Identification of L-asparaginases from *Streptomyces* strains with competitive activity and immunogenic profiles: a bioinformatic approach

Iván González-Torres [1], **Ernesto Pérez-Rueda** [2], **Zahaed Evangelista-Martínez** [3], **Andrés Zárate-Romero** [1,4], **Alejandro Huerta-Saquero** [Corresp. 1]

[1] Centro de Nanociencias y Nanotecnología, Universidad Nacional Autónoma de México, Ensenada, Baja California, México

[2] Instituto de Matemáticas Aplicadas y Sistemas, Universidad Nacional Autónoma de México, Mérida, Yucatán, México

[3] Subsede Sureste, Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco, AC, Mérida, Yucatán, México

[4] Consejo Nacional de Ciencia y Tecnología, Ciudad de México, Mexico

Corresponding Author: Alejandro Huerta-Saquero
Email address: saquero@cnyn.unam.mx

The enzyme L-asparaginase from *Escherichia coli* is a therapeutic enzyme that has been a cornerstone in the clinical treatment of acute lymphoblastic leukemia for the last decades. However, treatment effectiveness is limited by the highly immunogenic nature of the protein and its cross-reactivity towards L-glutamine. In this work, a bioinformatic approach was used to identify, select and computationally characterize L-asparaginases from *Streptomyces* through sequence-based screening analyses, immunoinformatics, homology modeling, and molecular docking studies. Based on its predicted low immunogenicity and excellent enzymatic activity, we selected a previously uncharacterized L-asparaginase from *Streptomyces scabrisporus*. Furthermore, two putative asparaginase binding sites were identified and a 3D model is proposed. These promising features allow us to propose L-asparaginase from *S. scabrisporus* as an alternative for the treatment of acute lymphocytic leukemia.

Identification of L-asparaginases from *Streptomyces* strains with competitive activity and immunogenic profiles: a bioinformatic approach

Iván González-Torres[1], Ernesto Perez-Rueda[2], Zahaed Evangelista-Martínez[3], Andrés Zárate-Romero[1,4], Alejandro Huerta-Saquero[1*]

[1] Centro de Nanociencias y Nanotecnología, Universidad Nacional Autónoma de México. Km. 107 Carretera Tijuana-Ensenada, Ensenada, Baja California, México.

[2] Instituto de Matemáticas Aplicadas y Sistemas, Universidad Nacional Autónoma de México, Mérida Yucatán, México.

[3] Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco, AC. Subsede Sureste, Mérida Yucatán, México.

[4] Consejo Nacional de Ciencia y Tecnología, Ciudad de México, México.

*Corresponding author:

Alejandro Huerta-Saquero[1]

Km. 107 Carretera Tijuana-Ensenada, Ensenada, Baja California, México.

Email address: saquero@cnyn.unam.mx

Abstract

The enzyme L-asparaginase from *Escherichia coli* is a therapeutic enzyme that has been a cornerstone in the clinical treatment of acute lymphoblastic leukemia for the last decades. However, treatment effectiveness is limited by the highly immunogenic nature of the protein and its cross-reactivity towards L-glutamine. In this work, a bioinformatic approach was used to identify, select and computationally characterize L-asparaginases from *Streptomyces* through sequence-based screening analyses, immunoinformatics, homology modeling, and molecular docking studies. Based on its predicted low immunogenicity and excellent enzymatic activity, we selected a previously uncharacterized L-asparaginase from *Streptomyces scabrisporus*. Furthermore, two putative asparaginase binding sites were identified and a 3D model is proposed. These promising features allow us to propose L-asparaginase from *S. scabrisporus* as an alternative for the treatment of acute lymphocytic leukemia.

Introduction

Acute lymphocytic leukemia (ALL) is a hematological disorder of the bone marrow and is characterized by abnormal proliferation of immature lymphoid line cells, blocked at an early stage of cell differentiation, that accumulate and replace healthy hematopoietic cells in the bone marrow (Pui, Relling & Downing, 2004; Onciu, 2009). ALL occurs predominantly in children of 1-4 years of age and represents approximately 25% of childhood cancers and about 80% of leukemias (Katz et al., 2015).

Although in most cases the risk factors and pathogenicity associated with ALL have not been clearly identified, the etiology of the disease has been mainly associated with a variety of conditions; cytogenetic alterations, mutations to key genes that regulate cellular proliferation,

differentiation and death; presence of oncogenic viruses, immunodeficiency, exposure to pesticides, solvents, and ionizing radiation (Pui, Relling & Downing, 2004; Bassan, Maino & Cortelazzo, 2016).

Treatment for ALL patients involve steroid drugs, prednisone, vincristine, and the enzyme L-asparaginase (ASNase) (Avramis, 2012; Schwab & Harrison, 2018). ASNase has been essential in the treatment of ALL since the 1970s, with demonstrated effectiveness as an individual drug with remission rates of up to 68% (Salzer et al., 2017). The combination of ASNase with other anticancer drugs has led to remission rates of up to 90% (Lanvers-Kaminsky, 2017).

Currently, there are four ASNase formulations available for the ALL treatment: two native forms of the enzyme, obtained from *Escherichia coli* (EcAII) and *Erwinia chrysanthemi* (ErAII), and pegylated *E. coli* ASNase (EcAII-PEG), as well as pegylated *E. chrysanthemi* ASNase (ErAII-PEG). Of these, EcAII-PEG has become the first-line treatments for ALL in the US, with EcAII the most widely used formulation. ErAII is administered to patients who have developed hypersensitivity to the above formulations (Pieters et al., 2011; Abribat, 2016; Barba et al., 2017). In recent years, evidence has been accumulating of its usefulness as an important component in the treatment of other hematological malignancies, such as acute myeloid leukemia, myelosarcoma, lymphosarcoma, Hodgkin's disease, and chronic lymphocytic leukemia (Emadi, Zokaee & Sausville, 2014; Lopes et al., 2015). Despite their high antileukemic potential, the use of ASNases in the treatment of ALL is limited by their toxicity. Among the adverse effects that have been reported are leukopenia, immune suppression, acute pancreatitis, liver dysfunction, hyperglycemia, abnormalities in hemostasis, and hemorrhages of the central nervous system (Schein et al., 1969; Ramya et al., 2012; Chan et al., 2014; Ali et al., 2016; Hijiya & van der Sluis, 2016; Kamal et al., 2019).

The generation of immune responses during treatment with ASNase is a common condition that has been reported in up to 75% of patients. These reactions depend on the formulation used, the mode of administration (intravenous or intramuscular), and the treatment protocol (Hijiya & van der Sluis, 2015). For example, between 30 and 75% of patients that receive the native form of the *E. coli* enzyme experience hypersensitivity reactions, and about 70% develop anti-EcAII antibodies after drug administration (Battistel et al., 2020); these antibodies lead to rapid inactivation of the enzyme (Walenciak et al., 2019).

Allergic reactions to ASNase, which are associated with its bacterial origin, range from mild urticaria to life-threatening anaphylactic shock. Irritation, fever, vomiting, gastrointestinal edema, and breathing difficulties are symptoms frequently reported (Lanvers-Kaminsky, 2017). On the other hand, adverse effects have been reported due to the toxicity produced by glutaminase cross activity, such as leukopenia, immune suppression, acute pancreatitis, hyperglycemia, thrombosis, neurotoxicity, and liver failure, among others (Ramya et al., 2012; Chan et al., 2014; Ali et al., 2016).

Different strategies to reduce the toxicity of ASNase have been reported, including modifications in the structure of the protein by mutagenesis, design of mutants with diminished ability to hydrolyze L-glutamine, chemical modifications in specific amino acids, and modifications to drug formulations (Ramya et al., 2012; Nguyen, Su & Lavie, 2016; Nguyen et al., 2018). Covalent conjugation of the enzyme with polyethylene glycol, known as PEGylation, reduces the incidence of hyperglycemia, pancreatitis, and anaphylaxis. This specific modification increase the half-life of the enzyme and reduces the frequency of drug administration (Thomas & Le Jeune, 2016).

On the other hand, the exploration of new sources of ASNases offers the possibility of finding versions of the enzyme with different pharmacological characteristics, potentially useful for the

treatment of ALL and other lymphomas (Krishnapura, Belur & Subramanya, 2016). In this sense, besides searching for less immunogenic asparaginases, it is essential to find those with high affinity for L-asparagine (in the micromolar range) in order to have the potential for therapeutic use. Some atypical ASNases, unrelated to EcAII and ErAII, such as *Rhizobium etli* type II ASNase (ReAII) (Ortuño-Olea & Durán-Vargas, 2000), have been proposed as alternatives with therapeutic potential; the *R. etli* ASNase has null glutaminase activity and a different immunogenic profile than *E. coli* and *E. chrysantemy* ASNases (Moreno-Enriquez et al., 2012; Huerta-Saquero et al., 2013). However, this enzyme has a low affinity for asparagine, which limits its use. Despite the success of *E. coli* and *E. chrysanthemi* ASNases in therapeutic regimens for ALL and other types of leukemia, the search for new ASNases that are less toxic and less immunogenic is necessary. In this sense, ASNases from phylogenetically distant microorganisms offer a specific target for the selection of variants with the appropriate characteristics. Among these, ASNases from *Streptomyces* are one potential group to be evaluated for immunogenicity, toxicity, and affinity for L-asparagin to obtain new ASNases with therapeutic potential. The main characteristics to select asparaginases with therapeutic potential are the high affinity for L-asparagine (in the micromolar range), null or low glutaminase cross-activity, as well as a different three-dimensional folding from the *E. coli* asparaginase, which suggests different immunogenicity**.** In this work, we develop a strategy based on bioinformatics tools to analyze and select ASNases from *Streptomyces* for ALL treatment, taking advantage of its phylogenetic distance from *E. coli*, looking for those candidates that meet the two fundamental criteria: asparaginases with high affinity for asparagine (using active site prediction tools and molecular docking), and that have lower immunogenicity (using antigenicity and protein structure prediction tools). As a reference, we selected the *E. coli* and *Streptomyces coelicolor* ASNases. The importance of this novel approach is discussed.

Materials & Methods

Identification and selection of homologous L-Asparaginases

Putative ASNases from *Streptomyces* were identified through a BLASTp search against the NR database of the NCBI using as seeds the amino acid sequences of EcAII (ID P00805) and *Streptomyces coelicolor* type II ASNase (ScAII; ID Q9K4F5). The search was restricted to the *Streptomyces* taxon (Taxid 1883), and an E-value less than 1e-06 was considered significant. Partial proteins and those from unidentified *Streptomyces* strains were excluded. In a posterior step, the set of protein sequences was filtered at 60% identity as cutoff to avoid redundancy, using the CD-Hit program (http://weizhongli-lab.org/cdhit_suite/cgi-bin/index.cgi) (Huang et al., 2010). Each cluster was analyzed using the HMMER program on the PFAM server (http://pfam.xfam.org/) to determine the protein family to which they belonged (Finn, Clements & Eddy, 2011; Finn et al., 2016).

1.2 Phylogenetic analysis

ASNases amino acid sequence alignments were performed using Clustal Omega (Sievers et al., 2011) with default parameters. The quality of the alignments was improved using the model PF06089.11 or PF00710.11 of ASNase, as required. Multiple sequence alignment statistics were computed with AliStat (http://www.csb.yale.edu/userguides/seq/hmmer/docs/node27.html).

Phylogenetic analyses were carried out using the maximum-likelihood method with the program Mega 7. The WAG model was chosen as substitution model, and 1000 replicates were performed. The best tree was calculated using the majority rule. Additionally, *E. coli type I ASNase* (EcAI) was included in the phylogenetic analysis of the PF00710.11 cluster. EcAI is closely related to

136 EcA but it does not have therapeutic potential. For the PF06089.11 cluster, *Rhizobium etli* type II
137 ASNase (ReAII) was included in the analysis.

