

Sequence analysis of enzymes with asparaginase activity[★]✉

Dominika Borek^{1★} and Mariusz Jaskólski^{1, 2 ✉}

¹*Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznań, Poland;* ²*Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland*

Received: 16 October, 2001; accepted: 17 November, 2001

Key words: L-asparagine, asparaginase, aspartylglucosaminidase, N-terminal nucleophile hydrolase, lysophospholipase, tRNA amidotransferase

Asparaginases catalyze the hydrolysis of asparagine to aspartic acid and ammonia. Enzymes with asparaginase activity play an important role both in the metabolism of all living organisms as well as in pharmacology. The main goal of this paper is to attempt a classification of all known enzymes with asparaginase activity, based on their amino acid sequences. Some possible phylogenetic consequences are also discussed using dendrograms and structural information derived from crystallographic studies.

L-Asparaginases (EC 3.5.1.1) hydrolyze L-asparagine to L-aspartate and ammonia (Fig. 1). Both the substrate and the product of this enzymatic reaction play important roles in a number of metabolic processes in all organisms, from bacteria to mammals. In plants, L-asparagine is the most abundant metabolite for storage and transport of nitrogen that is

utilized in protein biosynthesis [1–3]. In the human body, L-aspartate plays an important role as a precursor of ornithine in the urea cycle and in transamination reactions forming oxaloacetate in the gluconeogenic pathway leading to glucose. Initially, the interest in asparaginases was instigated by their antitumor activity [4–7]. Two of the bacterial

[★]Presented at the XXXVII Convention of the Polish Biochemical Society, Toruń, Poland, September, 10–14, 2001.

[✉]The research of Mariusz Jaskólski was supported in part by an International Research Scholar's award from the Howard Hughes Medical Institute.

[✉]Present address: Department of Biochemistry, UT Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9038, U.S.A.

[✉]Corresponding author: Prof. Mariusz Jaskólski, Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Grunwaldzka 6, 60-780 Poznań, Poland; tel.: (48 61) 829 1274; fax.: (48 61) 865 8008; e-mail: mariuszj@amu.edu.pl

Abbreviations: NCBI, National Center for Biotechnology Information; Ntn-hydrolase, N-terminal nucleophile hydrolase; ORF, open reading frame.

Apart from the lysophospholipase activity, the protein also possesses asparaginase and platelet-activating acetylhydrolase activities. It has been postulated that the same active center is involved in all these reactions [19]. The crystallographic structure is not known yet, but biochemical data indicate that the protein (both isolated and recombinant) is active in monomeric form [19].

In archaea, the synthesis of Gln-tRNA^{Gln} can proceed *via* an unusual pathway that is an alternative to the standard aminoacylation mechanism [20]. This pathway involves mischarging of tRNA^{Gln} with the noncognate amino acid Glu, with subsequent modification of the tRNA-bound glutamic acid in a transamidation reaction. This second process is catalyzed in archaea by a two-domain amidotransferase (Glu-AdT) [21, 22]. The α subunit of these enzymes shares about 20% identity and 30% similarity with EcAI. The α subunit probably provides the amino group to modify the tRNA-bound glutamic acid. The enzyme can use both L-glutamine and L-asparagine as amide donors [21]. The involvement of the amino-acid-metabolizing asparaginase in early protein biosynthesis machinery as well as the fact that in some organisms tRNA-dependent transamidation may be the sole biosynthetic route to asparagine, have led to an interesting implication of Asp and Glu as the earliest amino acids and lent support to the idea that amino-acid metabolism had an important role in the evolution of the protein synthesis system [21, 23, 24].

