Our Heritage

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

Identification of Escherichia Coli dedA Protein similarities with Different Organism Including Green Algae

Yogita Sharma [1], Shashank Rana[2], Nikunaj Bhardwaj [3], Vartika Sigh[4], Pratibha Teotia[1]

[1]Department of Biotechnology (SET), NIU, Greater Noida (UP), India
[2]Department of Microbiology, C.C.S. University Campus, Meerut (UP), India
[3]Department of Zoology, MS College, Saharanpur (UP), India
[4]Amity Institute of Global Warming & Ecological Studies, Amity University, Noida, (UP), India

Abstract

Membranous proteins organize approximately 30% of the predicted proteins encoded in all the genomes. Though, even in some fully characterized organism like *E. coli*, the roles of just about half of the expected membranous proteins are not completely understood. By the usage of genetic applications in isolating some new mutants, research laboratory has developed the techniques for understanding the applications of some earlier uncharacterized as well as conserved family of genes in the lipid biosynthesis in addition to cell division, the dedA family. DedA is an inner membranous protein present in eubacteria as well as in some Archaea. Genes for these membranous proteins are found in genomes of the few green algae. Presently, there are approximately 1000 genes in online database marked as being dedA family or having amino acid identity to DedA of E. coli. Till date, no role has been allocated to the DedA protein because of the incompetence to study proteins, by means of single mutant E. coli is without any noticeable phenotypic expression.

Key words: Membranous proteins, dedA family, Archaea, phenotypic expression

Background

We isolated the dedA protein from bacteria and check its similarities with different organism especially in green algae. To characterize the drug molecule several database like NCBI, PDB

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

Tools, BLAST, and SAVES have been used. We searched NCBI site to find out the protein sequence of dedA protein from *E.coli* in figure 1.

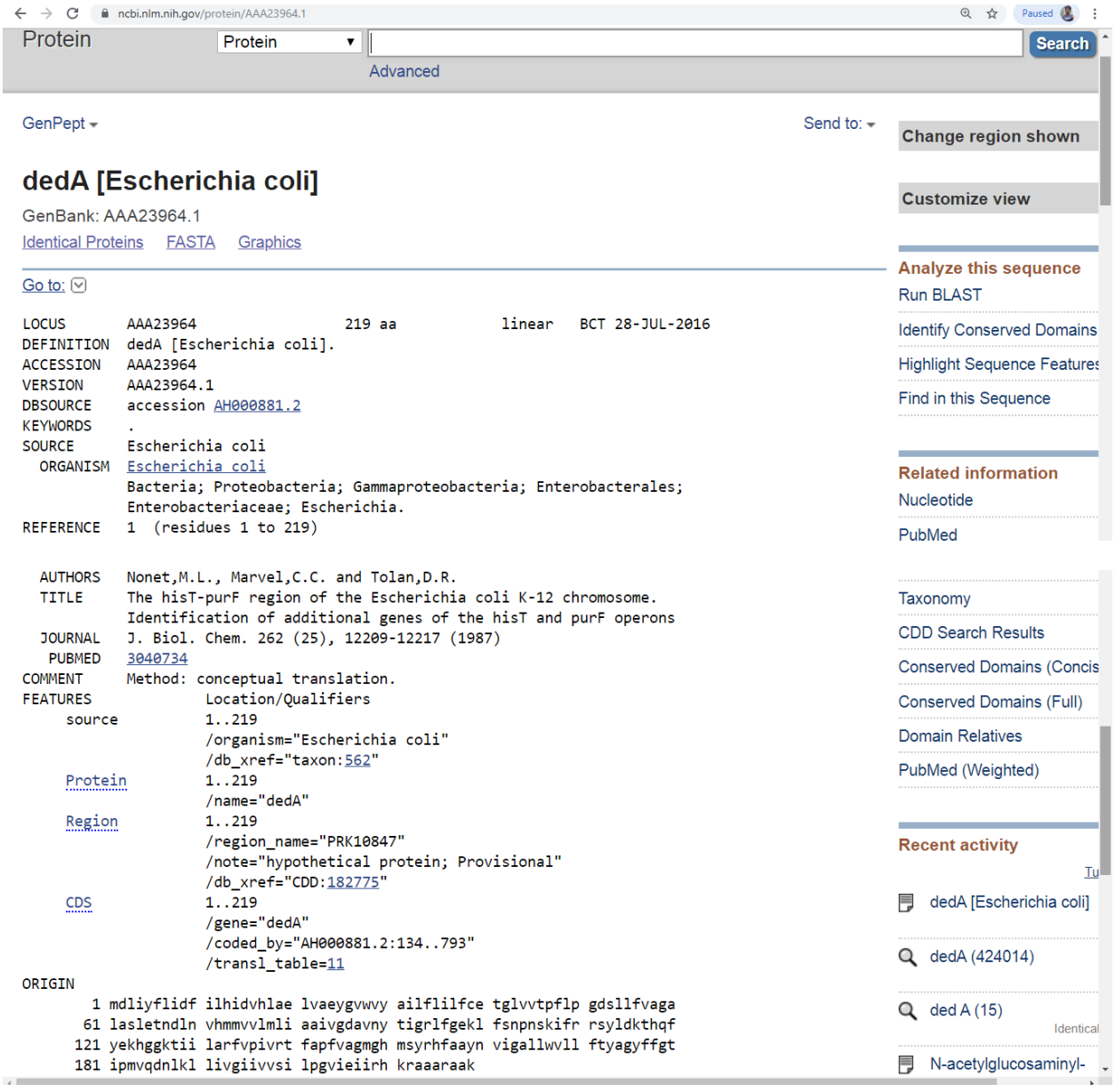


Figure1- Information of dedA [Escherichia coli] show in NCBI page

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

The dedA protein contains 219 amino acid sequences. Further with the help of BlastP we select the best similarity of dedA protein of *E.coli* to other organisms especially green plants in figure 2. We observed that the E.coli dedA protein was very similar to green algae and our best result was further confirmed by E value which was shown in figure 3. In figure 4 represent the amino acid sequence of *Micractinium conductrix.*

dedA [Escherichia coli]

GenBank: AAA23964.1

GenPept Identical Proteins Graphics

>AAA23964.1 dedA [Escherichia coli]

MDLIYFLIDFILHIDVHLAELVAEYGVWVYAILFLILFCETGLVVTPFLPGDSLLFVAGALASLETNDLN
VHMMVVLMLIAAIVGDAVNYTIGRLFGEKLFSNPNSKIFRRSYLDKTHQFYEKHGGKTIILARFVPIVRT
FAPFVAGMGHMSYRHFAAYNVIGALLWVLLFTYAGYFFGTIPMVQDNLKLLIVGIIVVSILPGVIEIIRH
KRAAARAAK

