International Journal of Biological Macromolecules 322 (2025) 146696

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

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

Mining the whole genome sequence of *Streptomyces cyaneofuscatus* strain CTM50504 isolated from the Aïn El-Atrous hot spring, Tunisia, for the discovery of extremozymes: Promising properties of protease activity

Sondes Mechri [a,b,c], Séverine Croze [b], Imen Rekik [a], Fawzi Allala [d], Fakher Frikha [e], Alexandre Noiriel [c], Marilize Le Roes-Hill [f], Abdelkarim Abousalham [c], Slim Tounsi [g], Joel Lachuer [b,h,*], Bassem Jaouadi [a,**]

[a] *Laboratory of Microbial and Enzymatic Biotechnologies and Biomolecules (LMEBB), Centre of Biotechnology of Sfax (CBS), University of Sfax (USF), Road of Sidi Mansour Km 6, P.O. Box 1177, Sfax 3018, Tunisia*

[b] *Univ Lyon, Université Claude Bernard Lyon 1 (UCBL), Plateforme ProfileXpert de Génomique et Microgénomique, SFR Santé-Lyon-Est, CNRS UMR-S3453, INSERM US7, Faculté de Pharmacie de Lyon, 8 avenue Rockefeller, 69373, Lyon, Cedex 08, France*

[c] *Univ Lyon, Université Claude Bernard Lyon 1 (UCBL), Institut de Chimie et de Biochimie Moléculaires et Supramoléculaires (ICBMS), UMR 5246 CNRS, Génie Enzymatique, Membranes Biomimétiques et Assemblages Supramoléculaires (GEMBAS), Bât Raulin, 43 Bd du 11 Novembre 1918, F-69622, Villeurbanne, Cedex, France*

[d] *Laboratoire de Biologie Cellulaire et Moléculaire (LBCM), Équipe de Microbiologie, Faculté des Sciences Biologiques, Université des Sciences et de la Technologie Houari Boumediene (USTHB), Alger, Algérie, Algiers 16111, Algeria*

[e] *Laboratory of Molecular and Cellular Screening Processes (LMCSP), Centre of Biotechnology of Sfax (CBS), University of Sfax (USF), Road of Sidi Mansour Km 6, P.O. Box 1177, Sfax 3018, Tunisia*

[f] *Applied Microbial and Health Biotechnology Institute (AMHBI), Cape Peninsula University of Technology (CPUT), P.O. Box 1906, Bellville 7535, South Africa*

[g] *Laboratory of Biopesticides (LB), Centre of Biotechnology of Sfax (CBS), University of Sfax (USF), Road of Sidi Mansour Km 6, P.O. Box 1177, Sfax 3018, Tunisia*

[h] *Univ Lyon, Université Claude Bernard Lyon 1 (UCBL), UMR INSERM 1052, CNRS 5286, Centre Léon Bérard, Centre de Recherche en Cancérologie de Lyon (CRCL), 28 Rue Laennec, 69008, Lyon, Cedex 08, France*

ARTICLE INFO

Keywords:
Whole-genome sequence
Genome assembly and annotation
Bioinformatic analyses
Protease
Homology modelling
Molecular dynamics simulations

ABSTRACT

Next-generation sequencing (NGS) methods allow for the generation of data leading to a greater understanding of the potential functionality and dynamics of microorganisms within their biotopes. The enormous volume of data that NGS produces necessitates understanding structural and functional genomics through the application of various omics techniques. *Streptomyces cyaneofuscatus* CTM50504 is a potential extracellular hydrolase producer isolated from a terrestrial hot spring, Aïn El-Atrous, Korbous (Nabeul, Tunisia). This strain can grow at 50 °C and a pH range of 6–9. It requires the presence of NaCl for growth and secretes proteases, lipases, phospholipases, amylases, and chitinases. Whole-genome sequence (WGS) analysis was performed on strain CTM50504 to identify protein-encoding genes, including hydrolases. The genome sequence was assembled into 858-contigs with an average G + C content of 71 % and a total length of 8,591,922-bp. Genome annotation revealed 770-protein-coding genes with 323 open reading frames encoding hydrolases, including 179-proteases, 20-lipases, 10-phospholipases, 5-amylases, and 5-chitinases. The gene encoding a serine alkaline protease (SCKP) was heterologously expressed in *Escherichia coli*. The recombinant SCKP (rSCKP) was purified by affinity chromatography, and its biochemical properties were determined. Molecular dynamics simulations provided deeper insights into how key amino acids contribute to substrate binding and elucidated the basis of substrate selectivity.

* Correspondence to: J. Lachuer, Univ Lyon, Université Claude Bernard Lyon 1 (UCBL), Plateforme ProfileXpert de Génomique et Microgénomique, SFR Santé-Lyon-Est, CNRS UMR-S3453, INSERM US7, Faculté de Pharmacie de Lyon, 8 avenue Rockefeller, 69373, Lyon, Cedex 08, France.

** Corresponding author.

E-mail addresses: joel.lachuer@univ-lyon1.fr (J. Lachuer), bassem.jaouadi@cbs.rnrt.tn (B. Jaouadi).

https://doi.org/10.1016/j.ijbiomac.2025.146696
Received 13 May 2025; Received in revised form 16 July 2025; Accepted 6 August 2025
Available online 9 August 2025
0141-8130/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Even though extremophiles and their associated enzymes are useful in industrial applications, comprehensive studies on their genomic adaptations are lacking [1,2]. This is surprising because genomic adaptations are often crucial to basic and applied investigations. Selected genomics-based studies have shown that microorganisms living in extreme environments restructure their genomes using various approaches such as insertion, expansion, or reduction of genome size; gene reshuffling through displacements and gene reorganization; G + C skewness in the genome; horizontal gene transfer; change in polyploidy level; and a preference for specific codon usage in genes whose products are required during adaptations to inhospitable ecological biotopes [3]. Therefore, genomics-based studies of prokaryotic extremophiles are needed to understand survival in extreme environments [4]. Comparative genomics studies have revealed a substantial loss of genes in acidophiles compared to alkaliphiles and slightly larger genome sizes in thermophiles than in psychrophiles. Genomic adaptations in halotolerant microorganisms have been identified as polyploidy, an improved mechanism for gaining inorganic osmolytes, a battery of genes for the biosynthesis of organic osmolytes, and an enhanced inorganic osmolyte transporter system [5]. Because of their improved catalytic properties, increased stability, and capacity to function under harsh conditions, extremozymes derived from extremophiles are a rich bioresource of industrially relevant enzymes for use in diverse biotechnological applications [6,7].

The availability of new genome sequences makes the process of looking for new industrial enzymes easier. The rapid advancements in next-generation sequencing (NGS) technologies have revolutionized the landscape of biochemical and biomolecular research. Next-generation sequencing offers unprecedented insights into the genetic composition of organisms, enabling researchers to explore complex biological systems with remarkable precision. This powerful tool, when integrated with cutting-edge computational techniques such as molecular docking, opens new avenues for understanding enzyme mechanisms at an atomic level.

Identifying specific enzymes from extremophilic resources is only as good as the current bioinformatic analyses, and many novel or unknown activities can be missed. It is accordingly vital to screen genomic clone libraries against chromogenic substrates of commercial interest for specific biocatalytic activities, especially if the turnover of a non-natural substrate is required. The recent commoditization of massively parallel DNA sequencing technologies provides new capabilities for generating large datasets for sequence-function pairings. Enzymes, as pivotal catalysts in biochemical processes, play a crucial role in a wide array of physiological functions. Understanding their structure-function relationships is essential for developing targeted therapeutic interventions and novel biotechnological applications.

In the era of omics and advances in high-throughput sequencing (HTS) technologies, a vast amount of molecular data is frequently submitted to international data repositories by groups of scientists from all over the world, which requires expert bioinformaticians and data scientists with strong theoretical and practical knowledge in the fields for their analysis. Despite the significant progress in recent years in HTS and omics approaches, exploring the vast functional extremozyme sequence space still needs improvement [8,9].

Streptomyces species are filamentous, spore-forming, and Gram-positive *Actinomycetota* [10]. Although *Streptomyces* species survive in various habitats, soil is their most significant resource bank. Unique characteristics of members of the genus *Streptomyces* include linear, relatively long genomes (8–10 Mb) and an extremely high G + C content (>70 %) [11]. In the domains of agronomy, medicine, and the food industry, *Streptomyces* species are well-known for their ability to produce a broad range of medically significant natural products with noteworthy biological activity [12]. Because of their extensive genomic structure, which includes several transposable elements close to the ends of the linear chromosome that permit the broad mobility and interchange of "accessory" genes, *Streptomyces* species produce a large variety of secondary metabolites such as antifungal, antiviral, antitumor, antihypertensive, immunosuppressant compounds, and enzymes [13]. Genome sequencing has significantly impacted *Streptomyces* studies, providing valuable insights into their ability to produce industrial enzymes. The annotation of *Streptomyces* genomes has revealed that members of this genus may harbour as many as 6000 to 10,000 genes, mainly due to gene duplication and lateral gene transfer. The *Streptomyces coelicolor* A3(2) genome was one of the first bacterial genomes to be analyzed [14]. The analysis highlighted the presence of several biosynthetic genes, including those that encode for industrial enzymes such as proteases, amylases, and cellulases. Similarly, other *Streptomyces* genome-based studies have also identified the genes required to produce industrially relevant enzymes such as xylanases and chitinases [15,16]. Furthermore, the genomes of *Streptomyces griseus* and *Streptomyces* sp. strain MWC3 contain genes necessary for producing an alkaline protease with possible application as a detergent additive and a new glycoside hydrolase with potential for application in the food industry, respectively [17]. To find unique genes encoding for proteases and chitinases, a comparison between the genomes of *Streptomyces lividans* and *S. coelicolor* was performed, highlighting the role of comparative genomics in discovering novel enzymes [18]. In the same way, Lee et al. [19] analyzed the whole genome sequences of three *Streptomyces* species, including *Streptomyces clavuligerus*, *Streptomyces avermitilis*, and *S. griseus* to identify the sequence variation in enzyme production pathways for the discovery of lipases and amylases for industrial production [19]. The genome sequences revealed how the genetic pathways associated with these enzymes are involved in their biosynthesis and helped identify new enzymes for possible use in industrial processes. As sequencing technologies develop, more research is expected to reveal other genetic data from members of the genus *Streptomyces*, which are known to produce a range of enzymes for potential industrial use.

Considering these points, the current study investigated strain CTM50504, a polyextremophilic actinomycete isolated from the Aïn El-Atrous hot spring at Korbous, Nabeul, Tunisia. Whole genome sequence analysis, using the NextSeq™ 500 Illumina platform, was performed on strain CTM50504. The sequence data obtained were assembled and annotated to understand the extremozyme machinery of strain CTM50504 and to discover potentially novel hydrolases for biotechnological applications. Accordingly, a nucleotide region (*sckp*) encoding a serine alkaline protease was identified and heterologously expressed in *Escherichia coli* BL21(DE3). The recombinant protease (rSCKP) was purified using affinity chromatography, and its biochemical properties were determined. Molecular modelling and docking, as well as molecular dynamics (MD) simulations of three synthetic peptides in the active site of rSCKP, were analyzed to examine the structural stability and catalytic mechanism of this enzyme.

2. Materials and methods

2.1. Sampling and isolation of extremophiles

To isolate hydrolase-producing bacteria, soil samples were collected from the terrestrial hot springs at Aïn El-Atrous, northeast Tunisia, Korbous-Nabeul (GPS coordinates: 36°810′N, 36°56′E). The temperature at the sampling site was 59 °C. The samples were collected under aerobic conditions from wet sediments inside the hot spring and were transported to the laboratory at ambient temperature (23 ± 2 °C). To isolate actinomycete strains, a serial dilution of the samples was carried out in a sterile saline solution (0.5 %, *w/v*, in distilled water), with a dilution prepared up to 10^{-5}. It was followed by the inoculation of 100 μL of each dilution onto five different International *Streptomyces* Project (ISP) media (ISP1, ISP2, ISP4, ISP5, and ISP7) at pH 8 [20,21]. Cycloheximide (25 μg/μL) and Ampicillin (20 μg/μL) were added to the isolation media to increase the ratio of actinomycete strains to other microorganisms.

Replicate plates were incubated at 40, 45, 50, 55, and 60 °C for ten days. After three rounds of purification, a 'typical' actinomycete colony, designated CTM50504 (sub-cultured onto ISP2, 45 °C), was selected and preserved at 4 °C. The CTM50504 culture was stored in a 20 % (*v/v*) glycerol solution at −32 °C and − 80 °C for long-term storage.

2.2. Screening for hydrolase enzymes

Protease activity was screened for by using skimmed milk-ISP2 agar containing (g/L): yeast extract (4), malt extract (10), glucose (4), agar (20), and 250 mL skimmed milk, pH 7.4. After incubation at 45 °C, a halo around colony growth, indicating casein degradation, was considered a positive indicator for protease activity [22]. Using azo-casein as a substrate, the proteolytic activity was determined as follows: 1 mL of 10 g/L azo-casein, suspended in 100 mM bicarbonate-NaOH (pH 11) supplemented with 2 mM Ca^{2+}, was mixed with 1 mL of a suitably diluted enzyme solution (1:1000). The sample was incubated for 30 min at 80 °C with shaking at 250 rpm. The assay mixture was cooled in ice-cold water for 5 min. The remaining substrate was discarded by centrifugation (10,000 ×*g*, 20 min) and filtration via Millipore cellulose filters (0.45 μm). The released azo dye was measured in the filtrate at 440 nm, and activity was expressed in casein units. The control consisted of the enzyme and buffer without substrate. In this case, one unit of protease activity was defined as the amount of enzyme causing an increase of 0.1 in absorbance at 440 nm in one min under the experimental conditions described. Protease activity on *N*-succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide (*N*-Suc-A-A-P-F-*p*NA or AAPF), *N*-succinyl-L-Ala-L-Ala-L-Phe-L-Phe-*p*NA (*N*-Suc-A-A-F-F-*p*NA or AAFF), and *N*-succinyl-L-Ala-Ala-Pro-Met-*p*NA (*N*-Suc-A-A-P-M-*p*NA or AAPM) was measured through the accumulation of free soluble peptides resulting from synthetic peptide degradation using the Folin-Ciocalteu reagent. To do this, a suitably diluted enzyme (1:1000, 0.5 mL) was mixed with 0.5 mL of 100 mM bicarbonate-NaOH buffer (pH 11) supplemented with 2 mM Ca^{2+}, containing 10 mg/mL for each synthetic peptide, and incubated for 15 min at 80 °C. Then, 2.5 mL of 200 g/L trichloroacetic acid was added to stop the reaction. Thirty minutes later, the cooled mixture was centrifuged (15,000 rpm for 20 min). A quantity of 0.5 mL of supernatant was mixed with 2.5 mL of 500 mM Na_2CO_3 and 0.5 mL of Folin-Ciocalteu reagent and incubated for 30 min at 23 ± 2 °C. One unit of protease activity was defined as the amount of enzyme yielding the equivalent of 1 μmol of tyrosine released per minute under the defined assay conditions.

