

BIOINFORMATIC STUDY OF AN ANTITUMOR PROTEIN, AZURIN

ABIRAAMI VALLI S, MYTHILI T*

Department of Plant Biology and Plant Biotechnology, Ethiraj College for Women, Egmore, Chennai - 600 008, Tamil Nadu, India.
Email:mythili22.t@gmail.com

Received: 16 October 2017, Revised and Accepted: 22 February 2018

ABSTRACT

Objective: The main objective of this study is to analyze the structure and function of an antitumor protein, azurin, thereby giving validation to the protein structure and existing physicochemical properties in the anticancer protein which are responsible for the anticancer activity.

Methods: Protein sequence analysis was done using Basic Local Alignment Search Tool (BLAST) with ten different randomly selected species of *Pseudomonas* obtained from GenBank. The physicochemical properties, prediction of secondary structure, identification of motifs and domains, three-dimensional (3-D) structure of the antitumor protein, validation through Ramachandran plot, multiple sequence alignment (MSA), and phylogenetic analysis were studied and functional property was confirmed through *in silico* docking.

Results: The similarity search (BLAST-P analysis) for the primary sequence from GenBank carried out showed 86% similarity to the second sequence, azurin (*Pseudomonas nitroreducens*). The ProtParam, ExPASy tool server indicated the presence of essential physicochemical properties in azurin. Secondary structure prediction revealed random coil, extended strand, alpha helix, and beta turn. The study on domains indicated the presence of one domain in azurin responsible for the anticancer activity. The 3-D structural analysis revealed azurin as metalloprotein, of length-128, and polymer-1 with α -helices, β -sheets, and β -barrels. The validation carried out through Ramachandran plot showed the presence of two outliers (ϕ and ψ). The biological relationship between the input sequences was studied through MSA and phylogenetic analysis. Further, azurin docked against the target protein (p53 tumor suppressor) showed the maximum binding affinity confirming its functional property of causing apoptosis.

Conclusion: All the properties analyzed in the present study revealed that the azurin protein can act as a very good anticancer agent, and through the phylogenetic analysis, it was identified that *Pseudomonas nitroreducens* was closely related to the test organism *Pseudomonas aeruginosa*.

Keywords: Azurin, *Pseudomonas aeruginosa*, p53 tumor suppressor protein, Apoptosis, Basic Local Alignment Search Tool, Multiple sequence alignment.

© 2018 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>) DOI: <http://dx.doi.org/10.22159/ajpcr.2018.v11i6.23339>

INTRODUCTION

Cancers are a large family of diseases involving abnormal cell growth with the potential to spread or invade to other parts of the body. Cancer represents one of the greatest burdens of our society and remains a serious and challenging major health problem in the future years. As a revolutionary method, investigations were carried out on the use of live bacteria and their products which were proposed a long time ago. Anticancer activity has been shown by several live, attenuated or genetically engineered organisms. However, there is a risk of using live organisms directly as it may cause certain infections. Therefore, their purified products such as proteins, enzymes, immunotoxins, antibiotics, or other secondary metabolites have been used in cancer therapy [1]. Moreover, some of these products have proven to cause significant results, such as tumor regression through growth inhibition, cell cycle arrest, or even apoptosis induction. The use of some bacterial products, especially proteins are able to target and lead to the death of tumor cells specifically, can be used in several cancer treatments [2].

Pseudomonas aeruginosa produces one such low molecular weight (M.Wt) redox protein, "azurin," one of the representative bacterial products used in the treatment of tumors [3]. The protein is a water-soluble, low M.Wt, type-I copper protein belonging to the cupredoxin family [4-6]. A number of features of azurin which includes intense blue color, a high reduction potential, and a small parallel hyperfine coupling in the electron spin resonance spectrum are mainly due to the presence of copper ion [7]. Azurin can be found in the cytosol and nuclear fractions after entering the cancer cells. Azurin binds with the tumor suppressor protein p53, forms a complex, and increases its intracellular level to raise, thereby triggering apoptosis [8].

Azurin has a significant role in a complex network which is mainly responsible for regulating the processes such as genomic stability, cell growth, and cell death. The protein is also a transcription factor which acts as a sequence-specific transcription regulator for several proapoptotic genes such as Bax and p21, involved in the inhibition of cell cycle progression and growth arrest [9]. Thus, the p53 has a fundamental role on azurin triggering apoptosis. The capability of azurin inducing apoptosis in tumor cells through p53 stabilization makes the azurin protein highly suitable for being employed as an anticancer agent. The p53 tumor suppressor is mainly involved in multiple central cellular processes including transcription, DNA repair, genomic stability, cell cycle control, and apoptosis but is functionally inactivated in several human cancers [10].

Interestingly, it was reported that azurin apart from inhibiting growth in cancer cells also has the capacity of inhibition in different pathogens such as virus, parasites (*Plasmodium falciparum*), and protozoans (*Toxoplasma gondii*). It clearly appears that azurin's ability for binding some pathogen surface proteins interferes in the entry and inhibits the growth of several pathogens and cancer cells [11]. Azurin is not only produced by *P. aeruginosa* but also by some strains of *Bordetella pertussis*, *Bordetella bronchiseptica*, and *Bordetella parapertussis* [12].

Apart from cancer studies, work has been carried out on the formation of redox protein azurin monolayers on functionalized oxygen exposing surfaces. These metalloproteins mediate the transfer of electrons in the denitrifying chain of *Pseudomonas* bacteria and exhibit self-assembly properties, clearly indicating its bioelectronic applications [13]. Azurin has also found its place in bio-sensing and bio-optoelectronics

applications. Several studies were made on the hybrid system constituted by azurin integrated with the semiconductor indium tin oxide (ITO) coated on quartz substrate. The system appears to be a good factor for bio-sensing and bio-optoelectronics applications, especially due to the coupling between the optical and electron transfer features of azurin with the conductive properties and optical transparency of ITO [14].

Bioinformatics is the information technology applied to molecular biology. Bioinformatics tools are used for several purposes such as predicting the secondary and tertiary structures of new sequences. It helps to detect the homology between the input sequences and existing families of sequences. The tools are used to find the diagnostic patterns, to characterize several protein families, and also to aid in suggesting oligonucleotide primers for the polymerase chain reaction. It also acts as an essential preface for molecular evolutionary analysis [15].

