

**In Silico Structural and Functional Analysis of *Bacillus megaterium*  
Asparaginase**

A

Thesis report

Submitted in partial fulfilment of the requirements for the award of the Degree of

Master of Science

In

Biotechnology

By

**Manisha Thakur (197817)**

Under the supervision

Of

**Dr Saurabh Bansal**



**MAY-2021**

**Department of Biotechnology and Bioinformatics**

**Jaypee University of Information Technology**

**Waknaghat, Solan- 173234,**

**Himachal Pradesh, India**

## DECLARATION

I hereby declare that the work presented in the thesis report entitled “**In Silico Structural and Functional Analysis of *Bacillus megaterium* Asparaginase**” submitted for partial fulfilment of the requirements for the degree of Master of Science in Biotechnology and Bioinformatics at Jaypee University of Information Technology, Waknaghat is an authentic record of my work carried out under the supervision of **Dr Saurabh Bansal**, Assistant Professor. This work has not been submitted elsewhere for the reward of any other degree/diploma. I am fully responsible for the contents of my Project report.



**Signature of Student**

Manisha Thakur

(197817)

**Department of Biotechnology and Bioinformatics**

**Jaypee University of Information Technology, Waknaghat, Solan- 173 234, H.P.,**

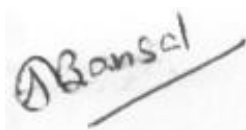
**May-2021**

## CERTIFICATE

This is to certify that the work which is being presented in the thesis report titled “**In Silico Structural and Functional Analysis of *Bacillus megaterium* Asparaginase**” in partial fulfilment of the requirements for the award of the degree of **Master of Science in Biotechnology** submitted to the **Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Wagnaghat** is an authentic record of work carried out by **Ms Manisha Thakur (197817)** during a period from July 2020 to May 2021 under my supervision at Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Wagnaghat. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

The above statement made is correct to the best of my knowledge.

Date: - 25-May-2021



**Dr Saurabh Bansal**

Assistant Professor,

Department of BT and BI

JUIT, Wagnaghat

## ACKNOWLEDGEMENT

The completion of any project depends upon cooperation, coordination, and combined efforts of several sources of knowledge. I am grateful to my project guide **Dr Saurabh Bansal**, Assistant Professor, Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, for his even willingness to give me valuable advice and direction whenever I approached him with any problem. I am thankful to him for providing immense guidance for this project.

I am also thankful to Professor Sudhir Kumar, Professor and Head, Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Solan - 173 234, HP and all the faculty members of the Department for their immense cooperation and motivation for the research of my project.



Manisha Thakur  
(197817)

## TABLE OF CONTENTS

Captions	Page No.
<b>STUDENT'S DECLARATION</b> .....	II
<b>CERTIFICATE</b> .....	III
<b>ACKNOWLEDGEMENT</b> .....	IV
<b>LIST OF FIGURES</b> .....	VIII-X
<b>LIST OF TABLES</b> .....	XI-XII
<b>LIST OF ABBREVIATIONS</b> .....	XIII
<b>ABSTRACT</b> .....	1
<b>CHAPTER- 1</b> .....	2
<b>1. INTRODUCTION</b> .....	
<b>CHAPTER- 2</b> .....	
<b>2. REVIEW OF LITERATURE</b> .....	
<b>2.1 Enzyme</b>	
<b>2.2 Asparaginase</b>	
2.2.1 Mechanism of action	
2.2.2 Source of asparaginase	
2.2.3 Production and Purification	
2.2.4 Asparaginase Importance in the medical industry and food industry	
<b>2.3 Commercial Available Source of Asparaginase</b>	
<b>2.4 Drawbacks of <i>E.coli</i> and <i>Erwinia chrysanthemi</i> Asparaginase</b>	
<b>2.5 <i>Bacillus megaterium</i></b>	
	3-34

2.5.1 Classification 2.5.2 Source of <i>B. megaterium</i> 2.5.3 Biotechnological importance of <i>B. megaterium</i> <b>2.6 In Silico Study</b> <b>2.7 Importance of My Work</b>	
<b>CHAPTER- 3.....</b> <b>3. OBJECTIVES.....</b>	35
<b>CHAPTER- 4.....</b> <b>4. MATERIALS AND METHODS.....</b> 4.1 Sequence retrieval 4.2 Sequence analysis 4.3 Multiple sequence alignment 4.4 Physicochemical characterization 4.5 Secondary structure prediction 4.6 Functional analysis 4.7 Structural classification 4.8 Homology Modeling of BmA sequence 4.9 Threading of BmA sequence 4.10 Evaluation of predicted protein	36-38
<b>CHAPTER- 5.....</b> <b>5. RESULTS AND DISCUSSION.....</b> 5.1 Sequence retrieval 5.2 Sequence analysis 5.3 Multiple sequence alignment 5.4 Physicochemical characterization 5.5 Secondary structure prediction 5.6 Functional analysis 5.7 Structural classification	39-68

5.8 Homology Modeling of BmA sequence	
5.9 Threading (by using I- TASSER) of BmA sequence	
5.10 Evaluation of predicted protein	
<b>CHAPTER- 6.....</b>	69
<b>6. CONCLUSION.....</b>	
<b>7. REFERENCES</b>	70- 84

## LIST OF FIGURES

Figure No.	Description	Page No
Figure 1	Basic reaction of the enzyme	3
Figure 2	Reactions catalyzed by LA	5
Figure 3	Ribbon (3-D) structure of <i>Helicobacter pylori</i> asparaginase	5
Figure 4	Scheme of the mechanism of L-Asparaginase	7
Figure 5	Flow chart of Production of the L-Asparaginase	12
Figure 6	Therapeutic applications of L-Asparaginase	15
Figure 7	Schematic representation of the antitumor activity of L-Asparaginase	16
Figure 8	Acrylamide formation in fried and baked foods	17
Figure 9	Stained <i>B. megaterium</i> cells.	20
Figure 10	Steps followed in the computational method for protein engineering.	23
Figure 11	UniProtKB web-based server for sequence retrieval.	25
Figure 12	BLAST User-Interface (UI) for the protein analysis.	26
Figure 13	Programmed multiple alignment by Clustal-W	26
Figure 14	ExPASy-ProtParam server for the physicochemical characterization of query protein sequence.	27
Figure 15	PSIPRED web-based server for the secondary structure prediction	28
Figure 16	Motif search tool to analyze the conserved region of the amino acids residues.	29
Figure 17	CATH server for the structural classification	29
Figure 18	SWISS-MODEL server for the protein structure prediction.	30



Figure 19	Steps followed in homology modeling.	31
Figure 20	I-TASSER server to identify structural templates from the PDB by multiple threading approach LOMETS	33
Figure 21	Phylogenetic tree of 10 Asparaginase protein sequences from <i>Bacillus</i> species.	41
Figure 22	Pie chart here represents the amino acid residues percentage in BmA.	43
Figure 23	Secondary structure map for BmA sequence.	44
Figure 24	Graphical representation of secondary structure by using SOPMA.	45
Figure 25	Secondary structure map representing the position of amino acids.	45
Figure 26	Graphical representation of amino acid position (trans-membrane, extracellular helices region)	46
Figure 27	Secondary structure map shows hydrophobic and hydrophilic regions of amino acid present in the BmA sequence.	46
Figure 28	Motif present in the <i>B. megaterium</i> asparaginase sequence	48
Figure 29	Identified motif graphical details.	48
Figure 30	Graphical representation of hydrophobicity of <i>B. megaterium</i> and their net charge density.	50
Figure 31	Sequence alignment of BmA	51
Figure 32	Superfamily Superposition structure of BmA using CATH server	52
Figure 33	Ribbon (3-D) Structure of BmA generated by CATH server	52
Figure 34	Sequence alignment of BmA with the selected template <i>Helicobacter pylori</i> asparaginase	53
Figure 35	Chemical structure of aspartic acid	54
Figure 36	Predicted protein structure of BmA based on sequence homology	54
Figure 37	Quality examination of predicted protein from different servers	55

Figure 38	Ramachandran plot generated by RAMPAGE	57
Figure 39	Steps followed by I-TASSER for protein prediction	58
Figure 40	Threading of BmA sequence for protein structure prediction.	59
Figure 41	Predicted Normalized B-Factor	60
Figure 42	Top 10 template-query alignments generated by LOMETS	61
Figure 43	Predicted 3D model and the estimated global and local accuracy.	61
Figure 44	Tertiary 3D modelled structure of asparaginase of <i>B. megaterium</i> viewed by I-TASSER	62
Figure 45	Ancestor chart of asparaginase metabolic process.	63
Figure 46	Evaluation of predicted protein from QMEAN and UCLA—DOE LAB SAVES	64
Figure 47	Ramachandran plot of Matrix protein of BmA generated using Procheck software.	66
Figure 48	Ramachandran plots for all residue types.	67
Figure 49	Chi1-chi2 plots	68

## LIST OF TABLES

<b>Table No.</b>	<b>Description</b>	<b>Page No</b>
Table No. 1	Sources of enzymes and their industrial applications	4
Table No. 2	Characteristics of the enzyme L-Asparaginase	5
Table No. 3	Development event of L-Asparaginase as anti-leukemic medication	6
Table No. 4	Sources of LA from gram-negative bacteria	8
Table No. 5	Sources of LA from gram-positive bacteria.	8-9
Table No. 6	Sources of LA from fungi.	9
Table No. 7	Sources of LA from algae	9-10
Table No.8	Sources of LA from yeast.	10
Table No. 9	Sources of LA from plant	11
Table No.10	Comparison of characteristics for SmF, SSF and SSC methods	12-13
Table No. 11	Production method, Enzyme yield, optimization temperature and pH of LA from various sources.	13
Table No. 12	Purification steps and kinetic properties of LA from various sources.	14-15
Table No. 13	Acrylamide contents of different food product	17
Table No. 14	Commercial LA used in food industry.	18
Table No.15	Commercial LA for therapeutic and pharmaceutical applications.	19
Table No.16	Side effects of <i>E.coli</i> and <i>Erwinia chrysanthemi</i> LA observed in ALL patients.	19
Table No.17	Classification of <i>B. megaterium</i> .	20
Table No.18	Industrially and pharmaceutically products produced by <i>B. megaterium</i>	21-22

Table No.19	Tools used for protein structure prediction.	24
Table No.20	Model Assessment Program	33-34
Table No.21	Details of selected sequences of LA from different <i>Bacillus</i> species.	39
Table No.22	Pairwise sequence alignment of BmA with L-asparaginases of various <i>Bacillus</i> species using BLASTp	40
Table No.23	Physicochemical property of selected <i>Bacillus</i> species sequences.	42
Table No.24	Comparison of physicochemical properties of BmA with the commercially available asparaginases; EcAII and ErA.	43
Table No.25	Calculated secondary structure elements of BmA sequence by using SOPMA.	44
Table No.26	Secondary structures of selected all sequences by using CFSSP.	47
Table No.27	Multi-level consensus sequence of motifs.	49
Table No.28	Details of the hydrophobicity of BmA using SOSUI Server.	49
Table No.29	Classification of BmA protein.	50
Table No.30	Details of template <i>H. pylori</i> asparaginase with respect to BmA.	54
Table No.31	<i>H. pylori</i> asparaginase ligands PLIP interactions with respect to BmA	54
Table No.32	Structure Assessment of predicted protein from RAMPAGE.	56
Table No.33	Consensus prediction derived based on the occurrence of the GO term among the selected templates.	63
Table No.34	Ramachandran plot- most favored, allowed, generously allowed, and disallowed regions percentage score.	65
Table No.35	All Ramachandran's, Chi1-chi2 plots and side-chain parameters of BmA generated using PROCHECK software.	66

## ABBREVIATIONS

LA	L-Asparaginase
ALL	Acute Lymphoblast Leukemia
BmA	<i>Bacillus megaterium</i> asparaginase
pI	Isoelectric Point
GRAVY	Grand Average Hydropathicity Index
MSA	Multiple Sequence Alignment
I- TASSER	Iterative threading assembly refinement
SIB	Swiss Institute of Bioinformatics
PIR	Protein Information Resource
EBI	European Bioinformatics Institute
BLAST	Basic Local Alignment Search Tool
SOPMA	Self- Optimized Prediction Method with Alignment
CFSSP	Chou and Fasman Secondary Structure Prediction
UniProt	Universal Protein Resource
PDB	Protein Data Bank
SCC	Solid State Cultivation
SSF	Solid-State Fermentation
SmF	Submerged Fermentation
PLIP	Protein-Ligand Interaction Profiler

## ABSTRACT

L-Asparaginase, which catalyzes L-asparagine, an essential amino acid for leukemic cells, into L-aspartic acid and ammonia, is used as a chemotherapeutic agent for the treatment of acute lymphoblast leukemia (ALL). Commercially available L-asparaginases from *Escherichia coli* and *Erwinia chrysanthemi* have a short half-life and several side effects on the administration to the patient. Therefore, there is a need for stable, more specific L-asparaginase with no or minimal side effects. This study aims to conduct an in silico structural and functional analysis of *B. megaterium* asparaginase (BmA) to explore its physicochemical properties using various computational tools. This study is an effort to find its potential as an alternative to commercially available asparaginases. In silico investigation of BmA suggests that the enzyme's homotetramer is thermostable due to higher aliphatic indices (i.e. 98.95). Its monomeric unit has a molecular weight of 35.15 kDa. The stability of the protein is due to the presence of a high percentage of  $\alpha$ -helices and  $\beta$  strands. In silico studies primarily suggest that BmA has all the desired properties for being used as a chemotherapeutic which further needs to be validated experimentally.

## CHAPTER- 1

### INTRODUCTION

L-asparaginase (LA) recognized as an amino-hydrolase, categories in the group of amidase [1] and involved in the hydrolysis of L-asparagine (L-Asn) to L-aspartate (L-Asp) and ammonia [2] LA has been used as an alternative for the treatment of different cancers such as acute lymphoblastic leukemia (ALL) and Hodgkin's lymphomas and other asparagine auxotrophic cancers [3]. LA is widely distributed and is present in bacteria, fungi, plants and many animal tissues [1] [4]. Microbials are known to be the most important source of LA [5] The bacterial LA has been categorized as type I and type II isozymes based on their subcellular location and properties. Out of them, type II LA showed favorable pharmacological effects in ALL [4].

Recently, this enzyme has found a novel application in the food industry to prevent the formation of a neurotoxin acrylamide (also classified as a potential carcinogen to humans) when foods are processed at high temperatures. They are potential candidates for leukemia treatment as they can effectively reduce blood plasma L-asparagine level [3]. To date, asparaginase from *Escherichia coli* (EcAII) and *Erwinia chrysanthemi* (ErA) have been used for therapy purposes.

However, these enzymes cause side effects owing to associated glutaminase activity. Furthermore, these enzymes possess low stability and a reduced half-life in the blood, necessitating multiple doses for effective treatment. It is, therefore, needed to search for novel LA to overcome the shortcomings of these commercially available asparaginases [6].

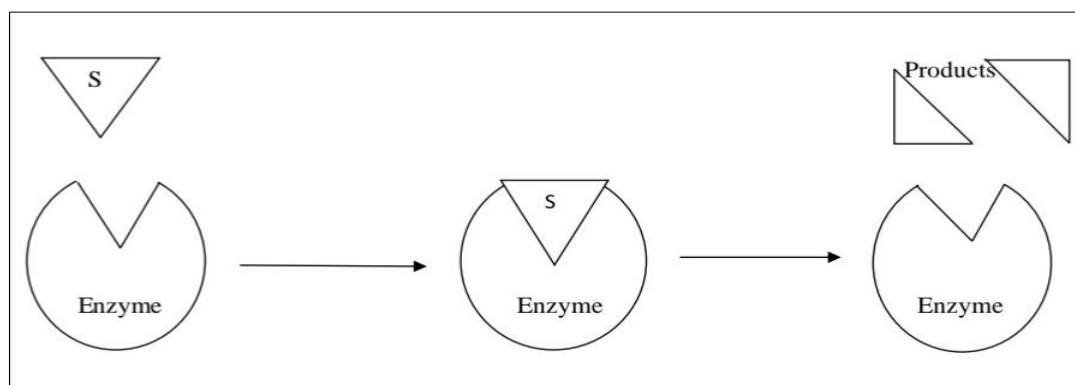
For this reason, the sequence of BmA is considered for computational analysis. *B. megaterium* is a good source of LA and a GRAS (Generally Recognized as Safe) organism. To our knowledge, there is no prior in-silico study on the properties of *B. megaterium* LA. Sequence analysis, physicochemical characterization, secondary structure prediction, functional analysis and structure classification of this BmA were studied and the protein structure based on sequence homology was then predicted. This study determines whether it will be a good alternative to ALL (acute lymphoblastic leukemia) treatment.

## CHAPTER- 2

### REVIEW OF LITERATURE

#### 2.1 Enzyme

Enzymes are protein in nature and capable of bringing about chemical and biochemical reactions within or outside of cells. They usually speed up the reaction and are highly particular to the substrate and working in mild pressure, temperature, and pH environments with high transfer rates, making them more efficient than conventional chemical catalysts. In most cases, undesirable by-products are formed which need to be removed through decontamination step to get the final products. Hydrolases are one of the enzymatic classes that stand out the most in industrial applications as they are inexpensive, highly specific and need the simple reaction conditions, and are also widely available [7], [8].



**Figure 1** Basic reaction of the enzyme [7]

Biomolecules, such as enzymes, are widely used in the foodstuff, textile commerce to make detergents, polymers and enantiomerically pure chemical intermediates (**Table 1**). The use of enzymes in industrial processes could be limited by the cost of their production, the difficulties in their recovery and disposal, or even the harsh conditions of some processes that could inactivate or denature the protein.



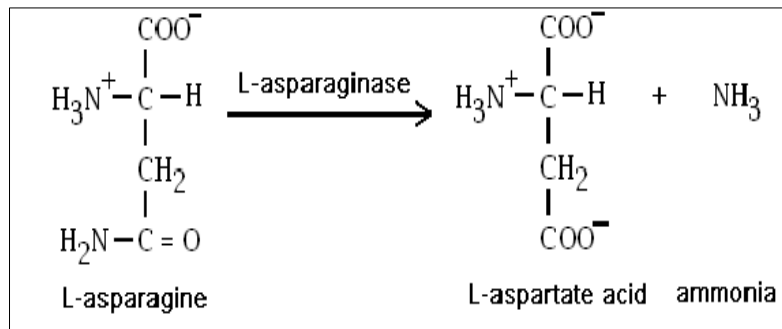
**Table 1** Sources of enzymes and their industrial applications

Enzyme source	Enzyme	Industrial application	Reference
<i>Rhizomucor miehei</i>	Alpha-amylase	Sweet-manufacturing, brewing, baking, detergent, and textile industries.	[9]
<i>Pyrococcus furiosus</i> DSM 3638	L-Asparaginase	Therapeutic and Industrial Use (in cancer therapy and prevents acrylamide formation of starchy food).	[4]
<i>Penicillium chrysogenum</i>	Lipase	Bioremediations	[10]
<i>Aeromonas sp.</i>	Chitinase	Biocontrol of phytopathogenic fungi and harmful insects.	[11]
<i>Erwinia carotovora</i> NCYC 1526	L-Asparaginase	Use in ALL and leukemia lymphosarcoma treatment.	[12]
<i>Thermotoga sp.</i>	Cellulase	Cellulose hydrolysis, polymer degradation in detergents	[13]
<i>Escherichia coli</i>	DNA polymerase	DNA Manipulation	[14]
<i>Helicobacter pylori</i>	L-Arginase	Use in the pathophysiology of various airways diseases such as asthma.	[15]
<i>Jatropha curcas L</i>	Lipase	Biodiesel	[16]
<i>Bacillus sp.</i>	Alkaline serine proteases	Food and Tannery manufacturing, medicinal formulations, cleansing agents, and some methods such as waste management, recovery of silver.	[17]
<i>Hypocrea jecorina</i>	L-Glutaminase	production of fermented foods and potent antileukemic agent	[6]
<i>Thermotoga sp.</i>	Xylanase	Pulp and paper industry	[18]
<i>Streptomyces tendae</i>	L-Asparaginase	ALL and tumour cells treatment	[19]

## 2.2 L-Asparaginase

LA recognized as an amino-hydrolase, categorized in the group of amidase [1] and involved in the L-Asn hydrolysis to L-aspartate and ammonia via the process of hydrolyzing the amide group present in the side chain of the molecule of L-Asn [2] (**Figure 2**). LA has been used to treat various kinds of leukemia and sarcomas such as ALL, Hodgkin's lymphomas,

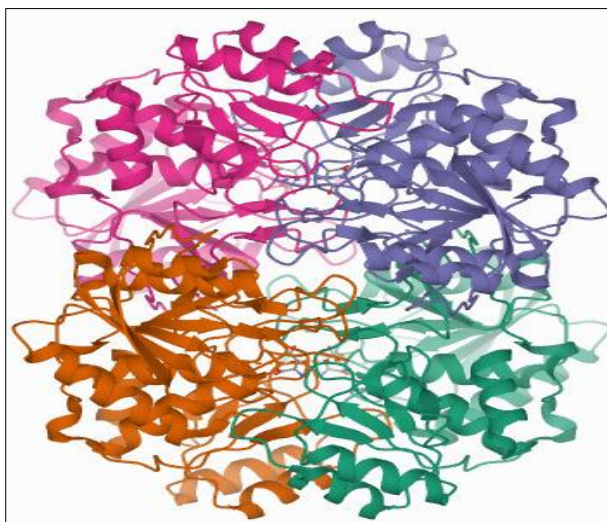
lymphosarcoma, reticulosarcoma [3]. LA is widely distributed and can be obtained from animal, plant, bacterial, algal and fungal sources [1], [4]. Bacterial LA are usually found as homo-tetramers with a molecular weight of 140–150 kDa. They are best expressed as dimers of intimate dimers (Table 2). Generally, bacterial LA monomer of 330 aa residues consist of 2 domains-N terminus and C-terminus linked through a linker (**Figure 3**). In LA, the active site is present at the inter-subunit interface between the N- and C-terminal domains of the intimate dimers [20].



**Figure 2** Reactions catalyzed by L-Asparaginase [21].

**Table 2** Characteristics of the enzyme L-Asparaginase.

- **Molecular Weight:** 140–150 kDa
- **Formula:** C<sub>1377</sub>H<sub>2208</sub>N<sub>382</sub>O<sub>442</sub>S<sub>17</sub>
- **Plasma Half-Life-** 39-49 hours (IM), 30 hrs (IV)
- **Bio Half-Life Clearance-** 0.035 ml/min/kg



**Figure 3** Ribbon (3-D) structure of *Helicobacter pylori* L-Asparaginase [130], [PMID: 19966411]

There are some limitations regarding LA (Pharmaceutical and pharmacokinetics issues). Pharmaceutical issues of LA are shelf life, thermal stability, aggregation and denaturation. Pharmacokinetics issues are low half-life immunogenicity toxicity.

**Table 3** Development event of LA as anti-tumor medication [adapted from [22]].

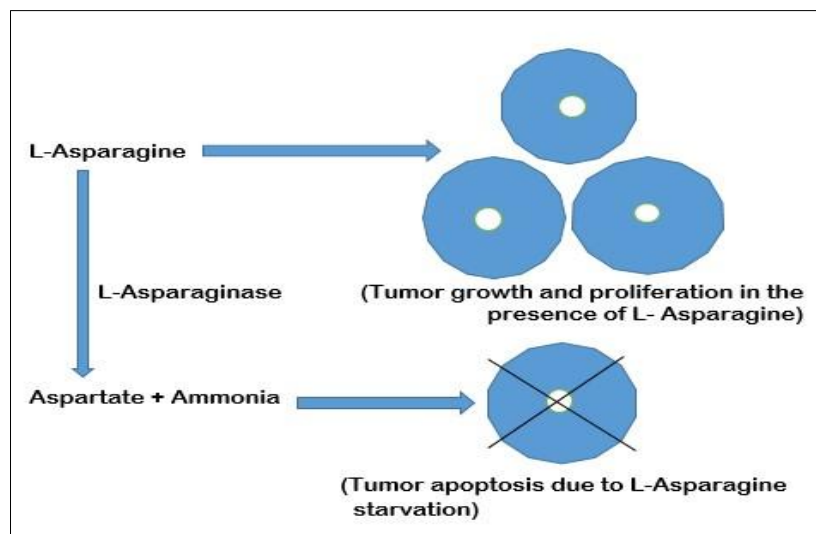
Year	Event	Scientist
1953	The sighting and development of LA as an anticancer drug began in 1953. The study described that serum from Guinea pig can slaughter resettled lymphomas in rats.	Kidd
1961	Feature accountable for the cell assassination capability in guinea pig serum is the LA activity.	Broome
1960	The earliest case of LA was reported in humans by converted LA from guinea and it is highly purified LA.	Dolowy and Hill
1963	Characterization of LA antileukemic agent in guinea pig serum	Broome
1964 - 1967	Destruction of cancer cell progress through <i>E. coli</i> - derived LA; seclusion or refinement of active <i>E. coli</i> isoform	-
1966	Major scientific use of LA as clinical tries	Dolowy
1968	Separation of LA from <i>Erwinia chrysanthemi</i>	Wade
1978	FDA approval for native <i>E. coli</i> LA for ALL treatment	-
1981	Preliminary expansion of PEG- <i>E. coli</i> - consequent LA	Kamisaki
1985	In UK, <i>Erwinia</i> LA got approval for the ALL treatment	-
1993	Identify the distinct pharmacokinetic properties	Asselin
1994	FDA gave sanction to PEG- <i>E. coli</i> LA (Oncaspar) for use in the treatment of ALL	-
2006	FDA approval to PEG- <i>E. coli</i> LA for first-line use for ALL treatment	-
2008	Start of COG ALL07P2 and compassionate use EMTP trials to evaluate LA <i>Erwinia chrysanthemi</i>	-
2011	FDA permission to <i>Er. chrysanthemi</i> LA for practice in patients with antipathy to <i>E. coli</i> - LA	-

### 2.2.1 Mechanism of L-Asparaginase

L-Asn is known to be a non-essential amino acid as the normal cells can synthesize for themselves as per their needs. However, the cancer cells are not able to synthesize L-Asn in sufficient amounts as per their demand due to a lack of or reduced activities of asparagine synthesis enzymes. As a consequence, they depend on serum L-Asn for its existence and

faster growth. L-asparaginase hydrolyzes L-Asn to L-Asp and ammonia by resulting in the unavailability of amino acids to develop tumor cubicles.

Cancer cells require a high amount of amino acids to support their proliferation rate. Enzymes targeting the amino acids essential for the proliferation of cancerous cells are one of the tools used in cancer therapy. It disrupts amino acids and disrupts the DNA replication in the cancer cells and stops the new blood vessels formation, preventing the progression of cancer and helps to kill cancer cells (**Figure 4**).



**Figure 4** Scheme of the mechanism of L-Asparaginase [adapted from [23]].

### 2.2.2 Sources of LA

The LA will be present in numerous organisms, which includes faunas, floras and microbes. The microorganisms are bacteria, fungi, yeast, algae and actinomycetes.

- **Bacteria Sources**

LA will be isolated from gram-positive and gram-negative microbial classes from the earthly and marine atmosphere. Gram-negative possesses high activity on LA. The two main microbial sources, *E. coli* and *Erwinia carotovora* are presently used for the clinical handling of ALL. Some examples of bacterial origins of LA are *Acinetobacter calcoaceticus*, *Azotobacter agilis*, *Erwinia aroideae*, *Klebsiella pneumonia*, *Thermus thermophiles*, and *B. subtilis* etc.

**Table 4** Sources of LA from gram-negative bacteria

<b>Sr. No.</b>	<b>Gram-negative bacteria</b>	<b>Reference</b>
1.	<i>Pyrococcus furiosus</i>	[4]
2.	<i>Azotobacter vinelandii</i>	[24]
3.	<i>Bacillus brevis</i>	[25]
4.	<i>Citrobacter sp.</i>	[26]
5.	<i>E. coli</i>	[27]
6.	<i>Enterbacter aerogenes</i>	[28]
7.	<i>Enterobacter cloacae</i>	[29]
8.	<i>Erwinia aroideae</i>	[30]
9.	<i>Er. Carotovara</i>	[12]
10.	<i>Er. Chrysanthemi</i>	[31]
11.	<i>Chryseobacterium taeanese</i> BCCS 016	[32]
12.	<i>Acinetobacter baumannii</i>	[33]
13.	<i>Pectobacterium carotovorum</i>	[34]
14.	<i>Pseudomonas aeruginosa 50071</i>	[35]
15.	<i>Pseudomonas fluorescens</i>	[36]

**Table 5** Sources of LA from gram-positive bacteria.

<b>Sr. No.</b>	<b>Gram-positive bacteria</b>	<b>Reference</b>
1.	<i>Bacillus circulans</i> (MTCC 8574)	[37]
2.	<i>B. coagulans</i>	[38]
3.	<i>B. firmus</i>	[39]
4.	<i>B. mesentericus</i>	[40]
5.	<i>B. velezensis</i>	[41]
6.	<i>B. subtilis</i> 168	[42]
7.	<i>B. licheniformis</i>	[43]
8.	<i>Streptomyces albidoflavus</i>	[44]
9.	<i>Corynebacterium glutamicum</i>	[45]
10.	<i>Mycobacterium bovis</i>	[46]
11.	<i>M. phlei</i>	[40]

12.	<i>Staphylococcus sp.</i>	[47]
13.	<i>S. aureus</i>	[48]
14.	<i>S. tendae</i>	[44]

- **Fungal Sources**

Fungi are another source of LA laterally with microbes. Production of LA was investigated in filamentous fungi. The molds were refined in different intermediates that contain diverse nitrogen foundations. As the fungal LA has produced extracellular, which makes its purification easier. *Alternaria sp.*, *Fusarium roseum*, *A. niger*, *Cyldrocapron obtusisporum* and *Mucor sp.* etc., are examples of LA's fungal sources.