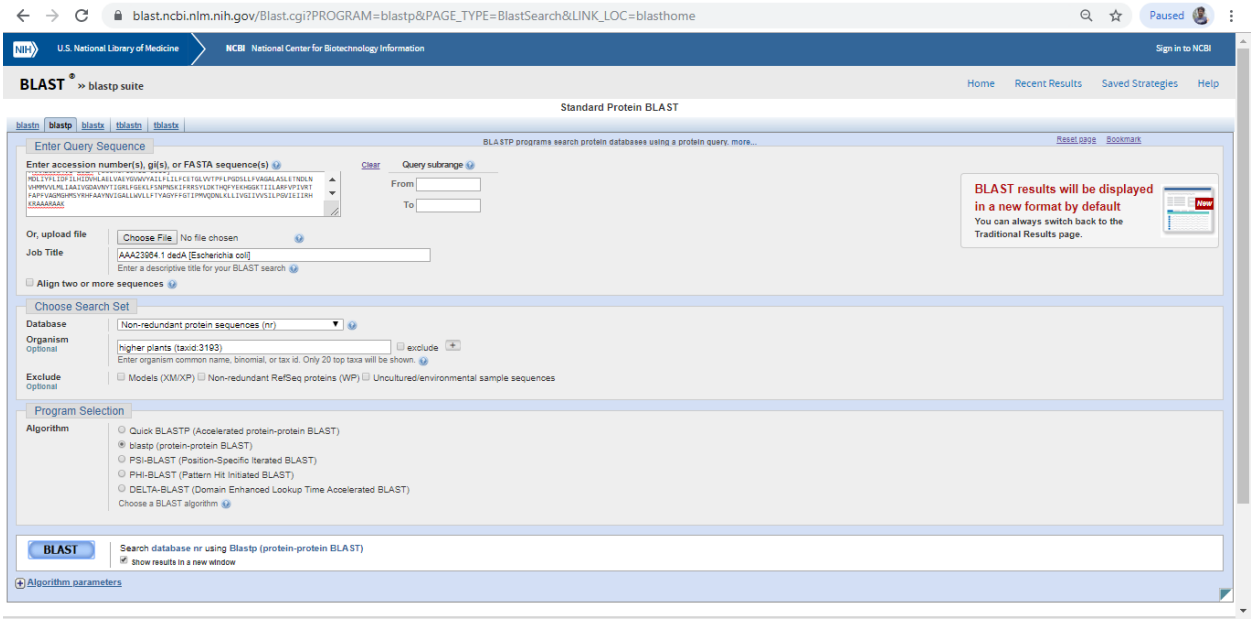
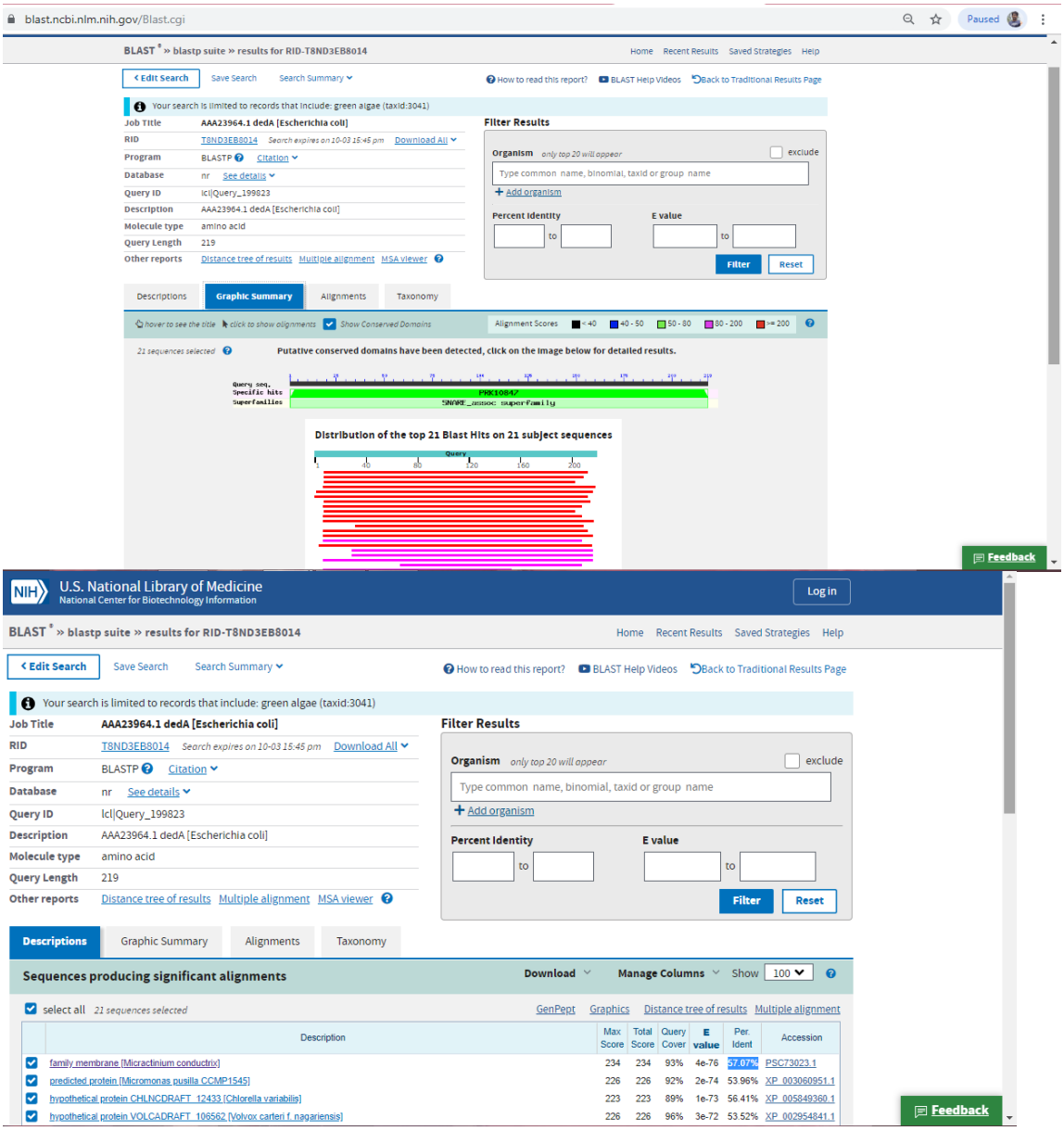


Figure 2- dedA [Escherichia coli] protein show in BLASTp with higher plant organism with non-redundant database

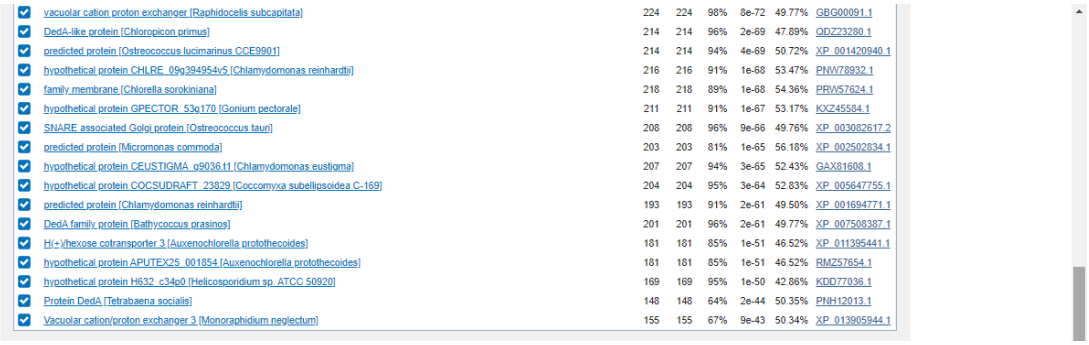
ISSN: 0474-9030

Vol-68-Issue-1-January-2020



ISSN: 0474-9030

Vol-68-Issue-1-January-2020



vacuolar cation proton exchanger [Raphidocelis subcapitata]	224	224	98%	8e-72	49.77%	GBG00091.1
DedA-like protein [Chloropicon primus]	214	214	96%	2e-69	47.89%	QDZ23280.1
predicted protein [Ostreococcus lucimarinus CCE9901]	214	214	94%	4e-69	50.72%	XP_001420940.1
hypothetical protein CHLRE_09g394954v5 [Chlamydomonas reinhardtii]	216	216	91%	1e-68	53.47%	PNW78932.1
family membrane [Chlorella sorokiniana]	218	218	89%	1e-68	54.36%	PRW57624.1
hypothetical protein GPECTOR_53g170 [Gonium pectorale]	211	211	91%	1e-67	53.17%	KXZ45584.1
SNARE associated Golgi protein [Ostreococcus tauri]	208	208	96%	9e-66	49.76%	XP_003082617.2
predicted protein [Micromonas commoda]	203	203	81%	1e-65	56.18%	XP_002502834.1
hypothetical protein CEUSTIGMA_g9036.t1 [Chlamydomonas eustigma]	207	207	94%	3e-65	52.43%	GAX81608.1
hypothetical protein COCSUDRAFT_23829 [Coccomyxa subellipsoidea C-169]	204	204	95%	3e-64	52.83%	XP_005647755.1
predicted protein [Chlamydomonas reinhardtii]	193	193	91%	2e-61	49.50%	XP_001694771.1
DedA family protein [Bathycoccus prasinos]	201	201	96%	2e-61	49.77%	XP_007508387.1
H(+)/hexose cotransporter 3 [Auxenochlorella protothecoides]	181	181	85%	1e-51	46.52%	XP_011395441.1
hypothetical protein APUTEX25_001854 [Auxenochlorella protothecoides]	181	181	85%	1e-51	46.52%	RMZ57654.1
hypothetical protein H632_c34p0 [Helicosporidium sp. ATCC 50920]	169	169	95%	1e-50	42.86%	KDD77036.1
Protein DedA [Tetrabaena socialis]	148	148	64%	2e-44	50.35%	PNH12013.1
Vacuolar cation/proton exchanger 3 [Monoraphidium neglectum]	155	155	67%	9e-43	50.34%	XP_013905944.1

