Biotechnology Reports 45 (2025) e00869

Contents lists available at ScienceDirect

Biotechnology Reports

journal homepage: www.elsevier.com/locate/btre

Research Article

Prospective identification of extracellular triacylglycerol hydrolase with conserved amino acids in *Amycolatopsis tolypomycina*'s high G+C genomic dataset

Supajit Sraphet [a], Bagher Javadi [b,*]

[a] *Institute of Molecular Biosciences, Mahidol University, Nakhon Pathom, 73170, Thailand*
[b] *Department of Sciences, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok, 10300, Thailand*

ARTICLE INFO

Keywords:
Amycolatopsis tolypomycina
Conserved-solvent accessibility
Enzyme stability
Computational proteogenomics
Structural biology analysis

ABSTRACT

Extracellular triacylglycerol hydrolases (ETH) play a critical role for microorganisms, acting as essential tools for lipid breakdown and survival in challenging environments. The pursuit of more effective ETH genes and enzymes through evolution holds significant potential for enhancing living conditions. This study employs a proteogenomic approach to identify high G+C ETH in a notable Gram-positive bacterium, *Amycolatopsis tolypomycina*. Utilizing knowledge from genome and machine learning algorithms, prospective ETH genes/enzymes were identified. Notably, the ETH structural conserved accessibility to solvent clearly indicated the specific sixteen residues (GLY50, PRO93, GLY141, ASP148, GLY151, ASP172, ALA176, GLY195, TYR196, SER197, GLN198, GLY199, GLY200, GLY225, PRO327, ASP336) with no frequency. By pinpointing key residues and understanding their role, this study sets the stage for enhancing ETH performance through computational proteogenomic and contributes to the broader field of enzyme engineering, facilitating the development of more efficient and versatile ETH enzymes tailored to specific industrial or environmental contexts.

1. Introduction

Finding the meaningful results from the biological data needs comprehensive statistical methods. Recently with computational approaches, this process cumulated lots of benefit especially with machine learning algorithms [1–3]. Modeling and learning from biological data towards classifying specific features with supervised and unsupervised data structure help biologists to find, explore, mimic and engineer the nature successfully. The protein/enzyme homology and evolution specifically have been received more attention with machine learning algorithms [4–7]. However, numerous questions remain to be addressed in order to tackle environmental pollution, particularly that caused by lipids, esters, and their derivatives. In both developed and developing countries, one significant source of pollution is petroleum hydrocarbons, which are prevalent in oil spills, industrial and municipal waste, as well as wastewater. These pollutants require careful and direct removal from the environment. The bioremediation of hydrocarbon contamination can be effectively achieved with the assistance of extracellular triacylglycerol hydrolases (ETHs), particularly those derived from microbial sources [8].

Biomolecular agents, such as triacylglycerol hydrolase enzymes, play a crucial role in addressing environmental pollution. Conversely, understanding the mechanisms of pathogenic actions, particularly enzymatic virulence factors like extracellular triacylglycerol hydrolases (ETHs) in *Candida albicans* [9–11], can be elucidated through structural studies of ETHs [12–16]. The necessity to address these challenges, combined with the potential and efficacy of enzymes, drives researchers to identify more efficient enzymes and unravel their structural complexities [13].

Enzymes derived from microorganisms represent promising candidates for study due to their remarkable ability to enable survival in harsh and diverse environments. Among these, the extraordinary extracellular enzyme known as triacylglycerol hydrolase (TH) (EC 3.1.1.3) is secreted into the surrounding environment to facilitate the processing and provision of essential nutrients for microbes [17,18]. It gained more attentions due to its diverse applications from food to medical and biosensor industries as well as environmental pollutions. The substrate of this enzyme is long chain triacylglycerol which hydrolyzed to fatty acids, diacylglycerol, monoacylglycerol and glycerol [19,20]. The mechanism of triacylglycerol hydrolase reaction is based on the

* Corresponding author at: Department of Sciences, Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok, 10300, Thailand.
E-mail address: Javadi.ba@ssru.ac.th (B. Javadi).

https://doi.org/10.1016/j.btre.2024.e00869
Received 12 March 2024; Received in revised form 3 November 2024; Accepted 6 December 2024
Available online 7 December 2024

alpha/beta hydrolase fold that possess catalytic triad included an acidic amino acid, a histidine and a nucleophilic serine residue. However, enzyme activation and enzyme-substrate-water interfacial and solvent accessibility of different triacylglycerol hydrolases need to be scrutinized in more details [21].

Different sources of triacylglycerol hydrolase enzyme from animals and plants to microbes provide lots of scientific reports in literatures [22]. However, triacylglycerol hydrolases of microbial origin offer distinct advantages, including high-yield production, ease of genetic modification, and a sustainable supply base, owing to the straightforward control of growth conditions, among other factors [23–25]. In addition to numerous research findings and application reports, it is essential to highlight the evolution of triacylglycerol hydrolase enzymes and the need for more detailed categorization. Specifically, further investigation into the distinctions between triacylglycerol hydrolases and esterases is required to develop improved methodologies for differentiating these two enzyme classes.

ETH main structure as well as specific features such as lid, oxyanion hole, substrate binding site would result the unique characterization and classification of different ETH family types. These structural features classified the ETH substrate specificities and selectivity [26]. The interaction between the substrate and total ETH conformation and specific residues provides the selectivity towards fatty acids chain lengths and finalized products. On the other hand, the shape of the ETH especially in scaffold and binding site; and the nature of amino acids that forming them dictate chemoselectivity, regioselectivity, and stereoselectivity [27]. Revealing the specific residue solvent accessibility (as one of the main features) would help to find the buried and exposed accessible area of the ETH structure and therefore provide the inevitable information for ETH engineering through rational design method [28]. Furthermore, accessibility to solvent can help in High throughput screening of the big library in engineering ETH with directed evolution method. Additionally, the conserved residues features involve in structural stability and functionality of ETH. It should be mentioned here that the ETH structural analysis is highly species-specific (even strain-specific) and should comprise the comprehensive specific proteogenomics dataset [29].

Many aspects of biology have influenced with machine learning approaches. The structural analysis of macromolecules had great advantages with different algorithms, from 3- d structure to structure-function prediction and analysis [30]. The details residues information of enzymes that revealed with machine learning methods could provide foreseeable help in evolution and engineering the enzymes [31]. The approaches help in more precise classification and therefore better understanding the functions [32]. This classification with insight structural information can be more useful when applying to the high G+C content species as they have more applications specially in finding drug and solving pollutions.

The importance of high GC genomic prokaryotes and the stability of the genome is considered to be associated with many mutational and selective forces in interaction with other ecological factors [33–46]. Elucidating precise information on extracellular triacylglycerol hydrolases from high G+C species will facilitate the engineering of novel enzymes in the laboratory and provide clearer insights into the evolutionary pathways of ETHs. The ETH from *Amycolatopsis tolypomycina* with more than 90 % similarity with other secreted triacylglycerol hydrolase from other species in the same genus need to be scrutinized in more details. *A. tolypomycina* is belonging to the genus *Amycolatopsis* and family Pseudonocardiaceae that have more than 70 % of the active chemical compounds that produced and commercially supplied in the market [47,48]. Despite the important position of this species and genus and ETH as an enzyme in the market, little was known about the *A. tolypomycina* extracellular triacylglycerol hydrolase structures. Therefore, the aim of this research article was to analyze the structural solvent accessibility of *A. tolypomycina* extracellular triacylglycerol hydrolases with the help of machine learning algorithm approaches.

2. Materials & methods

2.1. Physiochemical characteristics and cluster analysis

All sequences were retrieved from National Center of Biotechnology Information (https://www.ncbi.nlm.nih.gov/). The physiochemical characteristics (https://www.expasy.org/) [49] and similarity of sequences were calculated with Swiss server and Clustal Omega Multiple Sequence Alignment program [http://www.ebi.ac.uk/Tools/msa/clustalo/] and MultAlin server [http://multalin.toulouse.inra.fr/multalin/] respectively [50]. The molecular phylogenetic tree was presented with MEGAX software [51]. Phylogenetic tree of *A. tolypomycina* extracellular triacylglycerol hydrolase and other bacterial proteins was constructed by Neighbor-Joining method. Signal peptide and Subcellular localization were determined with SOSUI and CELLO v2.5 servers [52,53]. Protein- protein interactions provided with STRING webserver and database (https://version-11-5.string-db.org).

2.2. ETH enzyme modelling

Enzyme specific domain/s and secondary structures as well as 3-d structure were obtained from the Chou & Fasman Secondary Structure Prediction [54] and deep learning de novo modeling [55], respectively. In this method, the multiple sequence alignment was constructed followed by the prediction for distance and orientation distribution. Finally, with the coarse-grained structure modeling by energy minimization and full atom structure refinement, models were generated [55]. Ramachandran Map was applied to evaluate the quality of the models provided with deep learning de novo modeling [56].

2.3. Genome annotation

Putative ETH genes identified and annotated based on the whole genome sequencing of *A. tolypomycina* from NCBI (http://www.ncbi.nlm.nih.gov/). The genes sequences were retrieved, visualized and the syntheny of genes on the whole genome dataset were presented.

2.4. ETH solvent accessibility

The Solvent Accessible Surface Area (SASA) of extracellular triacylglycerol hydrolases models was calculated according to Fraczkiewicz and Braun [57]. Surface and nucleus environments of each 3-d cartesian coordinates of enzymes models were defined as the 50 % and 20 % of comparative analysis of SASA with methods described with [58]. Structurally conserved regions of each enzyme obtained by chimera software with default parameters [59]. With the help of machine learning unsupervised agglomerative hierarchical method the clustering of ETH presented.

Agglomerative hierarchical cluster analysis (HCA) as one of the important machine learning algorithms was applied to cluster the features and characteristics of ETH. Clustering scientifically considered as the process to partitioning the set of data points into the groups based on the similarity. Many types of clustering applied in computational protein analysis such as fuzzy clustering [60] and distribution model-based clustering [61]. In our research HCA that considered as the bottom-up strategy in the clustering of each data point was applied. In order to calculate the distance between the SASA (proteins), Euclidean distance and Ward method were applied.