138 **1.3 Antigenicity prediction**

139 The prediction of the probability of antigenicity of each ASNase was calculated with the server
140 ANTIGENpro (http://scratch.proteomics.ics.uci.edu/) (Magnan et al., 2010). ANTIGENpro is a
141 sequence-based, alignment-free, protein antigenicity predictor with an estimated accuracy of 82%.
142

143 **1.4 HLA class II binding prediction**

144 The amino acid sequence of each candidate ASNase was screened for T-cells epitopes with the
145 MHC II Analysis Resource at the Immune Epitope Data Base (IEDB) server
146 (http://tools.iedb.org/mhcii/). MHC II Analysis Resource parses sequences into 15-mer and
147 assesses the binding potential of each 15-mer to MHC class II molecules of one or more HLA
148 alleles. The IEDB recommended method was used for predictions for a set of 8 HLA alleles that
149 collectively represent >%95 world population: HLA-DRB1*01:01, HLA-DRB1*03:01, HLA-
150 DRB1*04:01, HLA-DRB1*07:01, HLA-DRB1*08:01, HLA-DRB1*11:01, HLA-DRB1*13:01
151 and HLA-DRB1*15:01. The IEDB-recommended method uses the consensus approach,
152 combining NN-align, SMM-align, CombLib, Sturniolo, and NetMHCIIpan (Wang et al., 2010).
153 For each peptide, a percentile rank is generated by comparing the peptide's score against the scores
154 of five million random 15-mer selected from SWISSPROT database, and the median percentile
155 rank is used to calculate a consensus percentile rank (CPR). Peptides with a CPR < 2 were defined
156 as high-affinity binders and thus selected for epitope density (ED) calculation. Multiple 9-mer
157 cores were identified in overlapped 15-mer peptides. To reduce overestimation of predicted
158 peptides, only the 9-mer cores, predicted by using the Sturniolo method (Sturniolo et al., 1999)

and with a CPR < 1, were considered for the analysis. Finally, epitope density (ED) was calculated using the follow equation, modified from (Santos et al., 2013):

$$ED = \frac{Predicted\ epitope * (2 - Affinity\ average\ (cpr))}{Protein\ length\ size - Epitope\ size + 1}$$

Where Predicted epitope is the number of epitopes with a CPR < 1.

Epitope coverage was calculated as the number of alleles covered by the epitope consensus, according to the following assumption: when a small number of alleles is covered, a lower percentage of the population will develop sensitivity to ASNase.

1.5 Protein structure prediction, refinement and quality assessment

The three-dimensional structures of the selected ASNases was modeled by homology using the I-Tasser server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Zhang, 2008). In brief, starting from an amino acid sequence, I-Tasser generates three-dimensional atomic models from multiple threading alignments and iterative structural assembly simulations. A C-score, provided as an estimate of the accuracy of the models generated, typically ranges between -5 to +2, with a higher value indicating higher confidence, and vice versa (Roy, Kucukural & Zhang, 2010).

For each ASNase, the model with the higher C-score was selected and then refined using the ModRefiner server (https://zhanglab.ccmb.med.umich.edu/ModRefiner/). ModRefiner improves the physical quality and structural accuracy of three-dimensional protein structures by a two-step, atomic-level energy minimization (Xu & Zhang, 2011). Finally, the quality of the models was evaluated by RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php), Qmean (https://swissmodel.expasy.org/qmean/), and Verify3D (http:/servicesn.mbi.ucla.edu/Verify3D).

182

183 **1.6 Molecular docking**

184 The molecular coupling was carried out using Autodock Tools software (Sanner, 1999; Morris et
185 al., 2009). EcAII (PDB ID: 3ECA) was recovered from the PDB protein database
186 (http://www.rcsb.org/) (Swain et al., 1993; Berman et al., 2000). Once refined, selected ASNase
187 structures were prepared using Dock prep at UCSF Chimera and refined using the Gasteiger
188 method (Gasteiger and Marsili, 1978).

189 The three-dimensional structures of the asparagine and glutamine ligands were obtained from the
190 DrugBank repository (https://www.drugbank.ca/; accession numbers DB00174 and DB00130,
191 respectively) (Wishart et al., 2018). The preparation of the ligands was carried out by the Gasteiger
192 method and, finally, the allocation of the rotation centers was determined (Gasteiger and Marsili,
193 1978).

194 For each ASNase, the search box was focused on previously proposed active sites. The box size
195 was defined to cover all residues of the ligand binding site, using a grid size of 0.375 Å.

196 Blind molecular docking was performed with Autodock 4.2 software, using the Lamarckian
197 genetic algorithm, with 1000 runs, for a population size equal to 150, with 2.5 x 10^6 evaluations,
198 a mutation rate equal to 0.02 in 27,000 generations.

199 In addition, the active site location was predicted by AutoLigand (Harris, Olson & Goodsell,
200 2008). Briefly, AutoLigand identifies sites of maximum affinity from maps generated by
201 AutoGrid, finding regions with better energy and a lower volume.

202

Results

2.1 L-Asparaginases from *Streptomyces* cluster into two type families according to its protein architecture

The Blast search against the *Streptomyces* taxon revealed 296 putative ASNases homologous to EcAII and 703 homologous to ScAII with a significant score. After manual examination of both groups, 136 and 311 complete sequences were kept for EcAII and ScAII groups, respectively. Protein domain analysis using PFAM server showed that 136 sequences are related to the PF00710.11 family of N-terminal ASNases. For sequences homologous to ScAII, PFAM analysis revealed that they belong to the PF06089.11 family of ASNases, a group of enzymes related to ReAII, a thermolabile enzyme induced by L-asparagine and repressed by the carbon source (Moreno-Enriquez et al., 2012; Huerta-Saquero et al., 2013). Representative clusters for PF00710.11 and PF06089.11 families obtained using the CD-Hit suite program were generated at a 60% identity cutoff, with 19 and 7 putative ASNases, respectively (Table 1). ASNases sequences showed similar lengths in both clusters, ranging from 320 to 420 amino acids.

The sequences belonging to the PF00710.11 family have conserved residues located at the ligand binding site necessary for L-asparagine hydrolysis: Thr 12, Tyr 25, Ser 58, Gln 59, Thr 89, Asp 90, and Lys 162 for subunit A; Asn 248 and Glu 283 for subunit C. In this regard, Thr 12–Lys 162–Asp 90 and Thr 12–Tyr 2–Glu 283 are the catalytic triads involved in L-asparagine hydrolysis, where Thr 12 and Thr 89 are involved in the nucleophilic attack of the substrate (Gesto et al., 2013; Sanches, Kraunchenko & Polikarpov, 2016).

Concerning the PF06089.11 family, we identified an N-terminal conserved motif, with sequences NCSGKHxAM, DGCGAPL, SHSGEx(2)H, and PRSx(2)KPxQ probably involved in asparagine hydrolysis. ReAII hydrolyzes L-asparagine at similar levels to *Erwinia chrysanthemi*, but with

lower affinity than L-asparaginases from both *E. coli* and *E. chrysanthemi* (Moreno-Enriquez et al., 2012). Furthermore, ReAII is the only ASNase characterized from the PF06089.11 family

2.2 Phylogenetic analysis of ASNases

For the PF00710.11 family, EcAI was added to the multiple sequence alignment in order to know the relationship between this ASNase and the candidate ASNases. EcAI belongs to the same family of proteins as EcAII, but it does not represent a therapeutic option for ALL treatment. It is noteworthy that asparaginases can also be classified according to subcellular localization, a) periplasmic asparaginases containing secretion signal peptide and, b) asparaginases with intracellular localization. The former generally have a higher affinity for asparagine. However, according to their architecture, both types of proteins can be found in the PF00710.11 or PF06089.11 families. This is the case of *E. coli* asparaginases I and II, both belonging to the PF00710.11 family (https://pfam.xfam.org/family/PF00710#tabview=tab1). We found that the ASNase with accession number WP_059134811.1 of *Streptomyces alboniger* is grouped in the same clade as EcAI, and so it was excluded from subsequent analyses (Figure 1A).

The phylogenetic reconstruction showed three well-defined clades (Figure 1A). The first clade includes ASNases from *Streptomyces* species *S. aureocirculatus* (WP_078965752.1), *S. cattleya* (WP_014151616.1), *S. thermoautotrophicus* (KWW98572.1), *S. himastatinicu* (EFL23513.1), *S. turgidiscabies* (ELP65653.1), *S. nanshensis* (WP_070201703.1), and *S. griseus* (WP_030748190.1).

The second clade includes ASNases from *S. albidoflavus* (WP_095730579.1), *S. kebangsaanensis* (WP_073950513.1), *S. fradiae* (WP_078649241.1), *S. himastatinicus* (WP_009718687.1), *S. purpureus* (WP_078513220.1), and *S. paucisporeus* (WP_079189481.1). Finally, the third clade

contains proteins from *S.purpurogeneiscleroticus* (WP_053609500.1), *S. purpurogeneiscleroticus* (WP_053610569.1), *S. phaeochromogenes* (WP_055617501.1), and *S. lavenduligriseus* (WP_051815467.1) where EcAII was included, suggesting that proteins clustered in this clade share similar properties to EcAII. In addition, two proteins, WP_053609500.1 and WP_055617501.1, exhibited the largest proportion of antigenic regions, with almost the same probability regions as the EcAII protein.

On the other hand, for the ASNases of PF06089.11, phylogenetic analysis included both the ASNase sequence of *R. etli* and *S. coelicolor* (ReAII and ScAII, respectively) (Figure 1B). The tree defines two clades. In the first one, where the ScAII was included, we also considered ARZ68596.1 from *S. albireticuli*, SOD64826.1 from *S. zhaozhouensis*, WP_078645645 from *S. varsoviensis*, CDR15801.1 from *S. iranensis*, and WP_020554088 from *S. scabrisporus.* In the second clade were included the following proteins: WP_044373749 from *S. ahygroscopicus* and WP_078980718.1 from *S. scabrisporus.*

Figure 1. Phylogenetic tree of PF00710.11 (A) and PF06089.11 (B) families. Blue dots highlight reference sequences added to each analysis. Red dots highlight sequences used as internal controls (asparaginases from *E. coli* and *R. etli*, respectively). A total of 1000 replicates were performed. Bootstrap values are indicated.

2.3 Antigenicity predictions

The results for antigenicity showed a likelihood of being antigenic for all ASNases in both sets that was lower than that of EcAII (Figure 2). Nevertheless, among selected *Streptomyces* ASNases, the candidates from *S. purpurogeneiscleroticus* (WP_053609500.1) and *S. phaeochromogenes*

(WP_055617501.1) showed a higher probability of being antigenic, whereas the rest of the ASNases showed very low antigenicity values in comparison with an *E. coli* ASNase (P00805_EcAII).

Figure 2. ASNase antigenicity predictions. The antigenicity scores for PF00710.11 family (A) and PF06089.11 family (B) of *Streptomyces* asparaginases were compared with the EcAII antigenicity score.

2.4 T-cell epitope analysis

After antigenicity prediction, the ED, the total number of high-affinity epitopes, the affinity epitopes, and the number of HLA alleles covered by each ASNase were calculated. The results showed that the ASNases with accession numbers WP_053609500.1, WP_053610569.1, EFL23513.1, WP_095730579.1, WP_078513220.1, and WP_052425051.1 have higher EDs than the reference (P00805_EcAII; ED=0.01114; 5 covered alleles) (Figure 3).

On the other hand, the ASNase with the lowest predicted ED was WP_044373749.1, with an ED of 0.0027 and a coverage of 4 alleles, following by WP_095730579.1 (2 alleles), ELP65653.1 (3 alleles), and Q9K4F5 (3 alleles) (Table 2).

Additionally, the distribution of epitopes was mapped into the sequences of the ASNases (Figure 3). ASNases of the PF06089.11 family tended to have a lower ED (Table 3) as well as lower allele coverage than those of the PF00710.11 family (Figure 3).