**Table 6** Sources of LA from fungi.

Sr. No.	Fungal Source	Reference
1.	<i>Alternaria sp.</i>	[49]
2.	<i>Aspergillus nidulans</i>	[50]
3.	<i>A. niger</i>	[51]
4.	<i>A. oryzae</i>	[52]
5.	<i>A. tamari</i>	[53]
6.	<i>A. terreus</i>	[54]
7.	<i>Cyldrocapron obtusisporum</i>	[40]
8.	<i>Mucor hiemalis</i>	[55]
9.	<i>Fusarium Sp.</i>	[49]

- **Algal Sources**

Algae are another source of LA. *Spirulina maxima*, *Chlamydomonas sp.*, *Sargassum sp.*, *Sargassum binderi* and a yellow-green alga (*Vaucheria uncinata*) are examples of algal sources of LA.

**Table 7** Sources of LA from algae

Sr. No.	Algal Source	Reference
1.	<i>Spirulina maxima</i>	[56]
2.	<i>Chlamydomonas sp</i>	[57]
3.	<i>Sargassum sp.</i>	[58]

4.	<i>Sargassum binderi</i>	[59]
5.	<i>Fucus serratus</i>	[60]

- **Actinomycete Sources**

Actinomycetes are unescapable wide-reaching in loam, aquatic and wildlife, but only those that originate in existing faunas, especially in fishes, are initiated to tolerate decent enzymatic movement. Actinomycetes sources of LA are *Actinomycetes sp*, *Streptomyces albidoflavus*, *S. collinus*, *S. griseus*, and *S. aurantiacus*.

- **Yeast Sources**

LA has been discovered in eukaryotic systems in a quest to discover new potential biopharmaceuticals. The yeast *Saccharomyces cerevisiae* expressed the genetic factor ASP1 constitutively, resulting in the intracellular cytoplasmic LA. LA with fewer distressing possessions conveyed from the yeast. *Candida utilis*, *Pichia polymorpha*, *Rhodotorula sp.* and *Saccharomyces cerevisiae* are examples of yeast sources of LA.

**Table 8** Sources of LA from yeast.

Sr. No.	Yeast Source	Reference
1.	<i>Candida utilis</i>	[61]
2.	<i>C. guilliermondii</i>	[62]
3.	<i>Streptomyces albidoflavus</i>	[44]
4.	<i>Pichia polymorpha</i>	[63]
5.	<i>Rhodospiridium toruloides</i>	[64]
6.	<i>Rhodotorula sp.</i>	[65]
7.	<i>Saccharomyces cerevisiae</i>	[66]
8.	<i>P. pastoris</i>	[66]

- **Plant Sources**

Plants are also a rich source of LA. Plant sources of LA are *Ocimum tenuiflorum*, *Pisum sativum*, *Withania somnifera*, *Asparagus racemosus*, and *Arabidopsis thaliana* etc.

**Table 9** Sources of LA from plant

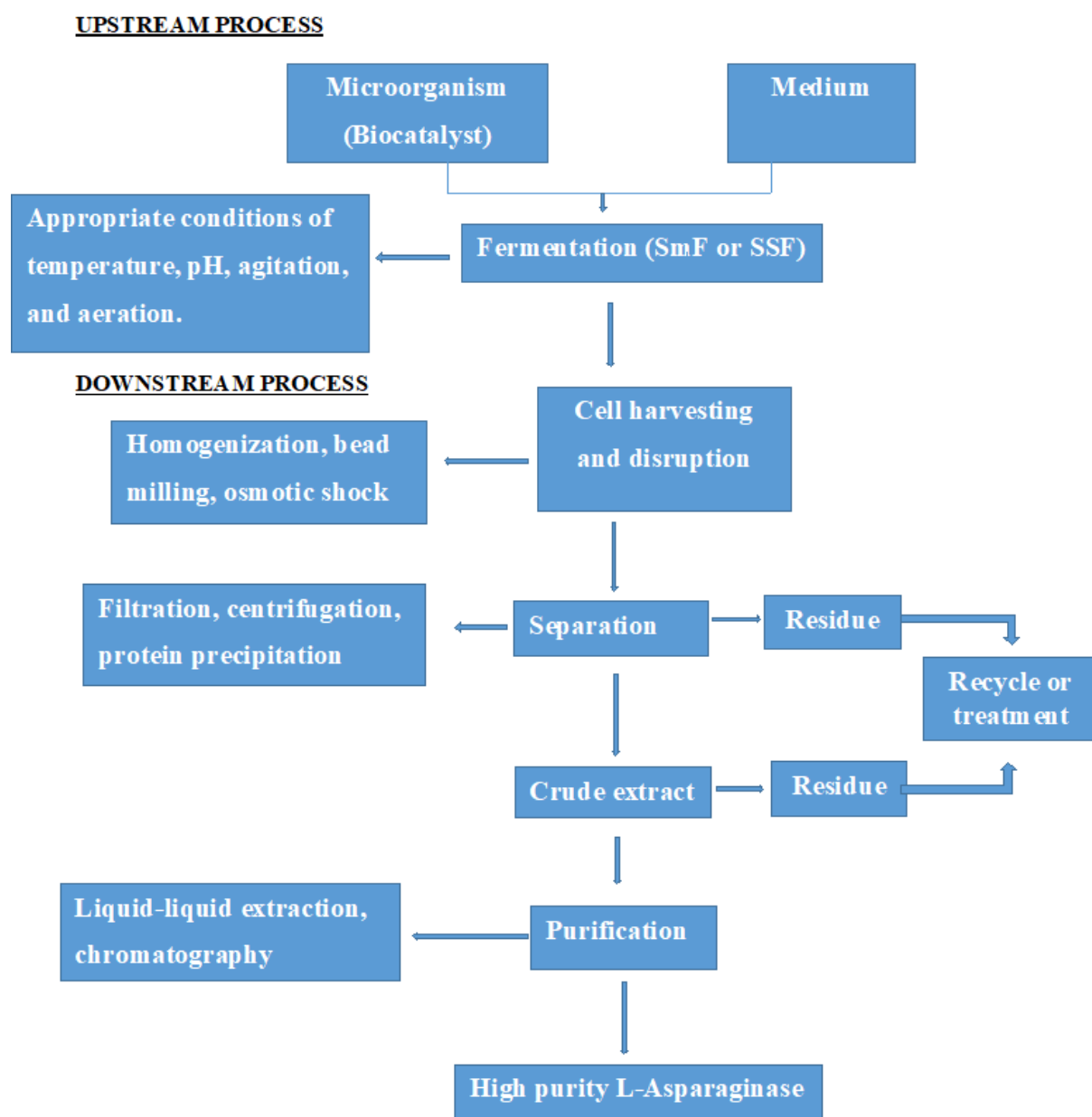
Sr. No.	Common Name (Plant source)	Scientific Name (Plant source)	Reference
1.	Holy basil	<i>Ocimum tenuiflorum</i>	[67]
2.	Green chillies	<i>Capsicum annum L</i>	[68]
3.	Tamarind	<i>Tamarindus Indica</i>	[68]
4.	Pea	<i>Pisum sativum</i>	[69]
5.	Lupin	<i>Lupin arboreus</i>	[70]
6.	Red chillies	<i>Capsicum annum L</i>	[71]
7.	Ashwagandha	<i>Withania somnifera L.</i>	[72]
8.	Bean	<i>Phaseolus vulgaris</i>	[55]
9.	Cowpea	<i>Vigna unguiculata</i>	[73]
10.	Soybean	<i>Glycine max</i>	[74]
11.	Mouse ear cress	<i>Arabidopsis thaliana</i>	[75]
12.	Shatavari	<i>Asparagus racemosus</i>	[76]
13.	Marijuana	<i>Cannabis Sativa</i>	[77]

### 2.2.3 Production and Purification

#### ❖ Production of LA

There are various methods for the production and LA optimization from different microorganisms. A schematic flowchart for the LA production followed by its purification is mentioned in Figure 5. Primarily solid-state fermentation (SSF) and submerged fermentation (SmF) are used conditions of reaction that differ from one microorganism to another to the production of the enzyme, and it is produced constitutively. SmF process has certain limitations such as very low net yield and also costly. At the same time, SSF economical and product yield is much more significant than the SmF (**Table 10**). LA has been produced by both SSF and SmF methods (**Table 11**).





**Figure 5** Flow chart of Production of the L-Asparaginase.

**Table 10** Comparison of characteristics for SmF, SSF and Solid-state Cultivation (SSC) methods [adapted from [78], [79]].

Sr. No.	Factor	SmF	SSF	SSC
1.	Energy consumed	High energy	Low energy	Low energy
2.	Water	High amount of water consumed along with effluents.	Low water requirement and Substrate are agricultural wastes.	Less water use and no effluent
3.	Concentration	30-80 g/l	-	100-300 g/l

4.	Equipment Volume	High volumes and high costs	High volume and high cost	Low volumes and lost costs
5.	Scale-up	Industrial equipment available	Difficult to control process parameters and Large-scale inoculums.	New design equipment needed
6.	Yield/product activity	low yield	High yield and product activity	Mild yield
7.	Mechanical agitation	Good homogenization	-	Static conditions preferred

**Table 11** Production method, Enzyme yield, optimization temperature and pH of LA from various sources.

Organism	Production Method	Medium	Opt. Temp. (°C)	Opt. pH	Nitrogen Source/ Carbon Source	Enzyme yield (IU/ml)	References
<i>Bacillus stratosphericus</i>	SSF	Nutrient Broth	25	7	Yeast extract / Maltose	7.81	[80]
<i>Aspergillus niger</i>	SSF	Glycine max Bran	30	6.5	N/A /Glucose	28.97	[51]
<i>Serratia marcescens</i>	SSF	coconut oil cake (COC)	35.5	7.5	-	-	[81]
<i>Aspergillus sp.</i>	SSF	Cotton seed oil cake, wheat bran, and red gram husk.	35	8	Yeast extract/ glucose	12.57	[82]
<i>Citrobacter sp.</i>	SSF	Tryptose Glucose Yeast Extract broth	37	8	Yeast extract/ glucose	82	[83]
<i>Enterobacter aerogenes</i> MTCC111	SmF	Modified M-9 agar medium	33	6.0	Ammonium chloride/Tri-sodium citrate	18.35	[28]
<i>Staphylococcus sp</i>	SmF	M-9 medium	37	7.5	Ammonium chloride/ Glucose	6-19	[47]

### Purification of LA

A need for purified enzyme and concentrated form of the enzyme is to know the stability and physiological activity of the enzyme. Enzyme purification is necessary for homogeneity level

to know enzyme conformation and structure. After purification of the enzyme, we can simply study the thermodynamic properties and kinetic properties and the enzyme mechanisms of the reaction. The importance of purified LA is to check enzyme antineoplastic activity and further therapeutic properties. Purified enzymes are of great importance for medicinal and industrial formulations. LA purified form is derived from crude microbial extract, with slight dissimilarities in it, according to the microorganism being used that is fungi or bacteria.

The extracellular LA is known to be the most useful/efficient compared to the intracellular LA because it contributes worthy yield in normal parameters. It has been discovered that different microbiological environments necessitate somewhat different conditions for its isolation and purification. There is definitely no distinct fixed composition of the medium from various bacterial sources yet known for the LA production. On the other hand, carbon and nitrogen sources play a vital role in LA production and microbial growth. For the production of LA, various types of microbes have variable growth parameters and optimum conditions such as temperature, pH, moisture content, etc. Various fungal and bacterial strains are used to produce the LA by using different necessities and drawbacks, purification steps, and kinetic properties of LA from multiple sources. All parameters are given in **Table 12**, [84]

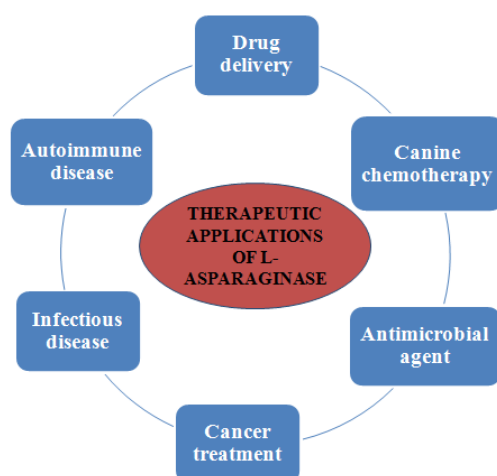
**Table 12** Purification steps and kinetic properties of LA from various sources.

Organism	Purification steps	$K_m$ (mM)	$V_{max}$	Specific activity (U/mg)	Fold / Yield (%)	Reference
<i>Pectobacterium carotovorum</i> MTCC 1428	DEAE cellulose chromatography, Sephadex G-100 chromatography, and ammonium sulfate precipitation.	0.657	4.45 U/ $\mu$ g	2020.91	72.12/ 42.05 %	[10]
<i>Cladosporium sp.</i>	Crude extract, Methanol precipitation, DEAE-cellulose column, Sepharose 6B	0.100	4.0 U/ $\mu$ g	83.3	867.7 fold	[85]
<i>Penicillium digitatum</i>	Ammonium sulfate precipitation and chromatography	0.01	-	833.15	60.94/ 4.35 %	[86]

<i>Aspergillus niger</i> AKV-MKBU	Ultra filtration, Ethanol Precipitation, and DEAE cellulose column	0.8141	6.228 M/mg/min.	46.75	10.3686 /24.88 %	[87]
<i>P. cyclopium</i>	Acetone Precipitation, Sephadex G-100 gel filtration	0.3	3333 U/mg	39480	52.3 / 4.5 %	[88]
<i>Erwinia chrysanthemi</i> 3937 (cloned in <i>Escherichia coli</i> BL21)	S-Sepharose FF column	0.058 ± 0.013	3.57 U/μg	118.7	15.4 / 69.8 %	[31]
<i>Streptomyces gulbargensis</i>	Sephacryl S-200 gel filtration, ammonium sulfate precipitation, and CM Sephadex C50 chromatography.	-	-	2053	82.12/ 32 %	[89]

### 2.2.4 L-Asparaginase Importance in medical industry and food industry

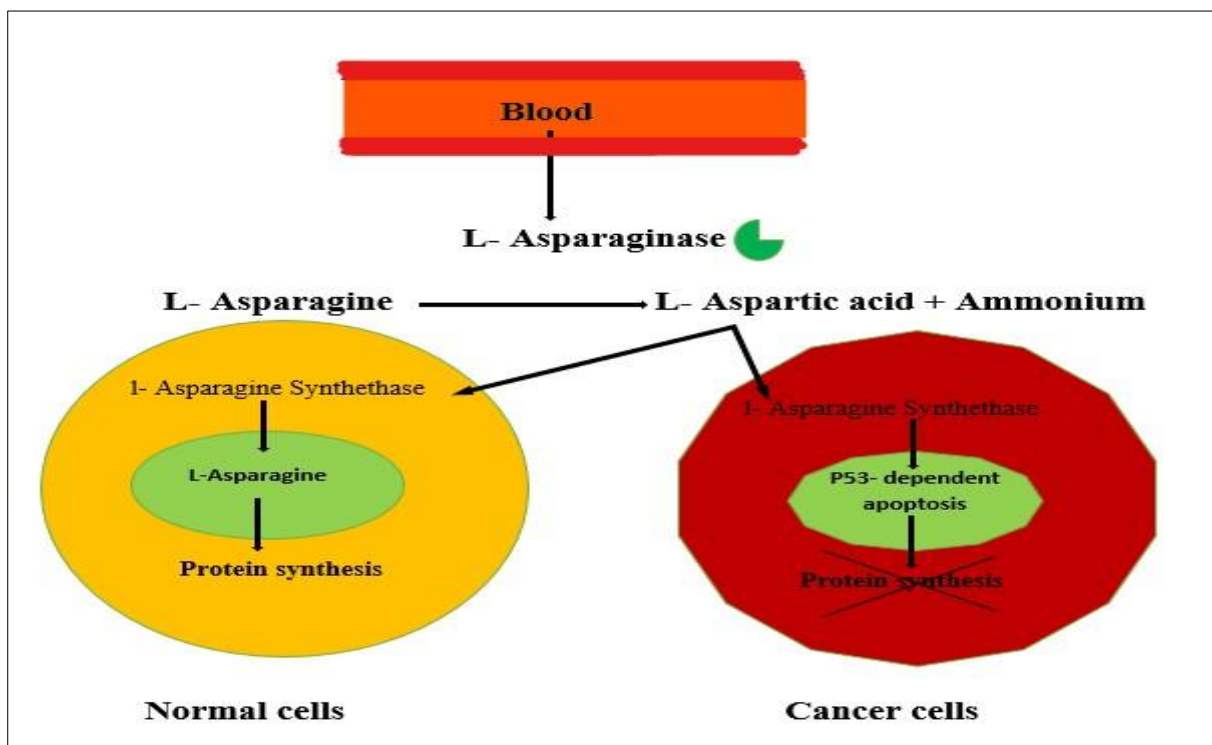
LA is a well-known enzyme for medical purposes and used in the drug to stop the growth of tumor cells. The enzyme has so much potential, it's not only limited as a chemotherapeutic agent, but it is also used in a wide range of medical applications such as auto immune disease, antimicrobial property, canine cancer, treatment of infectious disease and also in feline cancer treatment (**Figure 6**). In addition to this use, it is also used in the food industry to reduce the carcinogenic substance (i.e. acrylamide) concentration during the process of starchy food manufacturing [23].



**Figure 6** Therapeutic applications of L-Asparaginase [23].

### a) LA Importance in the medical industry

Enzymes play an important role as a therapeutic agent in the treatment of cancer. There are only two commercially available sources of LA that are *E. coli* and *Erwinia chrysanthemi*, which is used in the treatment of leukemia. It is now well known that the inhibitory effect of the enzyme is mediated by the depletion/ elimination of circulatory L-asparagine, which is a necessary nutrient for some tumor (leukemic) cells that are impaired in their capacity to synthesize L-asparagine but not for normal cells. The LA administration to leukemic individuals causes tumor cells to die selectively by hydrolyzing L-asparagine, leading in the malignant tumors treatment (Figure 7) [90].

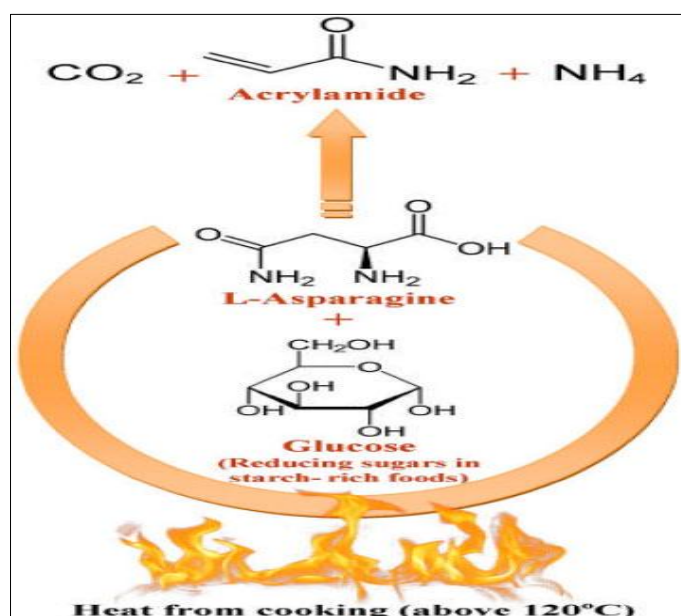


**Figure 7** Schematic representation of the antitumor activity of L-Asparaginase [adapted from [91]

### b) LA Importance in the food industry

Current developments in food technology have studied that when fried and baked food (especially starch-containing food) is heated at 120°C, it leads to acrylamide sugar (a cancer-causing agent) formation. This reaction occurs between the asparagine and the reducing sugar, recognized as the Maillard reaction (Figure 8). Bread dough and potato slices are

pretreated along with the LA enzyme before frying and baking, preventing acrylamide formation (**Table 13**).



**Figure 8** Acrylamide formation in fried and baked foods [92]

**Table 13** Acrylamide contents of different food products

Sr. No.	Food product	Acrylamide content	References
1.	Potato chips	173.51 ± 0.8 ug/g	[85]
2.	Gingerbread	890 mg/kg	[93]
3.	Chips (deep-fried)	351 mg/kg	[93]
4.	Cocktail snacks	1060 mg/kg	[93]
5.	Potato crisps	1249 mg/kg	[93]
6.	Commercial shortbread	278-326 mg/kg	[94]
7.	Roasted Barley and Corn	116 to 449 µg/kg	[95]
8.	Chips	12 to 3241 µg/kg	[95]

In this context, LA derived from *A. oryzae* and *A. niger* are utilized in the baking industry (**Table 14**). These enzymes perform optimally at temperatures ranging from 40 to 60°C and the pH levels ranging from 6.0 to 7.0. Since baking temperatures frequently exceeds 120°C, it is desirable to have enzymes that are stable and active across a wide temperature and pH range. As a result, LA derived from diverse sources (bacterial, fungal, plant, and animal) have been explored [4].

**Table 14** Commercial LA used in food industry.

Application	Form	Micro organism	Commercial Name	Commercial LA Activity	Manufacturer	References
Food Industry	Native LA	<i>A. niger</i>	1) PreventASe XR L	3500 ASNU/g	DSM, Heerlen, Netherlands	[91], [96]
			2) PreventASe L	6000 TASU/g		
		<i>A. oryzae</i>	3) Acrylaway L	47000 XRU/g	Novozyme, Bagsvaerd, Denmark	
			4) Acrylaway High T L	2500 ASPU/g		

### Other Importance of L-Asparaginase

- **LA in Amino Acid Metabolism**

The enzyme LA plays an important role in the biosynthesis of certain amino acids such as methionine, threonine, aspartic acid, and lysine. Aspartic acid, the precursors of threonine and lysine, is also formed by means of the LA enzyme action.

- **LA as Biosensor**

A biosensor technique is a reliable, eco-friendly and economical approach. For biosensors development, LA and L- glutaminase are recently used. Using LA (as a biosensor) to determine the asparagine and aspartate gives a precise measurement, compared to the older method such as nesslerization, which determined the ammonia liberation. The biosensor's mode of action is based on LA activity, which produces ammonium ions as a result of asparagine hydrolysis. It affects the pH, which is causing the alteration of the color and absorption. LA is used either in the food industry or leukemia by measuring the asparagine level [97].

### 2.3 Commercial Available Source of L-Asparaginase

At present, there are only two types of bacterial sources, such as *E. coli* and *Erwinia chrysanthemi*, which have been considered as the commercial source of asparaginase production for therapeutic and industrial applications. *E. coli* and *Er. chrysanthemi* LA form, half-life and commercial manufacturer names are given below (**Table 15**).

**Table 15** Commercial LA for therapeutic and pharmaceutical applications [14].

Application	LA Form	Source Organism	Half-life/ Asparagine depletion	Commercial Name	LA Manufacturer
Therapeutic/ Pharmaceutical	Native recombinant LA	<i>E. chrysanthemi</i>	0.65 ± 0.13 days/ 7–15 days	Erwinase	EUSA Pharma
		<i>E. coli</i>	-	Spectrila	Medac Gesellschaft
	Native LA	<i>E. coli</i>	1.28 ± 0.35 days/ 14–23 days	1) Kidrolas	EUSA Pharma
				2) Elspar	Ovation Pharmaceuticals
				3) Leukanase	Sanofi-aventis
PEGylated LA	<i>E. coli</i>	5.73 ± 3.24 days/ 26–34 days	Oncaspar	Enzon Pharmaceuticals	

#### 2.4 Drawbacks of *E.coli* and *E. chrysanthemi* LA

Currently, *E. coli* and *Er. chrysanthemi* asparaginases (Eca II and ErA) have been used for commercial purposes in the treatment of leukemia. But these LAs have been associated with several side effects, such as urticaria, laryngeal constriction, hypertension, oedema, loss of consciousness, diaphoresis, and asthma (hypersensitivity reactions) owing to associated L-glutaminase activity. Other adverse effects include vomiting, nausea, anorexia, hepatotoxicity, immunosuppression, acute pancreatitis, anemia, coma, lethargy, disorientation and drowsiness (Table 16). In addition, there is a need for the administration of multiple doses for effective leukemia treatment due to poor stability and a short half-life in the bloodstream [98].

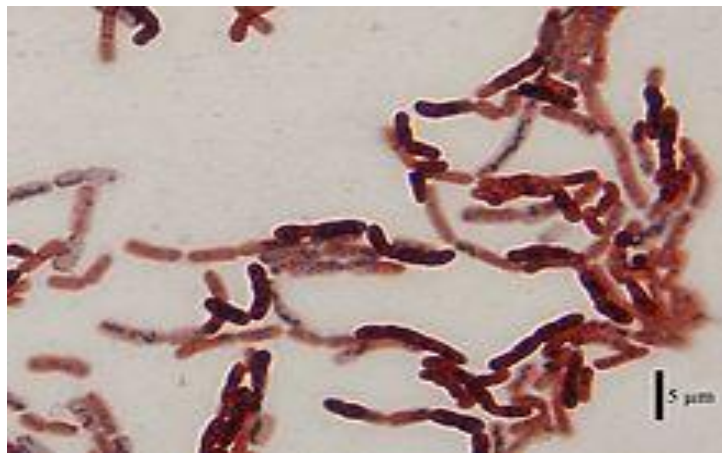
**Table 16** Side effects of *E.coli* and *Erwinia chrysanthemi* LA observed in ALL patients.

Sr. No.	Asparaginase type	Side effects	References
1.	<i>E. coli</i>	Allergy 2.5 %, Neurotoxicity 2.5%, Coagulation abnormalities 30.2 %, Convulsions 1.7%, Diabetes requiring insulin 1.4%, Pancreatitis 0.3%, Infection 5.1%, Liver toxicity 4.5%, and Death 0.3%.	[98]
2.	<i>Erwinia chrysanthemi</i>	Allergy 2.6 % , Neurotoxicity 1.4 %, Coagulation abnormalities 11.8%, Convulsions 0.3%, Diabetes requiring insulin 0.6%, Pancreatitis 0.9%, Infection 4.6%, Liver toxicity 3.8%, and Death 0.6%	



## 2.5 *Bacillus megaterium*

*B. megaterium* is a gram-positive bacterium that grows in the aerobic condition, forming an endospore that can resist the adverse condition (**Figure 9**). *B. megaterium* is widely distributed in diverse habitats such as rice paddies, soil, sediments and dried food etc. [99]. *B. megaterium* is fascinating, particularly due to bacterium morphology, rare or beneficial enzymes (for example, asparaginase) and its yield, and also have a wide-ranging biological habitation. [97,100]



**Figure 9** Stained *B. megaterium* cells [129].

### 2.5.1 Classification

**Table 17** Classification of *B. megaterium*.

Domain	Bacteria
Phylum:	Firmicutes
Class:	Bacilli
Order:	Bacillales
Family:	Bacillaceae
Genus:	<i>Bacillus</i>
Species:	<i>Megaterium</i>

### 2.5.2 Source of *B. megaterium*

Although bacterium is usually grown in the soil, it is also found in diverse habitats like seawater, paddies, normal flora, sediments, bee honey, and fish [99].

### 2.5.3 Biotechnological importance of *B. megaterium*

LA from *B. megaterium* was clinically much more beneficial as they possess less harmful activity, generally known as safe organisms. Like other species of *Bacillus*, *B. megaterium* having various industrial applications. Different industrially and pharmaceutically important products are produced by *B. megaterium* like heparosan, which is used in heparin synthesis, production of biomaterials, polyhydroxyalkanoates; an eco-friendly alternative to traditional petrochemical polymers like polyethylene and polypropylene, and xylanase (Used in bio-bleaching of pulp and paper industry).

