EXPERIMENTAL AND COMPUTATIONAL STUDIES ON URICASE AND ITS BIO-CONJUGATION WITH BOVINE SERUM ALBUMIN FOR HYPERURICEMIA

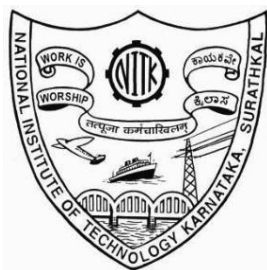Thesis

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

N ANAND KUMAR
Register No. 155082CH15F07

DEPARTMENT OF CHEMICAL ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY
KARNATAKA,SURATHKAL, MANGALURU - 575025

MARCH 2022

DECLARATION

I hereby *declare* that the Thesis entitled **"Experimental And Computational Studies on Uricase and its Bio-Conjugation with Bovine Serum Albumin for Hyperuricemia"** which is being submitted to the **National Institute of Technology Karnataka, Surathkal** in partial fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy** in the Department of Chemical Engineering, is a *bonafide report of the research work carried out by me*. The material contained in this Thesis has not been submitted to any University or Institution for the award of any degree.

Place: NITK, Surathkal

Date: 11|03|2022

N. Anand Kumar

Name: N Anand Kumar

Register number: 155082CH15F07

Department of Chemical Engineering

CERTIFICATE

This is to certify that the Thesis entitled **"Experimental And Computational Studies on Uricase and its Bio-Conjugation with Bovine Serum Albumin for Hyperuricemia"** submitted by **Mr. N Anand Kumar (Register Number: 155082CH15F07)** as the record of the research work carried out by him, is *accepted asthe Thesis submission* in partial fulfillment of the requirements for the award of degree of **Doctor of Philosophy.**

Chairman-DRPC

Dr. P.E. JagadeeshBabu

Head of the Department

Dept. of Chemical Engineering

NITK, Surathkal

Research Guide

Dr. P.E. JagadeeshBabu

Associate Professor

Dept. of Chemical Engineering

NITK, Surathkal

HEAD OF THE DEPARTMENT
CHEMICAL ENGINEERING
National Institute of Technology Surathkal
P.O. Srinivasnagar - 575 025, D.K., Mangalore

ACKNOWLEDGMENT

timely maintenance of the laboratory equipment and documentation work.

I want to express my sincere gratitude to **Dr. Debashree Chakraborty** and her student **Mr. Bratin Kumar Das,** department of chemistry, for their help and advice.

I sincerely thank NITK for providing me the financial support in the form of fellowship.

I am grateful to all of my former and current student colleagues **Dr.Basavaraj S, Dr. Kunal Kumar, Dr.Swapnali S Pawar, Dr.Smruthi G. Prabhu, Dr. Kishor Kumar M.J, Dr.Rohit Purushottam K, Atmuri Shourya, Sai Teja M V,** the M.Tech students **Miss. Shefali Mehra** and **Mrs. Kamala Meena S.** and the undergraduate students, namely **Mr.Sriwidya Ramana, Miss. Ananya Vasudevan, Mr. Boddu Vishal, Mr.Shubham Meena, Mr.Aditya Kumar Singh, Mr.Narsimha Bhakta**, and all those who have helped me during my dissertation work involved directly or indirectly in my endeavor.

I take this opportunity to thank my colleagues and friends **Dr.Abhinav K Nair, Dr. Sushma H, Dr.Diksha Sharma, Dr.Deepika D, Dr.Mugunthan E, Dr.Shiva Prasad K N, Dr.Venkata Ramana, Mr.Nageswar Rao B, Mr. Krishna Reddy P, Dr. Ravi, Dr.Reddiprasad R, Dr.Prasad Naik, Dr.Hanumanthu Rao, Mr.Subba Rao,** and **Mr.Vamsi** for their timely help and support whenever necessary. I also express my gratitude to all the fellow research scholars of the department.

Finally, I would like to thank my wife **Vinisha**, my parents, and members of my family who have supported me throughout the entire process by keeping me harmonious and helping me putting pieces together. Finally, I thank the almighty whose blessings have enabled me to accomplish my work successfully.

N Anand Kumar

DEDICATED TO MY BELOVED PARENTS

ABSTRACT

Hyperuricemia is a significant risk factor for many health conditions like gout, obesity, diabetes, hyperlipidemia, hypertension, and renal disease. Hyperuricemia is generally caused by increased blood uric acid due to a high intake of purine-rich food, decreased renal uric acid removal, or combining the two. Hyperuricemia is described as high blood uric acid level, which further results in the deposition of urate crystals in the joints and kidneys. When the blood uric acid concentration in adult men is above 7.0 mg/dL and in adult women of 6.0 mg/dL, they are said to have hyperuricemia (Maiuolo et al. 2016).

Hyperuricemia conditions, including refractory gout, are treated by uricases which effectively eliminate pre-existing uric acid crystals in the joints. Uricases have few drug-drug interactions. Though only uricases effectively treat refractory gout, the current uricase formulations are not appropriate for long-term use (Yang et al. 2012). Uricase is a naturally occurring enzyme (urate oxidase, E.C.1.7.3.3) that catalyzes the conversion of uric acid to allantoin and is a promising therapy for hyperuricemia. Rasburicase and pegloticase are the two major uricase formulations that have been approved for the treatment of hyperuricemia. However, unfortunately, prolonged intake of native form of uricase causes severe immunoreactions due to its foreignness (Garay et al. 2012).

In the present research work, we made efforts to use bioinformatics tools to characterize uricase protein sequences from different sources computationally. These protein sequences were subjected to multiple sequence alignment, homology search, domain architecture, motif search, and physiochemical properties. Multiple sequence analysis and homology search results revealed that the amino acid sequences of all the selected sequences have a high degree of similarity. The phylogenetic analysis of all the selected sequences from diverse sources of organisms revealed distinct clusters and demonstrated sequence similarity based on the source of the organism. Each sequence contains six motifs, and each of the twenty-five motifs is unique to its group of uricase sources. The computational physicochemical features of all the selected uricase proteins gave a complete understanding of their properties, namely pI, EC, Ai,

Ii, GRAVY, and are in the nature of basic properties of these enzymes with 33 kDa-39 kDa molecular weight. The amino acid valine has the highest average frequency of 8.79 percent in all the selected sources, indicating that it plays a critical role in the formation of uricase.

Literature survey shows that several *Bacillus* species can produce uricase with 25-30 U/ml of activity. The *Bacillus fastidious* uricase was commercialized by Sigma-Aldrich (product 94310, 9 U/mg) and used for various applications (Pustake et al. 2019a). To expand the usefulness of uricase, it is essential to screen more economical producers of unique properties of novel *Bacillus* uricase, considering the significance of the enzyme in treating hyperuricemia. The detection and identification of new strains capable of producing uricase have a high demand in the medical field. In this work, an attempt has been made to provide a comprehensive description of computational-based structural, functional, and phylogenetic analyses of uricase enzymes from various *Bacillus* species. Uricase protein sequences were analyzed for multiple sequence alignment, phylogenetic analysis, motif assessment, domain architecture review, basic physicochemical property understanding, and *in-silico* identification of uricase amino acid composition. Further, the structural and functional properties of uricase were analyzed. From the analysis, it has been observed that the selected *Bacillus* uricase proteins are active in an acidic to a neutral environment. CFSSP and PSIPRED were used to predict the secondary structure of uricase, which revealed that it is abundant in alpha helices and sheets. The tertiary structure model of the *Bacillus simplex* (WP_063232385.1) uricase protein was predicted and validated. Also, all *Bacillus* species of uricase enzyme and their corresponding genes showed a strong correlation from the phylogenetic comparison of the selected taxa.

Due to the antigenicity issue, the clinical application of uricase as an anti-hyperuricemia agent is limited. To develop less immunogenic uricase, *in-silico* mutagenesis of B-cell and T-cell epitopes have been proposed. The linear B-cell epitopes of *Arthrobacter globiformis (Ag)*-uricase and *Bacillus fastidious (Bf)*-uricase were predicted using the Emini surface accessibility, Parker hydrophilicity, and Karplus & Schulz flexibility methods. T159W, D169C, N264W, and Y203D mutations in *Ag*-uricase resulted in a decreased antigenic probability, whereas S139V,

K215W, G216F, and I172P mutations in *Bf*-uricase resulted in a decreased antigenic probability. Uric acid had a binding affinity of -48.71 kcal/mol for the catalytic pocket of *Ag*-uricase and *Bf*-uricase models, respectively. This energy is stabilized further in the mutant model by -6.36 kcal/mol for *Ag*-uricase and -1.45 kcal/mol for *Bf*-uricase. According to the 100ns MD simulation, both muteins are stable and retained their native-like structural characteristics. The outcome of the above analysis can be a guide for the experimental development of uricase to treat gout and related diseases.

Modifications of proteins are the critical biological tools for the production of a wide variety of proteins. Uricase from *Bacillus fastidious* was successfully conjugated to bovine serum albumin to improve its therapeutic properties. Various molar ratios of bovine serum albumin and glutaraldehyde were conjugated with uricase, and the maximum enzymatic activity of 91.85 percent was obtained at a ratio of 1:6 (mg/ml) uricase: BSA with 0.5 % glutaraldehyde concentration. As determined by the TNBSA assay, the degree of modification indicates that a 1:6 molar ratio of uricase and BSA could result in 76.69 percent of the enzymatic activity. The stability of the conjugated and native uricases was compared at different temperatures (20°C to 60 °C). Likewise, pH stability was investigated at pH values of 7.2 and 9.0. Both native and modified uricase at optimum pH 9.0 shows better retention in enzyme activity after 48 hrs of incubation, which indicates a steady decrease in enzyme activity. The findings of this study indicate that conjugated uricase is effective under physiological conditions, suggesting that it may be a helpful drug for treating hyperuricemia.

Considering the potency of the drug for hyperuricemia, this work aims to study the structure, function, and physiochemical properties of uricase by *in-silico* analysis, and to obtain uricase mutein, an enzyme with reduced immunogenicity, by *in-silico* mutagenesis. This study also aims to understand the various chemical modifications of the enzyme to enhance its efficacy in treating the disease.

Keywords: Uricase, hyperuricemia, physicochemical properties, *Bacillus* species, *in-silico* mutagenesis, immunogenicity, molecular dynamics simulation, bioconjugation, bovine serum albumin

iv

	amino acids. The sections are highlighted in red and pink colors are representing B-cell and T-cell epitopic peptides in uricase sequence. The black color represents identical amino acids, whereas the grey color represents similar amino acids	
5.3	Graphs represent different regions of the amino acid sequence of Uricase that can act as B-cell epitopes in the case of *Ag*-Uricase (A) Emini surface accessibility plot representing maximum antigenicity in *Ag*-Uricase at 167-172. (B) Plot presenting the change in surface accessibility score in *Ag*-Uricase at 167-172. (C) Parker Hydrophilicity prediction plot presenting maximum antigenicity in *Ag*-Uricase at 261-267. (D) The change in hydrophilicity score in *Ag*-Uricase at 261-267. (E) Karplus & Schulz Flexibility plot display maximum antigenicity in *Ag*-Uricase at 156-162. (F) Plot presenting the change in flexibility score in *Ag*-Uricase at 156-162. (G) Graph representing conformational B-cell epitopes from the 3D-structure of *Ag*-Uricase. The selected regions are marked by red circle	115
5.4	Graphs represent different regions of the amino acid sequence of Uricase that can act as B-cell epitopes in the case of *Bf*-Uricase (A) Emini surface accessibility plot representing maximum antigenicity in *Bf*-Uricase at 137-142. (B) Plot presenting the change in surface accessibility score in *Bf*-Uricase at 137-142. (C) Parker Hydrophilicity prediction plot presenting maximum antigenicity in *Bf*-Uricase at 212-218. (D) Plot presenting the change in hydrophilicity score in *Bf*-Uricase at 212-218. (E) Karplus & Schulz Flexibility plot display maximum antigenicity in *Bf*-Uricase at 213-219. (F) Plot presenting the change in flexibility score in *Bf*-Uricase at 213-219. (G) Representing conformational B-cell epitopes from the 3D-structure of *Bf*-	116

LIST OF ABBREVATIONS

3D	Three-dimensional
Ag^N	The hot-spot residue or mutation of *Arthrobacter globiformis*
AI	Aliphatic index
Bf^N	The hot-spot residue or mutation of *Bacillus fastidious*
BSA	Bovine serum albumin
CDD	Conserved domain database
cDNA	Complementary deoxyribonucleic acid
CD-4	Cluster of differentiation 4
CFSSP	Chou and Fasman Secondary Structure Prediction
E.C	Enzyme commission
EC	Extinction coefficient
EDTA	Ethylene diamine tetraacetic acid
EMEA	European Medicines Evaluation Agency
ExPASy	Expert protein analysis system
FDA	US Food and Drug Administration
GRAVY	Grand average of hydropathicity
HLA	Human leukocyte antigen
HMM	Hidden markov model
HPLC	High performance liquid chromatography
Ii	Instability index
PDB	Protein data bank
MEGA	Molecular evolutionary genetic analysis
MEME	Multiple EM for Motif Elicitation
MHC	Major histocompatibility complex
MM/GBSA	Molecular mechanics/ Generalized born surface area
MSA	Multiple sequence alignment
MUSCLE	**MU**ltiple **S**equence **C**omparison by **L**og- **E**xpectation
MW	Molecular weight
NCBI	National center for biotechnology information

NJ	Neighbor-joining
NMR	Nuclear magnetic resonance
PEG	Polyethylene glycol
Pfam	Protein families
Phyre	Protein Homology/AnalogY Recognition Engine
pI	Isoelectric point
PSIPRED	PSI-blast based secondary structure prediction
PSMA	Prostate specific membrane antigen
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RNA	Ribonucleic acid
SAVES	Structure analysis and verification server
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide electrophoresis
SIB	Swiss institute of bioinformatics
SOD	Superoxide dismutase
SOPMA	Self-optimized prediction method with alignment
TNBSA	2,4,6- trinitrobenzene sulfonic acid
UniProtKB	Universal protein resource knowledgebase
UV	Ultra-violet
A.III	Appendix III

NOMENCLATURE

Symbol	Symbol
°C	Celsius
Å	Angstrom
hr	Hour
mg	Milligram
min	Minute
mL	Milli liter
mM	Milli molar
U/mg	Units per milligram
mA	Milliampere
V	Volts
kDa	Kilo Dalton
µM	Micro molar
IU	International Unit
V_{max}	Maximal velocity
K_m	Michaelis-Menten *constant*
Ki	Inhibition constant
R^2	Correlation coefficient

CHAPTER 1

INTRODUCTION

1.1 Background of the research

The 'therapeutic enzymes' were first known about 40 years ago. Therapeutic enzymes can be safely used in medicine either in isolation or in adjunct form with other therapies to cure different diseases and medical conditions effectively and safely. Today, therapeutic enzymes in the native or recombinant form are used as oncolytic, anti-inflammatory agents, anti-coagulants, digestive aids, antimicrobial agents, a replacement for metabolic disorders, etc., to treat many diseases. Therapeutic enzymes were used extensively to cure various genetic and acquired human diseases by removing disease-causing metabolites (Shen and Sali 2006; Tan et al. 2010). Besides, high catalytic efficiency, high purity, greater affinity, unique selectivity, and good pharmacokinetics properties of these enzymes improve their utility in the current medical arena. Though these enzymes are only needed in minimal quantities, their high purity and specificity are essential. Protein size, immunogenicity, immune response, half-life, and other factors can significantly impact the therapeutic potential of these enzymes. Even though their pharmacology was only discovered in the last few decades, many enzymes have been used for a long time to treat a wide range of diseases. In the last ten years, biotechnological advancements have enabled the production of cheaper, safer enzymes with high potency and specificity at a lower cost (Vellard 2003). Enzymatic drugs are distinguished from other types of drugs based on two main characteristics: (i) binding affinity, specificity, and potency (ii) catalytic activity, meaning that they can convert multiple target molecules into products. The above characteristics mentioned are the reasons for their development as potent drugs to treat many diseases. It's also critical to fully comprehend the properties of enzymes and the catalytic activity to maximize their effectiveness and minimize potential side effects (Kumar et al. 2009). The following are some essential therapeutic enzymes

that are used for treating diseases: L-Asparaginase (catalyses the conversion of amino acid asparagine to aspartate: antitumour), L-Glutaminase (converts glutamine to glutamate: antitumor), superoxide dismutase (catalyses the dismutation of superoxide into oxygen and hydrogen peroxide: antioxidant, anti-inflammator), α-galactosidase (degradation of globotriaosylceramide: Fabry disease), Acid α- glucosidase (degradation of glycogen at 1.4- and 1.6-glycosidic linkages: Pompe disease), Lipase (catalyzes the hydrolysis of fats: Lipid digestion), Glucocerebrosidase ceredases (degradation of gluccosylceramide: Gaucher disease), α-glucosidase (catalyzes the breakdown of complex carbohydrates: Antitumor), β-lactamase (catalyzes the conversion of penicillin to penicilloate: Penicillin allergy), Urokinase (catalyzes the conversion of plasminogen to plasmin: Anticoagulant), N-acetyl- galactosamine (cleaving of terminal sulfate from sulfated glycoproteins, glycolipids and glycoaminoglycans: Mucopolysaccharidosis VI), Urate oxidase (catalyzes the conversion of uric acid to allantoin: Hyperuricemia), Collagenase (breaks the peptide bonds in collagen: Skin ulcers), Hyaluronidase (catalyse the degradation of hyaluronic acid: Heart attack), and β-Galactosidase (catalyzes the hydrolysis of galactosides: Antitumor) (Kang and Stevens 2009).

1.2. Gout and Hyperuricemia

Based on the clinical research by McCarty and Hollander, it has been found that up to 5 million people in the European Union and another 5 million in the United States were diagnosed with joint crystals of monosodium urate. Gout, an autoinflammatory disorder, is a common form of arthritis often found in men over 40 years of age and is observed in 1-2% of humans in developed countries. The prevalence of gout in India is unknown. According to a research conducted by the International League of Nations Against Rheumatism's Community-Oriented Program for the Control of Rheumatic Diseases (ILAR COPCORD) in India's Bhigwan hamlet, the prevalence is 0.12 percent. According to a study from Vellore, 15.8 percent of affected patients are under the age of 30; urban Indians are more affected than rural Indians, and due to the increased incidence of metabolic syndrome in the younger population, the first attack of gout happens a decade earlier in them (Matthew and Danda 2004; Smith et al.

2

2010). However, the most accurate method for estimating gout incidence and prevalence is debatable (Punnappuzha et al. 2014). Gout is a systemic disorder characterized by an increase in the levels of serum or plasma uric acid, which leads to hyperuricemia (child > 3.6 mg/dl, mature male > 7.3 mg/dl, adult female > 5.9 mg/dl) (Khade and Srivastava 2015). Elevated serum uric acid concentration can result in the deposition of insoluble monohydrate crystals of monosodium urate in body parts like joints, body tissues, and organs like kidneys, causing inflammation and pain (Scott 1978). The accumulation of uric acid crystals in the synovial membrane and synovial fluid is responsible for acute and chronic inflammation. Additionally, the deposition of uric acid can also induce tumor lysis syndrome and cardiovascular diseases. According to a recent report, the prevalence of gout in the USA is 4% (8.3 million adults), and that of hyperuricemia is 21% (43.3 million) (Zhu et al. 2011). The pathological conditions of gout and hyperuricemia are characterized by increased accumulation or decreased excretion of uric acid in the human body (Gliozzi et al. 2016). Hyperuricemia is an incapacitating problem in cancer patients, especially during chemotherapy for neoplastic diseases. It has a wide range of pathological effects in various organs like kidneys, brains, hypodermic tissues, and joints (da Silva Freitas et al. 2010). Moreover, the accumulation of uric acid crystals in blood serum is dangerous (Navolanic et al. 2003).

Gout classification

The disease known as gout has been studied in greater detail and is classified into primary and secondary gout. Primary gout is defined as an irregular uric acid metabolism in the absence of identifiable symptoms. Secondary gout is brought on as a result of an acquired illness.

1.3. History of Urate Oxidase

The idea of using uricase for gout and hyperuricemic state in cancer patients has been speculated for more than 50 years. The theory was published in Science in 1957, where London and Hudson demonstrated a temporary decrease in serum uric acid in gout patients and non-gout individuals following intravenous injection of small

amounts of purified porcine liver uricase (Figure 1.1). Oppenheimer and Kunkel had previously conducted experiments on chickens and a single patient in the 1940s. Urate oxidase is known to be generated from various sources over the years. Urate oxidase was initially isolated from plants and was used as a medicinal drug, but plant-based urate oxidase showed allergic reactions in humans. Later, microbial and animal origins were used to isolate uricase to avoid the side effects (Navolanic et al. 2003).

Hog liver uricase was obtained and purified from the liver of pigs, but it did not reach the clinical development stage. Uricozyme, a non-recombinant uricase purified from fungal cultures of *Aspergillus flavus*, was marketed in France in 1975 and in Italy in 1984, but its use was eventually stopped since it carried risks of renal failure and tumor lysis syndrome. Rasburicase (Fasturtec), which is a recombinant uricase cDNA from *Aspergillus flavus,* was approved by The European Medicines Agency (EMEA) in 2001 and by the Food Drug Administration in 2002 (Elitek) for prevention / treatment of tumor lysis in children. It is a non-PEGylated, recombinant protein produced using *Saccharomyces cerevisiae.* Rasburicase lowers serum acid levels quickly but less effectively than allopurinol while being well-tolerated than non-recombinant uricase. Due to its short biological half-life and extreme allergic reactions, the production of recombinant urate oxidase is challenging. Uricozyme was replaced by rasburicase, which is the purest form of recombinant uricase. Uricase-PEG5 (5-kDa PEG strands), a non-recombinant enzyme purified from *Candida utilis,* is no longer used in clinical development. Uricase-PEG5, a non-recombinant, bacterial protein (5-kDa PEG strands) from *Arthrobacter protophormiae,* is no longer in clinical development. Uricase-PEG 20 (20- kDa PEG strands linked via succinimidyl succinimide), Pegadricase (previously referred to as pegsiticase), a recombinant uricase cDNA from *Candida utilis.* It is currently in Phase II clinical trials for the treatment of refractory gout and tumor lysis syndrome. Pegloticase (9 strands 10-kDa PEG covalently attached to each subunit), Krystexxa, Puricase, obtained from mammalian cDNA (mainly porcine, with baboon C-terminal sequence) and approved by FDA in October 2010 for treatment of adults with chronic gout refractory to conventional therapy (Garay et al. 2012; Sherman et al. 2008; Terkeltaub 2007). The cost of a bottle containing 7.5mg of Rasburicase (Fasturtec®) is €300 in

France, and that of 8 mg of Pegloticase (Krystexxa®) is US$2300 in the USA. The annual treatment cost of rasburicase is about €7200, and Pegloticase is €41,240 (Garay et al. 2012).
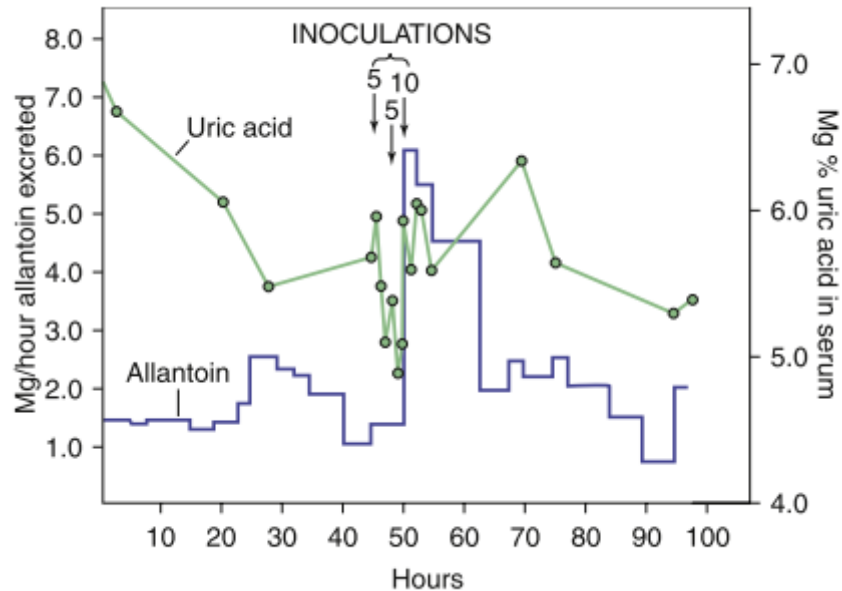
Figure 1.1: Pre- and post-intravenous injection of porcine uricase in 1957 with a gout patient having serum uric acid and urine allantoin

1.4. Diagnosis of gout

A general belief is that an examination of a tissue biopsy followed by monosodium urate crystals analysis is considered a gold standard for gout. Gout may be physically identified by examining the patient's big toe. The interphalangeal joint of the patient's big toe has osteophytic margins, cavitations, and erosions. Serum and urine metabolite concentrations are important markers of various pathogenic disorders such as xanthinuria, hyperuricemia, and gout. The serum urate level does not confirm or exclude the existence of acute gout. (Khade and Srivastava 2015).

Numerous diagnostic techniques to diagnose gout are available, which are as follows:

Conventional Radiography

Radiography is transmitting X-rays via a body part to a flat detector to project an

image. Plain radiography's diagnostic utility is constrained because radiographic shifts are relatively late manifestations of the disease, occurring years after the onset of gout flares.

Ultrasonography

Ultrasound is the preferred method for detecting all musculoskeletal conditions, including soft tissue and degenerative arthritis. This method is widely used to detect gout since it specifically demonstrates monosodium urate crystal deposition in early gout. Early-stages gout, even in patients with asymptomatic hyperuricemia, has been identified with ultrasound results. Ultrasound can be a beneficial adjunct in assessing patients with gout in its very early stages and asymptomatic hyperuricemia, including early gout attacks. It may also be helpful in undiagnosed patients with proven gout (Le Goff et al. 2008).

Magnetic Resonance Imaging (MRI)

Magnetic resonance imaging is an effective technique for detecting synovial involvement, bony erosions, and tophi early in the disease process. Due to the higher cost and portability of MRI, it has not been accepted as a routine diagnostic tool for gout. Despite the publication of established criteria and diagnoses for gout, the use of these criteria should be viewed with caution.

Apart from these, Numerous other techniques exist for determining the concentration of uric acid, including electrochemical method, chemiluminescence, high-performance liquid chromatography (HPLC), and UV-Visible spectroscopy.

Diagnosis and Quantification of Uric acid by Uricase

Based on the uricase enzymatic reaction, UV-Visible spectroscopy is used to assess the concentration of uric acid in serum and urine by analyzing the absorption spectra at 293 nm. After reacting to uricase, the absorption spectrum tests the decrease in uric acid concentration (Dalbeth and Doyle 2012; Khade and Srivastava 2015; Malik et al. 2009).

1.5. Treatment with Uricase

Since 1965, allopurinol has been the only medication approved by the FDA for the treatment of gout. In 2008 the European Medicines Agency (EMA) and in 2009, the US FDA approved febuxostat (non-xanthine oxidase inhibitor) to treat gout. Febuxostat is a potent xanthine oxidase inhibitor and decreases the serum urate concentration in humans predictably. New urate-lowering drugs use two conventional techniques: xanthine oxidase inhibition to reduce uric acid production and uricosuria promotion to enhance renal excretion; Pegloticase, a polymer-coupled type uricase, is a novel method for rapidly lowering serum urate concentrations. Pegloticase, the selective uricosuric drug RDEA594, and multiple interleukin-1 inhibitors (Anakinra, Rilonacept, Canakinumab) are the several other pipeline drugs for gout (Burns and Wortmann 2011).

In the late 1960s, *Aspergillus flavus* non-recombinant uricase was an effective treatment for human tumor lysis syndrome. The non-recombinant enzyme proved to be quite successful but was quite challenging to make, and severe allergic reactions were encountered frequently. In the 1990s, a recombinant uricase from *Aspergillus flavus* was developed and was approved by the US FDA in 2002. While this agent significantly reduces serum urate levels, excitement for its use in gout has been reduced by its lingering immunogenicity and short half-life (Garay et al. 2012). The promising research demonstrates that patients with gout that were treated with rasburicase observed side effects. Additionally, Repeated uricase injections in animals and humans have been shown to cause anaphylactic reactions and the formation of antibodies that neutralize uricase enzyme activity (Altman et al. 1949; Li et al. 2016; Pui et al. 1997).

To circumvent these adverse effects, strategies such as covalent pegylation of mammalian uricase with poly (ethylene glycol) were used. Pegloticase, a recombinant pegylated mammalian uricase (porcine-like), has been investigated into gout. Intravenous administration of pegloticase was found to be superior to subcutaneous administration in phase 1 studies. The uricase plasma activity increases linearly with a dose of up to 8 mg with an intravenous administration of 0·5 mg to 12 mg, explicitly

referring to uricase protein mass within the molecule. The half-life of this activity is 6·4–13·8 days (Ganson et al. 2006; Sundy et al. 2007). Gout prophylaxis was based on colchicine or non-steroidal anti-inflammatory drugs. Patients receiving pegloticase have experienced far more gout flares, infusion reactions, and severe adverse effects than other patients. Most often, infusion reactions resulted in withdrawal, and there were significant correlations between immunogenicity, infusion reactions, and efficacy. Antibodies against pegloticase with a high titer (greater than 1:7290) were associated with loss of response and infusion reactions. Poly(ethylene glycol) antibodies were also more predictive. These antibodies did not affect the function of uricase in vitro (Burns and Wortmann 2011).

Pegloticase, if accepted, will have a far smaller target population than febuxostat since careful selection is critical. Terkeltaub has proposed guidelines for the use of uricase in gout. Uricase is a better treatment for patients with tophaceous gout and have a considerable overabundance of all-out body uric acid and continuing gout assaults or harming arthropathy, and those who are bigoted to conventional medicines that turned ineffective.Glucose-6-phosphate-dehydrogenase deficient patients should not be administered with pegloticase as it can cause hemolysis. Probenecid, sulfinpyrazone, and benzbromarone are conventional uricosuric medications utilized in gout. Uric acid reabsorption can be inhibited by urate-anion exchanger transporter 1 (URAT1) and glucose transporter 9 (GLUT9) (Dalbeth and Merriman 2009; Terkeltaub 2007).

Additionally, Pegloticase can't be associated with any other urate-lowering medications. The primary unfavorable responses of pegloticase include serious cardiovascular events, immunologic response, infusion reactions, nasopharyngitis, contusion, backache, headache, nausea, elevated blood pressure, constipation, chest pain, fever, pruritus, vomiting, and dyspnea. Around 5% of the patients in Phase 3 clinical trial taking pegloticase has experienced a severe cardiovascular incident. Additionally, though pegloticase is less immunogenic than other uricases, approximately 40% of patients can develop pegloticase antibodies with the extension of the treatment period. Further, the high cost of pegloticase can restrict its widespread use in the clinic (Li et al. 2016).

In summary, compared to conventional urate-lowering medications, uricase, which is often used in refractory gout, has a more pronounced effect on decreasing urate levels. As a result, uricases provide intriguing therapeutic opportunities for chronic tophaceous gout when traditional xanthine oxidase inhibitors are ineffective. As per current clinical information, they could go about as a first-line treatment to permit quick urate store exhaustion to break up tophi. Additionally, animal and human studies have demonstrated that repeated uricase injections can result in anaphylactic reactions and the production of antibodies that neutralize uricase's activity. This difficulty can be avoided by conjugating polyethylene glycol to the enzyme's surface, a method that is thought to be necessary to minimize antigenicity and extend the half-life of enzymes. However, since the uricases currently used in clinical practice are derived from lower species, the primary adverse reaction is immunogenicity. Other frequently reported side effects include headache, nausea, fever, vomiting, rash, and other cardiovascular problems. The expression of pegloticase antibodies results in about half of the patients failing to respond to therapy, which increases the risk of anaphylaxis. Significant immunogenicity reduction is a critical goal in producing new uricases (Li et al. 2016).

In the present work, an attempt was made to analyze the individual amino acid sequences of uricase from four different sources, namely bacteria, fungi, plants, and animals, to elucidate uricase structure and physiochemical properties by several standard biocomputational tools. Then, the computational characterization of 70 uricase protein sequences from various *Bacillus* species and an investigation of their physical parameters, secondary and tertiary structure, functional properties, domains, motifs, and phylogenetic relationship using multiple bioinformatics tools was performed. The linear B-cell, conformational B-cell, and MHC-I-based T-cell epitopes were identified to reduce the immunogenicity of uricase sourced from *Arthrobacter globiformis* (*Ag*-Uricase) and *Bacillus fastidious* (*Bf*-Uricase). Emini surface accessibility, Parker hydrophilicity, and Karplus and Schulz flexibility methods were employed to detect the continuous B-cell epitopes and corresponding hot-spot residues. Similarly, the deimmunization method was used to identify T-cell epitopes. Next, the hot-spot residues were mutated to reduce the antigenic character of

9

the identified epitopes. Lastly, the impact of mutagenesis on the catalytic activity and the structural stability of uricase was assessed by molecular docking, free energy calculations, and molecular dynamics simulation.

Later, the bio-conjugation of uricase with BSA was studied to improve its properties. This is the first experimental study in which uricase from *Bacillus fastidious* is modified with BSA. Variables like the BSA concentration, glutaraldehyde concentration (cross-linker), pH, and temperature were optimized to achieve the desired degree of conjugation with desired residual activity. Further, the conjugate's stability with respect to temperature and pH was assessed, and the kinetic parameters were analyzed and discussed.

CHAPTER 2

LITERATURE REVIEW

2.1 Therapeutic enzymes

Each enzyme is a protein (or, in some cases, an RNA molecule) with a specific function like catalyzing reactions, aiding cell maintenance, and protecting against diseases in the ordinary course of development. Enzymes can function in intracellular and extracellular conditions or even on the cell membrane's surface. Due to their high activity and selectivity, enzymes have enormous therapeutic potential. However, due to their low stability, immunogenicity, and possible systemic toxicity, enzyme drugs are not widely used in clinical practice (Batool et al. 2016).

Protein therapeutics is the pharmaceutical industry's most valuable component. Therapeutic enzymes must exhibit pharmacokinetics, specificity, and a high degree of purity (Vidya et al. 2017). Bacteria is a significant source of therapeutic enzymes. Therapeutic proteins can also be obtained from various biological sources like organs, tissues, animal fluids, and genetically modified organisms and cells. During systematic administration of bacterial enzymes, the human body recognizes them as a foreign antigen, leading to the secretion of antibodies. Antibody secretion of B-cells is mainly governed by identifying antigenic epitopes on the surface of bacterial enzymes. Therefore, bacterial enzymes are limited due to their immunogenicity, poor stability, and toxicity.

Moreover, recent therapeutic enzymes are associated with common problems such as high degradation rates or rapid clearance. The diseases can be diagnosed with different enzyme assays and help to assess patients' responses to treatment. Enzymes are frequently used as biomarkers in a variety of disease states. Biotechnology has resulted in more potent and specific pharmaceuticals (Dean et al. 2017).

11

2.2 Uricase introduction

Urate oxidase (Uricase; EC 1.7.3.3) is a copper-binding enzyme that belongs to the class of oxidoreductases. It is involved in the purine destruction pathway and plays a vital role in nitrogen metabolism. It is a homotetramer with two subunits with a molecular weight of around 145-150 kDa and a 35 kDa subunit calculated based on cDNA sequence (Nanda and Jagadeesh Babu 2014; Schiavon et al. 2000). This enzyme catalyzes the conversion of uric acid into allantoin, carbon dioxide, and hydrogen peroxide in the presence of excess molecular oxygen (Dabbagh et al. 2015). Allantoin is more water-soluble (147 mg/dL) compared to uric acid (11 mg/dL) and can easily be excreted through the kidney (Khade and Srivastava 2015). Uricase is widely present in most vertebrates; however, higher primates (humans, apes, and certain monkeys) lack active uricase due to mutations during primate evolution. Uric acid is excreted as the final product of purine catabolism in these organisms (Wu et al. 1989). Hence, uricase can be used to treat hyperuricemia that usually develops during tumoral lysis and organ transplants (Howard et al. 2011; Nanda and JagadeeshBabu 2016).

The increasing importance of uricase is probably due to its potential use in medicinal chemistry and the treatment of several diseases. Generally, peroxisomes are the storehouses of uricase in various animal and plant tissues such as cowpea, bean, soybean, other legumes, and in rats, mice, fish (liver, kidney, and brain) tissue (Capote-Maínez and Sánchez 1997; Lucas et al. 1983). Bacterial organisms include *Arthrobacter globiformis*, *Bacillus fastidious*, *Bacillus subtilis*, *Pseudomonas aeruginosa*, *Escherichia coli*, *Proteus Vulgaris*, and *Streptomyces cyanogenus* are used for industrial production of the uricase enzyme. (Kai et al. 2008; Li et al. 2017; Pfrimer et al. 2010). It has also been synthesized from various natural sources such as fungi, yeast, plants, insects, mammals, and genetically engineered microorganisms (Zhou et al. 2005). Uricase was initially isolated from mammalian organisms (Adámek et al. 1989), as it was first found in the bovine kidney (Yazdi et al. 2006). Uricolytic organisms such as most vertebrates have uricase, but it is absent in man, anthropoid apes, uricotelic reptiles, birds, and almost all insects. The nonsense

mutation in exon two during the evolution of higher primates (apes and humans) is the cause for the absence of uricase (Schiavon et al. 2000). However, the exact reason for the loss of uricase activity in humans and some primates is unknown. The loss of uricase activity has resulted in a decreased cancer rate and increased life span of hominoids due to the required concentration of uric acid in serum, a potent scavenger of free radicals (Dabbagh et al. 2012). The loss of functional uricase is an ''evolutionary accident" in human beings that cause the accumulation of toxic uric acid resulting in gout (Keebaugh and Thomas 2010; Oda et al. 2002).

In 1957, London and Huston proposed using urate degrading enzyme uricases as a new way to treat hyperuricemia. As the efficiency of other therapies was at stake, the parenteral administration of uricase is exploited to treat hyperuricemia. The first attempt to use the uricase enzyme for gout therapy was by London and Hudson. They injected uricase into two patients with gout and observed a considerable reduction in blood urate level. Since then, uricase from various sources is used to cure hyperuricemia (London and Hudson 1957). Natural uricase, obtained from *Aspergillus flavus* cultures (Uricozyme [TM]), is used to prevent and treat hyperuricemia occurring during chemotherapy and treat hyperuricemia associated with organ transplants (Rozenberg et al. 1995). But, upon administering this native uricase in the human body, immunogenicity and hypersensitivity were the two dangerous problems faced, as it was obtained from a microbial source. Since 1996, the molecule being used is Rasburicase (Fasturtec [TM] in Europe, Elitek [TM] in the USA) brought by recombinant DNA technology. Rasburicase is obtained from a genetically modified strain of *Saccharomyces cerevisiae,* which expresses uricase cDNA, cloned from a strain of *Aspergillus flavus* (McDonnell et al. 2006). The use of Rasburicase in other hyperuricemia conditions, such as chronic gout and short-term use in tumor lysis syndrome management in paediatric patients with malignancies, is complex because it poses a short plasma half-life and requires daily living administration.

Humans do not express uricase, and the enzyme would be expected to be seen by the immune system as a foreign protein. Indeed, uricase from microorganisms and animals, when administered in patients, is highly antigenic, and the chronic treatment

with this enzyme frequently results in allergic reactions and anaphylactic shock (Bomalaski et al. 2002). The clinical application of uricase has been limited due to its undesirable biological properties (i.e., premature degradation and inactivation by endogenous proteases, elimination by the reticuloendothelial system, immunogenicity, and toxic side effects caused by the host immune system reacting toward foreign proteins) (Tan et al. 2012). For this reason, the development of a methodology to screen uricases from a different source of organisms diminishes its immunogenicity by *in silico* approaches and modify uricase, which is preserving biological activity, making it more suitable for therapeutic purposes, has been investigated.

2.3 Uricase structure

Bacterial uricase consists of two tetramers composed of four identical subunits. The overall dimension of one tetramer was reported to be $74\times86\times76$ Å^3. Each subunit of uricase contains 287 amino-acid residues and consists of four α-helices, two one-turn helices, eight long, and two short β-strands (Juan et al. 2008). Each monomer of uricase can be divided into two similar domains known as T-fold domains (Colloc'h et al. 1997), and each T-fold domains consists of antiparallel $\beta\beta\alpha\beta\beta$ superfold. A cylindrical tunnel with a rough diameter of 30 Å and length of 80 Å can be found at the center of the uricase tetramer. Figure 2.1 presents the complete structure of uricase (PDB ID: 4R8X). The enzyme contains no metal or other cofactors, but the catalytic process needs an oxygen molecule and a water molecule. The catalytic activity mainly depends on three highly conserved active site residues-arginine, glutamine, and phenylalanine (Wu et al. 1989).
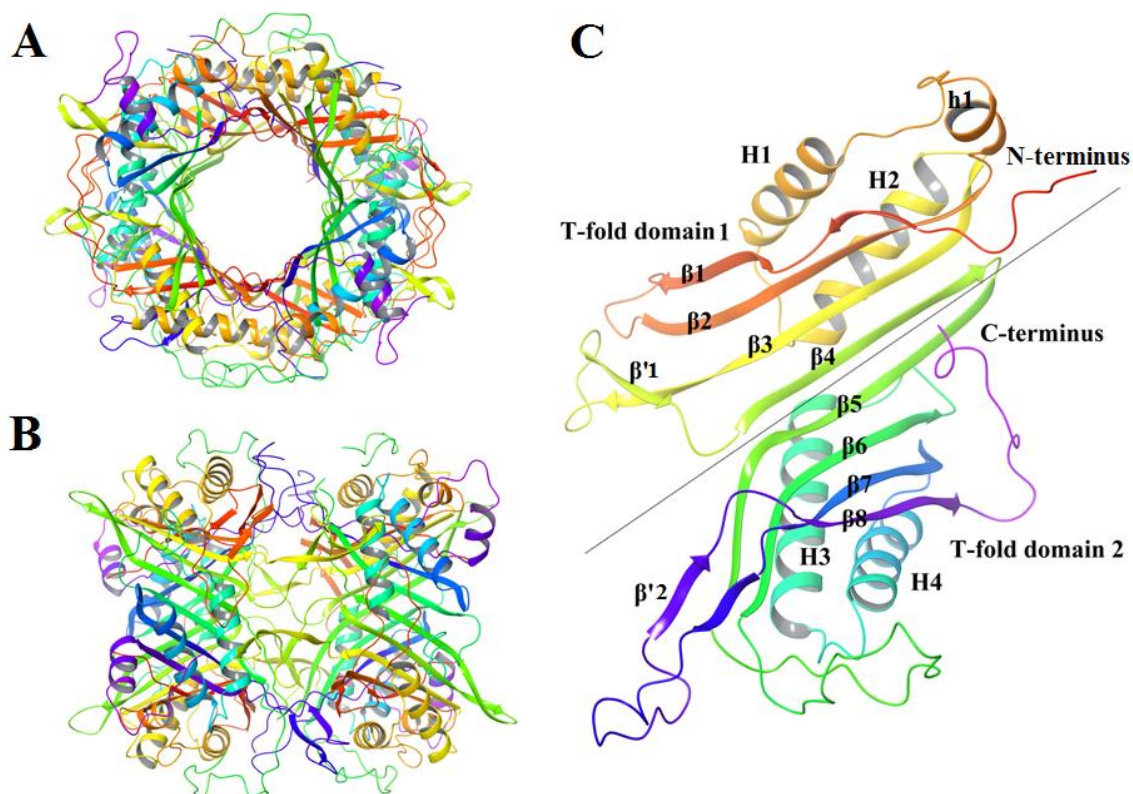
Figure 2.1: The complete structure of Uricase (PDB ID: 4R8X). (A) Front view of the tetramer of bacterial Uricase showing the big tunnel at the center of the protein. (B) Side view of Uricase. (C) The monomer of Uricase showing two similar T-domains. β, β', H, h symbols are used to specify the secondary structures like β-sheets and α-helices. (Sourced from: (Juan et al. 2008))

2.4 Mechanism of action of uricase

Uricase, which is found in the liver, catalyzes the oxidation of uric acid to allantoin in the presence of oxygen, resulting in the release of hydrogen peroxide (H_2O_2). In the past, several publications it is stated that the primary product of uric acid oxidation by uricase was allantoin. However, subsequent 13C-NMR monitoring revealed that the conversion of uric acid to allantoin involves three reactions: oxidation, hydrolysis, and decarboxylation (Bongaerts and Vogels 1979; Modrić et al. 1992). The first product produced during the study of NMR was 5-hydroxyizurate (HIU) in uric acid oxidation. HIU was a highly unstable compound that could degrade spontaneously or

in the presence of an enzyme to form 2-oxo4-hydroxy-4-carboxy-5-ureidoimidazoline (OHCU) (Kahn et al. 1997). The 5-hydroxyisourate hydrolase and OHCU decarboxylase enzymes catalyzed the hydrolysis and decarboxylation reactions more rapidly, respectively (Figure 2.2). These two enzymes in living organisms have a functional uricase. Few other enzymes have transformed allantoin into urea and glyoxylic acid. The substrate (uric acid) forms a strong bond with one of the enzyme's subunits through interaction with glutamine (Gln223), leucine (Leu222), and arginine. (Arg180). Although the enzyme's other subunit interacts with aspartate (Asp68) and threonine (Thr67) (Gabison et al. 2010; Juan et al. 2008)**.**
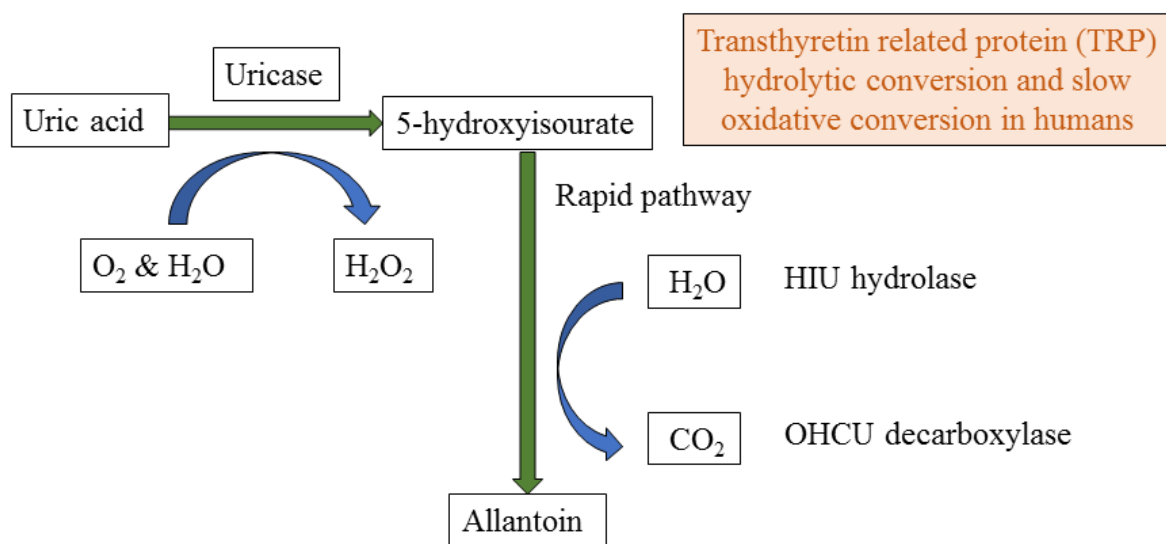
Figure 2.2: A way to degrade uric acid to allantoin by uricase in most mammals and other species (Nuki 2012)

2.5 Purine metabolism

Purine metabolism is the metabolic pathway that synthesizes and degrades purine. Purines are obtained in humans by diet, endogenous nucleotide degradation, and de novo synthesis. Purines are found in shellfish, sheep, salmon, rabbit, pork, perch, peas, oysters, sweet bread, sardines, liver, kidneys, and grains. Purine compounds are biosynthesized, converted, and degraded using a variety of enzymes (Doherty 2009).

Purine metabolism produces a variety of end products that vary by species. Since man, hominoids, birds, and reptiles lack uricase activity, the result of purine metabolism is uric acid (Figure 2.3). Uric acid is mainly produced in the liver and discharged into the urine by the kidney. Allantoin is a by-product of purine metabolism in non-primate mammals. The allantoin is further degraded by the effect of allantoinase and allantoicase in most fish and amphibians to urea and glyoxalate through allantoic acid. However, Mahler et al. proposed the hypothesis of the catabolic uric acid pathway in 1970, in which uric acid was converted to allantoin by urate oxidase and then hydrolyzed into allantoic acid by allantoinase (Mahler 1970; Oda et al. 2002). Purine plays a crucial role in nucleic acid production. Adenine, guanine, and hypoxanthine are the most frequently occurring purine bases in nucleotides. Purines, xanthine, and uric acid are the products of nucleotide catabolism, and xanthine monophosphate is the metabolic intermediate. Uric acid is the main product of purine metabolism in humans and apes. Primates excrete uric acid, but it is catabolized into other end products by mammals other than primates (Alvarez-Lario and Macarron-Vicente 2010).
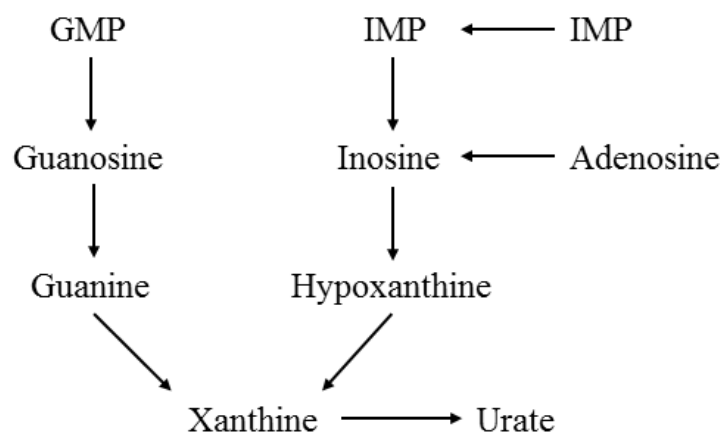
Figure 2.3: Metabolic pathway of purine degradation to uric acid in humans

2.6 Sources of uricase

Uricase has been detected in different sources such as bacteria, algae, fungi, plants, and animals. While having the presence of the gene, primates could not synthesize it due to gene point mutations. Numerous researchers from all over the world have investigated different sources of urate oxidase. Some researchers have done further research on each source of uricase. Numerous scientists favoured microbial uricase due to its low cost, more effectiveness, and widespread availability. While microorganisms such as bacteria, fungi, actinomycetes, yeast, and algae are all highly efficient enzyme producers, their enzymatic properties vary considerably.

Bacteria

The uricase preparations from various bacterial species like *Bacillus licheniformis* (Pawar and Rathod 2018), *Bacillus subtilis* (Pfrimer et al. 2010), *Bacillus fastidious* (Tan et al. 2012), *Arthrobacter globiformis* (Suzuki et al. 2004), *Bacillus thermocatenulatus* (Lotfy 2008), *Bacillus sp.* (Bongaerts and Vogels 1979), *Escherichia coli* (Li et al. 2006), *Enterobacter cloacae* (Machida and Nakanishi 1980), *Proteus Vulgaris* (Azab et al. 2014), *Pseudomonas aeruginosa* (Anderson and Vijayakumar 2011), *Microbacterium.* (Zhou et al. 2005), *Microbacterium* sp. (Kai et al. 2008), *Xanthomonas fuscans* (Ram et al. 2015) and *Comamonas* sp. BT UA (Ghosh and Sarkar 2014).

Fungi

Uricase activity has been detected in several fungi, namely, *Aspergillus niger* (Geweely and Nawar 2011), *Aspergillus welwitschia* strain 1–4 (El-Naggar et al. 2019), *Aspergillus flavus* (Bayol et al. 2002), *Mucor hiemalis* (Yazdi et al. 2006), *Gliocladium viride* (Nanda et al. 2012), *Cryptococcus sp.* (Lee et al. 2013) and *Gliomastix gueg* (Atalla et al. 2009).

Yeast

There was evidence of uricase activity in *Candida tropicalis* (Tanaka et al. 1977),

Candida sp. (Liu et al. 1994), and *Saccharopolyspora sp*. (Khucharoenphaisan and Sinma 2011).

Actinomycetes

The detection of uricase activity has been studied by many workers, e.g., *Streptomyces graminofaciens* (Azab et al. 2014), *Streptomyces albidoflavus* (Azab et al. 2014)*,* and *Streptomyces sp.* (Watanabe et al. 1969).

Plants

Uricase activity has been found in Cowpea (Rainbird and Atkins 1981), Soyabean root nodules (Bergmann et al. 1983; Lucas et al. 1983), *Triticum aestivum* L.*, Vicia faba* L., *and Cicer arietinum* L. (Montalbini et al. 1997), *Vigna unguiculata, Phaseolus* species and *Glycine max* (Capote-Maínez and Sánchez 1997).

Algae

Uricase has been detected in *Chlamydomonas reinhardtii* (Alamillo et al. 1991).

Animals

Uricase has been found in vertebrates, invertebrates such as mammals, reptiles, amphibians, and Pisces.

Mammals

Uricase activity is present in the bovine kidney, camel liver, rat liver, pork liver, porcine liver, and pig liver.

2.7 Bioconjugation of therapeutic enzymes

The enzymes administered parenterally will be recognized as a foreign antigen and elicit an immune response in the human system. The enzymes may also be removed from circulation by some of the proteolytic enzymes in plasma and renal ultra-filtration, thus decreasing its circulatory half-life, increasing the need for frequent

administration (Schellekens 2002). These limitations can be negated by a process called bioconjugation. i.e., disguising the enzyme surface by covalent modifications. Bioconjugation, a method of covalent modification of biomolecules (proteins, hormones, enzymes, growth factors, nucleotides, etc.) using a physiologically labile bond, helps in the stabilization of compounds in circulation, protects them from proteolytic enzymes, reduces the immunogenicity of the polypeptides and also provides new possibilities for drug targeting (Veronese and Morpurgo 1999).

Helmut Ringsdorf first proposed the concept of covalent polymer-drug conjugates in 1975. Bioconjugation involves the formation of different types of covalent attachments between a polymer and enzyme. Duncan and his colleagues produced the first successful active drug conjugate by using a polymer N-(2-hydroxypropyl) methyl acrylamide (HPMA) for the release of anticancer drug doxorubicin (Duncan et al. 1988). After this, various drug molecules were conjugated to polymers and got approval from the FDA for treatment. The conjugation of therapeutic enzymes to polymers started with the conjugation of streptokinase to dextran (35-50 kDa) with significant therapeutic success (Tochilin et al. 1982). Different types of polymer-drug conjugates are discussed extensively by Elvira et al., 2005. Some of the critical desired properties of polymers used for bioconjugation are: they should be non-toxic, non-immunogenic, can be quickly cleared from the body without accumulation in any organs, and highly pure. The conjugation can be done using different types of polymers like natural polymers (dextran, dextrin, pullulan, hyaluronic acid, etc.), synthetic polymers (PEG, polyacrylomorpholine, HPMA, polyvinyl alcohol, polystyrene co maleic acid, etc.), and semi-synthetic polymers(poly-L-lysine, polymaleic acid, PHEG, etc.) (Pasut and Veronese 2007).

Uricase enzyme was first conjugated to polyethylene glycol (PEG) by Nishimura et al. in 1979. They observed a considerable decrease in immunogenicity but had to compromise the residual enzymatic activity of the conjugate. PEG strands appear to protect a protein by causing steric hindrance. This causes a shield effect on the enzyme surface that will block its recognition by the immune system. PEG also inhibits interaction with cell-associated receptors and enzymes that may degrade the

20

protein and increase protein half-life. The uricase from different sources has been conjugated using other polymers, but a decrease in activity was observed. Freitas et al. reported that a conjugate of PEG derivatives with uricase showed only 75% residual activity compared to the native enzyme (Nishimura et al. 1979; da Silva Freitas et al. 2010).

Though PEGylation technology has been proved to be a very effective method, they also have some limitations. Some of the significant limitations are it is non-biodegradable and its accumulation in organs after prolonged usage. The studies on long-term side effects of these pegylated therapeutic enzymes are still under research. It has also been reported that anti-(PEG-enzyme) antibodies are detected in some patients' blood under treatment (Ganson et al. 2006; Garay et al. 2012). There is a need to find an alternative polymer that can be used for conjugation to improve the pharmaceutical properties of uricase. As a natural alternative to synthetic polymer, PEG, and its derivatives, polysaccharides are water-soluble and biocompatible polymers. The bioconjugation of polysaccharides to enzymes will help improve its circulatory half-life and decrease the enzyme's immunogenicity and retain overall therapeutic efficiency (Beesh et al. 2010).

2.8 Bovine Serum Albumin (BSA)

The use of bovine serum albumin in pharmaceuticals has a long history. Albumin has excellent biodegradability, biocompatibility, nontoxic and non-antigenic properties. Bovine serum albumin is a giant globular protein made up of a single polypeptide chain consisting of 583 amino acid residues and no carbohydrate residues, having a molecular weight of 66 kDa. It was reported that BSA could alter the heat denaturation of the protein by partial unfolding between 40°C and 50°C, exposing the non-polar residues on the surface and facilitating reversible protein-protein interactions. Thus, the unknown nature of interaction and the extended effects of BSA are of considerable interest since they could reveal a specific mechanism by which proteins can stabilize enzymes and predict whether BSA/protein could be a good enzyme modifier (Kishore et al. 2014).

Literature information about the conjugation of bovine serum albumin with different enzymes is discussed here.

Catalase-BSA conjugate was prepared using a heterobifunctional reagent. Here N-succinimidyl 3-(2-pyridyldithio) propionate (SPDP) was used as a heterobifunctional reagent. The molecular weight of the filtered conjugate was 380 kDa. After treatments of heat, urea and trypsin hydrolysis process, the conjugate's half-life times were 6, 69, and 65 minutes. So even though native catalase took 3, 22, and 28 minutes, respectively (Hu and Su 2002).

The function and thermal stability of amylase were analyzed using glutaraldehyde as a binder and BSA as a modifier. The enzyme's optimum temperature was discovered to be 50°C ± 2°C. Additional temperature rise resulted in the enzyme's thermal inactivation, which was irreversible. The level of thermal inactivation was reduced significantly after the enzyme was modified with BSA. Even after 3 hours of incubation, at 80°C, it was discovered that -amylase modified with BSA retained its function. On modification with BSA, the apparent thermal energy inactivity (Ed) of α-amylase increased markedly. At 70°C and 80°C, the conjugate half-life showed that 2.5 times greater than that of native α- amylase. The kinetic constants (Vm and Km) were also determined in this research work (Kishore et al. 2014).

α-1,4-glucosidase was conjugated to develop soluble polymers of albumin (rat or human) by using cross-linking agent glutaraldehyde. The polymer has an average molecular mass of eighty thousand, meaning that each enzyme molecule is made up of twelve albumin molecules. The enzyme-albumin polymer is more resistant to heat denaturation and trypsin proteolysis than an equal volume of free enzyme. The enzyme-albumin polymer's extreme level of resistance to bioinactivation is addressed in relation to the use of enzyme replacement therapy in a variety of metabolic diseases, including Pompe's disease (glycogenosis type II), in which alpha-1,4-glucosidase is the deficient enzyme (Poznansky and Bhardwaj 1980).

Bovine liver superoxide dismutase (SOD) was cross-linked to albumin results in soluble conjugates with up to 70% of the original enzyme activity retained. Compared

to native SOD (6 min), these conjugates have significantly longer plasma half-lives of enzymatic activity (15 h). They have strong anti-inflammatory and less antigenic properties compared to native SOD *in vivo*. These conjugates increase the potential for research into inflammation treatment (Wong et al. 1980).

Growth hormone's clinical application has been impeded by its low molecular weight and rapid elimination by the kidneys. In addition, different levels of nephrotoxicity have also been confirmed for specific proteins that are readily glomerulus. This limitation was tried to avoid by conjugating the somatotropin growth hormone to serum albumin to change the pharmacokinetics of peptides even as maintaining its activity (Poznansky et al. 1988).

Exogenous proteins and enzymes must be administered successfully for medicinal purposes. Immunogenicity should be adjusted without significant loss of the desired biological activity required for therapeutic reasons. To demonstrate this, the researchers used a model experiment in which the asparaginase from *Escherichia coli* was altered by copolymerization with isologous albumin, and changes in its immunogenicity have been investigated in the rat. The findings indicated that the conjugate was significantly less immunogenic while retaining adequate enzyme activity (Yagura et al. 1981).

1,4-glucosidase chemically cross-linked with antibody molecules and homologous albumin against isolated rat hepatocytes, a responsive and accurate soluble enzyme-polymer complex with an approximate molecular weight of 10^6. when injected intravenously, the ^{125}I-labelled complex is found to associate preferentially with hepatocytes, as opposed to free albumin or non-specific IgG-albumin, which also associate with the kupffer cells in high concentrations. The method has many benefits, including the ability to target enzymes to particular tissues and cells and the potential to reduce hepatocyte glycogen content in glycogenesis type II (Poznansky and Bhardwaj 1980).

The anticancer activity of polymer-based conjugates of asparaginase and homologous albumin was evaluated using the mouse models PANC-1 and 6C3HED (Human

pancreatic tumor cell line and lymphosarcoma). The conjugate seems more immune to proteolysis and is much more efficient as an anticancer agent in C3H/HeJ rice with 6C3HED lymphosarcoma. In view of anticancer activity, the enzyme is nearly twenty times more efficient, correlated to free enzyme. Comparably, the polymeric type of L-Asparaginase inhibits the cancer cells grown in tissue culture quite effectively. Here the polymeric-based form of L-Asparaginase's improved performance is most likely due to its biodegradation resistance. It's also shown that the surface of cell-particular MABs antibodies (Monoclonal) can be used to guide the polymer to cancer cells (Poznansky et al. 1982).

The study attempted to achieve increased stability and functionality of asparaginase. In this research, crosslinking BSA and ovalbumin with glutaraldehyde, mono-methoxy polyethylene glycol, and N-bromosuccinimide modify the enzyme. When contrasted to the native enzyme, the ovalbumin modification resulted in a ten-fold increase in enzyme activity. Whereas modification with BSA via glutaraldehyde cross-linking led to high stabilization of asparaginase, which would have been 8.5- and 7.62-fold higher at 28°C and 37°C by the end of one day compared to the native enzyme. So the quantity of conjugate formed affected these effects. Additionally, modification significantly increased L-asparaginase's half-life and serum stability (Mohan Kumar et al. 2014).

2.9 Various protein modification techniques

Posttranslational modifications (PTMs) are covalent processing processes that alter the characteristics of a protein via proteolytic cleavage and the addition of a modifying group to one or more amino acids, such as acetyl, phosphoryl, glycosyl, or methyl. PTMs exert a major influence on the structure and dynamics of proteins in a wide variety of biological processes. PTMs can occur in a single kind of amino acid or in a mixture of amino acids and alter the chemical characteristics of changed sites. The following subsections cover in greater detail the most extensively investigated PTMs.

24

Phosphorylation

The target sites for phosphorylation include Ser, Thr, Tyr, His, Pro, Arg, Asp, and Cys residues, but this change occurs most frequently on Ser, Thr, Tyr, and His residues. This alteration has the potential to significantly alter the function of proteins in a short period of time via one of two mechanisms: allostery or binding to interaction domains. It has been demonstrated that disruptions in the phosphorylation pathway can result in a variety of disorders, including cancer, Alzheimer's disease, Parkinson's disease, and heart disease.

Acetylation

The enzymes lysine acetyltransferase (KAT) and histone acetyltransferase (HAT) catalyse acetylation.Nα, Nε and O-acetylation occur at varying frequency on Lys, Ala, Arg, Asp, Cys, Gly, Glu, Met, Pro, Ser, Thr, and Val residues, however acetylation is most frequently recorded on Lysine residues. According to research, acetylated lysine is required for cell formation and its deficiency causes major diseases as cancer, ageing, immunological problems, neurological (Huntington's and Parkinson's) and cardiovascular diseases.

Glycosylation

In this modification, oligosaccharide chains are covalently bonded to particular residues. A glycosyltransferase enzyme is used to help with this enzymatic process, which usually happens in the side chain of residues like Trp, Ala, Arg, Asn, Asp, Ile, Lys, Ser, Thr, Val, Glu, Pro, Tyr, Cys and Gly. It is more typically found on Ser, Thr, Asn, and Trp residues in proteins and lipoproteins, though. It has been demonstrated that a failure in this mechanism contributes significantly to the development of several diseases such as cancer, cirrhosis of the liver, diabetes, HIV infection, Alzheimer's disease, and atherosclerosis.

Ubiquitylation

Ubiquitylation is a significant reversible PTM. This is a very versatile PTM because it

can occur on all twenty amino acids. This is a very versatile PTM because it can occur on all twenty amino acids. However, it happens more often on lysine. Ubiquitinated proteins can be acetylated on Lys residues or phosphorylated on Ser, Thr, or Tyr residues, significantly affecting the outcome of signalling. The ubiquitylation modification of substrate proteins can be eliminated by a group of specialist proteases known as deubiquitinase. Different types of cancers, metabolic syndromes, inflammatory disorders, type 2 diabetes, and neurodegenerative diseases can all be caused by dysfunction in the ubiquitin pathway.

Methylation

In target proteins, methylation occurs on the Lys, Arg, Ala, Asn, Asp, Cys, Gly, Glu, Gln, His, Leu, Met, Phe, and Pro residues. However, at least in eukaryotic cells, lysine and arginine are the two primary target residues for methylation. A deficiency in this modification can result in a variety of illnesses, including cancer, mental retardation (Angelman syndrome), diabetes, lipofuscinosis, and occlusive disease.

SUMOylation

SUMOylation occurs by SUMO (Small Ubiquitin-Related Modifier), a protein with a three-dimensional structure similar to that of ubiquitin that has been found in a wide variety of eukaryotic organisms. SUMOylation is a way to change the -amino group of lysine residues in a target protein through a multi-enzyme chain. SUMOylation changes frequently occur at the consensus motif WKxE. (where W represents Lys, Ile, Val or Phe and X any amino acid). Numerous data indicate that SUMOylation plays a significant role in the development of a range of human disorders, including cancer, Alzheimer's disease, Parkinson's disease, viral infections, cardiovascular disease, and diabetes.

Palmitoylation

Lipidation is a significant type of PTMs that involves the covalent binding of lipids to proteins. These PTMs occur via a range of lipids such as octanoic acid, myristic acid, palmitic acid, palmitoleic acid, stearic acid, and cholesterol, among others. Three

major forms of these lipid changes are myristoylation, palmitoylation, and prenylation. Palmitoylation is the chemical bonding of fatty acids, such as palmitic acid, to Cys, Gly, Ser, Thr, and Lys residues. Palmitoylation dysfunction has been associated with a variety of diseases, including neurological disorders (Huntington's disease, schizophrenia, and Alzheimer's disease) and several malignancies.

Myristoylation

Myristic acid, a 14-carbon saturated fatty acid, is covalently linked to the N-terminal glycine residue following the elimination of the starting Met in myristoylation. Met-Gly-X-X-X-X-Ser/Thr is a common motif for this attachment, which is mediated by N-myristoyl transferase (NMT). Myristoylation is more prevalent on Gly residues and less prevalent on Lys residues. Myristoylation has been implicated in the development and progression of a variety of disorders, including cancer, epilepsy, Alzheimer's disease, and Noonan-like syndrome, as well as viral and bacterial infections.

Prenylation

Prenylation reaction happens on cysteine and at the substrate protein's carboxyl-terminal end. Disruption of this alteration has been shown to be critical in the aetiology of cancer, cardiovascular and cerebrovascular illnesses, bone diseases, progeria, metabolic diseases, and neurodegenerative diseases.

Sulfation

Tyr, Cys, and Ser have been identified as prenylated protein target residues. This PTM is implicated in a variety of disorders, including autoimmune diseases, HIV infection, lung disease, and multiple sclerosis.

Proteolysis

Under normal conditions, peptide bonds stay together indefinitely. This means that cells need a way to break these bonds. Proteases are a group of enzymes that break down the peptide bonds in proteins. They are important for antigen processing, apoptosis, surface protein shedding, and cell signalling. Proteases are categorised

according to their mode of action, with aminopeptidases and carboxypeptidases cleaving at the amino or carboxy terminus of a protein, respectively. Another classification scheme is based on the active site groups of a particular protease that are involved in the proteolytic process (Ramazi and Zahiri 2021).

2.10 Computational studies of several enzymes

The assessment of amino acid sequence based on computational analysis of various enzymes like alpha-amylase (Vivek et al. 2012), β-propellerphytase (Mathew et al. 2014),β-galactosidases (Bose et al. 2013), fructosyl transferase (Alméciga-Díaz et al. 2011), glutaminase (Irajie et al. 2016), histidine acid phytase (Kumar et al. 2012), L-asparaginase (Dwivedi and Mishra 2014), manganese peroxidase (Yadav et al. 2017), phytases (Verma et al. 2016), alkaline proteases (Morya et al. 2012), pullulanase (Rahmatabadi et al. 2017), pectin lyase and pectinase (Yadav et al. 2009), pectate lyase (Dubey et al. 2010), polyphenol oxidase (Malviya et al. 2011), xanthine dehydrogenase (Dwivedi et al. 2013) have been described based on bioinformatics.

Wet-lab techniques are a time-consuming and expensive process compared to the application of bioinformatics tools and servers, which are more economical and time-saving methods (Rahmatabadi et al. 2017). Bioinformatics web-based servers and devices help understand unknown protein profiles with the aid of their sequence, structural, functional, and evolutionary data obtained by computational genomics and proteomics studies (Koteswara Reddy et al. 2017). For functional analysis of a protein, the 3D structure is required. Protein modelling is done to predict the structure of unknown proteins (Koteswara Reddy et al. 2017). Nowadays, the utilization of *in silico* analysis and characterization of various industrially important enzymes using their protein sequences has gained importance (Nezafat et al. 2015).

Sequence-based computation analysis of various enzymes has been studied for different purposes discussed here.

Pullulanase from various *Bacillus* species was characterized and functionally analyzed computationally. Pullulanase protein sequences were investigated for

phylogenetic analysis, motif assessment, and domain architecture. The pI values revealed that all the selected *Bacillus* pullulanase are active within an acidic to a neutral environment (Rahmatabadi et al. 2017).

Bioinformatics assessed the manganese peroxidase protein sequences from various fungal sources for phylogenetic analysis, multiple sequence alignment, domain architecture examination, and motif assessment. Also, the structural and functional analysis of manganese peroxidases was studied. The physiochemical analysis revealed that the amino acids ranged from 341-613, molecular mass from 319-580 kDa, and pI from 3.8-5.39, respectively (Yadav et al. 2017).

The study was performed to understand *Pseudomonas aeruginosa* genetic, structural, and functional properties in collagenase through various computational approaches. Collagenase's physicochemical properties and secondary structure have been investigated. Multiple methods were used to construct and verify the three-dimensional (3D) model of collagenase (Rani and Pooja 2018).

Various bio-computational methods were used to perform an *in silico* secondary structure estimation, homology modelling, and functional analysis of *mesorhizobium spp*. ACC deaminase to investigate physicochemical properties. For 3D modelling of the ACC deaminase enzyme, *M. loti* was chosen as a representative species of the *mesorhizobium* genera (Pramanik et al. 2017).

A study determined the structures of xylanase, salt bridge compositions, phylogenetic tree, functional domain and motifs, secondary structure, molar extinction coefficient, GRAVY, instability index, molecular weight, pI, and *in silico* composition of amino acids. Xylanase *in silico* analysis was carried out on 36 separate bacterial sources. The protein is highly stable and thermostable, indicated by the instability index and the aliphatic index values (Dutta et al. 2018).

The computational characterization of phytase protein sequences from *Klebsiella spp*. and analyzed the physical parameters, secondary and tertiary structure, functional properties, and phylogenetic relationship using various bioinformatics tools.

Klebsiella phytases were found to be alkalinity, thermostable, histidine phosphatase superfamily and molecular weight were 46-47 kDa. The protein structure prediction showed a higher percentage of α-helices and β-sheets (Pramanik et al. 2018).

An attempt was made to study the pectin lyase and pectinase sequences from various source organisms using various bio-computational tools and investigated from the view of homology search, multiple sequence alignment, and domain architecture phylogenetic tree construction and motif assessment. Alignment of various sequences showed a conserved region, signifying sequence homology and phylogenetic tree illustrating sequence similarity (Yadav et al. 2009).

The computational characterization of phytase protein sequences from *Enterobacter spp.* analyzed the physical parameters, secondary and tertiary structure, functional properties, and phylogenetic relationship using various bioinformatics tools. *Enterobacter* phytases were found to be acidic, thermostable, and the molecular weight was 48 kDa. The protein structure prediction showed the highest alpha-helical content, and it was a tetrameric protein. The conserved regions included "DG-DP-LG" in the *Enterobacter* phytases (Pramanik et al. 2018).

The study has assessed physicochemical characteristics, motif search, prediction of transmembrane region, three-dimensional structure analysis of histamine receptors with bioinformatics tools. It was found that histamine receptors were 55 kDa molecular weight, the aliphatic index was greater than 90, the instability index was 34.93-47.00, theoretical pI was 9.33-9.62, and all of these except histamine H1 receptors were found to be hydrophobic (Zobayer and Hossain 2018).

Computational analysis is used to understand better the structural, functional, physicochemical properties and the phylogenetic relationship of lipase protein from *Pseudomonas aeruginosa,* which is essential for lipase application in various fields. The primary, secondary and tertiary structures of lipase were determined. Lipase was a molecular weight range from 32 to 34 kDa and a thermostable protein, monomeric and acidic nature. The secondary structure of the protein was densely packed with random coils and alpha helices (Pramanik et al. 2018).

Several bacterial and fungal laccases were characterized for grand average hydropathicity, aliphatic index, instability index, extinction coefficient, and isoelectric point by bioinformatics approach. Also, information about the secondary structure, disulfide bridges, and cellular location. These selected bacterial and fungal laccases were analyzed individually and concurrently. Through homology modeling, structure prediction of laccases is also explored. The hydrophilic and acidic nature of all laccases of selected organisms. Bacterial laccases were found to be intracellular, while fungal laccases were found to be extracellular (Tamboli et al. 2015).

Computational analysis of asparaginase protein sequences from plants, fungi, and bacteria was explored. These sequences were subjected to multiple sequence alignment, domain identification, discovering individual amino acid composition, and phylogenetic tree construction. Compared to all other amino acids found in asparaginase, the amino acid alanine has a high average frequency of 10.77 percent in all selected sources (Dwivedi and Mishra 2014).

In silico analysis of protein and gene sequences of alkaline phosphatase were performed. Also, *Pseudomonas aeruginosa* phosphatase was chosen for physicochemical characteristics, phylogenetic relationship, structural and functional properties, and three-dimensional protein modeling. The protein was alkaline, thermostable, metalloenzyme superfamily, and molecular weight about 51 kDa (Pramanik et al. 2017).

Computational analysis of pectate lyase protein sequences from plants, fungi, bacteria, and nematodes was performed. Different source organisms were indicating various clusters in constructed phylogenetic tree. A sequence-level similarity was observed in multiple sequence alignment and uniformly observed pectate lyase C domain which was revealed by motif analysis in all sources of pectate lyase (Dubey et al. 2010).

Xanthine dehydrogenase protein sequences from fungi, bacteria, and animals were explored computationally. These sequences were subjected to multiple sequence alignment, domain identification, discovering individual amino acid composition, and

phylogenetic tree construction. Compared to all other amino acids found in xanthine dehydrogenase, the amino acid alanine has a high average frequency of 9.24 % percent in all selected sources. Two domains were conserved in all the selected sources (Dwivedi et al. 2013).

Computational analysis of polyphenol oxidase protein sequences from plants, fungi, and bacteria was explored. Multiple sequence alignment, domain identification, individual amino acid composition discovery, and phylogenetic tree construction were performed. Phylogenetic analysis was used to construct two major sequence clusters (Malviya et al. 2011).

Glutaminase protein sequences from various species of *Bacillus* and *Escherichia* were analyzed computationally. These sequences were characterized for multiple sequence alignment, superfamily search, phylogenetic relationship, physicochemical properties, and homology search. Various glutaminase enzyme groups showed the sequence level homology. Two main clusters were observed in the constructed phylogenetic tree for the glutaminases (Irajie et al. 2016).

β-galactosidase protein sequences from plants, fungi, and bacteria were computationally analyzed. Multiple sequence alignment, domain identification, individual amino acid composition discovery, and phylogenetic tree construction were performed. Three significant clusters were observed in the constructed phylogenetic tree for the β-galactosidase. Seven conserved motifs from various families were analyzed. These motifs demonstrated the evolutionary proximity of molecular species (Bose et al. 2013).

Alkaline protease protein sequences from various *Aspergillus* species were explored through computational tools and characterized for phylogenetic tree construction, domain identification, multiple sequence alignment, motif identification, superfamily search, and homology search. Between amino acid residues 69 and 110 and 130–204, a conserved region was observed. The sequence homology of various groups of alkaline protease enzymes was determined. There are ten common motifs for these proteases that have been observed (Morya et al. 2012).

32

Computational characterization of alpha-amylase protein sequences from plants, fungi, and bacteria was performed. These sequences were subjected to multiple sequence alignment, domain identification, discovering individual amino acid composition, and phylogenetic tree construction. Compared to all other amino acids found in alpha-amylase, the amino acid glycine has a high average frequency of 9.42 percent in all selected sources. Furthermore, nine motifs were also identified unique to their groups (Vivek et al. 2012).

Urate-oxidase bacteria phylogeny was studied based on 16S rRNA and uricase protein gene sequences. Most of the known species' representative and type strains (52 strains) were studied. The neighbor-joining method was used to create the phylogenetic trees, and then each sequence was bootstrapped with 500 replications (Dabbagh et al. 2012).

Bioinformatics tools were used to assess the physicochemical properties, motif search, multiple sequence alignment, and homology search of ascorbate peroxidase protein sequences. The constructed phylogenetic tree revealed distinct clusters based on the source of plant species. Various stretches of conserved regions with the highest homology in amino acid residues were found in multiple sequence alignment. The results of the motif analysis revealed a conserved peroxidase domain which is likely to be involved in both structural and enzymatic activities of all ascorbate peroxidase proteins (Pandey et al. 2011).

2.11 Immunogenicity of therapeutic proteins

Therapeutic proteins are critical components of modern medicine because they allow treating some of the most complex and incurable diseases. Immunogenicity refers to the elicitation of immune responses and is typically assessed using antibodies to therapeutic proteins. Responses such as these are likely to adversely affect the therapeutic effect and lead to the following problems: neutralization of a vital biotherapeutic agent and cross-reactivity with endogenous proteins that are not redundant and hypersensitive reactions. The majority of protein therapeutics can cause adverse immune responses in patients. Numerous patients produce anti-

therapeutic antibodies, impairing the therapeutic protein's protection and efficacy (Sauna et al. 2018).

Due to the immunogenicity of specific therapeutic proteins, adverse immune responses such as the development of anti-drug antibodies (ADA) occur, compromising drug effectiveness and patient safety. Therapeutic proteins immunogenicity is patient- and product-dependent. The therapeutic protein has a vital role in influencing immune responses. Non-human proteins produce more extended immune reactions than human proteins (Yari et al. 2017). Another strategy for reducing immunogenicity is identifying and excluding B cell or T cell epitopes in a protein. Numerous experiments, aided by bioinformatics techniques, have been conducted to determine, reduce, or exclude B-cell and T-cell epitopes from immunogenic proteins. The most common antigenic families are conformational B cell epitopes. There have been reports of some success in reducing the immunogenicity of the enzyme through random surface residual replacement. Mutation of particular large amino acids, including lysine, arginine, or glycine, can improve results to small ones such as alanine or glutamate (Yari et al. 2017). Due to advancements in bioinformatics, immunogenicity predictions can now be assisted by structure analysis and molecular dynamics simulations, molecular modelling, and docking studies due to improvements in the bioinformatics field. For example, there are currently many immunoinformatics databases and methods that are extremely useful for predicting immunogenicity and allergenicity in a variety of studies (Belén et al. 2019).

Bioinformatic tools allow for the development of protein engineering designs while using less labor and time. Nowadays, *in silico* approaches, identifying a wide variety of B or T cell epitopes can produce less immunogenic proteins or design vaccines. Ever since computer prediction techniques have been focused on one parameter like amino acid propensity, machine-learning approaches are increasingly evolving to improve prediction accuracy. However, few discontinuous B-cell epitope prediction methods are available (Yari et al. 2017). Bioinformatics provides numerous algorithms for predicting amino acid residues that elicit an immune response.

34

Although most B-cell epitopes are conformational or fold-dependent, Both continuous and discontinuous epitopes are detectable by B-cells, which are unique to the human protein. The detection of conformational B-cell epitopes, followed by the determination of hot spot epitopic residues, is the first step in eliminating B-cell epitopes. Modifications to these residues have the potential to alter their antigenicity. In comparison to time-consuming and expensive experimental methods, a range of publicly accessible computational tools is highly recommended for predicting B-cell epitopes to reduce or eliminate expense (Evander Emeltan Tjoa et al. 2018). Therefore, the removal of epitopes by protein engineering is known to be a fundamental solution in which antigenic motifs of the therapeutic protein are modified by site-directed mutagenesis process to reduce the antigenicity of the protein drug (JevsÌŒevar et al. 2010; Sherman et al. 2008; Zarei et al. 2018).

Numerous factors may affect the immunogenicity of therapeutic proteins, such as the length of duration of treatment, dose frequency, administration route, conjugates or fragments, type of protein, type of disease, and patient's patient genetic history. Numerous strategies for modifying therapeutic proteins to decrease their immunogenicity have been proposed, such as the humanization of monoclonal antibodies, exon shuffling, site-specific mutagenesis, and PEGylation.

PEGylation

PEGylation is the covalent bonding of polyethylene glycol (PEG) to the lysine molecules on the surface of a protein. PEG conjugation masks the protein's surface and increases the molecular size of the polypeptide, thus reducing its renal ultrafiltration, preventing the approach of antibodies or antigen processing cells, and reducing degradation by proteolytic enzymes. PEG is often used in manufacturing because it is nonimmunogenic, nontoxic, and insert. Compared to the native protein, pegylated proteins have a higher half-life, lower immunogenicity, and better biological activity. To treat hepatitis C, PEGylated (PEG)-IFN-alpha2a and PEG-IFN-alpha2b have been developed (Schellekens 2002).

The most common approach in reducing the antigenicity of bacteria-derived biopharmaceuticals was PEGylation. However, polyethylene glycol (PEG) coating of therapeutic protein reduces the efficiency by increasing the protein's size and water absorption properties. In addition, PEGylation can cause PEG-specific antibodies to be secreted in the human body (JevsÌŒevar et al. 2010).

Site-specific mutagenesis

Identification and removal of T-cell epitopes can lead to the development and reduction of immunogenicity of a designer molecule. This has been attempted in staphylokinase through site-specific mutagenesis, in which specific amino acids have been replaced with alanine (Schellekens 2002).

Exon shuffling

Exon shuffling is another technique for choosing and redesigning therapeutic proteins. Directed protein evolution by altering the location of protein-coding regions in the gene to select for better properties. Human protein shuffling can produce completely human antibody libraries that don't have point mutations that could elicit immunogenicity. But therapeutic proteins modified in this way have yet to be tested in clinical trials (Schellekens 2002).

Humanization of monoclonal antibodies

Humanization is a term that refers to the method of adapting monoclonal antibodies, which are often referred to as therapeutic proteins. The synthesis of human anti murine antibody response is currently a problem with murine monoclonal antibodies presently used in clinical applications. Monoclonal antibody immunogenicity can be reduced by substituting non-humans regions with human sequences (Schellekens 2002).

Here, research articles of different enzymes used for reducing immunogenicity through computational approaches were explained.

Mycoplasma hominis arginine deiminase (MhADI) was computationally analyzed to recognize and localize its immunoreactive regions. The three-dimensional structure of MhADI's bioactive shape was modelled. Epitope mapping of B-cells has been carried out using different servers with various algorithms. To minimize immune reactivity, the epitopic hot spot was modified. A high hydrophilicity score, flexibility, surface accessibility, convexity index, and B-cell epitope were used to pick the hot spot residue. Molecular dynamics simulations were used to test the structural stability of native and mutant proteins. The E304L mutein was proposed as a less antigenic and more stable derivative of the enzyme (Zarei et al. 2018).

In order to reduce the immunogenicity of *Erwinia chrysanthemi* L-Asparaginase, three different computational methods were used for predicting conformational B cell epitopes from their three-dimensional structure. Some residues were defined by point mutation as candidates for decreasing protein immunogenicity. The computational stability, binding energy, and hydrophobicity of mutants were analyzed along with immunogenicity. A molecular dynamics simulation was used to determine the stability of the best mutant. Mutant H240A, as well as Q239A, were found to be immunosuppressive. Molecular dynamics simulations demonstrated that the H240A mutation did not affect the target protein flexibility, stability, and overall structure (Yari et al. 2019).

Botulinum toxin serotype A initially screened for B-cell epitope residues which are linear and conformational. Seven residues were allowed to be mutated by overlapping the B-cell epitopes with the conserved sequence that was excluded. Two proposed muteins demonstrated a decrease in the probability of antigenicity. The hot-spot residue antigenicity score was much lower in 1079-1092. The results from the molecular dynamics simulation suggested that the flexibility of both proteins is greater than the protein structure alone. The antigenicity in both muteins is lower. Moreover, they are comparable to the native proteins in structure, stability, and functionality (Evander Emeltan Tjoa et al. 2018).

To improve understanding of slightly known aspects of the asparaginase immunogenicity, immunologic analysis of *Erichia carotora* and *Escherichia coli* was

used. In this regard, the structure of the asparaginases for immunogenic and allergic epitopes was predicted, using the relative frequency of the eight alleles mostly distributed worldwide. This study established that there are no discernible variations in the immunogenicity of the two enzymes, whereas *E. coli* asparaginase had a higher relative frequency of allergenic epitopes. These findings corroborate previously published research (Belén et al. 2019).

To keep the potency of the recombinant staphylokinase, site-directed mutagenesis was undertaken but decreased its antigenicity. K135R, K130T, D82A, E80A, K74Q, or K74R, E65D, and K35A variants exhibited increased enzymatic activity or decreased antibody binding to human staphylokinase. Eight variants with intact thrombolytic activity were identified through additive mutagenesis that absorbed less than a third of staphylokinase-specific antibodies (Collen et al. 1997).

The study demonstrated that site-directed mutagenesis decreased the antigenicity of *Erwinia chrysanthemi* L- asparaginase. The polyclonal antisera and synthetic hexapeptides from mice and rabbits were used to identify ten B-cell epitopes. The immunodominant epitope was in the region [282]GIVPPDEELPG[292], near C-terminus. Pro285 and Pro286 were required to bind two hexapeptides ([283]1VPPDE[288] and [287]DEELPG[292]) to antibodies since their substitution with nearly every other amino acid resulted in decreased binding. The remaining residues were less essential for antibody binding, as amino acid substitutions had little effect on binding. Three mutant enzymes were successfully generated in *E. coli* by site-directed mutagenesis: P285T, P286Q, and E286. The substitution of proline to threonine significantly decreased the enzyme's antigenicity against the wild-type enzyme (Moola et al. 1994).

To decrease the immunogenicity of L- Asparaginase from *Pectobacterium carotovorum* and *Escherichia coli, in silico* approaches were used. B-cell and T-cell epitopes were identified and reduced the immunogenicity score by 50 percent when mutated at the epitopic sites. The enzyme models have been developed and docked to the L-Asparagine substrate to assess its clinical effectiveness and have been shown to be equally effective in catalytic functioning. Additionally, molecular dynamics simulations using Gromacs were conducted for the models, which revealed that they

are stable in terms of structure and activity. K314C and D78L mutations in *E.coli* were found to decrease immunogenicity, while E231V and D145L mutations in *Pectobacterium* were found to reduce immunogenicity (Ramya and Pulicherla 2015).

The study aimed to recognize and locate immune-reactive regions of horseradish peroxidase homolog from *Lepidium draba* using computational analysis. Additionally, a variant sequence with reduced immunogenicity and improved stability was proposed. The enzyme's tertiary structure was expected. The study also discusses the functional and structural significance of the residues and the conservatory value of each one. Various software programs were used to predict the immune-dominant regions of proteins. The final four residues in the C-terminal region were expected to be the immunogenic consensus segments of *Lepidium draba* peroxidase. Changes to wild-type sequences have been applied to alleviate their immune-reactiveness. The new enzyme derivative was likely less immunogenic and more stable (Fattahian et al. 2017).

2.12 Applications of uricase

During the last few decades, the uricase enzyme has gained huge commercial importance due to its wide applications in chemical, medical, clinical chemistry, biotechnology, pharmaceutical, and many other industries. In recent years, uricase enzyme has been discovered to apply in therapeutic and diagnostic applications potentially.

Patients suffering from acute hyperuricemia and gout usually undergo chemotherapy, but recently, PEG-uricase was developed as a protein drug for the treatment (Rasburicase). A PEGylated form of uricase (poly (ethylene glycol) conjugates) is under clinical development for the treatment of chronic hyperuricemia in patients with "treatment-failure gout." Uricase enzyme synthesized from the *Aspergillus flavus* has been therapeutically for treating patients with hyperuricemia (Sherman et al. 2008).

Due to the activity of the uricase enzyme, it is used as a biosensor along with many other peroxidases, glucose oxidase, and catalase that were developed. These biosensors would help in the quantification of the uricase levels. Uricase enzyme has

been used for the symptoms of Lesch-Nyhan syndrome and has proven to be the enzyme capable of protecting neurological cells. It is also used as a peroxisomal marker and is potentially a sound system for studying protein sorting into peroxisomes (Khade and Srivastava 2015).

Uricase as a polymer-conjugated therapeutic agent for immunological studies. Polymer conjugation has been successfully used to enhance the therapeutic potential of many pharmacologically active proteins and peptides. It allows for altering their physicochemical and biological properties improving permanence in circulation, stability, solubility, and reducing immunogenicity. Poly(N-vinylpyrrolidone), poly(N-acryloilmorpholine) linear, and branched PEG can improve the immunogenic character of uricase. Immunological studies showed that antigenicity and immunogenicity of uricase were altered by polymer conjugation to such an extent that depended upon the polymer composition (Nyborg et al. 2016).

When the uricase enzyme reacts with a high concentration of H_2O_2, it causes severe damage to hair and skin due to the strong oxidizing activity of the H_2O_2. But, compared to the direct application of H_2O_2 to the dyeing of hair is more toxic, and a combination of the uricase and the H_2O_2 is milder and preferably better. Uricase also acts as a catalyst during the p-phenylenediamine oxidation used for hair –dyes (oxidative polymerization of monomeric precursors) (El-Naggar et al., 2019).

The biosensor can be developed to detect uric acid qualitatively as well as quantitatively by immobilization of uricase either in the matrix, e.g., polypyrrole, or on gold/amino acid nano-composites, or chitosan graft polyaniline composite film or is covalently immobilized with the help of glutaraldehyde as a cross-linker onto electrochemically synthesized polyaniline films. Due to the covalent attachment, the efficiency and half-life of the uricase are increased. The affinity of the uricase towards the substrate uric acid is increased from 3.4 x 10-1 mM/L to 5.1 x 10-3 mM/L. The number of electrons generated is directly proportional to the substrate uric acid conversion by uricase. The sensor is capable of detecting uric acid from serum at the rate of 100 serum samples in 94 min (Khade and Srivastava 2015).

The uricase enzyme acts against cancer by scavenging the serum-free radicals; this proves that though uricase is sensitive to the temperature, pH, and the source, it acts as an antioxidant (Dabbagh et al. 2012).

2.13 SCOPE AND OBJECTIVES

Characterization of enzymes in wet lab is a long step process, time-consuming, and also expensive compared with the available bioinformatics tools and servers which are more economical time-saving methods. Bioinformatics web-based servers and tools are useful for understanding unknown protein profiles through their sequential, structural, functional, and evolutionary data using computational genomics and proteomics studies. Computational approaches can be used to screen and investigate a uricase enzyme with desirable characteristics that can be employed in diverse industrial applications. The increasing importance of uricase is probably due to its potential use in medicinal chemistry and the treatment of several diseases. There is a pressing need for the cost-effective production of uricase from various sources with high purity. Uricase being an essential clinical enzyme, there is a great demand for highly active and pure forms of uricase.

Minimization of the antigenicity of uricase is necessary to use uricase as a protein drug to cure treatment-resistant gout. Therefore, removing the epitopes via protein engineering is reported to be the fundamental solution in which antigenic motifs of the therapeutic protein are modified by site-directed mutagenesis process for reducing the antigenicity of the protein drug. The experimental evolution of B-cell and T-cell epitopes of therapeutic proteins are limited because most of the approaches are expensive, time-consuming, and laborious (Potocnakova et al. 2016). Therefore, the widely accepted algorithms and tools of bioinformatics are highly recommended, which can reduce cost by predicting B-cell and T-cell epitopes from the amino acid sequence of uricase.

Polymer conjugation is increasing interest in pharmaceutical chemistry for delivering drugs of simple structure or complex compounds such as peptides, enzymes, and oligonucleotides. However, polypeptides and protein conjugation research are

especially active because new substances created by genetic engineering are being made in mass quantities, and in addition, patients find these drugs difficult to administer due to a number of inherent disadvantages. At the moment, the most widely used conjugation technology is PEGylation. Although it represents a historical breakthrough in pharmaceutical technology, this strategy shows several inherent limitations. So there is a need to find an alternative polymer that can be used for conjugation to improve the pharmaceutical properties of uricase.

Taking the above facts into account, the purpose of this work is to develop an enzymatic drug called uricase for the treatment of hyperuricemia.

The objectives for the present study are:

1. *In silico* characterization of amino acid sequences of uricase for understanding the conservation of amino acids, motifs, and identification of evolutionary relations, amino acid composition from different source organisms.

2. *In silico* based structural, functional, and phylogenetic analyses of uricase enzymes from various *Bacillus* species.

3. A computational approach to identify and mutate epitopic regions of uricase using *Arthrobacter globiformis* and *Bacillus fastidious* as model enzymes for reducing immunogenicity.

4. Studies of molecular dynamics for the stability analysis of the modelled enzyme.

5. Synthesis, purification, the effect of important reaction parameters (uricase to BSA ratio, and effect of crosslinking agent), and biochemical characterization of bovine serum albumin conjugates of uricase from *Bacillus fastidious.*

6. Studies for determining the stability of conjugates at various pH and temperature and kinetic parameters to determine Km and Vmax values.

2.14 Thesis outline

The current thesis is divided into seven chapters, namely

CHAPTER 1: Introduction

CHAPTER 2: Review of literature

CHAPTER 3: Computational analysis of therapeutic enzyme uricase from different microbial sources

This chapter is elucidating the structure and physicochemical properties of uricase by *in-silico* analysis. It also summarizes different studies like multiple sequence alignment (MSA), homology search, phylogenetic relation, motif search, domain architecture, and physiochemical properties including pI, EC, Ai, Ii, GRAVY of uricases from bacteria, fungi, yeast, plants, and animals' sources.

CHAPTER 4: *In silico* **structural and functional analysis of** *Bacillus* **uricases**

This chapter deals with computational-based structural, functional, and phylogenetic analyses of uricase enzymes from various *Bacillus* species. Additionally, it contains details about various analysis, such as multiple sequence alignment, phylogenetic analysis, motif assessment, domain architecture examination, understanding of basic physicochemical properties, and *in silico* identification of amino acid composition in uricase.

CHAPTER 5: *In-silico* **epitope identification and design of uricase mutein with reduced immunogenicity**

This chapter details the identification of the linear B-cell, conformational B-cell, and MHC-I-based T-cell epitopes to reduce the immunogenicity of uricase sourced from *Arthrobacter globiformis* and *Bacillus fastidious*. This section also studies identifying motifs and domains of uricase from various sources to describe this protein's structural, functional aspects in the evolutionary process. This section, in addition to the identification and mutation of hot-spot residues to reduce the antigenic character,

finally, studying the impact of mutagenesis on the catalytic activity and the structural stability of uricase by molecular docking and molecular dynamics simulation.

CHAPTER 6: Bio-conjugation of therapeutic enzyme uricase with BSA: An experimental investigation

This chapter elucidates the conjugation of bacterial uricase (*Bacillus fastidious)* with bovine serum albumin to improve its therapeutic property. This section also deals with the BSA concentration, glutaraldehyde concentration (cross-linker), pH, and temperature for optimization to achieve the desired degree of conjugation with desired residual activity. It also summarizes the details of conjugate's stability with respect to temperature and pH.

CHAPTER 7: Summary and Conclusion

This section reports a concise review of the research presented. The research work findings on computational studies on uricase from different sources, *Bacillus* species, immunogenicity reduction, and experimental study on chemical modification with bovine serum albumin, along with a few suggestions for future work were discussed.

CHAPTER 3

COMPUTATIONAL ANALYSIS OF THERAPEUTIC ENZYME URICASE FROM DIFFERENT MICROBIAL SOURCES

Enzyme research has developed over the years due to advancements in techniques such as directed evolution, metagenomics, and bioinformatics. The value of bioinformatics is being realized in the era of genomics, assisting in the identification and characterization of putative gene families of enzymes for varying industrial purposes. The available protein sequences for enzymes are being used to figure out the sequence-structure-function relationship. Prior to wet-lab experimentation, several studies on computer-based computational tools of enzyme design had been reported (Yadav et al. 2017). The main emphasis of *in silico* analysis of genes and proteins to find suitable biomarkers is to identify the pathogenic genera, design of drugs to combat the pathogenic microbes and superbugs, and diagnose infectious diseases (Pramanik et al. 2017). Nowadays, the utilization of *in silico* analysis and characterization of various industrially important enzymes using their protein sequences has gained importance (Nezafat et al. 2015). Currently, the available bioinformatical tools may help researchers in different areas start their experiments in certain projections instead of several expensive and lengthy practical steps (Rahmatabadi et al. 2017).