The reaction of L-asparagine hydrolysis in plants is catalyzed by a different class of enzymes, plant-type asparaginases, with no homology to the bacterial-type enzymes [25–31]. The most studied enzymes in this class are from legume plants and are involved in metabolic pathways connected with the assimilation of atmospheric nitrogen [28, 30, 31]. These enzymes are still awaiting their crystallographic characterization, but the existing biochemical data and sequence

homology indicate that they belong to the family of N-terminal nucleophile (Ntn) amidohydrolases [28, 30–36]. The affinity for L-asparagine is low ($K_m = 10^{-2}$ M), but the high concentration of L-asparagine (40 mM) presented to the enzyme when its activity is required, guarantees efficient processing [37]. The asparaginases from plants show about 60% sequence similarity to aspartylglucosaminidases (glycosylasparaginases). Aspartylglucosaminidases (EC 3.5.1.26) in humans catalyze the last stage of degradation of glycosylated proteins, i.e. the hydrolysis of the glycosidic bond between the sugar chains and the L-asparagine side chain [38]. They can also act as asparaginases, albeit with rather low substrate affinity ($K_m = 10^{-3}$ M) [39, 40]. The crystallographic structures determined for aspartylglucosaminidases from *Homo sapiens* [41] and *Flavobacterium meningosepticum* [42–44] show them to be heterotetramers created by autoproteolytic cleavage of two precursor protein chains. The first residue (Thr, Ser, or Cys) of the C-terminal (β) domain liberated in the autocatalytic process, serves the role of the nucleophilic agent. The nucleophilic character of the oxygen or sulfur atom is increased by the free amino group of the same amino acid [34]. It is intriguing that the *E. coli* genome contains a sequence, deposited in the genomic data bases as an open reading frame (ORF) with unknown function (NCBI access code: AAC73915) [45], that shows very high homology to plant-type asparaginases and glycosylasparaginases. We have shown earlier that the product of this *ybiK* gene is a functional protein and described some of its properties, demonstrating, for example, that it is related to aspartylglucosaminidases and plant-type asparaginases [46]. A similar protein from *Salmonella enterica* serovar Typhimurium that has recently been expressed and purified also reveals the characteristic properties of an Ntn hydrolase. It is interesting that this protein has no detectable aspartylglucosaminidase activity but instead has been re-

ported to have isoasparagine aminopeptidase activity [33].

The third class of asparaginase sequences is typified by the thermolabile asparaginase from *Rhizobium etli* (ReA), with no homology to other proteins with known function [47]. *Rhizobium etli* uses the enzyme in reactions that are a source of carbon and nitrogen in its metabolism.

MATERIALS AND METHODS

The amino-acid sequences of the three *Escherichia coli* asparaginases encoded by the *ansA*, *ansB*, and *ybiK* genes (NCBI access codes: AAC74837, AAC75994, and AAC73915, respectively) and the ReA sequence encoded by the *ansA* gene (NCBI access code: AAF00929), were used as probes for searching the GenBank [48] in BLAST [49, 50]. One file was created for those sequences that showed similarity to the *E. coli* sequences EcAI or EcAII (products of the *ansA* or *ansB* genes, respectively). A separate file held the sequences identified by the EcAIII probe (product of the *ybiK* gene) or by the ReA probe. Only complete sequences were used for alignments. Multiple sequence alignments for each file were performed using CLUSTALX [51]. Subsequently, phylogenetic trees were created using the neighbor-joining method [52] for the bacterial- and plant-type asparaginases. The sequences that showed similarity to *Rhizobium etli* asparaginase were too few for a reliable phylogenetic tree. Finally, the trees were drawn using the TreeView program [53].

RESULTS AND DISCUSSION

Bacterial-type asparaginases

All analyzed sequences show absolute conservation of the two threonine-containing motifs (Fig. 2) and of the lysine residue corresponding to EcAII Lys162 (*vide supra*). The

