Structural and functional diversity of asparaginases: Overview and recommendations for a revised nomenclature

Leonardo Schultz da Silva | Liam B. Doonan | Adalberto Pessoa Jr. | Marcos Antonio de Oliveira | Paul F. Long B.Sc (Hons), M.Sc, PhD (Cantab)

Abstract
Asparaginases (ASNases) are a large and structurally diverse group of enzymes ubiquitous amongst archaea, bacteria and eukaryotes, that catalyze hydrolysis of asparagine to aspartate and ammonia. Bacterial ASNases are important biopharmaceuticals for the treatment of acute lymphoblastic leukemia, although some patients experience adverse allergic side effects during treatment with these protein therapeutics. ASNases are currently divided into three families: plant-type ASNases, Rhizobium etli-type ASNases and bacterial-type ASNases. This system is outdated as both bacterial-type and plant-type families also include archaeal, bacterial and eukaryotic enzymes, each with their own distinct characteristics. Herein, phylogenetic studies allied to tertiary structural analyses are described with the aim of proposing a revised and more robust classification system that considers the biochemical diversity of ASNases. Accordingly, based on distinct peptide domains, phylogenetic data, structural analysis and functional characteristics, we recommend that ASNases now be divided into three new distinct classes containing subgroups according to structural and functional aspects. Using this new classification scheme, 25 ASNases were identified as candidates for future new lead discovery.

KEYWORDS
biopharmaceuticals, bioprospecting, classification, enzyme diversity, l-asparaginase, phylogeny

1 INTRODUCTION

L-Asparaginases (ASNases) (EC 3.5.1.1) are enzymes that catalyze the hydrolysis of asparagine (L-Asn) to aspartic acid and ammonia.1,2 ASNases are widely distributed in nature and are currently divided into three families: plant-type ASNases, Rhizobium etli-type ASNases and bacterial-type ASNases.3-5
The current classification scheme for ASNases is based on protein sequence, biochemical properties and crystallographic data, although only limited primary sequences and crystallographic structures were available at the time this classification was proposed. A later study expanded the phylogenetic analyses, but still with only a limited number of ASNase sequences and with no inclusion of any crystallographic data. The crystal structures for more ASNases have now been resolved and these structures, together with the primary amino acid sequences for the enzymes, are publicly available. This now affords the possibility for expanded sequences sampling to rebuild the existing ASNase phylogeny and to propose a more robust nomenclature for these clinically significant enzymes.

2 MATERIALS AND METHODS

2.1 Source of sequences

A dataset of 108 ASNase amino acid sequences from archaea, bacteria and eukaryotes (fungi, metazoans, and plants) was selected from the literature and downloaded from the NCBI (https://www.ncbi.nlm.nih.gov/protein) and UniProt (https://www.uniprot.org/) protein databases (Figs. S1–S3). The datasets were manually expanded to include an additional 229 ASNases sequences from: (1) nonredundant sequences subsequently deposited in UniProt (sequences with 100% identity were deleted); (2) sequences from model organisms (e.g., Escherichia coli, Danio rerio, Drosophila melanogaster, Mus musculus); and (3) sequences from a large diversity of taxa not previously sampled (e.g., cnidarians). A multiple sequence alignment of 337 sequences was performed using a plugin of Multiple Alignment using Fast Fourier Transform software v.7.402 as part of the Geneious bioinformatics tool with default settings, and the InterPro Scan was used to identify protein domains. To deepen the analyses of clades that contained enzymes with high sequence homology but with distinct affinity for L-Asn (different $K_M$ values), a separate dataset was constructed using 133 sequences from ASNases presenting more than 90% of sequence similarity to human and guinea pig ASNases. Additionally, in order to investigate enzymes with comparable affinity for L-Asn (similar $K_M$ values) but with distinct structural characteristics, a third dataset using 56 sequences with more than 60% sequence similarity to E. coli and Erwinia chrysanthemi ASNases was compiled.

2.2 ASNases phylogeny reconstruction

The ProtTest 3 program was used to select the best model for amino acid replacement and this protein model was applied to all phylogenetic analyses (LG+I+G). Maximum likelihood based phylogenetic analyses were carried out using Randomized Accelerated Maximum Likelihood. Ten initial randomized maximum parsimony starting trees were constructed and the “best tree” (best log likelihood score) was selected. One thousand bootstrap replications were subsequently applied to the tree. Bayesian analysis was carried out with the Geneious plugin of MrBayes v 3.2.6 using default settings. Finally, the trees were edited using FigTree v1.4.3 software and Inkscape (https://inkscape.org/). Tree topologies of both phylogenetic methods (ML and Bayesian) were compared and used to classify the ASNases sequences.

2.3 Primary and tertiary/quaternary structural analysis of ASNases

A literature search was performed to identify the cellular location, enzymatic activity and structural information for each sequence, where available. For analysis using available crystallographic structures, the coordinates were obtained from the PDB and the figures were generated using Pymol software (http://www.pymol.org). The MEME motif-based sequence analysis tools (v5.0.2) (http://meme-suite.org/tools/meme) was used to identify motifs present in sequence clusters. After compiling the data, a final division of enzymes into classes was performed using a combination of primary and tertiary structure, functional characteristics and inferred phylogenetic relationships.