3. Results

3.1. ETH clustering

The sequences of enzymes from *A. tolypomycina* were retrieved from NCBI with the high similarity with ETH from more than one hundred extracellular triacylglycerol hydrolases of other bacteria in different

genus such as *Streptomyces, Thalassolituus, Hydrocarboniclastica, Conexibacter, Nonomuraea, Thermomonospora, Frankia, Actinomadura, Saccharothrix, Alloactinosynnema, Kibdelosporangium* and *Kutzneria* (Fig. 1). These extracellular triacylglycerol hydrolase structures even observed in *Candida albicans* as opportunistic yeast from Saccharomycetaceae family of fungi and the ETH was part of its virulence mechanism. Therefore, the detail information of *A. tolypomycina* extracellular triacylglycerol hydrolases here would provide the better insight for understanding the mechanisms of virulence action and preparing the drug discovery too.

3.2. Physiochemical characteristics of ETH

Number of amino acids and molecular weight of ETH in this high G+C species were in the range of 366–407 (with average 393) and 37 kDa to 42 kDa respectively. Also, total numbers of negative and positive residues were in the range of 26–37 and 21–31. Average of molecular weight was 41 kDa beside the approximately equal number of negative and positive amino acids in the structure of ETH. The average of positive residues was 32 as well as negative residues 27. The isoelectric point (pI) of the ETH enzyme was more than 5.86 that indicated the acidic pH condition. The average of pI was 6. The purification of these enzymes would gain good advantage from this index (pI) for wet Lab experiments. The thermal stability of these enzymes that presented with aliphatic index was more than 83.38. The hydropathicity value of them was more than -0.034 with the average of 0.059. The physiochemical characteristics support the stability of the enzymes. Subcellular localization of all enzymes presented them as extracellular enzymes (Table 1). The percentage of different amino acids presented in the Supplementary table 1. It is clearly obvious that structures had many Ala, Gly, Leu, Pro and Val residues. Met and Cys were in the lowest percentage in ETH from *A. tolypomycina*. The average of Asp and Arg was the same ~ 4. Cys residues in the structure had disulfide s-s bridges. These bridges presented in the conserved region of ETH. All structures had at least one S-S bridge. The details information of disulfide bridge was presented in Table 2.

3.3. ETH enzyme modelling

All three-dimensional models of ETH enzymes from *A. tolypomycina* were obtained by de novo modeling. For more accuracy of models, the templates were applied in the de novo algorithm. All models were evaluated with template score or TM-score, QMEAN and Ramachandran map (Table 3). The stereo chemical evaluation of models showed that more than 94 % of residues were in favorite positions. The QMEAN and solvation indexes were more than 0.31 and -0.01 that indicated that all models were highly accepted. The secondary structure of 3-d models of ETH showed that the helices percentage was higher than beta sheets and turn loops (Table 3). This was one of the indicators of hydrolases enzymes. The secreted enzymes had signal peptide on the N-terminal of 3-d structures with minimum length of 20 amino acids. All ETHs confirmed as extracellular enzymes and all were soluble (Supplementary table 2).

3.4. ETH solvent accessibility

Clustering the extracellular triacylglycerol hydrolases of *A. tolypomycina* based on the Total Solvent Accessible Surface Area (SASA) showed that in overall all enzymes had at least 99 % similarity. Two distinct clusters here were cluster A and B with 99.2 % similarity. The cluster A included ETH enzymes ETH4 (SEB38271.1), ETH3 (SEC99911.1), ETH1 (SED48308.1), and ETH 5 (SEB35044.1) with similarity of 99.7 %. ETH enzymes of cluster B ETH2 (SED70235.1) had 99.2 % similarity with cluster A enzymes. The comparison of phylogenetic analysis of presented enzymes was in supported with SASA clustering. However, the easiness of this clustering compared to phylogenetic enzymes based on the sequences could be its advantage (Fig. 2A). Furthermore, clustering the surface and nucleus SASA in extracellular triacylglycerol hydrolases had the same result with total SASA in different similarity. The similarity percentage of cluster A was more than 95.5 % included the secreted enzymes ETH3, ETH1, ETH4 and ETH5. The two distinct clusters of dendrogram had more than 99.2 % similarity (Fig. 2B). The conserved-region-SASA cluster also provided more than 99.5 % similarity between the ETHs (Fig. 2C). The surface solvent accessible of extracellular triacylglycerol hydrolases in ETH5 had the highest accessibility to solvents. The average surface solvent accessible of secreted lipases was 9553.73 $Å^2$. Furthermore, results of surface accessibility showed that side chains of amino acids on the surface of the 3-d structures of the enzymes were more accessible to solvent or substrate compared to sidechain of buried amino acids (1795.81 $Å^2$ > 596.50 $Å^2$) (Table 4). To find more meaningful results from these homologous ETHs, the structural conservation regions were obtained and presented in Fig. 3. The details of these regions in 3-d structures were analyzed with hierarchical clustering of conserved region-SASA of ETHs.

The extracellular triacylglycerol hydrolases of ETH2 and ETH5 showed very high conserved-SASA similarity. There were three clusters based on the conserved SASA. Here the ETH1 and ETH4 had closer position and distinctly different from others. The second cluster had more than 98 % similarity based on the conserved SASA. The dissimilarity of these two clusters was less than 1 %. The dendrogram presented in Fig. 2C obtained only based on the conserved region SASA and gave better indication to choose the structure for mimicking or engineering.

The correlation of conserved region SASA presented in Table 4. The ETHs had high correlations. The highest correlation was related to ETH2 and ETH5. The extracellular triacylglycerol hydrolases ETH2 and ETH3 had the lower correlation base on the conserved region SASA.

This information beside the information of critical residues in the structures would lead to the potential structural stability predication for protein and enzyme engineering. The minimum conserved SASA of specific residues in the structures observed in our research was zero. The Fig. 3 showed the frequency of residues in these critical conserved-SASA positions. It could be important to mention that GLY and ALA, residues were the most frequent ones in the conserved regions of secreted enzymes with zero SASA in *A. tolypomycina*. The residues highlighted in the structure of ETHs were conserved based on the conserved–SASA and all of them had zero accessibility to solvent (Fig. 3). Therefore, they considered as the potential critical residue in the sense of stabilizing the structure of ETHs.

Moreover, the residues LYS, HIS, GLU, ASN, ARG and THR (Fig. 3) were conserved in all ETHs enzymes with highest solvent accessibility. The dendrogram presented in Fig. 2D was the result of hierarchical clustering of all features. The same clustering pattern here provided with three distinct clusters. ETH4 and ETH1 with 99.6 % of similarity shaped cluster A. ETH2 and ETH5 in cluster B. ETH3 had the minimum similarity with other clusters of secreted enzymes.

The structural analyses of ETHs were compared with other extracellular triacylglycerol hydrolases to find the specific residues with zero frequency. The results showed that 16 residues had no frequency and conserved in all sequences compared (GLY50, PRO93, GLY141, ASP148, GLY151, ASP172, ALA176, GLY195, TYR196, SER197, GLN198, GLY199, GLY200, GLY225, PRO327, ASP336). Among them GLY195, TYR196, SER197, GLN198, GLY199, GLY200 considered the residues with functional characteristics (Table 5). These residues were identical and had the same SASA. After superposition of the all ETHs, their positions were the same (Table 5). It should mention here that most of them had zero SASA in conserved area. The root mean square deviation (RMSD) between corresponding atoms of total models and theses specific segments provided in the Supplementary table 3.

3.5. Genome annotation

Further analysis of bacterial proteomics and protein-protein

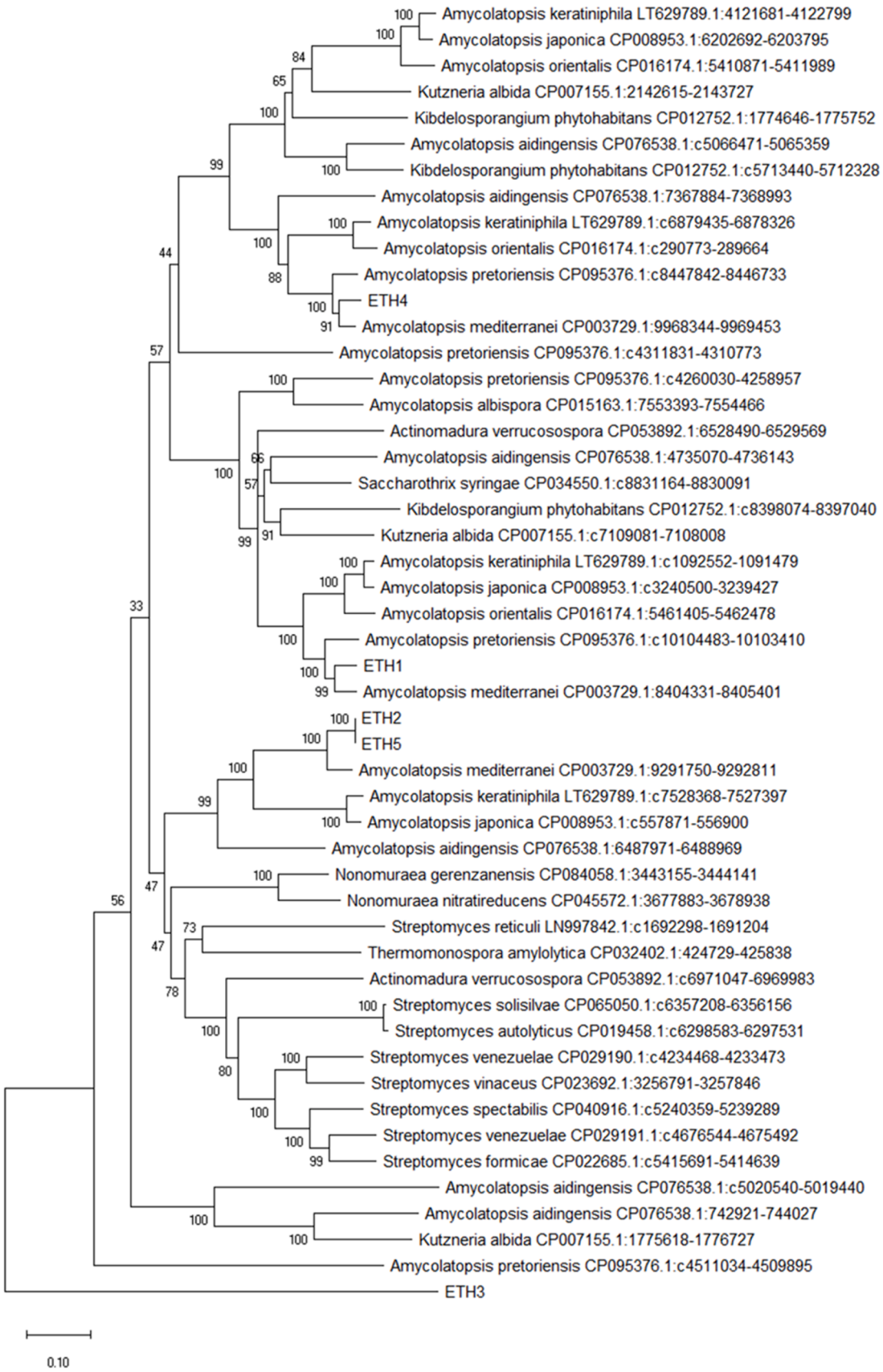

Fig. 1. Phylogenetic of *A. tolypomycina* extracellular triacylglycerol hydrolase and other bacteria.