In this study, an attempt was made to analyze the individual amino acid sequences of uricase from four different sources, namely bacteria, fungi, plant and animals, to elucidate uricase structure and physiochemical properties by several standard biocomputational tools. The complete computational exploration of all the selected protein sequences is aimed to find out the multiple sequence alignment, homology search, evolutionary relationships, motif search, physicochemical properties to compare all the protein sequences under common ground.

3.1 MATERIALS AND METHOD

3.1.1 Uricase enzyme sequence retrieval

The sequences of uricase from different sources like bacteria, fungi, plant, and animal were searched and retrieved from the National Centre for Biotechnology Information (NCBI) protein database (http://www.ncbi.nlm.nih.gov/protein) for *in silico* analysis. Fifteen protein sequences from each source and a total of sixty full-length FASTA formats of all the selected protein sequences were downloaded. Swiss-Prot or UniProtKB server (Apweiler et al. 2004) and ExPASy server (Artimo et al. 2012) are also used to retrieve protein sequences. For the computational investigation, these selected amino acid sequences were categorized into different categories like bacterial, fungal, plant, and animals.

3.1.2 Multiple sequence alignment

Four different sources of protein sequences were analyzed in multiple sequence alignment by the ClustalW tool available in MEGA7 software (version 7.0) (Kumar et al. 2016). Amino acid sequences of uricase from the selected category were executed to know about conserved amino acid residues and find the sequence-based similarity in the protein sequences. The considered parameters in multiple sequence alignment include a gap-opening penalty of 10, gap extension penalty of 0.2, protein weight matrix of gonnect, gap separation distance of 5, and no end gap separation. The programs of Seaview and Clustal2x (Larkin et al. 2007) were also used for multiple sequence alignment.

3.1.3 Phylogenetic analysis

Phylogenetic analysis is focused on evolutionary relationships among uricase sequences. From the above ClustalW sequence alignment, the molecular evolutionary genetic analysis software of MEGA7 (Kumar et al. 2016) was used to construct four different phylogenetic trees based on retrieved amino acid sequences. Here, Neighbor-joining (NJ) statistical method (Saitou and Nei 1987) was employed based on the p-distance model to determine the evolutionary history of uricase. 1000 bootstrap replications were used for each sequence to evaluate the topologies of phylogenetic

trees.

3.1.4 Motif identification

All the retrieved sequences of uricase were achieved as individual categories for conserved motif identification. MEME (Multiple EM for Motif Elicitation) server (version 5.0.2) (http://meme-suite.org/tools/meme) was employed based on the expectation-maximization approach for discovering motifs existing in the protein sequences (Bailey et al. 2009). The parameters considered here include the maximum number of motifs and motif width. The starting and ending point of the motifs were shown as blocks. Motif locations or sites, were shown in the MEME suite, which explains about conserved amino acid regions which are associated with some biological function of the enzyme. The MOTIF search was also used for motif identification.

3.1.5 Motif family identification

Based on the MEME suite identified, sequence motifs were employed to find their protein families using the Pfam database of version 31.0 (Bailey and Gribskov 1998). Pfam is a widely used protein family database that gives complete domains of uricase. The InterPro (http://www.ebi.ac.uk/interpro) was also used for the domain search. All the motif sequences were submitted as individuals to the sequence search option, available in the database for the domain organization associated with uricase protein. Also, the domain analysis of all retrieved sequences was performed.

3.1.6 Physicochemical characterization and amino acid composition

The physicochemical characteristics of retrieved 60 amino acid sequences of uricase were predicted by the ProtParam tool (https://web.expasy.org/protparam/). TheExPASy ProtParam is an *in silico* tool, which predicts the amino acid composition, various physical and chemical features of the protein. Individual amino acid sequences were submitted to the server. The calculated parameters include number of amino acids, molecular weight (MW), theoretical isoelectric point (pI), amino acid composition, the total number of negatively charged residues (Asp+Glu), the total number of positively charged residues (Arg+Lys), extinction coefficient (EC), instability index (Ii), aliphatic index (Ai) and grand average of hydropathicity

(GRAVY) (Gasteiger et al. 2005). Sequence Manipulation Suite (SMS) Version 2 was also used for the determination of protein theoretical isoelectric points (pI), molecular weights, and GRAVY (http://www.bioinformatics.org/sms2/). The ProtParam server by the Bjellqvist method was also used to calculate the theoretical pI of the protein (Bjellqvist et al. 1993, 1994). The individual amino acid composition was also discovered by Mega software (Kumar et al. 2016).

3.2 RESULTS AND DISCUSSION

The assessment of amino acid sequence based on computational analysis of various enzymes like alpha-amylase (Vivek et al. 2012), β-propellerphytase (Mathew et al. 2014), β-galactosidases (Bose et al. 2013), fructosyl transferase (Alméciga-Díaz et al. 2011), glutaminase (Irajie et al. 2016), histidine acid phytase (Kumar et al. 2012), L-asparaginase (Dwivedi and Mishra 2014), manganese peroxidase (Yadav et al. 2017), phytases (Verma et al. 2016) have been reported in the literature based on bioinformatics. In the present study, an attempt to assess the amino acid sequences of therapeutically essential enzyme uricase from different sources by employing different bioinformatics tools was made. The four different sources of uricase protein sequences were retrieved from NCBI. The number of amino acids are found to be in the range of 300-338, and several types of *in silico* analysis were carried out by using genuine database namely Swiss-Prot/UniProtKB, NCBI, from which the main information for further analysis is gathered (Irajie et al. 2016; Morya et al. 2016; Rahmatabadi et al. 2017; Yadav et al. 2009).

3.2.1 Multiple sequence alignment

The source of uricase selected for sequences along with accession number and the species name is mentioned in Table 3.1. The retrieved sequences of bacteria, fungi, plant, and animals sources of 60 uricase were subjected to multiple sequence alignment and homology search in ClustalW, and it has been observed that most of the amino acids are highly conserved. The multiple sequence alignment (MSA) of the selected bacterial uricase showed maximum conservation between 16 and 337 amino acids; conservation between 14 and 338 for fungal uricase; conservation between 12 and 317 for plant uricase, and 37-304 for animal uricase. The complete sequences of

all the uricase were found to be in between 51-314 amino acid residues. These conserved regions are the critical amino acid residues, and the positions of these conserved amino acids are functionally crucial places of a protein. The box shade server was used to represent the conserved amino acids in bacteria, fungi, plant, animal sourced sequences of uricase (Appendix-I).

The bacterial sequences have two glycine, one arginine, one threonine, and one glutamine conserved residues. Similarly, the fungal sequences have conserved six serine, five leucine, four threonine, three aspartic acid, three phenylalanine, three glycine, three histidine, three lysine, three asparagine, three proline, three tyrosine, two glutamic acid, two arginine, two valine, one alanine, one isoleucine, one glutamine, one methionine, and one tryptophan residues. The plant sequences have conserved fourteen valine, ten glycine, ten threonine, nine lysine, eight phenylalanine, eight glutamic acid, seven alanine, six leucine, five histidine, five serine, five arginine, five tyrosine, four aspartic acid, four asparagine, three tryptophan, three proline, two histidine, one cysteine, one isoleucine, one methionine, and one glutamine residues.

The animal sequences have conserved seven glycines, seven leucine, seven phenylalanine, four histidines, four lysines, four proline, four arginine, four serine, four threonine, four tyrosine, two aspartic acid, three glutamic acids, three asparagine, two glutamine, two valine, two tryptophan, and one alanine amino acid residues. Several identical conserved amino acids exist in fungi, plants, and animals compared to bacterial protein sequences. All the sequences of bacteria, fungus, plants, and animals have one glutamine, which was identically conserved in selected sources of uricase. The overall analysis of multiple sequence analysis shows that these amino acids play a role in the functional activity of the uricase protein. These results were corroborated using previous works done by several workers (Dwivedi et al. 2013; Dwivedi and Mishra 2014; Kumar et al. 2012; Malviya et al. 2011; Morya et al. 2012).

3.2.2 Phylogenetic analysis

Phylogenetic trees were constructed to compare the evolutionary relationship based on the uricase protein sequences of different sources of organisms. Multiple sequence alignments were used for the construction of phylogenetic trees based on the neighbor-joining method to know the evolutionary history with MEGA7 software (Kumar et al. 2016). This study mainly compares the evolutionary phylogenetic relationship and diversity among the selected 60 protein sequence taxa.

The bacterial phylogenetic tree has three distinct clusters, as shown in Figure 3.1. Cluster I consists of eight organisms, which were separated into two subclusters. Organisms like *Arthrobacter globiformis*, *Microbacterium sp*, *Singulisphaera acidiphila,* and *Microlunatus phosphovorus* were located in subcluster I, while other organisms are located in the other subcluster, namely *Saccharomonaspora cyanea*, *Thermobispora bispora*, *Hoyosella subflava,* and *Rhodococcus fascians,* was found to be closely related to each other. Cluster II consists of *Nakamurella multipartita*, *Streptomyces pratensis*, *Kitasatospora cheerisanensis,* and *Actinobacteria bacterium*. Cluster III consists of *Truepera radiovictrix,* and *Bacillus elenitireducens* appeared in the same cluster, showing sequence-level similarity. *Pseudomonas aeruginosa* was out-grouped and could not be included in any cluster.

The fungal phylogenetic tree has exhibited two distinct clusters, as shown in Figure 3.2. Various organisms, including *Trichoderma reesei, Cordyceps fumosorosea, Fusarium graminearum*, *Pseudogymnoascus sp.* were located in subcluster I. Subcluster II contains *Aspergillus parasiticus* and *Penicillium digitatum*. Subcluster III contains *Paracoccidioides lutzii* and *Blastomyces gilchristii. Aspergillus niger* and *Trametes coccinea* were found to be out-grouped from both the subclusters, and therefore these are distantly related. Cluster II consists of four organisms, namely *Ascoidea rubescens, Cyberlindnera jadinii, Lodderomyces elongisporus,* and *Hyphopichia burtonii. Conidiobolus coronatus* was distinct from both clusters.

Table 3.1: List of retrieved amino acid sequences of uricase from NCBI protein database and their respective accession number

Serial No.	Source	Accession No.	Species Name
1	Bacteria	AGA28823.1	*Singulisphaera acidiphila*
2	Bacteria	EHR61468.1	*Saccharomonospora cyanea*
3	Bacteria	ACV76680.1	*Nakamurella multipartita*
4	Bacteria	ADW02506.1	*Streptomyces pratensis*
5	Bacteria	KDN81792.1	*Kitasatospora cheerisanensis*
6	Bacteria	ADG89294.1	*Thermobispora bispora*
7	Bacteria	AEF42631.1	*Hoyosella subflava*
8	Bacteria	EYU05833.1	*Pseudomonas aeruginosa*
9	Bacteria	ADI14624.1	*Truepera radiovictrix*
10	Bacteria	BAK35228.1	*Microlunatus phosphovorus*
11	Bacteria	GAB16350.1	*Arthrobacter globiformis*
12	Bacteria	AMY51449.1	*Rhodococcus fascians*
13	Bacteria	KPI32983.1	*Actinobacteria bacterium*
14	Bacteria	ADH98118.1	*Bacillus selenitireducens*
15	Bacteria	KJQ52767.1	*Microbacterium sp.*
16	Fungi	KJK61270.1	*Aspergillus parasiticus*
17	Fungi	XP_006963697.1	*Trichoderma reesei*
18	Fungi	XP_001528662.1	*Lodderomyces elongisporus*
19	Fungi	XP_018702053.1	*Cordyceps fumosorosea*
20	Fungi	XP_001390131.1	*Aspergillus niger*
21	Fungi	XP_011321510.1	*Fusarium graminearum*
22	Fungi	KXN72467.1	*Conidiobolus coronatus*
23	Fungi	OSD05528.1	*Trametes coccinea*
24	Fungi	XP_020071435.1	*Cyberlindnera jadinii*
25	Fungi	XP_020075778.1	*Hyphopichia burtonii*
26	Fungi	XP_020049386.1	*Ascoidea rubescens*
27	Fungi	XP_002796429.1	*Paracoccidioides lutzii*
28	Fungi	OAT11654.1	*Blastomyces gilchristii*
29	Fungi	XP_014533620.1	*Penicillium digitatum*
30	Fungi	OBT42089.1	*Pseudogymnoascus sp.*

31	Plant	ABD03945.1	*Sorghum bicolor*
32	Plant	ABD03944.1	*Saccharum officinarum*
33	Plant	XP_007211669.1	*Prunus persica*
34	Plant	XP_020235639.1	*Cajanus cajan*
35	Plant	XP_002315419.2	*Populus trichocarpa*
36	Plant	ABD03946.1	*Triticum aestivum*
37	Plant	XP_022150541.1	*Momordica charantia*
38	Plant	XP_022025248.1	*Helianthus annuus*
39	Plant	NP_001267899.1	*Vitis vinifera*
40	Plant	AAB97726.1	*Phaseolus vulgaris*
41	Plant	BAA13184.1	*Glycine max*
42	Plant	XP_020870005.1	*Arabidopsis lyrata*
43	Plant	BAB18538.1	*Lotus japonicus*
44	Plant	ABD03939.1	*Medicago truncatula*
45	Plant	CAB77205.1	*Cicer arietinum*
46	Animal	NP_446220.1	*Rattus norvegicus*
47	Animal	NP_001121545.1	*Oryctolagus cuniculus*
48	Animal	NP_033500.1	*Mus musculus*
49	Animal	XP_006070940.1	*Bubalus bubalis*
50	Animal	XP_006919420.1	*Pteropus alecto*
51	Animal	XP_006086199.2	*Myotis lucifugus*
52	Animal	XP_015146362.1	*Gallus gallus*
53	Animal	XP_012657176.1	*Otolemur garnettii*
54	Animal	XP_012603035.1	*Microcebus murinus*
55	Animal	XP_021013952.1	*Mus caroli*
56	Animal	XP_025727815.1	*Callorhinus ursinus*
57	Animal	XP_025273620.1	*Canis lupus dingo*
58	Animal	NP_001037382.1	*Bombyx mori*
59	Animal	AFP60128.1	*Musca domestica*
60	Animal	XP_022352809.1	*Enhydra lutris kenyoni*
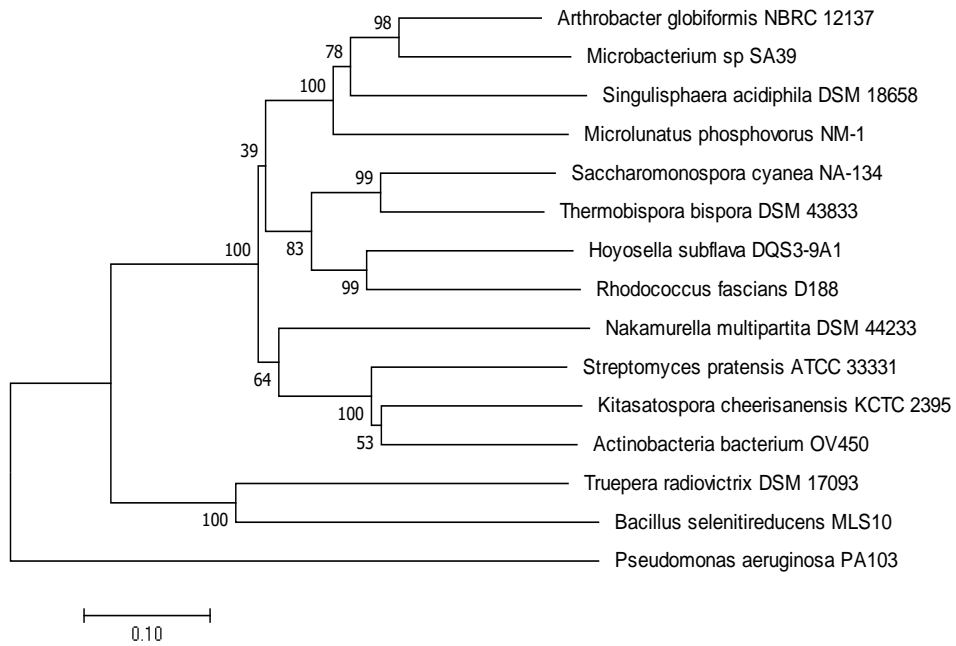
Figure 3.1: Phylogenetic tree of bacterial uricase sequences using neighbor-joining method
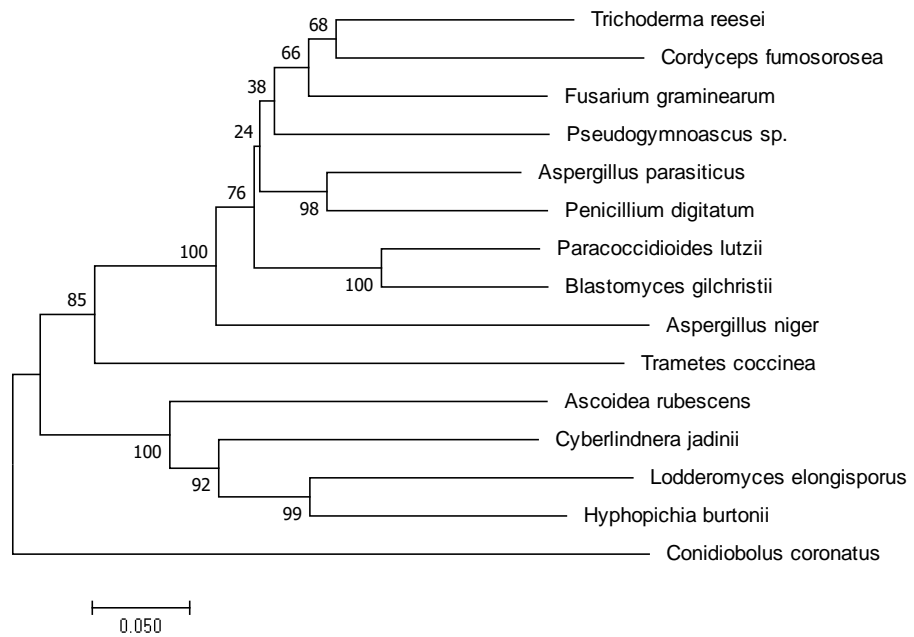
Figure 3.2: Phylogenetic tree of fungal uricase sequences using neighbor-joining method

The plant's phylogenetic tree has two distinct clusters, as shown in Figure 3.3. Various organisms, including *Phaseolus vulgaris, Glycine max, Cajanuscajan, Lotosjaponicus, Medicago truncatula,* and *Cicer arietinum,* were located in cluster I. The organisms in the other cluster II are *Prunus persica* and *Momordica charantia. Populus trichocarpa, Helianthus annuus, Vitis vinifera,* and *Arabidopsis lyrata* were out-grouped and could not be included in any clusters. *Triticum aestivum, Sorghum bicolor,* and *Saccharum officinarum* were found to be closely related to each other but out-grouped from both clusters.
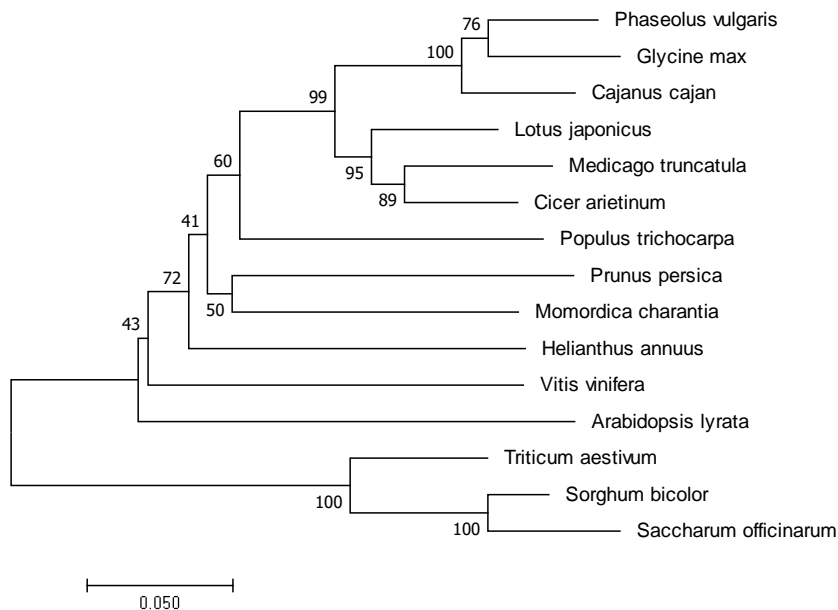
Figure 3.3: Phylogenetic tree of collected plant uricase amino acid sequences using neighbor-joining method

The animal phylogenetic tree has exhibited two distinct clusters, as shown in Figure 3.4. Various organisms, including *Callorhinus ursinus, Enhydra lutris kenyoni, Canis lupus dingo, Pteropus alecto, Myotis lucifugus, Otolemur garnettii,* and *Microcebus murinus,* were located in cluster I, while *Bubalus bubalis, Rattus norvegicus, Mus musculus,* and *Mus caroli* were located in cluster II. *Oryctolagus cuniculus* was found to be a distantly related organism and not included in any cluster. *Bombyx mori* and *Musca domestica* were observed to be closely related to each other but out-grouped from both clusters. *Gallus gallus* was out-grouped and could not be included in any clusters.

54
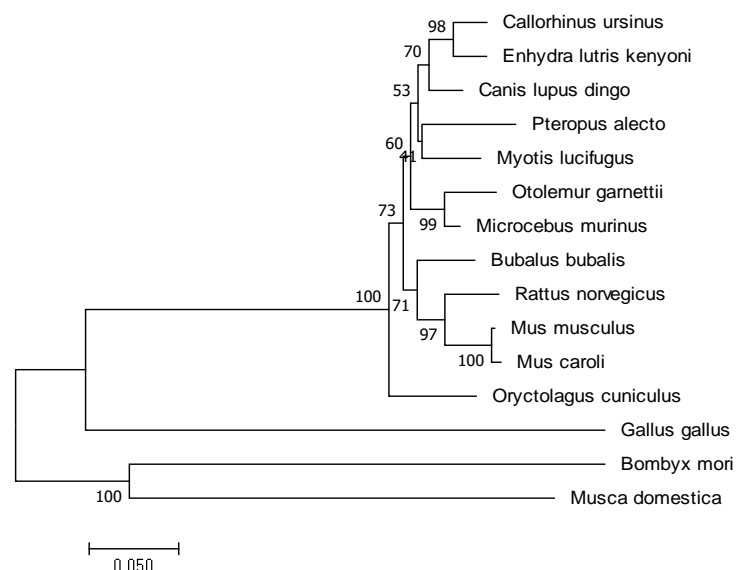
Figure 3.4: Phylogenetic tree of animal uricase sequences using neighbor-joining method

Phylogenetic tree construction of all the selected sources revealed separate clusters of bacteria, fungi, plant, and animal uricase, as shown in Figure 3.5. These distinct clusters show a sequence level similarity of a different source of organisms. Multiple accessions related to bacteria, fungi, plant, and animal uricase were located near the clusters denoting more sequence-level similarity. The organisms *Triticum aestivum, Sorghum bicolor*, and *Saccharum officinarum* showed distinct clusters among plants denoting sequence-level similarity. Also, *Bombyx mori* and *Musca domestica* showed distinct clusters among animals that are similar at the sequence level. *Gallus gallus* was found to be distantly related and therefore out-grouped from the animal cluster. Also, *Conidiobolus coronatus* was found to be distantly related and not included in the fungi cluster. In bacteria, *Truepera radiovictrix* and *Bacillus selenitireducens* showed different clusters denoting sequence-level similarity. *Pseudomonas aeruginosa* was found to be distantly related and therefore out-grouped from the clusters. The evolutionary history was deduced by the neighbor-joining method (Saitou and Nei 1987a). Similar phylogenetic trees were constructed by several authors to see evolutionary relationships among the taxa based on their amino acid sequences (Dwivedi et al. 2013; Dwivedi and Mishra 2014; Irajie et al. 2016; Malviya et al. 2011; Rahmatabadi et al. 2017; Yadav et al. 2017).

55
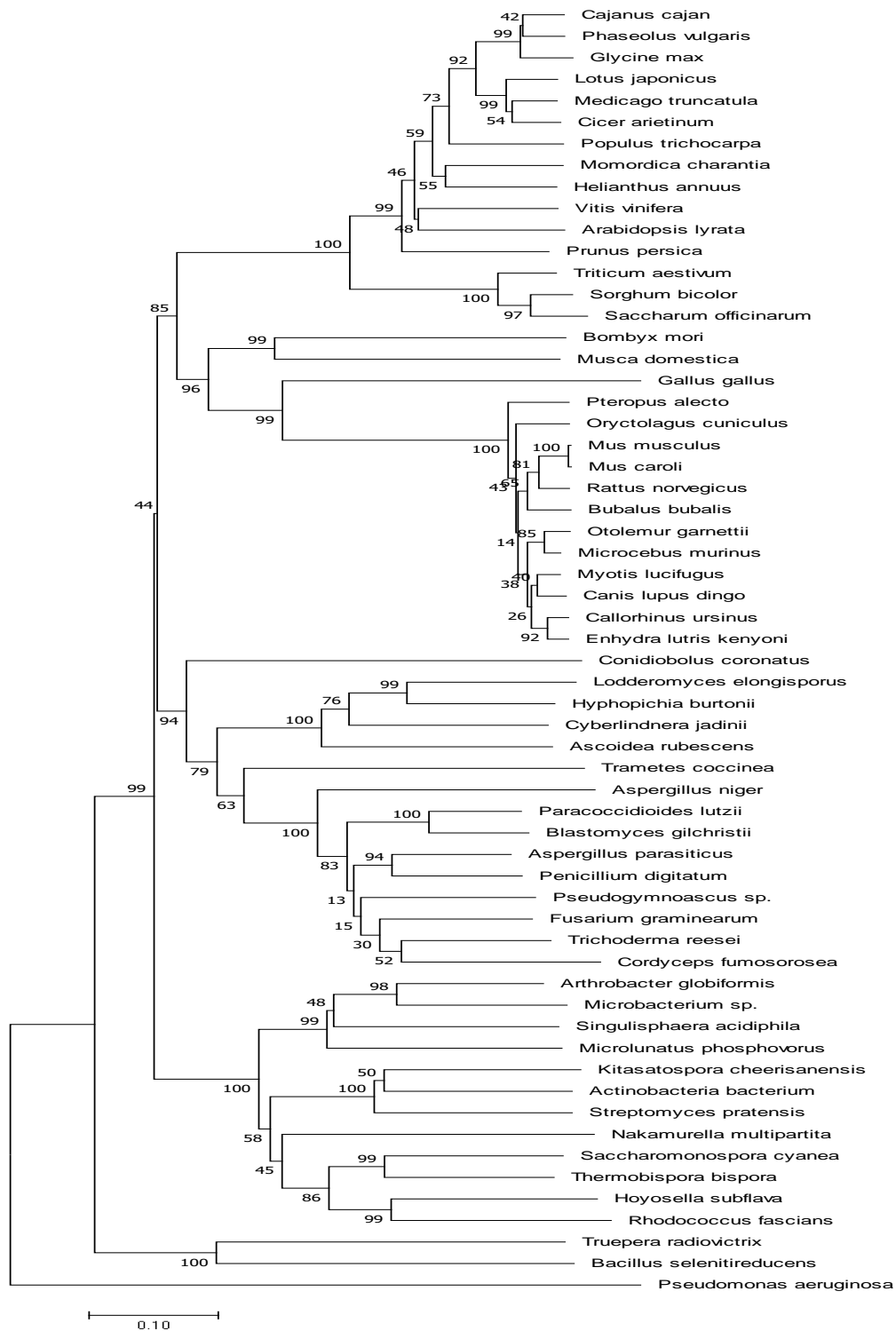
Figure 3.5: Phylogenetic tree of all the uricase amino acid sequences using neighbor-joining method

3.2.3 Motif identification

The MEME algorithm identifies and characterizes the shared motifs in a set of unaligned sequences. The Bayesian probabilistic model is used for finding the motifs for all the sequences and enhances the statistical parameters by Expectation-Maximization (EM) algorithm (Bailey et al. 2009). The motif-based sequence analysis tool (MEME) finds a total of six motifs denoted as 1,2,3,4,5,6 exists in the input protein sequences. Each of the six motifs was selected based on width and number of occurrences to minimize the E-value of the motif. The width of the motif in the range of 6-50 was specified. The distribution of all six motifs (regions) was present almost in uricase protein sequences from different sources, which is clearly mentioned in Table 3.2. Twenty-five motifs were found, and each of them was unique in their groups. The other details about motifs such as motifs width, protein sequence information, and best possible matches were also shown in Table 3.3. The motifs signify a possible role in structural and functional catalytic attributes of uricase. The combined block diagram of MEME-defined motifs in uricase sequences was represented in Figure 3.6. The Pfam database results of domain analysis indicated that the protein sequences from various sources of uricase have a two-domain organization associated with the uricase family. Similarly, motif identification of various proteins was studied and reported by several authors (Dwivedi et al. 2013; Dwivedi and Mishra 2014; Morya et al. 2012; Ramya and Pulicherla 2015).

3.2.4 Physicochemical characterization and amino acid composition

The physicochemical parameters of uricases include the number of amino acids, molecular weight, theoretical pI, number of negative residues, number of positive residues, extinction coefficient, instability index, aliphatic index, and GRAVY determined for various sources of organisms, which are listed in Table 3.4. These parameters were assessed using the Expasy ProtParam tool. The uricase from bacterial species *Bacillus selenitireducens* (ADH98118.1) has the highest molecular weight of 36 kDa, fungi *Trametes coccinea* (OSD05528.1) has 37 kDa, plant *Momordicac harantia*has 35 kDa, animal *Musca domestica* (AFP60128.1) has 39 kDa. The molecular weight of uricase ranged from 33-39 kDa.
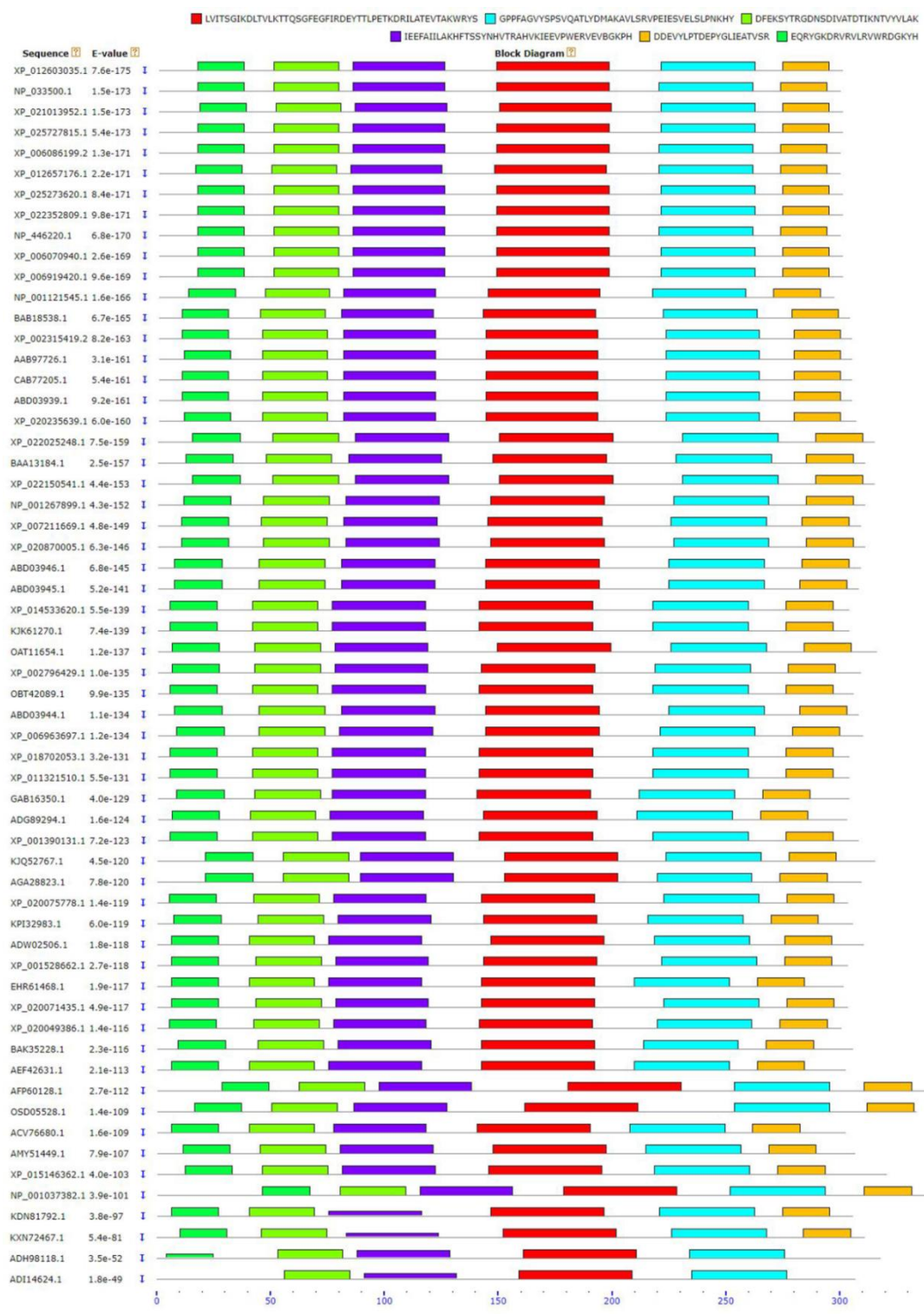
Figure 3.6: Block diagram of six motifs among 60 uricase proteins sequences from different source

Table 3.2: Motif distribution of all selected different source among 60 uricase protein sequences

S.NO	Species	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5	Motif 6
	(A) Bacteria						
1	*Singulisphaer aacidiphila*	+	+	+	+	+	+
2	*Saccharomonospora cyanea*	+	+	+	+	+	+
3	*Nakamurella multipartita*	+	+	+	+	+	+
4	*Streptomyces pratensis*	+	+	+	+	+	+
5	*Kitasatospora cheerisanensis*	+	+	-	+	+	+
6	*Thermobispora bispora*	+	+	+	+	+	+
7	*Hoyosella subflava*	+	+	+	+	+	+
8	*Pseudomonas aeruginosa*	-	-	-	-	-	-
9	*Truepera radiovictrix*	-	+	-	+	-	-
10	*Microlunatus phosphovorus*	+	+	+	+	+	+
11	*Arthrobacter globiformis*	+	+	+	+	+	+
12	*Rhodococcus fascians*	+	+	+	+	+	+
13	*Actinobacteria bacterium*	+	+	+	+	+	+
14	*Bacillus selenitireducens*	-	+	-	+	-	-
15	*Microbacterium sp.*	+	+	+	+	+	+
	(B) Fungi						
16	*Aspergillus parasiticus*	+	+	+	+	+	+
17	*Trichoderma reesei*	+	+	+	+	+	+
18	*Lodderomyces elongisporus*	+	+	+	+	+	+
19	*Cordyceps fumosorosea*	+	+	+	+	+	+
20	*Aspergillus niger*	+	+	+	+	+	+
21	*Fusarium graminearum*	+	+	+	+	+	+
22	*Conidiobolus coronatus*	+	+	-	+	+	+
23	*Trametes coccinea*	+	+	+	+	+	-
24	*Cyberlindnera jadinii*	+	+	+	+	+	+
25	*Hyphopichia burtonii*	+	+	+	+	+	+

26	*Ascoidea rubescens*	+	+	+	+	+	+
27	*Paracoccidioides lutzii*	+	+	+	+	+	+
28	*Blastomyces gilchristii*	+	+	+	+	+	+
29	*Penicillium digitatum*	+	+	+	+	+	+
30	*Pseudo gymnoascus sp.*	+	+	+	+	+	+
	(C) Plant						
31	*Sorghum bicolor*	+	+	+	+	+	+
32	*Saccharum officinarum*	+	+	+	+	+	+
33	*Prunus persica*	+	+	+	+	+	+
34	*Cajanus cajan*	+	+	+	+	+	+
35	*Populus trichocarpa*	+	+	+	+	+	+
36	*Triticum aestivum*	+	+	+	+	+	+
37	*Momordica charantia*	+	+	+	+	+	+
38	*Helianthus annuus*	+	+	+	+	+	+
39	*Vitis vinifera*	+	+	+	+	+	+
40	*Phaseolus vulgaris*	+	+	+	+	+	+
41	*Glycine max*	+	+	+	+	+	+
42	*Arabidopsis lyrata*	+	+	+	+	+	+
43	*Lotus japonicus*	+	+	+	+	+	+
44	*Medicago truncatula*	+	+	+	+	+	+
45	*Cicer arietinum*	+	+	+	+	+	+
	(D) Animal						
46	*Rattus norvegicus*	+	+	+	+	+	+
47	*Oryctolagus cuniculus*	+	+	+	+	+	+
48	*Mus musculus*	+	+	+	+	+	+
49	*Bubalus bubalis*	+	+	+	+	+	+
50	*Pteropus alecto*	+	+	+	+	+	+
51	*Myotis lucifugus*	+	+	+	+	+	+
52	*Gallus gallus*	+	+	+	+	+	-
53	*Otolemur garnettii*	+	+	+	+	+	+

54	*Microcebus murinus*	+	+	+	+	+	+
55	*Mus caroli*	+	+	+	+	+	+
56	*Callorhinus ursinus*	+	+	+	+	+	+
57	*Canis lupus dingo*	+	+	+	+	+	+
58	*Bombyx mori*	+	+	-	+	+	-
59	*Musca domestica*	+	+	-	+	+	-
60	*Enhydra lutris kenyoni*	+	+	+	+	+	+

Table 3.3: Distribution of motifs observed in uricase amino acid sequences of bacteria, fungi, plant and animals along with their pfam analysis

S.No.	Source	Motif width	E-value	Motif present in number of sequences	Motif	Pfam
1	Bacteria	32	1.0e-166	14	FEAAHLEGDNANVLPTDTQKNTVYAFAKEHGG	Uricase family
2	Bacteria	48	3.4e-280	14	HVVSGLKDLTVLKSTGSEFGHFLKDRYTTLEETTDRILATSVTARWRY	Uricase family
3	Bacteria	29	9.7e-157	12	IVLGQNQYGKAENRVVRITRDTDRHEIED	Pfam entry not found
4	Bacteria	50	2.4e-155	11	SPEAFGJRLADHFVSSFEPVDGARIEIEEYAWERIDVDGAEHDHSFVRKG	Uricase family
5	Bacteria	50	1.3e-289	12	HSLALQQTLYAMGKAVLEAHPEIAEIRFSLPNKHHFLVDLEPFGLENPNE	Uricase family
6	Bacteria	21	8.3e-107	13	VFYAADRPYGLIEATVLRDDV	Pfam entry not found
7	Fungi	21	2.2e-121	15	ARYGKDNVRVLKVHRDEKTGV	Pfam entry not found
9	Fungi	29	6.9e-225	15	LLEGDIETSYTKADNSVVVATDSIKNTIY	Uricase family
10	Fungi	50	5.8e-386	15	JTSSIKGLTVLKSTGSQFHGFVRDEYTTLPETWDRILSTDVDASWKWKNF	Uricase family
11	Fungi	45	3.3e-333	15	FAEDNSASVQATMYKMAEQILAAVPLVETVEYSLPNKHYFEIDLS	Uricase family
12	Fungi	41	5.4e-228	14	QNPVTPPELFASILGTHFIEKYKHIHAAHVDIITHRWTRMT	Pfam entry not found

61

13	Fungi	27	1.4e-183	14	LKNTGKDAEVYAPQSGPNGLIKCTVGR	Pfam entry not found
14	Plant	29	7.9e-212	15	EGFKFEQRHGKERVRARVWRSKDGRHFF	Pfam entry not found
15	Plant	50	8.8e-522	15	DCVNSYVRDDNSDIVATDTMKNTVYAKAKEC SEILSVEEFAILLAKHFTS	Uricase family
16	Plant	50	1.5e-481	15	FYKQVTTAIVKIVEKPWERVSVDGQPHEHGFK LGSEKHTTEVIVKKSGAL	Pfam entry not found
17	Plant	50	1.8e-505	15	TSGIEGLSLLKTTQSGFEGFIRDKYTALPDTRER MLATEVTALWRYSYES	Uricase family
18	Plant	50	4.6e-510	15	DTFFGPPKEGVYSPSVQNTLYLMAKAVLNRFP DIASVQLKMPNJHFLPVN	Uricase family
19	Plant	50	3.2e-269	15	VKFEDDVYLPTDEPHGSIEASLSRIWSKL	Pfam entry not found
20	Animal	50	1.1e-427	15	NDYKKNDEVEFVRTGYGKDMVKVLHIQRDGK YHSIKEVATSVQLTLSSKK	Pfam entry not found
21	Animal	50	4.3e-508	15	DYLHGDNSDIIPTDTIKNTVHVLAKFKGIKSIET FAMNICEHFLSSFNHV	Uricase family
22	Animal	41	2.7e-393	13	RAQVYVEEVPWKRFEKNGVKHVHAFIHTPTGT HFCEVEQMR	Pfam entry not found
23	Animal	50	1.8e-536	15	PPVIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPE VKDRCFATQVYCKWRY	Uricase family
24	Animal	50	6.8e-529	15	KFAGPYDKGEYSPSVQKTLYDIQVLSLSRVPEI EDMEISLPNIHYFNIDM	Uricase family
25	Animal	29	7.4e-252	12	KMGLINKEEVLLPLDNPYGRITGTVKRKL	Pfam entry not found

Theoretical pI of uricase from different organisms has ranged from 4.95-8.88. The pI value is ranged between 4-7, denoting that enzyme works best at acidic to neutral pH. Uricase from different sources showed pI values greater than 7, which indicated the basic properties of these enzymes. All the selected proteins from fungi, plants, and animals showed the pI value of more than 7, except bacterial species. The other characteristics of enzymes, namely negatively charged residues (Asp+Glu), positively charged residues (Arg+Lys), and GRAVY, also varied among the organisms. An Enzyme's overall charge depends upon the number of charged amino acids. A large number of acidic amino acids (glutamic acid and aspartic acid) exist in the negatively charged enzymes. In contrast, a large number of basic amino acids (arginine and lysine) exists in the positively charged enzymes. The negatively charged residues (Asp+Glu) are in large number in all the selected amino acid sequences of uricase compared to positively charged residues (Arg+Lys).

Extinction Coefficient (EC) denotes how much light a protein absorbs at a particular wavelength. The EC values of the selected bacterial uricase were ranging from 25,900-62,465 M^{-1} cm^{-1}, fungal uricase were in the range of 33,350 M^{-1} cm^{-1}, plant uricase ranged from 34,045-46,995 M^{-1} cm^{-1} and animal uricase in the range of 30,745-41,620 M^{-1} cm^{-1}. The instability index is an estimation of the stability of the desired protein (Artimo et al. 2012). A protein instability index value less than 40 indicates a stable protein, and a value beyond 40 indicates an unstable protein (Artimo et al. 2012). All selected sources of uricase had an instability index of less than 40 except for the species that belonged to the fungal source of *fusarium graminearum* (XP_011321510.1) and an animal source of *bombyx mori* (NP_001037382.1), which had an instability index of more than 40. Stable proteins are good candidates for industrial and medical applications.

Table 3.4: Physicochemical properties of uricase protein sequences from different sources of organisms computed using Expasy ProtParam tool

S.No.	Accession number	Species	Seq.length	MW	pI	-R	+R	EC	Ii	Ai	GRAVY
	(A) Bacteria										
1	AGA28823.1	*Singulisphaera acidiphila*	309	34,554.6	6.04	40	33	39,085	28.91	77.96	-0.375
2	EHR61468.1	*Saccharomonospora cyanea*	301	33,383.1	5.26	46	31	37,930	29.25	82.29	-0.369
3	ACV76680.1	*Nakamurella multipartita*	302	33,783.29	5.27	41	24	57,870	32.92	64.93	-0.523
4	ADW02506.1	*Streptomyces pratensis*	310	35,439.44	5.26	47	34	52,370	31.95	74.87	-0.594
5	KDN81792.1	*Kitasatospora cheerisanensis*	305	34,672.56	5.80	46	35	41,370	29.66	76.72	-0.536
6	ADG89294.1	*Thermobispora bispora*	301	33,811.97	5.70	43	33	46,410	33.59	86.84	-0.303
7	AEF42631.1	*Hoyosella subflava*	302	34,105.11	5.36	46	32	48,930	29.25	84.64	-0.375
8	EYU05833.1	*Pseudomonas aeruginosa*	308	35,664.17	5.55	45	33	62,465	38.99	70.68	-0.554
9	ADI14624.1	*Truepera radiovictrix*	307	33,857.32	6.15	37	33	25,900	31.65	82.70	-0.187
10	BAK35228.1	*Microlunatus phosphovorus*	305	33,852.06	5.08	43	29	47,900	34.77	89.21	-0.178
11	GAB16350.1	*Arthrobacter globiformis*	302	33,733.55	5.37	46	32	39,420	34.56	77.55	-0.387
12	AMY51449.1	*Rhodococcus fascians*	306	34,082.10	5.04	46	30	53,860	30.38	82.91	-0.331
13	KPI32983.1	*Actinobacteria bacterium*	305	34,692.04	6.61	40	38	45,505	30.99	79.93	-0.520
14	ADH98118.1	*Bacillus selenitireducens*	318	35,709.08	4.95	48	31	27,390	38.41	74.78	-0.353
15	KJQ52767.1	*Microbacterium sp.*	315	35,093.16	5.05	48	33	40,910	33.84	80.54	-0.301
	(B) Fungi										
16	KJK61270.1	*Aspergillus parasiticus*	302	34,240.7	7.18	37	37	53,525	37.53	80.33	-0.457
17	XP_006963697.1	*Trichoderma reesei*	308	34,173.6	6.60	37	35	43,555	28.15	80.10	-0.320
18	XP_001528662.1	*Lodderomyces elongisporus*	303	34,315.0	8.19	38	40	41,495	29.74	83.93	-0.454
19	XP_018702053.1	*Cordyceps fumosorosea*	302	33,506.9	7.90	36	37	39,420	32.95	79.11	-0.296
20	XP_001390131.1	*Aspergillus niger*	306	34,784.1	6.15	40	34	59,610	37.49	80.29	-0.418

21	XP_011321510.1	*Fusarium graminearum*	302	33,798.15	6.10	41	37	44,140	40.75	79.01	-0.378
22	KXN72467.1	*Conidiobolus coronatus*	311	35,204.57	7.87	33	34	33,350	37.03	80.55	-0.451
23	OSD05528.1	*Trametes coccinea*	333	36,951.91	6.19	42	37	38,850	38.63	88.41	-0.260
24	XP_020071435.1	*Cyberlindnera jadinii*	303	34,194.90	8.14	36	38	50,100	34.79	79.11	-0.392
25	XP_020075778.1	*Hyphopichia burtonii*	303	34,376.21	8.88	36	41	41,495	32.48	80.69	-0.426
26	XP_020049386.1	*Ascoidea rubescens*	300	34,113.90	8.71	34	38	40,005	35.07	86.70	-0.322
27	XP_002796429.1	*Paracoccidioides lutzii*	307	34,732.39	6.42	38	35	49,515	29.52	81.89	-0.341
28	OAT11654.1	*Blastomyces gilchristii*	314	35,032.64	7.25	34	34	46,535	25.51	75.45	-0.370
29	XP_014533620.1	*Penicillium digitatum*	302	33,683.29	6.51	35	33	49,055	25.44	86.19	-0.251
30	OBT42089.1	*Pseudogymnoascus sp.*	304	34,272.92	7.21	40	40	44,015	32.44	79.47	-0.434
	(C) Plant										
31	ABD03945.1	*Sorghum bicolor*	306	34,417.38	8.65	32	35	42,525	38.66	86.50	-0.189
32	ABD03944.1	*Saccharum officinarum*	306	34,179.37	8.85	31	35	42,525	33.20	88.43	-0.149
33	XP_007211669.1	*Prunus persica*	307	34,843.72	7.79	38	39	46,870	31.74	84.40	-0.366
34	XP_020235639.1	*Cajanus cajan*	310	35,014.08	8.31	37	39	45,505	26.28	86.39	-0.260
35	XP_002315419.2	*Populus trichocarpa*	308	35,032.97	8.67	35	38	41,035	31.86	81.30	-0.323
36	ABD03946.1	*Triticum aestivum*	307	34,391.09	8.33	34	36	42,525	36.30	81.17	-0.266
37	XP_022150541.1	*Momordica charantia*	313	35,560.58	8.75	37	41	41,495	33.48	85.85	-0.252
38	XP_022025248.1	*Helianthus annuus*	313	35,169.24	8.86	34	38	34,045	34.37	83.96	-0.295
39	NP_001267899.1	*Vitis vinifera*	309	34,701.50	8.36	35	37	38,515	36.48	83.30	-0.292
40	AAB97726.1	*Phaseolus vulgaris*	308	35,127.14	8.33	36	38	46,995	30.18	86.95	-0.291
41	BAA13184.1	*Glycine max*	309	35,138.12	8.31	36	38	46,995	25.77	84.79	-0.317
42	XP_020870005.1	*Arabidopsis lyrata*	309	34,840.76	8.30	36	38	38,515	26.55	82.59	-0.286
43	BAB18538.1	*Lotus japonicus*	307	34,985.05	7.76	39	40	40,005	29.77	85.67	-0.338
44	ABD03939.1	*Medicago truncatula*	308	34,974.94	8.52	39	42	40,005	33.80	81.27	-0.363
45	CAB77205.1	*Cicer arietinum*	308	35,133.05	8.61	40	43	38,515	37.54	79.71	-0.411
	(D) Animal										
46	NP_446220.1	*Rattus norvegicus*	303	34,907.94	8.20	39	41	34,630	33.55	82.24	-0.457
47	NP_001121545.1	*Oryctolaguscuniculus*	300	34,500.54	7.72	38	39	34,630	36.41	85.93	-0.374
48	NP_033500.1	*Mus musculus*	303	35,039.18	8.48	38	41	36,120	38.62	83.86	-0.460

49	XP_006070940.1	*Bubalus bubalis*	304	35,163.25	8.20	38	40	36,120	38.55	83.88	-0.438
50	XP_006919420.1	*Pteropus alecto*	304	35,240.54	8.13	40	42	41,620	37.26	82.96	-0.438
51	XP_006086199.2	*Myotis lucifugus*	303	34,772.74	8.48	39	42	33,140	30.58	79.04	-0.458
52	XP_015146362.1	*Gallus gallus*	320	36,603.20	7.05	37	37	35,505	36.21	82.75	-0.220
53	XP_012657176.1	*Otolemur garnettii*	303	34,808.77	6.97	40	39	34,630	32.43	82.87	-0.400
54	XP_012603035.1	*Microcebus murinus*	304	35,039.07	8.20	39	41	34,630	32.66	82.93	-0.457
55	XP_021013952.1	*Mus caroli*	304	35,207.29	7.75	40	41	36,120	38.05	82.93	-0.485
56	XP_025727815.1	*Callorhinus ursinus*	304	35,133.31	8.10	40	42	30,745	35.75	78.75	-0.464
57	XP_025273620.1	*Canis lupus dingo*	304	35,124.20	8.17	40	42	37,610	32.10	78.75	-0.480
58	NP_001037382.1	*Bombyx mori*	337	38,173.39	6.90	40	39	55,475	45.77	80.68	-0.347
59	AFP60128.1	*Musca domestica*	338	38,697.70	7.27	40	40	39,100	32.86	76.36	-0.519
60	XP_022352809.1	*Enhydra lutris kenyoni*	304	35,160.26	6.95	41	40	37,610	31.30	78.75	-0.454

MW=molecular weight (g/mol), pI= isoelectric point, -R=number of negative residues, +R=number of positive residues, EC=extinction coefficient (M^{-1}cm^{-1}), Ii=instability index, Ai=aliphatic index, GRAVY=grand average hydropathicity

The aliphatic index of protein means a measure of aliphatic groups (valine, acid alanine, leucine, and isoleucine) occupied in the total area of the protein. In this study, all the species of uricase showed an aliphatic index ranging from 70-89. These moderate-high aliphatic indices denote that the uricase is thermostable. The higher molecular weight and lower molecular weight of uricase showed a higher value of the aliphatic index. If the aliphatic index shows a higher value, then the protein is more thermostable than the low aliphatic index value of an enzyme (Artimo et al. 2012). The higher aliphatic index of protein may be considered a good indication of the enhanced thermostability of globular protein (Ikai 1980). The high thermostability of uricase is an excellent characteristic to use in industrial applications (Ikai 1980; Rawlings et al. 2010). The uricase of bacterial species *Microlunatus phosphovorus* (BAK35228.1) has the highest aliphatic index for fungi *Trametes coccinea* (OSD05528.1), plant *Phaseolus vulgaris* (AAB97726.1), and animal *Oryctolagus cuniculus* (NP_001121545.1).

The GRAVY (grand average hydropathicity) index is defined as the interaction of a given protein with water. The lower (negative) value of GRAVY shows better interaction between protein and water and also shows protein is hydrophilic, while a value above 0 denotes protein is hydrophobic. In this study, the GRAVY values were in the range between -0.149 and -0.594, denoting that uricase is a hydrophilic protein. Here, the uricase from *Saccharum officinarum* (ABD03944.1) has the lowest GRAVY value; it indicated that this protein has better interaction with water than other proteins. All the above characteristic data are presented in Table3.4, which were retrieved by the ProtParam tool. The proteins had better interactions with water molecules when they had a low range value of the GRAVY (Verma et al. 2016).

The amino acid composition of uricase protein sequences from different sources of organisms is shown in Appendix I, where twenty amino acids composition has been computed. The results showed that an average frequency of valine amino acid had the highest percent of 8.79, and cysteine had the lowest percent of 0.91 compared to other amino acids in all analyzed species. The average frequency of amino acid alanine showed the highest percent of 9.45 in bacterial species, whereas proline showed the lowest percent of 3.65. However, in fungal species, valine showed the highest percent

of 8.79, and proline showed the lowest percent of 4.09. Similarly, in plant species, valine showed the highest percent of 9.70, and proline showed the lowest percent of 4.56. In animal species, valine showed the highest percent of 8.81, and alanine showed the lowest percent of 3.67. The results exhibit the percentage of different amino acids that contributed to the formation of uricase sequences from different analyzed species. The present observations were confirmed by comparing it with other similar assessments of physicochemical features of several proteins by previous workers (Dubey et al. 2010; Kumar et al. 2012; Morya et al. 2012; Verma et al. 2016; Yadav et al. 2017).

Table 3.5: The list of software/databases used

Name of the software/databases	Input	Output
NCBI database	-	Protein sequences
ClustalW (MEGA7 Software)	Protein sequences	Multiple sequence alignment for conserved amino acid residues
Neighbor-joining statistical method (MEGA7 Software)	Protein sequences	Phylogenetic tree for the evolutionary history
Multiple EM for Motif Elicitation (MEME)	Protein sequences	Conserved motif identification
Pfam database	Sequence motifs	Motif family identification
ExPASy ProtParam tool	Protein sequences	Physicochemical characterization and amino acid composition

The theoretical pI of uricase derived from bacterial, fungal, plant, and animal sources ranged between 4.95-8.88.The pI value is ranged between 4-7 denoting that enzyme works best at acidic to neutral pH. Uricase from various sources had pI values greater than 7, indicating that these enzymes possessed required fundamental property. Except bacterial species, all other sources like fungi, plants, and animals have a pI value greater than 7. The bacterial source will be the best source since it has pI value less than 7. Under bacterial source, fifteen different bacterial species are present in this study and one of the best bacterial source is the *Bacillus* species. *Bacillus selenitireducens* (ADH98118.1) has a sequence length of 318, molecular weight of 35 kDa, pI of 4.95, -R of 48, +R of 31, EC of 27,390, Ii of 38.41, Ai of 74.78, and

GRAVY of -0.353. To explore the usefulness of bacterial uricase, *Bacillus* species are chosen for further study, to screen the best among them based on economical and unique properties by considering the significance of the enzyme to treat hyperuricemia.

Among the wide range of microorganisms, bacterial strains are preferred because of their well know properties like easy cultivation, very fast growth rate, very high protein yield, very low production cost, and for producing recombinant enzymes. In microbial fermentation, *Bacillus* species remain the primary bacterial workhorses. Certain *Bacillus* species are GRAS (generally recognized as safe) according to the Food and Drug Administration. *Bacillus* strains with the capacity to produce and secrete large amounts (20–25 g/L) of enzymes have achieved the prominence as industrial enzyme producers (Barros et al. 2013; Schallmey et al. 2004). The detection and identification of new strains of uricase have a high demand in the medical field. Computational approaches can be used to screen and investigate an uricase enzyme with desirable characteristics that can be employed in diverse industrial applications.

3.3 SUMMARY

In the present work, efforts were made to evaluate an overview of the computational characterization of uricase protein sequences from different sources using bioinformatics tools. In multiple sequence analysis and homology search findings of all the selected sequences, similarities between the protein sequences and maximum conservation of amino acids were observed to be between 51-314 residues. All the analyzed species of uricase possessed one glutamine residue, which was identically conserved in all the selected sources of sequences. Phylogenetic analysis of all the selected sequences from a different source of organisms showed separate clusters, and it showed the sequence similarity based on the source of the organism. This cluster analysis of all retrieved protein sequences gave a clear understanding of the evolutionary relationship among different groups of uricase at the molecular level. Six motifs exist in each of the sequences, and all twenty-five motifs are unique for their group belonging to different sources of uricase. From the computational

physicochemical features of all the selected uricase, proteins gave a complete understanding of properties, namely pI, EC, Ai, Ii, GRAVY, and are in the nature of basic properties of these enzymes with 33 kDa-39 kDa molecular weight. The amino acid valine has a high average frequency of 8.79 percent in all the selected sources compared with all other different amino acids that exist in uricase, denoting the amino acid valine to have a key lead in the formation of uricase.

CHAPTER 4

IN SILICO STRUCTURAL AND FUNCTIONAL ANALYSIS OF *BACILLUS* URICASES

Uricase is important because of its potential use in medicinal chemistry and the treatment of several diseases (Khade and Srivastava 2015). Microbial uricase is found to be inducible; therefore, uric acid or some other inducer is required for enzyme production in the medium (Adámek et al. 1989). Uricase from microorganisms and animals is highly antigenic, and repeated injections can result in anaphylactic shock and allergic reactions, which can be fatal (Bomalaski et al. 2002). Further, uricase is more expensive than allopurinol, which is the first choice of drug administered during conventional treatment of hyperuricemia. Therefore, the use of uricase as a therapeutic drug is highly restricted (Beedkar et al. 2012). Hence, there is a pressing need for a cost-effective method for uricase production (Beedkar et al. 2012). Therefore, being an essential clinical enzyme, there is a great demand for highly active and highly pure forms of uricase.

The advantages of the uricase enzyme include high selectivity and affinity towards its substrate uric acid, whereas the disadvantages include poor catalytic efficiency at physiological temperature ($37°C$) and pH (7.2), short half-life, inherent antigenicity, and limitations on effective treatment. Due to several serious drawbacks of uricase, its clinical usage has been limited (Tan et al. 2012). Yamamoto et al. have reported that thermophilic *Bacillus sp.*TB-90 was found to produce uricase that has been extensively studied for clinical purposes, and also it is thermally stable and has higher activity between pH 6-9 (Yamamoto et al. 1996). Hua Huang et al. developed a biochip system to detect uric acid by using purified recombinant *Bacillus subtilis* uricase (Huang and Wu 2004). Various uricase enzymes have been industrially produced by culturing several microorganisms (Feng et al. 2010). Discovering bacterial species that produce such enzymes can be used for the isolation of industrial enzymes (Rahmatabadi et al. 2017). Also, several *Bacillus* species have

produced uricase with 25-30 U/ml of activity, which shows the most important sources of industrial enzymes (Pustake et al. 2019). Uricase from *Bacillus fastidious* was commercialized by Sigma-Aldrich (product 94310, 9 U/mg) and has been used for various applications (Gruia et al. 2017; Nanda et al. 2016). It has a K_m value of approximately 65 μM and a K_i value of 4.8 μM, which is much lower than the intracellular uricase from *Bacillus fastidious* ATCC 29604 (Zhao et al. 2006). Considering its importance in treating diseases, it is still necessary to screen the new uricase producers that are more economical and may have unique properties to expand their usefulness. The detection and identification of new species capable of producing uricase have a high demand in the medical field.

Structural and functional analysis of enzymes using wet-lab techniques is a time-consuming and expensive process compared to the application of bioinformatics tools, which are more economical and time-saving methods (Rahmatabadi et al. 2017). Bioinformatic tools are useful for understanding the properties of unknown proteins with the aid of their sequence, structural, functional, and evolutionary data obtained by computational genomics and proteomics studies (Koteswara Reddy et al. 2017). For functional analysis of a protein, the 3D structure is required. Till now, there is no report on structural and functional characteristics of the uricase obtained from multiple species. The focus of the current study is the computational characterization of 70 uricase protein sequences from various *Bacillus* species and to investigate the physical parameters, secondary and tertiary structure, functional properties, domains, motifs, and phylogenetic relationship using various bioinformatics tools.

4.1 MATERIALS AND METHOD

4.1.1 Retrieval of uricase sequences

For the computational investigation, full-length amino acid sequences of uricase from various *Bacillus* species were searched and retrieved from the protein database of NCBI (National Center for Biotechnology Information) (http://www.ncbi.nl m.nih.gov/). Uricase proteins from 70 *Bacillus* species were selected and downloaded in FASTA format for further *in silico* analysis. Furthermore, UniProtKB (Universal Protein Resource Knowledgebase) Swiss Prot database (http://www.uniprot.org)

(Artimo et al. 2012; Pundir et al. 2017) was used for collecting functional information about proteins, and ExPASy (Expert Protein Analysis System) was also employed for obtaining amino acid sequences.

4.1.2 Multiple sequence alignment

The ClustalW tool available on the MEGA 7 software (Molecular Evolutionary Genetics Analysis) was used to perform the multiple sequence alignment (MSA) of retrieved protein sequences to identify the similarities between uricase among *Bacillus* species of the same family. All the parameters such as gap open, gap extension, and gap distance with end gaps were set as the default values. Clustalomega, MUSCLE, Sea view was also used to perform multiple sequence alignment.

4.1.3 Phylogenetic analysis

MEGA 7 tool (Kumar et al. 2016) was employed for constructing phylogenetic trees using the amino acid sequences and reverse translated sequences of uricase. Bioinformatics Reverse Translation Tool was used for converting the retrieved uricase protein sequences into gene sequences for the construction of cDNA (reverse translated) tree. The phylogenetic tree is a diagrammatic representation that shows the evolutionary relationships of various organisms (Yadav et al. 2009). The neighbor-joining (NJ) method (Saitou and Nei 1987) was employed in both cases for determining the evolutionary history of uricase and was based on the p- distance based model. Thousand bootstrap replications were used to test the phylogeny (Felsenstein 1985). The branching pattern in the evolutionary tree indicates how various biological species or other entities evolved from their common ancestors. Evolutionary distances and the distinct branches or groups of *Bacillus* uricase were observed for understanding the evolutionary ancestry.

4.1.4 Motif analysis

Multiple EM for Motif Elicitation (MEME) suite was used to perform motif discovery on DNA, RNA, or protein (Bailey et al. 2009). Uricase protein sequences of selected *Bacillus* species were submitted as the input for searching and analyzing conserved motifs in the MEME tool (http://meme-suite.org/tools/meme) (Bailey and Elkan

1994) using minimum motif width of 6, maximum of 50, and a maximum number of motifs as 6. This tool was employed for assessing vital signature sequences in *Bacillus* uricases. Pfam database (Finn et al. 2014) has a large collection of protein families, and domains and hence was used for the analysis of the uricase domains. The uricase sequences were submitted in the Pfam database at the sequence search option to find matches to the uricase family in the database. The domain organization of all the uricase sequences was analyzed.

4.1.5 Physicochemical characterization

The prediction of physicochemical characteristics of uricase protein sequences of various *Bacillus* species was computed using the ProtParam tool (Pooja et al. 2017; Rani et al. 2017). The ProtParam tool on the ExPASy server (Gasteiger et al. 2005) permits the determination of the physicochemical properties of a given protein. The calculated parameters such as amino acid composition, molecular weight (MW), theoretical pI, amino acids composition, the total number of negatively (Asp+Glu) and positively (Arg+Lys) charged residues, atomic composition, total number of atoms, extinction coefficient (EC) (Gill and von Hippel 1989), *in vivo* half-life, instability index (II) (Guruprasad et al. 1990), aliphatic index (AI) (Ikai 1980) and GRAVY (Grand average of hydropathicity) (Kyte and Doolittle 1982) were analyzed (Gasteiger et al. 2005). The server Akriti v1.0 and ProtScale, which computes pI/mW, are also useful to study the physicochemical features of the protein. Sequence manipulation suite2 was utilized for the characterization of theoretical pI and molecular weights of the protein. The Bjellqvist method was used for the determination of theoretical pI using the ProtParam server (Bjellqvist et al. 1993, 1994).

4.1.6 Secondary structure analysis

The web-based servers were used for predicting secondary structures of the retrieved *Bacillus* uricase sequences. The protein folding directly depends on the number of secondary structure elements. Hence, the presence of α-helices, extended strands, β-turns, random coils, and β-sheets in several *Bacillus* species of uricase was predicted using ExPASy SIB Bioinformatics. SOPMA (Self-Optimized Prediction Method with

Alignment), PSIPRED v3.3 protein sequence analysis workbench (Combet et al. 2000; Geourjon and Deléage 1995) and CFSSP server (Chou and Fasman Secondary Structure Prediction) (Jones 1999; McGuffin et al. 2000) were employed for the prediction of overall secondary structure. CFSSP is an online server that uses the Chou and Fasman algorithm for predicting the secondary structure of the protein from the amino acid sequence (Ashok Kumar 2013).

4.1.7 Phyre2 protein modeling, prediction and analysis

Phyre2 (Protein Homology/AnalogY Recognition Engine V 2.0) is a free web-based server available (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) for the prediction of protein structure, function, domain, domain boundary, site-directed mutagenesis, and evolutionary classification of proteins and protein crystal structures by molecular replacement. It is an easy protein bioinformatics tool and one of the most widely used protein structure prediction server, which has been cited over 1500 times. It can build reliable 3D protein models based on remote homology detection methods for predicting ligand binding sites and analyzing the impact of the given amino acid sequence variants (Kelley et al. 2015). It is used for elucidating the secondary and tertiary structure, composition of the domain, and quality of the model for the selected uricase protein sequences.

4.1.8 Tertiary structure analysis

The combination of overall secondary structure elements forms the tertiary structure, which characterizes the function of a protein. The amino acid sequence of *Bacillus simplex* was chosen as a standard among all 70 *Bacillus* species of uricase protein for tertiary structure prediction. The SWISS-MODEL server (Schwede et al. 2003) in automated mode was used to get 3D protein homology models of all 70 protein sequences then selected *Bacillus simplex* (WP_063232385.1) as its QMEAN value is close to 0 (zero) by selecting the most suited template (Pramanik et al. 2017). SWISS-MODEL is a homology-modeling server for predicting 3D protein structures and is one of the most widely used free web servers. The quality of the predicted model was evaluated and verified based on the QMEAN4 score, Z-score (Benkert et al. 2009, 2011), and The Structure Analysis and Verification Server v5.0 (SAVES). The

highest overall quality factor was produced by ERRAT (Colovos and Yeates 1993), Verify 3D (Bowie et al. 1991; Lüthy et al. 1992), Ramachandran plot and RAMPAGE (Lovell et al. 2003), and finally PROQ server (Cristobal et al. 2001) by submitting the predicted structure. The backbone conformational regions of the built model were investigated using the Ramachandran plot analysis (Lovell et al. 2003).

4.1.9 Functional analysis

Various tools were used for the evaluation of the functional characteristics of uricase protein sequences. CYS_REC tool was employed to analyze the presence of disulfide bonds and to identify the positions of cysteine residues. This tool also computed the SS bond pattern pairs in the protein sequence and calculated the presence of the SS bond in a protein (Roy et al. 2011). To determine the functional motifs and the superfamily to which the selected uricase protein sequence belongs, the Motif search tool (www.genome.jp/tools/motif/) (Singh et al. 2012) was used. The Conserved Domain Database (https://www.ncbi.nlm.nih.gov/cdd/) from NCBI was used to find the conserved domains of the selected uricase protein. STRING v11.0 (https://string-db.org/) web server performed the analysis of interacting partners of *Bacillus simplex* uricase with other closely related proteins (Szklarczyk et al. 2015). STRING is a database of known and predicted protein-protein interactions. SOSUI tool was used to distinguish whether the protein is a soluble or a transmembrane protein and also to determine transmembrane helices from a given amino acid sequence. All the retrieved *Bacillus* uricases and also the selected uricase protein were analyzed by this server (Hirokawa et al. 1998). The potential cleavage sites of proteases or chemicals in a given protein sequence were discovered by PeptideCuttertool (Appaiah and Vasu 2016).

4.2 RESULTS AND DISCUSSION

4.2.1 Retrieval of uricase sequences

Uricase amino acid sequences and gene sequences of 70 various *Bacillus* species were collected from the NCBI database in FASTA format. For *in silico* studies, the full-length sequences were used, and the variable lengths of the amino acids were found to

be in the range of 312-502. The list of obtained uricase protein sequences, along with their accession numbers, is listed in Table 4.1.

4.2.2 Multiple sequence alignment

The alignment of multiple related uricase protein sequences from various *Bacillus* species is performed by the ClustalW tool in MEGA 7 software to attain optimal matching of the sequences. This tool identified many highly conserved amino acids that are highlighted and indicated by a * mark (Appendix II) by the boxshade server. Among them, "YGK-RT-PL-IPES-SF-GDN," "ATDSMKN-EGF," "KV-SF," "YT-RPL-YVA-EQ," and "SIQ-IG-FPQL-TW-GFQ" are found to be the highly conserved sequences. Similar multiple sequence alignment studies were reported in the literature for other proteins (Dubey et al. 2010; Irajie et al. 2016; Niño-Gómez et al. 2017).

Table 4.1: List of enzyme uricase sequences from different *Bacillus* species

Sl. No.	Source organism	Accession number	No. of sequences
1	*Bacillus subtilis*	WP_101501434.1,BAM59327.1,APH66064.1,AII36988,CCU60286.1,EXF55358.1, WP_003222862.1, EME07777.1, AKC48817.1	9
2	*Bacillus Sp. Tb-90*	3WLV_A,1J2G_A,5AYJ_A,BAA08723.1, BAB20808.1	5
3	*Bacillus clausii*	KKI85158.1,BAD66267.1,WP_095326636.1,WP_095294289.1,WP_095236414.1, PAF09838.1, PAE88988.1, PAD14932.1	8
4	*Bacillus halotolerans*	KUP29050.1,PRS06588.1,PRP51591.1, WP_099043576.1	4
5	*Bacillus simplex*	PCD05853.1,PAL09042.1,CEG34811.1,ASS93773.1,WP_063232385.1,WP_061143228.1	6
6	*Bacillus siamensis*	PAD64173.1,WP_095241385.1,WP_045926035.1	3
7	*Bacillus gibsonii*	AOL30990.1	1
8	*Bacillus sp. BA3*	WP_101224285.1	1
9	*Bacillus intestinalis*	OWV36502.1, AJW84706.1, KFK78955.1	3
10	*Bacillus cereus*	AXJ21641.1, AUZ27736.1	2
11	*Bacillus filamentosus*	OXS68986.1, AKO95039.1, WP_081496159.1	3
12	*Bacillus fastidiosus*	ACR09749.1, 4R8X_A, 4R99_A	3
13	*Bacillus flexus*	AQX54882.1,WP_061784634.1,WP_0789	3

		89772.1	
14	*Bacillus smithii*	WP_040342081.1, WP_048623468.1	2
15	*Bacillus sp. AFS017274*	WP_098373266.1, PEZ74426.1	2
16	*Bacillus niacin*	KGM46460.1,WP_045524647.1,WP_0346 72575.1	3
17	*Bacillus beveridgei*	AOM84027.1	1
18	*Bacillus circulans*	SPT78254.1	1
19	*Bacillus licheniformis*	OLQ49074.1, WP_075749098.1	2
20	*Bacillus sp. JS*	AFI29791.1, WP_014665258.1	2
21	*Bacillus aryabhattai*	OZT14492.1, WP_094910043.1	2
22	*Bacillus sp.RU2C*	WP_083686476.1	1
23	*Bacillus sp. MD-5*	ASB62313.1	1
24	*Bacillus sp. mrc49*	PJN86603.1	1
25	*Bacillus sp. FJAT-22058*	KOR85772.1	1
	Total sequences		70

4.2.3 Phylogenetic analysis

The uricase amino acid and cDNA sequences are phylogenetically analyzed to study their evolutionary relationships among various *Bacillus* species using the NJ method. It is observed from the uricase protein phylogenetic tree that there are different clusters for species denoted as I, II, III, IV, V, VI, and VII, which consists of 27, 3, 9, 5, 13, 6, and 7 protein sequences respectively depicting the interrelationships within them (Figure 4.1). Multiple *Bacillus* species were grouped into distinct clusters exhibiting sequence similarity. Based on the phylogenetic tree, the uricase protein sequence of *Bacillus simplex* (WP_063232385.1) has a close evolutionary relationship with *Bacillus simplex* NBRC (ASS93773.1), which in turn is closely related to *Bacillus sp*.BA3 (WP_101224285.1) and *Bacillus sp.mrc*49 (PJN86603.1), and their closest neighbour is *Bacillus simplex* (WP_061143228.1). These species are clustered with each other and display 100% similarity. Cluster VII which includes *Bacillus smithii* (WP_040342081.1, WP_048623468.1), chain A of pdb|5AYJ|, pdb|3WLV|, pdb|1J2G| and *Bacillus sp*.TB-90 (BAA08723.1, BAB20808.1) is an outgroup compared to other clusters that indicate that they are distantly related.

The cDNA-based phylogenetic tree of uricase protein showed seven distinct clusters comprising 12, 10, 3, 11, 16, 5 and 13 sequences, respectively (Figure 4.2). The tree also showed that the uricase protein sequence of *Bacillus simplex* (WP_063232385.1) clustered with *Bacillus sp*.BA3 (WP_101224285.1) showed 100% similarity. These two sequences were closely related to *Bacillus smithii* (WP_040342081.1) and *Bacillus sp* TB-90 of the same cluster. This tree is constructed to know if there is any correlation between the respective cDNA of all the *Bacillus* uricase protein sequences. In addition, a similar type of phylogenetic analysis using protein sequences and their respective cDNA of bacterial xylanase, *Pseudomonas* lipases, and *mesorhizobium* ACC deaminase were analyzed and reported in the literature (Dutta et al. 2018; Pramanik et al. 2017, 2018; Yadav et al. 2009).

4.2.4 Motif analysis

The biological sequence characterization of uricase sequences from *Bacillus* species is identified by the MEME web-based program. A total of six motifs named as 1, 2, 3, 4, 5, and 6 were predicted in all the selected sequences (Appendix II). These motifs were identified in all *Bacillus* species sequences except for motif 5, which was absent in *Bacillus niacini* (KGM46460.1, WP_045524647.1, WP_034672575.1) and *Bacillus beveridgei* (AOM84027.1). The highest frame width of 50 was found in the first five motifs, with the exception of the sixth motif, which had a frame width of 29. Motif five has the lowest E value of 8.9e-1915 with 50 frame width and was observed in 66 sequences. All six motifs were present in 70 sequences except in motif 5, which was present in 66 sequences and E values in the range of 8.9e-1915 to 1.0e-2792 (Table 4.2). The observations from the analysis of protein sequences indicate that almost all the *Bacillus* species are a set of closely related sequences with conserved motifs, which indicates the possible role of these motifs in biological functions.

In addition, domain analysis by Pfam indicates that uricase enzyme sequences from several *Bacillus* species have two domain organizations that belong to the uricase family. These results revealed that the first two conserved motifs are a part of the uricase domain family, whereas the Pfam entry of other remaining conserved motifs was not available. 16S rRNA and uricase proteins were used to identify the phylogeny

of uricase-producing bacteria (Dabbagh et al. 2012). Similar methods of motif analysis of several enzymes have been previously reported (Bose et al. 2013; Dwivedi et al. 2013; Dwivedi and Mishra 2014; Pandey et al. 2011; Ramya and Pulicherla 2015).

It is believed that the residues that are conserved throughout evolution have a very crucial role in the structure and function of any enzyme. It is evident from multiple sequence alignment (MSA) (Appendix II) that 78.04% of the residues are conserved in all selected *Bacillus* uricases. All the six motifs obtained from the MEME database are shown in Figure 4.3. Among all the motifs, two motifs, i.e., motif1 and motif2, are found to be containing more number of conserved residues Figure 4.3 (A) and *Bacillu*s uricases (Appendix II). It can be noted here that the location of the catalytic pocket is within the domains. There is evidence suggesting that mutation within the conserved motifs can lead to a reduction in the activity of the enzyme. Imhoff et al. showed site-directed mutagenesis in Lys9Met, Thr69Ala of *Bacillus subtilis* and Lys22Met, and Thr67Ala of *Arthrobacter globiformis* uricase led to the loss of uricase activity (Imhoff et al. 2003). Furthermore, the reports provided by Ito et al. showed that Lys164Glu in the conserved amino acid sequence from Leu 160 to Lys164 of rat liver uricase reduces the activity significantly (Ito et al. 1992). In the case of *Bacillus simplex* (WP_063232385.1)*,* the junction of motif1 and 2 are found to be involved in making the active site, as shown in Figure 4.3 (A). The β-sheet of motif1 involve in making the central cavity of uricase. It is evident from Figure 4.3 (B) that the unstructured portion of motif3 acts as a cover which is crucial to reduce the solvent exposure of the catalytic pocket and provide a hydrophobic environment. The long β-sheet region in motif 3 is involved in making the central cavity of uricase. Therefore, it can be said that motif1, 2, and 3 are crucial for maintaining the catalytic activity of uricase. The rest motifs are mainly responsible for maintaining the enzyme structure, and the number of conserved residues are less compared to motif 1,2, 3. It is evident from MSA (Appendix II) that 21.96% of residues are unconserved, and they are mainly located at the surface of the *Bacillus* uricase. They mainly fold in unstructured form and have little contribution to maintaining the structure of the protein.

80
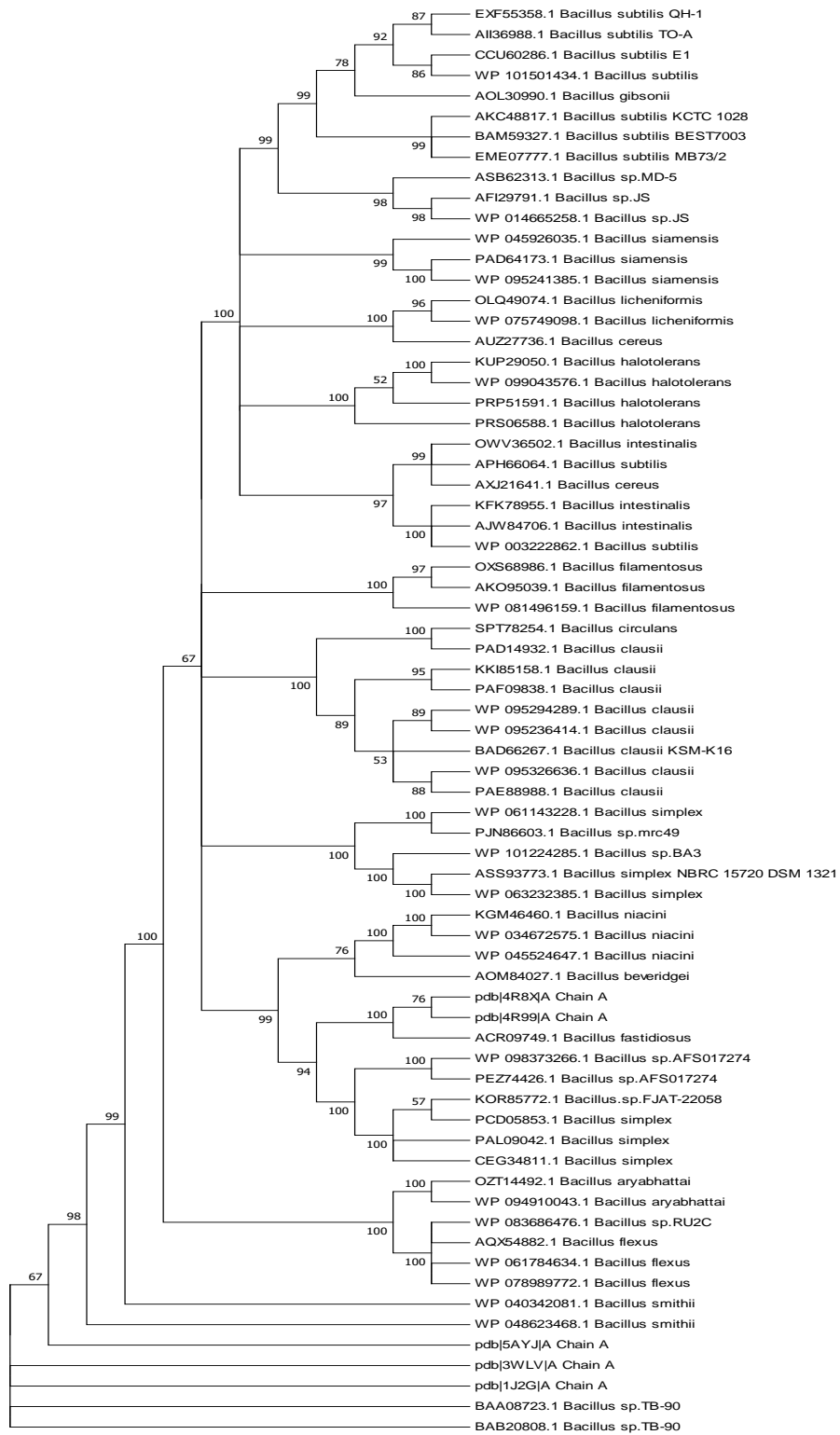
Figure 4.1: Phylogenetic tree of uricase amino acid sequences of different *Bacillus* species
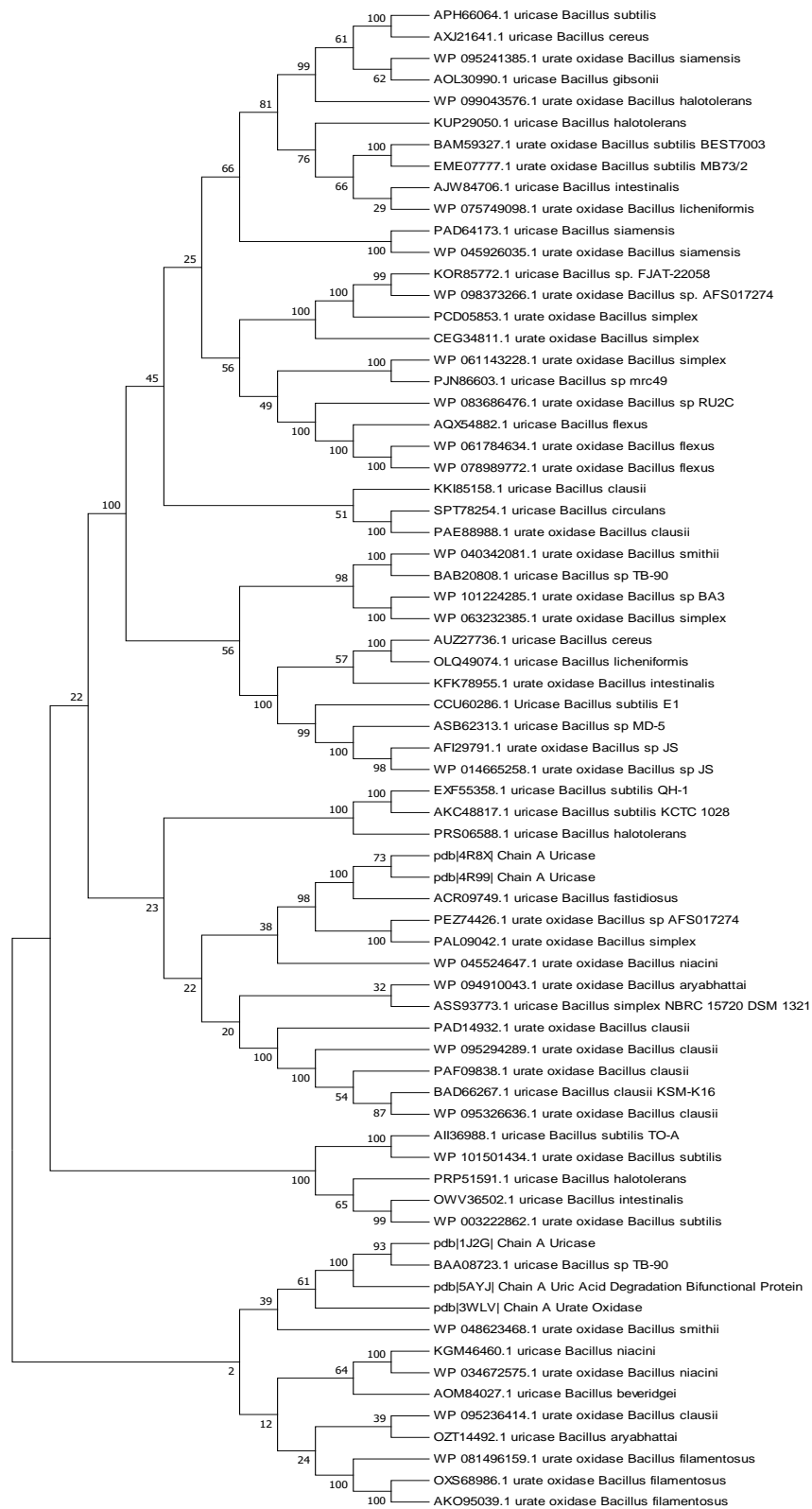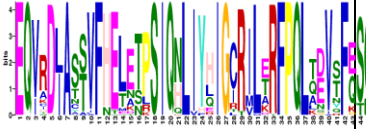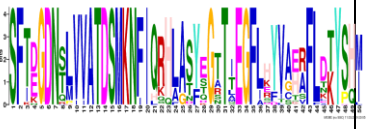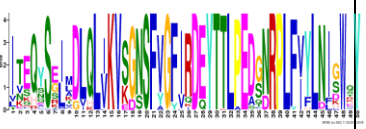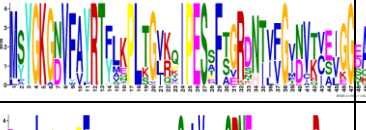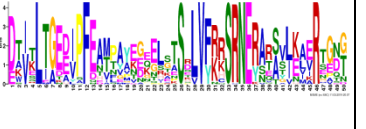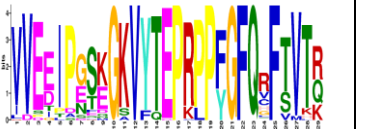
Figure 4.2: Phylogenetic tree of cDNA of uricase of different *Bacillus* species

Table 4.2 Six motifs best possible match information with sequence logo of uricase enzyme

Motif No	Logo	E-value	Sites	Width	Best Possible match	Pfam
1		2.0e-3081	70	50	EQVIDIATSIFHEMETPS IQNLIYEIGCRILTRFPQ LLEVTFESQNHTWD	Uricase
2		2.0e-2973	70	50	SFTEGDNSMVVATDSM KNFIQQHLATFKGATL EGFASYVSEAFLNKYP QI	Uricase
3		1.8e-2952	70	50	IVQQSSSILDLQLIKVSG NSFVGFVRDEYTTLPE DGNRPLFIYLNLHWVY	Not Found
4		1.0e-2792	70	50	LSYGKGNVFAYRTYSN PLTGIKQIPESTFSGRDH IIFGTNVKVSVGGSSF	Not Found
5		8.9e-1915	66	50	DTVKLIAEDIPFEAVTE ATDPQLKPSDLVFKKS RNERANAAVEIIRGEN G	Not Found
6		5.1e-1632	70	29	VVSEIPESKGKVYTEPR PPYGFQVFTVKK	Not Found
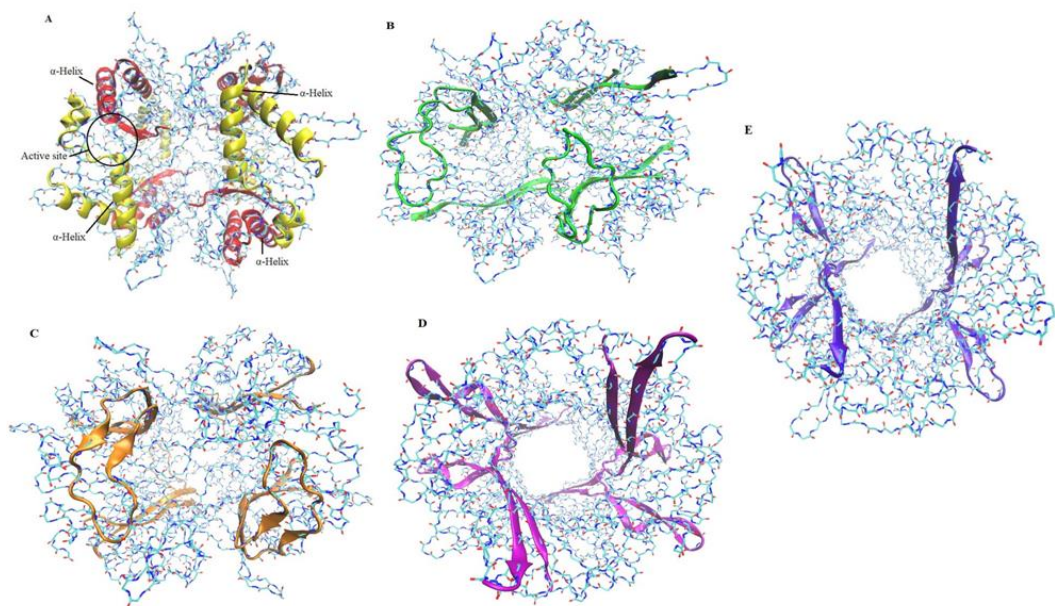
Figure 4.3: The locations of motifs and their tertiary structures of the uricase A) Motif1 (Red colour) & Motif2 (Yellow colour) B) Motif3 C) Motif4 D) Motif5 E) Motif6

4.2.5 Physicochemical characterization

The uniqueness of any given protein or enzyme molecules depends on a group of physicochemical properties such as theoretical isoelectric point (pI) value, molecular weight, instability index, aliphatic index, extinction coefficients, grand average of hydrophobicity, and the total number of amino acid residues. All these parameters of different *Bacillus* species of uricase were computed by the ProtParam tool and are listed in (Appendix II). Isoelectric point (pI) means the pH at which the net charge is zero. Theoretically, a pI value of greater than 7 depicts alkaline nature, and less than 7 value represents the acidic nature of the protein. The pI values of the selected uricase sequences of *Bacillus* species lie between 4.9 to 6.25, indicating that the nature of proteins varies from acidic to a neutral environment. The molecular weight of uricase has a range of 35.59 -59.85 kDa. The stability of proteins is indicated by the instability index. A value below 40 indicates that the protein structure is stable, whereas a value greater than 40 indicates that the protein structure may be unstable in nature according to the literature (Pramanik et al. 2017; Verma et al. 2016). The

instability index for all the species was between 23.01 to 45.63, indicating that most of the selected uricase proteins are stable in nature. The average extinction coefficient (EC) of the proteins is 40815.43. The EC value indicates the amount of light that may be absorbed by a selected protein at a certain wavelength.

The aliphatic index (AI) of a protein defines the relative volume occupied by aliphatic side chains, which may be considered to be a positive indication towards the thermostability of globular proteins. The AI of selected sequences varied between 70.92-85.67, suggesting that the proteins are thermostable in nature. A high aliphatic index indicates a higher thermostability of the protein. Uricase from *Bacillus aryabhattai* (OZT14492.1 and WP_094910043.1) has a high aliphatic index of 85.67, promising to be the most thermostable among all selected species. The suitable property for industrial use is the high thermostability of uricase. GRAVY (Grand average of hydrophobicity) values of all *Bacillus* uricase proteins were found to be negative, which indicates that these uricases have good interaction with the water molecule. A larger number of negatively charged amino acids (Asp+Glu) were observed in all uricase *Bacillus* species compared to positively charged amino acids (Arg+Lys). The overall charge of an enzyme depends on the number of charged amino acids. The amino acid composition of twenty amino acids was determined for uricase *Bacillus* species. The average frequency of mainly three amino acids, glutamine (8.585%), threonine (8.04%), and leucine (7.71%), were highly distributed. Other amino acids like leucine, serine, valine, alanine, and glycine were also rich in these protein sequences. Also, uricase from *Bacillus simplex* (WP_063232385.1) had glutamine (8.7%), serine (8.5%), and isoleucine (8.0%) predominantly as listed in Table 4.3. Various microbial proteins have reported physiological properties in the literature (Malviya et al. 2011; Morya et al. 2012; Nelapati and PonnanEttiyappan 2019; Yadav et al. 2017).

Table 4.3: Name, number and percentage of amino acids of uricase *Bacillus simplex* (WP_063232385.1)

S.No	Amino acid composition		
1	Ala (A)	31	6.2%
2	Arg (R)	17	3.4%
3	Asn (N)	29	5.8%
4	Asp (D)	24	4.8%
5	Cys (C)	2	0.4%
6	Gln (Q)	20	4.0%
7	Glu (E)	43	8.7%
8	Gly (G)	27	5.4%
9	His (H)	12	2.4%
10	Ile (I)	40	8.0%
11	Leu (L)	37	7.4%
12	Lys (K)	37	7.4%
13	Met (M)	9	1.8%
14	Phe (F)	30	6.0%
15	Pro (P)	21	4.2%
16	Ser (S)	42	8.5%
17	Thr (T)	29	5.8%
18	Trp (W)	3	0.6%
19	Tyr (Y)	15	3.0%
20	Val (V)	29	5.8%

4.2.6 Secondary structure analysis

The prediction of secondary structure elements in uricase from different *Bacillus* species was evaluated using CFSSP and SOPMA tools which are presented (Appendix II). SOPMA results show that four classes of protein secondary arrangements with alpha-helix (40.32%), extended strand (19.97%), beta-turn (5.70%), and random coil (33.99%) were observed. The increasing order of occurrence is as follows: alpha helix>random coil>extended strand> beta-turn in the protein. CFSSP result indicates that the proteins have alpha helix (71.88%), sheets (62.71%), and turns (13.97%). From the result, the alpha-helical conformation is high, which indicates that the secondary structure is more stable. Moreover, the secondary structure map of uricase was predicted using the PSIPRED protein analysis tool and shown in Figure 4.4. It was observed that uricase is mainly formed by helix structures and β-sheets. No disordered protein binding sites were discovered. The predicted

secondary structure of the amino acid sequences of uricase from *Bacillus simplex* (WP_063232385.1) is illustrated extensively (Appendix II). Similar secondary structure analysis of various enzymes have been performed and reported in prior studies (Rahmatabadi et al. 2017; Rani and Pooja 2018).

4.2.7 Phyre 2 structured modeling analysis

Uricase from *Bacillus simplex* (WP_063232385.1), which is the selected protein for the β sheet model, was constructed by the Phyre 2 tool (Kelley et al. 2015) and showed model dimensions of (Å): X:100.942 Y:71.576 Z=65.228 (Appendix II). Based on the BLAST results, clj2gC was selected as the best template for protein modeling. PDB input was the crystal structure of urate oxidase from *Bacillus sp*.tb-90 co-2 crystallized with 8-azaxanthine. The predicted secondary structure of uricase had 97% of residues modeled at >90% confidence obtained using the intensive modeling mode suggested by the server itself. From the observations of secondary structure, the analysis revealed that the chosen uricase enzyme has 13% disordered, 34% alpha-helix, and 23% beta strand regions, as shown in Figure 4.5.