2.4 R. etli ASNase structure prediction

A theoretical model was constructed using the Swiss Model Workspace (https://swissmodel.expasy.org). The

Highlights

- Asparaginases are a large and diverse family of enzymes, commercially significant in biotechnology and medicine.
- Currently, the family is divided into bacterial-type, plant-type and Rhizobium etli-type enzymes; but this nomenclature is outdated and has conflicts related to taxonomic origins and biochemical properties.
- A modernized analysis of available sequences and structures is presented, and a new updated and robust nomenclature is proposed. Application of this classification is demonstrated with the identification of asparaginases that may provide useful therapeutic agents.
FIG. 1  Phylogeny of Class 1 ASNases. (A) Maximum likelihood phylogenetic tree of Class 1 ASNases. The tree was circularized and midpoint rooted for display purposes. Black dots indicate bootstrap support values (1000 replicates) ≥ 75% and Bayesian posterior probability (PP) ≥ 0.9; red dots indicate values of bootstrap support ≥ 75% and PP ≤ 0.9; white dots indicate values of bootstrap support ≤ 75% and PP ≥ 0.9. The red arrow shows E. coli type 1 ASNase, the gray arrow shows E. coli type 2 ASNase and the black arrow shows E. chrysanthemi type 2 ASNase. (B–G) Crystallographic monomer structures of Class 1 ASNases. (B) Class 1 ASNases have two α/β globular type folding domains with a central β sheet surrounded by α-helices from both domains. The elements of the secondary structure are colored as follows: α helix = light blue; β sheet = yellow and loops/random coils = pink (e.g., EcAI, 2P2D). Representatives from Group 1 of Class 1 are the structures of (C) EcAII from E. coli (pink, 3ECA), (D) GATD of P. abyssi (yellow, 1ZQ1), (E) E. coli EcAI (green, 2P2D), and (F) guinea pig GpASNAse1 (light blue, 4R8K). (G) Overlap of monomers of Class 1 ASNases

model was generated using the ASNase protein sequence from R. etli (Uniprot code: Q2K0Z2) and glutaminase coordinates of E. coli (YBAS) (pdb code 1U60), which has 29% similarity with the R. etli ASNase.

3 | RESULTS AND DISCUSSION

3.1 | ASNases can be divided into three classes with distinct characteristics

Based on the alignment of 337 ASNases sequences and an InterProScan domain search, it was possible to separate the sequences into three ASNase classes, hereafter referred to as Class 1, 2, or 3. Class 1, previously considered as the bacterial-type ASNases (enzymes that presented “Asparaginase domain,” InterPro ID: IPR027474), contained 198 sequences from metazoans, bacteria, fungi, algae and archaea, but not plants (Fig. 1). Class 2, previously named as plant-type ASNases (enzymes containing “Peptidase T2, asparaginase 2 domain,” InterPro ID: IPR000246), contained 107 representatives from metazoans, bacteria, fungi, algae, archaea and plants (Fig. 2). Class 3, formerly referred to as R. etli-type ASNases (proteins from “Asparaginase II domain,” InterPro ID: IPR010349), contained only 32 sequences of bacterial and fungal origins only (Fig. 3).

3.1.1 | ASNase Class 1 proteins are highly diverse and include enzymes used as biopharmaceuticals

According to the UniProt database, proteins in Class 1 (InterPro ID: IPR027474) display distinct biological properties, such as cell location and affinity for the substrate, and can be further divided into four functionally distinct groups (Fig. 1(A)) which present a very conserved protein fold (Fig. 1(B)). Group 1: this group contains 38 ASNases homologues from the Dikarya sub-kingdom and the bacterial ASNases including from E. coli (EcAII) and E. chrysanthemi (ErAII). Both bacterial sequences contain an ~20 amino acid N-terminal signal peptide for export to the periplasm and, in the mature form, are ~330 amino acid in size (35 kDa) (Fig. S4). The EcAII and ErAII enzymes display Michaelis–Menten behavior (following the equation \( V = V_{\text{max}}[S]/K_M + [S] \), where \( V \) is the observed initial rate, \( V_{\text{max}} \) is its limiting value at substrate saturation, and \( K_M \) is the substrate concentration when \( V = V_{\text{max}}/2 \) and have a \( K_M \) value of 15 and 50 μM, respectively. Group 1 proteins can also be cytosolic enzymes, for example, ASNases of Helicobacter pylori, which has allosteric behavior and \( S_{0.5} = 290 \mu M \). ASNase 3 of S. cerevisiae is one of the representatives of the Dikarya sub-kingdom of this group, an enzyme of ~38 kDa in size and composed of 362
FIG. 2  Phylogeny of Class 2 ASNases. (A) Maximum likelihood phylogenetic tree of Class 2 ASNases. The tree was circularized and rooted on its midpoint for display purposes. Black points indicate bootstrap support values (1000 replicates) ≥ 75% and Bayesian posterior probability (PP) ≥ 0.9, white dots indicate values of bootstrap support ≥ 75% and PP ≤ 0.9 and red dots indicate values of bootstrap support ≤ 75% and PP ≥ 0.9. The phylogenetic tree of Class 2 presents three main groups. (B–I) Crystallographic structures of the polypeptide chain of Class 2 ASNases. In (B), the structure of a polypeptide chain of Class 2 ASNases, denoting an α/β type structure is presented, with a central β sheet derived from elements of both subunits are surrounded by α-helices. The elements of the secondary structure are colored as follows: α helix = light blue; ribbons β = yellow and loops/random coils = pink (PDB: 2ZAL). In (C–H), the polypeptide chain structures of Class 2 enzymes of (C) E. coli EcAIII (PDB: 2ZAL), (D) Phaseolus vulgaris K-dependent ASNase (PDB: 4PU6), (E) ASNase K-independent of Lupinus luteus (PDB: 2GEZ) and (F) human ASNase 3 (hASNase 3) (PDB: 4PVS). The enzymes of Group 2 of Class 3 are represented by (G) human threonine aspartase 1 (Taspase1) (PDB: 2A8J) and enzymes of Group 3 of Class 2 are represented by (H) human aspartyglycosaminidases (AGA) (PDB: 1APZ). (I) Overlap of cleaved polypeptide chain of Class 2 ASNases. In the representations of (B–H), the α subunit is colored blue with β in yellow and the N-terminal Thr residue is represented in red beads.