Table 1
Physiochemical characterization of *A. tolypomycina* extracellular triacylglycerol hydrolase.

ETHs	GenBank accession no.	Number of amino acids	Molecular weight (Da)	Theoretical pI	Total number of negatively charged residues (Asp + Glu)	Total number of positively charged residues (Arg + Lys)	Carbon	Hydrogen	Nitrogen	Oxygen	Sulfur	Formula	Total number of atoms	The instability index (II) is computed to be	Aliphatic index	Grand average of hydropathicity (GRAVY)	Subcellular Localization
ETH1	SED48308.1	395	41,485.99	5.39	36	29	1872	2885	499	553	8	C1872H2885N499O553S8	5817	(stable) 37.83	84.86	0.026	Extracellular
ETH2	SED70235.1	366	37,775.01	6.05	26	21	1719	2652	462	491	4	C1719H2652N462O491S4	5328	(stable) 37.89	92.87	0.215	Extracellular
ETH3	SEC99911.1	402	42,194.92	6.89	31	31	1888	2937	515	560	12	C1888H2937N515O560S12	5912	(stable) 39.02	83.38	-0.034	Extracellular
ETH4	SEB38271.1	407	42,636.24	5.86	37	31	1914	2972	524	568	7	C1914H2972N524O568S7	5985	(stable) 38.2	89.07	0.049	Extracellular
ETH5	SEB35044.1	399	41,572.05	6.2	31	27	1869	2887	519	545	7	C1869H2887N519O545S7	5827	(stable) 40.7	86.67	0.043	Extracellular

Table 2
Positions of disulfide s-s bonds presented in the conserved region of *A. tolypomycina* extracellular triacylglycerol hydrolase.

ETHs	Number of first Cys	Number of second Cys	Distance (Å)
ETH1	CYS (117)	CYS (283)	2.069
	CYS (352)	CYS (395)	2.041
ETH2	CYS (96)	CYS (261)	2.059
ETH3	CYS (16)	CYS (17)	2.337
	CYS (124)	CYS (290)	2.058
ETH4	CYS (121)	CYS (286)	2.044
	CYS (360)	CYS (404)	2.083
ETH5	CYS (129)	CYS (294)	2.057

The two cysteine considered as the disulfide bridge if the distance of sulfur was between 1.5 and 2.5 Ångstrom (Å).

interaction showed that ETHs interacted with triacylglycerol esterase, long chain acy-CoA synthetase, site specific DNA recombinase, S- formylglutathione hydrolase FrmB and regulatory proteins (Fig. 4). These proteins and enzymes should be scrutinized in more details to find the specific interactions.

It should be mentioned here that co-occurrence of these enzymes together gets the good chance for detection and inhibition of this characteristic in pathogenic bacteria and fungi (*Candida albicans*). Phylogenetic clustering earlier revealed that they were in the same cluster as *Streptomyces, Nonomuraea, Thermomonospora, Kutzneria, Actinomadura, Saccharothrix and Kibdelosporangium* genus (Fig. 1). More scrutinizing the genome showed that the gene synteny of ETHs in *A. tolypomycina* were in close proximity to genes serine phosphatase RsbU, acetyltransferase (GNAT) family protein. This detail patterns could be seen in the genome of other high G + C bacteria from different genus. It is noteworthy to mention that ETHs and their structure especially signal peptide and high conserved regions of them have the potential to find specific drug and inhibitor (Fig. 5).

4. Discussion

The significance of protein molecules for living organisms and their diverse applications across various industries is paramount. Studies in proteins and proteomics significantly enhance these applications. Consequently, investigating protein functions—directly related to their structure—is essential for understanding and leveraging their capabilities [62,63]. In this research the correlations of protein structure successfully deduced and categorized the specific characteristics, features and properties of amino acids in different ETHs models.

High-level production of extracellular triacylglycerol hydrolases (ETHs) for various industries necessitates a comprehensive understanding of the molecular mechanisms involved in their expression, folding, and subsequent secretion. The ideal industrial ETH should possess stability, selectivity, and high productivity to meet diverse application requirements [64]. To achieve these objectives, the isolation and purification of new enzymes can be both time-consuming and labor-intensive. Consequently, modifying existing enzymes with specific characterizations becomes essential for enhancing their efficacy and suitability for industrial applications [65]. Enhancing enzyme properties necessitates modifications to the enzyme's structure at the amino acid level. Two primary strategies, directed evolution and rational design, utilize various techniques—such as site-directed mutagenesis, iterative saturation mutagenesis, error-prone PCR, and DNA shuffling—to enable researchers to successfully attain these objectives [66]. However, without a comprehensive understanding of the enzyme's structure, stability, and conformation, along with the specific roles of amino acids, enhancing enzyme functionality becomes nearly impossible or exceedingly challenging. Therefore, enzyme analysis relies heavily on various physicochemical properties and the evolutionary context of structural residues [29].

In our research we showed that by combination of evolutionary and

Table 3
Features of *A. tolypomycina* extracellular triacylglycerol hydrolase models predicted from deep learning de novo modeling.

ETHs	Model evaluation features								Secondary structure			Nucleus amino acids			Surface amino acids		
	Oligo state	Ligand	QMEAN	Cβ	Solvation	Torsion	RRFR (%)	RROR (%)	Helix (%)	Sheet (%)	Turn (%)	Total	BB	SC	Total	BB	SC
ETH1	M	N	1.20	-0.68	0.46	1.20	96.95	0.00	66.3	37.5	12.7	1717.26	1150.87	611.83	11,247.96	7551.28	1918.39
ETH2	M	N	1.66	-1.70	0.55	1.90	97.80	0.27	65.0	38.5	12.0	1374.22	829.32	558.19	102.98	71.03	18.22
ETH3	M	N	1.22	-1.39	0.36	1.41	98.00	0.25	62.4	38.3	13.4	1621.45	1143.72	574.26	10,815.07	6883.51	1985.75
ETH4	M	N	0.31	-0.68	0.28	0.36	94.32	1.23	63.1	53.3	13.0	1852.41	1138.29	633.17	12,116.37	7615.38	2334.56
ETH5	M	N	0.92	-2.16	-0.01	1.41	97.48	0.00	66.3	37.5	12.7	1459.44	836.98	605.04	13,486.26	8689.18	2722.12

RRFR: Ramachandran plot-number of Residues in Favored Region
RROR: Ramachandran plot-number of Residues in Outlier Region
BB: Backbone
SC: Sidechain
M: Monomer, N: None

physicochemical properties of *A. tolypomycina*'s ETH the conserved 16 residues can successfully identify that would be essential for stability and conformation of this enzyme in this species (data set). The results of other researchers showed that only by changing the specific residues surface accessibility of ETH (in other bacteria) the functionality of the enzyme could improve. Modification of the buried (zero SASA) residues would result in deformation and denaturation of ETH. Thus, binding pocket and lid region of the ETH in literature showed more modification compared to other amino acids [26]. This modification applied in three steps: determination of specific protein changes through rational design method, making the changes in the structure and evaluation the engineered enzyme properties [67].

Recently, numerous concepts have been adopted to facilitate successful enzyme engineering, focusing on specific alterations in targeted regions such as the active site or flexible regions identified through structural modeling or sequence comparison. Our findings offer a valuable roadmap for understanding the diverse roles of amino acids in the ETH of *A. tolypomycina*. The modifications involve the deletion, addition, or alteration of individual amino acids at specific sites [68].

For instance, enhancing thermostability has been accomplished by either removing glycine or introducing proline in loop regions, as well as by adding extra salt bridges or disulfide bonds [69]. There are three primary methods for enzyme engineering: directed evolution, rational design, and a combined approach. Two significant techniques within rational design are site-directed mutagenesis and whole plasmid single-round PCR [27]. Computational analysis has played a crucial role in identifying specific regions and even individual amino acids suitable for modification. It is worth noting that computational methods can also pinpoint specific amino acids that contribute to the stabilization of the enzyme's structure [70]. Deep knowledge of structure and amino acids properties helped in redesigned the ETH with better thermostability [69], with replacing glycine, deleting the N-terminal domain or enhancing the stability with adding the surface hydrogen bonds [71] or replacing the high B factor region [72]. All of these improvements achieved based on the hot spot identification region/s in the structure. Finding hot spot regions of the structure need deep understanding on solvent accessibility and physicochemical properties of specific amino acids. The engineered process would be fulfilled with comprehensive analysis of these properties in combination of evolutionary factors. Finding the structural model of the different ETHs gained vital information for ETH engineering. Many researchers engineered new ETH based on above mentioned methods [73,74].

Our research on the structural hotspots of *A. tolypomycina* ETH provides valuable insights and serves as a blueprint for identifying the specific 16 amino acids within this proteogenomic dataset. The rational analysis of the ETH structure and its specific hotspots has been employed by other researchers in different genomic datasets; however, it is important to note that this method is species-specific. The Solvent Accessible Surface Area (SASA) of these specific amino acids plays a crucial role in guiding the modifications necessary for engineering ETH.