293 **Figure 3.** Epitope mapping of ASNases of the PF familes evaluated, PF00710.11 and PF06089.11.
294 The epitopes identified along with the ASNase sequences are shown. The color intensity represents
295 the number of hits for each of them.
296
297 Next, ASNases with lower allele coverage, lower ED, and lower probability of antigenicity were
298 selected for further analysis. *S. coelicolor* (Q9K4F5), *S. scabrisporus* (WP_078980718.1), and *S.*
299 *albireticuli* (ARZ68596.1) were selected as promising enzymes.
300 **2.5 Protein structure predictions**
301 From selected ASNases, homology-based models were generated (I-Tasser). For the subsequent
302 analysis, the *S. scabrisporus* asparaginase II model, which had the highest C-value, was chosen
303 (WP_078980718.1 SsAII-2) (Figure 4). The stereochemical quality of the models was evaluated
304 using Ramachandran plots. In order to improve the quality of the models, these were structurally
305 refined with ModRefiner and reassessed with RAMPAGE. In addition, the Verify3D server was
306 used to determine the compatibility of the three-dimensional model with the amino acid sequence.
307 Based on the predicted structure, ASNase WP_0789718.1 (PF06089.11 family) is related in terms
308 of folding to the beta-lactamase family. Beta-lactamases (SCOP data base, entry 56600) consist of
309 a cluster of alpha-helices and an alpha/beta sandwich. This folding is also found in transpeptidases,
310 esterases, penicillin receptors, D-aminopeptidases, and glutaminases (InterPro IPR012338).
311
312 **Figure 4.** 3D protein structure prediction of *S. scabrisporus* asparaginase II (WP_078980718.1;
313 SsAII-2).
314
315 **2.6 Active site prediction**

In order to identify the active site residues of the *S. scabrisporus* ASNase (WP_0789718.1), three approaches were used: genomic comparison, blind molecular coupling simulation, and search for high-affinity binding pockets with AutoLigand (active site). To our knowledge, there is no information regarding the active site of the family PF06089.11 ASNases, so genomic comparison was not possible. Using AutoLigand, two possible high affinity binding sites for L-asparagine were identified (Figure 5A). The first (site A) had a volume of 121 $Å^3$ and an energy per volume equal to -0.2149 kcal/mol $Å^3$; the second (site B) had a volume of 101 $Å^3$ and an energy per volume equal to -0. 2136 kcal/mol $Å^3$. Site A is located between an alpha-helix in the amino terminal containing the ^{57}PRSx(2)KPxQ^{65} motif, and a loop in the central region of the enzyme, containing the ^{141}NCSGKHxAML^{150} motif (Table 3). Site B is located in a pocket formed by a set of alpha-helices in the amino terminal of the protein, marked by the presence of the ^{87}SHTGQxHFV^{95} motif. On the other hand, by performing AutoDock 4.2 whole-protein molecular coupling simulations, we found that the best ligand-enzyme interaction (L-asparagine-ASNase), with a binding free energy of -4.17 kcal/mol, targeted residues corresponding to the ^{141}NCSGKHxAML^{150} motif, which correspond to the site A (Figure 5B).

Figure 5. SsAII-2 putative binding sites. A) Site A (orange) contains the NCSGKHxAML sequence and site B (blue) contains the SHTGQxHFV motif. B) Residues involved with asparagine through a direct interaction, obtained by blind molecular docking.

Additionally, in order to validate AutoLigand analysis searching active sites in the *S. scabrisporus* ASNases, a search for binding sites in EcAII was performed. To do this, the monomeric, dimeric, and tetrameric forms of the enzyme (the latter is the catalytically active form) were analyzed using

the same conditions used for SsAII-2. It was found that AutoLigand successfully identified the binding site of L-Asn, consisting of Thr 12, Tyr 25, Ser 58, Gln 59, Thr 89, Asp 90, and Lys 162 and also Asn 248 and Glu 283 (Figure 6), the latter two only for dimeric and tetrameric forms. The sites found (red squares curves) had energies by volume equal to -0.2119, -0.2242, and -0.2366 kcal/mol $Å^3$ and a volume of 136, 122, and 102 $Å^3$ for the monomer, dimer, and tetramer, respectively (Figure 7). It is relevant that for both the dimeric and the tetramer forms, AutoLigand successfully identified L-Asn binding pockets in EcAII: the pocket formed between the amino-terminal end of subunit A and the carboxy terminal of the subunit C, as well as equivalent pockets for dimer BD. In addition, several other solutions found by AutoLigand (curve with blue or green squares), using up to 90 filling points, converge in the different joint pockets formed by dimers.

Figure 6. EcAII dimer AutoLigand analysis. Cyan EcAII subunit C is shown in cyan and subunit A in magenta. The red mesh represents the highest-affinity pocket found by AutoLigand (putative active site). The site represented in the scheme corresponds to the residues located at a maximum distance of 5 Å using 20 points: Thr 12, Tyr 25, Ser 58, Gln 59, Thr 89, and Asp 90 from subunit C and Asn 248 and Glu 283 from subunit A.

Figure 7. AutoLigand results for EcAII. The minima observed in the total energy graphs per unit volume represent putative binding sites in the structures analyzed, for the monomer, dimer, and tetramer conformation. As more filling points are used, the binding sites, cavities, or grooves are filled and the affinity decreases. The best sites are the ones with the lowest energy and the lowest volume.

362 **2.7 Molecular docking**

363 Molecular docking simulations were performed at the putative sites found (Table 4). For EcAII,
364 as the reference ASNase, Thr 12, Tyr 25, Ser 58, Gln 59, Thr 89, Asp 90, Asn 248, and Glu 283
365 were established as flexible residues; meanwhile, molecular docking for *S. scabrisporus* ASNase
366 were performed using only the rigid structure of the protein, without defining flexible side chains
367 for L-asparagine binding.
368 Our results showed a higher affinity for L-asparagine of the *S. scabrisporus* ASNase site A than
369 site B; however, the affinity was lower than that for EcAII. For *S. scabrisporus* ASNase site A,
370 the L-asparagine interacts with residues Ser 59, Lys 62, Asn 141, Ser 143, Lys 145, His 146, Gly
371 237, Lys 255, and Gly 256 (Figure 8A); for site B, the residues that interact with L-asparagine are
372 Ala 84, Gly 78, Ser 87, Tyr 163, Leu 164, and Asp165 (Figure 8B). Interestingly, from site A, Lys
373 62, Asn 141, Ser 143, Lys 145, and His 146 are highly conserved in ASNases of the PF06089.11
374 family.
375
376 **Figure 8.** Interaction maps for sites A and B from *S. scabrisporus* ASNase. The black spheres
377 represent carbon atoms, the blue nitrogen and the red oxygen. Hydrogen bonds are represented by
378 green dotted lines and hydrophobic interactions are shown as red half-moons.
379

380 ## Discussion

381 In this work, a set of bioinformatics tools were used to identify, select, and characterize ASNases
382 from the *Streptomyces* genus. ASNase identification was carried out by searching sequences
383 homologous to EcAII and ScAII. EcAII is the best-characterized and most widely used ASNase
384 for ALL treatment, while ScAII is a homologous ASNase related to ReAII, an atypical ASNase

with no glutaminase activity and with a different immunogenic profile than EcAII (Huerta-Saquero et al., 2013). The search for homologous sequences resulted in two sets of sequences with a high probability of being ASNases (E value <1e-06). These sequence sets, in turn, were classified into two different protein families based on their homology, using HMMer: PF00710.11 and PF06089.11, according to the classification of the PFAM database. So far, most of the reported ASNases belong to the PF00710.11 family and have been extensively studied. EcAII and the *E. chrysanthemi* ASNase belong to this family. On the other hand, the PF06089.11 family represents a group of atypical ASNases that remain poorly characterized. Some representative reports about these ASNases include the *R. etli* ASNase (Ortuño-Olea & Durán-Vargas, 2000; Moreno-Enriquez et al., 2012; Huerta-Saquero et al., 2013).

Interestingly, the BLAST results showed a greater abundance of PF06089.11 family sequences compared to the PF00710.11 family in *Streptomyces*. In addition, we found that about 20% of species have ASNase isoforms. In that sense, many Gram-negative bacteria have at least two isozymes of the family PF00710.11 (Fernández & Zúñiga, 2006) and, in *E. coli*, the existence of a third isoenzyme has been recently reported (Correia da Silva et al., 2018). Historically, the genus *Streptomyces* has been attractive due to the wide repertoire of bioactive molecules produced. However, searching for ASNases of pharmacological interest has been done only rarely.

After the identification of two sets of ASNases, we chose T-cell ED as the immunogenicity indicator, according to Cantor et al. (2004), Fernández et al. (2012), and Galindo-Rodríguez et al. (2017), who proposed that HLA class II molecules play a critical role in the development of specific anti-ASNase antibodies and in hypersensitivity to the enzyme (Cantor et al., 2011; Fernandez et al., 2014; Galindo-Rodríguez et al., 2017). Additionally, it has been shown that proteins that are highly immunogenic generally contain a greater amount of T-cell epitopes, or

clusters thereof (Singh et al., 2012). In addition, the measurement and prediction of ED have generated interest as useful tools for comparisons between therapeutic proteins, allowing selection of the best candidate in terms of probable immunogenicity (De Groot & Martin, 2009). In this sense, our results showed that ASNases of the PF06089.11 family contain lower EDs than enzymes of the PF00710.11 family, as well as fewer epitope clusters throughout the sequence. In addition, the allele coverage, which is related to the percentage of the population that develops a significant immune response, showed *Streptomyces* ASNases to be potential pharmacological options. In other words, due to their low content of T-cell epitopes, low antigenicity profile, and low allele coverage, *Streptomyces* ASNases represent, in terms of immunogenicity, a pharmacological alternative for ALL treatment. In this sense, the *Streptomyces brollosae* NEAE-115 ASNase has better cytotoxicity and immunogenicity profiles for use in ALL treatment, based on evaluation in a murine model, compared with EcAII (El-naggar et al., 2018). Previously, anticancer activity of the *Streptomyces fradiae* NEAE-82 ASNase in colon cancer cell cultures was reported (El-Naggar et al., 2016).

For the PF06089.11 family of ASNases, the lack of information of the active site precludes direct comparison, as was used in the approach for the ASNase WP_078979039.1. However, the use of computational tools based on structure inspection and on the evaluation of affinity maps has proven highly effective in identifying probable binding sites in uncharacterized proteins (Harris, Olson & Goodsell, 2008). Based on the use of computational tools, it was possible to identify two putative binding sites in SsAII-2 (WP_078980718.1). Interestingly, in both sites the motifs NCSGKHxAM, PRSx(2)KPxQ, and SHTGQx(2)H were identified, and these motifs are highly conserved in the PF06089.11 family (Moreno-Enriquez et al., 2012). Of these., Borek et al. (2001)

proposed that some of the residues of the NCSGKHxAM motif could be involved in the hydrolytic deamidation of L-asparagine (Borek & Jaskólski, 2001).