In addition, *B. megaterium* is used for the synthesis of fungicidal toxins, herpes simplex corneal ulcers, HIV inhibitors, vitamin B12, hepatitis B virus, and pyruvate. Many of its industrial applications are given above in **Table 18**, [99]

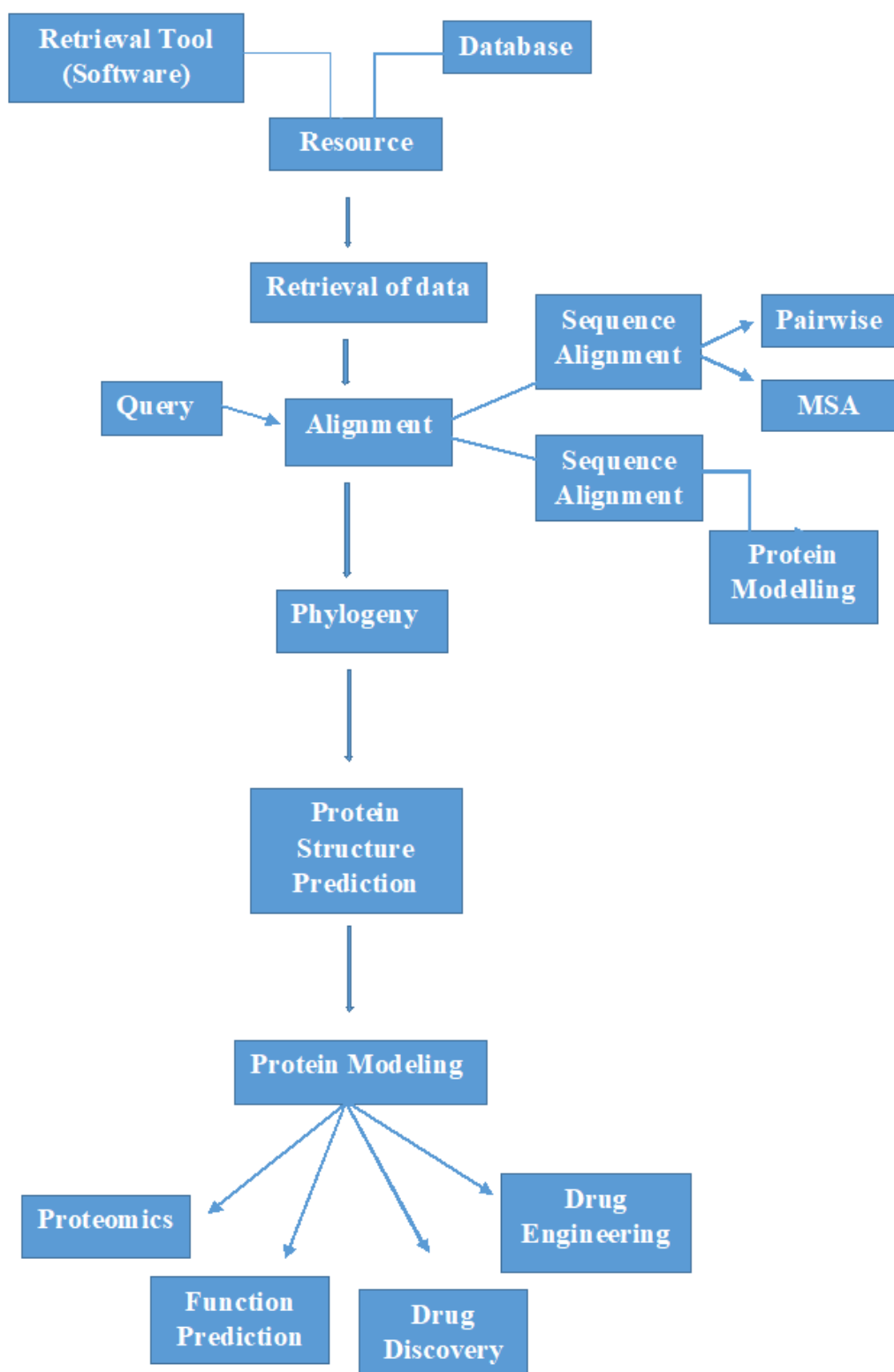
**Table 18** Industrially and pharmaceutically products produced by *B. megaterium*.

Product	Production strain	Expression plasmid	Medium	Product yield	Productivities	Application(s)	Reference
Heparosan	<i>B. megaterium</i> <i>MS941</i>	T7RNA polymerase (T7 RNAP)	M9+ medium	-	~ 250 mg/L and 2.74 g/L	Heparin synthesis, production of biomaterials	[101]
Glucose dehydrogenase	<i>B. megaterium</i> <i>M 1286</i>	-	Opt. medium	101 U/mg	-	catalyzes the oxidation of D-glucose	[102]
Polyhydroxy-alkanoates	<i>B. megaterium</i> <i>CCM 2037</i>	-	cheese whey	2.51 g/l and 0.79 g/l		Eco-friendly alternate to traditional petrochemical polymers like polyethylene and polypropylene	[103]
Alpha-amylase	<i>B. megaterium</i> <i>VUMB109</i>	-	Enriched culture media	3.2 mg/ml	-	Starch saccharification, used in baking, brewing, detergent, textile, paper and distilling industry	[104]

Xylanase	<i>B. megaterium</i>	-	MS2 medium	415.5 U/g	105 U/g	Used in bio-bleaching of pulp and paper industry	[105]
polyhydroxy butyrate (PHB)	<i>B. megaterium</i> B2	-	glycerol (7.6 g/L) and Na <sub>2</sub> HPO <sub>4</sub> (3 g/L)	1.20 g/L	-	Used to produce biodegradable, biocompatible and thermoplastic polyesters.	[106]
Keratinase	<i>B. megaterium</i> ATCC 14945	pWHK3	LB medium	-	166.2 U ml <sup>-1</sup>	-	[100]
Penicillin G amidase	<i>B. megaterium</i>	-	LB medium	~2000 U/L	-	used in the reverse synthesis of $\beta$ -lactam antibiotics	[107]

## 2.6 In Silico Study

In these few years, the enzyme industry is continuously developing exponentially. In-silico method can be used to discover a new natural enzyme and helping biotechnologists in the discovery of a novel (beneficial) enzyme at a meagre cost, in less time and by the eco-friendly method as compared to the traditional costly areas. In silico method is a highly successful method of searching for new enzymes followed by computer-based approaches and protein engineering. The exponent will develop computational capacity and biological data to make in-silico screening of enzymes to the competitive planned for finding new industrial enzymes [108]. The high-speed microprocessor used for the computational study is the best power, enabling the expert system to run multiple algorithms in parallel, providing an accessible interface and helps to generate the end results with a click of one button [109]. Computational methods are in-silico methods are help us to predict is the tertiary structure. Structure of proteins using in-silico method we can detect this structure of the protein and obviously it has no cost at all and takes very little time.



**Figure 10** Steps followed in the computational method for protein engineering.

**Table 19** Tools used for protein structure prediction.

<b>Tools</b>	<b>Basic function</b>
<b>UniProtKB</b>	Sequence retrieval
<b>BLASTp</b>	Sequence analysis
<b>Clustal-w</b>	Multiple sequence alignment
<b>ExPasy-ProtParam</b>	Physicochemical characterization
<b>PSIPRED or CFSSPS</b>	Secondary structure prediction
<b>Motif search tool</b>	Functional analysis
<b>CATH server</b>	Structural classification
<b>SWISS MODEL/ I- TASSER</b>	Homology Modeling/ Threading for protein prediction.
<b>QMEAN and UCLA</b>	Evaluation of predicted protein

### **2.6.1 Sequence retrieval**

Swiss Institute of Bioinformatics (SIB), the Protein Information Resource (PIR) and the European Bioinformatics Institute (EBI); these databases formed a knowledge base of the Universal Protein Resource (UniProt) consortium, which is found in the UniProtKB server [111]. In the UniProtKB, at least 120 million proteins sequences collections and annotations across all branches of life found [112]. There are two major sections: UniProtKB/TrEMBL, containing computer annotated entries and UniProtKB/Swiss-Prot, which contains manually annotated entries and contains data curated by scientists and offer users cross-links to approximately a hundred external databases and along with additional tools or information access [108].

The screenshot displays the UniProtKB web interface. At the top, there is a search bar with 'UniProtKB - bacillus megaterium' entered. Below the search bar, navigation links for 'BLAST', 'Align', 'Retrieve/ID mapping', 'Peptide search', and 'SPARQL' are visible. The main heading is 'UniProtKB 2021\_03 results'. A summary box explains that UniProtKB consists of two sections: 'Reviewed (Swiss-Prot) - Manually annotated' and 'Unreviewed (TrEMBL) - Computationally analyzed'. Below this, a filter bar shows 'Filter by' options and a search bar with the quote terms 'bacillus megaterium'. A table of results is displayed with columns for Entry, Entry name, Protein names, Gene names, Organism, and Length. Two entries are visible: P14779 (CPXB\_BACMB) and P40288 (DHG\_BACME).

Entry	Entry name	Protein names	Gene names	Organism	Length
<input type="checkbox"/> P14779	CPXB_BACMB	Bifunctional cytochrome P450/NADPH...	cyp102A1 cyp102, BG04_163	Bacillus megaterium (strain ATCC 14581 / DSM 32 / JCM 2506 / NBRC 15308 / NCIMB 9376 / NCTC 10342 / NRRL B-14308 / VKM B-512)	1,049
<input type="checkbox"/> P40288	DHG_BACME	Glucose 1-dehydrogenase		Bacillus megaterium	261

**Figure 11** UniProtKB web-based server for the sequence retrieval.

## 2.6.2 Sequence analysis

NCBI BLAST has a major accelerating significance due to its de facto standard in bioinformatics approximate string matching [114]. A BLAST algorithm [115] has originally 3 phases: identifying short sequences (words) with high match scores, extending those matches, and merging proximate extensions [110]

BLAST tools discover the sections of local similarity in the middle of sequences. BLAST tools can conclude evolutionary and functional relationships among sequences and advantages in the identification of the gene families members [116].

## 2.6.3 Multiple sequence alignment (MSA)

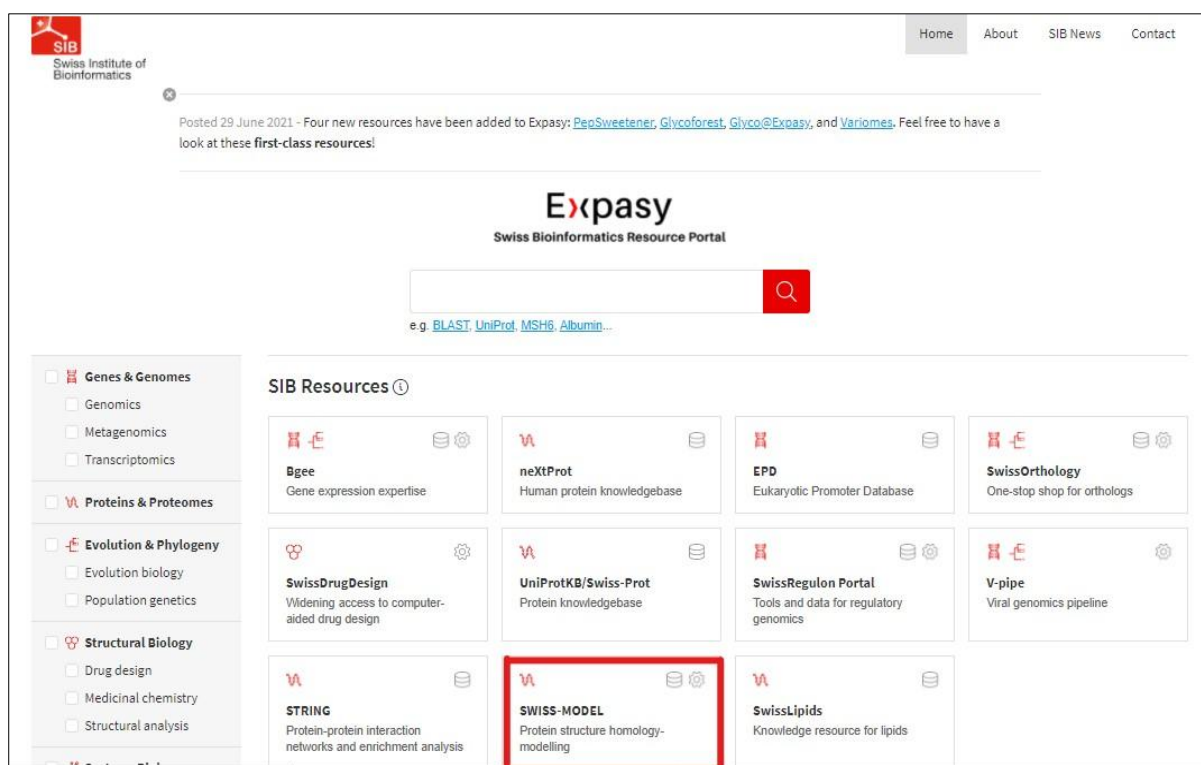
A Clustal server is extensively used for carrying out programmed multiple alignments of the sequences of amino acid or nucleotide. MSA be used to discover conserved residues and sections that can help recognize evolutionary conservation of functionally and structurally significant portions of the specific protein sequences [117]. As a result, predicted protein sequence important functionally and structurally region are determined using Clustal-W. It is very important that align sequences are related to each other to reduce error and make the result more useful [116].

Figure 12 BLAST User-Interface (UI) for the protein analysis.

Figure 13 Programmed multiple alignments by Clustal-W

## 2.6.4 Physicochemical characterization

Various physicochemical properties of LA include amino acid composition, isoelectric point (pI), aliphatic index, instability index (stability of proteins), extinction coefficient, aliphatic index, GRAVY or Grand Average Hydropathy and molecular weight were calculated by primary structure analysis by means of the ExPasy-ProtParam server. ExPasy (Bioinformatics Resource Portal) operated by SIB is an integrative and extensible portal that make available access to a range of scientific resources, databases, and software tools in various fields of science [118]

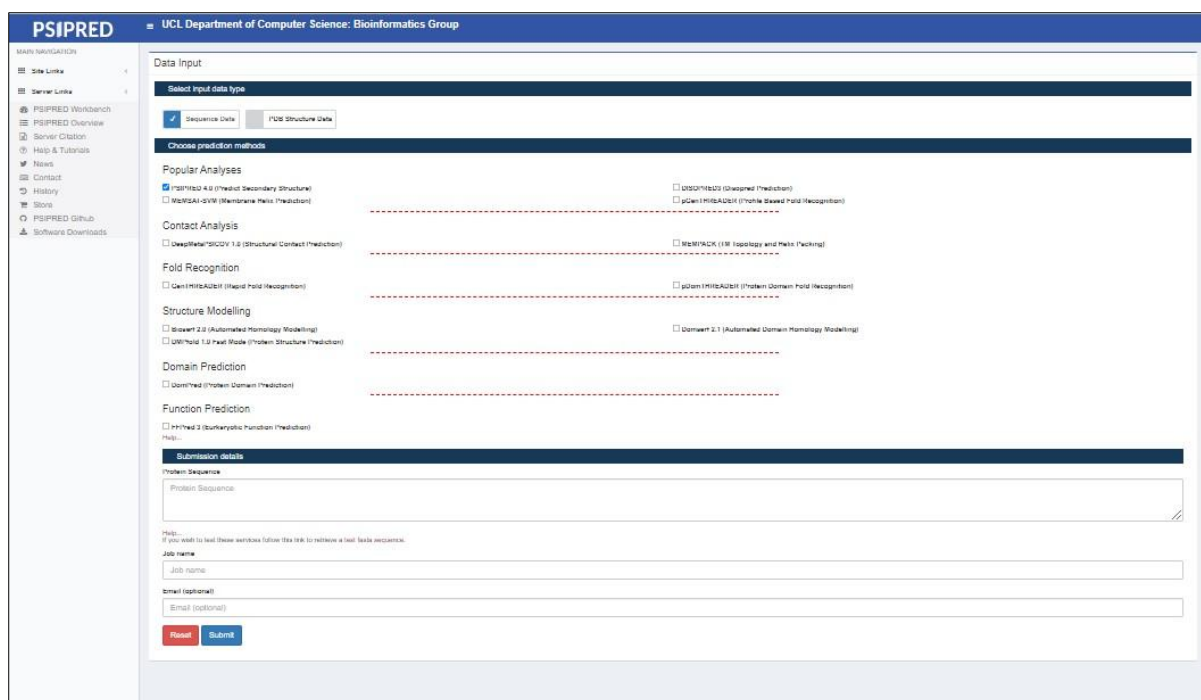


**Figure 14** ExPasy-ProtParam server for the physicochemical characterization of query protein sequence.

## 2.6.5 Secondary structure prediction

The secondary structural elements of a protein combine to generate its ultimate tertiary structure, which determines its functionality [113]. The secondary structural elements such as alpha helices, turns, beta-pleated sheet and loops were therefore predicted using these web-based servers, PSIPRED, SOPMA (Self-Optimized Prediction Method with Alignment) server, and CFSSP (Chou and Fasman Secondary Structure Prediction server) [120, 121].





**Figure 15** PSIPRED web-based server for the secondary structure prediction


### 2.6.6 Functional analysis

The Motif search tool analyzed the set of a conserved region of the amino acid residues, which can be accessed online. Whether the protein is a membrane-bound or a soluble protein can be identified using the SOSUI server, which is capable of predicting trans-membrane alpha-helices from the sequences of the amino acid using high accuracy and precision [113].

### 2.6.7 Structural classification

The CATH server was used for structural classification. The CATH website database clusters the protein structures based on their similarity or a common phylogenetic origin via physical curation and specific algorithms. CATH Classification is categorized into four different groups; these are (C) represents the class, (A) represents the architecture, (T) represents the topology and (H) represents the homologous family [109].

The class refers to the contents of the secondary structure of the proteins such as alpha, beta, mixed alpha/beta etc. In contrast, architecture illustrates the overall arrangement of the secondary structures, regardless of their connectivity, for example, an alpha/beta-sandwich. Topology also called the ‘fold level’, considers the connectivity of the secondary structures within the chain. The last hierarchical classification, i.e. homologous superfamily, pertains to a group of domains that share a common ancestor [109].



## MOTIF Search

Search Motif Library
Search Sequence Database
Generate Profile
KEGG2

[Help](#)

**Enter query sequence:** (in one of the three forms)

Sequence ID  (Example) mja:MJ\_1041

Local file name  No file chosen

Sequence data

**Select motif libraries:** ( [Help](#) )

<p><b>Databases</b></p> <p><input checked="" type="checkbox"/> Pfam</p> <p><input type="checkbox"/> NCBI-CDD</p> <p><input checked="" type="radio"/> All <input type="radio"/> COG</p> <p><input type="radio"/> TIGRFAM <input type="radio"/> SMART</p> <p><input type="checkbox"/> PROSITE Pattern</p> <p><input type="checkbox"/> PROSITE Profile</p> <p><input type="checkbox"/> User-defined Profile Library (may contain multiple profiles)</p> <p>Profile file name: <input type="button" value="Choose File"/> No file chosen</p> <p><input checked="" type="radio"/> PROSITE format</p> <p><input type="radio"/> HMMER format</p>	<p><b>Cut-off score</b> (Click each database to get help for cut-off score)</p> <p><input type="text" value="1.0"/> * E-value</p> <p><input type="text" value="1.0"/> * E-value</p> <p><input checked="" type="checkbox"/> Skip entries with SKIP-FLAG</p> <p><input checked="" type="checkbox"/> Skip frequently matching (unspecific) profiles</p> <p><input type="text"/></p>
---	--

[Feedback](#)
[KEGG](#)
[GenomeNet](#)

**Figure 16** Motif search tool to analyze the conserved region of the amino acids residues.

C A T H
Home Search Browse Download About Support

### Search CATH

Cut and paste a FASTA sequence in this box and click 'Search'

Search CATH by protein sequence (FASTA)

Locate the position of structural domains on your protein sequence.

Examples: FASTA sequence ID: A0A0Q0Y989

Search by Text or ID
Search by Sequence
Search by Structure

Progress

Matching Domains    Matching FunFams

[API](#) [Help](#)

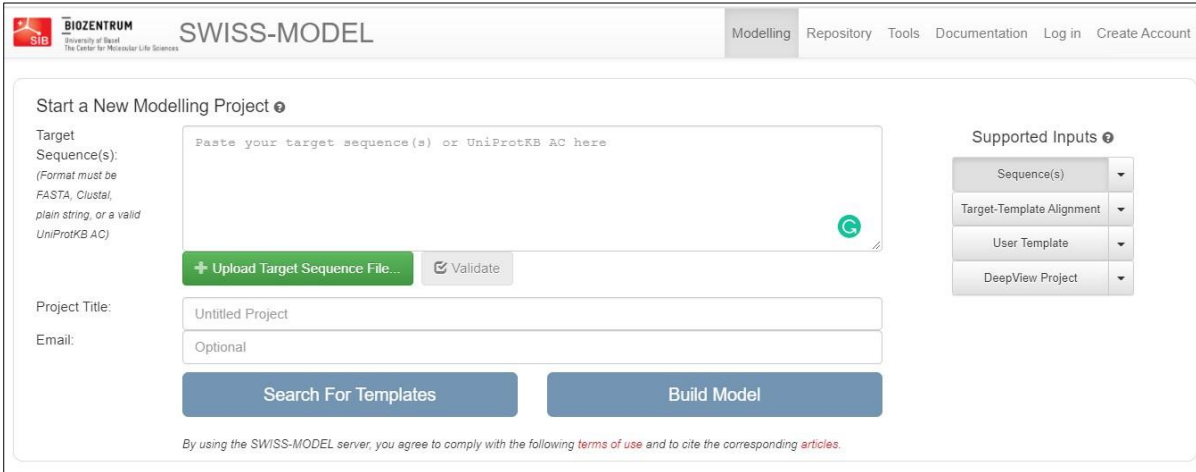
Paste your protein sequence into the text box above (or use an example) then click 'Search'.

- The progress bar below will let you know when your results are available.
- Click [Help](#) to find out more information on this search.
- Click [API](#) to find out how to use this service in your programs.

**Figure 17** CATH server for the structural classification

## 2.6.8 Homology Modeling

Homology modeling predicts protein structure based on sequence homology. For homology modeling, we used SWISS-MODEL [122], [123]. Initially, we performed a BLAST, pairwise alignment of the BmA sequence with all the other protein sequences procured from the databases like UniProtKB, Swiss-Prot. The sequence with a known structure having a higher similarity with our sequence is used to predict the three-dimensional structure. Homology modeling predicts the protein structure of the query protein sequence (in this case, BmA) based on the experimentally determined structure of the protein sequence that is closely related to the query sequence. Predicted protein structure based on sequence homology, also known as comparative modeling. Since, two proteins with a high sequence similarity possess a similar 3-D structure; one protein with a known structure is used to be copied for a protein with an unknown structure with a high degree of confidence.

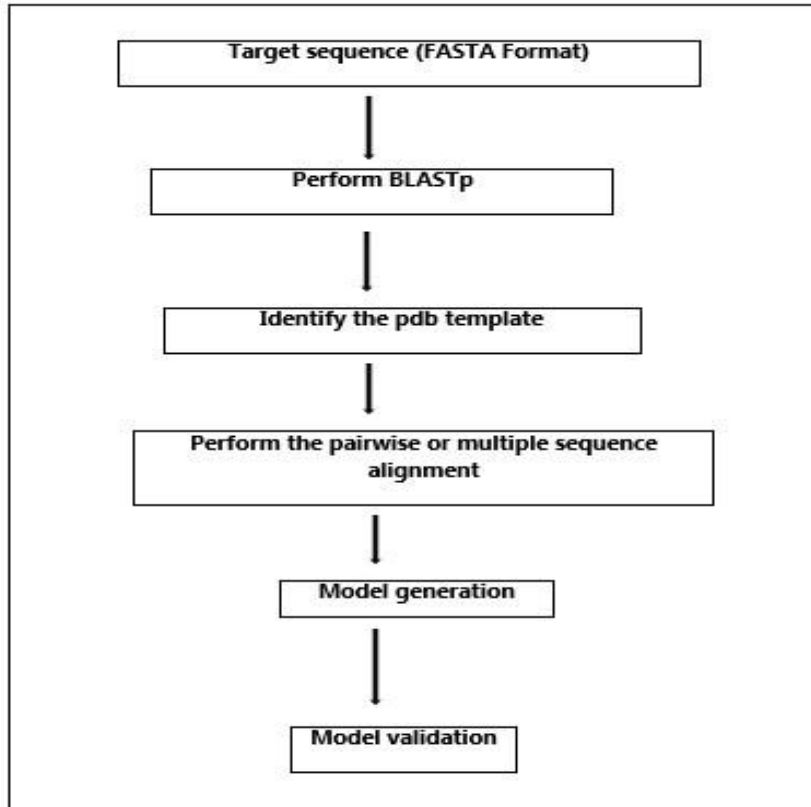


The screenshot shows the SWISS-MODEL web interface. At the top, there is a navigation bar with 'Modelling', 'Repository', 'Tools', 'Documentation', 'Log in', and 'Create Account'. The main heading is 'Start a New Modelling Project'. Below this, there is a 'Target Sequence(s)' input field with a placeholder 'Paste your target sequence(s) or UniProtKB AC here'. To the right of this field is a 'Supported Inputs' dropdown menu with options: 'Sequence(s)', 'Target-Template Alignment', 'User Template', and 'DeepView Project'. Below the input field are two buttons: '+ Upload Target Sequence File...' and 'Validate'. Further down, there are input fields for 'Project Title' (containing 'Untitled Project') and 'Email' (containing 'Optional'). At the bottom of the form are two large buttons: 'Search For Templates' and 'Build Model'. A small disclaimer at the very bottom reads: 'By using the SWISS-MODEL server, you agree to comply with the following terms of use and to cite the corresponding articles.'

**Figure 18** SWISS-MODEL server for the protein structure prediction.

### • Template selection

A SWISS-MODEL server template library extracts the template (known protein structure sequence) from the Protein Data Bank (PDB). The sequences from the template structure library are searched for the query protein to pick the templates for a query protein (unknown protein structure sequence). Then perform protein BLAST to the unknown protein structure sequence, then got a highly identical sequence. That highly identical sequence is called a template, which acts as a reference to the unknown protein structure sequence to predict tertiary protein structure. Highly identical sequences have probably similar 3-D structures or tertiary protein structures.



**Figure 19** Steps followed in homology modeling.

- **Alignment**

The alignment of the structure is created next to eliminating mismatched templates. The query sequence is aligned to the main template structures using local pair-wise alignment algorithm. In the next step, the alignment is further improved for modeling purposes. Therefore, location of the deletions and insertions is adjusted, allowing for the template structure context. In general, to help in the loop building process (in the alignment), isolated residues are relocated to the flanks.

- **Backbone and Loop Model Building**

In the process of protein structure prediction, the backbone atom positions of the template structure are averaged to create the core of the model. A scoring scheme best loop is selected, which accounts for the steric hindrance, favorable interactions like hydrogen bond formation, force, and field energy. If there is no appropriate loop to recognise, then to allow for more flexibility in the predicted structure, the flanking residues are involved in the reconstructed fragment.

- **Side Chain Refinement**

A rebuilding of the model side chains is constructed on the weighted positions of parallel residues present in the structures of the template. The model side chains are created through conserved residues by sterically interchanging template structure side chains. The scoring function evaluating unfavorably close contacts and favorable interactions (disulfide bridges and hydrogen bonds) is functional to choose the best possible conformation.

- **Model refinement using energy function**

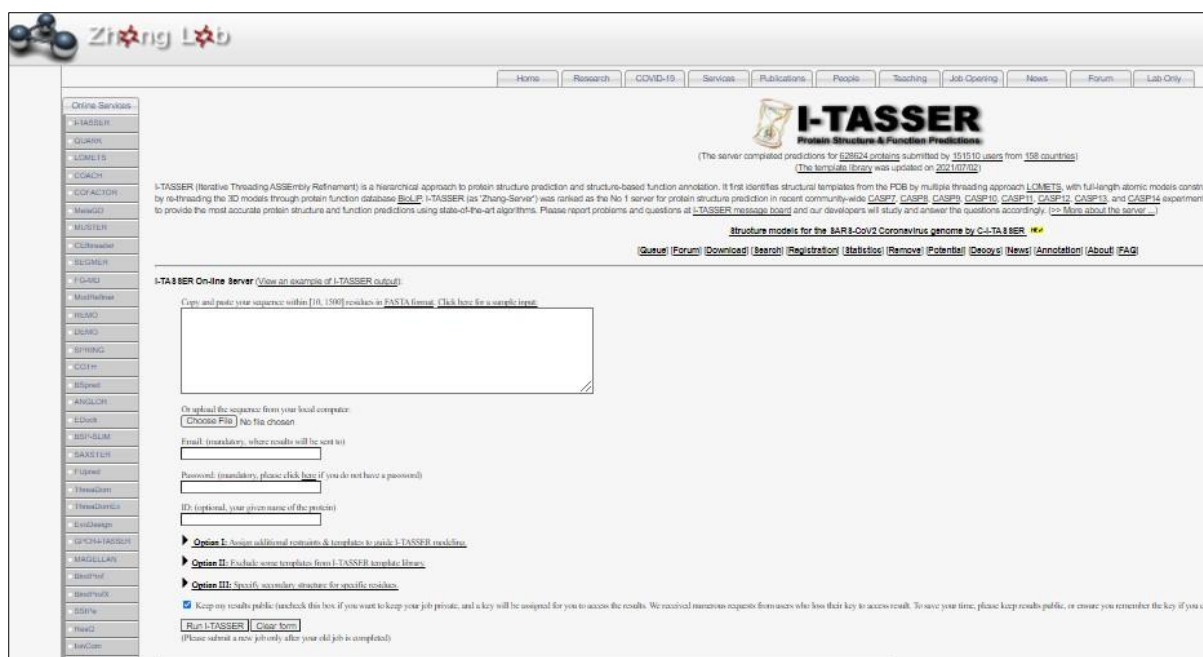
Molecular dynamics methods (energy minimization) are generally not capable of developing the built model accuracy, and it is only used in SWISS-MODEL to normalize the structure.