Early research on the enantioselectivity of *Candida antarctica* ETH, utilizing molecular modeling and rational design, demonstrated that virtually mutated amino acids within the modeling region exhibited enhanced interactions with target substrates. The observed changes in enantioselectivity resulting from these mutations in the laboratory correlated well with the virtual model, particularly regarding the chiral recognition of alcohol enantiomers. Additionally, the study identified unfavorable interactions between the halogen moiety of the rapidly reacting S enantiomer and a hotspot region located at the bottom of the stereoselectivity pocket [75].

In another study, Magnusson and colleagues employed rational design analysis to identify specific hotspot residues, successfully altering the size of the stereospecificity pocket in ETH by modifying Trp104. The engineered larger stereospecificity pocket accommodated secondary alcohols more effectively, resulting in a substrate selectivity shift of over 400,000 times for nonan-5-ol compared to propan-2-ol, facilitated by

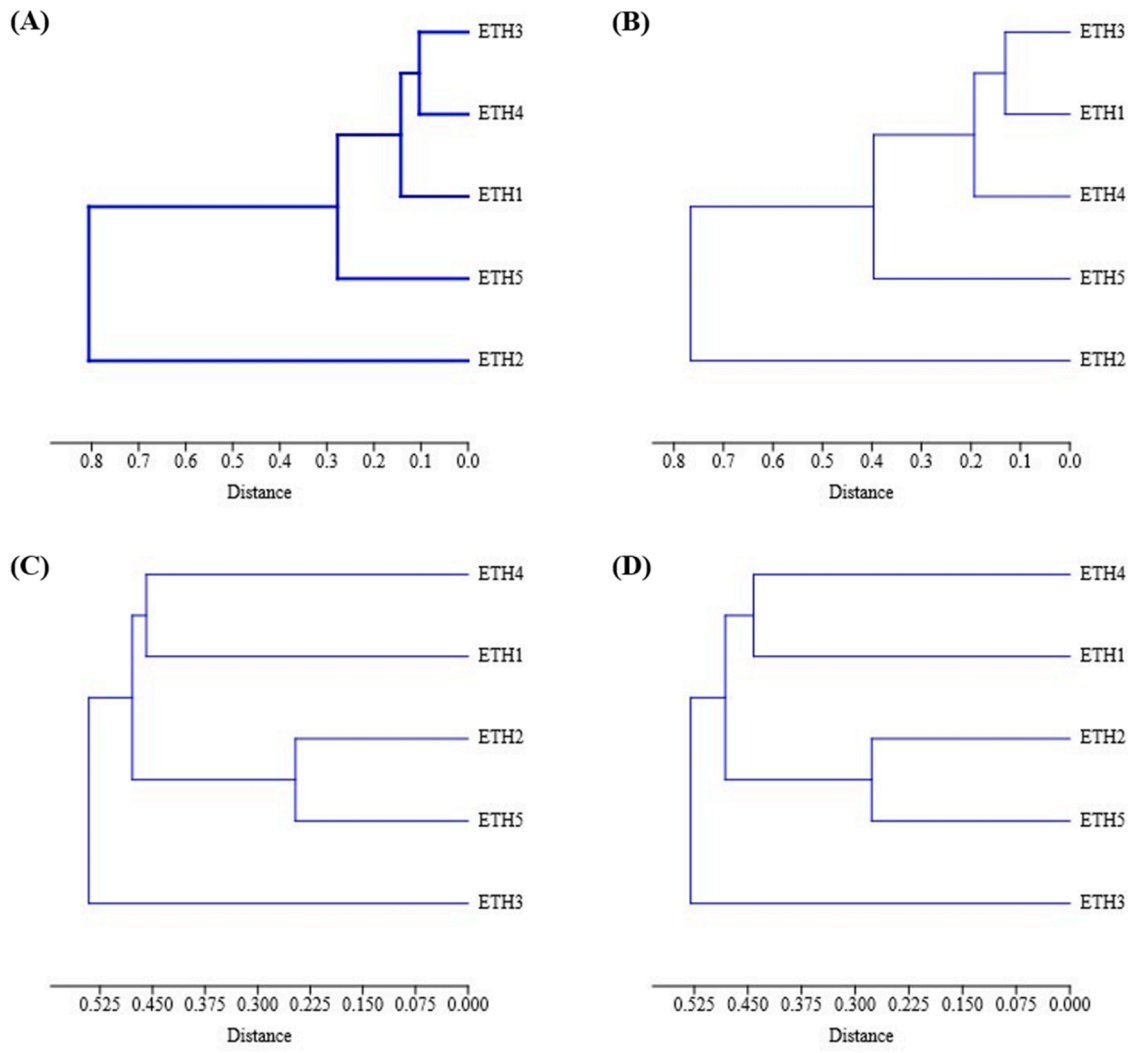

Fig. 2. Hierarchical clustering of total solvent accessible surface area (SASA) (A), surface and nucleus SASA (B), conserved region-SASA (C), all features (D) of *A. tolypomycina* extracellular triacylglycerol hydrolase.

Table 4
The correlation of conserved region SASA in *A. tolypomycina* extracellular triacylglycerol hydrolase.

ETHs	ETH1	ETH2	ETH 3	ETH 4	ETH 5
ETH1	1.000				
ETH2	0.764	1.000			
ETH3	0.718	0.646	1.000		
ETH4	0.797	0.749	0.714	1.000	
ETH5	0.747	0.934	0.667	0.746	1.000

both the Trp104Ala and Trp104Gln mutations [76]. Lafaquiere and colleagues, utilizing computational structural analysis and docking techniques, successfully achieved a 15-fold enhancement in the specific activity of *Burkholderia cepacia* ETH compared to the parental enzyme [26]. The redesign of the stereospecificity pocket in *Pseudozyma antarctica* ETH, predicted through molecular modeling, resulted in a final mutant that exhibited over a 30% increase in hydrolysis and transesterification activities [77].

Furthermore, rationally designed of the *Candida antarctica* ETH which expressed as the recombinant (mutant Asp122Leu) in *Escherichia coli* and *Pichia pastoris*, and purified, showed a significant decrease in the hydrolytic activity, thus reducing the side product yield during acylation [78]. It should be mentioned here that computational analysis and docking were applied to find the hot spot residues. In most research, one amino acid was mutated to achieve the target however the computational structural analysis could yield the result of landscape structural analysis and potential effect on protein function. Therefore, more than one amino acid can potentially mutated based on the analysis [79].

Numerous researchers have identified structural analysis and rational design as pivotal methods for pinpointing specific regions and residues—termed hot spots—for modifying ETHs to achieve desired targets. It is noteworthy that the results of these structural analyses reveal not only the residues responsible for enzyme stability but also those critical for specific functional activities [80–91]. The findings from our research also shed light on the intrinsic flexibility necessary for the motions of ETHs, as encoded in their three-dimensional structures. This understanding aids in determining the overall flexibility of ETHs, which is closely linked to the total surface area buried within their folds. Furthermore, this information can elucidate the relative solvent-accessible surface area (A_{rel}) of the enzymes [28].

The clustering ETHs with physiochemical and solvent accessible surface area beside conserved sequence and structure helped to find different aspects of ETH's residue-specific and conservation. The information provided in our research can be very helpful for simulation study of ETHs. Generally, laboratory experiments are costly and time consuming [92–94] thus computational study of protein sequence/structure could help to fulfil conservation structural study [95]. Machine

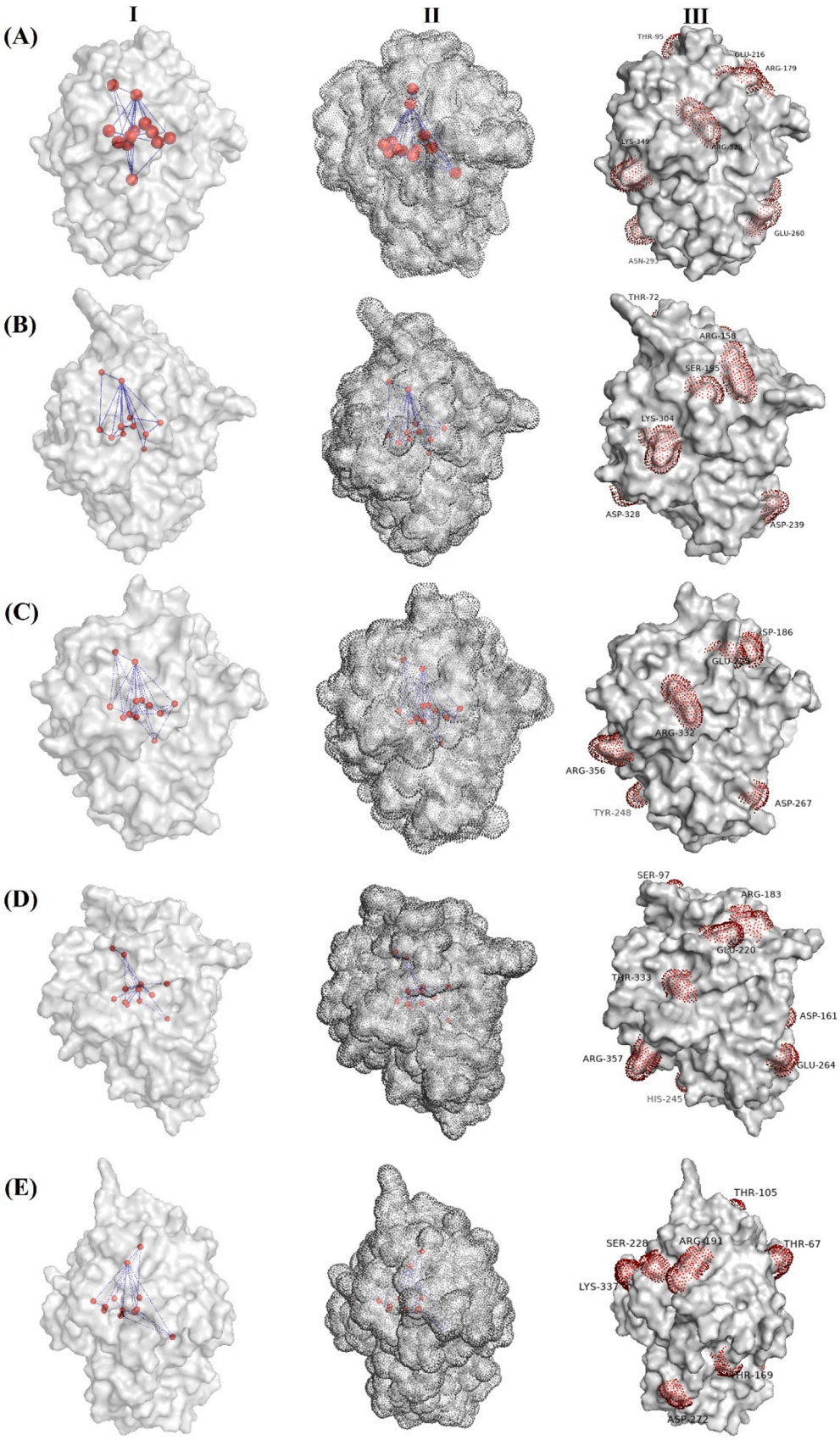

Fig. 3. The conserved residues with minimum SASA (zero) in the structure of *A. tolypomycina* Extracellular Triacylglycerol Hydrolase (I, II). The residues with maximum SASA in the structure of *A. tolypomycina* Extracellular Triacylglycerol Hydrolase (III). ETH1, ETH2, ETH3, ETH4 and ETH5 presented as (A), (B), (C), (D) and (E).