ExpASY server is a world-leading resource for protein information. Since less amount of *in silico* work is done in azurin, it is chosen for the present study. The genus *Pseudomonas* is well known for its genetic plasticity and metabolic versatility. In general, the species of *Pseudomonas* grow rapidly and are well known for their ability to metabolize an extensive number of substrates. In this study, *P. aeruginosa* is considered as the test organism since it is the major producer of azurin. *P. aeruginosa*, a highly versatile microorganism, has the capacity to tolerate low oxygen conditions. It can also survive with low levels of nutrients and grow in temperature ranging from 4 to 42°C.

The primary sequence was obtained from GenBank at National Center for Biotechnology Information (NCBI). Sequence analysis was done using Basic Local Alignment Search Tool (BLAST), and the following ten different species of *Pseudomonas* were randomly selected - *Pseudomonas aeruginosa*, *Pseudomonas nitroreducens*, *Pseudomonas denitrificans*, *Pseudomonas chlororaphis*, *Pseudomonas syringae*, *Pseudomonas fluorescens*, *Pseudomonas tolaasii*, *Pseudomonas veronii*, *Pseudomonas batumici*, and *Pseudomonas extremaustralis*.

The main objective of the study is to analyze the different physicochemical properties, prediction of secondary structure of the protein, identification of motifs and domains, predicting the three-dimensional (3-D) structure of the antitumor protein, validation of the protein structure through Ramachandran plot, multiple sequence alignment (MSA), and phylogenetic analysis were studied and functional property was confirmed through *in silico* docking.

The rationale and novelty is to study the antitumor protein, azurin, produced by *Pseudomonas aeruginosa*, whose antitumor properties have not been studied through *in silico* analysis. Hence, the present study emphasize the structure and function of azurin and also to analyze the closely related azurin producing organisms other than *P. aeruginosa* by using several bioinformatics tools and software.

METHODS

Sequence retrieval

The primary sequence of the hypothetical protein, azurin (Acc. No. B3EWN9.1), was retrieved from the GenBank [16] at NCBI [17].

BLAST-P analysis was performed (blast@ncbi.nlm.nih.gov). Based on similarity percentage and E-value, ten sequences of azurin protein from different organisms were selected.

Primary structure analysis

Using the primary sequence, the physicochemical properties of the protein from different species of *Pseudomonas* were analyzed with the tool ExpASY ProtParam server (<http://expasy.org/cgi-bin/protparam>). The various physicochemical properties such as number of amino acids, M.Wt, isoelectric point (pI), extinction coefficient (EC), instability index (II), aliphatic index (AI), grand average of hydropathy (GRAVY), number of negative residues (-R), number of positive residues (+R), and N-terminal end of the sequences were predicted.

Secondary structure analysis

Secondary structure analysis was done using a tool, self-optimized prediction method with alignment (SOPMA).

Identification of domains

To identify the domains of the protein, conserved domain database (CDD) was used. CDD is a protein annotation resource that consists of a collection of well-annotated MSA models for ancient domains and full-length proteins.

3-D-structural analysis

3-D structure and the various structural details of the azurin protein were obtained from Protein Data Bank (PDB) (<http://www.rcsb.org/pdb/>).

Validation of the antitumor protein azurin

Using MolProbity, validation of the azurin protein was carried through Ramachandran plot.

MSA

MSA was carried out with ten sequences of azurin using a tool, Clustal W.

Phylogenetic analysis

The ancestral analysis among the various species or taxa can be studied through Neighbor-joining (NJ) method. The cladogram was obtained, and the phylogenetic relationship among the organisms was studied.

In silico docking

In silico analysis through docking studies helps us to study the functional role of the antitumor protein. The target protein (p53 tumor suppressor) was retrieved from PDB (<http://www.rcsb.org/pdb/>). Azurin details were retrieved from the PubChem database. The antitumor protein was docked against the target protein in Mcule database. Mcule (<https://mcule.com/>) is an integrated drug discovery platform which provides drug discovery tools, high-quality compound databases, and delivery of several compounds. These components are integrated together and can be accessed through simple web interface. Docking results in interactions between the target and ligand molecules.

RESULTS

Sequence retrieval

The primary sequence of azurin protein produced by *P. aeruginosa* having 128 amino acids retrieved from GenBank, in the FASTA format, was analyzed for homology using BLAST-P.

(Primarysequence:>gi|408387573|sp|B3EWN9.1|AZUR_PSEAIRecName: full=Azurin AECVSDIQNDQMVFNTNATVVDKSKQFTVNLNLSHPGNLPKNV MGHNVVLLSTAADMQGVVTDGMASGLDKDYLLKPDSSRVIAHT KLIGSGEKDSVTFDVSKLKEGEQYMSFCTFPGHSALMKGTLTLK).

The sequences were arranged in descending order of identity percentage in the BLAST-P result page. From the sequences, ten sequences based on similarity percentage and E-value were obtained and tabulated (Table 1). E-value or expectation value shows that the alignment is not accidental or due to chance but has got biological significance. The similarity search for the primary sequence carried out with the help of BLAST tool showed 86% similarity to the second sequence, azurin (*P. nitroreducens*).

Primary structure analysis

The ProtParam, ExpASY tool server exhibited the essential physicochemical properties of the azurin protein of *P. aeruginosa*.

There are 128 amino acids in the sequence. The M.Wt was 13885.7 and the theoretical pI value (pI) was 5.72. The EC was found to be 8605/M/cm. The instability index (II) was 20.17 followed by an aliphatic index (AI) with the value of 70.78. Then, the grand average of hydropathy (GRAVY) in the protein was -0.350. The total number of negatively charged residues (-R) (Asp+Glu) was found to be 15, and the total

number of positively charged residues (+R) (Arg+Lys) was 12. Further, the N- terminal end of the sequence was found to be A (Ala) (Table 2).

