STUDIA UNIVERSITATIS BABEŞ-BOLYAI BIOLOGIA, 65, 1, 2020 (p. 5-18)
doi:10.24193/subbbiol.2020.1.01

In silico modeling and analysis of squalene synthase-like 1 (SSL-1) enzyme from green microalga *Botryococcus terribilis* AICB 872

Tiberiu Szöke-Nagy[1,2], Sebastian Alin Porav[1,2] and Nicolae Dragoş[1,3]✉

[1] Babeş-Bolyai University, Faculty of Biology and Geology, Cluj-Napoca, Romania;
[2] Department of Molecular and Biomolecular Physics, National Institute of R&D of Isotopic and Molecular Technologies, Cluj-Napoca, Romania;
[3] Institute of Biological Research, branch of the National Institute of R&D for Biological Sciences Bucharest, Cluj-Napoca, Romania;
✉ ***Corresponding author, E-mail: dragos.nicolae@ubbcluj.ro.***

Abstract. The genus *Botryococcus* contains a small number of green, colonial algae, with some taxa still uncertain. *B. braunii* is the most extensively studied species of the genus because of its hydrocarbon oils which can be used as an alternative energy source. Some *B. terribilis* AICB strains were previously described showing their ability to synthesize C30–C32 botryococcenes similar to those produced by chemical race B of *B. braunii* strains. The present study aimed to investigate the structural features of SSL-1 enzyme involved in the biosynthesis of presqualene diphosphate from *B. terribilis* AICB 872, and its functional conservation by means of computational proteomics and molecular biology techniques. Using PCR amplification we obtained a 3811bp contig containing the sequence of SSL-1 gene. The homology modeling analysis revealed the presence of alpha helical structures and a small beta sheet which are forming the SSL-1 catalytic core. Coil structures and both N and C terminus regions of the protein are characterized by highly disordered structural fragments. Finally, our data integrated within the available information in the literature allowed us to presume that the formation of presqualene diphosphate, the first step of hydrocarbon biosynthesis in *B. terribilis* strain occurs in a similar fashion with that described in *B. braunii*

Keywords: *Botryococcus terribilis,* squalene synthase-like 1, 3D structure prediction, AICB strain, biofuel.

Introduction

Botryococcus is a genus of green, colonial microalgae, consisting of 14 taxa according to AlgaeBase (Guiry and Guity, 2019), including *B. braunii* Kützing and *B. terribilis* Komárek & Marvan. *B. braunii* is the most extensively studied species of the *Botryococcus* genus because of its hydrocarbon-biosynthesis property.

Strains of *B. braunii* can synthesize large amounts of liquid hydrocarbons oils which can be used as an alternative energy source to fossil fuels (Al-Hothaly, 2018). *B. braunii* strains produce and accumulate various types of hydrocarbons, being classified in four chemical races (A, B, L and S).

Race A produces n-alkadiene and n-triene (Metzger *et al.*, 1986; Metzger and Largeau, 2005). Race B produces botryococcenes (Metzger *et al.*, 1985; Metzger and Largeau, 2005), squalenes and their methylated derivatives (Huang and Poulter, 1989; Achitouv *et al.*, 2004). Chemical race L produces C40 lycopadiene and small amounts of lycopatriene (Metzger and Casadevall, 1987; Metzger *et al.*, 1990). Recently, the lycopaoctaene production by lycopaoctaene synthase (LOS) has been reported in *B. braunii* Songkla Nakarin strain (Thapa *et al.*, 2016). Race S produces epoxy-n-alkane and saturated n-alkane (Kawachi *et al.*, 2012).

Niehaus *et al*. (2011), isolated the squalene synthase-like 1 (SSL-1) cDNA by screening the cDNA library with a radiolabeled *Botryococcus* squalene synthase probe under low stringency hybridization conditions. Further, by combining transcriptomic data from two independent projects and with deep mining computational screening another two SSL were discovered, namely SSL-2 and SSL-3. SSL-1, SSL-2 and SSL-3 enzymes are involved in the last steps of botryococcenes and squalenes biosynthesis pathway. SSL-1 catalyzes the formation of presqualene diphosphate (PSPP) intermediate by a head-head condensation of two moieties of farnesyl diphosphate (FPP). PSPP being further converted to squalenes (SSL-2) or botryococcenes (SSL-3). Squalene biosynthesis was observed under co-expression of SSL-1 and SSL-2, similar to botryococcenes biosynthesis under SSL-1 and SSL-3 co-expression.

The squalene synthase (SQS) genes were described among different taxa. For example, the overexpression of SQS in medicinal plants *Panax ginseng* (Lee *et al.*, 2004), *Eleutherococcus senticosus* (Seo *et al.*, 2005) and *Salvia miltiorrhiza* (Rong *et al.*, 2016) leads to enhanced biosynthesis of triterpenes and phytosterols. Human SQS is a microsomal enzyme involved in the first step of cholesterol biosynthesis (Park *et al.*, 2014).