The screening for lipase and phospholipase activities was carried out using ISP2 agar plate assays containing olive oil (OO) and phosphatidylcholine (PC), both at 1 % (*v/v*). Plates were inoculated and incubated at 45 °C for seven days. The plates were stained with 0.01 % (*w/v*) rhodamine B, and any colony displaying orange halos upon UV irradiation was considered a lipase-producing bacterium [23,24]. The liquid medium for lipase production consisted of (g/L): peptone casein (15), malt extract (5), K_2HPO_4 (1), KH_2PO_4 (1.75), $MgSO_4$ (0.5), and 1 % (*v/v*) OO, pH 7. The lipase activity was continuously assayed potentiometrically at pH 10 and 70 °C by measuring the free fatty acids (FFA) released from mechanically stirred triglycerides (TG) emulsion using 0.1 M NaOH and a pH-STAT device (Metrohm 902 Titrando, Herisau, Switzerland) as previously described [24,25]. The lipase enzymatic activity was expressed in international units (1 IU) corresponding to the amount of enzyme required for 1 μmol of FFA to be released per minute. The liquid medium for phospholipase D (PLD) production consisted of (g/L): glucose (8), soya flour (5), peptone (5), K_2HPO_4 (1), and $MgSO_4{\cdot}7H_2O$ (0.5), pH 7. The PLD activity corresponds to the amount of choline released from the PC, as previously described [26].

In addition, the isolates were also screened for their ability to produce Carbohydrate-Active enzymes (CAZymes). Amylase activity screening was performed using ISP2 agar supplemented with 1 % (*w/v*) starch. The amylolytic activity was detected by the formation of a clear zone around the isolates after adding iodine [27]. For the chitinase, pectinase, and cellulase activities, the screening was carried out on ISP2 agar supplemented with 1.5 % (*w/v*) colloidal chitin, 2 % (*w/v*) pectin, and 2 % (*w/v*) carboxymethyl cellulose (CMC), respectively, as previously described [28–30]. After incubation at 45 °C for seven days, plates were stained with 0.1 % (*w/v*) Congo red to visualize the extracellular chitinase, pectinase, and cellulase activities. The production of amylases, chitinases, pectinases, and cellulases was carried out using ISP2 liquid media supplemented with starch, colloidal chitin, pectin, and CMC, respectively (at 1.5 %, *w/v*). The CAZymes were assayed, and activity was expressed as previously described [27–30].

All liquid cultures were grown in 500-mL Erlenmeyer flasks with baffles containing 50 mL of broth medium and were supplemented with cycloheximide at 400 mg/L to minimize fungal contamination.

2.3. Genomic DNA extraction

Two bacterial DNA extraction kits were evaluated for high-quality bacterial genomic DNA extraction. The Bio Basic and NucleoSpin™ 96 (Macherey Nagel™, Strasbourg, France) kits were used with slight modifications. For the Bio Basic kit, after adding 200 μL of 96 % ethanol, the mixture was incubated overnight instead of immediately transferring the samples to the next step. For the Macherey Nagel™ kit, incubation of 1 h at 23 ± 2 °C after the addition of 180 μL of buffer T1 was added to the protocol before adding proteinase K. Extracted DNA was subjected to quality control (QC) using the Implen™ Nanophotometer™ N60/N50 (Implen GmbH Schatzbogen, München, Germany), and was quantified using the Quantus™ fluorometer (Promega, Madison, WI, USA). An Agilent 2100 Bioanalyzer DNA Kit (Agilent Technologies, Lawrence, Kansas, MO, USA) was used for sizing, quantification, and quality control of the extracted DNA.

2.4. Phylogenetic identification of strain CTM50504

Strain CTM50504 was characterized after 10 days of incubation at 45 °C. The phenotypic characteristics, such as the colour of the aerial and substrate mycelium, characteristics of the colony, colour, and morphology of the spore chain, and diffusible pigment production, were performed according to the recommendations of Shirling and Gottlieb and Bergey's manual of bacteriology [20]. The CTM50504 mycelium structures were observed using a light microscope. The enzymatic profiles of the CTM50504 strain were determined using the Analytical Profiling Index (API) ZYM gallery. For the molecular identification of strain CTM50504, the 16S rRNA gene sequence was amplified by Polymerase Chain Reaction (PCR) using *Actinomycetota*-specific primers [31]. The sequences obtained were compared with sequences available in the public sequence databases and the EzTaxon-e server (http://eztaxon-e.ezbiocloud.net/). Phylogenetic and molecular evolutionary genetic analyses were performed using the Molecular Evolutionary Genetics Analysis (MEGA) software v. 11.1. Distances and clustering were calculated using the neighbour-joining method. Bootstrap analysis was used to evaluate the tree topology of the neighbour-joining data by performing 1000 resamplings. Multiple nucleotide sequence alignment was performed using the BioEdit version 7.0.2 software program.

2.5. Whole genome sequence library preparation

To produce a genome sequence for strain CTM50504, the NextSeq™ 500 sequencing instrument at ProfileXpert, Genomics and Micro-Genomics Platform, of UCBL 1 was used. The sequence library was prepared using the NextSeq reagent kit version 2 and 1 μg of genomic DNA, as per the NextSeq™ 500/550 DNA library Prep Reference guide (Illumina, San Diego, CA, USA).

2.6. Bioinformatics pipeline

2.6.1. Genome assembly and analysis

FASTQ sequences were extracted from FAST5 files using the Guppy pipeline v. 2.3.1. The extracted FASTQ files were further demultiplexed by Porechop v. 0.2.4. The quality of the FASTQ files was verified using FastQC [25]. The reads were filtered using Trimmomatic version 0.39 [32] to ensure quality (Phred score > 30). De novo genome assembly was performed using SPAdes and Skesa with default parameters on the Proksee server. QUAST v5.0.2 was employed for assembly metrics determinations. The genome map of strain CTM50504 was constructed using Proksee [33] and the CGView server. The draft whole genome sequence of CTM50504 was uploaded to the Type Strain Genome Server (TYGS) for in silico-based taxonomic analysis [34]. The Average Nucleotide Identity (ANI) analysis was performed using JSpeciesWS [35].

The phylogenomic tree of CTM50504 and closely related strains was generated using the TYGS [32]. After genome assembly, gene prediction and annotation were performed using the Glimmer [36] and Bakta tools to identify regions of DNA containing protein-coding genes.

The Basic Local Alignment Search Tool (BLAST), InterProScan, Kyoto Encyclopedia of Genes and Genomes (KEGG), Blast2Go, and other tools were utilized to identify the gene function in the *S. cyaneofuscatus* CTM50504 genome. Blast2GO version 5.2.5 (2018) (http://www.blast2go.com) was used for automatic presentation to discover coding regions within the *S. cyaneofuscatus* CTM50504 genome. OmicsBox version 1.3 (https://www.biobam.com/omicsbox) was utilized for mapping, identification, visualization, and quantitative statistical analyses. AntiSMASH v7.0.1 was used (with the default parameters) to perform genome mining of putative secondary metabolite biosynthetic gene clusters [37]. The genes encoding shared and distinct proteins or pseudogenes between *S. cyaneofuscatus* CTM50504 and other streptomycetes, *Streptomyces cyaneofuscatus* NRRL B-2570^T [38], *Streptomyces arboris* TRM68085^T, *Streptomyces microflavus* JCM 4496^T, and *Streptomyces microflavus* DSM 40593^T for each sequenced genome could be viewed as a Venn diagram using OrthoVenn3 (https://orthovenn3.bioinfotoolkits.net/home). An oval of a distinct colour for each *Streptomyces* lineage indicates how many ortholog gene groups the strains under consideration share. The functional annotation of the genome via orthology assignment was carried out via the eggNOG-mapper v2 online portal (http://eggnog-mapper.embl.de) [39].

2.6.2. Nucleotide sequence and whole genome sequence accession numbers

The data reported in this work for the nucleotide sequence of the 16S rRNA gene (1526 bp) and Whole Genome Shotgun project of strain CTM50504 have been deposited in the DDBJ/ENA/GenBank databases under the accession numbers PP346394 and PRJNA1065629 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1065629), respectively.

2.7. Validation, comprehensive characterization, heterologous expression, and biochemical characterization of rSCKP

2.7.1. Validation of detected enzymes by PCR and Sanger sequencing

For validation of detected enzymes found by CTM50504 genome sequencing, the respective genomic regions encoding an extracellular protease, phospholipase, lipase, amylase, and chitinase were amplified by PCR using the oligonucleotides listed in Supplementary data, Table S1. The resulting PCR products were subjected to Sanger sequencing (ABI DNA SeqStudio Genetic Analyzer) and analyzed to verify the presence of the target enzymes.

2.7.2. Comprehensive analysis of rSCKP: Genetic, structural, and functional insights

The nucleotide sequence of the *sckp* gene encoding the recombinant alkaline serine protease, rSCKP, was obtained from the genome of *S. cyaneofuscatus* CTM50504. The translation of the nucleotide sequence into a protein sequence was performed using the Expasy Translate tool. Functional annotation and homology search were conducted using BLASTp (NCBI), enabling comparison with sequences in protein databases.

Functional domains and conserved motifs were identified using InterProScan. Multiple sequence alignment of homologous sequences was performed using ClustalW to assess the conservation of catalytic residues. The subcellular localization of the rSCKP protease was predicted using ProtComp. This prediction was complemented by signal peptide analysis with SignalP-5.0.

2.7.3. Heterologous expression of rSCKP

The *sckp* sequence was deduced by assembling and annotation of the CTM50504 genomic sequence, and a recombinant plasmid was prepared by CAYLA (Toulouse, France). Briefly, the *sckp* gene, containing the restriction sites, *Nde*I and *Eco*RI, was cut with the enzymes and was ligated into the pUT57 vector, linearized by the same enzymes, leading to the pSM200 plasmid. From the resulting plasmid, the *Nde*I and *Eco*RI insert carrying the *sckp* gene was ligated into a pET-28b vector linearized by the same enzymes, leading to the corresponding vector, pSM201. The pSM201 recombinant vector was then transformed into the *E. coli* TOP10, leading to the formation of the pSM202 vector. The latter plasmid construction was confirmed by sequencing. The pSM202 plasmid was transformed into an electrocompetent *E. coli* BL21(DE3) strain as described previously [40]. The culture and the IPTG induction of the rSCKP expression were carried out as detailed recently [41]. The recombinant protease was purified using a HisTrap™ HP chelating Ni-affinity chromatography column (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) as detailed by the authors [40]. The rSCKPs sample's purity was analyzed by sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The rSCKP proteins were visualized by Coomassie brilliant blue R-250 staining. In addition, a zymography technique was performed by integrating 0.1 % (*w/v*) azo-casein into the separating gel. The biochemical characterization and enzyme kinetics of the rSCKP protease were carried out according to standard conditions detailed previously by the authors [22,40] using azo-casein as modified protein substrate and AAPF, AAFF, and AAPM as synthetic peptide substrates. The effect of pH on the rSCKP activity was analyzed between pH 2 and 13 with different buffers at 100 mM at 80 °C. The pH stability was assessed by incubating the rSCKP at pH 6 to 12 in different buffers (100 mM each, supplemented with 2 mM $CaCl_2$): glycine-HCl for pH 3–5, MES for pH 5–6, PIPES for pH 6–7, HEPES for pH 7–8, Tris-HCl for pH 8–9, glycine-NaOH for pH 9–11, bicarbonate-NaOH for pH 11–11.5, Na_2HPO_4-NaOH for pH 11.5–12, and KCl-NaOH for pH 12–13 for 24 h at 60 °C. The optimum temperature for the rSCKP activity was determined from 30 to 100 °C, at pH 11. The rSCKP thermostability was explored by incubation at 70, 80, 90, and 100 °C and determination of the rSCKP activity at 2 h intervals. The compatibility of rSCKP, PREFERENZ® S 210 (Genecor International, Inc., a Danisco Division A/S, Rochester, NY, USA), and PROTEASE, Type XIV (Merck KGaA, Darmstadt, Germany; Sigma-Aldrich, protease from *Streptomyces griseus* Type XIV) with various commercial detergent formulations was also evaluated following the method described by Mechri et al. [42].

2.8. Structural analysis of rSCKP

2.8.1. Theoretical physico-chemical parameters of rSCKP

The theoretical physico-chemical parameters of the rSCKP protease were determined using ProtParam and EMBOSS pepstats. These tools calculated the molecular weight, theoretical isoelectric point (pI), instability index, and grand average of hydropathy (GRAVY). These values were compared to those of other microbial alkaline proteases to assess the rSCKP enzyme's stability and solubility.

2.8.2. Secondary structure prediction, 3D modelling, and structural model validation

Secondary structure prediction was performed using SecStrAnnotator, identifying α-helices and β-strands. A three-dimensional model of the rSCKP protease was generated using AlphaFold 2, an artificial intelligence tool specialized in protein structure prediction. Multiple models were generated, and the one with the highest pLDDT score was selected for validation.

The quality of the 3D model was evaluated using MolProbity, which analyzed the stereochemistry and overall quality of the predicted structure. Ramachandran plots were generated to assess the distribution of φ and ψ angles of the residues, and rotamers were evaluated to detect potential aberrant conformations. To further assess the functional relevance of the predicted structure, the conservation of the catalytic triad geometry by performing a local structural superposition was evaluated. The coordinates of the Asp27-His70-Ser278 triad from the AlphaFold-predicted model of rSCKP were aligned with those from a crystallized subtilisin BPN' (PDB ID: 1LW6), using the "Superimpose by Alignment" function in BIOVIA Discovery Studio Visualizer.

2.8.3. Model refinement and molecular docking of rSCKP

The molecular modelling of the rSCKP was performed using AlphaFold (https://alphafold.ebi.ac.uk/), which provided a reliable 3D structure. This structure was further refined by incorporating a Ca^{2+} and catalytically relevant water molecules through a molecular docking protocol. The COACH and PLIP tools were employed to confirm the potential presence of a calcium-binding site, given the importance of calcium in the structure of subtilisins, which AlphaFold did not model within the predicted rSCKP structure. Synthetic peptide candidates, subjected to docking, were selected based on observed relative in vitro activity. Their structures were retrieved from PubChem in PDB format and optimized using the Clean Geometry option in Biovia Discovery Studio. This ligand preparation ensured optimal initial conformations before docking. For charge and hydrogen assignment, both the rSCKP and ligands were prepared using the DockPrep tool with the AMBER ff14SB force field integrated into Chimera 1.14 to ensure a stable and biologically relevant conformation for molecular docking. Molecular docking was performed using AutoDock Vina, employing a Monte Carlo global optimization algorithm coupled with a hybrid empirical/force-field scoring function. The docking box (30 × 30 × 30 Å) was centered on the catalytic triad, and 10 poses per ligand were generated using an exhaustiveness value of 8. Validation of the docking results was assessed by calculating the root mean square deviation (RMSD) (lower and upper bounds) between the best-ranked pose and the alternative conformations. Poses with RMSD <2 Å were considered conformationally convergent. The resulting complexes were analyzed using Biovia Discovery Studio and the PLIP tool, which enabled the identification and characterization of key interactions between rSCKP and each peptide. Two-dimensional interaction diagrams of ligand-rSCKP interactions were generated to provide a clear visualization of the forces involved, including hydrogen bonds, hydrophobic interactions, π-π interactions, and van der Waals forces. PyMOL was utilized to produce the three-dimensional representations.