Secondary structure analysis

The secondary structure of the azurin protein from *P. aeruginosa* was predicted using a tool, SOPMA. It is made up of alpha helix, extended strand, beta turn, and random coil. When analyzed, the results showed that the random coil (Cc) and extended strand (Ee) were predominantly present with 36.72%, followed by alpha helix (Hh) with 15.62% and beta turn (Tt) 10.94%.

Identification of domains

Using CDD, identification of domains was done for the azurin protein of *P. aeruginosa*. The tool is specific for different proteins. It was found that one domain, with the name Azurin_like (domain), described as azurin and similar redox proteins were present at an interval 4–127 with E-value 2.43e-76 (Fig. 1).

3-D structural analysis

The 3-D structure of the azurin, *P. aeruginosa*, was obtained from PDB (Fig. 2). The structure showed α -helices, β -sheets, and β -barrels. The crystal structure of protein was obtained from X-ray diffraction method. The various details such as azurin as metalloprotein, length-128, and polymer-1 were obtained.

Validation of the antitumor protein azurin

The validation of the antitumor protein azurin was carried out through Ramachandran plot. It clearly showed the presence of two outliers (ϕ and ψ), namely, D 13 MET (22.6, 69.2) and G 3 CYS (55.5, -8.8).

95.9% of all residues were in favored regions, and 99.9% of all residues were in allowed regions (Figs. 3-8).

MSA

MSA performed with ten sequences of azurin using Clustal W showed the highest score value of 95.27 between the sequence 2 and sequence 3 and the lowest score value of 66.41 between the sequence 1 and sequence 8.

Phylogenetic analysis

To study the evolutionary relationship among the azurin protein, a phylogenetic tree was constructed by NJ method which resulted in a cladogram (Fig. 9).

The phylogenetic analysis showed four simple clusters. The cladogram constructed showed that the sequence of azurin from *P. nitroreducens* and *P. denitrificans* was closely related among the ten sets of sequences taken for the study. The sequence of azurin from *P. aeruginosa* and *P. extremaustralis* was distantly related.

In silico docking

The target protein (p53 tumor suppressor) was retrieved from (PDB ID-3LH0). Azurin details were recovered from the PubChem database. Azurin docked against the target protein showed the binding affinity of azurin to p53 tumor suppressor protein through the docking scores. More negative values are an indication of higher binding affinity. Docking scores of azurin were -6.9, -6.6, -5.9, and -5.8 (Fig. 10).

DISCUSSION

Primary structure analysis

The various physicochemical properties were examined to find the differences between ten azurin protein sequences using ExPASy ProtParam tool.

The pI plays an important role for protein purification as it is the pH at which the protein does not migrate in an electric field. In our study, the pI value of the protein azurin was 5.72, indicated that protein was acidic in nature. The other proteins showed an acidic and alkaline range of pH. Similar studies were carried out in dicer proteins. The pI value which was <7 (pI<7) indicates that proteins were considered as acidic and >7 (pI>7) reveals that these dicer proteins were basic in character. The computed pI value of XP_003535104 and ABS32306 that are >7 (pI>7) reveals that these proteins were basic in character [18].

EC of azurin at 280 nm is ranging from 8605 to 10095/M/cm. The EC clearly indicates the amount of light absorption of proteins at a particular wavelength. It indicates the presence of Cys, Trp, and Tyr. These amino acids (Trp, Tyr, and Cys) are considered to be an important parameter in the calculation of EC of proteins [19].

The instability index is used to determine whether it will be stable in the test tube. If the index is <40, it is probably stable in the test tube. If the value is >40, it is probably not stable. The instability index of azurin protein was shown to be 20.17, and this shows that the protein is stable. The instability index value for the dicer proteins was found to be ranging from 21.90 to 47.14. The results imply XP_003553805 as stable protein [20].

The aliphatic index of azurin protein 70.78 indicated that the protein may be stable for a wide range of temperatures. The aliphatic index is a measure of the relative volume occupied by aliphatic side chain of the amino acids such as alanine, valine, leucine, and isoleucine. The aliphatic index values of DCLs proteins were ranging from 80.19 to 97.03. The very high aliphatic index of all DCLs proteins supports the view that DCL proteins may be stable for a wide range of temperatures [18].

The GRAVY value was analyzed for the protein. It is calculated as the sum of hydrophathy values of all the amino acids. A hydrophathy scale which is

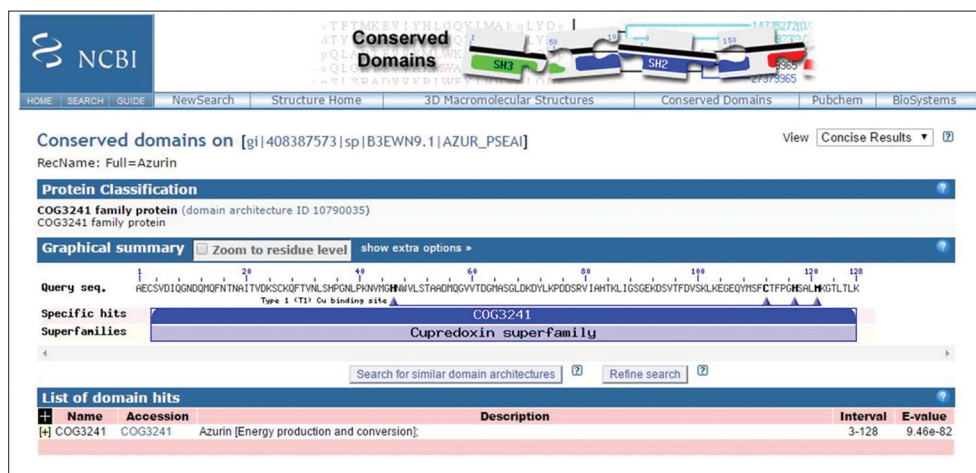


Fig. 1: Identification of domains in azurin protein

based on the hydrophobic and hydrophilic properties of the 20 amino acids is used. GRAVY values of our protein were -0.035 – 0.187 . The very low GRAVY index clearly indicated that the protein could better interact with water. The very low GRAVY index of DCLs EEE81952, XP_003553805, and XP_002268369 implies that these DCLs could result

in a better interaction with water [18]. These physicochemical properties might be responsible for the antitumor potential of the azurin protein.