The importance of studying the SSL enzymes from different strains of *Botryococcus* resides in the ability of these algal species to synthesize and secrete significant extracellular quantities of hydrocarbons similar to fossil fuels, the best-studied species of this genus from this point of view being *B. braunii* Showa (Hillen and Wake, 1979; Peterson *et al.*, 2008).

The major goal of this work was to investigate the structural and functional relationship between SSL-1 enzyme from *B. terribilis* AICB strain and *B. braunii* using molecular technique and in silico tridimensional structure prediction and analysis of amino acids.

Materials and methods

Strain cultivation

B. terribilis AICB 872 strain was isolated from Tăureni fishpond, Mureș County, Romania. This strain is deposited in the Algal and Cyanobacterial Culture Collection (AICB) at the Institute of Biological Research, Cluj-Napoca, Romania (Dragoș *et al.*, 1997). The algal culture was grown on BBM medium, under continuous irradiation (approx. 150 µmol photons $m^{-2}s^{-1}$), at 25°C±2°C, in 100 mL Erlenmeyer flask. The algal biomass was freshly harvested before every experiment, in the exponential growth phase by centrifugation at 10.000 x g for 2 minutes.

Genomic DNA (gDNA) purification

gDNA was isolated and purified from algal biomass using ZR Soil Microprobe DNA MiniPrep™ Kit (Zymo Research Corp., Irvine, CA, USA), according to the manufacturer's protocol. gDNA quantification was performed on a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

PCR amplification, isolation and sequencing of SSL-1 gene

The PCR was carried out using SSL1-F2 and SSL1-R3 primer pairs published in a previous paper (Szoke-Nagy *et al.*, 2015), as follows: each 20 µL reaction volumes containing 1U of Thermo Scientific™ DreamTaq™ DNA Polymerase in 2 µL of the manufacturer`s buffer, 0.2 mM dNTPs, 0.3 µM of each primer, and approximately 20 ng of gDNA template. The PCR was performed in a TProfessional TRIO Termocycler (Biometra, Göttingen, Germany). PCR cycling conditions were: initial denaturation at 95°C for 5 min followed by 34 cycles of denaturation at 95°C for 50 s, primer annealing at 59°C for 55 s, extension at 72°C for 2 min and a final extension at 72 °C for 10 min.

The PCR products were verified by electrophoresis on a 1% agarose gel in 1 X TAE running buffer, stained with ethidium bromide (1 µg mL^{-1}), and visualized on a UVP transilluminator (Analytik Jena AG, Germany). PCR products were purified from agarose gel with GeneJET Gel Extraction Kit (Thermo Fisher Scientific, Waltham, MA, USA) and sequenced by a commercial company (BaseClear B.V., Leiden, The Netherlands), with specific primers. SSL-1 gene was partially sequenced (3811 bp) using the primer walking technique with internal newly designed primers (data not shown).

The obtained SSL-1 fragments ranging from 700 to 1100 bp were manually corrected for mismatches and ambiguous nucleotides using Chromas Lite 2.1.1 software (Technelysium Pty, South Brisbane, Australia). Gene contig was generated using Contig editor from GeneStudio Pro 2.2.0.0 (GeneStudio Inc., USA).

3D structure prediction methodology

The obtained nucleotide sequence was virtually translated into amino acid sequences by Translate tool from ExPASy (Gasteiger *et al.*, 2003). The ORF (Open reading frame) showing long-chain amino acid sequence was subjected to NCBI protein BLAST and further used for protein structure prediction.

Phyre2 web portal (Lawrence *et al.*, 2015) was used to generate a primary 3D structure form the selected ORF. According to Lawrence *et al.*, (2015) the core method of Phyre2 for generating a 3D model consists of four stages: i) gathering homologous sequences using HHblits (Remmert *et al.*, 2012) to generate the evolutionary profile of amino acid residues from query sequence; ii) fold library scanning, data obtained in stage 1 was converted to a hidden Markov model (HMM) and subjected to HHsearch (Söding, 2005) for HMM-HMM matching against a database of known structures; iii) loop modeling using cyclic coordinate descent (Canutescu and Dunbrack, 2003); iv) side chain placement using R3 fitting protocol (Wei and Sahinidis, 2006).

Protein structure obtained from Phyre2 was subjected to quality assessment by RAMPAGE - Ramachandran plot analysis (Lovell *et al.*, 2003), ProSA-web Z-score (Wiederstein and Sippl, 2007) and MolProbity global score (Chen *et al.*, 2009).

Further, PREFMD (Protein REFinement via Molecular Dynamics) (Heo and Feig, 2018) was run locally for protein structure refinement. After refinement, the obtained model was subjected to another step of energy minimization process by employing the steepest descent algorithm with Charmm36 force field (Huang and MacKerell, 2013) and was allowed to run until it converged to machine precision or to a maximum force on each atom less than 100 kJ mol^{-1} nm^{-1}. For this step, we used the GROMACS (Abraham *et al.,* 2015) software package with GPU acceleration on a Linux workstation. The refined 3D model obtained at the end of the protocol was subjected to another round of quality assessment and validation.