Molecular dynamics (MD) simulations were prepared using CHARMM-GUI and executed with GROMACS 2025.0, employing the CHARMM36m force field. Each peptide–receptor complex (rSCKP) was embedded at the center of a rectangular TIP3P water box with a 10 Å buffer and neutralized with K^+ and Cl^- ions at a physiological concentration of 0.15 M. Energy minimization was performed using the steepest descent algorithm until the maximum force dropped below 10 kJ/mol, with a maximum of 100,000 steps. Van der Waals and electrostatic interactions were evaluated using a 1.2 nm cut-off under the Verlet scheme, with three-dimensional periodic boundary conditions applied throughout. Two equilibration steps followed: an NVT [number of particles (N), system volume (V), and temperature (T)] phase at 310 K using the V-rescale thermostat, and an NPT [number of particles (N), system pressure (P), and temperature (T)] phase at 1 bar using the Parrinello–Rahman barostat with an isothermal compressibility of 4.5×10^{-5} bar^{-1}. Bonds involving hydrogen atoms were constrained using the LINCS algorithm. The production phase lasted 100 ns, with a time step of 2 fs and no velocity regeneration, ensuring continuity from equilibration. Coordinates, energies, and logs were saved every 10 ps. Trajectory analyses were based on five structural parameters: receptor RMSD, ligand RMSD, radius of gyration (Rg), solvent-accessible surface area (SASA), and root mean square fluctuations (RMSF). Each simulation was repeated twice to assess robustness and reproducibility.

2.8.4. Molecular dynamics simulations and binding free energy analysis

Binding free energies of the peptide–rSCKP complexes were calculated using gmx_MMPBSA v1.5.0 [43]. The analysis used frames from 0 to 10,000 (0–100 ns) with a sampling interval of 500 frames, and was conducted on solvated trajectories to reflect aqueous conditions. The Poisson-Boltzmann radius was set to 2 Å, in accordance with CHARMM36m force field recommendations. Ionic strength was maintained at 0.15 M to simulate physiological conditions. Entropy was estimated using the interaction entropy method, with segment sizes of 25 frames and a temperature of 300 K. Energy decomposition was performed using mode 2 with verbosity level 3, targeting receptor residues located within 4 Å of the ligand (defined using the "within 4" clause), to identify key amino acids and quantify their individual energetic contributions.

The calculated thermodynamic parameters included the mean enthalpy of binding (ΔH), the entropic contribution (-TΔS), the binding free energy (ΔG), and the internal energy of the ligand (E_ligand). Together, these values provide a comprehensive view of the forces stabilizing the complex and the energetic trade-offs of peptide binding.

3. Results

3.1. Isolation of extremophile bacteria and screening of enzyme activities

The terrestrial hot spring samples were collected at Aïn El-Atrous, Korbous-Nabeul, Tunisia (GPS coordinates: 36°810′N, 36°56′E) to isolate hydrolase-producing strains. Among the seventeen actinomycete isolates, an isolate designated as strain CTM50504 exhibited the most promising hydrolase production. Strain CTM50504 tested positive for the extracellular production of lipase, phospholipase, protease, amylase, and chitinase. The ratio of the apparent zone diameter on solid media and the production of enzymatic activity (U/mL) in liquid media supported the selection of strain CTM50504 for further study (Table 1).

3.2. Genomic DNA extraction protocols and quality control

In this study, the modified NucleoSpin™ 96 (Macherey Nagel™)

Table 1
Qualitative and quantitative analysis of the enzymatic activities of strain CTM50504.

Hydrolase activity	Diameter ratio on solid media (mm)	Production of enzymatic activity (U/mL) in liquid media
Protease	+++	9000 ± 95
Phospholipase	+++++	6 ± 1
Lipase	+++++	80 ± 7
Amylase	++++	850 ± 12
Chitinase	+++++	40 ± 3
Pectinase	–	–
Cellulase	–	–

Note: The signs +++, ++++, and ++++++ correspond to a ratio of the halo diameter/colony diameter of 2, 2 to 3, and 5 to 6, respectively. The sign, –, corresponds to an absence of activity. All measurements and activity determinations were replicated at least three times. Values are reported with ± SE and represent the means of at least three replicates.

protocol resulted in genomic DNA of better quality and quantity. The 260/230 and 260/280 ratios were within the appropriate range for good-quality, pure nucleic acid. In addition, the Quantus™ fluorometer, Agilent 2100 Bioanalyzer, and 5200 Fragment Analyzer System analyses validated the extracted genomic DNA quality (quantity and size).

3.3. Phenotypic and molecular identification of strain CTM50504

Strain CTM50504 is an aerobic, non-acid-fast, Gram-positive, and catalase-positive actinomycete. In addition, strain CTM50504 was able to grow at 50 °C and a pH range of 6 to 9 and requires the presence of NaCl for its growth, suggesting that it is halophilic. These phenotypic characteristics are consistent with members of the genus *Streptomyces* (Table 2).

Based on the 16S rRNA gene sequence (GenBank accession no.: **PP346394**) analysis (GenBank and EzTaxon), the nearest phylogenetic neighbour of strain CTM50504 was identified as S. cyaneofuscatus NRRL B-2570^T with a sequence similarity of 99.93 %. Furthermore, when a 16S rRNA phylogenetic tree was constructed, the bootstrap analysis (1000 replicates) revealed that node values ranged from 50 to 98 % in all groups, demonstrating their robustness (Fig. 1), confirming the assignment of strain CTM50504 as a streptomycete.

Table 2
Phenotypic characteristics of strain CTM50504 and *Streptomyces cyaneofuscatus* NRRL B-2570^T [38] (GenBank accession no.: JOEM01000050).

Phenotypic characteristic		Strain CTM50504	*Streptomyces cyaneofuscatus* NRRL B 2570^T
Morphology	Colony colour	Brown	Sand yellow
	Cultivation medium used	ISP2	ISP2
	Incubation period (days)	10	10–14
Metabolite utilization	Arginine	+	+
	Citrate	+	+
	Gelatine	+	+
	Lysine	−	−
	Ornithine	−	−
	Tryptophan	+	+
	Urea	+	+
Metabolite production	Acetoin	Yes	Yes
	H_2S	No	No
	Indole	Yes	Yes
Enzymes	Acid phosphatase	+	+
	Alkaline phosphatase	+	+
	α-Chymotrypsin	+	+
	α-Fucosidase	−	−
	α -Galactosidase	−	−
	α-Glucosidase	+	+
	α-Mannosidase	+	+
	Arginine dihydrolase	+	+
	β-Galactosidase	+	+
	β-Glucosidase	+	+
Growth at:	5 % NaCl	+	−
	10 % NaCl	+	−
	15–20 % NaCl	+	−
	20 °C	−	+
	30 °C	−	+
	40 °C	+	+
	42 °C	+	−
	45 °C	+	−
	50 °C	+	−
	59 °C	+	−
	70 °C	+	−

Note: +: All strains are positive; −: All strains are negative. ISP: International *Streptomyces* Project.

3.4. General genomic features and phylogeny of Streptomyces cyaneofuscatus CTM50504

An extended genome analysis of *S. cyaneofuscatus* CTM50504 was performed to gain insights into the strain's ability to exhibit hydrolase activities. FastQC analysis of the Illumina NextSeq™ 500-generated data showed that 3,893,456 paired-end reads were generated at a length of 150 bp, resulting in a genome coverage of 200. After Trimmomatic quality filtering, 3,677,363 paired-end reads were obtained (Table 3).

Following assembly, a high-quality draft genome sequence of 858 contigs and an average G + C content of 71 % was obtained. The assembled genome's total size was 8,591,922 bp, with an N50 of 17,033 bp, with the longest contig length at 104,377 bp, which reflects the inherent limitations of Illumina short-read sequencing, which may hinder the full resolution of biosynthetic gene clusters and synteny relationships. Future incorporation of long-read technologies such as Oxford Nanopore or PacBio technologies could greatly improve assembly continuity and enhance the accuracy of functional and comparative genomic analyses.

3.5. Genome annotation

The assembled draft genome was annotated using the Bakta and Glimmer software tools, identifying 8100 coding sequences (CDSs), seven rRNAs, 76 tRNAs, 49 other RNA genes, and 7 CRISPR regions [44] (Table 3). A functional categorization of the CDSs by gene ontology (GO) terms was performed based on the Blastx hits from the non-redundant (nr) database when using Blast2GO annotation in OmicsBox 2.0.10. A total of 8100 genes were annotated in the GO database and could be divided into three subcategories according to the function of each gene: biological process (49 branches), cellular component (6 branches), and molecular function (9 branches) (Supplementary data, Fig. S1 and Excel File 1). In the biological process subcategories, the three largest branches are cellular process, metabolic process, and organic substance metabolic process. In the cellular component subcategories, the branch regarding the cell membrane accounts for the most significant proportion. In the molecular function subcategories, the top four are the catalytic activity, hydrolase activity, binding, and transferase activity branches (Supplementary data, Fig. S1).

Analysis of the subsystems at the “function” level (Supplementary data, Excel File 2) revealed the presence of proteins with potential biotechnological interest. For example, we found genes encoding lipases, proteases, and galactosidases. Thermostable variants of these enzymes are currently in high demand in industry.

Functional annotation of CDSs by assignment of orthology (Clusters of Orthologous Groups, COG) was performed by eggNOG-mapper v2. The annotation by eggNOG-mapper provided a detailed description of the GO term, Enzyme Commission (EC) number, annotation level, COG category, and description (Supplementary data, Excel File 3). A total of 7197 proteins (88.85 %) out of 8100 CDSs were queried by eggNOG-mapper v2 for orthology mapping. The COG annotation of the *S. cyaneofuscatus* CTM50504 genome showed that the most abundant role was Function Unknown ‘S' (20,8 %), transcription ‘K' (13.07 %), followed by amino acid metabolism and transport ‘E' (11.69 %), and carbohydrate transport and metabolism ‘G' (6.15 %) (Supplementary data, Fig. S2). The details of COG functional classification are shown in Supplementary data, Fig. S2, and Excel File 3. The distribution of the annotated genes was similar to that of various other *Streptomyces* strains [45], as transcription and primary metabolism, including carbohydrate and amino acid metabolism, are among the essential features in living organisms.

3.6. Phylogenetic and OrthoVenn analysis

The TYGS was used to perform phylogenomic analysis. Comparison with other *Streptomyces* strains showed that CTM50504 is a strain of

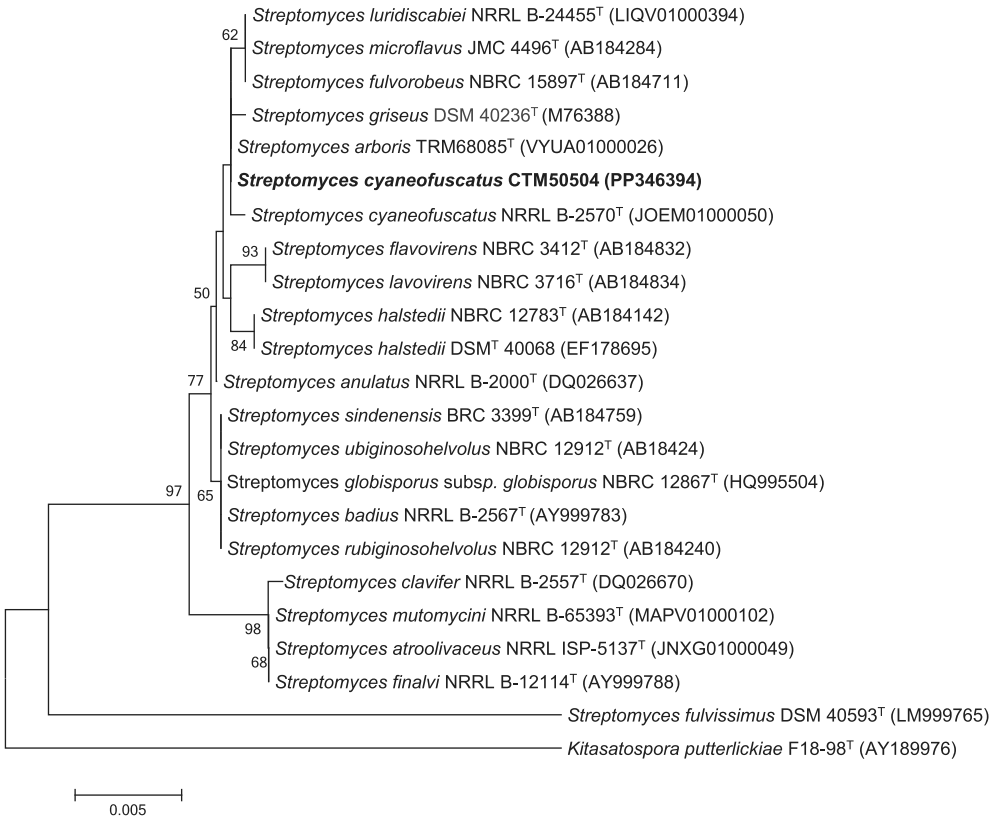

Fig. 1. Neighbour-joining phylogenetic tree based on the 16S rRNA gene sequences of strain CTM50504 and representatives of related taxa. Phylogenetic tree showing the relationship of strain CTM50504 to other species within the genus *Streptomyces* based on 16S rRNA gene sequence analysis (GenBank accession no.: **PP346394**). Bar, 0.005 substitutions per nucleotide position. Bootstrap values were expressed as a percentage of 1000 replicates, and only values >50 % are shown at the branch points. The sequence of *Kitasatospora putterlickiae* F18-98^T (GenBank accession no.: **AY189976**) was selected at random as an outgroup.

S. cyaneofuscatus NRRL B-2570^T (GenBank accession no.: JOEM01000050; digital DNA:DNA hybridization >70 %) and is related to *S. arboris* TRM68085^T (GenBank accession no.: VYUA01000026), *S. griseus* DSM 40236^T (GenBank accession no.: M76388), *S. fulvorobeus* NBRC 15897^T (GenBank accession no.: AB184711), *S. microflavus* NBRC 13062^T (GenBank accession no.: AB184284), and *S. luridiscabiei* NRRL B-24455^T (GenBank accession no.: LIQV01000394) (Fig. 2 and Supplementary data, Excel File 4).

The taxonomic position of *S. cyaneofuscatus* CTM50504 was confirmed as a strain of *S. cyaneofuscatus* (Fig. 2 and Supplementary data, Excel File 3) by TYGS and JSpiecesWG server analysis: the average nucleotide identity (ANI) scores, as the ANIm values between *S. cyaneofuscatus* CTM50504 and other *Streptomyces* species ranged from 92 to 99 % (Fig. 2 and Supplementary data, Excel File 4).

The BLAST atlas of *S. cyaneofuscatus* CTM50504 was created using the CGView server [46]. BLAST atlas was utilized to compare the genome of CTM50504 with that of the genome of the type strain of *S. cyaneofuscatus* (NRRL B-2570^T) (Fig. 3).

To examine the differences among the five *Streptomyces* strains in the phylogenomic cluster of the CTM50504 genome, the compositions of orthologous gene clusters were analyzed using OrthoVenn3 [39]. A collective of 7831 orthologous gene clusters was identified within the five *Streptomyces* strains. Among these, 5108 clusters were shared across all five strains, with 5044 clusters of single-copy genes. A total of 28 clusters were found to be unique to *S. cyaneofuscatus* CTM50504 (Fig. 4).