Secondary structure analysis

Alpha helix (α -helix) is a common secondary structure of proteins and is a right-handed coiled or spiral conformation in which every backbone



Fig. 2: 3-D structure of azurin, *Pseudomonas aeruginosa* from protein data bank

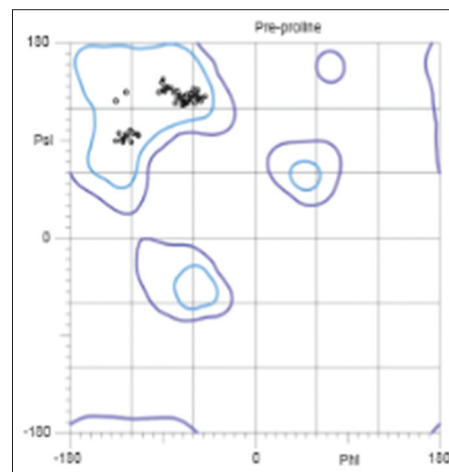


Fig. 5: Pre-proline

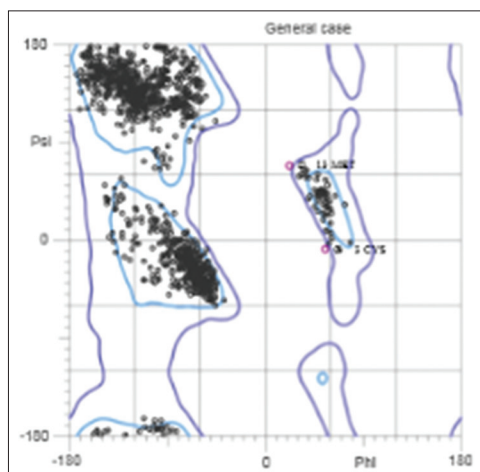


Fig. 3: General case

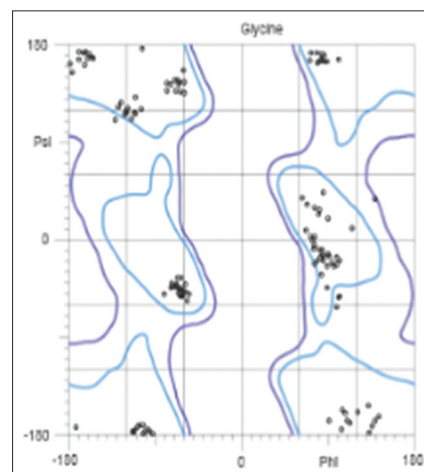


Fig. 6: Glycine

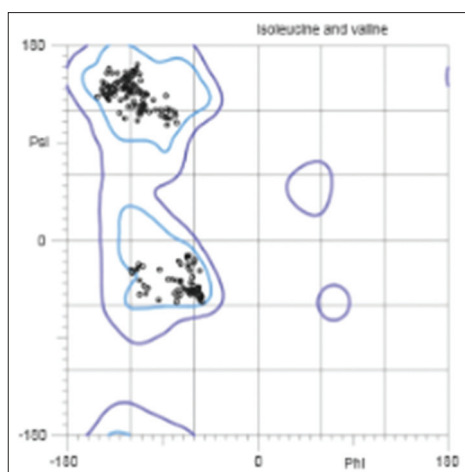


Fig. 4: Isoleucine and valine

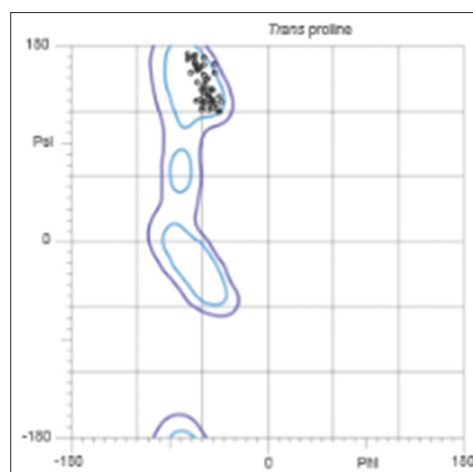


Fig. 7: Transproline

Table 1: Ten azurin protein sequences selected from BLAST-P

S. No.	Description	Max score	Total score	Query cover (%)	E value	Ident. (%)	Accession number
1.	<i>P. aeruginosa</i>	267	267	100	1e-89	100	B3EWN9.1
2.	<i>P. nitroreducens</i>	237	237	100	2e-77	86	WP_024766001.1
3.	<i>P. denitrificans</i>	236	236	100	7e-77	86	WP_015479120.1
4.	<i>P. chlororaphis</i>	213	213	100	7e-68	74	WP_023970237.1
5.	<i>P. syringae</i>	206	206	100	6e-65	72	WP_024660069.1
6.	<i>P. fluorescens</i>	204	204	100	2e-64	72	WP_016976925.1
7.	<i>P. tolaasii</i>	229	229	100	4e-74	72	WP_016972585.1
8.	<i>P. veronii</i>	229	229	100	5e-74	72	WP_017849732.1
9.	<i>P. batumici</i>	234	234	100	1e-75	72	WP_040063868.1
10.	<i>P. extremaustralis</i>	232	232	100	5e-75	74	WP_010563346.1

P. nitroreducens: *Pseudomonas nitroreducens*, *P. aeruginosa*: *Pseudomonas aeruginosa*, *P. denitrificans*: *Pseudomonas denitrificans*, *P. chlororaphis*: *Pseudomonas chlororaphis*, *P. syringae*: *Pseudomonas syringae*, *P. fluorescens*: *Pseudomonas fluorescens*, *P. tolaasii*: *Pseudomonas tolaasii*, *P. veronii*: *Pseudomonas veronii*, *P. batumici*: *Pseudomonas batumici*, *P. extremaustralis*: *Pseudomonas extremaustralis*, BLAST: Basic Local Alignment Search Tool