Finally, the refined protein structure was used for subsequent analysis, including: multiple sequence alignment using ClustalW from Mega7 (Kumar *et al.*, 2016); identification of functional regions and calculating of evolutionary conservation by ConSurf (Ashkenazy *et. al.*, 2016), ligand binding site prediction by P2Rank (Krivák and Hoksza, 2018; Jendele *et al.*, 2019). UCSF CHIMERA was used for protein visualization and analysis (Pettersen *et al.,* 2004).

Results

Molecular analysis

Total DNA extracted from *B. terribilis* AICB 872 strain was 387.7 ng μL^{-1}. Nucleic acids purity was estimated by analysing the A_{260}/A_{280} and A_{260}/A_{230} ratios. The A_{260}/A_{280} for DNA isolated using ZR Soil Microprobe DNA MiniPrep™ Kit was 2.10 and A_{260}/A_{230} ratio was 1.68. The isolated gDNA was further subjected to gel electrophoresis to verify the integrity of extracted DNA (Fig. 1A). DNA gel electrophoresis yielded a strong band above the 10000 bp marker band and a faintly smear consisting of DNA fragments under 500 bp.

PCR amplification of SSL-1 gene using gDNA as template and SSL1-F2 and SSL1-R3 primer pair, revealed a band of approx. 3500 bp. The PCR product was isolated and sequenced. SSL1-F2 (820 bp) and SSL1-R3 (777 bp) fragments were subjected to NCBI Nucleotide BLAST. SSL1-F2 nucleotide sequence BLAST report showed a high percentage of identity with 57-172 and 173-333 regions from *Botryococcus braunii* squalene synthase-like 1 mRNA (HQ585058.1); similar SSL1-R3 showed 100% identity with 951-1058 and 1059-1114 region from the same mRNA sequence of *B. braunii*. Further, we designed a new primer pair (data not shown) and used the primer waking techniques to partially sequence the SSL-1 gene. Finally, using Contig editor from GeneStudio Pro we managed to assemble a partial SSL-1 contig of 3811 bp. The partial CDS of SSL-1 was obtained by multiple alignments with similar squalene synthase CDS fragments in MEGA 7 and BLAST from NCBI.

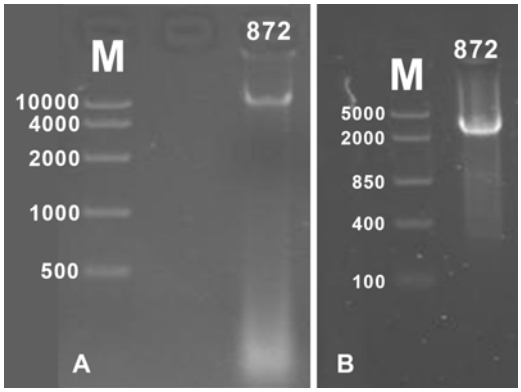

Figure 1. Gel electrophoresis of gDNA extracted from *B. terribilis* AICB 872 (A) and PCR product amplification of SSL-1 gene using SSL1-F2 and SSL1-R3 primer pair (B). DNA sample was electrophoresed on 1.6% (A) or 1% (B) agarose gel in 1X TAE Buffer. M - FastRuler High Range (A) / Middle Range (B) DNA Ladder (Thermo Fisher Scientific, Waltham, MA, USA).

Based on Phyre2 secondary structure and disorder prediction (Fig. 2), the SSL-1 of *B. terribilis* AICB 872 strain consists of 18 α-helix structure (70% of total protein structure) and a short β-sheet at Met216 and Phe217 residues position. The C-terminal transmembrane domain was not observed according to Phyre2 secondary structure prediction. Based on domain prediction, high hits were obtained with the chain C of *Trypanosoma cruzi* squalene synthase (PDB: 3wcc) and chain A of human squalene synthase (PDF: 1ezf).