- **Modeling results and evaluation**

Potential applications of the protein models can be determined by generally on the built models quality. The quality (accuracy) of the built model can differ significantly, even inside altered sections of the identical protein. Generally, highly-conserved core regions can be able to model considerable or dependably than surface residues or variable loop regions. The users can evaluate the consistency of the built model using numerous tools provided by the SWISS-MODEL server. The predicted protein structure model can be assessed by QMEAN and UCLA—DOE LAB SAVES server using six programs. These are ERRAT, Verify3D, PROVE, WHATCHECK, PROCHECK and CRYST [110]. If the quality of the built model is not good, then the prediction of the protein model can be created by using I-TASSER from the threading of unknown protein structure sequence.

### **2.6.9 Threading**

I-TASSER is a categorized methodology to predict protein structure. I-TASSER aimed to model single-domain globular proteins by using several approaches. It has been used to identifies structural templates from the PDB by multiple threading method LOMETS [124]. Therefore, BmA predicted structure was obtained using a web-based server, I- TASSER (Figure 20). Threading determines the model by structural similarity with or without detectable sequence similarity.



**Figure 20** I-TASSER server to identify structural templates from the PDB by multiple threading approach LOMETS

### 2.6.10 Evaluation of predicted protein

QMEAN and UCLA—DOE LAB SAVES (The Structure Analysis and Verification Server version 6) servers were used to assess the predicted protein model. These servers execute six programmes to examine and validate protein structures during and after model refining (Table 20).

**Table 20** Model Assessment Program

Programs	Description
<b>ERRAT</b>	Analyzes the statistics of non-bonded interactions between distinct atom types and shows the value of the error function vs position of a 9-residue sliding window, which is obtained by comparing statistics from highly refined structures.
<b>Verify3D</b>	Determines an atomic model (3D) compatibility with its amino acid sequence (1D) by assigning a structural class based on its location and environment (alpha, beta, loop, polar, non-polar etc.) and comparing the findings to good structures.
<b>PROVE</b>	Calculates the volumes of atoms in macromolecules using an algorithm that treats the atoms as hard spheres and computes a statistical Z-score deviation for the model from PDB-deposited structures that are well resolved (2.0 Å or higher) and refined (R-factor of 0.2 or better).
<b>WHATCHECK</b>	Derived from a subset of protein verification tools from the WHATIF program, The tool

	performs comprehensive checks on numerous stereo-chemical parameters of the residues in the model.
<b>PROCHECK</b>	Analyzes residue-by-residue geometry and overall structural geometry to determine the stereo-chemical quality of a protein structure
<b>CRYST</b>	This program searches the Protein Data Bank for entries with a unit cell comparable to the one in your input file. CRYST1 record is needed. For more options, the standalone CRYST server may be used.

## 2.7 Importance of Current Work

Commercially available LA from *E. coli* and *Erwinia chrysanthemi* have a short half-life and several side effects on the administration to the patient. Therefore, there is a need for stable, more specific LA with no or minimal side effects. The current work aims to conduct an in-silico structural and functional analysis of BmA to explore its physicochemical properties using various computational tools. For this reason, the sequence of BmA is considered for computational analysis. *B. megaterium* is a good source of LA and a GRAS organism (Generally Recognized as Safe). To our knowledge, no prior in-silico studies on the properties of the *B. megaterium* LA are available. Sequence analysis, physicochemical characterization, secondary structure prediction, functional analysis and structure classification of this BmA were studied and the protein structure based on sequence homology was then predicted. This study gives an insight into the general characteristics of the BmA compared with its peer asparaginases. It will help in identifying whether BmA can be a good alternative to other asparaginases.

## CHAPTER- 3

### OBJECTIVES

- To analyze the structure and function correlation of *B. megaterium* asparaginase for potential application in cancer therapy using bioinformatics tools.
  - Sequence retrieval
  - Comparative sequence analysis of BmA with other *Bacillus* asparaginases
  - Physicochemical characterization based on BmA primary structure
  - Correlation establishment of BmA structure with its function



## CHAPTER- 4

### MATERIALS AND METHODS

#### 4.1 Sequence retrieval

The LA sequences of 10 different *Bacillus* sp. including *B. megaterium* were retrieved from the UniProtKB server, a knowledge base of UniProt consortium constructed by SIB, EBI, and the PIR [111]. The UniProt databases can be used online (<http://www.uniprot.org/>) or downloaded in different formats (<ftp://ftp.uniprot.org/pub>).

For computational investigation, the full-length asparaginase protein sequences of 10 different *Bacillus* species were downloaded in FASTA format [113].

#### 4.2 Sequence analysis

BLAST discovers areas of local similarity in the middle of the sequences. BLAST can be used to understand evolutionary and functional connections among sequences, as well as to identify members of gene families. The NCBI BLASTp databases can be accessed online (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) [116].

#### 4.3 Multiple sequence alignment (MSA)

The MSA can help recognize phylogenetic conservation of functionally and structurally significant regions of the protein sequences [117]. This can help to discover the conserved residues and sections that can advantage to understand the functional and evolutionary relationships. Therefore the structurally and functionally significant regions of the protein sequence predicted using Clustal-w (<https://www.genome.jp/tools-bin/clustalw>)

#### 4.4 Physicochemical characterization

Various physicochemical properties of LA such as molecular weight, instability index (stability of proteins), isoelectric point (pI), amino acid composition, extinction coefficient, and aliphatic index, and GRAVY or Grand Average Hydropathy were calculated based on primary structure analysis by means of the Expasy-ProtParam tool (<http://web.expasy.org/protparam/>). Expasy (Bioinformatics Resource Portal) run through SIB, which is an integrative and extensible portal that provides access to a range of scientific resources, databases, and software tools in various fields of science [118]

The isoelectric point (pI) values vary from 4–7, showing that the proteins are acidic and will migrate towards a positive electrode in a 2-D gel. If the protein's instability index is more than 40, it implies that it is unstable. The higher value of aliphatic index suggests that the proteins are thermostable. The lower range of GRAVY indicates that the proteins have greater interactions with water [113]. The extinction coefficient shows how much light the protein absorbs at a specific wavelength [119].

#### 4.5 Secondary structure prediction

The secondary structural elements of a protein association to form its final tertiary structure that determines its functionality [113]. Secondary structural elements such as helix, turn, sheet and coil were therefore predicted using different web-based servers, PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>), SOPMA server, ([https://npsa-prabi.ibcp.fr/cgi-bin/secpred\\_sopma.pl](https://npsa-prabi.ibcp.fr/cgi-bin/secpred_sopma.pl)) and CFSSP server, (<http://cho-fas.sourceforge.net/>) [120, 121].

#### 4.6 Functional analysis

A set of conserved amino acid residues were examined by means of the Motif search tool, which can be accessed online (<https://www.genome.jp/tools/motif/>), soluble proteins and membrane were distinguished using SOSUI server (<https://harrier.nagahama-bio.ac.jp/sosui/cgi-bin/msosui.cgi>), which predict trans-membrane helices from sequences of the amino acid with high accuracy and precision [113]

#### 4.7 Structural classification

The CATH server helps to identify the structural classification of the predicted protein structure. The CATH website database clusters the protein structures based on their common phylogenetic origin or similarity using physical curation and assured algorithms [109]. The CATH server databases can be accessed online ([http://www.cathdb.info/search/by\\_sequence](http://www.cathdb.info/search/by_sequence)).

#### 4.8 Homology Modeling of BmA sequence

The 3-D structure of BmA was predicted based on homology modeling using SWISS-MODEL (<https://swissmodel.expasy.org/interactive>) [122, 123].

#### **4.9 Threading of BmA sequence**

I-TASSER is a categorized methodology for the prediction of the protein structure. I-TASSER aimed to modeling single-domain globular proteins by using several approaches. It has been used to classifies structural templates as of the PDB through the multiple threading approach LOMETS [124]. Therefore, BmA predicted structure was obtained using a web-based server, I-TASSER server database can be accessed online (<https://zhanglab.dcmf.med.umich.edu/I-TASSER/>).

#### **4.10 Evaluation of predicted protein**

The quality of model of the predicted protein structure was evaluated using QMEAN and UCLA—DOE LAB SAVES (Verification and a Structure Analysis Server version 6) servers (<https://saves.mbi.ucla.edu/>), which inspected and validated the protein structure by running six programs.

## CHAPTER- 5

### RESULTS AND DISCUSSION

#### 5.1 Sequence Retrieval

A total of 10 asparaginase sequences were retrieved for different *Bacillus* species (*B. subtilis*, *B. altitudinis*, *B. mycooides*, *B. cereus*, *B. wiedmannii*, *B. safensis*, *B. circulans*, *B. sonorensis*, and *B. licheniformis*) along with *B. megaterium*. After phylogenetic analysis, it was discovered that *B. megaterium* is more closely related to *B. mycooides*, *B. cereus*, and *B. wiedmannii* [109] (**Table 21**).

**Table 21** Details of selected sequences of LA from different *Bacillus* species.

Sr. No.	Protein Accession Number	Organism	Length
1	KNH26430.1	<i>B. megaterium</i>	323 aa
2	KFF56126.1	<i>B. subtilis</i>	329 aa
3	KLV14741.1	<i>B. altitudinis</i>	324 aa
4	TKI83183.1	<i>B. mycooides</i>	324 aa
5	TKI94247.1	<i>B. cereus</i>	324 aa
6	TKI15105.1	<i>B. wiedmannii</i>	324 aa
7	AYJ88275.1	<i>B. safensis</i>	329 aa
8	KLV25750.1	<i>B. circulans</i>	332 aa
9	RHJ11726.1	<i>B. sonorensis</i>	329 aa
10	RHL19359.1	<i>B. licheniformis</i>	322 aa

#### 5.2 Sequence Analysis

BLAST tools discover the sections of local similarity in the middle of sequences. BLAST tools can be used to conclude evolutionary and functional relationships among sequences in addition to the advantage in the identification of the gene families members. BLAST<sub>p</sub> of all the ten asparaginase sequences of *Bacillus* species (**Table 22**), [116].

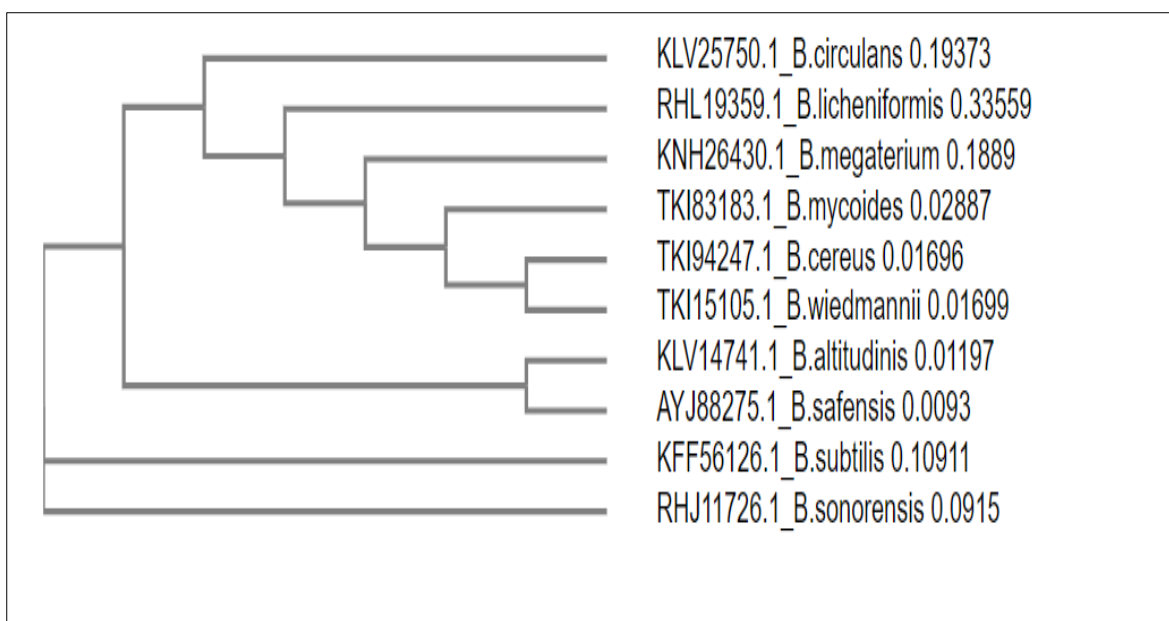
**Table 22** Pairwise sequence alignment of BmA with L-asparaginases of various *Bacillus* species using BLASTp

<b>Organism Name</b>	<b>Amino Acid Length</b>	<b>Similarity</b>	<b>Identity</b>	<b>Query Coverage</b>	<b>E-Value</b>
<i>B. megaterium</i>	323 aa	100%	100%	-	-
<i>B. subtilis</i>	329 aa	46%	26%	95%	1e-22
<i>B. altitudinis</i>	324 aa	45%	25%	99%	8e-22
<i>B. mycoides</i>	324 aa	78%	64%	100%	1e-156
<i>B. cereus</i>	324 aa	78%	64%	100%	5e-158
<i>B. wiedmannii</i>	324 aa	78%	64%	100%	2e-152
<i>B. safensis</i>	329 aa	45%	27%	99%	2e-23
<i>B. circulans</i>	332 aa	43%	24%	95%	2e-20
<i>B. sonorensis</i>	329 aa	45%	24%	99%	3e-19
<i>B. licheniformis</i>	322 aa	50%	34%	95%	1e-56

### 5.3 Multiple Sequence Alignment

MSA can discover conserved residues and sections that can help recognize evolutionary conservation of functionally and structurally significant portions of the specific protein sequences [117]. As a result, predicted protein sequence important functionally and structurally region are determined by means of Clustal-W. It is very important that align sequences are related to each other in order to reduce error and make the result more useful [116].

MSA and phylogenetic tree construction were performed to analyze the diversity of selected proteins and their evolutionary relationship (**Figure 21**). Additional, these sequences helped to analyse the physicochemical characterization [109]



**Figure 21** Phylogenetic tree of 10 Asparaginase protein sequences from *Bacillus* species.

#### 5.4 Physicochemical Characterization

- Theoretical pI:** An isoelectric point (pI) of the protein symbolizes the pH at which a given amino acid sequence will have a total neutral charge. The theoretical pI parameter used to determine the 2-D gel of a specific protein will migrate. PI values lie in the middle of the ranges of 4–7 that confirmations the proteins are acidic and will migrate towards a positive electrode in a 2-D gel.

Therefore, all sequences under study have pI ranges of 4–7; all asparaginases are acidic.

- Instability index:** It measures the stability of a given protein in vitro. It is designed based on the presence of certain di-peptides, which confer a degree of stability of the protein. If the protein instability index is below 40, it is then considered stable in atmospheric conditions. If the instability index is greater than 40, it indicates that the protein is unstable. Since all sequences under study have an instability index of less than 40, all asparaginases are stable in atmospheric conditions.

- Aliphatic index:** The aliphatic index of the protein is defined as the comparative volume occupied by means of aliphatic side chains (valine, alanine, leucine and isoleucine,). It may possibly be regarded as a positive factor in the rise of the thermostability of the globular proteins [122]. A high aliphatic index confers a degree of thermal stability to a given protein.

All the selected sequences have a high aliphatic index. BmA shows a 98.95 aliphatic index that means BmA has a high degree of thermal stability.

- **Extinction coefficient:** It shows how much light the protein absorbs at a specific wavelength; it determines the purified protein concentration using a spectrophotometer [109].
- **GRAVY:** This parameter represents the overall hydrophilic or hydrophobic character of a given protein. A negative GRAVY value indicates hydrophilic protein, whereas a positive GRAVY value indicates hydrophobic protein.

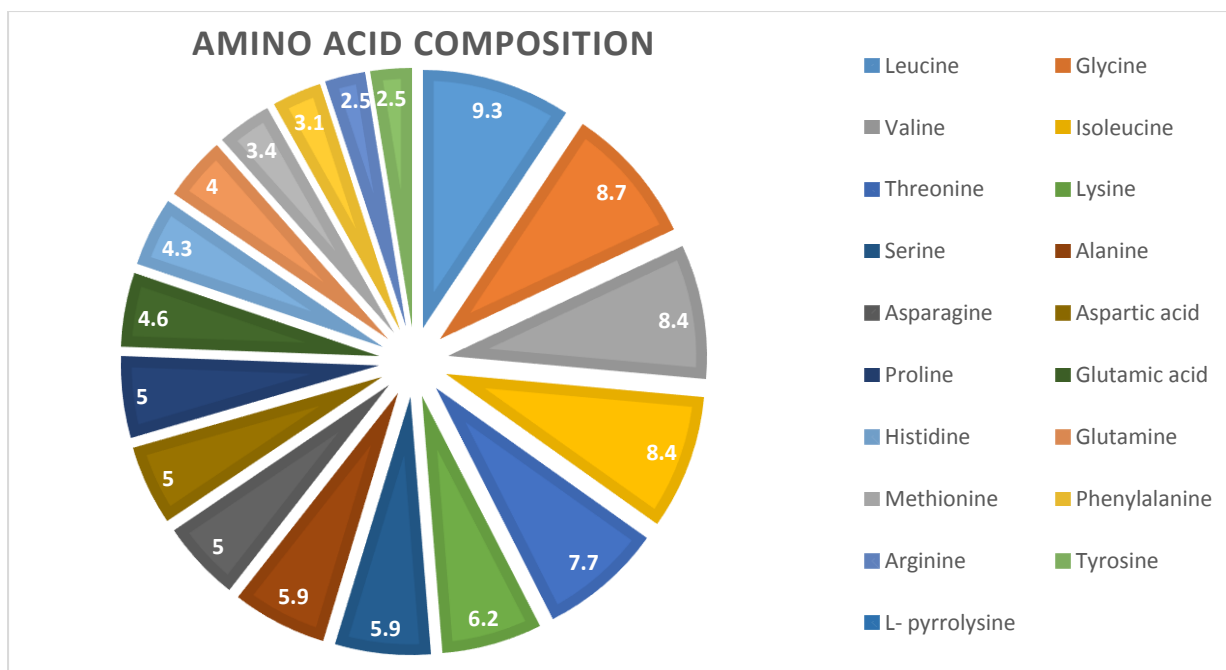
*B. wiedmannii*, *B. cereus*, and *B. mycooides*. have reported a positive value of GRAVY. All the other sequences have been reported here to have negative GRAVY values.

**Table 23** Physicochemical property of selected *Bacillus* species sequences.

Organism Name	Protein Accession Number	Theoretical pI	Extinction coefficient	Instability index	Aliphatic index	GRAVY
<i>B. megaterium</i>	KNH26430.1	6.53	11920	26.85	98.95	-0.044
<i>B. subtilis</i>	KFF56126.1	4.65	33600	39.33	95.38	-0.074
<i>B. altitudinis</i>	KLV14741.1	5.02	34840	29.92	95.11	-0.182
<i>B. mycooides</i>	TKI83183.1	5.57	10555	29.07	105.49	0.059
<i>B. cereus</i>	TKI94247.1	5.32	12045	30.46	105.49	0.046
<i>B. wiedmannii</i>	TKI15105.1	5.41	12045	30.03	105.19	0.057
<i>B. safensis</i>	AYJ88275.1	5.03	34965	28.59	93.92	-0.184
<i>B. circulans</i>	KLV25750.1	4.87	37025	33.60	102.47	-0.005
<i>B. sonorensis</i>	RHJ11726.1	5.09	33600	35.37	95.68	-0.157
<i>B. licheniformis</i>	RHL19359.1	5.49	27070	34.60	89.60	-0.181

### **Amino acid composition and Molecular weight:**

BmA has 323 amino acid residues and possesses a molecular weight of 35154.51 Da. There are (Asp + Glu) 31 total no. of negatively charged residues and the total number of positively charged residues is (Arg + Lys) 28, as shown in **figure 22**.



**Figure 22** Pie chart here represent the amino acid residues percentage in BmA.

### Comparison of physicochemical properties with commercially available asparaginases:

When we did a comparative study of *B. megaterium* physicochemical properties with a commercially available source of asparaginase, i.e. *E. coli* and *E. chrysanthemi*, *B. megaterium* found a good aliphatic index .and theoretical pI of BmA lies in between *E. coli* and *E. chrysanthemi* (as shown in **table 24**). The theoretical pI of BmA suggests that the BmA can be more effective at the blood pH range.

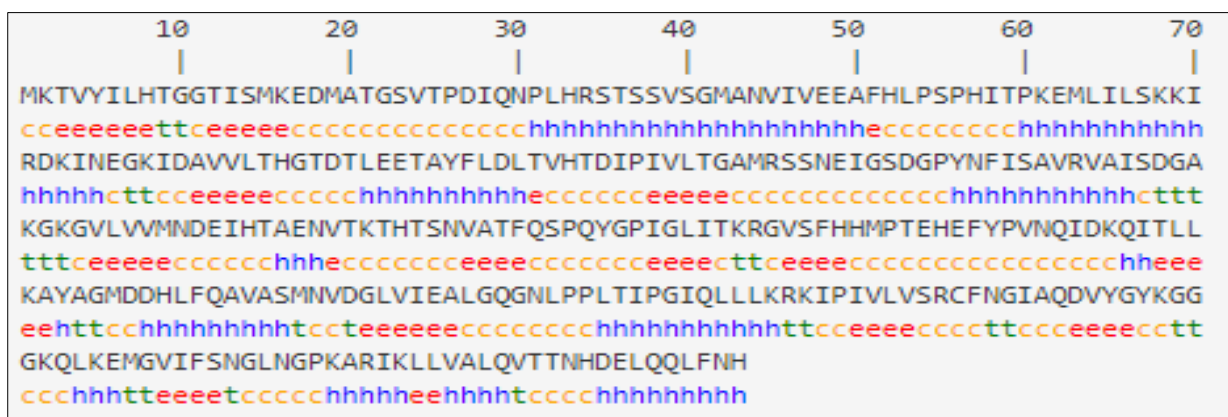
**Table 24** Comparison of physicochemical properties of BmA with the commercially available asparaginases; EcAII and ErA

Organism Name	Protein Accession Number	Theoretical pI	Extinction coefficient	Instability index	Aliphatic index	GRAVY
<i>B. megaterium</i>	KNH26430.1	6.53	11920	26.85	98.95	-0.044
<i>E. coli</i>	BAE77020.1	5.96	23505	18.27	85.46	-0.114
<i>E. chrysanthemi</i>	P06608.1	7.84	24870	17.20	98.30	0.042



## 5.5 Structure Prediction

Based on the secondary organization ( $\alpha$ -helices,  $\beta$ -sheets and coils), it has been observed that asparaginase proteins can be categorized into three main groups:  $\alpha$ -helices,  $\beta$ -sheets and coils. (Figure 23), (Table 25). This illustration has been supported by various structural findings of asparaginase from multiple organisms. Crystal structure calculations have revealed a variation of the structural parameters to a number of asparaginases [109, 125].

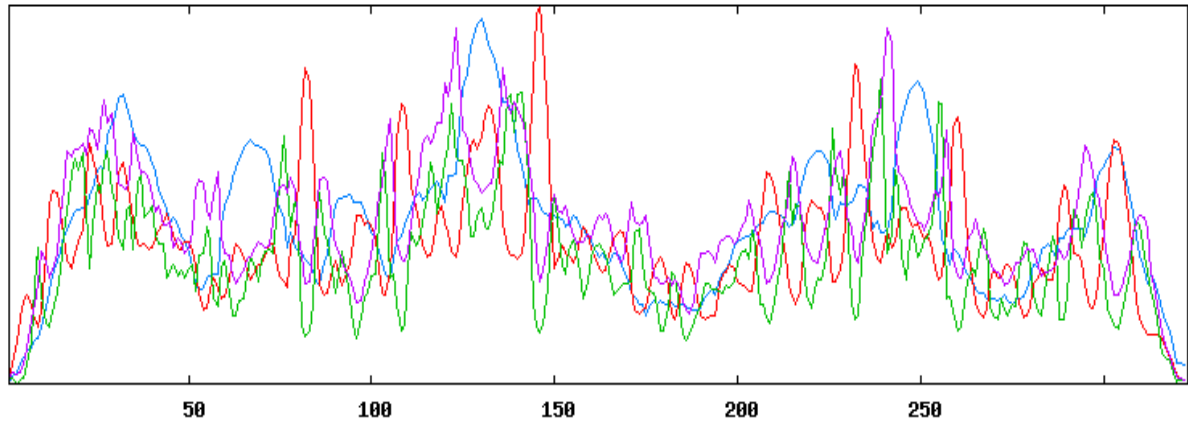


**Figure 23** Secondary structure map for BmA sequence.

**Table 25** Calculated secondary structure elements of BmA sequence by using SOPMA.

Secondary structure	Number of Residues	Number of Residues (In Percentage)
Alpha helix (Hh)	104	32.20%
Extended strand (Ee)	66	20.43%
Beta turn (Tt)	26	8.05%
Random coil (Cc)	127	39.32%

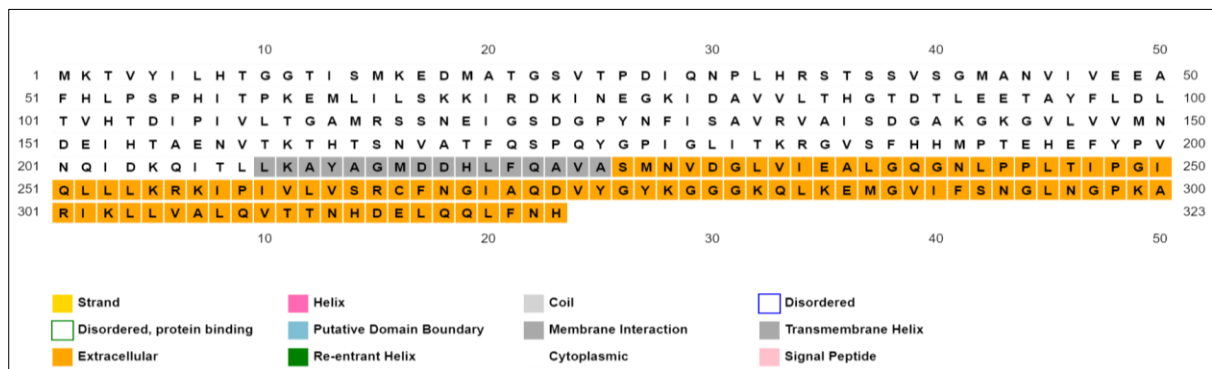
SOPMA calculates the %age of  $\alpha$ -helices,  $\beta$ -sheets and loops present in between these  $\alpha$ -helices and  $\beta$ -sheets that is also called coils. SOPMA is using a homology approach for the prediction of secondary structure. There are 70 amino acid residues in each line because it shows output width as 70 by default (Figure 23). Graphical representation of  $\alpha$ -helix (blue), extended strand (red), beta-turn (green) and random coils (yellow) shown in different colors (Figure 24), [126], [122]



**Figure 24** Graphical representation of secondary structure by using SOPMA.

The graphical representation shows the number of residues of helices, sheets and coils present in the BmA sequence.

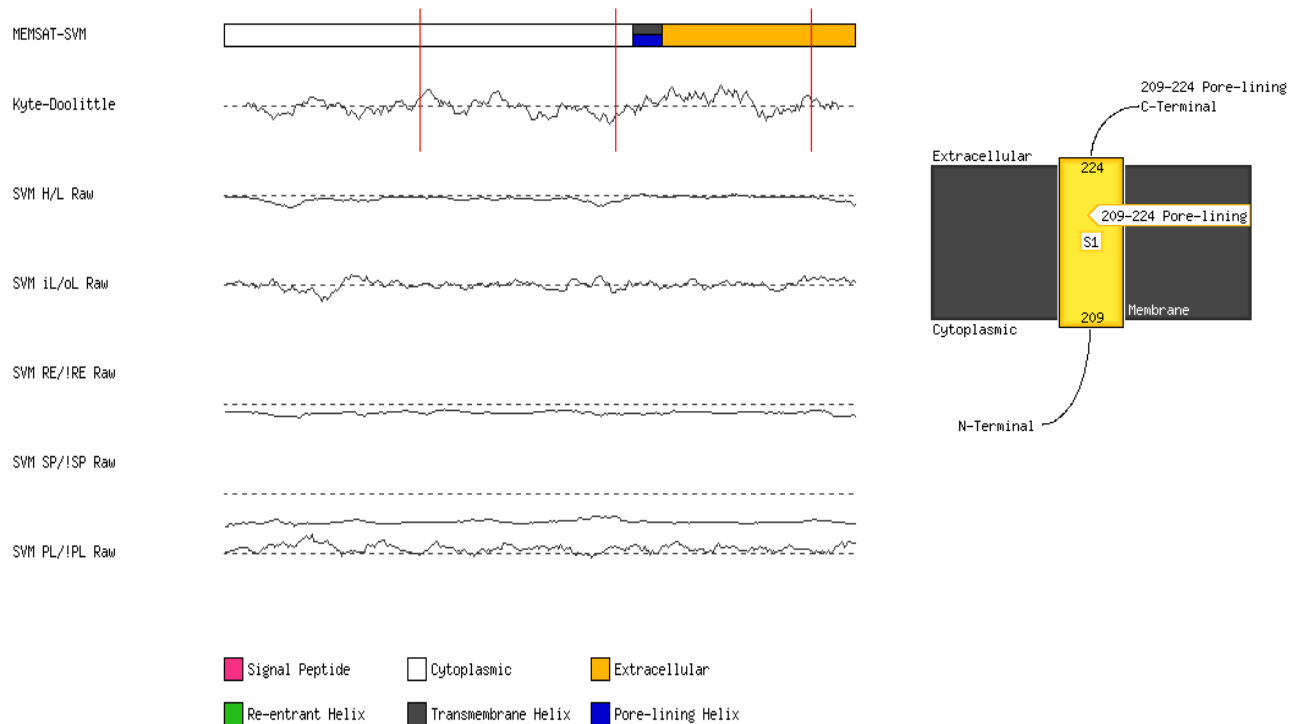
❖ **Secondary structure map (amino acid position)**



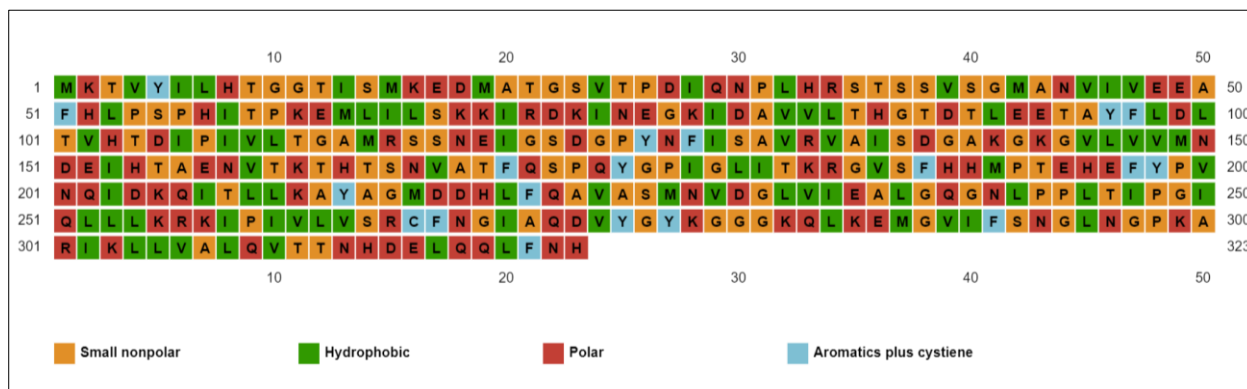
**Figure 25** Secondary structure map representing the position of amino acids.