Table 2: Physicochemical properties computed using ProtParam tool

S. No.	Description	Accession number	Number of amino acids	M.Wt.	pI	EC	II	AI	GRAVY	-R	+R	N-terminal end of the sequence
1.	<i>P. aeruginosa</i>	B3EWN9.1	128	13885.7	5.72	8605/M/cm	20.17	70.78	-0.350	Asp+Glu-15	Arg+Lys-12	A (Ala)
2.	<i>P. nitroreducens</i>	WP_024766001.1	148	15854.1	6.89	8605/M/cm	9.60	78.38	-0.035	Asp+Glu-14	Arg+Lys-14	M (Met)
3.	<i>P. denitrificans</i>	WP_015479120.1	148	15737.1	7.73	8605/M/cm	10.35	81.08	0.070	Asp+Glu-13	Arg+Lys-14	M (Met)
4.	<i>P. chlororaphis</i>	WP_023970237.1	148	15847.2	6.04	8605/M/cm	22.98	80.41	0.016	Asp+Glu-17	Arg+Lys-15	M (Met)
5.	<i>P. syringae</i>	WP_024660069.1	148	16003.4	6.05	8605/M/cm	23.70	87.64	-0.014	Asp+Glu-18	Arg+Lys-16	M (Met)
6.	<i>P. fluorescens</i>	WP_016976925.1	148	15724.1	6.40	8605/M/cm	25.77	85.07	0.109	Asp+Glu-16	Arg+Lys-15	M (Met)
7.	<i>P. tolaasii</i>	WP_016972585.1	148	15653.0	6.39	8605/M/cm	26.28	87.03	0.187	Asp+Glu-15	Arg+Lys-14	M (Met)
8.	<i>P. veronii</i>	WP_017849732.1	148	15823.2	6.40	10095/M/cm	20.59	86.96	0.107	Asp+Glu-16	Arg+Lys-15	M (Met)
9.	<i>P. batumici</i>	WP_040063868.1	148	15725.1	6.40	8605/M/cm	22.43	87.70	0.144	Asp+Glu-16	Arg+Lys-15	M (Met)
10.	<i>P. extremaustralis</i>	WP_010563346.1	148	15797.2	6.40	10095/M/cm	19.54	84.32	0.081	Asp+Glu-15	Arg+Lys-14	M (Met)

M.Wt: Molecular weight, pI: Isoelectric point, EC: Extinction coefficient, II: Instability index, AI: Aliphatic index, GRAVY: Grand average of hydropathy, -R: Number of negative residues, +R: Number of positive residues, *P. nitroreducens*: *Pseudomonas nitroreducens*, *P. aeruginosa*: *Pseudomonas aeruginosa*, *P. denitrificans*: *Pseudomonas denitrificans*, *P. chlororaphis*: *Pseudomonas chlororaphis*, *P. syringae*: *Pseudomonas syringae*, *P. fluorescens*: *Pseudomonas fluorescens*, *P. tolaasii*: *Pseudomonas tolaasii*, *P. veronii*: *Pseudomonas veronii*, *P. batumici*: *Pseudomonas batumici*, *P. extremaustralis*: *Pseudomonas extremaustralis*

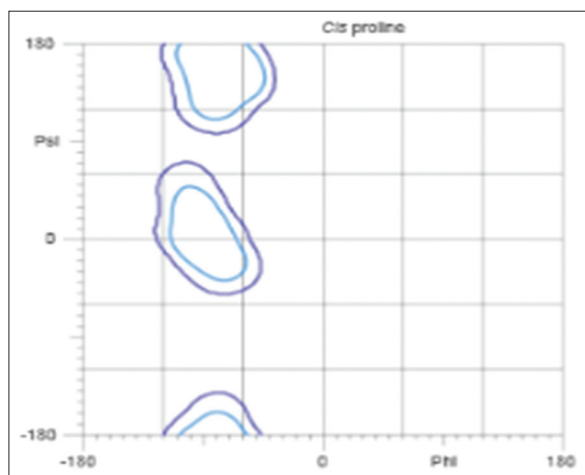


Fig. 8: Cis-proline

N-H group donates a hydrogen bond to the backbone C=O group of the amino acid four residues. Extended strand (E-strand) in protein is often associated with the formation of β -sheet [21]. Beta turns (β -turns) are very common motifs in proteins and polypeptides. Each consists of four amino acid residues. Random coil is a polymer conformation where the monomer subunits are oriented randomly while still being bonded to adjacent cells.

The secondary structure is composed of alpha helix and beta sheets, and the secondary structure is predicted using SOPMA [22]. In our analysis, the secondary structure of prediction done expressed random coil (36.67%) and α -helix (15.62%) followed by extended strand (36.72%).

Identification of domains

CDD clearly indicated the specific hit, E-value=2.43e-76 (cd13843)-Azurin and similar redox proteins, through which it clearly showed that the azurin is a bacterial blue copper-binding protein. It serves as a redox partner to enzymes such as nitrite reductase or arsenite oxidase. The copper of azurin is tetrahedrally coordinated by a cysteine, 2 histidines, and a methionine residue. The electron transfer reactions are carried out with the Cu center transitioning between the oxidized Cu (II) form and the reduced Cu (I) form. Azurin can function as tumor suppressor as it forms a complex with p53 that triggers apoptosis in various human cancer cells. Auracyanins A and B are from photosynthetic bacteria. They are very similar blue copper proteins with 38% sequence identity, and they are homologous to the bacterial redox protein azurin. However, auracyanin A is expressed only when *C. aurantiacus* cells are grown in light, whereas auracyanin B is expressed under dark and in light. Thus, auracyanin A may function as a redox partner in photosynthesis, while auracyanin B may function in aerobic respiration.

The result coincides with the domain search which was done on the BLAST site by Conserved Domain Search. It showed two domains - the medium chain reductase/dehydrogenases/zinc-dependent alcohol dehydrogenase-like family and L-ironate 5 dehydrogenase family [22].