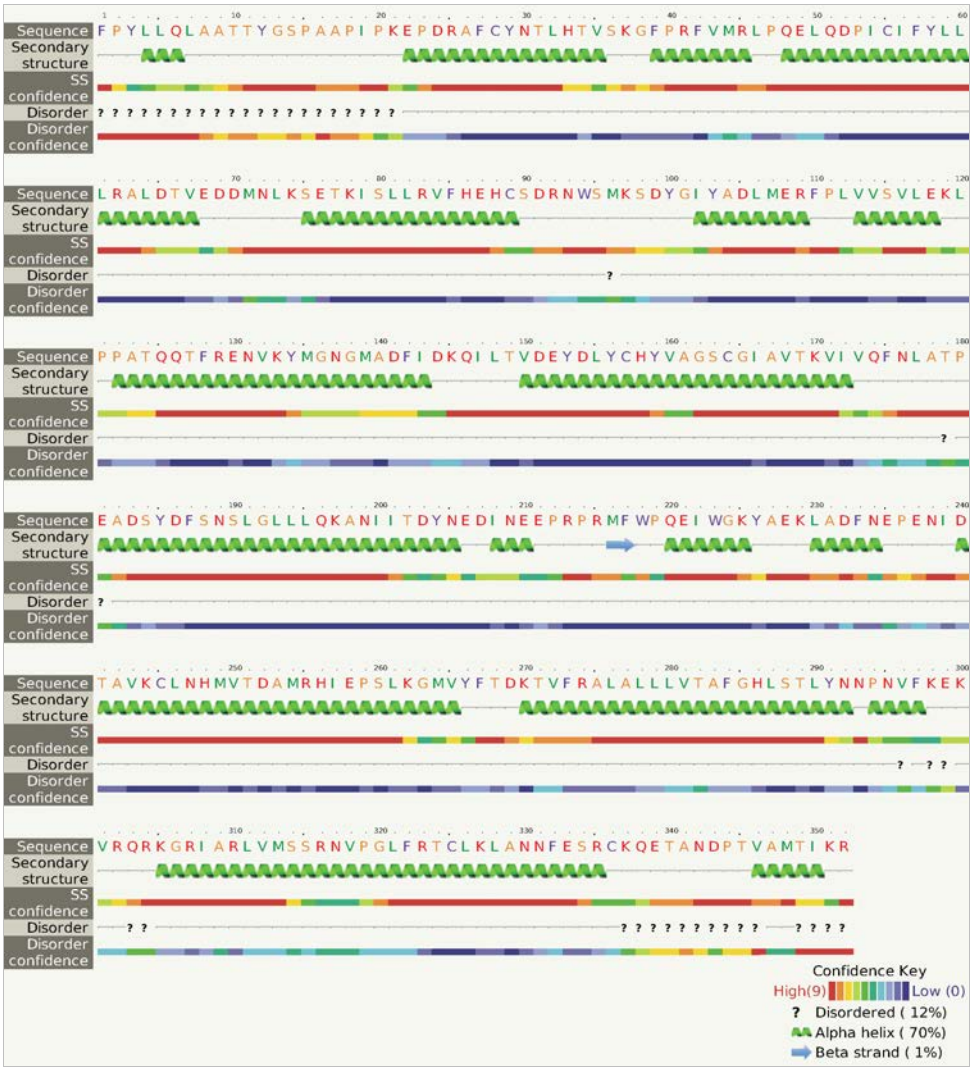

Figure 2. Protein secondary structure and disorder prediction of SSL-1 enzyme of *B. terribilis* AICB 872 strain generated by Phyre2.

Structure prediction and refinement.

The 3D structure of squalene synthase-like 1 from AICB 872 was generated by Phyre2 using the PDB structures of 3wcc chain C and 1ezf chain A as templates. The squalene from *Trypanosoma cruzi* 3wcc was found to be most similar to our protein, having a sequence identity of 38%. Based on these two templates 329 aa were modeled with 100% confidence while 23 residues were modeled by *ab initio*. Before further analysis of the protein structure, we performed a refinement step using PREFMD protocol followed by energy minimization in GROMACS.

Based on model quality score for both unrefined and refined structures, the refined model was found to be more accurate based on various parameters including MolProbity, Ramachandran plot and Prosa Z score. Protein dihedral angle plotted as a Ramachandran diagram (Fig. 3) showed better stereochemical properties for the refined model with 96% of the residues in the favoured regions while for the initial model just 92% of the residues were found in this region. The amino acid residues Thr30, Leu65, Ser117, Gly120, Glu137, Leu139, Pro199, Glu229, Met235, Ala246, Asn257 and Lys317 from the initial structure, Tyr3, Lys21, Tyr100 and Arg213 from the refined structure, respectively were found in the disallowed region of the Ramachandran map.