dendrogram for the bacterial-type asparaginases (Fig. 3) is in good agreement with the tree of life [54] and is clearly divided into two main parts. One of them comprises mainly sequences from archaea and eukaryota, while the other one includes eubacterial asparaginases. In the archaeal part of the dendrogram, the α subunit of the Glu-tRNA-modifying amidotransferases creates one subset. Another group of proteins, lysophospholipases, also form a discernible subset. It is interesting that differences in quaternary structure can be correlated with the results of the sequence alignments. In all the enzymes labeled in green in Fig. 3, but not in the other sequences, there is a characteristic extra insertion, about 25 amino acids long and with some sequence conservation, which follows the position of residue 42 in EcAII. Interestingly, all those "green" sequences consistently correspond to eukaryotic organisms, even though the branch where they are located on the dendrogram contains bacterial species as well. The 42 site in EcAII marks the end of a long, flexible loop that in all type II enzymes creates the environment of the active site. The presence of a longer loop in lysophospholipases explains why these enzymes can catalyze a different reaction and why they are active as monomers. For archaeal amidotransferases, the alignment reveals not only the presence of an additional archaeal motif at the N-terminus reported earlier [21], but also a short motif after residue 151 (EcAII sequence). In EcAII, residue 151 is the first amino acid of a β hairpin loop that is located at the interface between monomers A and B, and that seems to play a role in tetramer stability.

The second branch of the dendrogram comprises mainly eubacterial sequences known as type II asparaginases. Typically, these enzymes are secreted into the periplasm, in contrast to type I bacterial L-asparaginases, which are cytosolic. Yeasts, such as *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, also produce two biologically and genetically distinct L-asparaginases. One of

these is a cell-wall glycoprotein, while the other is a constitutively expressed cellular enzyme. Their amino-acid sequences clearly classify both these enzymes in the bacterial type II branch (Fig. 3), in line with the earlier suggestion that the asparaginase genes were inde-

four branches. The aspartylglucosaminidase sequences form one of the branches. The enzymes from plants and their close homologs form another branch. The product of the *E. coli ybiK* gene (EcAIII) is also placed on this branch supporting the hypothesis based on