This orange region shows an extracellular region of amino acid and the grey colored amino acid residues are belonging to the trans-membrane helix (as shown in **figure 25**). That means BmA have certain interaction with the membrane.

The white region represents the cytoplasmic amino acids and is an extracellular region shown (**figure 26**) by orange. Similarly, the blue-colored region shows the gap present in the membrane.



**Figure 26** Graphical representation of amino acid position (trans-membrane, extracellular helices region)



**Figure 27** Secondary structure map shows a hydrophobic and hydrophilic region of amino acid present in the BmA sequence.

In this secondary structure map, there are four distinct groups of amino acids at the molecular behavior level: non-polar, polar, hydrophobic, and aromatic amino acids. As shown (in **figure 27**) green region shows hydrophobic amino acids and the red region shows a hydrophilic region of the amino acids. As you have shown in the map, non-polar (orange region) and aromatic (blue region) amino acids are very small in number.

**Table 26** gives the data of all the selected sequences, the number of residues to the total residues present in the secondary structure (helix, sheet and coils) of the sequence.

**Table 26** Secondary structures of selected all sequences by using CFSSP.

Sr. No.	Protein Accession Number	Organism Name	Helix (%)	Sheet (%)	Turn (%)
1	KNH26430.1	<i>B. megaterium</i>	74.6	46.1	12.4
2	KFF56126.1	<i>B. subtilis</i>	79.9	50.8	12.5
3	KLV14741.1	<i>B. altitudinis</i>	76.9	46.2	12.5
4	TKI83183.1	<i>B. mycoides</i>	77.8	54.9	10.2
5	TKI94247.1	<i>B. cereus</i>	77.5	50.6	9.6
6	TKI15105.1	<i>B. wiedmannii</i>	77.2	46.6	9.6
7	AYJ88275.1	<i>B. safensis</i>	77.5	46.2	11.9
8	KLV25750.1	<i>B. circulans</i>	83.7	54.8	12.3
9	RHJ11726.1	<i>B. sonorensis</i>	79.3	43.5	12.8
10	RHL19359.1	<i>B. licheniformis</i>	71.4	69.3	14.3

## 5.6 Functional Analysis

### a) Determination of the motif present in the BmA sequence.

Motifs are the conserved region of the sequence or protein which determines the function of the protein. Two domains are present in the *B. megaterium* asparaginase sequence, i.e. TPF00710, Asparaginase, N-terminal, and PF17763, Asparaginase C-terminal domain, as shown in **Figure 28**.

Two motifs are existing in sequences through a number of motifs 10, a min. width of 100 and maximum width of 150. Conserved sequences in these motifs can be used to design definite degenerate primers to identify the type, isolate, and a class of asparaginase. Here, all the sequences under study are found to be soluble in nature [109].

KEY:  Family Pfam A domain  Secondary structure

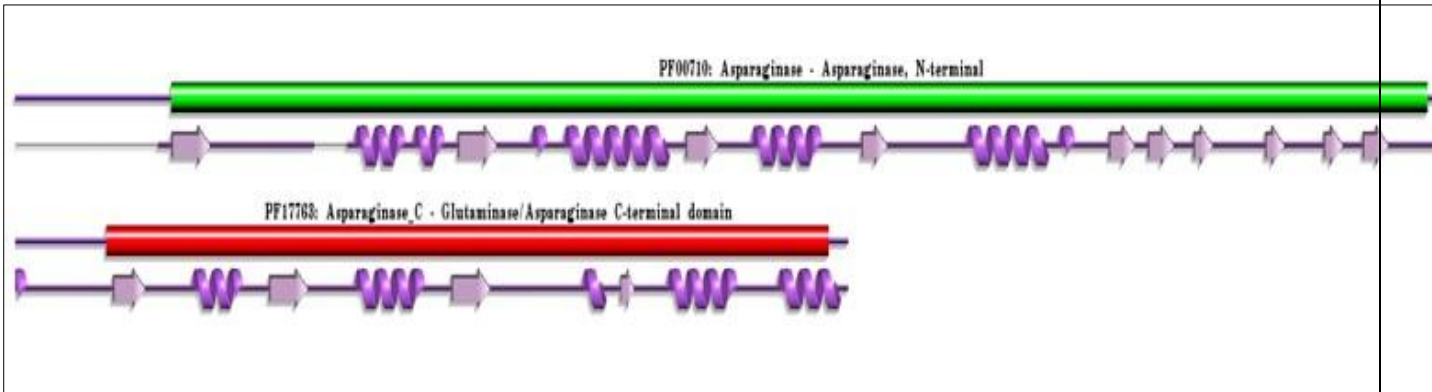


Figure 28 Motif present in the BmA sequence.

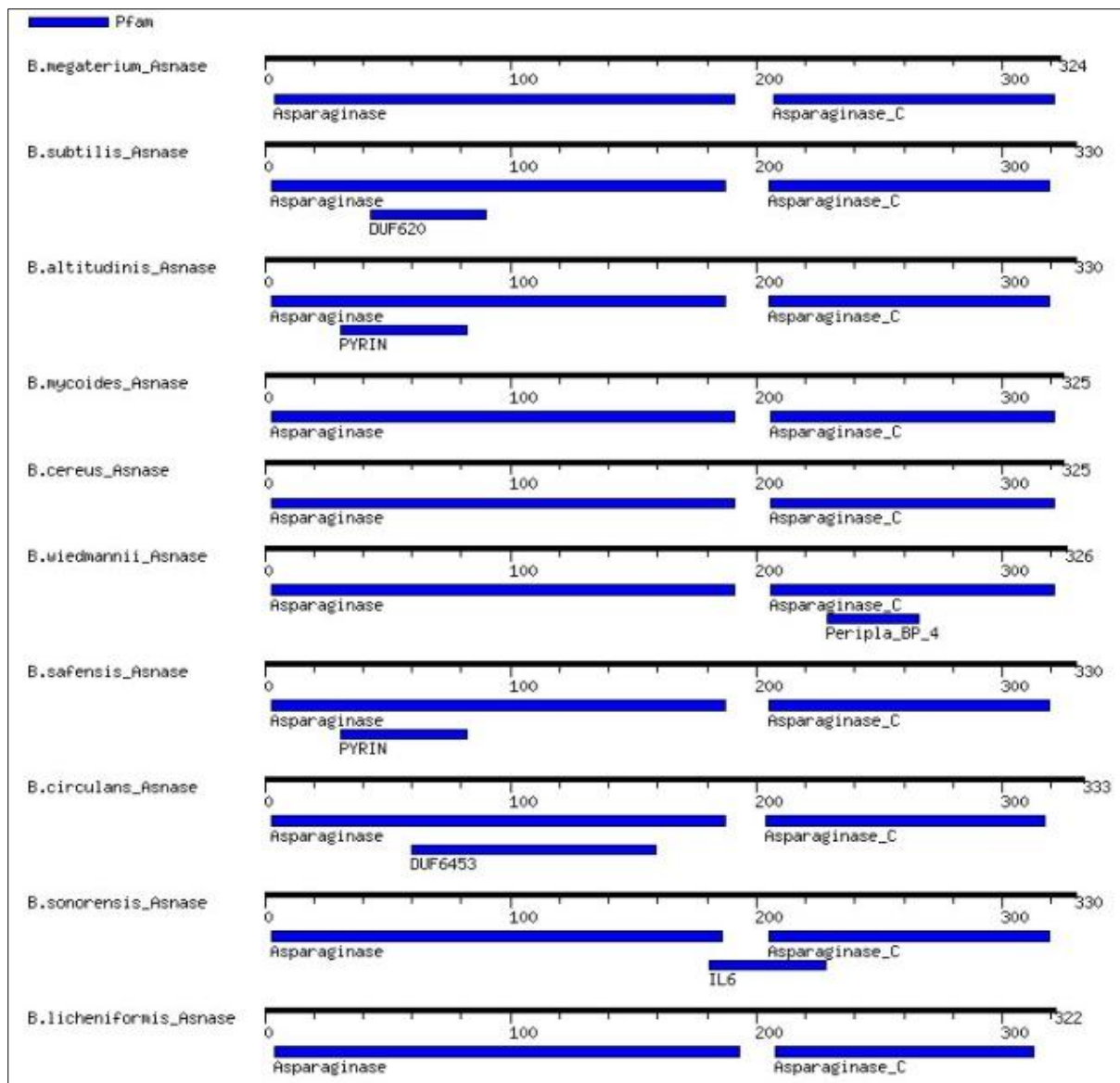


Figure 29 Identified motif graphical details.

**Table 27** Multi-level consensus sequence of motifs

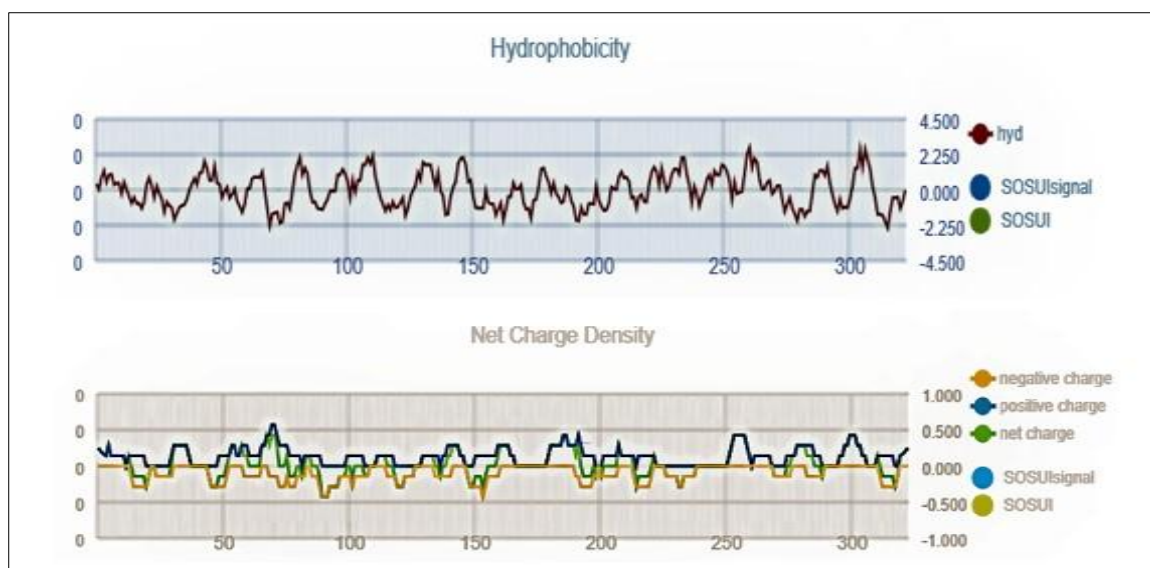
Motif No.	Multi-level consensus sequence
1.	MNKKVALITGGTIASRKTESGRLAAGAISGPELAEMCSLPEDVQIDVY PAFQLPSMHITFEHLLELKQTIERVFQNGGYDGAVVTHGTDLTLEETAYF LDLTIEDERPVVVTGSQRAPEQQGTDAYTNIRHAVYTACSPDIKGAGTV VVFNERIFNARYVKKVHASNLQGFDFVFGFGYLGIIIDNDKVVYVYQ
2.	AVDIVKCYLDGDGKFIIRAAVREGVEGIVLEGVGRGQVPPNMMADIEQA LNQGVYIVITTSAAEEGEVYTTYDYAGSSYDLAKKGVILGKDYDSKKAR MKLAVLLASYKEGIKDKFCY

**b) Determination of the hydrophobicity of BmA using SOSUI Server**

The average hydrophobicity of *B. megaterium* is - 0.043963, suggesting that BmA is a soluble protein and has a cytoplasmic subcellular localization site (**Table 28, Figure 30**).

**Table 28** Details of the hydrophobicity of BmA using SOSUI Server.

Organism	Total Length	Average of Hydrophobicity	Subcellular localization site	Nature of protein
<i>B. megaterium</i>	323 aa	-0.043963	Cytoplasmic	Soluble protein



**Figure 30** Graphical representation of hydrophobicity of *B. megaterium* and their net charge density.

### 5.7 Structural Classification

Structural classification of BmA was done using CATH server, which categorizes the protein based on four different groups; these are (C) represents the class, (A) represents the architecture, (T) represents the topology and (H) represents the homologous family [109]. CATH server categorized BmA protein as an alpha-beta class, 3-layer sandwich architecture, Rossmann fold topology and the homologous family is L-Asparaginase (N-terminal domain).

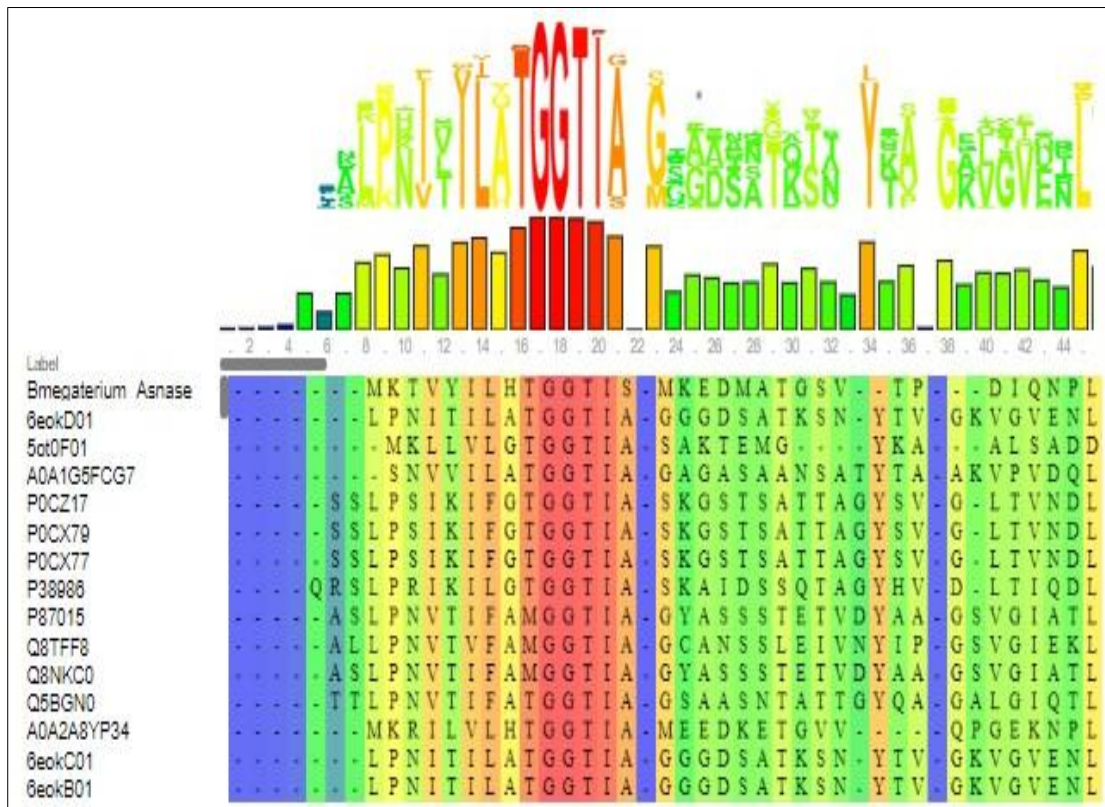
**Table 29** Classification of BmA protein.

Level	Description
C	Alpha Beta
A	3 layer sandwich
T	Rossmann fold
H	L- Asparaginase, N- terminal domain

For the classification of BmA protein, the CATH server firstly predicted its secondary structure based on sequence alignment (**Figure 31**)

## Prediction of the secondary structure of BmA by using CATH Server

Match	E- Value
Cytoplasmic L-asparaginase I	1.3e-66



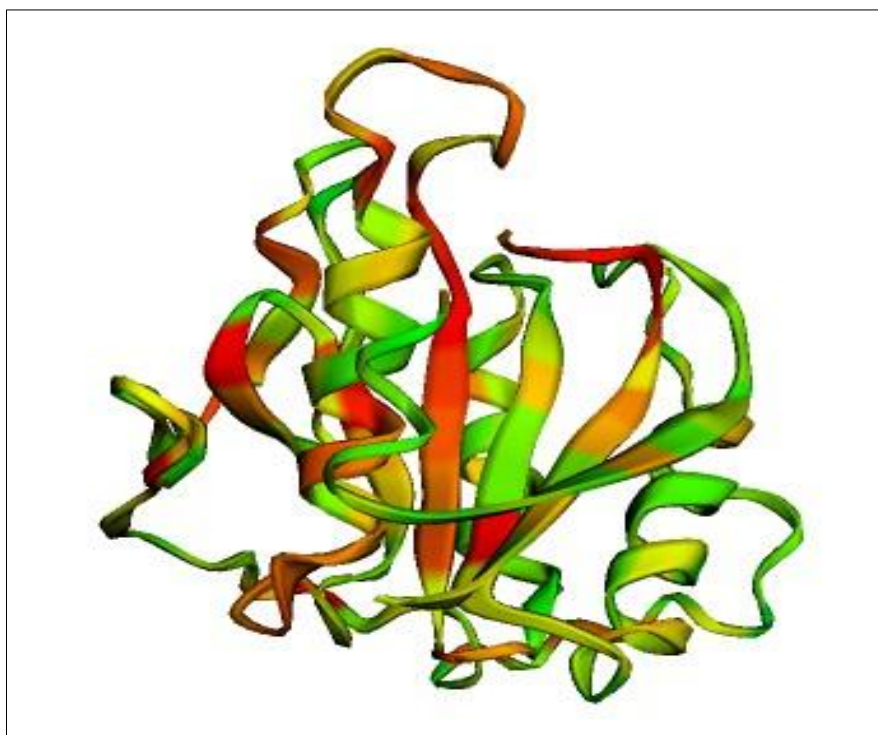
**Figure 31** Sequence alignment of BmA.

Further CATH server superposes four representative domains within this superfamily when choosing how to superpose the structure; more emphasis is given to structurally similar residues (according to SSAP) to generate a superfamily superposition structure of BmA (Figure 32). The structure highlights a highly conserved "core" even in super-families with substantial structural diversity. At the same time, it generates 3D structure of BmA (**Figure 32**).





**Figure 32** Superfamily Superposition structure of BmA using CATH server.



**Figure 33** Ribbon (3-D) Structure of BmA generated by CATH server.

## 5.8 Homology Modeling of BmA

The comparative protein model of BmA was built through the SWISS-Model using the selected template. *Helicobacter pylori* has a 33.12% similarity with *B. megaterium* asparaginase sequence. There are four ligands (aspartic acid) present in this predicted structure and they have 0.76 GMQE values and -1.95 QMEAN values, respectively. All the data related to homology modeling are given in **Table 30**.

### 5.8.1 SWISS-Model

#### Model-Template Alignment

Model_01:A	MKTVYLLPTGGTISMKDMAT-SSVTF-DI-QNPLHRSTSSVSGMANVIVEEAFHPPSPHITPKEMLILSKKIRDKINEG	77
Model_01:B	MKTVYLLPTGGTISMKDMAT-SSVTF-DI-QNPLHRSTSSVSGMANVIVEEAFHPPSPHITPKEMLILSKKIRDKINEG	77
Model_01:C	MKTVYLLPTGGTISMKDMAT-SSVTF-DI-QNPLHRSTSSVSGMANVIVEEAFHPPSPHITPKEMLILSKKIRDKINEG	77
Model_01:D	MKTVYLLPTGGTISMKDMAT-SSVTF-DI-QNPLHRSTSSVSGMANVIVEEAFHPPSPHITPKEMLILSKKIRDKINEG	77
2wlt.1.A	LPTIALLDTGGTLAGGVVDASLGSVSGGELGPKELLRKIPSLNKIARIQGEQVSNIGSQDMNKEEIFKLAQRAQELDDSDS	84
Model_01:A	IDAVVLTHTGDTLEETAYFLDLTVHDDIPIVLTGAVRSSNEIGSDGPYNFISAVRVAISDGAKGKGVLVVMNDETHTAE	157
Model_01:B	IDAVVLTHTGDTLEETAYFLDLTVHDDIPIVLTGAVRSSNEIGSDGPYNFISAVRVAISDGAKGKGVLVVMNDETHTAE	157
Model_01:C	IDAVVLTHTGDTLEETAYFLDLTVHDDIPIVLTGAVRSSNEIGSDGPYNFISAVRVAISDGAKGKGVLVVMNDETHTAE	157
Model_01:D	IDAVVLTHTGDTLEETAYFLDLTVHDDIPIVLTGAVRSSNEIGSDGPYNFISAVRVAISDGAKGKGVLVVMNDETHTAE	157
2wlt.1.A	RIQGGVLTHTGDTLEESAYFLNLVHSTKPEVVLVQAMRNASSLSADGALNLYEAVSVAVNEKSANKGVLVVMNDETHTAE	164
Model_01:A	NVTKTHSNVATFOSPOYGPIGLITKRGVSHHMPDEH-----EFYE--VNQIDKQITLLKAYAGMDHLEQAVASMNVD	230
Model_01:B	NVTKTHSNVATFOSPOYGPIGLITKRGVSHHMPDEH-----EFYE--VNQIDKQITLLKAYAGMDHLEQAVASMNVD	230
Model_01:C	NVTKTHSNVATFOSPOYGPIGLITKRGVSHHMPDEH-----EFYE--VNQIDKQITLLKAYAGMDHLEQAVASMNVD	230
Model_01:D	NVTKTHSNVATFOSPOYGPIGLITKRGVSHHMPDEH-----EFYE--VNQIDKQITLLKAYAGMDHLEQAVASMNVD	230
2wlt.1.A	EVVQTHHTHVSTFKALNSGMLGSMVYCKTRYVQPLRKHHTTESFSLSQLKTPLPVVDIIVTPAGMTEDLFOASLNSHAK	244
Model_01:A	GLVIEALGQGNLPPLDIPGIQLLKRKIPIVLVSBFCNGIAODVYGYKGGKQLKEMGVI(SNGLN)GPKARPKLLVALQV	310
Model_01:B	GLVIEALGQGNLPPLDIPGIQLLKRKIPIVLVSBFCNGIAODVYGYKGGKQLKEMGVI(SNGLN)GPKARPKLLVALQV	310
Model_01:C	GLVIEALGQGNLPPLDIPGIQLLKRKIPIVLVSBFCNGIAODVYGYKGGKQLKEMGVI(SNGLN)GPKARPKLLVALQV	310
Model_01:D	GLVIEALGQGNLPPLDIPGIQLLKRKIPIVLVSBFCNGIAODVYGYKGGKQLKEMGVI(SNGLN)GPKARPKLLVALQV	310
2wlt.1.A	GVVIALVGNQVNSAGFLKAMQEASOMGVIVVRSRVGSGGVVTSG-----EIDDKAYGEISDNLNPKARVLLQALTK	318
Model_01:A	TTNHDELQQLFNH	323
Model_01:B	TTNHDELQQLFNH	323
Model_01:C	TTNHDELQQLFNH	323
Model_01:D	TTNHDELQQLFNH	323
2wlt.1.A	TNDKAKIQEMFEE	331

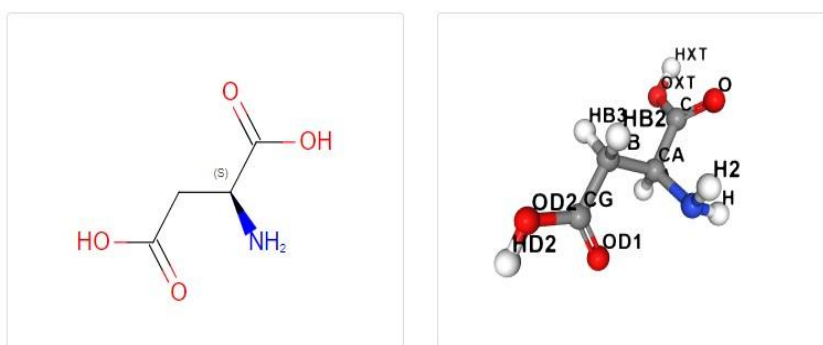
**Figure 34** Sequence alignment of BmA with the selected template *Helicobacter pylori* asparaginase

**Table 30** Details of template *H. pylori* asparaginase with respect to BmA.

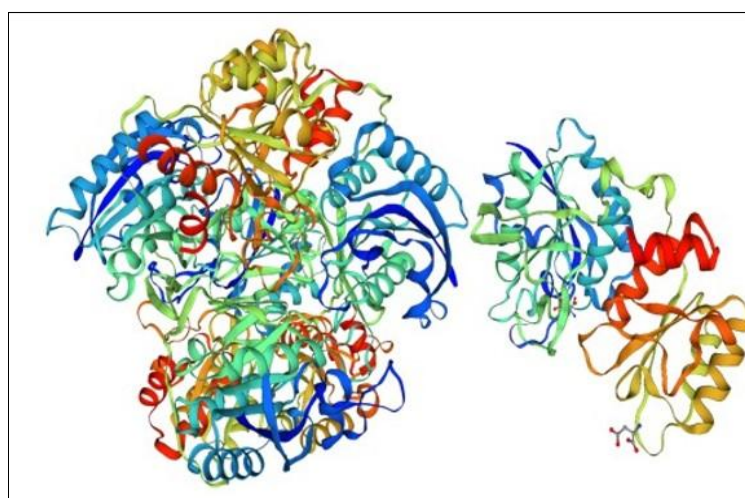
Template	Coverage	GMQE	QMEAN	Identity	Ligands	Oligo State	Method
<i>Helicobacter pylori</i>	1-323 (0.98%)	0.76	-1.95	33.12%	4x Aspartic acid	Homo-tetramer	X-Ray Diffraction 1.40 Å

**Table 31** *H. pylori* asparaginase ligands PLIP interactions with respect to BmA

Ligands (4x Aspartic acid)	Number of Residues (within 4 Å)	PLIP Interactions
ASP.1	13	15
ASP.2	13	16
ASP.3	13	16
ASP.4	13	15



**Figure 35** Chemical structure of aspartic acid



**Figure 36** Predicted protein structure of BmA based on sequence homology

## 5.8.2 Quality estimation of predicted protein

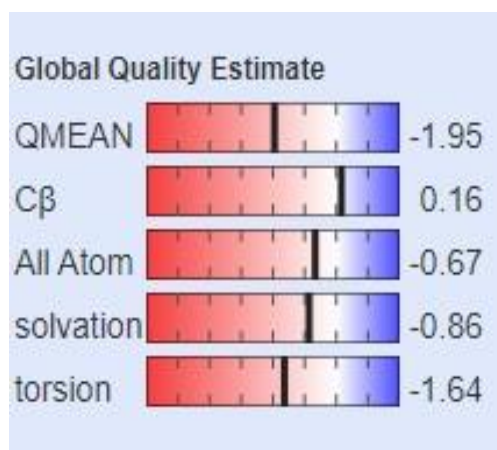


Figure (a)



Figure (b)

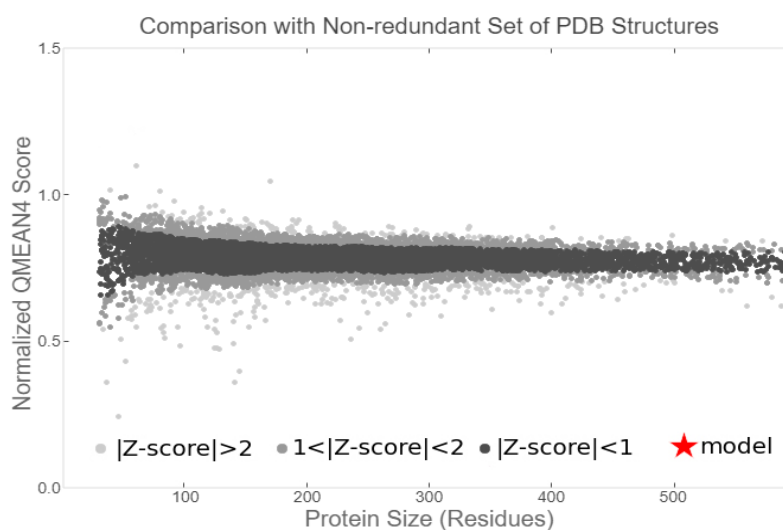


Figure (c)

**Figure 37** Quality examination of predicted protein from different servers: **Figure (a)** Global estimation score of predicted protein, **Figure (b)** Using QMEAN server estimate the local quality of the predicted protein structure and **Figure (c)** Built model comparison with a non-redundant set of PDB structures from QMEAN server.