Table 5
The structural, sequence conserved SASA residues in *A. tolypomycina* extracellular triacylglycerol hydrolase.

ETH1		ETH2		ETH3		ETH4		ETH5	
GLY	50	GLY	25	GLY	57	GLY	47	GLY	58
PRO	93	PRO	70	MET	80	PRO	95	PRO	103
GLY	141	GLY	120	GLY	148	GLY	145	GLY	153
ASP	148	ASP	127	ASP	155	ASP	152	ASP	160
GLY	151	GLY	130	GLY	158	GLY	155	GLY	163
ASP	172	ASP	151	ASP	179	ASP	176	ASP	184
ALA	176	ALA	155	ALA	183	ALA	180	ALA	188
GLY	195	GLY	174	GLY	202	GLY	199	GLY	207
TYR	196	TYR	175	TYR	203	TYR	200	TYR	208
SER	197	SER	176	SER	204	SER	201	SER	209
GLN	198	GLN	177	GLN	205	GLN	202	GLN	210
GLY	199	GLY	178	GLY	206	GLY	203	GLY	211
GLY	200	GLY	179	GLY	207	GLY	204	GLY	212
GLY	225	GLY	204	GLY	232	GLY	229	GLY	237
PRO	327	PRO	306	PRO	334	PRO	335	PRO	339
ASP	336	ASP	315	ASP	343	ASP	344	ASP	348

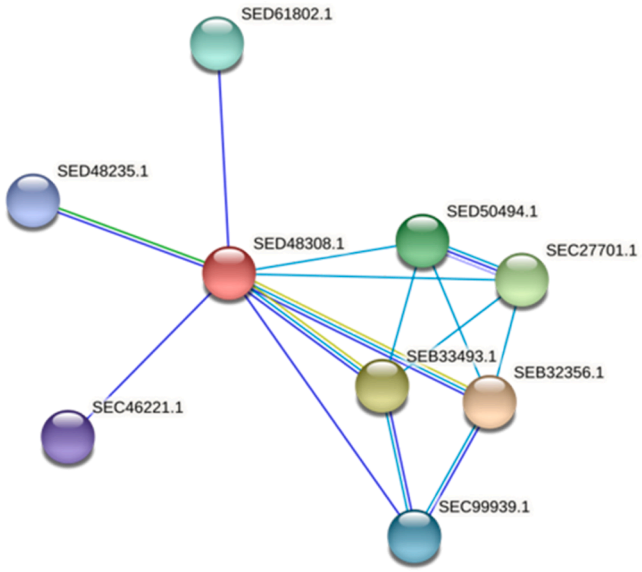

Fig. 4. Protein-protein interaction; SED48308.1: Extracellular Triacylglycerol Hydrolase, SEB32356.1: Lipase, SEB33493.1: Triacylglycerol esterase/lipase EstA, alpha/beta hydrolase fold, SEC27701.1: Long-chain acyl-CoA synthetase, SED50494.1: Long-chain acyl-CoA synthetase, SED61802.1: Site-specific DNA recombinase, SEC99939.1: S-formylglutathione hydrolase FrmB, SED48235.1: Uncharacterized protein, SEC46221.1: Regulatory protein, tetR family.

learning approaches recently provided great pathway to classify and deduce the protein structure for specific aims and applications [96]. In our study, these approaches effectively facilitated the clustering of ETHs and the identification of crucial segments and residues, paving the way for future laboratory experiments. The primary, secondary, tertiary, and quaternary structures of ETHs, derived from various bioinformatics analyses, particularly the three-dimensional structures, elucidate the complete functionality of these enzymes. The de novo modeling of ETHs demonstrated enhanced advantages when integrated with a homology model template. The identification of 16 key residues (GLY50, PRO93, GLY141, ASP148, GLY151, ASP172, ALA176, GLY195, TYR196, SER197, GLN198, GLY199, GLY200, GLY225, PRO327, ASP336) within the three-dimensional structure provides valuable insights into the functionality and stability of ETH structures. It is noteworthy that the majority of the amino acids exhibited zero solvent accessibility, categorizing them as interior residues. While numerous characteristics of amino acids and their three-dimensional structures can be utilized for protein analysis, [97] our research highlights that the solvent-accessible surface area, in conjunction with conserved structures and sequences, effectively delineates key residues critical for protein structural stability. In other words, the significance of the solvent-accessible surface area (SASA) features in relation to the functionality and stability of ETHs is clearly demonstrated by the presence of zero SASA residues, which form the foundational skeleton of each ETH structure. We have highlighted the zero SASA residues that are crucial for the stability of ETHs, specifically GLY112, GLY167, ALA169, GLY199, GLY200, ALA206, and ALA224. Notably, the conservation of glycine (GLY) and alanine (ALA) residues with zero SASA across all ETH structures is particularly remarkable. Utilizing SASA features, we identified the maximum SASA in each ETH structure, highlighting their potential significance for enzyme function. The high degree of similarity in amino acids and SASA values in both stability and function will facilitate successful mimicking and engineering of these enzymes. Consequently, our research advocates for a combined approach that leverages conserved -SASA to pinpoint species-specific residues in ETHs, which are likely to have a substantial impact on both stability and functionality.

The research on ETHs from *A. tolypomycina* as one of the species in *Amycolatopsis* genus and Actinobacteria phylum highlighted the importance of this genus again as the high guanine and cytosine content genome for industry and medicine [48,98]. Thus, besides the organic matter decomposing and important biomedicinally compounds researchers should scrutinize more suitable bioproduct in this genus [99–101]. Interestingly antifungal, antiviral and many pesticides produced from another species in this phylum, the *Streptomyces* spp. [102, 103]. The important of *Amycolatopsis* genus with production of aminoglycosides, anthracyclines, chloramphenicol, macrolide, tetracyclines compound opened the new opportunity to scrutinized the enzymes specially triacylglycerol hydrolases and particularly ETHs from the *A. tolypomycina* of this genus [100,102,104]. Introducing the new species such as *A. tolypomycina* or new structure of ETHs from *A. tolypomycina* could help to introduce them to industry for potential lipase production [99,105,106]. As clearly showed in phylogenetic tree the ETHs had very high similarity with other ETHs and in case of pathogenic microorganism the structural information here would help to combat the pathogenicity of species.

From a structural perspective, it is important to note that ETHs exhibit specificity towards multi-molecular aggregates of esters, such as micelles and emulsified particles [107]. Furthermore, soluble ETHs become activated through conformational changes induced by their interaction with oil-water interfaces [22]. Molecular packing, surface tension at the lipid-water interface, and the physicochemical properties of substrates can significantly influence the activity of ETHs [108]. The details presented in our research contribute to a clearer understanding of these mechanisms.

Three dimensional structures of many ETHs supported the conformational mechanism of actions hypothesis. The determination of crystal structure of THs started with human pancreas and the fungus *Rhizomucor miehei* lipases [109]. It was observed that catalytic triad (Ser--HisAsp/Glu) active site pocket was buried under the alpha loop and therefore prevented the substrate accession. This conformation was confirmed with complex X-ray crystal determination of human pancreatic lipase and colipase. Thus, the helical loop that acted as lid should remove to activate the enzyme [110].

The three-dimensional structures of complex ETHs and their substrates reveal a key characteristic that distinguishes ETHs from esterases: the presence of a lid domain that covers the catalytic active site. Notably, the transition from an inactive to an active state is contingent upon the interfacial binding of ETHs to the lipid-water interface. Our study provides valuable insights into the mechanisms of action of ETHs, facilitating their differentiation and clustering from esterases. While numerous three-dimensional structures of ETHs and esterases from both animal and microbial origins have been elucidated, no structure from *A. tolypomycina* has been resolved to date.