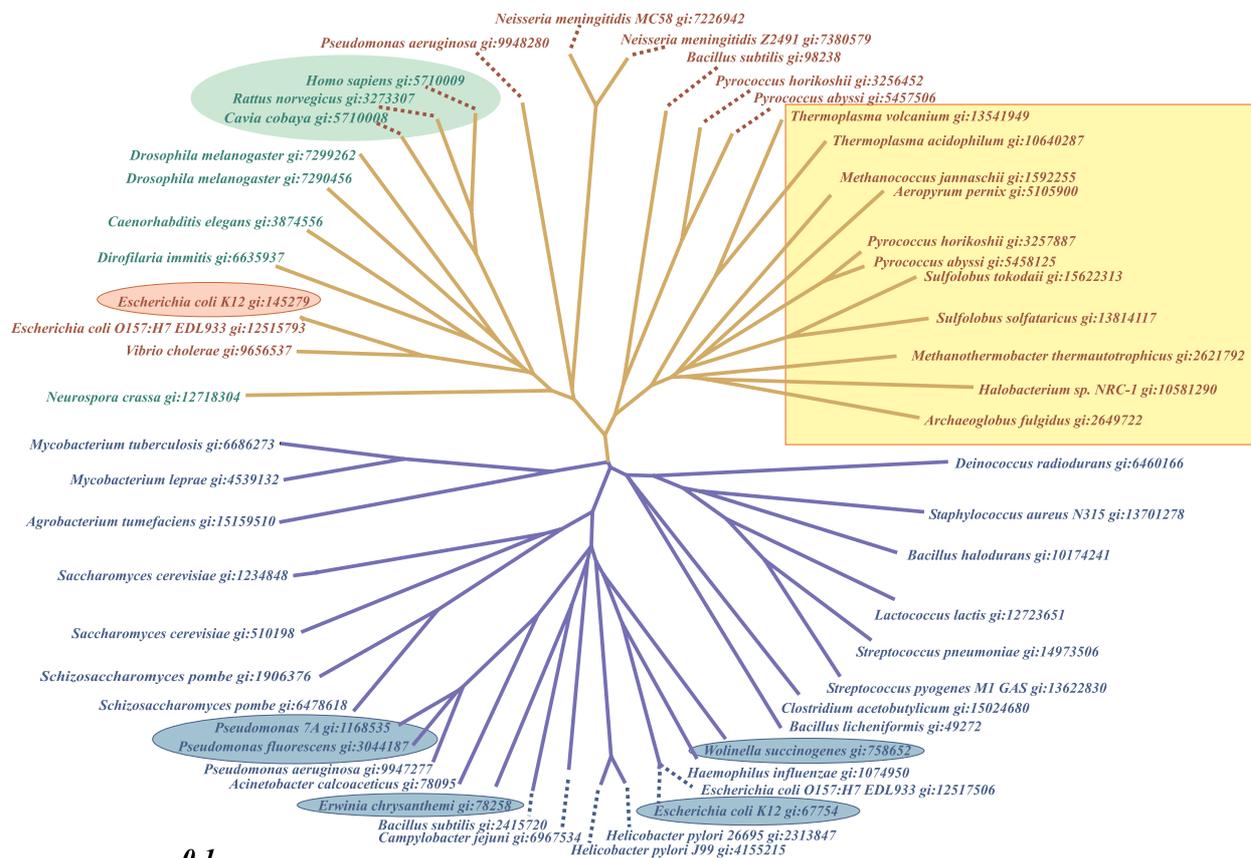


Figure 3. A dendrogram of bacterial-type asparaginases.

Orange branch – type I bacterial asparaginases, blue branch – type II bacterial asparaginases. Lettering colors: blue – sequences with high homology to EcAII, brown – sequences with high homology to EcAI, green – sequences with high homology to lysophospholipases. Yellow box – subunit α of Glu-AdT amidotransferases, green oval – lysophospholipases, pink oval – EcAI, reported to be active in dimeric form [18], blue ovals – tetrameric bacterial type II asparaginases with antitumor activity. *gi* is a unique sequence identifier assigned by the NCBI [48]. The scale bar represents 10% divergence.

pendently duplicated in prokaryota and eukaryota after the divergence of these lineages [15].

Plant-type asparaginases

All sequences in this group show an absolute conservation of the catalytic threonine residue [40–43]. The dendrogram for plant-type asparaginases (Fig. 4) is clearly divided into

biochemical data that it is a close relative of plant asparaginases [55]. The third branch comprises mainly archaeal sequences with unknown function, while the fourth one includes sequences from eukaryota for which no function is known either. It is interesting that sequences from *Homo sapiens*, *Drosophila melanogaster*, and *Arabidopsis thaliana* are found on all branches, except the archaeal branch, sometimes in more than one example.

Table 1. Sequence identity (red) and similarity (blue) for the *Rhizobium etli* asparaginase family

	<i>Rhizobium etli</i> gi:6007496	<i>Streptomyces griseus</i> gi:1139576	<i>Agrobacterium tumefaciens</i> gi:15155819	<i>Sinorhizobium meliloti</i> gi:15073886	<i>Mesorhizobium loti</i> gi:13476327	<i>Streptomyces coelicolor A3(2)</i> gi:8388734	<i>Synechococcus sp. PCC 7002</i> gi:3184558	<i>Synechocystis sp. PCC 6803</i> gi:1208456
<i>Rhizobium etli</i> gi:6007496		25	28	27	26	25	27	24
<i>Streptomyces griseus</i> gi:1139576	45		31	31	33	74	27	26
<i>Agrobacterium tumefaciens</i> gi:15155819	41	40		67	52	32	28	25
<i>Sinorhizobium meliloti</i> gi:15073886	41	35	78		56	32	29	24
<i>Mesorhizobium loti</i> gi:13476327	43	45	65	71		31	29	24
<i>Streptomyces coelicolor A3(2)</i> gi:8388734	34	82	45	46	45		27	28
<i>Synechococcus sp. PCC 7002</i> gi:3184558	40	37	42	43	44	38		59
<i>Synechocystis sp. PCC 6803</i> gi:1208456	37	37	42	42	41	40	75	

Sequences with homology to *Rhizobium etli* asparaginase

The alignment of eight sequences similar to *Rhizobium etli* asparaginase (Table 1) reveals the presence of a highly conserved motif NCSGKH. Since this is the only fragment with high conservation, it could be implicated for a role in the catalytic mechanism. It is interesting to note that no homologs of ReA are present in many of the genomes that have been sequenced to-date.