The Swiss model web-based tool calculates the QMEAN score function to estimate global and local model quality based on the geometry interaction and the solvent potential of a query protein model. The global estimation score of predicted protein over here provided the individual scores for each of the different quality estimations given in **figure 37 (a)**. Predicted local quality estimate similarity to the crystal structure that has been used as a target template. Z-score ranges from 0 to 1 (Z-score which are compared to the expected

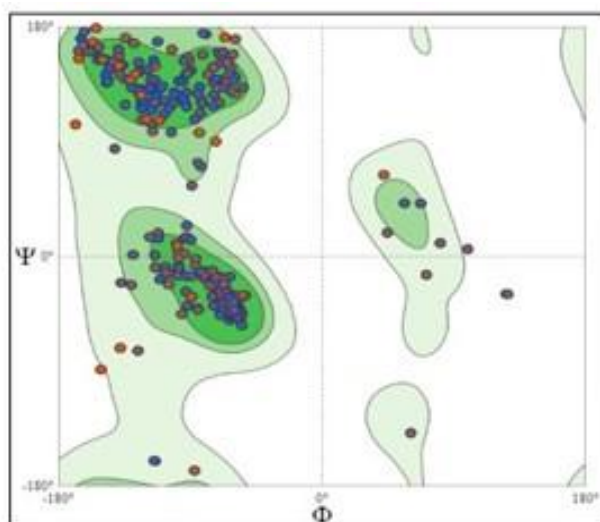
value from the native structure). If the value close to 1, then the quality of the particular region is higher. As shown in **figure 37 (b)**, most of the residues are below 1, which means the quality of the amino acid residues in the regions is not good.

**Figure 37 (c)** compares the predicted protein model to the other protein crystal structure in the PDB. Since the red star lies in the white region (outside the other crystal structures) of the plot, the structure built may be unreliable or less stable.

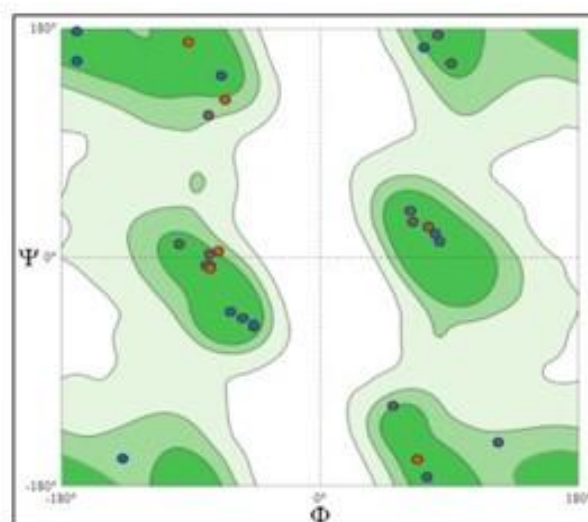
**RAMPAGE** provides the stereochemical properties of the predicted protein structure (**Figure 38**). RAMPAGE also provides the residues percentage in various regions such as favored region, outlier region and allowed region. If more residues are in the favored region, the protein structure will be more stable.

**Table 32** Structure Assessment of predicted protein from RAMPAGE.

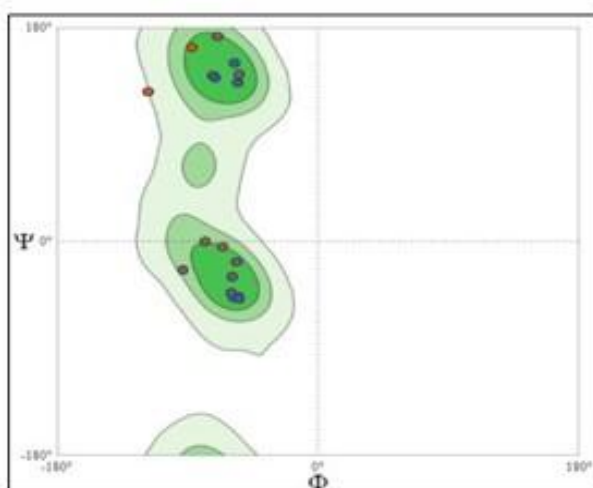
<b>MolProbity Score</b>	<b>2.05</b>
<b>Clash Score</b>	12.94
<b>Favored region</b>	94.08%
<b>Outliers region</b>	1.56%
<b>Rotamer Outliers</b>	1.09%
<b>C-<math>\beta</math> Deviations</b>	20
<b>Bad Bonds</b>	0/10068
<b>Bad Angles</b>	120/13660
<b>Cis Non-Proline</b>	12/1224
<b>Twisted Non-Proline</b>	4/1224



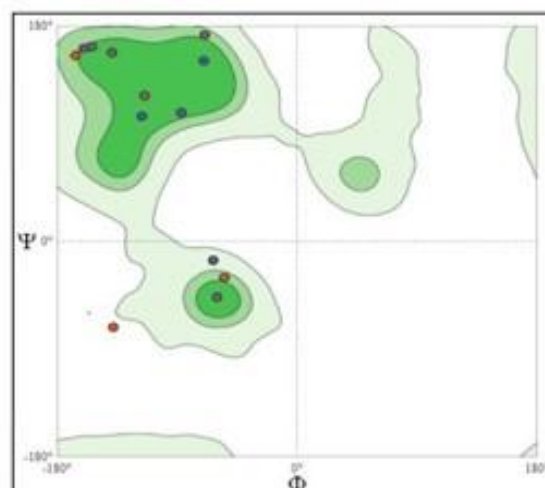
**Figure (a)**



**Figure (b)**



**Figure (c)**



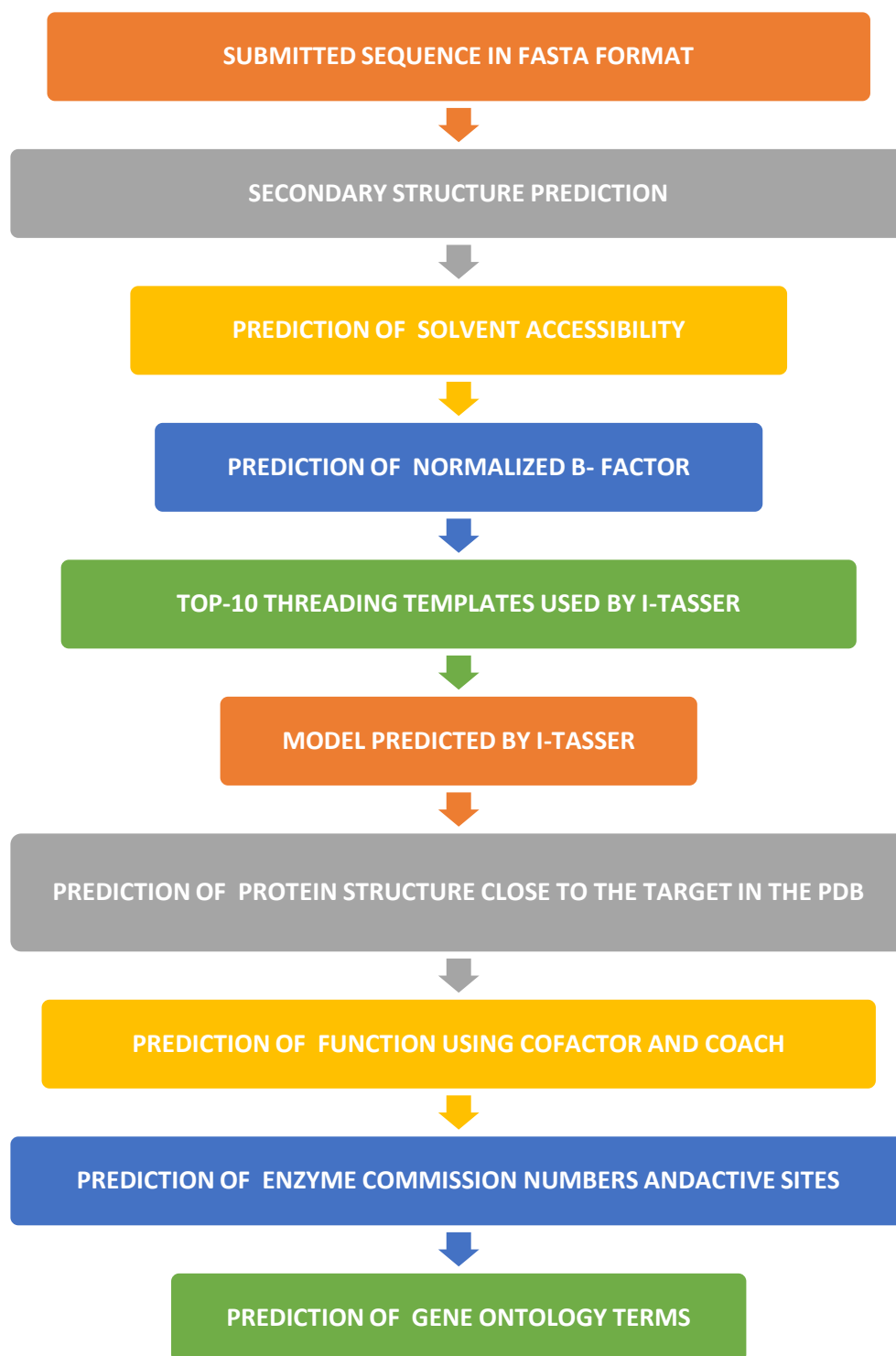
**Figure (d)**

**Figure 38** Ramachandran plot generated by RAMPAGE (**Figure a**) Show general Ramachandran plot, (**figure b**) No. of glycine residues, (**Figure c**) No. of proline residues, and (**Figure d**) no. of Pre-Proline residues.

We conclude that the structure we get from homology modeling was not a good quality model from this data. So we further perform threading for model building and evaluation [127]

### 5.9 Threading of BmA sequence

The detailed steps involved in predicting the BmA structure using I-TASSER is given in **Figure 39**.



**Figure 39** Steps followed by I-TASSER for protein prediction

**Figure 40 (b)** shows the secondary structure elements of BmA based on sequence alignment. Here, H represents the alpha helices and S represents the beta sheets and C represents the coils (which is present between helices and sheets). If the secondary structure score is close

to 9, the predicted structure is quite reliable, but if the score is closed to 0-2, we can't rely on the structure.

There are 232 amino acid residues present in the BmA sequence. Some of the sections of these predicted secondary structures are more reliable than the residues of the other regions (as shown in **Figure 40 (b)**).

Solvent accessible protein is the surface area of an amino acid that is accessible to a solvent. If the predicted score of the solvent accessibility to the residue is close to 9, the residues are highly exposed and so present over the protein's surface. If the score is close to zero, that means the residues are buried in the core of the protein. As shown in **figure 40 (c)**, many of the residues have the predicted solvent accessibility score is close to 0, which means these residues are buried in the 3D structure of the protein and make the core of the protein.

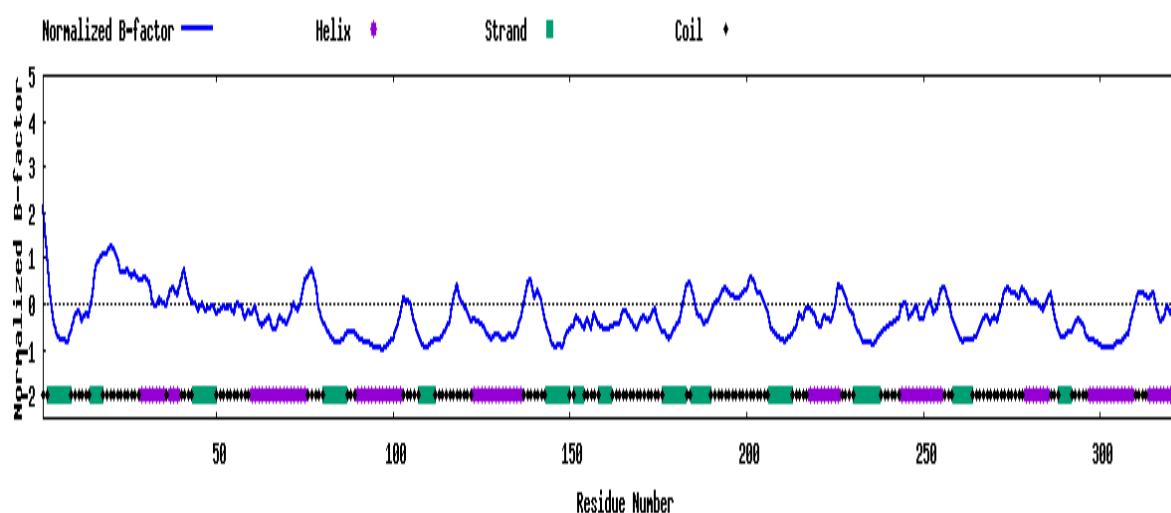


**Figure 40** Threading of BmA sequence for protein structure prediction: **Figure (a)** FASTA Format of BmA sequence, **Figure (b)** sequence-based prediction of secondary structure and **Figure (c)** Predicted solvent accessibility.



B-factor is a value to indicate the extent of the inheritance thermal mobility of residues/atoms in proteins. Amino acid absorbs heat energy to convert into kinetic energy, making the amino acid mobile; that phenomenon is called thermal mobility.

In this graph, alpha-helices (red region), beta sheets (green region) and coils (black region) are shown (**Figure 41**). In the beginning, there are two coils present and these coils have a high B-factor that means they are more mobile as they move towards alpha helices and beta sheets. This blue line falls to the negative value and when it reaches again in the coil region, the mobility of the amino acid or residues is greater. That means the protein structure is more flexible in the coil region, making the amino acid residues more mobile. Based on the B-factor profile (BFP) distributions and predictions, residues with BFP values higher than 0 are less stable in experimental structures [124].



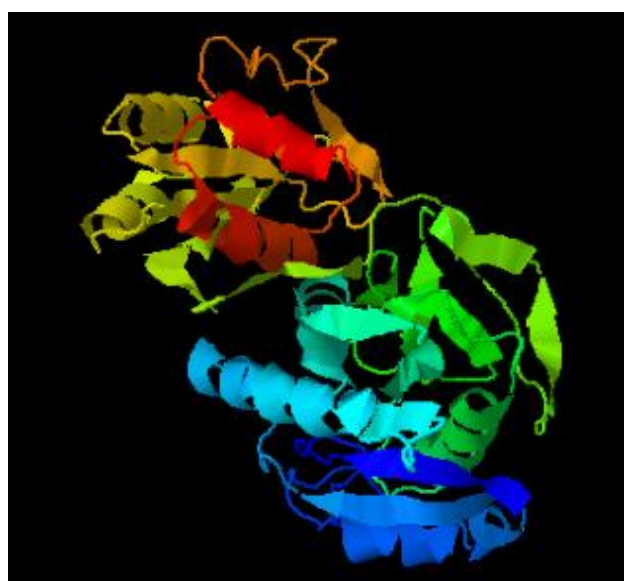
**Figure 41** Predicted Normalized B-Factor

I-TASSER picks ten templates from the PDB, which are pretty close to the query protein. By using those templates, it can predict protein models. These templates are used for the prediction of the protein model. If Z-score is higher than 1, the predicted models are quite close to the native one. If it is less than 1, that means the predicted structure is not reliable.

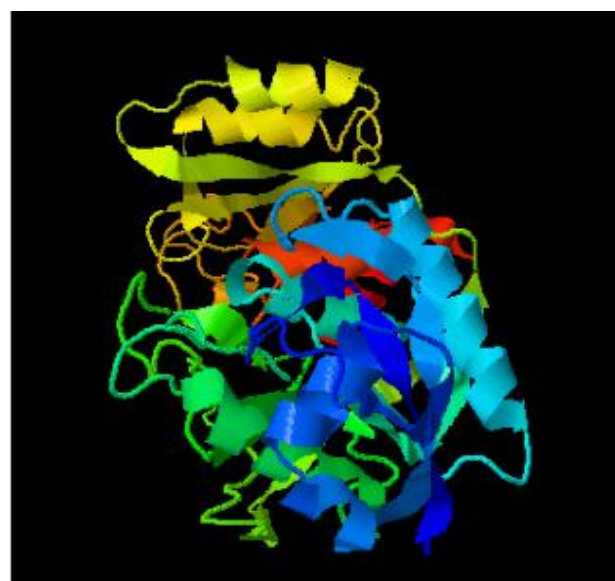
The predicted structures obtained from I-TASSER show Z-score greater than one (as shown in **figure 42**), which means the predicted structure may be very close to the native structure of BmA.

Rank	PDB Hit	Iden1	Iden2	Cov	Norm. Z-score	Download Align.	20	40	60	80
							Sec.Str C C S S S S S S C C C C C C S S S C C C C C C C C C H H H H H H H C H H C C C C S S S S S S C C C C C C C C C H H H H H H H H H H H H C C C C C S S S S Seq M K T V Y I L H T G G T I S M K E D M A T G S V T P D I Q N P L H R S T S S V S G M A N V I V E E A F H L P S P H I T P K E M L I L S K K I R D K I N E G K I D A V V L			
1	<a href="#">2witA</a>	0.32	0.33	0.98	2.70	<a href="#">Download</a>	L P T I A L L A T G G T I A G S G A S L G S Y S G E L G V K E L L K A I P S L N K I A R I Q G E Q V S N I G S Q D M N E E I W F K L A Q R A Q E L L D D S R I Q G V V I	T S G E I D D K A Y G F I T S D N L N P Q K A R V L L Q L A L T K T N D K A K I Q E M F E E		
2	<a href="#">1wsaA</a>	0.33	0.33	0.97	4.08	<a href="#">Download</a>	K P Q V T I L A T G G T I A G - - - - - Y S A G A V T V D K L L A A V P A I N D L A T I K G E Q I S S I G S Q E M T G K V W L K L A K R V N E L L A Q K E T E A V I I	D D K L G F V A T E S L N P Q K A R V L L M L A L T K T S D R E A I Q K I F S T		
3	<a href="#">5k3oA</a>	0.32	0.34	0.98	3.71	<a href="#">Download</a>	K P Q V T I L A T G G T I A G S G E S S V Y S A G A V T V D K L L A A V P A I N D L A T I K G E Q I S S I G S Q E M T G K V W L K L A K R V N E L L A Q K E T E A V I I	D K K L G F V A T E S L N P Q K A R V L L M L A L T K T S D R E A I Q K I F S T		
4	<a href="#">2wit</a>	0.32	0.33	0.98	2.81	<a href="#">Download</a>	L P T I A L L A T G G T I A G S G A S L G S Y S G E L G K E L L K A I P S L N K I A R I Q G E Q V S N I G S Q D M N E E I W F K L A Q R A Q E L L D D S R I Q G V V I	D D K A Y G F I T S D N L N P Q K A R V L L Q L A L T K T N D K A K I Q E M F E E		
5	<a href="#">2wit</a>	0.32	0.33	0.98	2.05	<a href="#">Download</a>	L P T I A L L A T G G T I A G S G A S L G S Y S G L G V K E L L K A I P S L N K I A R I Q G E Q V S N I G S Q D M N E E I W F K L A Q R A Q E L L D D S R I Q G V V I	D D K A Y G F I T S D N L N P Q K A R V L L Q L A L T K T N D K A K I Q E M F E E		
6	<a href="#">2witA</a>	0.32	0.33	0.98	3.43	<a href="#">Download</a>	L P T I A L L A T G G T I A G S G A S L G Y K S G E L G V K E L L K A I P S L N K I A R I Q G E Q V S N I G S Q D M N E E I W F K L A Q R A Q E L L D D S R I Q G V V I	D K A Y G F I T S D N L N P Q K A R V L L Q L A L T K T N D K A K I Q E M F E E		
7	<a href="#">2wit</a>	0.32	0.33	0.98	3.09	<a href="#">Download</a>	L P T I A L L A T G G T I A G S G A S L G S Y S G L G V K E L L K A I P S L N K I A R I Q G E Q V S N I G S Q D M N E E I W F K L A Q R A Q E L L D D S R I Q G V V I	D D K A Y G F I T S D N L N P Q K A R V L L Q L A L T K T N D K A K I Q E M F E -		
8	<a href="#">2witA</a>	0.21	0.33	0.98	6.21	<a href="#">Download</a>	L P T I A L L A T G G T I A G S G A S L G Y K S G E L G V K E L L K A I P S L N K I A R I Q G E Q V S N I G S Q D M N E E I W F K L A Q R A Q E L L D D S R I Q G V V I			
9	<a href="#">2witA</a>	0.32	0.33	0.98	2.94	<a href="#">Download</a>	L P T I A L L A T G G T I A G S G A S L G S Y S G E L G K E L L K A I P S L N K I A R I Q G E Q V S N I G S Q D M N E E I W F K L A Q R A Q E L L D D S R I Q G V V I	T S G E I D D K A Y G F I T S D N L N P Q K A R V L L Q L A L T K T N D K A K I Q E M F E E		
10	<a href="#">2witA</a>	0.32	0.33	0.98	4.37	<a href="#">Download</a>	L P T I A L L A T G G T I A G S G A S L G S Y S G E L G K E L L K A I P S L N K I A R I Q G E Q V S N I G S Q D M N E E I W F K L A Q R A Q E L L D D S R I Q G V V I	D K A Y G F I T S D N L N P Q K A R V L L Q L A L T K T N D K A K I Q E M F E E		

**Figure 42** Top 10 template-query alignments generated by LOMETS.



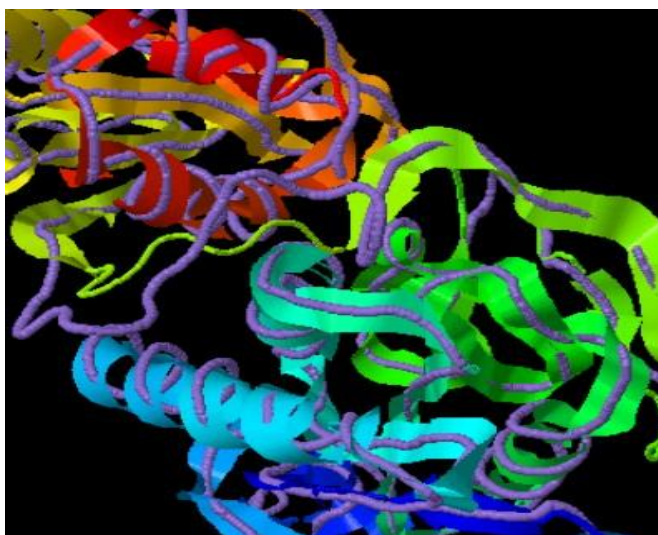
C-score= 1.22



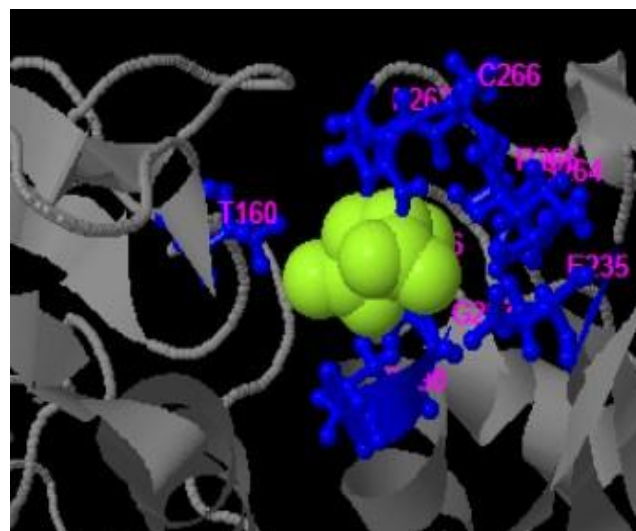
C-score = -1.26

**Figure 43** Predicted 3D model and the estimated global and local accuracy.

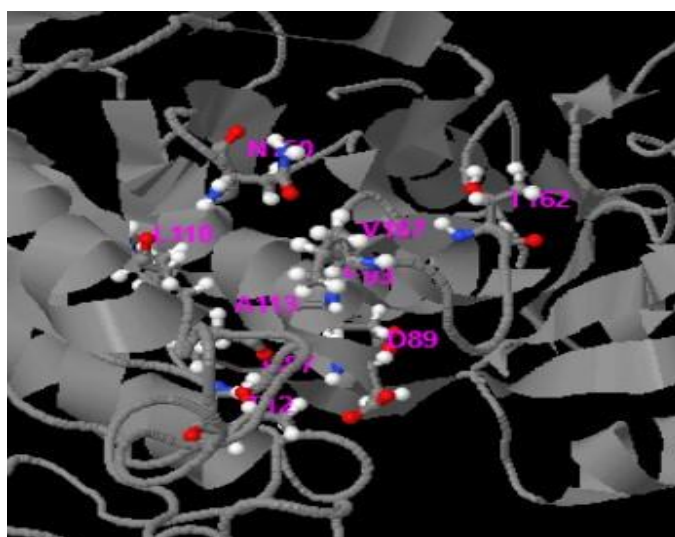
C- Score range lies between [-5 to 2] and C-score > -1.5 indicates a model of correct global topology. The structure will be more reliable if it has a positive value, close to 2. The first predicted model has a C-score of 1.22 (**Figure 43**), which means the predicted 3D model has good global and local accuracy.



**Figure (a)** Structure alignment between the first I-TASSER Model and top 10 most similar structure templates in PDB



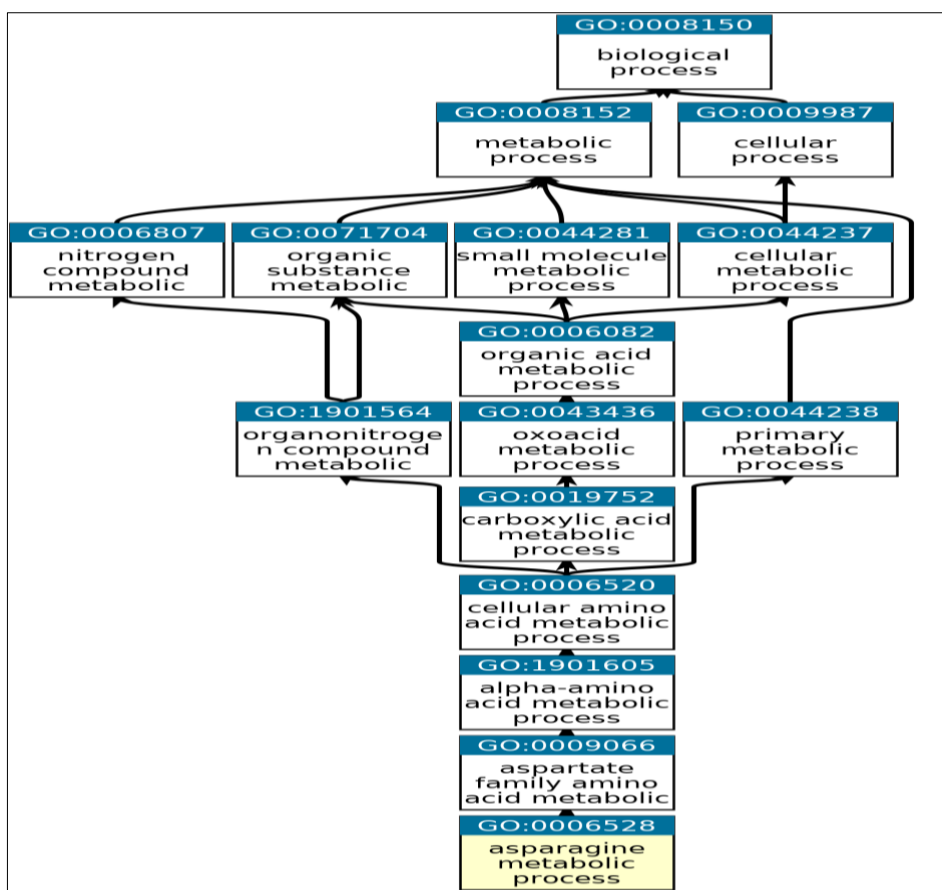
**Figure (b)** Predicted ligand-binding sites.



**Figure (c)** Predicted enzyme commission numbers and active sites

**Figure 44** Tertiary 3D modelled structure of asparaginase of *B. megaterium* viewed by I-TASSER, **Fig (a)** Structure superposition of query protein structure (shown in cartoon) and template proteins (shown in the backbone), **Fig (b)** Predictive binding ligand is shown in the green-yellow sphere, **Fig (c)** Binding residues are shown in blue ball and stick. Predicted active site residues involved are shown in colored balls and stick.