The functional classification of the proteins belonging to the 28 clusters of ortholog proteins unique to *S. cyaneofuscatus* CTM50504 is presented in Supplementary data, Fig. S3. Most proteins were classified as involved in biological, metabolic, 'cellular metabolic processes', and 'response to stress processes' (Supplementary data, Fig. S3A). In this case, the biological processes represent a specific objective for which the organism is genetically programmed, such as cell division. On the other hand, metabolic processes involve chemical reactions and the pathways through which living organisms transform chemicals. In contrast, cellular metabolic processes are chemical reactions, and their pathways occur within the cell to transform chemicals. In contrast, 3 % of the 28 clusters were associated with the 'response to stress' ontology that includes feedback to abiotic and biotic stresses, chemical stimuli, temperature, and inorganic stimuli. The presence of these proteins is not surprising since strain CTM50504 was isolated from a high-temperature and high-salt environment. When analyzed from the point of view of molecular functions (Supplementary data, Fig. S3B), 8.05 % of the proteins represent hydrolase activity, which is a class of enzymes responsible for hydrolysing a chemical bond and dividing a large molecule into two smaller ones; 15 % of the proteins were predicted to be involved in a catalytic process; and another 15 % for binding with DNA, for transcription of genes responsible for stress response, and other molecules.

3.7. Metabolic pathway analysis

Protein-coding sequences were screened against the KEGG database to analyse the predicted gene products of metabolic processes and related cellular processes, resulting in 336 KEGG pathways assigned to CDSs of the CTM50504 genome. Primarily among the 336 pathways, the pathway most represented is related to the biosynthesis of secondary metabolites (321), microbial metabolism in diverse environments (201), biosynthesis of cofactors (132), ABC transporters (132), and the biosynthesis of amino acids (100) (Supplementary data, Fig. S4).

Table 3
Summary information for the *Streptomyces cyaneofuscatus* CTM50504 genome assembly and annotation.

Genome assembly features	Statistic
Genome coverage	200
Contigs (≥ 0 bp)	4399
Contigs (≥ 500 bp)	858
Total length (≥ 0 bp)	9,550,108
Total length (≥ 500 bp)	8,591,922
Largest contig	104,377
Total length	8,591,922
GC (%)	71.5
N50	17,033
N75	8619
L50	151
L75	325
Gene_num	8231
Gene_total_length (bp)	7,115,427
Gene_average_length (bp)	902.40
GC_content_in_gene_region (%)	71.50
Gene/Genome (%)	82.82
Intergenetic_region_length (bp)	1,476,495
GC_content_in_intergenetic_region (%)	68.62
Intergenetic/Genome (%)	17.18
Trna	76
tmRNA	1
rRNA	7
ncRNA	23
ncRNA regions	25
CRISPR	7
CDSs	8100
ORFs	12
BioProject	PRJNA1065629
BioSample	SAMN39465713
Accession Number	JAYXBL010000000
Sequence Read Archive (SRA)	–

3.8. Identification of secondary metabolite gene clusters using antiSMASH

The antiSMASH platform predicted putative biosynthetic gene clusters (BGCs) within the CTM50504 genome [47]. The analysis showed 49 putative BGCs (Supplementary data, Excel File 5). Among these, there were three NRP-metallophores, 10 NRPS, PKS-like, six T1PKS, transAT-PKS gene, three melanin, seven terpenes, two NI-siderophore, two hydrogen-cyanide, lassopeptide, two butyrolactone, four RiPP-like, two T2PKS, two T3PKS, three NRPS-like, ectoine, two LAP, two RRE-containing, two thiopeptide, and three lanthipeptide-class-ii BGCs (Supplementary data, Excel File 5). Nine gene clusters showed 100 % similarity to known gene clusters predicted to encode for NRP-metallophore, NRPS, PKS-like, T1PKS, *trans*-AT-PKS, ectoine, and a terpene. In contrast, some gene clusters revealed a low similarity percentage with other known clusters, indicating that *S. cyaneofuscatus* CTM50504 could be a valuable biological resource for producing novel bioactive compounds.

3.9. Annotation of enzyme-encoding genes and their biotechnological potential

Annotations from the three methods, BLAST2GO, KEGG, and eggNOG-mapper, were used as a reference in the search for commercially relevant enzymes as defined by the list used by the Association of Manufacturers and Formulators of Enzyme Products (AMFEP). The enzyme classification code (EC number) and class distributions (level 3) generated by BLAST2GO annotation are summarized in Supplementary data, Fig. S5A. Among the 770 identified enzymes, 323 (41.94 %) were related to hydrolases or hydrolytic enzymes, while 237 (30.78 %) sequences were related to transferases (Supplementary data, Fig. S5A).

The search resulted in 268 biotechnologically relevant enzymes, including 179 proteases (61 % of total predicted commercial enzymes), 59 glycosyltransferases (20 %), 5 amylases (2 %), 5 chitinases (2 %), and 20 lipases (7 %) (Supplementary data, Fig. S5B). In the lipase family, we found two phospholipase A2, nine triacylglycerol lipases, two phospholipase C, five lysophospholipases, one sphingomyelin phosphodiesterase, and one phospholipase D-like domain-containing protein (Supplementary data, Fig. S5B).

3.10. Validation of detected enzymes by PCR

The genes encoding the protease (1539 bp), phospholipase (1234 bp), lipase (924 bp), amylase (1737 bp), and chitinase (2370 bp) in the CTM50504 genome were amplified by PCR using the oligonucleotides (listed in Supplementary data, Table S1) and were validated by Sanger sequencing.

3.11. Validation, comprehensive characterization, heterologous expression, and biochemical characterization of rSCKP

3.11.1. Validation and comprehensive genetic, structural, and functional characterization of rSCKP

InterProScan analysis confirmed the presence of the peptidase S9/S53 domain (IPR000209) and the characteristic subtilase signature (PR00723). The catalytic site was identified with the classical catalytic residues Asp27, His70, and Ser278, indicating a subtilisin-type enzymatic activity. These residues are involved in the cleavage mechanism of peptide bonds via a catalytic triad (D-H-S), confirming that rSCKP adopts a typical serine protease mode of action.

The ProtComp tool predicted that rSCKP is an extracellular protease with a secretion score of 8.8. The presence of a 21 amino acid signal peptide was confirmed using SignalP-5.0, suggesting the enzyme's involvement in the degradation of environmental proteins. This result aligns with other bacterial subtilases, which are often secreted to facilitate nutrient acquisition by breaking down surrounding protein substrates.

3.11.2. Nucleotide and protein sequence analysis, and heterologous expression of rSCKP

The analysis of the gene encoding the recombinant alkaline serine protease, rSCKP, was performed using the Expasy Translate tool, yielding a 1539-base-pair sequence translated into a 355-amino-acid protein. BLASTP analysis revealed strong similarities with S9 family proteases, particularly those from the genus *Streptomyces* (Supplementary data, Table S2). The rSCKP protein includes a well-defined pre-pro-sequence of 157 amino acids and has a predicted molecular mass of 52 kDa. The pSM202 construct was designed to express in *E. coli* the SP-Pro-M rSCKP enzyme. The culture growth and the subtilisin activity were monitored for 72 h. The optimum time to produce subtilisin activity was obtained 22 h after IPTG induction. While no protease activity was present intracellularly in the recombinant *E. coli* cells, substantial activity was measured in the culture media. The BL21(DE3)/pSM202 strain was kept for further protease purification. The maximum protease activity reached was 4400 U/mL after 3 h of induction with IPTG, with a specific activity of 10,621 U/mg (Table 4). After purification using a HiTrap™ chelating HP column, rSCKP exhibited a specific activity of 168,82 U/mg, representing a 15.9-fold purification with an 80 % recovery of initial activity (Table 4). These values indicate a high expression level and purification efficiency of the recombinant enzyme.

The rSCKP enzyme consists of a single band of ~45 kDa (Fig. 5), while the theoretical mature one corresponds to ~36 kDa. The difference is attributed to the pET-28b polylinker extension of hexahistidine and to the post-translational changes of the protein, which may include glycosylation. The azo-casein zymogram activity staining revealed one clear zone of proteolytic activity at 45 kDa against the blue background for the purified rSCKP. The optimal activity of rSCKP was achieved at pH 11 (Fig. 6A). The relative activity at pH 7 and 13 was 90 % and 10 %, respectively. At pH levels of 5 and 6, the rSCKP activity was 38 % and 58 %, respectively. The rSCKP pH stability profile indicated that the

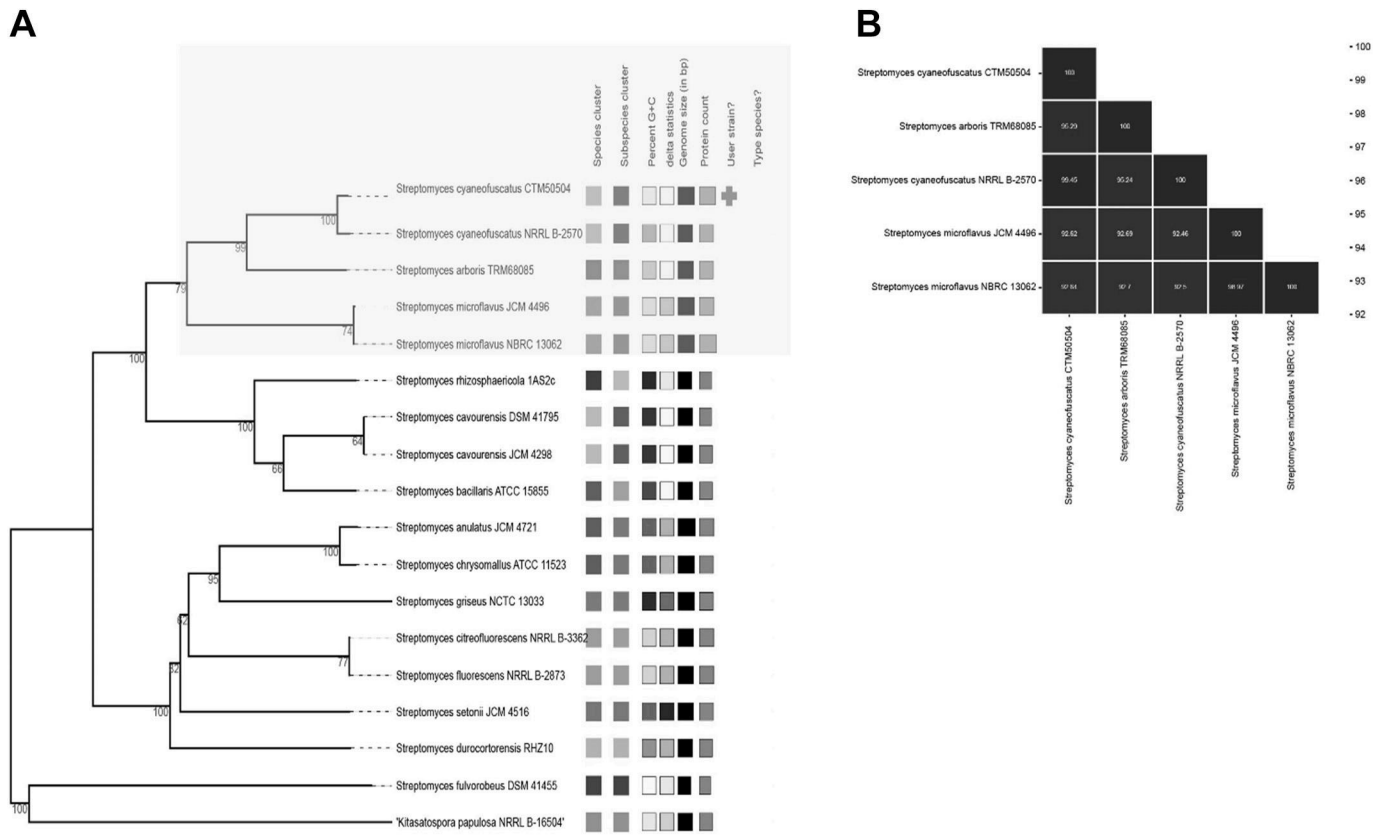

Fig. 2. Phylogenomic tree and heat map of pairwise ANI. (A) Phylogenomic tree inferred using the Type Strain Genome Server (https://tygs.dsmz.de/). Result of Genome BLAST Distance Phylogeny approach (GBDP), using distances calculated from genome sequences. The branch lengths are scaled in terms of the GBDP distance formula *d5*. The numbers below branches are GBDP pseudo-bootstrap support values >60 % from 100 replications, with an average branch support of 79.8 %. The tree was rooted at the midpoint. The light pink box group groups all the strains very close to strain CTM50504, according to the classification proposed by Sangal et al. [91]. Leaf labels are annotated by affiliation to ① species and ② subspecies clusters, predicted by the digital DNA:DNA hybridization (dDDH) calculator service; ③ genomic G + C content (Min. 61.7 % , Max. 70.7 %); ④ δ values (Min. 0.1 , Max. 0.3); ⑤ overall genome sequence length (Min. 3.9 Mb , Max. 10.4); and ⑥ number of proteins (Min. 3611 , Max. 9472). (B) The heatmap shows the distribution of pairwise ANI values across the pink group in the phylogenetic tree. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

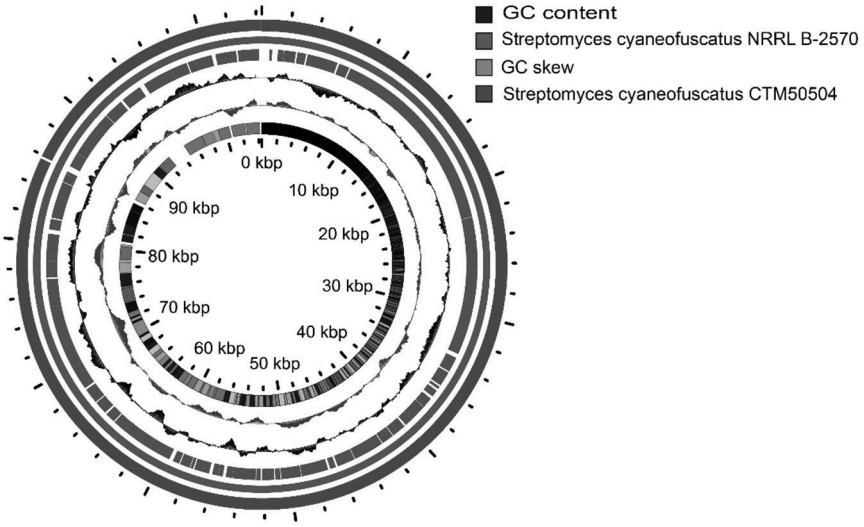

Fig. 3. Overview of the circular genome map. The circular genome map of the *Streptomyces cyaneofuscatus* strain CTM50504 was mapped against the *Streptomyces cyaneofuscatus* NRRL B-2570^{T} genome. The circular genome was generated with the CGView server.