On the other hand, the residues we found conserved in this family of asparaginases resemble those of the active site of the Ntn amidotransferases, in which, among the important residues for glutamine deamidation are found Cys, Asn, and Gly, and the deamidation mechanism proceeds with an oxyanion formation with the substrate. Although this mechanism is described for glutamine amidohydrolases, it may be a mechanism similar to that of this family of asparaginases, whose active site is different from those of the PF00710.11 family (Isupov et al., 1996). In that sense, the *E. coli* GLMS protein (1xfg) several catalytic residues have been identified, among which Cys1 is the catalytic nucleophile, and the nucleophilic character of its thiol group appears to be increased through general base activation by its own alpha-amino group. The authors propose that when a nitrogen acceptor is present Cys1 is kept in the active conformation, explaining the phenomenon of substrate-induced activation of the enzyme and that Arg26 is central in this coupling (Isupov et al., 1996). To determine if the catalytic residues reported in GLMS overlap with the WP_078980718 protein, we performed a structural alignment using a flexible structure alignment approach, POSA (Partial order structure alignment (http://posa.godziklab.org/) (Li et al., 2014) between the structure 1xfg (Glutaminase domain of glucosamine 6-phosphate synthase of *E. coli*) and our three-dimensional model from WP_078980718. The alignment showed an overlap with an RMSD of 6.55 Å with 121 equivalent positions and a p-value of 3.46e-01, which suggests that the alignment can be considered non-significant at global scale (random structural similarity), therefore the two proteins are not homologous (as we expected). However, at local overlapping, the residues in this region strongly suggest that the catalytic mechanism could be similar (Figure 9). The R26 residue of 1xfga which participates in substrate coupling, overlaps

453 with the site identified in WP_07898071 (NCSGKHxAM), and the other residues are equivalent
454 in both proteins, reinforcing the notion that at the NCSGKHxAM site, the substrate can be coupled,
455 and thus be the catalytic site.
456
457 **Figure 9**. Structural alignment of *S. scabrisporus* asparaginase II (WP_07898071) and *E. coli*
458 GLMS protein sequences. The residues involved in catalytic activity are shown in bold. Capital
459 letters indicate equivalent residues according to POSA program.
460
461 Although site A showed higher affinity for L-asparagine binding, additional studies are needed to
462 confirm the best site for ligand binding. Additionally, molecular dynamics simulations can provide
463 more evidence of the characteristics of the binding site and, together with *in vitro* studies, will be
464 useful for understanding the mechanism of enzymatic reaction (Karplus & Kuriyan, 2005).
465 Although our results predicted that SsAII-2 has a lower affinity than EcAII, its different folding
466 and immunogenic characteristics place it as a good candidate. Identifying catalytic site residues
467 will allow us to perform site-directed modifications to increase its affinity.
468 The strategy developed here can be applied to the search for asparaginases in other clades of
469 microorganisms, and even in eukaryotes, specifically mammalian asparaginases, whose
470 evolutionary proximity to humans predicts less immunogenicity.
471

472 ## Conclusions

473 In summary, the search for ASNases in phylogenetically distant microorganisms and the
474 application of bioinformatic tools to assess their toxicity and affinity for L-asparagine are viable
475 approaches to obtain new ASNases with therapeutic potential. Based on its low immunogenicity

and excellent enzymatic activity predicted, we have identified the *S. scabrisporus* ASNase as a potential alternative for the treatment of ALL. The subsequent enzymatic and immunogenic characterization of the *S. scabrisporus* ASNase is necessary for the validation of this bioinformatic approach.

Acknowledgements

We acknowledge Katrin Quester and Itandehui Betanzo for technical assistance.

References

Abribat T. 2016. Pegylated L-asparaginase. 1:1–5.

Ali U, Naveed M, Ullah A, Ali K, Shah SA, Fahad S, Mumtaz AS. 2016. L-asparaginase as a critical component to combat Acute Lymphoblastic Leukaemia (ALL): A novel approach to target ALL. *European Journal of Pharmacology* 771:199–210. DOI: 10.1016/j.ejphar.2015.12.023.

Avramis VI. 2012. Asparaginases: biochemical pharmacology and modes of drug resistance. *Anticancer research* 32:2423–37.

Barba P, Dapena JL, Montesinos P, Rives S. 2017. Asparaginasas en el tratamiento de la leucemia linfoblástica aguda. *Medicina Clínica* 148:225–231. DOI: 10.1016/j.medcli.2016.12.006.

Bassan R, Maino E, Cortelazzo S. 2016. Lymphoblastic lymphoma: an updated review on biology, diagnosis, and treatment. *European Journal of Haematology* 96:447–460. DOI: 10.1111/ejh.12722.

500 Battistel AP, Rocha BS da, Santos MT dos, Daudt LE, Michalowski MB. 2020. Allergic
501 reactions to asparaginase: Retrospective cohort study in pediatric patients with acute
502 lymphoid leukemia. *Hematology, Transfusion and Cell Therapy*. DOI:
503 10.1016/j.htct.2019.10.007.
504 Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne
505 PE. 2000. The Protein Data Bank. *Nucleic Acids Research* 28:235–242. DOI:
506 10.1093/nar/28.1.235.
507 Borek D, Jaskólski M. 2001. Sequence analysis of enzymes with asparaginase activity. *Acta*
508 *Biochimica Polonica* 48:893–902.
509 Cantor JR, Yoo TH, Dixit A, Iverson BL, Forsthuber TG, Georgiou G. 2011. Therapeutic
510 enzyme deimmunization by combinatorial T-cell epitope removal using neutral drift.
511 *Proceedings of the National Academy of Sciences* 108:1272–1277. DOI:
512 10.1073/pnas.1014739108.
513 Chan WK, Lorenzi PL, Anishkin A, Purwaha P, Rogers DM, Sukharev S, Rempe SB, Weinstein
514 JN. 2014. The glutaminase activity of L- Asparaginase is not required for anticancer activity
515 against ASNS-negative cells. *Blood* 123:3596–3606. DOI: 10.1182/blood-2013-10-535112.
516 Correia da Silva R, Santos Siqueira A, Ranieri Jerônimo Lima A, de Melo Lima A, Silva Santos
517 A, Cristina Figueira Aguiar D, Costa Gonçalves E. 2018. In silico characterization of a
518 cyanobacterial plant-type isoaspartyl aminopeptidase/asparaginase. DOI: 10.1007/s00894-
519 018-3635-6.
520 El-naggar NE, Deraz SF, El-ewasy SM, Suddek GM. 2018. Purification , characterization and
521 immunogenicity assessment of glutaminase free L-asparaginase from Streptomyces
522 brollosae NEAE-115. :1–15.

El-Naggar NE-A, Deraz SF, Soliman HM, El-Deeb NM, El-Ewasy SM. 2016. Purification, characterization, cytotoxicity and anticancer activities of L-asparaginase, anti-colon cancer protein, from the newly isolated alkaliphilic Streptomyces fradiae NEAE-82. *Scientific Reports* 6:32926. DOI: 10.1038/srep32926.

Emadi A, Zokaee H, Sausville EA. 2014. Asparaginase in the treatment of non-ALL hematologic malignancies. *Cancer Chemotherapy and Pharmacology* 73:875–883. DOI: 10.1007/s00280-014-2402-3.

Fernandez CA, Stewart E, Panetta JC, Wilkinson MR, Morrison AR, Finkelman FD, Sandlund JT, Pui CH, Jeha S, Relling M V., Campbell PK. 2014. Successful challenges using native E. coli asparaginase after hypersensitivity reactions to PEGylated E. coli asparaginase. *Cancer Chemotherapy and Pharmacology* 73:1307–1313. DOI: 10.1007/s00280-014-2464-2.

Fernández M, Zúñiga M. 2006. Amino acid catabolic pathways of lactic acid bacteria. *Critical Reviews in Microbiology* 32:155–183. DOI: 10.1080/10408410600880643.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* 39:W29–W37. DOI: 10.1093/nar/gkr367.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* 44:D279–D285. DOI: 10.1093/nar/gkv1344.

Galindo-Rodríguez G, Jaime-Pérez JC, Salinas-Carmona MC, González-Díaz SN, Castro-Corona Á, Cavazos-González R, Treviño-Villarreal H, Heredia-Salazar AC, Gómez-Almaguer D. 2017. Do immunoglobulin G and immunoglobulin E anti- l -asparaginase antibodies have

546 distinct implications in children with acute lymphoblastic leukemia? A cross-sectional
547 study. *Revista Brasileira de Hematologia e Hemoterapia* 39:202–209. DOI:
548 10.1016/j.bjhh.2016.11.006.
549 Gasteiger J, Marsili M. 1978. A new model for calculating atomic charges in molecules.
550 *Tetrahedron Letters* 19:3181–3184. DOI: 10.1016/S0040-4039(01)94977-9.
551 Gesto DS, Cerqueira NMFSA, Fernandes PA, Ramos MJ. 2013. Unraveling the enigmatic
552 mechanism of L-asparaginase II with QM/QM calculations. *Journal of the American*
553 *Chemical Society* 135:7146–58. DOI: 10.1021/ja310165u.
554 De Groot AS, Martin W. 2009. Reducing risk, improving outcomes: Bioengineering less
555 immunogenic protein therapeutics. *Clinical Immunology* 131:189–201. DOI:
556 10.1016/j.clim.2009.01.009.
557 Harris R, Olson AJ, Goodsell DS. 2008. Automated prediction of ligand-binding sites in
558 proteins. *Proteins: Structure, Function and Genetics* 70:1506–1517. DOI:
559 10.1002/prot.21645.
560 Hijiya N, van der Sluis IM. 2015. *Asparaginase-associated toxicity in children with acute*
561 *lymphoblastic leukemia*. DOI: 10.3109/10428194.2015.1101098.
562 Hijiya N, van der Sluis IM. 2016. Asparaginase-associated toxicity in children with acute
563 lymphoblastic leukemia. *Leukemia & lymphoma* 57:748–57. DOI:
564 10.3109/10428194.2015.1101098.
565 Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: A web server for clustering and
566 comparing biological sequences. *Bioinformatics* 26:680–682. DOI:
567 10.1093/bioinformatics/btq003.
568 Huerta-Saquero A, Evangelista-Martínez Z, Angélica ME, Perez-Rueda E. 2013. Rhizobium etli

asparaginase II: an alternative for acute lymphoblastic leukemia (ALL) treatment. *Bioengineered* 4:1–7. DOI: 10.1128/JB.00714-13.

Isupov MN, Obmolova G, Butterworth S, Badet-Denisot M-A, Badet B, Polikarpov I, Littlechild JA, Teplyakov A. 1996. Substrate binding is required for assembly of the active conformation of the catalytic site in Ntn amidotransferases: evidence from the 1.8 å crystal structure of the glutaminase domain of glucosamine 6-phosphate synthase. *Structure* 4:801–810. DOI: 10.1016/S0969-2126(96)00087-1.

Kamal N, Koh C, Samala N, Fontana RJ, Stolz A, Durazo F, Hayashi PH, Phillips E, Wang T, Hoofnagle JH, Drug-Induced Liver Injury Network. 2019. Asparaginase-induced hepatotoxicity: rapid development of cholestasis and hepatic steatosis. *Hepatology international* 13:641–648. DOI: 10.1007/s12072-019-09971-2.

Karplus M, Kuriyan J. 2005. Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences of the United States of America* 102:6679–85. DOI: 10.1073/pnas.0408930102.

Katz AJ, Chia VM, Schoonen WM, Kelsh MA. 2015. Acute lymphoblastic leukemia: an assessment of international incidence, survival, and disease burden. *Cancer Causes & Control* 26:1627–1642. DOI: 10.1007/s10552-015-0657-6.

Krishnapura PR, Belur PD, Subramanya S. 2016. A critical review on properties and applications of microbial l-asparaginases. *Critical reviews in microbiology* 42:720–37. DOI: 10.3109/1040841X.2015.1022505.

Lanvers-Kaminsky C. 2017. Asparaginase pharmacology: challenges still to be faced. *Cancer Chemotherapy and Pharmacology* 0:1–12. DOI: 10.1007/s00280-016-3236-y.