The predictions are made for all the 3 GO categories- molecular function for predicting asparaginase activity, the biological process for predicting asparagine metabolic process (**Figure 45**), and cellular components (Cytosol/ periplasmic space). GO- score should be in the range of 0 to 1; if it's close to 1, that information is more reliable.



**Figure 45** Ancestor chart of asparagine metabolic process.

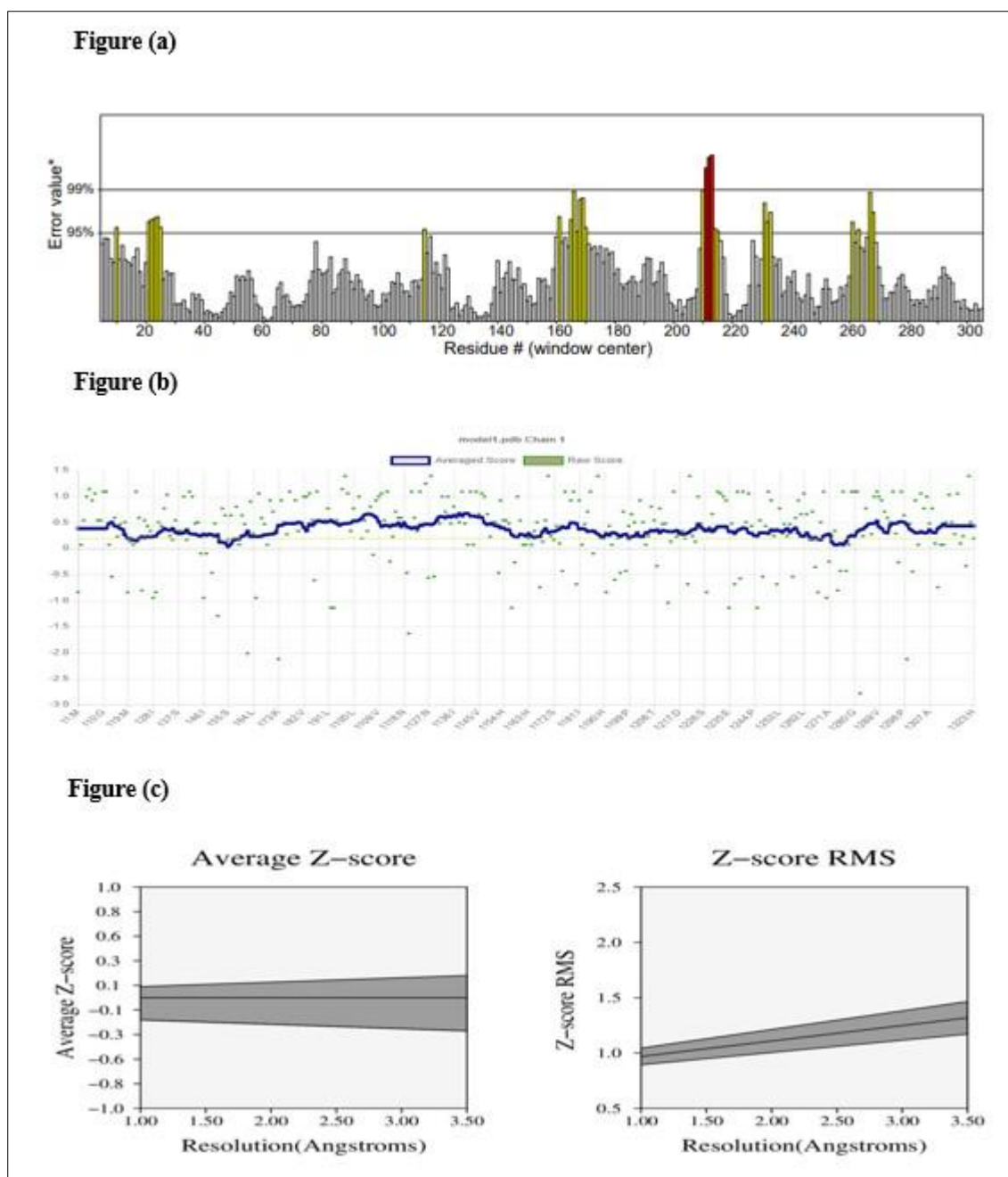
As given in **Table 33** GO score for molecular function, biological process and cellular component are 1.0, 0.99 and 0.70, respectively, which means all the gene ontology functions are reliable.

**Table 33** Consensus prediction derived based on the occurrence of the GO term among the selected templates.

Gene ontology terms	Molecular Function	Biological Process	Cellular Component
<b>GO-Score:</b>	1.00	0.99	0.70/0.70

### 5.10 Evaluation of predicted protein from QMEAN and UCLA—DOE LAB SAVES

A predicted protein structure was evaluated using the two servers, QMEAN and UCLA—DOE LAB SAVES (Verification and a Structure Analysis Server version6) server.



**Figure 46** Evaluation of predicted protein from QMEAN and UCLA—DOE LAB SAVES: **Figure (a)** Graphically represented quality estimation of predicted BmA structure by ERRAT server, **Figure (b)** Averaged score and a raw score of predicted BmA structure by VERIFY3D, and **Figure (c)** Analysis of Entire predicted structure of BmA.

ERROR plot function is based on the statistics of non-bonded atom interaction in the reported structure compared to a database of reliable high-resolution structures. **Figure 46 (a)** shows the overall quality factor of the predicted structure is 91.4286, which means that the predicted protein structure model has been passed.

Verify3D output contains a 3D or 1D profile of the protein model. The 3D- 1D profile describe the residue and environments from a three-dimensional structure to a one-dimensional string. It describes whether the area of the side chain of amino acid residue is buried in the protein and represents the fraction of the side chain area exposed to polar atoms. 94.12 % of the residues have averaged 3D-1D score  $\geq 0.2$  as shown in **Figure 46 (b)**, which means quality estimation from Verify3d of the predicted protein structure has been passed.

PROVE calculates the atom's volume in macromolecules using an algorithm that treats the atoms like hard-sphere and calculates a statistical z-score derivation for the model from high resolution (2.0 Å) and refined (R- factor of 0.2) PDB deposited structure. Buried outlier protein atoms total from 1 Model: 5.6%, which means quality estimation from PROVE has failed (**Figure 46 (c)**).

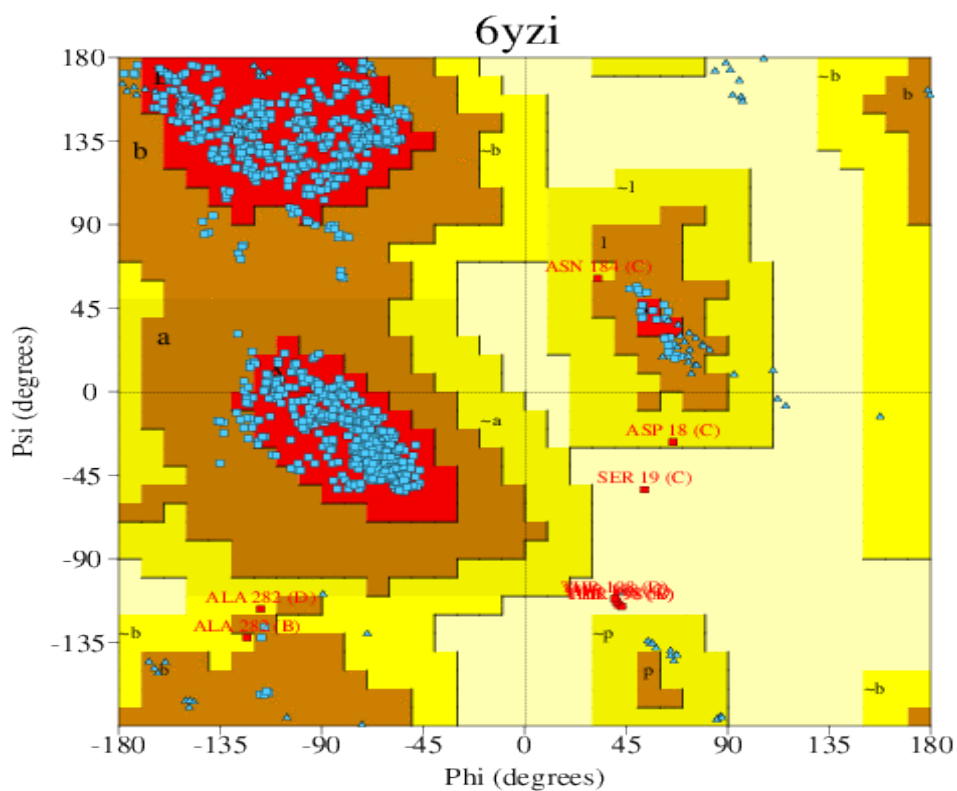
- **Evaluation of predicted protein from PROCHECK**

PROCHECK tool is used to validate the 3-D organization of the predicted protein structures. The program made Ramachandran plot which contains a plot of phi-psi torsional angles, where the red region is the most favored region consisting of more than 90% of residues.

Ramachandran plot demonstrates the phi-psi torsion angles to all amino acids (residues) present in the structure. The colored region shown in the plot represents the diverse regions (**figure 47**) as follow darkest areas (red region) relate to the "core" regions demonstrating the most favorable phi-psi value combinations [127].

**Table 34** Ramachandran plot- most favored, allowed, generously allowed, and disallowed regions percentage score.

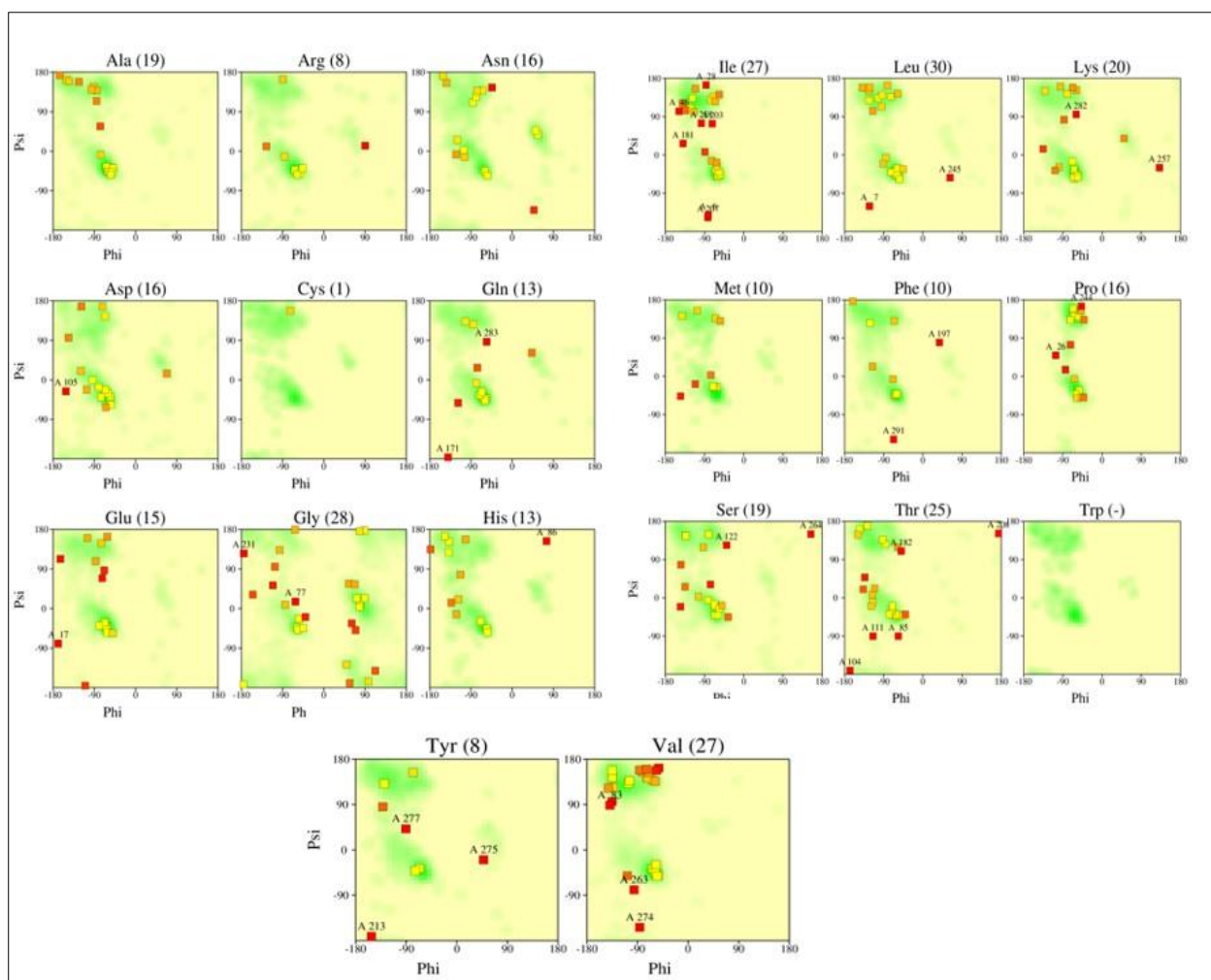
<b>Ramachandran plot statistics</b>	<b>No. of Residue</b>	<b>Score Percentage</b>
<b>Favored regions</b> [A,B,L]	1007	91.4%
<b>Allowed regions</b> [a, b, l, p]	86	7.8%
<b>Generously allowed regions</b> [~a, ~b, ~l, ~p]	4	0.4%
<b>Disallowed regions</b> [XX]	5	0.5%



**Figure 47** Ramachandran plot of Matrix protein of BmA generated using PROCHECK software.

**Table 35** All Ramachandran's, Chi1-chi2 plots and side-chain parameters of BmA generated using PROCHECK software.

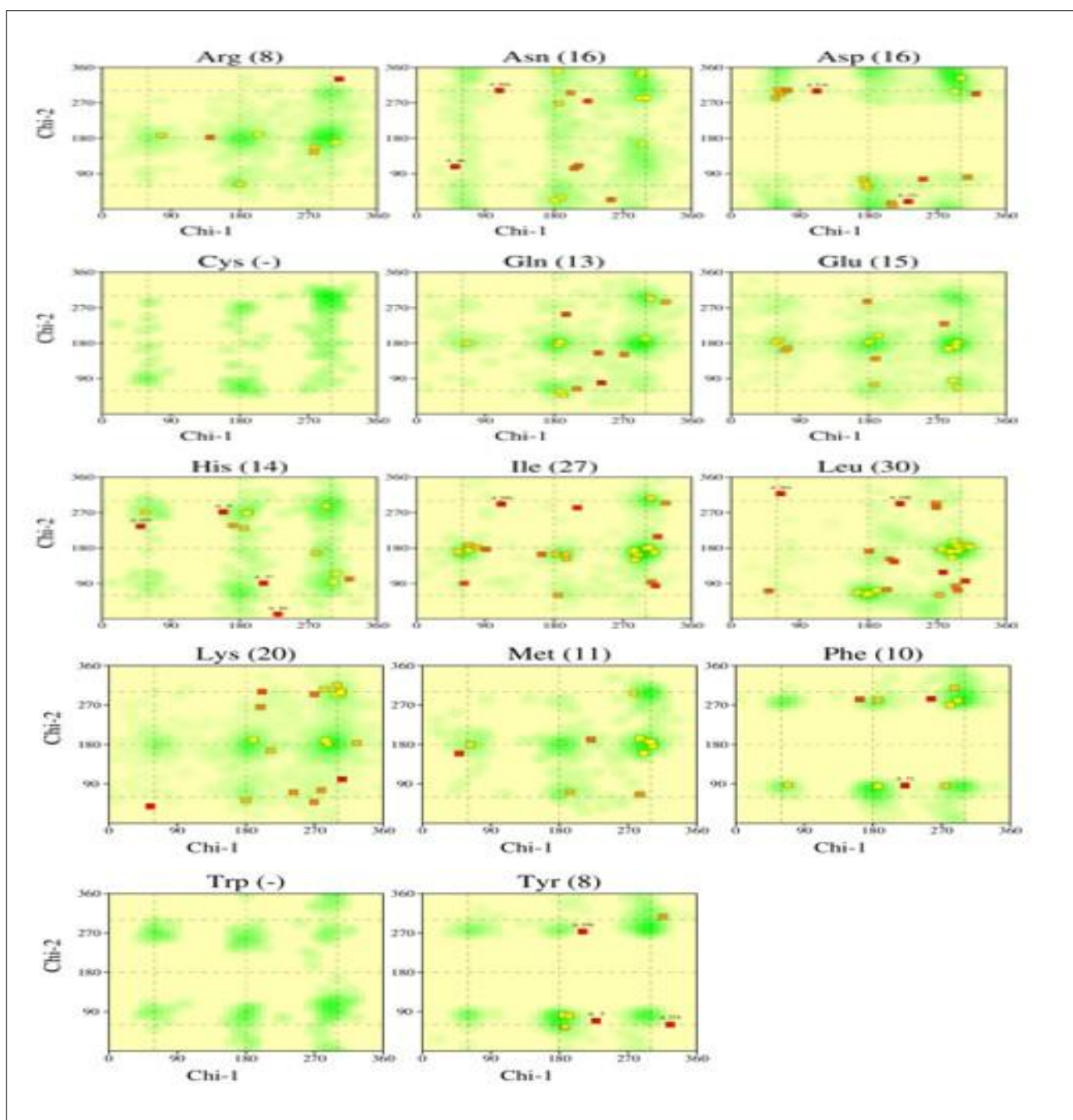
All Ramachandran	Chi1-chi2 plots	Side-chain parameters	Residue properties	G-factors Dihedrals	Planar groups
35 labelled residues (out of 321)	15 labelled residues (out of 188)	5 better 0 inside 0 worse	Max. deviation: 16.2 Bad contacts: 0 Bond Length/angle: 6.8	Dihedrals: 0.79 Covalent: 0.08 Overall: 0.48	76.0% within limits 24.0% highlighted 5 off graph



**Figure 48** Ramachandran plots for all residue types.

This plot shows all Ramachandran's plots to the twenty different amino acid types. The darkest shaded region in the given plot is the most favorable region (**Figure 48**).





**Figure 49** Chi1-chi2 plots

Chi1-Chi 2 plots demonstrate the chi1-chi2 side-chain torsion angle combinations

The shaded region shows the favorable region in the plot, the darkest shaded region in the given plot is the most favorable region (**Figure 49**).

## CHAPTER- 6

### CONCLUSION

L-Asparaginase is an enzyme that hydrolyses the L-asparagine to L-aspartate and ammonia. LA activity has been studied in bacteria, yeast, fungus, algae, and plant tissues and certain rodents' serum. The current study involves numerous bio-computational tools to analyze the structural and functional parameters of BmA. Ten asparaginase protein sequences of different *Bacillus* species were retrieved from the UniProtKB database based on amino acid numbers. The evolutionary connections between different *Bacillus* species were analyzed based on the phylogenetic tree created through multiple sequence analysis of 10 asparaginase sequences. The alignment data suggested that *B. megaterium* asparaginase has similarities with the asparaginase of majority of *Bacillus* species. *B. megaterium* asparaginase (BmA) shows a 78% similarity with the asparaginases of *B. mycooides*, *B. cereus*, and *B. wiedmannii*.

The physicochemical properties of the asparaginases were derived using ProtParam server based on the primary structure of the proteins. The isoelectric point (pI) values for all the 10 protein sequences lie between the ranges of 4–6, showing the acidic nature of the proteins. The high range of aliphatic indices of the protein under study indicates the thermostable nature of the proteins.

The predicted secondary structure of the BmA using PSIPRED and CFSSP, suggested that BmA mainly consists of helix and sheets secondary elements. The Motif search tool and SOSUI server were utilized for functional analysis of the BmA whereas the CATH server was utilized for the classification based on 3D structure of the BmA. This study indicated that the BmA has the pI value of 6.53 and is stable. These properties can assist BmA to keep longer in blood circulation. This preliminary investigation of BmA leads to more experimental research to better understand its potential application in cancer treatment.

## REFERENCES

- [1] J. J. M. Cachumba, F. A. F. Antunes, G. F. D. Peres, L. P. Brumano, J. C. Dos Santos, and S. S. Da Silva, "Current applications and different approaches for microbial L-asparaginase production," *Brazilian J. Microbiol.*, vol. 47, pp. 77–85, 2016, doi: 10.1016/j.bjm.2016.10.004.
- [2] S. M. Chohan, N. Rashid, M. Sajed, and T. Imanaka, "Pcal\_0970: an extremely thermostable l-asparaginase from *Pyrobaculum calidifontis* with no detectable glutaminase activity," *Folia Microbiol. (Praha)*, vol. 64, no. 3, pp. 313–320, Oct. 2018, doi: 10.1007/s12223-018-0656-6.
- [3] V. D. Dwivedi and S. K. Mishra, "In silico analysis of L-asparaginase from different source organisms," *Interdiscip. Sci. Comput. Life Sci.*, vol. 6, no. 2, pp. 93–99, Jun. 2014, doi: 10.1007/s12539-012-0041-0.
- [4] S. Bansal, D. Gnaneswari, P. Mishra, and B. Kundu, "Structural stability and functional analysis of L-asparaginase from *pyrococcus furiosus*," *Biochem.*, vol. 75, no. 3, pp. 375–381, 2010, doi: 10.1134/S0006297910030144.
- [5] S. Zhang *et al.*, "Biochemical characterization of a novel l-asparaginase from *Bacillus megaterium* H-1 and its application in French fries," *Food Res. Int.*, vol. 77, pp. 527–533, Nov. 2015, doi: 10.1016/j.foodres.2015.08.031.
- [6] S. Zhang *et al.*, "Biochemical characterization of a novel l-asparaginase from *Bacillus megaterium* H-1 and its application in French fries," *Food Res. Int.*, vol. 77, pp. 527–533, 2015, doi: 10.1016/j.foodres.2015.08.031.
- [7] J. B. va. Beilen and Z. Li, "Enzyme technology: An overview," *Current Opinion in Biotechnology*, vol. 13, no. 4. Elsevier Ltd, pp. 338–344, Aug. 01, 2002, doi: 10.1016/S0958-1669(02)00334-8.
- [8] D. Gonçalves Filho, A. Gonçalves Silva, and C. Zanella Guidini, "Lipases: sources, immobilization methods, and industrial applications," doi: 10.1007/s00253-019-10027-6.
- [9] Y. chuan Wang, H. fang Hu, J. wen Ma, Q. juan Yan, H. jie Liu, and Z. qiang Jiang,

- “A novel high maltose-forming  $\alpha$ -amylase from *Rhizomucor miehei* and its application in the food industry,” *Food Chem.*, vol. 305, no. May 2019, p. 125447, 2020, doi: 10.1016/j.foodchem.2019.125447.
- [10] S. Kumar, V. Venkata Dasu, and K. Pakshirajan, “Purification and characterization of glutaminase-free L-asparaginase from *Pectobacterium carotovorum* MTCC 1428,” *Bioresour. Technol.*, vol. 102, no. 2, pp. 2077–2082, 2011, doi: 10.1016/j.biortech.2010.07.114.
- [11] Saima, M. Kuddus, Roohi, and I. Z. Ahmad, “Isolation of novel chitinolytic bacteria and production optimization of extracellular chitinase,” *J. Genet. Eng. Biotechnol.*, vol. 11, no. 1, pp. 39–46, 2013, doi: 10.1016/j.jgeb.2013.03.001.
- [12] M. Pourhossein and H. Korbekandi, “Cloning, expression, purification and characterisation of *Erwinia carotovora* L-asparaginase in *Escherichia coli*,” *Adv. Biomed. Res.*, vol. 3, no. 1, p. 82, 2014, doi: 10.4103/2277-9175.127995.
- [13] S. Acharya and A. Chaudhary, “Bioprospecting thermophiles for cellulase production: A review,” *Brazilian J. Microbiol.*, vol. 43, no. 3, pp. 844–856, 2012, doi: 10.1590/S1517-83822012000300001.
- [14] W. L. Salzer, B. L. Asselin, P. V. Plourde, T. Corn, and S. P. Hunger, “Development of asparaginase *Erwinia chrysanthemi* for the treatment of acute lymphoblastic leukemia,” *Ann. N. Y. Acad. Sci.*, vol. 1329, no. 1, pp. 81–92, 2014, doi: 10.1111/nyas.12496.
- [15] A. P. Gobert *et al.*, “*Helicobacter pylori* Induces Macrophage Apoptosis by Activation of Arginase II ,” *J. Immunol.*, vol. 168, no. 9, pp. 4692–4700, 2002, doi: 10.4049/jimmunol.168.9.4692.
- [16] R. D. Abigor, P. O. Uadia, T. A. Foglia, M. J. Haas, K. Scott, and B. J. Savary, “Partial purification and properties of lipase from germinating seeds of *Jatropha curcas* L,” *JAOCs, J. Am. Oil Chem. Soc.*, vol. 79, no. 11, pp. 1123–1126, 2002, doi: 10.1007/s11746-002-0614-3.
- [17] B. Padmapriya, T. Rajeswari, R. Nandita, and F. Raj, “Production and purification of alkaline serine protease from marine *Bacillus* species and its application in detergent industry,” *Eur. J. Appl. Sci.*, vol. 4, no. 1, pp. 21–26, 2012.

- [18] H. D. Simpson, U. R. Haufler, and R. M. Daniel, "An extremely thermostable xylanase from the thermophilic eubacterium *Thermotoga*," *Biochem. J.*, vol. 277, no. 2, pp. 413–417, 1991, doi: 10.1042/bj2770413.
- [19] A. Kavitha and M. Vijayalakshmi, "Optimization and Purification of L-Asparaginase Produced by *Streptomyces tendae* TK-VL\_333," *Zeitschrift fur Naturforsch. - Sect. C J. Biosci.*, vol. 65, no. 7–8, pp. 528–531, 2010, doi: 10.1515/znc-2010-7-817.
- [20] D. Cappelletti, L. R. Chiarelli, M. V. Pasquetto, S. Stivala, G. Valentini, and C. Scotti, "Helicobacter pylori l-asparaginase: A promising chemotherapeutic agent," *Biochem. Biophys. Res. Commun.*, vol. 377, no. 4, pp. 1222–1226, Dec. 2008, doi: 10.1016/j.bbrc.2008.10.118.
- [21] Y. Singh and S. K. Srivastava, "Screening and characterization of microorganisms capable of producing antineoplastic drug , L-asparaginase," *Int. J. Biol. Med. Res.*, vol. 3, no. 1, pp. 2548–2554, 2012.
- [22] T. Batool, E. A. Makky, M. Jalal, and M. M. Yusoff, "A Comprehensive Review on l-Asparaginase and Its Applications," *Appl. Biochem. Biotechnol.*, vol. 178, no. 5, pp. 900–923, 2016, doi: 10.1007/s12010-015-1917-3.
- [23] A. Vimal and A. Kumar, "Biotechnological production and practical application of L-asparaginase enzyme," *Biotechnol. Genet. Eng. Rev.*, vol. 33, no. 1, pp. 40–61, 2017, doi: 10.1080/02648725.2017.1357294.
- [24] S. A. Gaffar and Y. I. Shethna, "Purification and some biological properties of asparaginase from *Azotobacter vinelandii*," *Appl. Environ. Microbiol.*, vol. 33, no. 3, pp. 508–514, 1977, doi: 10.1128/aem.33.3.508-514.1977.
- [25] U. Narta, S. Roy, S. S. Kanwar, and W. Azmi, "Improved production of l-asparaginase by *Bacillus brevis* cultivated in the presence of oxygen-vectors," *Bioresour. Technol.*, vol. 102, no. 2, pp. 2083–2085, 2011, doi: 10.1016/j.biortech.2010.07.118.
- [26] T. Banks, "The Properties and Large-scale Production of from *Citrobacter*," no. 2, pp. 1–16, 1975.
- [27] J. Netrval, "Stimulation of L-asparaginase production in *Escherichia coli* by organic and amino acids," *Folia Microbiol. (Praha)*, vol. 22, no. 2, pp. 106–116, 1977, doi:

10.1007/BF02881635.