4.2.8 Tertiary structure analysis

The 3D structure of *Bacillus simplex* uricase (WP_063232385.1) was modeled through the SWISS-MODEL server using the most suitable template. The 3D structure of *Bacillus simplex* uricase was not available in PDB. The alignment of the selected template (1j2g.1.A) with the target sequence of *Bacillus simplex* (WP_063232385.1) is illustrated (Appendix II), and three-dimensional protein models are shown in Figure 4.6. The best-matched template ((1j2g.1) was a crystal structure of urate oxidase from *Bacillus sp*. TB-90 co-crystallized with 8-azaxanthine (AZA). It has been identified that 64.24%, similarity 0.49, four AZA ligands, a resolution of 2.20A determined by X-ray diffraction method and oligo state of homo-tetramer, were found by BLAST. The representative species of *Bacillus* to illuminate the protein structure of *Bacillus simplex* uricase was chosen as per the QMEAN score (Benkert et al. 2011; Berman et al. 2000; Pramanik et al. 2017). The QMEAN score of the built model was -0.75, which indicates the global quality of the entire model (Figure 4.7). The Z-score of the predicted uricase was -6.46, which indicates the absolute quality of

the model based on protein structures determined by X-ray crystallography as reference. Generally, a good quality modeled 3D structure is defined by QMEAN values close to 0 and Z-score value <1.
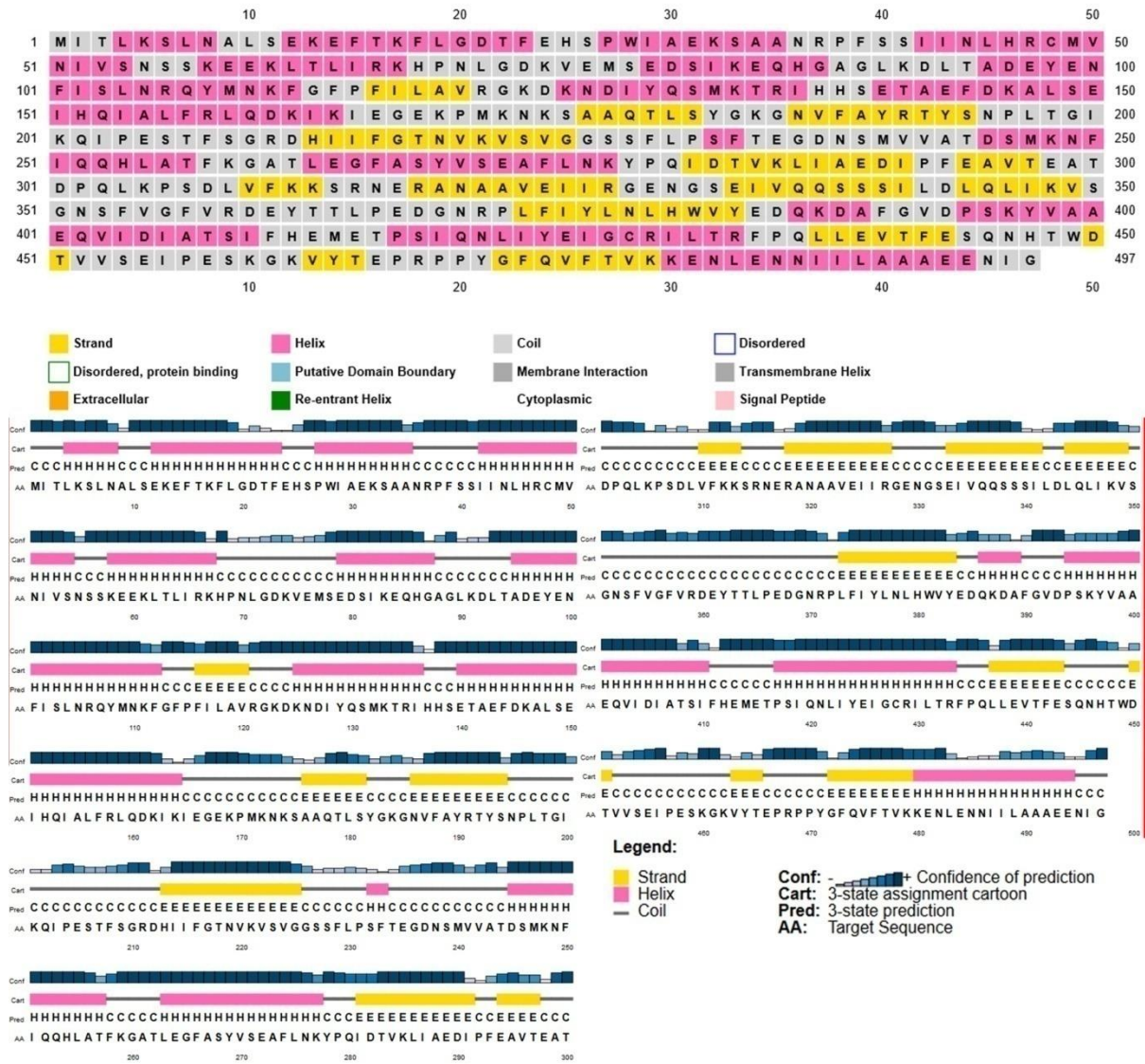
Figure 4.4: Secondary structure analysis of uricase from *Bacillus simplex* (WP_063232385.1) as revealed by PSIPRED map
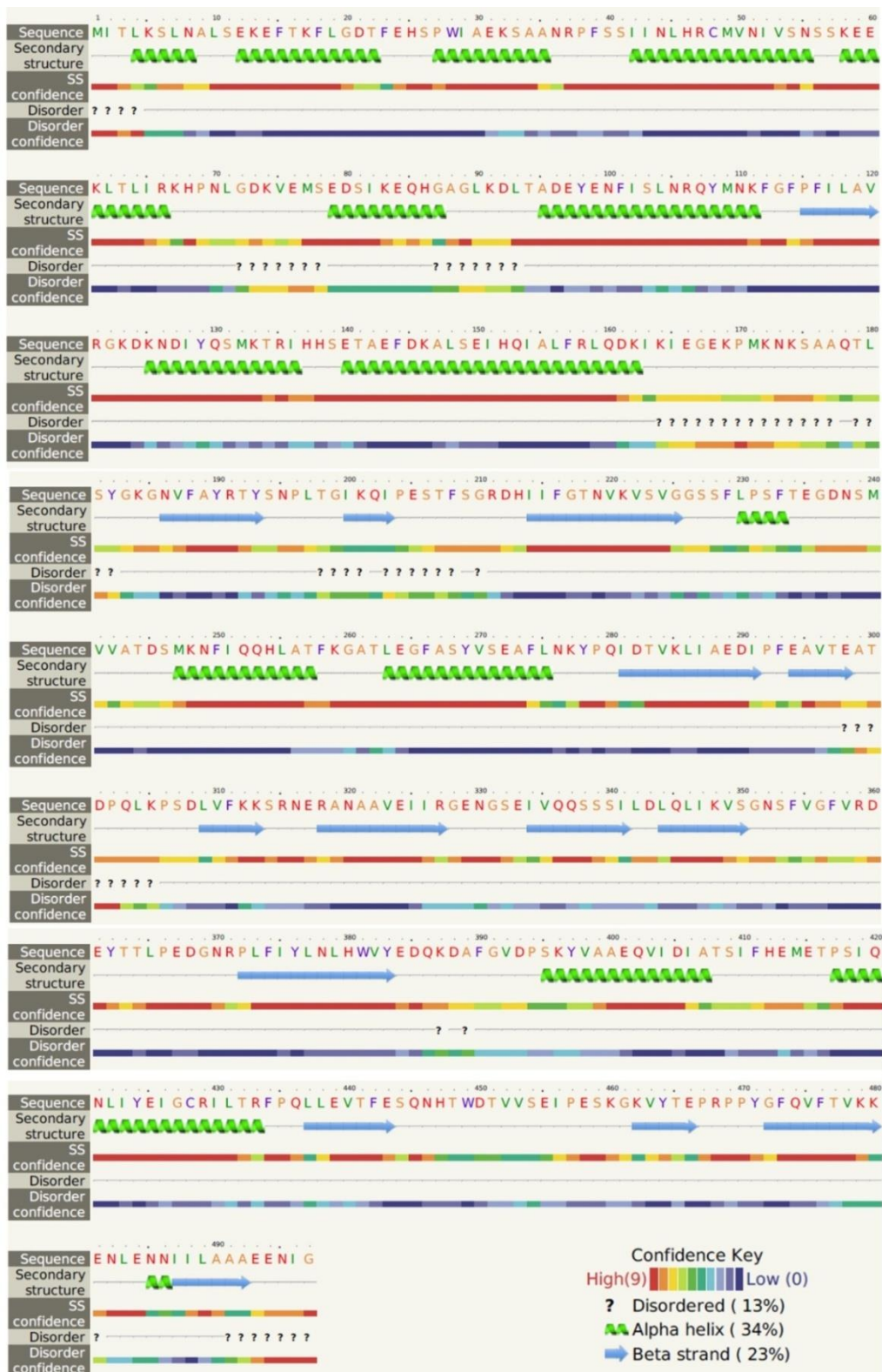
Figure 4.5: Secondary structure and disorder prediction of selected uricase *Bacillus simplex* (WP_063232385.1) from Pyre2 server
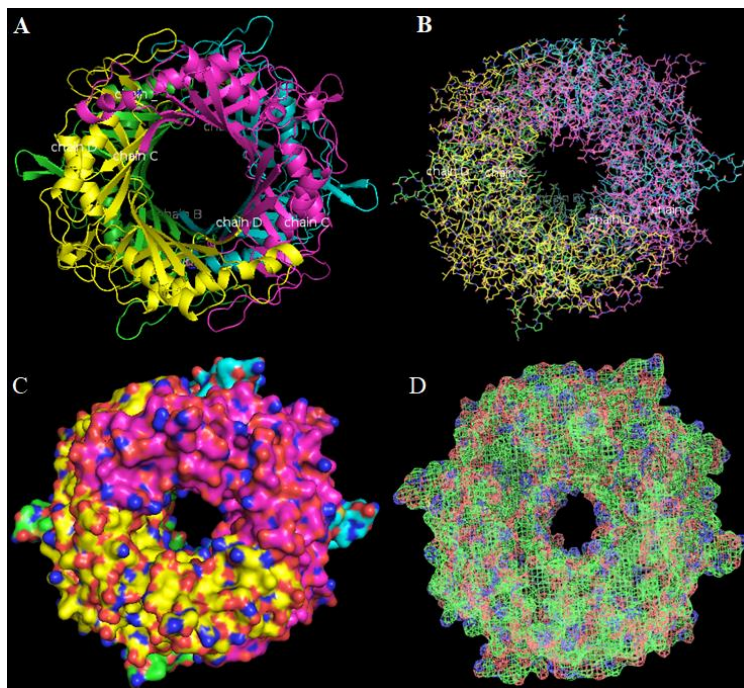
Figure 4.6: Predicted 3D model structures of uricase protein of *Bacillus simplex* (WP_063232385.1) with different views showed by PyMol. (A) and (B) showing four distinct chains of the protein (C) Surface view (D) Mess view
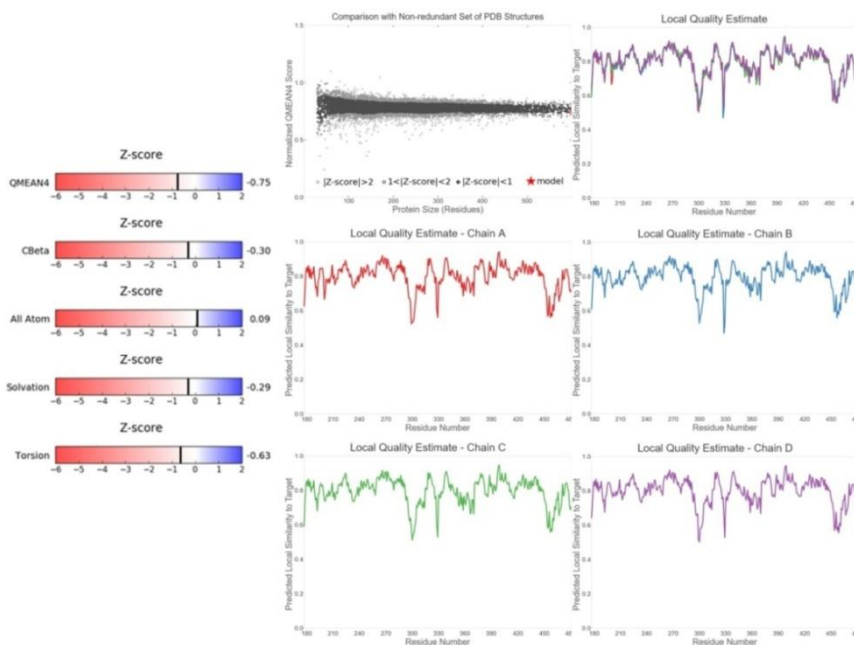
Figure 4.7: Quality analysis of the built protein model for *Bacillus simplex* (WP_063232385.1) from QMEAN server
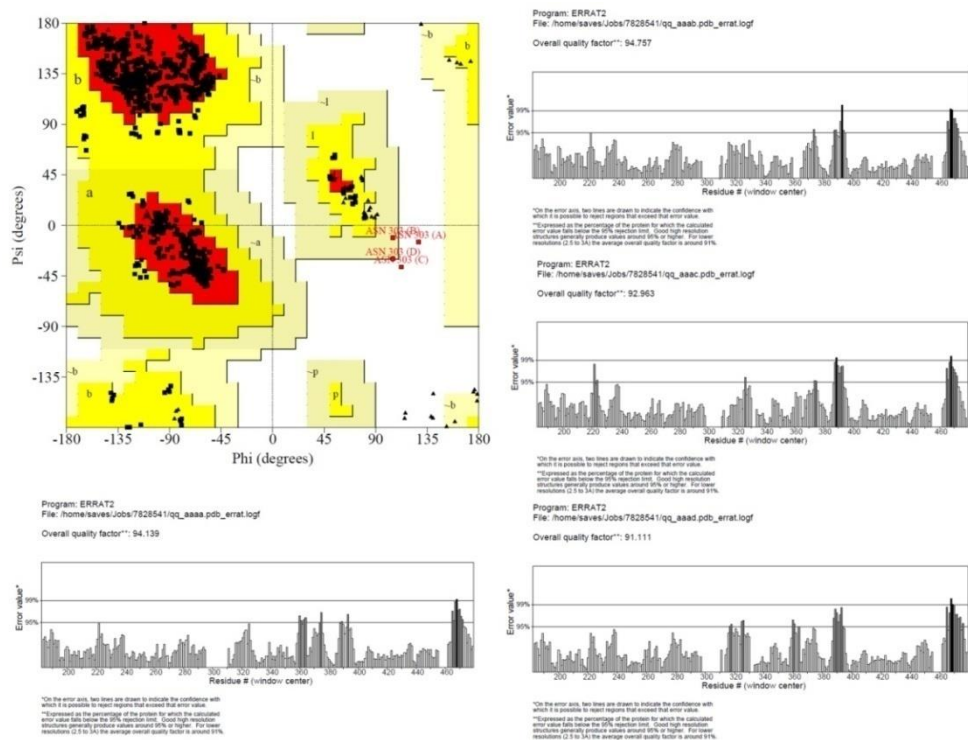
Figure 4.8: Validation of modeled uricase protein of *Bacillus simplex* (WP_063232385.1) from SAVES server (Ramachandran plot and ERRAT).

The results from Ramachandran plot analysis showed that the favorable region consists of 96.9% residues, 2.4% of residues in the allowed region, and the remaining 0.7% of residues resided in the outlier region (Figure 4.8). Generally, the best model has greater than 90% of residues in the favorable region of the Ramachandran plot. The quality assessment of the uricase protein model was evaluated using the SAVES server. 81.64% of the residues had an average 3D-1D score of >=0.2, as observed from Verify 3D. The modeled protein was acceptable as at least 80% of the amino acids had scored >=0.2 in the 3D/1D profile (Appendix II). The highest overall quality factor value of 94.64 was observed in chain B of the selected uricase protein model using SAVES ERRAT (Figure 4.7). The quality of the generated model protein was also analyzed by the PROQ server, which is based on LGscore and MaxSub scores. This analysis showed that the predicted LGscore and MaxSub score was 6.686 and 0.579, respectively (Appendix II). The criteria for a good model are

LGscore>4 and MaxSub>0.8. Predictions 3D protein models built by *in silico* homology modeling were performed by various authors (Beedkar et al. 2012; Pramanik et al. 2017, 2018; Rani and Pooja 2018; Zobayer and Hossain 2018).

4.2.9 Functional analysis

All selected species of *Bacillus* uricase displayed soluble nature based on the result obtained from the SOSUI web server, which is documented (Appendix II). The formation of disulfide bonds in proteins due to oxidation of thiol groups of cysteine residues is important to provide thermostability to the protein (Tamboli et al. 2015). The prediction of SS-bond states of cysteines and locating the disulfide bridges in *Bacillus* uricases were performed by the CYS_REC tool. The results indicated that all sequences of uricase having cysteine residues between 1-4 were not SS-bonded, and this may obstruct the stability of these sequences (Appendix II). The protein sequence of PDB 5AYJ, which is a hyper thermostable mutant of *Bacillus* sp. TB-90 urate oxidase showed two cysteines that were found at residue positions 297 and 304. CYS 297 is probably SS-bonded, whereas CYS 304 is probably not SS-bonded. This is the only sequence out of 70 species of uricase that was found to have disulfide bonds. The chosen uricase of *Bacillus simplex* showed two cysteines in the 48th and 428th positions which were not SS-bonded. In addition, the functional study identified three functional motifs in the selected protein sequence: OHCU decarboxylase, uricase, and DUF2383 (Figure 4.9). In the protein-protein interaction network, a total of ten predicted functional partners were observed in the *Bacillus simplex* uricase using the STRING tool (Appendix II).

Pfam (3 motifs)		
Pfam	**Position (Independent E-value)**	**Description**
OHCU_decarbox	8..160(4.9e-41)	PF09349, OHCU decarboxylase
Uricase	182..313(3.2e-15) 338.472(6.9e-12)	PF1014, Uricase
DUF2383	122..180(0.078)	PF09537, Domain of unknown function
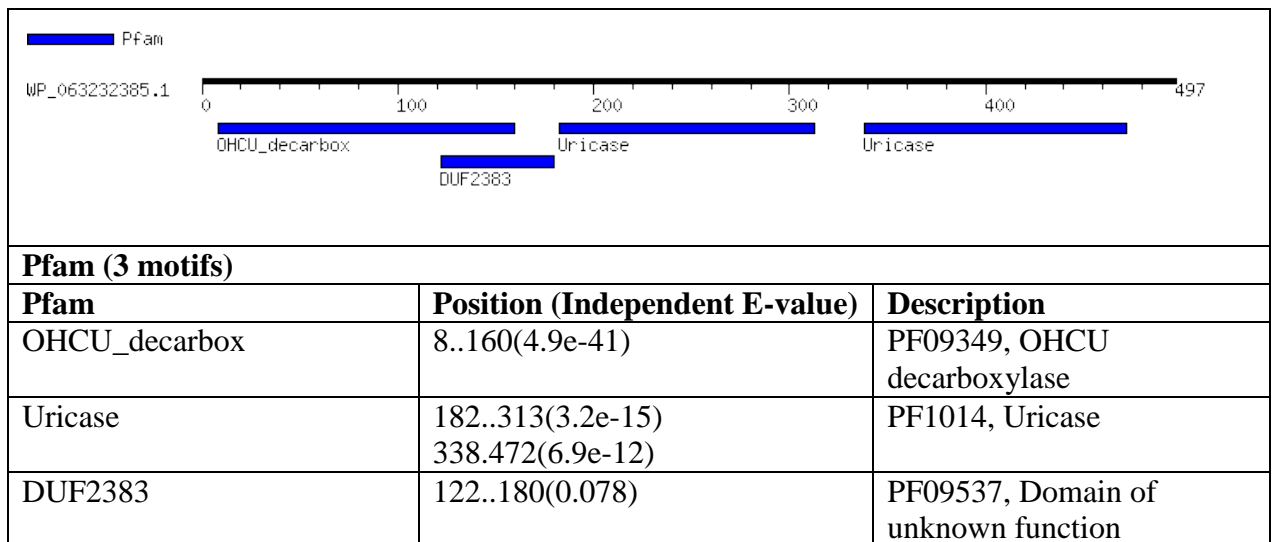
Figure 4.9: Motif finder tool result showing three functional motifs and their positions for the uricase from *Bacillus simplex* (WP_063232385.1)
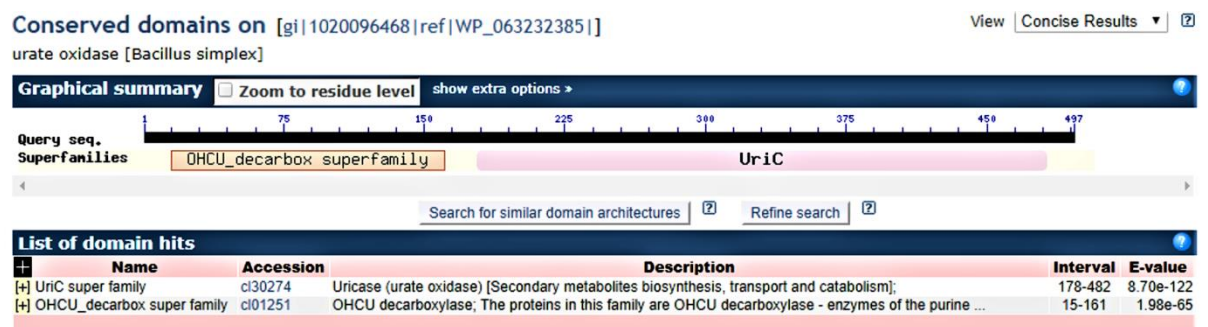
Figure 4.10: Conserved domain database (CDD) search result of uricase *Bacillus simplex* (WP_063232385.1) showing two types of domains

The *Bacillus simplex* uricase protein has mainly two types of conserved domains that belong to the two protein superfamilies: Uricase and OHCU decarboxylase superfamily. OHCU decarboxylase is an enzyme involved in purine catabolism that catalyzes the breakdown of OHCU into S(+)-allantoin (third step). The first step is catalyzed by urate oxidase, and the second step is catalyzed by HIUases, as shown in Figure 4.10. *In silico* hydrolysis of selected uricase protein, digestive enzymes such as proteinase K, pepsin, chymotrypsin, and thermolysin were performed. The results

from Peptide Cutter identified the average number of cutting sites as 260.72 for proteinase K, 248.53 for pepsin (pH>2), 231.59 for chymotrypsin-low specificity, and 246.72 for thermolysin. Computational-based functional analysis of several microbial proteins ha*ve* been studied by various authors (Dutta et al. 2018; Pramanik et al. 2018; Tamboli et al. 2015; Verma et al. 2016).

Table 4.4: The list of Softwares/Databases used

Name of the Softwares/Databases	Input	Output
NCBI database	Protein name	Protein sequences
ClustalW (MEGA7 Software)	Protein sequence	Multiple sequence alignment for conserved amino acid residues
Neighbor-joining statistical method (MEGA7 Software)	Protein sequence	Phylogenetic tree for the evolutionary history
Multiple EM for Motif Elicitation (MEME)	Protein sequence	Conserved motif identification
Pfam database	Sequence motifs	Motif family identification
ExPASyProtParam tool	Protein sequence	Physicochemical characterization and amino acid composition
Self-Optimized Prediction Method with Alignment (SOPMA)	Protein sequence	Secondary structure prediction of protein
PSIPRED v3.3 tool	Protein sequence	Secondary structure prediction of protein
CFSSP server (Chou and Fasman Secondary Structure Prediction)	Protein sequence	Secondary structure prediction of protein
Phyre2 (Protein Homology/AnalogY Recognition Engine V 2.0)	Protein sequence	Secondary structure of the protein modeling and prediction
SWISS-MODEL server	Protein sequence	Modeling of three-dimensional (3D) protein structure
SAVES (Structure Analysis and Verification Server v5.0)	Predicted 3D protein structure	The quality of the predicted model evaluated and verified
ERRAT	Predicted 3D protein structure	The highest overall quality factor of the predicted model evaluated and verified
Verify 3D	Predicted 3D	The quality of the predicted

	protein structure	model evaluated and verified
Ramachandran plot	Predicted 3D protein structure	The quality of the predicted model evaluated and verified
PROQ server	Predicted 3D protein structure	The quality of the predicted model evaluated and verified
CYS_REC tool	Protein sequence	Prediction of the SS-bonding States of Cysteines in Protein Sequences
Motif search tool	Protein sequence/3D protein structure	Determination of the functional motifs and the superfamily of the protein
The Conserved Domain Database	Protein sequence	Identification of the conserved domains of the protein
STRING v11.0	Protein sequence	Networks functional enrichment analysis of Protein-Protein interaction
SOSUI tool	Protein sequence	Determines whether the protein is soluble or transmembrane
Peptide Cutter tool	Protein sequence	Provides information about the predicted protease cleavage sites and sites cleaved by chemicals in the protein sequence

The pI values of the selected uricase sequences of *Bacillus* species lie between 4.9 to 6.25, indicating that the nature of proteins varies from acidic to a neutral environment. The instability index for all the *Bacillus* species was between 23.01 to 45.63, indicating that most of the selected uricase proteins are stable in nature. The Ai of selected sequences varied between 70.92-85.67, suggesting that the proteins are thermostable in nature. *Bacillus* species remain the primary bacterial workhorses in microbial fermentation. Certain *Bacillus* species are GRAS (generally recognized as safe) according to the Food and Drug Administration. *Bacillus* strains with the capacity to produce and secrete large amounts (20–25 g/L) of enzymes have risen to prominence as industrial enzyme producers (Barros et al. 2013; Schallmey et al. 2004). *Bacillus fastidious* is one of the organisms in the total of seventy *Bacillus* species of uricase *Bacillus fastidious* (4R8X_A) has a sequence length of 322, molecular weight of 37 kDa, pI of 4.99, -R of 46, +R of 31, EC of

34,840, Ii of 24.16, Ai of 79.56, and GRAVY of -0.271. Uricase from *Bacillus fastidious* was commercialized by Sigma-Aldrich (product 94310, 9 U/mg) and has been used for various applications.

An intracellular uricase from *Bacillus fastidious* with high catalytic capacity (Zhao et al. 2009). The intracellular uricase is more stable in gastrointestinal system (Handayani et al. 2018). The use of intracellular enzymes for analytical and medical purposes is becoming more common (Aly et al. 2013). *Bacillus fastidious* was first isolated in 1929 by den Doorn de Jong and described as an aerobic, rod shaped organism. *Bacillus fastidious* has attracted little attention during the past four decades. Mahler et al 1970 employed urate oxidase for analytical application, using strains of *Bacillus fastidious* isolated by soil enrichment with uric acid (Mahler 1970). Single polypeptides are found in the majority of bacterial uricases. *Bacillus fastidious* SMG 83 has two polypeptides. Uricase from *Bacillus fastidious* may be made up of four identical subunits based on the link between the molecular weights of active uricases and the polypeptides they contained. Other microbial uricases were found to include two distinct polypeptides. The variation in the composition of *Bacillus fastidious* uricases could be attributed to differences in the biological properties of different strains (Zhao et al. 2006).

Due to the incidence of hypersensitivity reactions during therapy, the uricase formulation displayed an immunogenic response, resulting in poor patient compliance. A significant hypersensitive reaction was seen in bacterial uricase (anaphylaxis, hemolysis, methemoglobinemia). The clinical utilization of uricase against gout is limited due to immunogenicity. Uricase from *Bacillus fastidious* was chosen further to decrease the immunogenicity by computational approaches. Along with *Bacillus fastidious*, uricase sourced from *Arthrobacter globiformis* also chosen to reduce the immunogenicity. Uricase from *Arthrobacter globiformis* was also commercialized by Sigma-Aldrich (product U7128, 15-30 units/mg protein) and has been used for various applications. Because immunogenic and allergic reactions compromise uricase efficacy and safety, the scientific community primary goal is to develop an effective replacement for this biopharmaceutical.

4.3 SUMMARY

Uricase has pharmaceutical importance as a biodrug for the treatment of acute hyperuricemia and refractory gout. Currently, it is being used in clinical laboratories for diagnostic purposes to quantify uric acid concentration. In the present work, attempts have been made to provide a complete description of the structural and functional aspects of various *Bacillus* species having uricase activity using bio-computational web-based servers and tools. The evolutionary relationships among uricases of various species have been evaluated using multiple sequence analysis and phylogenetic tree construction. Phylogenetic analysis was performed, and it was revealed that the amino acid and cDNA sequences of *Bacillus simplex* uricase are closely related to *Bacillus sp*.BA3. The selected *Bacillus* uricase proteins are active within an acidic to a neutral environment, and it is thermally stable with molecular mass ranging from 35.59-59.85 kDa, which was determined by *in silico* physicochemical analysis of the protein sequences. The prediction of secondary structure was performed by CFSSP and PSIPRED, which showed that uricase is rich in alpha-helices and sheets. All the selected proteins have been subjected to functional analysis using CYS_REC, STRING server analysis, and PeptideCutter tool. The CDD tool identified two conserved domains of the Uricase and OHCU decarboxylase superfamily. Also, the motif search tool revealed that OHCU decarboxylase, uricase, and DUF2383 were three functional motifs. The tertiary structure model of the *Bacillus simplex* (WP_063232385.1) uricase protein was predicted and validated. The quality estimation was done as a cross-evaluation for the predicted uricase protein using various servers, and this exhibited a high overall quality factor score of 94.64. Hence, the present study may be helpful in the field of computational proteomics to get a better understanding of the uricase protein. This investigation would be useful to future researchers to conduct wet-lab studies regarding the structure, function, isolation, and characterization of *Bacillus* uricase enzyme for potential industrial applications.

CHAPTER 5

IN-SILICO EPITOPE IDENTIFICATION AND DESIGN OF URICASE MUTEIN WITH REDUCED IMMUNOGENICITY

Bacteria are the major source of therapeutic enzymes, and therapeutic enzymes can also be obtained from various biological sources like organs, tissues, animal fluids, and genetically modified organisms and cells (Rodríguez et al. 2014; Valderrama-Rincon et al. 2012). During systematic administration of bacterial enzymes, the body recognizes them as a foreign antigen, and this leads to the secretion of antibodies. Antibody secretion of B-cells is mainly governed by the identification of antigenic epitopes on the surface of bacterial enzymes. Therefore, the use of bacterial enzymes is limited due to their immunogenicity, poor stability, and toxicity (De Duve 1966; Dean et al. 2017). Moreover, recent uses of therapeutic enzymes are associated with common problems such as high degradation rates or rapid clearance (Mumtaz and Bachhawat 1992).

There have been numerous attempts to treat gout and other hyperuricemia-related diseases through the systematic administration of uricase extracted from various sources (London and Hudson 1957). The first recombinant form of uricase from *Aspergillus flavus* is Rasburicase. Due to its high immunogenicity and short half-life, Rasburicase therapy is stated to be limited (Bayol et al. 2002; Coiffier et al. 2003; Garay et al. 2012; Nuki 2012; Szczurek et al. 2017; Tan et al. 2012). Additionally, the therapeutic potential of recombinant uricase for the treatment of gout is associated with pharmacologic tolerance and potency problems (Baraf et al. 2008; Guttmann et al. 2017). Therefore, it is important to reduce the immunogenicity of uricase as a protein-drug to cure treatment-resistant gout. The combination of polyethylene glycol with uricase was reported as the first clinical study to successfully reduce plasma uric acid concentration over 32 hours (Davis et al. 1981; Sherman et al. 2008). However, due to several limitations of PEGylation of therapeutic enzymes, the identification of

hot spot B-cell and T-cell epitopic residues is crucial for the preparation of uricase mutein, which can be easily administered in the human body without immunological effect. Presently, the uricase used for the therapeutic purpose (Punnappuzha et al. 2014; Tan et al. 2012; Zhao et al. 2006) is mainly sourced from *Arthrobacter globiformis* and *Bacillus fastidious* due to high specific activity. Therefore, we aim to identify the epitopic regions and decrease the immunogenicity of uricase from the above-mentioned species. The experimental evolution of B-cell and T-cell epitopes of therapeutic proteins are limited because most of the approaches are expensive, time-consuming, and labor-intensive (Potocnakova et al. 2016). Therefore, the widely accepted algorithms and tools of bioinformatics are highly recommended, which can reduce cost by predicting B-cell and T-cell epitopes from the amino acid sequence of uricase (Kolaskar and Tongaonkar 1990; Saha and Raghava 2006; Singh et al. 2013).

In the present study, we aim to identify the linear B-cell, conformational B-cell, and MHC-I-based T-cell epitopes to reduce the immunogenicity of uricase sourced from *Arthrobacter globiformis* (*Ag*-Uricase) and *Bacillus fastidious* (*Bf*-Uricase). Multiple sequence alignment (MSA) was performed to detect the conserved and identical residues of the uricase from different sources. Motifs and domains of uricase from various sources were also identified to describe the structural, functional aspects of this protein in the evolutionary process. Emini surface accessibility, Parker hydrophilicity, and Karplus & Schulz flexibility methods were employed to detect the continuous B-cell epitopes and corresponding hot-spot residues. Similarly, the deimmunization method was used to identify T-cell epitopes. Next, the hot-spot residues were mutated to reduce the antigenic character of the identified epitopes. Lastly, the impact of mutagenesis on the catalytic activity and the structural stability of uricase was assessed by molecular docking, free energy calculations, and molecular dynamics simulation. To the best of our knowledge, this *in-silico* study to reduce the immunogenicity of bacterial uricase is presented here for the first time.

5.1 METHODS AND COMPUTATIONAL DETAILS

The overall computational methodology which was used for obtaining Uricase mutein with reduced antigenicity is illustrated in Figure 5.1
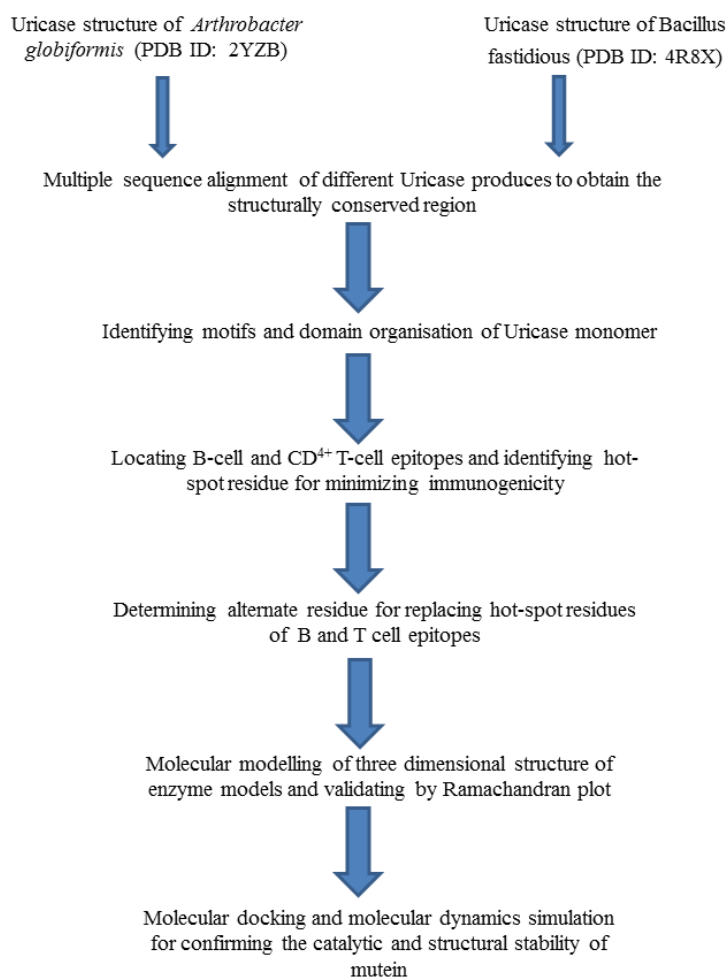
Figure 5.1: Flow chart showing the methodology employed for generating enzyme models to diminished immunogenicity of uricase through *in silico* approaches.

5.1.1 Uricase sequences retrieval

The amino acid sequences of uricase were chosen from the following thirteen mass producers, *Drosophila melanogaster* (Friedman and Barker 1982), *Oryctolagus cuniculus* (Oda et al. 2002), *Rattus norvegicus* (Ito et al. 1991), *Mus musculus* (Lee et al. 2006), *Cavia porcellus* (Fujiwara et al. 1987), *Papio hamadryas* (Xiong et al.

2013), *Bacillus fastidious* (Zhao et al. 2006), *Arthrobacter globiformis* (Suzuki et al. 2004), *Camelus dromedarius* (Osman et al. 1989), *Chlamydomonas reinhardtii* (Alamillo et al. 1991), *Aspergillus flavus* (Leplatois et al. 1992), *Phaseolus vulgaris* (Papadopoulou et al. 1995) and *Cyberlindnera jadinii* (Adámek et al. 1990). The amino acid sequences of uricase obtained from the above-mentioned prokaryotic and eukaryotic producers were subjected to sequence similarity to understand the conservation and evolutionary relatedness of the taxa. The full-length FASTA sequences of the above-mentioned uricase were collected from the National center for biotechnology information database (NCBI) (https://www.ncbi.nlm.nih.gov/).

5.1.2 Multiple sequence alignment and phylogenetic comparison

Multiple sequence alignment (MSA) of all the selected amino acid sequences of uricase from different habitats were performed by ClustalW tool (Thompson et al. 1994) of MEGA (Molecular Evolutionary Genetics Analysis, V-7.0) software (Kumar et al. 2001, 2008, 2016) to identify the conserved residues of uricase throughout the process of evolution. ClustalW is a widely used matrix-based algorithm that implements progressive alignment methods (Chatzou et al. 2016) to align the multiple proteins, DNA, or RNA sequences from different sources. The parameters used for MSA include gap opening penalty of 10, gap extension penalty of 0.2, gonnect protein weight matrix, and gap distance separation penalty of 5 with no end gap separation. The evolutionary relationship of uricase was determined by constructing the phylogenetic tree of all thirteen sequences employing the maximum parsimony statistical method (Mount 2008). The topologies of the phylogenetic tree were evaluated by applying 1000 bootstrap replicas (Kumar et al. 2008, 2016).

5.1.3 Motif analysis

Multiple EM for Motif Elicitation (MEME) (http://meme-suite.org/tools/meme) is a widely used tool for discovering motifs in a set of related DNA/RNA or protein sequences (Bailey et al. 2009, 2015). In proteins, a motif may possibly relate to the enzyme active site or structural unit required for correct folding. Sequence motifs are therefore known as the essential functional units for molecular evolution. The

identified motifs and their locations recovered by MEME elucidate the conserved regions associated with structural and functional properties of uricase in the evolution process. The starting and ending point of the motifs were displayed as blocks.

Pfam (El-Gebali et al. 2019), a web-based tool, was used for accurate classification of protein families and domains using HMM (hidden Markov model). All the uricase sequences were submitted in the form of accession numbers to analyze the domain organization associated with uricase (Finn et al. 2014; Sammut et al. 2008).

5.1.4 Antigenic epitopes prediction

The immune epitope database (http://www.iedb.org/) is a standard and organized database with a large collection of experimentally characterized immune epitopes (Kim et al. 2012). The amino acid sequences of *Ag*-Uricase and *Bf*-Uricase were retrieved from NCBI and submitted in FASTA format to immune epitope database and analysis resource (IEDB-AR) tool (Zhang et al. 2008) for predicting and analyzing both B-cell and T-cell peptide epitopes.

Surface accessibility, hydrophilicity, and mobility are considered critical criteria for assessing the antigenicity of any protein or peptide (Parker et al. 1986). Hence, the continuous B-cell epitopes were identified based on Emini surface accessibility, Parker Hydrophilicity, Karplus & Schulz Flexibility prediction methods in the IEDB database (Emini et al. 1985; Karplus and Schulz 1985; Larsen et al. 2006; Parker et al. 1986). Thresholds of 1.00, 1.63, and 0.996 were employed for surface accessibility, hydrophilicity, and mobility, respectively, in the above-mentioned methods to determine antigenicity. The FASTA format of *Ag*-Uricase and *Bf*-Uricase sequences were imported to each epitope prediction panel and submitted for predicting B-cell epitopes. All the predicted epitopes where ranked according to their corresponding antigenic scores. The highly immunogenic amino-acid residue inside each epitope peptide was also documented. B-cell linear epitope prediction methods anticipate the immune response according to the characteristics of the amino-acid sequence of the antigen using amino acid scales (AASs) and Hidden Markov Models (HMMs) (Moutaftsi et al. 2006; Peters et al. 2006; Yu et al. 2002). The conformational B-cell

epitopes of *Ag*-Uricase and *Bf*-Uricase were identified using the Discotope tool (V-2.0) (Kringelum et al. 2012). The server predicts B-cell epitopes based on the spatial information, surface accessibility, and amino acid statistics of discontinuous epitopes identified from the crystal structure of the antigen-antibody complex (Haste Andersen et al. 2006). The 3D structure of uricase from both species was imported to the conformational B-cell epitope prediction panel, and a threshold of -3.7 with 17% sensitivity and 95% specificity was employed.

Next, both *Ag*-Uricase and *Bf*-Uricase protein sequences were screened to identify T-cell epitopes using the deimmunization method. The human leukocyte antigen (HLA)-B*5801 allele that is strongly associated with hyperuricemia and gout was chosen for predicting MHC-II based T-cell epitope (Ko et al. 2015; Park et al. 2015).

5.1.5 *In-silico* mutagenesis

The crystal structure of *Ag*-Uricase (PDB ID: 2YZB, resolution 1.9Å) (Juan et al. 2008) and *Bf*-Uricase (PDB ID: 4R8X, resolution 1.401 Å) (Feng et al. 2015) was retrieved from protein data bank (PDB) for modeling studies. Both crystal structures of uricase were prepared using protein preparation workflow (Sastry et al. 2013) in Maestro. The missing hydrogen atoms were added to both structures. It was reported that the functional or active form of uricase can exist as a homotetramer (Kratzer et al. 2014). Therefore, 2YZB was kept in tetrameric form by deleting the extra chains to reduce the size, making it comparable with 4R8X. The uricase activity is found to be optimal at a pH of 9.0 (Juan et al. 2008; Tan et al. 2012). Hence, the protonation state of the amino acid residues of both 2YZB and 4R8X were optimized at a pH of 9.0. The orientations of the hydroxyl group of Asn and Gln residues were also optimized for both crystal structures of uricase. Next, the structures of both proteins were minimized using the OPLS-2005 force field (Shivakumar et al. 2010; William L. Jorgensen et al. 1996) with RMSD (protein heavy atoms) convergence criteria of 0.30 Å.

The hot-spot amino acids for *in-silico* mutagenesis were chosen based on the score obtained from B and T cell epitopes prediction of uricase from both bacterial sources.

In-silico site-directed mutagenesis (SDM) was performed on both optimized structures of 2YZB and 4R8X, using Pymol software (v 1.6). The obtained mutant proteins were validated using the Ramachandran plot and I-MUTANT (Capriotti et al. 2005) web-server. The structures were further used for molecular docking to determine the impact of side-directed mutagenesis on their structural and catalytic aspects.

5.1.6 Ligand preparation

The 3D structure of the uric acid was constructed using the builder panel in Maestro (v-11.7.011, Schrödinger, LLC, New York, 2018). The possible ionization state was generated, and partial charges were assigned at a pH of 9.0 prior to docking. The geometry of the structure was optimized, and its energy was minimized using OPLS-2005 force-field (Shivakumar et al. 2010; William L. Jorgensen et al. 1996) in the Ligprep module (Schrödinger Release 2018-3: LigPrep, Schrödinger, LLC, New York, 2018). The resulting structure was considered for further modeling studies.

5.1.7 Molecular docking

The molecular docking approach may be used to simulate the atomic level interaction between a small molecule and a protein, allowing us to define the behavior of small molecules at the binding region of target proteins as well as elucidate essential biochemical processes. Docking is a two-step method that begins with the prediction of the ligand conformation as well as its position and orientation within these sites (often referred to as pose) and ends with the determination of the binding affinity. Prior to docking procedures, knowing the location of the binding site considerably improves docking efficiency. Often, the binding site is known prior to docking ligands into it. Additionally, information about the sites can be obtained by comparing the target protein to a family of proteins with comparable functions or to proteins co-crystallized with other ligands. When no binding sites are known, cavity detection tools or web services can be used to find probable active sites within proteins. Blind docking is the process of docking without making any assumptions about the binding location (Meng et al. 2011).

Glide (Friesner et al. 2006) (Schrödinger Release 2018-3: Glide, Schrödinger, LLC, New York, 2018) was used to perform docking (Extra Precision mode) of uric acid at the active sites of both 2YZB and 4R8X. Glide uses a hierarchical array of filters to investigate possible ligand locations at the catalytic pocket of uricase (Friesner et al. 2004). The geometry of the uric acid was kept in flexible mode while the receptor was depicted as rigid. The receptor grid was generated with a partial charge cut-off of 0.25e, and the van- der -walls scaling factor was kept at 1.00. The active site residues of 2YZB, i.e., Asn249, Gln223, Leu222, Arg180, Phe163 (Chain A), and Asp68-Ala66 (Chain D) were selected to generate the grid box suitable for accommodating uric acid, and the grid center was placed at the centroid of the interacting amino acids (Juan et al. 2008). Since the catalytic pocket residues of 4R8X are unclear, the amino acids located at the interface of two identical subunits, such as Phe179, Ala193, Arg196, Ile244, Gln244, Asn271, Gln299 (Chain C), and Ala68-Asp70 (Chain A) (Feng et al. 2015; Kratzer et al. 2014) were selected to build the grid box. The binding affinity of uric acid towards both the wild and mutated uricase was calculated by MM/GBSA method (Genheden and Ryde 2015). The detailed methodology of MM/GBSA calculation is given in Appendix III. The accuracy of docking was assessed by measuring the RMSD (RMSD = 0.16Å) between the co-crystal and redocked position of uric acid at the catalytic pocket (Figure A.III.4-A) of uricase. The binding pose of uric acid at the active sight of uricase was further confirmed by molecular dynamics simulation.

5.1.8 The MD protocol

The native, mutated form of both tetrameric 2YZB and 4R8X (in association with substrate uric acid) was subjected to atomic molecular dynamic simulation in order to compare their conformational stability under motion. Simulation with all the subsequent calculations was carried out using the Desmond (Bowers et al. 2006) package, and Maestro GUI was used for visualization. OPLS-2005 (Shivakumar et al. 2010; William L. Jorgensen et al. 1996) force field was used to generate the necessary parameters required for energy minimization and MD simulations of 2YZB and 4R8X. All four structures of uricase (including both mutated and normal from the two

aforementioned species) were solvated separately in orthorhombic periodic unit with SPC (simple-point charge) water molecules (Jorgensen et al. 1983). The resulting systems were neutralized by adding counter ions. In addition, 0.15 M NaCl was added to imitate physiological conditions. Next, the systems were minimized under the steepest descent algorithm (Averill and Painter 1992) with a maximum of 2000 iterations until a gradient threshold of 25 kcal/mol/Å is reached.

All the systems of solvated uricase (mutated and wild structure of both 2YZB and 4R8X) were initially heated up at 300K for 1 ns and subsequently equilibrated under canonical (NVT) ensemble for 4 ns. Next, the system condition was changed from NVT to isothermal-isobaric (NPT) ensemble at a temperature of 300K for 5 ns each in order to equilibrate the pressure at 1 atm. During each equilibration step, protein-ligand heavy atoms were restrained. The temperature and pressure of the system were controlled, respectively, using the Noose-Hoover thermostat (Martyna et al. 1992) and Martyna-Tobias-Klein barostat (Martyna et al. 1994). Temperature and pressure relaxation time of 2ps was assigned throughout the equilibration time. Lastly, all protein-ligand complexes were subjected to production simulation for 100ns with a time step of 2 fs. The restrains on solute heavy atoms were removed and allowed to move freely throughout the production run. For long-range electrostatic interactions, smooth Particle Mesh Ewald (PME) method (Essmann et al. 1995) was used with a tolerance of $1 \times e^{-9}$, and for short-range electrostatic interactions, a cut-off radius of 9.0Å was applied. A multiple-time step RESPA (Reversible Reference System Propagator Algorithm) integrator algorithm was employed throughout with a time step of 2 fs for bonded, 2 fs for 'near' bonded, and 6 fs for 'far' non bonded interactions. The trajectories of the solute atoms of all four solvated protein-ligand complexes were retrieved at each 20 ps interval for analyzing the data. The conformational stability of uricase from both species was assessed by calculating the time evolution of protein backbone RMSD, ligand RMSD, residue-wise RMSF, and radius of gyration. Change in binding free energy during the course of the simulation was also performed using MM/GBSA calculation using Prime (Genheden and Ryde 2015). The detailed procedure for calculating the above-mentioned quantities is documented in Appendix III.

5.2 RESULTS AND DISCUSSION

5.2.1 Multiple sequence alignment and phylogenetic analysis

The multiple sequence alignment of thirteen uricase from different sources was carried out by the ClustalW approach. It was reported that the conserved residues are important to explain the structural and functional aspects of uricase. As shown in Figure 5.2 and Table A.III.1, most of the amino acids are conserved between uricases from mammalian sources (78.3-94.6 % sequence identity) compared to those from other sources (21.3-48.2% sequence identity). For example, the bacterial uricases from *A.globiformis* and *B.fastidious* showed only 25.26% identity. In the mutagenesis process, it is advised to substitute residues outside the conserved region in order to preserve the structural and functional characteristics of the therapeutic drug (Sun et al. 2011).

Uricase has a variety of metabolic activities that vary depending on the host organism. A cross-reaction exits between different species of uricases, possessing similar molecular weight, same cell location, and tissue specificity. This recommends, therefore, that diverse species of uricases may have a common evolutionary origin (Oda et al. 2002; Varela-Echavarría et al. 1988).

```
Oryctolagus cun    1 ---------------------MATTKKNEDVEFVRIGYGKDMVKVLHIQRDG---KYHS
Rattus norvegic    1 -----------MAH-------YHDDYGKNDEVEFVRIGYGKDMVKVLHIQRDG---KYHS
Mus musculus       1 -----------MAH-------YHDNYGKNDEVEFVRIGYGKDMVKVLHIQRDG---KYHS
Papio_hamadryas    1 -----------MAD-------YHNNYKKNDELEFVRIGYGKDMVKVLHIQRDG---KYHS
Cavia_porcellus    1 --------MESEAESQRKKFLYNNDYEVEFVRIGYGKDMVKVLHIQRDG---KYHS
Drosophila_mela    1 MFATPLRQPAAANHQTPKNSAGMDEHGKPYQYEITDHGYGKDAVKVLHVSRNC---PVHA
Camelus dromeda    1 ---------------------------------QIGYGKDMVKVLRVQRNG---KYHS
Phaseolus vulga    1 ---------------------MAQEVVEGFKFEQRHGKDRVRVARVWRTPQ--GRHF
Chlamydomonas r    1 ---------------------MATYALP----LHHHGKSKVRLGRVWRDG---TVHH
Bacillus_fastid    1 -------------------AERTMFYGKGDVYVFRTYNPIKGLKQIPESNFTEKHNT
Arthrobacter_gl    1 ------------------MTATAETSTGTKVVLGQNQYGKAEVRLVKVTRNT---ARHE
Cyberlindnera_j    1 --------------------------MSTTLSSSTYGKDNVKFLVKKDPQNPKKQE
Aspergillus fla    1 --------------------------MS-AVKAARYGKDNVRVYKVHKD-EKTGVQT
consensus          1                           ............. ...   . ..

Oryctolagus_cun   36 IKEVATSVQLTLS-SKQDYVYGDNSDIIPTDTIKNTVHVLAKFKG-IKSIEVFAMNICEH
Rattus_norvegic   40 IKEVATSVQLTLR-SKKDYLHGDNSDIIPTDTIKNTVHVLAKFKG-IKSIETFAMNICEH
Mus musculus      40 IKEVATSVQLTLR-SKKDYLHGDNSDIIPTDTIKNTVHVLAKLRG-IRNIETFAMNICEH
Papio hamadryas   40 IKEVATSVQLTLS-SKKDYLHGDNSDIIPTDTIKNTVHVLAKFKG-IKSIEAFGVNICEY
Cavia_porcellus   50 IKEVATSVQLTLS-SKKDYLHGDNSDIIPTDTIKNTVHVLAKFKG-IKSIETFALNICQY
Drosophila_mela   58 IQEFEVGTHLKLY-SKKDYYQGNNSDIVATDSQKNTVYLLAKKHG-IESPEKFALLLARH
Camelus_dromeda   23 IKEVATSVELNLS-SRKDYIYGDNADIIPTDTMKNTVHVLAKTKG-IKSIETFAMHVCEH
Phaseolus_vulga   35 VVEWRVGITLFSD-CVNSYIRDDNSDIVATDTMKNTVYAKAKECSDILSVEDFAILIAKH
Chlamydomonas r   30 MVEWNVNTMIDSD-MEHAFLNGDNTGMTATDTQKNTVYVAQRMSQKCTIEQAIAIAQH
Bacillus_fastid   40 IFGVNAKVALKGEQLLTSHTEGDNSLVVATDSMKNFIQRHAASYE-GATLEGFLQYVCEA
```

```
Arthrobacter_gl    39  IQDLNVTSQLRGD-FEAAHTAGDNAHVVATDTQKNTVYAFARDGF--AITEEFLLRLGKH
Cyberlindnera_j    32  VMEATVTCLLEGG-FDTSYTEADNSSIPTDTVKNTILVLAKTTE-IWPIERFAAKLATH
Aspergillus_fla    30  VYEMTVCVLLEGE-IETSYTKADNSVIVATDSIKNTIYITAKQNP-VTPPEIFCSILGTH
consensus          61  . ..  .. .  .  ...... ..*......**..**.. ..*. .. .. ...* .. ... .

Oryctolagus_cun    94  FLSGFNHVVRVHVYVEEVPWKRLE-----------KNGVQ------------HVHAFI
Rattus_norvegic    98  FLSSFSHVTRAHVHVEEVPWKRFE-----------KNGVK------------HVHAFI
Mus_musculus       98  FLSFNHVTRAHVYVEEVPWKRFE------------KNGIK------------HVHAFI
Papio_hamadryas    98  FLSFNHVIRAQVYVEEIPWKRLE------------KNGVK------------HVHAFI
Cavia_porcellus   108  FLSSFCHVIRAHVYLEEVPWKRFE-----------KNGAK------------HVHAFI
Drosophila_mela   116  FLNKFSHVEEAHVHVBAYPWQRVCQEETRTNVNGKCENGVQGNCDFSSIDNRSLHNHAFI
Camelus_dromeda    81  FLSFNHVIRVGVCVEEVPWKRLE-----------KNGVK------------HVHAFI
Phaseolus_vulga    94  FVSFYKKVTGAIVNIVEKPWERVI-----------VDGQP------------HQHCFT
Chlamydomonas_r    89  FVRTYPLVSKAKTYVEQKPWTRVQ-----------QSGIP------------HDHCFA
Bacillus_fastid    99  FLAKLSHLDAVRLEAKEYAFDDIQVG---------TDKGVVT----------SDLVFR
Arthrobacter_gl    96  FTEGEDWVTGGRWAAQQFFWDRIN-----------------------------DHDHAFS
Cyberlindnera_j    90  FVEKLSHVSGVSVKIVQDRWVKYA-----------VDGKP-----------HDHSFI
Aspergillus_fla    88  FLEKVNHIHAAHVNIVCHRWTRMD-----------IDGKP-----------HPHSFI
consensus         121  *.... .. .... ... .. ...              ...            . ...*.

Oryctolagus_cun   129  HTPTGTHICEVEQRRSGL----PVIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRC
Rattus_norvegic   133  HTPTGTHFCDVEQVRNGP----PIIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRC
Mus_musculus      133  HTPTGTHFCEVEQMRNGP----PVIHSGIKDLKVLKTTQSGFEGFLKDQFTTLPEVKDRC
Papio_hamadryas   133  HTPTGTHFCEVEQLRSGP----PVIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRC
Cavia_porcellus   143  HTPTGTHFCEVEQRRSEP----PVIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRC
Drosophila_mela   176  FTPTALHYCDVVIRRIDPK---QTVILGIKGLRVLKTTQSSFVNFVNDEFRSLPDQYDRI
Camelus_dromeda   116  YTPTEKRFCEVEQMRSGR----PVIHSGIKDLKVLKTTQSGFEGFIRDQFSTLPETKDRI
Phaseolus_vulga   129  LG-GEKHTTEAIVQRSGS----LQITSGIEGLSVLKTTQSGFENFIRNKYTALPLTRERI
Chlamydomonas_r   124  LQGTEVRTVYVTYDRAGT----LDVTAGLKDLSLLKTTQSGVEGFLRDQYILLPETNDRI
Bacillus_fastid   138  KSRNEYVTATVEVARTASGTEVVEQASGIADIQLIKVSGSSFYGYIIDEYTTIAEATDRP
Arthrobacter_gl   127  RNKSEVRTAVLEISGSEQ-----AIVAGIEGLTVLKSTCSEFHGFPRDKYTTIQETTDRI
Cyberlindnera_j   125  HEGGEKRITDLYYKRSGD----YKLSSAIKDLTVLKSTGSMFYGLNQCDFTTIQPTTDRI
Aspergillus_fla   123  RDSEEKRNVQVDVVEGKG----IDISKSSTSGLTVLKSTNSQFWGFLRDEYTTLKETWDRI
consensus         181  ..... ...... ...           ............*....*...........*.. ..*.

Oryctolagus_cun   185  FATQVYCKWRYQHS------------QDVDFEATWDIVRDTVLEKFAGPYDRGEYSPSVQ
Rattus_norvegic   189  FATQVYCKWRYQN-------------RDVDFEATWGAVRDIVLKKFAGPYDRGEYSPSVQ
Mus_musculus      189  FATQVYCKWRYQR-------------RDVDFEAIWGAVRDIVLQKFAGPYDRGEYSPSVQ
Papio_hamadryas   189  FATQVYCKWRYHQC------------RDVDFEATWGTIRDLVLEKFAGPYDRGEYSPSVQ
Cavia_porcellus   199  FATQVYCKWRYHQS------------RDVDFEATWDTVRDIIKFAGPYDCRGEYSPSVQ
Drosophila_mela   233  FSTVVDCSWEYSDT------------ENLDFLRAWQTVKNIIRNFAGDPQVGVSSPSVQ
Camelus_dromeda   172  FATQVYCRWRYHPG------------RDVDFDATWDTVRDIILKKFAGPFDRGEYSPSVQ
Phaseolus_vulga   184  LATEVTALWRYSYESLYNLP-----QKPLYFTDKYLEVKKVLADTFFGPPNEGVYSPSVQ
Chlamydomonas_r   180  MASTVTCTWKY--------A-----AAPACYDAAYAAAKQGLLDAELGPGARRVQPSVQ
Bacillus_fastid   198  LYIFLNIGWAYENQ------------FDAKGDNPANYVAAEQVRDIAASVFTLDNKSIQ
Arthrobacter_gl   182  LATDVSARWRYNT-------------VEVDFDAVYASVRGLLLKAFAET-----HSLALQ
Cyberlindnera_j   181  LSTDVDATWVWDNKKIGSVYDIAKAADKGIFDNVYNQARETTLTTFAL-----ENSPSVQ
Aspergillus_fla   179  LSTDVDATWQWKN--FSGLQEVRSHVPK-FDATWATAREVTLKTFAE-----DNSASVQ
consensus         241  ... . ..*.. ..               ...... .  ....... ....  .......*

Oryctolagus_cun   233  KTLYDIQVLTLSRVPQIEDMEISLPNIHYFNIDMS----KMGLINKEEVLPIDNPYGKI
Rattus_norvegic   236  KTLYDIQVLTLSQLPIEDMEISLPNIHYFNIDMS-----KMGLINKEEVLLPIDNPYGKI
Mus_musculus      236  KTLYDIQVLSLSQLPIEDMEISLPNIHYFNIDMS-----KMGLINKEEVLLPIDNPYGKI
Papio_hamadryas   237  KTLYDIQVLSLSRQVPIEDIEISLPNIHYFNIDMS----KMGLINKEEVLLPIDNPYGKI
Cavia_porcellus   247  KTLYDIQVLSLRQVPIEDIEISLPNIHYVNIDMS-----KMGLINKEEVLLPIDNPYGKI
Drosophila_mela   281  HTLYLSERQVLDVPQVSVISMTMPNKHYFNFDTKP-FQKIAPGDNNEVFIPYDKPHGTI
Camelus_dromeda   220  RTLCDIQMLSLSQLPIEDVELSLPNIHYIEIDMS-----KMGLINKKEVLPIDNPYGRI
Phaseolus_vulga   239  NTIYLVAKATLNRFPLIAYVHLKMPNIHELPVNISS-KDGPIVKFEDDVLLPTDEPHGSI
Chlamydomonas_r   227  YTLYDVAKNILDRVPTSESIFLNMPNHLIPCNPVG-SS-----FNNDVYVATSEPHGNI
Bacillus_fastid   246  HLIYHTGLTILDRFPQLTEVNFGTNNRTWDTVVEG------TDGFKGAVFTEPRPPFGFQ
Arthrobacter_gl   224  QTLYEMGRAVIETHPLIDEIKMSLPNKHHFLVDLQP---FGQDNPNEVGYAADRPYGLI
Cyberlindnera_j   236  ATMENVATQILEKACSVYSVSYALPNKHYFLIDLKW-KGLEN---DNELFYESPHPNGLI
Aspergillus_fla   230  ATIYKVAEQILARQQLIETVEYSLPNKHYFEIDISWHKGLQNTGKNAEVTAPQSDPNGLI
consensus         301  ...... . .. .............*....  ...      .... .. ....... *.* .

Oryctolagus_cun   289  TGTVK------------------
Rattus_norvegic   292  TGTVRRK---LPSRL--------
Mus_musculus      292  TGTVRRK---LPSRL--------
```

109

```
Papio_hamadryas   293  TCTVKRK---LSSRL--------
Cavia porcellus   303  TCTVKRK---LSSKL--------
Drosophila mela   340  YAQLARKN--LNSIL--------
Camelus_dromeda   276  TGKLKRK---PTSKL--------
Phaseolus_vulga   298  EASLSR----VWSKL--------
Chlamydomonas_r   281  EAVLTRKGIAPHCKL--------
Bacillus_fastid   300  GFSVFQEDLAREKASANSEYVAL
Arthrobacter gl   280  EATIQREGSRADHPIWSNIAGFC
Cyberlindnera j   292  KCTVVRK---EKTKL--------
Aspergillus_fla   290  KCTVGRSS--LKSKL--------
consensus         361  ......  . ...
```

Figure 5.2: Multiple sequence alignment shows maximum conservation exhibiting in between 37 and 373 amino acids of uricase protein sequences from different sources. '*' indicates fully conserved residue, and '.' indicates moderately conserved amino acids. The sections are highlighted in red and pink colors are representing B-cell and T-cell epitopic peptides in the uricase sequence. The black color represents identical amino acids, whereas the grey color represents similar amino acids.

The dendrogram shows that there are two clusters in which one cluster contains uricase from animals (*R.norvegicus, M.musculus*, *C.porcellus*, *P.hamadryas*, *O.cuniculus*, *C.dromedarius,* and *D.melanogaster*), and the second cluster comprises of uricase from the plant (*P.vulgaris*), algal (*C.reinhardtii*), bacterial (*A. globiformis* and *B.fastidious*) and fungal sources (*C.jadinii* and *A.flavus*) (Figure A.III.1). In the first cluster, it is observed that uricase from *R.norvegicus*, *M.musculus*, *C.porcellus,* and *P.hamadryas* have similarities and identities at the sequence level. In the second cluster, the *P.vulgaris* uricase show sequence-level similarity with the eukaryotic algae *C.reinhardtii*. The fungal source of uricase from *C.jadinii* and *A.flavus* was observed to be in the same cluster, while the bacterial uricases from *Arthrobacter globiformis* and *Bacillus fastidious* appeared in the same cluster with a similar sequence level. The antigenicity problem was reported to be highly present in therapeutic enzymes from plant source (Sánchez et al. 1987). Therefore, the clinical uses of such plant-derived therapeutic enzymes are limited.

5.2.2 Motifs conservation

MEME (Motif-based sequence analysis tool) was reported to be based on the Bayesian probabilistic model, which uses an expectation-maximization algorithm to obtain the motifs for all the sequences and optimizes the statistical parameters (Bailey et al. 2009). The maximum number of motifs obtained from uricase sequences was six and is documented in Table 5.1. The motifs 2, 3, and 4 are common to all the thirteen uricases indicating the enzyme's major function is conserved. These three motifs clearly indicate their potential role in structural and catalytic functional attributes of uricase. However, motif 1 and 5 are absent in uricases from *A.globiformis*, *B.fastidious*, *C.jadinii*, and *A.flavus,* and motif 6 is absent in *D.melanogaster*, *P.vulgaris, C.reinhardtii*, *B.fastidious* and *A.flavus*. The combined block diagram of motifs is displayed in Figure A.III.2. Other details, including motif widths, information about the sequence, and the best possible matches, are listed in Table A.III.2.

Table 5.1: Conserved motifs locations for uricase protein from different source organisms

S.No	Organism	Genbank ID	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5	Motif 6
1	*Oryctolagus cuniculus*	189303536	7-47	48-76	83-123	147-196	207-236	257-285
2	*Rattus norvegicus*	20127395	11-51	52-80	87-127	151-200	210-239	260-288
3	*Mus musculus*	6678509	11-51	52-80	87-127	151-200	210-239	260-288
4	*Papio hamadryas*	20513624	11-51	52-80	87-127	151-200	211-240	261-289
5	*Cavia porcellus*	884943374	21-61	62-90	97-137	161-210	221-250	271-299
6	*Drosophila melanogaster*	17136576	29-69	70-98	105-145	195-244	255-284	-
7	*Camelus dromedarius*	339716249	-	35-63	70-110	134-183	194-223	244-272
8	*Phaseolus vulgaris*	2809326	-	47-75	83-123	146-195	213-242	-
9	*Chlamydomo nas reinhardtii*	11066111	-	42-70	78-118	142-191	201-230	-

10	*Bacillus fastidious*	823631078	-	53-81	88-128	160-209	-	-
11	*Arthrobacter globiformis*	187609193	-	51-79	85-125	144-193	-	248-276
12	*Cyberlindnera jadinii*	1147426164	-	44-72	79-119	143-192	-	260-288
13	*Aspergillus flavus*	137100	-	42-70	77-117	141-190	-	-

The results obtained from Pfam clearly indicate that the monomeric uricase sequences from various sources have two domain organizations that belong to the uricase family. Each domain consists of 286 amino-acid residues; however, the starting and ending amino acid number varies for each organism. The double domain organization of the uricase monomer indicates the T-fold domains.

5.2.3 Antigenic epitopes prediction

IEDB epitope database and prediction resource were used to determine the antigenic epitopes of *Ag*-Uricase and *Bf*-Uricase. Continuous sequential regions or various antigenic determinant groups were reported as the major contributors for the formation of antigenic sites or epitopes in a protein (Ramya and Pulicherla 2015). In order to reduce the clinical immune-reactiveness of the therapeutic enzyme uricase, primarily the continuous B-cell epitopes were predicted based on the important antigenic parameters such as Emini surface accessibility, Parker hydrophilicity, and Karplus & Schulz flexibility. The majority of the antigenic regions of the proteins were reported to have more polar and charged residues rather than hydrophobic residues (Zarei et al. 2019), based on propensity scales for each of the 20 amino acids (Ahmad et al. 2016; Bull and Breese 1974; Malherbe 2009; Roseman 1988). The predicted B-cell epitope peptides obtained from Emini surface accessibility, Parker hydrophilicity and Karplus & Schulz Flexibility scores of *Ag*-Uricase and *Bf*-Uricase were illustrated in Table A.III.3

In the case of *Ag*-Uricase, the B-cell epitopic peptides ^{167}PR**D**KYT^{172}, ^{261}GQD**N**PE^{267,} and ^{156}LKS**T**GSE^{162} have the highest surface accessibility score (4.904), parker hydrophilicity score (6.514), Karplus & Schulz flexibility score (1.108), respectively

(Table 5.2), in which the polar residues D169 (Figure 5.3-A), N264 (Figure 5.3-C) and T159 (Figure 5.3-E) made the main contribution to such high score were selected as hot-spot residues for mutagenesis. Similarly, in the case of *Bf*-Uricase, the B-cell epitopic peptide ^{137}RK**S**RNE^{142}, ^{212}DDAK**G**DN^{218} and ^{213}DAK**G**DNP^{219} have the highest surface accessibility score (6.575), parker hydrophilicity score (7.214), and Karplus & Schulz flexibility score (1.088), respectively (Table 5.2) in which the polar residues S139 (Figure 5.4-A), K215 (Figure 5.4-C) and G216 (Figure 5.4-E) were selected as a hot-spot residue for mutagenesis. In the case of *Ag*-Uricase, the conformational B-cell epitopes are found to be overlapped with the peptides predicted as linear B-cell epitopes (region ~167-172 and 260-267), indicating the mutations of those two regions are crucial to mask the immunogenicity (Figure 5.3-G). In contrast, *Bf*-Uricase is found to have very less conformational B-cell epitopes (Figure 5.4-G). It can be observed from the highlighted portion of Figure 5.2 that all the selected hot spot residues of B-cell epitopes are located at either non-conserved or moderately conserved regions indicating that they may have comparatively less significance for preserving the structural and functional characteristics of uricase (Zarei et al. 2019). Additionally, all the selected hot-spot residues located at the B-cell epitopic region are polar residues. Therefore, replacing these residues with hydrophobic residues was considered the best way to reduce adverse allergic reactions in the human body (Bander et al. 2005).

T-cell immune responses are induced by identifying T-cell epitopes attached to MHC molecules displayed at the surface of antigen presenting cells. T-cell epitope prediction is based on identifying peptide lengths within an antigen that can stimulate CD4+ T-cells, which ultimately elicits the immune response in the human body (Ahmed and Maeurer 2009). Deimmunization is a new technology that locates and mutates polypeptide sequences using immunological and molecular biology techniques, which helps in reducing protein immunogenicity that does not affect the protein function (Cantor et al. 2011; Jones et al. 2009; Macfarlane et al. 2006). The success of reduced immunogenicity has been observed in humanized and chimeric antibodies with the removal of potential T-cell epitopes through a mutagenesis approach (Bander et al. 2005; Holgate and Baker 2009; Macfarlane et al. 2006).

A consensus prediction approach is one of the widely used techniques for identifying variable-length peptides related to T-cell epitopes (Moola et al. 1994; Moutaftsi et al. 2006). Therefore, identifying potential immunogenic T-cell epitopes of *Ag*-Uricase and *Bf*-Uricase is essential for locating the binding sight of MHC-II molecules. The T-cell epitopic peptides obtained in the deimmunization analysis are given in Table A.III.7. The lower the median percentile rank, the higher the propensity to act as an epitope (Dhanda et al. 2018). The top-scored epitopic peptides are documented in Table 5.3. The hot-spot residues for the T-cell epitopic peptides are Tyr203, Ile172 for *Ag*-Uricase, and *Bf*-Uricase, respectively. The identified hot-spot residues are next subjected to the *in-silico* mutagenesis process to obtain less immunogenic candidates of uricase.
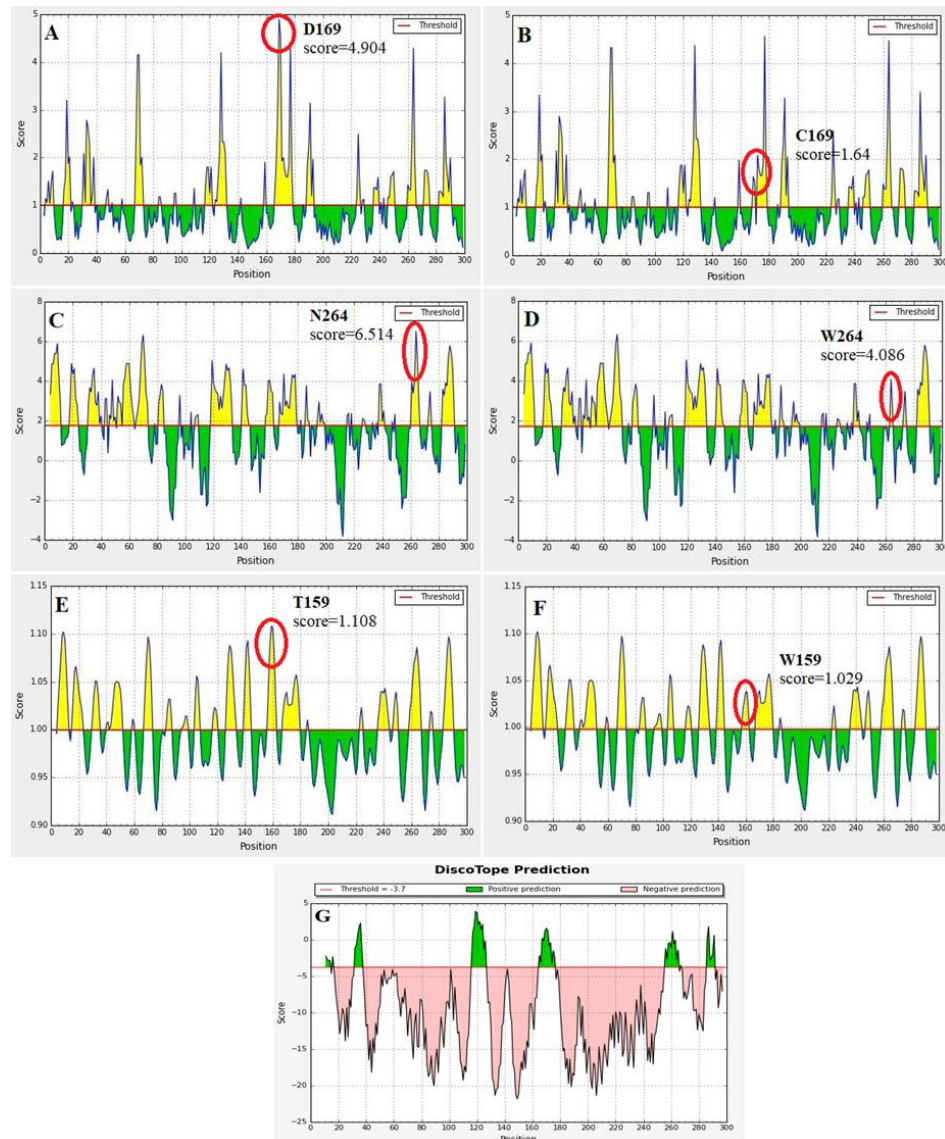
Figure 5.3: Graphs represent different regions of the amino acid sequence of uricase that can act as B-cell epitopes in the case of *Ag*-Uricase (A) Emini surface accessibility plot representing maximum antigenicity in *Ag*-Uricase at 167-172. (B) Plot presenting the change in surface accessibility score in *Ag*-Uricase at 167-172. (C) Parker Hydrophilicity prediction plot presenting maximum antigenicity in *Ag*-Uricase at 261-267. (D) The change in hydrophilicity score in *Ag*-Uricase at 261-267. (E) Karplus & Schulz Flexibility plot display maximum antigenicity in *Ag*-Uricase at 156-162. (F) Plot presenting the change in flexibility score in *Ag*-Uricase at 156-162. (G) Graph representing conformational B-cell epitopes from the 3D structure of *Ag*-Uricase. The selected regions are marked by a red circle.

Figure 5.4: Graphs represent different regions of the amino acid sequence of Uricase that can act as B-cell epitopes in the case of *Bf*-Uricase (A) Emini surface accessibility plot representing maximum antigenicity in *Bf*-Uricase at 137-142. (B) Plot presenting the change in surface accessibility score in *Bf*-Uricase at 137-142. (C) Parker Hydrophilicity prediction plot presenting maximum antigenicity in *Bf*-Uricase

116

at 212-218. (D) Plot presenting the change in hydrophilicity score in *Bf*-Uricase at 212-218. (E) Karplus & Schulz Flexibility plot display maximum antigenicity in *Bf*-Uricase at 213-219. (F) Plot presenting the change in flexibility score in *Bf*-Uricase at 213-219. (G) Representing conformational B-cell epitopes from the 3D-structure of *Bf*-Uricase. The selected regions are marked by red circle.

Table 5.2: B-cell epitopic scores of *Ag*-Uricase and *Bf*-Uricase. The bold letters are representing the hot spot residues

Organisms	Method	Peptide	Region	Score
Ag-Uricase	Surface accessibility	PR**D**KYT	167-172	4.904
	Surface accessibility (after-mutation)	PR**C**KYT	167-172	1.64
	Hydrophilicity	GQD**N**PNE	261-267	6.514
	Hydrophilicity (after-mutation)	GQD**W**PNE	261-267	4.086
	Flexibility	LKS**T**GSE	156-162	1.108
	Flexibility (after-mutation)	LKS**W**GSE	156-162	1.029
Bf-Uricase	Surface accessibility	RK**S**RNE	137-142	6.575
	Surface accessibility (after-mutation)	RK**V**RNE	137-142	3.743
	Hydrophilicity	DDA**K**GDN	212-218	7.214
	Hydrophilicity (after-mutation)	DDA**W**GDN	212-218	4.971
	Flexibility	DAK**G**DNP	213-219	1.088
	Flexibility (after-mutation)	DAK**F**DNP	213-219	1.039

Table 5.3: T-cell epitopic scores of *Ag*-Uricase and *Bf*-Uricase. The bold letters are representing the hot spot residues

S.NO	organism	Peptide	Start position	End position	Median percentile rank
1	*Ag*-Uricase	AVYASVRGLLLKAFA	201	215	10.78
2	*Bf*-Uricase	IADIQLIKVSGSSFY	166	180	8.77

5.2.4 Residual modification

The hot spot residue of the B-cell, T-cell epitopes of *Ag*-Uricase and *Bf*-Uricase were identified and mutated using Pymol software. The B-cell and T-cell epitopes on the 3D structure of uricase from both sources are shown in Figure 5.6 (A, B). In the case of *Ag*-Uricase, Asp located at 169 position is found to have high surface accessibility characteristics (Figure 5.3-A, Table 5.2), whereas maximal reduction of antigenic probability is obtained for Ag^{D169C} mutation (Figure 5.3-B, Table 5.2, Table A.III.8). It is evident from Parker hydrophilicity analysis (Figure 5.3-C, D, Table 5.2, TableA.III.8) that Ag^{N264W} mutation causes optimal reduction of immunogenicity of *Ag*-Uricase. The Ag^{T159W} mutation is found to decrease the flexibility characteristics in the case of *Ag*-Uricase (Figure 5.3-E, F). Similarly, Bf^{S139V} (Figure 5.4-A, B, Table A.III.8), Bf^{K215W} (Figure 5.4-C, D, Table A.III.8), and Bf^{G216F} (Figure 5.4-E, F, Table A.III.8) mutations are found to reduce the surface accessibility and hydrophilicity characteristics of the B-cell epitope of *Bf*-Uricase, respectively.

In the case of T-cell epitopes, Ag^{Y203D} mutation gives an optimal reduction in immunogenicity for *Ag*-Uricase, whereas Bf^{I172P} mutation in *Bf*-Uricase is seen as the best result (Table 5.4, Table A.III.9, and A.III.10). In each subunit, three B-cell (based on Emini surface accessibility, Parker hydrophilicity, and Karplus & Schulz flexibility) and one T-cell mutation (based on Deimmunization technique) were carried out. A total of 16 mutations were carried out in each tetrameric form of uricase sourced from the above-mentioned species. All four protein models were validated using ΔΔG values (Table A.III.11) and Ramachandran plot (Figure A.III.3, Table

118

A.III.12). All the ΔΔG were found in the permissible range. It is evident from Table 5.1 that mutations are done inside motif 4 and motif 6, which can vary the structural and functional aspects of mutant uricase models. It can be noticed from the highlighted part of Figure 5.2 that the impact for amino acid changes at particular sites in the sequence are less likely to vary the protein structure and function because mutations are mainly done at the non-conserved and moderately conserved portion of the uricase sequence (Sun et al. 2011). Both the wild-type and mutated protein models are subjected to molecular docking to have further insights into the functional characteristics of uricase.

5.2.5 Molecular docking of uricase

Molecular docking was performed to assess the influence of mutagenesis on the functional aspect of uricase. The non-bonded interactions between the uric acid and the amino-acid residues at the catalytic pocket of *Ag*-Uricase, *Bf*-Uricase are illustrated in Figure 5.5. The binding affinity of docked pose of uric acid towards both the wild type, mutated form of 2YZB (*Ag*-Uricase), and 4R8X (*Bf*-Uricase) are documented in Table 5.5. The uric acid is found to interact with amino acid residues located at the junction of the identical monomers of 2YZB, indicating a pronounced binding pocket of uricase (Figure 5.6-A). This binding pocket residues away from the epitopic regions that are located at the surface of uricase. The uric acid at the catalytic pocket of normal or wild type 2YZB exhibited non bonded interactions with Asp68 of chain A and Phe163, Arg180, Leu222, Gln223 of chain D (Figure 5.6-B).

Table 5.4: T-cell epitopic scores obtained from mutation of all the amino acids with the hot spot residue located at 203 of *Ag*-Uricase and 172 of *Bf*-Uricase. In the case of *Ag*-Uricase, Tyr is present in 203 position whereas Ile is present at 172 position in the case of *Bf*-Uricase

S.No	Amino acids	*Ag*-Uricase	*Bf*-Uricase
1	Native	10.78	8.77
2	Ala	19.685	17.49
3	Cys	19.345	28.97

4	Asp	24.585	30.42
5	Glu	22.9	27.72
6	Phe	8.525	12.25
7	Gly	22.47	35.87
8	His	16.035	27.895
9	Ile	10.84	-
10	Lys	13.555	24.3
11	Leu	12.52	7.325
12	Met	15.245	10.335
13	Asn	21.235	21.675
14	Pro	21.025	51.955
15	Glu	16.465	15.88
16	Arg	14.75	24.715
17	Ser	21.11	23.14
18	Thr	19.15	24.06
19	Val	14.72	12.675
20	Trp	11.105	18.84
21	Tyr	-	16.255

Table 5.5: The docking scores of uric-acid at the catalytic pocket of wild and mutated uricase

S. No	Organism	Docking score	Binding Energy MM/GBSA (kcal/mol)	Total no of non-bonded interactions	Possible Number of H-bonds
1.	*Ag*-Uricase *(native)* (PDB ID: 2YZB)	-8.414	-47.71	7	6
2.	*Ag*-Uricase *(mutated)* (PDB ID:2YZB)	-8.570	-48.60	7	6

3.	Bf-Uricase(native) (PDB ID :4R8X)	-5.221	-39.40	5	4
4.	Bf-Uricase(mutated) (PDB ID :4R8X)	-5.389	-41.44	5	4
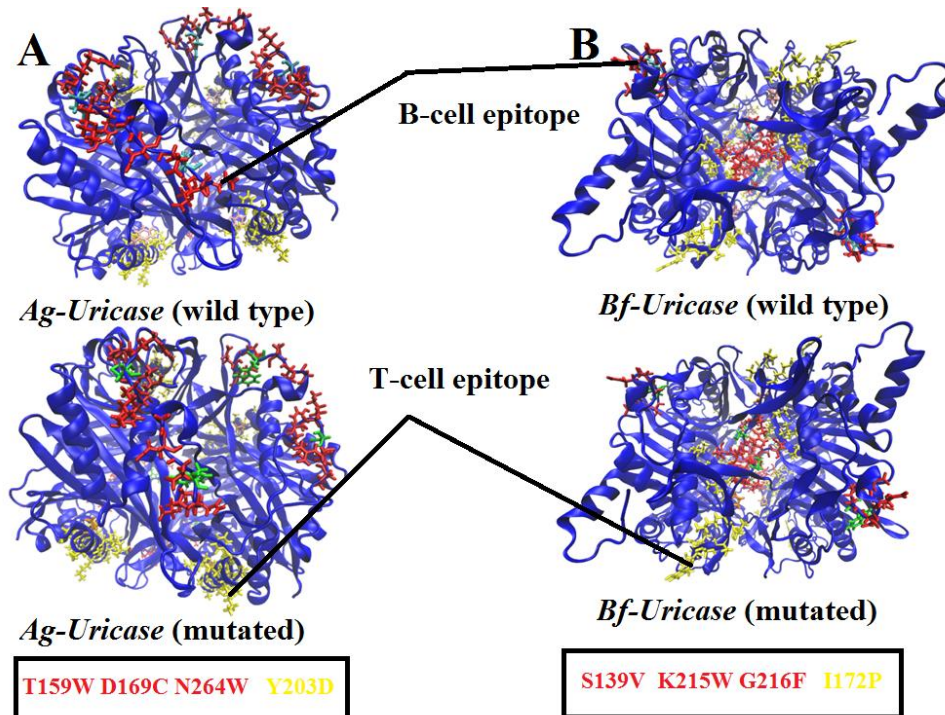
Figure 5.5: The locations of B-cell and T-cell epitopes on (A) *Ag-Uricase* and (B) *Bf-*Uricase. B-cell epitopes are represented in red color, and T-cell epitopes are marked in yellow color. The mutated residues are shown in green color. The mutations done in each monomer of *Ag-Uricase* and *Bf-Uricase* are listed below the enzyme structure.
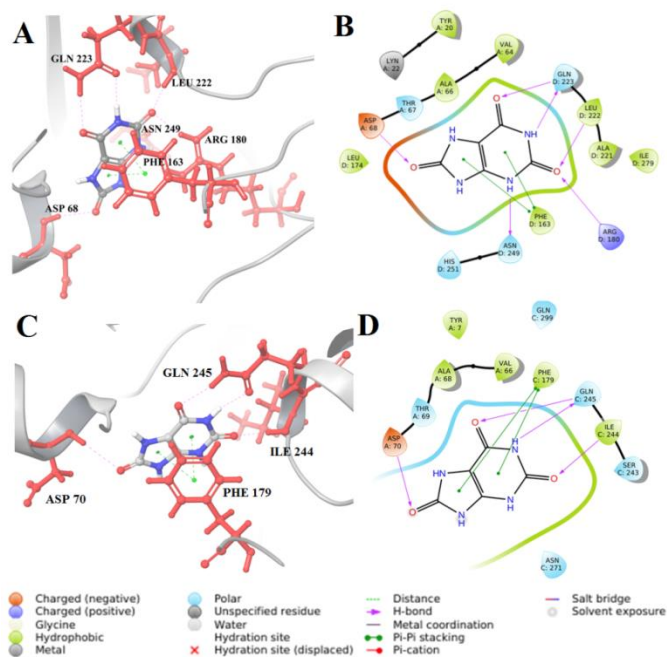
Figure 5.6: The docking pose and two-dimensional (2D)-ligand interaction diagram of *Ag*-Uricase (A & B) and *Bf*-Uricas*e* (C & D) 4R8X. The interacting amino acids are represented in red.
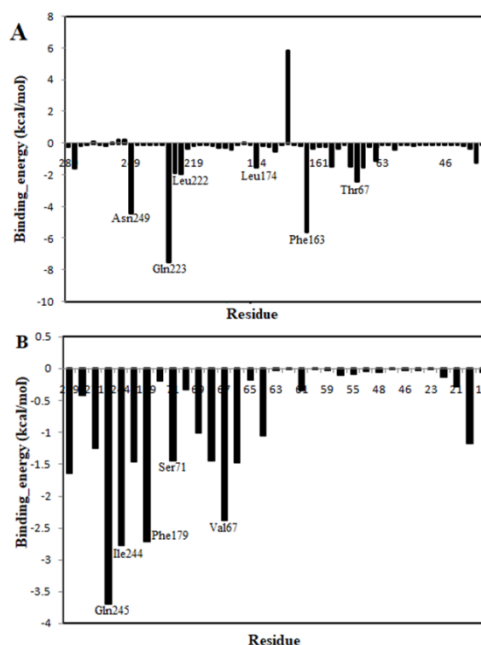
Figure 5.7: Residue-wise decomposition of the binding energy of uric acid towards the catalytic pocket of uricase. (A) Showing the binding energy decomposition of uric

acid in the case of *Ag*-Uricase. (B) Showing the residue-wise decomposition of uric acid for *Bf*-Uricase.

The oxygen atom located at the five-membered ring of uric acid is found to accept one hydrogen bond (C=O-----HO-Asp68, hydrogen bond length = 2.39 Å) with the side chain of Asp68. Two hydrogen bonds are formed between the uric acid and Gly223 (NH----O=CGly223 and C=O---HNGly223, hydrogen bond length = 1.94 Å and 2.15 Å, respectively). The nitrogen atom located at the peptide bond between Ala221, Leu222 donates one hydrogen bond to the oxygen atom of uric acid (C=O-----HNLeu222, hydrogen bond length 1.83 Å). Another hydrogen bond interaction is found to be present between the oxygen atom of uric acid and the side-chain of Arg180 (C=O-----NH-Arg180, hydrogen bond length= 2.20 Å). Phe163 is found to exhibit π-π stacking interaction with the heterocyclic rings of uric acid. It is evident from Figure 5.6-B that one hydrogen bond is present between the uric acid and the Asn249 (NH-----O=C-Asn249, hydrogen bond length = 2.43Å), which plays an important role in stabilizing the substrate inside the catalytic pocket. It can be observed from Table 5.5 that the numbers of interactions are unchanged after *in-silico* mutagenesis of 2YZB. However, the binding energy is found to vary after mutation. The binding affinity of uric acid towards wild type 2YZB is found to be -47.7 kcal/mol (Table 5.5). The per-residue energy contribution of the amino-acid residues towards the binding of uric acid at the catalytic pocket of 2YZB is plotted in Figure 5.7-A to understand the protein-ligand association at molecular level. It is clear from Figure 5.7-A that the binding affinity of uric acid is mainly dependent on the interaction between Asn249, Gln223, Leu222, Phe163, and Thr67 of 2YZB. Especially, Asn249 (-4.04 kcal/mol), Gln223 (-7.47 kcal/mol), and Phe163 (-5.57 kcal/mol) are the key residues that are responsible for anchoring the ligand at the active site of 2YZB. After the mutation in the B-cell and T-cell epitope region of 2YZB, we observed that the binding energy did not change remarkably (-48.60 kcal/mol). This phenomenon implies that mutation in the backbone of uricase did not affect the binding of uric acid at the active sight of uricase and preserved the catalytic activity of the enzyme.

The non-bonded interaction of uric acid with the catalytic pocket of 4R8X is displayed in Figure 5.6-C and D. The uric acid is located at the interface of chain-C and chain-A of 4R8X. It is clear from Figure 5.6-A and C that the binding pose of uric acid in 4R8X is similar to 2YZB. Uric acid is found to have hydrogen bonding interaction with Ile244 (C=O-----HN-lle244, hydrogen bond length=2.02 Å respectively), Gln245 (C=O----NHGln245, C=O----NHGln245, hydrogen bond length 2.21 Å, 2.00 Å, respectively) and exhibit π- π stacking interaction with Phe179 of chain C. Additionally, the oxygen atom located at the five-membered ring of uric acid accepts one hydrogen bond from Asp70 of chain A (C=O-----NH-Asp70, hydrogen bond length= 2.72 Å). The binding energy of uric acid at the catalytic pocket of wild type 4R8X is found to be -39.40 kcal/mol. The residue wise decomposition of binding energy is illustrated in Figure 5.7-B, which suggests that Gln245 (-3.704 kcal/mol), Ile244 (-2.783 kcal/mol), Phe179 (-2.720 kcal/mol), and Val67 (-2.377 kcal/mol) play key roles in stabilizing the ligand at the binding pocket of 4R8X. After site-directed mutagenesis of 2YZB, the binding energy of uric acid is found to be -41.44 kcal/mol. It is evident from the binding energy data that the catalytic activity of 2YZB did not vary much after the *in-silico* mutagenesis process. Thus, it is confirmed from molecular docking and MM/GBSA studies that both the protein model retains the functionality after the reduction of antigenicity.

5.2.6 Molecular dynamics simulation

100 ns molecular dynamics simulation was performed in each case to confirm the stability of the native structure of the protein after *in-silico* mutagenesis. The conformational stability of uricase was assessed by computing the time evolution of the root mean square deviation (RMSD) and root mean square fluctuation (RMSF) of the MD simulation trajectory. The RMSD of the normal protein backbone and mutated protein of both the species are shown in Figure 5.8 and compared to study the effect of mutagenesis on the native structure.

In the case of native 2YZB, the RMSD value gradually increased up to 1.82 Å around 1000 frames and became stable with an average value of 1.8Å (Figure 5.8-A). It is clear from Figure 5.8-A that the RMSD of mutated 2YZB increased up to 2.07 Å until

864 frames and similarly became stable like the wild type 2YZB. It is evident from the above-mentioned plot that both the native, mutated form of 2YZB displayed less deviation from the starting (t=0) structure and were found to be stable throughout the simulation time. An RMSD graph between the normal and mutated 4R8X is illustrated in Figure 5.8-B. The backbone RMSD of wild type 4R8X increased up to 1.97 Å around 1500 frames of the MD trajectory and then reached to steady-state. In contrast, the RMSD value of the mutated 4R8X increased up to 2.24 Å around 1950 frames and stabilized with an average RMSD value of 2.07 Å. The difference between the average RMSDs of both the normal and mutated 4R8X is found to be ~ 0.06 Å at the last phase of the simulation. It is evident from Figure 5.8-B that the native, as well as the mutated 4R8X, are structurally stable. The RMSD of uric acid (ligand) with respect to protein backbone is found to be comparable in the two species, both in its native and mutated form (Figure 5.8-A, B). This finding confirms the stabilization of the ligand at the catalytic site after mutation. The root mean square fluctuation (RMSF) measures the residue-wise fluctuation during simulation. The RMSF of the mutated protein and the normal one was compared to see the fluctuation at the mutated sights (Figure 5.8). The average backbone RMSF of native 2YZB is found to be 0.695 Å, whereas the mutated 2YZB displayed an average RMSF value of 0.693 Å. This indicates that both the native and mutated 2YZB are stable. It is clear from Figure 5.8-C that the B-cell epitopic regions 156-162, 167-172, 261-267 of wild type 2YZB have average RMSF values of 0.77 Å, 0.81 Å, and 0.87 Å, respectively. In the case of mutated 2YZB, the above-mentioned regions have average RMSF values of 0.62 Å, 0.79 Å, and 0.755 Å, respectively, which indicates less fluctuation after mutation. The average RMSF value of the T-cell epitopic region (201-215) of wild type 2YZB is found to be 0.66 Å, whereas the average RMSF value increased to 0.77 Å after mutation.
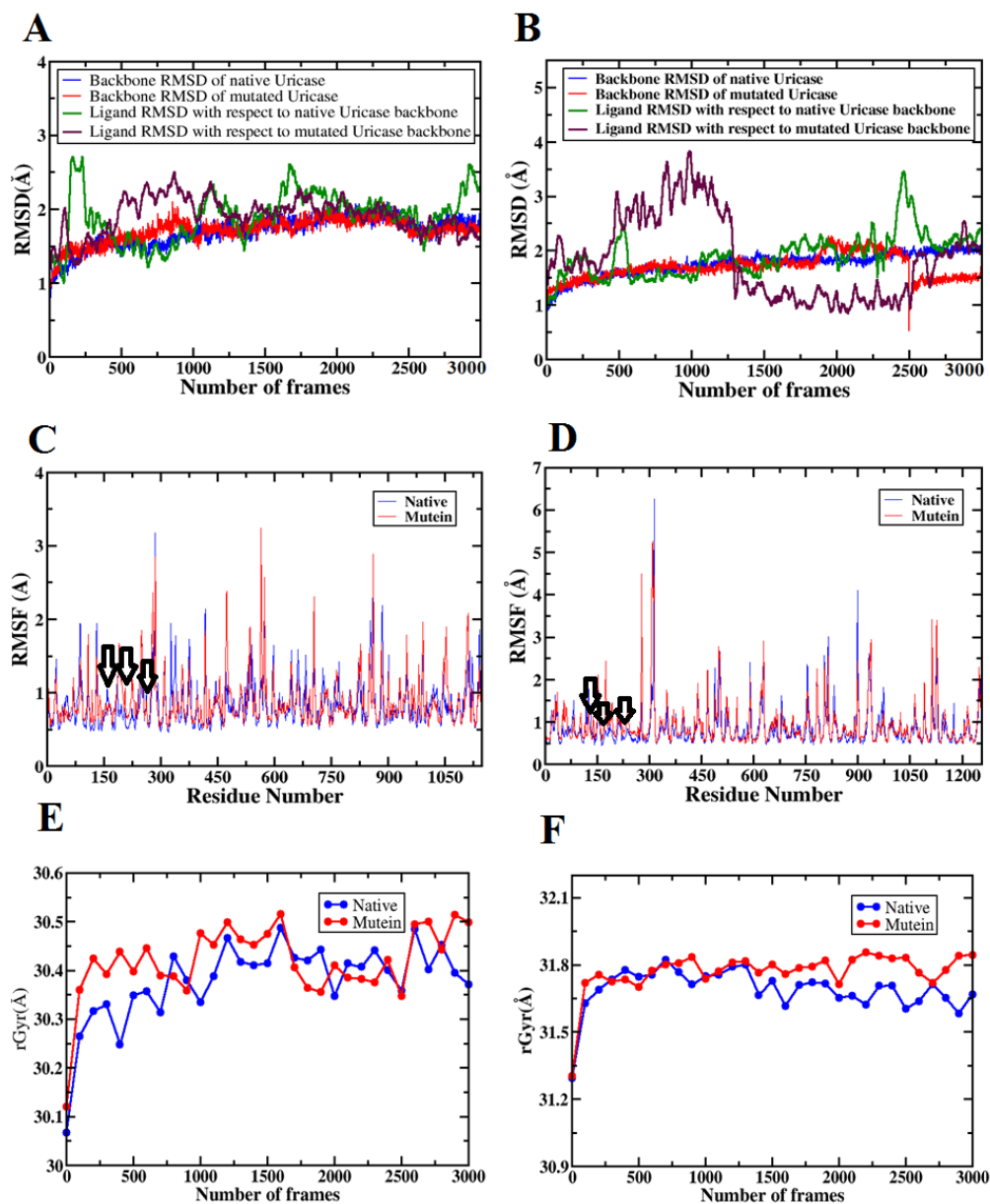
Figure 5.8: Comparisons between the backbone RMSDs, RMSFs of the wild type, and mutated uricase. (A) Showing the time evolution of RMSD of *Ag*-Uricase and (B) Showing time evolution of the RMSD of *Bf*-Uricase. (C) Showing the RMSF in case of *Ag*-Uricase. The arrow sign showing the fluctuation in the epitopic region (D) Showing RMSF in the case of *Bf*-Uricase. The arrow sign showing the fluctuation in the epitopic region. (E) The time evolution of the Radius gyration of native (blue) and mutated (red) *Ag*-Uricase. (F) The time evolution of the Radius gyration of native (blue) and mutated (red) *Bf*-Uricase. The blue color represents the native, and the red

color represents the mutein backbone. Green color represents the ligand RMSD in the native protein, and the maroon color represents ligand RMSD in mutated protein.
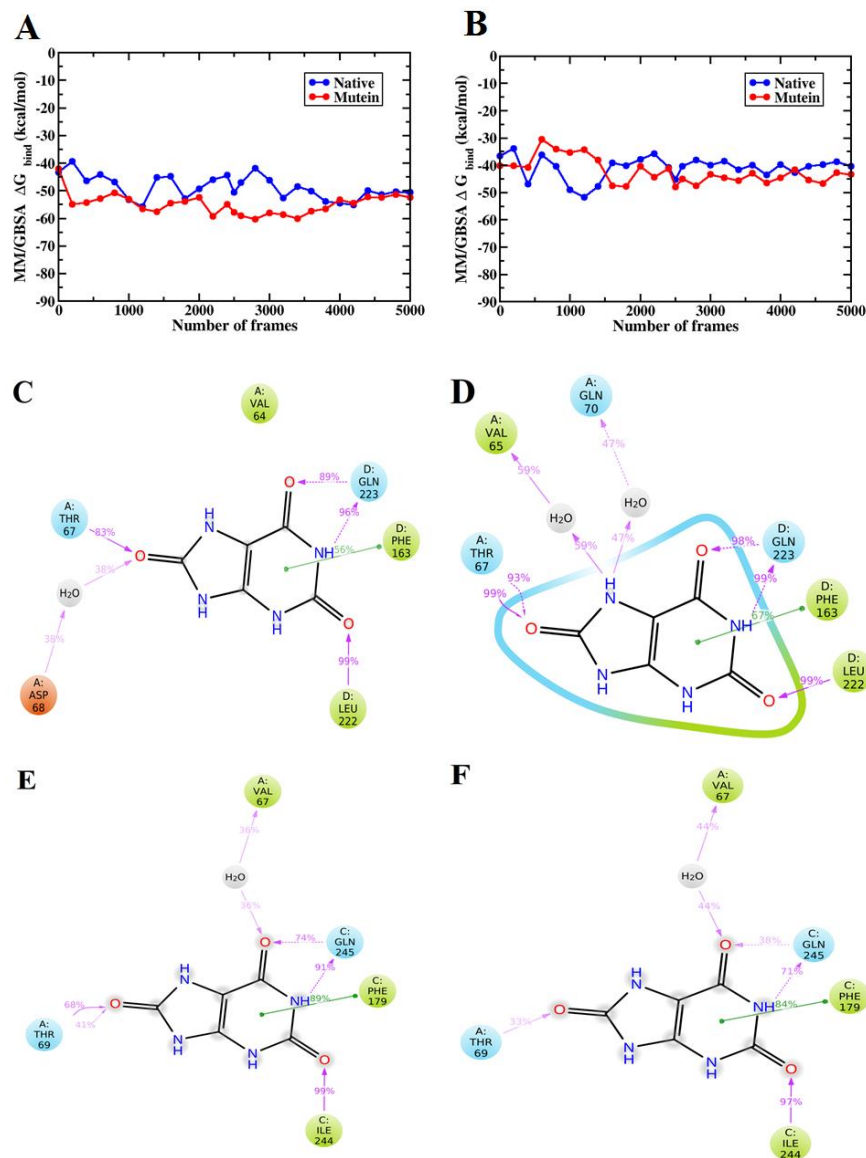
Figure 5.9: **(**A) The change of the binding free energy of uric acid in 2YZB throughout the 100 ns MD simulation. (B) The change of the binding free energy of uric acid in 4R8X throughout the MD simulation. The interaction percentage between the uric acid and the amino-acid residues at the binding pocket of uricase is represented. (C) & (D) Showing the interaction in case of wild and mutated *Ag*-Uricase. (E) & (F) Showing the interaction in case of wild and mutated *Bf*-Uricase

It is clear from the above discussion that fluctuation of the B-cell epitopes reduces after mutation. In contrast, the fluctuation of T-cell RMSF increases after mutation, although negligible. These phenomena indicate that mutated 2YZB has less susceptibility to interact with antibodies and CD4+ cells. Residues 287, 599, 864 are found to be more fluctuated and away from the antigenic region. In the case of both native and mutated 4R8X, the average RMSF value is found to be almost similar (around 0.668 Å), indicating the stability of both structures (Figure 5.8-D). However, the flexibility of the mutated region of 4R8X is found to be higher compared to its native type. The average RMSF of T-cell epitope 166-180 displayed a higher value (1.94 Å) compared to its native form (0.49 Å). The B-cell epitope region (137-142) showed similar (0.75 Å) fluctuation in mutated from compared to its native form (0.75 Å) (Figure 5.8-D). However, the increment of the flexibility of antigenic epitope is comparable with its native form. The higher fluctuations are observed at the C-terminal and N-terminal ends of uricase due to solvent exposure. The time evolution of compactness or the overall size of uricase was measured by means of the radius of gyration (R_g) and is illustrated in Figure 5.8-E, F. It is found from Figure 5.8-E that both the native and mutated uricase backbone of 2YZB is found to be stabilized at an average R_g score of 30.3 Å. Similarly, in the case of 4R8X, both the native and mutated uricase backbone maintained an average R_g value of 31.8 Å (Figure 5.8-F). The closeness of R_g value in both the native and mutein uricase indicates that the compactness of the protein backbone was unchanged after the mutation in the antigenic region. The secondary structure contents of both the species are found to be similar to their corresponding mutant model (Figure A.III.5-A, B, C, and D). Hence, the mutations are not responsible for the remarkable secondary structure loss in uricase.