FIG. 3  Phylogeny of Class 3 ASNases. (A) Maximum likelihood phylogenetic tree of Class 3 ASNases. The tree was circularized and midpoint rooted for display purposes. Black points indicate bootstrap support values (1000 replicates) ≥ 75% and Bayesian posterior probability (PP) ≥ 0.9, white dots indicate values of bootstrap support ≥ 75% and PP ≤ 0.9 and red dots indicate values of bootstrap support ≤ 75% and PP ≥ 0.9. (B) Representations of the R. etli ASNase model in the monomeric form. Motifs 1 and 2 are colored in red and blue, respectively and presented as spheres. The elements of the secondary structure are colored as follows: α helix = light blue, ribbons β = yellow and loops/random coils = pink. The figures were generated using PyMol.
amino acids. Group 2: this group comprises 40 ASNase sequences from the Dikarya sub-kingdom and includes proteins that contain ~50 additional N-terminal amino acid residues of unknown function, resulting in an overall molecular mass of ~40 kDa (Fig. S5). Bioinformatics analyses revealed that these N-terminal sequences did not encode for any recognized signal peptide motifs. The only enzyme in Group 2 that has been characterized to date is AspI from S. cerevisiae, which has allostery behavior and $S_{0.5} = 75 \mu M$. Group 2 does not include any enzymes for which a crystal structure has been resolved. Group 3: this group consists of 34 enzymes from Archaea named Glutamyl-tRNA aminotransferase D (GatD), and have an N-terminal extension of ~90 amino acids residues, giving an overall primary sequence of ~450 residues (e.g., GatD of Pyrococcus abyssi; Fig. 1(D)) and molecular mass of ~50 kDa (Fig. S6). The N-terminal extensions form a heterodimer in Glutamyl-tRNA aminotransferase E (GatE). These enzymes are aminotransferases, using acylated Glutamyl-tRNA as substrates to produce Gln-tRNA. However, L-Asn can also be used as an amide donor in addition to Gln.

Group 4: this group is composed of 86 enzymes, 41 of which contain a 33 amino-acid ankyrin domain sequence in the C-terminal region that mediates protein–protein interactions. These proteins can be cytosolic or secreted (Figs. S7 and S8) and includes the 338 amino acid 37 kDa E. coli type I cytosolic ASNase (Fig. 1(E)) but does not contain any recognized ankyrin domains. This ASNase exhibits allostery behavior, with low substrate affinity ($S_{0.5}$ in range of mM). Using ASNase of Cavia porcellus as an example (Fig. 1(F)), the polypeptide chain consists of ~560 residues and has a molecular mass of 60 kDa. In addition to ASNase activity ($K_M$ in the $\mu M$–mM range), the enzymes of Group 4 can also exhibit lysophospholipase, transacylase and acetyldihydrolyase activities.

Using the MEME motif discovery software, it was possible to detect regions of highly conserved amino acids containing the following previously known motifs: motif 1: T-G-G-T and motif 2: H-G-T-D-T (Figs. S4–S7). Motifs 1 and 2 were found in all 198 Class 1 ASNases and are of catalytic importance. For example, in E. coli ASNase 2, Motif 1 ($^{67}TGGT^{12}$) contains a $^{12}$Thr residue, which stabilizes the substrate at the active site and Motif 2 ($^{87}HTGT^{91}$) that contains a $^{91}$Thr residue, which promotes the protonation of the leaving amino group and activates a nucleophilic water molecule. A third but previously undescribed motif (equivalent to $^{297}L-N-P-X-K-X-R$ in E. coli ASNase 2) was found in Groups 1 and 2 ASNases and is located in the C-terminal region of the proteins. It has been reported that residues $^{300}$Gln, $^{303}$Arg and $^{281}$Tyr, when substituted by Ala, affect the stability of the intimate dimers. In this context, Motif 3 may contain residues necessary for the formation of intimate dimers (Fig. S9).

### 3.1.2 Class 2 ASNases: functional diversity and structural conservation

Previous work classified enzymes from this class as plant type ASNases and contain an “asparaginase 2” domain (InterPro ID: IPR000246). However, phylogenetic analyses revealed that in addition to plants, this group of enzymes can also be found in bacteria, for example, EcAIII from E. coli, and in archaea and eukaryotes including metazoans (Fig. 2(A) and Fig. S1). Based on the revised classification proposed herein, this group should now be designated as Class 2 ASNases and comprises 107 enzymes with very diverse substrate specificities. Although all the enzymes display ASNase activity, biologically L-Asn may not be the major substrate for these enzymes. For example, some hydrolyze dipeptides of $\beta$ aspartyl and function in the metabolism of oligosaccharides or are Thr endopeptidase.

In this context, it is important to note that not all the enzymes present $K_M$ for L-Asn in the mM range. Class 2 ASNases also retain high conservation in secondary protein folding (Fig. 2(B)) further dividing the class into three distinct functional subgroups. Group 1: 60 isoaspartyl peptidases, $\beta$-aspartyl peptidases and ASNases including EcAIII (Fig. 2(C)), a protein of 321 amino-acid residues and molecular weight of 35 kDa. The substrates for this group are L-Asn and L-isoAsp, some representatives can also hydrolyze spontaneous damage products of proteins and cyanoformic acid (a polymer commonly found in cyanobacteria). In plants (Fig. 2(D) and 2(E)), this group of ASNases causes release of ammonia from Asn, which becomes the main nitrogen source for transport and storage.