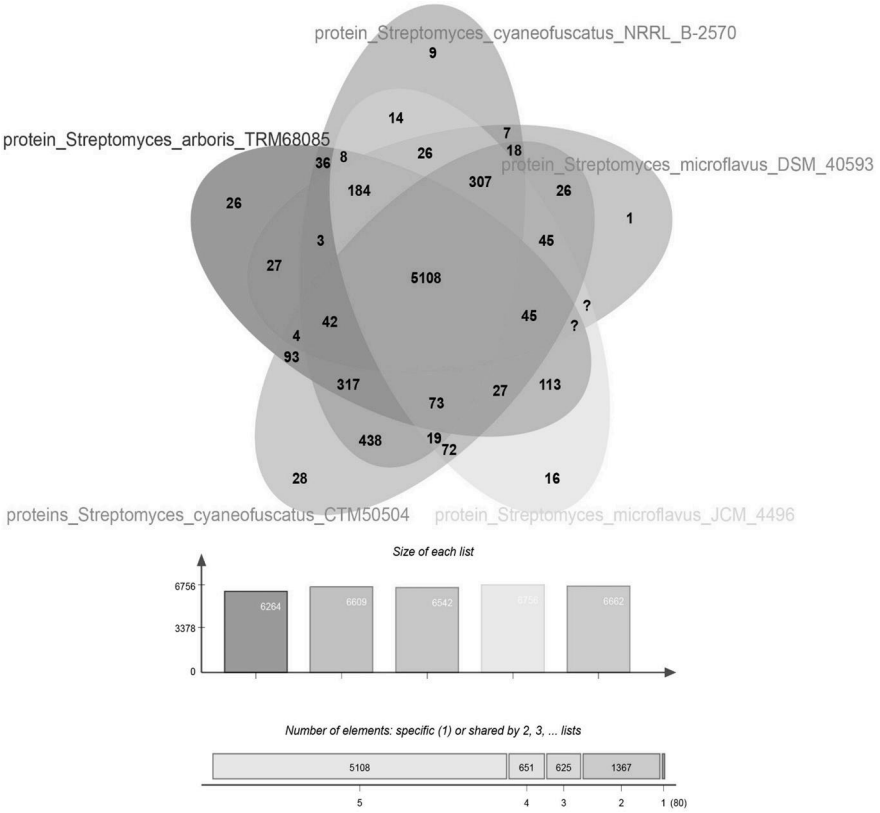

Fig. 4. Genome-wide comparison and analysis of the orthologous clusters. The genome-wide comparison and analysis of the orthologous clusters obtained by OrthoVenn3 between five different *Streptomyces* species: *Streptomyces cyaneofuscatus* strain CTM50504, *Streptomyces cyaneofuscatus* NRRL B-2570^{T}, *Streptomyces arboris* TRM68085^{T}, *Streptomyces microflavus* JCM 4496^{T}, and *Streptomyces microflavus* DSM 40593^{T}.

Table 4
Purification of the recombinant enzyme, rSCKP, and resultant specific activity of the purified enzyme, and the final degree of purification obtained.

Step of purification	Full of activity (units)[a,b] × 10^3	Full of protein (mg)[a,c]	Specific activity (U/mg of protein)[a]	Activity recovery rate (%)	Purification factor (*n*-fold)
Crude preparation (200 mL)	880 ± 14	8,28 ± 09	10,621	100	1.0
HiTrap™ chelating HP column	704 ± 32	4,17 ± 02	168,82	80	15.9

rSCKP: Recombinant serine alkaline protease from *Streptomyces cyaneofuscatus* strain CTM5050.
[a] These data are the average of at least three assays, and ± standard errors are reported.
[b] The purification was started from a 200 mL culture.
[c] Protein concentration was ascertained by the Bradford's technique [89].

enzyme showed a maximum activity at pH 7, 8, 9, and 10, respectively, for up to 75, 60, 45, and 15 min (Fig. 6B). With 2 mM of Ca^{2+}, rSCKP activity was enhanced to 235 and 125 %, respectively, at 70 and 90 °C, with an optimum at 80 °C (Fig. 6C). Without 2 mM of Ca^{2+}, rSCKP was observed to retain only 97 % and 35 % of activity, respectively, at 60 and 90 °C, with an optimum at 70 °C. rSCKP half-life times at 70, 80, 90, and 100 °C were 22, 14, 8, and 6 h in the presence of 2 mM Ca^{2+} (Fig. 6D).

As shown in Fig. 6E, rSCKP maintained high proteolytic activity in the presence of the tested commercial detergents, retaining 85–100 % of its original activity. In contrast, PREFERENZ® S 210 and PROTEASE Type XIV exhibited reduced stability when exposed to the same liquid and solid detergents (Fig. 6E). These findings highlight the excellent compatibility and robustness of rSCKP, supporting its suitability for use in laundry detergent formulations, particularly in products such as Ariel and Class.

The analysis of the relative activity of rSCKP towards various peptide substrates reveals a clear specificity for certain peptide sequences. No enzymatic activity was observed for substrates containing basic (Lys and Arg) or aliphatic, hydrophobic (Leu) residues at the P1 position, suggesting that rSCKP does not exhibit typical trypsin- or chymotrypsin-like activity. In contrast, the enzyme shows a strong preference for substrates featuring aromatic residues such as Phe at the P1 position, particularly when preceded by small neutral residues such as Ala or Pro at P2 and P3. The substrate AAPF exhibited the highest activity (100 %) and was used as a reference point. This pattern indicates that rSCKP favors hydrophobic and aromatic sequences, reflecting the specificity of its active site. These findings provide valuable insight into the cleavage profile of rSCKP and highlight its potential for targeted applications in peptide or protein modification (Table 5).

3.12. Structural analysis of rSCKP

3.12.1. Theoretical physico-chemical parameters

ProtParam analysis determined a molecular weight of 36.6 kDa and an isoelectric point of 5.47. The instability index is 24.70 with a GRAVY score of −0.157. Additionally, EMBOSS pepstats confirmed these values, with a very close isoelectric point (5.5867) and a low probability of inclusion body formation (Fig. 7A).

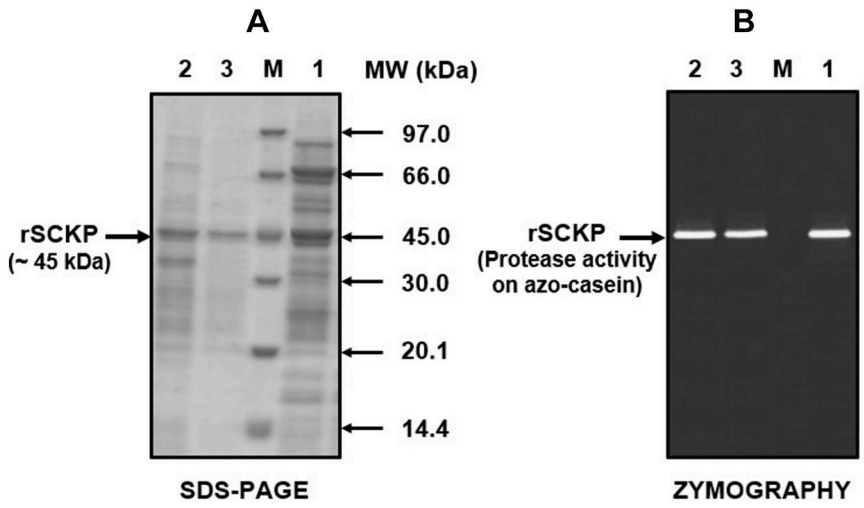

Fig. 5. Electrophoretic analysis of rSCKP. (A) SDS-PAGE of rSCKP (30 μg). Lanes 1–4, total cell extract, protein marker, rSCKP after Ni-NTA affinity chromatography (200 mM imidazole), and rSCKP after Ni-NTA affinity chromatography (180 mM imidazole). (B) Zymogram caseinolytic activity staining of rSCKP at 20 μg. Lanes 1–4, total cell extract, protein marker, rSCKP after Ni-NTA affinity chromatography (200 mM imidazole), and rSCKP after Ni-NTA affinity chromatography (180 mM imidazole).

3.12.2. Secondary structure prediction, 3D structure elucidation, and structural model validation

The SecStrAnnotator tool revealed a structure composed of 9 α-helices and 25 β-strands, consistent with the typical features of subtilases. 3D modelling using AlphaFold 2 resulted in a highly reliable model with an average pLDDT score of 97.8, indicating a high confidence in the predicted structure (Fig. 7B). MolProbity analysis of the 3D model indicated a high-quality structural conformance with 97.5 % of residues in favorable regions of the Ramachandran plot, no aberrant residues, 98.15 % favorable rotamers, and only 0.37 % incorrect rotamers. No significant Cβ deviations were predicted (Fig. 8). Geometric analysis of the 'D27, H70, and S278' catalytic triad indicated an arrangement consistent with classical subtilisins. Indeed, structural superposition of the catalytic triad residues between the AlphaFold-predicted model of rSCKP and a crystallized subtilisin BPN' (PDB ID: 1LW6) yielded a low RMSD of 0.65 Å across Cα atoms, supporting the structural conservation of the active site geometry (Supplementary data, Fig. S6).

3.12.3. Molecular docking experiments

The calcium ion was included in the predicted enzyme model following an analysis and elucidation of the calcium-binding site using the COACH tool, which identified residues in positions 210, 212, 213, 215, 217, 237, and 239 as being involved in calcium binding. Validation with PLIP (Supplementary data, Table S3) revealed a direct metal complexation interaction, with the calcium ion (CA) adopting a square pyramidal geometry coordinated by the main chain oxygen of Val215 (2.63 Å), the side chain hydroxyl group of Thr217 (2.63 Å), the main chain oxygen of Val237 (2.56 Å), and both carboxylate oxygens of Asp239 (2.56 and 2.51 Å).

3.12.4. Comparative analysis of docking performance and validation through molecular dynamics simulations

The docking analysis of three synthetic peptide substrates (AAPF, AAFF, and AAPM) with rSCKP revealed significant differences in binding energetics and structural stability. The results demonstrate a clear correlation between computational docking scores and experimental catalytic activity (where peptides with the best binding energies exhibited the highest enzymatic activity in vitro). The RMSD analysis across all docking poses revealed that for each ligand, only 1 out of 10 poses (10 %) had RMSD values <2 Å compared to the best-ranked pose, suggesting low conformational occupancy but a consistent top pose (Supplementary data, Table S4). This may reflect the flexibility of the binding site and limitations of rigid-body docking methods.

The AAPF emerged as the most favorable substrate, exhibiting the strongest binding affinity (−8.4 kcal/mol) in its optimal pose (Supplementary data, Table S5). This finding aligns perfectly with its observed 99 % catalytic activity. The binding energy landscape showed considerable variation (−8.4 to −6.9 kcal/mol), with RMSD values of 0.000 Å, for the best pose, and up to 26.930 Å in weaker poses, indicating multiple potential binding modes of decreasing quality (Supplementary data, Table S5).

The AAFF displayed slightly reduced but still robust binding characteristics, with its top pose scoring −8.1 kcal/mol (RMSD = 0.000 Å) (Supplementary data, Tables S3 and S5. The narrower energy range (−8.1 to −6.8 kcal/mol) and more consistent RMSD values across poses suggest a more stable interaction profile compared to AAPF. This correlates with its 98 % experimental activity, slightly below AAPF but significantly above AAPM.

The AAPM demonstrated the weakest binding profile among the three substrates, with its best pose achieving only −7.0 kcal/mol (Supplementary data, Table S3). The wider RMSD distribution in poorer poses (up to 20.341 Å) indicates greater structural variability in binding modes, consistent with its 95 % experimental activity.

MD simulations further confirmed stable substrate binding and conserved enzymatic geometry, supporting docking predictions (See Section 3.12.10).

3.12.5. Structural determinants of binding affinity

Atomic-level analysis of the top-performing ligand poses reveals critical structural determinants of substrate specificity. Tables S4 and S5 (Supplementary data) quantify the binding affinities and key interactions driving structural complementarity, while Figs. 9 and 10 provide 3D/2D visualizations and mechanistic interpretations of these specificity determinants.

3.12.6. Hydrogen bonding network analysis

The catalytic triad D27-H70-S278 (Figs. 9 and 10) forms optimal hydrogen bonds with AAPF (2.22–2.26 Å distances, 129–150° angles) (Supplementary data, Table S5), creating a geometrically perfect configuration for nucleophilic attack. AAFF maintains similar interactions (Fig. 10C) but with slightly longer bond distances (2.28–4.62 Å), while AAPM (Fig. 10B) shows the most compromised hydrogen bonding pattern (Supplementary data, Tables S6 and S5).

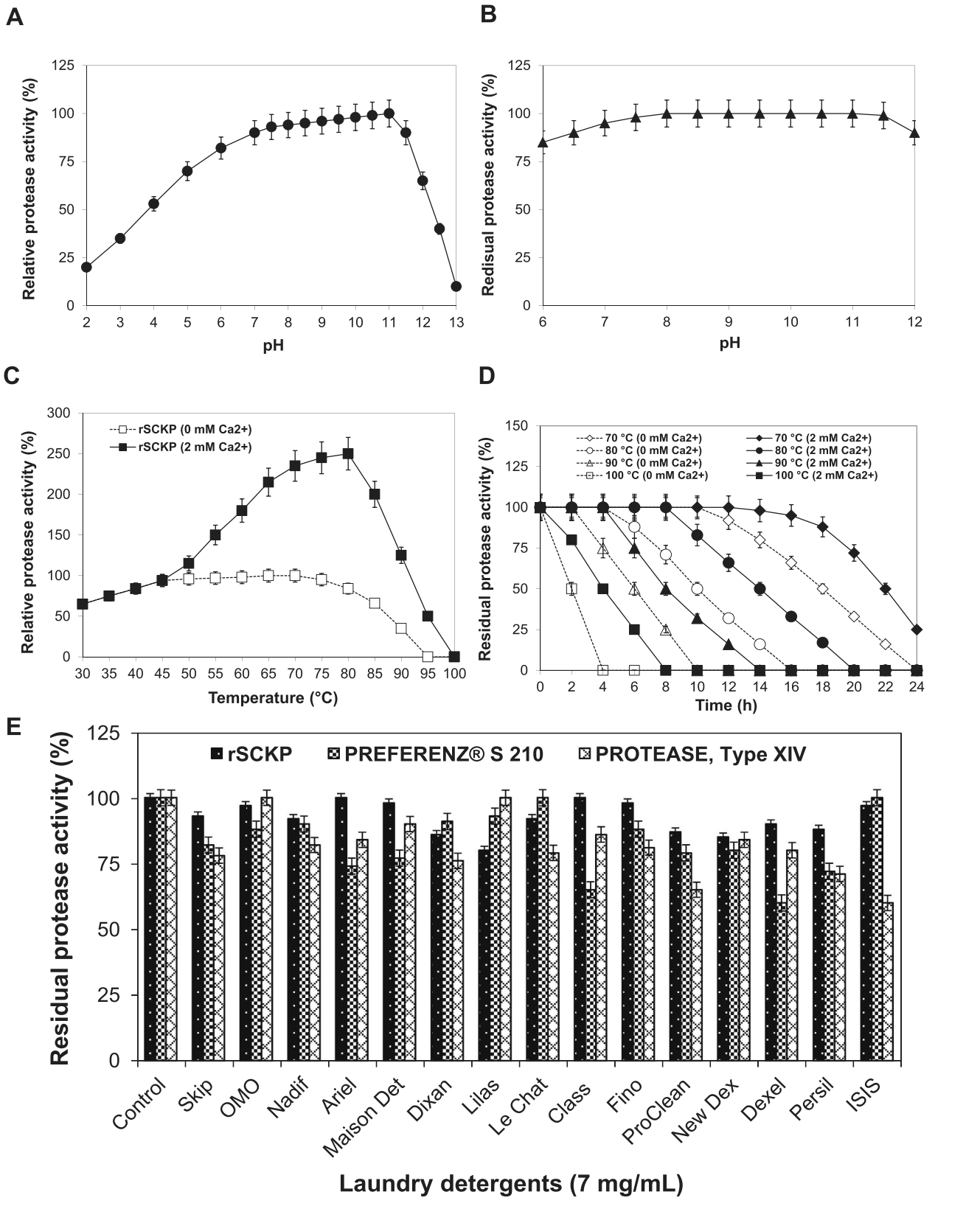

Fig. 6. Physico-chemical properties of rSCKP. The activity of the enzyme at pH 11 and 80 °C was taken as 100 %. Effects of pH on the (A) activity and (B) stability of rSCKP. Effects of the (C) thermoactivity and (D) thermostability of rSCKP. The enzyme was pre-incubated in the presence and absence of $CaCl_2$ at temperatures ranging from 70 to 100 °C. The activity of the non-heated enzyme was taken as 100 %. Each point represents the mean of three independent trials. Data represent means ± standard deviation (SD) ($n = 3$). (E) Stability of rSCKP, PREFERENZ® S 210, and PROTEASE, Type XIV with various commercial detergents formulations. The protease activity of each tested enzyme in the control trial (without detergent supplement) under identical conditions was considered as 100 %. Each point represents the mean of three independent trials. Vertical bars indicate the SD of the mean (n = 3).