The time evolution of the binding free energy of uric acid at the active site of uricase is presented in Figure 5.9-A, B. In both species, the binding free energy is found to be stabilized at the last of the MD trajectory. This further indicates the stability of the ligand at the catalytic pocket of uricase after mutation. The average binding energy of uric acid at the catalytic pocket of uricase is found to be -48.71 kcal/mol and -40.93 kcal/mol for 2YZB and 4R8X, respectively. The average binding energy difference

(between the mutated and wild type uricase model) is found to be -6.36 kcal/mol for 2YZB and -1.45 kcal/mol for 4R8X. During the course of simulation, four stable hydrogen bonds are found to be intact among the docking predicted hydrogen bonds between 2YZB and uric acid. The histograms illustrated in Figure A.III.6-A reveal that the regions Glu162-Arg180, Leu222-Asn249, and Val64-Asp68 have major interaction with uric acid during the simulation. After mutation of 2YZB, there was no substantial change in the hydrogen bonding interaction profile with the substrate. The interaction of Gln70 with the uric acid has increased, whereas the interaction with Asp68 is found to be decreased (Figure A.III.6-B). It is clear from Figure 5.9-C and Figure A.III.5-A that Thr67, Leu222, and Gln223 formed hydrogen bonding interaction for 83%, 99%, and 96% of the trajectory. One water-mediated hydrogen bond is observed with Asp68 for 38% of the simulated trajectory (Figure 5.9-C). It can be found from Figure A.III.6-A, B that hydrogen bonding interactions between uric acid and Thr67, Leu222, Gln223 are similar after mutation. One extra water-mediated hydrogen bond is present, which anchors the ligand with Gln70 for 47% of the trajectory (Figure 5.9-D). The contacts between uric acid and the interacting amino acid residue in both the wild and mutated 2YZB are found to match with the plot of interaction fractions (Figure A.III.7-A, B). In the case of native 4R8X, the main interacting regions are Val67-Asp70 and Ile244-Asn271 (Figure A.III.6-C). The non-bonded interactions between binding pocket residues and uric acid are found to be unchanged after mutation (Figure A.III.6-D). The NH and O atoms of the six-membered ring of uric acid formed two hydrogen bonds with Gln245 for 74% and 91% of the simulated trajectory (Figure 5.9-E). Additionally, Ile244 formed a hydrogen bond with uric acid for 99% of the trajectory. In the case of mutated 4R8X, Gln245 formed a hydrogen bond with 71% of the trajectory, and Ile244 formed a hydrogen bond with 97% (Figure 5.9-F). The contacts between uric acid and Val67, Thr69, Phe179, Ile244, and Gln245 are found to be similar in both the native and mutated 4R8X (Figure A.III.7-C, D). Therefore, it is clear from the MD simulation results that the enzyme variants remain stable after mutating immunogenic amino-acids, distributed throughout the protein surface without much change in their catalytic activity.

Using *in silico* techniques, the study focused on protein engineering as an alternative to PEGylated enzymes, with reduced immunogenicity as an intrinsic character of protein. Practical execution of the results obtained from *in-silico* analysis will be a time consuming and costly process. Using *in-silico* analysis, linear B-cell, conformational B-cell and MHC-I based T-cell epitopes were identified and reduced the immunogenicity of uricase sourced from *Bacillus fastidious* and *Arthrobacter globiformis.* This is the first work on *in-silico* identification of epitopes, hot-spot residues of uricase. Conjugating the drugs with polyethylene glycol (PEG) is one of the most successful approaches to address the problem of clinical immunogenicity. However, polyethylene glycol (PEG) coating of therapeutic protein reduces the efficiency by increasing the protein's size and water absorption properties. The combination of polyethylene glycol with uricase was reported as the first clinical study to successfully reduce plasma uric acid concentration.

5.3 SUMMARY

The clinical application of uricase as an anti-hyperuricemia agent is limited due to antigenicity problem. In order to generate less immune reactive therapeutic drug, *in-silico* mutagenesis of B-cell and T-cell epitope has been proposed. Multiple sequence alignment of thirteen uricases from different sources was performed to identify the conserved sequence. Out of the six motifs obtained, three were found to be common for all uricase producers. Motif2, motif3, motif4 are expected to preserve most of the structural and functional aspects of uricase. In the case of *Ag*-Uricase, the epitopic peptide ^{167}PR**D**KYT^{172} was found to be highly surface accessible due to the presence of Asp at 169 position. According to the Parker hydrophilicity method, the peptide sequence ^{261}GQD**N**PNE^{267} had the highest antigenic probability due to the placement of Asn at 264 position. Peptide sequence ^{156}LKS**T**GSE^{162} was also considered as immunogenic due to the relatively higher flexibility. Similarly, two epitopic peptides ^{137}RK**S**RNE^{142} and ^{212}DDA**K**GDN^{219} were found to influence antibody secretion in the human body for *Bf*-Uricase. The antigenic property of these epitopes was high due to the presence of Ser at 139 and Lys at 215 positions, respectively. Deimmunization studies were carried out to locate the T-cell epitopes for both

species. The epitopic peptides ^{201}AV**Y**ASVRGLLLKAFA^{215} and ^{166}IADIQL**I**KVSGSSFY^{180} were found to have a high propensity to activate CD4$^+$ cells for both species. Four hot-spot amino acid residues were identified for each monomer of uricase. The maximal reduction of immunogenicity was obtained for T159W, D169C, N264W, and Y203D mutations in *Ag*-Uricase and S139V, K215W, G216F, I172P mutations in *Bf*-Uricase. All the amino acid mutations were done in non-conserved and moderately conserved region of the uricase sequence, which is less likely to alter the structural and functional characteristics of the therapeutic drug. The stabilization in the binding affinity of uric acid in mutein model of uricase confirmed that their catalytic activity is unchanged. The MD simulation indicates that both the muteins are stable and they preserve their native-like structural characteristics. The insights obtained from the study provide a guideline for the experimental development of uricase drug for treating gout and related diseases.

CHAPTER 6

BIO-CONJUGATION OF THERAPEUTIC ENZYME URICASE WITH BSA: AN EXPERIMENTAL INVESTIGATION

Intake of native uricase can trigger an immune response in the human immune system. Consequently, the human immune system considers native uricase as a foreign substance and produces antibodies against the enzyme. Due to this adversity, the half-life of the enzyme becomes shorter and reduces the treatment's efficiency. To overcome these adverse effects, the enzyme surface can be covered with a polymer, such as polyethylene glycol (PEG), by a bioconjugation process. Pegloticase is a commercially available conjugated uricase with polyethylene glycol for the treatment of gout. Still, the usage of polymer conjugated uricase has certain limitations. About 77% of Pegloticase treated patients showed adverse effects such as gout flares and infusion reaction, and also it is strictly prohibited to take pegloticase during pregnancy (Edwards 2008; Nanda and Jagadeesh Babu 2017; Sherman et al. 2008).

In this study, uricase was conjugated with bovine serum albumin, which is also a protein. BSA has been used as a potential bio-conjugating compound with different enzymes like amylase, L-Asparagines, and catalase (Hu and Su 2002; Mohan Kumar et al. 2014). Though literature says that BSA can be successfully used as bio-conjugating element, there is no study or report on Uricase's bio-conjugation with BSA to improve its properties. This is the first experimental study in which uricase from *Bacillus fastidious* is modified with BSA. In this study, variables like the BSA concentration, glutaraldehyde concentration (cross-linker), pH, and temperature were optimized to achieve the desired degree of conjugation with desired residual activity. Further, the conjugate's stability with respect to temperature and pH was assessed, and the kinetic parameters were analyzed and discussed.

6.1 MATERIALS AND METHODS

6.1.1 Materials

Uricase from *Bacillus fastidious* (specific activity: 9 U/mg), uric acid sodium salt, 2,4,6- trinitrobenzene sulfonic acid (TNBSA) were purchased from Sigma-Aldrich chemical company. Bovine Serum Albumin (BSA), Sodium chloride, Glycine, molecular weight markers, Bradford reagent, Dialysis membrane (20 kDa cutoff) were obtained from Hi-Media Laboratories, Mumbai. Millipore water was used throughout the experiment. Glutaraldehyde, Ultra-filtration tubes 100 kDa (Millipore Amicon Ultra-4) was supplied by Merck Millipore.

6.1.2 Preparation of Uricase-BSA conjugates

Uricase enzyme solution (1mg/mL), uric acid solution (0.5mg/ml) and BSA solution of different concentration (1, 1.5, 2, 4, 6, 8 and 10 (mg/mL)) were prepared using 100mM borate buffer (pH 9). Uric acid is used as an active site protector during bio-conjugation reaction. Glutaraldehyde solutions at different concentration (0.25%, 0.5%, 0.75%, 1%, and 1.25%) were prepared and used for bio-conjugation reaction. For conjugation of uricase with bovine serum albumin, 0.5ml of enzyme solution, 0.5mL of uric acid solution, and 0.5ml of BSA solution were added and kept at 4°C for 5mins. Different concentration ratios of uricase vs. BSA like 1:1, 1:1.5, 1:2, 1:4, 1:6, 1:8 and 1:10 were used to optimise the reaction. Then, 50μL of glutaraldehyde solution was added as a crosslinking agent and incubated at 4°C with gentle mixing for 2hrs. Glutaraldehyde concentration was optimised by using different ratios of uricase-glutaraldehyde like 1:0.25, 1:0.5, 1:0.75, 1:1, and 1:1.25. After incubation, glycine was added twice the BSA amount, to stop the cross-linking reaction (Kishore et al. 2014).

0.5mL of uricase + 0.5mL of uric acid + 0.5 mL of BSA

$\downarrow$ Mixing for 5mins at 4℃

Addition of 50μL glutaraldehyde

$\downarrow$ Mixing for 2 hrs at 4℃

Addition of glycine (twice the amount of BSA)

$\downarrow$

Dialysis against 1% NaCl and 1% glycine

$\downarrow$

Ultracentrifugation at 5000rpm for 20 mins
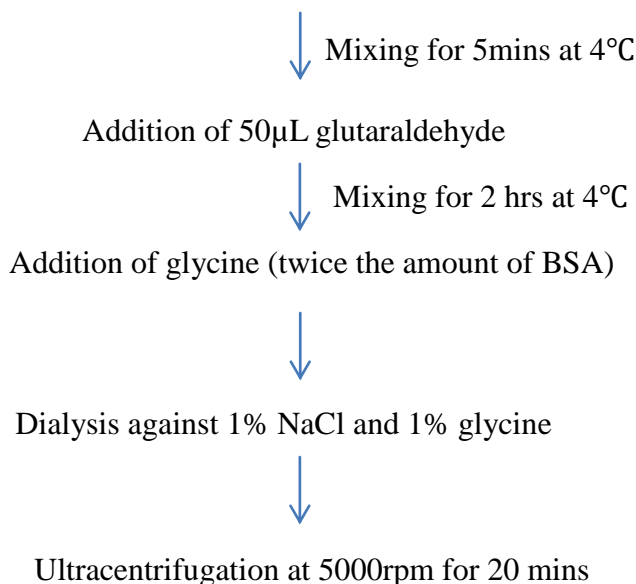
Figure 6.1: Schematic representation of method

6.1.3. Purification of bioconjugates

After conjugation, the sample was subjected to dialysis against 1% glycine and 1% NaCl using a dialysis membrane (20 kDa molecular cutoff) for about 24hrs. Before initiating dialysis, the membrane was activated by immersing it in 5g of sodium bicarbonate and 0.073g of EDTA solution and incubated at 80°C for 20 mins. After incubation, the membrane was washed with 25% ethanol solution and washed twice with distilled water. During dialysis, unreacted uric acid (168.11 Da), excess glycine (75.07 Da), and glutaraldehyde (100.11 Da) were removed. After 24hrs of dialysis, the sample was collected and centrifuged using 100 kDa molecular weight cut-off filter at 5000rpm for 20mins to remove unreacted bovine serum albumin (66 kDa). The supernatant was discarded, and the concentrated sample was collected for further assay.

6.1.4 Determination of enzymatic activity

Enzyme activity was measured aerobically by observing the reduction absorbance at 293 nm due to the enzymatic oxidation of uric acid. Uricase activity was estimated as

described previously (Mahler 1970). The reaction mixture contained 3ml of 20 mM sodium borate buffer (pH 9.0), 75 µl of 3.57 mM uric acid, and 20 µl of uricase sample. In blank, instead of an uricase enzyme, 20 µL of buffer was used. An International Unit (IU) of uricase is defined as the amount of enzyme required to transform 1 micromole of uric acid into allantoin in one minute under the standard assay conditions of pH 9.0 at 25°C. The protein concentration was estimated according to the method described by Bradford (1976) with bovine serum albumin as standard (Bradford 1976; Nanda and JagadeeshBabu 2016).

6.1.5 Electrophoresis (SDS-PAGE)

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis was conducted according to the method of Laemmli (1970). Briefly, 12% acrylamide solution (pH 8.8) and 5% bisacrylamide solution (pH 6.8) were used as resolving and stacking gels. Electrophoresis was performed at a constant volt of 75 V, and 20 mA current and the gel was stained with Coomassie brilliant blue R-250 dye in acetic acid/ethanol/water solution (1:4:5) and destained in the same staining solution without Coomassie brilliant blue till the bands were clearly visible. The samples were calibrated against a wide range of molecular weight markers (Laemmli 1970; Nanda et al. 2016).

6.1.6 Optimization of bovine serum albumin for bioconjugation

Studies showed that bovine serum albumin concentration affects the activity of the enzyme with which it was conjugate (Gowda et al. 2014). So, to optimise the bovine serum albumin concentration, bovine serum albumin was conjugated with uricase at different concentrations (uricase (mg/mL) : BSA (mg/mL)) in the ratio of 1:1, 1:1.5, 1:2, 1:4, 1:6, 1:8 and 1:10 at 4°C. After conjugation and purification, enzyme assay and protein estimation was performed to all the samples to determine the highest activity of the enzyme.

6.1.7 Optimization of glutaraldehyde for bioconjugation

After determining the optimum bovine serum albumin concentration at which the enzyme shows the highest activity, by keeping that BSA concentration as constant, conjugation was performed by varying glutaraldehyde concentrations as (uricase (mg/mL) : glutaraldehyde (%)) 1:0.25, 1:0.5, 1:0.75, 1:1, 1:1.25 at 4°C (Zhang et al.

2005). After conjugation and purification, enzyme assay and protein estimation was performed to all the samples to determine the highest activity of the enzyme.

6.1.8 Degree of modification

The degree of modification of ε-amino groups was determined by measuring the number of free amino groups using the trinitrobenzene sulfonic acid (TNBS) method, according to the literature. In short, 1ml of sodium borate buffer (20 mM, pH 9.0), 1ml of $NaHCO_3$ (4 w/v %), and 1ml of aqueous (0.01 w/v %) TNBS solution was added to 1 ml of aliquot of the reaction mixture and color development was allowed to proceed for 2 h at 40 °C. The reaction was terminated by adding 4 ml of 10% sodium dodecyl sulfate and 50 ml of 1N HCl, and the absorbance was recorded at 340 nm at room temperature (Habeeb 1966; Hu et al. 2018). Finally, the amino group consumption by the reaction was calculated as follows:

$$\% \text{ Substitution} = \left(\frac{\text{Absorbance of native protein} - \text{Absorbance of conjugate}}{\text{Absorbance of native protein}}\right) * 100$$

6.1.9 Stability analysis

Effect of both temperature and pH on the stability of uricase was analyzed as follows. The effect of temperature on free and conjugated enzyme activity was evaluated at pH 9.0 by incubating the sample at different temperatures (20°C, 30°C, 40°C, 50°C, and 60°C). The effect of pH on free and conjugated enzyme activity was determined at ambient temperature.

6.1.10 Kinetic analysis

The apparent Km of the free and conjugated uricase was determined using the Lineweaver- Burk Plot method. The enzyme activity was assayed with varying concentration of uric acid (0.5-4.5 μM) in sodium borate buffer (mM, pH 9.0) at 25 °C and the Km was estimated by the double reciprocal plot according to the Michaelis-Menten equation based on uric acid disappearance rate evaluated at 292 nm (Lineweaver and Burk 1934; Punnappuzha et al. 2014; Wu et al. 2016).

6.2 RESULTS AND DISCUSSION

6.2.1 BSA optimization

Studies have shown that conjugation with bovine serum albumin (BSA) can affect enzyme activity (Kishore et al. 2014). In this study, BSA concentration was optimized to achieve the required degree of conjugation with maximum residual activity. From Table 6.1, it has been observed that we could able to achieve maximum residual activity of 91.85% while using 6mg/mL of BSA concentration. The higher concentration of BSA (>6mg/mL) led to a reduction in enzyme activity due to reduced availability of surface accessible active sites. For further optimization, 6mg/mL BSA concentration was used.

Table 6.1: Uricase activity with respect to BSA concentration

BSA Concentration	% Residual activity	Activity
1mg/ml	41.38	3.91 U/mg
1.5mg/ml	49.31	4.66 U/mg
2mg/ml	51.53	4.87 U/mg
4mg/ml	53.12	5.02 U/mg
6mg/ml	91.85	8.68 U/mg
8mg/ml	53.86	5.09 U/mg
10mg/ml	33.76	3.19 U/mg

6.2.2 Glutaraldehyde optimization

Glutaraldehyde acts as a cross-linking agent in the bioconjugation of uricase with BSA, where the Glutaraldehyde concentration significantly affects the activity of the

enzyme (Zhang et al. 2005b). Glutaraldehyde concentration was optimised to achieve maximum possible residual activity by varying its concentration as follows 0.1%, 0.25%, 0.5%, 0.75%, 1% and 1.25%. The highest enzymatic activity of 91.85% was achieved at 0.5% concentration, but beyond 0.5% concentration, activity was reduced**.** Figure 6.1 shows the effect of glutaraldehyde concentration on enzymatic activity. Increasing and decreasing trends of activity could be due to the limitation in the surface accessible active sites and limitation in mass transfer between the conjugate's inner and outer regions. Beyond 0.5%, due to the increase in these limitations, residual activity reduces (Zhang et al. 2005b).

Table 6.2: Enzyme activity with respect to glutaraldehyde concentration

Glutaraldehyde concentration	% Residual activity	Activity
0%	67.2	6.35 U/mg
0.25%	80.32	7.59 U/mg
0.5%	91.85	8.68 U/mg
0.75%	51.64	4.88 U/mg
1%	41.38	3.91 U/mg
1.25%	40.74	3.85 U/mg
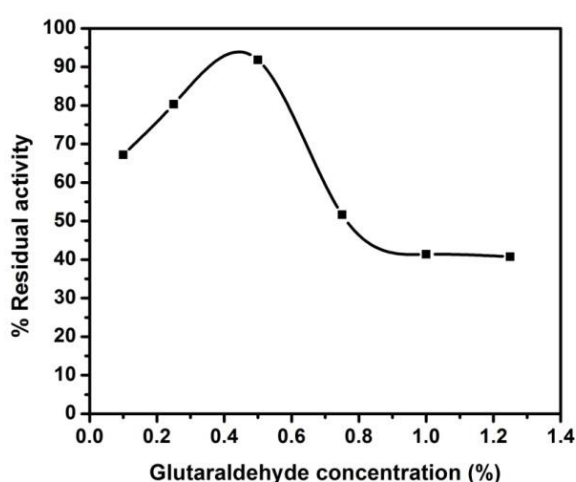
Figure 6.2: Effect of glutaraldehyde concentration on uricase activity

6.2.3 Determination of molecular weight of the conjugate by SDS-PAGE

To confirm the success of conjugation, SDS PAGE analysis was performed. Figure 6.3 shows the SDS PAGE gel stained in a commasive blue solution. Lane 1 represents the protein marker, Lane 3 represents native uricase, and Lane 4 represents Bovine serum albumin. Lane 6 to 10 represents conjugates prepared with different molar ratios of uricase and BSA (1:2, 1:4, 1:6, 1:8, 1:10). From the figure, it is clear that the conjugation between uricase and BSA has happed successfully. Since the conjugated sample is partially purified, bands of higher molecular weight conjugates were also observed in the lanes. It is further observed that the thickness of the bands in Lane 6-10 is high compared to the thickness of the bands in Lane3 and Lane4. This could be due to the high concentration of the conjugated enzyme in the sample.
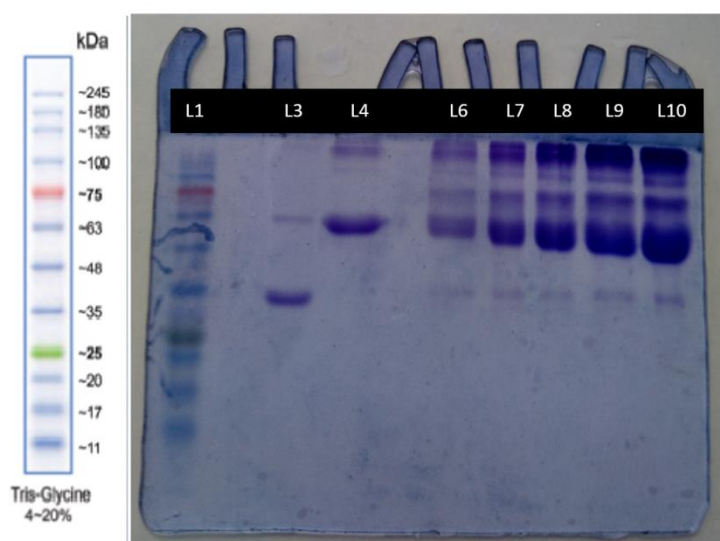
Lane 1: Protein marker; Lane 3: Native uricase; Lane 4: Bovine serum albumin; Lane 6 to 10: uricase:BSA conjugates with the concentration ratio of (1:2, 1:4, 1:6, 1:8, 1:10).

Figure 6.3: Confirmation of conjugate by SDS gel electrophoresis

6.2.4 Degree of substitution

TNBSA (2, 4, 6 trinitrobenzene sulfonic acid) assay reacts with primary amines, and that can be used to determine the free amino groups present in the enzyme (Degree of substitution). Table 6.3 shows an increase in the percentage substitution as the concentration of Bovine Serum Albumin increases. This increasing trend was also

observed in the SDS-PAGE analysis, where the band's width increases as the BSA concentration increases. Both enzymatic activity (91.85%) and the substitution percentage were higher in the 1:6 (Uricase: BSA) ratio. Therefore, an Uricase: BSA (mg/ml) ratio of 1:6 was chosen as the best composition for bioconjugation. Similar results were observed in the literature for L-asparaginase-BSA conjugate, where the degree of substitution was reported as 64 % (Mohan Kumar et al. 2014; Sashidhar et al. 1994).

Table 6.3: Degree of Modification

Uricase:BSA(mg/ml)	% Substitution
1:2	71.12
1:4	74.36
1:6	76.69
1:8	78.24
1:10	79.31

6.2.5 Stability Analysis

Unmodified uricase and modified uricase were incubated at different temperatures (20℃, 30℃, 40℃, 50℃, 60℃) for 1 hour to check the thermal stability of the enzyme (da Silva Freitas et al. 2010). The enzymatic assay analysis showed that the native uricase is more stable at lower temperatures (20℃, 30℃, and 40℃), but at temperature 50℃ and 60℃, modified uricase retained its activity slightly higher than an unmodified uricase. This could be due to two reasons; one is conjugation, i.e., BSA's chemical binding has improved the thermal stability of uricase, and second is the masking of uricase with BSA has given physical barrier for uricase. Residual activity of native uricase and conjugated uricase corresponds to their temperature are listed in Figure 6.4. Literature shows that at 50℃ ± 2℃, the native and BSA-α-Amylase enzyme could be more stable, which supported our findings (Kishore et al. 2014).
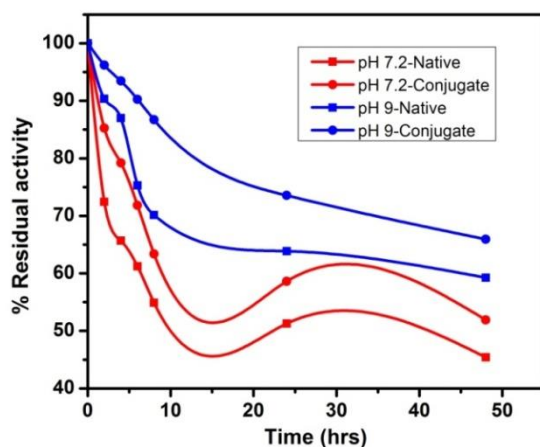
Figure 6.4: Stability of modified and unmodified uricase at different pH
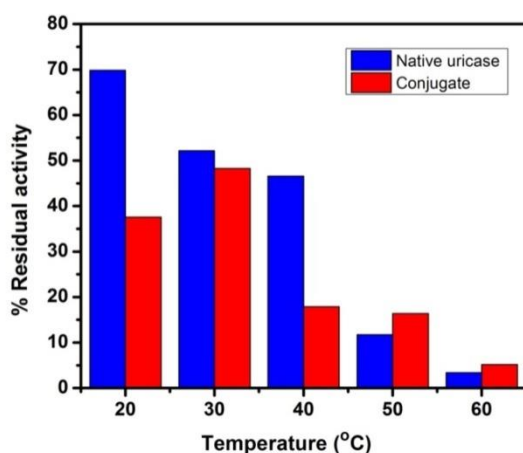
Figure 6.5: Stability of modified and unmodified uricase at different temperature

The effect of pH on native and modified uricase was studied at both physiological pH (7.2) and optimum pH (9.0). The optimum pH of the enzyme uricase from *Bacillus fastidious* is 9.0. It is inevitable that the conjugate should be stable at physiological pH. The graph shows that the activity of modified uricase was maximum than the native one at physiological pH 7.2 (Figure 6.4). This indicates that modified uricase becomes stable when exposed to a physiological atmosphere, a privilege for pharmaceutical application. On the other hand, exposure of both native and modified uricase to optimum pH 9.0 shows better retention in enzyme activity after 48 hrs of

incubation, which clearly indicates the steady the decrease in the enzyme activity (Figure 6.4).

6.2.6 Enzyme kinetics

Michaelis–Menten kinetic constant 'Km' determines the enzyme uricase's affinity to its substrate uric acid. Km values were obtained from non-linear regression analysis (Figure 6.6). Unmodified uricase showed a Km value of 6.6135 µM/mL, whereas modified uricase showed a Km value of 6.4639 µM/mL, which is slightly higher, which could be due to the modification of enzyme with bovine serum albumin (Punnappuzha et al. 2014). The residual activity retained after conjugation and its stability could make it as an effective drug delivery system.
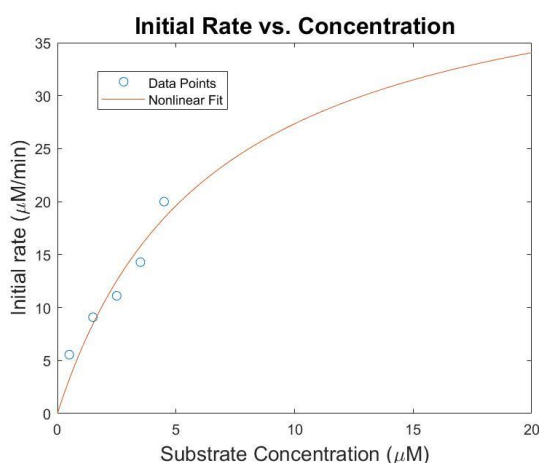
Figure 6.6: Kinetic parameters plot of native uricase (Non-linear regression analysis)
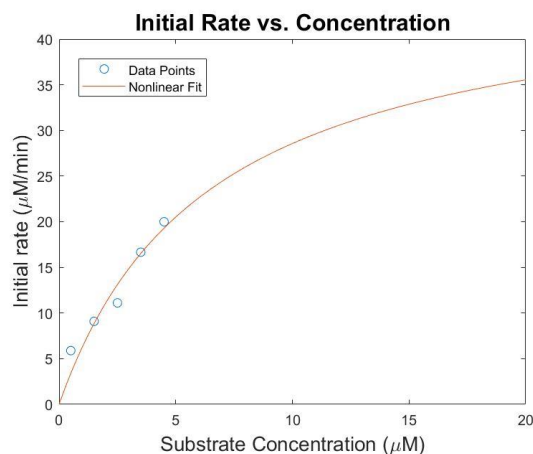
Figure 6.7: Kinetic parameters plot of conjugate (Non-linear regression analysis)

6.3 SUMMARY

Uricase from *Bacillus fastidious* was successfully conjugated with bovine serum albumin to achieve better therapeutically property. Different bovine serum albumin and glutaraldehyde ratios were conjugated with uricase, where 1:6 (mg/ml) uricase: BSA ratio and 0.5% glutaraldehyde concentration showed a maximum enzymatic activity of 91.85% after conjugation. The degree of modification, which was determined using the TNBSA assay, showed that 1:6 ratio could give the maximum activity of 76.69%. SDS gel electrophoresis was performed to analyze the molecular weight distribution in the conjugated sample. The conjugated and native uricase stability was compared at different temperatures (20-60℃) for 72 hrs. Stability analysis showed that native uricase was more stable at physiological temperature whereas conjugated uricase was more stable at higher temperatures, i.e., 50℃ and 60℃. Similarly, pH stability was studied at pH of 7.2 and 9.0. Both native and modified uricase at optimum pH 9.0 shows better retention in enzyme activity after 48 hrs of incubation, which indicates a steady decrease in enzyme activity. The findings of this study suggest that conjugated uricase is sustained effectively at the physiological condition, which could make it a potent drug to treat hyperuricemia.

CHAPTER 7

7. SUMMARY AND CONCLUSION

7.1 SUMMARY

Heterologous enzymes are believed to be a significant source of biopharmaceuticals due to their substrate specificity (De Duve 1966). Therapeutic enzymes were therefore used extensively to cure variety of genetic and acquired human diseases by the removal of disease causing metabolites (Shen and Shen 2006; Tan et al. 2010). In addition, high catalytic efficiency, high purity, greater affinity, unique selectivity, and good pharmacokinetics properties of these enzymes improve their utility in the current medical arena. Uricase or Urate oxidase (E.C 1.7.3.4) is an enzyme that catalysis the oxidation of uric acid into 5-hydroxy-isourate and subsequently forms allantoin (Colloc'h and Prangé 2014). In the human body, the elimination of uric acid occurs in the kidney (70%) and gastrointestinal tract (30%). The excess of uric acid is associated with the pathogenesis of systemic hypertension (Htn), obesity, chronic kidney disease (CKD), cardiovascular disorders, hyperuricemia, metabolic syndrome (MS), type 2 diabetes mellitus (DM), non-alcoholic fatty liver disease and coronary heart disease (CHD) (Sharaf El Din et al. 2017). An imbalance in the production rate and excretion rate of uric acid in humans is known as hyperuricemia. In humans, longstanding hyperuricemia results in deposition of monosodium urate monohydrate crystals in the joints and tissues leading to a condition called Gout that causes inflammatory pain (Sherman et al. 2008). Anti-hyperuricemia drugs employed to control the concentration of uric acid in serum, urine or soft tissues have several associated side effects. The use of the therapeutic enzyme, uricase, is considered as an alternative for the regulation of uric acid in the serum, urine and biological fluids without much side effects (Khade and Srivastava 2015a). Uricase enzyme is present in bacteria, fungi, plants and animals, genetically engineered microorganisms and yeast (Chen et al. 2008; Ishikawa et al. 2004).

While uricase has been used to treat hyperuricemia and tumour lysis syndrome, it has been associated with allergic, hypersensitivity, and anaphylactic reactions in patients (Ali and Lally 2009). Uricase is growing importance is likely due to its potential use in medicinal chemistry and the treatment of a variety of diseases. Bacterial organisms such as *Arthrobacter globiformis*, *Bacillus fastidious* were used for industrial production of the uricase enzyme. These bacterial enzymes were commercialized and has been used for various applications (Tan et al. 2012; Zhao et al. 2006). Presently, the uricase used for therapeutic purpose are mainly sourced from bacteriadue to high specific activity. Bacteria are the major sources of therapeutic enzymes (Valderrama-Rincon et al. 2012). Further, uricase is more expensive in comparison to allopurinol which is the first choice of drug administered during conventional treatment of hyperuricemia. Therefore, the use of uricase as a therapeutic drug is highly restricted. Hence, there is a pressing need for uricase produced from various sources to be cost-effective and have qualities that helps to extend its utility.

Keeping the capacity of the enzyme in view, especially in the treatment of hyperuricemia, the present work is planned to evaluate an overview of the computational characterization of uricase protein sequences from different sources using bioinformatics tools. Also deals with the complete description of the structural and functional aspects of various *Bacillus* species having uricase activity using bio-computational web-based servers and tools. These computational approaches can be used for the screening and investigation of an uricase enzyme with desirable characteristics that can be employed in diverse industrial applications. Due to its high immunogenicity and short half-life, Rasburicase therapy is stated to be limited (Coiffier et al. 2003; Garay et al. 2012). Additionally, the therapeutic potential of recombinant uricase for the treatment of gout is associated with pharmacologic tolerance and potency problem (Guttmann et al. 2017). The antigenic determinants of uricase were identified for reducing the immunogenicity via *in-silico* mutagenesis. The uricase enzyme was conjugated with bovine serum albumin (BSA) in order to improve its therapeutic properties.

Elucidating the structural and physiochemical properties of uricase by *insilico* analysis was carried out. A total number of sixty amino acid sequences of uricase belongs to different sources were obtained from NCBI and different analysis like multiple sequence alignment (MSA), homology search, phylogenetic relation, motif search, domain architecture and physiochemical properties including pI, EC, Ai, and Ii, were performed. Multiple sequence alignment of all the selected protein sequences exhibited the distinct difference between bacterial, fungal, plant and animal sources based on the position-specific existence of conserved amino acid residues. The maximum homology of all the selected protein sequences is between 51-388. In singular category, homology is between 16-337 for bacterial uricase, 14-339 for fungal uricase, 12-317 for plants uricase, and 37-361 for animals uricase. The phylogenetic tree, constructed based on the amino acid sequences disclosed different clusters indicating uricase from a different source. The physiochemical features revealed that the uricase amino acid residues are in between 300-338 with a molecular weight as 33-39 kDa and theoretical pI ranging from 4.95-8.88. The amino acid composition results showed that valine amino acid has a higher average frequency of 8.79 percentage compared to different amino acids in all analyzed species.

Computational-based structural, functional, and phylogenetic analyses of uricase enzymes from various *Bacillus* species were performed. Seventy uricase protein sequences from *Bacillus* species were selected for multiple sequence alignment, phylogenetic analysis, and motif assessment, domain architecture examination, understanding of basic physicochemical properties and *in silico* identification of the composition of amino acids in uricase. Further, structural (secondary and tertiary structure prediction), and functional (CYS_REC, MOTIF scan, CD-search, STRING, SOSUI, and PeptideCutter) analysis of uricase were performed. *Bacillus simplex* (WP_063232385.1) was chosen as the representative species of the *Bacillus* genera. The three-dimensional (3D) structure of *Bacillus simplex* uricase is predicted and validated using QMEAN, RAMPAGE, ERRAT, Verify 3D and PROQ servers. Analysis revealed that the tertiary structure of the selected uricase has good quality and acceptability. Various *Bacillus* sources revealed that all the selected *Bacillus* uricases are active within acidic to a neutral environment, thermally stable with

molecular weight ranging from 35.59-59.85 kDa. The secondary structure analysis showed that all the uricases are rich in alpha helices and sheets. The CDD tool identified two conserved domains, one of which belongs to OHCU decarboxylase and another to Uricase superfamily. The quality estimation of 3D modelled protein gave a high overall quality factor score of 94.64. Also, all *Bacillus* species of uricase enzyme and their corresponding genes showed a strong correlation from the phylogenetic comparison of the selected taxa.

The clinical utilization of uricase against gout is limited due to the immunogenicity. The antigenic determinants of uricase were identified and reduced their immunogenicity via *in-silico* mutagenesis. Multiple sequence alignment and motif analysis werecarried out to identify the conserved residuesin evolutionary process. Emini surface accessibility, Parker hydrophilicity, and Karplus & Schulz flexibility methods were employed to predict the linear B-cell epitopes of both *Ag*-uricase and *Bf*-uricase. Deimmunization approach identified T-cell epitopes and the hot spot residues. Reduced antigenic probability was obtained in case of T159W, D169C, N264W and Y203D mutations for *Ag*-uricase, while S139V, K215W, G216F and I172P mutations for *Bf*-uricase. The binding affinity values of uric acid towards the catalytic pocket of *Ag*-uricase and *Bf*-uricase models were found to be -48.71 kcal/mol and -40.93 kcal/mol, respectively. This energy is further stabilized in the mutant model by -6.36 kcal/mol and -1.45 kcal/mol for *Ag*-uricase and *Bf*-uricase, respectively. About 100ns molecular dynamics simulation was performed to evaluate the conformational stability of both native and mutated uricase.

Uricase from *Bacillus fastidiosus* was conjugated with bovine serum albumin. Different ratios of bovine serum albumin and glutaraldehyde was conjugated with uricase and 1:6 (mg/ml) uricase: BSA ratio and 0.5% glutaraldehyde was considered as the optimum BSA concentration for bioconjugation as it retained 91.85% activity after conjugation. Therefore, bioconjugation of uricase with a natural protein BSA retained highest activity than the conjugation of uricase with polymers. The degree of modification of 1:6 ratio conjugate was found to be 76.69%. SDS gel electrophoresis was performed to confirm the presence of conjugate. Stability of conjugated uricase

and native uricase were compared from the temperature range (20°C, 30 °C, 40°C, 50°C, 60°C) for 1 hr and found that conjugated and native uricase was more stable at physiological temperature whereas conjugated uricase was more stable at higher temperatures 50°C and 60°C when compared to native uricase. On the other hand, exposure of both native and modified uricase to optimum pH 9.0 shows better retention in enzyme activity after 48 hrs of incubation, which clearly indicates the steady the decrease in the enzyme activity. Kinetic parameters (K_m and V_{max}) were determined to study native and modified uricase affinity to its substrate uric acid.

7.2 CONCLUSION

- Computational studies of amino acid sequences of various organisms like bacteria, fungi, yeast, plants, and animals were correlated successfully on the basis of multiple sequence alignment, phylogenetic tree construction, domain identification, and individual amino acid composition.

- These sequences from various sources like bacteria, fungi, yeast, plant, and animals have revealed sequence-based similarities that differed depending on the sources.

- The physicochemical features of all the selected uricase proteins showed differences with their respective groups.

- It was found that the amino acid valine plays a crucial role in uricase composition as the average frequency of valine in all of the selected sources was 8.79 %, which was very high when compared to other amino acids.

- This work might be informative and a stepping-stone to other researchers to get an idea about the physicochemical features, evolutionary history, and structural motifs of uricase that can be widely used in biotechnological and pharmaceutical industries.

- All *Bacillus* uricase proteins were active in acidic to neutral environments and thermally stable, with the molecular mass ranging between 35.59-59.85 kDa as found by the physicochemical *in-silico* analysis and can be used as suitable candidates in the medical industry.

- The secondary structure analysis displayed a high alpha-helical conformation, which showed a more stable secondary structure as compared to other secondary structure elements in all *Bacillus* uricases.

- The detailed computational structural and functional investigation on the uricase protein could help in screening a suitable uricase-producing microbe with desirable characteristics for industrial application.

- This is the first study on *in-silico* identification of epitopes, hot-spot residues of uricase. Reduced immunogenicity for *Ag*-uricase by T159W, D169C, N264W, and Y203D mutations and S139V, K215W, G216F, and I172P mutations for *Bf*-uricase.

- The insights obtained by *in silico* approach for reduced clinical immunogenicity can serve as a guideline for the experimental development of uricase drug for treating gout and related diseases.

- Under optimized conditions, conjugated uricase obtained by chemical modification with BSA showed maximum residual uricase activity of 91.85% compared to native uricase.

- It was found that conjugated uricase is stable at higher temperature and physiological pH.

- The Km values displayed by the native enzyme were 6.6135 µM/mL, while the modified enzyme had a slightly higher value of 6.4639 µM/mL. This indicates that the substrate affinity for the enzyme has decreased only slightly following conjugation.

- This chemical modification suggests that conjugated uricase is sustained effectively at physiological conditions, which could make it as a potent drug to treat hyperuricemia.

- Overall, these computational and experimental studies can be considered for protein engineering work, as well as for hyperuricemia and gout.

7.3 FUTURE SCOPE

- To predict a molecular solution to reduce the toxicity of the enzyme, reduce certain side effects and improve stability of the enzyme during the bioinformatics work through manipulation of its amino acids to achieve a better degree of treatment.

- *In vitro* analysis of immunogenicity of the conjugates to evaluate its therapeutic efficiency.

- Further structural elucidation of the bioconjugate and evaluation of its polydispersity and purity of the drug conjugate is essential for its complete characterization.

- Also, studies on the pharmacokinetics and pharmacodynamics of the drug conjugate are essential for a complete understanding of its efficiency and can add to the pool of information for formulating it for medical applications.

REFERENCES

Adámek, V., Králová, B., Šüchová, M., Valentová, O., and Demnerová, K. (1989). "Purification of microbial uricase." *J. Chromatogr. B Biomed. Appl*, 497, 268–275.

Adámek, V., Suchová, M., Demnerová, K., Králová, B., Fořt, I., and Morava, P. (1990). "Fermentation of Candida utilis for uricase production." *J. Ind. Microbiol.*, 6(2), 85–89.

Ahmad, T. A., Eweida, A. E., and El-Sayed, L. H. (2016). "T-cell epitope mapping for the design of powerful vaccines." *Vaccine Rep.*, 6, 13–22.

Ahmed, R. K. S., and Maeurer, M. J. (2009). "T-Cell Epitope Mapping." *Epitope Mapping Protocols*, M. Schutkowski and U. Reineke, eds., Totowa, NJ: Humana Press, 427–438.

Alamillo, J. M., Cárdenas, J., and Pineda, M. (1991). "Purification and molecular properties of urate oxidase from *Chlamydomonas reinhardtii*." *Biochim Biophys Acta.*, 1076(2), 203–208.

Ali, S., and Lally, E. V. (2009). "Treatment failure gout." *Med Health R I*, 92(11), 369–371.

Alméciga-Díaz, C. J., Gutierrez, Á. M., Bahamon, I., Rodríguez, A., Rodríguez, M. A., and Sánchez, O. F. (2011). "Computational analysis of the fructosyltransferase enzymes in plants, fungi and bacteria." *Gene*, 484(1–2), 26–34.

Altman, K. I., Smull, K., and Barron, E. S. G. (1949). "A new method for the preparation of uricase and the effect of uricase on the blood uric acid levels of the chicken." *Arch Biochem*, 21(1), 158–165.

Alvarez-Lario, B., and Macarron-Vicente, J. (2010). "Uric acid and evolution." *Rheumatology*, 49(11), 2010–5.

Anderson, A., and Vijayakumar, S. (2011). "Purification and optimization of uricase enzyme produced by *Pseudomonas aeruginosa*." *J.Exp.Sci.*, 2(11), 5–8.

Appaiah, P., and Vasu, P. (2016). "*In silico* Designing of Protein Rich in Large Neutral Amino Acids Using Bovine αs1 Casein for Treatment of Phenylketonuria." *J Proteomics Bioinform*, 9(11), 287-297.

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., and Yeh, L.-S. L. (2004). "UniProt: the Universal Protein knowledgebase." *Nucleic Acids Res.*, 32(Database issue), D115-119.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., Castro, E. de, Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., and Stockinger, H. (2012). "ExPASy: SIB bioinformatics resource portal." *Nucleic Acids Res.*, 40(Web Server issue), W597-603.

Ashok Kumar, T. (2013). "CFSSP: Chou and Fasman Secondary Structure Prediction server." *Wide spectr Res J.,* 1(9): 15-19.

Atalla, M., Farag, M., Eman, R. H., Abd-El-Lataif, M. S., and Nehad, E. A. (2009). "Optimum conditions for uricase enzyme production by Gliomastix gueg." *Malays. J. Microbiol.*, 5(1), 45–50.

Averill, F. W., and Painter, G. S. (1992). "Steepest-descent determination of occupation numbers and energy minimization in the local-density approximation." *Phys. Rev. B*, 46(4), 2498–2502.

Azab, E., Ali, M., and Fareed, M. (2014). "Studies on uricase induction in certain bacteria." *Egypt. J. Biol.*, 44(8), 811–821.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. (2009). "MEME SUITE: tools for motif discovery and searching." *Nucleic Acids Res.*, 37(Web Server issue), W202-208.

Bailey, T. L., and Elkan, C. (1994). "Fitting a mixture model by expectation maximization to discover motifs in biopolymers." *Proc Int Conf Intell Syst Mol Biol*, 2, 28–36.

Bailey, T. L., and Gribskov, M. (1998). "Combining evidence using p-values: application to sequence homology searches." *Bioinform.*, 14(1), 48–54.

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). "The MEME Suite." *Nucleic Acids Res.*, 43(W1), W39-49.

Bander, N. H., Milowsky, M. I., Nanus, D. M., Kostakoglu, L., Vallabhajosula, S., and Goldsmith, S. J. (2005). "Phase I trial of 177lutetium-labeled J591, a monoclonal antibody to prostate-specific membrane antigen, in patients with androgen-independent prostate cancer." *J. Clin. Oncol.*, 23(21), 4591–4601.

Baraf, H. S. B., Matsumoto, A. K., Maroli, A. N., and Waltrip, R. W. (2008). "Resolution of gouty tophi after twelve weeks of pegloticase treatment." *Arthritis Rheum.*, 58(11), 3632–3634.

Barros, F. F. C., Simiqueli, A. P. R., Andrade, C. J. de, and Pastore, G. M. (2013). "Production of Enzymes from Agroindustrial Wastes by Biosurfactant-Producing Strains of *Bacillus subtilis*." *Biotechnol Res Int.*, 2013, 103960.

Batool, T., Makky, E. A., Jalal, M., and Yusoff, M. M. (2016). "A Comprehensive Review on l-Asparaginase and Its Applications." *Appl Biochem Biotechnol*, 178(5), 900–923.

Bayol, A., Capdevielle, J., Malazzi, P., Buzy, A., Claude Bonnet, M., Colloc'h, N., Mornon, J.-P., Loyaux, D., and Ferrara, P. (2002). "Modification of a reactive

cysteine explains differences between rasburicase and Uricozyme, a natural *Aspergillus flavus* uricase." *Biotechnol. Appl. Biochem.*, 36(Pt 1), 21–31.

Beedkar, S. D., Khobragade, C. N., Bodade, R. G., and Vinchurkar, A. S. (2012). "Comparative structural modeling and docking studies of uricase: Possible implication in enzyme supplementation therapy for hyperuricemic disorders." *Comp.Biol.Med*, 42(6), 657–666.

Beesh, M., Majewska, P., and Vandamme, T. (2010). "Synthesis and characterization of dextran esters as coating or matrix systems for oral delivery of drugs targeted to the colon." *Int. J. Drug Deliv*, 2(1), 22–31.

Belén, L. H., Lissabet, J. B., Oliveira Rangel-Yagui, C. de, Effer, B., Monteiro, G., Pessoa, A., and Farías Avendaño, J. G. (2019). "A structural *in silico* analysis of the immunogenicity of l-asparaginase from *Escherichia coli* and *Erwinia carotovora*." *Biologicals*, 59, 47–55.

Benkert, P., Biasini, M., and Schwede, T. (2011). "Toward the estimation of the absolute quality of individual protein structure models." *Bioinform*, 27(3), 343–350.

Benkert, P., Künzli, M., and Schwede, T. (2009). "QMEAN server for protein model quality estimation." *Nucleic Acids Res.*, 37(Web Server issue), W510-514.

Bergmann, H., Preddie, E., and Verma, D. P. (1983). "Nodulin-35: a subunit of specific uricase (uricase II) induced and localized in the uninfected cells of soybean nodules." *EMBO J*, 2(12), 2333–2339.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). "The Protein Data Bank." *Nucleic Acids Res.*, 28(1), 235–242.

Bjellqvist, B., Basse, B., Olsen, E., and Celis, J. E. (1994). "Reference points for comparisons of two-dimensional maps of proteins from different human cell types

defined in a pH scale where isoelectric points correlate with polypeptide compositions." *Electrophoresis*, 15(3–4), 529–539.

Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J. C., Frutiger, S., and Hochstrasser, D. (1993). "The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences." *Electrophoresis*, 14(10), 1023–1031.

Bomalaski, J. S., Holtsberg, F. W., Ensor, C. M., and Clark, M. A. (2002). "Uricase formulated with polyethylene glycol (uricase-PEG 20): biochemical rationale and preclinical studies." *J. Rheumatol.*, 29(9), 1942–1949.

Bongaerts, G. P., and Vogels, G. D. (1979). "Mechanism of uricase action." *Biochim Biophys Acta*, 567(2), 295–308.

Bose, R., Arora, S., Dwivedi, V. D., and Pandey, A. (2013). "Amino acid sequence based in silico analysis of β- galactosidases." *Int. J.Bioinforma.Biosci.*, 3(2), 37–44.

Bowers, K. J., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossvary, I., Moraes, M. A., Sacerdoti, F. D., Salmon, J. K., Shan, Y., and Shaw, D. E. (2006). "Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters." *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, SC '06, New York, NY, USA: ACM.

Bowie, J. U., Lüthy, R., and Eisenberg, D. (1991). "A method to identify protein sequences that fold into a known three-dimensional structure." *Science*, 253(5016), 164–170.

Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." *Anal.Biochem.*, 72(1–2), 248–254.

Bull, H. B., and Breese, K. (1974). "Surface tension of amino acid solutions: a hydrophobicity scale of the amino acid residues." *Arch. Biochem. Biophys.*, 161(2), 665–670.

Burns, C. M., and Wortmann, R. L. (2011). "Gout therapeutics: new drugs for an old disease." *The Lancet*, 377(9760), 165–177.

Cantor, J. R., Yoo, T. H., Dixit, A., Iverson, B. L., Forsthuber, T. G., and Georgiou, G. (2011). "Therapeutic enzyme deimmunization by combinatorial T-cell epitope removal using neutral drift." *Proc. Natl. Acad. Sci. U.S.A.*, 108(4), 1272–1277.

Capote-Maínez, N., and Sánchez, F. (1997). "Characterization of the common bean uricase II and its expression in organs other than nodules." *Plant Physiol.*, 115(4), 1307–1317.

Capriotti, E., Fariselli, P., and Casadio, R. (2005). "I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure." *Nucleic Acids Res.*, 33(Web Server issue), W306-310.

Chatzou, M., Magis, C., Chang, J.-M., Kemena, C., Bussotti, G., Erb, I., and Notredame, C. (2016). "Multiple sequence alignment modeling: methods and applications." *Brief. Bioinformatics*, 17(6), 1009–1023.

Chen, Z., Wang, Z., He, X., Guo, X., Li, W., and Zhang, B. (2008). "Uricase production by a recombinant Hansenula polymorpha strain harboring *Candida utilis* uricase gene." *Appl Microbiol Biotechnol*, 79(4), 545–554.

Coiffier, B., Mounier, N., Bologna, S., Fermé, C., Tilly, H., Sonet, A., Christian, B., Casasnovas, O., Jourdan, E., Belhadj, K., and Herbrecht, R. (2003). "Efficacy and tolerability of pegloticase for the treatment of chronic gout in patients refractory to conventional treatment: two randomized controlled trials." *J.Clin.Oncol.*, 21(23), 4402–4406.

158

Collen, D., Stockx, L., Lacroix, H., Suy, R., and Vanderschueren, S. (1997). "Recombinant staphylokinase variants with altered immunoreactivity. IV: Identification of variants with reduced antibody induction but intact potency." *Circulation*, 95(2), 463–472.

Colloc'h, N., Hajji, M. E., Bachet, B., L'Hermite, G., Schiltz, M., Prangé, T., Castro, B., and Mornon, J.-P. (1997). "Crystal Structure of the protein drug urate oxidase-inhibitor complex at 2.05 Å resolution." *Nat Struct Mol Biol*, 4(11), 947–952.

Colloc'h, N., and Prangé, T. (2014). "Functional relevance of the internal hydrophobic cavity of urate oxidase." *FEBS Lett.*, 588(9), 1715–1719.

Colovos, C., and Yeates, T. O. (1993). "Verification of protein structures: patterns of nonbonded atomic interactions." *Protein Sci.*, 2(9), 1511–1519.

Combet, C., Blanchet, C., Geourjon, C., and Deléage, G. (2000). "NPS@: network protein sequence analysis." *Trends Biochem. Sci.*, 25(3), 147–150.

Cristobal, S., Zemla, A., Fischer, D., Rychlewski, L., and Elofsson, A. (2001). "A study of quality measures for protein threading models." *BMC Bioinformatics*, 2, 5.

Dabbagh, F., Ghoshoon, M. B., Hemmati, S., Zamani, M., Mohkam, M., and Ghasemi, Y. (2015). "Engineering Human Urate Oxidase: Towards Reactivating It as an Important Therapeutic Enzyme." *Curr Pharm Biotechnol*, 17(2), 141–146.

Dabbagh, F., Moradpour, Z., Ghasemian, A., and Ghasemi, Y. (2012). "Phylogeny of urate oxidase producing bacteria on the basis of gene sequences of 16S rRNA and uricase protein." *Iran.J.Pharm. Sci.*, 8(2), 99–102.

Dalbeth, N., and Doyle, A. J. (2012). "Imaging of gout: an overview." *Best Pract Res Clin Rheumatol*, 26(6), 823–838.

Dalbeth, N., and Merriman, T. (2009). "Crystal ball gazing: new therapeutic targets for hyperuricaemia and gout." *Rheumatology (Oxford)*, 48(3), 222–226.

Davis, S., Park, Y. K., Abuchowski, A., and Davis, F. F. (1981). "Hypouricaemic effect of polyethyleneglycol modified urate oxidase." *Lancet*, 2(8241), 281–283.

De Duve, C. (1966). "The significance of lysosomes in pathology and medicine." *Proc Inst Med Chic*, 26(4), 73–76.

Dean, S. N., Turner, K. B., Medintz, I. L., and Walper, S. A. (2017). "Targeting and delivery of therapeutic enzymes." *Ther Deliv*, 8(7), 577–595.

Dhanda, S. K., Grifoni, A., Pham, J., Vaughan, K., Sidney, J., Peters, B., and Sette, A. (2018). "Development of a strategy and computational application to select candidate protein analogues with reduced HLA binding and immunogenicity." *Immunology*, 153(1), 118–132.

Doherty, M. (2009). "New insights into the epidemiology of gout." *Rheumatology (Oxford)*, 48 Suppl 2, ii2–ii8.

Dubey, A. K., Yadav, S., Kumar, M., Singh, V. K., Sarangi, B. K., and Yadav, D. (2010). "*In silico* characterization of pectate lyase protein sequences from different source organisms." *Enzyme Res.*, 2010, 37–44.

Duncan, R., Kopecková, P., Strohalm, J., Hume, I. C., Lloyd, J. B., and Kopecek, J. (1988). "Anticancer agents coupled to N-(2-hydroxypropyl)methacrylamide copolymers. II. Evaluation of daunomycin conjugates in vivo against L1210 leukaemia." *Br J Cancer*, 57(2), 147–156.

Dutta, B., Banerjee, A., Chakraborty, P., and Bandopadhyay, R. (2018). "*In silico* studies on bacterial xylanase enzyme: Structural and functional insight." *J Genet Eng Biotechnol*, 16(2), 749–756.

Dwivedi, V. D., Arora, S., Kumar, A., and Mishra, S. K. (2013). "Computational analysis of xanthine dehydrogenase enzyme from different source organisms." *Netw. Model. Anal. Health Inform.Bioinform.*, 2(4), 185–189.

Dwivedi, V. D., and Mishra, S. K. (2014). "*In silico* analysis of L-asparaginase from different source organisms." *Interdiscip Sci.*, 6(2), 93–99.

Edwards, N. L. (2008). "Treatment-failure gout: A moving target." *Arthritis Rheum*, 58(9), 2587–2590.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., and Finn, R. D. (2019). "The Pfam protein families database in 2019." *Nucleic Acids Res.*, 47(D1), D427–D432.

El-Naggar, N. E.-A., Haroun, S. A., El-Weshy, E. M., Metwally, E. A., and Sherief, A. A. (2019). "Mathematical modeling for bioprocess optimization of a protein drug, uricase, production by *Aspergillus welwitschiae* strain 1–4." *Sci Rep*, 9(1), 12971.

Emini, E. A., Hughes, J. V., Perlow, D. S., and Boger, J. (1985). "Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide." *J. Virol.*, 55(3), 836–839.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). "A smooth particle mesh Ewald method." *J.Chem.Phys.*, 103(19), 8577–8593.

Evander Emeltan Tjoa, S., Maria Vianney, Y., and Emantoko Dwi Putra, S. (2018). "*In silico* mutagenesis: decreasing the immunogenicity of botulinum toxin type A." *J.Biomol.Struct. Dyn.*, 37(18), 4767-4778

Fattahian, Y., Riahi-Madvar, A., Mirzaee, R., Asadikaram, G., and Rahbar, M. R. (2017). "*In silico* locating the immune-reactive segments of Lepidium draba peroxidase and designing a less immune-reactive enzyme derivative." *Comput Biol Chem.*, 70, 21–30.

Felsenstein, J. (1985). "Confidence Limits on Phylogenies: An Approach Using the Bootstrap." *Evolution*, 39(4), 783.

Feng, J., Li, X., Yang, X., Zhang, C., Yuan, Y., Pu, J., Zhao, Y., Xie, Y., Yuan, H., Bu, Y., and Liao, F. (2010). "A new practical system for evaluating the pharmacological properties of uricase as a potential drug for hyperuricemia." *Arch.Pharm.Res.*, 33(11), 1761–1769.

Feng, J., Wang, L., Liu, H., Yang, X., Liu, L., Xie, Y., Liu, M., Zhao, Y., Li, X., Wang, D., Zhan, C.-G., and Liao, F. (2015). "Crystal structure of *Bacillus fastidious* uricase reveals an unexpected folding of the C-terminus residues crucial for thermostability under physiological conditions." *Appl. Microbiol. Biotechnol.*, 99(19), 7973–7986.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., and Punta, M. (2014). "Pfam: the protein families database." *Nucleic Acids Res.*, 42(Database issue), D222-230.

Friedman, T. B., and Barker, A. P. (1982). "Purification and partial characterization of urate oxidase from *Drosophila melanogaster*." *Insect Biochem.*, 12(5), 563–570.

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004). "Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy." *J. Med. Chem.*, 47(7), 1739–1749.

Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., and Mainz, D. T. (2006). "Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes." *J. Med. Chem.*, 49(21), 6177–6196.

Fujiwara, S., Ohashi, H., and Noguchi, T. (1987). "Comparison of intraperoxisomal localization form and properties of amphibian (Rana catesbeiana) uricase with those of other animal uricases." *Comp. Biochem.Physiol.B.,* 86(1), 23–26.

Gabison, L., Chiadmi, M., El Hajji, M., Castro, B., Colloc'h, N., and Prangé, T. (2010). "Near-atomic resolution structures of urate oxidase complexed with its substrate and analogues: the protonation state of the ligand." *Acta Crystallogr D Biol Crystallogr*, 66(6), 714–724.

Ganson, N. J., Kelly, S. J., Scarlett, E., Sundy, J. S., and Hershfield, M. S. (2006). "Control of hyperuricemia in subjects with refractory gout, and induction of antibody against poly(ethylene glycol) (PEG), in a phase I trial of subcutaneous PEGylated urate oxidase." *Arthritis Res Ther*, 8(1), R12.

Garay, R. P., El-Gewely, M. R., Labaune, J.-P., and Richette, P. (2012). "Therapeutic perspectives on uricases for gout." *Joint Bone Spine*, 79(3), 237–242.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005). "Protein Identification and Analysis Tools on the ExPASy Server." *The Proteomics Protocols Handbook*, J. M. Walker, ed., Totowa, NJ: Humana Press, 571–607.

Genheden, S., and Ryde, U. (2015). "The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities." *Expert Opin Drug Discov*, 10(5), 449–461.

Geourjon, C., and Deléage, G. (1995). "SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments." *Comput. Appl. Biosci.*, 11(6), 681–684.

Geweely, N., and Nawar, L. (2011). "Production, optimization, purification and properties of uricase isolated from some fungal flora in Saudi Arabian soil." *Aust. J. Basic & Appl. Sci*, 5(10), 220–230.

Ghosh, T., and Sarkar, P. (2014). "Isolation of a novel uric-acid-degrading microbe *Comamonas sp*. BT UA and rapid biosensing of uric acid from extracted uricase enzyme." *J Biosci*, 39(5), 805–819.

Gill, S. C., and Hippel, P. H. von. (1989). "Calculation of protein extinction coefficients from amino acid sequence data." *Anal Biochem.*, 182(2), 319–326.

Gliozzi, M., Malara, N., Muscoli, S., and Mollace, V. (2016). "The treatment of hyperuricemia." *Int.J.Cardio.*, 213, 23–27.

Gruia, F., Parupudi, A., Baca, M., Ward, C., Nyborg, A., Remmele, R. L., and Bee, J. S. (2017). "Impact of Mutations on the Higher Order Structure and Activity of a Recombinant Uricase." *J.Pharm. Sci.*, 106(4), 1018–1024.

Guruprasad, K., Reddy, B. V. B., and Pandit, M. W. (1990). "Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence." *Protein Eng. Des. Sel.*, 4(2), 155–161.

Guttmann, A., Krasnokutsky, S., Pillinger, M. H., and Berhanu, A. (2017). "Pegloticase in gout treatment - safety issues, latest evidence and clinical considerations." *Ther Adv Drug Saf.*, 8(12), 379–388.

Habeeb,A.F.S.A.(1966)."Determination of free amino groups in proteins by trinitrobenzenesulfonic acid." *Anal Biochem.*, 14(3), 328–336.

Haste Andersen, P., Nielsen, M., and Lund, O.(2006). "Prediction of residues in discontinuous B-cell epitopes using protein 3D structures." *Protein Sci.*, 15(11), 2558–2567.

Hirokawa, T., Boon-Chieng, S., and Mitaku, S. (1998). "SOSUI: classification and secondary structure prediction system for membrane proteins." *Bioinformatics*, 14(4), 378–379.

Holgate, R. G. E., and Baker, M. P. (2009). "Circumventing immunogenicity in the development of therapeutic antibodies." *IDrugs*, 12(4), 233–237.

Howard, S. C., Jones, D. P., and Pui, C.-H. (2011). "The tumor lysis syndrome." *N. Engl. J. Med.*, 364(19), 1844–1854.

Hu, T., and Su, Z. (2002). "Preparation and characterization of a bovine serum albumin-catalase conjugate." *Biotechnol. Lett.*, 24(12), 1027–1030.

Hu, X., Wang, Y., Liu, C., Jin, Z., and Tian, Y. (2018). "Dextrin-uricase conjugate: Preparation, characterization, and enzymatic properties." *Int.J.Biol.Macromol.*, 111, 28–32.

Huang, S.-H., and Wu, T.-K. (2004). "Modified colorimetric assay for uricase activity and a screen for mutant *Bacillus subtilis* uricase genes following StEP mutagenesis." *Eur. J. Biochem.*, 271(3), 517–523.

Ikai, A. (1980). "Thermostability and aliphatic index of globular proteins." *J. Biochem.*, 88(6), 1895–1898.

Imhoff, R. D., Power, N. P., Borrok, M. J., and Tipton, P. A. (2003). "General Base Catalysis in the Urate Oxidase Reaction: Evidence for a Novel Thr−Lys Catalytic Diad [†]." *Biochemistry*, 42(14), 4094–4100.

Irajie, C., Mohkam, M., Nezafat, N., Hosseinzadeh, S., Aminlari, M., and Ghasemi, Y. (2016a). "*In Silico* analysis of glutaminase from different species of *Escherichia* and *Bacillus*." *Iranian J Med Sci.*, 41(5), 406–414.

Ishikawa, J., Yamashita, A., Mikami, Y., Hoshino, Y., Kurita, H., Hotta, K., Shiba, T., and Hattori, M. (2004). "The complete genomic sequence of *Nocardia farcinica* IFM 10152." *Proc. Natl. Acad. Sci. U.S.A.*, 101(41), 14925–14930.

Ito, M., Kato, S., Nakamura, M., Mitiko, G., and Takagi, Y. (1992). "Identification of an amino acid residue involved in the substrate-binding site of rat liver uricase by site-directed mutagenesis." *Biochem.Biophys.Res.Commun.,* 187(1), 101–107.

Ito, M., Nakamura, M., Ogawa, H., Kato, S., and Takagi, Y. (1991). "Structural analysis of the gene encoding rat uricase." *Genomics*, 11(4), 905–913.

JevsÌŒevar, S., Kunstelj, M., and Porekar, V. G. (2010). "PEGylation of therapeutic proteins." *Biotechnol. J.*, 5(1), 113–128.

Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." *J. Mol. Biol.*, 292(2), 195–202.

Jones, T. D., Crompton, L. J., Carr, F. J., and Baker, M. P. (2009). "Deimmunization of monoclonal antibodies." *Methods Mol. Biol.*, 525, 405–423.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). "Comparison of simple potential functions for simulating liquid water." *J. Chem. Phys.*, 79(2), 926–935.

Juan, E. C. M., Hoque, M. M., Shimizu, S., Hossain, M. T., Yamamoto, T., Imamura, S., Suzuki, K., Tsunoda, M., Amano, H., Sekiguchi, T., and Takénaka, A. (2008). "Structures of *Arthrobacter globiformis* urate oxidase-ligand complexes." *Acta Crystallogr.D.*, 64(8), 815–822.

Kahn, K., Serfozo, P., and Tipton, P. A. (1997). "Identification of the True Product of the Urate Oxidase Reaction." *J. Am. Chem. Soc.*, 119(23), 5435–5442.

Kai, L., Ma, X.-H., Zhou, X.-L., Jia, X.-M., Li, X., and Guo, K.-P. (2008). "Purification and characterization of a thermostable uricase from *Microbacterium sp*. strain ZZJ4-1." *World J Microbiol Biotechnol*, 24(3), 401–406.

Kang, T. S., and Stevens, R. C. (2009). "Structural aspects of therapeutic enzymes to treat metabolic disorders." *Hum. Mutat.*, 30(12), 1591–1610.

Karplus, P. A., and Schulz, G. E. (1985). "Prediction of chain flexibility in proteins."*Sci.Nat.*, 72(4), 212–213.

Keebaugh, A. C., and Thomas, J. W. (2010). "The evolutionary fate of the genes encoding the purine catabolic enzymes in hominoids, birds, and reptiles." *Mol. Biol. Evol.*, 27(6), 1359–1369.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). "The Phyre2 web portal for protein modeling, prediction and analysis." *Nat Protoc*, 10(6), 845–858.

Khade, S., and Srivastava, S. K. (2015). "Uricase and its clinical applications." *Int J Biol Med Res*, 6(3), 5211–5215.

Khucharoenphaisan, K., and Sinma, K. (2011). "Production and partial characterization of uric acid degrading enzyme from new source *Saccharopolyspora sp*. PNR11." *Pak J Biol Sci*, 14(3), 226–231.

Kim, Y., Ponomarenko, J., Zhu, Z., Tamang, D., Wang, P., Greenbaum, J., Lundegaard, C., Sette, A., Lund, O., Bourne, P. E., Nielsen, M., and Peters, B. (2012). "Immune epitope database analysis resource." *Nucleic Acids Res.*, 40(Web Server issue), W525-530.

Kishore, V., Gowda, S., Krishna, S., Sharma, K., M, R., and K P, N. (2014). "Bovine Serum Albumin a Potential Thermostabilizer: a Study on α-Amylase." *J. Appl. Environ. Microbiol*, 2(2), 37–41.

Ko, T.-M., Tsai, C.-Y., Chen, S.-Y., Chen, K.-S., Yu, K.-H., Chu, C.-S., Huang, C.-M., Wang, C.-R., Weng, C.-T., Yu, C.-L., Hsieh, S.-C., Tsai, J.-C., Lai, W.-T., Tsai, W.-C., Yin, G.-D., Ou, T.-T., Cheng, K.-H., Yen, J.-H., Liou, T.-L., Lin, T.-H., Chen, D.-Y., Hsiao, P.-J., Weng, M.-Y., Chen, Y.-M., Chen, C.-H., Liu, M.-F., Yen, H.-W., Lee, J.-J., Kuo, M.-C., Wu, C.-C., Hung, S.-Y., Luo, S.-F., Yang, Y.-H., Chuang, H.-P., Chou, Y.-C., Liao, H.-T., Wang, C.-W., Huang, C.-L., Chang, C.-S., Lee, M.-T. M., Chen, P., Wong, C.-S., Chen, C.-H., Wu, J.-Y., Chen, Y.-T., and Shen, C.-Y. (2015). "Use of HLA-B*58:01 genotyping to prevent allopurinol induced severe cutaneous adverse reactions in Taiwan: national prospective cohort study." *BMJ*, 23, 351: h4848.

Kolaskar, A. S., and Tongaonkar, P. C. (1990). "A semi-empirical method for prediction of antigenic determinants on protein antigens." *FEBS Lett.*, 276(1–2), 172–174.

Koteswara Reddy, G., Nagamalleswara Rao, K., and Yarrakula, K. (2017). "Insights into structure and function of 30S Ribosomal Protein S2 (30S2) in *Chlamydophila pneumoniae*: A potent target of pneumonia." *Comput Biol Chem*, 66, 11–20.

Kratzer, J. T., Lanaspa, M. A., Murphy, M. N., Cicerchi, C., Graves, C. L., Tipton, P. A., Ortlund, E. A., Johnson, R. J., and Gaucher, E. A. (2014). "Evolutionary history and metabolic insights of ancient mammalian uricases." *Proc Natl Acad Sci USA*, 111(10), 3763.

Kringelum, J. V., Lundegaard, C., Lund, O., and Nielsen, M. (2012). "Reliable B Cell Epitope Predictions: Impacts of Method Development and Improved Benchmarking." *PLOS Comput. Biol.*, 8(12), e1002829.

Kumar, S., Nei, M., Dudley, J., and Tamura, K. (2008). "MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences." *Brief. Bioinform.*, 9(4), 299–306.

Kumar, S., Pakshirajan, K., and Venkata Dasu, V. (2009). "Development of medium for enhanced production of glutaminase-free L-asparaginase from *Pectobacterium carotovorum* MTCC 1428." *Appl Microbiol Biotechnol*, 84(3), 477–486.

Kumar, S., Stecher, G., and Tamura, K. (2016). "MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets." *Mol. Biol. Evol.*, 33(7), 1870–1874.

Kumar, S., Tamura, K., Jakobsen, I. B., and Nei, M. (2001). "MEGA2: molecular evolutionary genetics analysis software." *Bioinformatics*, 17(12), 1244–1245.

Kumar, V., Singh, G., Verma, A. K., and Agrawal, S. (2012). "*In silico* characterization of histidine Acid phytase sequences." *Enzyme Res*, 2012, 845465.

Kyte, J., and Doolittle, R. F. (1982). "A simple method for displaying the hydropathic character of a protein." *J. Mol. Biol.*, 157(1), 105–132.

Laemmli, U. K. (1970). "Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4." *Nature*, 227(5259), 680–685.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007). "Clustal W and Clustal X version 2.0." *Bioinformatics*, 23(21), 2947–2948.

Larsen, J. E. P., Lund, O., and Nielsen, M. (2006). "Improved method for predicting linear B-cell epitopes." *Immunome Res*, 2, 2.

Le Goff, B., Berthelot, J.-M., André, V., Guillot, P., and Maugars, Y. (2008). "Ultrasonography for diagnosing atypical gout. Two case reports." *Joint Bone Spine*, 75(5), 610–612.

Lee, I. R., Yang, L., Sebetso, G., Allen, R., Doan, T. H. N., Blundell, R., Lui, E. Y. L., Morrow, C. A., and Fraser, J. A. (2013). "Characterization of the Complete Uric Acid Degradation Pathway in the Fungal Pathogen *Cryptococcus neoformans*." *PLoS ONE*, (K. Nielsen, ed.), 8(5), e64292.

Lee, Y., Park, B. C., Lee, D. H., Bae, K.-H., Cho, S., Lee, C. H., Lee, J. S., Myung, P. K., and Park, S. G. (2006). "Mouse transthyretin-related protein is a hydrolase which degrades 5-hydroxyisourate, the end product of the uricase reaction." *Mol. Cells*, 22(2), 141–145.

Leplatois, P., Le Douarin, B., and Loison, G. (1992). "High-level production of a peroxisomal enzyme: Aspergillus flavus uricase accumulates intracellularly and is active in *Saccharomyces cerevisiae*." *Gene*, 122(1), 139–145.

Li, J., Chen, Z., Hou, L., Fan, H., Weng, S., Xu, C., Ren, J., Li, B., and Chen, W. (2006). "High-level expression, purification, and characterization of non-tagged *Aspergillus flavus* urate oxidase in *Escherichia coli*." *Protein Expr. Purif.*, 49(1), 55–59.

Li, W., Xu, S., Zhang, B., Zhu, Y., Hua, Y., Kong, X., Sun, L., and Hong, J. (2017). "Directed evolution to improve the catalytic efficiency of urate oxidase from *Bacillus subtilis*." *PLoS ONE*, 12(5), e0177877.

Li, X., Wang, B., Li, Y., and Li, C. (2016). "Gout and Hyperuricemia Evaluation of the safety and efficacy of urate oxidase." *Gout. Hyperuricemia*, 3(1), 9–13.

Lineweaver, H., and Burk, D. (1934). "The Determination of Enzyme Dissociation Constants." *J. Am. Chem. Soc.*, 56(3), 658–666.

Liu, J., Li, G., Liu, H., and Zhou, X. (1994). "Purification and properties of uricase from *Candida sp*. and its application in uric acid analysis in serum." *Appl. Biochem. Biotechnol.*, 47(1), 57–63.

London, M., and Hudson, P. B. (1957). "Uricolytic activity of purified uricase in two human beings." *Science*, 125(3254), 937–938.

Lotfy, W.(2008)."Production of a thermostable uricase by a novel *Bacillus thermocatenulatus* strain." *Bioresour. Technol.*, 99(4), 699–702.

Lovell, S. C., Davis, I. W., Arendall, W. B., Bakker, P. I. W. de, Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003). "Structure validation by Calpha geometry: phi,psi and Cbeta deviation." *Proteins*, 50(3), 437–450.

Lucas, K., Boland, M. J., and Schubert, K. R. (1983). "Uricase from soybean root nodules: purification, properties, and comparison with the enzyme from cowpea." *Arch Biochem Biophys*, 226(1), 190–197.

Lüthy, R., Bowie, J. U., and Eisenberg, D. (1992). "Assessment of protein models with three-dimensional profiles." *Nature*, 356(6364), 83–85.

Macfarlane, D. J., Smart, R. C., Tsui, W. W., Gerometta, M., Eisenberg, P. R., and Scott, A. M. (2006). "Safety, pharmacokinetic and dosimetry evaluation of the proposed thrombus imaging agent 99mTc-DI-DD-3B6/22-80B3 Fab′." *Eur.J.Nucl.Med.Mol.Imaging*, 33(6), 648–656.

Machida, Y., and Nakanishi, T. (1980). "Purification and Properties of Uricase from *Enterobacter cloacae*." *Agric.Biol.Chem.*, 44(12), 2811–2815.

Mahler, J. L. (1970). "A new bacterial uricase for uric acid determination." *Anal. Biochem.*, 38(1), 65–84.

Maiuolo, J., Oppedisano, F., Gratteri, S., Muscoli, C., and Mollace, V. (2016). "Regulation of uric acid metabolism and excretion." *Int. J.Cardiol.*, 213, 8–14.

Malherbe, L. (2009). "T-cell epitope mapping." *Ann. Allergy Asthma Immunol.*, 103(1), 76–79.

Malik, A., Schumacher, H. R., Dinnella, J. E., and Clayburne, G. M. (2009). "Clinical diagnostic criteria for gout: comparison with the gold standard of synovial fluid crystal analysis." *J Clin Rheumatol*, 15(1), 22–24.

Malviya, N., Srivastava, M., Diwakar, S. K., and Mishra, S. K. (2011). "Insights to sequence information of polyphenol oxidase enzyme from different source organisms." *Appl. Biochem.Biotechnol.*, 165(2), 397–405.

Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992). "Nosé–Hoover chains: The canonical ensemble via continuous dynamics."*J.Chem.Phys.*, 97(4), 2635–2643.

Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994). "Constant pressure molecular dynamics algorithms."*J. Chem. Phys.*, 101(5), 4177–4189.

Mathew, A., Verma, A., and Gaur, S. (2014). "An *in-silico* insight into the characteristics of β-propeller phytase." *Interdiscip Sci*, 6(2), 133–139.

McDonnell, A. M., Lenz, K. L., Frei-Lahr, D. A., Hayslip, J., and Hall, P. D. (2006). "Single-dose rasburicase 6 mg in the management of tumor lysis syndrome in adults." *Pharmacotherapy*, 26(6), 806–812.

McGuffin, L. J., Bryson, K., and Jones, D. T. (2000). "The PSIPRED protein structure prediction server." *Bioinformatics*, 16(4), 404–405.

Modrić, N., Derome, A. E., Ashcroft, S. J. H., and Poje, M. (1992). "Tracing and identification of uricase. Reaction intermediates." *Tetrahedron Lett.*, 33(44), 6691–6694.

Mohan Kumar, N. S., Kishore, V., and Manonmani, H. K. (2014). "Chemical modification of L-Asparaginase from *Cladosporium sp.* for improved activity and thermal stability." *Prep. Biochem.Biotechnol.*, 44(5), 433–450.

Montalbini, P., Redondo, J., Caballero, J. L., Cárdenas, J., and Pineda, M. (1997). "Uricase from leaves: its purification and characterization from three different higher plants." *Planta*, 202(3), 277–283.

Moola, Z. B., Scawen, M. D., Atkinson, T., and Nicholls, D. J. (1994). "Erwinia chrysanthemi L-asparaginase: epitope mapping and production of antigenically modified enzymes." *Biochem. J.*, 302 (Pt 3), 921–927.

Morya, V. K., Yadav, S., Kim, E. K., and Yadav, D. (2012). "In silico characterization of alkaline proteases from different species of *aspergillus*." *Appl.Biochemis.Biotechnol.*, 166(1), 243–257.

Morya, V. K., Yadav, V. K., Yadav, S., and Yadav, D. (2016). "Active Site Characterization of Proteases Sequences from Different Species of *Aspergillus*." *Cell Biochem Biophys*, 74(3), 327–335.

Mount, D. W. (2008). "Maximum Parsimony Method for Phylogenetic Prediction." *Cold Spring Harbor Protocols*, 2008(5), pdb.top32.

172

Moutaftsi, M., Peters, B., Pasquetto, V., Tscharke, D. C., Sidney, J., Bui, H.-H., Grey, H., and Sette, A. (2006). "A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus." *Nat. Biotechnol.*, 24(7), 817–819.

Mumtaz, S., and Bachhawat, B. K. (1992). "Enzyme engineering and its application in lysosomal storage disease." *Pure Appl.Chem.*, 64(8), 1055–1060.

Nanda, P., and Jagadeesh Babu, P. E. (2014). "Isolation, screening and production studies of uricase producing bacteria from poultry sources." *Prep. Biochem.Biotechnol.*, 44(8), 811–821.

Nanda, P., and Jagadeesh Babu, P. E. (2017). "Solid Phase PEGylation of Uricase." *Materials Today: Proceedings*, 4(9), 10494–10497.

Nanda, P., and JagadeeshBabu, P. E. (2016). "Studies on the Site-specific PEGylation Induced Interferences Instigated in Uricase Quantification Using the Bradford Method." *Int J Pept Res Ther*, 22(3), 399–406.

Nanda, P., P.E., J., and Raju, J. R. (2016). "Production and Optimization of Site-Specific monoPEGylated Uricase Conjugates Using mPEG-Maleimide Through RP–HPLC Methodology." *J Pharm Innov*, 11(4), 279–288.

Nanda, P., PonnanEttiyappan, J., Jenifer Fernandes, and Pranita Hazarika. (2012). "Studies on Production, Optimization and Purification of Uricase from Gliocladium viride." *Res Biotechnol.*, 3(4), 35–46.

Navolanic, P. M., Pui, C.-H., Larson, R. A., Bishop, M. R., Pearce, T. E., Cairo, M. S., Goldman, S. C., Jeha, S. C., Shanholtz, C. B., Leonard, J. P., and McCubrey, J. A. (2003). "Elitek-rasburicase: an effective means to prevent and treat hyperuricemia associated with tumor lysis syndrome, a Meeting Report, Dallas, Texas, January 2002." *Leukemia*, 17(3), 499–514.

Nelapati, A. K., and PonnanEttiyappan, J. (2019). "Computational Analysis of Therapeutic Enzyme Uricase from Different Source Organisms." *Curr.Proteom.,* 17(1),59-77.

Nezafat, N., Negahdaripour, M., Gholami, A., and Ghasemi, Y. (2015). "Computational analysis of collagenase from different Vibrio, Clostridium and Bacillus strains to find new enzyme sources." *Trends Pharm.Sci*., 1(4), 213-222

Niño-Gómez, D. C., Rivera-Hoyos, C. M., Morales-Álvarez, E. D., Reyes-Montaño, E. A., Vargas-Alejo, N. E., Ramírez-Casallas, I. N., Erkan Türkmen, K., Sáenz-Suárez, H., Sáenz-Moreno, J. A., Poutou-Piñales, R. A., González-Santos, J., and Arévalo-Galvis, A. (2017). "'In Silico' Characterization of 3-Phytase A and 3-Phytase B from *Aspergillus niger*." *Enzyme Res.,* 2017, 1–23.

Nishimura, H., Ashihara, Y., Matsushima, A., and Inada, Y. (1979). "Modification of yeast uricase with polyethylene glycol: disappearance of binding ability towards anti-uricase serum." *Enzyme*, 24(4), 261–264.

Nuki, G. (2012). "Uricase Therapy of Gout." *Gout & Other Crystal Arthropathies*,174–186.

Nyborg, A. C., Ward, C., Zacco, A., Chacko, B., Grinberg, L., Geoghegan, J. C., Bean, R., Wendeler, M., Bartnik, F., O'Connor, E., Gruia, F., Iyer, V., Feng, H., Roy, V., Berge, M., Miner, J. N., Wilson, D. M., Zhou, D., Nicholson, S., Wilker, C., Wu, C. Y., Wilson, S., Jermutus, L., Wu, H., Owen, D. A., Osbourn, J., Coats, S., and Baca, M. (2016). "A Therapeutic Uricase with Reduced Immunogenicity Risk and Improved Development Properties." *PLOS ONE*, (M. Rito-Palomares, ed.), 11(12), e0167935.

Oda, M., Satta, Y., Takenaka, O., and Takahata, N. (2002). "Loss of urate oxidase activity in hominoids and its evolutionary implications." *Mol. Biol. Evol.*, 19(5), 640–653.

Osman, A. M., Corso, A. D., Ipata, P. L., and Mura, U. (1989). "Liver uricase in *Camelusdromedarius*:purificationandproperties."*Comp.Biochem.Physiol.B,Comp.Bio chem.*, 94(3), 469–474.

Pandey, S., Kumar Negi, Y., and S. Marla, S. (2011). "Comparative Insilico Analysis of Ascorbate Peroxidase Protein Sequences from Different Plant Species."*J.Bioeng.Biomed Sci.*, 01(02), 1000103.

Papadopoulou, K., Roussis, A., Kuin, H., and Katinakis, P. (1995). "Expression pattern of uricase II gene during root nodule development in Phaseolus vulgaris." *Experientia*, 51(1), 90–94.

Park, D.-J., Kang, J.-H., Lee, J.-W., Lee, K.-E., Wen, L., Kim, T.-J., Park, Y.-W., Park, S.-H., and Lee, S.-S. (2015). "Cost-effectiveness analysis of HLA-B5801 genotyping in the treatment of gout patients with chronic renal insufficiency in Korea." *Arthritis Care Res (Hoboken)*, 67(2), 280–287.

Parker, J. M. R., Guo, D., and Hodges, R. S. (1986). "New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and x-ray-derived accessible sites." *Biochemistry*, 25(19), 5425–5432.

Pasut, G., and Veronese, F. M. (2007). "Polymer-drug conjugation, recent achievements and general strategies." *Prog. Polym. Sci.*, 32(8), 933–961.

Pawar, S. V., and Rathod, V. K. (2018). "Optimization of novel and greener approach for the coproduction of uricase and alkaline protease in *Bacillus licheniformis* by Box–Behnken model." *Prep.Biochem.Biotechnol.*, 48(1), 24–33.

Peters, B., Bui, H.-H., Frankild, S., Nielson, M., Lundegaard, C., Kostem, E., Basch, D., Lamberth, K., Harndahl, M., Fleri, W., Wilson, S. S., Sidney, J., Lund, O., Buus, S., and Sette, A. (2006). "A community resource benchmarking predictions of peptide binding to MHC-I molecules." *PLoS Comput. Biol.*, 2(6), e65.

Pfrimer, P., Moraes, L. M. P. de, Galdino, A. S., Salles, L. P., Reis, V. C. B., De Marco, J. L., Prates, M. V., Bloch, C., and Torres, F. A. G. (2010). "Cloning, purification, and partial characterization of Bacillus subtilis urate oxidase expressed in Escherichia coli." *J Biomed Biotechnol*, 2010, 674908.

Pooja, K., Rani, S., Kanwate, B., and Pal, G. K. (2017). "Physico-chemical, Sensory and Toxicity Characteristics of Dipeptidyl Peptidase-IV Inhibitory Peptides from Rice Bran-derived Globulin Using Computational Approaches." *Int J Pept Res Ther.*, 23(4), 519–529.

Potocnakova, L., Bhide, M., and Pulzova, L. B. (2016). "An Introduction to B-Cell Epitope Mapping and In Silico Epitope Prediction." *J Immunol Res*, 2016, 6760830.

Poznansky, M. J., and Bhardwaj, D. (1980). "α-1,4-Glucosidase–albumin polymers : in vitro properties and advantages for enzyme replacement therapy." *Can. J. Physiol. Pharmacol.*, 58(3), 322–325.

Poznansky, M. J., Halford, J., and Taylor, D. (1988). "Growth hormone-albumin conjugates. Reduced renal toxicity and altered plasma clearance." *FEBS Lett*, 239(1), 18–22.

Poznansky, M. J., Shandling, M., Salkie, M. A., Elliott, J. F., and Lau, E. (1982). "Advantages in the use of L-asparaginase-albumin polymer as an antitumor agent." *Cancer Res.*, 42(3), 1020–1025.

Pramanik, K., Ghosh, P. K., Ray, S., Sarkar, A., Mitra, S., and Maiti, T. K. (2017). "An in silico structural, functional and phylogenetic analysis with three dimensional protein modeling of alkaline phosphatase enzyme of *Pseudomonas aeruginosa*." *J Genet Eng Biotechnol*, 15(2), 527–537.

Pramanik, K., Kundu, S., Banerjee, S., Ghosh, P. K., and Maiti, T. K. (2018). "Computational-based structural, functional and phylogenetic analysis of *Enterobacter* phytases." *3 Biotech*, 8(6), 262.

Pramanik, K., Pal, P., Soren, T., Mitra, S., Ghosh, P. K., Sarkar, A., and Maiti, T. K. (2018). "In silico structural, functional and phylogenetic analysis of *Klebsiella* phytases." *J.Plant Biochem.Biotechnol.*, 27(3), 362–372.

Pramanik, K., Saren, S., Mitra, S., Ghosh, P. K., and Maiti, T. K. (2018). "Computational elucidation of phylogenetic, structural and functional characteristics of *Pseudomonas* Lipases." *Comput Biol Chem*, 74, 190–200.

Pramanik, K., Soren, T., Mitra, S., and Maiti, T. K. (2017). "In silico structural and functional analysis of Mesorhizobium ACC deaminase." *Comput Biol Chem*, 68, 12–21.

Pui, C. H., Relling, M. V., Lascombes, F., Harrison, P. L., Struxiano, A., Mondesir, J. M., Ribeiro, R. C., Sandlund, J. T., Rivera, G. K., Evans, W. E., and Mahmoud, H. H. (1997). "Urate oxidase in prevention and treatment of hyperuricemia associated with lymphoid malignancies." *Leukemia*, 11(11), 1813–1816.

Pundir, S., Martin, M. J., and O'Donovan, C. (2017). "UniProt Protein Knowledgebase." *Methods Mol. Biol.*, 1558, 41–55.

Punnappuzha, A., PonnanEttiyappan, J., Nishith, R. S., Hadigal, S., and Pai, P. G. (2014). "Synthesis and Characterization of Polysialic Acid–Uricase Conjugates for the Treatment of Hyperuricemia." *Int J Pept ResTher*, 20(4), 465–472.

Pustake, S. O., Bhagwat, P. K., and Dandge, P. B. (2019). "Statistical media optimization for the production of clinical uricase from *Bacillus subtilis* strain SP6." *Heliyon*, 5(5), e01756.

Rahmatabadi, S. S., Sadeghian, I., Nezafat, N., Negahdaripour, M., Hajighahramani, N., Hemmati, S., and Ghasemi, Y. (2017). "*In silico* Investigation of Pullulanase Enzymes from Various *Bacillus* Species." *Curr. Proteom.*, 14(3), 175–185.

Rainbird, R. M., and Atkins, C. A. (1981). "Purification and some properties of urate oxidase from nitrogen-fixing nodules of cowpea." *Biochim Biophys Acta Enzymol.*, 659(1), 132–140.

Ram, S. K., Raval, K., and JagadeeshBabu, P. E. (2015). "Enhancement of a Novel Extracellular Uricase Production by Media Optimization and Partial Purification by Aqueous Three-Phase System." *Prep. Biochem.Biotechnol.*, 45(8), 810–824.

Ramya, L. N., and Pulicherla, K. K. (2015). "Studies on Deimmunization of Antileukaemic L-Asparaginase to have Reduced Clinical Immunogenicity- An *in silico* Approach." *Pathol.Oncol.Res.*, 21(4), 909–920.

Ramya, L. N., and Pulicherla, K. K. (2015). "Molecular insights into cold active polygalacturonase enzyme for its potential application in food processing." *J.Food Sci. Technol.*, 52(9), 5484–5496.

Rani, S., and Pooja, K. (2018). "Elucidation of structural and functional characteristics of collagenase from *Pseudomonas aeruginosa.*" *Process Biochem*, 64, 116–123.

Rani, S., Pooja, K., and Pal, G. K. (2017). "Exploration of potential angiotensin converting enzyme inhibitory peptides generated from enzymatic hydrolysis of goat milk proteins." *Biocatal.Agric.Biotechnol.*, 11, 83–88.

Rawlings, N. D., Barrett, A. J., and Bateman, A. (2010). "MEROPS: the peptidase database." *Nucleic Acids Res*, 38(Database issue), D227-233.

Rodríguez, V., Asenjo, J. A., and Andrews, B. A. (2014). "Design and implementation of a high yield production system for recombinant expression of peptides." *Microb. Cell Fact.*, 13, 65.

Roseman, M. A. (1988). "Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds." *J. Mol. Biol.*, 200(3), 513–522.

Roy, S., Maheshwari, N., Chauhan, R., Sen, N. K., and Sharma, A. (2011). "Structure prediction and functional characterization of secondary metabolite proteins of Ocimum." *Bioinformation*, 6(8), 315–319.

Rozenberg, S., Roche, B., Dorent, R., Koeger, A. C., Borget, C., Wrona, N., and Bourgeois, P. (1995). "Urate-oxidase for the treatment of tophaceous gout in heart transplant recipients. A report of three cases." *Rev Rhum Engl Ed*, 62(5), 392–394.

Saha, S., and Raghava, G. P. S. (2006). "Prediction of continuous B-cell epitopes in an antigen using recurrent neural network." *Proteins*, 65(1), 40–48.

Saitou, N., and Nei, M. (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees." *Mol. Biol. Evol.*, 4(4), 406–425.

Sammut, S. J., Finn, R. D., and Bateman, A. (2008). "Pfam 10 years on: 10,000 families and still growing." *Brief. Bioinformatics*, 9(3), 210–219.

Sánchez, F., Campos, F., Padilla, J., Bonneville, J. M., Enríquez, C., and Caput, D. (1987). "Purification, cDNA Cloning, and Developmental Expression of the Nodule-Specific Uricase from *Phaseolus vulgaris L.*" *Plant Physiol.*, 84(4), 1143–1147.

Sashidhar, R. B., Capoor, A. K., and Ramana, D. (1994). "Quantitation of ϵ-amino group using amino acids as reference standards by trinitrobenzene sulfonic acid." *J.Immunol. Methods*, 167(1–2), 121–127.

Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013). "Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments." *J. Comput. Aided Mol. Des.*, 27(3), 221–234.

Sauna, Z. E., Lagassé, D., Pedras-Vasconcelos, J., Golding, B., and Rosenberg, A. S. (2018). "Evaluating and Mitigating the Immunogenicity of Therapeutic Proteins." *Trends Biotechnol.*, 36(10), 1068–1084.

Schallmey, M., Singh, A., and Ward, O. P. (2004). "Developments in the use of *Bacillus* species for industrial production." *Can. J. Microbiol.*, 50(1), 1–17.

Schellekens, H. (2002). "Immunogenicity of therapeutic proteins: clinical implications and future prospects." *Clin Ther*, 24(11), 1720–1740.

Schiavon, O., Caliceti, P., Ferruti, P., and Veronese, F. M. (2000). "Therapeutic proteins: a comparison of chemical and biological properties of uricase conjugated to linear or branched poly(ethylene glycol) and poly(N-acryloylmorpholine)." *Farmaco*, 55(4), 264–269.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003). "SWISS-MODEL: An automated protein homology-modeling server." *Nucleic Acids Res.*, 31(13), 3381–3385.

Scott, J. T. (1978). "New knowledge of the pathogenesis of gout." *J Clin Pathol Suppl (R Coll Pathol)*, 12, 205–213.

Sharaf El Din, U. A. A., Salem, M. M., and Abdulazim, D. O. (2017). "Uric acid in the pathogenesis of metabolic, renal, and cardiovascular diseases: A review." *J Adv Res*, 8(5), 537–548.

Shen, L.-J., and Shen, W.-C. (2006). "Drug evaluation: ADI-PEG-20--a PEGylated arginine deiminase for arginine-auxotrophic cancers." *Curr. Opin. Mol. Ther.*, 8(3), 240–248.

Shen, M.-Y., and Sali, A. (2006). "Statistical potential for assessment and prediction of protein structures." *Protein Sci.*, 15(11), 2507–2524.