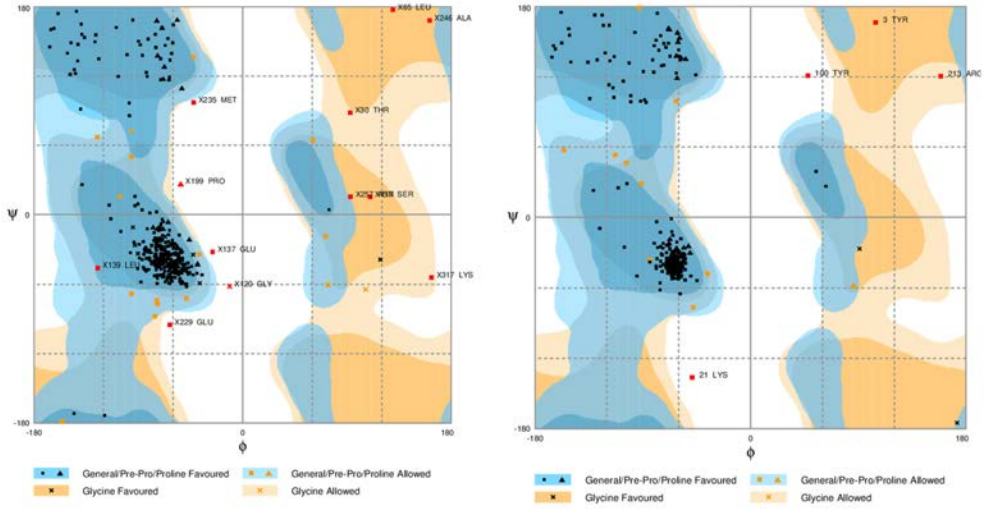

Figure 3. RAMPAGE - Ramachandran plot of both initial (left) and refined (right) SSL-1 protein structure from *B. terribilis* AICB 872 strain.

The overall quality score calculated by ProSA for the two models were very similar 9.23 for initial model and 8.98 for the refined, respectively. Graphic representation of the Z-score (Fig. 4) in the context of all known, experimentally determined protein structures, is placing the models in the characteristic range for

native proteins with similar molecular weight. A more drastic improvement of the refined model was observed based on MolProbity, where the overall score was 1.66 with 0 clashes between atoms, 0 bad bonds and 60 bad angles compared to the crude model that has a general quality score of 3.02 with 126 clashes, 21 bad bonds and 135 bad angles. The only score better for the crude model consists of 92% favoured rotamers over 85%.

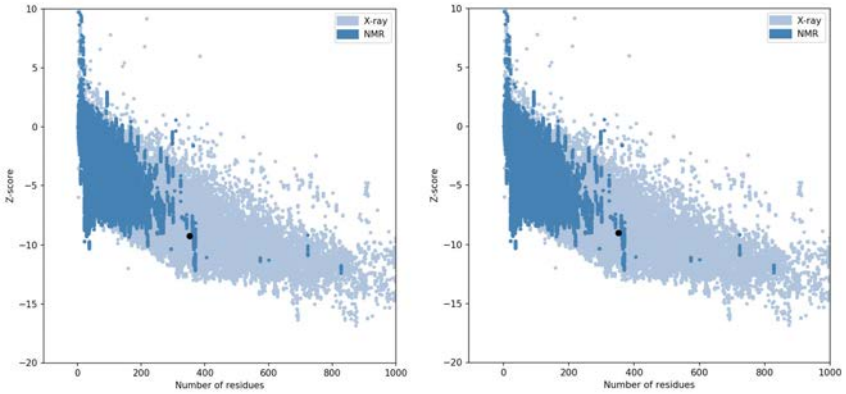

Figure 4. ProSA-web Z-score graphic representation of initial SSL-1 (left) and refined (right) from *B. terribilis* AICB 872 strain.

Visual analysis and comparison of the two models by superimposing them with MatchMaker from UCSF Chimera showed good alignment of all a-helices, with no visible modifications. Most differences between the models are among coil structures (Fig. 5) and both N and C terminus regions of the protein. These regions, as shown by 2D structure prediction, are characterized by the presence of highly disorder structural fragments. The results are not unexpected, because the disordered regions usually present intrinsic flexibility, which is problematic even for experimentally methods not just for computational methods. Also, these fragments were not covered by the template provided, the residues being modeled *ab initio.*

We used the refined model to predict the ligand-binding site and to identify the amino acids that form the pocket. This analysis was performed by P2Rank with default settings. We predicted 5 pockets, of which only one had high confidence score (>50). The predicted pocket is comprised of 48 residues being located in the middle part of the protein (Fig. 6 centre). Also, based on the protein conservation analysis done with ConSurf, the predicted pocket is localized in a well-conserved region (Fig. 6 left). Based on these facts we concluded that this pocket is responsible for specific binding of the ligand. The ligand prediction based on the most conserved binding site revealed two possible candidate: PS7 and FPS - Farnesyl thiopyrophosphate.

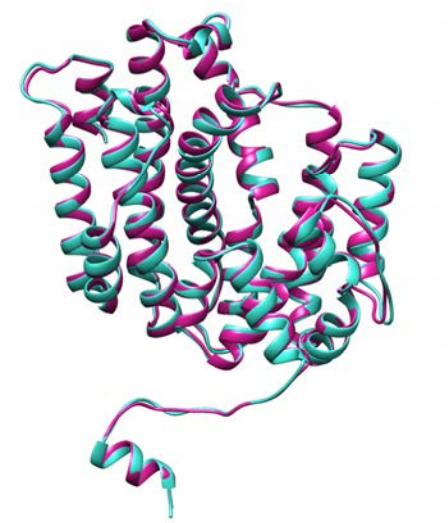

Figure 5. The superimposed 3D structure of both initial (magenta) and refined (turquoise) model of the SSL-1 enzyme from *B. terribilis* AICB 872 strain. The superimposed 3D structure was obtained using UCSF CHIMERA.