Table 5
Substrate specificity of the rSCKP protease from *Streptomyces cyaneofuscatus* strain CTM50504 using various synthetic peptides (*p*NA).

Substrate (5 mM) ↓	Relative rSCKP protease activity (%)[a,b] at 410 nm
P4 - P3 - P2 - P1 - P'1	
Ac-Lys-*p*NA	0 ± 0.0
Ac-Leu-*p*NA	0 ± 0.0
Benz-Arg-*p*NA	0 ± 0.0
Suc-Phe-Leu-Phe-*p*NA	0 ± 0.0
Benz-Phe-Val-Arg-*p*NA	0 ± 0.0
Suc-Phe-Leu-Phe-*p*NA	0 ± 0.0
Benz-Phe-Val-Arg-*p*NA	0 ± 0.0
Suc-Tyr-Leu-Val-*p*NA	15 ± 0.5
Suc-Ala-Ala-Ala-*p*NA	23 ± 0.9
Suc-Ala-Pro-Ala-*p*NA	36 ± 1.0
Suc-Ala-Ala-Val-*p*NA	51 ± 1.3
Suc-Ala-Ala-Phe-*p*NA	65 ± 1.6
Suc-Ala-Ala-Val-Ala-*p*NA	68 ± 1.6
Met-Ala-Ala-Pro-Val-*p*NA	77 ± 1.8
Met-Ala-Ala-Pro-Leu-*p*NA	89 ± 2.1
Suc-Ala-Ala-Pro-Met-*p*NA	95 ± 2.3
Suc-Ala-Ala-Phe-Phe-*p*NA	98 ± 2.4
Suc-Ala-Ala-Pro-Phe-*p*NA	99 ± 2.5
Suc-Phe-Ala-Ala-Phe-*p*NA	100 ± 2.5

[a] Values represent means of five independent replicates, and ± standard errors are reported.
[b] The activity of each substrate was determined by measuring absorbance at specified wavelengths according to the relative method reported elsewhere [90].

3.12.7. Hydrophobic complementarity

As shown in Tables S4 and S5 (Supplementary data) and in Fig. 10, AAPF benefits from extensive π-π stacking interactions between its phenylalanine residues and Tyr230/Tyr133 (4.57–4.87 Å distances) (Fig. 10A). AAFF maintains similar hydrophobic contacts, though with reduced geometric perfection (Fig. 10C). The AAPM methionine side chain shows weaker hydrophobic packing (Fig. 10B), explaining its lower affinity. The hydrophobic interactions appear to contribute to complex stabilization. Mapping of the rSCKP binding site hydrophobic residues revealed a predominantly hydrophobic environment, which favors aromatic side chains such as phenylalanine (P1). This explains the strong binding affinity observed for AAPF (Supplementary data, Fig. S7).

3.12.8. Water-mediated interactions

The hydration shell analysis reveals AAPF forms three stable water bridges (WAT1–3) that enhance binding stability (Fig. 10A). AAFF maintains two water-mediated interactions (Fig. 10C), while AAPM shows only one (Fig. 10B), contributing to their respective affinity differences. Table S4 (Supplementary data) summarizes water-mediated interactions with the three ligands.

3.12.9. Mechanistic implications for catalytic efficiency

The structural data provide compelling explanations for the observed activity differences.

3.12.9.1. Transition state stabilization. The COFACTOR tool predicted that the active site residues of SCKP are located at D27, H70, N174, and S278, highlighting the importance of N174 in enzymatic catalysis (see Fig. 9A and B). N174 is suggested to play a key role in stabilizing the tetrahedral intermediate of the enzymatic reaction, thereby promoting substrate hydrolysis (see Fig. 10A, B, and C). Indeed, the AAPF optimal positioning allows N174 to perfectly stabilize the oxyanion transition state (2.26 Å H-bond). In AAFF and AAPM, this interaction becomes progressively less ideal (3.10 Å and 2.51 Å, respectively), reducing catalytic efficiency (see Supplementary data, Table S3).

3.12.9.2. Substrate positioning. An interesting finding revealed by PLIP is the presence of salt bridges between certain substrates and charged residues of the protease, notably His70 and Asp246. These interactions may influence the optimal positioning of the substrate and facilitate its hydrolysis. The conserved salt bridge with H70 (4.75 Å in AAPF), therefore, ensures proper substrate orientation. This interaction weakens in AAFF (6.03 Å) and is absent in AAPM, leading to suboptimal positioning for nucleophilic attack (see Supplementary data, Table S4).

3.12.10. Molecular dynamics and MM/PBSA for binding stability and peptide ranking confirmation

In the MD simulations, each peptide-rSCKP complex was subjected to two independent 100 ns simulations to evaluate structural stability and the reproducibility of dynamic behavior. The analysis focused on five key parameters: receptor RMSD (Fig. 11A and B), ligand RMSD (Fig. 11C), radius of gyration (Rg) (Fig. 11D), solvent-accessible surface area (SASA) (Fig. 11E), and residue-wise fluctuations (RMSF) (Fig. 11F). Mean values and standard deviations are summarized in the Table 6.

The receptor RMSD values indicated good structural stability across all complexes, ranging from 0.198 to 0.278 nm. Notably, the AAFF peptides exhibited moderate fluctuations in both replicates (0.211 and 0.230 nm), suggesting that receptor architecture remained well-conserved. In contrast, AAPM_1 showed a higher RMSD (0.278 nm), hinting at local instabilities or pronounced conformational adjustments (Fig. 11A).

Ligand RMSD values assessed the peptide's stability within the binding pocket. The lowest RMSD was observed for AAFF_1 (0.454 nm), indicative of strong, stable anchoring. Conversely, AAPM_2 had the highest RMSD (1.283 nm), reflecting greater mobility and potentially weaker or transient interactions with the binding site (Fig. 11B). These observations are further supported by the kernel density estimates of RMSD values, which highlight distinct conformational stability profiles for each peptide (Fig. 11C).

The Rg values showed minimal variation (1.954 to 1.973 nm), indicating preserved compactness of the protein across all systems. No unfolding or collapse occurred over the 100 ns timescale, supporting the structural robustness of the receptor (Fig. 11D).

SASA values revealed moderate differences. AAPF_2 had the highest solvent exposure (159.726 nm^2), suggesting a slightly more open conformation or superficial ligand interaction. AAPM_2 had the lowest SASA (156.856 nm^2), possibly due to deeper insertion of the peptide into a buried cavity (Fig. 11E).

RMSF analysis revealed regions of flexible and rigid behavior. AAPF_1 and AAPM_1 exhibited the highest residue fluctuations (0.117 and 0.116 nm), likely due to local rearrangements induced by ligand binding. AAFF_1 (0.092 nm) and AAPM_2 (0.091 nm) displayed minimal fluctuations, supporting a more stabilized receptor-ligand interaction (Fig. 11F).

Simulation reproducibility was generally high. AAFF replicates yielded closely matched data across all parameters, confirming stability and robustness. AAPF simulations also showed consistent trends, though ligand RMSD differed slightly (0.995 vs. 0.762 nm), suggesting somewhat more variable dynamics. AAPM replicates displayed greater variability, particularly in receptor (0.278 vs. 0.198 nm) and ligand RMSD (1.173 vs. 1.283 nm), indicating higher sensitivity to initial conditions and potentially less stable interactions.

Overall, AAFF and AAPF showed good reproducibility, reinforcing the reliability of the observations. AAPM exhibited greater fluctuations, warranting further energetic analysis via MM/PBSA to understand the determinants of binding stability. The MM/PBSA analysis estimated binding free energies by integrating enthalpic, entropic, and detailed energy decomposition terms (van der Waals, electrostatic, polar, and nonpolar solvation contributions). The complete data for all six simulations are presented in Table 6.

Enthalpy values (ΔH) ranged from −11.66 to −17.07 kcal/mol, indicating favorable binding interactions. Van der Waals interactions (VDWAALS) dominated, especially in AAFF_2 (−37.74 kcal/mol) and AAPM_1 (−33.22 kcal/mol), reflecting strong peptide embedding in hydrophobic receptor regions. Electrostatic interactions (EEL) varied

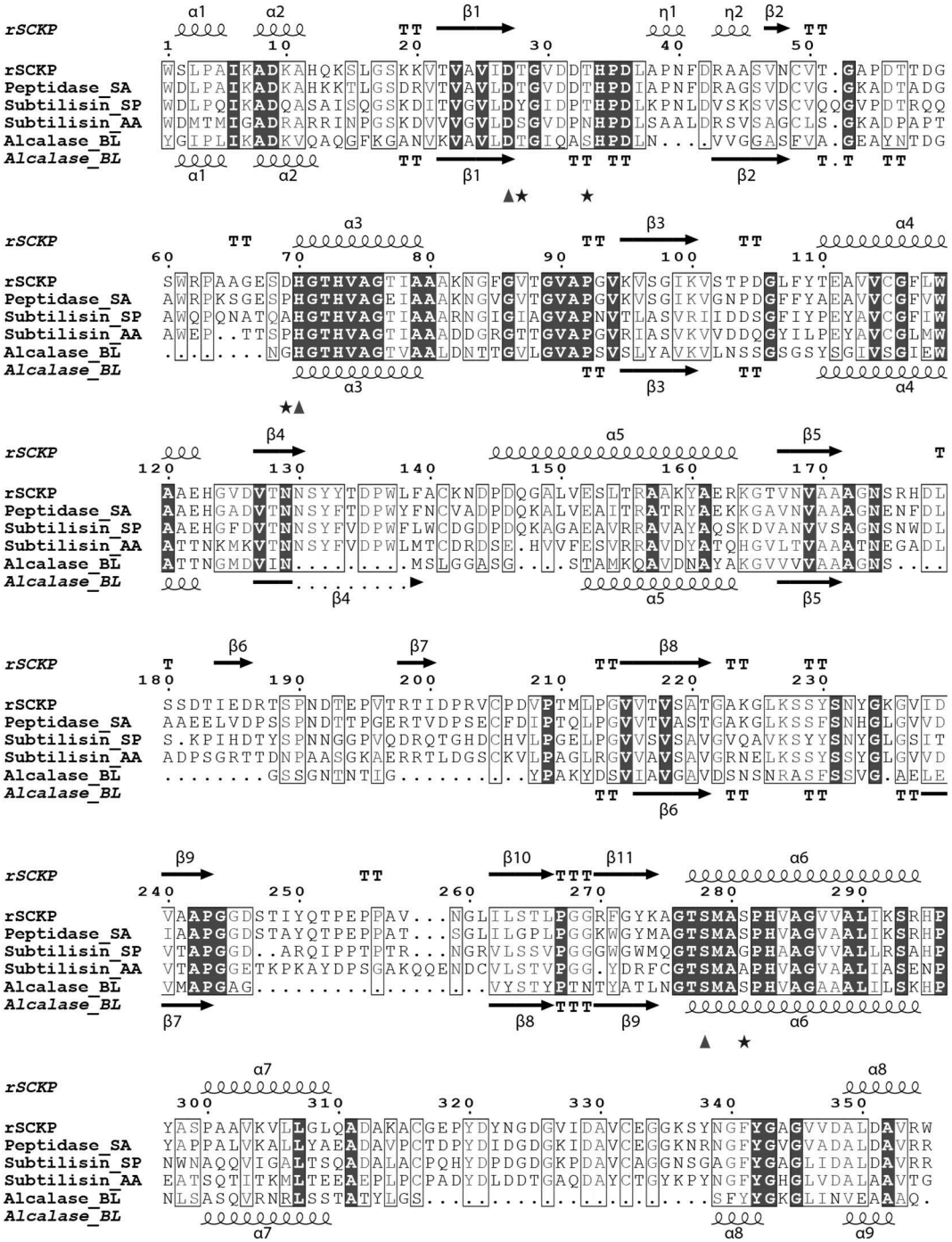

Fig. 7. Multiple alignment and structural analysis of alkaline proteases: rSCKP from *Streptomyces cyaneofuscatus* CTM50504 compared to subtilisins from *Streptomyces albus* ATCC21838, *Streptomyces phatthalungensis*, *Amycolatopsis algeriensis*, and Alcalase from *Bacillus licheniformis*. – Red triangles (▲) for the catalytic triad (D27, H70, and S278) of rSCKP, and blue stars (★) for non-similar residues (T28, T33, D69, and S281) potentially impacting the structure-function of rSCKP – Secondary structures of rSCKP (AlphaFold 2 generated model) and Alcalase (crystal structure, PDB ID: 1BE6) shown for comparison. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

widely, reaching up to 80 kcal/mol in AAPM_2 and AAPF_2, suggesting major charge reorganization upon binding (Table 7).

Polar solvation energies (EPB), being positive and high, reflected the energetic cost of exposing charged groups in water. Nonpolar solvation (ENPOLAR) was modest (~ −3 to −4 kcal/mol), partially offsetting the polar penalty. Total solvation energies (GSOLV) ranged from −28 to −67 kcal/mol, with the most favorable values in AAPM_2 and AAPF_1 (Table 7).

Gas-phase interaction energies (GGAS = VDWAALS + EEL) showed that AAPF_2 (50.32 kcal/mol) and AAPM_2 (55.53 kcal/mol) had energetically rich complexes, but this benefit was mitigated by solvation penalties.

The entropic contribution (-TΔS) revealed large differences. AAFF_1 and AAPF_2 showed the highest entropy penalties (58.8 and 42.2 kcal/mol), which significantly reduced their thermodynamic favorability (Table 7). These penalties may reflect reduced flexibility at the binding interface or internal ligand constraints.