3-D structural analysis

Proteins are one of the essential biological macromolecules performing several biological functions such as enzymatic catalysis, transport, regulation of metabolism, immune response, and nerve conduction. The 3-D structure of the protein is of great importance in studying the molecular function. Within a particular cell, a newly synthesized protein chain spontaneously folds into the compact globular structure to perform the corresponding function. Thus, there is an algorithm to fold proteins to their native structures. For the past four decades, efforts have been directed to discover nature's algorithm and several

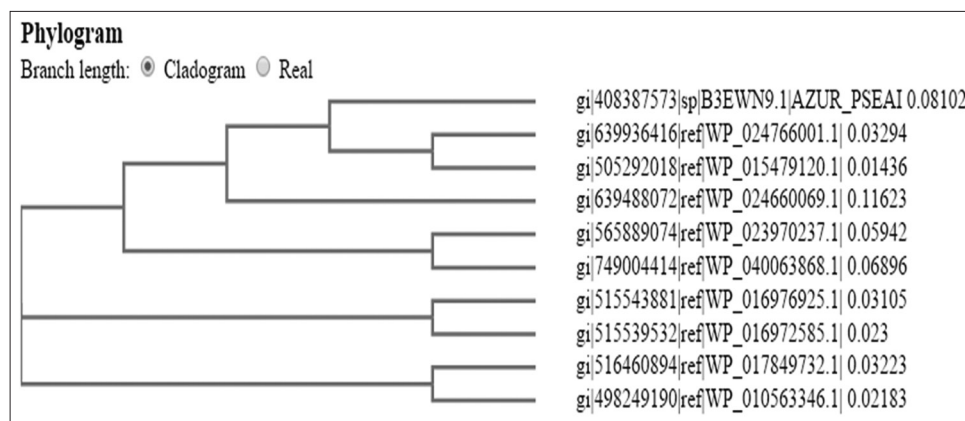


Fig. 9: Cladogram of azurin protein of ten sequences

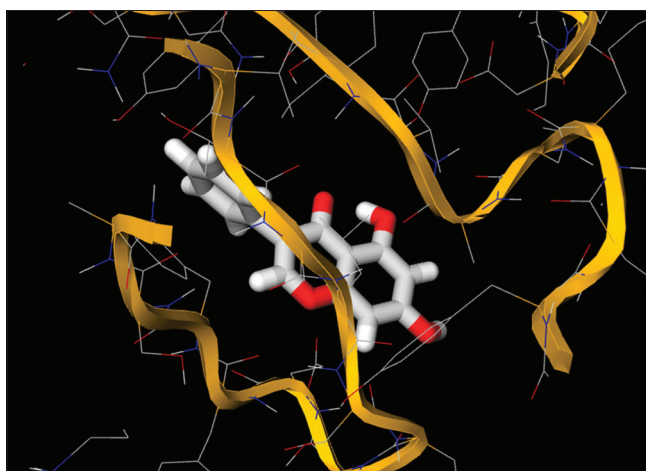


Fig. 10: Illustration of azurin docked with p53 tumor suppressor

computational methods have been developed to predict the structure of proteins from their sequences [23].

Thus, the 3-D structure of the protein can be studied through the presence of α -helices, β -sheets, and β -barrels which clearly showed that the protein has several folds which might be responsible for the antitumor activity.

Validation of the antitumor protein azurin

The three important backbone torsion angles ϕ , ψ , and ω are the main determinants of a particular protein fold. The allowed range of co-angles are very restrictive [24], and hence, the variations in this torsion angle do not give much conformational variety. This led to the creation of 2-D scatter plots ϕ and ψ pairs, comparing them to a predicted distribution [25]. These scatter plots are now commonly known as Ramachandran plots.

The major principle underlying the formation of Ramachandran map is that the atoms in a particular structure could not approach each other closer than the sum of their Van der Waals radii. However, this was turned out in the study of collagen that the belief is not strictly true and the limiting distances are to be reformulated. Using the analysis on the crystal structure data on small molecules, two sets of limiting distances of closest approach were separately obtained, which was named as normal and outer limits. Any specific conformation of a pair of linked peptide units could be generated for a given pair of phi and psi using analytical methods, and the distances between the several pairs of non-bonded atoms could be examined using these limiting distances. The conformation was then designated as fully allowed or partially allowed or disallowed depending on whether all the distances were greater than the normal limit or one or more distances were less than the outer limit (short contact) or some distances lie between the two limits [26].

Simple polymer models can be used to make a predicted distribution of pairs of ϕ and ψ angles using a particular volume exclusion model - no two non-bonded atoms can overlap in this model. Results of these calculations show a considerable conformational freedom in the two torsion angles but with certain clear restrictions. The deviations from the expected distribution for a new protein structure can now be used to judge the quality of that structure. A distinctive advantage of the ϕ and ψ distribution over many other diagnostics for structure quality is that it is very hard to improve the ϕ and ψ distribution using just structure refinement software. Thus, the Ramachandran plot is a significant indicator of the intrinsic quality of the structure. At present, instead of volume exclusion models, many modern programs to make Ramachandran plots use database statistics to create the reference distribution. A big advantage of these statistical techniques is that there are no simplifications involved and the distributions thus represent the real conformational preference of a protein chain [27].

Azurin protein clearly showed the presence of two outliers (phi and psi), namely, D 13 MET (22.6, 69.2) and G 3 CYS (55.5, -8.8). 95.9% of all residues were in favored regions and 99.9% of all residues were in allowed regions.

MSA

In bioinformatics, sequence alignment is a standard technique for visualizing the relationships between residues in a collection of related proteins. Given the amino acid sequences of a set of proteins to be compared, an alignment displays the residues for each protein on a single line, with gaps (" - ") inserted such that "equivalent" residues appear in the same column. Hence, sequence alignment is so far the most common task in bioinformatics. Several procedures relying on sequence comparison are diverse and usually range from database searches [28] to secondary structure prediction [29].

The sequence alignment of three or more biological sequences is known as the MSA [30]. Protein sequence alignment is the task of identifying evolutionarily or structurally related positions in a collection of amino acid sequences. MSA constitutes an extremely powerful means of revealing the constraints imposed by structure and function on the evolution of a protein family.

MSA was done using Clustal W. The main objective of doing MSA was to analyze how far the input azurin sequences are biologically related in which highest score value of 95.27 was found between the sequence 2 and sequence 3. This clearly indicates the biological relationship between these two sequences. Similar studies were carried out in the population study of selected protein sequences using genetic algorithm to find an enhanced algorithm for MSA of protein sequences [31].