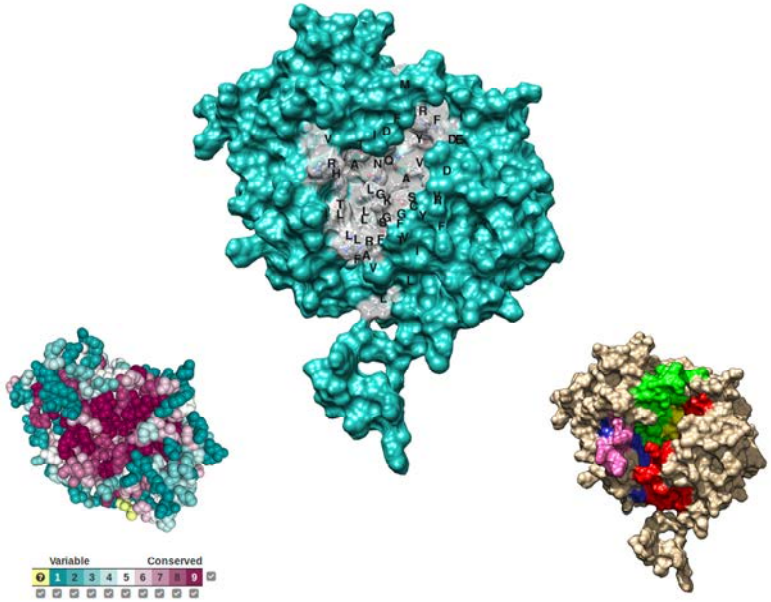

Figure 6. Amino acid residue conservation according to ConSurf (left), predicted pocket from P2Rank (centre) and conserved domain (Lee & Poulter, 2008) (right) representation of the SSL-1 enzyme from *B. terribilis* AICB 872 strain. Domain I – red, Domain II – yellow, Domain III – green, Domain IV – blue, NADPH binding motif – pink.

Multiple sequence alignments were performed as previously described (Szoke-Nagy *et al.*, 2015) and four conserved domains, two Mg^{2+} binding motifs and one NADPH binding motif were also identified in SSL-1 enzyme from *B. terribilis* AICB 872 strain (Fig. 6 right). Domain I and III presented each other one aspartate-rich motif being possibly involved in substrate binding. The conserved domains and motifs were previously described and identified by Gu *et al.* (1998), Pandit *et al.* (2000), and Lee and Poulter (2008).

Discussion

Squalene synthase-like proteins are enzymes involved in hydrocarbon biosynthesis in *B. braunii*, being isolated and described for the first time by Niehaus *et al.* (2011). Hydrocarbon biosynthesis in *B. braunii* Showa starts with a head-head condensation of two moieties of FPP to form an intermediate PSPP, a reaction catalysed by SSL-1, further squalenes are produced by rearrangement of PSPP by SSL-2 in an NADPH-dependent reaction, while SSL-3 converts PSPP to botryococcenes.

In this study, we identified and partially sequenced SSL-1 gene (3811 bp) from a local strain of *Botryococcus terribilis* AICB 872 using previously-designed primer pairs. The stain *B. terribilis* AICB 872 was isolated from a fishpond located in Mureş country from the region of Transylvania, Romania. The homology modeling analysis of SSL-1 enzyme from *B. terribilis* AICB 872 revealed the presence of 18 alfa helix structures and a small beta-sheet. Based on model quality score for both unrefined and refined structures, the refined model was found to be more accurate having less amino acid residues in the disallowed region of Ramachandran plot, and a better Z-score. The superimposed 3D structure of both SSL-1 enzymes from *B. braunii* and *B. terribilis* AICB 872 models showed that the coil structures and both N and C terminus regions of the protein are characterized by the presence of highly disordered structural fragments. Based on the most conserved binding site two possible ligand candidates were identified as FPS and PS7.

The 3D structure of squalene synthase-like enzymes from *B. braunii* was previously investigated by Bell *et al.* (2014) and Elumalai *et al.* (2018). Bell *et al.* (2014) used tridimensional structure prediction and directed mutagenesis of SSL-1 and SSL-3 from *B. braunii* Showa and for a better understanding of 1′–1 and 1′–3 linkages specificity for PSPP rearrangement, also they attempted to identify the functional residues and domain responsible for this critical step.

The hydrocarbon content of *B. terribilis* AICB strains was published for the first time by Hegedűs *et al.* (2014), showing the ability of *B. terribilis* AICB strains to synthesize C_{30}–C_{32} botryococcenes similar with those produced by chemical race B of *B. braunii* strain. Our observations related to the similarities

between SSL-1 from both *B. terribilis* AICB 872 and *B. braunii* unravelled the high degree of conservation of the catalytic situs from both strains. On these facts, we believe that the first step of hydrocarbon biosynthesis starting from two moieties of FPP is similarly in both species. Until these findings, the hydrocarbon biosynthesis was limited to one species, being described only at *B. braunii*.