Replicate consistency was generally good, with ΔH, GSOLV, and

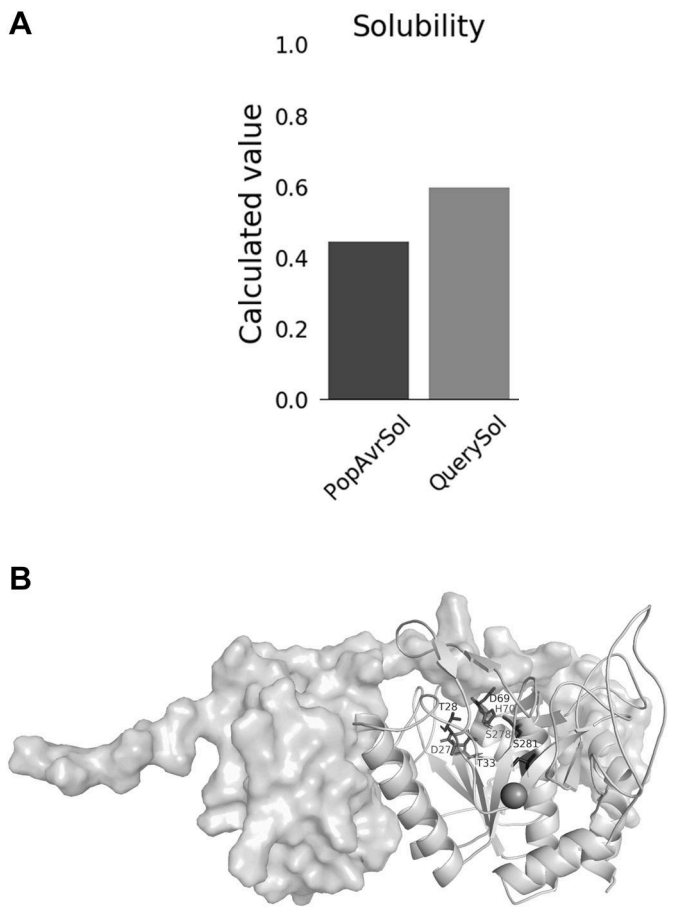

Fig. 8. Structural insights into rSCKP catalysis: key residues, water networks, and ligand binding modes. (A–B) 3D representation of rSCKP highlighting catalytic residues (D27, H70, and S278 in *red sticks*; and N174 in *magenta sticks*) and structural waters (WAT1–3 as *blue spheres*), shown within (A) a *solvent-accessible surface* and (B) the *full macromolecular structure* (calcium ion as *orange sphere*). (C—D) Positioning of ligands AAPF (cyan), AAFF (yellow), and AAPM (green) in the catalytic cavity, depicted as (C) a *solvent-accessible surface* and (D) *ribbon secondary structure*. All figures were generated with PyMOL. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

GGAS values showing acceptable standard deviations. Some fluctuations in EPB and EEL were noted, likely reflecting structural flexibility and solvent exposure.

Importantly, binding free energy (ΔG) depends not just on enthalpy, but also on solvation and entropy. AAPF_1, despite having the most favorable enthalpy (−17.07 kcal/mol), showed a moderate ΔG (15.43 kcal/mol) due to high entropy cost (Table 7).

AAFF_2 emerged as a strong candidate, combining high van der Waals interactions, low entropy penalty (-TΔS = 23.11 kcal/mol), and favorable ΔG (6.66 kcal/mol). It also had the least unfavorable ligand internal energy (E_ligand = −2.36 kcal/mol), suggesting low strain upon binding, and the highest number of interacting receptor residues (nAA = 9), indicative of a broad and stable interface (Table 7).

In contrast, AAPF_1 and AAPM_1 had much more negative E_ligand values (around −8 kcal/mol), implying greater conformational cost. AAPF_1 and AAPM_2 had the lowest nAA values (4), reflecting more superficial or limited contact with the receptor. These two indicators, E_ligand and nAA, complement each other by revealing both the conformational stability of the ligand and the extent of its interaction interface. AAFF_2 clearly stands out as the best-performing complex in both respects.

4. Discussion

As part of the Horizon Europe, Coordination and support action (CSA) Twinning project, NGS-4-ECOPROD, https://cordis.europa.eu/project/id/101079425, we became interested in addressing the networking gaps in the field of NGS for the development of eco-friendly biotech-products, particularly extremozymes. The diversity of hot spring microbial enzymes is enormous, and they hold great potential for use in biotechnology-based companies. Exploring the synergy between NGS, biochemistry, biomolecular techniques, and molecular docking highlights the collective potential in unravelling the intricacies of hot spring extremozyme activity and application.

The hot spring biotopes harbour various microorganisms, such as bacteria, fungi, and archaea [48–50]. The hydrolytic enzymes from extremophilic actinomycetesare of immense use in basic and applied research [51]. In this study, an isolate designated as strain CTM50504 was isolated from a hot spring environment and was shown to produce extracellular enzymes, including lipases, phospholipases, proteases, amylases, and chitinases. These enzymes, especially as extremozymes, are crucial in several industries besides academic and biotechnological research. Results from 16S rRNA gene sequence and phylogenetic analyses revealed that strain CTM50504 is closely related to *S. cyaneofuscatus* NRRL B-2570^T with 99.93 % sequence similarity (Fig. 1). The phylogenetic position of strain CTM50504 was further confirmed via phylogenomics analysis (Fig. 4) [34]. The genome-wide Average Nucleotide Identity (gANI) value between *S. cyaneofuscatus* CTM50504 and related *Streptomyces* strains further confirmed the genomic relatedness of strain CTM50504 to *S. cyaneofuscatus* NRRL B-2570^T. In conjunction with digital DNA:DNA hybridization, this method is used for taxonomic resolution among closely related genomes (80–100 % ANI) [52]. However, the present study mainly focused on discovering extremozymes from *S. cyaneofuscatus* CTM50504, and whole genome sequence analysis can more comprehensively deepen our understanding of molecular mechanisms and genetic characteristics [53,54].

To better understand the ability of strain CTM50504 to produce the enzymes of interest, whole genome sequencing was performed using the NextSeq™ 500 platform.

Among bacteria, *Streptomyces* species are well-known for their relatively large, linear genomes (8–10 Mb) and remarkably high G + C% (>70 %). The genome of strain CTM50504 comprises one chromosome of 8,59 Mbp with 71 % GC content and is predicted to contain 8100 open reading frames (ORFs). It has been reported that genomes with a high G + C content are generally large and possess a large number of ORFs [55]. The type strain of *S. cyaneofuscatus* was initially described by Pridham et al. in 1958 [56]. The G + C content of the type-strain genome is 71.6 %, and its approximate size is 7,90 Mbp. In the genome sequencing of *S. cyaneofuscatus* CTM50504, 8100 protein-coding sequences (CDSs) were found. The proteins with functional assignments included 770 proteins with EC numbers, 6159 with Gene Ontology (GO) assignments, and 2625 proteins mapped to KEGG pathways.

Members of the genus *Streptomyces* are known to produce a wide range of secondary metabolites with antibacterial, antitumor, and antiparasitic properties (to name a few). The whole genome analysis of strain CTM50504 revealed 49 predicted biosynthetic gene clusters encoding for secondary metabolites, such as Non-Ribosomal Peptide Synthetases (NRPS) and Polyketide Synthases (PKS). Sharma et al. [57] showed that *Streptomyces* species possessing antifungal activity are positively related to the presence of NRPS and PKS biosynthetic pathways, highlighting the potential of other strains that contain NRPS and PKS gene clusters to also produce antifungal agents [57].

Nine gene clusters showed 100 % similarity with known gene clusters linked to the synthesis of NRP-metallophore, NRPS, PKS-like, T1PKS, *trans*AT-PKS, ectoine, terpene, melanin, and NI-siderophore in *Streptomyces* strains [58]. Specifically, some gene clusters in the genome of *S. cyaneofuscatus* CTM50504 revealed a low similarity percentage

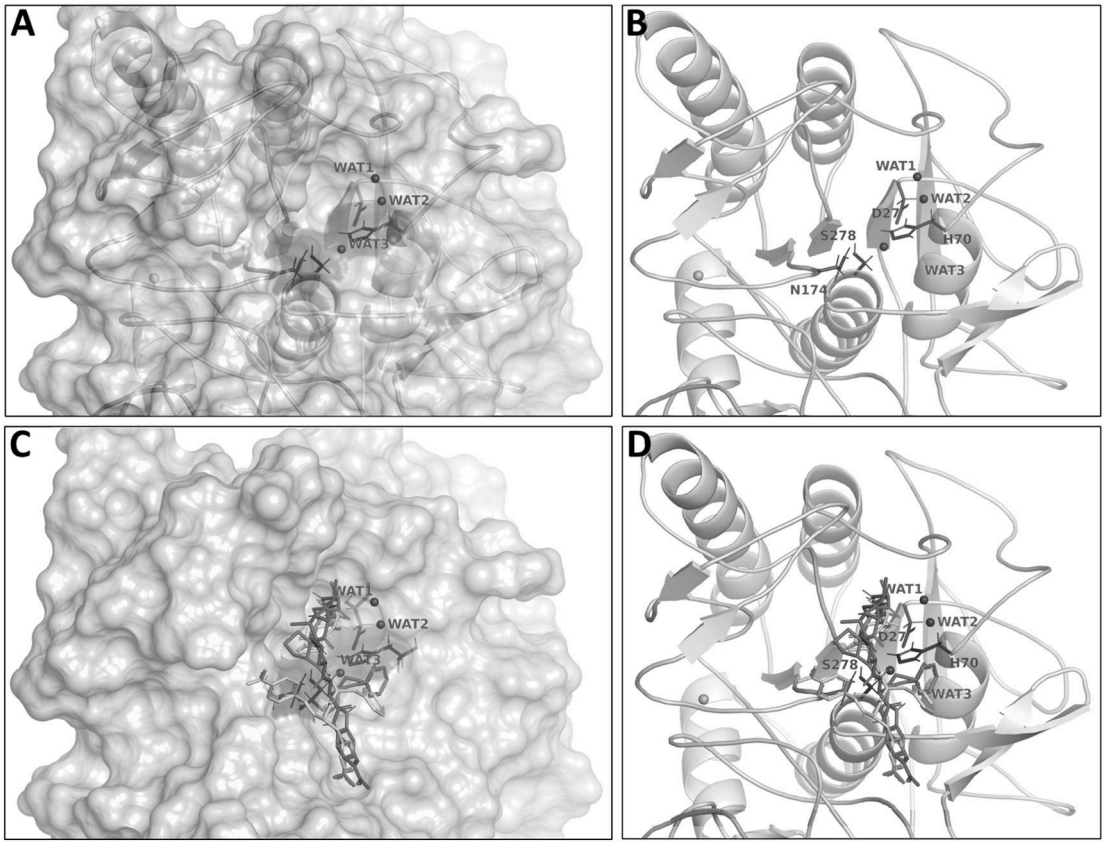

Fig. 9. Calculated solubility value of rSCKP (QuerySol) compared to the average of 0.45 for soluble proteins from *E. coli* (PopAvrSol) and structural model of rSCKP protease generated by AlphaFold 2, highlighting the catalytic site and the pro-peptide. (A) Mature rSCKP shows a predicted normalized solubility of 0.599. The original figure was generated by the Protein-Sol tool. (B) Secondary structures are depicted in distinct colors: helices in light blue, sheets in beige, and loops in light gray. The catalytic residues D27, H70, and S278 are highlighted in red. Non-similar residues (T28, T33, D69, and S281) potentially impacting the structure-function of rSCKP are highlighted in blue. The pro-peptide, essential for regulating enzymatic activity, is represented as a surface to emphasize its role in inhibiting the protease before maturation. This visualization provides insight into the structural organization of the protease and the key elements of its catalytic mechanism. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with other known clusters. This highlights the potential for new secondary metabolites to be isolated from the fermentation of *S. cyaneofuscatus* CTM50504. In this context, genome mining has proven to be a fundamental tool for genome-based natural product discovery and has previously guided the discovery of novel natural products from several *Actinomycetota* [59–61].

Various extracellular enzymes secreted by *Streptomyces* species, such as cellulases, amylases, proteases, and lipases, are involved in biomass degradation [62,63]. These enzymes support the complex life cycle of these bacteria, which can be exploited for application in biotechnology-based research, such as biocatalysis and bioremediation [53]. Enzymes as biocatalysts to synthesize chemicals in various industries, such as food, pharmaceutical, fuel, and fine chemicals, have substantial economic and environmental benefits [64,65].

Hot springs are one of the primary sources of microbial-derived industrially important enzymes [48,50]. These enzymes are typically stable and active even at temperatures higher than the optimal growth temperature of the enzyme-producing strain, showing potential for numerous industrial applications. Furthermore, thermostable enzymes have been reported to be more stable in the presence of solvents and detergents and can function under extreme conditions of pressure or acidity [66]. An analysis of the CTM50504 genome showed the presence of genes encoding industrially important enzymes. The results show that seven functions are annotated, such as transferase, oxidoreductase, lyase, ligase, isomerase, and translocase (Fig. 7), with the significant category being hydrolase enzymes (41,94 %). Hydrolases are a diverse group of enzymes that catalyse the hydrolysis of various chemical bonds, including ester, amide, glycosidic, and peptide bonds. Based on their catalytic mechanism, hydrolases are classified into several families: lipases, proteases, glycosidases, and esterases.

Hydrolases are enzymes that catalyse the hydrolysis of various substrates. They are classified into different families based on functional properties and sequence similarity. *Streptomyces* species have been found to produce hydrolases from several families, including lipases (EC 3.1.1), proteases (EC 3.4.21–24), and cellulases (EC 3.2.1). These enzymes play a crucial role in the degradation of complex biomolecules and have immense biotechnological significance. For instance, *S. coelicolor* produces a thermostable lipase that can withstand high temperatures, making it suitable for industrial applications in detergent formulation and biofuel production [67].

Similarly, *Streptomyces* sp. strain AB1 produces a keratinolytic alkaline proteinase (KERAB) that exhibits remarkable stability under extreme pH conditions, making it an ideal candidate for future applications in detergent formulations, dehairing during leather processing, and non-aqueous peptide biocatalysis [68]. On the other hand, *S. lividans* is known for producing cellulases capable of efficiently degrading cellulose, offering potential in biofuel production and waste management [69]. The hydrolases produced by *Streptomyces* species have found numerous applications in various industries. Lipases have been extensively utilized in the food, pharmaceutical, and detergent industries due to their ability to catalyse the hydrolysis and synthesis of ester bonds [70]. Proteases have been employed in cheese production, meat tenderization, and removing protein-based stains [71]. Similarly, cellulases, capable of breaking down cellulose into glucose, have