- [28] R. R. Erva, A. N. Goswami, P. Suman, R. Vedanabhatla, and S. B. Rajulapati, "Optimization of L-asparaginase production from novel *Enterobacter* sp., by submerged fermentation using response surface methodology," *Prep. Biochem. Biotechnol.*, vol. 47, no. 3, pp. 219–228, 2017, doi: 10.1080/10826068.2016.1201683.
- [29] I. Husain, A. Sharma, S. Kumar, and F. Malik, "Purification and characterization of glutaminase free asparaginase from *enterobacter cloacae*: In-vitro evaluation of cytotoxic potential against human myeloid leukemia HL-60 cells," *PLoS One*, vol. 11, no. 2, pp. 1–27, 2016, doi: 10.1371/journal.pone.0148877.
- [30] R. E. Peterson and A. Ciegler, "L-asparaginase production by *Erwinia aroideae*," *Appl. Microbiol.*, vol. 18, no. 1, pp. 64–67, 1969, doi: 10.1128/aem.18.1.64-67.1969.
- [31] G. A. Kotzia and N. E. Labrou, "l-Asparaginase from *Erwinia Chrysanthemi* 3937: Cloning, expression and characterization," *J. Biotechnol.*, vol. 127, no. 4, pp. 657–669, 2007, doi: 10.1016/j.jbiotec.2006.07.037.
- [32] A. Ebrahiminezhad, S. Rasoul-Amini, and Y. Ghasemi, "l-Asparaginase Production by Moderate Halophilic Bacteria Isolated from Maharloo Salt Lake," *Indian J. Microbiol.*, vol. 51, no. 3, pp. 307–311, 2011, doi: 10.1007/s12088-011-0158-6.
- [33] N. M. Ali, N. H. Hussein, A. S. Dwaish, B. T. El Haboby, S. N. Muslim, and S. abbas abid, "Extraction and Purification of L-Asparaginase Produced by *Acinetobacter Baumannii* and their Antibiofilm Activity Against Some Pathogenic Bacteria," *Int. J. Biotechnol.*, vol. 5, no. 1, pp. 7–14, 2016, doi: 10.18488/journal.57/2016.5.1/57.1.7.14.
- [34] L. N. Ramya, M. Doble, V. P. B. Rekha, and K. K. Pulicherla, "L-asparaginase as potent anti-leukemic agent and its significance of having reduced glutaminase side activity for better treatment of acute lymphoblastic leukaemia," *Appl. Biochem. Biotechnol.*, vol. 167, no. 8, pp. 2144–2159, 2012, doi: 10.1007/s12010-012-9755-z.
- [35] A. A. El-Bessoumy, M. Sarhan, and J. Mansour, "Production and purification of L-asp," vol. 37, no. 4, pp. 387–393, 2004.
- [36] V. Kishore, K. P. Nishita, and H. K. Manonmani, "Cloning, expression and characterization of l-asparaginase from *Pseudomonas fluorescens* for large scale

- production in *E. coli* BL21,” *3 Biotech*, vol. 5, no. 6, pp. 975–981, 2015, doi: 10.1007/s13205-015-0300-y.
- [37] M. Hymavathi, T. Sathish, C. S. Rao, and R. S. Prakasham, “Enhancement of l-asparaginase production by isolated bacillus circulans (MTCC 8574) using response surface methodology,” *Appl. Biochem. Biotechnol.*, vol. 159, no. 1, pp. 191–198, 2009, doi: 10.1007/s12010-008-8438-2.
- [38] A. S. Law and J. C. Wriston, “Purification and properties of Bacillus coagulans l-asparaginase,” *Arch. Biochem. Biophys.*, vol. 147, no. 2, pp. 744–752, 1971, doi: 10.1016/0003-9861(71)90434-6.
- [39] S. Wakil and A. Adelegan, “Screening, Production and Optimization of L-Asparaginase From Soil Bacteria Isolated in Ibadan, South-western Nigeria,” *J. Basic Appl. Sci.*, vol. 11, no. February, pp. 39–51, 2015, doi: 10.6000/1927-5129.2015.11.06.
- [40] S. Yadav, S. K. Verma, J. Singh, and A. Kumar, “Industrial Production and Clinical Application of L-Asparaginase: A Chemotherapeutic Agent,” *Int. J. Biotechnol. Bioeng.*, vol. 8, no. 1, pp. 54–60, 2014, doi: 10.5281/zenodo.1091554.
- [41] Y. Mostafa *et al.*, “Enhanced production of glutaminase-free L-asparaginase by marine Bacillus velezensis and cytotoxic activity against breast cancer cell lines,” *Electron. J. Biotechnol.*, vol. 42, pp. 6–15, 2019, doi: 10.1016/j.ejbt.2019.10.001.
- [42] G. Sanghvi *et al.*, “Mitigation of acrylamide by l-asparaginase from Bacillus subtilis KDPS1 and analysis of degradation products by HPLC and HPTLC,” *Springerplus*, vol. 5, no. 1, 2016, doi: 10.1186/s40064-016-2159-8.
- [43] S. A. Alrumman, Y. S. Mostafa, K. A. Al-izran, M. Y. Alfaifi, T. H. Taha, and S. E. Elbehairi, “Production and Anticancer Activity of an L-Asparaginase from Bacillus licheniformis Isolated from the Red Sea, Saudi Arabia,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–14, 2019, doi: 10.1038/s41598-019-40512-x.
- [44] K. J. P. Narayana, K. G. Kumar, and M. Vijayalakshmi, “L-asparaginase production by Streptomyces albidoflavus,” *Indian J. Microbiol.*, vol. 48, no. 3, pp. 331–336, 2008, doi: 10.1007/s12088-008-0018-1.
- [45] N. Asthana and W. Azni, “Microbial L-Asparaginase: A Potent Antitumour Enzyme,”

2003. Accessed: Jun. 15, 2021. [Online]. Available: <http://nopr.niscair.res.in/handle/123456789/11297>.

- [46] A. Sharma, ““ EXTRACTION OF THE ENZYME ASPARAGINASE FROM *Mycobacterium smegmatis* ’ Dissertation / project report submitted for partial fulfillment of the By :,” no. May, pp. 1–63, 2018.
- [47] R. S. Prakasham, C. S. Rao, R. S. Rao, G. S. Lakshmi, and P. N. Sarma, “L-asparaginase production by isolated *Staphylococcus* sp. - 6A: Design of experiment considering interaction effect for process parameter optimization,” *J. Appl. Microbiol.*, vol. 102, no. 5, pp. 1382–1391, 2007, doi: 10.1111/j.1365-2672.2006.03173.x.
- [48] A. MATHEW, K. DHEVENDARAN, M. GEORGEKUTTY, and P. NATARAJAN, “L-Asparaginase activity in antagonistic streptomycetes associated with clam *Villorita cyprinoides* (Hanley),” *Indian Journal of Geo-Marine Sciences (IJMS)*, vol. 23, no. 4, pp. 204–208, 1994.
- [49] S. Hatamzadeh *et al.*, “Isolation and identification of L-asparaginase-producing endophytic fungi from the Asteraceae family plant species of Iran,” *PeerJ*, vol. 2020, no. 1, pp. 1–16, 2020, doi: 10.7717/peerj.8309.
- [50] C. Drainas, J. R. Kinghorn, and J. A. Pateman, “Aspartic hydroxamate resistance and asparaginase regulation in the fungus *Aspergillus nidulans*,” *J. Gen. Microbiol.*, vol. 98, no. 2, pp. 493–501, 1977, doi: 10.1099/00221287-98-2-493.
- [51] A. Mishra, “Production of L-asparaginase, an anticancer agent, from *Aspergillus niger* using agricultural waste in solid state fermentation,” *Appl. Biochem. Biotechnol.*, vol. 135, no. 1, pp. 33–42, 2006, doi: 10.1385/ABAB:135:1:33.
- [52] F. F. G. Dias, A. L. T. G. Ruiz, A. Della Torre, and H. H. Sato, “Purification, characterization and antiproliferative activity of L-asparaginase from *Aspergillus oryzae* CCT 3940 with no glutaminase activity,” *Asian Pac. J. Trop. Biomed.*, vol. 6, no. 9, pp. 785–794, 2016, doi: 10.1016/j.apjtb.2016.07.007.
- [53] M. I. de M. Sarquis, E. M. M. Oliveira, A. S. Santos, and G. L. da Costa, “Production of L-asparaginase by filamentous fungi,” *Mem. Inst. Oswaldo Cruz*, vol. 99, no. 5, pp. 489–492, 2004, doi: 10.1590/S0074-02762004000500005.



- [54] A. M. Farag, S. W. Hassan, E. A. Beltagy, and M. A. El-Shenawy, "Optimization of production of anti-tumor l-asparaginase by free and immobilized marine *Aspergillus terreus*," *Egypt. J. Aquat. Res.*, vol. 41, no. 4, pp. 295–302, 2015, doi: 10.1016/j.ejar.2015.10.002.
- [55] M. Thakur, L. Lincoln, F. N. Niyonzima, and S. M. Sunil, "Biotransformation Isolation , Purification and Characterization of Fungal," *J. Biocatal. Biotransformation*, vol. 2, no. 2, pp. 1–9, 2014.
- [56] A. El Baky. and G. S. El Baroty., "Optimization of growth conditions for purification and production of L-asparaginase by *Spirulina maxima*. Evidence-Based Complementary and Alternative Medicine," *Evidence-Based Complement. Altern. Med.*, vol. 2016, p. 7, 2016, [Online]. Available: <https://www.hindawi.com/journals/ecam/contents/>.
- [57] J. H. Paul, "Isolation and characterization of a *Chlamydomonas* L-asparaginase," *Biochem. J.*, vol. 203, no. 1, pp. 109–115, 1982, doi: 10.1042/bj2030109.
- [58] B. R. Mohapatra, R. K. Sani, and U. C. Banerjee, "Characterization of L- asparaginase from *Bacillus* sp. isolated from an intertidal marine alga (*Sargassum* sp.)," *Lett. Appl. Microbiol.*, vol. 21, no. 6, pp. 380–383, 1995, doi: 10.1111/j.1472-765X.1995.tb01086.x.
- [59] H. Karim *et al.*, "Optimization of enzyme activity of l-asparaginase derived from enterobacter agglomerans sb 221 bacterial symbiont of brown algae sargassum sp.," *Rasayan J. Chem.*, vol. 13, no. 3, pp. 1571–1579, 2020, doi: 10.31788/RJC.2020.1335691.
- [60] A. Thangavel, G. Krishnamoorthy, M. Subramanian, and M. Maruthamuthu, "Seaweed Endophytic Fungi: A Potential Source for Glutaminase Free L-Asparaginase," *Che Sci Rev Lett*, vol. 2, no. 5, pp. 348–354, 2013.
- [61] J. O. Kil, G. N. Kim, and I. Park, "Extraction of Extracellular L-Asparaginase from *Candida utilis*," *Biosci. Biotechnol. Biochem.*, vol. 59, no. 4, pp. 749–750, 1995, doi: 10.1271/bbb.59.749.
- [62] N. Ahmad, N. P. Pandit, and S. K. Maheshwari, "L-asparaginase gene-a therapeutic approach towards drugs for cancer cell," vol. 2, no. 4, pp. 1–11, 2012, [Online].

Available: <http://www.innspub.net>.

- [63] M. Salah Foda, H. H. Zedan, S. Abd El-Megeed, and A. Hashem, "Formation and properties of L-glutaminase and L-asparaginase activities in *Pichia polymorpha*," *Acta Microbiol. Pol.*, vol. 29, no. 4, pp. 343–352, Jan. 1980, Accessed: Jun. 13, 2021. [Online]. Available: <https://europepmc.org/article/med/6164254>.
- [64] M. S. Ramakrishnan and R. Joseph, "Characterization of an extracellular asparaginase of *Rhodospiridium toruloides* CBS14 exhibiting unique physicochemical properties," *Can. J. Microbiol.*, vol. 42, no. 4, pp. 316–325, 1996, doi: 10.1139/m96-047.
- [65] A. Sarkar, A. M. Philip, D. P. Thakker, M. S. Wagh, and K. V. B. Rao, "In vitro Antioxidant activity of extracellular L-glutaminase enzyme isolated from marine yeast *Rhodotorula* sp. DAMB1," *Res. J. Pharm. Technol.*, vol. 13, no. 1, p. 209, 2020, doi: 10.5958/0974-360x.2020.00042.6.
- [66] M. A. Ferrara *et al.*, "Asparaginase production by a recombinant *Pichia pastoris* strain harbouring *Saccharomyces cerevisiae* ASP3 gene," *Enzyme Microb. Technol.*, vol. 39, no. 7, pp. 1457–1463, 2006, doi: 10.1016/j.enzmictec.2006.03.036.
- [67] M. Pola, C. P. Durthi, S. B. Rajulapati, and R. R. Erva, "Modelling and optimization of L-asparaginase production from *Bacillus stratosphericus*," *Curr. Trends Biotechnol. Pharm.*, vol. 12, no. 4, pp. 390–405, 2018.
- [68] M. Bano and V. M. Sivaramakrishnan, "Preparation and properties of L-asparaginase from green chillies (*Capsicum annum* L.)," *J. Biosci.*, vol. 2, no. 4, pp. 291–297, 1980, doi: 10.1007/BF02716861.
- [69] B. Zena Abdullah Khalaf and M. Hussein Al-Gelawi, "Extraction and Purification of Asparaginase enzyme from *Pisum sativum* plant and studying their cytotoxicity against L20B tumor cell line Master of Science in Biotechnology," no. March, 2012.
- [70] T. J. Lough, B. D. Reddington, M. R. Grant, D. F. Hill, P. H. S. Reynolds, and K. J. F. Farnden, "The isolation and characterisation of a cDNA clone encoding L-asparaginase from developing seeds of lupin (*Lupinus arboreus*)," *Plant Mol. Biol.*, vol. 19, no. 3, pp. 391–399, 1992, doi: 10.1007/BF00023386.
- [71] K. Kumar and S. Walia, "Sensors & Transducers and Development of Asparagine

- Biosensor for Leukemia,” vol. 144, no. 9, pp. 192–200, 2012.
- [72] V. P. Oza, S. D. Trivedi, P. P. Parmar, and R. B. Subramanian, “Withania somnifera (Ashwagandha): A novel source of L-asparaginase,” *J. Integr. Plant Biol.*, vol. 51, no. 2, pp. 201–206, 2009, doi: 10.1111/j.1744-7909.2008.00779.x.
- [73] E. M. Mohamed Ali, “Purification and characterization of *Vigna unguiculata* cultivar asparaginase,” *J. Biol. Res.*, vol. 11, pp. 29–36, 2009, doi: 10.4314/ejbmb.v27i1.43196.
- [74] J. G. Streeter, “Asparaginase and Asparagine Transaminase in Soybean Leaves and Root Nodules,” *Plant Physiol.*, vol. 60, no. 2, pp. 235–239, 1977, doi: 10.1104/pp.60.2.235.
- [75] E. Boutet *et al.*, *Uniprotkb/swiss-prot, the manually annotated section of the uniprot knowledgebase: How to use the entry view*, vol. 1374, no. January. 2016.
- [76] K. Beulah and K. P. J. Hemalatha, “Screening of Medicinal Plants for Potential Source of l-Asparaginase and Optimization of Conditions for Maximum Extraction and Assay of l-Asparaginase from *Asparagus racemosus*,” *Iran. J. Sci. Technol. Trans. A Sci.*, vol. 43, no. 1, 2019, doi: 10.1007/s40995-017-0359-x.
- [77] P. Teena, K. Raman, K. Jagjit, and K. Kuldeep, “Isolation of l-asparaginase from *cannabis sativa* and development of biosensor for detection of asparagine in leukemic serum samples,” *Res. J. Pharm. Technol.*, vol. 7, no. 8, pp. 850–855, 2014.
- [78] I. Adamsons, J. Hutzelmann, and D. Heitjan, “Comparison of summary measures for analyzing longitudinal visual field data,” *Investig. Ophthalmol. Vis. Sci.*, vol. 37, no. 3, pp. 235–258, 1996.
- [79] K. Doriya, N. Jose, M. Gowda, and D. S. Kumar, *Solid-State Fermentation vs Submerged Fermentation for the Production of L-Asparaginase*, 1st ed., vol. 78. Elsevier Inc., 2016.
- [80] M. Pola, S. B. Rajulapati, C. Potla Durthi, R. R. Erva, and M. Bhatia, “In silico modelling and molecular dynamics simulation studies on L-Asparaginase isolated from bacterial endophyte of *Ocimum tenuiflorum*,” *Enzyme Microb. Technol.*, vol. 117, no. June, pp. 32–40, 2018, doi: 10.1016/j.enzmictec.2018.06.005.

- [81] S. Ghosh, S. Murthy, S. Govindasamy, and M. Chandrasekaran, "Optimization of L-asparaginase production by *Serratia marcescens* (NCIM 2919) under solid state fermentation using coconut oil cake," *Sustain. Chem. Process.*, vol. 1, no. 1, p. 9, 2013, doi: 10.1186/2043-7129-1-9.
- [82] K. Doriya and D. S. Kumar, "Optimization of solid substrate mixture and process parameters for the production of L-asparaginase and scale-up using tray bioreactor," *Biocatal. Agric. Biotechnol.*, vol. 13, no. January, pp. 244–250, 2018, doi: 10.1016/j.bcab.2018.01.004.
- [83] A. J. Shah, R. V. Karadi, and P. P. Parekh, "Isolation, Optimization and Production of L-asparaginase from Coliform Bacteria," *Asian J. Biotechnol.*, vol. 2, no. 3, pp. 169–177, 2010, doi: 10.3923/ajbkr.2010.169.177.
- [84] F. Muneer *et al.*, "Microbial l-asparaginase: purification, characterization and applications," *Arch. Microbiol.*, vol. 202, no. 5, pp. 967–981, 2020, doi: 10.1007/s00203-020-01814-1.
- [85] N. S. Mohan Kumar and H. K. Manonmani, "Purification, characterization and kinetic properties of extracellular l-asparaginase produced by *Cladosporium* sp.," *World J. Microbiol. Biotechnol.*, vol. 29, no. 4, pp. 577–587, 2013, doi: 10.1007/s11274-012-1213-0.
- [86] A. Shrivastava, A. A. Khan, A. Shrivastav, S. K. Jain, and P. K. Singhal, "Kinetic studies of L-asparaginase from *Penicillium digitatum*," *Prep. Biochem. Biotechnol.*, vol. 42, no. 6, pp. 574–581, 2012, doi: 10.1080/10826068.2012.672943.
- [87] A. K. Vala *et al.*, "Characterization of L-asparaginase from marine-derived *Aspergillus niger* AKV-MKBU, its antiproliferative activity and bench scale production using industrial waste," *Int. J. Biol. Macromol.*, vol. 108, no. December, pp. 41–46, 2018, doi: 10.1016/j.ijbiomac.2017.11.114.
- [88] M. S. Shafei *et al.*, "Purification, characterization and kinetic properties of penicillium cyclopium L-asparaginase: Impact of lasparaginase on acrylamide content in potato products and its cytotoxic activity," *Curr. Trends Biotechnol. Pharm.*, vol. 9, no. 2, pp. 132–140, 2015.
- [89] S. Amena, N. Vishalakshi, M. Prabhakar, A. Dayanand, and K. Lingappa, "Production,

- purification and characterization of L-asparaginase from *Streptomyces gulbargensis*,” *Brazilian J. Microbiol.*, vol. 41, no. 1, pp. 173–178, 2010, doi: 10.1590/S1517-83822010000100025.
- [90] R. Tomar, P. Sharma, A. Srivastava, S. Bansal, Ashish, and B. Kundu, “Structural and functional insights into an archaeal l-asparaginase obtained through the linker-less assembly of constituent domains,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 70, no. 12, pp. 3187–3197, 2014, doi: 10.1107/S1399004714023414.
- [91] J. C. F. Nunes *et al.*, “Recent Strategies and Applications for l-Asparaginase Confinement,” *Molecules*, vol. 25, no. 24, 2020, doi: 10.3390/molecules25245827.
- [92] N. E. A. El-Naggar, S. M. El-Ewasy, and N. M. El-Shweihy, “Microbial L-asparaginase as a potential therapeutic agent for the treatment of acute lymphoblastic leukemia: The pros and cons,” *International Journal of Pharmacology*, vol. 10, no. 4, pp. 182–199, 2014, doi: 10.3923/ijp.2014.182.199.
- [93] E. J. M. Konings *et al.*, “Acrylamide exposure from foods of the Dutch population and an assessment of the consequent risks,” *Food Chem. Toxicol.*, vol. 41, no. 11, pp. 1569–1579, 2003, doi: 10.1016/S0278-6915(03)00187-X.
- [94] O. Marconi, E. Bravi, G. Perretti, R. Martini, L. Montanari, and P. Fantozzi, “Acrylamide risk in food products: The shortbread case study,” *Anal. Methods*, vol. 2, no. 11, pp. 1686–1691, 2010, doi: 10.1039/c0ay00191k.
- [95] B. K. Koh, “Determination of acrylamide content of food products in Korea,” *J. Sci. Food Agric.*, vol. 86, no. 15, pp. 2587–2591, Dec. 2006, doi: 10.1002/jsfa.2652.
- [96] E. Rottmann, K. F. Hauke, U. Krings, and R. G. Berger, “Enzymatic acrylamide mitigation in French fries – An industrial-scale case study,” *Food Control*, vol. 123, p. 107739, 2021, doi: 10.1016/j.foodcont.2020.107739.
- [97] H. M. Orabi, E. M. El-Fakharany, E. S. Abdelkhalek, and N. M. Sidkey, “L-Asparaginase and L-glutaminase: Sources, production, and applications in medicine and industry,” *J. Microbiol. Biotechnol. Food Sci.*, no. 2, pp. 179–190, 2019, doi: 10.15414/jmbfs.2019.9.2.179-190.
- [98] M. Duval *et al.*, “Comparison of *Escherichia coli*-asparaginase with *Erwinia*-

- asparaginase in the treatment of childhood lymphoid malignancies: Results of a randomized European Organisation for Research and Treatment of Cancer - Children's Leukemia Group phase 3 trial," *Blood*, vol. 99, no. 8, pp. 2734–2739, 2002, doi: 10.1182/blood.V99.8.2734.
- [99] P. S. Vary *et al.*, *Bacillus megaterium*-from simple soil bacterium to industrial protein production host, vol. 76, no. 5. 2007.
- [100] C. Korneli, F. David, R. Biedendieck, D. Jahn, and C. Wittmann, "Getting the big beast to work-Systems biotechnology of *Bacillus megaterium* for novel high-value proteins," *J. Biotechnol.*, vol. 163, no. 2, pp. 87–96, 2013, doi: 10.1016/j.jbiotec.2012.06.018.
- [101] A. Williams *et al.*, "Metabolic engineering of *Bacillus megaterium* for heparosan biosynthesis using *Pasteurella multocida* heparosan synthase, PmHS2," *Microb. Cell Fact.*, vol. 18, no. 1, pp. 1–13, 2019, doi: 10.1186/s12934-019-1187-9.
- [102] H. E. Pauly and G. Pfeleiderer, "D-Glucose Dehydrogenase from *Bacillus megaterium* M 1286 : Purification , Properties and Structure Brought to you by | Purdue University Lib Authenticated Brought to you by | Purdue University Lib Authenticated," *Physiol. Chem.*, vol. 356, pp. 1613–1624, 1975.
- [103] S. Obruca, I. Marova, S. Melusova, and L. Mravcova, "Production of polyhydroxyalkanoates from cheese whey employing *Bacillus megaterium* CCM 2037," *Ann. Microbiol.*, vol. 61, no. 4, pp. 947–953, 2011, doi: 10.1007/s13213-011-0218-5.
- [104] M. Jana *et al.*, "Salt-independent thermophilic  $\alpha$ -amylase from *Bacillus megaterium* VUMB109: An efficacy testing for preparation of maltooligosaccharides," *Ind. Crops Prod.*, vol. 41, no. 1, pp. 386–391, 2013, doi: 10.1016/j.indcrop.2012.04.048.
- [105] I. Sindhu, S. Chhibber, N. Capalash, and P. Sharma, "Production of cellulase-free xylanase from *Bacillus megaterium* by solid state fermentation for biobleaching of pulp," *Curr. Microbiol.*, vol. 53, no. 2, pp. 167–172, 2006, doi: 10.1007/s00284-006-0051-4.
- [106] P. Moreno, C. Yañez, N. S. M. Cardozo, H. Escalante, M. Y. Combariza, and C. Guzman, "Influence of nutritional and physicochemical variables on PHB production

- from raw glycerol obtained from a Colombian biodiesel plant by a wild-type *Bacillus megaterium* strain,” *N. Biotechnol.*, vol. 32, no. 6, pp. 682–689, 2015, doi: 10.1016/j.nbt.2015.04.003.
- [107] Y. Yang *et al.*, “High yield recombinant penicillin G amidase production and export into the growth medium using *Bacillus megaterium*,” *Microb. Cell Fact.*, vol. 5, pp. 1–14, 2006, doi: 10.1186/1475-2859-5-36.
- [108] Y. Zhang, A. N. Aryee, and B. K. Simpson, “Current role of in silico approaches for food enzymes,” *Curr. Opin. Food Sci.*, vol. 31, pp. 63–70, 2020, doi: 10.1016/j.cofs.2019.11.003.
- [109] S. A. Kulkarni, J. Zhu, and S. Blechinger, “In silico techniques for the study and prediction of xenobiotic metabolism: A review,” *Xenobiotica*, vol. 35, no. 10–11, pp. 955–973, 2005, doi: 10.1080/00498250500354402.
- [110] T. Schwede, J. Kopp, N. Guex, and M. C. Peitsch, “SWISS-MODEL: An automated protein homology-modeling server,” *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3381–3385, Jul. 2003, doi: 10.1093/nar/gkg520.
- [111] E. Boutet, D. Lieberherr, M. Tognolli, and A. Bairoch, “Plant Bioinformatics,” *Plant Bioinforma.*, no. May 2014, 2007, doi: 10.1007/978-1-59745-535-0.
- [112] A. Bateman, “UniProt: A worldwide hub of protein knowledge,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D506–D515, 2019, doi: 10.1093/nar/gky1049.
- [113] A. Mathew, A. Verma, and S. Gaur, “An in-silico insight into the characteristics of  $\beta$ -propeller phytase,” *Interdisciplinary Sciences: Computational Life Sciences*, vol. 6, no. 2, pp. 133–139, 2014, doi: 10.1007/s12539-013-0010-2.
- [114] A. Mahram and M. C. Herbordt, “NCBI BLASTP on high-performance reconfigurable computing systems,” *ACM Trans. Reconfigurable Technol. Syst.*, vol. 7, no. 4, 2015, doi: 10.1145/2629691.
- [115] J. M. Mesas, J. A. Gil, and J. F. M. N., “Characterization and partial purification of L-asparaginase from *Corynebacterium glutamicum*,” pp. 3–7, 1990.
- [116] S. Pundir, M. J. Martin, and C. O’Donovan, “UniProt Tools,” *Curr. Protoc. Bioinforma.*, vol. 53, no. 1, pp. 1.29.1-1.29.15, 2016, doi:

- 10.1002/0471250953.bi0129s53.
- [117] J. D. Thompson, T. J. Gibson, and D. G. Higgins, "Multiple Sequence Alignment Using ClustalW and ClustalX," *Curr. Protoc. Bioinforma.*, vol. 00, no. 1, pp. 1–22, 2003, doi: 10.1002/0471250953.bi0203s00.
- [118] V. K. Garg *et al.*, "MFPPi – Multi FASTA ProtParam Interface," *Bioinformatics*, vol. 12, no. 2, pp. 74–77, Apr. 2016, doi: 10.6026/97320630012074.
- [119] W. W. Yu, L. Qu, W. Guo, and X. Peng, "Experimental Determination of the Extinction Coefficient of CdTe, CdSe, and CdS Nanocrystals," *ACS Publ.*, vol. 15, no. 14, pp. 2854–2860, Jul. 2003, doi: 10.1021/cm034081k.
- [120] L. J. McGuffin, K. Bryson, and D. T. Jones, "The PSIPRED protein structure prediction server," 2000. Accessed: Jun. 14, 2021. [Online]. Available: <http://globin.bio.warwick.ac.uk/psipred/>.
- [121] A. Kumar and T. A. Kumar, "CFSSP: Chou and Fasman Secondary Structure Prediction server," *Wide Spectr.*, vol. 1, no. 9, pp. 15–19, 2013, doi: 10.5281/zenodo.50733.
- [122] K. Pramanik *et al.*, "In silico structural, functional and phylogenetic analysis of Klebsiella phytases," *J. Plant Biochem. Biotechnol.*, vol. 27, no. 3, pp. 362–372, 2018, doi: 10.1007/s13562-018-0445-y.
- [123] N. Ciucx and M. C. Peitsrh Urctrophuresis, "SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling," *IS*, vol. 21, no. 15, pp. 14–2723, Dec. 1997, doi: 10.1002/elps.1150181505.
- [124] J. Yang, Y. Wang, and Y. Zhang, "ResQ: An Approach to Unified Estimation of B-Factor and Residue-Specific Error in Protein Structure Prediction," *J. Mol. Biol.*, vol. 428, no. 4, pp. 693–701, 2016, doi: 10.1016/j.jmb.2015.09.024.
- [125] P. Dhavala and A. C. Papageorgiou, "Structure of Helicobacter pylori L-asparaginase at 1.4 Å resolution," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 65, no. 12, pp. 1253–1261, 2009, doi: 10.1107/S0907444909038244.
- [126] N. Singh, S. Upadhyay, A. Jaiswar, and N. Mishra, "Central Bringing Excellence in Open Access In silico Analysis of Protein," *Anal. Protein. J Bioinform*, vol. 1, no. 2, p.



1007, 2016.

[127] F. Carrascoza, S. Zaric, and R. Silaghi-Dumitrescu, “Computational study of protein secondary structure elements: Ramachandran plots revisited,” *J. Mol. Graph. Model.*, vol. 50, pp. 125–133, 2014, doi: 10.1016/j.jmgm.2014.04.001.

[128] J. Yang and Y. Zhang, “Protein Structure and Function Prediction Using I-TASSER,” *Curr. Protoc. Bioinforma.*, vol. 52, no. 1, pp. 5.8.1-5.8.15, 2015, doi: 10.1002/0471250953.bi0508s52.

[129] <https://commons.wikimedia.org/w/index.php?curid=31778633>

[130] DOI: 10.1107/S0907444909038244