Figure 3- BLASTP result show graphical summary and hit identification of dedA [Escherichia coli] protein

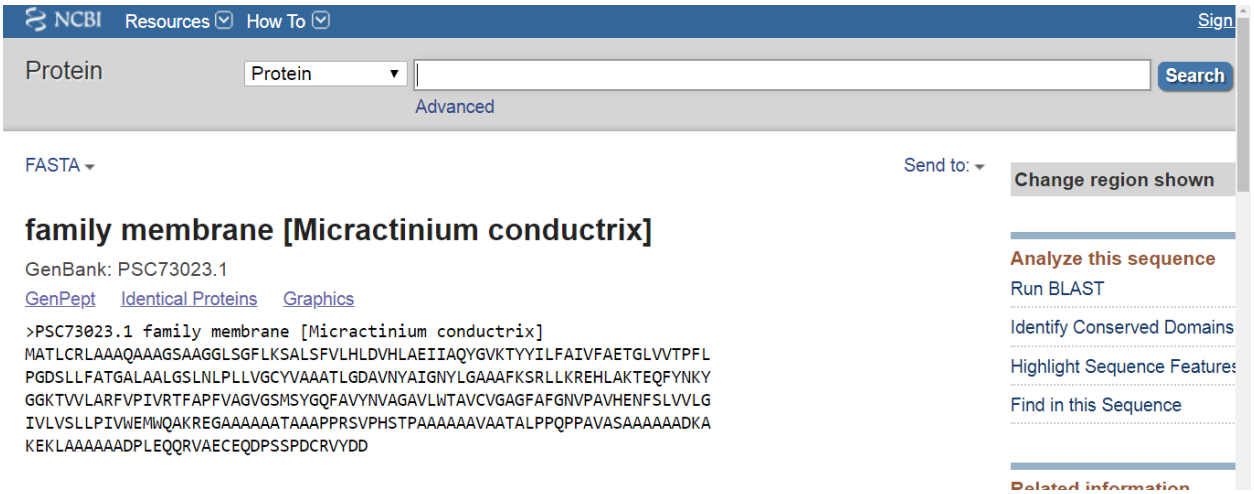


Figure 4- Amino acid sequence of *Micractinium conductrix*

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

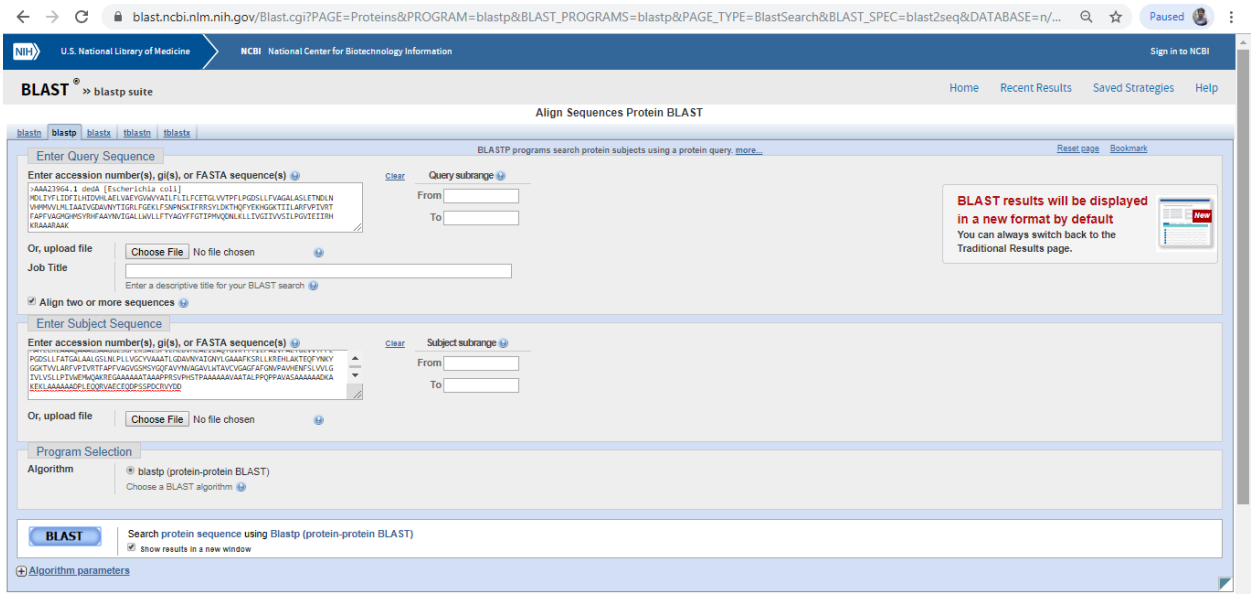


Figure 5-BLAST2 show *Micractinium conductrix* and dedA *[Escherichia coli]*

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

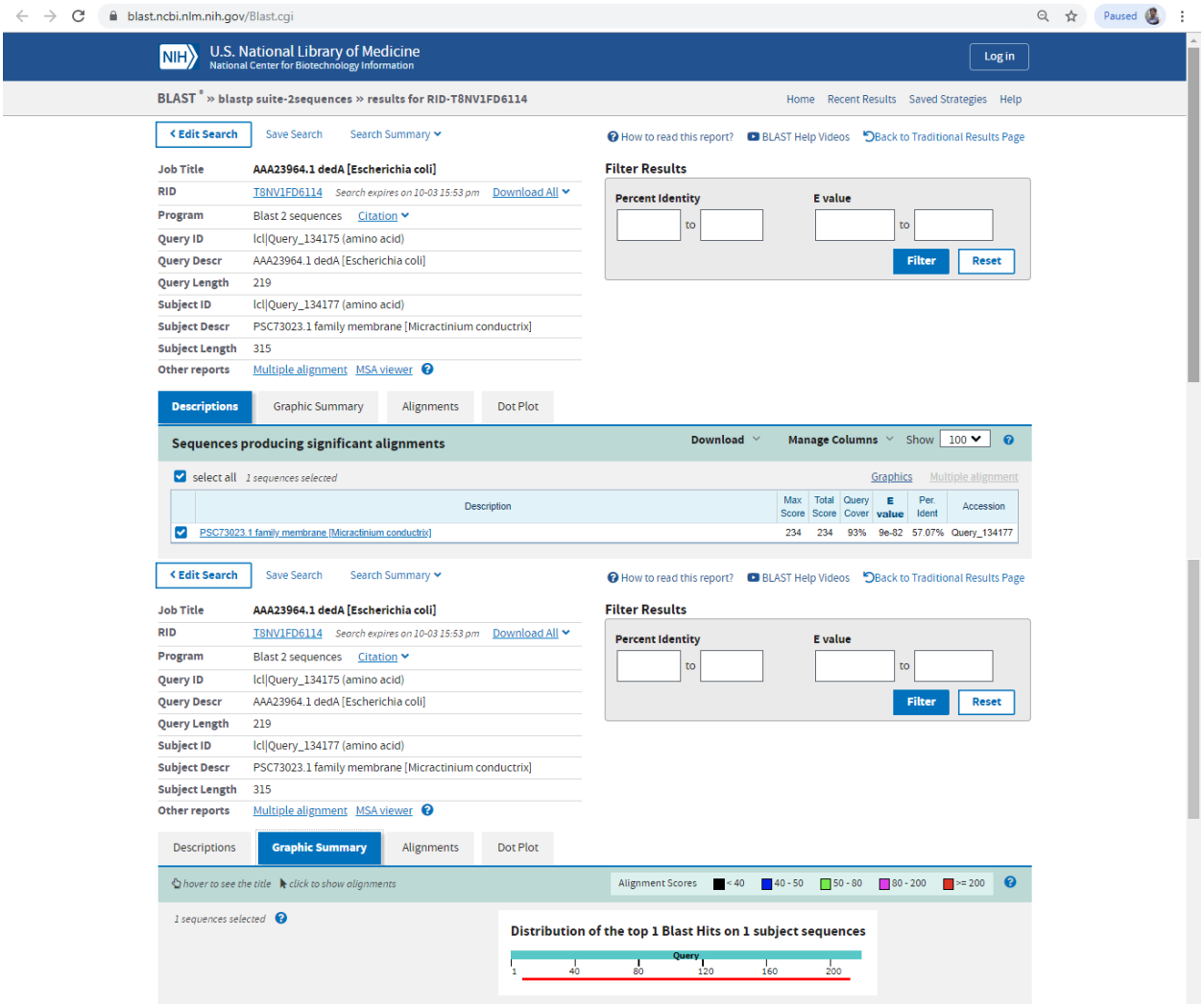


Figure 6-BLAST2 result show *Micractinium conductrix* and dedA *[Escherichia coli]* graphical summary with hit identification

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

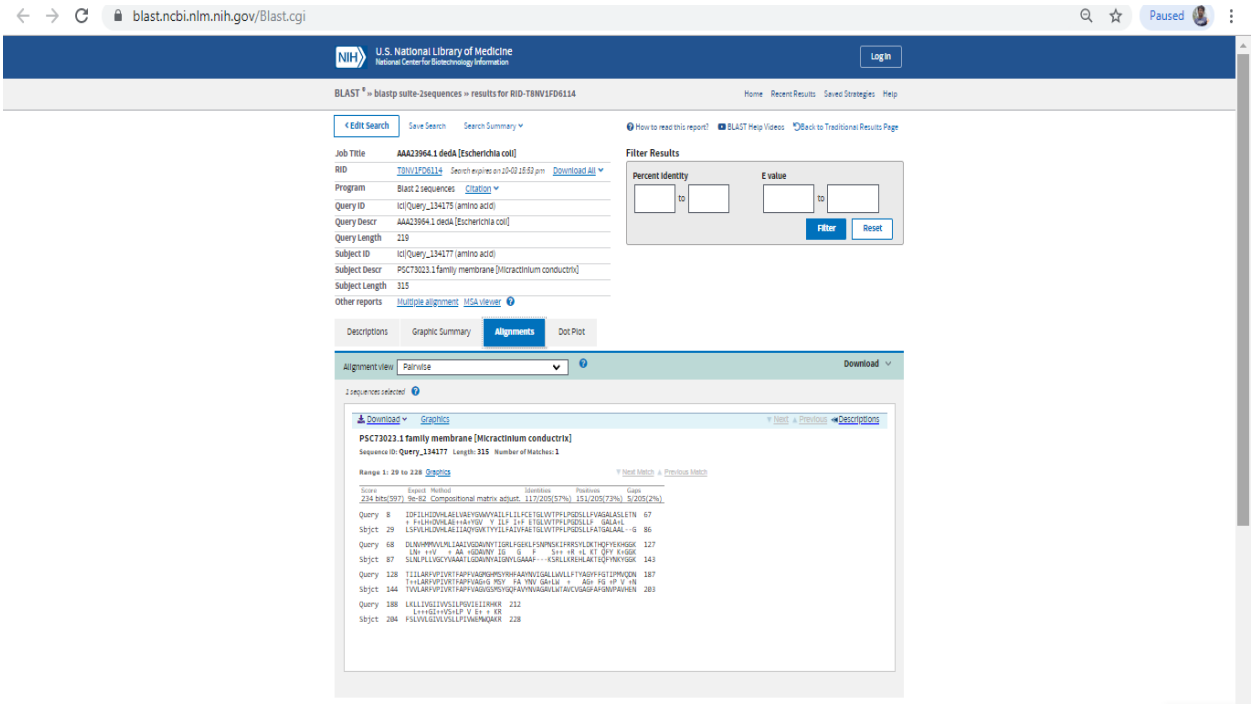


Figure 7- Alignment of Micractinium *conductrix* and dedA *[Escherichia coli] protein*

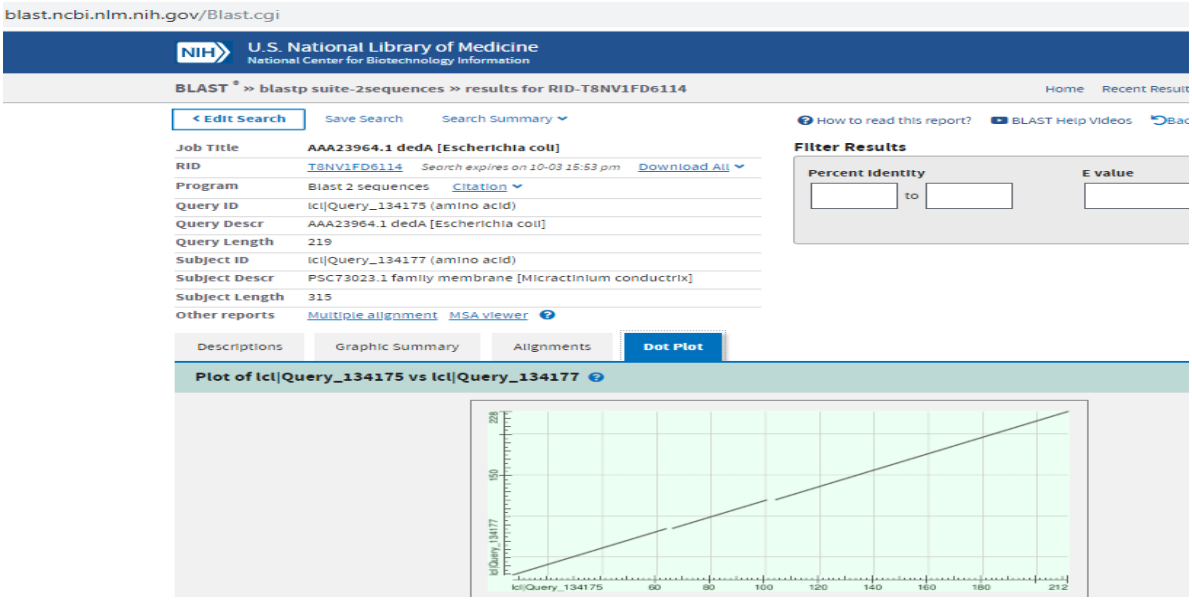


Figure 8- graphical plot of *Micractinium conductrix* and dedA *[Escherichia coli] protein*

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

In figure 5, show the next step we have followed similarity check of dedA protein of *E.coli* with *Micractinium conductrix,* a green algae which showed higher similarity after Blastp. *Micractinium conductrix has* 315 amino acid proteins. We used BLAST2 program to check the similarity of dedA protein in both the organism (The result shown in above figure). Result shows the 57.7 percent similarity in both the protein and was shown in graphical representation in figure 6. The 7 figure show alignment of Micractinium *conductrix* and dedA *[Escherichia coli] protein.*

In Figure 8, the dot matrix plot view shows regions of similarity based upon the BLAST results. The query sequence is represented on the X-axis and the numbers represent the bases/residues of the query. The subject is represented on the Y-axis and again the numbers represent the bases/residues of the subject. Alignments are shown in the plot as lines. Plus strand and protein matches are slanted from the bottom left to the upper right corner, minus strand matches are slanted from the upper left to the lower right. The number of lines shown in the plot is the same as the number of alignments found by BLAST.