Conclusions

In the present work, we successfully identified and sequenced the partial SSL-1 gene (3811 bp) encoding squalene synthase-like 1 enzyme from *B. terribilis* AICB 872 strain using PCR assay with previously described primer pairs.

The procedures for *in silico* tridimensional structure prediction and analysis of amino acids sequence from *B. terribilis* AICB 872 strain revealed high-quality tridimensional structure which can be used for structure prediction for other SSL enzymes.

The homology modeling analysis of tridimensional structure reveals a high degree of conservation of the catalytic situs between the predicted structure of SSL-1 enzyme from *B. terribilis* AICB 872 strain with those obtained by Elumalai *et. al.* (2018) and SSL-1 from *B. braunii*. Moreover, hydrocarbon content of *B. terribilis* AICB strains published by Hegedűs *et al.* (2014), which reveals the ability of *B. terribilis* AICB strains to synthesize C_{30}–C_{32} botryococcenes similar with those found in chemical race B of *B. braunii*, we presume that the first step of hydrocarbon biosynthesis in *B. terribilis* occurs similar way with the one described in *B. braunii* strains.

REFERENCES

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2, 19-25.

Achitouv, E., Metzger, P., Rager, M.N., & Largeau, C. (2004). C31 - C34 methylated squalenes from a bolivian strain of *Botryococcus braunii*. *Phytochem*., 65(23), 3159-3165.

Al-Hothaly, A. K. (2018). An optimized method for the bio-harvesting of microalgae, *Botryococcus braunii*, using *Aspergillus* sp. in large-scale studies. *MethodsX*, 5, 788-794.

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.*, 44:W344-W350.

Bell, S. A., Niehaus, T. D., Nybo, E., & Chapell, J. (2014). Structure-function mapping of key determinants for hydrocarbon biosynthesis by squalene and squalene synthase-like enzymes from the green alga *Botryococcus braunii* race B. *Biochemistry*, 53(48), 7570-7581.

Canutescu, A. A., & Dunbrack, R. L. (2003). Cyclic coordinate descent: a robotics algorithm for protein loop closure. *Protein Sci.,* 12(5), 963–972.

Chen, V. B., Arendall III, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapra, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2009). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D*, 66(Pt 1), 12-21.

Dragoş, N., Peterfi, L. Ş., Momeu, L., & Popescu, C. (1997). An introduction to the algae and the culture collection of algae – At the Institute of Biological Research Cluj-Napoca. Cluj University Press, Cluj-Napoca, 267 pp.

Elumalai, S., Sangeetha, T., & Rajesh Kanna, G. (2018). *In silico* modeling and characterization of squalene synthase and botryococcene synthase enzymes from a green photosynthetic microalga *Botryococcus braunii. J. Pet. Environ. Biotechnol.*, 9, 3.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, 31(13), 3784-3788.

Gu, P., Ishii, Y., Spencer, T. A, & Shechter, I. (1998). Function-structure studies and identification of three enzyme domain involved in the catalytic activity in rat hepatic squalene synthase. *J. Biol. Chem.*, 273(20), 12515-12525.

Guiry, M. D., & Guiry, G. M. (2019). AlgaeBase, World-wide electronic publication, National University of Ireland, Galway, http://www.algaebase.org/

Hegedűs, A., Mocan, A., Barbu-Tudoran, L., Coman, C., Drugă, B., Sicora, C., & Dragoş, N. (2014). Morphological, biochemical, and phylogenetic assessments of eight *Botryococcus terribilis* strains collected from freshwaters of Transylvania. *J. Appl. Phycol.*, 27(2), 865-878.

Heo, L., & Feig, M. (2018). PREFMD: a web server for protein structure refinement via molecular dynamics simulations. *Bioinformatics*, 34(6), 1063-1065.

Hillen, L. W., & Wake, L. V. (1979). Solar oil [microform]: liquid hydrocarbon fuels from solar energy via algae, First National Conference of Australian Institute of Energy, The University of Newcastle, 5th-9th February.

Hunag, J., & MacKerell, A. D. (2013). CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.*, 34(25), 2135-2145.

Huang, Z., & Poulter, C. D. (1989). Tetramethylsqualene, a triterpene from *Botryococcus braunii* var. Showa. *Phytochem*., 28(5), 1467-1470.

Jendele, L., Krivák, R., Škoda, P., Novotný, M., & Hoksza, D. (2019). PrankWeb: a web server for ligand binding site prediction and visualization, *Nucleic Acids Res.*, 47(W1), W345-W349.