In metazoans, the human hASRGL1 (Fig. 2(F)) is a 308 amino acid protein with a molecular weight of 32 kDa, which also displays $\beta$-aspartyl peptidase activity. Group 2: this group comprises 14 endopeptidases, which cleave substrates after Asp residues and are localized within the nucleus of eukaryotic cells (Fig. 2(A)). The human isoform of Thr aspartase 1 (Fig. 2(G), Taspase1) is composed of 420 amino acids and has a molecular mass of 45 kDa. Taspase1 cleaves KMT2A (Histone-lysine N-methyltransferase) involved in the regulation of gene expression during early stages of hematopoiesis. Its dysregulation is involved in hematological malignancies, including ALL. Group 3: this group contains 33 enzymes with amidohydrolase activity involved in catabolism of oligosaccharides (e.g., aspartyl-glucosaminidases) (Fig. 2(A)). The AGA (Fig. 2(H)) of Homo sapiens is composed of 346 amino acids; however, the first 23 amino acids correspond to a signal peptide, thereby giving a mass of 36 kDa to the mature protein. In eukaryotes, Group 3 ASNases are localized in lysosomes and are implicated in the decomposition of glycosylated proteins by cleavage of the glycosidic bond between the carbohydrate and an Asn residue.
group of enzymes is well studied and is found in bacteria and eukaryotes.42–44

Using MEME, two highly conserved amino acid motifs were identified in Class 2 enzymes. Motif 1 is represented by the sequence T/Q-V/I-L-G-X-V/I-L-X(2)-D/H-X(2)-G/N.168 Thr in the human enzyme Isoaspartyl peptidase (ASGL1), where cleavage occurs between the α and β chains, is located within this motif making it essential for the function of the enzyme. The second motif, T/S-T/S-X-G-X(3)-K/R-X(2)-G-R-V/I-G, contains 186 Thr (in ASGL1), which has catalytic activity45 (Fig. S10).

### 3.1.3 Class 3 ASNases

Proteins belonging to Class 3 were traditionally classified as *R. etli* ASNases (a symbiotic bacterium of leguminous plants46 and contain “asparaginase II” domains (InterPro ID: IPR010349) (not to be confused with “asparaginase 2” domains found within Class 2 proteins). Only 32 nonredundant proteins could be grouped into this class (Fig. 3(A)). Class 3 ASNases include representatives from fungi, bacteria (including *R. etli* ASNases) and cyanobacteria. Using the enzyme of *R. etli* as an archetype to describe Class 3 enzymes, this ASNase is composed of 370 amino acids and has a mass of ~40 kDa. The *K₉₄* for this enzyme is 8.9 mM and the catalytic efficiency measured by a *k₅₆/*₉₄ is 1 × 10⁴ M⁻¹ Sec⁻¹ for the l-Asn substrate. No glutaminase activity has been detected.46 Class 3 ASNases are also found in cyanobacteria, actinobacteria and *Firmicutes* bacteria (Fig. 3). To date, there have been few studies that have characterized these enzymes and, amongst the proteins used in this analysis, only the ASNase from *Streptomyces griseus* has been studied and was found to be an extracellular enzyme with ASNase but no glutaminase activity.47

It was possible to identify two conserved amino acid motifs in enzymes grouped into this class—Motif 1: R/S-X(2)-K-P/A-X-Q-A-L and Motif 2: N/G-C-S-G/A-K-H-X-G/A-M/P (Fig. S11). Although the active site of Class 3 enzymes has not been characterized, in other ASNases catalysis involves at least two Thr residues and in some cases Ser.48,49 Ser residues are found in both motifs, which have a carbonyl γ that conceivably could function in catalysis. To investigate this further, a theoretical model was constructed for the ASNase of *R. etli*. The model was generated using the glutaminase coordinates of *E. coli* (YBAS) (PDB code: 1U60),13 which has 21% identity and 29% similarity with the ASNase of *R. etli* (Fig. 3(B)). The model covered from residue 22 to 316 (from a total of 367 residues, ~80%) providing a quantitative model energy analysis (QMEAN) = −6.35 and global model quality estimate between the ASNase of *R. etli* model and the *E. coli* crystallographic of 0.35, suggesting the model was of good quality. Glutaminase enzymes encode a motif S-X-X-K (66-S-X-X-K-69 in YBAS), which contains the catalytic nucleophile Ser66, very similar to motif 1 identified for class 3 asparaginases.13 In addition, the theoretical model showed that motifs 1 and 2 were spatially close and that the amino acids of these motifs could interact thereby, in principle, forming an interaction network.

#### 3.2 Phylogenetic considerations supporting the revised classification

Phylogeny of Class 1 ASNases revealed low conservation of primary sequence (<30%) in agreement with previous work.3 However, conservation of distinct sequence motifs allowed the class to be further divided into four groups. Groups 1 and 2 contained enzymes from the yeast Dikarya sub-kingdom and from bacteria, where the primary sequence is composed of an exclusive 50 amino acids N-terminal domain. Group 3 is composed of ASNases of archaeal origin and where the N-terminal of the primary sequence is composed of ~90 amino acids, which function in tRNA repair by transamidation of misacylated Glu-tRNA(Gln) to produce the correct Gln-tRNA(Gln).19 Group 4 is the most diverse, containing ASNases and lysophospholipase from protozoa, bacteria, fungi and metazoans. Additionally, the ASNases primary sequences of metazoan and fungal origin contain ankyrin domain repeats, suggesting formation of multi-domain proteins common to metazoans.50 Repetitive ankyrin domain sequences may also provide a regulatory role for these enzymes.19 Enzymes of Group 4 can also exhibit lysophospholipase, transacetylase and acetylhydrolase activities,24–26 in addition to the ASNase function (*K₃* in the μM-mM range). Class 2 ASNases represent enzymes of the Ntn-hydrolases family, which are enzymes with an N-terminal Thr, Ser or Cys residue, which acts as a nucleophile during catalysis.4 Phylogenetic analyses provided further sub-division of Class 2 ASNases: Group 1 is composed of isoaspartyl-peptidases, β-aspartyl-peptidases and ASNase enzymes derived from multiple taxa; Group 2 is composed of endopeptidases of plant, fungal and metazoan origins; and Group 3 contains aspartyl-glucosaminidases enzymes from bacteria, plants and metazoans. Only 32 nonredundant proteins could be grouped by phylogeny into a revised Class 3 and were traditionally classified as *R. etli* ASNases.46 Class 3 ASNases presented fungi, cyanobacteria and bacteria representatives. The conserved motifs suggested that the catalytic sites are quite distinct from Class 1 and Class 2 ASNases, with Ser replacing Thr. Future studies to expand, define and further refine Class 3 ASNases are now especially warranted.
3.3 Structural considerations supporting the revised classification