Li Z, Natarajan P, Ye Y, Hrabe T, Godzik A. 2014. POSA: a user-driven, interactive multiple

592 protein structure alignment server. *Nucleic acids research* 42:W240-5. DOI:
593 10.1093/nar/gku394.
594 Lopes AM, Oliveira-Nascimento L de, Ribeiro A, Tairum CA, Breyer CA, Oliveira MA de,
595 Monteiro G, Souza-Motta CM de, Magalhães P de O, Avendaño JGF, Cavaco-Paulo AM,
596 Mazzola PG, Rangel-Yagui C de O, Sette LD, Converti A, Pessoa A. 2015. Therapeutic L-
597 asparaginase: upstream, downstream and beyond. *Critical Reviews in Biotechnology*
598 8551:1–18. DOI: 10.3109/07388551.2015.1120705.
599 Magnan CN, Zeller M, Kayala MA, Vigil A, Randall A, Felgner PL, Baldi P. 2010. High-
600 throughput prediction of protein antigenicity using protein microarray data. *Bioinformatics*
601 26:2936–2943. DOI: 10.1093/bioinformatics/btq551.
602 Moreno-Enriquez A, Evangelista-Martinez Z, Gonzalez-Mondragon EG, Calderon-Flores A,
603 Arreguin R, Perez-Rueda E, Huerta-Saquero A. 2012. Biochemical characterization of
604 recombinant L-asparaginase (AnsA) from Rhizobium etli, a member of an increasing
605 rhizobial-type family of L-asparaginases. *Journal of microbiology and biotechnology*
606 22:292–300.
607 Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. 2009.
608 AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.
609 *Journal of Computational Chemistry* 30:2785–2791. DOI: 10.1002/jcc.21256.
610 Nguyen HA, Su Y, Lavie A. 2016. Design and Characterization of Erwinia Chrysanthemi l-
611 Asparaginase Variants with Diminished l-Glutaminase Activity. *The Journal of biological*
612 *chemistry* 291:17664–76. DOI: 10.1074/jbc.M116.728485.
613 Nguyen HA, Su Y, Zhang JY, Antanasijevic A, Caffrey M, Schalk AM, Liu L, Rondelli D, Oh
614 A, Mahmud DL, Bosland MC, Kajdacsy-Balla A, Peirs S, Lammens T, Mondelaers V, De

615 Moerloose B, Goossens S, Schlicht MJ, Kabirov KK, Lyubimov A V, Merrill BJ,
616 Saunthararajah Y, Van Vlierberghe P, Lavie A. 2018. A Novel l-Asparaginase with low l-
617 Glutaminase Coactivity Is Highly Efficacious against Both T- and B-cell Acute
618 Lymphoblastic Leukemias In Vivo. *Cancer research* 78:1549–1560. DOI: 10.1158/0008-
619 5472.CAN-17-2106.
620 Onciu M. 2009. Acute Lymphoblastic Leukemia. *Hematology/Oncology Clinics of North*
621 *America* 23:655–674. DOI: 10.1016/j.hoc.2009.04.009.
622 Ortuño-Olea L, Durán-Vargas S. 2000. The L-asparagine operon of Rhizobium etli contains a
623 gene encoding an atypical asparaginase. *FEMS Microbiology Letters* 189:177–182. DOI:
624 10.1016/S0378-1097(00)00275-5.
625 Pieters R, Hunger SP, Boos J, Rizzari C, Silverman L, Baruchel A, Goekbuget N, Schrappe M,
626 Pui CH. 2011. L-asparaginase treatment in acute lymphoblastic leukemia. *Cancer* 117:238–
627 249. DOI: 10.1002/cncr.25489.
628 Pui C-H, Relling M V., Downing JR. 2004. Acute Lymphoblastic Leukemia. *New England*
629 *Journal of Medicine* 350:1535–1548. DOI: 10.1056/NEJMra023001.
630 Ramya LN, Doble M, Rekha VPB, Pulicherla KK. 2012. L-asparaginase as potent anti-leukemic
631 agent and its significance of having reduced glutaminase side activity for better treatment of
632 acute lymphoblastic leukaemia. *Applied Biochemistry and Biotechnology* 167:2144–2159.
633 DOI: 10.1007/s12010-012-9755-z.
634 Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein
635 structure and function prediction. *Nature Protocols* 5:725–738. DOI: 10.1038/nprot.2010.5.
636 Salzer W, Bostrom B, Messinger Y, Perissinotti AJ, Marini B. 2017. Asparaginase activity levels
637 and monitoring in patients with acute lymphoblastic leukemia. *Leukemia and Lymphoma*

0:1–10. DOI: 10.1080/10428194.2017.1386305.

Sanches M, Kraunchenko S, Polikarpov I. 2016. Structure , Substrate Complexation and Reaction Mechanism of Bacterial Asparaginases. *Current chemical biology* 1:75–86. DOI: 10.2174/187231307779814057.

Sanner MF. 1999. Python: a programming language for software integration and development. *Journal of molecular graphics & modelling* 17:57–61.

Santos AR, Pereira VB, Barbosa E, Baumbach J, Pauling J, Röttger R, Turk MZ, Silva A, Miyoshi A, Azevedo V. 2013. Mature Epitope Density - A strategy for target selection based on immunoinformatics and exported prokaryotic proteins. *BMC Genomics* 14:S4. DOI: 10.1186/1471-2164-14-S6-S4.

Schein PS, Rakieten N, Gordon BM, Davis RD, Rall DP. 1969. The toxicity of Escherichia coli L-asparaginase. *Cancer research* 29:426–34.

Schwab C, Harrison CJ. 2018. Advances in B-cell Precursor Acute Lymphoblastic Leukemia Genomics. *HemaSphere* 2:1. DOI: 10.1097/HS9.0000000000000053.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* 7:539. DOI: 10.1038/msb.2011.75.

Singh SK, Cousens LP, Alvarez D, Mahajan PB. 2012. Determinants of immunogenic response to protein therapeutics. *Biologicals* 40:364–368. DOI: 10.1016/j.biologicals.2012.06.001.

Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, Braxenthaler M, Gallazzi F, Protti MP, Sinigaglia F, Hammer J. 1999. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nature*

Biotechnology 17:555–561.

Swain AL, Jaskólski M, Housset D, Rao JK, Wlodawer A. 1993. Crystal structure of Escherichia coli L-asparaginase, an enzyme used in cancer therapy. *Proceedings of the National Academy of Sciences of the United States of America* 90:1474–8. DOI: 10.1073/pnas.90.4.1474.

Thomas X, Le Jeune C. 2016. Erythrocyte encapsulated l -asparaginase (GRASPA) in acute leukemia. *International Journal of Hematologic Oncology* 5:11–25. DOI: 10.2217/ijh-2016-0002.

Walenciak J, Wyka K, Janczar S, Młynarski W, Zalewska-Szewczyk B. 2019. Dynamic changes in specific anti-L-asparaginase antibodies generation during acute lymphoblastic leukemia treatment. *Pharmacological Reports* 71:311–318. DOI: 10.1016/j.pharep.2018.11.002.

Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B. 2010. Peptide binding predictions for HLA DR , DP and DQ molecules. *BMC Bioinformatics* 11:568. DOI: 10.1186/1471-2105-11-568.

Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research* 46:D1074–D1082. DOI: 10.1093/nar/gkx1037.

Xu D, Zhang Y. 2011. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical journal* 101:2525–34. DOI: 10.1016/j.bpj.2011.10.024.

Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics* 9:40.

684 DOI: 10.1186/1471-2105-9-40.

685

Table 1(on next page)

Table 1. Representative *Streptomyces* ASNases of the PF00710.11 and PF06089.11 families, at 60% identity cutoff.

Table 1. Representative *Streptomyces* ASNases of the PF00710.11 and PF06089.11 families, at 60% identity cutoff.

ASNase ID	Organism	Length (amino acids)	Family
PF00710.11 family			
WP_053609500.1	*S.purpurogeneiscleroticus*	373	PF00710.11
WP_053610569.1	*S. purpurogeneiscleroticus*	338	PF00710.11
WP_055617501.1	*S. phaeochromogenes*	380	PF00710.11
WP_051815467.1	*S. lavenduligriseus*	363	PF00710.11
WP_078649241.1	*S. fradiae*	350	PF00710.11
EFL23513.1	*S. himastatinicu ATCC 53653*	351	PF00710.11
WP_014151616.1	*S. cattleya*	331	PF00710.11
WP_095730579.1	*S. albidoflavus*	333	PF00710.11
WP_078965752.1	*S. aureocirculatus*	343	PF00710.11
WP_078513220.1	*S. purpureus*	421	PF00710.11
WP_009718687.1	*S. himastatinicus*	347	PF00710.11
WP_079189481.1	*S. paucisporeus*	384	PF00710.11
WP_052425051.1	*S. fulvoviolaceus*	340	PF00710.11
ELP65653.1	*S. turgidiscabies Car8*	358	PF00710.11
WP_070201703.1	*S. nanshensis*	347	PF00710.11
KWW98572.1	*S. thermoautotrophicus*	333	PF00710.11
WP_073950513.1	*S. kebangsaanensis*	333	PF00710.11
WP_030748190.1	*S. griseus*	329	PF00710.11
WP_059134811.1	*S. alboniger*	332	PF00710.11
PF06089.11 family			
ARZ68596.1	*S. albireticuli*	428	PF06089.11
CDR15801.1	*S. iranensis*	387	PF06089.11
SOD64826.1	*S. zhaozhouensis*	316	PF06089.11
WP_020554088	*S. scabrisporus*	332	PF06089.11
WP_044373749	*S. ahygroscopicus*	330	PF06089.11
WP_078645645	*S. varsoviensis*	348	PF06089.11
WP_078980718.1	*S. scabrisporus*	327	PF06089.11

Table 2(on next page)

Table 2. High-affinity epitope prediction. Epitope number, CPR value, allele coverage, and ED of ASNases.

Table 2. High-affinity epitope prediction. Epitope number, CPR value, allele coverage, and ED of ASNases.

ASNase ID	Epitope number	CPR value	Allele number	ED
P00805_EcAII	10	0.6383	5	0.0114
WP_053609500.1	12	0.5174	5	0.0171
WP_053610569.1	14	0.5381	8	0.0196
WP_055617501.1	7	0.4532	7	0.0112
WP_051815467.1	6	0.6673	5	0.0060
WP_078649241.1	7	0.4554	6	0.0111
EFL23513.1	10	0.6054	8	0.0115
WP_014151616.1	3	0.4024	5	0.0056
WP_095730579.1	8	0.5346	2	0.0115
WP_078965752.1	3	0.4987	4	0.0045
WP_078513220.1	9	0.4551	6	0.0119
WP_009718687.1	5	0.6480	4	0.0052
WP_079189481.1	4	0.5217	5	0.0051
WP_052425051.1	10	0.4369	6	0.0170
ELP65653.1	5	0.6717	3	0.0047
WP_070201703.1	7	0.6637	6	0.0069
KWW98572.1	4	0.7254	4	0.0034
WP_073950513.1	6	0.7424	4	0.0048
WP_030748190.1	3	0.5125	6	0.0046
Q9K4F5_ScAII	3	0.4167	3	0.0053
ARZ68596.1	5	0.6283	5	0.0044
SOD64826.1	5	0.5046	5	0.0080
WP_078645645.1	7	0.6404	5	0.0074
CDR15801.1	5	0.5510	5	0.0059
WP_078980718.1	6	0.7003	6	0.0056
WP_044373749.1	3	0.7114	4	0.0027

Epitope number refers to the number of epitopes with CPR < 1. Allele number is the number of allele coverage for high affinity epitopes (with a CPR < 1).

Table 3(on next page)

Table 3. SsAII-2 putative binding site residues. PF06089.11 family conserved residues are shown in bold.

Table 3. SsAII-2 putative binding site residues. PF06089.11 family conserved residues are shown in bold.

Site	Ligand-binding positions predicted by AutoLigand	Ligand-binding positions predicted by blind molecular docking
A	Arg 58, Ser 59, Lys 62, **Asn 141, Ser 143, Gly 144, Lys 145, His 146, Ala 147**, Gly 236, Gly 237, Asp 238, Gly 239, Lys 255, Gly 256, Gly 257, Ala 258, Pro 281, Leu 326	Asn 135, Thr 136, Arg 137, Arg 139**, Asn 141, Gly 144, His 146,** Asp 192
B	Ala 85, Gly 86**, Ser 87, His 88, Thr 89, Gly 90, Gln 91, His 94,** Leu 164, Asp 165, Pro 166, Gly 167, His 168, Leu 173, Glu 177, Gly 178, Asp 180	-

Table 4(on next page)

Table 4. Molecular docking energies of ASNases.

Table 4. Molecular docking energies of ASNases.

Table 4. Molecular docking energies of ASNases.

ASNase	Free energy binding (kcal/mol)	Inter-molecular energy (kcal/mol)	van der Waals – hydrogen bonds (kcal/mol)	Electrostatic energy (kcal/mol)	Hydrogen bonds
E. coli EcAII; tetramer	-9.81	-11.30	-5.88	-3.61	9
E. coli EcAII; monomer	-8.46	-9.95	-7.02	-2.35	6
S. scabrisporus WP_078980718.1 – site A	-6.67	-8.17	-5.25	-2.91	6
S. scabrisporus WP_078980718.1 – site B	-4.62	-6.11	-4.39	-1.72	2

Figure 1

Phylogenetic tree of PF00710.11 (A) and PF06089.11 (B) families.