General classification

Based on the above alignments, as well as on biochemical and crystallographic data, the known asparaginase sequences can be divided into three families (Fig. 5). The first family corresponds to bacterial-type asparaginases, the second to plant-type asparaginases and the third one to enzymes similar to *Rhizobium etli* asparaginase. The enzymes included in

this classification catalyze a number of different enzymatic reactions, including reactions connected with protein biosynthesis [21].

REFERENCES

1. Atkins, C.A., Pate, S.J. & Sharkey, P.J. (1975) Asparaginase metabolism – key to the nitrogen nutrition of developing legume seeds. *Plant Physiol.* **56**, 807–812.
2. Gomes, M.A.F. & Sodek, L. (1984) Allantoinase and asparaginase activities in maturing fruits of nodulated and non-nodulated soybeans. *Physiol. Plant* **62**, 105–109.
3. Sieciechowicz, K.A., Joy, K.W. & Ireland, R.J. (1988) The metabolism of asparagine in plants. *Phytochemistry* **27**, 663–671.
4. Broome, J. (1961) Evidence that the L-asparaginase activity of guinea pig serum is re-

- sponsible for its antilymphoma effects. *Nature* **191**, 1114–1115.
5. Mashburn, L. & Wriston, J. (1963) Tumor inhibitory effect of L-asparaginase. *Biochem. Biophys. Res. Commun.* **12**, 50–55.
 6. Mashburn, L. & Wriston, J. (1964) Tumor inhibitory effect of L-asparaginase from *Escherichia coli*. *Arch. Biochem. Biophys.* **105**, 450–452.
 7. Campbell, H.A., Mashburn, L.T., Boyse, A.E. & Old, L.J. (1967) Two L-asparaginases from *Escherichia coli* B. Their separation, purification, and antitumor activity. *Biochemistry* **6**, 721–730.
 8. Campbell, H.A. & Mashburn, L.T. (1969) L-Asparaginase EC-2 from *Escherichia coli*. Some substrate specificity characteristics. *Biochemistry* **8**, 3768–3775.
 9. Roberts, J., Prager, M.D. & Bachynsky, N. (1966) The antitumor activity of *Escherichia coli* L-asparaginase. *Cancer Res.* **26**, 2213–2217.
 10. Boyse, E., Old, L., Campbell, H. & Mashburn, L. (1967) Suppression of murine leukemias by L-asparaginase, incidence of sensitivity among leukemias of various types: Comparative inhibitory activities of guinea pig serum L-asparaginase and *Escherichia coli* L-asparaginase. *J. Exp. Med.* **125**, 17–31.
 11. Lay, H., Ekert, H. & Colebatch, J. (1975) Combination chemotherapy for children with acute lymphocytic leukemia who fail to respond to standard remission-induction therapy. *Cancer* **36**, 1220–1222.
 12. Bodey, G., Hewlett, J., Coltman, C., Rodriguez, V. & Freireich, E. (1974) Therapy of adult acute leukemia with daunorubicin and L-asparaginase. *Cancer* **33**, 626–630.
 13. Swain, A.L., Jaskólski, M., Housset, D., Rao, J.K.M. & Wlodawer, A. (1993) Crystal structure of *Escherichia coli* L-asparaginase, an enzyme used in cancer therapy. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1472–1478.
 14. Lubkowski, J., Palm, G.J., Gilliland, G.L., Derst, C. & Rohm, K.H. (1996) Crystal structure and amino acid sequence of *Wolinella succinogenes* L-asparaginase. *Eur. J. Biochem.* **241**, 201–207.
 15. Bonthron, D.T. & Jaskólski, M. (1997) Why a “benign” mutation kills enzyme activity. Structure-based analysis of the A176V mutant of *Saccharomyces cerevisiae* L-asparaginase I. *Acta Biochim. Polon.* **44**, 491–504.
 16. Lubkowski, J., Wlodawer, A., Ammon, H.L., Copeland, T.D. & Swain, A.L. (1994) Structural characterization of *Pseudomonas* 7A glutaminase-asparaginase. *Biochemistry* **33**, 10257–10265.
 17. Lubkowski, J., Wlodawer, A., Housset, D., Weber, I.T., Ammon, H.L., Murphy, K.C. & Swain, A.L. (1994) Refined crystal structure of *Acinetobacter glutaminasificans* glutaminase-asparaginase. *Acta Crystallogr. D* **50**, 826–832.
 18. Jerlstrom, P.G., Bejjak, D.A., Jennings, M.P. & Beacham, I.R. (1989) Structure and expression in *Escherichia coli* K-12 of the L-asparaginase I-encoding *ansA* gene and its flanking regions. *Gene* **78**, 37–46.
 19. Sugimoto, H., Odani, S. & Yamashita, S. (1998) Cloning and expression of cDNA encoding rat liver 60-kDa lysophospholipase containing an asparaginase-like region and ankyrin repeat. *J. Biol. Chem.* **273**, 12536–12542.
 20. Curnow, A.W., Ibba, M. & Söll, D. (1996) tRNA-dependent asparagine formation. *Nature* **382**, 589–590.
 21. Tumbula, D.L., Becker, H.D., Chang, W. & Söll, D. (2000) Domain-specific recruitment of amide amino acids for protein synthesis. *Nature* **407**, 106–110.