From a structural perspective, Class 1, 2 and 3 enzymes are very distinct suggesting evolutionary divergence within the ASNases. Concerning the crystallographic structures determined for members of Class 1 ASNases, there is remarkable conservation of the 3D structures of the monomers of enzymes in each of the different sub-groups. Although the enzymes have low sequence identity (<30%), the three-dimensional structure of the monomers give high spatial identity. In all cases, ASNases have two α/β globular domains and, in both domains, have a β sheet surrounded by α helices (Fig. 1(B)). When EcAI (Fig. 1(C)) and EcAII (Fig. 1(D)) are compared, the structures are very similar, giving a root mean square deviation (r.m.s.d.) of the Cα amino acid chain of 2.5 Å. The enzymes also shared high similarity with guinea pig GpASNase (r.m.s.d. ~1.2 Å) (Fig. 1(B)–1(G)). However, it is important to note that the graphical representations of GpASNase do not include C-terminal ankyrin repeats, as these were not present in the crystal structures (Fig. 1(E)). When EcAII was compared with a Group 3 ASNase (GatD from P. abyssi), a high level of structural identity (r.m.s.d. ~1.5 Å) was also observed. In general, these results indicated that even with low amino-acid identity, the tertiary structure of Class 1 ASNases are highly conserved, suggestive of a common evolutionary ancestry. Class 2 enzymes show more complex spatial organization compared with ASNases of Class 1 and resemble N-terminal nucleophilic hydrolases. Class 2 ASNases are translated as a single inactive polypeptide precursor, which undergoes an autocatalytic intramolecular cleavage, dividing the polypeptide chain into two subunits: α subunit and β subunit, exposing the N-terminal nucleophile amino acid (Thr, Ser, or Cys) of the β subunit. Therefore, a single polypeptide chain can form an α/β heterodimer. Despite members of this class displaying quite distinct biological functions and exhibiting only moderate similarity (25–40%), analysis of the tertiary folding of the polypeptide chain reveals a very high structural similarity (r.m.s.d. ~0.8–1.0 Å²), even when comparing proteins from different organisms such as bacteria, plants or humans (Figure. 2(B)–(I)). The representatives of this class form α/β folds, having two β sheets superimposed within the molecule and surrounded by an α helix (Fig. 2(B)). This folding is quite different from Class 1 enzymes (Fig. 1(B)–1(G)). Concerning Class 3 enzymes, no crystallographic structures are available in Protein Data Bank. However, our theoretical model generated using the glutaminase coordinates of E. coli (YBAS) revealed a tertiary structure distinct from Class 1 and 2 enzymes.

3.4 Application of the revised nomenclature

The organization of structural and functional features of the different ASNases classes and groups provided by the revised nomenclature, could facilitate bioprospecting for exciting new ASNases. The most interesting enzymes from a therapeutic point of view can be found in Class 1, with examples of enzymes that show K_M to L-Asn in micromolar range, for example, EcAII and ErA2 in Group 1, Asp1 from S. cerevisiae from Group 2 and GpASNase1 from C. porcellus from Group 3. The most studied ASNases are EcAII (Fig. 1, blue arrow) and ErAII (Fig. 1, black arrow) because these enzymes have both been developed into biopharmaceuticals for the treatment of ALL, and provide excellent prototypes to explore discovery of new ASNases with therapeutic potential.

Briefly, both enzymes have high sequence similarity (63%) and the crystallographic structures display remarkable conservation of amino acids involved in catalysis. The reaction mechanism involves two catalytic triads (Thr, Tyr, Glu) and (98-Thr, Asp, Lys) in EcAII, and 15-Thr, Tyr, Glu and 95-Thr, Asp in ErAII) and a movement of a loop containing residues 12–31 in EcAII and 14–35 in ErAII, including the catalytic Thr residue of motif 1 (residue 12 in E. coli and residue 15 in E. chrysanthemi). However, in E. coli ASNase, the Glu283 residue of the adjacent monomer (Thr, Ser, or Cys) of the β subunit. Therefore, a single polypeptide chain can form an α/β heterodimer. Despite members of this class displaying quite distinct biological functions and exhibiting only moderate similarity (25–40%), analysis of the tertiary folding of the polypeptide chain reveals a very high structural similarity (r.m.s.d. ~0.8–1.0 Å²), even when comparing proteins from different organisms such as bacteria, plants or humans (Figure. 2(B)–(I)). The representatives of this class form α/β folds, having two β sheets superimposed within the molecule and surrounded by an α helix (Fig. 2(B)). This folding is quite different from Class 1 enzymes (Fig. 1(B)–1(G)). Concerning Class 3 enzymes, no crystallographic structures are available in Protein Data Bank. However, our theoretical model generated using the glutaminase coordinates of E. coli (YBAS) revealed a tertiary structure distinct from Class 1 and 2 enzymes.
FIG. 4  Phylogenetic and structural analyses of the Class 1 ASNases from bacteria. (A) Alignment of sequence representatives of each clade: blue clade = ASNases of *E. coli*, *Salmonella typhimurium*, *Citrobacter rodentium* and *Shigella boydii*; green clade = *E. chrysanthemi*, *Klebsiella aerogenes*, *Pectobacterium carotovorum* and *Serratia plymuthica*. Numbers I, II and III represent the regions of the sequences analyzed. (B) Crystal structure of EcAII (with the different monomers represented as blue and yellow) with the network of interactions formed between 283E, 25Y, 12Y and 117P; and (C) crystal structure of ErAII (with the different monomers represented as cyan and green) with the network of interactions formed between 29Y, 123P and 15Y.