Phylogenetic analysis

Molecular phylogenetics applies a combination of molecular and statistical techniques to study the evolutionary relationships among the

organisms or genes. The primary objective of phylogenetic studies is to mainly recover the order of evolutionary events and represent them in evolutionary trees that graphically depict relationships among species or genes over time [32].

The cladogram helps us to infer the evolutionary history of the selected sequence. A cladogram is a branching tree-like diagram in which the endpoints of the branches represent specific species of organisms. Phylogenetic tree helps us to determine the functionality of a gene and also helps to trace the origin of a gene.

The phylogenetic tree construction involves the steps:

- Collection of a particular set of orthologous sequences in a database,
- MSA of the sequences [33].

In the present study, phylogenetic analysis indicated four simple clusters. The cladogram constructed showed that the sequence of azurin from *P. nitroreducens* and *P. denitrificans* is closely related among the ten sets of sequences taken for the study. The sequence of azurin from *P. aeruginosa* and *P. extremaustralis* are distantly related. The study coincides with the studies on the construction of phylogenetic tree for butyrylcholinesterase [34].

In silico docking

Protein-protein interactions are highly essential for biochemical processes in several biological systems. The 3-D structure of the macromolecular complex resulting from the protein-protein association is a very useful source for understanding its specific functions.

Docking methodology identifies the preferred orientation of one particular molecule to the other when they bind with each other to form a stable complex [35]. It aims to predict the experimental binding modes and affinities of small molecules within the binding site of particular receptor targets and is currently used as a standard computational tool in drug design for optimization of several lead compounds and also in virtual screening studies to find novel biologically active molecules. The basic tools of a docking methodology include a search algorithm and an energy scoring function for generating and evaluating ligand poses [36]. The study is mainly done to predict compounds with prominent inhibitory or binding property [37].

Molecular docking consists of three main aspects, namely, pose prediction, virtual screening, and binding affinity estimation [38]. A standard docking methodology must be able to exactly predict the native ligand pose within the receptor binding site (i.e., to find the experimental ligand geometry within a certain tolerance limit) and the associated physical-chemical molecular interactions. Moreover, when investigating large compound libraries, the method must be able to successfully distinguish binding from non-binding molecules and to correctly rank these ligands among the best compounds in the database [39].

A search algorithm and an energy scoring function are the two main basic tools of docking for generating and evaluating the ligand conformations. The ability to handle the intrinsic molecular flexibility of a system and to exactly describe the energetics of receptor-ligand interactions is critical to the development of predictive docking methodologies that are useful in prospective drug design studies. The basic concept of docking involves the formation of protein-ligand complexes [40]. The native geometry of the complex can generally be assumed to reflect the global minimum of the binding free energy, thus indicating that docking is also regarded as an energy-optimization technique.

Protein-protein docking is mainly divided into rigid docking, flexible docking, and full flexible docking strategies. In the rigid docking, the interacting proteins are considered as rigid bodies without any conformational changes as they interact with each other; however, in the flexible docking, the conformational flexibility of the protein molecules upon the protein association is considered as an essential factor. In the rigid docking, the total conformational space is mainly represented by six variables, which consists of relative translational

vectors (x , y , and z) and rotational angles (ϕ , θ , and ψ) of the protein components. While in the flexible docking, torsional angle changes of each protein are added to the particular conformational space. In the full flexible docking, the ligand is flexed through its torsion angles as well as the side chains of active sites residues [41].

The antitumor protein azurin docked against the target protein using rigid docking method showed high binding affinity. More negative values clearly indicated the binding affinity of the azurin protein with p53 tumor suppressor indicating the functional role which may be responsible for the azurin protein for being an antitumor protein. The study coincides with the computational study for protein-protein docking using global optimization and empirical potentials [42].

CONCLUSION

The *in silico* analysis of the azurin protein revealed knowledge about the structure and function of the protein. The azurin produced by different species of *Pseudomonas* showed similarity to the test organism, *P. aeruginosa*. The study showed that the proteins are related evolutionarily and also have got similar conserved regions. The various features like presence of intense blue color, random coils, motifs, and domains make the protein to induce apoptosis through p53 stabilization. The validation of the antitumor protein indicated the outliers which may be responsible for the antitumor activity. *In silico* docking confirmed the functional role of the protein which causes apoptosis in various human cancer cells. Thus, through identification of these properties, azurin produced from the several species of *Pseudomonas* can be suggested to act against the tumor growth.

ACKNOWLEDGMENT

The authors thank Mrs. Prema Sampathkumar, Former Associate Professor and Head, Mrs. V. Manimozhi, Associate Professor and Head, the Faculty members, and supporting staff of Department of Plant Biology and Plant Biotechnology, Ethiraj College for Women (Autonomous), Chennai - 600 008, and Dr. Mrs. A. Nirmala, Principal, Ethiraj College for Women (Autonomous), for their valuable support, encouragement throughout the entire period of research.

CONFLICTS OF INTERESTS

The authors declared that they had no conflicts of interests.

REFERENCES

1. Bernardes N, Chakrabarty AM, Fialho AM. Engineering of bacterial strains and their products for cancer therapy. *Appl Microbiol Biotechnol* 2013;97:5189-99.
2. Fialho AM, Bernardes N, Chakrabarty AM. Recent patents on live bacteria and their products as potential anticancer agents. *Recent Pat Anticancer Drug Discov* 2012;7:31-55.
3. Mohamed MS, Fattah Howayada A, Mostafa M. Azurin as antitumor protein and its effect on the cancer cell lines. *Curr Res J Biol Sci* 2010;2:396-401.
4. Bernardes N, Seruca R, Chakrabarty AM, Fialho AM. Microbial-based therapy of cancer: Current progress and future prospects. *Bioeng Bugs* 2010;1:178-90.
5. Naguleswaran A, Fialho AM, Chaudhari A, Hong CS, Chakrabarty AM, Sullivan WJ Jr, et al. Azurin-like protein blocks invasion of *Toxoplasma gondii* through potential interactions with parasite surface antigen SAG1. *Antimicrob Agents Chemother* 2008;52:402-8.
6. Yamada T, Mehta RR, Lekmine F, Christov K, King ML, Majumdar D, et al. A peptide fragment of azurin induces a p53-mediated cell cycle arrest in human breast cancer cells. *Mol Cancer Ther* 2009;8:2947-58.
7. Adman ET. Copper protein structures. *Adv Protein Chem* 1991;42:145-97.
8. Yamada T, Goto M, Punj V, Zaborina O, Chen ML, Kimbara K, et al. Bacterial redox protein azurin, tumor suppressor protein p53, and regression of cancer. *Proc Natl Acad Sci U S A* 2002;99:14098-103.
9. Yamada T, Hiraoka Y, Ikehata M, Kimbara K, Avner BS, Das Gupta TK, et al. Apoptosis or growth arrest: Modulation of tumor suppressor p53's specificity by bacterial redox protein azurin. *Proc Natl Acad Sci U S A* 2004;101:4770-5.