Sherman, M. R., Saifer, M. G. P., and Perez-Ruiz, F. (2008). "PEG-uricase in the management of treatment-resistant gout and hyperuricemia." *Adv. Drug Deliv. Rev.*, 60(1), 59–68.

Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., and Sherman, W. (2010). "Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field." *J. Chem. Theory Comput.*, 6(5), 1509–1519.

Silva Freitas, D. da, Spencer, P. J., Vassão, R. C., and Abrahão-Neto, J. (2010). "Biochemical and biopharmaceutical properties of PEGylated uricase." *Int.J.Pharm.*, 387(1–2), 215–222.

Singh, H., Ansari, H. R., and Raghava, G. P. S. (2013). "Improved Method for Linear B-Cell Epitope Prediction Using Antigen's Primary Sequence." *PLOS ONE*, 8(5), e62216.

Singh, S., Singh, G., Sagar, N., Yadav, P. K., Jain, P. A., Gautam, B., and Wadhwa, G. (2012). "Insight into trichomonas vaginalis genome evolution through metabolic pathways comparison." *Bioinformation*, 8(4), 189–195.

Sun, J., Xu, T., Wang, S., Li, G., Wu, D., and Cao, Z. (2011). "Does difference exist between epitope and non-epitope residues?" *J.Immunol. Res.*, 7,3,1.

Sundy, J. S., Ganson, N. J., Kelly, S. J., Scarlett, E. L., Rehrig, C. D., Huang, W., and Hershfield, M. S. (2007). "Pharmacokinetics and pharmacodynamics of intravenous PEGylated recombinant mammalian urate oxidase in patients with refractory gout." *Arthritis Rheum*, 56(3), 1021–1028.

Suzuki, K., Sakasegawa, S.-I., Misaki, H., and Sugiyama, M. (2004). "Molecular cloning and expression of uricase gene from *Arthrobacter globiformis* in Escherichia coli and characterization of the gene product." *J.Biosci.Bioeng.*, 98(3), 153–158.

Szczurek, P., Mosiichuk, N., Woliński, J., Yatsenko, T., Grujic, D., Lozinska, L., Pieszka, M., Święch, E., Pierzynowski, S. G., and Goncharova, K. (2017). "Oral uricase eliminates blood uric acid in the hyperuricemic pig model." *PLOS ONE*, (J. A. Joles, ed.), 12(6), e0179195.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., and Mering, C. von. (2015). "STRING v10: protein-protein interaction networks, integrated over the tree of life." *Nucleic Acids Res.*, 43(Database issue), D447-452.

Tamboli, A. S., Rane, N. R., Patil, S. M., Biradar, S. P., Pawar, P. K., and Govindwar, S. P. (2015). "Physicochemical characterization, structural analysis and homology modeling of bacterial and fungal laccases using *in silico* methods." *Netw Model Anal Health Inform Bioinform.*, 4(1), 17.

Tan, Q., Zhang, J., Wang, N., Li, X., Xiong, H., Teng, Y., He, D., Wu, J., Zhao, C., Yin, H., and Zhang, L. (2012). "Uricase from *Bacillus fastidious* loaded in alkaline enzymosomes: Enhanced biochemical and pharmacological characteristics in hypouricemic rats." *Eur J Pharm Biopharm.*, 82(1), 43–48.

Tan, Q.-Y., Zhang, J.-Q., Wang, N., Yang, H., Li, X., Xiong, H.-R., Wu, J.-Y., Zhao, C.-J., Wang, H., and Yin, H.-F. (2012). "Improved biological properties and hypouricemic effects of uricase from *Candida utilis* loaded in novel alkaline enzymosomes." *Int J Nanomedicine*, 7, 3929–3938.

Tan, Y., Xu, M., and Hoffman, R. M. (2010). "Broad selective efficacy of recombinant methioninase and polyethylene glycol-modified recombinant methioninase on cancer cells In Vitro." *Anticancer Res.*, 30(4), 1041–1046.

Tanaka, A., Yamamura, M., Kawamoto, S., and Fukui, S. (1977). "Production of uricase by Candida tropicalis using n-alkane as a substrate." *Appl. Environ. Microbiol.*, 34(4), 342–346.

Terkeltaub, R. (2007). "Learning how and when to employ uricase as bridge therapy in refractory gout." *J Rheumatol*, 34(10), 1955–1958.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." *Nucleic Acids Res.*, 22(22), 4673–4680.

Tochilin, V. P., Voronkov, I. I., and Mazaev, A. V. (1982). "Use of immobilized streptokinase (streptodecase) for treating thromboses." *Ter Arkh*, 54(11), 21–25.

Valderrama-Rincon, J. D., Fisher, A. C., Merritt, J. H., Fan, Y.-Y., Reading, C. A., Chhiba, K., Heiss, C., Azadi, P., Aebi, M., and DeLisa, M. P. (2012). "An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*." *Nat. Chem. Biol.*, 8(5), 434–436.

Varela-Echavarría, A., Montes de Oca-Luna, R., and Barrera-Saldaña, H. A. (1988). "Uricase protein sequences: conserved during vertebrate evolution but absent in humans." *FASEB J.*, 2(15), 3092–3096.

Vellard, M. (2003). "The enzyme as drug: application of enzymes as pharmaceuticals." *Curr. Opin. Biotechnol.*, 14(4), 444–450.

Verma, A., Singh, V. K., and Gaur, S. (2016). "Computational based functional analysis of *Bacillus* phytases." *Comput Biol Chem*, 60, 53–58.

Veronese, F. M., and Morpurgo, M. (1999). "Bioconjugation in pharmaceutical chemistry." *Farmaco*, 54(8), 497–516.

Vidya, J., Sajitha, S., Ushasree, M. V., Binod, P., and Pandey, A. (2017). "Therapeutic Enzymes: l-Asparaginases." *Curr. Dev. Biotechnol Bioeng.*, 249–265.

Vivek, D. D., Tanuj, S., Amit, P., and Sarad Kumar, M. (2012). "Insights to Sequence Information of Alpha Amylase Enzyme From Different Source Organisms." *Int J Adv Biotechnol Bioinform*, 1(1), 87–91.

Watanabe, Y., Yano, M., and Fukumoto, J. (1969). "Studies on the Formation of Uricase by Streptomyces : Part I. The Effect of Purine Bases on the Induced Formation of Uricase by the Cultured Cells." *Agric.Biol.Chem.*, 33(9), 1282–1290.

William L. Jorgensen, David S. Maxwell, and, and Tirado-Rives, J. (1996). "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids." *J.Am.Chem.Soc.*,118 (45), 11225-11236.

Wong, K., Cleland, L. G., and Poznansky, M. J. (1980). "Enhanced anti-inflammatory effect and reduced immunogenicity of bovine liver superoxide dismutase by conjugation with homologous albumin." *Agents and Actions*, 10(3), 231–239.

Wu, J., Lu, S., Zheng, Z., Zhu, L., and Zhan, X. (2016). "Modification with polysialic acid–PEG copolymer as a new method for improving the therapeutic efficacy of proteins." *Prep. Biochem.Biotechnol.,* 46(8), 788–797.

Wu, X. W., Lee, C. C., Muzny, D. M., and Caskey, C. T. (1989). "Urate oxidase: primary structure and evolutionary implications." *Proc. Natl. Acad. Sci. U.S.A.*, 86(23), 9412–9416.

Xiong, R., Umar, S., and Chen, J. (2013). "Process for Production of Recombinant Baboon Uricase in *Escherichia Coli Rosetta* (DE3)." *Biotechnol.Biotechnol.Equip.*, 27(5), 4141–4144

Yadav, M., Yadav, S., Yadav, D., and Yadav, K. (2017). "*In- silico* Analysis of Manganese Peroxidases from Different Fungal Sources." *Curr.Proteom.*, 14(3), 201–213.

Yadav, P. K., Singh, V. K., Yadav, S., Yadav, K. D. S., and Yadav, D. (2009). "*In silico* analysis of pectin lyase and pectinase sequences." *Biochemistry.*, 74(9), 1049–1055.

Yagura, T., Kamisaki, Y., Wada, H., and Yamamura, Y. (1981). "Immunological studies on modified enzymes. I. soluble L-asparaginase/mouse albumin copolymer with enzyme activity and substantial loss of immunogenicity." *Int Arch Allergy Appl Immunol*, 64(1), 11–18.

Yamamoto, K., Kojima, Y., Kikuchi, T., Shigyo, T., Sugihara, K., Takashio, M., and Emi, S. (1996). "Nucleotide sequence of the uricase gene from *Bacillus sp.* TB-90." *J. Biochem.*, 119(1), 80–84.

Yang, X., Yuan, Y., Zhan, C.-G., and Liao, F. (2012). "Uricases as Therapeutic Agents to Treat Refractory Gout: Current States and Future Directions: Formulated Uricase to Treat Refractory Gout." *Drug Dev.Res.*, 73(2), 66–72.

Yari, M., Eslami, M., Ghoshoon, M. B., Nezafat, N., and Ghasemi, Y. (2019). "Decreasing the immunogenicity of *Erwinia chrysanthemi* asparaginase via protein engineering: computational approach." *Mol Biol Rep*, 46(5), 4751–4761.

Yari, M., Ghoshoon, M. B., Vakili, B., and Ghasemi, Y. (2017). "Therapeutic Enzymes: Applications and Approaches to Pharmacological Improvement." *Curr Pharm Biotechnol.*, 18(7).

Yazdi, M. T., Zarrini, G., Mohit, E., Faramarzi, M. A., Setayesh, N., Sedighi, N., and Mohseni, F. A. (2006). "*Mucor hiemalis*: a new source for uricase production." *World J Microbiol Biotechnol*, 22(4), 325–330.

Yu, K., Petrovsky, N., Schönbach, C., Koh, J. Y. L., and Brusic, V. (2002). "Methods for prediction of peptide binding to MHC molecules: a comparative study." *Mol. Med.*, 8(3), 137–148.

Zarei, M., Nezafat, N., Rahbar, M. R., Negahdaripour, M., Sabetian, S., Morowvat, M. H., and Ghasemi, Y. (2019). "Decreasing the immunogenicity of arginine deiminase enzyme via structure-based computational analysis." *J. Biomol. Struct. Dyn.*, 37(2), 523–536.

Zhang, Q., Wang, P., Kim, Y., Haste-Andersen, P., Beaver, J., Bourne, P. E., Bui, H.-H., Buus, S., Frankild, S., Greenbaum, J., Lund, O., Lundegaard, C., Nielsen, M., Ponomarenko, J., Sette, A., Zhu, Z., and Peters, B. (2008). "Immune epitope database analysis resource (IEDB-AR)." *Nucleic Acids Res.*, 36(Web Server issue), W513-518.

Zhang, Y. Q., Zhou, W. L., Shen, W. De, Chen, Y. H., Zha, X. M., Shirai, K., and Kiguchi, K. (2005). "Synthesis, characterization and immunogenicity of silk fibroin-L- asparaginase bioconjugates." *J.Biotechnol.*, 120(3), 315–326.

Zhao, Y., Zhao, L., Yang, G., Tao, J., Bu, Y., and Liao, F. (2006). "Characterization of uricase from *Bacillus fastidious* A.T.C.C. 26904 and its application to serum uric acid assay by a patented kinetic uricase method." *Biotechnol.Appl.Biochem.*, 45(2), 75.

Zhou, X., Ma, X., Sun, G., Li, X., and Guo, K. (2005). "Isolation of a thermostable uricase-producing bacterium and study on its enzyme production conditions." *Process Biochem*, 40(12), 3749–3753.

Zhu, Y., Pandya, B. J., and Choi, H. K. (2011). "Prevalence of gout and hyperuricemia in the US general population: the National Health and Nutrition Examination Survey 2007-2008." *Arthritis Rheum.*, 63(10), 3136–3141.

Zobayer, N., and Hossain, A. B. M. A. (2018). "*In silico* Characterization and Homology Modeling of Histamine Receptors." *J.Biol.Sci.*, 18(4), 178–191.

APPENDIX I

Multiple sequence alignment of uricase from different microbial sources

```
Singulisphaera    1 MTTPSGANPADTAPDGKIVLGKNQYGKAGNHVVRITRDIAR--------------HEIED
Saccharomonospo   1 ---------------VAIVLGPNQYGKAQNHLVRVYRGISR--------------HEIRD
Nakamurella       1 ---------------VSVTLGHNQYGKAETRVVRVYRDSDP--------------HEIVD
Streptomyces      1 ---------------VPTTLGQNQYGKAENRVVRITRDGDT--------------HHIKD
Kitasatospora     1 ---------------VAHVLGQNQYGKAENRIVRVHRDADR--------------HHIKD
Thermobispora     1 ---------------VAILGRNQYGKAEVRVFRVYRDIPR--------------HEVRD
Hoyosella         1 ---------------VNISLGENQYGKAENRVVRIYRDIPR--------------HEIRD
Pseudomonas       1 --------------VSADYPRDLIGYGNNPPHPHWPGDAR---------------IALS
Truepera          1 -------------MTHRPTLSAIDYGKGDIGVYR-EAAPTYRVPAIPESPFTGRAHDLFA
Microlunatus      1 ----------MSDVTVSLGANKYGKAECRLVKITRDIAR--------------HEIED
Arthrobacter      1 -------------MSNKIVLGHNQYGKAEVRVVKITRDIDR--------------HEIED
Rhodococcus       1 ----------MTDLTGKIVLGANQYGKAENRVVRIYRDSPR--------------HEIHD
Actinobacteria    1 -------------MSKHILAQNQYGKAENRIVKVTRKGADGS----------WHEIRD
Bacillus          1 -------------MSERTML---YGKGDVFVYRTMATPLKGVRQIPESNFTGRSNTIFG
Microbacterium    1 MTDILTRTEAATTPTDRIVLGANQYGKAEVRVVKVTRDSDR--------------HEIED
consensus         1                  .  ....  ...*............ .        ...  .

Singulisphaera   47 LTVISQLRGD-FESCHTEGDNSHCVATDTQKNTIFSLARD-GVGS--PEAFLIRLGKHFT
Saccharomonospo  32 LTVSSALRGD-FSAAHIDGDQQDVLPTDTQKNTVFSFAKDKGVGA--IEDFALTLADHFV
Nakamurella      32 YNVSVALSGD-FEEIHRTGDNSNCLTTDATKNTVNAFAKEYSEAARQPESFGIALAKHFV
Streptomyces     32 LNVSVALSGD-MDDVHYSGSNANVLPTDTTKNTVYAFAKEHGIES--AEQFGIHLARHFV
Kitasatospora    32 LNVSVSLRGD-FEDVHLTGSNANCLPTDSTKNTVYAFAKRHGIES--AEGFAMLLARHFV
Thermobispora    32 LNVWIALRGD-FTDAHVTGDQSHVLPTDTQKNTVYALAKKEGIRA--IEDFALTLGDHFL
Hoyosella        32 LTVSIALRGD-FTVAHTEGDQSPVLPTDTQKQTVYAFAKTHGGGT--IEEYGIALARHFV
Pseudomonas      31 FVLNYEEGGE--RCVLHGDKESEAFLSEMVAAQPLQGVRHMSMES--LYEYCSRACVWRL
Truepera         47 AEVGVQVLGGPFAAAYTEGDNRNVVATDTMKNFVLKHALAFEGAT--LEAFLDSLGRAFF
Microlunatus     35 LNVTSQLRGARLVDSYLTGDNSLVVATDTQKNTVYAFARHGVGS--PEELLIRLGDHFV
Arthrobacter     34 LNVTSQLRGD-FEAAHLEGDNAHVVATDTQKNTIYAFARE-GVGS--PBAFLIRLCEHFT
Rhodococcus      37 LNVSSALRGD-FSAAHIAGDQSAVLPTDTQKQTAYAVAKTKGILH--LESYGIDLARHMV
Actinobacteria   36 LNVSVALRGE-FRDVHLTGDNANCLPTDTKNTVYAFGKEHGIAS--PEAFGIILAKHFV
Bacillus         44 VDLQVSLAGEAFLTSFTEGDNAKVVATDSMKNFILHHAGBYEGTT--MEGELAYVSACFL
Microbacterium   47 LNVSSQLRGD-FAAAHIAGDNAHVVPTDTQKNTVFAFARD-GIGS--PEKFLIRLADHFT
consensus        61 .....  ..*.    ..   ...  ....................     .  .  ....

Singulisphaera  103 TSFEW--VSGGRWAAEQETWQRIKVDG--------KEHNHAFVQNRTETRTAVILIDKDG
Saccharomonospo  89 AQTPA--ADGARIETDEMPWQRIDVDG--------RDHDHSFVQGGLGTRTTVVNVDGRG
Nakamurella      91 DDTAP--VTRARIKLEMYPWNRISHDG--------TPHPHAFARDGYVRTATVTYDGTN
Streptomyces     89 TSQEP--IEVARIRLEEYAWERIATSDGNSKFIGADEVHKHSFVRNGQEVRTTQITFDGEK
Kitasatospora    89 ENTERGVVHSARIRLEEYSWDRIAAASDRP--GETADTGHSFVRNGQEVRTTEVEYDGGR
Thermobispora    89 RQVPA--ATGARIALEEYAWDRIDVDG--------TGHDHGFVRRGQGTRTTVVTVEGRG
Hoyosella        89 DDVAP--VNEARIEIDEHAWDRAIVDG--------AEHEHAWTRRGPDVRTAAVTVSGSG
Pseudomonas      87 LKLFK----RRNVPLTVEAVAMAAQRN--------PEVIRAMVADGHEICSHGYRWIDYQ
Truepera        105 EAYPD--METLRITGREVPEAAAPVRQGET----FVPSPVLLSRSRGDCGAAAILVTREG
Microlunatus     93 SSFDW--IEGGLWQAEQYSWDRILLDG--------LEHIHSFVRKGQATRLATVQKVDGE
Arthrobacter     90 SSFDW--VTGGRWEAESYAWERIQAHG--------SAHIHSFVRKGQEVRTAVIVRDGAA
Rhodococcus      94 DDVAP--VDGARIELEEYAWTRAVDG--------VEHDHTWTRKGEEIRTSAVTVEGKG
Actinobacteria   93 SSQSP--IREAQVRLEEYAWERIPVP--------TRKEQHSFVRKGQEVRTAQITYSETT
Bacillus        102 DTYDH--ISEVEVSGRQHPFEAVEVGSPAG----VVPSFTVEREGMLEKPGTRISVGRDE
Microbacterium  103 GSFEW--VDGGRWSAESYEWERIPVAG--------AGHIHSFVRKGQETRTAVVLAEGAD
consensus       121              .   ...  ... ....             ........... .  ...  ..

Singulisphaera  153 ------PHLFAGLKDLTVLKSTGSEFHGPQDKYTTLVETSDRILSTDVAARWRYNT---
Saccharomonospo  139 PD--RTAHVVSGLKDLTLLKSTGSEERGFLKDRYTTLEETDDRILATSLVARWRYEG---
Nakamurella     141 ------LWVVSGLQDYVVLKITDSEFWGYLQDKYTTLKPTHDRVMATSVTCQWWH-----
Streptomyces    147 ------WEVISGLKDLTVNSTNSEFWGYVKDRYTTLKEAYDRILATDVEARWRY--NW-
Kitasatospora   147 ------FRVVSGLKDLVVVNTTDSEFWGYLKDPYTTLPEAYDRILATQVEARWAEGFAG-
Thermobispora   139 DE--RRAWVLSGLSDLIIAKITGSEFHGFLKDEYTTLEETHDRILATSIHTRWRYLT---
Hoyosella       139 AA--QRIWVVGGIRDLVLLKSTGSEFRDELRDEFTVLEPTTDRVMATSLVAQWRLAD---
Pseudomonas     135 Y-----MDEAQEREHVLEAIRILIELTGQRPVGWYTGRTGPNTRRLVMEEGGELYDSDT-
Truepera        159 -----VRELESGRLGLQLLKITGSSFASFARDAHTTLPEMHDRPLLIYLDAFWRYRDPAA
Microlunatus    143 ------THVTAGLKDLVVLKSTGSEFRGFPRDRYTTLIETDDRILATSVICRWRYLPE--
```

187

```
Arthrobacter      140 ------THLISGLKDLIVLKSTQSGFVGYPKDKYTTLPETTDRILATDVSARWRERSG--
Rhodococcus       144 EA--QRTWVVSGFKEMVILKSTGSEFHGFIEDDLTILEPTTDRVMATSLIAQWRYTT---
Actinobacteria    143 G-----LQVISGLKDLIVVNSTNSEFHGYIKDRYTTLKEAYDRILATKVIARWAH--SA-
Bacillus          156 NGEAQIGSLESAVHDLHLIVKGSFAGFIRDEYTTLPETKDRPLFIYLNISWTYKETEH
Microbacterium    153 ------RHVISGLTGLIVLKITQSGFVGYPRDRYTSLQETTDRILATEVIARWRYAQG--
consensus         181        ........ ........ .......... .. ...... .......
```

```
Singulisphaera    204 ------AELDFNAVYADVRRILLETFASVHSLALQQTLFQMGKNVLQAHRNIVDIRFSMP
Saccharomonospo   194 ------TDVGWDETFESIRSLLLEQFAEIHSRALQQTLYGMGHAVLEKHPEVAEIKFSAP
Nakamurella       190 ----TDSEADWAKSYDSAVATMSSVFAGHHSLALQQTMEAMGEAMIAEQPEIGEVRFSLP
Streptomyces      198 -TGDEQRMPNWEKSYQARKHMLQAFAETYSLSLQQTLQQMGSRIINSRSEIDEIRFSLP
Kitasatospora     200 -RDAEDAAPDWNHSYREVRRHLLEAFAQTYSYSLQQTLHAMGTRVLDHRAEVDEVRLELP
Thermobispora     194 ------TDVDWDKTEASVRSILLRQFATVHSLALQQTLYAMGSAVLEAHPEIAEIRLSAP
Hoyosella         194 ------INCDWDEIEASVRRTIVETFATHHSLALQQTLYTIGKNVLEAHRELVEVRLSAP
Pseudomonas       189 -YDDDLPYWDPASIAEKPHLVIPYTLDTNDMRFTQVVQGHNNGEQFFQYLKDAFDVLYEEG
Truepera          214 ALQTGTGEITGYVASEQVRDLLAATFHDVNSRIQHLVYEMGVRLLARFPQIREVRFEA-
Microlunatus      195 ---AVEAGIDFNALYAGVSEVFLATFASVHSLALQQTQVEMGKAALEAFPEIAEVKFAMP
Arthrobacter      192 ---TDFSSLDFNKSYDIVRSLLLEGFTEKYSHALQQTLDMGAKVLEAHSEIEEIKFSMP
Rhodococcus       199 ------TDVEWDAVVAGVANAMIERFANLQSLALQQTLYAMGEAVLEKYPFIAEISMSAP
Actinobacteria    195 -LAGDDDAFDWDQSYKKVRKNMLEAFAETYSYSLQQTINQMAERVLDNCPRVNEVRLNLP
Bacillus          216 ALGATP---EQYVAAEQVKHIAQTLFHQEDSPSIQKLIYDIGIRVLERFPQIETVSFES-
Microbacterium    205 ---VVLDDLDFTAVYDDVKRLLLEEFTRRYSAALQTTLEDMGRVVLEAYPEIAEIRMSMP
consensus         241        ...  ...  .. ....  ..   . ..*.... .. ...    ... .......
```

```
Singulisphaera    258 NKHHFVVDLAHFGQDNP---NVVFWAADRPEGLIEGTVKRDD--AADDLWESVAGEC---
Saccharomonospo   248 NKHHFLVDLSPFGVDNP---GEVFYAADRPYGLIEASVVRDDASERGSAWHAVPAWV---
Nakamurella       246 NKHHFVIDLSPYGLENP---NEVFHADDRPYGFIEGTIHNDLHKDAAPAPNQAFDPGQGW
Streptomyces      257 NNHHFLVDLEPFGLENDTADGAVYFAADRPYGLIEATVLRDGVEPKIPVDMTNL------
Kitasatospora     259 NKHHFLVDLEPFGLKNE---NEVYARMYGLIEGTVHRDGVTPVIPVS----------
Thermobispora     248 NKHHFLVDLQPFGLDNP---GEVFYASDRPYGLIEASVVRDDVPEAPEAWLATPGEC---
Hoyosella         248 NKHHFLYDLERFGIENR---NEVFHADDRPYGLIQATVLRDDAADAGPAWMQQLGWLT--
Pseudomonas       248 ATAPKMISTGLHCRLIGRPARMAALERFIQYAQSHDKVWFARREDIARHWHREHPEQETE
Truepera          273 QNRLWDHAFSDDETG-----RKVHTDPRPPYGRIGLTLTA--------------------
Microlunatus      252 NKHHFLVDLAPYGLENP---NEVFHADDRPYGLIEGTIVRDGVAAAPQAWTDVPTFV---
Arthrobacter      249 NKHHFLVDLSPFGLDNP---NEVFFAADRPYGLIEATVLRDDAEAADAAWSGIAGEC---
Rhodococcus       253 NKHHFVYDLGKFGIENN---LEVFNADDRPYGLIQATVEREDAPDAGSAWRTYSVVG---
Actinobacteria    254 NKHHFLVDLEPFGLKND---NEVYFAADRPMYGLIEGTVHRDGVQPVIATSDWIVA-----
Bacillus          272 NNRTWETVRDEIPASGE---GKVFTDPRPEIGFQKFTVLQEDVKREEEALR----------
Microbacterium    262 NKHHFVVDLAPFELDNP---NEVFFASDRSYGLIEAAVTREGQADAPQAWEEATAFC---
consensus         301 .........  ......      ......  .........  .  .   .  ..    .
```

```
Singulisphaera       -
Saccharomonospo      -
Nakamurella       303 -
Streptomyces         -
Kitasatospora        -
Thermobispora        -
Hoyosella            -
Pseudomonas       308 A
Truepera             -
Microlunatus         -
Arthrobacter         -
Rhodococcus          -
Actinobacteria       -
Bacillus             -
Microbacterium       -
consensus         361
```

Multiple sequence alignment of bacterial species of uricase protein sequences showing maximum conservation in between 16 and 337 amino acids

```
Aspergillus       1 MSA-----------VKAARYGKDNVRVYKVHKDEKT-GVQTVYEMTVCVLLEG-EIETSY
Trichoderma       1 MASSA--------YVSAARYGKDNVRVLKTDRDAAT-GVHTVTEMTVSCLLEG-DIDVSY
Lodderomyces      1 MSV-----------EIVKSSYGKANVKFLKVRKDANNPKVQEILEANVKVLLRG-KFDTSY
Cordyceps         1 MPS-----------ITAARYGKDNVRVFKAHRDPKT-GVHSITETVVCVLLEG-DIETSY
Aspergillus       1 MYN-----------LSDAQYGKDNVRLYKVIRDAGT-GVQTVYEITVCVLLEG-DIETSY
Fusarium          1 MPY-----------VSAARYGKDNVRVCKVDRDSST-GVQTVTEMTVCCLLEG-EIETSY
Conidiobolus      1 MASQIR------HKLDFQEYGKDNVRFVFVFKQAGG--QQSLVEYTVTVLLSGPRFTASY
Trametes          1 MSKVYEPEGEGLSYLSHARYGKDKVRVFVVRDG---AWHNIVEYNVTALVEG-DIEVSY
```

188

```
Cyberlindnera     1  MST----------TLSSSTYGKDNVKFLKVKKDPQNPKKQEVMEATVTCLLEG-GFDTSY
Hyphopichia       1  MS-----------QLLESSYGKANVKFLKVKKNPSNPTEQEVLEANVQVLLRG-GFDVSY
Ascoidea          1  MS-----------VLVDSIYGKDNVKFLKIKRDPSNPKLQDVIEVTVKVLLKG-AFDVSY
Paracoccidioide   1  MASR---------LAAARYGKDNVRVYKVYRNEVT-GVHTVVEMTVCVLLEG-EIESSY
Blastomyces       1  MTSR---------LAAARYGKDNVRVHRNEKT--GVHTVVEMTVCVLLEG-DIETSY
Penicillium       1  MSA----------LAAARYGKDNVRVCKVQRDEKT-GIQTVVEMTICVLLEG-DIETSY
Pseudogymnoascu   1  MAA----------IVEARYGKDNVRVYKVTRDEKS-GVQTVTEMTVCALLEG-DIEPSY
consensus         1  *.           .    ***  *         *      *    * . *  . **

Aspergillus      48  TKADNSVIVATDSIKNTIYITAKQNP-VTPPELFGSILGIHFIEKY-NIHAAHVNIITCH
Trichoderma      51  TKGDNSVVVATDSIKNTIIITAKQNP-VNPPELFAAILGAHFIDTY-SIHAANVKVITH
Lodderomyces     50  TEADNSSIVPTDTVKNTILVEAKTHN-VWPIESFAAHLAKHFTSKY-EQVEGIEVSIVQA
Cordyceps        48  TRADNSVVVATDSMKNTVEILAKQHP-VTPPELFAATVGAHFVDTY-AHIHVANVRIVTT
Aspergillus      48  TKEDNSVLVTTDAIKNTCYIVAKHNP-VHPPELFGSILGKHFITQY-KHIHTAHIDITCH
Fusarium         48  TQADNSVVVATDSIKNTIYITAKENP-VNPPELYASILGHFIEKY-KIHVANVSVKTV
Conidiobolus     53  TEADNNDVIATDTVKNTIYVVANQSKAIPNAEKFSVELAEFFLNKYPSHVQRVNVVVEQQ
Trametes         57  TEADNSVVVATDSIKNITYYLAKTSPHVLHPERFALHLGIHLVSKY-AHLKKAFITIEQL
Cyberlindnera    50  TEADNSSIVPTDTVKNTLVLAKTTE-IWPIERFAAKLAIHFVEKY-SHVSGVSVKIVQD
Hyphopichia      49  TKADNSPIVPTDTVKNTILVEAKKTD-VWPIERFAAHLAKHFSSKY-AHVMGVEVNIVQA
Ascoidea         49  TKADNSPIVPTDTVKNILIKARTED-PFLVENFANILASHFLAKY-DHISSVDVDIIQL
Paracoccidioide  49  TQADNSVVVATDSVKNTIYIMAKLHP-VTPPELFASILGIHFVKTY-KHIHTAHVDIITH
Blastomyces      49  TKADNSVVVATDSVKNTIYIMAKLHP-VTPPELFASILGHFTTTY-PHIHTAHVDITAI
Penicillium      48  TKADNSVVVATDSIKNTIFIKAKQNP-VTPPELFGSILGHFIEKY-NIHAAHVTIVTH
Pseudogymnoascu  48  TDADNSVVVATDSIKNTIYIKAKENP-VTPPELYASILASHFVDTY-KIHAAHVKVIVH
consensus        61  *..**.....**..**....  *.      .*  . .. ...   *     . .   .  ..

Aspergillus     106  RWTRVDTDG------KPHPHSFIRDSEEKRNVQVDVVEGKG-------IDIKSSISGLTV
Trichoderma     109  RWTRVTTRG------KPHPHSFLRDGQETRNVEARVSRKDG-------IAITSGIEGLTV
Lodderomyces    108  IWKKIQLEG------KPHDHSFKHEGPETRQTYLNYEKASKK------IQLSSSIKDLKV
Cordyceps       106  RWARVQVDG------RPHPHSFVKEGEEKRTVTARVSRGG-------LAINSGIADLTV
Aspergillus     106  TWERVTIEN------EPHPHSFYQDGSVKRQIRVDVSKASG-------INITSAIFGLSV
Fusarium        106  RWARLDVDG-----KPHPHSFFKDGEETRNVEVRVSRQEG-------IELKSSLVGLTV
Conidiobolus    113  SWDRQSFNG------AEHQHAFVGNNGEKHSSQVIAIQETPQSP--IKYEITSKLDGLTV
Trametes        116  RWQRIPTAQGEGQTPKPHSHAFYRDGDEKRTVEAETDASACKDK--IVASVTSGLKDLLV
Cyberlindnera   108  RWVKYAVDG------KPHDHSFIHEGGEKRITDLYYKRSGD-------YKLSSAIKDLTV
Hyphopichia     107  KWSRFNVND------APHPHSFKHEGPETRRVFLDYCKRKDK------LVVKSSIKDLTV
Ascoidea        107  RWSKYLVND------KFHDHSFIKDGDEKRICHLFRSRVGD-------FKLTSGIKDLTL
Paracoccidioide 107  RWTRVTIDD------EPHQHSFLRDGTETRNVSATVTKRTG-------ISLTSAIAGLSV
Blastomyces     107  RWTRVTIDG------KPHPHSFFRDGTITRNVSATVIDGTTTKHGTPTIKLTSSIAALSV
Penicillium     106  RWVRLDVDG------KPHPHSFIKPGSETRNVQVDVIEGKG-------IDINSSINGLTV
Pseudogymnoascu 106  RWTRVTIDG------KPHPHSFFRDGNETRNVEVTAREGKG-------IEVRSSISGLLV
consensus       121  .*  ..  ...        ..*.*.*..  .....  ...         . ..*  . .*.

Aspergillus     153  LKSTNSQFWGFLRDEYTTLKETWDRILSTDVDAIWQWKNESG----------------IQ
Trichoderma     156  LKSTGSAFHGFVRDEITTLPETWDRILSTDVDASWKWKPEAN----------------VQ
Lodderomyces    156  LKSTGSMFYGYNVCDYTTLKPTKDRILSTDVEAIWTEDSNKIGT------------LDDI
Cordyceps       153  LKSTGSAFHGFVRDEITTLPETWDRIMSTDVDASWDWKVFPD----------------IA
Aspergillus     153  LKSTRSEFWGYIKDEYTTLPETWDRILSSDVDIKWRWKHFGD----------------VS
Fusarium        153  LKSTGSAFHGFVRDEITTLPETWDRIFSTDVDAIWKWKKFEDS---------------VD
Conidiobolus    165  LKITGSSFKHFYADEYRTLPDAEDRILSTVVTLEWNYSTENN-----------------
Trametes        174  LKITGSAFESFVRDEYTTLMEVNDRIFSTSIDLSYKESPLRIPPPADGEGKVFDLPLTAE
Cyberlindnera   155  LKSTGSMFYGYNQCDETTIQPTTDRILSTDVDAIWVWDNKKIGS------------VYDI
Hyphopichia     155  LKSTGSMFYGYNVCDYTTLKPTKDRILSTDVDASWTYSSKALSN-----------FEDV
Ascoidea        154  LKSTGSMFYGYNKCDYTTLKPTNDRILSTDVDASWVWSNQITS--------------SNI
Paracoccidioide 154  LKSTNSQFHGFIRDEITTLPETWDRILSTDVDAKWTWTPEAS----------------IE
Blastomyces     161  LKSTNSQFHNFIRDEYTTLKETWDRILSTDVDASWTWQPEAS----------------VD
Penicillium     153  LKSTGSQFWGFVRDEYTTLKETWDRILSTDVAANWQWRRFTG----------------LT
Pseudogymnoascu 153  LKSTGSQFHGFVRDEYTTLPEVWDRILSTEIDSTWKWNNEES----------------LR
consensus       181  **.*.* *  ........**....**..*.......   .            .

Aspergillus     197  EVRSHVPKFDATWATAREVTLKTFAEDNSASVQATMYKMAEQILARQQLVETVEYSLPNK
Trichoderma     200  AVREAVSKFNPAWEEAREITLTRFAEDESASVQNTMYKMCEQILAAVPDTELVIYSLPNK
Lodderomyces    204  IKSC--SIFDNAYNQAKDVTLELFCKENSPSVQATMYNNMSQKILDNVSEIAVVRYVLPNI
Cordyceps       197  AVRAAPRFDAAHEAARAITFKRFADDNSASVQATMYKMSEDILAAVPEVNKVYYSLPNK
Aspergillus     197  DVEANVSHFIEAFASARNASFKAFADENSMSAQATIYKMAEKFLDCVPLASAVEYSWPNK
Fusarium        197  AVKQFAPKFDTVREAARNITLKTFAEDASASVQATMYKMSDLTILESVPEVAIVTYSLPNK
Conidiobolus    207  SYSVYETPFDEIRAAAKHFALDEFANKPSPSVQATIYDTANRIINKYHPVDKVSISLPNR
Trametes        234  QAKCTAWDGDSVATVARNITLEIFADDESASVQATIYKMAQRIVSEHPHVQSVSYNLPNK
Cyberlindnera   203  AKAADKGIFDNVYNQAREITLTTFALENSPSVQATMENMATQILEKACSVYSVSYALPNK
Hyphopichia     203  FKQADAGLFDKTYENARKVTLDLFALENSASVQATMYNNMSHKILELVPEVNVTYILPNK
Ascoidea        200  SDLSNNGLFDKTMASARKITLDLFATQNSPSVQATMYDMASEIIKTSSKIKSVTYILPNK
Paracoccidioide 198  DVKAIVPKFDETWEVARDTTLRVFAQDNSASVQNSMYKMAEQILDAQPLLSCVEYSLPNK
```

189

```
Blastomyces      205 AVKRAAPQFDETWRAVRDITMRTFAHDNSASVQNSMYKMAEQIIAVQPLLETVEYALPNK
Penicillium      197 EVKTNFEKFNAAWEEARAITLKTFAEDNSASVQATMYKMCEQIIAAVPLLDTVEYALPNI
Pseudogymnoascu  197 SVRASVPKFDKAWAAARPITMKLFAERNSPSVQNTMYKMCEHILKAVPDVETVEYSLPNK
consensus        241 ...   ...  .  .. ....  *. ..*.*.*.......... ..  .. ...* ...**.

Aspergillus      257 HYFEIDLSWHKGLQNTG-KNAEVEAPQSDPNGLIKCTVGRSS----LKSKL
Trichoderma      260 HFFEIDLSWHKGIKNTG-KDAEVYVPQSNPNGLIKCEVARGSPP-TRGSHL
Lodderomyces     262 HYILFNFEWKGIKNND-----DLFYPSSDPNGLITCTVGRKG----DKAKL
Cordyceps        257 HYFEIDLSWHKGIKNTG-KDAEVYVPQSNPNGLIKVEVSRDG----HSSKL
Aspergillus      257 HYVEIDLSWHKGIRNTD-KDAEVYLPQSAPNGLIKGTLTRSTLDRSRSAKL
Fusarium         257 HYFEIDLSWHKGIKNTG-KDAEVYAPQSPNGLIKCEVSRDS----LQSKL
Conidiobolus     267 HYEAADLPIFKQLGQDK-KVHDLYVPIAHPNGLIKATVKRN-----VAPKL
Trametes         294 HYVPVDMKY-IGIDNMSPPKAEVFIPLSAPSGLISATVSRK-----
Cyberlindnera    263 HYFLIDLKWKGLENDN-----ELFYPSPHPNGLIKCTVVRK-----EKTKL
Hyphopichia      263 HYILFNFEWKGIKDNS-----ELFYPSPDPNGLIKSTVGRK-----ESAKL
Ascoidea         260 HYFLVDFSWFNNLKND-----EVFYPSPHPNGLINCTVTRD-----PKSKL
Paracoccidioide  258 HYFEIDLSWYKGLKNTG-KDAEVYVPQSGPNGLIKCTVARREGERRETPKL
Blastomyces      265 HYFEIDLSWHKGLKNTG-KDAEVYAPQSPNGLIKCTVGRPGVDGRETPKL
Penicillium      257 HFFEVDLSWHNGLKNTG-KDAEVYAPQSNPNGLIKCTVARAG----QMAKL
Pseudogymnoascu  257 HYFEIDLSFHKGIKNTG-ADTTVYAPQSGPNGLIKCTVGRKDK--LEKGKL
consensus        301 *................ ......  *..  *.***.....*          ..
```

Multiple sequence alignment of fungal species of uricase protein sequences showing maximum conservation in between 14 and 338 amino acids

```
Sorghum bicolor    1 ----------MAENFQLQSRHGKSRVRVSRVWRRPAAAGGHIIVEWNVAVSIVSDCLPSY
Saccharum offic    1 ----------MAENFQLQSRHGKSRVRVSRVWRRPAAAGGHVIVEWNVAVSIVSDCLPSY
Prunus_persica     1 --MSN-----KIEGFNFDQRHGKQRVRVARVWRSENG--RHSIVEWNVGISLLIDSVVAY
Cajanus_cajan      1 -MAQE-----VVEGFKDQSHGKERVRVARVWRTKTKQG--RHFVVEWRVGIILLSDCVNSY
Populus_trichoc    1 --MAN-----ELDGFNFEQRHGKARVRVARVWRNKSDH-IHSMVEWGVSIILLSDCVNSY
Triticum aestiv    1 ----------MAGRFDLQGRHGKSRVRVSRVWRRPAEAGGHLFVEWSVAVSVVSDCLPSY
Momordica chara    1 --MAADDAVVAVNGFNFQQRHGKERVRVARVWRNRDG--RHFIVEWSVGISLILSDCVSAY
Helianthus_annu    1 --MATTKDQIQIGGNIFEQRHGKERVRVGRVWRSGDN--RHNFVEWNVSISLLSDCVNAY
Vitis_vinifera     1 --MAS----DKIDGFTLEQRHGKARVRLGRVWRSQSG--RHIFVEWTVSISLLSNCLAAY
Phaseolus_vulga    1 -MAQE-----VVEGFKFEQRHGKERVRVARVWRTPQG--RHFVVEWRVGILFSDCVNSY
Glycine max        1 MAKQE-----VVEGFKFEQRHGKERVRVARVWRTKRQG--QHFIVEWRVGIILFSDCVNSY
Arabidopsis lyr    1 --MAQ-----EAHGIRLEQRHGKARVRVGRVWRHDHDG-SEHHFVEWNVSISLLSHCLSSY
Lotus japonicus    1 --MAK-----EVEGFKFEQRHGKERVRVARVWKNNQG--RHFVVEWRVSISLLSDCLNSY
Medicago_trunca    1 --MAK-----NVEGFEFEQRHGKERVRVARVWKSKDGK-QQFVVEWRVSINLLSDCVNSY
Cicer_arietinum    1 --MAK-----NVEGFEFEQRHGKERVRVARVWKSKDGK-RHFVVEWRVSINLLSDCVNSY
consensus          1  ..       ....  ....***.***..***.. .  .. ***  *..........*

Sorghum bicolor   51 ISSDNSAIVATDSIKNTVYVKAKECTEVSMEEFAVILGRHFTSLYPQVSEATVTIVERP
Saccharum_offic   51 ISSDNSAIVATDSIKNTVYVKAKECTEVSMEEFAVILGRHFTSLYPQVSEATVTIVERP
Prunus_persica    52 TRDDNSDIVATDTMKNTVYAKAKECVEELSVEDFAIRLAKHFTSLYQQVTAALVKIVEKP
Cajanus_cajan     53 LRDDNSDIVATDTMKNTVYAKAKECSEILSVEDFAIVLAKHFVSFYKQVTGAIVNIVEKP
Populus_trichoc   53 VRDDNSDIVATDTMKNTVYVKAKECSEQLSAENFAILLARHFTSFYKQVTTAIVKIVEKP
Triticum aestiv   51 TSDDNSAIVATDSIKNTVYVKAKECIEVSMEEFAVILGRHFTSLYSQVSEATVTIVERP
Momordica chara   57 VRDDNSDIVATDTMKNTVYAKAKECSEQLSVEDFAILLAKHFTSYYKQVTTAIVKIVEKS
Helianthus_annu   57 VSADNSDIVATDTMKNTVYVKAKECKEQVSVEEFAILLAKHFTSFYPQVTTAIVKIVEKP
Vitis_vinifera    53 VRDDNSDIVATDSMKNTVYAKAKECAQQLSVEEFAIKLAKHFTSFYQQVTTAIVTVEEKP
Phaseolus vulga   53 LRDDNSDIVATDTMKNTVYAKAKECSDILSVEDFAILLAKHFSFYKKVTGAIVNIVEKP
Glycine max       54 LRDDNSDIVATDTMKNTVYAKAKECSDILSAEEFAILLAKHFVSFYQKVTGAIVNIVEKP
Arabidopsis lyr   53 YRDDNSDIVATDTMKNTVYVKAKECGDRLSVEEFAILICKHFCSFYPQVFTAIVSIIEKP
Lotus_japonicus   52 LRDDNSDIVATDTMKNTVYAKAKECSEILSVEDFAILLAKHFTSFYKQVTTAIVKIVEKP
Medicago_trunca   53 VRDDNSDIVATDTMKNTVYAKAKECSEILSVEDFAILLAKHFTSFYSQVTTAIVKIVEKP
Cicer arietinum   53 LRDDNSDIVATDTMKNTVYAKAKECSEILSAEDFAILLAKHFTSFYRQVTTAIVNIVEKP
consensus         61 ...***..****..*****.*****.....*.*.**......**.*.* .*...* ..*..

Sorghum bicolor  111 WERVIVDGKPHSHGFKVGVEKHSTEVIVKKSGSLLINSGIQGYSLLKTTQSGFEGFVTDR
Saccharum_offic  111 WERVIVDGKPHSHGFKVGVEKHSTEVIVKKSGSLLINSGIQGYSLLKTTQSGFEGFVTDR
Prunus_persica   112 WERVSIDGQPHEHGFKLGSEKHTTEVILKKSGALKVTSGIEGLALLKTTKSGFEGFIRDK
Cajanus cajan    113 WERVVVDGQPHEHGFKLGSEKHTAEAVQKSGALQLTSGIEGLSVLKTTKSGFEGFIRDK
Populus trichoc  113 WERVHLNGQPHEHGFKLGSEKHTAEVTQKSGVLKLTSGIEGLSVLKTTMSGFEGFIRDQ
Triticum aestiv  111 WERVAVDGKPHSHGFKLGSEKHTTEVNVKKSGSLLINSGIQGYSLLKTTQSGFEGFVRDK
Momordica_chara  117 WERVSVNGQPHNHGFKLGSEKHTTEVIVKKSGSLQLTSGIQGYSLLKTTQSGFERFIRDN
Helianthus_annu  117 WERISMNGQPHDHGFKLGSEKHTTEVIVQKNGSLGVTSGVVGLSLLKTTQSGFEGFIRDQ
Vitis_vinifera   113 WERAYLDGQPHDHGFKRGSEPTHTTEVIVEKSGALQVTSGIQGIALLKTTQAGFEGFIRDK
Phaseolus vulga  113 WERVIVDGQPHQHGFTLGSEKHTTEAIVQKSGSLQLTSGIEGLSVLKTTQSGFENFIRNK
Glycine_max      114 WERVIVDGQPHEHGFKLGSEKHTTEAIVQKSGSLQLTSGIEGLSVLKTTQSGFVNFIRDK
```

190

```
Arabidopsis_lyr   113 WERVSIDGKPHLHGFKLGSENHTAEARVEKSGALNLTSGIGGLALLKTTQSGFERFIRDK
Lotus japonicus   112 WERVNVDGQPHEHGFKLGSEKHTAEAIVQKSGALQLTSGIEGLSLLKTTKSGFEGFIRDK
Medicago trunca   113 WERVNVDGQPHDHGFKLGSEKHTTEAIVKKSGALQLTSGIEGLSVLKTTKSGFEGFVRDK
Cicer_arietinum   113 WERVSVDGQPHEHGFKLGSEKHTTEAIVKKSGALQLTSGIEGLSLLKTTKSGFEGFIRDK
consensus         121 *.*.....*.**.***..*.*.*.*.....  *.*.*..**..*...****..**..*...


Sorghum_bicolor   171 YRLLPDTRERIVATEVTAWWRYPFEHVSQLPSKPFCFTQLYQDVKRVLTETFFGPADVGV
Saccharum offic   171 YRLLPDTRERIVATEVTAWWRYPFEHVSQLPSKPFCFTQRYQDVKKVLAETFFGPADVGV
Prunus persica    172 YTVLPDTRERILATDLTASWTYPYESLYSIPQKPLYFTERYISVKKVLADTFYGPPKEGV
Cajanus_cajan     173 YTALPDTRERMLATEVTALWRYSYESLYSIPQKPLYFTKKYLEVKKVLADTFFGPPNGGV
Populus_trichoc   173 YTALPETRERMLATEVTALWRYSYESASSIPKNPLYFTEKYLDVKKSLANTFFGPPKEGV
Triticum_aestiv   171 YTLLPETRERIVATEVTAWWRYPFEHISQLPSKPFCFTQRYQDVKKVLADTFFGPSDVGV
Momordica_chara   177 YTVLPETRERILATEVSASWRYSIEALYSIPKKQFYFTETYLDVKKVLVDTFFGPPKEGV
Helianthus annu   177 NTILPETRERMLATEVSASWRYQFKSLSISNKPLQFTEKYLRVKKVLMDTFFGPPKEGV
Vitis vinifera    173 YTALADTRERIVATEVTASWRYPFESLSGIPLQPLYFTEKYLDVKKVLADTFFGPPRGGV
Phaseolus_vulga   173 YTALPDTRERILATEVTALWRYSYESLYNLPQKPLYFTKKYLEVKKVLADTFFGPPNRGV
Glycine_max       174 YTALPDTRERMVATEVTALWRYSYESLYSIPQKPLYFTEKYQEVKKVLADTFFGPPKGGV
Arabidopsis lyr   173 YTVLPETGERMLATEVNASWRYSYESVASIPTKGLYFTEKFMDVKKVLVDTFFGPPETGV
Lotus japonicus   172 YTVLPETRERMLATEVTALWRYSYESLYSIPQKPLYFTEKYLDVKKVIVDTFFGHPKEGV
Medicago trunca   173 YTLLPDTRERMLATEVTALWRYSYESLYSVPKKPLYFTEKYLDVKRVLLDTFFGSPKEGV
Cicer arietinum   173 FTALPETRERMLATEVTALWRYIYESFYSIPQKPLYFTLKYLDVKRVLVDTFFGSPKEGV
consensus         181 ...*..*.**..**...*  *.*...........  ....**.....**......**.*... **


Sorghum bicolor   231 YSPPVQNTLYLMAKFVLTRFPDISSVQLLMPNLHFLPVNLGSKETP-IVKFADDVYLPTD
Saccharum offic   231 YSPSVQNTLYLMCQEVLTRVAEISSIXLKMPNLHFLAVNLGSKETP-LVKIADDVYPATD
Prunus_persica    232 YSPSVQSTLYHMATNVLKGFPDIATVQLKMPNLHFLPVNLSNKDN-TIVKFDDVYLPTD
Cajanus_cajan     233 YSPSVQNTLYLMAKATLNRFPDIAFVHLKMPNLHFIPVNLSNKDG-PIVKFDDVYLPTD
Populus trichoc   233 YSASVQRTLFQMAKAVLNRFPDISSSIQLKMPNIHFLPVNLSSKEN-TIVKFNDVYLPTD
Triticum aestiv   231 YSPSVQNTLYLMARFVLTRFPDIASVQLLMPNLHFLPVNLGGKENPGIVKFADDVYMPTD
Momordica chara   237 YSPSVQYTLYDMAKSVLSRFPVISLVKLKMPNLHFLPVNLSTKDNRSIVKFDDVYLPTD
Helianthus_annu   237 YSPSVQATLYDMAKAVLGRFPDISSVHLKMPNIHFLPVNLSSKVNPVIVKFDDVYLPTD
Vitis_vinifera    233 YSPSVQSTLYQMAKNVLNRFPDISSIQLKMPNLHFLPVNISSKDNPAIVKFDDVYLPTS
Phaseolus_vulga   233 YSPSVQNTLYLMAKATLNRFPDIAYVHLKMPNLHFLPVNISSKDG-PIVKFDDVYLPTD
Glycine max       234 YSPSVQNTLYLMAKATLNRFPDIAYVSLKLPNLHFIPVNISNQDG-PIVKFDDVYLPTD
Arabidopsis lyr   233 YSPSVQRTLYLMCSAVLKRFADVSSIHLKMPNIHFLPVNLSTKENPSVVKFKDDVYLPTD
Lotus japonicus   232 YSPSVQSTLYQMAKATLNRFPDVASIQLKMPNIHFIPVNLSNKDG-PIVKFDDVYLPTD
Medicago_trunca   233 YSPSVQATLYQMAKAALNRFPDIASIQLKMPNIHFIPVNLSNKNG-QFVKFDDVYLPTD
Cicer_arietinum   233 YSPSVQSTLYQMCKATLNRFPDIASIQLKMPNIHFLPVNLSNKDG-HIVKFDDVYLPTD
consensus         241 **..** **. *.....*...........*..**.**..**.....    .**..***...*.


Sorghum bicolor   290 EPHGIIEAILSRPMS--KL
Saccharum offic   290 EPHRLIGAPLIRPMS--KL
Prunus_persica    291 EPHGSIEAILSRFWS--KM
Cajanus_cajan     292 EPHGSIKASLSRLWSNSKL
Populus_trichoc   292 EPHGSIEASLSRFWS--KM
Triticum aestiv   291 EPHGIIEAILSRANS--KL
Momordica chara   297 EPHGSIEACLSRFSS--KL
Helianthus_annu   297 EPHGSIEASLSRPLS--KM
Vitis_vinifera    293 EPHGSIEATVSRIRA--KI
Phaseolus_vulga   292 EPHGSIEASLSRVWS--KL
Glycine max       293 EPHGSIQASLSRLWS--KL
Arabidopsis lyr   293 EPHGSIEATVSRITS--KL
Lotus_japonicus   291 EPHGSIEASLSRILS--KM
Medicago_trunca   292 EPHGSIEASLSCSRS--KM
Cicer_arietinum   292 EPHGSIEASLSRTRS--KM
consensus         301 ***..*.*....  .  *.
```

Multiple sequence alignment of plant species of uricase protein sequences showing maximum conservation in between 12 and 317 amino acids

```
Rattus norvegic   1 -----MAHYHDD----------------------YGKNDEVEFVRTGYGKDMVKVLHI
Oryctolagus_cun   1 -----MATTK-----------------------KNEDVEFVRTGYGKDMVKVLHI
Mus_musculus      1 -----MAHYHDN----------------------YGKNDEVEFVRTGYGKDMVKVLHI
Bubalus_bubalis   1 -----MAHYHND----------------------YQKNDEVEFVRTGYGKDMVKVLHI
Pteropus alecto   1 -----MAHYHND----------------------YKKNDEVEFVRTGYGKDMVKVLHI
Myotis lucifugu   1 -----MAHYHND----------------------YKKNNEVEFVRTGYGKDMVKVLRI
Gallus_gallus     1 -----MSQVTIK-----------------------DIEVLNCEYGKNTIKFLRL
Otolemur_garnet   1 -----MAHYN-E----------------------YKKDDEVEFVQTGYGKDMVKVLHI
Microcebus_muri   1 -----MAHYNNE----------------------YKKNDEVEFVRTGYGKDMVKVLHI
```

191

```
Mus_caroli          1  -----MVHYHDDD----------------------YGKNDEVEFVRTGYGKDMVKVLHI
Callorhinus urs     1  -----MAHYHND-----------------------YKKNDEVEFVRTGYGKDMVRVLHI
Canis lupus din     1  -----MAHYHND-----------------------YKKNDEVEFVRTGYGKDMVKVLHI
Bombyx_mori         1  MPWTSTNRIYAKPSSTSANETPSPSLASPRVALTAAEDSGGRFELCDHGYGKSSVKLLHV
Musca_domestica     1  MFANPLQKPTAKGKSFQDREAP-----------------HQYAISDYGYGKDAVKILHV
Enhydra_lutris_     1  -----MAHYDHD-----------------------YKKNDEVEFVRTGYGKDMVRVLCI
consensus           1   .....  .              .  ...........***......*..

Rattus norvegic    32  QRDGKYHSIKEVATSVQLTLRSKKDYLHGDNSDIIPTDTIKNTVHVLAKFKGIKSIETFA
Oryctolagus_cun    28  QRDGKYHSIKEVATSVQLTLSSKQDYVYGDNSDIIPTDTIKNTVHVLAKFKGIKSIEVFA
Mus_musculus       32  QRDGKYHSIKEVATSVQLTLRSKKDYLHGDNSDIIPTDTIKNTVHVLAKLRGIRNIETFA
Bubalus_bubalis    32  QRDGKYHSIKEVATSVQLTLNSRREYLHGDNSDIIPTDTIKNTVHVLAKFKGIKSIETFA
Pteropus_alecto    32  QRDGKYHTIKEVATSVQLTLSSKKDYLTGDNSDIIPTDTIKNTVHVLAKLKGIKSIETFS
Myotis lucifugu    32  QRDGKTHSIKEVATSVQLTLSSKKDYLHGDNSDIIPTDTIKNTVHVLAKFKGIKSIETFA
Gallus gallus      27  HREGKKHFVKEVEVCTHLRLISAHEYLDGNNSFMIPTDTIKNIVLVLAKKNGISSIEQFA
Otolemur_garnet    31  QRDGKYHSIKEVATSVQLTLSSKKDYLHGDNSDIIPTDTIKNTVHVLAKFKGIKSIEAFA
Microcebus_muri    32  QRDGKYHSIKEVATSVQLTLSSKKDYLHGDNSDIIPTDTIKNTVHVLAKFKGIKSIEAFA
Mus caroli         33  QRDGKYHSIKEVATSVQLTLRSKKDYLHGDNSDIIPTDTIKNTVHVLAKHRGIRNIETFA
Callorhinus urs    32  QRDGKYHNIKEVATSVQLTLSSKKDYVHGDNSDIIPTDTIKNTVHVLAKFKGIKSIETFA
Canis lupus din    32  QRDGKYHSIKEVATSVQLTLSSKKDYVYGDNSDIIPTDTIKNTVHVLAKFKGIKSIETFA
Bombyx mori        61  HRDEGHHFVIREFEVFTELKIASETAYILGDNKEVVATDSQKNTVYLLAKKYGVKLPEEFC
Musca_domestica    43  KRDGPVHSIKEFEVGTHLKLYSKKDYFHGDNSDIVATDSQKNTVYLLAKKHGIENPENFA
Enhydra_lutris_    32  QRDGKYHSIKEVATSVQLTLSSKKDYVHGDNSDIIPTDTIKNTVHVLAKFKGIKSIETFA
consensus          61  .*.....*...*......*.*.*....*..*.*......**..**.*..***..*.....*.

Rattus norvegic    92  MNICEHFLSSFSHVTRAHVHVEEVPWKRFEKNGVK------------------HVHAFIH
Oryctolagus_cun    88  MNICEHFLSSFNHVVRVHVYVEEVPWKRLEKNGVQ------------------HVHAFIH
Mus_musculus       92  MNICEHFLSSFNHVTRAVYVEEVPWKRFEKNGIK------------------HVHAFIH
Bubalus bubalis    92  MNICEHFLSSFNHVIRVQVYVEEVPWKRFEKNGVK------------------HVHAFIH
Pteropus alecto    92  MNICEHFLSSFNHVIRAQVYVEEVPWKRFEKKGAK------------------HVHAFIH
Myotis lucifugu    92  MNICEHFLSSFNHVIRAQVYVEEVPWKRFEKNGAK------------------HVHAFIH
Gallus_gallus      87  IDICKHFVMTTFCQVAYVKTYIQEVPWQRQYQNGVP------------------HIHSFIL
Otolemur_garnet    91  VTICEHFLSSFNHVIRAQVFVEEVPWKRFEKNGVK------------------HVHAFIH
Microcebus_muri    92  LNICEHFLSSFNHVIRAQVFVEEVPWKRFEKNGVK------------------HVHAFIH
Mus caroli         93  MNICEHFLSSFNHVTRAVYVEEVPWKRFEKNGIK------------------HVHAFIH
Callorhinus urs    92  MNICEHFLSSFNHVIRTQVYVEEVPWKRFEKNGVK------------------HVHAFIH
Canis lupus din    92  MNICEHFLSSFNHVIRAQVYVEEVPWKRFEKNGVK------------------HVHAFIH
Bombyx_mori       121  AVVVNHFIYMYKQVLEAKCRVIEYPWERLQAGTP------------------HSHAFVF
Musca_domestica   103  TLLIARHFMQKYAHVEEVHTHVEEYPWQRVSAEETGCQDGNGNCNYTNINNRARHNHAFIF
Enhydra_lutris_    92  MNICEHFLSSFNHVIRVQVYVEEVPWKRFEKNGVK------------------HVHAFIH
consensus         121  .....**......*.........*.**.*.......                   *.*.*..

Rattus norvegic   134  TPTGTHFCDVEQVRNGP-PLIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRCFATQ
Oryctolagus_cun   130  TPTGTHFCEVEQRRSGL-PVIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRCFATQ
Mus_musculus      134  TPTGTHFCEVEQMRNGP-PVIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRCFATQ
Bubalus_bubalis   134  TPTGTHFCEVEQIRSGP-PVIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRCFATQ
Pteropus alecto   134  TPTGTHFCEVEQKRSGF-PVIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRCLATQ
Myotis lucifugu   134  TPTGTHFCEVEQMRNGP-PVIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRCFATQ
Gallus_gallus     129  VPDGIRFCEAEQCRNGP-LVVCAGIKDLKLMKTTQSGFEGFYRNEHTTLPERNDRILCGE
Otolemur_garnet   133  TSTGTHFCDVEQVRNGP-PVIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRCFATQ
Microcebus_muri   134  TPTGTHFCEVEQMRNGP-PVIHSGIKDLKVLALTQSGFEGFIKDQFTTLPEVKDRCFATQ
Mus caroli        135  TPTGTHFCEVEQMRNGP-PVIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRCFATQ
Callorhinus urs   134  TPTGTHFCEVEQMRGGP-PVIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRCFATQ
Canis_lupus_din   134  NPTGTHFCEVEQMRSGP-PVIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRCFATK
Bombyx_mori       162  SPIATRVSEVSQTRHEA-VVVKSGLSGLRVLKTTQSAFVDFVQDEYTTLTDAVERIFSTI
Musca_domestica   163  TPTEIRVCDVVLRRTDEKQTVISGIRGLRVLKTTQSFVNFVNDEFRSLPDQYDRIFSTV
Enhydra_lutris_   134  TPTGTHFCEIEQMRGGP-PVIHSGIKDLKVLKTTQSGFEGFIKDQFTTLPEVKDRCFATQ
consensus         181  ............*  ..  .....*...*...*****.*..*........*.......

Rattus_norvegic   193  VYCKWRYQN-RDVDFEATWGAVRDIVLKKFAGPYDRGEYSPSVQKTLYDIQVLILSQLPE
Oryctolagus_cun   189  VYCKWRYQHSQDVDFEATWDIVRDTVLEKFAGPYDKGEYSPSVQKTLYDIQVLILSRVPQ
Mus_musculus      193  VYCKWRYQR-RDVDFEATWGAVRDIVLQKFAGPYDKGEYSPSVQKTLYDIQVLSLSQLPE
Bubalus_bubalis   193  VYCKWRYFQGRDVDFEATWEAVRSIVLKKFAGPYDKGEYSPSVQKTLYDIQVLSLSQLPE
Pteropus alecto   193  VYCKWRYHWCRDVDFEATWDAVRDIVLEKFAGPYDKGEYSPSVQKTLYDIQVLSLSRVPE
Myotis lucifugu   193  VYCKWRYRGS-DVDFEATWDTVRDIVLEKFAGPHDKGEYSPSVQKTLYDIQALSLSRVPE
Gallus_gallus     188  FHCKWSYGECRDFDFDCIWSKVRECILEAFSGPPDCGEYSPSYQRTVNCIQMCVLSRVPQ
Otolemur_garnet   192  VYCKWRYHCGRSVDFEATWDTVRDIVLEKFAGPYDKGEYSPSVQKTLYDIQVLSLSHVPE
Microcebus muri   193  VYCKWRYHQGRNVDFEATWDTVRDIVLEKFAGPYDKGEYSPSVQKTLYDIQVLSLSRVPE
Mus caroli        194  VYCKWRYQR-RDVDFEATWGAVRDIVLQKFAGPYDKGEYSPSVQKTLYDIQVLSLSQLPE
Callorhinus urs   193  VYCKWRYHRCRDVDFDAICDTVRDIVLEKFAGPYDKGEYSPSVQKTLYDIQVHSLSRVPE
Canis_lupus_din   193  VYCKWRYHQGRDVDFEATWDTVRDIVLEKFAGPYDKGEYSPSVQKTLYDIQVHSLSRVPE
```

192

```
Bombyx_mori      221 VEAEWIYDNMRTADFLNAWLTVKDALLKFAGPPDTGVYSPSVQHTLYQAEKTVLEKVSE
Musca domestica  223 VDCSWEYSNTDKVLFCKSWNIVKNILTRKFAGDPNVGTSSPSVQHTLYLTEKEVLDTLPE
Enhydra lutris   193 VYCKWRYHQYRDVDFTATWDTVRDIVLEKFAGPYDKGEYSPSVQKTLYDIQVHSLSQVPE
consensus        241 ....*.*.  ....*..............*.*.....*...***.*.*.........*.....


Rattus_norvegic  252 IEDMEISLPNIHYFNIDMSKMGLINK-----EEVLLPLDNPYGKITGTVKRK----LPSR
Oryctolagus_cun  249 IEDMEISLPNIHYFNIDMSKMGLINK-----EEVLLPLDNPYGKITGTVKRK----LSSR
Mus musculus     252 IEDMEISLPNIHYFNIDMSKMGLINK-----EEVLLPLDNPYGKITGTVKRK----LPSR
Bubalus bubalis  253 IEDMEISLPNIHYFNIDMSKMGLINK-----EEVLLPLDNPYGRITGTVKRK---LTSR
Pteropus_alecto  253 IELMEISLPNIHYFNIDMSKMGLINK-----EEVLLPTDNPYGRITGTVKRK----LLSK
Myotis_lucifugu  252 IEDMEISLPNIHYFNIDMSKMGLVNK-----DEVLLPSDNPYGRITGTVKRK----LASR
Gallus_gallus    248 VQVLEVILNNNFYNVVDMKALGCTND-----KEVLVPVETPYGSCACTLGRKKYLEAQSH
Otolemur_garnet  252 IEDMEISLPNIHYFNIDMTKMGLINK-----EEVLLPLDNPYGRITGTVKRK----LSSR
Microcebus muri  253 IEDMEISLPNIHYFNIDMSKMGLINK-----EEVLLPLDNPYGRITGTVKRK----LSSR
Mus caroli       253 IEDMEISLPNIHYFNIDMSKMGLINK-----EEVLLPLDNPYGKITGTVKRK----LPSR
Callorhinus_urs  253 IEDMEISLPNMHYFNIDMSKMGLINK-----EEVLLPLDNPYGKITGTVKRK----LASK
Canis_lupus_din  253 VEDMEISLPNIHYFNIDMSKMGLINK-----EEVLLPLDNPYGRITGTVKRK---LASK
Bombyx mori      281 ITWLKMTMPNKHYLNIDMSKFPVNVTKGDPRHFIYHPLDPAGLIYAQLRR----PKSK
Musca domestica  283 VSVISMTMPNKHYFNDTKPFQAVVP-GE-NNEVFIPVDKPHGTIYAQLARK----DIAS
Enhydra lutris   253 IEDMEISLPNMHYFNIDMSKMGLINK-----EEVLLPLDNPYGRITGTVKFK---LASK
consensus        301 .........*..*...*.........     .....*...*.*.........    . ..


Rattus norvegic  303 L----------------
Oryctolagus cun  300 L----------------
Mus_musculus     303 L----------------
Bubalus_bubalis  304 L----------------
Pteropus alecto  304 L----------------
Myotis lucifugu  303 L----------------
Gallus gallus    303 MIKDEKQSQFGLVAAQGK
Otolemur_garnet  303 L----------------
Microcebus_muri  304 L----------------
Mus_caroli       304 L----------------
Callorhinus urs  304 L----------------
Canis lupus din  304 L----------------
Bombyx mori      337 L----------------
Musca_domestica  337 HL----------------
Enhydra_lutris_  304 L----------------
consensus        361 .
```

Multiple sequence alignment of animal species of uricase protein sequences showing maximum conservation in between 37 and 304 amino acids

```
Singulisphaera   1 -------------------------MTTPSGANPADTAPDGKIVLGKNQYGKAGNHVVR
Saccharomonospo  1 ----------------------------------------MAIVLGPNQYGKAQNHLVR
Nakamurella      1 ----------------------------------------MSVTLGHNQYGKAETRVVR
Streptomyces     1 ----------------------------------------MPTILGQNQYGKAENRVVR
Kitasatospora    1 ----------------------------------------MAHVLGQNQYGKAENRIVR
Thermobispora    1 ----------------------------------------MAIVLGRNQYGKAEVRVFR
Hoyosella        1 ----------------------------------------MNISLGENQYGKAENRVVR
Pseudomonas      1 ---------------------------------------MSADYPRDLIGYGN
Truepera         1 --MTHRPTLSAID-----------------------YGKGDIGVYR-FAAPTYRVPA
Microlunatus     1 ------------------------------------MSDVTVSLGANKYGKAECRLVK
Arthrobacter     1 ------------------------------------MSNKIVLGHNQYGKAEVRVVK
Rhodococcus      1 -----------------------------MTDLTGKIVLGANQYGKAENRVVR
Actinobacteria   1 --------------------------------MSKHILAQNQYGKAENRIVK
Bacillus         1 --MSERTML-----------------------YGKGDVFVYRTYATPLKGVRQ
Microbacterium   1 ------------------------MTDILTRTEAATTPTDRIVLGANQYGKAEVRVVK
Aspergillus      1 --MSA--------------------------------------VKAARYGKDNVRVYK
Trichoderma      1 --MASSA-------------------------------------YVSAARYGKDNVRVLK
Lodderomyces     1 --MSV---------------------------------------EIVKSSYGKANVKFLK
Cordyceps        1 --MPS----------------------------------------ITAARYGKDNVRLFK
Aspergillus      1 --MYN----------------------------------------LSDAQYGKDNVRLYK
Fusarium         1 --MPY----------------------------------------VSAARYGKDNVRVCK
Conidiobolus     1 --MASQIR-----------------------------------HKLDFQEYGKQNVRFVR
Trametes         1 --MSKVYEP----------------------------EGEGLSYLSHARYGKDKVRVFR
Cyberlindnera    1 --MST---------------------------------------TLSSSTYGKDNVKFLK
Hyphopichia      1 --MS-----------------------------------------QLLESSYGKANVKF
Ascoidea         1 --MS-----------------------------------------VLVDSIYGKDNVKFLK
```

```
Paracoccidioide   1 --MASR---------------------------------------LAAARYGKDNVRVYK
Blastomyces       1 --MTSR---------------------------------------LAAARYGKDNVRVCK
Penicillium       1 --MSA----------------------------------------LAAARYGKDNVRVCK
Pseudogymnoascu   1 --MAA----------------------------------------IVEARYGKDNVRVYK
Sorghum           1 ------------------------------------------MAENFQLQSRHGKSRVRVSR
Saccharum         1 ------------------------------------------MAENFQLQSRHGKSRVRVSR
Prunus            1 --MSN-------------------------------------KIEGFNFDQRHGKQRVRVAR
Cajanus           1 -MAQE-------------------------------------VVEGFKFDQSHGKERVRVAR
Populus           1 --MAN-------------------------------------ELDGFNFEQRHGKARVRVAR
Triticum          1 ------------------------------------------MAGRFDLQGRHGKSRVRVSR
Momordica         1 --MAADDAV---------------------------------VAVNGFNFQQRHGKDRVRVGR
Helianthus        1 --MATTKDQ-------------------------------IQIGGNIFEQRHGKERVRVGR
Vitis             1 --MAS-------------------------------------DKIDGFTLEQRHGKARVRLGR
Phaseolus         1 -MAQE-------------------------------------VVEGFKFEQRHGKERVRVAR
Glycine           1 MAKQE-------------------------------------VVEGFKFEQRHGKDRVRVAR
Arabidopsis       1 --MAQ-------------------------------------EAHGIRLEQRHGKARVRVGR
Lotus             1 --MAK-------------------------------------EVEGFKFEQRHGKERVRVAR
Medicago          1 --MAK-------------------------------------NVEGFEFEQRHGKDRVRVAR
Cicer             1 --MAK-------------------------------------NVEGFEFEQRHGKDRVRVAR
Rattus            1 --MAHYDD-YGKND------------------------EVEFVRTGYGKDMVKVLH
Oryctolagus       1 --MATTK-----KNE-----------------------DVEFVRTGYGKDMVKVLH
Mus               1 --MAHYDN-YGKND------------------------EVEFVRTGYGKDMVKVLH
Bubalus           1 --MAHYND-YQKND------------------------EVEFVRTGYGKDMVKVLH
Pteropus          1 --MAHYND-YKKND------------------------EVEFVRTGYGKDMVKVLH
Myotis            1 --MAHYND-YKKNN------------------------EVEFVRTGYGKDMVKVLR
Gallus            1 --MSQVTIK-----------------------------DIEVLNCEYGKNTIKFLR
Otolemur          1 --MAHYN-E-YKKDD-----------------------EVEFVQTGYGKDMVKVLH
Microcebus        1 --MAHYNNE-YKKND-----------------------EVEFVRTGYGKDMVKVLH
Mus               1 --MVHYHDDDYGKND-----------------------EVEFVRTGYGKDMVKVLH
Callorhinus       1 --MAHYHND-YKKND-----------------------EVEFVRTGYGKDMVKVLH
Canis             1 --MAHYHND-YKKND-----------------------EVEFVRTGYGKDMVKVLH
Bombyx            1 -MPWTSTNRIYAKPSSTSANETPSPSLASPRVALTAAEDSGGRFELCDHGYGKSSVKLLH
Musca             1 -MFANPLQKPTAKGKSFQDREAP------------------HQYAISDYGYGKDAVKILH
Enhydra           1 --MAHYDHD-YKKND-----------------------EVEFVRTGYGKDMVRVLC
consensus         1 .                                    .   .... .....


Singulisphaera    35 ITRDTAR---HEIEDITVISQLRGD-FESCHTEGDNSHCVATDTQKNTIFSLARD---GV
Saccharomonospo   20 VYRGTSR---HEIRDLTVSSALRGD-FSAAHIDGDQQDVLPTDTQKNTVFSFAKE--KGV
Nakamurella       20 VYRDSDP---HEIVDYNVSVALSGD-FEEIHRTGDNSNCITTDATKNTVNAFAKEYSEAA
Streptomyces      20 ITRDGDT---HHIKDINVSVALSGD-MDDVHYSGSNANVIPTDTTKNTVYAFAKE--HGI
Kitasatospora     20 VHRDADR---HHLKDINVSVSLRGD-FEDVHIDGDSNANCIPTDSTKNTVYAFAKR--HGI
Thermobispora     20 VYRDTPR---HEVRDINVWTALRGD-FTDAHVTGDQSHVIPTDTQKNTYAIAKK--EGI
Hoyosella         20 IYRDTPR---HEIRDTVSTALRGD-FTVAHTEGDQSPVIPTDTQKQTVYAFAKT--HGG
Pseudomonas       15 NPPHPHWP-GDARIALSFVINYEEGGERCVLHGDKESEAFLSEMVAAQPLQGVRHMS--M
Truepera          32 IPESPFTGRAHDIFAAEVGVQVLGGPFAAAYTEGDNRNVVATDTMKNFVLKHALAFE--G
Microlunatus      23 ITRDTAR---HEIEDIVISQLRGARLVDSYITGDNSLVVATDTQKNTVYAFAREH--GV
Arthrobacter      22 ITRDTDR---HEIEDINVISQLRGD-FEAAHIEGDNAHVVATDTQKNTIYAFARE---GV
Rhodococcus       25 IYRDSPR---HEIHDINVSSALRGD-FSAAHLAGDSAVIPTDTQKQTAYAYART--KGL
Actinobacteria    21 VTRKGADGSWHEIRDINVSVALRGE-FRDVHITGDNANCLPTDTTKNTVAFCKE--HGI
Bacillus          29 IPESNFTGRSNTIFGVDIQVSLAGEAFLTSETEGDNAKVVATDSMKNFTLHHAGEYE--G
Microbacterium    35 VTRDSDR---HEIEDINVSSQLRGD-FAAAHLAGDNAHVVPTDTQKNTVYAFARD---GI
Aspergillus       19 VHKDEKT-GVQTVYEMTVCVLLEG-EIETSYTKADNSVIVATDSIKNTIYITAKQNP--V
Trichoderma       22 TDRDAAT-GVHTVTEMTVSCLLEG-DIDVSYTKGDNSVVVATDSIKNTIFITAKQNP--V
Lodderomyces      20 VRKDANNPKVQEILEANVKVLLRG-KFDTSYTEADNSSIVPTDTVKNTILVEAKTHN--V
Cordyceps         19 AHRDPKT-GVHSITETVCVLLEG-DIETSYTRADNSVVATDSMKNTVFILAKQHP--V
Aspergillus       19 VHRDAGT-GVQTVYEITVCVLLEG-DIETSYTKEDNSLVTTDAIKNTCYIVAKHNP--V
Fusarium          19 VDRDSST-GVQTVTEMTVCCLLEG-EIETSYTQADNSVVATDSIKNTIYITAKENP--V
Conidiobolus      24 VFKQAGG--QQSLVEYTVTVLLSGPRFTASYTEADNNDVIATDTVKNIIYVVANQSK-AL
Trametes          30 VVRDG---AWHNIVEYNVTALVEG-DIVSYTEADNSVVVATDSIKNITYYLARTSP-HV
Cyberlindnera     20 VKLDPQNPKKQEVMEATVICLLEG-GFDTSYTEADNSSIVPTDTVKNTILVLAKTTE--I
Hyphopichia       19 VKKNPSNPTEQEVLEANVQVLLRG-GFDVSYTKADNSPIVPTDTVKNTILVEAKKTD--V
Ascoidea          19 IKRDPSNPKIQDVIEVTVKVLLKG-AFDVSYTKADNSPIVPTDTVKNTILIKARTED--P
Paracoccidioide   20 VYRNEVT-GVQTVVEMTVCVLLEG-EIESSYTQADNSVVATDSMKNTIYIMAKLHP--V
Blastomyces       20 VHRNEKT-GVHTVVEMTVCVLLEG-DIETSYTKADNSVVATDSMKNTIYIMAKLHP--V
Penicillium       19 VQRDEKT-GIQTVVEMTICVLLEG-DIETSYTKADNSVVATDSIKNTIFIKAKQNP--V
Pseudogymnoascu   19 VHRDEKS-GVQTVTEMTVCALLEG-DIEPSYTDADNSVVATDSIKNTIYIKARENP--V
Sorghum           21 VWRRPAAAGGHIIVEWNVAVSIVS-DCLPSYISSDNSAIVATDSIKNTVYVKAKECT-EV
Saccharum         21 VWRRPAAAGGHVIVEWNVAVSIVS-DCLPSYISSDNSAIVATDSIKNTVYVKAKECT-EV
Prunus            24 VWRSENG--RHSIVEWNVGISLLT-DSVVAYTRDDNSDIVATDTMKNTVYAKAKECV-EE
Cajanus           25 VWKTKQG--RHFVVEWRVGITLLS-DCVNSYIRDDNSDIVATDTMKNTVYAKAKECS-EI
```

194

```
Populus         24 VWRNKSDH-IHSYVEWGVSIILLS-DCVNSYVRDDNSDIVATDTMKNTVYVKAKECS-EQ
Triticum        21 VWRRPAEAGGHLFVEWSVAVSVVS-DCLPSYTSDDNSAIVATDSIKNTVYVKAKECT-EV
Momordica       29 VWRNRDG--RHFIVEWSVGISILS-DCVSAYVRDDNSDIVATDTMKNTVYAKAKECS-EQ
Helianthus      29 VWRSGDN--RHNFVEWNVSISLLS-DCVNAYVSADNSDIVATDTMKNTVYVKAKECK-EQ
Vitis           25 VWRSQSG--RHIFVEWTVSISLLS-NCLAAYVRDDNSDIVATDSMKNTVYAKAKECA-QQ
Phaseolus       25 VWRTPQG--RHFVVEWRVGITLFS-DCVNSYLRDDNSDIVATDTMKNTVYAKAKECS-DT
Glycine         26 VWKTRQG--QHFIVEWRVGITLFS-DCVNSYLRDDNSDIVATDTMKNTVYAKAKECS-DI
Arabidopsis     24 VWRHDHDG-SHHFVEWNVSISLLS-HCLSSYYRDDNSDIVATDTMKNTVYVKAKECG-DR
Lotus           24 VWKNNQG--HHFVVEWRVSISLLS-DCLNSYLRDDNSDIVATDTMKNTVYAKAKECS-ET
Medicago        24 VWKSKDGK-QQFVVEWRVSINLLS-DCVNSYVRDDNSDIVATDTMKNTVYAKAKECS-EI
Cicer           24 VWKSKDGK-RHFVVEWRVSINLLS-DCVNSYVRDDNSDIVATDTMKNTVYAKAKECS-EI
Rattus          31 IQRDGKY---HSIKEVATSVQLTL-RSKKDYLHGDNSDIIPTDTIKNTVHVLAKFK--GI
Oryctolagus     27 IQRDGKY---HSIKEVATSVQLTL-SSKQDYVYGDNSDIIPTDTIKNTVHVLAKFK--GI
Mus             31 IQRDGKY---HSIKEVATSVQLTL-RSKKDYLHGDNSDIIPTDTIKNTVHVLAKLR--GI
Bubalus         31 IQRDGKY---HSIKEVATSVQLTL-NSRREYLHGDNSDIIPTDTIKNTVHVLAKFK--GI
Pteropus        31 IQRDGKY---HTIKEVATSVQLTL-SSKKDYLIGDNSDIIPTDTIKNTVHVLAKLK--GI
Myotis          31 IQRDGKF---HSIKEVATSVQLTL-SSKKDYLHGDNSDIIPTDTIKNTVHVLAKFK--GI
Gallus          26 LHRDGKK--HFVKEVEVCTHLRL-TSAHEYLDGNNSFVIPTDTIKNIVLVLAKKN--GI
Otolemur        30 IQRDGKY---HSIKEVATSVQLTL-SSKKDYLHGDNSDIIPTDTIKNTVHVLAKFK--GI
Microcebus      31 IQRDGKY---HSIKEVATSVQLTL-SSKKDYLHGDNSDIIPTDTIKNTVHVLAKFK--GI
Mus             32 IQRDGKY---HSIKEVATSVQLTL-RSKKDYLHGDNSDIIPTDTIKNTVHVLAKHR--GI
Callorhinus     31 IQRDGKY---HNIKEVATSVQLTL-SSKKDYVHGDNSDIIPTDTIKNTVHVLAKFK--GI
Canis           31 IQRDGKY---HSIKEVATSVQLTL-SSKKDYVYGDNSDIIPTDTIKNTVHVLAKFK--GI
Bombyx          60 VHRDEGH---HVIREFEVFTELKL-ASETAYILGDNKEVVATDSQKNTVYLAKKY--GV
Musca           42 VKRDGPV---HSIKEFEVGTHLKL-YSKKDYFHGDNSDIVATDSQKNTVYLAKKH--GI
Enhydra         31 IQRDGKY---HSIKEVATSVQLTL-SSKKDYVHGDNSDIIPTDTIKNTVHVLAKFK--GI
consensus       61 . ..        . .. . ... .        .. .... ............ ..       .


Singulisphaera  88 GSPEAELIRLGKHFTLSEEW--VSGCRWAAEQFTWQRIKVD------------------G
Saccharomonospo 74 GALEDFAITLADHFVAQTPA--ADGARLEIDEMPWQRIDVD------------------G
Nakamurella     76 RQPESFGIALAHFVDDTAP--VTRARLKIEMYPWNRLSHD------------------G
Streptomyces    74 ESAEQFGIHLARHFVTSQEP--LEVARLRIEEYAWERLATSDGNSKF----------IGA
Kitasatospora   74 ESAEQFAMLLARHFVENTERGVVHSARLRIEEYSWDRLAAASDRP------------GET
Thermobispora   74 RALEDFAITLGDHFLRQVPA--ATGARLAIEEYAWDRIDVD------------------G
Hoyosella       74 GTLEYGIALARHFVDDVAP--VNEARLEIDEHAWDRAIVD------------------G
Pseudomonas     72 ESLYEYGSRAGVWRLLKL----FKRRNVPLTVFAVAMAAQRNPE----------------
Truepera        90 ATLEAFLDSLGRAFFEAY--PDMETLRLTGREVPEAAAPVRQGE---------------TF
Microlunatus    78 GSPEELLIRLGDHFVSSFDW--IEGGLWQAEQYSWDRIILD------------------G
Arthrobacter    75 GSPEAFLIRLGEHFTSSFDW--VTGGRWEAESYAWERIQAH------------------G
Rhodococcus     79 LHLESYGIDLARHYVDDVAP--VDGARLELEEYAWTRAVVD------------------G
Actinobacteria  78 ASPEAFGILLAKHFLSSQSP--LREAQVRIEEYAWERIPVP------------------T
Bacillus        87 TTMEGFLAYVSACFLDTY--DHISEVEVSGRQFPEEAVEVGSPA--------------GV
Microbacterium  88 GSPEKFLIRLADHFTGSFEW--VDGGRWSAESYEWERIPVA------------------G
Aspergillus     75 TPPELFCSILGTHFLEKY--NHIHAAHVNTICHRWTRMDLDG-----------------
Trichoderma     78 NPPELFAAILGAHFLDTY--SHIHAANVKVITHRWTRMTLRG-----------------
Lodderomyces    77 WPLESFAAHLAKHFTSKY--EQVEGIEVSTVQALWKKIQLEG-----------------
Cordyceps       75 TPPELFAATVGAHFVDTY--AHIHVANRIVTTRWARMQVDLG-----------------
Aspergillus     75 HPPELFCSILDITCHTWERMTIEN--------------------------
Fusarium        75 NPPELYASILGSHFLEKY--KHIHVANVSVKTVRWARLDVDG-----------------
Conidiobolus    81 PNAEKFSVELAEFFINKYP-SHVQRVNVVVEQQSWDRQSFNG-----------------
Trametes        85 LHPERFAIHLGTHLVSKY--AHLKKAFLTIEQLRWQRIPIAQGEGQTP------------
Cyberlindnera   77 WPLERFAAKLATHFVEKY--SHVSGVSVKIVQDRWVKYAVDG-----------------
Hyphopichia     76 WPLERFAAHLAKHFSSKY--AHVMGVEVNIVQAKWSRFNVND-----------------
Ascoidea        76 FLVENFANILASHFLAKY--DHISSVDVDLIQLRWSKYLVND-----------------
Paracoccidioide 76 TPPELFASILGTHFVKTY--KHIHTAHVDIITHRWTRMTLDD-----------------
Blastomyces     76 TPPELFASILGHFTHTY--PHIHTAHVDITAIRWTRMTLDG-----------------
Penicillium     75 TPPELFCSILGTHFLEKY--NHIHAAHVTIVTHRWVRLDVDG-----------------
Pseudogymnoascu 75 TPPELYASILASHFVDTY--KHIHAAHVKVIVHRWTRMTLDG-----------------
Sorghum         79 VSLEEFAVILGRHFTSLY--PQVSEATVTIVERPWDRVTVDGKP---------------
Saccharum       79 VSLEEFAVILGRHFTSLY--PQVSEATVTIVERPWERVTVDGKP---------------
Prunus          80 LSVEDFAIRLAKHFTSLY--QQVTAALVKIVEKPWERVSIDGQP---------------
Cajanus         81 LSVEDFAIVLAKHFVSFY--KQVTGAIVNIVEKPWERVVVDGQP---------------
Populus         81 LSAENFAILLARHFTSFY--KQVTTAIVKIVEKPWERVHLNGQP---------------
Triticum        79 VSLEEFAVILGRHFTSLY--SQVSEATVTIVERPWERVAVDGKP---------------
Momordica       85 ISVEDFAILLAKHFTSYY--KQVTTAIVKIVEKSWERVSVNGQP---------------
Helianthus      85 VSLEEFAIILAKHFTSFY--PQVTTAIVKIVEKPWERLSVNGQP---------------
Vitis           81 LSVEEFAIKLAKHFTSFY--QQVTTAIVTVEEKPWERAYIDGQP---------------
Phaseolus       81 LSVEDFAILLAKHFVSFY--KKVTGAIVNIVEKPWERVIVDGQP---------------
Glycine         82 LSAEEFAILLAKHFVSFY--QKVTGAIVNIVEKPWERVTVDGQP---------------
Arabidopsis     81 LSVEEFAILIGKHFCSFY--PQVFTAIVSIIEKPWERVSIDGKP---------------
```

```
Lotus            80 LSVEDFAILLAKHFTSFY--KQVTTAIVKIVEKPWERVNVDGQP----------------
Medicago         81 LSVEDFAILLAKHFTSFY--SQVTTAIVKIVEKPWERVNVDGQP----------------
Cicer            81 LSAEDFAILLAKHFTSFY--RQVTTAIVNIVEKPWERVSVDGQP----------------
Rattus           85 KSTETFAVNICEHFLSSF--SHVTRAHVHVEEVPWKRFEKNCVK----------------
Oryctolagus      81 KSTETFAVNICEHFLSSF--NHVVRVHVHVEEVPWKRFEKNCVQ----------------
Mus              85 RNTETFAVNICEHFLSSF--NHVTRAHVYVEEVPWKRFEKNCIK----------------
Bubalus          85 KSTETFAVNICEHFLSSF--NHVIRVQVYVEEVPWKRFEKNCVK----------------
Pteropus         85 KSTETFSVNICEHFLSSF--NHVIRAQVYVEEVPWKRFEKKCAK----------------
Myotis           85 KSTETFAVNICEHFLSSF--NHVIRAQVYVEEVPWKRFEKNCAK----------------
Gallus           80 SSIEQFAIDICKHFMTTF--CQVAYVKTYIQEVPWQRQYQNCVP----------------
Otolemur         84 KSTEAFAVNICEHFLSSF--NHVIRAVHVEEVPWKRFEKNCVK----------------
Microcebus       85 KSTEAFAVNICEHFLSSF--NHVIRAHVHVEEVPWKRFEKNCVK----------------
Mus              86 RNTETFAVNICEHFLSSF--NHVTRAHVYVEEVPWKRFEKNCIK----------------
Callorhinus      85 KSTETFAVNICEHFLSSF--NHVIRTQVYVEEVPWKRFEKNCVK----------------
Canis            85 KSTEAFAVNICEHFLSSF--NHVIRAQVYVEEVPWKRFEKNCVK----------------
Bombyx          114 KTPEEFCAVVVNHFLYMY--KQVLEAKCRVIEYPWERLQAGTP----------------
Musca            96 ENPENFAILLARHFVQKY--AFVEEVHIHVEEYPWQRVSAHETGCQDGNGNCNYTNINNR
Enhydra          85 KSTETFAVNICEHFLSSF--NHVIRVQVYVEEVPWKRFEKNCVK----------------
consensus       121    ...   ...   ..   ....   ..    .   .  .   ..  ..  ...

Singulisphaera  128 KEHNHAFVQNRTETRTAVLLIDKDG--------PHLFAGLKDLIVLKSTGSEHGFPQDK
Saccharomonospo 114 RDHDHSFVQGGLGTRTTVVNVDGRGPDRT----AHVVSGIKDLTLLKSTGSEHRGFLKDR
Nakamurella     116 TPHPHAFARDGGYVRTATVTYDGTN--------LWVVSGIQDYVLKTTDSEFWGYLQDK
Streptomyces    122 DEVKHSFVRKGQETRTTQITFDGEK--------WEVISGIKDLIVNSTISEFWGYVKDR
Kitasatospora   122 ADTGHSFVRNGQEVRTTEVEYDDGV--------FRVVSGIKDLVVMNTTDSEFWGYLKDP
Thermobispora   114 TGHDHGFVRRGQGTRTTVVTVEGRGDERR----AWVLSGISDLIIAKTTGSEHGFLRDE
Hoyosella       114 AEHEHAWTRRGPDVRTAAVTVSGSCAAQR----IWVVGGIRDIVLLKTGSEFRDFLRDE
Pseudomonas     112 --VIRAMVADGHEICSHGYRWIDYQY-------MDEAQEREHVLEAIRILDELTGQRPVG
Truepera        134 VPSPVLLSRSRGDCGAAALLVTREG-------VRELESGRLGLQLLKTTGSSFASFARDA
Microlunatus    118 LEHDHSFVRKGQATRLATVQKVDGE--------THVTAGIKDLVLKSTGSEHRGFPRDR
Arthrobacter    115 SAHDHSFVRKGQEVRTAVIVRDGAA--------THLISGIKDLIVLKSTISGFVGAPKDK
Rhodococcus     119 VEHDHTWTRKGEEIRTSAVTVEGKGEAQR----TWVVSGFKEMVILKSTGSEHGFIEDD
Actinobacteria  118 RKEQHSFVRKGQEVRTAQITYSETTG-------LQVISGIKDLTIVMSTNSEFHGYIKDR
Bacillus        131 VPSETVFREGMLEKPGTRISVGRDENGEA--QIGSLESAVHDLHIKVKGSSFAGFIRDE
Microbacterium  128 AGHDHSFVRKGQETRTAVVLAEGAD--------RHVISGITGLIVLKTTSGFVGVPRDR
Aspergillus     115 KPHPHSFIRDSEEKRNVQVDVVEGKG-------TDIKSSLSGLIVLKSTNSQFWGFIRDE
Trichoderma     118 KPHPHSFLRDGQETRNVEARVSRKDG-------IAITSGIEGLIVLKSTGSAHGFVRDE
Lodderomyces    117 KEHDHSFKHEGPETRQTYINYEKASKK------LQLSSSIKDLKVLKSTGSMFYGNVCD
Cordyceps       115 RPHPHSFVKEGEIEKRTVTARVSRGG-------LAINSGIADLIVLKSTGSAHGFVRDE
Aspergillus     115 EPHPHSFYQDGSVKRQIRVDVSKASG-------INITSIIFGLIVLKSTRSEFWGYIKDE
Fusarium        115 KPHPHSEFKDGEETRNVEVRVSRQEG-------IEIKSSLVGLIVLKSTSLEHGFVRDE
Conidiobolus    122 AEHQHAFVGNNGEKHSSQVIATQETPQSP--IKYEITSKIDGLIVLKTTGSSFKHFYADE
Trametes        131 KPHSHAFYRDGDEKRTVEAEIDASAGKDK--IVASVTSGLKDLIVLKTTGSAFESFVRDE
Cyberlindnera   117 KPHDHSFTHEGGEKRITDL-YYKRSGD------YKLSSAIKDLIVLKSTGSMFYGNQCD
Hyphopichia     116 APHPHSFKHEGPETRRVFIDYCKRKDK------LVVKSSIKDLIVLKSTGSMFYGNVCD
Ascoidea        116 KFHDHSFIKDGDEKRICHL-FRSRVGD------FKLTSGIKDLILLKSTGSMFYGNKCD
Paracoccidioide 116 EPHQHSFLRDGTETRNVSATVTKRTG-------ISLTSAIAGLSVLKSTNSQFHGFIRDE
Blastomyces     116 KPHPHSEFRDGTDTRNVSATVTDGTTTKHGTPTIKLTSSIAALSVLKSTNSQFHNFIRDE
Penicillium     115 KPHPHSFIKPGSETRNVQVDVIEGKG-------IDINSSINGLIVLKSTGSQFWGFVRDE
Pseudogymnoascu 115 KPHPHSFFRDGNETRNVEVTAREGKG-------IEVRSSISGLIVLKSTGSQFWGFVRDE
Sorghum         121 --HSHGFKVG-VEKHSTEVIVKKSGS-------LLINSGIQGYSLLKTTQSGFEGFVTDR
Saccharum       121 --HSHGFKVG-VEKHSTEVIVKKSGS-------LLINSGIQGYSLLKTTQSGFEGFVTDR
Prunus          122 --HEHGFKLG-SEKHTEVILKKSGA-------LKVTSGIEGTALLKTTKSGFEGFIRDK
Cajanus         123 --HEHGFKLG-SEKHTAEVVQKSGA-------LQLTSGIEGLSVLKTTKSGFEGFIRDK
Populus         123 --HEHGFKLG-SEKHTAEVIVQKSGV-------LKLTSGIEGLSVLKTTMSGFEGFIRDQ
Triticum        121 --HSHGFKLG-SEKHTEVNVKKSGS-------LLINSGIQGYSLLKTTQSGFEGFVRDR
Momordica       127 --HNHGFKLG-SEKHTEVIVKKSGA-------LQVTSGIEGLSLLKTTQSGFERFIRDN
Helianthus      127 --HDHGFKLG-SEKHTEVIVQKNGS-------LGVTSGVVGLSLLKTTQSGFEGFIRDQ
Vitis           123 --HDHGFKRG-SETHTEVIVEKSGA-------LQVTSGIQGLALLKTTQAGFEGFIRDK
Phaseolus       123 --HQHGFTLG-SEKHTEAIVQKSGS-------LQLTSGIEGLSVLKTTQSGFENFIRNK
Glycine         124 --HEHGFKLG-SEKHTEAIVQKSGS-------LQLTSGIEGLSVLKTTQSGFVNFIRDK
Arabidopsis     123 --HLHGFKLG-SENHTAEARVEKSGA-------LNLTSGIGGLALLKTTQSGFERFIRDK
Lotus           122 --HEHGFKLG-SEKHTEAIVQKSGA-------LQLTSGIEGLSLLKTTKSGFEGFIRDK
Medicago        123 --HDHGFKLG-SEKHTEAIVKKSGA-------LQLTSGIEGLSVLKTTQSGFEGFVRDK
Cicer           123 --HEHGFKLG-SEKHTEAIVKKSGA-------LQLTSGIEGLSLLKTTKSGFEGFIRDK
Rattus          127 --HVHAFTHTPTGTHFCEVEQVRNGP-------PIIHSGIKDLKVLKTTQSGFEGFIKDQ
Oryctolagus     123 --HVHAFTHTPTGTHFCEVEQRRSGL-------PVIHSGIKDLKVLKTTQSGFEGFIKDQ
Mus             127 --HVHAFTHTPTGTHFCEVEQMRNGP-------PVIHSGIKDLKVLKTTQSGFEGFIKDQ
Bubalus         127 --HVHAFTHTPTGTHFCEVEQLRSGP-------PVIHSGIKDLKVLKTTQSGFEGFIKDQ
Pteropus        127 --HVHAFTHTPTGTHFCEVEQKRSGF-------PVIHSGIKDLKLLKTTQSGFEGFIKDQ
```