The –COOH group is positioned toward the N-terminal of the substrate main chain, increasing the nucleophilicity of the active site and decreasing the volume of the active site in comparison with ErAII, indicating a strong influence on substrate choice.\(^\text{57–59}\) However, \(^\text{283}\)Glu positioning occurs due to interactions with the NH\(_2\) groups from the side chains of the \(^\text{59}\)Gln (Region I) and \(^\text{250}\)Asn (Region II) residues, which would be lost by performing substitution with Glu and Ser, respectively.\(^\text{15,58}\) In addition, the volume of the active site of ErA2 is larger in comparison with EcAII, which justifies its higher affinity for Gln (K\(_M\) = 360 \(\mu\)M). Analyses of site-directed mutations and the crystallographic structures of EcAII and ErAII revealed that the main modulator of glutaminase activity is due to the presence or absence of \(^\text{283}\)Glu located in region III (Fig. 4(B-III)), and a natural deletion in the ErAII sequence group probably resulted in natural substitutions of key residues (Gln by Glu at region I, and Asn by Ser at region II) that would allow catalysis of both substrates, even in the absence of \(^\text{283}\)Glu.

Treatment using an ASNase lacking glutaminase activity may provide beneficial effects in most cases of ALL, in which the cancerous cells do not produce Asn-synthetase (ASNS\(^+\)), thereby offering an approach to decrease possible adverse side effects. Some authors point out that glutaminase activity may also be important for cytotoxicity, especially in tumor cells where the Asn synthetase gene remains activated, even after neoplastic transformation (ASNS\(^+\) tumor types).\(^\text{53}\) Nevertheless, utilizing the analysis of the two sub-groups obtained in the phylogenetic tree of Class 1 ASNases (Fig. 1(A)), it was possible to select 15 ASNases for that may have reduced affinity for Gln, but ASNase activity with K\(_M\) in the range of \(\mu\)M, which is useful for future preclinical development (EcAII homologs: ASNases from *Trabulsiella guamensis*, *Hafnia paralvei*, *Escherichia fergusonii*, *Shigella dysenteriae*, *Shigella boydii*, *Escherichia marmotae*, *Shigella sonneti*, *Klebsiella pneumoniae*, *Vibrio parahaemolyticus*, *Enterobacter cloacae*, *Citrobacter rodentium*, *Salmonella anatum*, *Salmonella typhimurium* and *Salmonella choleraesuis*) or 10 ASNases with moderated glutaminase affinity to treat ASNS\(^+\) tumor types (ErA2 homologs: *Deinococcus marmoris*, *Caballotonia arvi*, *Pectobacterium carotovorum*, *Erwinia solani*, *Serratia plymuthica*, *Klebsiella*...
Sequence analyses of the Class 1 ASNases from vertebrates. Alignment of the amino acid sequences of Class 1 ASNases of primates, rodents and a lagomorph. Conserved motifs 1 and 2 and regions 1 and 2 are related to the low $K_M$ from $C.\porcellus$. Identical residues are highlighted in dark gray and similar residues in light gray. The red boxes denote motifs 1 and 2, blue boxes denote residues that are different between species, and black boxes denote residues that differ only in GpASNase1. The alignments were performed using ClustalW.

The same analyses can be performed with GpASNase1 from $C.\porcellus$ (domestic guinea pig). The GpASNase1 has a $K_M$ over Asn in the $\mu$M range and exhibits Michaelis–Menten behavior. Despite sharing high sequence similarity with the human protein (HsASNase1), HsASNase has a value of $S_{0.5}$ between 2.9 and 11.5 mM toward Asn, and displays allosteric behavior as demonstrated by a Hill coefficient between 2.5 and 3.9. The reasons for the differences between GpASNase1 and HsASNase1 remains elusive. Recent studies using a DNA shuffling technique that generated chimeric enzymes containing segments of hASNase1 and GpASNase1 revealed two regions in the GpASNase1 enzyme (region 1: between amino acids 1 and 38 and region 2: between amino acids 305 and 357) (Fig. 5(A)) essential for maintaining the low $K_M$ of the enzyme. To investigate this point, we analyzed a multiple sequence alignment of ASNase sequences from primates and rodents, and it was possible to observe that Regions 1 and 2 are highly conserved in primates but not in rodents, suggesting that these amino acid substitutions may be related to the observed differences in $K_M$ between hASNase1 and GpASNase1. Thus, rodent ASNases may be more interesting from a biotechnological point of view, as there is a higher diversity of enzymes when analyzing the two regions related to the $K_M$ of the enzymes (Fig. 5). Structural analysis using the crystallographic structure of GpASNase revealed a tetrameric enzyme formed by a dimer of intimate dimers (Fig. S13A), where regions 1 and 2 is in the region of the active site (Fig. S13B). Both regions are related to the open/close dynamics of the loop containing 308Tyr (region 2) of the adjacent monomer, responsible for correct positioning of 19Thr (region 1) and stabilization of the substrate, acting as a lid to enclose the substrate at the active site. It is also possible to observe a $\beta$-sheet twist in region 1, which allows for correct positioning of 308Tyr (Fig. S13B). However, in GpASNase1, the loop that facilitates correct positioning of the Tyr residue has two deletions and four substitutions (Ala for Thr, Ala for Ser, Met for Leu, and Gly for Asn), which may cause faster lid closure over the substrate at the active site (Fig. S13). In addition, homologs of GpASNase1 in rodents, for example, *Ictidomys tridecimlineatus* (13-lined ground squirrel) and *Dipodomys ordii* (Ord’s kangaroo rat) and the lagomorph *Oryctolagus cuniculus* (European rabbit) showed some substitutions in regions 1 and 2, which could also affect loop dynamics compared with the
enzymes of Class 1 from primates (Figure S14), although only GpASNase1 has been well characterized.20,60 Thus, structural aspects can be related to differences in the \( K_M \) of eukaryotic ASNases and, together with the phylogenetic analyses, has revealed that Class 1 ASNases other than those in Group 1 (e.g., ASNases of Group 4) could be interesting enzymes with favorable features that can be exploited for future clinical or other biotechnological applications.