It is noteworthy that numerous researchers have endeavored to

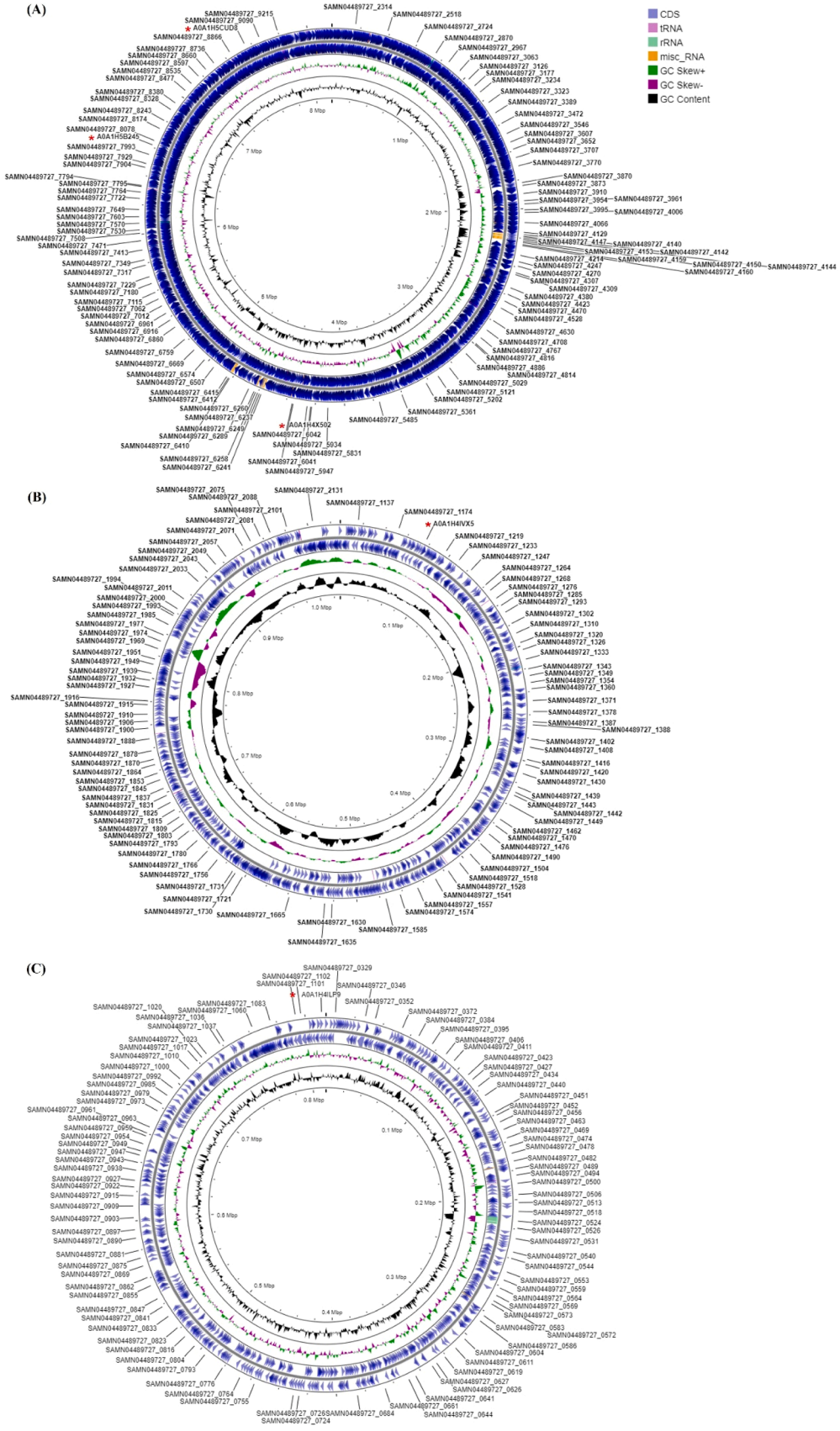

Fig. 5. Full genome presentations of *A. tolypomycina* sequence datasets (GenBank: FNSO01000004 (A), FNSO01000003 (B), FNSO01000002 (C), ETH highlighted with red asterisks.

identify distinct features for clustering esterases and extracellular triacylglycerol hydrolases (ETHs). Certain ETHs exhibited a lid-like structure, while intriguingly, some of these enzymes even displayed a peptide loop that functions as a mini lid [111]. The no-lid structure was further analyzed using tributyrin and other substrates in laboratory experiments. However, some ETHs lacked a lid structure altogether, and for certain enzymes, the lid appeared to have no clear role in interfacial activation, as exemplified by the pancreatic ETH from *Myocastor coypu*, which features a 23-residue lid [112]. Then these characteristics (interfacial activation and the presence of the lid domain) cannot differentiate between the ETH and esterases. In our research the potentiality of 16 keys residues for ETHs as the key feature was presented in *A. tolypomycina*. However other methods introduced based on pH and active site residues [113]. The active site and pH couldn't clearly answer the clustering carboxylesterases (ETHs, esterases) [114].

5. Conclusions

In conclusion, the computational structural analysis of extracellular triacylglycerol hydrolases (ETHs) derived from the high-GC bacterium *A. tolypomycina* has been elucidated. This study highlights the physicochemical properties and conserved solvent accessibility of specific residues that significantly influence the stability and conformation of these enzymes. The specific amino acids identified through computational analysis can inform the design of subsequent wet laboratory experiments. Furthermore, this computational investigation contributes additional evidence regarding the evolutionary dynamics and clustering of ETHs. Ultimately, the insights gained from solvent accessibility and the structural-sequence conservation in the three-dimensional models provide critical features for understanding the evolutionary trajectory of ETHs within the *A. tolypomycina* dataset.

Funding

This work was supported by Mahidol University, Basic Research Fund: fiscal year 2021 [grant numbers BRF1-004/2564] and Suan Sunandha Rajabhat University [grant number งป11773/2566].

CRediT authorship contribution statement

Supajit Sraphet: Writing – original draft, Investigation, Funding acquisition, Formal analysis. **Bagher Javadi:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the Faculty of Science and Technology, Suan Sunandha Rajabhat University and Mahidol university, Thailand.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.btre.2024.e00869.

Data availability

No data was used for the research described in the article.

References

[1] L.A. Abriata, M. Dal Peraro, State-of-the-art web services for de novo protein structure prediction, Brief. Bioinform. 22 (2021) bbaa139.
[2] F. Pucci, M. Schwersensky, M. Rooman, Artificial intelligence challenges for predicting the impact of mutations on protein stability, Curr. Opin. Struct. Biol. 72 (2022) 161–168.
[3] M. Veit-Acosta, A. de, W.F. Junior, Computational prediction of binding affinity for cdk2-ligand complexes. A protein target for cancer drug discovery, Curr. Med. Chem. 29 (2022) 2438–2455.
[4] G.O. Almutairi, A. Malik, M. Alonazi, J.M. Khan, A.S. Alhomida, M.S. Khan, et al., Expression, purification, and biophysical characterization of recombinant mers-cov main (mpro) protease, Int. J. Biol. Macromol. 209 (2022) 984–990.
[5] B. Banaganapalli, C.M. Anuradha, C. Mulakayala, G. D, S. Chitta, *In silico* structural characterization of *mycobacterium tuberculosis* h37rv udp-n-acetylmuramate dehydrogenase, Int. J. Integr. Biol. 6 (2009) 12–16.
[6] S.V. Reshma, N. Sathyanarayanan, H.G. Nagendra, Characterization of hypothetical protein vng0128c from *halobacterium* nrc-1 reveals gale like activity and its involvement in leloir pathway of galactose metabolism, J. Biomol. Struct. Dyn. 33 (2015) 1743–1755.
[7] Z. Shafat, A. Hamza, A. Islam, M.S. Al-Dosari, M.K. Parvez, S. Parveen, Structural exploration of y-domain reveals its essentiality in hev pathogenesis, Protein Express. Purif. 187 (2021) 105947.
[8] S. Ali, S.A. Khan, M. Hamayun, I.J. Lee, The recent advances in the utility of microbial lipases: a review, Microorganisms. 11 (2023) 510.
[9] A. Gácser, F. Stehr, C. Kröger, L. Kredics, W. Schäfer, J.D. Nosanchuk, Lipase 8 affects the pathogenesis of *candida albicans*, Infect. Immun. 75 (2007) 4710–4718.
[10] M.G. Paraje, S.G. Correa, M.S. Renna, M. Theumer, C.E. Sotomayor, Candida albicans-secreted lipase induces injury and steatosis in immune and parenchymal cells, Can. J. Microbiol. 54 (2008) 647–659.
[11] D.A. Schofield, C. Westwater, T. Warner, E. Balish, Differential *candida albicans* lipase gene expression during alimentary tract colonization and infection, FEMS Microbiol. Lett. 244 (2005) 359–365.
[12] R. Dumache, A. Enache, I. Macasoi, C.A. Dehelean, V. Dumitrascu, A. Mihailescu, et al., Sars-cov-2: an overview of the genetic profile and vaccine effectiveness of the five variants of concern, Pathogens. 11 (2022) 516.
[13] M. Santibáñez-Andrade, Y.I. Chirino, I. González-Ramírez, Y. Sánchez-Pérez, C. M. García-Cuellar, Deciphering the code between air pollution and disease: The effect of particulate matter on cancer hallmarks, Int. J. Mol. Sci. 21 (2019) 136.
[14] C. Tran, I.E. Cock, X. Chen, Y. Feng, Antimicrobial *bacillus*: metabolites and their mode of action, Antibiot. ChemOther. 11 (2022) 88.
[15] T. Vanzolini, M. Bruschi, A.C. Rinaldi, M. Magnani, A. Fraternale, Multitalented synthetic antimicrobial peptides and their antibacterial, antifungal and antiviral mechanisms, Int. J. Mol. Sci. 23 (2022) 545.
[16] C. Zhang, M. Yang, Antimicrobial peptides: From design to clinical application, Antibiot. ChemOther. 11 (2022) 349.
[17] M. Holtof, C. Lenaerts, D. Cullen, J. Vanden Broeck, Extracellular nutrient digestion and absorption in the insect gut, Cell Tissue Res. 377 (2019) 397–414.
[18] M. Lübeck, P.S. Lübeck, Fungal cell factories for efficient and sustainable production of proteins and peptides, Microorganisms. 10 (2022) 753.
[19] D.L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S.M. Franken, et al., The α/β hydrolase fold, Protein Eng. Design Select. 5 (1992) 197–211.
[20] J.L. Sussman, M. Harel, F. Frolow, C. Oefner, A. Goldman, L. Toker, et al., Atomic structure of acetylcholinesterase from *torpedo californica*: a prototypic acetylcholine-binding protein, Science 253 (1991) 872–879.
[21] Y. Bourne, C. Martinez, B. Kerfelec, D. Lombardo, C. Chapus, C. Cambillau, Horse pancreatic lipase. The crystal structure refined at 2.3 a resolution, J. Mol. Biol. 238 (1994) 709–732.
[22] J. Pliego, J.C. Mateos, J. Rodriguez, F. Valero, M. Baeza, R. Femat, et al., Monitoring lipase/esterase activity by stopped flow in a sequential injection analysis system using *p*-nitrophenyl butyrate, Sensors 15 (2015) 2798–2811.
[23] T. Iftikhar, M. Niaz, E. Ali, R. Jabeen, R. Abdullah, Production process of extracellular lipases by *fusarium* sp. Using agricultural by products, Pak. J. Bot. 44 (2012) 335–339.
[24] A. Kumar, S. Parihar, N. Batra, Enrichment, isolation and optimization of lipase-producing *staphylococcus* sp. From oil mill waste (oil cake), J. Exp. Sci. 3 (2012) 26–30.
[25] A. Suyambu, R. Ramasubburayan, A. Palavesam, G. Immanuel, Optimization and purification of lipase through solid state fermentation by bacillus cereus msu as isolated from the gut of a marine fish *sardinella longiceps*, Int. J. Pharm. Sci. 6 (2014) 291–298.
[26] V. Lafaquière, S. Barbe, S. Puech-Guenot, D. Guieysse, J. Cortés, P. Monsan, et al., Control of lipase enantioselectivity by engineering the substrate binding site and access channel, Chembiochem. 10 (2009) 2760–2771.
[27] N.M. Antikainen, S.F. Martin, Altering protein specificity: Techniques and applications, Bioorg. Med. Chem. 13 (2005) 2701–2716.
[28] J.A. Marsh, Buried and accessible surface area control intrinsic protein flexibility, J. Mol. Biol. 425 (2013) 3250–3263.
[29] G.K. Meghwanshi, S. Verma, V. Srivastava, R. Kumar, Archaeal lipolytic enzymes: Current developments and further prospects, Biotechnol. Adv. 61 (2022) 108054.
[30] J. Damborsky, J. Brezovsky, Computational tools for designing and engineering enzymes, Curr. Opin. Chem. Biol. 19 (2014) 8–16.
[31] S. Sraphet, B. Javadi, Computational characterizations of gdp-mannose 4,6-dehydratase (noel) rhizobial proteins, Curr. Genet. 67 (2021) 769–784.