National Center for Biotechnology Information (NCBI)

The NCBI houses a series of databases including GenBank for DNA sequences, NCBI Epigenomics database and PubMed, a bibliographic database for the biomedical literature. It is an important resource for bioinformatics tools and services.

BIOINORMATICS TOOL FOR PROTEIN STRUCTURE PREDICTION AND ANALYSIS

BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi)

BLAST is a service of the National Center for Biotechnology Information (NCBI). A nucleotide or protein sequence sent to the BLAST server is compared against databases at the NCBI and a summary of matches is returned to the user. The www BLAST server can be accessed through the home page of the NCBI at www.ncbi.nlm.nih.gov. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches.

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

BLAST P (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins)

It is one of the most widely used bioinformatics programs because it addresses a fundamental problem and the heuristic algorithm it uses is much faster than calculating an optimal alignment. This emphasis on speed is vital to making the algorithm practical on the huge genome database currently available; although subsequent algorithm can be even faster. It is available on the web on the NCBI website protein blast or blastp, Compares an amino acid query sequence against a protein sequence database.

BLAST2

METHODOLOGY

Collection for of 5-HT (Serotonin receptor) Protein retrieval from NCBI

Access NCBI Home Page
▼
Search *dedA* Protein from *E.coli* in NCBI
▼
Take dedA protein Sequence
▼
Search the protein database by using BLASTp
▼
Enter dedA Protein Query sequence
▼
Choose Non-Redundant Protein Database
▼
Search database Non-redundant protein sequences (nr)
▼
Choose organism higher plants (taxid: 3193)

Methodology for BLAST 2 for sequence similarity of both proteins

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

Access Home page of BLAST
▼
Choose Align two or more sequences option
▼
Enter query sequence dedA [*Escherichia coli*]
▼
Enter subject sequence
▼
Enter BLAST [show in new window]

RESULTS AND DISCUSSION

3D structure of a protein is constructing by using homology modeling. 3D structure of the *Mus musculus*5-HT Receptor was completed based on experimentally solved structural homologues. An amino acid sequence of 5-HT Receptor of *Mus musculus* was retrieved from NCBI. On the basis of BLAST result we predict template against target, in graphical summary of blast result red colour represented above 200% similarity between query and subject show best result. The BLAST analyses is was accepted on the base of *E*-value, query coverage Table I, and identities of query and subject. We choose as a template 4IARA against target 5-HT receptor protein on the bases of that result shown in below table (Table I).

Three dimensional structures find out dedA protein [E. coli] using I-TASSER server:

Prediction of 3-dimensional protein structures from amino acid sequences represents one of the most important problems in computational structural biology. I-TASSER (Iterative Threading ASSEmbly Refinement) is a hierarchical approach to protein structure and function prediction. It first identifies structural templates from the PDB by multiple threading approach LOMETS, with full-length atomic models constructed by iterative template-based fragment assembly simulations. https://zhanglab.ccmb.med.umich.edu/I-TASSER/ LOMETS (Local Meta-Threading Server) is meta-threading method for template-based protein structure prediction. (https://zhanglab.ccmb.med.umich.edu/I-TASSER/)

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

It detects structure templates from the Protein Data Bank by a technique called fold recognition (or threading). The full-length structure models are constructed by reassembling structural fragments from threading templates using replica exchange Monte Carlo simulations. I-TASSER has been extended for structure-based protein function predictions, which provides annotations on ligand binding site, gene ontology and enzyme commission by structurally matching structural models of the target protein to the known proteins in protein function databases *(*Roy A, Yang J, Zhang Y, 2012; Zhang C, Freddolino PL, Zhang Y, 2017).

Generate structure and function of dedA protein using I-TASSER

The input to the I-TASSER server is the primary amino acid sequence of the query protein (dedA ***E. coli)*** protein. When user submits an amino acid sequence, the server first tries to retrieve template proteins of similar folds (or super-secondary structures) from the PDB library by LOMETS, a locally installed meta-threading approach.

In the second step, the continuous fragments excised from the PDB templates are reassembled into full-length models by replica-exchange Monte Carlo simulations with the threading unaligned regions (mainly loops) built by ab initio modeling. In cases where no appropriate template is identified by LOMETS, I-TASSER will build the whole structures by ab initio modeling. The low free-energy states are identified by SPICKER through clustering the simulation decoys.

In the third step, the fragment assembly simulation is performed again starting from the SPICKER cluster centroids, where the spatial restrains collected from both the LOMETS templates and the PDB structures by TM-align are used to guide the simulations. The purpose of the second iteration is to remove the steric clash as well as to refine the global topology of the cluster centroids. The decoys generated in the second simulations are then clustered and the lowest energy structures are selected. The final full-atomic models are obtained by REMO which builds the atomic details from the selected I-TASSER decoys through the optimization of the hydrogen-bonding network (see Figure 1).

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

For predicting the biological function of the protein (the last column at Figure 1), the I-TASSER server matches the predicted 3D models to the proteins in 3 independent libraries which consist of proteins of known enzyme classification (EC) number, gene ontology (GO) vocabulary, and ligand-binding sites. The final results of function predictions are deduced from the consensus of top structural matches with the function scores calculated based on the confidence score of the I-TASSER structural models, the structural similarity between model and templates as evaluated by TM-score, and the sequence identity in the structurally aligned regions. https://zhanglab.ccmb.med.umich.edu/I-TASSER/about.html