Blue dots highlight reference sequences added to each analysis. Red dots highlight sequences used as internal controls (asparaginases from *E. coli* and *R. etli*, respectively). A total of 1000 replicates were performed. Bootstrap values are indicated.

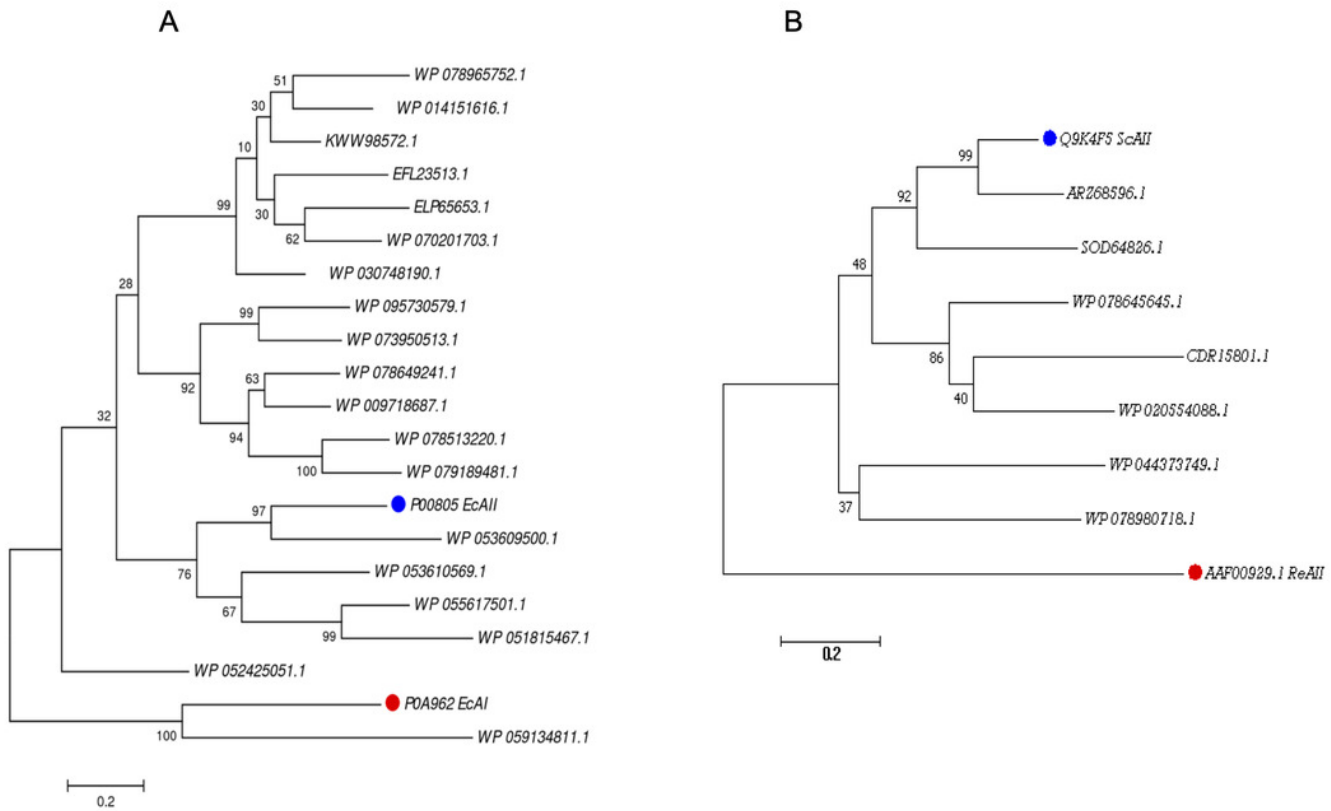
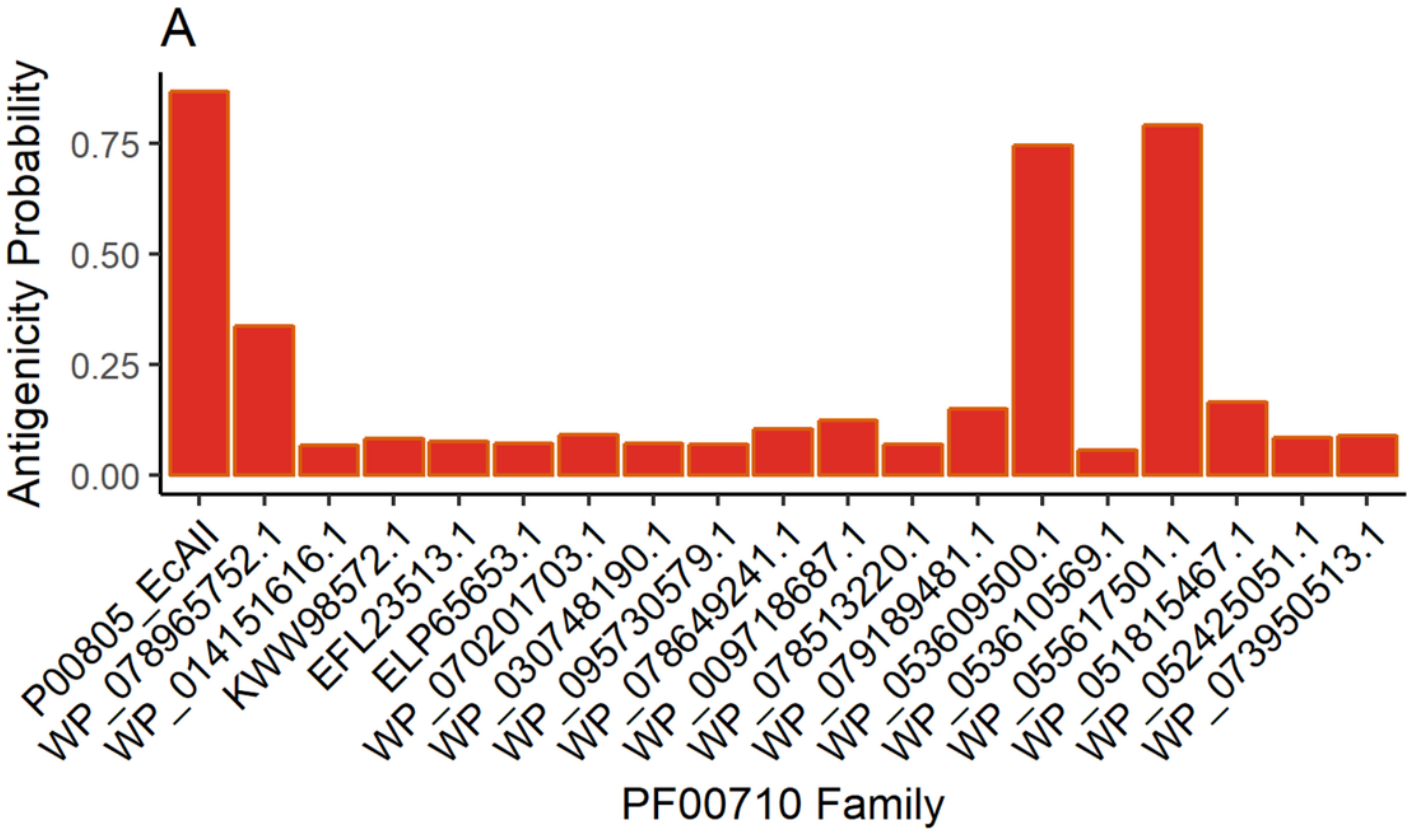


Figure 2

ASNase antigenicity predictions.

The antigenicity scores for PF00710.11 family (A) and PF06089.11 family (B) of *Streptomyces* asparaginases were compared with the EcAII antigenicity score.


A
Antigenicity Probability
0.00
0.25
0.50
0.75
P00805_EcAII
WP_078965752.1
WP_014151616.1
KWW98572.1
EFL23513.1
ELP65653.1
WP_070201703.1
WP_030748190.1
WP_095730579.1
WP_078649241.1
WP_009718687.1
WP_078513220.1
WP_079189481.1
WP_053609500.1
WP_053610569.1
WP_055617501.1
WP_051815467.1
WP_052425051.1
WP_073950513.1
PF00710 Family

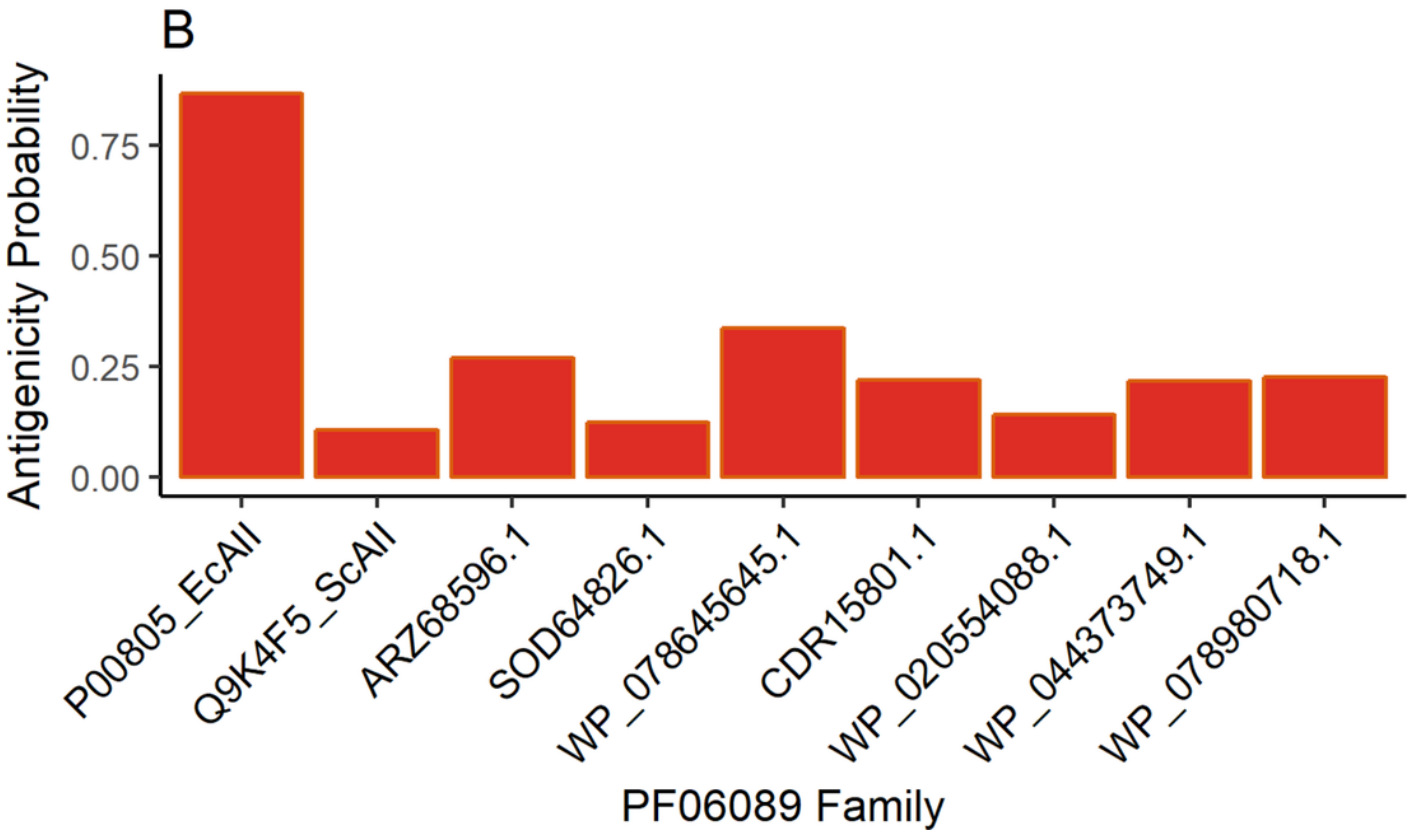

B
Antigenicity Probability
0.00
0.25
0.50
0.75
P00805_EcAII
Q9K4F5_ScAII
ARZ68596.1
SOD64826.1
WP_078645645.1
CDR15801.1
WP_020554088.1
WP_044373749.1
WP_078980718.1
PF06089 Family

Figure 3

Epitope mapping of ASNases of the PF familes evaluated, (A) PF00710.11 and (B) PF06089.11.