4 CONCLUSIONS

The existing scheme used to classify ASNases is contradictory and confusing because the current divisions of bacterial-type, plant-type and \( R. \) elliptolytica enzymes is not homogenous but, instead, contain representatives from a wide range of other taxa.4 The current classification separates ASNases according to biochemical properties, for example, bacterial-type II enzymes are localized within the periplasmic space, have high affinity for L-Asn and exhibit Michaelis–Menten behavior.3–5 However, some enzyme currently classified as bacterial-type II enzymes have opposing biochemical properties, for example, the Asp3 ASNase from the yeast \( S. \) cerevisiae is located at the cell wall and has a low affinity for L-Asn (\( K_M = 270 \mu M \)).61 Using phylogenetic tree constructions that adopted enhanced taxon sampling supported by structural and biochemical data, it was possible to propose a revised ASNases nomenclature at two levels: (1) ASNases divided into three classes based upon distinct domains (InterPro Scan) and phylogenetic and structural features, and (2) further division of the classes into groups with shared functional, mechanistic and structural features. The revised nomenclature is underpinned by recognition of conserved sequence motifs that are distinctive for each of the three classes, which could be related to enzyme structure and catalytic activity, strongly inferring discrete evolutionary origins of each Class (Fig. 1, 2, and 3).

In summary, a revision of the current ASNase classification was made using primary protein sequence phylogeny, tertiary protein structure analysis and characteristics of enzyme catalysis. Using this updated classification scheme, it was possible to select ASNases with desirable properties suitable for further assessment as alternative biopharmaceuticals to treat ALL. Additionally, expanded taxon sampling for the phylogenetic analysis provided previously poorly characterized ASNases with high primary sequence diversity mainly in regions related to substrate affinity, suggesting that enzymes with interesting catalytic properties have still to be discovered.

FUNDING

Funding was from Fundação de Amparo à Pesquisa do Estado de São Paulo (2013/08617-7, 2014/2039-9, 2017/19942-7, 2018/04685-1), the Universidade de São Paulo (13.1.1502.9.8), and the Leverhulme Trust (RPG-2016-037).

ORCID

Paul F. Long B.Sc (Hons), M.Sc, PhD (Cantab) https://orcid.org/0000-0001-6698-4602