196

```
Myotis          127 --HVHAFIHTPIGTHFCEVEQMRSGP-------PVIHSGIKDLKVLKTTQSGFEGFIKDQ
Gallus          122 --HIHSFILVPDGIRFCEAEQCRNGP-------LVVCAGIKDLKLMKTTQSGFEGFYRNE
Otolemur        126 --HVHAFIHTSIGTHFCDVEQMRNGP-------PVIHSGIKDLKVLKTTQSGFEGFIKDQ
Microcebus      127 --HVHAFIHTPIGTHFCEVEQMRNGP-------PVIHSGIKDLKVLKTTQSGFEGFIKDQ
Mus             128 --HVHAFIHTPIGTHFCEVEQMRNGP-------PVIHSGIKDLKVLKTTQSGFEGFIKDQ
Callorhinus     127 --HVHAFIHTPIGTHFCEVEQMRGGP-------PVIHSGIKDLKVLKTTQSGFEGFIKDQ
Canis           127 --HVHAFIHNPIGTHFCEVEQMRSGP-------PVIHSGIKDLKVLKTTQSGFEGFIKDQ
Bombyx          155 --HSHAFVFSPIATRWSEVSQTRHEA-------VVVKSGLSGLRVLKTTQSAFVDFVQDE
Musca           154 ARHNHAFIFTPLEIRYCLVVLRRTDPK------QTVISGIRGLRVLKTTQSSFVNFVNDE
Enhydra         127 --HVHAFIHTPIGTHFCEIEQMRGGP-------PVIHSGIKDLKVLKTTQSGFEGFIKDQ
consensus       181    . .....  .. .. . . .         .  . ... ............ .....


Singulisphaera  180 YTTLVETSDRIILSTDVAARWRYNT------------------------AELIDENAVYAD
Saccharomonospo 170 YTTLEETDDRILATSLVARWRY--------------------------EGTDVGWDETFES
Nakamurella     168 YTTLKPTHDRVMATSVTGQWWH--------------------------TDSEADWAKSYDS
Streptomyces    174 YTTLKEAYDRILATDVSARWRY--NWTGD-------------------EQRMPNWEKSYEQ
Kitasatospora   174 YTTLPEAYDRILATQVTARWAFGFAGRDA------------------EDAAPDWNHSYRE
Thermobispora   170 YTTLEETHDRILATSIHTRWRY--------------------------LTTDVDWDKTFAS
Hoyosella       170 HTVLEPTTDRVMATSLVAQWRF--------------------------ADINCDWDEIFAS
Pseudomonas     163 WYTGRTGPNTRRLVMEECGELYDSDTYDD------------------DLPYWDPASTA
Truepera        187 HTTLPEMHDRPLLIYLDAFWRYRDPAAAL----------------QTGTGELTGYVASEQ
Microlunatus    170 YTTLIETDDRILATSVTGRWRYLPE---A------------------VEAGLDENALYAG
Arthrobacter    167 YTTLPETTDRILATDVSARWRFRSG---T------------------DFSSLDFNKSYED
Rhodococcus     175 ITILEPTTDRVMATSLTAQWRY-------------------------TTTDVEWDAVYAG
Actinobacteria  171 YTTLKEAYDRILATKVTARWAH--SALAG------------------DDDAFDWDQSYKK
Bacillus        189 YTTLKDRPLFIYLNISWTYKETEHAL---------------GATP---EQYVAAEQ
Microbacterium  180 YTSLQETTDRILATDVTARWRYAQG---V------------------VLDDLDFTAVYDD
Aspergillus     168 YTTLKETWDRILSTDVDATWQWKNFSG----------------LQEVRSHVPKFDATWAT
Trichoderma     171 FTTLPETWDRILSTDVDASWKWKPFAN----------------VQAVREAVSKFNPAWEE
Lodderomyces    171 YTTLKPTKDRILSTDVEATWTFDSNKIGT-----------LDDIIKSG--SIFDNAYNQ
Cordyceps       168 FTTLPETWDRIMSTDVDASWDWKVFPD----------------LAAVRAAAPRFDAAHEA
Aspergillus     168 YTTLPETWDRILSSDVDIKWRWKHFGD----------------VSDVEANVSHFIEAFAS
Fusarium        168 YTTLPETWDRILSTDVTAWKWKKFDS----------------VDAVKQFAPKFDTVREA
Conidiobolus    180 YRTLPDAEDRILSTVVTLEWNYSTFNN-----------------SYSVYETPFLEIRAA
Trametes        189 YTTIMEVNDRIFSTSLDLSYKFSPLRIPPPADGEGKVFDLPLTAEQAKGTAWDGLSVATV
Cyberlindnera   170 FTTIQPTTDRILSTDVDATWVWDNKKIGS-----------VYDIAKAADKGIFDNVYNQ
Hyphopichia     170 YTTLKPTKDRILSTDVDASWTYSSKALSN-----------FEDVFKQADAGLFDKTYEN
Ascoidea        169 YTTLKPTNDRILSTDVDASWVWSNQITS-------------SNISDLSNNGLFDKTYAS
Paracoccidioide 169 FTTLPETWDRILSTDVDAKWTWTPFAS----------------LEDVKAIVPKFDETWEV
Blastomyces     176 YTTLPETWDRILSTDVDASWTWQPFAS----------------VDAVKRAAPQFDETWRA
Penicillium     168 YTTLKETWDRILSTDVAANWQWRRFTG----------------LTEVKTNFEKENAAWEE
Pseudogymnoascu 168 YTTLPEVWDRILSTEIDSTWKWNNFES----------------LRSVRASVPKFDKAWAA
Sorghum         171 YRLLPDTRERIVATEVTAWWRYPFEHVSQ----------------LPSKPFCETQRYQD
Saccharum       171 YRLLPDTRERIVATEVTAWWRYPFEHVSQ----------------LPSKPFCETQRYQD
Prunus          172 YTVLPDTRERILATDLTASWTYPYESIYS----------------IPQKPIYFTERYLS
Cajanus         173 YTALPETRERMLATEVTALWRYSYESLYS----------------LPQKPLYFTDKYLE
Populus         173 YTALPETRERMLATEVTALWRYSYESASS----------------IPKNPLYFTERYLD
Triticum        171 YTLLPETRERIVATEVTAWWRYPFEHISQ----------------LPSKPFCETQRYQD
Momordica       177 YTVLPETRERILATEVSASWRYSFEALYS----------------IPKKQFYFTETYLD
Helianthus      177 NTILPETRERMLATEVSASWRYQFKSLSS----------------ISNKPLQFTEKYLR
Vitis           173 YTALADTRERIVATEVTASWRYPFESLSG----------------IPLQPLYFTEKYLD
Phaseolus       173 YTALPDTRERILATEVTALWRYSYESLYN----------------LPQKPLYFTDKYLE
Glycine         174 YTALPDTRERMVATEVTALWRYSYESLYS----------------LPQKPLYFTEKYQE
Arabidopsis     173 YTVLPETGERMLATEVNASWRYSYESVAS----------------IPTKGLYFTEKEMD
Lotus           172 YTVLPETRERMLATEVTALWRYSYESLYS----------------IPQKPLYFTEKYLD
Medicago        173 YTILPDTRERMLATEVTALWRYSYESLYS----------------VPKKPLYFTEKYLD
Cicer           173 FTALPETRERMLATEVTALWRYTYESFYS----------------IPQKPLYFTDKYLD
Rattus          178 FTTLPEVKDRCFATQVYCKWRYQN----------------------RDVDFEATWGA
Oryctolagus     174 FTTLPEVKDRCFATQVYCKWRYQHS---------------------QDVDFEATWDI
Mus             178 FTTLPEVKDRCFATQVYCKWRYQR----------------------RDVDFEAIWGA
Bubalus         178 FTTLPEVKDRCFATQVYCKWRYHQG---------------------EGTDVDFEATWEA
Pteropus        178 FTTLPEVKDRCLATQVYCKWRYHWC---------------------RDVDFEATWDT
Myotis          178 FTTLPEVKDRCFATQVYCKWRY-RG---------------------SDVDFEATWDT
Gallus          173 HTTLPERNDRILCGEFFCKWSYGEC---------------------RDFDFDCIWSK
Otolemur        177 FTTLPEVKDRCFATQVYCRWRYHCG---------------------RSVDFEATWDT
Microcebus      178 FTTLPEVKDRCFATQVYCKWRYHQG---------------------RNVDFEATWDT
Mus             179 FTTLPEVKDRCFATQVYCKWRYQR----------------------RDVDFEAIWGA
Callorhinus     178 FTTLPEVKDRCFATQVYCKWRYHRC---------------------RDVDFEAICDT
Canis           178 FTTLPEVKDRCFATKVYCKWRYHQG---------------------RDVDFEATWDT
Bombyx          206 YTTLTDAVERIFSTIVEAEWIYDNM---------------------RTADFDNAWLT
```

197

```
Musca          208 FRSLPLQYDRIFSTVVDCSWEYSNT----------------------DKVLECKSWNI
Enhydra        178 HTTLPEVKDRCFATQVYCKWRYHQY----------------------RDVDFDATWDT
consensus      241 ................ . ...                         . ... .


Singulisphaera 215 VRRIILLTFASVH-----SLALQQTLFQMCKNVLQAHRNIVDIRFSM--PNKHHFVVDLA
Saccharomonospo 205 IRSLLILFQFAEIH-----SRALQQTLYGVCHAVLEKHPEVAEIKFSA--PNKHHFLVDLS
Nakamurella    203 AVATVSSVFAGHH-----SLALQQTMFAMGEAMIAEQPLIGEVRFSL--PNKHHFVIDLS
Streptomyces   214 ARKHVLQAFAETY-----SLSLQQTLYQMGSRIINSRSLIDEIRFSL--PNNHHFLVDLE
Kitasatospora  216 VRRHILEAFAQTY-----SYSLQQTLHAVCTRVLDHRAEVDEVRLEL--PNKHHFLVDLE
Thermobispora  205 VRSILLRQFATVH-----SLALQQTLYAMGSAVLEAHPLIAEIRLSA--PNKHHFLVDLQ
Hoyosella      205 VRRTIVETFATHH-----SLALQQTLYIGKNVLEAHRLIVEWRLSA--PNKHHELYDLE
Pseudomonas    203 EKPHLVIPYTLDTN-----DMRFTQVQGHNCEQFFQYLKLAFDVLYEEGATAPKMLSIGLH
Truepera       231 VRDLIAATFHDVN-----SRSTQHLVYEVCVRLLARFPQLREVRFEA--QNRLWDHAFSD
Microlunatus   209 VSEVFLATFASVH-----SLALQQTQWEVCKAAIEAFPLIAEVKFAM--PNKHHFLVDLA
Arthrobacter   206 VKSLLILDGETEKY-----SHALQQTLFDVCAKVLEAHSLIEEIKFSM--PNKHHFLVDLS
Rhodococcus    210 VKAAMIERFANLQ-----SLALQQTLYAMCEAVLEKYPFIAEISMSA--PNKHHFVYDLG
Actinobacteria 211 VRKNVILAFAETY-----SYSLQQTLNQMAERVLDNCPRVNEVRLNL--PNKHHFLVDLE
Bacillus       230 VKHIAQTLFHQED-----SPSIQKLLYDICIRVLERFPQLETVSFES--NNRTWETVRDE
Microbacterium 219 VKRLILLEFTRRY-----SAALQTTLFDVCRVVLEAYPLIAEIRMSM--PNKHHFVVDLA
Aspergillus    212 AREVTLKTFAEDN-----SASVQATLYKMAEQILARQQLVETVEYSL--PNKHYFEIDLS
Trichoderma    215 AREITLTRFAEDE-----SASVQNTMYKVCEQILAAVPLTELVIYSL--PNKHHFELDLS
Lodderomyces   217 AKDVTLDLFCKEN-----SPSVQATVYNMSQKILDNVSLIATVRYVL--PNIHYILFNFE
Cordyceps      212 ARALTFKRFADDN-----SASVQATLYKMSEDILAAVPEVNKVYYSL--PNKHYFELDLS
Aspergillus    212 ARNASFKAFAEDN-----SMSAQATLYKMAEKFLDCVPLADAVEYSW--PNKHYVEIDLS
Fusarium       212 ARNITLKTFAEDA-----SASVQATMYKMSDLILESVPEVATVTYSL--PNKHFEIDLS
Conidiobolus   222 AKHFALDEFANKP-----SPSVQATLYDTANRIINKYHPVDKVSISL--PNRHFEAADLP
Trametes       249 ARNITLEIFALDE-----SASVQATLYKMAQRIVSEHPHVQSVSYNL--PNKHYVPVDMK
Cyberlindnera  218 AREITLTTFALEN-----SPSVQATMFNVATQILEKACSVYSVSYAL--PNKHYFLIDLK
Hyphopichia    218 ARKVTLDIFALEN-----SASVQATVYNMSHKILELVPEVENVTYIL--PNKHYILFNFE
Ascoidea       215 ARKITLDLFATQN-----SPSVQATMYDVASEILKTSSKIKSVTYTL--PNKHYFLMDFS
Paracoccidioide 213 ARDTTLRVFAQDN-----SASVQNSMYKMAEQILDAQPLLSCVEYSL--PNKHYFEIDLS
Blastomyces    220 VRDITVRTFAHDN-----SASVQNSMYKMAEQILAVQPLLETVEYAL--PNKHYFEIDLS
Penicillium    212 ARAITLKTFAEDN-----SASVQNTMYKVCEHILKAVPDVETVEYSL--PNKHYFEIDLS
Pseudogymnoascu 212 AREITVKLFAEEN-----SPSVQNTMYKVCEHILKAVPDVETVEYSL--PNKHYFEIDLS
Sorghum        214 VKRVITDTFFGPADVGVYSPFVQNTLYLMAKEVLTRFPDISSVQLRM--PNLHFLPVNLG
Saccharum      214 VKKVLAETFFGPADVGVYSPSVQNTLYLVCQEVLTRVAFISSIXLKM--PNLHFLAVNLG
Prunus         215 VKKVLATFYGPPKEGVYSPSVQSTLYHMATNVLKGFPLIATVQLKM--PNIHFLPVNLS
Cajanus        216 VKKVLATFFGPPNGGVYSPSVQNTLYLMAKATLNRFPLIAFVHLKM--PNLHFIPVNIS
Populus        216 VKKSLANTFFGPPKEGVYSASVQRTLFQMAKAVLNRFPLISSTQLKM--PNIHFLPVNIS
Triticum       214 VKKVLADTFFGPSDVGVYSPSVQNTLYLMAREVLTRFPLIASVQLRM--PNLHFLPVNLG
Momordica      220 VKKVLVDTFFGPPKEGVYSPSVQYTLYDMAKSVLSRFPVISLVKLKM--PNLHFLPVNLS
Helianthus     220 VKKVLMDTFFGPPKEGVYSPSVQATLYDMAKAVLGRFPDISSVHLKM--PNIHFLPVNLS
Vitis          216 VKKVLAETFFGPPRGGVYSPSVQSTLYQMAKNVLNRFPDISSIQLKM--PNLHFLPVNIS
Phaseolus      216 VKKVLATFFGPPNRCVYSPSVQNTLYLMAKATLNRFPDIAYVHLKM--PNLHFLPVNIS
Glycine        217 VKKVLATFFGPPKGGVYSPSVQNTLYLMAKATLNRFPDIAYVSLKL--PNLHFIPVNIS
Arabidopsis    216 VKKVLVDTFFGPPETGVYSPSVQRTLYLMCSAVLKRFADVSSIHLKM--PNIHFLPVNLS
Lotus          215 VKKVIVDTFFGHPKEGVYSPSVQSVSTLYQMAKATLNRFPDVASIQLKM--PNIHFIPVNLS
Medicago       216 VKRVLDTFFGSPKEGVYSPSVQATLYQMAKAALNRFPLIASIQLKM--PNIHFIPVNLS
Cicer          216 VKRVLVDTFFGSPKEGVYSPSVQSTLYQMCKATLNRFPLIASIQLKM--PNLHFLPVNLS
Rattus         213 VRDIVLKKFAGPYDRGEYSPSVQKTLYDIQVLTLSQLPLIEDVEISL--PNIHFNIDMS
Oryctolagus    210 VRDTVLBKFAGPYDKGEYSPSVQKTLYDIQVLTLSRVPQIEDVEISL--PNIHFNIDMS
Mus            213 VRDIVLQKFAGPYDKGEYSPSVQKTLYDIQVLSLSQLPLIEDVEISL--PNIHFNIDMS
Bubalus        214 VRSIVLKKFAGPYDKGEYSPSVQKTLYDIQVLSLSQLPLIEDVEISL--PNKHFNIDMS
Pteropus       214 IRDIVLBKFAGPYDKGEYSPSVQKTLYDIQVLSLSQVPLIEEVEISL--PNIHFNIDMS
Myotis         213 VRDIVLBKFAGPYDKGEYSPSVQKTLYDIQALSLSRVPLIEDVEISL--PNIHFNIDMS
Gallus         209 VRECILEAFSGPPDCGEYSPSYQRTVNCLQMCVLSRVPQVQVIEVIL--NNNFYNVVDMK
Otolemur       213 VRDIVLBKFAGPYDKGEYSPSVQKTLYDIQVLSLSHVPLIEDVEISL--PNIHFNIDMT
Microcebus     214 VRDIVLBKFAGPYDKGEYSPSVQKTLYDIQVLSLSRVPLIEDVEISL--PNIHFNIDMS
Mus            214 VRDIVLQKFAGPYDKGEYSPSVQKTLYDIQVLSLSQLPLIEDVEISL--PNIHFNIDMS
Callorhinus    214 VRDIVLBKFAGPYDKGEYSPSVQKTLYDIQVHSLSRVPLIEDVEISL--PNMHFNIDMS
Canis          214 VRDIVLBKFAGPYDKGEYSPSVQKTLYDIQVHSLSRVPEMEDVEISL--PNIHFNIDMS
Bombyx         242 VKDAILBKFAGPPDTGVYSPSVQHTLYLMQAEKTVLEKVSDITWIKMTM--PNKHYLNIDMS
Musca          244 VKNIILRKFAGDPNVGTSSPSVQHTLYLTEKEVLDTLPEVSVISMTM--PNKHYFNFDTK
Enhydra        214 VRDIVLBKFAGPYDKGEYSPSVQKTLYDHQVHSLSQVPLIEDVEISL--PNMHFNIDMS
consensus      301 .. .... ...       . ....* ... .. .. ..... .... .. ... ....


Singulisphaera 268 HFGQDNP-----NVVEWAADRPFGLIECTVK----RDD--AADDLWESVAGFC-------
Saccharomonospo 258 PFGVDNP-----GEVEYAADRPYGLIEASVV----RDDASERGSAWHAVPAWV-------
Nakamurella    256 PYGLENP-----NEVEHADDRPYGFIECTHNDLHKDAAPAPNQAFDPGQGW--------
```

198

```
Streptomyces      267 PFGLENDTA--DGAVYFAADRPYGLIEATVL----RDGVEPKIPVDMTNL----------
Kitasatospora     269 PFGLKNE-----NEVFYAADRMYGLIEGTVH----REGVTPVIPVS-------------
Thermobispora     258 PFGLDNP-----GEVEYASDRPYGLIEASVV----RDDVPEAPEAWLATPGFC-------
Hoyosella         258 RFGIENR-----NEVPHADDRPYGLIQATVL----RDDAADAGPAWMQQLGWLT------
Pseudomonas       260 CRLIGRPARMAALERFIQYAQSHDKVWFARREDIARHWHREHPFQETEA----------
Truepera          284 DETG-------RKVHTDPRPPYGRIGLTTA-----------------------------
Microlunatus      262 PYGLENP-----NEVFYAADRPYGLIEGTIV----RDGVAAAPQAWTDVPTFV-------
Arthrobacter      259 PFGLDNP-----NEVFFAADRPYGLIEATVL----RDDAEAADAAWSGIAGFC-------
Rhodococcus       263 KFGIENN-----LEVFNADDRPYGLIQATVE----REDAPDAGSAWRTYSVVG-------
Actinobacteria    264 PFGLKND-----NEVFFAADRMYGLIEGTVH----RDGVQPVTATSDWIVA---------
Bacillus          283 IPASGE------GKVFTDPRPPFGFQKFTVLQEDVKREEALE-----------------
Microbacterium    272 PFELDNP-----NEVFFASDRSYGLIEAAVT----REGQADAPQAWEAATAFC-------
Aspergillus       265 WHKGLQNTG-KNAEVFAPQSDPNGLIKCTVGRSS----LKSKL----------------
Trichoderma       268 WHKGIKNTG-KDAEVVVPQSNPNGLIKCEVARGSPP-TRGSHL----------------
Lodderomyces      270 WKGIKNND-----DLFYESSDPNGLITCTVGRKG----DKAKL----------------
Cordyceps         265 WHKGIKNTG-KDAEVVVPQSNPNGLIKVEVSRDG----HSSKL----------------
Aspergillus       265 WHKGIRNTD-KDAEVYLPQSAPNGLIKCTLTRSTLDRSRSAKL----------------
Fusarium          265 WHKGIKNTG-KDAEVVAPQSGPNGLIKCEVSRDS----LQSKL----------------
Conidiobolus      275 IFKQLGQDK-KVHDLYVPIAHPNGLIKATVKRN-----VAPKL----------------
Trametes          302 Y-IGIDNMSPPKAEVFIPLSAPSGLISATVSRK--------------------------
Cyberlindnera     271 WKGLENDN-----ELFYPSPHPNGLIKCTVVRK-----EKTKL----------------
Hyphopichia       271 WKGIKDNS-----ELFYPSPDPNGLIKSTVGRK-----ESAKL----------------
Ascoidea          268 WFNNLKND-----EVFYPSPHPNGLINCTVTRD-----PKSKL----------------
Paracoccidioide   266 WYKGLKNTG-KDAEVVVPQSGPNGLIKCTVARREGERRETPKL----------------
Blastomyces       273 WHKGLKNTG-KDAEVVAPQSGPNGLIKCTVGRPGVDGRETPKL----------------
Penicillium       265 WHNGLKNTG-KDAEVVAPQSNPNGLIKCTVARAG----QMAKL----------------
Pseudogymnoascu   265 FHKGIKNTG-ADTTVUAEQSGPNGLIKCTVGRKDK--LEKGKL----------------
Sorghum           272 -SKETP-LVKFADDVYLPTDEPHGTIEATISRPMS------KL----------------
Saccharum         272 -SKETP-LVKIADDVVPATDEPHRTIGAPLTRPMS------KL----------------
Prunus            273 -NKDN-TIVKFEDDVYLPTDEPHGSIEATISRFWS------KM----------------
Cajanus           274 -NKDG-PIVKFDDDVYIPTDEPHGSIKASISRLWSN----SKL----------------
Populus           274 -SKEN-TIVKFNDDVFLPTDEPHGSIEASISRFWS------KM----------------
Triticum          272 -GKENPGLVKFADDVYMPTDEPHGTIEATLSRANS------KL----------------
Momordica         278 -TKDNRSIVKFEDDVYLPTDEPHGSIEACISRFSS------KL----------------
Helianthus        278 -SKVNPVIVKFEDDVYLPTDEPHGSIEASISRPLS------KM----------------
Vitis             274 -SKDNPAIVKFDDDVYLPTSEPHGSIEATVSRIRA------KI----------------
Phaseolus         274 -SKDG-PIVKFEDDVYLPTDEPHGSIEASISRVWS------KL----------------
Glycine           275 -NQDG-PIVKFEDDVYLPTDEPHGSIQASISRLWS------KL----------------
Arabidopsis       274 -TKENPSMVKFKDDVYLPTDEPHGSIEATVSRITS------KL----------------
Lotus             273 -NKDG-PIVKFDDDVYLPTDEPHGSIEASISRILS------KM----------------
Medicago          274 -NKNG-QFVKFDDDVYLPTDEPHGSIEASLCSRS-------KL----------------
Cicer             274 -NKDG-HIVKFDDDVYIPTDEPHGSIEASISRTRS------KM----------------
Rattus            271 KMGLINK-----EEVLILPLDNPYGKITCTVRRK-----LPSRL---------------
Oryctolagus       268 KMGLINK-----EEVLILPLDNPYGKITCTVKRK-----LSSRL---------------
Mus               271 KMGLINK-----EEVLILPLDNPYGKITCTVKRK-----LPSRL---------------
Bubalus           272 KMGLINK-----EEVLILPTDNPYGRITCTVKRK-----LTSRL---------------
Pteropus          272 KMGLINK-----EEVLILPTDNPYGRITCTVKRK-----LLSKL---------------
Myotis            271 KMGLVNK-----DEVLILPSDNPYGRITCTVKRK-----LASRL---------------
Gallus            267 ALGCTND-----KEVLVPVETPYGSCACTLGRKKYL-EAQSHMIKDEKQSQFGLVAAQGK
Otolemur          271 KMGLINK-----EEVLILPLDNPYGRITCTVKRK-----LSSRL---------------
Microcebus        272 KMGLINK-----EEVLILPLDNPYGRITCTVKRK-----LSSRL---------------
Mus               272 KMGLINK-----EEVLILPLDNPYGKITCTVKRK-----LPSRL---------------
Callorhinus       272 KMGLINK-----EEVLILPLDNPYGKITCTVKRK-----LASKL---------------
Canis             272 KMGLINK-----EEVLILPLDNPYGRITCTVKRK-----LASKL---------------
Bombyx            300 KFPVNVTKGDPRHFLYHPIDQPAGLIYAQIRRR----PKSKL----------------
Musca             302 PFQAVVP-GE-NNEVTIPVDKPHGTIYAQIARK----DIASHL---------------
Enhydra           272 KMGLINK-----EEVLILPLDNPYGRITCTVKKK-----LASKL---------------
consensus         361             ..... . . . . ... .       ..
```

Multiple sequence alignment of all selected uricase protein sequences showing maximum conservation in between 51 and 314 amino acids.

Amino acid composition of uricase protein sequence in different microbial sources

	Ala(A) %	Arg(R) %	Asn(N) %	Asp(D) %	Cys(C) %	Gln(Q) %	Glu(E) %	Gly(G) %	His(H) %	Ile(I)%	Leu(L) %	Lys(K) %	Met(M) %	Phe(F) %	Pro(P) %	Ser(S)%	Thr(T) %	Trp(W) %	Tyr(Y) %	Val(V) %	Total
AGA28823.1	8.7	6.1	4.5	8.1	1.0	3.9	4.9	7.1	4.5	3.9	8.1	4.5	1.0	6.1	2.9	5.2	8.4	1.9	1.3	7.8	99.9
EHR61468.1	9.3	7.0	2.0	9.0	0.0	3.7	6.3	8.3	4.3	4.3	8.0	3.3	1.0	4.7	3.7	6.0	6.6	1.7	2.3	8.6	100.1
ACV76680.1	9.6	4.3	4.3	7.9	0.3	3.3	5.6	6.6	5.3	3.0	5.3	3.6	2.3	4.3	5.0	6.3	8.3	2.3	4.3	7.9	99.8
ADW02506.1	7.1	6.5	5.2	7.1	0.0	3.9	8.1	6.1	3.2	6.1	6.5	4.5	2.3	3.9	2.9	6.5	7.7	1.9	4.2	6.5	100.2
KDN81792.1	9.2	8.9	4.3	6.6	0.3	2.3	8.5	6.9	5.2	2.6	7.9	2.6	1.6	3.9	3.0	4.9	6.6	1.3	4.3	9.2	100.1
ADG89294.1	9.6	8.3	1.7	8.0	0.3	3.0	6.3	7.3	4.3	4.7	9.0	2.7	0.7	4.3	3.7	4.0	9.0	2.0	3.0	8.3	90.6
AEF42631.1	10.3	8.6	3.0	7.9	0.3	3.6	7.3	6.6	4.3	5.3	7.6	2.0	1.3	4.0	3.0	3.6	8.3	2.3	2.3	8.3	89.6
EYU05833.1	8.8	8.4	2.9	6.2	1.0	4.2	8.4	7.1	4.2	3.9	8.1	2.3	3.9	4.2	5.5	3.9	4.2	2.3	5.2	5.2	91.1
ADI14624.1	12.1	9.1	1.6	5.9	0.3	2.6	6.2	8.1	2.9	2.0	11.1	1.6	1.6	5.9	5.9	4.6	7.8	70.0	3.3	6.8	157.3
BAK35228.1	10.2	5.9	2.6	6.9	0.3	3.3	7.2	7.5	3.0	3.6	9.8	3.6	1.0	4.6	3.3	5.2	7.5	2.0	3.3	9.2	89.8
GAB16350.1	9.6	5.6	3.0	7.6	0.3	2.6	7.6	7.3	4.3	3.6	8.6	5.0	1.0	6.0	2.3	7.3	7.0	1.7	2.6	7.0	90.4
AMY51449.1	11.1	5.6	2.6	7.2	0.0	3.3	7.8	7.2	3.3	4.2	7.8	4.2	2.0	2.9	2.6	5.2	7.8	2.0	4.6	8.5	88.8
KPI32983.1	8.5	6.6	5.2	6.2	0.7	4.3	6.9	5.6	3.6	4.9	7.5	5.9	1.6	3.6	3.0	5.6	6.9	1.6	3.9	7.9	91.5
ADH98118.1	6.3	5.3	2.5	4.7	0.3	3.8	10.4	7.9	2.5	5.3	6.6	4.4	2.2	6.6	4.7	6.9	7.9	0.6	3.5	7.5	93.6

	Ala(A) %	Arg(R) %	Asn(N) %	Asp(D) %	Cys(C) %	Gln(Q) %	Glu(E) %	Gly(G) %	His(H) %	Ile(I)%	Leu(L) %	Lys(K) %	Met(M) %	Phe(F) %	Pro(P) %	Ser(S)%	Thr(T) %	Trp(W) %	Tyr(Y) %	Val(V) %	Total
KJQ52767.1	11.4	7.9	2.2	7.9	0.3	3.2	7.3	6.7	2.9	3.2	7.9	2.5	1.3	5.1	3.2	4.8	8.9	1.6	2.9	8.9	88.7
KJK61270.1	6.3	4.0	4.6	5.6	1.0	4.0	6.6	5.0	3.6	6.0	6.6	8.3	1.7	3.3	3.0	7.6	8.6	2.3	3.3	8.6	93.7
XP_006963697.1	9.1	5.5	4.5	5.8	1.0	1.9	6.2	5.5	3.6	5.8	5.8	5.8	1.6	3.9	4.5	7.5	8.8	1.9	2.3	8.8	90.7
XP_001528662.1	5.3	2.0	6.6	6.3	1.0	3.0	6.3	4.3	2.3	6.6	7.9	11.2	1.3	4.0	3.0	8.6	7.3	1.3	4.3	7.6	94.9
XP_018702053.1	10.9	6.3	3.6	7.0	0.3	1.3	5.0	5.0	4.3	4.3	5.3	6.0	2.0	4.3	4.6	7.0	7.9	1.7	2.6	10.6	89.1
XP_001390131.1	7.2	4.2	4.2	7.2	1.3	2.3	5.9	4.6	4.6	6.5	6.9	6.9	1.3	3.6	3.3	8.5	7.5	2.3	4.6	7.2	92.9
XP_011321510.1	7.0	4.3	3.6	6.6	1.3	2.3	7.0	4.6	2.6	5.0	5.6	7.9	1.3	3.6	4.0	8.9	8.3	1.7	3.6	10.6	92.8
KXN72467.1	8.4	3.9	6.4	5.1	0.0	5.8	5.5	3.5	3.5	4.8	6.8	7.1	0.3	5.5	4.8	7.4	6.4	0.6	4.8	9.3	91.5
OSD05528.1	8.7	4.5	2.7	6.0	0.0	2.4	6.6	5.4	3.6	6.0	7.5	6.6	1.5	3.3	5.4	8.7	6.3	0.9	4.5	9.3	91.2
XP_020071435.1	5.9	2.3	4.6	6.9	1.3	2.6	5.0	4.6	2.3	5.3	7.6	10.2	1.7	4.0	3.6	7.9	9.6	1.7	5.0	7.9	94.1
XP_020075778.1	6.3	3.0	6.3	5.9	0.7	2.0	5.9	3.6	2.6	3.6	8.6	10.6	1.7	5.0	4.6	8.6	6.3	1.3	4.3	9.2	93.8
XP_020049386.1	4.0	3.0	6.7	9.3	1.0	1.7	2.0	3.0	2.3	7.0	9.0	9.7	1.7	5.3	4.0	10.3	7.7	1.3	4.0	7.0	96.0
XP_002796429.1	7.2	5.5	3.6	5.9	1.0	2.6	6.5	4.2	3.6	4.9	7.5	5.9	2.3	3.3	3.9	7.2	10.4	2.0	3.6	9.1	93.0
OAT11654.1	8.0	5.1	3.8	6.4	1.0	2.2	4.5	4.8	4.5	5.1	6.1	5.7	2.5	3.2	5.1	6.7	12.	1.9	2.9	8.3	92.2

	Ala(A)%	Arg(R)%	Asn(N)%	Asp(D)%	Cys(C)%	Gln(Q)%	Glu(E)%	Gly(G)%	His(H)%	Ile(I)%	Leu(L)%	Lys(K)%	Met(M)%	Phe(F)%	Pro(P)%	Ser(S)%	Thr(T)%	Trp(W)%	Tyr(Y)%	Val(V)%	Total
																	4				
XP_014533620.1	8.6	4.0	5.6	5.3	1.0	3.3	6.3	6.0	3.0	6.3	7.0	7.0	1.7	4.0	3.3	5.3	8.9	2.3	2.3	8.9	91.5
OBT42089.1	6.9	4.9	4.3	5.6	1.0	1.3	7.6	5.3	3.9	5.6	5.9	8.2	2.0	3.3	4.3	7.2	7.9	1.6	3.6	9.5	93.0
ABD03945.1	5.2	6.2	2.6	4.2	1.0	3.3	6.2	5.2	2.9	4.6	7.5	5.2	1.6	4.6	6.2	9.2	7.5	1.6	3.3	11.8	94.7
ABD03944.1	6.2	5.9	2.6	3.6	1.0	3.3	6.5	5.6	2.9	4.9	7.2	5.6	1.6	3.9	5.6	9.2	7.2	1.6	3.3	12.1	93.6
XP_007211669.1	5.5	4.6	3.6	5.5	0.3	2.6	6.8	5.2	2.9	5.2	8.5	8.1	1.6	4.2	4.6	7.8	8.1	1.6	4.2	8.8	94.2
XP_020235639.1	6.1	4.2	3.5	5.8	0.6	2.9	6.1	5.8	2.9	4.2	8.7	8.4	1.6	5.2	4.5	7.4	6.1	1.6	3.9	10.3	93.7
XP_002315419.2	6.2	5.2	4.9	4.2	0.6	2.9	7.1	4.5	3.2	5.5	7.5	7.1	2.6	5.5	3.9	9.4	6.5	1.6	2.9	8.4	93.5
ABD03946.1	5.9	6.5	2.9	4.6	1.0	2.9	6.5	6.5	2.9	3.3	7.8	5.2	1.6	4.9	5.2	9.1	7.2	1.6	3.3	11.1	94.1
XP_022150541.1	5.8	5.4	3.5	5.4	1.0	2.9	6.4	4.5	2.6	5.4	7.0	7.7	1.3	5.4	3.5	9.6	6.4	1.3	4.2	10.9	94.4
XP_022025248.1	4.5	4.5	4.5	4.5	0.6	3.8	6.4	5.8	2.9	5.8	7.0	7.7	2.6	4.8	4.8	9.6	6.4	1.3	2.6	10.2	95.8
NP_001267899.1	7.8	5.5	2.6	5.2	0.6	4.5	6.1	5.5	2.6	6.1	7.4	6.5	1.6	4.9	4.5	8.4	7.4	1.3	3.6	7.8	92.1
AAB97726.1	5.5	5.2	3.9	5.2	0.6	3.2	6.5	5.2	2.9	4.9	8.8	7.1	1.3	5.2	4.9	7.1	6.8	1.6	4.2	9.7	94.3
BAA13184.1	5.8	4.5	3.2	4.9	0.6	4.5	6.8	5.5	2.6	4.9	8.4	7.8	1.3	5.2	4.5	7.4	6.8	1.6	4.2	9.4	94.1

	Ala(A) %	Arg(R) %	Asn(N) %	Asp(D) %	Cys(C) %	Gln(Q) %	Glu(E) %	Gly(G) %	His(H) %	Ile(I)%	Leu(L) %	Lys(K) %	Met(M) %	Phe(F) %	Pro(P) %	Ser(S)%	Thr(T) %	Trp(W) %	Tyr(Y) %	Val(V) %	Total
XP_0208700 05.1	5.5	5.5	2.9	4.9	1.0	1.6	6.8	6.5	4.5	5.2	8.1	6.8	2.3	4.9	4.2	9.4	6.5	1.3	3.6	8.7	94.7
BAB18538.1	5.5	4.2	3.3	5.2	0.7	3.3	7.5	4.9	3.3	5.5	8.5	8.8	2.0	4.9	4.2	8.1	6.2	1.3	3.9	8.8	94.6
ABD03939.1	5.8	4.5	3.6	5.8	1.0	3.2	6.8	4.9	2.3	4.5	7.8	9.1	1.9	5.2	3.9	8.8	6.2	1.3	3.9	9.4	94.1
CAB77205.1	5.5	5.5	3.2	5.8	0.6	2.6	7.1	5.2	2.9	5.2	7.8	8.4	1.9	5.5	3.9	8.4	7.1	1.3	3.6	8.1	94.1
NP_446220.1	3.3	4.6	3.6	6.9	1.3	3.3	5.9	5.9	5.3	6.9	6.9	8.9	2.0	5.0	4.3	5.3	6.9	1.0	4.0	8.6	96.6
NP_0011215 45.1	2.7	4.0	3.3	6.3	1.3	4.3	6.3	5.3	4.3	6.7	7.0	9.0	2.0	4.7	3.7	6.3	7.3	1.0	4.0	10. 3	97.1
NP_033500.1	3.3	5.0	4.3	6.3	1.3	3.6	6.3	5.9	5.0	6.9	7.6	8.6	2.3	4.6	4.3	5.0	6.3	1.0	4.3	8.3	96.9
XP_0060709 40.1	3.0	4.6	3.9	5.6	1.3	3.9	6.9	5.6	4.9	6.6	7.6	8.6	2.0	4.9	3.9	5.9	6.6	1.0	4.3	8.9	97.0
XP_0069194 20.1	3.0	3.6	3.3	6.2	1.6	3.3	6.9	5.3	4.6	7.6	7.6	10. 2	2.6	4.6	3.6	5.9	7.2	1.3	4.3	7.2	96.9
XP_0060861 99.2	4.0	4.6	4.0	6.6	1.3	3.0	6.3	5.6	4.6	6.6	6.3	9.2	2.3	5.3	4.0	6.6	6.6	1.0	3.6	8.6	96.1
XP_0151463 62.1	3.8	4.4	5.0	4.4	5.6	5.3	7.2	5.9	2.8	6.9	6.9	7.2	1.9	5.3	3.8	5.0	5.3	0.9	3.8	8.8	96.4
XP_0126571 76.1	3.3	4.0	3.3	6.9	1.7	3.3	6.3	5.6	5.3	6.9	6.6	8.9	2.0	5.0	3.6	6.3	6.9	1.0	4.0	9.2	96.8
XP_0126030 35.1	3.3	4.3	4.6	6.2	1.3	3.0	6.6	5.6	5.3	6.9	6.9	9.2	2.0	4.9	3.9	5.9	6.2	1.0	3.9	8.9	96.6
XP_0210139 52.1	3.0	4.9	3.9	6.9	1.3	3.6	6.2	5.9	5.3	6.9	7.2	8.6	2.3	4.6	4.3	4.9	6.2	1.0	4.3	8.6	96.9
XP_0257278	3.0	4.3	3.9	6.9	2.0	3.0	6.2	5.6	5.3	6.9	5.9	9.5	3.0	4.9	3.9	5.3	6.6	0.7	4.3	8.9	97.1

	Ala(A)%	Arg(R)%	Asn(N)%	Asp(D)%	Cys(C)%	Gln(Q)%	Glu(E)%	Gly(G)%	His(H)%	Ile(I)%	Leu(L)%	Lys(K)%	Met(M)%	Phe(F)%	Pro(P)%	Ser(S)%	Thr(T)%	Trp(W)%	Tyr(Y)%	Val(V)%	Total
15.1																					
XP_025273620.1	3.3	3.9	3.9	6.6	1.3	3.0	6.6	5.6	4.9	6.6	5.9	9.9	2.6	4.9	3.9	5.9	6.2	1.0	4.6	9.2	96.5
NP_001037382.1	7.7	4.7	3.0	5.0	0.6	3.0	6.8	4.5	3.9	4.5	7.4	6.8	1.8	4.2	5.3	7.4	8.0	1.8	4.5	9.2	92.4
AFP60128.1	5.3	4.4	6.5	6.2	1.5	3.8	5.6	4.7	4.7	5.9	5.9	7.4	1.2	5.3	5.0	6.2	6.2	0.9	4.4	8.6	94.4
XP_022352809.1	3.0	3.6	3.3	7.2	1.6	3.6	6.2	5.6	4.9	6.9	5.9	9.5	3.0	4.9	3.9	5.6	6.6	1.0	4.6	8.9	96.8
Average	6.56	5.15	3.82	6.19	0.91	3.17	6.53	5.65	3.73	5.31	7.39	6.85	1.82	4.62	4.10	6.90	7.31	2.63	3.74	8.79	94.58

Multiple sequence alignment of uricase from different *Bacillus* species

```
3WLVA           1 ------------------------------------------------------------
KGM46460.1      1 ------------------------------------------------------------
WP_045524647.1  1 ------------------------------------------------------------
WP_034672575.1  1 ------------------------------------------------------------
AOM84027.1      1 ------------------------------------------------------------
1J2GA           1 ------------------------------------------------------------
ACR09749.1      1 ------------------------------------------------------------
4R8XA           1 ------------------------------------------------------------
SPT78254.1      1 ---------------------------------------------------------MEN
KKI85158.1      1 ---------------------------------------------------------MEN
OXS68986.1      1 -------------------------------------------------------MMKNN
KOR85772.1      1 ------------------------------------------------------------
WP_098373266.1  1 ------------------------------------------------------------
PEZ74426.1      1 ------------------------------------------------------------
PCD05853.1      1 ------------------------------------------------------------
PAL09042.1      1 ------------------------------------------------------------
CEG34811.1      1 ------------------------------------------------------------
5AYJA           1 ------------------------------------------------------------
BAA08723.1      1 ------------------------------------------------------------
AKO95039.1      1 -------------------------------------------MTNKGEIKVMKNN
4R99A           1 ------------------------------------------------------------
BAD66267.1      1 -------------------------------------MVSLRSVHQTTERVGVEVEN
WP_095326636.1  1 -------------------------------------MVSLRSVHQTTERVGVEVEN
WP_095294289.1  1 -------------------------------------MVSLRSVHQTTERVGVEVEN
WP_095236414.1  1 -------------------------------------MVSLRSVHQTTERVGVEVEN
PAF09838.1      1 -------------------------------------MVSLRSVHQTTERVGVEVEN
PAE88988.1      1 -------------------------------------MVSLRSVHQTTERVGVEVEN
PAD14932.1      1 -------------------------------------MVSLRSVHQTTERVGVEVEN
WP_081496159.1  1 -------------MITKKFSDIFEDSPLVASRIYQLYKIARYFVYMTNKGEIKVMKNN
OWV36502.1      1 --MFTVDDLNQMDIQTLTDTLGSIFEHSSWIAEKAAALRPFSSLSDLHHKVAGIVKAADR
EXF55358.1      1 --MFTVDDLNQMDTQTLTDTLGSIFEHSSWIAERSAALRPFSSLSDLHRKVTGIVKAADR
CCU60286.1      1 --MFTVDDLNQMDTQTLTGTLGSIFEHSSWIAERSAALRPFSSLSDLHRKVTGIVKAADR
AII36988.1      1 --MFTVDDLNQMDTQTLTDTLGSIFEHSSWIAERSAALRPFSSLSDLHRKVTGIVKAADR
APH66064.1      1 --MFTVDDLNQMDIQTLTDTLGSIFEHSSWIAEKAAALRPFSSLSDLHHKVAGIVKAADR
BAM59327.1      1 --MFTVDDLNQMDTQTLTDTLGSIFEHSSWIAERSAALRPFSSLSDLHRKVTGIVKAADR
WP_101501434.1  1 --MFTVDDLNQMDTQTLTDTLGSIFEHSSWIAERSAALRPFSSLSDLHRKVTGIVKAADR
AKC48817.1      1 --MFTVDDLNQMDTQTLTDTLGSIFEHSSWIAERSAALRPFSSLSDLHRKVTGIVKAADR
PAD64173.1      1 --MFTVDDLNQMERHTLTDTLGSIFEHSSWIAEEAAELRPFSSLSDLHRKVAGIVKAADR
WP_095241385.1  1 --MFTIDDLNQMERHTLTDTLGSIFEHSSWIAEEAAELRPFSSLSDLHRKVAGIVKAADR
WP_045926035.1  1 --MFTIDDLNQMERHTLTDTLGSIFEHSSWIAEEAAELRPFSSLSDLHRKVAGIVKAADR
AJW84706.1      1 --MFTVDDMNQMDIQTLTDTLESIFEHSSWIAEKAAALRPFSSLSDLHHKVASIVKAADR
KFK78955.1      1 --MFTVDDMNQMDIQTLTDTLESIFEHSSWIAEKAAALRPFSSLSDLHHKVASIVKAADR
AXJ21641.1      1 --MFTVDDLNQMDIQTLTDTLGSIFEHSSCIAEKAAALRPFSSLSDLHHKVAGIVKAADR
WP_003222862.1  1 --MFTVDDMNQMDIQTLTDTLESIFEHSSWIAEKAAALRPFSSLSDLHHKVASIVKAADR
EME07777.1      1 --MFTVDDLNQMDTQTLTDTLGSIFEHSSWIAERSAALRPFSSLSDLHRKVAGIVKAADR
AUZ27736.1      1 --MFTVDDLNQMDRHTLTDTLGSIFEHSSWIAQETAALRPFSSLSDLHHKVAGIVKAADR
KUP29050.1      1 --MFTVDDLNQMDRQTLTDTLGSIFEHSAWIAEEAAALRPFSSLSDLHQKVSRLVKAADR
PRS06588.1      1 --MFTVDDLNQMDRQTLTDTLGSIFEHSAWIAEEAAALRPFSSLSDLHQKVSRLVKAADR
PRP51591.1      1 --MFTVDDLNQMDRQTLTDTLGSIFEHSAWIAEEAAALRPFSSLSDLHQKVSRLVKAADR
WP_099043576.1  1 --MFTVDDLNQMDRQTLTDTLGSIFEHSAWIAEEAAALRPFSSLSDLHQKVSRLVKAADR
OLQ49074.1      1 --MFTVDDLNQMDRHTLTDTLGSIFEHSSWIAEEAAALRPFSSLSDLHHKVAGIVKAADR
WP_075749098.1  1 --MFTVDDLNQMDTQTLTDTLGSIFEHSSWIAERSAALRPFSSLSDLHHKVAGIVKAADR
AFI29791.1      1 --MLTVDDLNQMDTQTLTDTLGSIFEHSSWIAERSAALRPFSSLSDLHRKVTGIVKAADR
WP_014665258.1  1 --MLTVDDLNQMDTQTLTDTLGSIFEHSSWIAERSAALRPFSSLSDLHRKVTGIVKAADR
ASB62313.1      1 --MLTVDDLNQMDTQTLTDTLGSIFEHSSWIAERSAALRPFSSLSDLHRKVTGIVKAADR
AOL30990.1      1 --MFTVDDLNQMDTQTLTDTLGSIFEHSSWIAERSAALRPFSSLSDLHRKVTGIVKAADR
WP_083686476.1  1 MMKQTIHEVNELPEQTFIGLFQDIYEHSSWIVKKVAPLRPFSSLQEFHHQTIRVIDEASN
OZT14492.1      1 MIKRTIHEVNKLPEQTFIELFQDIYEHSSWIVEKIVPLRPFISLQEFHHRTIKIINEASD
WP_094910043.1  1 MIKRTIHEVNKLPEQTFIELFQDIYEHSSWIVEKIVPLRPFISLQEFHHRTIKIINEASD
```

```
AQX54882.1          1 MMKQTIHEVNELPEQTFIGLFQDIYEHSSWIVKKVAPLRPFSSIQEFHHRTIRVIDEASN
WP_061784634.1      1 MMKQTIHEVNELPEQTFIGLFQDIYEHSSWIVKKVAPLRPFSSIQEFHHRTIRVIDEASN
WP_078989772.1      1 MMKQTIHEVNELPEQTFIGLFQDIYEHSSWIVKKVAPLRPFSSIQEFHHRTIRVIDEASN
WP_040342081.1      1 MMR--LKQLNEMSASEFIHLLGGVFENSSWVAERAEPNRPYSSFQSLYNKMVEIVETASE
WP_101224285.1      1 --MITLKSINAISEKDFTMFLGDTFEHSPWIAEKAAANRPFSSIILHQCVVDIVSNSSN
ASS93773.1          1 --MITLKSINAISEKEFTKFLGDTFEHSPWIAEKSAANRPFSSIINLHRCVVNIVSNSSK
WP_063232385.1      1 --MITLKSINAISEKEFTKFLGDTFEHSPWIAEKSAANRPFSSIINLHRCVVNIVSNSSK
WP_061143228.1      1 --MLTLAELNAQSEKEFTKFLGDTFEHSPWIAQKAAASRPFHSIMNLHQCLEAIVRNSSK
PJN86603.1          1 --MITLAELNAQSEKEFTKFLGDTFEHSPWIAKKAAASRPFSSIMNLHQCLEAIVRNSSK
WP_048623468.1      1 MMR--LKQINEMSASEFIHLLGGVFENSSWVAERAEPNRPYSSFQSLYNKMVEIVETASD
BAB20808.1          1 MMR--LKQINEMSASEFIHLLGGVFENSSWVAERAEPNRPYSSFQSLYNKMVEIVETASD
consensus           1      .. ... .       .  ..... ..    .  ...... ..  .. ..  ..


3WLVA               1 ------------------------------------------------------------
KGM46460.1          1 ------------------------------------------------------------
WP_045524647.1      1 ------------------------------------------------------------
WP_034672575.1      1 ------------------------------------------------------------
AOM84027.1          1 ------------------------------------------------------------
1J2GA               1 ------------------------------------------------------------
ACR09749.1          1 ------------------------------------------------------------
4R8XA               1 ------------------------------------------------------------
SPT78254.1          4 ------------------------------------------------------------
KKI85158.1          4 ------------------------------------------------------------
OXS68986.1          6 ------------------------------------------------------------
KOR85772.1          1 ------------------------------------------------------------
WP_098373266.1      1 ------------------------------------------------------------
PEZ74426.1          1 ------------------------------------------------------------
PCD05853.1          1 ------------------------------------------------------------
PAL09042.1          1 ------------------------------------------------------------
CEG34811.1          1 ------------------------------------------------------------
5AYJA               1 ------------------------------------------------------------
BAA08723.1          1 ------------------------------------------------------------
AKO95039.1         14 ------------------------------------------------------------
4R99A               1 ------------------------------------------------------------
BAD66267.1         21 ------------------------------------------------------------
WP_095326636.1     21 ------------------------------------------------------------
WP_095294289.1     21 ------------------------------------------------------------
WP_095236414.1     21 ------------------------------------------------------------
PAF09838.1         21 ------------------------------------------------------------
PAE88988.1         21 ------------------------------------------------------------
PAD14932.1         21 ------------------------------------------------------------
WP_081496159.1     46 ------------------------------------------------------------
OWV36502.1         59 RTQLDLINKHPRLGTKKTMSVTSVREQQNAGISKLEQEEYEEFIKLNEHYYERFGFPFIL
EXF55358.1         59 ETQLDLIKKHPRLGTKKTMSDDSVREQQNAGIGKLEQQEYEEFIMLNEHYYDRFGFPFIL
CCU60286.1         59 ETQLDLIKKHPRLGTKKTMSDDSVREQQNAGIGKLEQQEYEEFIMLNEHYYDRFGFPFIL
AII36988.1         59 ETQLDLIKKHPRLGTKKTMSDDSVREQQNAGIGKLEQQEYEEFIMLNEHYYDRFGFPFIL
APH66064.1         59 RTQLDLIKKHPRLGTKKTMSVTSVREQQNAGISKLEQEEYEEFIKLNEHYYERFGFPFIL
BAM59327.1         59 ETQLDLIKKHPRLGTKKTMSDDSVREQQNAGIGKLEQQEYEEFIMLNEHYYDRFGFPFIL
WP_101501434.1     59 ETQLDLIKKHPRLGTKKTMSNDSVREQQNAGISKLEQQEYEEFIKLNEHYYDRFGFPFIL
AKC48817.1         59 ETQLDLIKKHPRLGTKKTMSDDSVREQQNAGISKLEQQEYEEFIMLNEHYYDRFGFPFIL
PAD64173.1         59 QTQLDLINKHPRLGTKKNIMSDASVSEQRNAGISKLEQEEYEEFIKLNEHYDERFGFPFIL
WP_095241385.1     59 QTQLDLINKHPRLGTKNIMSDASVSEQRNAGISKLEQEEYEEFIKLNEHYDERFGFPFIL
WP_045926035.1     59 QTQLDLINKHPRLGTKNIMSDASVSEQRNAGISKLEQEEYEEFIKLNEHYERFGFPFIL
AJW84706.1         59 QTQLDLINKHPRLGTKNTMSVTSVREQQNAGISKLEQEEYEEFIKLNERYERFGFPFIL
KFK78955.1         59 QTQLDLINKHPRLGTKKTMSVTSVREQQNAGISKLEQEEYEEFIKLNEHYYERFGFPFIL
AXJ21641.1         59 RTQLDLINQHPRLGTKKTMSVTSVREQQNAGISKLEQEEYEEFIKLNEHYYERFGFPFIL
WP_003222862.1     59 QTQLDLINKHPPLGTKNTMSVTSVREQQNAGISKLEQEEYEEFIKLNERYYERFGFPFIL
EME07777.1         59 ETQLDLIKKHPRLGTKKTMSDDSVREQQNAGIGKLEQQEYEEFIMLNEHYYDRFGFPFIL
AUZ27736.1         59 QTQLDLINKHPRLGTKKAMSDTSVREQQNAGISKLEQPEYAEFIKLNEHYYARFGFPFIL
KUP29050.1         59 KTQLELICKHPRLGTKKTMSASSVKEQQNAGISKLEQQEYEEFIKQNEDYYHNFGFPFIL
PRS06588.1         59 KTQLELICKHPRLGTKKPVSASSVKEQQNAGISKLEQQEYEEFIKRNEDYYHNFGFPFIL
PRP51591.1         59 KTQLELICKHPRLGTKKTMSASSVKEQQNAGISKLEQQEYEEFIKQNEDYYHNFGFPFIL
WP_099043576.1     59 KTQLELICKHPRLGTKKTMSASSVKEQNAGISQLEQHEYEEFIKQNEDYYHNFGFPFIL
OLQ49074.1         59 QTQLDLISKHPRLGTKKAMSDTSVREQQNAGISKLEQPEYAEFIKLNEHYYARFGFPFIL
WP_075749098.1     59 QTQLDLISKHPRLGTKKAMSDTSVREQQNAGISKLEQPEYAEFIKLNEHYYARFGFPFIL
AFI29791.1         59 ETQLDVINKHPRLGTKKTMSSESIKEQNAGISKLEQQEYEEFIKLNEHYYDRFGFPFIL
WP_014665258.1     59 ETQLDVINKHPRLGTKKTMSSESIKEQNAGISKLEQQEYEEFIKLNEHYYDRFGFPFIL
ASB62313.1         59 ETQLEVINKHPRLGTKKTMSSESIKEQNAGISKLEQQEYEEFIKLNEHYYDRFGFPFIL
AOL30990.1         59 ETQLDLIVKHPRLGTKKTMSGDSVREQQNAGISKLEQQEYEEFIMLNEHYYDRFGFPFIL
WP_083686476.1     61 QRKLDLLQAHPNLGAKIAMTHSINEQTGAGLTSLTAEEYEHFTNANKTYMNRFGFPFIV
```

206

```
OZT14492.1        61 QRKLDLLQAHPNLGAKIAMTTHSINEQKGAGLTSLTAREYEDFTHANKAYMNRFGFPFII
WP_094910043.1    61 QRKLDLLQAHPNLGAKIAMTTHSINEQKGAGLTSLTAREYEDFTHANKAYMNRFGFPFII
AQX54882.1        61 QRKLDLLQAHPNLGAKIAMTTHSINEQTGAGLTSLTAEEYEHFTNANKTYMNRFGFPFIV
WP_061784634.1    61 QRKLDLLQAHPNLGAKIAMTTHSINEQTGAGLTSLTAEEYEHFTNANKTYMNRFGFPFIV
WP_078989772.1    61 QRKLDLLQAHPNLGAKIAMTTHSINEQTGAGLTSLTAEEYEHFTNANKTYMNRFGFPFIV
WP_040342081.1    59 NEQLKLIQMHPHLGTNVKLTDFSQEEQKHAGLNELTEDEHNHLMLLNKEYMDKFGFPFVM
WP_101224285.1    59 EEKLTLIKKHPNLGDKVEMSDDSTKEQHGAGLKDLTAEEYESFTSLNRQYMNKFGFPFIL
ASS93773.1        59 EEKLTLIRKHPNLGDKVEMSEDSTKEQHGAGLKDLTADEYENFTSLNRQYMNKFGFPFIL
WP_063232385.1    59 EEKLTLIRKHPNLGDKVEMSEDSTKEQHGAGLKDLTADEYENFTSLNRQYMNKFGFPFIL
WP_061143228.1    59 EEKLALIRKHPNLGDKVKMSNESILEQHGAGLGDLTAAEYENFTSLNRQYMQKFGFPFIL
PJN86603.1        59 EEKLALIRKHPNLGDKVKMSNDSILEQHGAGLGDLTAAEYENFTSLNRQYMKKFGFPFIL
WP_048623468.1    59 NEQLKLIQMHPHLGTNVKLTDFSQEEQKHAGLNELTKDEQNHLILLNQKYKDKFGFPFVM
BAB20808.1        59 NEQLKLIQMHPHLGTNVKLTDFSQEEQKHAGLNELTKDEQNHLILLNQKYKDKFGFPFVM
consensus         61  ..  .. ...  .. .. ...  ...   ... ...  .  .  .......


3WLVA              1 --------------------------------------------------------V
KGM46460.1         1 -----------------------------------------------------MKR
WP_045524647.1     1 -----------------------------------------------------MKR
WP_034672575.1     1 -----------------------------------------------------MKR
AOM84027.1         1 ----------------------------------------------------MTER
1J2GA              1 -------------------------------------------------MTKHKERV
ACR09749.1         1 -------------------------------------------------------R
4R8XA              1 ----------------------------------------------------AER
SPT78254.1         4 -----------------------------------------------------KR
KKI85158.1         4 -----------------------------------------------------KR
OXS68986.1         6 -----------------------------------------------------NR
KOR85772.1         1 -------------------------------------------------MTAQNENR
WP_098373266.1     1 -------------------------------------------------MTVQNENR
PEZ74426.1         1 -------------------------------------------------MTVQNENR
PCD05853.1         1 -------------------------------------------------MTAQNENR
PAL09042.1         1 -------------------------------------------------MTAQNENR
CEG34811.1         1 -------------------------------------------------MTAQNENR
5AYJA              1 --------------------------------------------------TKHKERV
BAA08723.1         1 -------------------------------------------------MTKHKERV
AKO95039.1        14 -----------------------------------------------------NR
4R99A              1 ----------------------------------------------------AER
BAD66267.1        21 -----------------------------------------------------KR
WP_095326636.1    21 -----------------------------------------------------KR
WP_095294289.1    21 -----------------------------------------------------KR
WP_095236414.1    21 -----------------------------------------------------KR
PAF09838.1        21 -----------------------------------------------------KR
PAE88988.1        21 -----------------------------------------------------KR
PAD14932.1        21 -----------------------------------------------------KR
WP_081496159.1    46 -----------------------------------------------------NR
OWV36502.1       119 AVKGKTKQDIHQALLERLENERETEFQQALEEIYRIARFRLADIITEKGETQM-----KR
EXF55358.1       119 AVKGKTKQDIHQALLARLESERETEFQQALIEIYRIARFRLADIITEKGETQM-----KR
CCU60286.1       119 AVKGKTKQDIHQALLARLESERETEFQQALIEIYRIARFRLADIITEKGETQM-----KR
AII36988.1       119 AVKGKTKQDIHQALLARLENERETEFQQALEEIYRIARFRLADIITEKGETQM-----KR
APH66064.1       119 AVKGKTKQDIHQALLARLESERETEFQQALIEIYRIARFRLADIITEKGETQM-----KR
BAM59327.1       119 AVKGKTKQDIHQALLARLESERETEFQQALIEIYRIARFRLADIITEKGETQM-----KR
WP_101501434.1   119 AVKGKTKQDIHQALLARLESERETEFQQALIEIYRIARFRLADIITEKGETQM-----KR
AKC48817.1       119 AVKGKTKQDIHQALLARLESERETEFQQALIEIYRIARFRLADIITEKGETQM-----KR
PAD64173.1       119 AVKGKTKQDIHRALVKRLENEQETEFQQALIEIYRIARFRLADIITEKGETQM-----KR
WP_095241385.1   119 AVKGKTKQDIHRALVKRLENEQETEFQQALIEIYRIARFRLADIITEKGETQM-----KR
WP_045926035.1   119 AVKGKTKQDIHRALVKRLENEQETEFQQALIEIYRIARFRLADIITEKGETQM-----KR
AJW84706.1       119 AVKGKTKQDIYQALLERLENERETEFHQALKEIYRIARFRLADIITEKGETQM-----KR
KFK78955.1       119 AVKGKTKQDIYQALLERLENERETEFHQALKEIYRIARFRLADIITEKGETQM-----KR
AXJ21641.1       119 AVKGKTKQDIHQALLERLENERETEFQQALEEIYRIARFRIVDIITEKGETQM-----KR
WP_003222862.1   119 AVKGKTKQDIYQALLERLENERETEFHQALKEIYRIARFRLADIITEKGETQM-----KR
EME07777.1       119 AVKGKTKQDIHQALLARLESERETEFQQALIEIYRIARFRLADIITEKGETQM-----KR
AUZ27736.1       119 AVKGKAKQGIRRALLARLENDPETEFQQALIEIYRIARFRLADIITEKGETQM-----KR
KUP29050.1       119 AVKEKTKQEIRQALLTRLKNKPETEFQQALIEIYRIARFRLEDIITEKGEIQM-----KR
PRS06588.1       119 AVKKKTKQEIRQALLTRLKNKPETEFQQALIEIYRIARFRLEDIITEKGAIQM-----KR
PRP51591.1       119 AVKKKTKQEIRQALLTRLKNKPETEFQQALIEIYRIARFRLEDIITEKGAIQM-----KR
WP_099043576.1   119 AVKEKTKQEIRQALLTRLKNKPETEFQQALIEIYRIARFRLEDIITEKGEIQM-----KR
OLQ49074.1       119 AVKGKTKQDIRRALLARLENDPETEFQQALIEIYRIARFRLADIITEKGETQM-----KR
WP_075749098.1   119 AVKGKTKQDIRRALLARLENDPETEFQQALIEIYRIARFRLADIITEKGETQM-----KR
AFI29791.1       119 AVKGKAKQDIHQALLERLENERESEFQQALIEIYRIARFRLADIITEKGETQM-----KR
WP_014665258.1   119 AVKGKAKQDIHQALLERLENERESEFQQALIEIYRIARFRLADIITEKGETQM-----KR
ASB62313.1       119 AVKGKTKQDIHQALLERLENERESEFQQALIEIYRIARFRLADIITEKGETQM-----KR
```

207

```
AOL30990.1       119 AVKGTKQDIHQALIARLESERETEFQQALIEIYRIARFRIADIITEKGETQM-----KR
WP_083686476.1   121 AVRGTKSTIYQSLIDRLQNDKKTEFTTALAEVYKIAYFRLVDKIKTEESVTMTNQSNRQ
OZT14492.1       121 AVRGTKSTIYQILLDRLKNDKNTEFTTALAEVYKIAYFRLTDKIKTEERVTMTNHSNRQ
WP_094910043.1   121 AVRGTKSTIYQILLDRLQNDKKTEFTTALAEVYKIAYFRLTDKIKTEERVTMTNHSNRQ
AQX54882.1       121 AVRGTKSTIYQSLINRLQNDKKTEFTTALAEVYKIAYFRLVDKIKTEERVTMTNQSNRQ
WP_061784634.1   121 AVRGTKSTIYQSLINRLQNDKKTEFTTALAEVYKIAYFRLVDKIKTEERVTMTNQSNRQ
WP_078989772.1   121 AVRGTKSTIYQSLIDRLQNDKKTEFTTALAEVYKIAYFRLVDKIKTEERVTMTNQSNRQ
WP_040342081.1   119 AVRGTKQDIYRTIKERLKNNYRTEFEQALEEIKKIAMFRLQEIINGGEMISMTNYKERV
WP_101224285.1   119 AVRGDKNDIYQSMKTRLHHSETIEFDKALSEIHKIALFRLQDKIKIEGENSMKNKSAAQ
ASS93773.1       119 AVRGDKNDIYQSMKTRLHHSETAEFDKALSEIHQIALFRLQDKIKIEGEKPMKNKSAAQ
WP_063232385.1   119 AVRGDKNDIYQSMKTRLHHSETAEFDKALSEIHQIALFRLQDKIKIEGEKPMKNKSAAQ
WP_061143228.1   119 AVRGDKNDIYQSMLTRLNHTTTIEFDKALAEIYRIALFRLQDKIKIEGKNSMKNKSAAQ
PJN86603.1       119 AVRGDKNDIYQSMLTRLNHTTTAEYDKALAEIYRIALFRLQDKIKIEGKNSMKNKSAAQ
WP_048623468.1   119 AVRGKIKQEIFRTIKERLQNNHQTEFKQALEEIKKIAMFRLQEIFREGENNSMTKHKERV
BAB20808.1       119 AVRGKIKQEIFRTIKERLQNNHQTEFKQALEEIKKIAMFRLQEIFREGENNSMTKHKERV
consensus        121 ..... .. .     ... ..        ... ..   ...... ... .        ..


3WLVA              2 -MYYGKGDVFAYRTYLKPLTGVRTIPESPFSGRDHILFGVNVKISVGGTKLLISFTKGDN
KGM46460.1         4 TMSYGKADVWVYRTYAKPLTVIRNIPESTFIGRKNILFGVNVKVAVEGEAFFSSFKNGDN
WP_045524647.1     4 TMSYGKADVWVYRTYAKPLTSIRSIPESSFIGRKNILFGVNVKVAIEGEAFFSSFKDGDN
WP_034672575.1     4 TMSYGKADVWVYRTYAKPLTVIRNIPESTFIGRKNILFGVNVKVAVEGEAFFSSFKNGDN
AOM84027.1         5 TMLYGKGDVFVYRTYATPLTGIAMIPESDEKGRSNTILGMDIQVSLAGEAFFSFTEGDN
1J2GA              9 -MYYGKGDVFAYRTYLKPLTGVRTIPESPFSGRDHILFGVNVKISVGGTKLLISFTEGDN
ACR09749.1         2 TMFYGKGDVYVERTYANPIKGLKQIPESNFTEKHNTIFGVNAKVALKGEQLLISFTEGDN
4R8XA              4 TMFYGKGDVYVERTYANPIKGLKQIPESNFTEKHNTIFGVNAKVALKGEQLLISFTEGDN
SPT78254.1         6 TMSYGKGNVFAYRTFVEPLSGLQPIPESAFIVRDNTVFGINVTVEGGNAFLSSFTDGDN
KKI85158.1         6 TMSYGKGNVFAYRTFVEPLSGLQPIPESAFAVRDNTVFGINVTVEGGNAFLSSFTDGDN
OXS68986.1         8 IMAYGKGDVFAYRTHLAPIKGVKEIPESTFIGRSNIIFGANLRVEGGSAFLSSFTEGDN
KOR85772.1         9 TMYYGKGDVFVYRTFVQPLTGLRKIPESSFIERDNTIFGFNCQISLKGEAFLSSFTEGDN
WP_098373266.1     9 TMYYGKGDVFVYRTFVQPLTGLRKIPESGFTERDNTIFGFNCRISLKGEAFLSSFTEGDN
PEZ74426.1         9 TMYYGKGDVFVYRTFVQPLTGLRKIPESGFTERDNTIFGFNCRISLKGEAFLSSFTEGDN
PCD05853.1         9 TMYYGKGDVFVYRTFVQPLTGLRKIPESSFTERDNTIFGFNCQISLKGEAFLSSFTEGDN
PAL09042.1         9 TMYYGKGDVFVYRTFVQPLTGLRKIPESSFTERDNTIFGFNCQISLKGEAFLSSFTEGDN
CEG34811.1         9 TMFYGKGDVFVYRTFVQPLTGLRKIPESSFTERDNTIFGFNCQISLKGEAFLSSFTEGDN
5AYJA              8 -MYYGKGDVFAYRTYLKPLTGVRTIPESPFSGRDHILFGVNVKISVGGTKLLISFTKGDN
BAA08723.1         9 -MYYGKGDVFAYRTYLKPLTGVRTIPESPFSGRDHILFGVNVKISVGGTKLLISFTKGDN
AKO95039.1        16 IMAYGKGDVFAYRTHLAPIKGVKEIPESTFIGRSNIIFGANLRVEGGSAFLSSFTEGDN
4R99A              4 TMFYGKGDVYVERTYANPIKGLKQIPESNFTEKHNTIFGVNAKVALKGEQILLISFTEGDN
BAD66267.1        23 TMSYGKGNVFAYRTFVEPLSGLQPIPESAFIVRDNTVFGINVTVEGGNAFLSSFTDGDN
WP_095326636.1    23 TMSYGKGNVFAYRTFVEPLSGLQPIPESAFIVRDNTVFGINVTVEGGNAFLSSFTDGDN
WP_095294289.1    23 TMSYGKGNVFAYRTFVEPLSGLQPIPESAFIVQDNTVFGINVTVEGGNAFLSSFTDGDN
WP_095236414.1    23 TMSYGKGNVFAYRTFVEPLSGLQPIPESAFIVQDNTVFGINVTVEGGNAFLSSFTDGDN
PAF09838.1        23 TMSYGKGNVFAYRTFVEPLSGLQPIPESAFAVRDNTVFGINVTVEGGNAFLSSFTDGDN
PAE88988.1        23 TMSYGKGNVFAYRTFVEPLSGLQPIPESAFTVRDNTVFGINVTVEGGNAFLSSFTDGDN
PAD14932.1        23 TMSYGKGNVFAYRTFVEPLSGLQPIPESAFTVRDNTVFGINVTVEGGNAFLSSFTDGDN
WP_081496159.1    48 IMAYGKGDVFAYRTHLAPIKGVKEIPESTFIGRSNIIFGANLRVEGGSAFLSSFTEGDN
OWV36502.1       174 TMSYGKGNVFAYRTFLKPLTGVRQIPESSFIGRDNTVVGIDVTCEIGGDAFLPSFTDGDN
EXF55358.1       174 TMSYGKGNVFAYRTFLKPLTGVKQIPESSFAGRSNTVVGIDVTCEIGGEAFLPSFTDGDN
CCU60286.1       174 TMSYGKGNVFAYRTFLKPLTGVKQIPESSFAGRSNTVVGIDVTCEIGGEAFLPSFTDGDN
AII36988.1       174 TMSYGKGNVFAYRTFLKPLTGVKQIPESSFAGRSNTVVGIDVTCEIGGEAFLPSFTDGDN
APH66064.1       174 TMSYGKGNVFAYRTFLKPLTEVRQIPESSFIGRDNTVVGIDVTCEIGGDAFLPSFTDGDN
BAM59327.1       174 TMSYGKGNVFAYRTYLKPLTGVKQIPESSFAGRDNTVVGIDVTCEIGGEAFLPSFTDGDN
WP_101501434.1   174 TMSYGKGNVFAYRTFLKPLTGVKQIPESSFAGRSNTVVGIDVTCEIGGEAFLPSFTDGDN
AKC48817.1       174 TMSYGKGNVFAYRTYLKPLTGVKQIPESSFAGRDNTVVGIDVTCEIGGEAFLPSFTDGDN
PAD64173.1       174 TMSYGKGNVFAYRTFLKPLTGVKQIPESSFSGRDNTVVGIDVTCEIGGDAFLPSFTDGDN
WP_095241385.1   174 TMSYGKGNVFAYRTFLKPLTGVKQIPESSFSGRDNTVVGIDVTCEIGGDAFLPSFIDGDN
WP_045926035.1   174 TLSYGKGNVFAYRTFLKPLTGVKQIPESSFAGRSNTVVGIDVTCEIGGEAFLPSFIDGDN
AJW84706.1       174 TMSYGKGNVFAYRTFLKPLTGVRQIPESSFIGRDNTVVGIDVTCEIGGDAFLISFTDGDN
KFK78955.1       174 TMSYGKGNVFAYRTFLKPLTGVRQIPESSFIGRDNTVVGIDVTCEIGGDAFLISFIDGDN
AXJ21641.1       174 TMSYGKGNVFAYRTFLKPLTGVRQIPESSFIGRDNTVVGIDVTCEIGGDAFLISFIDGDN
WP_003222862.1   174 TMSYGKGNVFAYRTFLKPLTGVRQIPESSFIGRDNTVVGIDVTCEIGGDAFLISFIDGDN
EME07777.1       174 TMSYGKGNVFAYRTYLKPLTGVKQIPESAFSGRDNTVVGIDVTCEIGGEAFLPSFTDGDN
AUZ27736.1       174 TMSYGKGNVFAYRTFLKPLTGVKQIPESAFSGRDNTVVGIDVTCEIEGDAFLPSFTDGDN
KUP29050.1       174 TMSYGKGNVFAYRTFLKPLTGIKKIPESSFIGRTNTVVGIDVTCEIGGEAFLPSFTDGDN
PRS06588.1       174 TMSYGKGNVFAYRTFLKPLTGIKKIPESSFIGRTNTVVGIDVTCEIGGDAFLQSFTDGDN
PRP51591.1       174 TMSYGKGNVFAYRTFLKPLTGINQIPESSFIGRTNTVVGIDVTCEIGGDAFLPSFTEGDN
WP_099043576.1   174 TMSYGKGNVFAYRTFLKPLTGIKKIPESSFIGRTNTVVGIDVTCEIGGDAFLPSFTDGDN
OLQ49074.1       174 TMSYGKGNVFAYRTFLKPLTGVKQIPESAFSGRDNTVVGIDVTCEIEGDAFLPSFTDGDN
WP_075749098.1   174 TMSYGKGNVFAYRTFLKPLTGVKQIPESAFSGRDNTVVGIDVTCEIEGDAFLPSFTDGDN
AFI29791.1       174 TMSYGKGNVFAYRTFLKPLTGVKQIPESSFIGRDNTVVGIDVTCEIGGDAFLPSFTDGDN
```

```
WP_014665258.1  174  TMSYGKGNVFAYRTFLKPLTGVKQIPESSFTGRDNTVVGVDVTCEIGGDAFLPSFTDGDN
ASB62313.1      174  TMSYGKGNVFAYRTFLKPLTGVKQIPESSFIGRDNTIVGVDVTCEIGGDAFLPSFTDGDN
AOL30990.1      174  TMSYGKGNVFAYRTFLKPLTGVKQIPESSFAGRSNTIVGVDVTCEIGGDAFLPSFTDGDN
WP_083686476.1  181  -MYYGKDVFAYRTYLKPLTGVKVIPESEFSGRNNIVFGVNVKVAIGGSQFLSSFIEGDN
OZT14492.1      181  -MYYGKDVFAYRTYLKPLTGVKVIPESEFSGRNNIVFGVKVKVAIGGSQFLSSFTEGDN
WP_094910043.1  181  -MYYGKDVFAYRTYLKPLKGVKVIPESEFSGRNNIVFGVNVKVAIGGSQFLSSFTEGDN
AQX54882.1      181  -MYYGKDVFAYRTYLKPLKGVKVIPESEFSGRNNIVFGVNVKVAIGGSQFLSSFTEGDN
WP_061784634.1  181  -MYYGKDVFAYRTYLKPLKGVKVIPESEFSGRNNIVFGVNVKVAIGGSQFLSSFIEGDN
WP_078989772.1  181  -MYYGKDVFAYRTYLKPLKGVKVIPESEFSGRNNIVFGVNVKVAIGGSQFLSSFIEGDN
WP_040342081.1  179  -MYYGKGDVFAYRTYLKPLTGVRTIPESPFSGRDHILFGVNVKISVGGTKLLLSFTKGDN
WP_101224285.1  179  TLSYGKGNVFAYRTYSKPLTGIKQIPESIFSGRDHIIFGTNVKVSVGGSSFLPSFTEGDN
ASS93773.1      179  TLSYGKGNVFAYRTYSNPLTGIKQIPESIFSGRDHIIFGTNVKVSVGGSSFLPSFTEGDN
WP_063232385.1  179  TLSYGKGNVFAYRTYSNPLTGIKQIPESIFSGRDHIIFGTNVKVSVGGSSFLPSFTEGDN
WP_061143228.1  179  SLSYGKGNVFAYRTYSIPLTGIKQIPESIFSGRDHIIFGTNVKVSVGGSAFLPSFTEGDN
PJN86603.1      179  SLSYGKGNVFAYRTYSIPLTGIKQIPESIFSGRDHIIFGTNVKVSVGGSAFLPSFTEGDN
WP_048623468.1  179  -MYYGKDVFAYRTYLKPLTGVRTIPESPFSGRDHILFGVNVKISVGGTKLLLSFTKGDN
BAB20808.1      179  -MYYGKDVFAYRTYLKPLTGVRTIPESPFSGRDHILFGVNVKISVGGTKLLLSFTKGDN
consensus       181  ...***..*....**...**....  ****.*........*.......*......**..***


3WLVA            61  SLVVATDSMKNFIQKHLASYTGTTIEGFLEYVATSFLKKYSHIEKISLIGEIPFETTFA
KGM46460.1       64  TLVVATDSMKNFILKHACDYQGATQEGFLEFVAQRFLETYSHMTGINISGEQISFEELPV
WP_045524647.1   64  TLVVATDSMKNFILKHACDYQGATQEGFLEFVAQRFLETYSHMSGINISGEQISFEELPV
WP_034672575.1   64  TLVVATDSMKNFILKHACDYQGATQEGFLEFVAQRFLETYSHMTGINISGEQISFEELPV
AOM84027.1       65  SKVVATDSMKNFILHHACEYTGSTVEGFIAYVSAQFLKKYDHTSSVDMSGKQFPFEAAMV
1J2GA            68  SLVVATDSMKNFIQKHLASYTGTTIEGFLEYVATSFLKKYSHIEKISLIGEIPFETTFA
ACR09749.1       62  SLVVATDSMKNFIQRHAASYEGATIEGFLQYVCEAFLAKYSHIDAVRLEAKEYAFDDIQV
4R8XA            64  SLVVATDSMKNFIQRHAASYEGATIEGFLQYVCEAFLAKYSHIDAVRLEAKEYAFDDIQV
SPT78254.1       66  QMVVATDSMKNFIQRHLATFSGRTIEGFIRYVGEAFLTTYSHIDWVKLTGEAVPFENTTY
KKI85158.1       66  QMVVATDSMKNFIQRHLATFSGRTIEGFIRYVGEAFLTTYSHIDWVKLTGEAVPFENTTY
OXS68986.1       68  RLVVATDSMKNFIQQHLASYSGTTMEGFLDYVGKAFLLKYSQIDSVKLTGEHLSFQEANM
KOR85772.1       69  SLVVATDSMKNIIHRQAANYQGNTAEGFLKFICEVFLDKYSHIDAVELTATEVPFDHVLV
WP_098373266.1   69  SLVIATDSMKNIIHRQAANYQGNTAEGFLKFLCEVFLDKYSHIDAVELTANEVPFNHVLV
PEZ74426.1       69  SLVIATDSMKNIIHRQAANYQGNTAEGFLKFLCEVFLDKYSHIDAVELTANEVPFNHVLV
PCD05853.1       69  SLVVATDSMKNIIHRQAANYQGNTAEGFLKFICEVFLDKYSHIDAVELTATEVPFDHVLV
PAL09042.1       69  SLVVATDSMKNIIHRQAANYQGNTAEGFLKFICEVFLDKYSHIDAVELTATEVPFDHVLV
CEG34811.1       69  SLVVATDSMKNIIHRQAANYQGNTAEGFLKFICEVFLDKYSHIDAVELTATEVPFDHVLV
5AYJA            67  SLVVATDSMKNFIQKHLASYTGTTIEGFLEYVATSFLKKYSHIEKISLIGEIPFETTFA
BAA08723.1       68  SLVVATDSMKNFIQKHLASYTGTTIEGFLEYVATSFLKKYSHIEKISLIGEIPFETTFA
AKO95039.1       76  RLVVATDSMKNFIQQHLASYSGTTMEGFLDYVGKAFLLKYSQIDSVKLTGEHLSFQEANM
4R99A            64  SLVVATDSMKNFIQRHAASYEGATIEGFLQYVCEAFLAKYSHIDAVRLEAKEYAFDDIQV
BAD66267.1       83  QMVVATDSMKNFIQRHLATFSGRTIEGFIRYVGEAFLTTYSHIDWVKLTGEAVPFENTTY
WP_095326636.1   83  QMVVATDSMKNFIQRHLATFSGRTIEGFIRYVGEAFLTTYSHIDWVKLTGEAVPFENTTY
WP_095294289.1   83  QMVVATDSMKNFIQRHLATFSGRTIEGFIRYVGEAFLTTYSHIDWVKLTGEAVPFENTTY
WP_095236414.1   83  QMVVATDSMKNFIQRHLATFSGRTIEGFIRYVGEAFLTTYSHIDWVKLTGEAVPFENTTY
PAF09838.1       83  QMVVATDSMKNFIQRHLATFSGRTIEGFIRYVGEAFLTTYSHIDWVKLTGEAVPFENTTY
PAE88988.1       83  QMVVATDSMKNFIQRHLATFSGRTIEGFIRYVGEAFLTTYSHIDWVKLTGEAVPFENTTY
PAD14932.1       83  QMVVATDSMKNFIQRHLATFSGRTIEGFIRYVGEAFLTTYSHIDWVKLTGEAVPFENTTY
WP_081496159.1  108  RLVVATDSMKNFIQQHLASYSGTTMEGFLDYVGRAFLLKYSQIDSVKLTGEHLSFQEANM
OWV36502.1      234  TLVVATDSMKNFIQRHLASYEGTTTEGFLHYVAHRFLDTYSHMDTITLTGEDIPFEAMPA
EXF55358.1      234  TLVVATDSMKNFIQRHLASYEGTTTEGFLHYVAHRFLDTYSHMDKITLTGEDIPFEAMPA
CCU60286.1      234  TLVVATDSMKNFIQRHLASYEGTTTEGFLHYVAHRFLDTYSHMDKITLTGEDIPFEAMPA
AII36988.1      234  TLVVATDSMKNFIQRHLASYEGTTTEGFLHYVAHRFLDTYSHMDKITLTGEDIPFEAMPA
APH66064.1      234  TLVVATDSMKNFIQRHLASYEGMTTEGFLHYVAHRFLDTYSHMDTITLTGEDIPFEAMPA
BAM59327.1      234  TLVVATDSMKNFIQRHLASYEGTTTEGFLHYVAHRFLDTYSHMDTITLTGEDIPFEAMPA
WP_101501434.1  234  TLVVATDSMKNFIQRHLASYEGTTTEGFLHYVAHRFLDTYSHMDKITLTGEDIPFEAMPA
AKC48817.1      234  TLVVATDSMKNFIQRHLASYEGTTTEGFLHYVAHRFLDTYSHMDTITLTGEDIPFEAMPA
PAD64173.1      234  TLVVATDSMKNFIQRHFASYEGATIEGFLDVAHRFLNTYSHMDTITLTGEDIPFEAMPA
WP_095241385.1  234  TLVVATDSMKNFIQRHFASYEGATIEGFLDVAHRFLNTYSHMDTITLTGEDIPFEAMPA
WP_045926035.1  234  TLVVATDSMKNFIQRHFASYEGTTIEGFLDVAHRFLNTYSHMDTITLTGEDIPFEAMPA
AJW84706.1      234  TLVVATDSMKNFIQSHLASYEGTTTEGFLHYVAHRFLDTYSHMDTITLTGEDIPFEAMPA
KFK78955.1      234  TLVVATDSMKNFIQSHLASYEGTTTEGFLHYVAHRFLDTYSHMDTITLTGEDIPFEAMPA
AXJ21641.1      234  TLVVATDSMKNFIQRHLASYEGTTTEGFLHYVAHRFLDTYSHMDTITLTGEDIPFEAMPA
WP_003222862.1  234  TLVVATDSMKNFIQSHLASYEGTTTEGFLHYVAHRFLDTYSHMDTITLTGEDIPFEAMPA
EME07777.1      234  TLVVATDSMKNFIQRHLASYEGTTTEGFLHYVAHRFLDTYSHMDTITLTGEDIPFEAMPA
AUZ27736.1      234  SPVVATDSMKNFIQRHLASYEGTTTEGFLHYTAHRFLDTYSHMDAIITLTGEDIPFEAMPA
KUP29050.1      234  TLIVATDSMKNFIQRHLASYEGTTSEGFLHYVAHRFFDTYSHMDTITLTGEDIQEAMPA
PRS06588.1      234  ILIVATDSMKNFIQRHLASYEGTTTEGFLHYVAHRFLDTYSHMDTITLTGEDIPFEAMPA
PRP51591.1      234  MLIVATDSMKNFIQRHLASYEGTTTEGFLHYVAHRFLDTYSHMDTITLTGEDIPFEAMPA
WP_099043576.1  234  ILIVATDSMKNFIQRHLASYEGTTSEGFLHYVAHRFFDTYSHMDTITLTGEDIPFEAMPA
OLQ49074.1      234  SPVVATDSMKNFIQRHLASYEGTTTEGFLHYTAHRFLDTYSHMDAIITLTGEDIPFEAMPA
```

```
WP_075749098.1   234 SPVVATDSMKNFIQRHLASYEGTTTEGFLYTAHRFLDTYSHMDAITLTGEDIPFEAMPA
AFI29791.1       234 TLVVATDSMKNFIQRHLASYEGTTTEGFLYVAHRFLDTYSHMDTITLTGEDIPFEAMPA
WP_014665258.1   234 TLVVATDSMKNFIQRHLASYEGTTTEGFLYVAHRFLDTYSHMDTITLTGEDIPFEAMPA
ASB62313.1       234 TLVVATDSMKNFIQRHLASYEGTTTEGFLYVAHRFLDTYSHMDTIALTGEDIPFEAMPA
AOL30990.1       234 TLVVATDSMKNFIQRHLASYEGTTTEGFLYVAHRFLDTYSHMDTITLTGEDIPFEAMPA
WP_083686476.1   240 SLVVATDSMKNFIQRHLGSYDGSTIEGFLYVAEAFLDKYPQMETVQLTGDEVPFEATNG
OZT14492.1       240 SLVVATDSMKNFIQRHLGSYEGSTIEGFLYVAEAFLDKYPQMETVELIGDEVPFEATNG
WP_094910043.1   240 SLVVATDSMKNFIQRHLGSYDGSTIEGFLYVAEAFLDKYPQMETVELIGDEVPFEATNG
AQX54882.1       240 SLVVATDSMKNFIQRHLGSYDGSTIEGFLYVAEAFLDKYPQMETVQLTGDEVPFEATNG
WP_061784634.1   240 SLVVATDSMKNFIQRHLGSYDGSTIEGFLYVAEAFLDKYPQMETVQLTGDEVPFEATNG
WP_078989772.1   240 SLVVATDSMKNFIQRHLGSYDGSTIEGFLYVAEAFLDKYPQMETVQLTGDEVPFEATNG
WP_040342081.1   238 SLVVATDSMKNFIQKHLASYTGTTIEGFLEYVATSFLKKYSHIEKISLIGEDIPFETTFA
WP_101224285.1   239 SMVVATDSMKNFIQRHLATFKGATLEGFASYVSEAFLNKYPQIDTVKLIAEDIPFEAVTE
ASS93773.1       239 SMVVATDSMKNFIQQHLATFKGATLEGFASYVSEAFLNKYPQIDTVKLIAEDIPFEAVTE
WP_063232385.1   239 SMVVATDSMKNFIQQHLATFKGATLEGFASYVSEAFLNKYPQIDTVKLIAEDIPFEAVTE
WP_061143228.1   239 SMVLATDSMKNFIQRHLATFTGTTLEGFASYVSEAFLNKYPQIDTVKLMAEDIPFEAVTE
PJN86603.1       239 SMVLATDSMKNFIQRHLATFTGTTLEGFASYVSEAFLNKYPQIDTVKLMAEDIPFEAVTE
WP_048623468.1   238 SLVVATDSMKNFIQKHLASYTGTTIEGFLEYVATSFLKKYSHIEKISLIGEDIPFETTFA
BAB20808.1       238 SLVVATDSMKNFIQKHLASYTGTTIEGFLEYVATSFLKKYSHIEKISLIGEDIPFETTFA
consensus        241 ....*******.*.........*.*.***..... *....*...... .......*... .


3WLVA            121 VK-NGNRAASELVFKKSRNEYATAYLNVVRNEDNTLNITEQQSGIAGLQLIKVSGNSFVG
KGM46460.1       124 PTANGSFEPSPLVFRYSLNEQAGASIEVKRENDDII-TTLHLSTLKGLKLIKVKGSSFYG
WP_045524647.1   124 RT-NESFEPSPLVFRYSLNEQAGTTIEVKREKDEII-TSHHISSLKGLKLIKVKGSSFYG
WP_034672575.1   124 PTANGSFEPSPLVFRYSLNEQAGASIEVKRENDDII-TTLHLSTLKGLKLIKVKGSSFYG
AOM84027.1       125 GS-DAGVQESETVFREGVIEKPGVTLSTGRSESGEAKLNALECSVHDLHLIKVKGSSFAG
1J2GA            128 VK-NGNRAASELVFKKSRNEYATAYLNVVRNEDNTLNITEQQSGIAGLQLIKVSGNSFVG
ACR09749.1       122 GT-DKGVVTSDLVFRKSRNEYATATVEVARTASGTE-VVEQASGIADIQLIKVSGSSFYG
4R8XA            124 GT-DKGVVTSDLVFRKSRNEYVTATVEVARTASGTE-VVEQASGIADIQLIKVSGSSFYG
SPT78254.1       126 AN-GEEPSTSTLVYKHSRNERNEASIELVREGNGWR-INRQNSALLDLQLVKVKDNSFVG
KKI85158.1       126 AN-GEESSTSTLVYKNSRNERNEASIELVREGNGWQ-INSQNSRLLDLQLVKVKDNSFVG
OXS68986.1       128 FE-DEEIKKSPLVFNHSRNEKAVSSLHIIRHGQDIN-VESRESSICDLQLIKVSGNSFVG
KOR85772.1       129 PK-GEGHENSDVVYRCSRNERATTTIEVKRLPTGSK-IVKHSSGIVDVQLIKVKGSSFYG
WP_098373266.1   129 PK-GEGHENSDVVFRCSRNERATTTIEVKRLPTGSK-IVKHSSGIVDVQLIKVKGSSFYG
PEZ74426.1       129 PK-GEGHENSDVVFRCSRNERATTTIEVKRLPTGSK-IVKHSSGIVDVQLIKVKGSSFYG
PCD05853.1       129 PK-GEGHEKSDVVYRCSRNERATTTIEVKRLPTGSK-IVKHSSGIVDVHLIKVKGSSFYG
PAL09042.1       129 PK-GKGHENSDVVYRCSRNERATTTIEVKRLPTGSK-IVKHSSGIVDVHLIKVKGSSFYG
CEG34811.1       129 PK-GEGHENSDVVYRCSRNERATTTIEVKRLPTGSK-IVKHSSGIVDVHLIKVKGSSFYG
5AYJA            127 VK-NGNRAASELVFKKSRNEYATAYLNVVRNEDNTLNITEQQSGIAGLQLIKVSGNSFYG
BAA08723.1       128 VK-NGNRAASELVFKKSRNEYATAYLNVVRNEDNTLNITEQQSGIAGLQLIKVSGNSFVG
AKO95039.1       136 FE-DEEIKKSPLVFNHSRNEKAVSSLHIIRHGQDIN-VESRESSICDLQLIKVSGNSFVG
4R99A            124 GT-DKGVVTSDLVFRKSRNEYVTATVEVARTASGTE-VVEQASGIADIQLIKVSGSSFYG
BAD66267.1       143 AN-GEESSTSSLVYKHSRNERNEASIELVREGNGWQ-INSQNSRLLDLQLVKVKDNSFVG
WP_095326636.1   143 AN-GEESSTSSLVYKHSRNERNEASIELVREGNGWQ-INSQNSRLLDLQLVKVKDNSFVG
WP_095294289.1   143 AN-GEESSTSTLVYKHSRNERNEASIELVREGNGWQ-INSQNSRLLDLQLVKVKDNSFVG
WP_095236414.1   143 AN-GEESSTSTLVYKHSRNERNEASIELVREGNGWQ-INSQNSRLLDLQLVKVKDNSFVG
PAF09838.1       143 AN-GEESSTSTLVYKNSRNERNEASIELVREGNGWQ-INSQNSRLLDLQLVKVKDNSFVG
PAE88988.1       143 AN-GEESSTSTLVYKHSRNERNEASIELVREGNGWQ-INSQNSRLLDLQLVKVKDNSFVG
PAD14932.1       143 AN-GEEPSTSTLVYKHSRNERNEASIELVREGNGWR-INRQNSALLDLQLVKVKDNSFVG
WP_081496159.1   168 FE-DEEIKKSPLVFNHSRNEKAVSSLHIIRHGQDIN-VESRESSICDLQLIKVSGNSFVG
OWV36502.1       294 YE-EQELGTSHLVFRRSRNERARSVLKAERIGDTIT-IKEQYSEIMDLQLVKVSGNSFVG
EXF55358.1       294 YE-EKELSTSRLVFRRSRNERSRSVLKAERSGNTIT-ITEQYSEIMDLQLIKVSGNSFVG
CCU60286.1       294 YE-EKELSTSRLVFRRSRNERSRSVLKAERSGNTIT-ITEQYSEIMDLQLIKVSGNSFVG
AII36988.1       294 YE-EKELSTSRLVFRRSRNERSRSVLKAERSGNTIT-ITEQYSEIMDLQLVKVSGNSFVG
APH66064.1       294 YE-EQELGTSHLVFRRSRNERARSVLKAERIGDTIT-IKEQYSEIMDLQLVKVSGNSFVG
BAM59327.1       294 YE-EKELSTSRLVFRRSRNERSRSVLKAERSGNTIT-ITEQYSEIMDLQLVKVSGNSFVG
WP_101501434.1   294 YE-EKELSTSRLVFRRSRNERSRSVLKAERSGNTIT-ITEQYSEIMDLQLVKVSGNSFVG
AKC48817.1       294 YE-EKELSTSRLVFRRSRNERSRSVLKAERSGNTIT-ITEQYSEIMDLQLVKVSGNSFYG
PAD64173.1       294 YE-AQELRTSRLVFRRSRNERARSVLKAERGETIT-ITEQYSEMMDLQLVKVSGNSFVG
WP_095241385.1   294 YE-AQELRTSRLVFRRSRNERARSVLKAERGETIT-ITEQYSEMMDLQLVKVSGNSFVG
WP_045926035.1   294 YE-AQELRTSRLVFRRSRNERARSVLKAERIEDTIT-IKEQYSEIIDLQLVKVSGNSFVG
AJW84706.1       294 YE-GQELGTSHLVFRRSRNERARSVLKAERIGDTIT-ITEQYSEIMDLQLVKVSGNSFVG
KFK78955.1       294 YE-GQELGTSHLVFRRSRNERARSVLKAERIGDTIT-ITEQYSEIMDLQLVKVSGNSFVG
AXJ21641.1       294 YE-EQELGTSHLVFRRSRNERARSVLKAERIGDTIT-IKEQYSEIMDLQLIKVSGNSFVG
WP_003222862.1   294 YE-GQELGTSHLVFRRSRNERARSVLKAERIGDTIT-ITEQYSEIMDLQLVKVSGNSFVG
EME07777.1       294 YE-EKELSTSRLVFRRSRNERSRSVLKAERSGNTIT-ITEQYSEIMDLQLVKVSGNSFVG
AUZ27736.1       294 YE-EQGLGTSQLVFRRSRNERARSVLKAARIGDTIT-ITEQYSEITDLQLVKVSGNSFVG
KUP29050.1       294 YD-DQELGTSQLVFRRSRNERARSVLKAERIGDTIT-LKEQYSEITDLQLVKVSGNSFVG
PRS06588.1       294 YE-DQELGTSQLVFRRSCNERARSVLKAERMGDTIT-LKEQYSEITDLQLVKVSGNSFVG
PRP51591.1       294 YE-DHELGTSQLVFRRSRNERARSVLKAERMGDTIT-LKEQYSEITDLQLVKVSGNSFVG
```