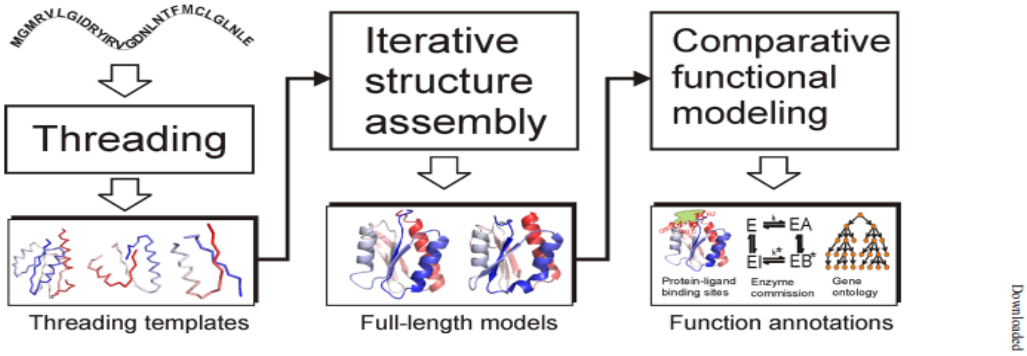


Figure 9 step wise protocol of I-TASSER server

Output

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

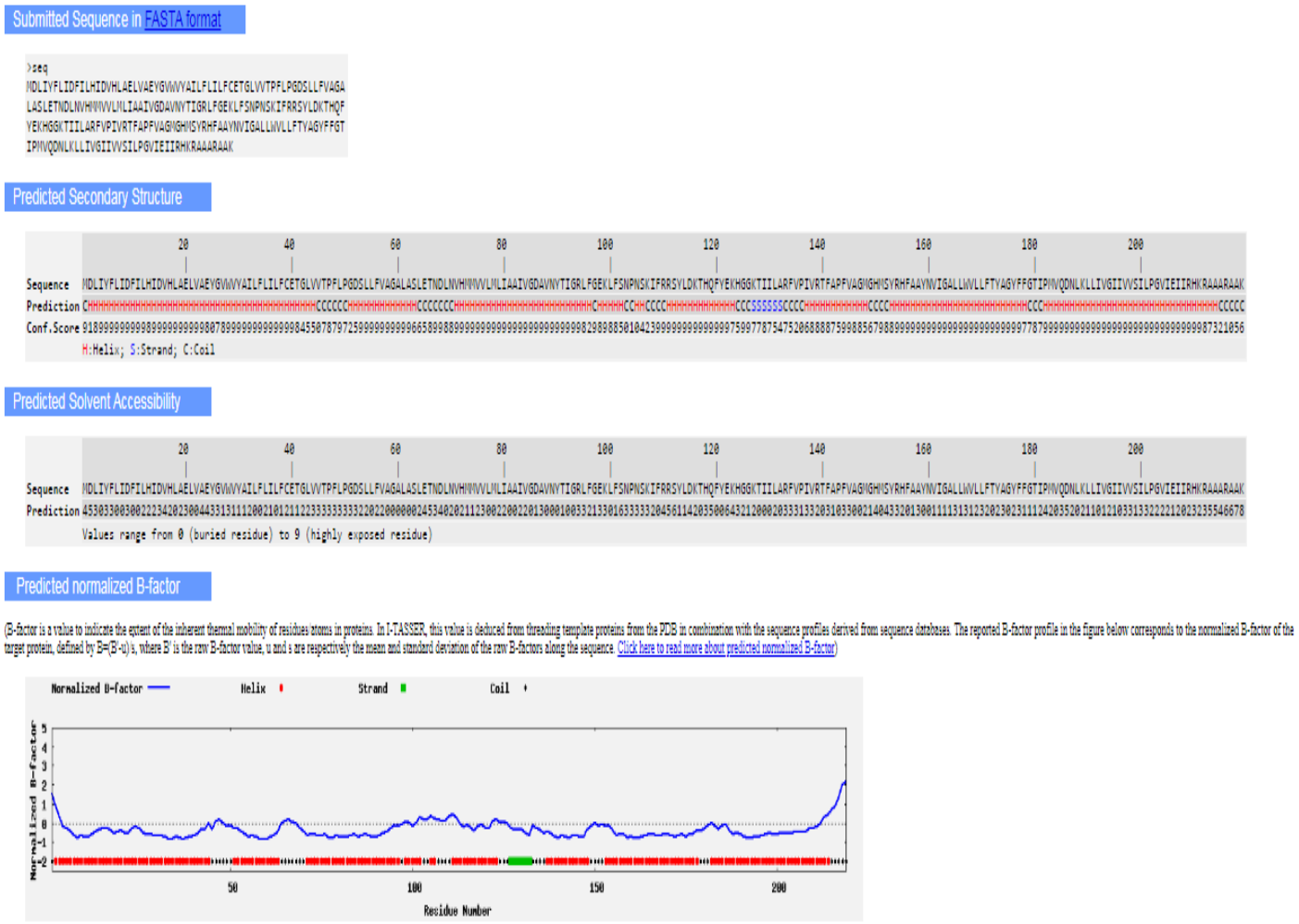


Figure 10 - Illustration of submitted sequence and predicted local structure feature in the I-TASSER server output- Amino acid sequence in FASTA format, predicted secondary structure showing Helix, Strand and Coils, predicted solvent accessibility, and predicted normalized B-factor.

For each submission, one unique job ID and one URL are assigned to track its modeling status. The user will be notified by email when the modeling has completed, and the resulting data are reported on a webpage at the URL assigned. An example output page is available at: http: //zhanglab.ccmb.med.umich.edu/I-TASSER/example. The output data include: (i) a summary of

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

the submitted sequence and local structural feature prediction, (ii) the top 10 threading templates used, (iii) the top-ranked 3D structure models with global and local accuracy estimations, (iv) the top 10 proteins with similar structures to the query and structure-based function annotations on ligand-binding site, EC number and GO terms. The modeling results are kept on the server for 90 days and will be deleted after that to save disc space in our system. All the modeling results listed on the result page are collected together in a tar ball file, which is provided for download on the same page. Users are encouraged to download this file to their computer to store the results permanently. A graphical explanation of the I-TASSER output is provided at the results annotation page: http:// zhanglab.ccmb.med.umich.edu/I-TASSER/annotation.

Submitted sequence and predicted structural features.

The first four sections of the I-TASSER result page summarize the submitted amino acid sequence and the predicted local structure features including secondary structure, solvent accessibility and normalized B-factor, which are illustrated in Figure 10. In general, positive B-factor values indicate that the residues are more flexible in the structure, while negative values suggest that the residues are relatively more stable. The predicted secondary structure is also shown in the B-factor plot. Residues located in loop or tail regions tend to have higher predicted B-factor values, as they are usually less stable compared with residues located at other regular secondary structure regions. The user-specified restraints, including template alignments and secondary structure restraints, are also listed in these sections when provided.

Top 10 templates used by I-TASSER. With the current set of 14 threading programs in LOMETS (Wu, S. and Zhang, Y., 2007), up to 140 templates are used by I-TASSER to extract distance restraints. However, the top 10 templates ranked by LOMETS are the most relevant ones because they are given a higher weight in restraints collection and are used as the starting models in the low-temperature replicas in replica-exchange Monte Carlo simulations. The information of these tem-plates is listed in the fifth section of the results page, which includes: (i) the template PDB IDs, (ii) normalized threading Z-scores, (iii) coverage of alignments, (iv) sequence identities and

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

(v) alignments between the query and the tem-plates. While the Z-score corresponds to the difference be-tween the raw alignment score and the mean in units of standard deviation, a normalized Z-score is defined as the Z-score divided by the program-specific Z-score cutoff.

A normalized Z-score >1 indicates a confident alignment. The query protein is classified as an 'Easy' target if there are on average at least one template per threading program having the normalized Z-score >1; otherwise, it is considered a 'Hard' target.

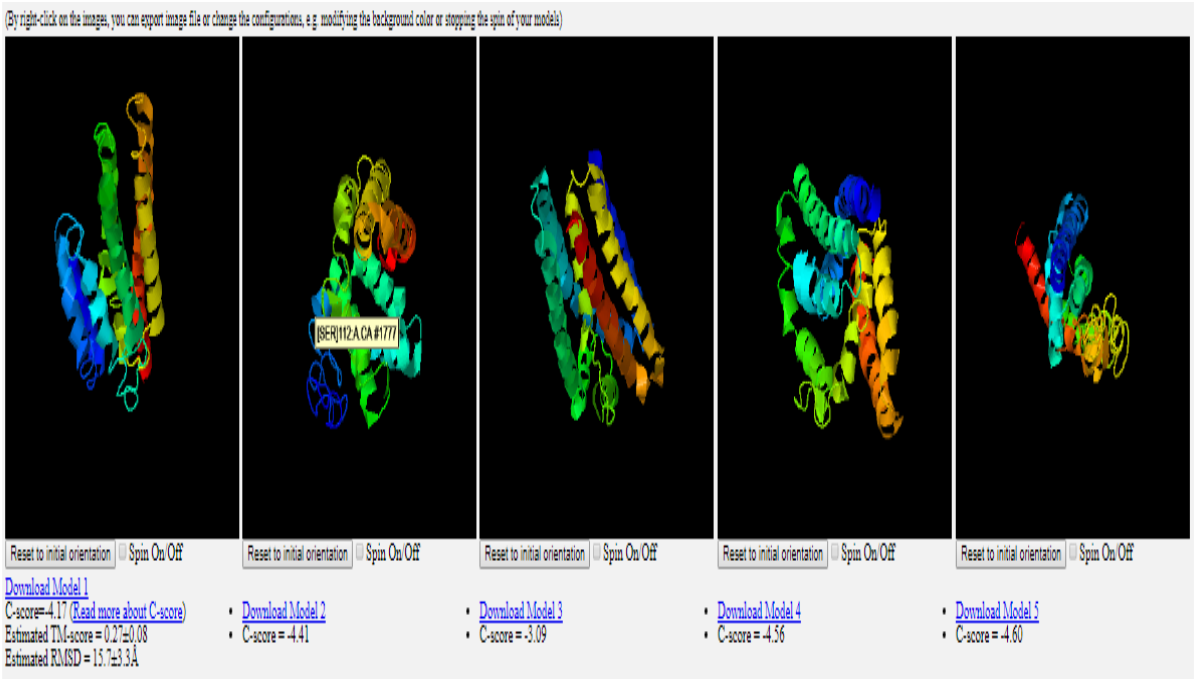


Figure 11 - Top five final model predicted by I-TASSER With estimated TM-score and RMSD value