The epitopes identified along with the ASNase sequences are shown. The color intensity represents the number of hits for each of them.

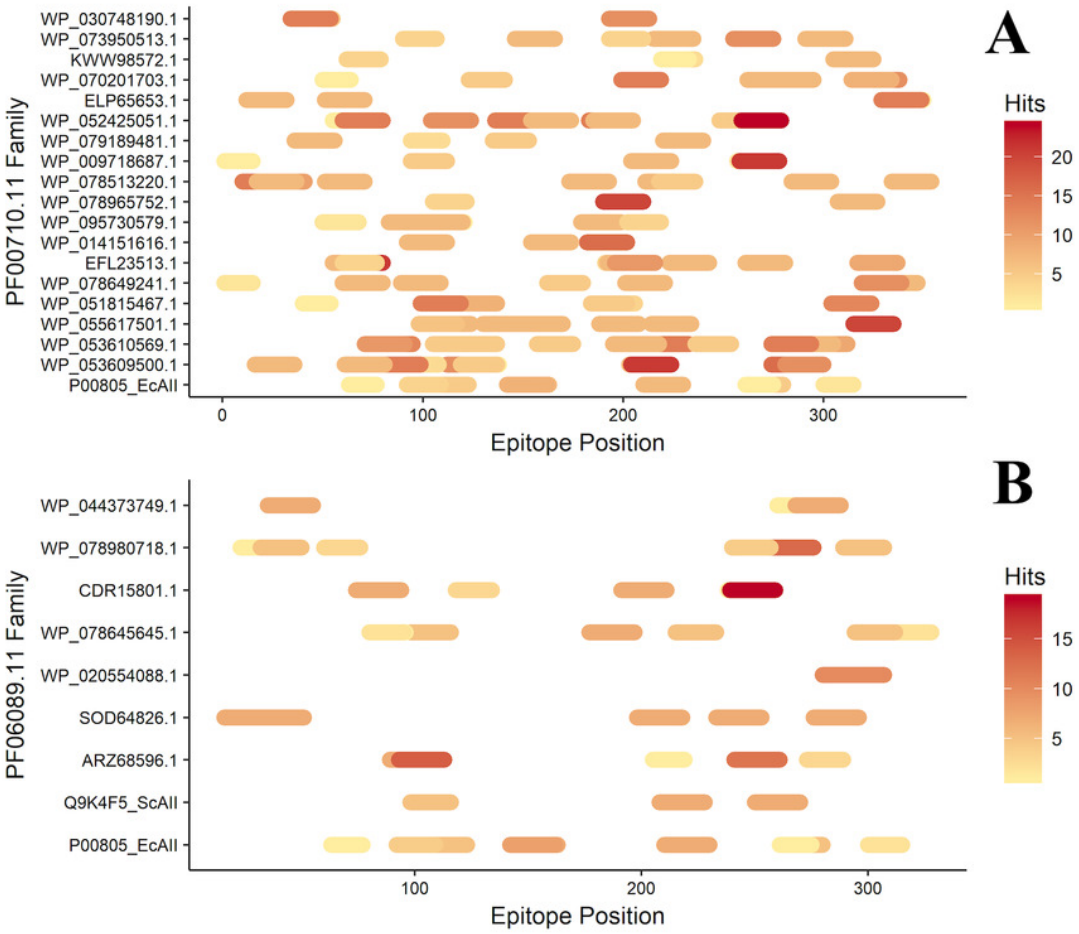

A
PF00710.11 Family
WP_030748190.1
WP_073950513.1
KWW98572.1
WP_070201703.1
ELP65653.1
WP_052425051.1
WP_079189481.1
WP_009718687.1
WP_078513220.1
WP_078965752.1
WP_095730579.1
WP_014151616.1
EFL23513.1
WP_078649241.1
WP_051815467.1
WP_055617501.1
WP_053610569.1
WP_053609500.1
P00805_EcAII
0
100
200
300
Epitope Position
Hits
20
15
10
5
B
PF06089.11 Family
WP_044373749.1
WP_078980718.1
CDR15801.1
WP_078645645.1
WP_020554088.1
SOD64826.1
ARZ68596.1
Q9K4F5_ScAII
P00805_EcAII
100
200
300
Epitope Position
Hits
15
10
5

Figure 4

3D protein structure prediction of *S. scabrisporus* asparaginase II (WP_078980718.1; SsAII-2).

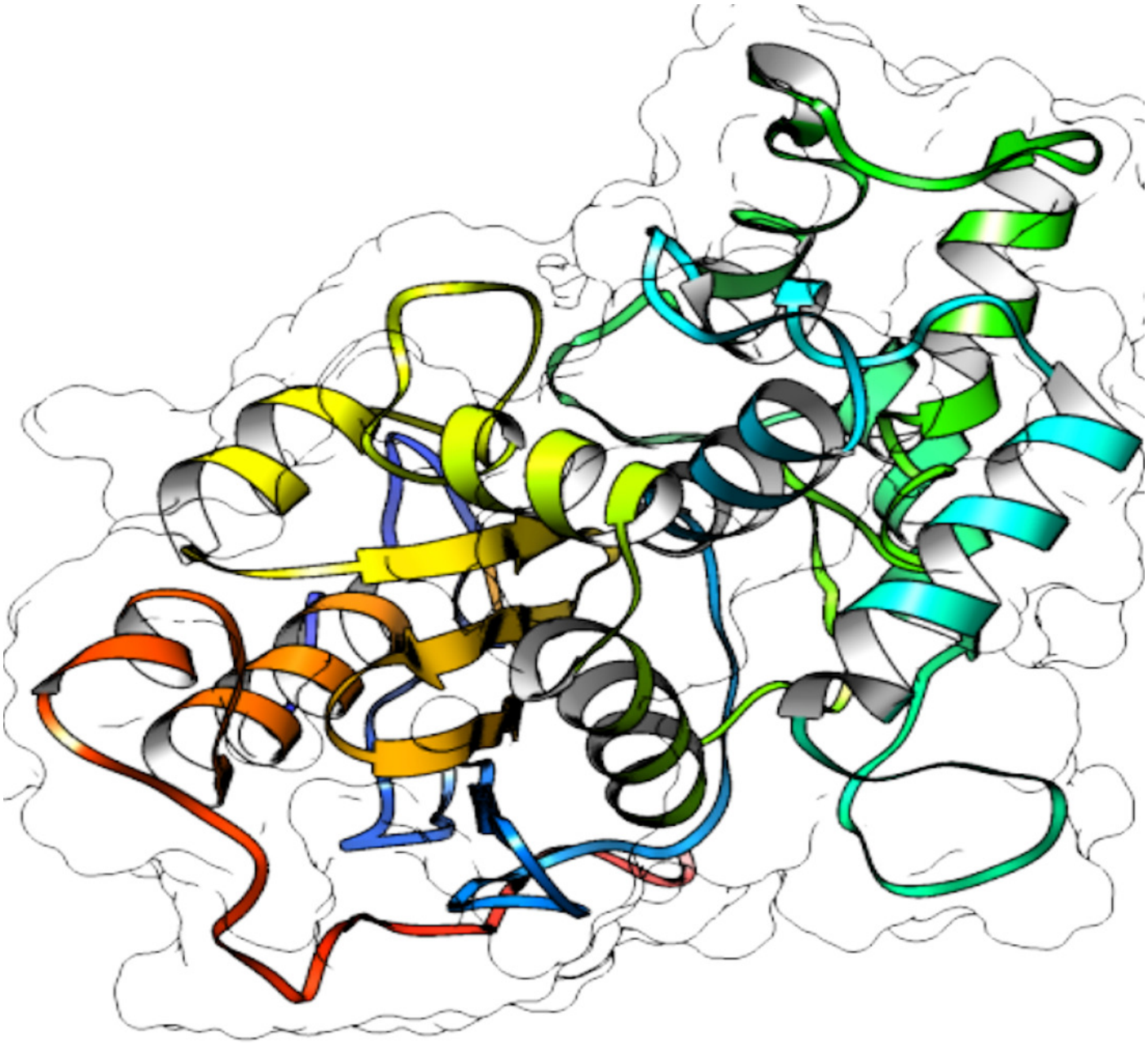


Figure 5

SsAII-2 putative binding sites.

A) Site A (orange) contains the NCSGKHxAML sequence and site B (blue) contains the SHTGQxHFV motif. B) Residues involved with asparagine through a direct interaction, obtained by blind molecular docking.

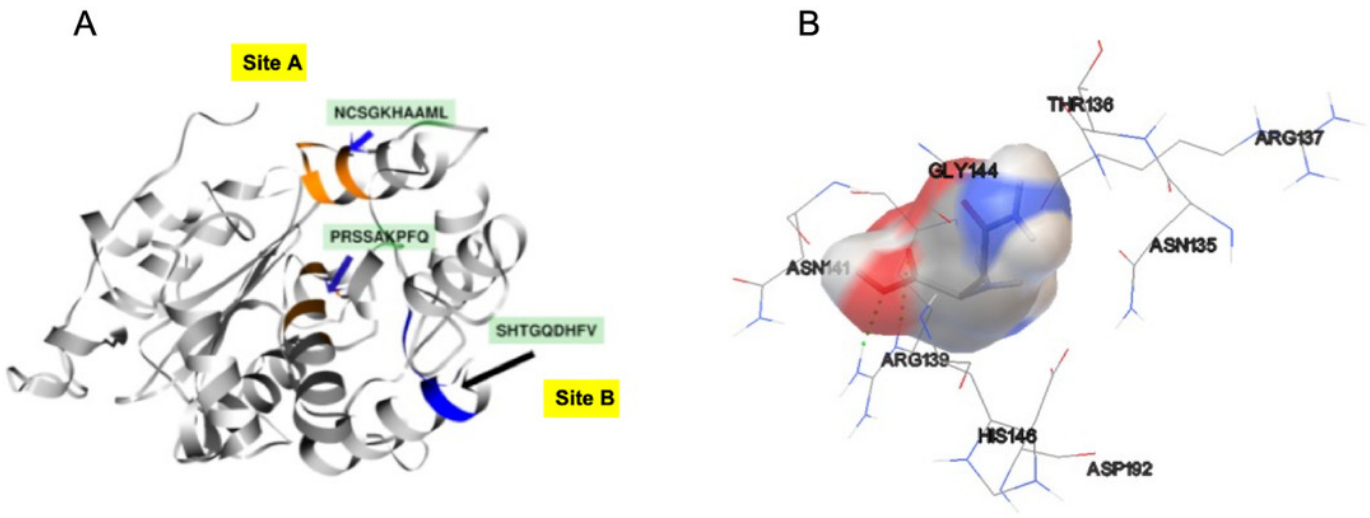


Figure 6

EcAII dimer AutoLigand analysis.

EcAII subunit A is shown in cyan and subunit C in magenta. The red mesh represents the highest-affinity pocket found by AutoLigand (putative active site). The site represented in the scheme corresponds to the residues located at a maximum distance of 5 Å using 20 points: Thr 12, Tyr 25, Ser 58, Gln 59, Thr 89, and Asp 90 from subunit C and Asn 248 and Glu 283 from subunit A.