210

```
WP_099043576.1  294  YD-DQEIGTSQLVFRRSRNERARSVLKAERIGDTIT-IKEQYSEITDLQLVKVSGNSFVG
OLQ49074.1      294  YE-EQGIGTSQLVFRRSRNERARSVLKAARIGDTIT-ITEQYSEITDLQLVKVSGNSFVG
WP_075749098.1  294  YE-EQGIGTSQLVFRRSRNERARSVLKAARIGDTIT-ITEQYSEITDLQLVKVSGNSFVG
AFI29791.1      294  YE-EKEITASRLVFRRSRNERSRSVLKAERSGNTIT-ITEQYSEIMDLQLVKVSGNSFAG
WP_014665258.1  294  YE-EKEITASRLVFRRSRNERSRSVLKAERSGNTIT-ITEQYSEIMDLQLVKVSGNSFAG
ASB62313.1      294  YE-EKEITASRLVFRRSRNERSRSVLKAERSGNTIT-ITEQYSEIMDLQLVKVSGNSFAG
AOL30990.1      294  YE-EKEISTSRLVFRRSRNERSRSVLKAERSGNTIT-ITEQYSEIMDLQLVKVSGNSFVG
WP_083686476.1  300  MV-GNTLTESKLVYKRSRNEYAQAGIKIERIVQGQQ-ITEQYSKLKDLQLIKVEGNSFVG
OZT14492.1      300  MV-GNTLTESKLVYKRSRNEYAQAGIKIERTAQGQQ-ITEQYSKLKDLQLIKVEGNSFVG
WP_094910043.1  300  MV-GNTLTESKLVYKRSRNEYAQAGIKIERTAQGQQ-ITEQYSKLKDLQLIKVEGNSFVG
AQX54882.1      300  MV-GNTLTESKLVYKRSRNEYAQAGIKIERIVQGQQ-ITEQYSKLKDLQLIKVEGNSFVG
WP_061784634.1  300  MV-GNTLTESKLVYKRSRNEYAQAGIKIERIVQGQQ-ITEQYSKLKDLQLIKVEGNSFVG
WP_078989772.1  300  MV-GNTLTESKLVYKRSRNEYAQAGIKIERIVQGQQ-ITEQYSKLKDLQLIKVEGNSFVG
WP_040342081.1  298  VK-NGNRAASELVFKKSRNEYATAYLNVVRNEDNTLNITEQQSGLADLQLIKVSGNSFVG
WP_101224285.1  299  AT-DLQLKPSDLVFKKSRNERAKAAVEIIRRENGSE-IIQQSSSIHDLQLIKVSGNSFVG
ASS93773.1      299  AT-DPQLKPSDLVFKKSRNERANAAVEIIRGENGSE-IVQQSSSILDLQLIKVSGNSFVG
WP_063232385.1  299  AT-DPQLKPSDLVFKKSRNERANAAVEIIRGENGSE-IVQQSSSILDLQLIKVSGNSFVG
WP_061143228.1  299  AT-GTPIKPSDLVFKKSRNERAQSSLEIIRGENGNE-ITQQCSSILDLQLIKVSGNSFVG
PJN86603.1      299  AT-GTPIKQSDLVFKKSRNERAQSSLETIRGENGNE-ITQQCSSILDLQLIKVSGNSFVG
WP_048623468.1  298  VK-NGNRAASELVFKKSRNEYATAYLNVVRNEDNTLNITEQQSGLADLQLIKVSGNSFVG
BAB20808.1      298  VK-NGNRAASELVFKKSRNEYATAYLNVVRNEDNTLNITEQQSGLAGLQLIKVSGNSFVG
consensus       301   ...*.*.........*.. ... *..    ....  .....*.**...**.*


3WLVA           180  FIRDEYTTLPEDSNRPLFVYLNIKWKYKNTEDSFGTNPENYVAAEQIRDIATSVFHETET
KGM46460.1      183  YVKDEYTTLPESFDRPLFIFLNIDWRYDDLEDARGNTSDGYVAAEQVRDIAYTVFHEENS
WP_045524647.1  182  YFKDEYTTLPESYDRPLFIFLNIDWRYNTDDARGNSSDGYVAAEQVRDIAYSVFHEQNS
WP_034672575.1  183  YVKDEYTTLPESFDRPLFIFLNIDWRYDDLEDARGNTSDGYVAAEQVRDIAYTVFHEENS
AOM84027.1      184  FIRDEYTKLPETKDRPLFTYLNISWTYKNMDAFGDQPELYVAAEQVKHIAQTIFHQEAS
1J2GA           187  FIRDEYTTLPEDSNRPLFVYLNIKWKYKNTEDSFGTNPENYVAAEQIRDIATSVFHETET
ACR09749.1      180  YIIDEYTTLAEATDRPLYIFLNIGWAYENQDDAKGDNPANYVAAEQVRDLAASVFHTLDN
4R8XA           182  YIIDEYTTLAEATDRPLYIFLNIGWAYENQDDAKGDNPANYVAAEQVRDLAASVFHTLDN
SPT78254.1      184  FIRDQYTTLPEDSNRPLFYYLNIGWSYETDDDALGEEPRYVAGEQVADLASSVFHELAS
KKI85158.1      184  FIRDQYTTLPEDSNRPLFTYLNIGWSYETDDDALGETPARYVADEQVADLASSVFHELAS
OXS68986.1      186  FIRDQYTTLPEDSNRPLFTFLNIGWNYENSKDALGWNYENSKDAL
KOR85772.1      187  FIQDEYTTLPEAQDRPLFTYLDFDWEYSRWEDGTGADPEIYVAAEQICDIANTVFHELNN
WP_098373266.1  187  FIQDEYTTLPEAQDRPLFTYLDFDWEYSRWEDATGSNPEKYVAAEQISDIANTVFHELNN
PEZ74426.1      187  FIQDEYTTLPEAQDRPLFTYLDFDWEYSRWEDATGSNPEKYVAAEQISDIANTVFHELNN
PCD05853.1      187  FIQDEYTTLPEAQDRPLFTYLDFDWEYSRWEDGTGADPEKYVAAEQICDIANTVFHELNN
PAL09042.1      187  FIQDEYTTLPEAQDRPLFTYLDFDWEYSRWEDGTGADPEKYVAAEQICDIANTVFHELNN
CEG34811.1      187  FIQDEYTTLPEAQDRPLFTYLDFDWEYSRWEDGTGADPENYVAAEQICDIANTAFHELNN
5AYJA           186  FIRDEYTTLPEDSNRPLFVYLNIKWKYKNTEDSFGTNPENYVAAEQIRDIATSVFHETET
BAA08723.1      187  FIRDEYTTLPEDSNRPLFVYLNIKWKYKNTEDSFGTNPENYVAAEQIRDIATSVFHETET
AKO95039.1      194  FIRDEYTTLPEDGNRPLFIFLNIGWNYENSKDALGETPSLYVASEQVHDIASSVFHELT
4R99A           182  YIIDEYTTLAEATDRPLYIFLNIGWAYENQDDAKGDNPANYVAAEQVRDLAASVFHTLDN
BAD66267.1      201  FIRDQYTTLPEDSNRPLFTYLNIGWSYETDDDALGEEPARYVAGEQVADLASSVFHELTS
WP_095326636.1  201  FIRDQYTTLPEDSNRPLFTYLNIGWSYETDDDALGEEPARYVAGEQVADLASSVFHELAS
WP_095294289.1  201  FIRDQYTTLPEDSNRPLFTYLNIGWSYETNDDALGEEPARYVAGEQVADLASSVFHELTS
WP_095236414.1  201  FIRDQYTTLPEDSNRPLFTYLNIGWSYETNDDALGEEPARYVAGEQVADLASSVFHELAS
PAF09838.1      201  FIRDQYTTLPEDSNRPLFTYLNIGWSYETDDDALGEEPARYVAGEQVADLASSVFHELAS
PAE88988.1      201  FIRDQYTTLPEDSNRPLFTYLNIGWSYETDDDALGEEPARYVAGEQVADLASSVFHELAS
PAD14932.1      201  FIRDQYTTLPEDSNRPLFTYLNIGWSYETDDDALGEEPRYVAGEQVADLASSVFHELAS
WP_081496159.1  226  FIRDEYTTLPEDGNRPLFIFLSIGWNYENSKDALGETPSLYVASEQVHDLASSVFHELET
OWV36502.1      352  FIRDEYTTLPEDGNRPLFVYLNISWQYENTDDARATDPARYVAAEQVRDLASTVFHELET
EXF55358.1      352  FIRDEYTTLPEDGNRPLFVYLNISWQYENTNDAYAFDPARYVAAEQVRDLASTVFHELET
CCU60286.1      352  FIRDEYTTLPEDGNRPLFVYLNISWQYENTNDAYAFDPSRYVAAEQVRDLASTVFHELET
AII36988.1      352  FIRDEYTTLPEDGNRPLFVYLNISWQYENTNDAYAFDPARYVAAEQVRDLASTVFHELET
APH66064.1      352  FIRDEYTTLPEDGNRPLFVYLNISWQYENTDDARATDPARYVAAEQVRDLASTVFHELET
BAM59327.1      352  FIRDEYTTLPEDGNRPLFVYLNISWQYENTNDSYASDPARYVAAEQVRDLASTVFHELET
WP_101501434.1  352  FIRDEYTTLPEDGNRPLFVYLNISWQYENTNDAYAFDPSRYVAAEQVRDLASTVFHELET
AKC48817.1      352  FIRDEYTTLPEDGNRPLFVYLNISWQYENTNDSYASDPARYVAAEQVRDLASTVFHELET
PAD64173.1      352  FIRDEYTTLPEDGNRPLFVHLNIGWHYENTNDAYASDPARYVAAEQVRDLASAVFHELET
WP_095241385.1  352  FIRDEYTTLPEDGNRPLFVHLNIGWHYENTNDAYASDPARYVAAEQVRDLASAVFHELET
WP_045926035.1  352  FIRDEYTTLPEDGNRPLFVHLNIGWHYENTNDAYASDPARYVAAEQVRDLASAVFHELET
AJW84706.1      352  FIRDEYTTLPEDGNRPLFVYLNICWQYENTDDARASDPARYVAAEQVRDLASTVFHELKT
KFK78955.1      352  FIRDEYTTLPEDGNRPLFVYLNICWQYENTDDARASDPARYVAAEQVRDLASTVFHELKT
AXJ21641.1      352  FIRDEYTTLPEDGNRPLFVYLNISWQYENTDDARATDPARYVAAEQVRDLASTVFHELET
WP_003222862.1  352  FIRDEYTTLPEDGNRPLFVYLNICWQYENTDDARASDPARYVAAEQVRDLASTVFHELKT
EME07777.1      352  FIRDEYTTLPEDGNRPLFVYLNISWQYENTNDSYASDPARYVAAEQVRDLASTVFHELET
AUZ27736.1      352  FIRDEYTTLPEDGNRPLFVYLNIGWHYENTNDAYASDPARYVAAEQVRDLASTVFHELET
KUP29050.1      352  FIRDEYTTLPEDGNRPLFVYLNISWRYEHADDAYAADPARYVAAEQVRDLASTVFHELKT
PRS06588.1      352  FIRDEYTTLPEDGNRPLFVYLNISWRYEHTEDAYAADPARYVAAEQVRDLASTVFHELET
```

```
PRP51591.1        352  FIRDEYTTLPEDGNRPLFVYLNISWRYEHADDAYAADPARYVAAEQVRDLASTVFHELET
WP_099043576.1    352  FIRDEYTTLPEDGNRPLFVYLNISWRYEHADDAYAADPARYVAAEQVRDLASTVFHELKT
OLQ49074.1        352  FIRDEYTTLPEDGNRPLFVYLNIGWHYENTNDAYTSDPARYVAAEQVRDLASTVFHELET
WP_075749098.1    352  FIRDEYTTLPEDGNRPLFVYLNIGWHYENTNDAYTSDPARYVAAEQVRDLASTVFHELET
AFI29791.1        352  FIRDEYTTLPEDGNRPLFVYLNISWQENTNDAYASDPARYVAAEQVRDLASTVFHELET
WP_014665258.1    352  FIRDEYTTLPEDGNRPLFVYLNISWQENTNDAYASDPARYVAAEQVRDLASTVFHELET
ASB62313.1        352  FIRDEYTTLPEDGNRPLFVYLNISWQYENTNDAYASEPARYVAAEQVRDLASTVFHELET
AOL30990.1        352  FIRDEYTTLPEDGNRPLFVYLNISWQYENTNDAYAFDPARYVAAEQVRDLASTVFHELET
WP_083686476.1    358  FVRDEYTTLPEDSNRPLFVYLNIGWTYTQPEDAIGDAPLSYVAAEQVKDIACSVFNETET
OZT14492.1        358  FVRDEYTTLPEDSNRPLFVYLNIGWTYTQSEDAIGDAPLSYVAAEQVRDIACSVFNETET
WP_094910043.1    358  FVRDEYTTLPEDSNRPLFVYLNIGWTYTQSEDAIGDAPLSYVAAEQVKDIACSVFNETET
AQX54882.1        358  FVRDEYTTLPEDSNRPLFVYLNIGWTYTQPEDAIGDAPLSYVAAEQVKDIACSVFNETET
WP_061784634.1    358  FVRDEYTTLPEDSNRPLFVYLNIGWTYTQPEDAIGDAPLSYVAAEQVKDIACSVFNETET
WP_078989772.1    358  FVRDEYTTLPEDSNRPLFVYLNIGWTYTQPEDAIGDAPLSYVAAEQVKDIACSVFNETET
WP_040342081.1    357  FIRDEYTTLPEDTNRPLFVYLNIKWKYKNTEDSFGDNPEYYVAAEQIRDIATSVFHETET
WP_101224285.1    357  FVRDEYTTLPEDGNRPLFTYLNLHWVYEDQKDAFGVDPSKYVAAEQVIDIATSVFHEMET
ASS93773.1        357  FVRDEYTTLPEDGNRPLFTYLNLHWVYEDQKDAFGVDPSKYVAAEQVIDIATSIFHEMET
WP_063232385.1    357  FVRDEYTTLPEDGNRPLFTYLNLHWVYEDQKDAFGVDPSKYVAAEQVIDIATSIFHEMET
WP_061143228.1    357  FVRDEYTTLPEDGNRPLFTYLNLHWIYEDQKDAFGTDPSKYVAAEQVIDIATSVFHELET
PJN86603.1        357  FVRDEYTTLPEDGNRPLFTYLNLHWIYEDQKDAFGTDPSKYVAAEQVTDIATSVFHELET
WP_048623468.1    357  FIRDEYTTLPEDSNRPLFVYLNIKWKYKNTEDSFGMNPENYVAAEQIRDIATSVFHETET
BAB20808.1        357  FIRDEYTTLPEDSNRPLFVYLNIKWKYKNTEDSFGTNPENYVAAEQIRDIATSVFHETET
consensus         361  ...*.**.*.*....***...*.. * *.....*. . ....***.**.....*.*.....
```

```
3WLVA             240  LSIQHLIYLIGRRILERFPQLEVYFESQNHTWDKIVEEIPES-EGKVYTEPRPPYGFQC
KGM46460.1        243  PSIQNLIYRIGTRILERFPQLAEVRFESNNRTWETILDEIPASQEGKVFTEPRPPYGFQC
WP_045524647.1    242  PSIQNLIYRIGTRILERFPQLLEVRFESNNRTWETILDEIPDSLEGKVFTEPRPPYGFQC
WP_034672575.1    243  PSIQNLIYRIGTRILERFPQLAEVRFESNNRTWETILDEIPASQEGKVFTEPRPPYGFQC
AOM84027.1        244  PSIQKLIYDIGIRVMQRFPQLETVSFESNNRTWETIRDEIPASKEGKVFTDPRPPYGFQK
1J2GA             247  LSIQHLIYLIGRRILERFPQLQEVYFESQNHTWDKIVEEIPES-EGKVYTEPRPPYGFQC
ACR09749.1        240  KSIQHLIYHIGLTILDRFPQLTEVNFGTNNRTWDTVVEGTDCF-KGAVFTEPRPPFGFQG
4R8XA             242  KSIQHLIYHIGLTILDRFPQLTEVNFGTNNRTWDTVVEGTDCF-KGAVFTEPRPPFGFQG
SPT78254.1        244  PSIQHLVYQIGCRMLKRFPQLQEVTFESQNRTWDTVVEDIPET-EGKVYTEPRLPFGFQR
KKI85158.1        244  PSIQHLVYQIGCRMLKRFPQLQEVTFESQNRTWDTVVEDIPET-EGKVYTEPRLPFGFQR
OXS68986.1        246  PSIQNLIYQIGCRILERFPQLQSVNFQSQNHTWETVVEEIPGE-KGKVYTEPKPPFGFQL
KOR85772.1        247  RSIQQLIYHIGTRILERFPQLANVQFKTNNRTWETVVETIPNS-EGSVYQEPRPPFGFQG
WP_098373266.1    247  RSIQQLIYHIGIRILERFPQLANVQFKTNNRTWETVVDTIPNS-EGSVYQEPRPPFGFQS
PEZ74426.1        247  RSIQQLIYHIGIRILERFPQLANVQFKTNNRTWETVVDTIPNS-EGSVYQEPRPPFGFQS
PCD05853.1        247  RSIQQLIYHIGIRILERFPQLANVQFKTNNRTWETVVETIPNS-EGSVYQEPRPPFGFQG
PAL09042.1        247  RSIQQLIYHIGIRILERFPQLANVQFKTNNRTWETVVETIPNS-EGSVYQEPRPPYGFQG
CEG34811.1        247  RSIQQLIYHIGIRILERFPQLANVQFKTNNRTWETVVETIPNS-EGSVYQEPRPPYGFQG
5AYJA             246  LSIQHLIYLIGRRILERFPQLQEVYFESQNHTWDKIVEEIPES-EGKVYTEPCPPYGFQC
BAA08723.1        247  LSIQHLIYLIGRRILERFPQLQEVYFESQNHTWDKIVEEIPES-EGKVYTEPRPPYGFQC
AKO95039.1        254  PSIQNLIYQIGCRILERFPQLQSVNFQSQNHTWETVVEEIPGE-KGKVYTEPKPPFGFQL
4R99A             242  KSIQHLIYHIGLTILDRFPQLTEVNFGTNNRTWDTVVEGTDCF-KGAVFTEPRPPFGFQG
BAD66267.1        261  PSIQHLIYQIGCRMKRFPQLQEVTFESQNRTWDTVVEDIPET-EGKVYTEPRLPFGFQR
WP_095326636.1    261  PSIQHLIYQIGCRMKRFPQLQEVTFESQNRTWDTVVEDIPET-EGKVYTEPRLPFGFQR
WP_095294289.1    261  PSIQHLIYQIGCRMKRFPQLQEVTFESQNRTWDTVVEDIPET-EGKVYTEPRLPFGFQR
WP_095236414.1    261  PSIQHLIYQIGCRMKRFPQLQEVTFESQNRTWDTVVEDIPET-EGKVYTEPRLPFGFQR
PAF09838.1        261  PSIQHLIYQIGCRMKRFPQLQEVTFESQNRTWDTVVEDIPET-EGKVYTEPRLPFGFQR
PAE88988.1        261  PSIQHLIYQIGCRMKRFPQLQEVTFESQNRTWDTVVEDIPET-EGKVYTEPRLPFGFQR
PAD14932.1        261  PSIQHLVYQIGCRMKRFPQLQEVTFESQNRTWDTVVEDIPET-EGKVYTEPRLPFGFQR
WP_081496159.1    286  PSIQNLIHQIGCRILERFPQLQSVNFQSQNHTWETVVEEIPGE-KGKVYTEPKPPFGFQL
OWV36502.1        412  PSIQNLIHHIGCRILTRFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQR
EXF55358.1        412  PSIQNLIYHIGCRILARFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPYGFQH
CCU60286.1        412  PSIQNLIYHIGCRILARFPQLTDVSFQSQNHTWDAVVEEIPGS-QGKVYTEPRPPYGFQH
AII36988.1        412  PSIQNLIYHIGCRILARFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPYGFQH
APH66064.1        412  PSIQNLIHHIGCRILTTFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQH
BAM59327.1        412  PSIQNLIYHIGCRILARFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPYGFQH
WP_101501434.1    412  PSIQNLIYHIGCRILARFPQLTDVSFQSQNHTWDTVVEEIPGS-QGKVYTEPRPPYGFQH
AKC48817.1        412  PSIQNLIYHIGCRILARFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPYGFQR
PAD64173.1        412  PSIQNLIYHIGCRILTRFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQR
WP_095241385.1    412  PSIQNLIYHIGCRILTRFPQLTGVTFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQR
WP_045926035.1    412  PSIQNLIYHIGCRILTRFPQLTEVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQR
AJW84706.1        412  PSIQNLIYHIGCRILTRFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQR
KFK78955.1        412  PSIQNLIYHIGCRILTRFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQR
AXJ21641.1        412  PSIQNLIHHIGCRILTRFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQR
WP_003222862.1    412  PSIQNLIYHIGCRILTRFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQR
EME07777.1        412  PSIQNLIYHIGCRILARFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPYGFQR
AUZ27736.1        412  PSIQNLIYHIGCRILTRFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQR
```

```
KUP29050.1      412  PSIQNLIYHIGCRILTRFPQLADVSFQSQNHTWDTVVEEIPGT-KGKVYTEPRPPFGFQR
PRS06588.1      412  PSIQNLIYHIGCRILTRFPQLADASFQSQNHTWDTVVEEIPGT-KGKVYTEPRPPFGFQR
PRP51591.1      412  PSIQNLIYHIGCRILTRFPQLADVSFQSQNHTWDTVVEEIPGT-KGKVYTEPRPPFGFQR
WP_099043576.1  412  PSIQNLIYHIGCRILTRFPQLADVSFQSQNHTWDTVVEEIPGT-KGKVYTEPRPPFGFQR
OLQ49074.1      412  PSIQNLIYHIGCRILTRFPQLTDISFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQR
WP_075749098.1  412  PSIQNLIYHIGCRILTRFPQLTDISFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQR
AFI29791.1      412  PSIQHLIYHIGCRILARFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQR
WP_014665258.1  412  PSIQHLIYHIGCRIIARFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQR
ASB62313.1      412  PSIQHLIYHIGCRILARFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPFGFQR
AOL30990.1      412  PSIQNLIYHIGCRILARFPQLTDVSFQSQNHTWDTVVEEIPGS-KGKVYTEPRPPYGFQY
WP_083686476.1  418  PSIQNLIYLIGIRVLERFPQLKDVTFESQNHTWDAVVEDIPNS-DGKVYTEPKPPFGFQV
OZT14492.1      418  PSIQNLIYLIGIRVLERFPQLKDVTFESQNHTWDSVVEDIPNS-NGKVYTEPKPPFGFQV
WP_094910043.1  418  PSIQNLIYLIGIRVLERFPQLKDVTFESQNHTWDSVVEDIPNS-NGKVYTEPKPPFGFQV
AQX54882.1      418  PSIQNLIYLIGIRVLERFPQLKDVTFESQNHTWDAVVEDIPNS-DGKVYTEPKPPFGFQV
WP_061784634.1  418  PSIQNLIYLIGIRVLERFPQLKDVTFESQNHTWDAVVEDIPNS-DGKVYTEPKPPFGFQV
WP_078989772.1  418  PSIQNLIYLIGIRVLERFPQLKDVTFESQNHTWDAVVEDIPNS-DGKVYTEPKPPFGFQV
WP_040342081.1  417  LSIQHLIYLIGCRILERFPQLQEVNFESQNHTWDKIVEEIPGS-QGKVYTEPRPPYGFQC
WP_101224285.1  417  PSIQNLIYEIGCRILTRFPQLLEVTFESQNHTWDTVVSEIPES-KGKVYTEPRPPYGFQV
ASS93773.1      417  PSIQNLIYEIGCRILTRFPQLLEVTFESQNHTWDTVVSEIPES-KGKVYTEPRPPYGFQV
WP_063232385.1  417  PSIQNLIYEIGCRILTRFPQLLEVTFESQNHTWDTVVSEIPES-KGKVYTEPRPPYGFQV
WP_061143228.1  417  PSIQNLIFEIGCRILTRFPQLLEVTFESQNHTWDTVVPIISDS-MGKVYTEPRPPFGFQV
PJN86603.1      417  PSIQNLIFEIGCRILTRFPQLLEVTFESQNHTWDTVVPIISDS-MGKVYTEPRPPFGFQV
WP_048623468.1  417  LSIQHLIYLIGRRILERFPQLQEVYFESQNHTWDKIVEEIPES-KGKVYTEPRPPYGFQC
BAB20808.1      417  LSIQHLIYLIGRRILERFPQLQEVYFESQNHTWDKIVEEIPES-EGKVYTEPRPPYGFQC
consensus       421  .****.*...**.... .**** ...* ..*.**......... .*.*...*..*.***.

3WLVA           299  FTVTQEDLPHENIL--------------------
KGM46460.1      303  FSMTRADLEAKEEPNK-------------------
WP_045524647.1  302  FSMTRDDLESEGESKK-------------------
WP_034672575.1  303  FSMTRADLEAKEEPNK-------------------
AOM84027.1      304  FTVLQEDVRREEGRS--------------------
1J2GA           306  FTVTQEDLPHENIL--------------------
ACR09749.1      299  FSVHQEDLAREKASANSEYVAL-------------
4R8XA           301  FSVHQEDLAREKASANSEYVAL-------------
SPT78254.1      303  FSVTRADLATQTTSNRTETTRL-------------
KKI85158.1      303  FSVTRADLGPQTTSARTEKARL-------------
OXS68986.1      305  FSVTREDLKTNSSKEKKANLK--------------
KOR85772.1      306  FVVTQEDVKQKKAADSTVGASR-------------
WP_098373266.1  306  FVVTQEDVKQKKAADSTVGVSR-------------
PEZ74426.1      306  FVVTQEDVKQKKAADSTVGVSR-------------
PCD05853.1      306  FVVTQEDVKQKKAADSTVGASR-------------
PAL09042.1      306  FVVTQEDVKQKKAADSTVGASR-------------
CEG34811.1      306  FVVTQEDVKQKKAADSTVGASR-------------
5AYJA           305  FTVTQEDLPHENILMFSDEPDHKGALK--------
BAA08723.1      306  FTVTQEDLPHENILMFSDEPDHKGALK--------
AKO95039.1      313  FSVTREDLKTNSSKEKKANLK--------------
4R99A           301  FSVHQEDLAREKASANSEYVALKLAAALKHHHHHH
BAD66267.1      320  FSVTRADLGPQTTSARTEKARL-------------
WP_095326636.1  320  FSVTRADLGPQTTFTRTETARL-------------
WP_095294289.1  320  FSVTRADLGPQTTFTRTETARL-------------
WP_095236414.1  320  FSVTRADLGPQTTFTRTETARL-------------
PAF09838.1      320  FSVTRADLGPQTTSARTEKARL-------------
PAE88988.1      320  FSVTRADLGPQTTFTRTETARL-------------
PAD14932.1      320  FSVTRADLATQTTSNRTETTRL-------------
WP_081496159.1  345  FSVTREDLKTNSSKEKKANLK--------------
OWV36502.1      471  FTVTREDTEKEKQKAAE--GSLKA-----------
EXF55358.1      471  FTVTREDAEKEKQKAAEKCRSLKA-----------
CCU60286.1      471  FTVTREDAEKEKQKAAEKCRSLKA-----------
AII36988.1      471  FTVTREDAEKEKQKAAEKCRSLKA-----------
APH66064.1      471  FTVTREDAEKEKQKAAEALGSLKA-----------
BAM59327.1      471  FTVTREDAEKEKQKAAEKCRSLKA-----------
WP_101501434.1  471  FTVTREDAEKEKQKAAEKCRSLKA-----------
AKC48817.1      471  FTVTREDAEKEKQKAAEKCRSLKA-----------
PAD64173.1      471  FTVTREDAEKEKQKADEALGSLKA-----------
WP_095241385.1  471  FTVTREDAEKEKQKADEALGSLKA-----------
WP_045926035.1  471  FTVTREDAEKEKQKADEALGGLKA-----------
AJW84706.1      471  FTVTREDAEKEKQKAVEKLGSLKA-----------
KFK78955.1      471  FTVTREDAEKEKQKAVEKLGSLKA-----------
AXJ21641.1      471  FTVTREDAEKEKQKAAEALGSLKA-----------
WP_003222862.1  471  FTVTREDAEKEKQKAVEKLGSLKA-----------
```

213

```
EME07777.1        471 FTVTREDAEKEKQKAAEKCRSLKA-----------
AUZ27736.1        471 FTVTREDAEKEKQKASEALGSLKS-----------
KUP29050.1        471 FTVTREDAEKEKRKAGEALGSLNA-----------
PRS06588.1        471 FTVTREDAEKEKRKAGGALGSLNA-----------
PRP51591.1        471 FTVTREDAEKEKRKAGEALGSLNA-----------
WP_099043576.1    471 FTVTREDAEKEKRKAGEALGSLNA-----------
OLQ49074.1        471 FTVTREDAEKEKQKATEALGSLKS-----------
WP_075749098.1    471 FTVTREDAEKEKQKATEALGSLKS-----------
AFI29791.1        471 FTVTREDAEKEKQKAAETAGSLKA-----------
WP_014665258.1    471 FTVTREDAEKEKQKAAETAGSLKA-----------
ASB62313.1        471 FTVTREDAEKEKQKAAETAGSLKA-----------
AOL30990.1        471 FTVTREDAEKEKQKAAEKCRSLKA-----------
WP_083686476.1    477 FTVTQEDVKIAVTSALEESN---------------
OZT14492.1        477 FTVTQEDVKIAVTSALEESN---------------
WP_094910043.1    477 FTVTQEDVKIAVTSALEESN---------------
AQX54882.1        477 FTVTQEDVKIAVTSALEESN---------------
WP_061784634.1    477 FTVTQEDVKIAVTSALEESN---------------
WP_078989772.1    477 FTVTQEDVKIAVTSALEESN---------------
WP_040342081.1    476 FTVTQEDLQHKNIPMLSAEIQ--------------
WP_101224285.1    476 FTVKKENLENNKILAAAEENIG-------------
ASS93773.1        476 FTVKKENLENNIILAAAEENIG-------------
WP_063232385.1    476 FTVKKENLENNIILAAAEENIG-------------
WP_061143228.1    476 FTVKREHLENDNILAAAKAAGKEWI----------
PJN86603.1        476 FTVKREHLENDNIFAAAKAAGKEWI----------
WP_048623468.1    476 FTVTQEDLPHENILMFSDEPDHKGALK--------
BAB20808.1        476 FTVTQEDLPHENILMFSDEPDHKGALK--------
consensus         481 *...........  .      .
```

Multiple sequence alignment of 70 *Bacillus* species of uricase sequences indicating highly conserved amino acid residues. Asterisks (*) shows strongly conserved amino acid residue. Highlighted area in the above alignment denotes highly conserved sequences. The parts indicated in yellow and red represent the catalytic residues Ala68, Phe179, Arg196, Ile244, Gln244, Asn271, and Gln299 in the uricase sequence, as well as the catalytic traid residues Lys9, Thr69, and Asp70.

Secondary Structure:

```
            *          *          *          *          *          *
Query   1   MITLKSLNALSEKEFTKFLGDTFEHSPWIAEKSAANRPFSSIINLHRCMVNIVSNSSKEEKLTLIRKHPN   70
Helix   1   HHHHHHHHHHHHHHHHH      HHHHHHHHH                HH        HHHHHHHHH        70
Sheet   1   EEEEEEE        EEEEEEE              EEEEEEEEEEEEEEE         EEEE          70
Turns   1            TT        T     T      T     T                   T      T       70
Struc   1   EEEHHHHHHHHHHHEEEEEEETECCHTHHHHHHHHCCTCEEEEEEEEEEEEEEEECTCCHHHHEEEHHCCCC   70


            *          *          *          *          *          *
Query  71   LGDKVEMSEDSIKEQHGAGLKDLTADEYENFISLNRQYMNKFGFPFILAVRGKDKNDIYQSMKTRIHHSE  140
Helix  71   HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH  140
Sheet  71                      EEEE     EEEEEEEEEEEEEEEEEEEEEEEE        EEEEEEEEE       140
Turns  71      T    TT   T              T           T            T   T       T        140
Struc  71   HHHHHHHHTHHHTHHHHHHHHHHHHHHHEHEEEEEEEEEHEEEEEEEEEEEHTHHHTHEEHHHEEEEHHH  140


            *          *          *          *          *          *
Query 141   TAEFDKALSEIHQIALFRLQDKIKIEGEKPMKNKSAAQTLSYGKGNVFAYRTYSNPLTGIKQIPESTFSG  210
Helix 141   HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH       HH          HHHHHHHHHH       210
Sheet 141            EEEEEEEEEEEEEEEE            EEEE       EEEEEEEEEEEEEEEEEEEEEE     210
Turns 141        T              T      T T    TT          TT           T       TT    T 210
Struc 141   HHHHHHHHHEHEEHEEHHHHEHHHHTHHHHHTHHHHEEEECCTTCEEEEEEEEETEEEEEEHHHHTECCT  210


            *          *          *          *          *          *
Query 211   RDHIIFGTNVKVSVGGSSFLPSFTEGDNSMVVATDSMKNFIQQHLATFKGATLEGFASYVSEAFLNKYPQ  280
Helix 211   HHHHHHHHHHH                  HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH H  280
Sheet 211      EEEEEEEEE        EEEEE        EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE  280
Turns 211   T       T          TT         TT T        T        T        T          T    280
Struc 211   THHEEEEEEEEEECCTTCEEEEECCTTCTCEEHHHTHHEEEEHHHHEHEHHEHHHHEEEEHHHHEEEEEEE  280


            *          *          *          *          *          *
Query 281   IDTVKLIAEDIPFEAVTEATDPQLKPSDLVFKKSRNERANAAVEIIRGENGSEIVQQSSSILDLQLIKVS  350
Helix 281   HHHHHHHHHHHHHHHHH       HHHHHHHHHHHHHHHHHHHHHH      HHHHHHHHHHHHHHH      350
Sheet 281   EEEEEEEEEEEEEEEEEEEEEEE       EE          EEE       EEEEEEEEEEEEEEE       350
Turns 281             T          T   TT     T   T          T   T     T T            350
Struc 281   EEEEHHHHTEHHHHHHHHEEETECHTHHEEHHTHHHHHHHHHHEHEHHTCCCTHEEHTEEEEHHEEEEEEC  350


            *          *          *          *          *          *
Query 351   GNSFVGFVRDEYTTLPEDGNRPLFIYLNLHWVYEDQKDAFGVDPSKYVAAEQVIDIATSIFHEMETPSIQ  420
Helix 351      HHHHHHH              HHHHHHHHHHHHHHHHH      HHHHHHHHHHHHHHHHHHHHHHHHH  420
Sheet 351   EEEEEEEEEEEE          EEEEEEEEEEEE                EEEEEEEEEEEEEEEEEEEEEEEE  420
Turns 351   TT        T      T    T            T TT     TT                        T     420
Struc 351   TTEEEEEHHTEEEECCTCCCTCEEEEEEEEEEHHHTHHTHHHHCTTHEHHHEEEHHEEEEHHHHHETEEE  420


            *          *          *          *          *          *
Query 421   NLIYEIGCRILTRFPQLLEVTFESQNHTWDTVVSEIPESKGKVYTEPRPPYGFQVFTVKKENLENNIILA  490
Helix 421   HHHHH        HHHHHHHHHHH       HHHHHHHHHHHH              HHHHHHHHHHHHHHH  490
Sheet 421   EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE        EE       EEEEEEE      EEEEEEE  490
Turns 421                            TT          T T     TT T        TT    T          490
Struc 421   EEEEEEEEEEEEEEEEHHHEEHHHTTEEEEEEEEHHHHTHHTHEECTTCTCEEEEEEEHHHHHHHEEEHHH  490


Query 491   AAEENIG  497
Helix 491   HHHHH    497
Sheet 491            497            Total Residues: H: 374   E: 311   T: 72
Turns 491       T    497               Percent: H: 75.3   E: 62.6   T: 14.5
Struc 491   HHHTHCC  497
```

Figure A II (A) Predicted secondary structure of selected protein uricase *Bacillus simplex* (WP_063232385.1)

215

Figure A II (B) β sheet structure model of uricase *Bacillus simplex* (WP_063232385.1) from Phyre 2 server

Figure A II (C) Result of target- template alignment from SWISS-MODEL
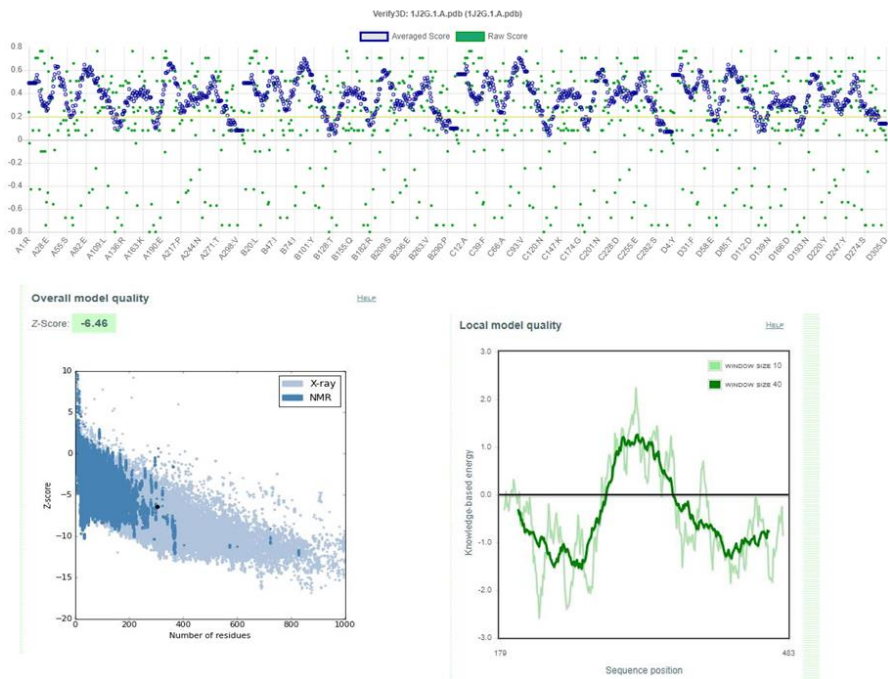
Figure A II (D) Validation results of built model protein of *Bacillus simplex* (WP_063232385.1) from verify 3D and ProSA-web
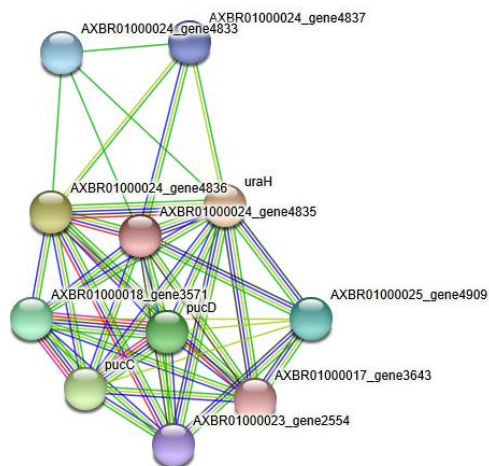
Figure A II (E) STRING server result analysis of protein-protein interaction map for the uricase of *Bacillus simplex* (WP_063232385.1)

Table A II (A) Distribution of motifs in uricase protein sequences

S.NO	Species	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5	Motif 6
1	3WLV_A	+	+	+	+	+	+
2	KGM46460.1	+	+	+	+	-	+
3	WP_045524647.1	+	+	+	+	-	+
4	WP_034672575.1	+	+	+	+	-	+
5	AOM84027.1	+	+	+	+	-	+
6	1J2G_A	+	+	+	+	+	+
7	ACR09749.1	+	+	+	+	+	+
8	4R8X_A	+	+	+	+	+	+
9	SPT78254.1	+	+	+	+	+	+
10	KKI85158.1	+	+	+	+	+	+
11	OXS68986.1	+	+	+	+	+	+
12	KOR85772.1	+	+	+	+	+	+
13	WP_098373266.1	+	+	+	+	+	+
14	PEZ74426.1	+	+	+	+	+	+
15	PCD05853.1	+	+	+	+	+	+
16	PAL09042.1	+	+	+	+	+	+
17	CEG34811.1	+	+	+	+	+	+
18	5AYJ_A	+	+	+	+	+	+
19	BAA08723.1	+	+	+	+	+	+
20	AKO95039.1	+	+	+	+	+	+
21	4R99_A	+	+	+	+	+	+
22	BAD66267.1	+	+	+	+	+	+
23	WP_095326636.1	+	+	+	+	+	+
24	WP_095294289.1	+	+	+	+	+	+
25	WP_095236414.1	+	+	+	+	+	+
26	PAF09838.1	+	+	+	+	+	+
27	PAE88988.1	+	+	+	+	+	+
28	PAD14932.1	+	+	+	+	+	+
29	WP_081496159.1	+	+	+	+	+	+
30	OWV36502.1	+	+	+	+	+	+
31	EXF55358.1	+	+	+	+	+	+
32	CCU60286.1	+	+	+	+	+	+
33	AII36988.1	+	+	+	+	+	+
34	APH66064.1	+	+	+	+	+	+
35	BAM59327.1	+	+	+	+	+	+
36	WP_101501434.1	+	+	+	+	+	+
37	AKC48817.1	+	+	+	+	+	+
38	PAD64173.1	+	+	+	+	+	+
39	WP_095241385.1	+	+	+	+	+	+
40	WP_045926035.1	+	+	+	+	+	+

41	AJW84706.1	+	+	+	+	+	+
42	KFK78955.1	+	+	+	+	+	+
43	AXJ21641.1	+	+	+	+	+	+
44	WP_003222862.1	+	+	+	+	+	+
45	EME07777.1	+	+	+	+	+	+
46	AUZ27736.1	+	+	+	+	+	+
47	KUP29050.1	+	+	+	+	+	+
48	PRS06588.1	+	+	+	+	+	+
49	PRP51591.1	+	+	+	+	+	+
50	WP_099043576.1	+	+	+	+	+	+
51	OLQ49074.1	+	+	+	+	+	+
52	WP_075749098.1	+	+	+	+	+	+
53	AFI29791.1	+	+	+	+	+	+
54	WP_014665258.1	+	+	+	+	+	+
55	ASB62313.1	+	+	+	+	+	+
56	AOL30990.1	+	+	+	+	+	+
57	WP_083686476.1	+	+	+	+	+	+
58	OZT14492.1	+	+	+	+	+	+
59	WP_094910043.1	+	+	+	+	+	+
60	AQX54882.1	+	+	+	+	+	+
61	WP_061784634.1	+	+	+	+	+	+
62	WP_078989772.1	+	+	+	+	+	+
63	WP_040342081.1	+	+	+	+	+	+
64	WP_101224285.1	+	+	+	+	+	+
65	ASS93773.1	+	+	+	+	+	+
66	WP_063232385.1	+	+	+	+	+	+
67	WP_061143228.1	+	+	+	+	+	+
68	PJN86603.1	+	+	+	+	+	+
69	WP_048623468.1	+	+	+	+	+	+
70	BAB20808.1	+	+	+	+	+	+

Table A II (B) Conserved amino acid residues inside the motifs obtained from MEME server for all the selected *Bacillus* species of uricase

Name	Region	Amino acids	Number of conserved residues
Motif 1	401-450	E401,Q402,A407,F411,S418,I419,Q420,L422,I426,G427,F434,P435,Q436,L437,F442,N446,T448,W449	18
Motif 2	232-281	S232,F233,G236,D237,N238,A243,T244,D245,S246,M247,K248,N249,I251,G260,T262,E264,G265,F266,F274,Y278	20
Motif3	334-383	L346,K348,V349,S353,F354,G356,D360,Y362,T363,L365,E367,R371,P372,L373,L377,W381,Y383	17
Motif4	180-229	Y182,G183,K184,V187,R191,T192,P196,L197,I203,P204,E205,S206,F208,G217,G226	15
Motif5	282-331	F293,S307,V310,E317,R327	5
Motif6	452-482	G461,V463,P467,P470,G472,F473,Q474,F476	8

Table A II (C) Physiochemical characterization of different *Bacillus* species of uricase protein sequences

Sl. No.	Accession No	Source organism	No. of Amino Acids	Mol. Wt	Theoritical Pi	Negative Charge	Positive Charge	Extinction coeff.	Instability Index	Aliphatic index	GRAVY
1	3WLV_A	*Bacillus Sp. Tb-90*	312	35627.25	5.61	39	32	36330	43.18	83.69	-0.354
2	KGM46460.1	*Bacillus niacini*	318	36273.83	5.11	45	35	37360	36.95	76.04	-0.414
3	WP_045524647.1	*Bacillus niacini*	317	36427.87	5.34	46	37	38850	41.82	74.1	-0.492
4	WP_034672575.1	*Bacillus niacini*	318	36273.83	5.11	45	35	37360	36.95	76.04	-0.414
5	AOM84027.1	*Bacillus beveridgei*	318	35753.31	5.22	45	34	28880	35.67	72.04	-0.374
6	1J2G_A	*Bacillus Sp. Tb-90*	319	36538.34	5.96	40	35	36330	43.1	81.85	-0.403
7	ACR09749.1	*Bacillus fastidiosus*	320	35595.72	4.96	43	29	34840	23.4	80.19	-0.267
8	4R8X_A	*Bacillus fastidiosus*	322	37908.92	4.99	46	31	34840	24.16	79.56	-0.271
9	SPT78254.1	*Bacillus circulans*	324	39758.69	5.1	44	33	39880	41.9	75.24	-0.374
10	KKI85158.1	*Bacillus clausii*	324	39403.41	5.1	44	34	39880	42.96	78.48	-0.328
11	OXS68986.1	*Bacillus filamentosus*	325	40937.82	5.39	45	35	26025	41.86	80.42	-0.333
12	KOR85772.1	*Bacillus sp. FJAT-22058*	327	37118.55	5.37	44	34	36120	33.82	74.77	-0.435
13	WP_098373266.1	*Bacillus sp. AFS017274*	327	40312.7	5.4	45	36	34505	39.48	75.96	-0.393
14	PEZ74426.1	*Bacillus sp. AFS017274*	327	40384.35	5.32	46	36	34505	39.32	75.75	-0.409
15	PCD05853.1	*Bacillus simplex*	327	40567.26	5.17	49	36	36120	38.26	75.88	-0.385
16	PAL09042.1	*Bacillus simplex*	327	40534.26	5.31	47	36	37610	37.89	77.25	-0.378
17	CEG34811.1	*Bacillus simplex*	327	40499.1	5.11	49	35	36120	36.96	75.35	-0.387
18	5AYJ_A	*Bacillus sp. TB-90*	331	37810.73	5.78	43	36	36455	42.29	80.36	-0.415
19	BAA08723.1	*Bacillus sp. TB-90*	332	40327	5.92	44	38	36330	45.58	80.68	-0.386

20	AKO95039.1	*Bacillus filamentosus*	333	41808.69	5.69	46	38	26025	40.51	80.79	-0.349
21	4R99_A	*Bacillus fastidiosus*	335	39316.34	5.52	46	33	34840	23.01	70.92	-0.302
22	BAD66267.1	*Bacillus clausii*	341	41770.55	5.19	47	36	39880	43.87	77.95	-0.348
23	WP_095326636.1	*Bacillus clausii*	341	41973.35	4.94	48	33	39880	45	77.95	-0.327
24	WP_095294289.1	*Bacillus clausii*	341	41988.36	4.94	47	32	39880	44.59	77.68	-0.331
25	WP_095236414.1	*Bacillus clausii*	341	41958.34	4.94	47	32	39880	44.59	77.95	-0.324
26	PAF09838.1	*Bacillus clausii*	341	41903.3	4.95	48	34	39880	44.08	78.54	-0.317
27	PAE88988.1	*Bacillus clausii*	341	41987.33	4.9	49	33	39880	45.63	78.01	-0.329
28	PAD14932.1	*Bacillus clausii*	341	42036.41	4.96	49	34	39880	43.32	77.74	-0.354
29	WP_081496159.1	*Bacillus filamentosus*	365	45519	5.88	49	42	30495	41.13	82.08	-0.314
30	OWV36502.1	*Bacillus intestinalis*	492	59789.67	5.39	76	59	40465	38.44	79.04	-0.455
31	EXF55358.1	*Bacillus subtilis*	494	56586.75	5.73	71	60	44935	40.56	74.64	-0.521
32	CCU60286.1	*Bacillus subtilis*	494	56558.69	5.73	70	59	44935	41.84	74.84	-0.515
33	AII36988.1	*Bacillus subtilis*	494	56586.75	5.73	71	60	44935	40.56	74.64	-0.521
34	APH66064.1	*Bacillus subtilis*	494	56535.64	5.39	75	57	40465	37.92	77.21	-0.5
35	BAM59327.1	*Bacillus subtilis*	494	56559.59	5.58	72	59	46425	39.6	74.43	-0.541
36	WP_101501434.1	*Bacillus subtilis*	494	56628.72	5.81	70	60	44935	41.34	74.43	-0.538
37	AKC48817.1	*Bacillus subtilis*	494	59852.23	5.74	73	62	46425	42.67	75.95	-0.472
38	PAD64173.1	*Bacillus siamensis*	494	56381.42	5.5	74	58	38975	42.21	77.41	-0.493
39	WP_095241385.1	*Bacillus siamensis*	494	56381.42	5.5	74	58	38975	42.21	77.41	-0.493
40	WP_045926035.1	*Bacillus siamensis*	494	56500.62	5.57	74	59	40465	43.59	80.57	-0.476
41	AJW84706.1	*Bacillus intestinalis*	494	56508.67	5.45	72	57	43445	34.16	76.6	-0.483
42	KFK78955.1	*Bacillus intestinalis*	494	56508.67	5.45	72	57	43445	34.16	76.6	-0.483
43	AXJ21641.1	*Bacillus cereus*	494	56447.53	5.44	74	57	34965	36.78	77.79	-0.495

44	WP_003222862.1	*Bacillus subtilis*	494	56508.67	5.45	72	57	43445	34.16	76.6	-0.483
45	EME07777.1	*Bacillus subtilis*	494	56559.59	5.58	72	59	46425	39.6	74.43	-0.541
46	AUZ27736.1	*Bacillus cereus*	494	55982.9	5.69	68	56	43445	42.75	75.26	-0.487
47	KUP29050.1	*Bacillus halotolerans*	494	56536.74	5.84	71	62	43445	38.55	75.06	-0.536
48	PRS06588.1	*Bacillus halotolerans*	494	56472.75	5.76	71	61	43570	40.72	76.26	-0.516
49	PRP51591.1	*Bacillus halotolerans*	494	56500.86	5.81	70	60	43445	39.03	77.63	-0.492
50	WP_099043576.1	*Bacillus halotolerans*	494	56514.69	5.81	71	61	43445	38.82	75.06	-0.531
51	OLQ49074.1	*Bacillus licheniformis*	494	56028.86	5.55	70	56	43445	41.76	75.06	-0.498
52	WP_075749098.1	*Bacillus licheniformis*	494	56028.86	5.55	70	56	43445	41.76	75.06	-0.498
53	AFI29791.1	*Bacillus sp. JS*	494	56345.33	5.59	72	59	43445	39.81	75.24	-0.527
54	WP_014665258.1	*Bacillus sp. JS*	494	56345.33	5.59	72	59	43445	39.81	75.24	-0.527
55	ASB62313.1	*Bacillus sp. MD-5*	494	56399.42	5.53	72	58	43445	41.3	76.23	-0.515
56	AOL30990.1	*Bacillus gibsonii*	494	56541.7	5.68	70	59	46425	40.49	74.84	-0.504
57	WP_083686476.1	*Bacillus sp. RU2C*	496	56193.49	5.34	63	50	49280	35.37	82.5	-0.357
58	OZT14492.1	*Bacillus aryabhattai*	496	56535.15	5.81	64	56	49280	32.89	85.67	-0.356
59	WP_094910043.1	*Bacillus aryabhattai*	496	56535.15	5.81	64	56	49280	32.89	85.67	-0.356
60	AQX54882.1	*Bacillus flexus*	496	56290.65	5.48	63	52	49280	34.98	85.5	-0.367
61	WP_061784634.1	*Bacillus flexus*	496	56289.67	5.56	62	52	49280	35.13	85.5	-0.367
62	WP_078989772.1	*Bacillus flexus*	496	56290.65	5.48	63	52	49280	34.98	82.5	-0.367
63	WP_040342081.1	*Bacillus smithii*	496	57251.05	5.73	66	55	50895	43.63	81.35	-0.448
64	WP_101224285.1	*Bacillus sp. BA3*	497	56108.59	5.65	68	56	38975	38.11	83.58	-0.355
65	ASS93773.1	*Bacillus simplex*	497	56016.39	5.52	67	54	38975	35.88	83.58	-0.365
66	WP_063232385.1	*Bacillus simplex*	497	56016.39	5.52	67	54	38975	35.88	83.58	-0.365

67	WP_061143228.1	*Bacillus simplex*	500	56270.91	5.91	61	53	42985	33.26	82.54	-0.302
68	PJN86603.1	*Bacillus sp. mrc49*	500	56245.86	6.06	61	55	44475	33.68	81.38	-0.317
69	WP_048623468.1	*Bacillus smithii*	502	58050.94	6.25	67	61	46300	43.89	79.6	-0.518
70	BAB20808.1	*Bacillus sp. TB-90*	502	57977.78	6.15	67	60	46300	43.76	79.6	-0.516

Table A II (D) Secondary structures of selected uricase proteins with their accession number

Sl. No.	Accession No	SOPMA				CFSSP		
		Helix (%)	Extended strand (%)	Beta turn (%)	Random coil (%)	Helix (%)	Sheets (%)	Turns (%)
1	3WLV_A	30.45	24.04	5.13	40.38	65.4	71.2	14.1
2	KGM46460.1	31.45	23.58	5.97	38.99	64.8	65.1	14.5
3	WP_045524647.1	28.08	24.92	4.1	42.9	59	64.7	14.2
4	WP_034672575.1	31.45	23.58	5.97	38.99	64.8	65.1	14.5
5	AOM84027.1	29.56	25.79	4.4	40.25	81.1	42.1	16
6	1J2G_A	29.15	26.33	5.64	38.87	66.5	69.6	14.1
7	ACR09749.1	32.81	23.75	3.44	40	73.4	65.3	12.2
8	4R8X_A	31.99	25.47	4.97	37.58	73.6	43.5	12.4
9	SPT78254.1	27.47	25.93	4.32	42.28	64.2	51.2	14.8
10	KKI85158.1	34.57	23.77	4.63	37.04	66	49.7	14.8
11	OXS68986.1	32.31	22.15	6.46	39.08	73.2	63.1	16.3
12	KOR85772.1	28.44	23.24	5.2	43.12	69.4	65.7	15
13	WP_098373266.1	32.72	24.46	5.5	37.31	65.7	65.7	15
14	PEZ74426.1	32.72	24.46	5.5	37.31	65.7	65.7	15
15	PCD05853.1	33.33	23.55	5.81	37.31	71.6	65.7	15
16	PAL09042.1	32.11	24.46	5.2	38.23	68.8	65.7	15
17	CEG34811.1	29.66	22.63	5.81	41.9	69.4	64.5	15
18	5AYJ_A	30.21	23.87	6.95	38.97	68	69.8	13.9
19	BAA08723.1	27.41	25.9	5.72	40.96	68.4	67.8	14.5
20	AKO95039.1	33.93	23.72	5.71	36.64	73.6	62.5	16.5
21	4R99_A	34.93	22.39	4.48	38.21	74	43	11.9
22	BAD66267.1	32.26	24.34	5.57	37.83	66	72.1	15

23	WP_095326636.1	28.15	24.93	6.74	40.18	65.4	72.7	14.7
24	WP_095294289.1	34.9	22.58	4.99	37.54	65.4	74.2	14.7
25	WP_095236414.1	32.84	21.99	6.16	39	66.9	72.7	14.4
26	PAF09838.1	30.79	25.51	4.11	39.59	68	70.1	14.4
27	PAE88988.1	28.15	24.93	6.74	40.18	65.4	72.7	14.7
28	PAD14932.1	29.62	22.87	4.99	42.52	66.3	72.7	14.4
29	WP_081496159.1	32.05	23.56	6.85	37.53	77	39.2	16.2
30	OWV36502.1	47.56	17.28	5.89	29.27	74	67.3	12.8
31	EXF55358.1	47.17	16.8	5.87	30.16	74.9	65.6	14.2
32	CCU60286.1	48.99	16.19	5.67	29.15	73.1	66.2	14.4
33	AII36988.1	47.17	16.8	5.87	30.16	74.9	65.6	14.2
34	APH66064.1	48.58	17	6.07	28.34	74.1	67	12.3
35	BAM59327.1	48.38	17.61	5.87	28.14	71.7	65.6	14.4
36	WP_101501434.1	48.58	17	6.07	28.34	71.9	66.2	14.4
37	AKC48817.1	48.38	17.61	5.87	28.14	71.7	65.6	14.4
38	PAD64173.1	47.77	16.6	6.48	29.15	73.3	63.6	13
39	WP_095241385.1	47.77	16.6	6.48	29.15	73.3	63.6	13
40	WP_045926035.1	47.77	16.6	6.28	29.35	74.5	65.2	13
41	AJW84706.1	45.95	16.8	5.87	31.38	72.1	66.8	12.8
42	KFK78955.1	45.95	16.8	5.87	31.38	72.1	66.8	12.8
43	AXJ21641.1	48.79	17	5.47	28.74	74.1	67.2	12.3
44	WP_003222862.1	45.95	16.8	5.87	31.38	72.1	66.8	12.8
45	EME07777.1	48.38	17.61	5.87	28.14	71.7	65.6	14.4
46	AUZ27736.1	46.96	17.21	5.67	30.16	72.9	64.8	13.2
47	KUP29050.1	48.38	17.21	5.67	28.74	75.7	69.8	13.6
48	PRS06588.1	47.98	17.41	6.28	28.34	77.9	65	13.8

49	PRP51591.1	48.38	16.6	6.07	28.95	78.9	67.4	12.8
50	WP_099043576.1	48.18	16.4	5.87	29.55	75.7	67.8	13.2
51	OLQ49074.1	48.58	17	5.67	28.74	72.9	66.2	13
52	WP_075749098.1	48.58	17	5.67	28.74	72.9	66.2	13
53	AFI29791.1	47.77	16.19	5.67	30.36	73.9	63.2	14.4
54	WP_014665258.1	47.77	16.19	5.67	30.36	73.9	63.2	14.4
55	ASB62313.1	47.37	16.19	6.28	30.16	75.5	65.4	14.4
56	AOL30990.1	46.76	17.41	5.67	30.16	72.9	65.6	14.4
57	WP_083686476.1	44.96	17.34	6.25	31.45	75.4	51.4	12.9
58	OZT14492.1	46.17	16.94	5.85	31.05	78	50.6	13.3
59	WP_094910043.1	46.17	16.94	5.85	31.05	78	50.6	13.3
60	AQX54882.1	45.97	17.94	5.85	30.24	75.4	50.6	12.9
61	WP_061784634.1	44.35	17.74	5.85	32.06	75.4	50.6	12.9
62	WP_078989772.1	45.97	17.94	5.85	30.24	75.4	50.6	12.9
63	WP_040342081.1	44.76	17.74	7.26	30.24	80.3	61.2	14.7
64	WP_101224285.1	45.88	17.1	5.43	31.59	69.2	52.6	12.9
65	ASS93773.1	45.47	16.5	5.23	32.8	75.3	62.6	14.5
66	WP_063232385.1	45.47	16.5	5.23	32.8	75.3	62.6	14.5
67	WP_061143228.1	43.8	17.6	6	32.6	71.8	64.8	13.4
68	PJN86603.1	45	17	5.4	32.6	72	65.2	13.4
69	WP_048623468.1	43.03	17.53	6.57	32.87	75.3	51.6	13.9
70	BAB20808.1	42.23	18.53	6.37	32.87	75.3	52.6	14.1

Table A II (E) SOSUI server results of all selected uricase protein sequences from various *Bacillus* species

Sl. No.	Accession no.	Nature
1	pdb|3WLV|A	soluble
2	KGM46460.1	soluble
3	WP_045524647.1	soluble
4	WP_034672575.1	soluble
5	AOM84027.1	soluble
6	pdb|1J2G|A Chain A	soluble
7	ACR09749.1	soluble
8	pdb|4R8X|A Chain A	soluble
9	SPT78254.1	soluble
10	KKI85158.1	soluble
11	OXS68986.1	soluble
12	KOR85772.1	soluble
13	WP_098373266.1	soluble
14	PEZ74426.1	soluble
15	PCD05853.1	soluble
16	PAL09042.1	soluble
17	CEG34811.1	soluble
18	pdb|5AYJ|A Chain A	soluble
19	BAA08723.1	soluble
20	AKO95039.1	soluble
21	pdb|4R99|A Chain A	soluble
22	BAD66267.1	soluble
23	WP_095326636.1	soluble
24	WP_095294289.1	soluble
25	WP_095236414.1	soluble
26	PAF09838.1	soluble
27	PAE88988.1	soluble
28	PAD14932.1	soluble
29	WP_081496159.1	soluble
30	OWV36502.1	soluble
31	EXF55358.1	soluble
32	CCU60286.1	soluble
33	AII36988.1	soluble
34	APH66064.1	soluble
35	BAM59327.1	soluble

36	WP_101501434.1	soluble
37	AKC48817.1	soluble
38	PAD64173.1	soluble
39	WP_095241385.1	soluble
40	WP_045926035.1	soluble
41	AJW84706.1	soluble
42	KFK78955.1	soluble
43	AXJ21641.1	soluble
44	WP_003222862.1	soluble
45	EME07777.1	soluble
46	AUZ27736.1	soluble
47	KUP29050.1	soluble
48	PRS06588.1	soluble
49	PRP51591.1	soluble
50	WP_099043576.1	soluble
51	OLQ49074.1	soluble
52	WP_075749098.1	soluble
53	AFI29791.1	soluble
54	WP_014665258.1	soluble
55	ASB62313.1	soluble
56	AOL30990.1	soluble
57	WP_083686476.1	soluble
58	OZT14492.1	soluble
59	WP_094910043.1	soluble
60	AQX54882.1	soluble
61	WP_061784634.1	soluble
62	WP_078989772.1	soluble
63	WP_101224285.1	soluble
64	WP_040342081.1	soluble
65	ASS93773.1	soluble
66	WP_063232385.1	soluble
67	WP_061143228.1	soluble
68	PJN86603.1	soluble
69	WP_048623468.1	soluble
70	BAB20808.1	soluble

Table A II (F) CYS-REC results of various *Bacillus* species from uricase proteins

Sl. No.	Accession no.	No.of Cysteines	Positions	Disulphides
1	pdb|3WLV|A	1	298	none
2	KGM46460.1	1	302	none
3	WP_045524647.1	1	301	none
4	WP_034672575.1	1	166	none
5	AOM84027.1	1	305	none
6	pdb|1J2G|A	1	94	none
7	ACR09749.1	1	96	none
8	pdb|4R8X|A	1	255	none
9	SPT78254.1	1	255	none
10	KKI85158.1	1	255	none
11	OXS68986.1	2	171 257	none
12	KOR85772.1	4	50 101 143 234	none
13	WP_098373266.1	3	50 101 143	none
14	PEZ74426.1	3	50 101 143	none
15	PCD05853.1	4	50 101 143 234	none
16	PAL09042.1	4	50 101 143 234	none
17	CEG34811.1	4	50 101 143 234	none
18	pdb|5AYJ|A	2	297 304	ss bond at 297
19	BAA08723.1	1	305	none
20	AKO95039.1	2	179 265	none
21	pdb|4R99|A	1	272	none
22	BAD66267.1	1	272	none
23	WP_095326636.1	1	272	none
24	WP_095294289.1	1	272	none
25	WP_095236414.1	1	272	none
26	PAF09838.1	1	272	none
27	PAE88988.1	1	272	none
28	PAD14932.1	1	272	none
29	WP_081496159.1	2	211 297	none
30	OWV36502.1	2	271 423	none
31	EXF55358.1	3	217 423 489	none
32	CCU60286.1	3	217 423 489	none
33	AII36988.1	3	217 423 489	none
34	APH66064.1	2	217 423	none
35	BAM59327.1	3	217 423 489	none
36	WP_101501434.1	3	217 423 489	none
37	AKC48817.1	3	217 423 489	none
38	PAD64173.1	2	217 423	none
39	WP_095241385.1	2	217 423	none

40	WP_045926035.1	2	217 423	none
41	AJW84706.1	3	217 375 423	none
42	KFK78955.1	3	217 375 423	none
43	AXJ21641.1	3	28 217 423	none
44	WP_003222862.1	3	217 375 423	none
45	EME07777.1	3	217 423 489	none
46	AUZ27736.1	2	217 423	none
47	KUP29050.1	3	66 217 423	none
48	PRS06588.1	4	66 217 310 423	none
49	PRP51591.1	3	66 217 423	none
50	WP_099043576.1	3	66 217 423	none
51	OLQ49074.1	2	217 423	none
52	WP_075749098.1	2	217 423	none
53	AFI29791.1	2	217 423	none
54	WP_014665258.1	2	217 423	none
55	ASB62313.1	2	217 423	none
56	AOL30990.1	3	217 423 489	none
57	WP_083686476.1	1	409	none
58	OZT14492.1	1	409	none
59	WP_094910043.1	1	409	none
60	AQX54882.1	1	409	none
61	WP_061784634.1	1	409	none
62	WP_078989772.1	1	409	none
63	WP_040342081.1	2	428 475	none
64	WP_101224285.1	2	48 428	none
65	ASS93773.1	2	48 428	none
66	WP_063232385.1	2	48 428	none
67	WP_061143228.1	3	47 338 428	none
68	PJN86603.1	3	48 338 428	none
69	WP_048623468.1	1	475	none
70	BAB20808.1	1	475	none

APPENDIX III

RMSD (Root Mean Square Deviation) calculation:

The RMSD is used to quantify the average change in displacement of atoms in protein or ligand for a particular frame with respect to reference frame (Kufareva and Abagyan 2012). The RMSD for frame x can be defined as:

$$RMSD_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i'(t_x) - r_i(t_{ref}))^2} \quad \ldots\ldots\ldots (2)$$

Where N is the number of atoms in protein or ligand molecule, t_{ref} is the reference time (usually the first frame i.e. t=0 is used as the reference frame), r' is the position of protein atoms in frame x after superimposing on the reference frame, where frame x is recorded as t_x. This procedure is iterated for every frame in the simulation trajectory. Monitoring the RMSD of protein provide the insights into it's structural conformation throughout the simulation trajectory, whereas ligand RMSD shows the stability of ligand with respect to protein and its catalytic pocket.

Residue wise RMSF (Root Mean Square Fluctuation) calculation:

The residue wise RMSF is crucial for monitoring fluctuation of atomic or residue position in the trajectory after fitting to the reference frame. The RMSF is defined as:

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T} < (r_i'(t) - r_i(t_{ref}))^2 >} \quad \ldots\ldots\ldots\ldots (3)$$

Where T is the total time of the trajectory over which RMSF is calculated, t_{ref} is the reference time, r_i is the position of atoms in residue r' is the position of atoms in residue i after superimposing on the reference structure. The angular brackets indicate the average of the square distance is taken over the selection of atoms in the residue.

Radius of Gyration calculation:

Radius of gyration is useful to observe the time evolution of the compactness of protein structure. The equation of radius of gyration (R_g) is given as:

$$R_g = \left(\frac{\sum_i r_i^2 m_i}{\sum_i m_i}\right)^{1/2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

Where m_i and r_i are the mass, position of atom i respectively with respect to the centre of mass of protein.

MM/GBSA calculation:

In order to estimate the binding free energy of uric acid at the catalytic pocket of both wild type uricase during the course of simulation, MM/GBSA calculation was performed in Prime module (v-3.0). OPLS-2005 force field was used to compute the binding free energy of uricase. Last five frames were extracted from the MD simulation (interval 5 ns) and MM/GBSA was performed in each frame to compare time evolution of free energy of uric acid in both the wild type, mutein model of uricase. The water molecules were removed and VSGB solvation model (Li et al. 2011) was used to solvate the protein-ligand complex in each frame extracted for MM/GBSA calculation. The dielectric constants of solute, solvent were assigned to 1 and 80 respectively. Binding free energy of uric acid further displays the effect of mutagenesis on the functional aspect of uricase. The Free energy of binding can be calculated as:

$$\Delta G_{bind} = G_{complex} - (G_{receptor} + G_{ligand}) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad (5)$$

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (6)$$

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{vdW} + \Delta E_{ele} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (7)$$

$$\Delta G_{solv} = \Delta G_{GB} + \Delta G_{SA} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

$$\Delta G_{SA} = \gamma\Delta A + \beta \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (9)$$

Where $G_{complex}$, $G_{receptor}$, G_{ligand} are the free energy of protein-ligand complex, receptor and ligand respectively. The free energy of binding is composed of molecular mechanics energy (ΔE_{MM}) and solvation energy (ΔG_{solv}). Molecular mechanics energy has three components such as: internal energy ($\Delta E_{internal}$), van-der-walls energy (ΔE_{vdW}) and electrostatic energy whereas, solvation energy consists of polar solvation energy term (ΔG_{GB}) (Onufriev et al. 2004) and nonpolar solvation energy term (ΔG_{SA}) (Amidon et al. 1975). The hydrophobic solvation term is a liner combination of both

surface tension proportionality constant (γ) and free energy of non-polar solvation of point solute (β). The values of γ, β were set to 0.0072 kcal/mol.Å^2 and 0 respectively (Tsui and Case 2000). The entropic contribution can be calculated by normal mode analysis (Genheden et al. 2012) which is avoided here due to high computational cost.

Table A.III.1 Percent identity matrix of thirteen uricases from different sources

		1	2	3	4	5	6	7	8	9	10	11	12	13
1	*Bacillus fastidious*	100.0	25.2	21.9	21.2	23.9	24.5	21.8	24.1	23.6	22.9	24.1	22.8	22.5
2	*Arthrobacter globiformis*	25.2	100.0	33.8	34.6	37.6	37.2	32.1	37.2	33.5	33.2	34.2	32.7	33.1
3	*Phaseolus vulgaris*	21.9	33.8	100.0	43.9	32.0	34.0	35.3	37.4	35.1	34.1	35.4	34.3	35.0
4	*Chlamydomonas reinhardtii*	21.2	34.6	43.9	100.0	29.2	30.8	33.4	34.7	32.2	32.6	31.9	31.8	31.8
5	*Cyberlindnera jadinii*	23.9	37.6	32.0	29.2	100.0	48.1	33.3	38.0	36.0	35.3	35.3	34.4	35.1
6	*Aspergillus flavus*	24.5	37.2	34.0	30.8	48.1	100.0	33.4	36.3	35.0	34.7	35.1	34.9	34.2
7	*Drosophila melanogaster*	21.8	32.1	35.3	33.4	33.3	33.4	100.0	42.8	44.5	43.2	43.0	41.1	42.0
8	*Camelus dromedarius*	24.1	37.2	37.4	34.7	38.0	36.3	42.8	100.0	79.0	78.3	79.2	79.4	79.7
9	*Rattus norvegicus*	23.6	33.5	35.1	32.2	36.0	35.0	44.5	79.0	100.0	94.0	88.3	89.4	88.1
10	*Mus musculus*	22.9	33.2	34.1	32.6	35.3	34.7	43.2	78.3	94.0	100.0	87.6	89.7	87.7
11	*Oryctolagus cuniculus*	24.1	34.2	35.4	31.9	35.3	35.1	43.0	79.2	88.3	87.6	100.0	89.4	87.7
12	*Papio hamadryas*	22.8	32.7	34.3	31.8	34.4	34.9	41.1	79.4	89.4	89.7	89.4	100.0	90.1
13	*Cavia porcellus*	22.5	33.1	35.0	31.8	35.1	34.2	42.0	79.7	88.1	87.7	87.7	90.1	100.0

Table A.III.2 Sequence logo of motif information and regular expression along with their pfam analysis
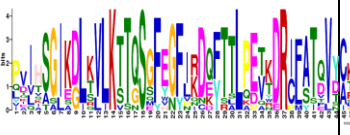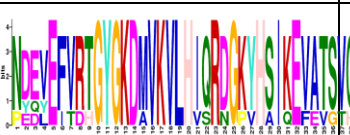
Motif number	Sequence Logo	E-value	Sites	Width	Best possible match	Pfam
1		3.3e-294	13	50	PVIHSGIKDLKV LKTTQSGFEGFI KDZFTTLPETKD RIFATQVYCKW RYQ	Uricase family
2		2.5e-146	13	29	SSKKDYLHGDN SDIIPTDTIKNTV HVLAK	Uricase family
3		6.8e-117	8	29	PNIHYFNIDMSK MGLINKEEVLLP LDNPY	Uricase family
4		2.7e-161	13	41	IETFAMNJCEHF LSSFNHVTRAHV YVEEVPWKRLE KBGVKH	Uricase family
5		3.9e-109	6	41	NDEVEFVRTGY GKDMVKVLHIQ RDGKYHSIKEVA TSVQLTL	Uricase family
6		3.6e-106	9	30	WGTVRDIVLEKF AGPYDKGEYSPS VQKTLY	Uricase family

Table A.III.3 Emini surface accessibility of uricase from both the sources of predicted peptides and their residue scores

S.NO	4R8X Wild type			2YZB Wild type		
	Peptide	Residue	Score	Peptide	Residue	Score
1	1AERTMF6	R	0.973	1MTATAE6	A	0.781
2	2ERTMFY7	T	1.509	2TATAET7	T	1.139
3	3RTMFYG8	M	0.862	3ATAETS8	A	1.058
4	4TMFYGK9	F	0.881	4TAETST9	E	1.511
5	5MFYGKG10	Y	0.604	5AETSTG10	T	1.036
6	6FYGKGD11	G	1.019	6ETSTGT11	S	1.48
7	7YGKGDV12	K	0.873	7TSTGTK12	T	1.709
8	8GKGDVY13	G	0.873	8STGTKV13	G	0.879
9	9KGDVYV14	D	0.655	9TGTKVV14	T	0.487

10	10GDVYVF15	V	0.284	10GTKVVL15	K	0.278
11	11DVYVFR16	Y	0.561	11TKVVLG16	V	0.278
12	12VYVFRT17	V	0.485	12KVVLGQ17	V	0.334
13	13YVFRTY18	F	1.024	13VVLGQN18	L	0.268
14	14VFRTYA19	R	0.66	14VLGQNQ19	G	0.626
15	15FRTYAN20	T	1.431	15LGQNQY20	Q	1.322
16	16RTYANP21	Y	2.555	16GQNQYG21	N	1.587
17	17TYANPL22	A	1.076	17QNQYGK22	Q	3.207
18	18YANPLK23	N	1.491	18NQYGKA23	Y	1.871
19	19ANPLKG24	P	0.941	19QYGKAE24	G	2.015
20	20NPLKGL25	L	0.769	20YGKAEV25	K	0.863
21	21PLKGLK26	K	0.956	21GKAEVR26	A	1.079
22	22LKGLKQ27	G	1.07	22KAEVRL27	E	0.899
23	23KGLKQI28	L	0.91	23AEVRLV28	V	0.334
24	24GLKQIP29	K	0.704	24EVRLVK29	R	0.661
25	25LKQIPE30	Q	1.231	25VRLVKV30	L	0.283
26	26KQIPES31	I	2.001	26RLVKVT31	V	0.551
27	27QIPESN32	P	1.609	27LVKVTR32	K	0.551
28	28IPESNF33	E	0.804	28VKVTRN33	V	1.074
29	29PESNFT34	S	1.656	29KVTRNT34	T	2.088
30	30ESNFTE35	N	1.855	30VTRNTA35	R	1.055
31	31SNFTEK36	F	2.142	31TRNTAR36	N	2.783
32	32NFTEKH37	T	2.175	32RNTARH37	T	2.624
33	33FTEKHN38	E	2.175	33NTARHE38	A	2.32
34	34TEKHNT39	K	3.625	34TARHEI39	R	1.011
35	35EKHNTI40	H	1.761	35ARHEIQ40	H	1.214
36	36KHNTIF41	N	0.88	36RHEIQD41	E	2.006
37	37HNTIFG42	T	0.436	37HEIQDL42	I	0.845
38	38NTIFGM43	I	0.317	38EIQDLN43	Q	0.998
39	39TIFGMN44	F	0.317	39IQDLNV44	D	0.428
40	40IFGMNA45	G	0.222	40QDLNVT45	L	0.881
41	41FGMNAK46	M	0.633	41DLNVTS46	N	0.682
42	42GMNAKV47	N	0.542	42LNVTSQ47	V	0.707
43	43MNAKVA48	A	0.554	43NVTSQL48	T	0.707
44	44NAKVAL49	K	0.461	44VTSQLR49	S	0.861
45	45AKVALK50	V	0.574	45TSQLRG50	Q	1.148
46	46KVALKG51	A	0.562	46SQLRGD51	L	1.328
47	47VALKGE52	L	0.487	47QLRGDF52	R	0.858
48	48ALKGEQ53	K	1.136	48LRGDFE53	G	0.858
49	49LKGEQL54	G	0.927	49RGDFEA54	D	1.051
50	50KGEQLL55	E	0.927	50GDFEAA55	F	0.542
51	51GEQLLT56	Q	0.669	51DFEAAH56	E	0.746
52	52EQLLTS57	L	0.906	52FEAAHT57	A	0.644
53	53QLLTSF58	L	0.453	53EAAHTA58	A	0.752
54	54LLTSFT59	T	0.377	54AAHTAG59	H	0.43
55	55LTSFTE60	S	0.793	55AHTAGD60	T	0.71
56	56TSFTEG61	F	0.951	56HTAGDN61	A	1.131
57	57SFTEGD62	T	1.101	57TAGDNA62	G	0.839
58	58FTEGDN63	E	1.321	58AGDNAH63	D	0.791
59	59TEGDNS64	G	2.044	59GDNAHV64	N	0.581
60	60EGDNSL65	D	1.168	60DNAHVV65	A	0.436
61	61GDNSLV66	N	0.501	61NAHVVA66	H	0.264
62	62DNSLVV67	S	0.375	62AHVVAT67	V	0.237
63	63NSLVVA68	L	0.227	63HVVATD68	V	0.391

236

64	64SLVVAT69	V	0.204	64VVATDT69	A	0.415
65	65LVVATD70	V	0.254	65VATDTQ70	T	0.968
66	66VVATDS71	A	0.413	66ATDTQK71	D	2.609
67	67VATDSM72	T	0.55	67TDTQKN72	T	4.154
68	68ATDSMK73	D	1.483	68DTQKNT73	Q	4.154
69	69TDSMKN74	S	2.36	69TQKNTV74	K	1.846
70	70DSMKNF75	M	1.416	70QKNTVY75	N	2.004
71	71SMKNFI76	K	0.594	71KNTVYA76	T	1.169
72	72MKNFIQ77	N	0.768	72NTVYAF77	V	0.506
73	73KNFIQR78	F	1.52	73TVYAFA78	Y	0.318
74	74NFIQRH79	I	1.035	74VYAFAR79	A	0.432
75	75FIQRHA80	Q	0.65	75YAFARD80	F	0.971
76	76IQRHAA81	R	0.758	76AFARDG81	A	0.613
77	77QRHAAS82	H	1.45	77FARDGF82	R	0.526
78	78RHAASY83	A	1.312	78ARDGFA83	D	0.613
79	79HAASYE84	A	1.16	79RDGFAT84	G	0.876
80	80AASYEG85	S	0.843	80DGFATT85	F	0.646
81	81ASYEGA86	Y	0.843	81GFATTE86	A	0.67
82	82SYEGAT87	E	1.205	82FATTEE87	T	1.172
83	83YEGATL88	G	0.741	83ATTEEF88	T	1.172
84	84EGATLE89	A	0.819	84TTEEFL89	E	0.956
85	85GATLEG90	T	0.468	85TEEFLL90	E	0.547
86	86ATLEGF91	L	0.41	86EEFLLR91	F	0.742
87	87TLEGFL92	E	0.334	87EFLLRL92	L	0.353
88	88LEGFLQ93	G	0.401	88FLLRLG93	L	0.202
89	89EGFLQY94	F	0.763	89LLRLGK94	R	0.466
90	90GFLQYV95	L	0.327	90LRLGKH95	L	0.769
91	91FLQYVC96	Q	0.177	91RLGKHF96	G	0.808
92	92LQYVCE97	Y	0.354	92LGKHFT97	K	0.595
93	93QYVCEA98	V	0.434	93GKHFTE98	H	1.25
94	94YVCEAF99	C	0.217	94KHFTEG99	F	1.25
95	95VCEAFL100	E	0.114	95HFTEGF100	T	0.541
96	96CEAFLA101	A	0.155	96FTEGFD101	E	0.664
97	97EAFLAK102	F	0.58	97TEGFDW102	G	0.806
98	98AFLAKY103	L	0.524	98EGFDWV103	F	0.415
99	99FLAKYS104	A	0.696	99GFDWVT104	D	0.346
100	100LAKYSH105	K	1.093	100FDWVTG105	W	0.346
101	101AKYSHL106	Y	1.093	101DWVTGG106	V	0.395
102	102KYSHLD107	S	1.807	102WVTGGR107	T	0.463
103	103YSHLDA108	H	0.913	103VTGGRW108	G	0.463
104	104SHLDAV109	L	0.432	104TGGRWA109	G	0.63
105	105HLDAVR110	D	0.632	105GGRWAA110	R	0.441
106	106LDAVRL111	A	0.383	106GRWAAQ111	W	0.772
107	107DAVRLE112	V	0.804	107RWAAQQ112	A	1.352
108	108AVRLEA113	R	0.487	108WAAQQF113	A	0.598
109	109VRLEAK114	L	0.963	109AAQQFF114	Q	0.492
110	110RLEAKE115	E	2.247	110AQQFFW115	Q	0.512
111	111LEAKEY116	A	1.798	111QQFFWD116	F	0.847
112	112EAKEYA117	K	2.203	112QFFWDR117	F	0.958
113	113AKEYAF118	E	1.101	113FFWDRI118	W	0.388
114	114KEYAFD119	Y	1.82	114FWDRIN119	D	0.72
115	115EYAFDD120	A	1.52	115WDRIND120	R	1.388
116	116YAFDDI121	F	0.615	116DRINDH121	I	1.796
117	117AFDDIQ122	D	0.68	117RINDHD122	N	1.796

118	118FDDIQV123	D	0.5	118INDHDH123	D	1.248
119	119DDIQVG124	I	0.571	119NDHDHA124	H	1.799
120	120DIQVGT125	Q	0.493	120DHDHAF125	D	0.969
121	121IQVGTD126	V	0.493	121HDHAFS126	H	0.777
122	122QVGTDK127	G	1.408	122DHAFSR127	A	1.119
123	123VGTDKG128	T	0.805	123HAFSRN128	F	1.077
124	124GTDKGV129	D	0.805	124AFSRNK129	S	1.583
125	125TDKGVV130	K	0.603	125FSRNKS130	R	2.1
126	126DKGVVT131	G	0.603	126SRNKSE131	N	4.201
127	127KGVVTS132	V	0.484	127RNKSEV132	K	2.326
128	128GVVTSD133	V	0.404	128NKSEVR133	S	2.326
129	129VVTSDL134	T	0.337	129KSEVRT134	E	2.088
130	130VTSDLV135	S	0.337	130SEVRTA135	V	1.055
131	131TSDLVF136	D	0.393	131EVRTAV136	R	0.584
132	132SDLVFR137	L	0.533	132VRTAVL137	T	0.278
133	133DLVFRK138	V	0.796	133RTAVLE138	A	0.649
134	134LVFRKS139	F	0.639	134TAVLEI139	V	0.232
135	135VFRKSR140	R	1.517	135AVLEIS140	L	0.216
136	136FRKSRN141	K	3.287	136VLEISG141	E	0.211
137	137RKSRNE142	C	6.575	137LEISGS142	I	0.382
138	138KSRNEY143	R	5.26	138EISGSE143	S	0.801
139	139SRNEYV144	N	1.952	139ISGSEQ144	G	0.801
140	140RNEYVT145	E	2.102	140SGSEQA145	S	1.155
141	141NEYVTA146	Y	1.084	141GSEQAI146	E	0.604
142	142EYVTAT147	V	0.973	142SEQAIV147	Q	0.453
143	143YVTATV148	T	0.417	143EQAIVA148	A	0.341
144	144VTATVE149	A	0.461	144QAIVAG149	I	0.195
145	145TATVEV150	T	0.461	145AIVAGI150	V	0.079
146	146ATVEVA151	V	0.323	146IVAGIE151	A	0.135
147	147TVEVAR152	E	0.626	147VAGIEG152	G	0.191
148	148VEVART153	V	0.626	148AGIEGL153	I	0.212
149	149EVARTA154	A	0.851	149GIEGLT154	E	0.303
150	150VARTAS155	R	0.659	150IEGLTV155	G	0.228
151	151ARTASG156	T	0.879	151EGLTVL156	L	0.268
152	152RTASGT157	A	1.255	152GLTVLK157	T	0.309
153	153TASGTE158	S	1.11	153LTVLKS158	V	0.419
154	154ASGTEV159	G	0.571	154TVLKST159	L	0.733
155	155SGTEVV160	T	0.419	155VLKSTG160	K	0.502
156	156GTEVVE161	E	0.542	156LKSTGS161	S	0.907
157	157TEVVEQ162	V	0.948	157KSTGSE162	T	1.905
158	158EVVEQA163	V	0.664	158STGSEF163	G	0.825
159	159VVEQAS164	E	0.514	159TGSEFH164	S	0.837
160	160VEQASG165	Q	0.685	160GSEFHG165	E	0.574
161	161EQASGI166	A	0.647	161SEFHGF166	F	0.502
162	162QASGIA167	S	0.377	162EFHGFP167	H	0.58
163	163ASGIAD168	G	0.364	163FHGFPR168	G	0.656
164	164SGIADI169	I	0.252	164HGFPRD169	F	1.264
165	165GIADIQ170	A	0.326	165GFPRDK170	P	1.858
166	166IADIQL171	D	0.272	166FPRDKY171	R	2.943
167	167ADIQLI172	I	0.272	167PRDKYT172	C	4.904
168	168DIQLIK173	Q	0.538	168RDKYTT173	K	4.577
169	169IQLIKV174	L	0.239	169DKYTTL174	Y	1.927
170	170QLIKVS175	I	0.457	170KYTTLQ175	T	1.999
171	171LIKVSG176	K	0.261	171YTTLQE176	T	1.731

172	172IKVSGS177	V	0.425	172TTLQET177	L	1.594
173	173KVSGSS178	S	0.812	173TLQETT178	Q	1.594
174	174VSGSSF179	G	0.351	174LQETTD179	E	1.845
175	175SGSSFY180	S	0.742	175QETTDR180	T	4.381
176	176GSSFYG181	S	0.548	176ETTDRI181	T	1.773
177	177SSFYGY182	F	0.868	177TTDRIL182	D	0.844
178	178SFYGYI183	Y	0.454	178TDRILA183	R	0.591
179	179FYGYII184	G	0.237	179DRILAT184	I	0.591
180	180YGYIID185	Y	0.458	180RILATD185	L	0.591
181	181GYIIDE186	I	0.506	181ILATDV186	A	0.224
182	182YIIDEY187	I	0.801	182LATDVS187	T	0.428
183	183IIDEYT188	D	0.738	183ATDVSA188	D	0.525
184	184IDEYTT189	E	1.519	184TDVSAR189	V	1.017
185	185DEYTTL190	Y	1.787	185DVSARW190	S	0.741
186	186EYTTLA191	T	1.081	186VSARWR191	A	0.869
187	187YTTLAE192	T	1.081	187SARWRY192	R	1.835
188	188TTLAEA193	L	0.697	188ARWRYN193	W	2.202
189	189TLAEAT194	A	0.697	189RWRYNT194	R	3.145
190	190LAEATD195	E	0.807	190WRYNTV195	Y	1.192
191	191AEATDR196	A	1.916	191RYNTVE196	N	1.963
192	192EATDRP197	T	2.932	192YNTVEV197	T	0.744
193	193ATDRPL198	D	1.396	193NTVEVD198	V	0.793
194	194TDRPLY199	R	2.166	194TVEVDF199	E	0.427
195	195DRPLYI200	P	1.052	195VEVDFD200	V	0.494
196	196RPLYIF201	L	0.545	196EVDFDA201	D	0.672
197	197PLYIFL202	Y	0.23	197VDFDAV202	F	0.288
198	198LYIFLN203	I	0.239	198DFDAVY203	D	0.608
199	199YIFLNI204	F	0.203	199FDAVYA204	A	0.368
200	200IFLNIG205	L	0.128	200DAVYAS205	V	0.57
201	201FLNIGW206	N	0.192	201AVYASV206	Y	0.253
202	202LNIGWA207	I	0.224	202VYASVR207	A	0.491
203	203NIGWAY208	G	0.426	203YASVRG208	S	0.654
204	204IGWAYE209	W	0.459	204ASVRGL209	V	0.344
205	205GWAYEN210	A	1.053	205SVRGLL210	R	0.281
206	206WAYENQ211	Y	1.843	206VRGLLL211	G	0.173
207	207AYENQD212	E	2.928	207RGLLLK212	L	0.466
208	208YENQDD213	N	4.84	208GLLLKA213	L	0.24
209	209ENQDDA214	Q	3.12	209LLLKAF214	L	0.21
210	210NQDDAK215	D	3.603	210LLKAFA215	K	0.258
211	211QDDAKG216	D	2.217	211LKAFAE216	A	0.541
212	212DDAKGD217	A	2.138	212KAFAET217	F	0.947
213	213DAKGDN218	K	2.059	213AFAETH218	A	0.644
214	214AKGDNP219	G	1.907	214FAETHS219	E	0.855
215	215KGDNPA220	D	1.907	215AETHSL220	T	0.814
216	216GDNPAN221	N	1.533	216ETHSLA221	H	0.814
217	217DNPANY222	P	2.427	217THSLAL222	S	0.388
218	218NPANYV223	A	1.079	218HSLALQ223	L	0.465
219	219PANYVA224	N	0.678	219SLALQQ224	A	0.592
220	220ANYVAA225	Y	0.443	220LALQQT225	L	0.638
221	221NYVAAE226	V	0.759	221ALQQTM226	Q	0.765
222	222YVAAEQ227	A	0.817	222LQQTMY227	Q	1.187
223	223VAAEQV228	A	0.387	223QQTMYE228	T	2.492
224	224AAEQVR229	E	1.022	224QTMYEM229	M	1.424
225	225AEQVRD230	Q	1.689	225TMYEMG230	Y	0.814

226	226EQVRDI231	V	1.172	226MYEMGR231	E	1.104
227	227QVRDIA232	R	0.684	227YEMGRA232	M	1.127
228	228VRDIAA233	D	0.399	228EMGRAV233	G	0.534
229	229RDIAAS234	I	0.72	229MGRAVI234	R	0.216
230	230DIAASV235	A	0.273	230GRAVIE235	A	0.378
231	231IAASVF236	A	0.141	231RAVIET236	V	0.552
232	232AASVFH237	S	0.275	232AVIETH237	I	0.383
233	233ASVFHT238	V	0.392	233VIETHP238	E	0.587
234	234SVFHTL239	F	0.32	234IETHPE239	T	1.369
235	235VFHTLD240	H	0.399	235ETHPEI240	H	1.369
236	236FHTLDN241	T	0.865	236THPEID241	P	1.32
237	237HTLDNK242	L	1.997	237HPEIDE242	E	1.584
238	238TLDNKS243	D	1.967	238PEIDEI243	I	0.816
239	239LDNKSI244	N	0.955	239EIDEIK244	D	1.055
240	240DNKSIQ245	K	2.006	240IDEIKM245	E	0.603
241	241NKSIQH246	S	1.635	241DEIKMS246	I	1.153
242	242KSIQHL247	I	0.838	242EIKMSL247	K	0.569
243	243SIQHLI248	Q	0.294	243IKMSLP248	M	0.508
244	244IQHLIY249	H	0.344	244KMSLPN249	S	1.166
245	245QHLIYH250	L	0.667	245MSLPNK250	L	1.166
246	246HLIYHI251	I	0.27	246SLPNKH251	P	1.603
247	247LIYHIG252	Y	0.196	247LPNKHH252	N	1.628
248	248IYHIGL253	H	0.196	248PNKHHF253	K	1.71
249	249YHIGLT254	I	0.404	249NKHHFL254	H	0.912
250	250HIGLTI255	G	0.181	250KHHFLV255	H	0.421
251	251IGLTIL256	L	0.11	251HHFLVD256	F	0.351
252	252GLTILD257	T	0.261	252HFLVDL257	L	0.213
253	253LTILDR258	I	0.517	253FLVDLQ258	V	0.271
254	254TILDRF259	L	0.543	254LVDLQP259	D	0.484
255	255ILDRFP260	D	0.581	255VDLQPF260	L	0.508
256	256LDRFPQ261	R	1.436	256DLQPFG261	Q	0.678
257	257DRFPQL262	F	1.436	257LQPFGQ262	P	0.703
258	258RFPQLT263	P	1.241	258QPFGQD263	F	1.423
259	259FPQLTE264	Q	1.098	259PFGQDN264	G	1.321
260	260PQLTEV265	L	0.941	260FGQDNP265	Q	1.321
261	261QLTEVN266	T	0.978	261GQDNPN266	D	2.454
262	262LTEVNF267	E	0.489	262QDNPNE267	N	4.294
263	263TEVNFG268	V	0.587	263DNPNEV268	P	1.84
264	264EVNFGT269	N	0.587	264NPNEVF269	N	0.954
265	265VNFGTN270	F	0.545	265PNEVFY270	E	0.93
266	266NFGTNN271	G	1.181	266NEVFYA271	V	0.608
267	267FGTNNR272	T	1.438	267EVFYAA272	F	0.382
268	268GTNNRT273	N	2.397	268VFYAAD273	Y	0.368
269	269TNNRTW274	N	2.547	269FYAADR274	A	0.971
270	270NNRTWD275	R	2.948	270YAADRP275	A	1.734
271	271NRTWDT276	T	2.645	271AADRPY276	D	1.734
272	272RTWDTV277	W	1.221	272ADRPYG277	R	1.699
273	273TWDTVV278	D	0.463	273DRPYGL278	P	1.387
274	274WDTVVE279	T	0.555	274RPYGLI279	Y	0.582
275	275DTVVEG280	V	0.523	275PYGLIE280	G	0.515
276	276TVVEGT281	V	0.452	276YGLIEA281	L	0.336
277	277VVEGTD282	E	0.523	277GLIEAT282	I	0.31
278	278VEGTDG283	G	0.697	278LIEATI283	E	0.219
279	279EGTDGF284	T	0.813	279IEATIQ284	A	0.461

280	280GTDGFK285	D	0.939	280EATIQR285	T	1.287
281	281TDGFKG286	G	0.939	281ATIQRE286	I	1.287
282	282DGFKGA287	F	0.657	282TIQREG287	Q	1.261
283	283GFKGAV288	K	0.292	283IQREGS288	R	1.171
284	284FKGAVF289	G	0.256	284QREGSR289	E	3.272
285	285KGAVFT290	A	0.426	285REGSRA290	G	1.908
286	286GAVFTE291	V	0.369	286EGSRAD291	S	1.627
287	287AVFTEP292	F	0.576	287GSRADH292	R	1.279
288	288VFTEPR293	T	1.117	288SRADHP293	A	1.998
289	289FTEPRP294	E	2.327	289RADHPI294	D	1.045
290	290TEPRPP295	P	4.156	290ADHPIW295	H	0.561
291	291EPRPPF296	R	2.494	291DHPIWS296	P	0.744
292	292PRPPFG297	P	1.425	292HPIWSN297	I	0.717
293	293RPPFGF298	P	0.798	293PIWSNI298	W	0.369
294	294PPFGFQ299	F	0.706	294IWSNIA299	S	0.241
295	295PFGFQG300	G	0.452	295WSNIAG300	N	0.34
296	296FGFQGF301	F	0.253	296SNIAGF301	I	0.28
297	297GFQGFS302	Q	0.391	297NIAGFC302	A	0.112
298	298FQGFSV303	G	0.294			
299	299QGFSVH304	F	0.461			
300	300GFSVHQ305	S	0.461			
301	301FSVHQE306	V	0.807			
302	302SVHQED307	H	1.557			
303	303VHQEDL308	Q	0.958			
304	304HQEDLA309	E	1.304			
305	305QEDLAR310	D	1.877			
306	306EDLARE311	L	1.877			
307	307DLAREK312	A	2.167			
308	308LAREKA313	R	1.311			
309	309AREKAS314	E	2.13			
310	310REKASA315	K	2.13			
311	311EKASAN316	A	1.749			
312	312KASANS317	S	1.354			
313	313ASANSE318	A	1.172			
314	314SANSEY319	N	1.818			
315	315ANSEYV320	S	1.007			
316	316NSEYVA321	E	1.007			
317	317SEYVAL322	Y	0.516			

Table A.III.4 Parker hydrophilicity of uricase from both the sources of predicted peptides and their residue scores

S.NO	4R8X Wild type			2YZB Wild type		
	Peptide	Residue	Score	Peptide	Residue	Score
1	1AERTMFY7	T	0.571	1MTATAET7	T	3.343
2	2ERTMFYG8	M	1.086	2TATAETS8	A	4.871
3	3RTMFYGK9	F	0.786	3ATAETST9	E	4.871
4	4TMFYGKG10	Y	1	4TAETSTG10	T	5.386
5	5MFYGKGD11	G	1.686	5AETSTGT11	S	5.386
6	6FYGKGDV12	K	1.757	6ETSTGTK12	T	5.9
7	7YGKGDVY13	G	2.8	7TSTGTKV13	G	4.257
8	8GKGDVYV14	D	2.543	8STGTKVV14	T	2.986