[32] F. García-Guevara, M. Avelar, M. Ayala, L. Segovia, Computational tools applied to enzyme design – a review, Biocatalysis 1 (2016) 109–117.
[33] D. Agashe, N. Shankar, The evolution of bacterial DNA base composition, J. Experim. Zool. Part B Mol. Develop. Evol. 322 (2014) 517–528.
[34] S. Aslam, X.-R. Lan, B.-W. Zhang, Z.-L. Chen, L. Wang, D.-K. Niu, Aerobic prokaryotes do not have higher gc contents than anaerobic prokaryotes, but obligate aerobic prokaryotes have, BMC. Evol. Biol. 19 (2019) 35.
[35] J. Bohlin, L. Snipen, S.P. Hardy, A.B. Kristoffersen, K. Lagesen, T. Dønsvik, et al., Analysis of intra-genomic gc content homogeneity within prokaryotes, BMC. Genom. 11 (2010) 464.
[36] A.-K. Dietel, H. Merker, M. Kaltenpoth, C. Kost, Selective advantages favour high genomic at-contents in intracellular elements, PLoS. Genet. 15 (2019) e1007778.
[37] K.U. Foerstner, C. von Mering, S.D. Hooper, P. Bork, Environments shape the nucleotide composition of genomes, EMBO Rep. 6 (2005) 1208–1213.
[38] S. Glémin, Y. Clément, J. David, A. Ressayre, Gc content evolution in coding regions of angiosperm genomes: a unifying hypothesis, Trends. Genet. 30 (2014) 263–270.
[39] F. Hildebrand, A. Meyer, A. Eyre-Walker, Evidence of selection upon genomic gc-content in bacteria, PLoS. Genet. 6 (2010) e1001107.
[40] R. Kogay, Y.I. Wolf, E.V. Koonin, O. Zhaxybayeva, Selection for reducing energy cost of protein production drives the gc content and amino acid composition bias in gene transfer agents, mBio 11 (2020) e01206–e01220.
[41] F. Lassalle, S. Périan, T. Bataillon, X. Nesme, L. Duret, V. Daubin, Gc-content evolution in bacterial genomes: the biased gene conversion hypothesis expands, PLoS. Genet. 11 (2015) e1004941.
[42] S. Mann, Y.-P.P. Chen, Bacterial genomic g+c composition-eliciting environmental adaptation, Genomics 95 (2010) 7–15.
[43] R. Raghavan, Y.D. Kelkar, H. Ochman, A selective force favoring increased g+c content in bacterial genes, Proc. Natl Acad. Sci. 109 (2012) 14504–14507.
[44] E.R. Reichenberger, G. Rosen, U. Hershberg, R. Hershberg, Prokaryotic nucleotide composition is shaped by both phylogeny and the environment, Genome Biol. Evol. 7 (2015) 1380–1389.
[45] J.L. Weissman, W.F. Fagan, P.L.F. Johnson, Linking high gc content to the repair of double strand breaks in prokaryotic genomes, PLoS. Genet. 15 (2019) e1008493.
[46] H. Wu, Z. Zhang, S. Hu, J. Yu, On the molecular mechanism of gc content variation among eubacterial genomes, Biol. Direct. 7 (2012) 2.
[47] I. Nouioui, L. Carro, M. García-López, J.P. Meier-Kolthoff, T. Woyke, N. C. Kyrpides, et al., Genome-based taxonomic classification of the phylum *actinobacteria*, Front. Microbiol. 9 (2018) 2007.
[48] M. Ventura, C. Canchaya, A. Tauch, G. Chandra, G.F. Fitzgerald, K.F. Chater, et al., Genomics of *actinobacteria*: tracing the evolutionary history of an ancient phylum, Microbiol. Mol. Biol. Rev. 71 (2007) 495–548.
[49] E. Gasteiger, C. Hoogland, A. Gattiker, D. Se, M.R. Wilkins, R.D. Appel, et al., Protein identification and analysis tools on the expasy server. The proteomics protocols handbook, J. M. Walker. Totowa (2005) 571–607. NJ.
[50] F. Corpet, Multiple sequence alignment with hierarchical clustering, Nucleic. Acids. Res. 16 (1988) 10881–10890.
[51] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, Mega x: molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549.
[52] C.S. Yu, C.J. Lin, J.K. Hwang, Predicting subcellular localization of proteins for gram-negative bacteria by support vector machines based on n-peptide compositions, Protein Sci. 13 (2004) 1402–1406.
[53] C.S. Yu, Y.C. Chen, C.H. Lu, J.K. Hwang, Prediction of protein subcellular localization, Proteins. 64 (2006) 643–651.
[54] T.A. Kumar, Cfssp: Chou and fasman secondary structure prediction server, Wide Spectr. 1 (2013) 15–19.
[55] J. Yang, I. Anishchenko, H. Park, Z. Peng, S. Ovchinnikov, D. Baker, Improved protein structure prediction using predicted interresidue orientations, Proc. Natl Acad. Sci. 117 (2020) 1496–1503.
[56] G.N. Ramachandran, V. Sasisekharan, Conformation of polypeptides and proteins, Adv. Protein Chem. 23 (1968) 283–438.
[57] R. Fraczkiewicz, W. Braun, Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules, J. Comput. Chem. 19 (1998) 319–333.
[58] B. Rost, C. Sander, Conservation and prediction of solvent accessibility in protein families, Proteins. 20 (1994) 216–226.
[59] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E. C. Meng, et al., Ucsf chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.
[60] H. Wu, L. Gao, J. Dong, X. Yang, Detecting overlapping protein complexes by rough-fuzzy clustering in protein-protein interaction networks, PLoS. One 9 (2014) e91856.
[61] K. Inoue, W. Li, H. Kurata, Diffusion model based spectral clustering for protein-protein interaction networks, PLoS. One 5 (2010) e12623.
[62] H. Do, J.H. Lee, M.H. Kwon, H.E. Song, J.Y. An, S.H. Eom, et al., Purification, characterization and preliminary x-ray diffraction analysis of a cold-active lipase (cpslip) from the psychrophilic bacterium *colwellia psychrerythraea* 34h, Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. 69 (2013) 920–924.
[63] W. Yang, Y. He, L. Xu, H. Zhang, Y. Yan, A new extracellular thermo-solvent-stable lipase from *burkholderia ubonensis* sl-4: identification, characterization and application for biodiesel production, J. Mol. Catal. B: Enzym. 126 (2016) 76–89.
[64] V.W. Dolinsky, D. Gilham, M. Alam, D.E. Vance, R. Lehner, Triacylglycerol hydrolase: Role in intracellular lipid metabolism, Cell Mol. Life Sci. 61 (2004) 1633–1651.
[65] A.D. Quiroga, R. Lehner, Role of endoplasmic reticulum neutral lipid hydrolases, Trends. Endocrinol. Metab. 22 (2011) 218–225.
[66] V. Nagaroor, S.N. Gummadi, An overview of mammalian and microbial hormone-sensitive lipases (lipolytic family iv): biochemical properties and industrial applications, Biotechnol. Genet. Eng. Rev. (2022) 1–30.
[67] S. Yaginuma, H. Kawana, J. Aoki, Current knowledge on mammalian phospholipase a(1), brief history, structures, biochemical and pathophysiological roles, Molecules. 27 (2022).
[68] F. Akram, A.S. Mir, H. Iu, A. Roohi, An appraisal on prominent industrial and biotechnological applications of bacterial lipases, Mol. Biotechnol. 65 (2023) 521–543.
[69] G. Li, X. Fang, F. Su, Y. Chen, L. Xu, Y. Yan, Enhancing the thermostability of *rhizomucor miehei* lipase with a limited screening library by rational-design point mutations and disulfide bonds, Appl. Environ. Microbiol. 84 (2018).
[70] M.T. Reetz, J.D. Carballeira, A. Vogel, Iterative saturation mutagenesis on the basis of b factors as a strategy for increasing protein thermostability, Angew. Chem. Int. Ed. Engl. 45 (2006) 7745–7751.
[71] H.J. Park, J.C. Joo, K. Park, Y.J. Yoo, Stabilization of *candida antarctica* lipase b in hydrophilic organic solvent by rational design of hydrogen bond, Biotechnol. Bioprocess Eng. 17 (2012) 722–728.
[72] Q.A. Le, J.C. Joo, Y.J. Yoo, Y.H. Kim, Development of thermostable *candida antarctica* lipase b through novel in silico design of disulfide bridge, Biotechnol. Bioeng. 109 (2012) 867–876.
[73] S.V. Sharma, S. Hooda, An analysis upon various strategies for redesign and direct evolution of enzyme engineering, Adv. Sci. Technol. 13 (2017) 144–150.
[74] Zaugg J., Gumulya Y., Gillam E.M.J., Bodén M., 2014. Computational tools for directed evolution: a comparison of prospective and retrospective strategies. Directed Evolution Library creation: Methods and Protocols. E. M. J. Gillam, J. N. Copp and D. Ackerley. New York, Springer New York: 315–333.
[75] D. Rotticci, J.C. Rotticci-Mulder, S. Denman, T. Norin, K. Hult, Improved enantioselectivity of a lipase by rational protein engineering, Chembiochem. 2 (2001) 766–770.
[76] A.O. Magnusson, J.C. Rotticci-Mulder, A. Santagostino, K. Hult, Creating space for large secondary alcohols by rational redesign of *candida antarctica* lipase b, Chembiochem. 6 (2005) 1051–1056.
[77] D. Liu, P. Trodler, S. Eiben, K. Koschorreck, M. Müller, J. Pleiss, et al., Rational design of *pseudozyma antarctica* lipase b yielding a general esterification catalyst, Chembiochem. 11 (2010) 789–795.
[78] J. Müller, M.A. Sowa, B. Fredrich, H. Brundiek, U.T. Bornscheuer, Enhancing the acyltransferase activity of *candida antarctica* lipase a by rational design, Chembiochem. 16 (2015) 1791–1796.
[79] E. Durmaz, S. Kuyucak, U.O. Sezerman, Modifying the catalytic preference of tributyrin in *bacillus thermocatenulatus* lipase through in-silico modeling of enzyme-substrate complex, Protein Eng. Des. Sel. 26 (2013) 325–333.
[80] W. Tang, D. Lan, Z. Zhao, S. Li, X. Li, Y. Wang, A thermostable monoacylglycerol lipase from marine *geobacillus* sp. 12amor1: biochemical characterization and mutagenesis study, Int. J. Mol. Sci. 20 (2019).
[81] I. Oroz-Guinea, K. Zorn, U.T. Bornscheuer, Enhancement of lipase cal-a selectivity by protein engineering for the hydrolysis of erucic acid from *crambe* oil, Eur. J. Lipid Sci. Technol. 122 (2020) 1900115.
[82] K. Zorn, I. Oroz-Guinea, H. Brundiek, M. Dörr, U.T. Bornscheuer, Alteration of chain length selectivity of *candida antarctica* lipase a by semi-rational design for the enrichment of erucic and gondoic fatty acids, Adv. Synth. Catal. 360 (2018) 4115–4131.
[83] X.W. Yu, S.S. Zhu, R. Xiao, Y. Xu, Conversion of a *rhizopus chinensis* lipase into an esterase by lid swapping, J. Lipid Res. 55 (2014) 1044–1051.
[84] G. Santarossa, P.G. Lafranconi, C. Alquati, L. DeGioia, L. Alberghina, P. Fantucci, et al., Mutations in the "lid" region affect chain length specificity and thermostability of a *pseudomonas fragi* lipase, FEBS Lett. 579 (2005) 2383–2386.
[85] J. Damnjanović, H. Nakano, Y. Iwasaki, Acyl chain that matters: Introducing sn-2 acyl chain preference to a phospholipase d by protein engineering, Protein Eng. Des. Sel. 32 (2019) 1–11.
[86] X.W. Yu, N.J. Tan, R. Xiao, Y. Xu, Engineering a disulfide bond in the lid hinge region of *rhizopus chinensis* lipase: Increased thermostability and altered acyl chain length specificity, PLoS. One 7 (2012) e46388.
[87] N. Willems, M. Lelimousin, J. Skjold-Jørgensen, A. Svendsen, M.S.P. Sansom, The effect of mutations in the lid region of *thermomyces lanuginosus* lipase on interactions with triglyceride surfaces: a multi-scale simulation study, Chem. Phys. Lipids 211 (2018) 4–15.
[88] Brundiek H., Padhi S.K., Kourist R., Evitt A., Bornscheuer U.T. (2012) Altering the scissile fatty acid binding site of *candida antarctica* lipase a by protein engineering for the selective hydrolysis of medium chain fatty acids. 114: 1148–1153.
[89] S.K. Padhi, R. Fujii, G.A. Legatt, S.L. Fossum, R. Berchtold, R.J. Kazlauskas, Switching from an esterase to a hydroxynitrile lyase mechanism requires only two amino acid substitutions, Chem. Biol. 17 (2010) 863–871.
[90] F. Secundo, G. Carrea, C. Tarabiono, P. Gatti-Lafranconi, S. Brocca, M. Lotti, et al., The lid is a structural and functional determinant of lipase activity and selectivity, J. Mol. Catal. B: Enzym. 39 (2006) 166–170.
[91] R.R. Klein, G. King, R.A. Moreau, M.J. Haas, Altered acyl chain length specificity of *rhizopus delemar* lipase through mutagenesis and molecular modeling, Lipids 32 (1997) 123–130.
[92] S. Javed, F. Azeem, S. Hussain, I. Rasul, M.H. Siddique, M. Riaz, et al., Bacterial lipases: a review on purification and characterization, Prog. Biophys. Mol. Biol. 132 (2018) 23–34.
[93] P. Chandra, S.R. Enespa, P.K. Arora, Microbial lipases and their industrial applications: a comprehensive review, Microb. Cell Factories 19 (2020) 169.