REFERENCES

1. Costa-Silva, TA, Costa, IM, Biasoto, HP, Lima, GM, Silva, C, Pessoa, A, Monteiro, G. Blood Rev. 2020;39:100651 https://doi.org/10.1016/j.blre.2020.100651
2. Lopes, AM, Oliveira-Nascimento, L, Ribeiro, A, Tairum, CA Jr, Breyer, CA, Oliveira, MA, et al. Crit. Rev. Biotechnol. 2017;37:82–99.
3. Borek, D, Jaskolski, M. Acta Biochim. Pol. 2001;48:893–902.
4. Michalska, K, Jaskolski, M. Acta Biochim. Pol. 2006;53:627–640.
5. Dwivedi, VD, Mishra, SK. Interdiscip. Sci. 2014;6:93–99.
6. Katoh, K, Standley, DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 2013;30:772–780.
7. Kearse, M, Moir, R, Wilson, A, Stones-Havas, S, Cheung, M, Sturrock, S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–1649.
8. Darriba, D, Taboada, GL, Doallo, R, Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution Bioinformatics. 2011;27:1164–1165.
9. Stamatakis, A. Bioinformatics. 2014;30:1312–1313.
10. Warangkar, SC, Khobragade, CN. Enzyme Res. 2010;165878 https://doi.org/10.4061/2010/165878
11. Huexesnbeck, JP, Ronquist, F. Bioinformatics. 2001;17:754–755.
12. Berman, HM, Westbrook, J, Feng, Z, Gilliland, G, Bhat, TN, Weissig, H, et al. Nucleic. Res. 2000;28:235–242.
13. Brown, G, Singer, A, Proudfoot, M, Skarina, T, Kim, Y, Chang, C, et al. Biochemistry. 2008;47:5724–5735.
14. Jennings, MP, Beacham, IR. J. Bacteriol. 1990;172:1491–1498.
15. Nguyen, HA, Su, Y, Lave, A, J. Biol. Chem. 2016;291:17664–17676.
16. Maggi, M, Chiarelli, LR, Valentini, G, Scotti, C. PLoS One. 2015;10:e0117025.
17. Dunlop, PC, Meyer, GM, Roon, RJ. J. Biol. Chem. 1980;255:1542–1546.
18. Costa, IM, Schultz, L, de Araujo Bianchi Pedra, B, Silva Moreira Leite, M, Farsky, SHP, de Oliveira, MA, et al. Sci. Rep. 2016;6:36239.
19. Schmitt, E, Panvert, M, Blanquet, S, Mechulam, Y. Structure 2005;13:1421–1433.
20. Schalk, AM, Lave, A. Biochemistry. 2014;53:2318–2328.
21. Menniti, M, Iuliano, R, Füsser, M, Sopjani, M, Alesutan, I, Maggioli, S, et al. Cell. Physiol. Biochem. 2010;26:587–596.
22. Belviso, S, Iuliano, R, Amato, R, Perrotti, N, Menniti, M. PLoS One 2017;12:e0178174.
23. Yun, MK, Nourse, A, White, SW, Rock, CO, Heath, RJ. J. Mol. Biol. 2007;369:794–811.
24. Rigouin, C, Nguyen, HA, Schalk, AM, Lavie, A. Sci. Rep. 2017;7:10224.
25. Sugimoto, H, Yamashita, S. Biochim. Biophys. Acta. 1999;1438:264–272.
26. Karamitros, CS, Konrad, M. J. Biol. Chem. 2014;289:12962–12975.
27. Sanches, M, Barbosa, JA, de Oliveira, RT, Abrahao Neto, J., Polikarpov, I. Acta Crystallogr. D. Biol. Crystallogr. 2003;59:416–422.
28. Mezentsev, YU, Molnar, AA, Sokolov, NN, Lisitsina, VB, Shatskaya, MA, Ivanov, AS, Archakov, AI. Biochemistry (Moscow). 2011;5:124–134.
29. Hejazi, M, Piotukh, K, Mattow, J, Deutzmann, R, Volkmer-Engert, R, Lockau, W. Isoaspartyl dipeptidase activity of plant-type asparaginases. Biochem. J. 2002;364:129–136.
30. Khan, JA, Dunn, BM, Tong, L. Structure 2005;13:1443–1452.
31. Cantor, JR, Stone, EM, Chantranupong, L, Georgiou, G. Biochemistry. 2009;48,11026–11031.
32. Borek, D, Michalska, K, Brzezinski, K, Kisiel, A, Podkowinski, J, Bonthron, DT, et al. Eur. J. Biochem. 2004;271:3215–3226.
33. Niizuma, H, Cheng, EH, Hsieh, JJ. Mol. Cell. Oncol. 2015;2:e999513.
34. Richter, R, Hejazi, M, Kraft, R, Ziegler, K, Lockau, W. Eur. J. Biochem. 1999;263:163–169.
35. Michalska, K, Brzezinski, K, Jaskolski, M. J. Biol. Chem. 2005;280:28484–28491.
36. Bruneau, L, Chapman, R, Marsolais, F. Planta 2006;224:668–679.
37. Atkins, CA, Pate, JS, Sharkey, PJ. Plant Physiol. 1975;56:807–812.
38. Zhao, Z, Wang, L, Volk, AG, Birch, NW, Stoltz, KL, Bartom, ET, et al. Genes Dev. 2019;33:61–74.
39. Z.-v.d. Poel, S, McE, N, R, Gill, H, J, Espinosa, R, 3rd, Patiel, YT, Harden, A, et al. Proc. Natl. Acad. Sci. USA. 1991;88:10735–10739.
40. Makino, M, Kojima, T, Ohgushi, T, Yamashina, I. J. Biochem. 1968;63:186–192.
41. Enomaa, N, Heiskanen, T, Hallia, R, Sormunen, R, Seppala, R, Vihinen, M, Peltonen, L. Biochem. J. 1992;286:613–618.
42. Tollerud, OK, Aronson, NN Jr. Biochem. J. 1989;260, 101-108.
43. Xuan, J, Tarentino, AL, Grimwood, BG, Plummer, TH Jr, Cui, T, Guan, C, Van Roey, P. Protein Sci. 1998;7:774–781.
44. Qian, X, Guan, C, Guo, IC. Structure 2003;11:997–1003.
45. Li, W, Irani, S, Crutchfield, A, Hodge, K, Matthews, W, Patiel, P, Zhang, YJ, Stone, E. Biochemistry 2016;55:960–969.
46. Moreno-Enriquez, A, Evangelista-Martinez, Z, Gonzalez-Mondragon, EG, Calderon-Flores, A, Arreguin, R, Perez-Rueda, E, Huerta-Saquero, A. J. Microbiol. Biotechnol. 2012;22:292–300.
47. Meena, B, Anburajan, L, Sathish, T, Vijaya, RR, Dharani, G, Vinithkumar, NV, Kirubagaran, R. Sci. Re. 2015;5:12404.
48. Nomme, J, Su, Y, Konrad, M, Lavie, A. Biochemistry 2012;51:6816–6826.
49. Nomme, J, Su, Y, Lavie, A. J. Mol. Biol. 2014;426:2471–2485.
50. Ekman, D, Bjorklund, AK, Elofsson, A. J. Mol. Biol. 2007;372:1337–1348.
51. Michalska, K, Hernandez-Santoyo, A, Jaskolski, M. J. Biol. Chem. 2008;283:13388–13397.
52. Avramis, VI, Tiwari, PN. Int J Nanomedicine 2006;1:241–254.
53. Offman, MN, Krol, M, Patel, N, Krishnan, S, Liu, J, Saha, V, Bates, PA. Blood 2011;117:1614–1621.
54. Anishkin, A, Vanegas, JM, Rogers, D, Lorenzi, PL, Chan, WC, Purwaha, P, et al. J. Mol. Biol. 2015;427:2867–2885.
55. Nguyen, H. A, Su, Y, Lavie, A. Biochemistry 55:2016;1246–1253.
56. Borek, D, Kozak, M, Pei, J, Jaskolski, M. FEBS J. 2014;281:4097–4111.
57. Derst, C, Henseling, J, Rohm, KH. Protein Sci. 2000;9:2009–2017.
58. Nguyen, HA, Durden, DL, Lavie, A. Sci. Rep. 2017;7:41643.
59. Nguyen, HA, Su, Y, Zhang, JY, Antanasijevic, A, Caffrey, M. Cancer Res. 2018;78:1549–1560.
60. Schalk, AM, Antanasijevic, A, Caffrey, M, Lavie, A. J. Mol. Biol. Chem. 2016;291:5088–5100.
61. Dunlop, PC, Meyer, GM, Ban, D, Roon, RJ. J. Biol. Chem. 1978;253:1297–1304.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

---

**How to cite this article:** da Silva LS, Doonan LB, Pessoa A, de Oliveira MA, Long PF. Structural and functional diversity of asparaginases: Overview and recommendations for a revised nomenclature. 2021;1-11. https://doi.org/10.1002/bab.2127