9	9KGDVYVF15	V	0.414	9TGTKVVL15	K	0.743
10	10GDVYVFR16	Y	0.2	10GTKVVLG16	V	0.814
11	11DVYVFRT17	V	0.129	11TKVVLGQ17	V	0.857
12	12VYVFRTY18	F	-1.571	12KVVLGQN18	L	1.114
13	13YVFRTYA19	R	-0.743	13VVLGQNQ19	G	1.157
14	14VFRTYAN20	T	0.529	14VLGQNQY20	Q	1.414
15	15FRTYANP21	Y	1.357	15LGQNQYG21	N	2.757
16	16RTYANPL22	A	1.357	16GQNQYGK22	Q	4.886
17	17TYANPLK23	N	1.571	17QNQYGKA23	Y	4.371
18	18YANPLKG24	P	1.643	18NQYGKAE24	G	4.629
19	19ANPLKGL25	L	0.6	19QYGKAEV25	K	3.1
20	20NPLKGLK26	K	1.114	20YGKAEVR26	A	2.843
21	21PLKGLKQ27	G	0.971	21GKAEVRL27	E	1.8
22	22LKGLKQI28	L	-0.471	22KAEVRLV28	V	0.457
23	23KGLKQIP29	K	1.143	23AEVRLVK29	R	0.457
24	24GLKQIPE30	Q	1.443	24EVRLVKV30	L	-0.371
25	25LKQIPES31	I	1.557	25VRLVKVT31	V	-0.743
26	26KQIPESN32	P	3.871	26RLVKVTR32	K	0.386
27	27QIPESNF33	E	1.743	27LVKVTRN33	V	0.786
28	28IPESNFT34	S	1.629	28VKVTRNT34	T	2.843
29	29PESNFTE35	N	3.886	29KVTRNTA35	R	3.671
30	30ESNFTEK36	F	4.4	30VTRNTAR36	N	3.457
31	31SNFTEKH37	T	3.586	31TRNTARH37	T	4.286
32	32NFTEKHN38	E	3.657	32RNTARHE38	A	4.657
33	33FTEKHNT39	K	3.4	33NTARHEI39	R	2.914
34	34TEKHNTI40	H	3.571	34TARHEIQ40	H	2.771
35	35EKHNTIF41	N	1.514	35ARHEIQD41	E	3.457
36	36KHNTIFG42	T	1.214	36RHEIQDL42	I	1.843
37	37HNTIFGM43	I	-0.2	37HEIQDLN43	Q	2.243
38	38NTIFGMN44	F	0.5	38EIQDLNV44	D	1.414
39	39TIFGMNA45	G	-0.2	39IQDLNVT45	L	1.043
40	40IFGMNAK46	M	-0.129	40QDLNVTS46	N	3.114
41	41FGMNAKV47	N	0.486	41DLNVTSQ47	V	3.114
42	42GMNAKVA48	A	2.1	42LNVTSQL48	T	0.371
43	43MNAKVAL49	K	-0.029	43NVTSQLR49	S	2.286
44	44NAKVALK50	V	1.386	44VTSQLRG50	Q	2.1
45	45AKVALKG51	A	1.2	45TSQLRGD51	L	4.057
46	46KVALKGE52	L	2.014	46SQLRGDF52	R	2
47	47VALKGEQ53	K	2.057	47QLRGDFE53	G	2.186
48	48ALKGEQL54	G	1.271	48LRGDFEA54	D	1.629
49	49LKGEQLL55	E	-0.343	49RGDFEAA55	F	3.243
50	50KGEQLLT56	Q	1.714	50GDFEAAH56	E	2.943
51	51GEQLLTS57	L	1.829	51DFEAAHT57	A	2.871
52	52EQLLTSF58	L	-0.3	52FEAAHTA58	A	1.743
53	53QLLTSFT59	T	-0.671	53EAAHTAG59	H	3.871
54	54LLTSFTE60	S	-0.414	54AAHTAGD60	T	4.186
55	55LTSFTEG61	F	1.714	55AHTAGDN61	A	4.886
56	56TSFTEGD62	T	4.457	56HTAGDNA62	G	4.886
57	57SFTEGDN63	E	4.714	57TAGDNAH63	D	4.886
58	58FTEGDNS64	G	4.714	58AGDNAHV64	N	3.614
59	59TEGDNSL65	D	4.714	59GDNAHVV65	A	2.786
60	60EGDNSLV66	N	3.443	60DNAHVVA66	H	2.271
61	61GDNSLVV67	S	1.8	61NAHVVAT67	V	1.586
62	62DNSLVVA68	L	1.286	62AHVVATD68	V	2.014

63	63NSLVVAT69	V	0.6	63HVVATDT69	A	2.457
64	64SLVVATD70	V	1.029	64VVATDTQ70	T	3.014
65	65LVVATDS71	A	1.029	65VATDTQK71	D	4.357
66	66VVATDSM72	T	1.743	66ATDTQKN72	T	5.886
67	67VATDSMK73	D	3.086	67TDTQKNT73	Q	6.329
68	68ATDSMKN74	S	4.614	68DTQKNTV74	K	5.057
69	69TDSMKNF75	M	3	69TQKNTVY75	N	3.357
70	70DSMKNFI76	K	1.114	70QKNTVYA76	T	2.914
71	71SMKNFIQ77	N	0.543	71KNTVYAF77	V	0.743
72	72MKNFIQR78	F	0.214	72NTVYAFA78	Y	0.229
73	73KNFIQRH79	I	1.114	73TVYAFAR79	A	-0.171
74	74NFIQRHA80	Q	0.6	74VYAFARD80	F	0.514
75	75FIQRHAA81	R	-0.1	75YAFARDG81	A	1.857
76	76IQRHAAS82	H	2.143	76AFARDGF82	R	0.814
77	77QRHAASY83	A	3.014	77FARDGFA83	D	0.814
78	78RHAASYE84	A	3.271	78ARDGFAT84	G	2.871
79	79HAASYEG85	S	3.486	79RDGFATT85	F	3.314
80	80AASYEGA86	Y	3.486	80DGFATTE86	A	3.829
81	81ASYEGAT87	E	3.929	81GFATTEE87	T	3.514
82	82SYEGATL88	G	2.314	82FATTEEF88	T	1.386
83	83YEGATLE89	A	2.5	83ATTEEFL89	E	1.386
84	84EGATLEG90	T	3.586	84TTEEFLL90	E	-0.229
85	85GATLEGF91	L	1.157	85TEEFLLR91	F	-0.371
86	86ATLEGFL92	E	-0.971	86EEFLLRL92	L	-2.429
87	87TLEGFLQ93	G	-0.414	87EFLLRLG93	L	-2.729
88	88LEGFLQY94	F	-1.429	88FLLRLGK94	R	-3.029
89	89EGFLQYV95	L	-0.643	89LLRLGKH95	L	-1.414
90	90GFLQYVC96	Q	-1.557	90LRLGKHF96	G	-1.414
91	91FLQYVCE97	Y	-1.257	91RLGKHFT97	K	0.643
92	92LQYVCEA98	V	0.357	92LGKHFTE98	H	1.157
93	93QYVCEAF99	C	0.357	93GKHFTEG99	F	3.286
94	94YVCEAFL100	E	-1.814	94KHFTEGF100	T	1.157
95	95VCEAFLA101	A	-1.243	95HFTEGFD101	E	1.771
96	96CEAFLAK102	F	0.1	96FTEGFDW102	G	0.043
97	97EAFLAKY103	L	-0.371	97TEGFDWV103	F	0.829
98	98AFLAKYS104	A	-0.557	98EGFDWVT104	D	0.829
99	99FLAKYSH105	K	-0.557	99GFDWVTG105	W	0.529
100	100LAKYSHL106	Y	-0.557	100FDWVTGG106	V	0.529
101	101AKYSHLD107	S	2.186	101DWVTGGR107	T	2.443
102	102KYSHLDA108	H	2.186	102WVTGGRW108	G	-0.414
103	103YSHLDAV109	L	0.843	103VTGGRWA109	G	1.314
104	104SHLDAVR110	D	1.714	104TGGRWAA110	R	2.143
105	105HLDAVRL111	A	-0.529	105GGRWAAQ111	W	2.257
106	106LDAVRLE112	V	0.286	106GRWAAQQ112	A	2.3
107	107DAVRLEA113	R	1.9	107RWAAQQF113	A	0.171
108	108AVRLEAK114	L	1.286	108WAAQQFF114	Q	-1.743
109	109VRLEAKE115	E	2.1	109AAQQFFW115	Q	-1.743
110	110RLEAKEY116	A	2.357	110AQQFFWD116	F	-0.614
111	111LEAKEYA117	K	2.057	111QQFFWDR117	F	-0.314
112	112EAKEYAF118	E	2.057	112QFFWDRI118	W	-2.314
113	113AKEYAFD119	Y	2.371	113FFWDRIN119	D	-2.171
114	114KEYAFDD120	A	3.5	114FWDRIND120	R	0.571
115	115EYAFDDI121	F	1.543	115WDRINDH121	I	2.186
116	116YAFDDIQ122	D	1.286	116DRINDHD122	N	5.043

117	117AFDDIQV123	D	1.029	117RINDHDH123	D	3.914
118	118FDDIQVG124	I	1.543	118INDHDHA124	H	3.614
119	119DDIQVGT125	Q	3.6	119NDHDHAF125	D	3.443
120	120DIQVGTD126	V	3.6	120DHDHAFS126	H	3.371
121	121IQVGTDK127	G	2.986	121HDHAFSR127	A	2.543
122	122QVGTDKG128	T	4.943	122DHAFSRN128	F	3.243
123	123VGTDKGV129	D	3.557	123HAFSRNK129	S	2.629
124	124GTDKGVV130	K	3.557	124AFSRNKS130	R	3.257
125	125TDKGVVT131	G	3.486	125FSRNKSE131	N	4.071
126	126DKGVVTS132	V	3.671	126SRNKSEV132	K	4.857
127	127KGVVTSD133	V	3.671	127RNKSEVR133	S	4.529
128	128GVVTSDL134	T	1.543	128NKSEVRT134	E	4.671
129	129VVTSDLV135	S	0.2	129KSEVRTA135	V	3.971
130	130VTSDLVF136	D	-0.586	130SEVRTAV136	R	2.629
131	131TSDLVFR137	L	0.543	131EVRTAVL137	T	0.386
132	132SDLVFRK138	V	0.614	132VRTAVLE138	A	0.386
133	133DLVFRKS139	F	0.614	133RTAVLEI139	V	-0.229
134	134LVFRKSR140	R	-0.214	134TAVLEIS140	L	0.1
135	135VFRKSRN141	K	2.1	135AVLEISG141	E	0.171
136	136FRKSRNE142	S	3.743	136VLEISGS142	I	0.8
137	137RKSRNEY143	R	4.786	137LEISGSE143	S	2.443
138	138KSRNEYV144	N	3.657	138EISGSEQ144	G	4.614
139	139SRNEYVT145	E	3.586	139ISGSEQA145	S	3.8
140	140RNEYVTA146	Y	2.957	140SGSEQAI146	E	3.8
141	141NEYVTAT147	V	3.1	141GSEQAIV147	Q	2.343
142	142EYVTATV148	T	1.571	142SEQAIVA148	A	1.829
143	143YVTATVE149	A	1.571	143EQAIVAG149	I	1.714
144	144VTATVEV150	T	1.314	144QAIVAGI150	V	-0.543
145	145TATVEVA151	V	2.143	145AIVAGIE151	A	-0.286
146	146ATVEVAR152	E	2	146IVAGIEG152	G	0.229
147	147TVEVART153	V	2.443	147VAGIEGL153	I	0.057
148	148VEVARTA154	A	2	148AGIEGLT154	E	1.329
149	149EVARTAS155	R	3.457	149GIEGLTV155	G	0.5
150	150VARTASG156	T	3.157	150IEGLTVL156	L	-1.629
151	151ARTASGT157	A	4.429	151EGLTVLK157	T	0.329
152	152RTASGTE158	S	5.243	152GLTVLKS158	V	0.143
153	153TASGTEV159	G	4.114	153LTVLKST159	L	0.071
154	154ASGTEVV160	T	2.843	154TVLKSTG160	K	2.2
155	155SGTEVVE161	E	3.657	155VLKSTGS161	S	2.386
156	156GTEVVEQ162	V	3.586	156LKSTGSE162	T	4.029
157	157TEVVEQA163	V	3.071	157KSTGSEF163	G	4.029
158	158EVVEQAS164	E	3.257	158STGSEFH164	S	3.514
159	159VVEQASG165	Q	2.957	159TGSEFHG165	E	3.4
160	160VEQASGI166	A	2.343	160GSEFHGF166	F	1.343
161	161EQASGIA167	S	3.171	161SEFHGFP167	H	0.829
162	162QASGIAD168	G	3.486	162EFHGFPR168	G	0.5
163	163ASGIADI169	I	1.486	163FHGFPRD169	F	0.814
164	164SGIADIQ170	A	2.043	164HGFPRDK170	P	2.943
165	165GIADIQL171	D	-0.2	165GFPRDKY171	R	2.371
166	166IADIQLI172	I	-2.157	166FPRDKYT172	D	2.3
167	167ADIQLIK173	Q	-0.2	167PRDKYTT173	K	4.357
168	168DIQLIKV174	L	-1.029	168RDKYTTL174	Y	2.743
169	169IQLIKVS175	I	-1.529	169DKYTTLQ175	T	3
170	170QLIKVSG176	K	0.429	170KYTTLQE176	T	2.686

171	171LIKVSGS177	V	0.5	171YTTLQET177	L	2.614
172	172IKVSGSS178	S	2.743	172TTLQETT178	Q	3.629
173	173KVSGSSF179	G	2.571	173TLQETTD179	E	4.314
174	174VSGSSFY180	S	1.486	174LQETTDR180	T	4.171
175	175SGSSFYG181	S	2.829	175QETTDRI181	T	4.343
176	176GSSFYGY182	F	1.629	176ETTDRIL182	D	2.171
177	177SSFYGYI183	Y	-0.329	177TTDRILA183	R	1.357
178	178SFYGYII184	G	-2.4	178TDRILAT184	I	1.357
179	179FYGYIID185	Y	-1.9	179DRILATD185	L	2.043
180	180YGYIIDE186	I	0.529	180RILATDV186	A	0.086
181	181GYIIDEY187	I	0.529	181ILATDVS187	T	0.414
182	182YIIDEYT188	D	0.457	182LATDVSA188	D	1.857
183	183IIDEYTT189	E	1.471	183ATDVSAR189	V	3.771
184	184IDEYTTL190	Y	1.3	184TDVSARW190	S	2.043
185	185DEYTTLA191	T	2.743	185DVSARWR191	A	1.9
186	186EYTTLAE192	T	2.429	186VSARWRY192	R	0.2
187	187YTTLAEA193	L	1.614	187SARWRYN193	W	1.729
188	188TTLAEAT194	A	2.629	188ARWRYNT194	R	1.543
189	189TLAEATD195	E	3.314	189RWRYNTV195	Y	0.714
190	190LAEATDR196	A	3.171	190WRYNTVE196	N	1.229
191	191AEATDRP197	T	4.786	191RYNTVEV197	T	2.129
192	192EATDRPL198	D	3.171	192YNTVEVD198	V	2.957
193	193ATDRPLY199	R	1.786	193NTVEVDF199	E	1.914
194	194TDRPLYI200	P	0.343	194TVEVDFD200	V	2.343
195	195DRPLYIF201	L	-1.714	195VEVDFDA201	D	1.9
196	196RPLYIFL202	Y	-4.457	196EVDFDAV202	F	1.9
197	197PLYIFLN203	I	-4.057	197VDFDAVY203	D	0.514
198	198LYIFLNI204	F	-5.5	198DFDAVYA204	A	1.343
199	199YIFLNIG205	L	-3.371	199FDAVYAS205	V	0.843
200	200IFLNIGW206	N	-4.529	200DAVYASV206	Y	1.629
201	201FLNIGWA207	I	-3.086	201AVYASVR207	A	0.8
202	202LNIGWAY208	G	-2.043	202VYASVRG208	S	1.314
203	203NIGWAYE209	W	0.386	203YASVRGL209	V	0.529
204	204IGWAYEN210	A	0.386	204ASVRGLL210	R	-0.514
205	205GWAYENQ211	Y	2.386	205SVRGLLL211	G	-2.129
206	206WAYENQD212	E	3	206VRGLLLK212	L	-2.243
207	207AYENQDD213	N	5.857	207RGLLLKA213	L	-1.414
208	208YENQDDA214	Q	5.857	208GLLLKAF214	L	-3.329
209	209ENQDDAK215	D	6.943	209LLLKAFA215	K	-3.843
210	210NQDDAKG216	D	6.643	210LLKAFAE216	A	-1.414
211	211QDDAKGD217	A	7.071	211LKAFAET217	F	0.643
212	212DDAKGDN218	K	7.214	212KAFAETH218	A	2.257
213	213DAKGDNP219	G	6.086	213AFAETHS219	E	2.371
214	214AKGDNPA220	D	4.957	214FAETHSL220	T	0.757
215	215KGDNPAN221	N	5.657	215AETHSLA221	H	2.371
216	216GDNPANY222	P	4.571	216ETHSLAL222	S	0.757
217	217DNPANYV223	A	3.229	217THSLALQ223	L	0.5
218	218NPANYVA224	N	2.1	218HSLALQQ224	A	0.614
219	219PANYVAA225	Y	1.4	219SLALQQT225	L	1.057
220	220ANYVAAE226	V	2.214	220LALQQTM226	Q	-0.471
221	221NYVAAEQ227	A	2.771	221ALQQTMY227	Q	0.571
222	222YVAAEQV228	A	1.243	222LQQTMYE228	T	1.386
223	223VAAEQVR229	E	2.114	223QQTMYEM229	M	2.1
224	224AAEQVRD230	Q	4.071	224QTMYEMG230	Y	2.057

225	225AEQVRDI231	V	2.629	225TMYEMGR231	E	1.8
226	226EQVRDIA232	R	2.629	226MYEMGRA232	M	1.357
227	227QVRDIAA233	D	1.814	227YEMGRAV233	G	1.429
228	228VRDIAAS234	I	1.886	228EMGRAVI234	R	0.557
229	229RDIAASV235	A	1.886	229MGRAVIE235	A	0.557
230	230DIAASVF236	A	-0.029	230GRAVIET236	V	1.9
231	231IAASVFH237	S	-1.157	231RAVIETH237	I	1.386
232	232AASVFHT238	V	0.729	232AVIETHP238	E	1.086
233	233ASVFHTL239	F	-0.886	233VIETHPE239	T	1.9
234	234SVFHTLD240	H	0.243	234IETHPEI240	H	1.286
235	235VFHTLDN241	T	0.314	235ETHPEID241	P	3.857
236	236FHTLDNK242	L	1.657	236THPEIDE242	E	3.857
237	237HTLDNKS243	D	3.9	237HPEIDEI243	I	1.971
238	238TLDNKSI244	N	2.457	238PEIDEIK244	D	2.486
239	239LDNKSIQ245	K	2.571	239EIDEIKM245	E	1.586
240	240DNKSIQH246	S	4.186	240IDEIKMS246	I	1.4
241	241NKSIQHL247	I	1.443	241DEIKMSL247	K	1.229
242	242KSIQHLI248	Q	-0.7	242EIKMSLP248	M	0.1
243	243SIQHLIY249	H	-1.786	243IKMSLPN249	S	-0.014
244	244IQHLIYH250	L	-2.414	244KMSLPNK250	L	1.943
245	245QHLIYHI251	I	-2.414	245MSLPNKH251	P	1.429
246	246HLIYHIG252	Y	-2.457	246SLPNKHH252	N	2.329
247	247LIYHIGL253	H	-4.071	247LPNKHHF253	K	0.086
248	248IYHIGLT254	I	-2.014	248PNKHHFL254	H	0.086
249	249YHIGLTI255	G	-2.014	249NKHHFLV255	H	-0.743
250	250HIGLTIL256	L	-3.057	250KHHFLVD256	F	-0.314
251	251IGLTILD257	T	-1.929	251HHFLVDL257	L	-2.443
252	252GLTILDR258	I	-0.186	252HFLVDLQ258	V	-1.886
253	253LTILDRF259	L	-2.314	253FLVDLQP259	D	-1.886
254	254TILDRFP260	D	-0.7	254LVDLQPF260	L	-1.886
255	255ILDRFPQ261	R	-0.586	255VDLQPFG261	Q	0.243
256	256LDRFPQL262	F	-0.757	256DLQPFGQ262	P	1.629
257	257DRFPQLT263	P	1.3	257LQPFGQD263	F	1.629
258	258RFPQLTE264	Q	0.986	258QPFGQDN264	G	3.943
259	259FPQLTEV265	L	-0.143	259PFGQDNP265	Q	3.386
260	260PQLTEVN266	T	2.171	260FGQDNPN266	D	4.086
261	261QLTEVNF267	E	0.557	261GQDNPNE267	N	6.514
262	262LTEVNFG268	V	0.514	262QDNPNEV268	P	5.171
263	263TEVNFGT269	N	2.571	263DNPNEVF269	N	3
264	264EVNFGTN270	F	2.829	264NPNEVFY270	E	1.3
265	265VNFGTNN271	G	2.714	265PNEVFYA271	V	0.6
266	266NFGTNNR272	T	3.843	266NEVFYAA272	F	0.6
267	267FGTNNRT273	N	3.586	267EVFYAAD273	Y	1.029
268	268GTNNRTW274	N	3.471	268VFYAADR274	A	0.514
269	269TNNRTWD275	R	4.086	269FYAADRP275	A	1.343
270	270NNRTWDT276	T	4.086	270YAADRPY276	D	2.386
271	271NRTWDTV277	W	2.557	271AADRPYG277	R	3.471
272	272RTWDTVV278	D	1.029	272ADRPYGL278	P	1.857
273	273TWDTVVE279	T	1.543	273DRPYGLI279	Y	0.414
274	274WDTVVEG280	V	1.614	274RPYGLIE280	G	0.1
275	275DTVVEGT281	V	3.786	275PYGLIEA281	L	-0.2
276	276TVVEGTD282	E	3.786	276YGLIEAT282	I	0.243
277	277VVEGTDG283	G	3.857	277GLIEATI283	E	-0.629
278	278VEGTDGF284	T	3.071	278LIEATIQ284	A	-0.586

246

279	279EGTDGFK285	D	4.414	279IEATIQR285	T	1.329
280	280GTDGFKG286	G	4.114	280EATIQRE286	I	3.586
281	281TDGFKGA287	F	3.6	281ATIQREG287	Q	3.286
282	282DGFKGAV288	K	2.329	282TIQREGS288	R	3.914
283	283GFKGAVF289	G	-0.414	283IQREGSR289	E	3.771
284	284FKGAVFT290	A	-0.486	284QREGSRA290	G	5.214
285	285KGAVFTE291	V	1.943	285REGSRAD291	S	5.786
286	286GAVFTEP292	F	1.429	286EGSRADH292	R	5.486
287	287AVFTEPR293	T	1.214	287GSRADHP293	A	4.671
288	288VFTEPRP294	E	1.214	288SRADHPI294	D	2.714
289	289FTEPRPP295	P	2.043	289RADHPIW295	H	0.357
290	290TEPRPPF296	R	2.043	290ADHPIWS296	P	0.686
291	291EPRPPFG297	P	2.114	291DHPIWSN297	I	1.386
292	292PRPPFGF298	P	-0.314	292HPIWSNI298	W	-1.186
293	293RPPFGFQ299	F	0.243	293PIWSNIA299	S	-1.186
294	294PPFGFQG300	G	0.457	294IWSNIAG300	N	-0.671
295	295PFGFQGF301	F	-1.157	295WSNIAGF301	I	-0.843
296	296FGFQGFS302	Q	-0.529	296SNIAGFC302	A	0.786
297	297GFQGFSV303	G	0.257			
298	298FQGFSVH304	F	-0.257			
299	299QGFSVHQ305	S	1.914			
300	300GFSVHQE306	V	2.171			
301	301FSVHQED307	H	2.786			
302	302SVHQEDL308	Q	2.786			
303	303VHQEDLA309	E	2.157			
304	304HQEDLAR310	D	3.286			
305	305QEDLARE311	L	4.1			
306	306EDLAREK312	A	4.057			
307	307DLAREKA313	R	3.243			
308	308LAREKAS314	E	2.743			
309	309AREKASA315	K	4.357			
310	310REKASAN316	A	5.057			
311	311EKASANS317	S	5.386			
312	312KASANSE318	A	5.386			
313	313ASANSEY319	N	4.3			
314	314SANSEYV320	S	3.471			
315	315ANSEYVA321	E	2.843			
316	316NSEYVAL322	Y	1.229			

Table A.III.5 Karplus & Schulz Flexibility scores of 2YZB sequence to predict peptides which can act as B-cell epitope (Uricase from *Arthrobacter globiformis*)

Position	Residue	Start	End	Peptide	Score
4	T	1	7	MTATAET	0.996
5	A	2	8	TATAETS	1.025
6	E	3	9	ATAETST	1.05
7	T	4	10	TAETSTG	1.076
8	S	5	11	AETSTGT	1.099
9	T	6	12	ETSTGTK	1.102
10	G	7	13	TSTGTKV	1.094
11	T	8	14	STGTKVV	1.065
12	K	9	15	TGTKVVL	1.03

13	V		10	16	GTKVVLG	0.997
14	V		11	17	TKVVLGQ	0.988
15	L		12	18	KVVLGQN	1.006
16	G		13	19	VVLGQNQ	1.035
17	Q		14	20	VLGQNQY	1.062
18	N		15	21	LGQNQYG	1.066
19	Q		16	22	GQNQYGK	1.055
20	Y		17	23	QNQYGKA	1.041
21	G		18	24	NQYGKAE	1.03
22	K		19	25	QYGKAEV	1.023
23	A		20	26	YGKAEVR	1.007
24	E		21	27	GKAEVRL	0.986
25	V		22	28	KAEVRLV	0.965
26	R		23	29	AEVRLVK	0.953
27	L		24	30	EVRLVKV	0.958
28	V		25	31	VRLVKVT	0.973
29	K		26	32	RLVKVTR	0.995
30	V		27	33	LVKVTRN	1.016
31	T		28	34	VKVTRNT	1.037
32	R		29	35	KVTRNTA	1.051
33	N		30	36	VTRNTAR	1.05
34	T		31	37	TRNTARH	1.037
35	A		32	38	RNTARHE	1.012
36	R		33	39	NTARHEI	0.985
37	H		34	40	TARHEIQ	0.975
38	E		35	41	ARHEIQD	0.98
39	I		36	42	RHEIQDL	0.994
40	Q		37	43	HEIQDLN	1.006
41	D		38	44	EIQDLNV	1.008
42	L		39	45	IQDLNVT	1
43	N		40	46	QDLNVTS	1.003
44	V		41	47	DLNVTSQ	1.016
45	T		42	48	LNVTSQL	1.032
46	S		43	49	NVTSQLR	1.048
47	Q		44	50	VTSQLRG	1.05
48	L		45	51	TSQLRGD	1.05
49	R		46	52	SQLRGDF	1.048
50	G		47	53	QLRGDFE	1.042
51	D		48	54	LRGDFEA	1.025
52	F		49	55	RGDFEAA	0.993
53	E		50	56	GDFEAAH	0.964
54	A		51	57	DFEAAHT	0.94
55	A		52	58	FEAAHTA	0.935
56	H		53	59	EAAHTAG	0.952
57	T		54	60	AAHTAGD	0.973
58	A		55	61	AHTAGDN	0.998
59	G		56	62	HTAGDNA	1.011
60	D		57	63	TAGDNAH	1.005
61	N		58	64	AGDNAHV	0.987
62	A		59	65	GDNAHVV	0.958
63	H		60	66	DNAHVVA	0.936
64	V		61	67	NAHVVAT	0.932
65	V		62	68	AHVVATD	0.944
66	A		63	69	HVVATDT	0.971

67	T		64	70	VVATDTQ	1.012
68	D		65	71	VATDTQK	1.046
69	T		66	72	ATDTQKN	1.077
70	Q		67	73	TDTQKNT	1.097
71	K		68	74	DTQKNTV	1.089
72	N		69	75	TQKNTVY	1.062
73	T		70	76	QKNTVYA	1.016
74	V		71	77	KNTVYAF	0.962
75	Y		72	78	NTVYAFA	0.925
76	A		73	79	TVYAFAR	0.915
77	F		74	80	VYAFARD	0.929
78	A		75	81	YAFARDG	0.959
79	R		76	82	AFARDGF	0.985
80	D		77	83	FARDGFA	0.995
81	G		78	84	ARDGFAT	0.997
82	F		79	85	RDGFATT	0.993
83	A		80	86	DGFATTE	1.002
84	T		81	87	GFATTEE	1.022
85	T		82	88	FATTEEF	1.032
86	E		83	89	ATTEEFL	1.031
87	E		84	90	TTEEFLL	1.01
88	F		85	91	TEEFLLR	0.975
89	L		86	92	EEFLLRL	0.952
90	L		87	93	EFLLRLG	0.947
91	R		88	94	FLLRLGK	0.957
92	L		89	95	LLRLGKH	0.978
93	G		90	96	LRLGKHF	0.993
94	K		91	97	RLGKHFT	1.001
95	H		92	98	LGKHFTE	1.002
96	F		93	99	GKHFTEG	1.006
97	T		94	100	KHFTEGF	1.014
98	E		95	101	HFTEGFD	1.014
99	G		96	102	FTEGFDW	1.002
100	F		97	103	TEGFDWV	0.978
101	D		98	104	EGFDWVT	0.959
102	W		99	105	GFDWVTG	0.962
103	V		100	106	FDWVTGG	0.989
104	T		101	107	DWVTGGR	1.031
105	G		102	108	WVTGGRW	1.056
106	G		103	109	VTGGRWA	1.052
107	R		104	110	TGGRWAA	1.022
108	W		105	111	GGRWAAQ	0.986
109	A		106	112	GRWAAQQ	0.965
110	A		107	113	RWAAQQF	0.961
111	Q		108	114	WAAQQFF	0.967
112	Q		109	115	AAQQFFW	0.966
113	F		110	116	AQQFFWD	0.962
114	F		111	117	QQFFWDR	0.965
115	W		112	118	QFFWDRI	0.974
116	D		113	119	FFWDRIN	0.996
117	R		114	120	FWDRIND	1.015
118	I		115	121	WDRINDH	1.023
119	N		116	122	DRINDHD	1.021
120	D		117	123	RINDHDH	1.01

121	H		118	124	INDHDHA	0.988
122	D		119	125	NDHDHAF	0.963
123	H		120	126	DHDHAFS	0.95
124	A		121	127	HDHAFSR	0.946
125	F		122	128	DHAFSRN	0.969
126	S		123	129	HAFSRNK	1.007
127	R		124	130	AFSRNKS	1.045
128	N		125	131	FSRNKSE	1.078
129	K		126	132	SRNKSEV	1.088
130	S		127	133	RNKSEVR	1.083
131	E		128	134	NKSEVRT	1.063
132	V		129	135	KSEVRTA	1.037
133	R		130	136	SEVRTAV	1.016
134	T		131	137	EVRTAVL	0.995
135	A		132	138	VRTAVLE	0.973
136	V		133	139	RTAVLEI	0.961
137	L		134	140	TAVLEIS	0.961
138	E		135	141	AVLEISG	0.977
139	I		136	142	VLEISGS	1.015
140	S		137	143	LEISGSE	1.055
141	G		138	144	EISGSEQ	1.086
142	S		139	145	ISGSEQA	1.093
143	E		140	146	SGSEQAI	1.066
144	Q		141	147	GSEQAIV	1.023
145	A		142	148	SEQAIVA	0.976
146	I		143	149	EQAIVAG	0.939
147	V		144	150	QAIVAGI	0.93
148	A		145	151	AIVAGIE	0.939
149	G		146	152	IVAGIEG	0.959
150	I		147	153	VAGIEGL	0.984
151	E		148	154	AGIEGLT	0.996
152	G		149	155	GIEGLTV	0.994
153	L		150	156	IEGLTVL	0.98
154	T		151	157	EGLTVLK	0.971
155	V		152	158	GLTVLKS	0.981
156	L		153	159	LTVLKST	1.007
157	K		154	160	TVLKSTG	1.05
158	S		155	161	VLKSTGS	1.09
159	T		156	162	LKSTGSE	1.108
160	G		157	163	KSTGSEF	1.106
161	S		158	164	STGSEFH	1.081
162	E		159	165	TGSEFHG	1.036
163	F		160	166	GSEFHGF	0.992
164	H		161	167	SEFHGFP	0.969
165	G		162	168	EFHGFPR	0.963
166	F		163	169	FHGFPRD	0.979
167	P		164	170	HGFPRDK	1.012
168	R		165	171	GFPRDKY	1.026
169	D		166	172	FPRDKYT	1.035
170	K		167	173	PRDKYTT	1.039
171	Y		168	174	RDKYTTL	1.026
172	T		169	175	DKYTTLQ	1.025
173	T		170	176	KYTTLQE	1.026
174	L		171	177	YTTLQET	1.027

175	Q	172	178	TTLQETT	1.042
176	E	173	179	TLQETTD	1.053
177	T	174	180	LQETTDR	1.057
178	T	175	181	QETTDRI	1.048
179	D	176	182	ETTDRIL	1.024
180	R	177	183	TTDRILA	0.993
181	I	178	184	TDRILAT	0.969
182	L	179	185	DRILATD	0.967
183	A	180	186	RILATDV	0.98
184	T	181	187	ILATDVS	0.996
185	D	182	188	LATDVSA	1.01
186	V	183	189	ATDVSAR	1
187	S	184	190	TDVSARW	0.982
188	A	185	191	DVSARWR	0.965
189	R	186	192	VSARWRY	0.943
190	W	187	193	SARWRYN	0.94
191	R	188	194	ARWRYNT	0.948
192	Y	189	195	RWRYNTV	0.962
193	N	190	196	WRYNTVE	0.98
194	T	191	197	RYNTVEV	0.99
195	V	192	198	YNTVEVD	0.984
196	E	193	199	NTVEVDF	0.97
197	V	194	200	TVEVDFD	0.958
198	D	195	201	VEVDFDA	0.945
199	F	196	202	EVDFDAV	0.938
200	D	197	203	VDFDAVY	0.929
201	A	198	204	DFDAVYA	0.921
202	V	199	205	FDAVYAS	0.913
203	Y	200	206	DAVYASV	0.911
204	A	201	207	AVYASVR	0.927
205	S	202	208	VYASVRG	0.948
206	V	203	209	YASVRGL	0.972
207	R	204	210	ASVRGLL	0.989
208	G	205	211	SVRGLLL	0.989
209	L	206	212	VRGLLLK	0.984
210	L	207	213	RGLLLKA	0.976
211	L	208	214	GLLLKAF	0.972
212	K	209	215	LLLKAFA	0.97
213	A	210	216	LLKAFAE	0.967
214	F	211	217	LKAFAET	0.971
215	A	212	218	KAFAETH	0.979
216	E	213	219	AFAETHS	0.987
217	T	214	220	FAETHSL	0.992
218	H	215	221	AETHSLA	0.98
219	S	216	222	ETHSLAL	0.963
220	L	217	223	THSLALQ	0.953
221	A	218	224	HSLALQQ	0.961
222	L	219	225	SLALQQT	0.989
223	Q	220	226	LALQQTM	1.012
224	Q	221	227	ALQQTMY	1.023
225	T	222	228	LQQTMYE	1.006
226	M	223	229	QQTMYEM	0.973
227	Y	224	230	QTMYEMG	0.957
228	E	225	231	TMYEMGR	0.953

229	M	226	232	MYEMGRA	0.965
230	G	227	233	YEMGRAV	0.981
231	R	228	234	EMGRAVI	0.976
232	A	229	235	MGRAVIE	0.966
233	V	230	236	GRAVIET	0.962
234	I	231	237	RAVIETH	0.964
235	E	232	238	AVIETHP	0.987
236	T	233	239	VIETHPE	1.01
237	H	234	240	IETHPEI	1.023
238	P	235	241	ETHPEID	1.039
239	E	236	242	THPEIDE	1.04
240	I	237	243	HPEIDEI	1.038
241	D	238	244	PEIDEIK	1.043
242	E	239	245	EIDEIKM	1.032
243	I	240	246	IDEIKMS	1.018
244	K	241	247	DEIKMSL	1.003
245	M	242	248	EIKMSLP	0.987
246	S	243	249	IKMSLPN	0.99
247	L	244	250	KMSLPNK	1.01
248	P	245	251	MSLPNKH	1.031
249	N	246	252	SLPNKHH	1.039
250	K	247	253	LPNKHHF	1.023
251	H	248	254	PNKHHFL	0.988
252	H	249	255	NKHHFLV	0.952
253	F	250	256	KHHFLVD	0.93
254	L	251	257	HHFLVDL	0.924
255	V	252	258	HFLVDLQ	0.934
256	D	253	259	FLVDLQP	0.957
257	L	254	260	LVDLQPF	0.976
258	Q	255	261	VDLQPFG	0.998
259	P	256	262	DLQPFGQ	1.021
260	F	257	263	LQPFGQD	1.03
261	G	258	264	QPFGQDN	1.052
262	Q	259	265	PFGQDNP	1.07
263	D	260	266	FGQDNPN	1.076
264	N	261	267	GQDNPNE	1.086
265	P	262	268	QDNPNEV	1.072
266	N	263	269	DNPNEVF	1.049
267	E	264	270	NPNEVFY	1.01
268	V	265	271	PNEVFYA	0.963
269	F	266	272	NEVFYAA	0.929
270	Y	267	273	EVFYAAD	0.915
271	A	268	274	VFYAADR	0.931
272	A	269	275	FYAADRP	0.967
273	D	270	276	YAADRPY	1.004
274	R	271	277	AADRPYG	1.019
275	P	272	278	ADRPYGL	1.017
276	Y	273	279	DRPYGLI	0.99
277	G	274	280	RPYGLIE	0.961
278	L	275	281	PYGLIEA	0.95
279	I	276	282	YGLIEAT	0.942
280	E	277	283	GLIEATI	0.953
281	A	278	284	LIEATIQ	0.971
282	T	279	285	IEATIQR	0.982

283	I		280	286	EATIQRE	1.005
284	Q		281	287	ATIQREG	1.029
285	R		282	288	TIQREGS	1.057
286	E		283	289	IQREGSR	1.084
287	G		284	290	QREGSRA	1.097
288	S		285	291	REGSRAD	1.089
289	R		286	292	EGSRADH	1.06
290	A		287	293	GSRADHP	1.026
291	D		288	294	SRADHPI	0.986
292	H		289	295	RADHPIW	0.958
293	P		290	296	ADHPIWS	0.946
294	I		291	297	DHPIWSN	0.945
295	W		292	298	HPIWSNI	0.952
296	S		293	299	PIWSNIA	0.963
297	N		294	300	IWSNIAG	0.962
298	I		295	301	WSNIAGF	0.95

Table A.III.6 Karplus & Schulz Flexibility scores of 4R8X sequence to predict peptides which can act as B-cell epitope (Uricase from *Bacillus fastidious*)

Position	Residue	Start	End	Peptide	Score
4	T	1	7	AERTMFY	0.986
5	M	2	8	ERTMFYG	0.96
6	F	3	9	RTMFYGK	0.959
7	Y	4	10	TMFYGKG	0.981
8	G	5	11	MFYGKGD	1.019
9	K	6	12	FYGKGDV	1.05
10	G	7	13	YGKGDVY	1.051
11	D	8	14	GKGDVYV	1.024
12	V	9	15	KGDVYVF	0.975
13	Y	10	16	GDVYVFR	0.94
14	V	11	17	DVYVFRT	0.934
15	F	12	18	VYVFRTY	0.944
16	R	13	19	YVFRTYA	0.967
17	T	14	20	VFRTYAN	0.982
18	Y	15	21	FRTYANP	0.986
19	A	16	22	RTYANPL	0.991
20	N	17	23	TYANPLK	1.003
21	P	18	24	YANPLKG	1.018
22	L	19	25	ANPLKGL	1.024
23	K	20	26	NPLKGLK	1.034
24	G	21	27	PLKGLKQ	1.033
25	L	22	28	LKGLKQI	1.027
26	K	23	29	KGLKQIP	1.035
27	Q	24	30	GLKQIPE	1.038
28	I	25	31	LKQIPES	1.052
29	P	26	32	KQIPESN	1.069
30	E	27	33	QIPESNF	1.069
31	S	28	34	IPESNFT	1.066
32	N	29	35	PESNFTE	1.049
33	F	30	36	ESNFTEK	1.037
34	T	31	37	SNFTEKH	1.038

35	E	32	38	NFTEKHN	1.038
36	K	33	39	FTEKHNT	1.041
37	H	34	40	TEKHNTI	1.027
38	N	35	41	EKHNTIF	1.005
39	T	36	42	KHNTIFG	0.981
40	I	37	43	HNTIFGM	0.953
41	F	38	44	NTIFGMN	0.939
42	G	39	45	TIFGMNA	0.938
43	M	40	46	IFGMNAK	0.95
44	N	41	47	FGMNAKV	0.968
45	A	42	48	GMNAKVA	0.98
46	K	43	49	MNAKVAL	0.98
47	V	44	50	NAKVALK	0.973
48	A	45	51	AKVALKG	0.981
49	L	46	52	KVALKGE	1.003
50	K	47	53	VALKGEQ	1.035
51	G	48	54	ALKGEQL	1.058
52	E	49	55	LKGEQLL	1.056
53	Q	50	56	KGEQLLT	1.038
54	L	51	57	GEQLLTS	1.018
55	L	52	58	EQLLTSF	1.004
56	T	53	59	QLLTSFT	1.006
57	S	54	60	LLTSFTE	1.013
58	F	55	61	LTSFTEG	1.026
59	T	56	62	TSFTEGD	1.046
60	E	57	63	SFTEGDN	1.065
61	G	58	64	FTEGDNS	1.079
62	D	59	65	TEGDNSL	1.075
63	N	60	66	EGDNSLV	1.058
64	S	61	67	GDNSLVV	1.023
65	L	62	68	DNSLVVA	0.985
66	V	63	69	NSLVVAT	0.964
67	V	64	70	SLVVATD	0.958
68	A	65	71	LVVATDS	0.974
69	T	66	72	VVATDSM	0.995
70	D	67	73	VATDSMK	1.013
71	S	68	74	ATDSMKN	1.022
72	M	69	75	TDSMKNF	1.014
73	K	70	76	DSMKNFI	1.002
74	N	71	77	SMKNFIQ	0.985
75	F	72	78	MKNFIQR	0.971
76	I	73	79	KNFIQRH	0.968
77	Q	74	80	NFIQRHA	0.969
78	R	75	81	FIQRHAA	0.967
79	H	76	82	IQRHAAS	0.958
80	A	77	83	QRHAASY	0.947
81	A	78	84	RHAASYE	0.945
82	S	79	85	HAASYEG	0.954
83	Y	80	86	AASYEGA	0.977
84	E	81	87	ASYEGAT	0.998
85	G	82	88	SYEGATL	1.005
86	A	83	89	YEGATLE	1.006
87	T	84	90	EGATLEG	1
88	L	85	91	GATLEGF	0.994

89	E	86	92	ATLEGFL	0.992
90	G	87	93	TLEGFLQ	0.98
91	F	88	94	LEGFLQY	0.962
92	L	89	95	EGFLQYV	0.946
93	Q	90	96	GFLQYVC	0.924
94	Y	91	97	FLQYVCE	0.915
95	V	92	98	LQYVCEA	0.914
96	C	93	99	QYVCEAF	0.913
97	E	94	100	YVCEAFL	0.919
98	A	95	101	VCEAFLA	0.924
99	F	96	102	CEAFLAK	0.934
100	L	97	103	EAFLAKY	0.948
101	A	98	104	AFLAKYS	0.959
102	K	99	105	FLAKYSH	0.969
103	Y	100	106	LAKYSHL	0.966
104	S	101	107	AKYSHLD	0.959
105	H	102	108	KYSHLDA	0.953
106	L	103	109	YSHLDAV	0.945
107	D	104	110	SHLDAVR	0.939
108	A	105	111	HLDAVRL	0.938
109	V	106	112	LDAVRLE	0.934
110	R	107	113	DAVRLEA	0.939
111	L	108	114	AVRLEAK	0.958
112	E	109	115	VRLEAKE	0.98
113	A	110	116	RLEAKEY	1.003
114	K	111	117	LEAKEYA	1.009
115	E	112	118	EAKEYAF	0.993
116	Y	113	119	AKEYAFD	0.976
117	A	114	120	KEYAFDD	0.97
118	F	115	121	EYAFDDI	0.979
119	D	116	122	YAFDDIQ	0.998
120	D	117	123	AFDDIQV	1.006
121	I	118	124	FDDIQVG	1.002
122	Q	119	125	DDIQVGT	0.995
123	V	120	126	DIQVGTD	0.999
124	G	121	127	IQVGTDK	1.016
125	T	122	128	QVGTDKG	1.037
126	D	123	129	VGTDKGV	1.044
127	K	124	130	GTDKGVV	1.035
128	G	125	131	TDKGVVT	1.017
129	V	126	132	DKGVVTS	1.007
130	V	127	133	KGVVTSD	1.015
131	T	128	134	GVVTSDL	1.033
132	S	129	135	VVTSDLV	1.044
133	D	130	136	VTSDLVF	1.032
134	L	131	137	TSDLVFR	1.006
135	V	132	138	SDLVFRK	0.985
136	F	133	139	DLVFRKS	0.99
137	R	134	140	LVFRKSR	1.018
138	K	135	141	VFRKSRN	1.056
139	S	136	142	FRKSRNE	1.082
140	R	137	143	RKSRNEY	1.077
141	N	138	144	KSRNEYV	1.057
142	E	139	145	SRNEYVT	1.02

143	Y	140	146	RNEYVTA	0.987
144	V	141	147	NEYVTAT	0.969
145	T	142	148	EYVTATV	0.962
146	A	143	149	YVTATVE	0.966
147	T	144	150	VTATVEV	0.962
148	V	145	151	TATVEVA	0.959
149	E	146	152	ATVEVAR	0.956
150	V	147	153	TVEVART	0.959
151	A	148	154	VEVARTA	0.978
152	R	149	155	EVARTAS	1.001
153	T	150	156	VARTASG	1.029
154	A	151	157	ARTASGT	1.051
155	S	152	158	RTASGTE	1.066
156	G	153	159	TASGTEV	1.066
157	T	154	160	ASGTEVV	1.046
158	E	155	161	SGTEVVE	1.023
159	V	156	162	GTEVVEQ	0.999
160	V	157	163	TEVVEQA	0.99
161	E	158	164	EVVEQAS	1.002
162	Q	159	165	VVEQASG	1.016
163	A	160	166	VEQASGI	1.023
164	S	161	167	EQASGIA	1.016
165	G	162	168	QASGIAD	0.995
166	I	163	169	ASGIADI	0.971
167	A	164	170	SGIADIQ	0.955
168	D	165	171	GIADIQL	0.946
169	I	166	172	IADIQLI	0.941
170	Q	167	173	ADIQLIK	0.943
171	L	168	174	DIQLIKV	0.95
172	I	169	175	IQLIKVS	0.963
173	K	170	176	QLIKVSG	0.993
174	V	171	177	LIKVSGS	1.028
175	S	172	178	IKVSGSS	1.06
176	G	173	179	KVSGSSF	1.078
177	S	174	180	VSGSSFY	1.067
178	S	175	181	SGSSFYG	1.034
179	F	176	182	GSSFYGY	0.99
180	Y	177	183	SSFYGYI	0.951
181	G	178	184	SFYGYII	0.924
182	Y	179	185	FYGYIID	0.922
183	I	180	186	YGYIIDE	0.934
184	I	181	187	GYIIDEY	0.954
185	D	182	188	YIIDEYT	0.987
186	E	183	189	IIDEYTT	1.006
187	Y	184	190	IDEYTTL	1.015
188	T	185	191	DEYTTLA	1.017
189	T	186	192	EYTTLAE	1.001
190	L	187	193	YTTLAEA	0.988
191	A	188	194	TTLAEAT	0.985
192	E	189	195	TLAEATD	0.988
193	A	190	196	LAEATDR	1.005
194	T	191	197	AEATDRP	1.025
195	D	192	198	EATDRPL	1.034
196	R	193	199	ATDRPLY	1.025

197	P		194	200	TDRPLYI	1
198	L		195	201	DRPLYIF	0.961
199	Y		196	202	RPLYIFL	0.928
200	I		197	203	PLYIFLN	0.913
201	F		198	204	LYIFLNI	0.916
202	L		199	205	YIFLNIG	0.93
203	N		200	206	IFLNIGW	0.942
204	I		201	207	FLNIGWA	0.946
205	G		202	208	LNIGWAY	0.936
206	W		203	209	NIGWAYE	0.934
207	A		204	210	IGWAYEN	0.947
208	Y		205	211	GWAYENQ	0.979
209	E		206	212	WAYENQD	1.022
210	N		207	213	AYENQDD	1.061
211	Q		208	214	YENQDDA	1.083
212	D		209	215	ENQDDAK	1.083
213	D		210	216	NQDDAKG	1.083
214	A		211	217	QDDAKGD	1.077
215	K		212	218	DDAKGDN	1.079
216	G		213	219	DAKGDNP	1.088
217	D		214	220	AKGDNPA	1.083
218	N		215	221	KGDNPAN	1.075
219	P		216	222	GDNPANY	1.051
220	A		217	223	DNPANYV	1.016
221	N		218	224	NPANYVA	0.977
222	Y		219	225	PANYVAA	0.947
223	V		220	226	ANYVAAE	0.934
224	A		221	227	NYVAAEQ	0.94
225	A		222	228	YVAAEQV	0.96
226	E		223	229	VAAEQVR	0.985
227	Q		224	230	AAEQVRD	1.006
228	V		225	231	AEQVRDI	1.012
229	R		226	232	EQVRDIA	1.006
230	D		227	233	QVRDIAA	0.99
231	I		228	234	VRDIAAS	0.96
232	A		229	235	RDIAASV	0.941
233	A		230	236	DIAASVF	0.929
234	S		231	237	IAASVFH	0.923
235	V		232	238	AASVFHT	0.928
236	F		233	239	ASVFHTL	0.933
237	H		234	240	SVFHTLD	0.947
238	T		235	241	VFHTLDN	0.971
239	L		236	242	FHTLDNK	1.005
240	D		237	243	HTLDNKS	1.041
241	N		238	244	TLDNKSI	1.061
242	K		239	245	LDNKSIQ	1.056
243	S		240	246	DNKSIQH	1.033
244	I		241	247	NKSIQHL	0.996
245	Q		242	248	KSIQHLI	0.957
246	H		243	249	SIQHLIY	0.929
247	L		244	250	IQHLIYH	0.906
248	I		245	251	QHLIYHI	0.889
249	Y		246	252	HLIYHIG	0.884
250	H		247	253	LIYHIGL	0.89

251	I	248	254	IYHIGLT	0.902
252	G	249	255	YHIGLTI	0.916
253	L	250	256	HIGLTIL	0.929
254	T	251	257	IGLTILD	0.941
255	I	252	258	GLTILDR	0.957
256	L	253	259	LTILDRF	0.976
257	D	254	260	TILDRFP	0.999
258	R	255	261	ILDRFPQ	1.011
259	F	256	262	LDRFPQL	1.014
260	P	257	263	DRFPQLT	1.021
261	Q	258	264	RFPQLTE	1.018
262	L	259	265	FPQLTEV	1.016
263	T	260	266	PQLTEVN	1.016
264	E	261	267	QLTEVNF	1
265	V	262	268	LTEVNFG	0.987
266	N	263	269	TEVNFGT	0.984
267	F	264	270	EVNFGTN	0.994
268	G	265	271	VNFGTNN	1.023
269	T	266	272	NFGTNNR	1.052
270	N	267	273	FGTNNRT	1.074
271	N	268	274	GTNNRTW	1.072
272	R	269	275	TNNRTWD	1.056
273	T	270	276	NNRTWDT	1.042
274	W	271	277	NRTWDTV	1.021
275	D	272	278	RTWDTVV	1.013
276	T	273	279	TWDTVVE	1.003
277	V	274	280	WDTVVEG	0.997
278	V	275	281	DTVVEGT	1.009
279	E	276	282	TVVEGTD	1.027
280	G	277	283	VVEGTDG	1.049
281	T	278	284	VEGTDGF	1.052
282	D	279	285	EGTDGFK	1.046
283	G	280	286	GTDGFKG	1.034
284	F	281	287	TDGFKGA	1.016
285	K	282	288	DGFKGAV	1.01
286	G	283	289	GFKGAVF	0.993
287	A	284	290	FKGAVFT	0.976
288	V	285	291	KGAVFTE	0.974
289	F	286	292	GAVFTEP	0.982
290	T	287	293	AVFTEPR	1.009
291	E	288	294	VFTEPRP	1.035
292	P	289	295	FTEPRPP	1.052
293	R	290	296	TEPRPPF	1.051
294	P	291	297	EPRPPFG	1.036
295	P	292	298	PRPPFGF	1.013
296	F	293	299	RPPFGFQ	0.986
297	G	294	300	PPFGFQG	0.975
298	F	295	301	PFGFQGF	0.971
299	Q	296	302	FGFQGFS	0.974
300	G	297	303	GFQGFSV	0.975
301	F	298	304	FQGFSVH	0.963
302	S	299	305	QGFSVHQ	0.957
303	V	300	306	GFSVHQE	0.962
304	H	301	307	FSVHQED	0.983

305	Q		302	308	SVHQEDL	1.01
306	E		303	309	VHQEDLA	1.027
307	D		304	310	HQEDLAR	1.028
308	L		305	311	QEDLARE	1.02
309	A		306	312	EDLAREK	1.018
310	R		307	313	DLAREKA	1.028
311	E		308	314	LAREKAS	1.035
312	K		309	315	AREKASA	1.039
313	A		310	316	REKASAN	1.03
314	S		311	317	EKASANS	1.024
315	A		312	318	KASANSE	1.031
316	N		313	319	ASANSEY	1.034
317	S		314	320	SANSEYV	1.036
318	E		315	321	ANSEYVA	1.013

Table A.III.7 Peptides identified which have propensity to act as CD4+ T-cell epitope from 2YZB and 4R8X

S.NO	Organism	Peptide	Start position	End position	Median percentile rank
1	*Arthrobacter globiformis*	AVYASVRGLLLKAFA	201	215	10.78
		LKAFAETHSLALQQT	211	225	11.59
		VRGLLLKAFAETHSL	206	220	12.56
		EVDFDAVYASVRGLL	196	210	14.835
		AFARDGFATTEEFLL	76	90	18.76
2	*Bacillus fastidious*	IADIQLIKVSGSSFY	166	180	8.77
		DVYVFRTYANPLKGL	11	25	9.205
		HLIYHIGLTILDRFP	246	260	9.325
		RPLYIFLNIGWAYEN	196	210	10.99
		GSSFYGYIIDEYTTL	176	190	13.46
		ATLEGFLQYVCEAFL	86	100	13.65
		FLQYVCEAFLAKYSH	91	105	13.92
		KVALKGEQLLTSFTE	46	60	15.715
		EQVRDIAASVFHTLD	226	240	15.915
		NKSIQHLIYHIGLTI	241	255	16.87
		SMKNFIQRHAASYEG	71	85	17.05

		LIKVSGSSFYGYIID	171	185	18.385
		GYIIDEYTTLAEATD	181	195	18.625
		TSDLVFRKSRNEYVT	131	145	19.485

Table A.III.8 IEDB scores obtained after replacing all the amino acid with the hot-spot residues from B-cell epitopes

Amino acid	Predicted residue scores					
	2YZB Surface-accessibility	2YZB Hydrophilicity	2YZB Flexibility	4R8X Surface-accessibility	4R8X Hydrophilicity	4R8X Flexibility
Alanine (A)	3.037	5.814	1.058	5.032	6.7	1.068
Cysteine (C)	**1.64**	5.714	1.038	2.729	6.6	1.048
Aspartic acid (D)	4.904	6.943	1.098	8.073	7.829	1.06
Glutamic acid (E)	5.075	6.629	1.113	8.349	7.514	1.076
Phenylalanine (F)	2.617	4.2	1.031	4.342	5.086	1.04
Glycine (G)	2.978	6.329	1.125	4.934	7.214	1.088
Histidine (H)	4.04	5.814	1.044	6.67	6.7	1.053
Isoleucine (I)	2.131	4.371	1.049	3.542	5.257	1.058
Lysine (K)	5.805	6.329	1.113	9.527	7.214	1.075
Leucine (L)	2.496	4.2	1.04	4.143	5.086	1.05
Methionine (M)	2.978	4.914	1.035	4.934	5.8	1.045
Asparagine (N)	4.733	6.514	1.119	7.795	7.4	1.081
Proline (P)	4.561	5.814	1.103	7.516	6.7	1.066
Glutamine (Q)	5.075	6.371	1.131	8.349	7.257	1.093
Arginine (R)	5.694	6.114	1.099	9.348	7.0	1.062
Serine (S)	3.982	6.443	1.132	6.575	7.329	1.094
Threonine (T)	4.272	6.257	1.108	7.048	7.143	1.07
Valine(V)	2.253	4.986	1.044	**3.743**	5.871	1.053
Tryptophan (W)	3.157	**4.086**	**1.029**	5.227	**4.971**	**1.039**
Tyrosine (Y)	4.618	5.243	1.038	7.61	6.129	1.048

Table A.III.9 T-cell epitopic regions and mutant analogues obtained from *Arthrobacter globiformis* (2YZB)

Protein number	Peptide	Peptide ID	Start position	End position	Median percent-ile rank	Median differe-nce	C termina-l neighbo-r 1(Media-n)	C terminal neighbo r 2 (Media-n)	N terminal neighbor 1(Median)	N terminal Neighbor 2 (Median)	Deimmunization score
1	AVYASVRG LLLKAFA	wild	201	215	10.78	0	12.56	11.59	14.835	39.98	NA
1	AVDASVRG LLLKAFA	Y203D	201	215	24.585	13.805	NA	NA	26.435	54.31	3
1	AVYASPRG LLLKAFA	V206P	201	215	23.615	12.835	14.045	NA	18.23	NA	3
1	AVYASVRG CLLKAFA	L209C	201	215	23.345	12.565	18.85	NA	17.26	NA	3
1	AVYASVRG GLLKAFA	L209G	201	215	23.045	12.265	19.4	NA	18.63	NA	3
1	AVEASVRG LLLKAFA	Y203E	201	215	22.9	12.12	NA	NA	20.865	44.255	3
1	AVGASVRG LLLKAFA	Y203G	201	215	22.47	11.69	NA	NA	23.655	48.955	3
1	AVYASRRG LLLKAFA	V206R	201	215	22.06	11.28	14.16	NA	16.78	NA	3
1	AVNASVRG LLLKAFA	Y203N	201	215	21.235	10.455	NA	NA	19.555	48.215	3
1	AVSASVRG LLLKAFA	Y203S	201	215	21.11	10.33	NA	NA	24.285	45.64	3
1	AVPASVRG LLLKAFA	Y203P	201	215	21.025	10.245	NA	NA	19.165	43.74	3
1	AVYASKRG LLLKAFA	V206K	201	215	20.57	9.79	14.205	NA	19.11	NA	3
1	AVYASGRG LLLKAFA	V206G	201	215	19.72	8.94	15.74	NA	20.53	NA	3
1	AVAASVRG LLLKAFA	Y203A	201	215	19.685	8.905	NA	NA	16.265	42.27	3
1	AVCASVRG LLLKAFA	Y203C	201	215	19.345	8.565	NA	NA	24.585	56.215	3

Table A.III.10 T-cell epitopic regions and mutant analogues obtained from *Bacillus fastidious* (4R8X)

Protein number	Peptide	Peptide ID	Start position	End position	Median percent-ile rank	Median differen-ce	C termina-l neighbor 1(Median)	C termin-al neighb-or 2 (Media-n)	N termina-l neighbo-r 1(Media-n)	N terminal Neighbor 2 (Median)	Deimmunizatio n score
1	IADIQLIKVS GSSFY	wild	166	180	8.77	0	18.385	13.46	27.24	38.8	NA
1	IADIQLPKVS GSSFY	I172P	166	180	51.955	43.185	32.875	NA	29.955	NA	3
1	IADIQLGKVS GSSFY	I172G	166	180	35.87	27.1	29.555	NA	39.445	NA	3
1	IADIQLDKVS GSSFY	I172D	166	180	30.42	21.65	32.42	NA	43.01	NA	3
1	IADIQLCKVS GSSFY	I172C	166	180	28.97	20.2	28.985	NA	42.27	NA	3
1	IADIQLHKVS GSSFY	I172H	166	180	27.895	19.125	27.55	NA	31.925	NA	3
1	IADIQLEKVS GSSFY	I172E	166	180	27.72	18.95	30.355	NA	34.16	NA	3
1	IADIQDIKVS GSSFY	L171D	166	180	26.44	17.67	23.335	NA	35.485	NA	3
1	IADIQLRKVS GSSFY	I172R	166	180	24.715	15.945	23.345	NA	31.5	NA	3
1	IADIQLKKVS GSSFY	I172K	166	180	24.3	15.53	25.625	NA	34.245	NA	3
1	IADIQLTKVS GSSFY	I172T	166	180	24.06	15.29	28.96	NA	28.06	NA	3
1	IADIQLSKVS GSSFY	I172S	166	180	23.14	14.37	29.91	NA	28.805	NA	3
1	IADPQLIKVS GSSFY	I169P	166	180	21.98	13.21	NA	NA	43.7	41.13	3
1	IADIQLNKVS GSSFY	I172N	166	180	21.675	12.905	26.945	NA	29.045	NA	3
1	IADIQEIKVS GSSFY	L171E	166	180	21.56	12.79	22.135	NA	34.71	NA	3
1	IADIQLIKDS GSSFY	V174D	166	180	20.955	12.185	27.88	NA	36.17	NA	3
1	IADIQLIGVS GSSFY	K173G	166	180	20.885	12.115	24.15	NA	36.59	NA	3
1	IADIQLICVSG SSFY	K173C	166	180	20.645	11.875	24.835	NA	36.09	NA	3
1	IADIQLIDVS GSSFY	K173D	166	180	20.325	11.555	29.525	NA	36.825	NA	3
1	IADIQLIKVS GDSFY	S177D	166	180	19.895	11.125	28.89	14.775	NA	NA	3
1	IADIQLIKVS GCSFY	S177C	166	180	19.845	11.075	28.48	14.09	NA	NA	3
1	IADEQLIKVS GSSFY	I169E	166	180	19.77	11	NA	NA	42.84	41.37	3
1	IADIQLIKVD GSSFY	S175D	166	180	19.67	10.9	22.64	NA	27.525	NA	3
1	IADIQNIKVS GSSFY	L171N	166	180	19.435	10.665	22.525	NA	32.105	NA	3
1	IADIQLIKVC GSSFY	S175C	166	180	19.05	10.28	19.315	NA	27.525	NA	3
1	IADQQLIKVS GSSFY	I169Q	166	180	19.025	10.255	NA	NA	37.77	40.395	3
1	IADNQLIKVS GSSFY	I169N	166	180	18.845	10.075	NA	NA	40.225	42.06	3
1	IADIQLWKVS GSSFY	I172W	166	180	18.84	10.07	24.395	NA	27.77	NA	3
1	IADIQLIKVW	S175W	166	180	18.59	9.82	22.445	NA	27.27	NA	3

1	GSSFY										
1	IADIQLIKVG GSSFY	S175G	166	180	18.19	9.42	24.43	NA	27.585	NA	3
1	IADDQLIKVS GSSFY	I169D	166	180	18.03	9.26	NA	NA	49.46	40.755	3
1	IADIQLIEVSG SSFY	K173E	166	180	17.755	8.985	27.075	NA	40.98	NA	3
1	IADIQLIKWS GSSFY	V174W	166	180	17.695	8.925	19.45	NA	33.54	NA	3
1	IADIQLIKVS GSSDY	F179D	166	180	17.69	8.92	38.095	24.015	NA	NA	3
1	IADIQLIKVS GSSEY	F179E	166	180	17.66	8.89	36.84	23.21	NA	NA	3
1	IADIQLIKVS GSSPY	F179P	166	180	17.65	8.88	28.275	25.375	NA	NA	3
1	IADIDLIKVS GSSFY	Q170D	166	180	17.565	8.795	NA	NA	38.095	39.93	3
1	IADIQLAKVS GSSFY	I172A	166	180	17.49	8.72	28.885	NA	27.525	NA	3
1	IADIQGIKVS GSSFY	L171G	166	180	17.46	8.69	22.345	NA	29.825	NA	3
1	IADIQLIKVS GESFY	S177E	166	180	17.41	8.64	25.385	13.605	NA	NA	3
1	IADIQLIKCSG SSFY	V174C	166	180	17.385	8.615	27.265	NA	40.68	NA	3
1	IADIQPIKVSG SSFY	L171P	166	180	18.68	9.91	22.74	NA	26.855	NA	8
1	IADIQTIKVS GSSFY	L171T	166	180	17.625	8.855	20.63	NA	25.995	NA	8

Table A.III.11 Predictor of effects of single point protein mutation

Mutations	ΔΔG value (Kcal/mol)	Stability
D169C	-0.01	Neutral stability
N264W	0.17	Neutral stability
S139V	-0.03	Neutral stability
K215W	0.15	Neutral stability
I172P	-1.94	Neutral stability
Y203D	-1.06	Neutral stability
G216F	-0.20	Neutral stability
T159W	-0.16	Neutral stability

Stability Predictors:

ΔΔG: ΔG (New Protein)-ΔG (Wild type) in Kcal/mole, ΔΔG<-0.5: Large Decrease of Stability, ΔΔG>0.5: Large Increase of Stability, -0.5<=ΔΔG<=0.5: Neutral Stability

Table A.III.12 Statistics of Ramachandran plot analysis

Ramachandran plot analysis				
	4R8X Native	4R8X Mutant	4R99 Native	4R99 Mutant
Residues in most favoured regions [A,B,L]	1030 (91.8%	1020 (91.2%)	885 (85.4%)	887 (85.6%)
Residues in additional allowed regions [a,b,l,p]	92 (8.2%)	97 (8.7%)	151 (14.6%)	149 (14.4%)
Residues in generously allowed regions [~a,~b,~l,~p]	0 (0.0%)	1 (0.1%)	0 (0.0%)	0 (0.0%)
Residues in disallowed regions	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Number of non-glycine and non-proline residues	1122 (100.0%)	1118 (100.0%)	1036 (100.0%)	1036 (100.0%)
Number of end-residues (excl. Gly and Pro)	18	18	9	9
Number of glycine residues (shown as triangles)	85	85	76	76
Number of proline residues	32	36	28	28
Total number of residues	1257	1257	1149	1149
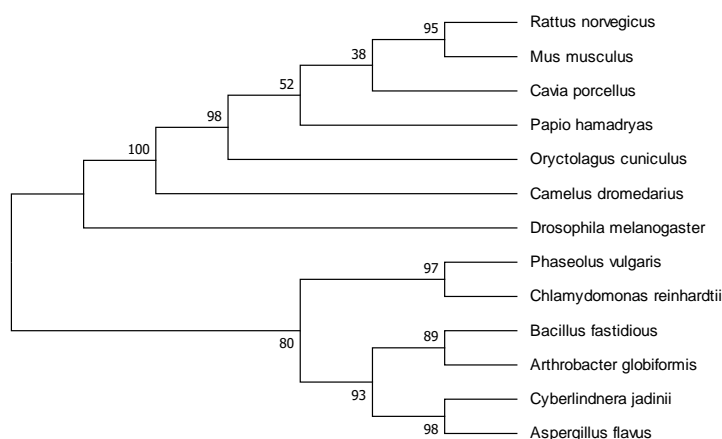
Figure A.III.1 Phylogenetic tree of uricase sequences of different organisms constructed by maximum parsimony method
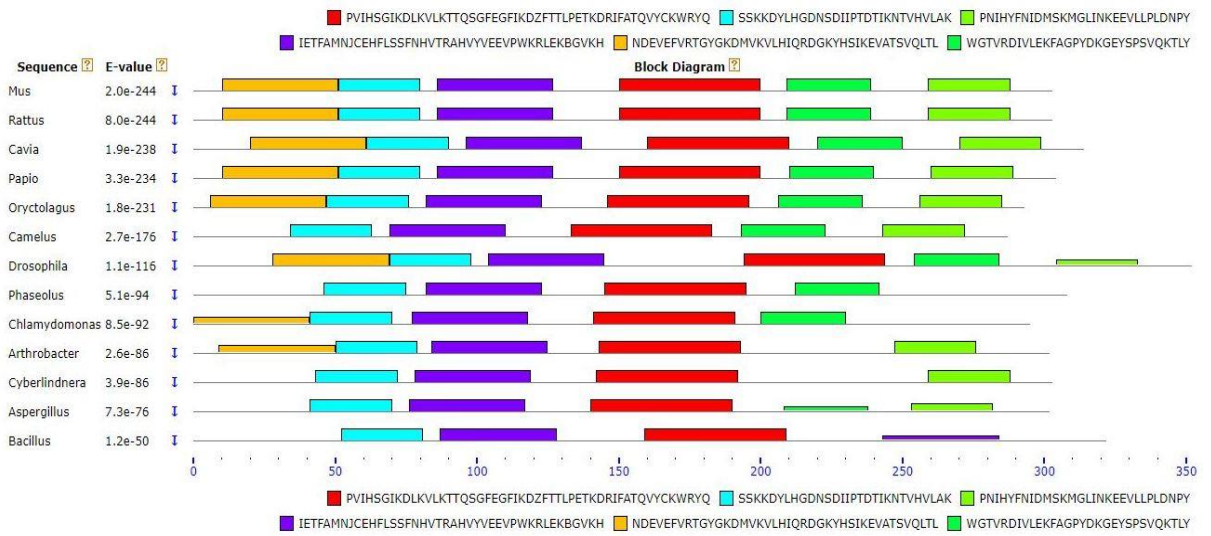
Figure A.III.2 Motifs obtained in uricase sequences from various sources represented in combined block diagram
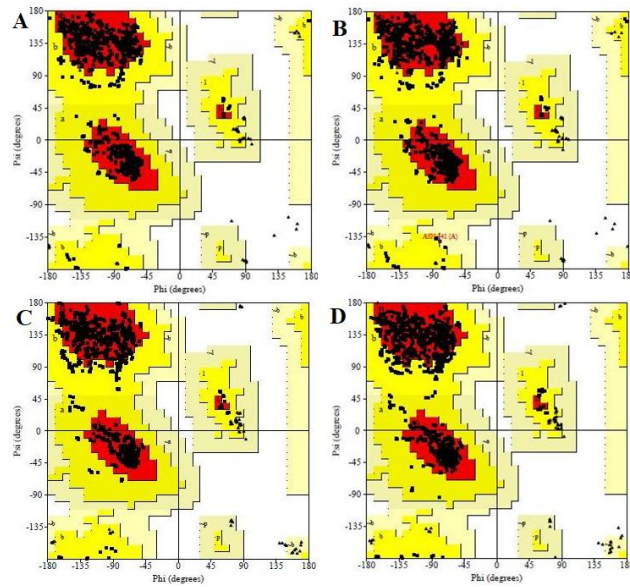
Figure A.III.3 Ramachandran plot of (A) Native uricase of *Bacillus fastidious* (B) Mutant uricase of *Bacillus fastidious* (C) Native uricase of *Arthrobacter globiformis* (D) Mutated uricase of *Arthrobacter globiformis*

265
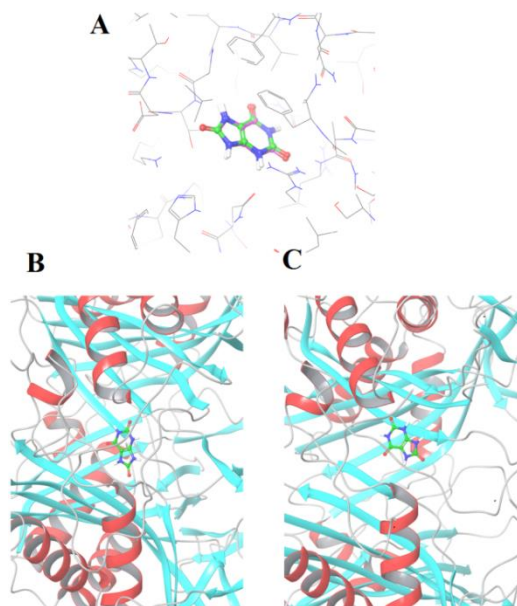
Figure A.III.4 (A) Superposition of cocrystal and re-docked uricase (MSD = 0.16 Å). (B) Docking pose of uric acid in the case of *Arthrobacter globiformis* (C) Docking pose of uric acid in the case of *Bacillus fastidious*
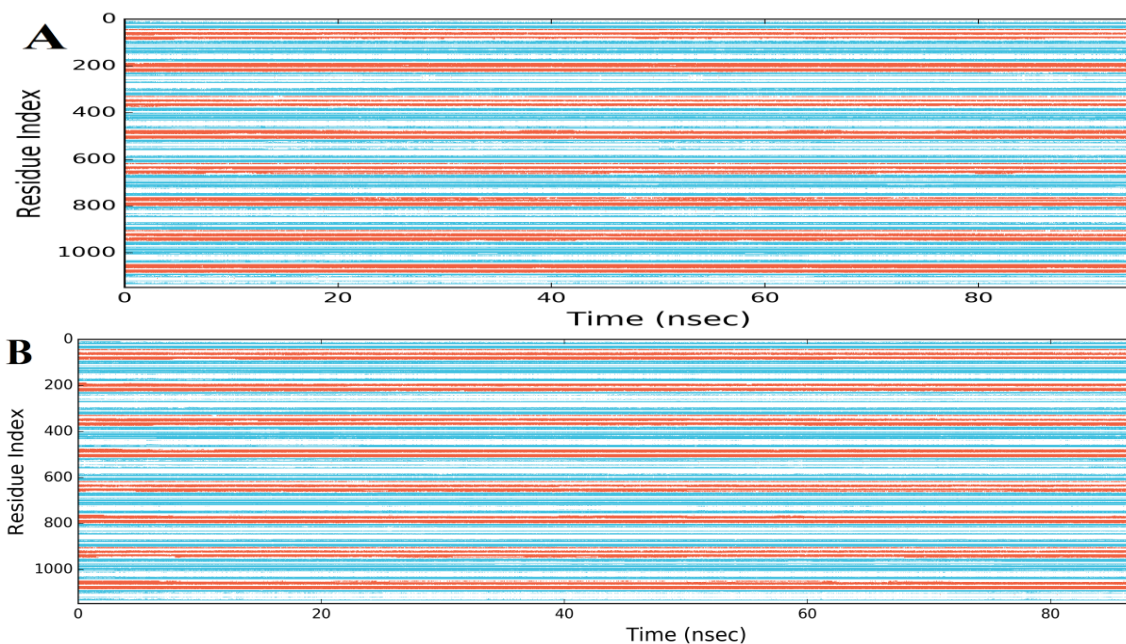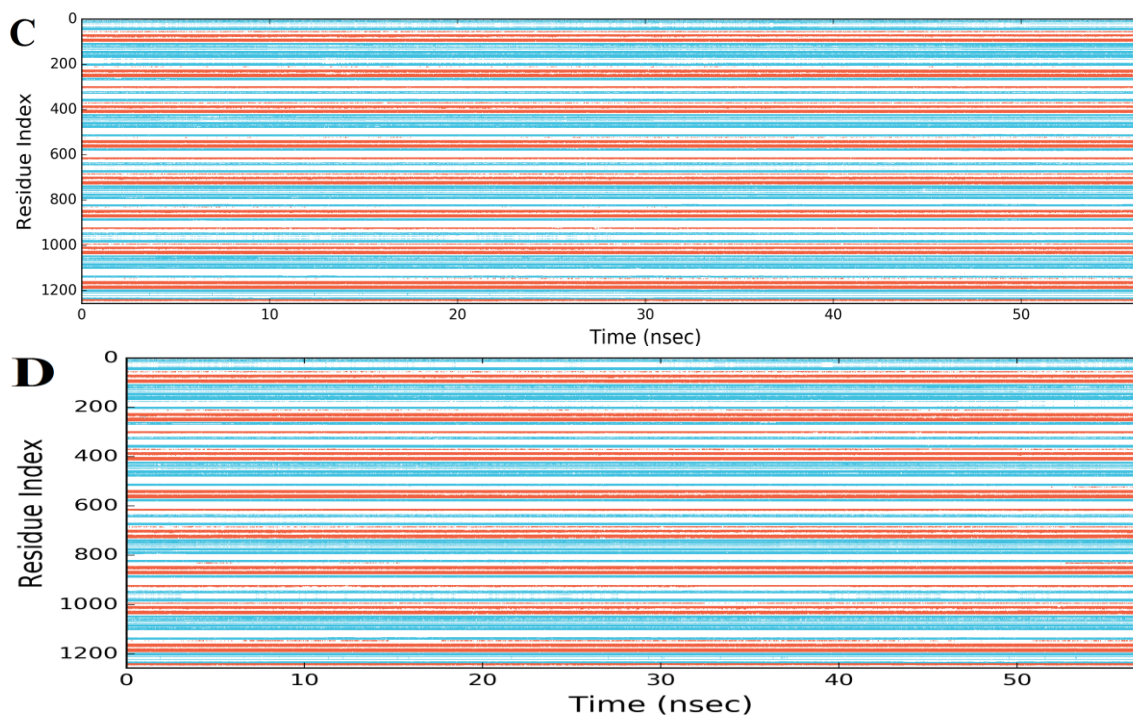
Figure A.III.5 The secondary structure content of uricase throughout the MD trajectory is illustrated (A) Wild *Arthrobacter*, (B) Mutated *Arthrobacter*, (C) Wild *Bacillus*, (D) Mutated *Bacillus*. The alpha helices, beta sheet and loops are shown in orange, blue, white colour respectively
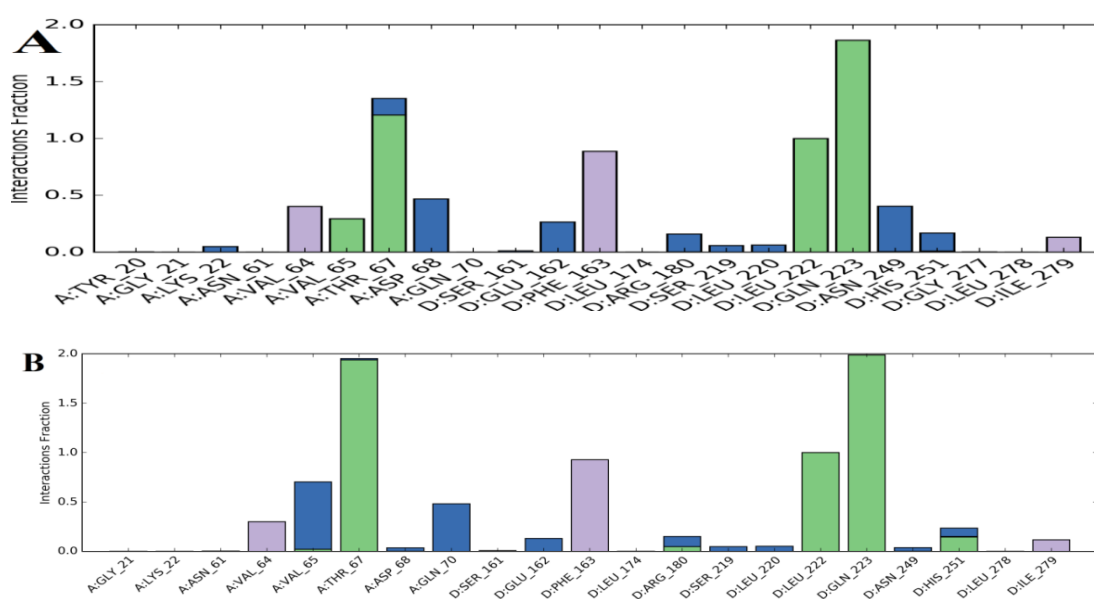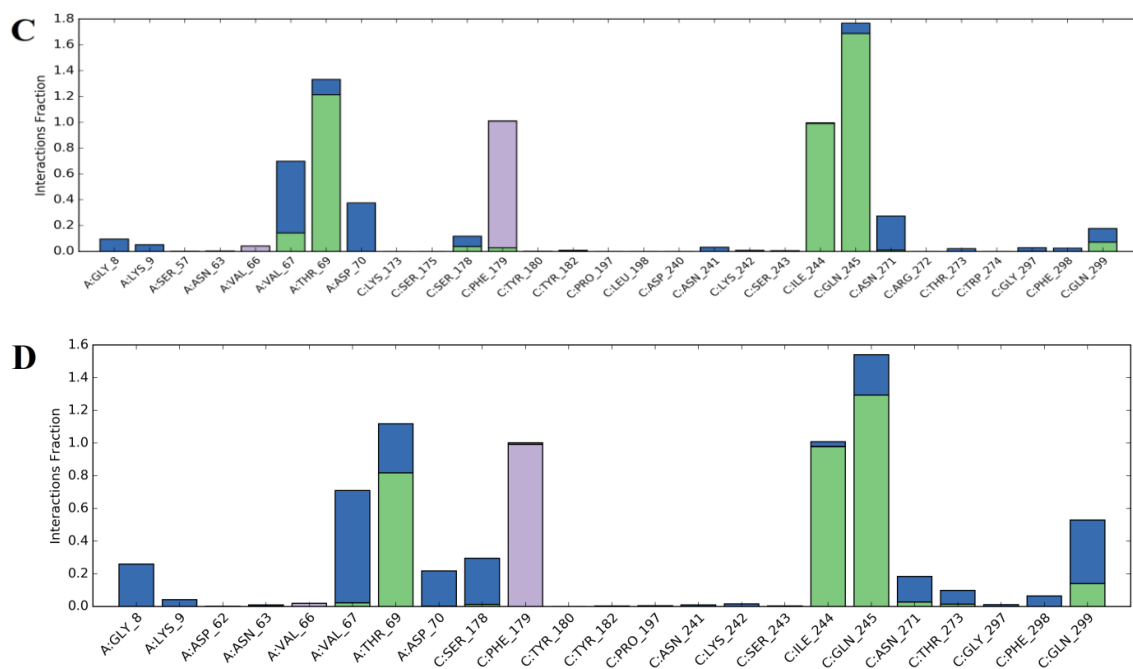
Figure A.III.6 Illustrating the interaction fraction in both wild type, mutated uricase sourced from (A) Native uricase of *Arthrobacter globiformis* (B) Mutated uricase of *Arthrobacter globiformis* (C) Native uricase of *Bacillus fastidious* (D) Mutated uricase of *Bacillus fastidious*. The green colour is representing the hydrogen bonding interaction, blue colour is representing the water mediated hydrogen bond, red colour is representing the ionic interaction and the violet colour is representing the hydrophobic interaction
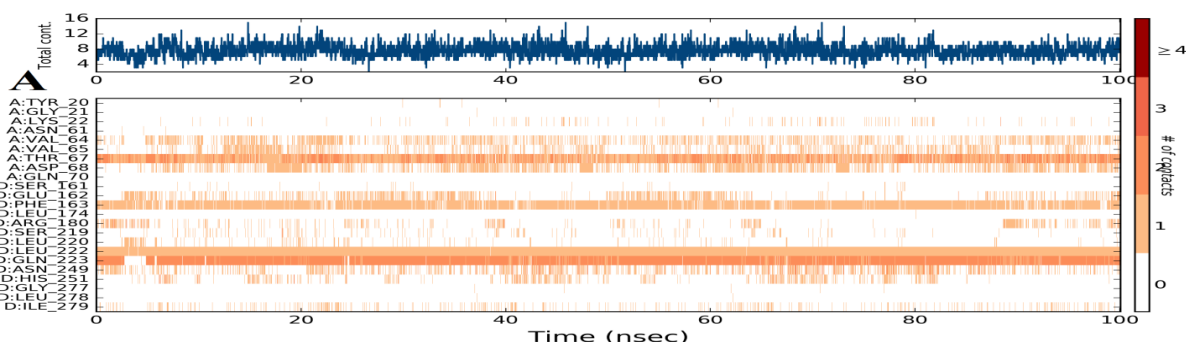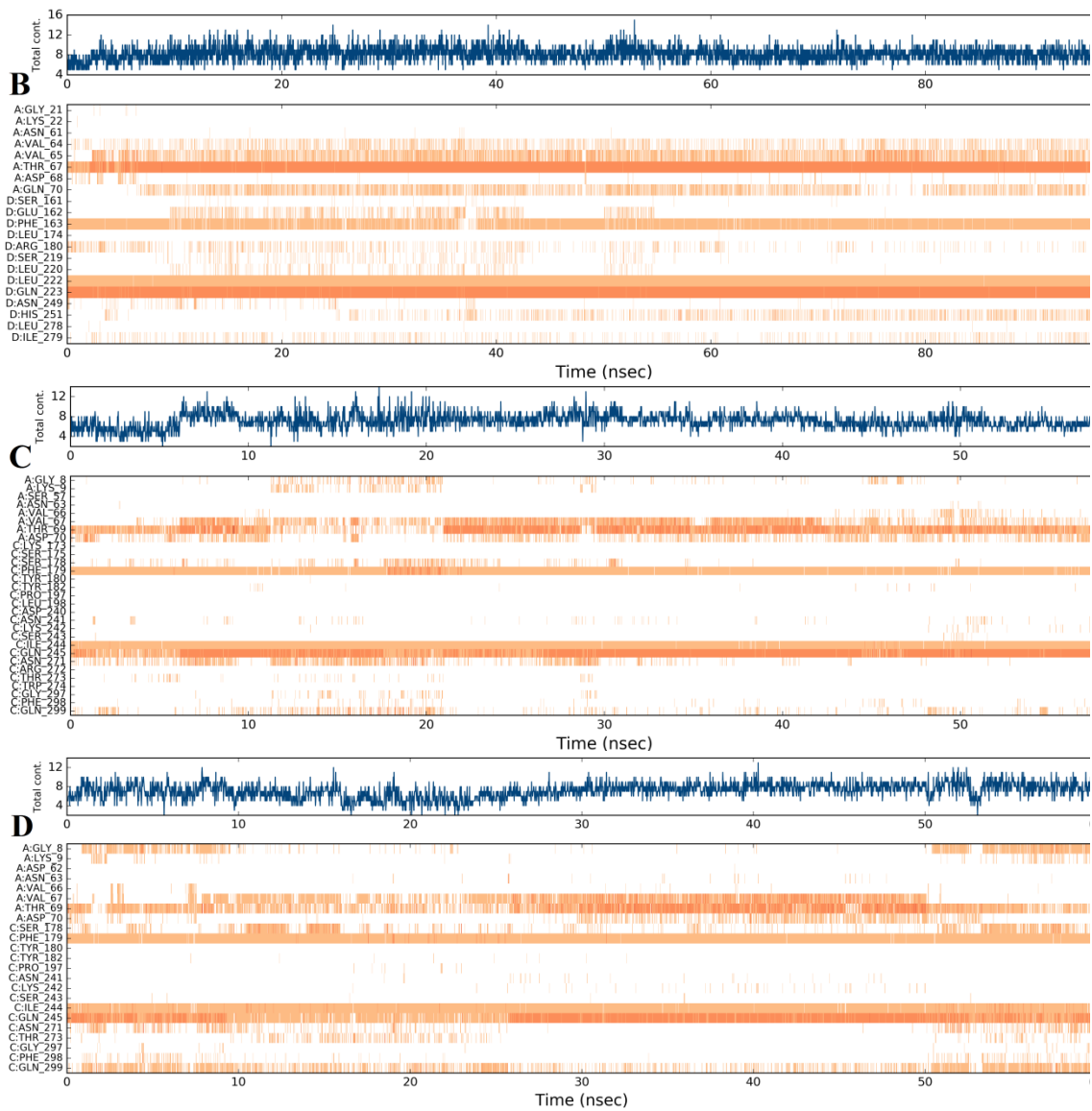
Figure A.III.7 The contact plot between uric acid and (A) Native uricase from native *Arthrobacter globiformis,* (B) Mutated uricase from *Arthrobacter globiformis*, (C) Native uricase from *Bacillus fastidious*, (D) Mutated uricase from *Bacillus fastidious*

APPENDIX IV

Assay of uricase activity:

The activity of uricase was determined by spectrophotometrically monitoring the disappearance of uric acid at 293 nm under standard assay conditions of pH 9.0 and 25 °C (Mahler 1970). The assay components were added to a quartz cuvette, inverted immediately, and the decrease in absorbance at 293 nm was monitored for approximately 5 minutes. In order to calculate the A293/minute, the maximum linear rate for both the Test and the Blank was used.

Uricase assay components:

Components	Blank (mL)	Test (mL)
20 mM Borate buffer (pH 9.0)	3.020	3.000
3.57 mM Uric Acid solution	0.100	0.100
Enzyme (in borate buffer)	---	0.020

The enzyme activity is calculated using the following equation:

$$units/mLenzyme = \frac{(\Delta A_{293}/\min Test - \Delta A_{293}/\min Blank)(3.12)(DF)}{(12.6)(0.02)}$$

Where,

    3.12 = Total volume (in milliliters) of assay

    DF = Dilution factor

    12.6 = Milli molar extinction coefficient of Uric Acid at 293nm

    0.02 = Volume (in milliliter) of enzyme used

One unit (U) of uricase is defined as the amount of uricase enzyme that is required to convert 1 μmol of uric acid to allantoin per minute at pH 9.0 at 25 °C.

The specific activity of enzyme was calculated as given below:

Units/mg protein = $\dfrac{\text{Units / mL enzyme}}{\text{mg protein / mL enzyme}}$

Where, mg protein / mL enzyme was determined from protein estimation using Brardford assay

Total protein Estimation by Bradford Analysis:

Bradford is used to determine the sample protein concentration. The basis of this assay is the complex formation of the dye and the protein in the solution. The complex protein colouring causes a shift in the maximum absorption rate between 465 and 595nm. The standard solution was BSA (bovine serum albumin).

Procedure

A Bradford assay of 1 part protein mixed with 30 parts of the Bradford reagent yielded the standard 3.1 mL volume. The blank sample was made up of buffer devoid of protein. The standard for protein concentration was a known amount of protein, which was mixed with the unknown sample to be tested. Bradford tests were done at room temperature by default. The colour development process began immediately and was monitored at 595 nm, with the protein concentration determined using a standard curve. This test is conducted in test tubes. Each tube contains 0.1mL of the protein sample and 3mL of the Bradford Reagent.

S.No	BSA (µL)	Distilled water (µL)	Bradford reagent (mL)	OD @ 595nm
1	20	80	3	0.246
2	40	60	3	0.443
3	60	40	3	0.631
4	80	20	3	0.825
5	100	-	3	0.991
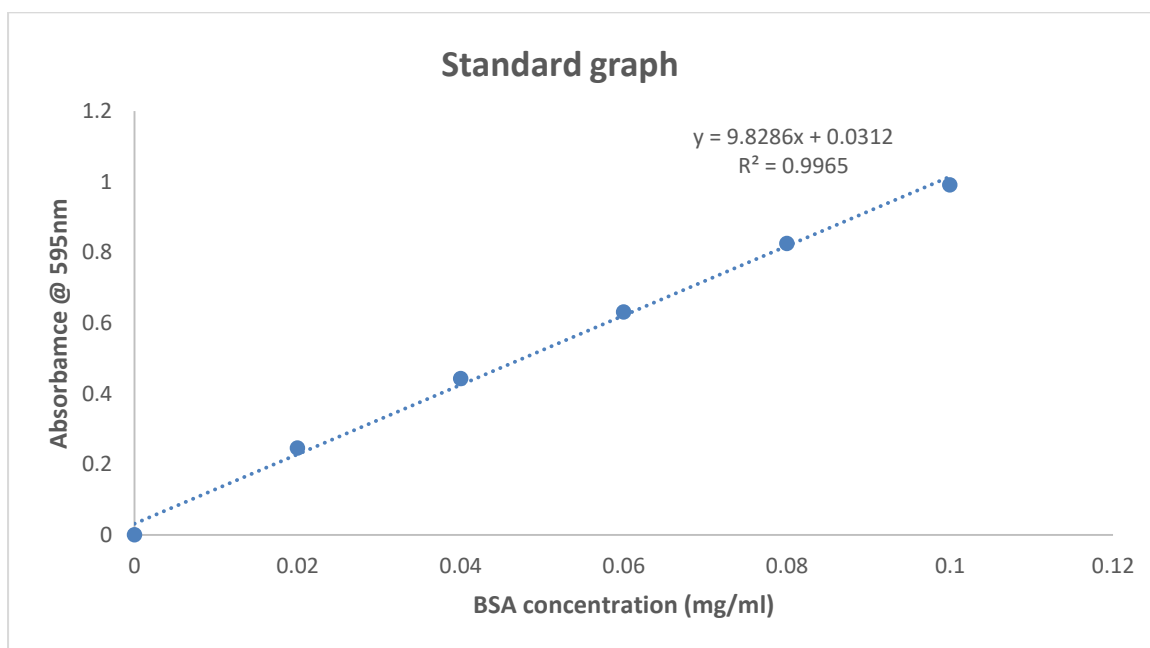
Standard graph of Bradford assay

SDS-PAGE of proteins for the determination of the molecular weight:

SDS PAGE (Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis) is a technique that is used to separate proteins according to their molecular weight. It is a widely used technique in biotechnology, genetics, forensics and molecular biology for sorting protein molecules according to their electrophoretic mobility.

Principle of SDS-PAGE

When an electric field is applied to a charged molecule, it migrates in the opposite direction to the electrode. The separation of charged molecules is determined by their relative mobility. Due to the lower resistance encountered during electrophoresis, the smaller molecules migrate more rapidly. The rate of migration is also influenced by the protein structure and charge. The impact of protein structure and charge is eliminated using sodium dodecyl sulphate and polyacrylamide and the proteins are separated by length of polypeptide chain.

Detailed procedure

1. All the marker protein vials from the manufacturer were run through a centrifuge before use (short spin 10,000 rpm for 1 min).

2. According to the criteria, the gel casting assembly was prepared. By filling water between the plates, it was determined that the assembly is leakproof. The amount of separating and stacking gel that was used was approximately 6mL and 4mL respectively. The 12% and 5% composition of the gels is provided below.

3. 1ml of water was mixed with 100mg of ammonium per sulphate (APS)

4. The plate assembly was attached to the PAGE apparatus.

5. The top of the separating gel was washed with distilled water, drained, and replaced with distilled water before setting it (approximately 45 minutes later).

6. In 5 ml of Stacking gel, 20µl of APS and 2µl of TEMED were added and thoroughly mixed and poured over the polymerized separating gel. Allow the gel assembly to solidify.

7. Teflon comb was gently placed in the gel solution, avoiding trapping air bubbles, and the stacking gel polymerized for about 45 minutes.

8. In the meantime, unmodified uricase and conjugates samples were prepared. The samples were mixed in the loading buffer in order to make the samples ready for loading on the gel. Five minutes of boiling was followed by five minutes of rapid cooling in an ice bath. The samples were ready to be loaded after 10 minutes at room temperature.

9. The comb was removed once the stacking gel had set. Non-polymerized acrylamide was rapidly removed by rinsing the wells with distilled water.

10. 1X gel running buffer was used to fill the gel running apparatus.

11. 20 µL of conjugated uricase, 20 µL of unmodified uricase, 20 µL of protein markers were pipetted into separate vials and labelled. The samples were put into various wells in the gel matrix.

12. According to the standard red for anode and black for cathode, the cords were connected to the power supply.

13. 20 mA current and 70 V voltage were chosen. The electrophoresis technique took

4-5 hours, depending on the time required for the dye front to reach the separating gels end.

14. The cords were dislodged and the panels gently removed from the PAGE apparatus following the completion of the electrophoresis process.

15. Using a spatula, the glass plates were opened, and the gels were recovered and transferred to a trough of distilled water. The gels were then rinsed twice in distilled water.

16. After discarding the water, a solution of 25 mL of Coomassie brilliant blue R-250 stain was added and the sample was stained for 45-60 minutes.

17. The gel was destaining by washing it in distilled water. The gel was rinsed in distilled water many times until the bands were visible and the excess staining colour had vanished.

Composition of SDS-PAGE gels

1. Resolving gel (12%), 8 mL

- 2.6 mL Water
- 3.2 mL 30% Acrylamide-Bisacrylamide Gel mix
- 2.0 mL 1.5 M Tris Buffer (pH 8.8)
- 80 μL 10% Sodium dodecyl sulphate
- 80 μL 10% Ammonium per sulphate
- 8 μL TEMED

2. Stacking gel (5%), 8 mL

- 4.5 mL Water
- 1.33 mL 30% Acrylamide-Bisacrylamide Gel mix
- 2.0 mL 0.5 M Tris Buffer (pH 6.8)
- 80 μL 10% Sodium dodecyl sulphate
- 80 μL 10% Ammonium per sulphate
- 8 μL TEMED

LIST OF PAPERS BASED ON THIS RESEARCH WORK

International Journal

Anand Kumar Nelapati and JagadeeshBabu P.E. (2020). "Computational Analysis of Therapeutic Enzyme Uricase from Different Source Organisms."*Current Proteomics*, 17 (1), 59-77.

Anand Kumar Nelapati, Bratin Kumar Das, JagadeeshBabuPonnanEttiyappan, Debashree Chakraborty (2020)."*In-silico* epitope identification and design of Uricase mutein with reduced immunogenicity."*Process Biochemistry*, 92, 288-302.

Anand Kumar Nelapati and JagadeeshBabu P.E (2020). "*In silico* structural and functional analysis of *Bacillus* uricases."*Current Proteomics*, 18(2), 124-142.

Papers under correction

Anand Kumar Nelapati, Kamala Meena S and JagadeeshBabu P.E, "Bioconjugation of therapeutic enzyme uricase with BSA: An experimental investigation.

Conferences

Anand Kumar Nelapati and JagadeeshBabu P.E. "Study of the docking of competitive urate oxidase inhibitors". International conference on advanced chemical and structural biology (ICACSB), PRIST University, Thanjavur, Chennai, February 19-21,2019.

Anand Kumar Nelapati and JagadeeshBabu P.E."*In silico* Engineering of Uricase for Effective Treatment of Hyperuricemia". International conference on advances in chemical and materials sciences, department of chemistry, Mangalore university, Mangalore, October 17-19, 2019.

BIO-DATA

NELAPATI ANAND KUMAR

Email ID**:** nelapatianandkumar@gmail.com and Contact no: +91-9949586996

Personal Details:

Fathers Name: NelapatiYesuratnam

Date of Birth: 02 May, 1991

Gender: Male

Nationality: Indian

Permanent Address: Achanta Vemavaram (Village),Achanta (Mandalam), West Godavari (District), Andhra Pradesh, INDIA. Pin -534267.

Educational qualification:

- Ph.D. in Chemical Engineering, 2015-2022
  Department of Chemical Engineering,
  National Institute of Technology Karnataka,
  Surathkal, Mangalore, Karnataka, India.
- M.Tech in Biotechnology, 2012-2014
  Andhra University College of Engineering(A),
  Visakhapatnam. Andhra Pradesh, India.
- B.Tech. in Biotechnology, 2008-2012
  RVR&JC College of Engineering,
  Guntur, Andhra Pradesh, India.

Declaration: I hereby declare that the above-written particulars are true to the best of my knowledge.

Place: Surathkal, India

Date: **NELAPATI ANAND KUMAR**