22. Becker, H.D. & Kern, D. (1998) *Thermus thermophilus*: A link in evolution of the tRNA-dependent amino acid amidation pathways. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12832–12837.
23. Wong, J.T.F. (1975) A co-evolution theory of the genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1909–1912.
24. Wong, J.T.F. (1976) The evolution of a universal genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2336–2340.
25. Lea, P.J., Farnden, L. & Mifflin, B.J. (1978) The purification and properties of asparaginase from *Lupinus* species. *Phytochemistry* **17**, 217–222.
26. Lea, P.J. & Mifflin, B.J. (1980) *The Biochemistry of Plants*; vol. 5, pp. 290–293, Academic Press, New York.
27. Lea, P.J., Festenstein, G., Hughes, J.S. & Mifflin, B.J. (1984) An immunological and enzymological survey of asparaginase in seeds of *Lupinus*. *Phytochemistry* **3**, 511–514.
28. Dickson, J.M.J.J., Vincze, E., Grant, M.R., Smith, L.A., Rodber, K.A., Farnden, K.J.F. & Reynolds, P.H.S. (1992) Molecular cloning of the gene encoding developing seed L-asparaginase from *Lupinus angustifolius*. *Plant Mol. Biol.* **20**, 333–336.
29. Lough, T.J., Chang, K.S., Carne, A., Monk, B.C., Reynolds, P.H. & Farnden, K.J. (1992) L-Asparaginase from developing seeds of *Lupinus arboreus*. *Phytochemistry* **31**, 1519–1527.
30. Lough, T.J., Reddington, B.D., Grant, M.R., Hill, D.F., Reynolds, P.H. & Farnden, K.J. (1992) The isolation and characterisation of a cDNA clone encoding L-asparaginase from developing seeds of lupin (*Lupinus arboreus*). *Plant Mol. Biol.* **19**, 391–399.
31. Borek, D., Podkowinski, J., Kisiel, A. & Jaskolski, M. (1999) Isolation and characterization of cDNA encoding L-asparaginase from *Lupinus luteus* (Accession No. AF112444). (PGR99-050) *Plant Physiol.* **119**, 1568.
32. Mononen, I., Fisher, K.J., Kaartinen, V. & Aronson, N.N., Jr. (1993) Aspartylglycosaminuria: Protein chemistry and molecular biology of the most common lysosomal storage disorder of glycoprotein degradation. *FASEB J.* **7**, 1247–1256.
33. Larsen, R.A., Knox, T.M. & Miller, C.G. (2001) Aspartic peptide hydrolases in *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.* **183**, 3089–3097.
34. Brannigan, J.A., Dodson, G., Duggleby, H.J., Moody, P.C., Smith, J.L., Tomchick, D.R. & Murzin, A.G. (1995) A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* **378**, 416–419.
35. Duggleby, H.J., Tolley, S.P., Hill, C.P., Dodson, E.J., Dodson, G. & Moody, P.C. (1995) Penicillin acylase has a single-amino-acid catalytic centre. *Nature* **373**, 264–268.
36. Lowe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. & Huber, R. (1995) Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* **268**, 533–539.
37. Bewley, J.D. & Black, M. (1994) *Seeds. Physiology of Development and Germination*; 2nd edn., pp. 334–335, Plenum Press, New York.
38. Tarentino, A.L. & Maley, F. (1969) The purification and properties of a beta-aspartyl *N*-acetylglucosylamine amidohydrolase from hen oviduct. *Arch. Biochem. Biophys.* **130**, 295–303.
39. Noronkoski, T., Stoineva, I.B., Petkov, D.D. & Mononen, I. (1997) Recombinant human glycosylasparaginase catalyzes hydrolysis of L-asparagine. *FEBS Lett.* **412**, 149–152.
40. Liu, Y., Guan, C. & Aronson, N.N. (1998) Site-directed mutagenesis of essential residues involved in the mechanism of bacterial