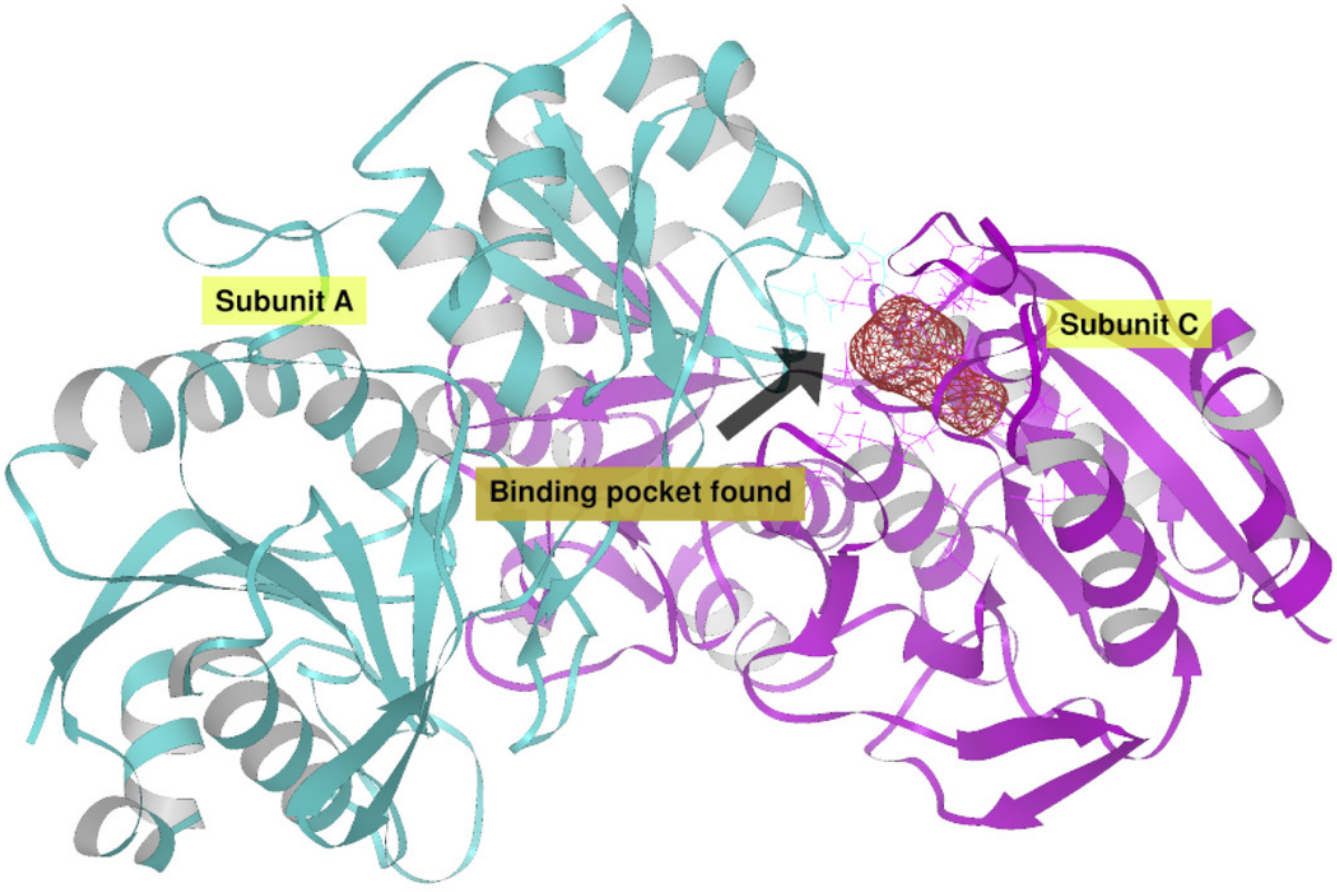


Figure 7

AutoLigand results for EcAII.

The minima observed in the total energy graphs per unit volume represent putative binding sites in the structures analyzed, for the monomer, dimer, and tetramer conformation. As more filling points are used, the binding sites, cavities, or grooves are filled and the affinity decreases. The best sites are the ones with the lowest energy and the lowest volume.

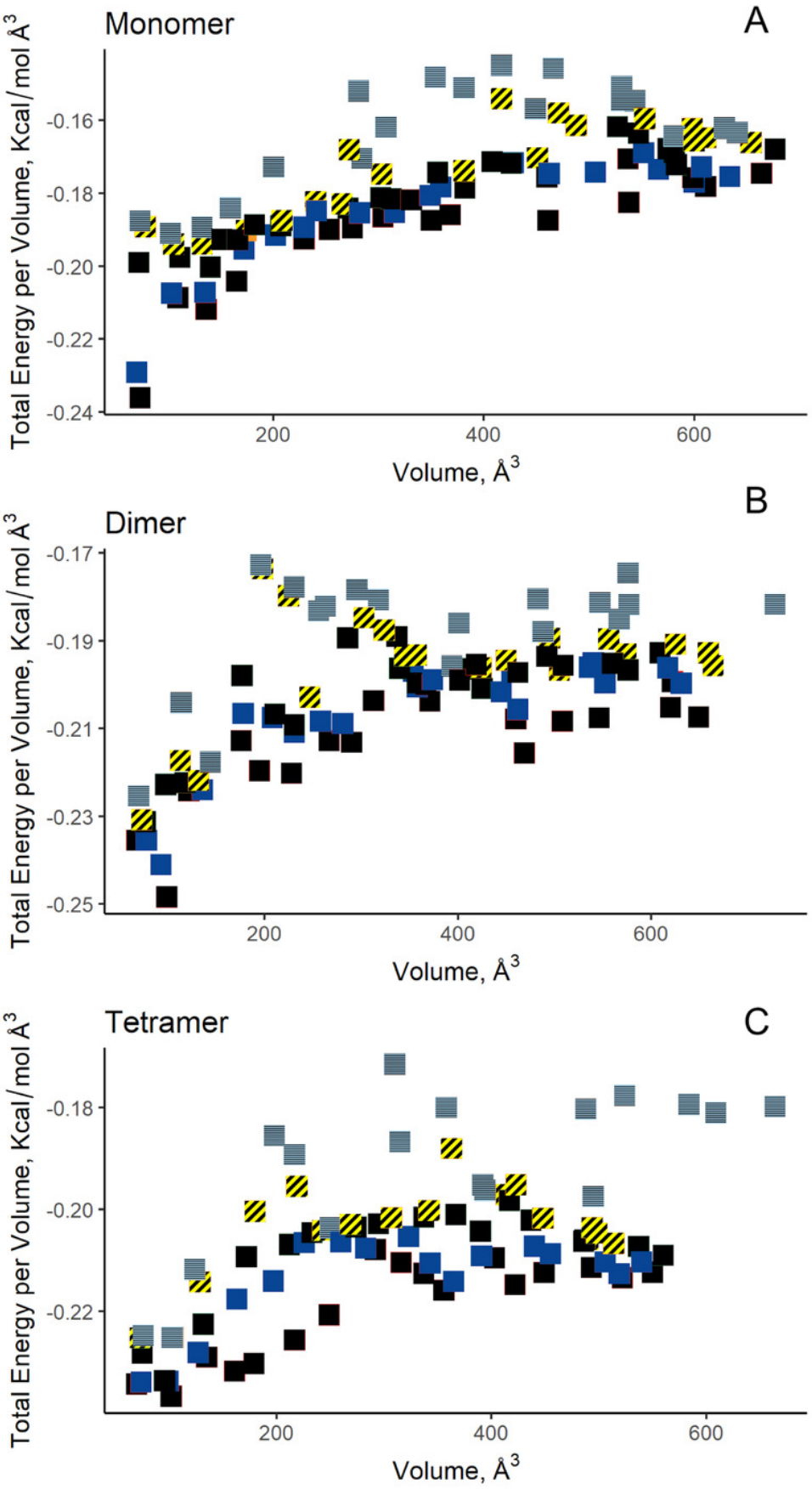

Monomer
A
Total Energy per Volume, Kcal/mol Å3
-0.16
-0.18
-0.20
-0.22
-0.24
200
400
600
Volume, Å3
Dimer
B
Total Energy per Volume, Kcal/mol Å3
-0.17
-0.19
-0.21
-0.23
-0.25
200
400
600
Volume, Å3
Tetramer
C
Total Energy per Volume, Kcal/mol Å3
-0.18
-0.20
-0.22
200
400
600
Volume, Å3

Figure 8

Interaction maps for sites A and B from *S. scabrisporus* ASNase.

The black spheres represent carbon atoms, the blue nitrogen and the red oxygen. Hydrogen bonds are represented by green dotted lines and hydrophobic interactions are shown as red half-moons.

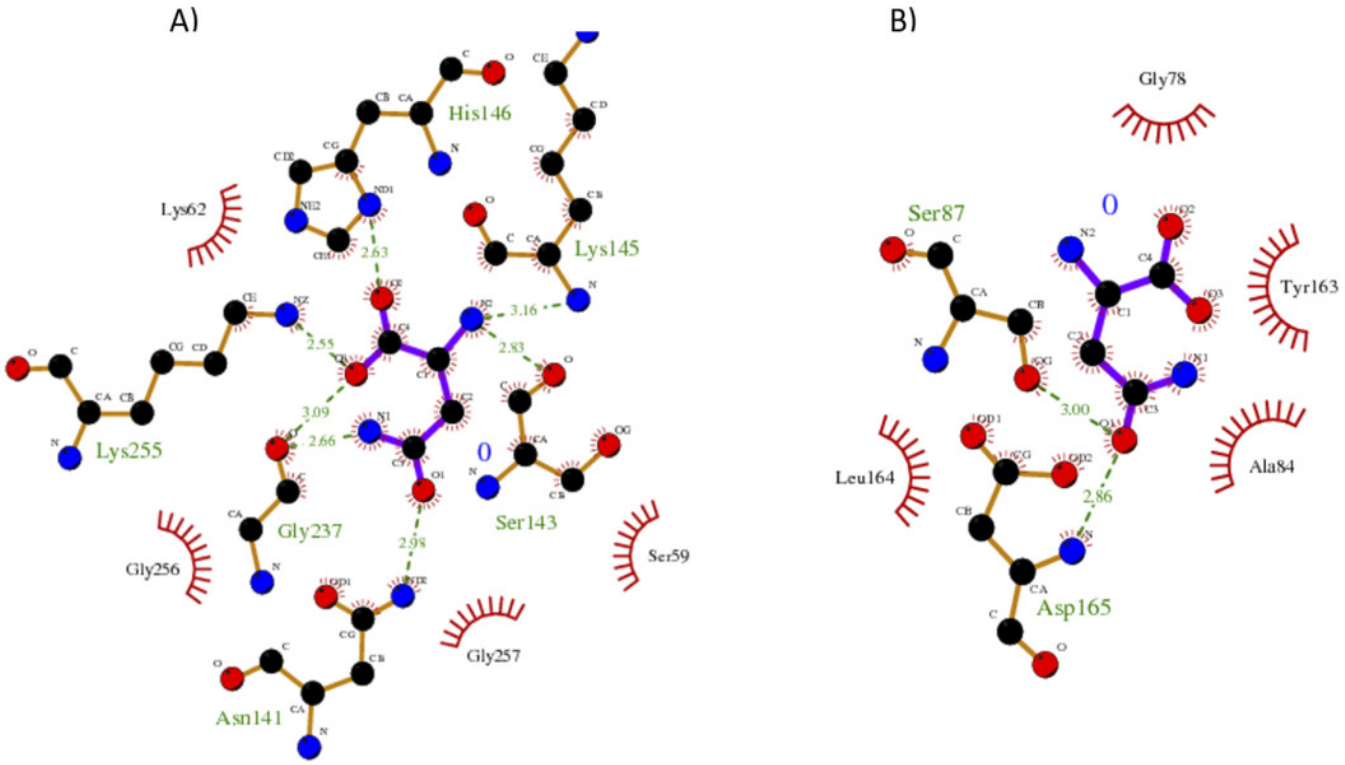


Figure 9

Figure 9. Structural alignment of *S. scabrisporus* asparaginase II (WP_07898071) and *E. coli* GLMS protein sequences.

The residues involved in catalytic activity are shown in bold. Capital letters indicate equivalent residues according to POSA program.

```
118 -----RPEDEDTYQEMIARGeenTRERMNCSGKHAAML-

  1 CgivgAIAQRDVAEILLEGL---RRLEYRGYDSAGLAVv
```