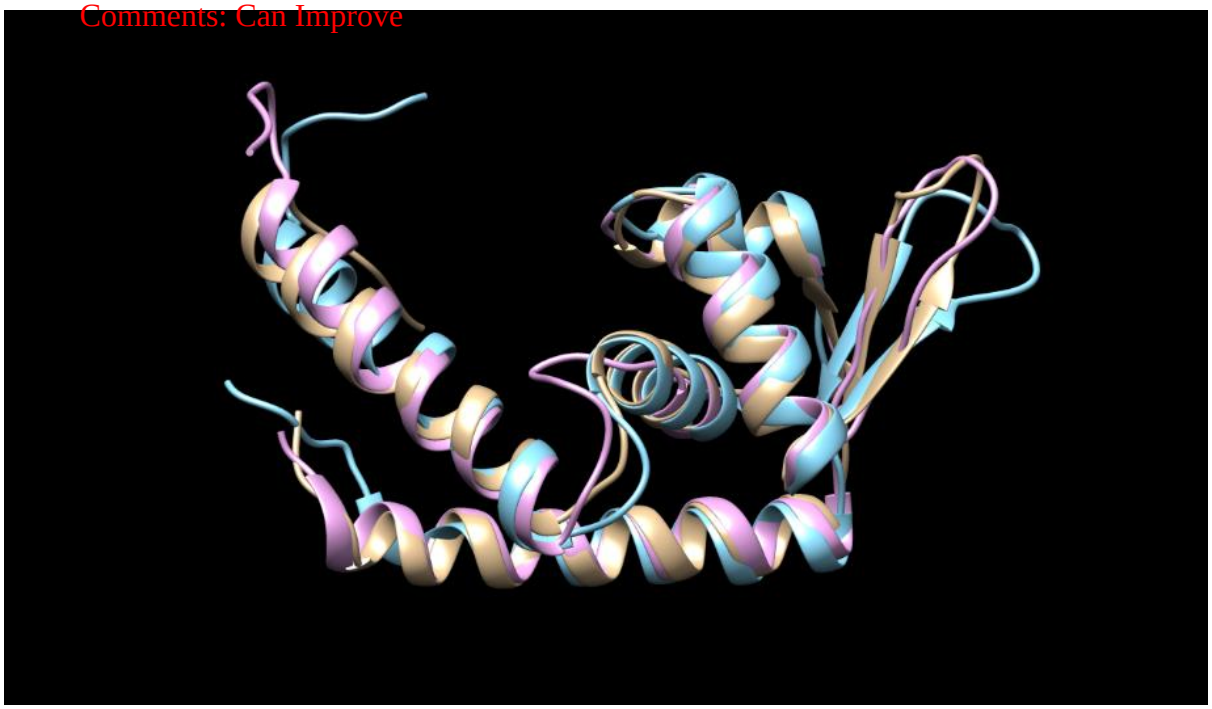


Fig 5. The modelled protein from advance modelling is aligned with the reference sequence and basic modelled protein in *chimera*

Inference

The protein Staphylococcus aureus (PDB ID: 2fnp) was modeled using both basic and advanced modeling techniques. In the basic modeling approach, 1p4x was selected as the template, and the protein sequence was modeled accordingly. In contrast, the advanced modeling process employed three templates: 2frh, 1px4, and 1hsj. The sequences of these templates were aligned using Chimera for comparative modeling.

The resulting model was visually evaluated, and Figure 4 represents the sequence alignment/match alignment of the templates. This figure provides insights into how closely the model resembles the reference structures. The Ramachandran plot highlights that VAL 92 (A), ALA 38 (A), and LEU 87 (A) are outside the most favored regions, with VAL 92 being in a disallowed region. Additionally, loop refinement is necessary to further align loop regions that were not properly aligned during the initial modeling.

Total:8

DOCKING PRELIMS

Comments: Can Improve

1. Protocol for docking

Remove non-interacting ions using chimera from the receptor protein, save it in pdb format

Convert the ligand from sdf to pdb

Adt>file> (open the receptor.pdb)
rec> edit> hydrogens> polar only
> charges> compute gasteiger

Open grid> macromolecule > choose> rece >save as pdbqt

Click ligand> open> select the ligand.pdbqt(the kollman charges would be added)

Save the ligand as pdbqt

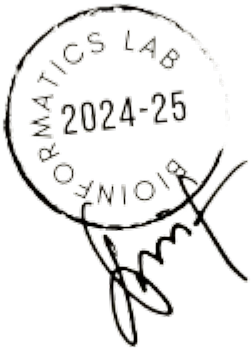
Go to castp> type your pdbcode> save the active sites of the chain

in adt> select chain A> select manually > active sites

Open grid> macromolecule > choose> rece
>set map types > choose >lig
>grid box > (cover the active sites) > file > save the close saving current

create a config file with the receptor and ligand details and mention the size and position of the grid box

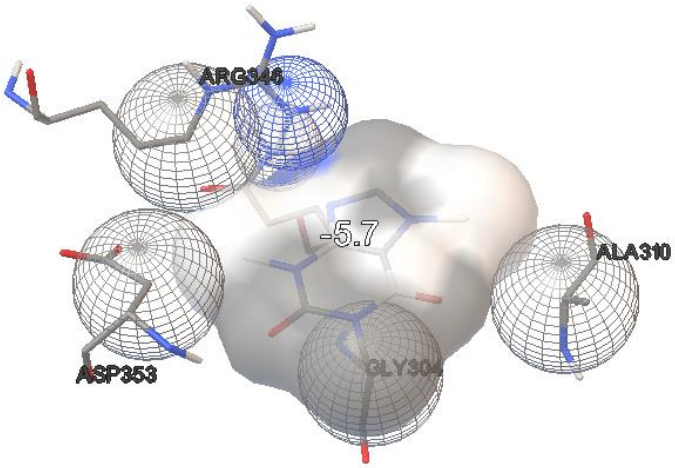
open pmv> analyse



Marks: 2-2-2-1-1 / Date :04:10:2024 [14:59:35]

2. Result from the analysis

Comments: Can Improve



The number -5.7 represents the binding affinity between the ligand and the receptor, measured in kilocalories per mole (kcal/mol). A more negative value indicates stronger binding. In this case, a value of -5.7 suggests moderate binding affinity.
Amino acids surrounding the ligand are labeled, indicating key residues involved in the interaction- ARG346, ASP353, GLY304, ALA310