- glycosylasparaginase. *J. Biol. Chem.* **273**, 9688–9694.
41. Oinonen, C., Tikkanen, R., Rouvinen, J. & Peltonen, L. (1995) Three-dimensional structure of human lysosomal aspartylglucosaminidase. *Nat. Struct. Biol.* **2**, 1102–1108.
42. Xuan, J., Tarentino, A.L., Grimwood, B.G., Plummer, T.H., Jr., Cui, T., Guan, C. & Van Roey, P. (1998) Crystal structure of glycosylasparaginase from *Flavobacterium meningosepticum*. *Protein Sci.* **7**, 774–781.
43. Guo, H.C., Xu, Q., Buckley, D. & Guan, C. (1998) Crystal structures of *Flavobacterium glycosylasparaginase*. An N-terminal nucleophile hydrolase activated by intramolecular proteolysis. *J. Biol. Chem.* **273**, 20205–20212.
44. Xu, Q., Buckley, D., Guan, C. & Guo, H.C. (1999) Structural insights into the mechanism of intramolecular proteolysis. *Cell* **98**, 651–661.
45. Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. & Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474.
46. Borek, D. & Jaskolski, M. (2000) Crystallization and preliminary crystallographic studies of a new L-asparaginase encoded by the *Escherichia coli* genome. *Acta Crystallogr. D* **56**, 1505–1507.
47. Ortuño-Olea, L. & Durán-Vargas, S. (2000) The L-asparagine operon of *Rhizobium etli* contains a gene encoding an atypical asparaginase. *FEMS Microbiol. Lett.* **189**, 177–182.
48. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A. & Wheeler, D.L. (2000) GenBank. *Nucleic Acids Res.* **28**, 15–18.
49. Karlin, S. & Altschul, S.F. (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2264–2268.
50. Karlin, S. & Altschul, S.F. (1993) Applications and statistics for multiple high-scoring segments in molecular sequences. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5873–5877.
51. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997) The CLUSTALX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882.
52. Saitou, N. & Nei, M. (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
53. Page, R.D.M. (1996) TreeView: An application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* **12**, 357–358.
54. Woese, C. (2000) Interpreting the universal phylogenetic tree. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8392–8396.
55. Borek, D. (2001) *Ph.D. Thesis*. A. Mickiewicz University, Poznań, Poland.
56. Hofmann, K., Bucher, P., Falquet, L. & Bairoch, A. (1999) The PROSITE database, its status in 1999. *Nucleic Acids Res.* **27**, 215–219.