10. Harris CC. Structure and function of the p53 tumor suppressor gene: Clues for rational cancer therapeutic strategies. *J Natl Cancer Inst* 1996;88:1442-55.
11. Chaudhari A, Fialho AM, Ratner D, Gupta P, Hong CS, Kahali S, *et al.* Azurin, *Plasmodium falciparum* malaria and HIV/AIDS: Inhibition of parasitic and viral growth by azurin. *Cell Cycle* 2006;5:1642-8.
12. Sutherland IW, Wilkinson JF. Azurin: A copper protein found in bordetella. *J Gen Microbiol* 1963;30:105-12.
13. Biasco A, Maruccio G, Visconti P, Bramanti A, Calogiuri P, Cingolani R, *et al.* Self-chemisorption of azurin on functionalized oxide surfaces for the implementation of biomolecular devices. *Mater Sci Eng* 2004;24:563-7.
14. Bizzarri AR, Andolfi L, Taranta M, Cannistraro S. Optical and electronic coupling of the redox copper azurin on ITO-coated quartz substrate. *Biosens Bioelectron* 2008;24:204-9.
15. Sundaralingam R, Kumaresan V. *Bioinformatics*. 2nd ed. Nagercoil: Saras Publication; 2012.
16. Lipman D, Ostell J, Benson D, Boguski M. Genbank. *Nucleic Acids Res* 1994;22:3441-4.
17. Pruitt KD, Tatusova T, Klimke W, Maglott DR. NCBI reference sequences: Current status, policy and new initiatives. *Nucleic Acids Res* 2009;37:D32-6.
18. Filiz E, Koc I. *In silico* analysis of dicer-like protein (DCLs) sequences from higher plant species. *IUFS J Biol* 2013;72:53-63.
19. Kumar N, Bhalla TC. *In silico* analysis of amino acid sequences in relation to specificity and physicochemical properties of some aliphatic amidases and kynurenine form amidases. *J Bioinf Seq Anal* 2011;3:116-23.
20. Guruprasad K, Reddy BV, Pandit MW. Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Eng Des Sel* 1990;4:155-64.
21. Eswar N, Ramakrishnan C, Srinivasan N. Stranded in isolation: Structural role of isolated extended strands in proteins. *Protein Eng* 2003;16:331-9.
22. Vidhya VG, Bhaskar A, Purushothaman P. *In silico* analysis and 3D modelling of SORD protein in diabetic retinopathy. *J Comput Method Mol Des* 2011;1:22-7.
23. Selvaraj S. Recent research developments in protein folding, stability and design. *Res Signpost* 2002;1:3-5.
24. MacArthur MW, Thornton JM. Deviations from planarity of the peptide bond in peptides and proteins. *J Mol Biol* 1996;264:1180-95.
25. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain conformations. *J Mol Biol* 1963;7:95-9.
26. Ramakrishnan C. Ramachandran and his map resonance. *Journal of Science Education* 2001;6:48-56.
27. Laskowski RA, MacArthur MW, Moss DS, Thornton JM- Procheck: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993;26:283-91.
28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403-10.
29. Rost B, Sander C, Schneider R. PHD- An automatic server for protein structure prediction. *Can Assoc BIOS* 1994;10:53-60.
30. Hamidi S, Naghibzadeh M, Sadri J. Protein Multiple Sequence Alignment Based on Secondary Structure Similarity. *International Conference on Advances in Computing, Communications and Informatics*; 2013. p. 1224-9.
31. Kumar M. An enhanced algorithm for multiple sequence alignment of protein sequences using genetic algorithm. *EXCLI J* 2015;14:1232-55.
32. Dowell K. An introduction to computational methods and tools for analyzing evolutionary relationships. *Molecular Phylogenetics. Technical report*, University of Maine; 2008. p. 1-18.
33. Saitou N, Nei M. The Neighbour-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406-25.
34. Srinivas K, Rao AA, Sridhar GR, Gedela S. Methodology for phylogenetic tree construction. *J Proteomics Bioinf* 2008;1:5-11.
35. Chandran D, Pappachen LK, Prathap M, Jinsha MJ, Jilsha G. *In silico* drug design and molecular docking studies of some novel benzothiazole derivatives as anti-cancer and anti-inflammatory agents. *Int J Pharm Pharm Sci* 2014;6:203-8.
36. Guedes IA, de Magalhães CS, Dardenne LE. Receptor-ligand molecular docking. *Biophys Rev* 2014;6:75-87.
37. Priyadarshini SE, Ramalingam C, Ramesh B. Designing a novel B-Lactamase inhibitor by using QSAR and docking studies. *Int J Pharm Pharm Sci* 2014;6:516-23.
38. Jain AN, Nicholls A. Recommendations for evaluation of computational methods. *J Comput Aided Mol Des* 2008;22:133-9.
39. Kolb P, Irwin JJ. Docking screens: Right for the right reasons? *Curr Top Med Chem* 2009;9:755-70.
40. Donald AJ. Docking and scoring functions/Virtual screening. *Burger's Med Chem Drug Discov* 1998;1:281-330.
41. Mukesh B, Rakesh K. Molecular docking. *Int J Res Ayur Pharm* 2011;2:1746-51.
42. Lee K. Computational study for protein-protein docking using global optimization and empirical potentials. *Int J Mol Sci* 2008;9:65-77.