[94] N.B. Melani, E.B. Tambourgi, E. Silveira, Lipases: From production to applications, Separ. Purif. Rev. 49 (2020) 143–158.
[95] B. Javadi, In silico characterization of lipase architectural structure in *rhizobium leguminosarum*, Plant Cell Biotechnol. Mol. Biol. 21 (2020) 14–26.
[96] S. Sraphet, B. Javadi, Application of hierarchical clustering to analyze solvent-accessible surface area patterns in *amycolatopsis lipases*, Biology 11 (2022) 652.
[97] N.A. Petushkova, M.A. Pyatnitskiy, V.A. Rudenko, O.V. Larina, O.P. Trifonova, J. S. Kisrieva, et al., Applying of hierarchical clustering to analysis of protein patterns in the human cancer-associated liver, PLoS. One 9 (2014) e103950.
[98] Z. Song, T. Xu, J. Wang, Y. Hou, C. Liu, S. Liu, et al., Secondary metabolites of the genus *amycolatopsis*: Structures, bioactivities and biosynthesis, Molecules. 26 (2021) 1884.
[99] S. Uttatree, P. Winayanuwattikun, J. Charoenpanich, Isolation and characterization of a novel thermophilic-organic solvent stable lipase from *acinetobacter baylyi*, Appl. Biochem. Biotechnol. 162 (2010) 1362–1376.
[100] M. Verma, J. Kaur, M. Kumar, K. Kumari, A. Saxena, S. Anand, et al., Whole genome sequence of the rifamycin b-producing strain *amycolatopsis mediterranei* s699, J. Bacteriol. 193 (2011) 5562–5563.
[101] C. Peano, F. Damiano, M. Forcato, A. Pietrelli, C. Palumbo, G. Corti, et al., Comparative genomics revealed key molecular targets to rapidly convert a reference rifamycin-producing bacterial strain into an overproducer by genetic engineering, Metab. Eng. 26 (2014) 1–16.
[102] M. Sharma, P. Dangi, M. Choudhary, Actinomycetes: Source, identification, and their applications, Int. J. Curr. Microbiol. App. Sci. 3 (2014) 801–832.
[103] R. Kumari, P. Singh, R. Lal, Genetics and genomics of the genus *amycolatopsis*, Indian J. Microbiol. 56 (2016) 233–246.
[104] W. Zhao, Y. Zhong, H. Yuan, J. Wang, H. Zheng, Y. Wang, et al., Complete genome sequence of the rifamycin sv-producing *amycolatopsis mediterranei* u32 revealed its genetic characteristics in phylogeny and metabolism, Cell Res. 20 (2010) 1096–1108.
[105] A.E. Todd, C.A. Orengo, J.M. Thornton, Evolution of function in protein superfamilies, from a structural perspective, J. Mol. Biol. 307 (2001) 1113–1143.
[106] X. Zheng, X. Chu, W. Zhang, N. Wu, Y. Fan, A novel cold-adapted lipase from *acinetobacter* sp. Xmz-26: gene cloning and characterisation, Appl. Microbiol. Biotechnol. 90 (2011) 971–980.
[107] A. Baldessari, L. Iglesias, Lipases and phospholipases, Methods and Protocols, Humana Press, New York, 2012.
[108] H. Chahinian, L. Sarda, Distinction between esterases and lipases: Comparative biochemical properties of sequence-related carboxylesterases, Protein Pept. Lett. 16 (2009) 1149–1161.
[109] Z.S. Derewenda, U. Derewenda, G.G. Dodson, The crystal and molecular structure of the *rhizomucor miehei* triacylglyceride lipase at 1.9 a resolution, J. Mol. Biol. 227 (1992) 818–839.
[110] H. van Tilbeurgh, M.P. Egloff, C. Martinez, N. Rugani, R. Verger, C. Cambillau, Interfacial activation of the lipase-procolipase complex by mixed micelles revealed by x-ray crystallography, Nature 362 (1993) 814–820.
[111] J. Uppenberg, N. Oehrner, M. Norin, K. Hult, G.J. Kleywegt, S. Patkar, et al., Crystallographic and molecular-modeling studies of lipase b from *candida antarctica* reveal a stereospecificity pocket for secondary alcohols, Biochemistry 34 (1995) 16838–16851.
[112] K. Thirstrup, R. Verger, F. Carrière, Evidence for a pancreatic lipase subfamily with new kinetic properties, Biochemistry 33 (1994) 2748–2756.
[113] A. Aloulou, J.A. Rodriguez, S. Fernandez, D. van Oosterhout, D. Puccinelli, F. Carrière, Exploring the specific features of interfacial enzymology based on lipase studies, Biochim. Biophys. Acta 1761 (2006) 995–1013.
[114] U.T. Bornscheuer, Microbial carboxyl esterases: classification, properties and application in biocatalysis, FEMS Microbiol. Rev. 26 (2002) 73–81.