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

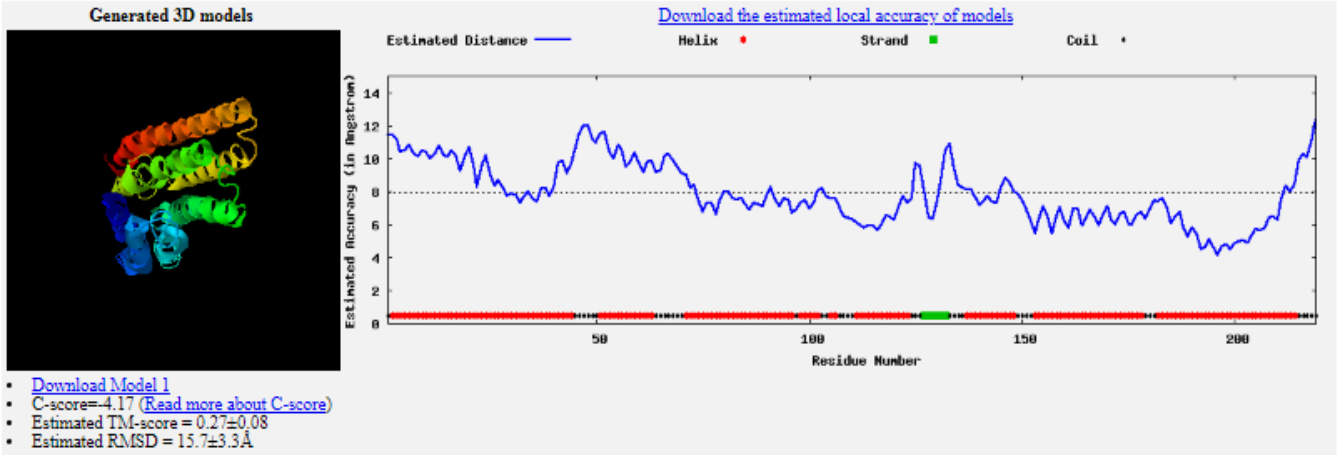


Figure 12 -Top five model predicted by I-TASSER.

An excerpt of the predicted structure model with global and local accuracy estimations
The structure is visualized in rainbow cartoon by the JSmol applet on the left panel. The estimated local accuracy is shown as a plot on the right panel, which indicates that the N-terminal and the residues between 40 and 50, 130 and 140 have relatively higher modeling error while most of other regions are accurate with estimated distance to native smaller than 2 A˚ in this example.

In figure 11, show top five models predicted by I-TASSER. Up to five full-length structural models, together with the estimated global and local accuracy, are reported in the sixth section of the result page. In the event that the modeling simulations converge, there may be less than five models reported, which is usually an indication that the models have a relatively high confidence. Figure 12, shows the first I-TASSER model of an example protein; it has a global C-score of 0.9, and the estimated TM-score and RMSD are 0.84 and 2.4 A, respectively. Users can download the PDB-formatted structure file of the model to their own computers in order to visualize the structure locally. The data file for the residue-specific local accuracy estimation and

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

the predicted B-factor values are also available for download by clicking the link 'Estimated local accuracy of models' provided on the webpage.

Protein Structure close to the target in the PDB (as identified by TM-align)

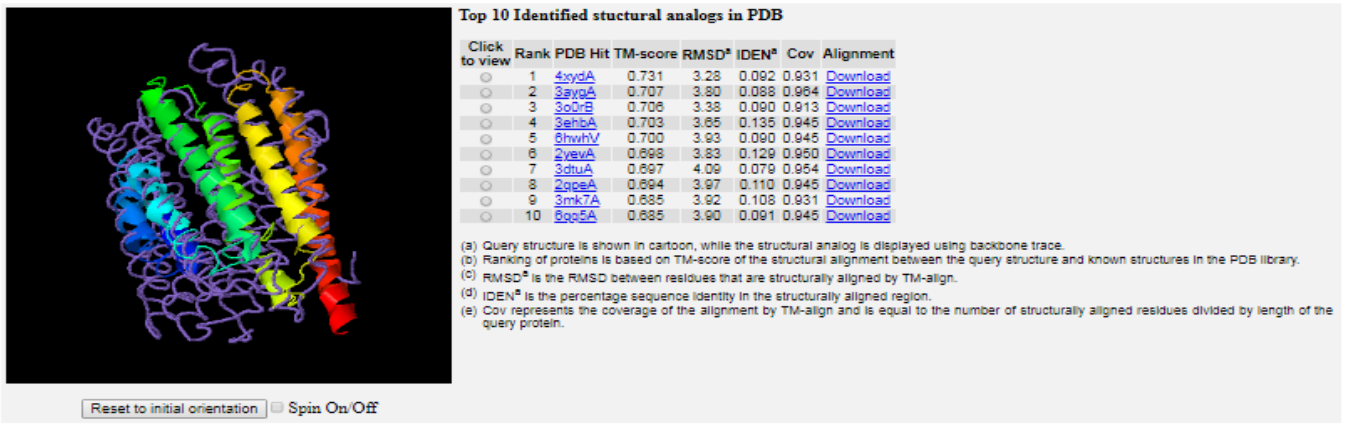


Figure 13 -The top 10 PDB proteins that are structurally close to the example protein. The query structure and the PDB proteins are shown in cartoon and backbone, respectively. Each of the structural alignments can be visualized interactively by clicking the corresponding radio buttons.

Structure analogs in PDB

The first I-TASSER model is searched against the PDB library by TM-align (Zhang, Y. and Skolnick, J., 2005) to find the analogs that are structurally similar to the query proteins. Figure 13, shows an example of the searching results. The structural alignments between the query and the 10 closest proteins are ranked by TM-score (Zhang, Y. and Skolnick, J., 2004). The table provides the numerical details of the structural alignments, including the TM-score, alignment coverage, RMSD and the sequence identity in the structurally aligned region. The links for

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

downloading the coordinate files of the superimposed structures are provided in the same table. Note that the proteins listed in this section can be different from those listed at the section 'Top 10 threading templates used by I-TASSER' because they are detected by different methods. The former is detected by structural alignment based on the first I-TASSER model, while the latter is found by threading from query sequence.