Kawachi, M., Tanoi, T., Demura, M., Kaya, K., & Watanabe, M. M. (2012). Relationship between hydrocarbons and molecular phylogeny of *Botryococcus braunii*. *Algal Res.*, 1(2), 114-119.

Krivák, R., & Hoksza, D. (2018). P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *J. Cheminformatics*, 10(1), 39.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.*, 33(7), 1870-1874.

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

Lee, S., & Poulter, C. D. (2008). Cloning, solubilization, and characterization of squalene synthase from *Thermosynechococcus elongates* BP-1, *J. Bacteriol.,* 190(11), 3808 – 3816.

Lee, M. H., Jeong, J. H., Seo, J. W., Shin, C. G., & Kim, Y. S. (2004). Enhanced triterpene and phytosterol biosynthesis in *Panax ginseng* overexpressing squalene synthase gene. *Plant Cell Physiol.*, 45(8), 976- 984.

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

Metzger, P., & Largeau, C. (2005). *Botryococcus braunii*: a rich source for hydrocarbons and related ether lipids. *Appl. Microbiol. Biot.*, 66(5), 486-496.

Metzger, P., Allard, B., Casadevall, E., Berkaloff, C., & Coute, A. (1990). Structure and chemistry of a new chemical race of *Botryococcus braunii* that produces lycopadiene, a tetraterpenoid hydrocarbon. *J. Phycol.*, 26(2), 258-266.

Metzger, P., & Casadevall, E. (1987). Lycopadiene, a tetraterpenoid hydrocarbon from new strains of the green alga *Botryococcus braunii*. *Tetrahedron Lett.*, 28(34), 3931-3934.

Metzger, P., Templier, J., Largeau, C., & Casadevall, E. (1986). An *n*-alkatriene and some *n*-alkadienes from the A race of the green alga *Botryococcus braunii*. *Phytochem*., 25(8), 1869-1872.

Metzger, P., Casadevall, E., Pouet, M. J., & Pouet, Y. (1985). Structures of some botryococcenes: branched hydrocarbons from the B race of the green alga *Botryococcus braunii*, *Phytochem*., 24(12), 2995-3002.

Niehaus, T. D., Okada, S., Devarenne, T. P., Watt, D. S., Sviripa, V., & Chappell, J. (2011). Identification of unique mechanisms for triterpene biosynthesis in *Botryococcus braunii*. *PNAS*, 108(30), 12260-12265.

Pandit, J., Danley, D. E., & Schulte, G. K. (2000). Crystal structure of human squalene synthase. A key enzyme in cholesterol biosynthesis. *J. Biol. Chem.,* 275(39), 30610-30617.

Park, J., Matralis, A. N., Berghuis, A. M., & Tsantrizos, Y. S. (2014). Human isoprenoid synthase enzymes as therapeutic targets. *Front. Chem.*, 2: 10.3389/fchem.2014. 00050.

Petersen, H. I., Rosenberg, P., & Nytoft, H. P. (2008). Oxygen groups in coals and alginite-rich kerogen revisited. *Int. J. Coal Geol.*, 74(2), 93-113.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.*, 25(13), 1605-1612.

Remmert, M., Biegert, A., Hauser, A., & Söding, J. (2012). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods*, 9, 173–175.

Rong, Q., Jiang, D., Chen, Y., Shen, Y., Yuan, Q., Lin, H., Zha, L., Zhang, Y., & Hunag, L. (2016). Molecular cloning and functional analysis of squalene synthase 2(SQS2) in *Salvia miltiorrhiza* Bunge. *Front. Plant Sci.*, 7:1274.

Seo, J. W., Jeong, J. H., Shin, C. G., Lo, S. C., Han, S. S., Yu, K. W., Harada, E., Han, J. Y., & Choi, Y. E., (2005). Overexpression of squalene synthase in *Eleutherococcus senticosus* increases phytosterol and triterpene accumulation. *Phytochem*., 66(8), 869-877.

Söding, J. (2005). Protein homology detection by HMM-HMM comparison,. *Bioinformatics*, 21(7), 951–960.

Szoke-Nagy, T., Hegedűs, A., Baricz, A., Chiriac, C., Szekeres E., Coman, C., & Dragoş, N. (2015). Identification, isolation and bioinformatic analysis of squalene synthase-like cDNA fragments in *Botryococcus terribilis* AICB 870 Strain. *Studia UBB Biologia*. LX(1), 23-37.

Thapa, H. R., Naik, M. T., Okada, D., Takada, K., Molnár, I., Xu. Y., & Devarenne, T. P. (2016). A squalene synthase-like enzyme initiates production of tetraterpenoid hydrocarbons in *Botryococcus braunii* Race L. *Nat. Commun.*, 7:e11198.

Wei, X., & Sahinidis, N. V. (2006). Residue-rotamer-reduction algorithm for the protein side-chain conformation problem. *Bioinformatics*, 22(2), 188–194.

Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*., 35, W407-W410.