Predicted function using COFACTOR and COACH

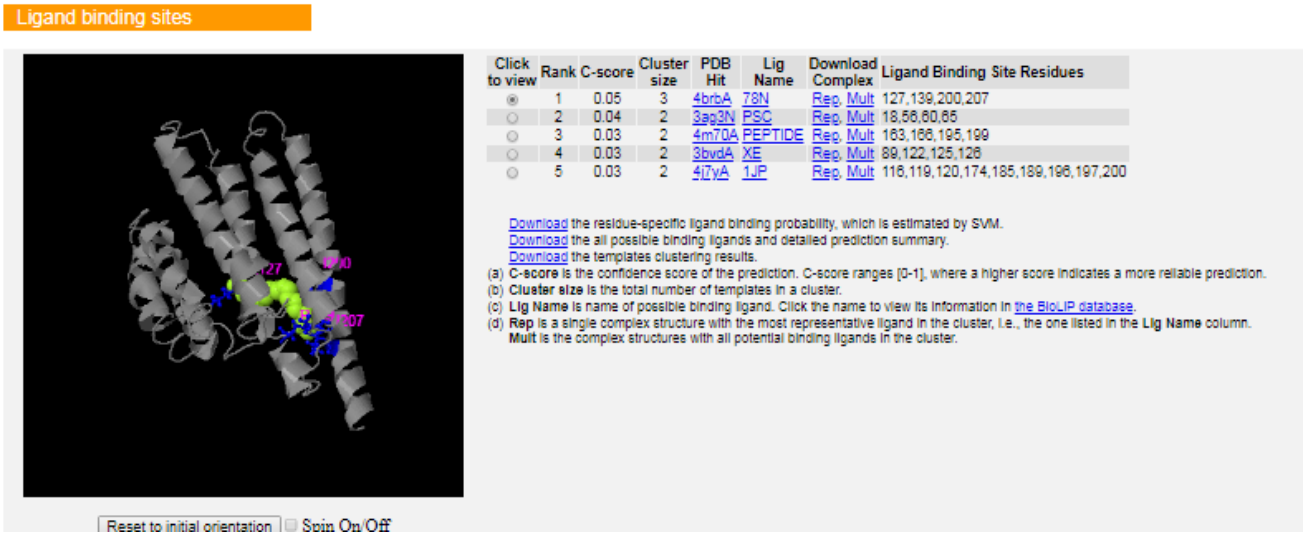


Figure 14 -Illustration of the predicted ligand-binding site, enzyme commission number and active site. The query structure is shown in gray cartoon.

The predicted ligand-binding site

The figure 14 shows, predicted binding ligands and ligand-binding residues are highlighted in yellow-green spheres and blue ball-and-sticks, respectively. For each prediction, two types of complex structures are provided for download, one containing a representative ligand (i.e. the 'Rep' link) and the other containing multiple ligands (i.e., the 'Mult' link), respectively.

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

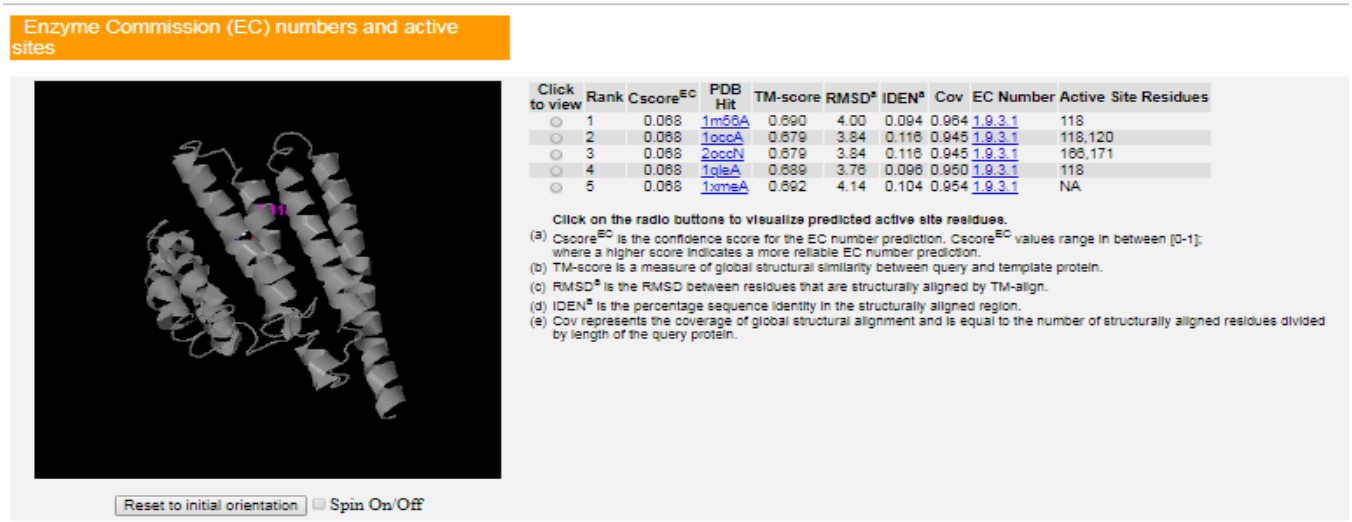


Figure 15 -The predicted EC number and active site residues are shown in colored ball-and-sticks.

Structure-based function annotation by COACH

The first I-TASSER model, which in general has the highest confidence score, is submitted to COACH (Yang, J., Roy, A. and Zhang, Y., 2013) to predict its biological function, including ligand-binding site, EC number and GO terms in figure 15. . The predicted GO terms are available in the last section of the results page, which are presented in two parts. The first part lists the top 10 ranked template proteins that are annotated with GO terms. As the template proteins may have additional functional domains, the most frequently-occurring GO terms in each of the three functional aspects (molecular function, biological process and cellular component) are reconciled from the top five homologs, with the resulting consensus GO terms presented in the second part.

CONCLUSIONS

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

On the basis of BLAST result we predict the similarity of dedA protein of *E.coli* with *Micractinium conductrix* and it represent approximately 200 similarities with dedA of *E. coli.* On the bases of that result we have predict blast 2 graphical summary on both the species. We further recommend that because both the species matches great similarity so we can use *Micractinium conductrix* as an alternative source to isolate dedA protein

References

Brylinski,M. and Skolnick,J. (2008) A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. Proc. Natl. Acad. Sci. U.S.A., 105, 129–134.

C Zhang, PL Freddolino, Y Zhang. COFACTOR: improved protein function prediction by combining structure, sequence and proteinâ€“protein interaction information. Nucleic Acids Research, 45: W291-W299, 2017.

C Zhang, PL Freddolino, Y Zhang. COFACTOR: improved protein function prediction by combining structure, sequence and proteinâ€“protein interaction information. Nucleic Acids Research, 45: W291-W299, 2017.

Capra,J.A., Laskowski,R.A., Thornton,J.M., Singh,M. and Funkhouser,T.A. (2009) Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. PLoS Comput. Biol., 5, e1000585.

https://zhanglab.ccmb.med.umich.edu/I-TASSER/about.html

https://zhanglab.ccmb.med.umich.edu/papers/2015_9.pdf

J Yang, Y Zhang. I-TASSER server: new development for protein structure and function predictions, Nucleic Acids Research, 43: W174-W181, 2015.

J Yang, Y Zhang. I-TASSER server: new development for protein structure and function predictions, Nucleic Acids Research, 43: W174-W181, 2015.

Roy A, Yang J, Zhang Y (2012). "COFACTOR: An accurate comparative algorithm for structure-based protein function annotation". Nucleic Acids Research. 40 (Web Server issue): W471–W477.

Roy,A., Yang,J. and Zhang,Y. (2012) COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res., 40, W471–W477.

Sherwood,D. and Cooper,J. (2011) Crystals, X-rays and Proteins: Comprehensive Protein Crystallography. Oxford Univ Press, London.

Wu,S. and Zhang,Y. (2007) LOMETS: a local meta-threading-server for protein structure prediction. Nucleic Acids Res., 35, 3375–3382.

Xu,D. and Zhang,Y. (2011) Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. Biophys. J., 101, 2525–2534.

ISSN: 0474-9030

Vol-68-Issue-1-January-2020

Yang,J., Roy,A. and Zhang,Y. (2013) Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment.Bioinformatics, 29, 2588–2595.

Zhang C, Freddolino PL, Zhang Y (2017). "COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information". Nucleic Acids Research. 45 (W1): W291–W299.

Zhang,J., Liang,Y. and Zhang,Y. (2011) Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. Structure, 19, 1784–1795.

Zhang,Y. and Skolnick,J. (2004) Scoring function for automated assessment of protein structure template quality. Proteins, 57, 702–710.

Zhang,Y. and Skolnick,J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res., 33, 2302–2309.