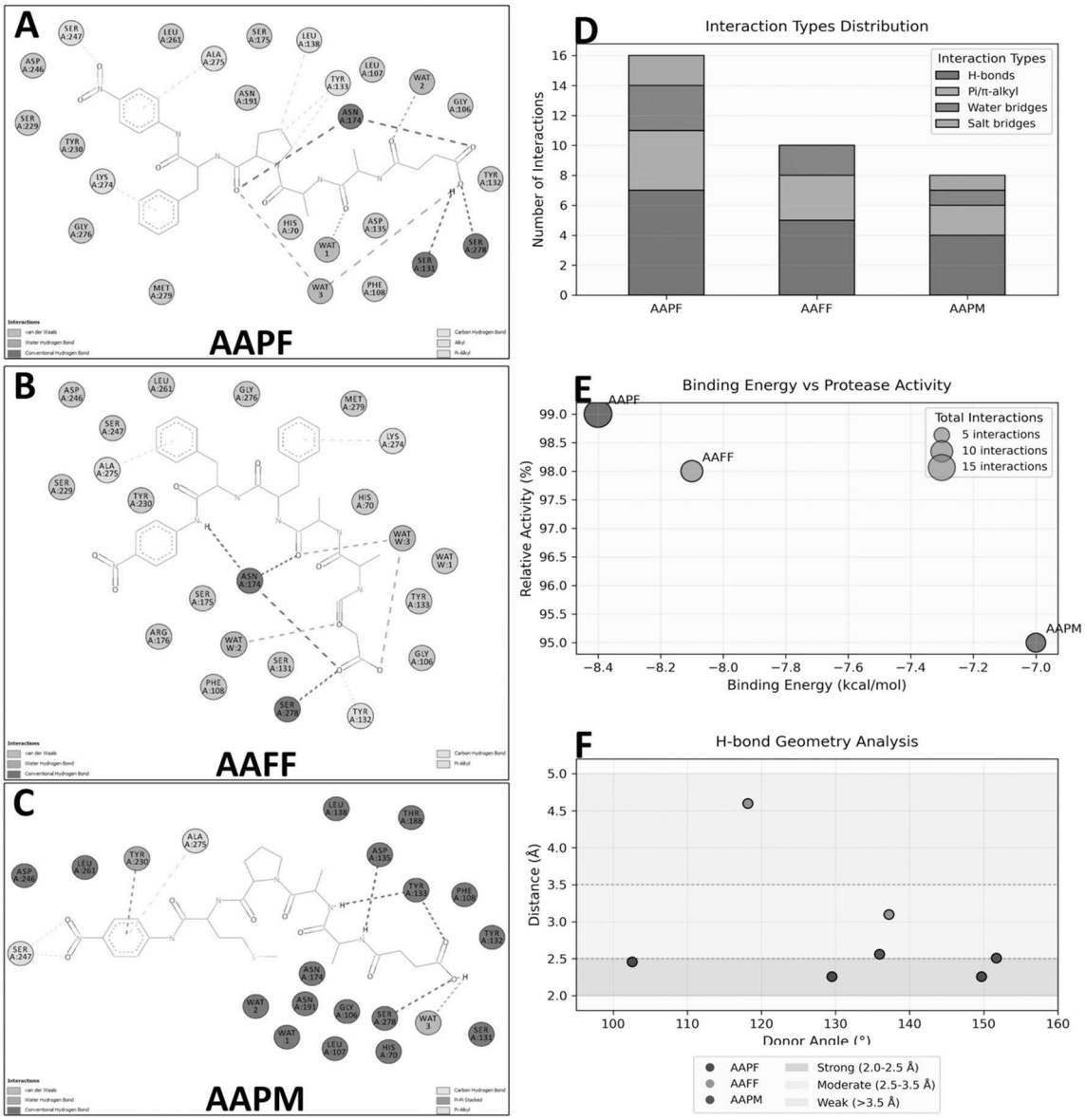

Fig. 10. Computational analysis of ligand docking with rSCKP: 2D interaction diagram of ligands (sticks). (A) AAPF, (B) AAFF, and (C) AAPM with rSCKP (residue circles): Hydrogen bonds (green/light green/cyan dashes), π-π stacked (magenta), π-alkyl (pink), and Van der Waals interactions (no dashes). Images were generated using Biovia Discovery Studio 2020. (D) Distribution of molecular interaction types (hydrogen bonds, π/π-alkyl interactions, water bridges, and salt bridges) for each ligand. (E) Correlation between the binding energy (ΔG) and relative activity, with point sizes scaled to total interaction counts. (F) Hydrogen bond geometry analysis showing donor angle-distance relationship, colour-coded by interaction strength (green: strong [2.0–2.5 Å], gold: moderate [2.5–3.5 Å], salmon: weak [>3.5 Å]). Ligands are shown in distinct colors (AAPF: blue, AAFF: orange, and AAPM: green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

immense potential in the biofuel industry and the production of paper and textiles [72]. In this study, genome mining revealed the presence of 268 enzymes relevant to biotechnology. These include 179 proteases, 59 glycosyltransferases, 5 amylases, 5 chitinases, and 20 lipases (Fig. 9). Within the lipase family, we identified 2 phospholipase A2, 9 triacylglycerol lipases, 2 phospholipases C, 5 lysophospholipases, sphingomyelin phosphodiesterase, and phospholipase D-like domain-containing protein.

The high number of proteases in the genome can be explained by the great variety of proteins with different structures found in this group, evidenced by the diversity of conserved domains found in proteases in the *Streptomyces* genome. It can also be explained by the ubiquity of proteins in the hot spring environment [63] and by the importance of proteases in the general functioning of microbial cells [73]. Among the proteases found, we highlight those involved in metabolic systems to repair and maintain proteins, such as enzymes containing an ATP-dependent Lon domain [74].

Glycosyltransferases are enzymes that play a crucial role in the biosynthesis of carbohydrates [75]. They catalyse the transfer of a sugar moiety from a donor molecule onto an acceptor molecule, forming glycosidic bonds. These enzymes are essential for bacteria as they are involved in the biosynthesis of exopolysaccharides, which contribute to cell surface properties and cell interactions. This enzyme is among the most represented enzymes in the CTM50504 genome. Furthermore, glycosyltransferases have significant utility in biotechnology. They are used to produce various products in the food manufacturing, cosmetics, and animal nutrition industries and are valuable tools in synthetic carbohydrate chemistry. With advances in molecular biology and biotechnology, more than 30 glycosyltransferases have been cloned and are now readily available in sufficient quantities for in vitro synthesis

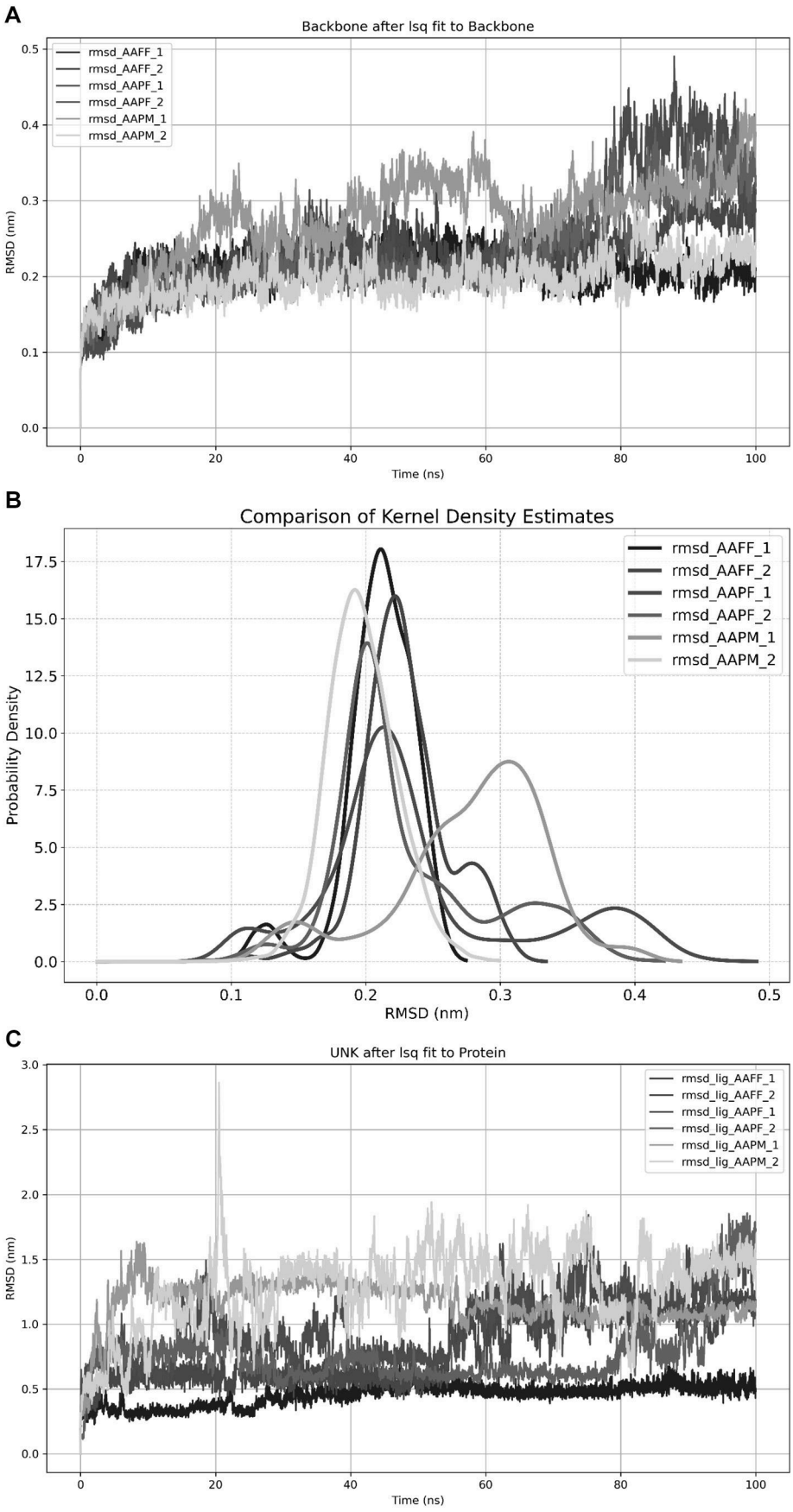

Fig. 11. Molecular dynamics simulations explained rSCKP's substrate binding selectivity. (A) Backbone RMSD Stability After LSQ Fit to Protein Backbone. Cα atom deviations (nm) over time, derived from GROMACS 2025.0 simulations. The curves represent simulation replicates for each peptide: AAFF (dark blue and blue),

AAPF (dark green and green), and AAPM (yellow and light green). Data obtained from GROMACS 2025.0 simulations. (B) Kernel Density Estimates of Conformational RMSD. Probability densities of peptide RMSD (0.1–0.5 nm) showing stability differences. The curves represent simulation replicates for each peptide: AAFF (dark blue and blue), AAPF (dark green and green), and AAPM (yellow and light green). Data obtained from GROMACS 2025.0 simulations. (C) Ligand RMSD after alignment to the protein. Root-mean-square deviation (RMSD, in nm) of peptide ligands over time (0–100 ns) following least-squares (LSQ) fitting to the binding site backbone. The curves represent simulation replicates for each peptide: AAFF (dark blue and blue), AAPF (dark green and green), and AAPM (yellow and light green). RMSD values >1.0 nm (AAPM) indicate partially unstable binding, whereas values <0.5 nm (AAFF) reflect a stable interaction. Data obtained from GROMACS 2025.0 simulations (CHARMM36m, TIP3P). (D). Total and Axial Radius of Gyration Over Time. Rg (nm) trends from 0 to 100 ns, computed using GROMACS tools under NPT conditions (1 bar, 310 K). The curves represent simulation replicates for each peptide: AAFF (dark blue and blue), AAPF (dark green and green), and AAPM (yellow and light green). Data obtained from GROMACS 2025.0 simulations. (E) Kernel Density Estimates of Solvent-Accessible Surface Area (SASA) for Peptide-rSCKP Complexes. Probability distributions of SASA (nm^2) across replicates, calculated from 100 ns CHARMM36m/TIP3P trajectories. The curves represent simulation replicates for each peptide: AAFF (dark blue and blue), AAPF (dark green and green), and AAPM (yellow and light green). Data obtained from GROMACS 2025.0 simulations. (F) Per-Residue Fluctuations (RMSF) Across Peptide Complexes. Residue-wise flexibility (nm) averaged over replicates. The curves represent simulation replicates for each peptide: AAFF (dark blue and blue), AAPF (dark green and green), and AAPM (yellow and light green). Data obtained from GROMACS 2025.0 simulations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

[76].

This study is the first to target the hydrolase repertoire of a *S. cyaneofuscatus* strain, providing insight into the tremendous biotechnological potential of the enzymes identified. Different thermostable enzymes, such as α-amylase and chitinase, were predicted in the CTM50504 genome. These enzymes have been used extensively in industrial processes. These thermophilic and hyperthermophilic enzymes are part of the enzyme category called extremozymes. These enzymes can function under extreme conditions such as high salt levels, alkaline conditions, extreme pressure, or acidity [77]. In addition, many lipases were predicted to be present in the CTM50504 genome. Their physiological function is still unknown. However, they are considered valuable biocatalysts since they are active and highly enantioselective on industrially important or rare compounds not accepted as substrates by standard lipolytic enzymes. Some examples of these compounds are bulky esters of tertiary alcohols for pharmaceutical applications [78,79], (+)-methyl acetate for the food industry and the production of perfumes [80], aryl-carboxylic acid esters for the production of aromas or preservatives, removal of environmental pollutants, or the selective removal of protecting groups [81]. Besides, family VIII enzymes are often remarkably stable and even exhibit enhanced activity in the presence of organic solvents [82].

The deduced amino acid sequences for the protease, phospholipase, lipase, amylase, and chitinase activities exhibited high similarity with enzymes from *Streptomyces* strains. For instance, thermostable lipase from *Streptomyces exfoliatus* strain DSMZ 41693 and related species have demonstrated activity at elevated temperatures (55–60 °C) and alkaline pH, making them suitable for applications in biodiesel production, detergents, and enantioselective synthesis in the pharmaceutical industry [83,84]. Phospholipases from *Streptomyces violaceoruber* A-2688 and *Streptomyces racemochromogenes* 10–3 have been reported to catalyse lipid modifications with industrial relevance in oil degumming, food emulsification, and the synthesis of phospholipid derivatives [85]. Chitinases from thermophilic species such as *Streptomyces thermoviolaceus* strain OPC-520 exhibit high stability and activity on chitin-rich substrates at elevated temperatures, with promising applications in biocontrol, chitin waste valorization, and the production of chitooligosaccharides with biomedical properties [86]. Furthermore, thermostable and alkaline-tolerant α-amylases from *Streptomyces* sp. strain SLBA-08 and *Streptomyces erumpens* MTCC 7317 have been shown to efficiently hydrolyze starch under industrial conditions relevant to food processing and detergent formulations [87,88]. The co-occurrence of these enzymes with rSCKP in the same genome highlights the potential of *S. cyaneofuscatus* strain CTM50504 as a source of robust, multifunctional biocatalysts suitable for a wide range of environmentally friendly industrial processes.

The rSCKP has the highest sequence identity (67 %) to the peptidase from *Amycolatopsis alba* (TR: A0A3R9DTQ9) (listed in Supplementary data, Table S2). The CTM50504 phospholipase sequence showed an identity of 96 % with the phospholipase from *Streptomyces arboris* with an 18-residue difference. The lipase from the strain CTM50504 has a sequence identity of 96 % with the lipase of *Streptomyces microflavus* with a 6-residue difference, while the CTM50504 amylase sequence showed a sequence identity of 97 % with the amylase from *Streptomyces* sp. MT29 with a 16-residue difference. Lastly, the chitinase from the *Streptomyces cyaneofuscatus* strain CTM50504 has an identity of 98 % with the lipase of *S. arboris* with an 18-residue difference.

rSCKP from strain CTM50504 features a 34 amino acid signal peptide (spanning from Met to Ala) that displays hallmark characteristics of subtilisins: a short positively charged N-terminal region (comprising one Arg and two Lys), followed by a highly hydrophobic stretch of approximately 20 amino acids, and a C-terminal sequence (Ala-Ala-Pro) that is cleaved by signal peptidase during maturation. The pro-peptide sequence comprises the following 123 amino acids, with a theoretical mass of 12 kDa. This domain is thought to function not only as a competitive inhibitor of the rSCKP enzyme but also plays a crucial role in ensuring proper folding of the protein. Once folding is complete, subtilisin initiates intramolecular degradation of its chaperone domain, leading to autoproteolytic processing and the formation of the mature, active enzyme. The resulting subtilisin consists of 274 amino acids (from Trp1 to Trp355) and has a predicted molecular mass of 36 kDa.

Compared to previous data [40,51], the rSCKP specific activity was higher when the enzyme expression was under the control of the strong promoter T7 (pSM202) than the Ptrc and lac. The rSCKP enzyme consists of a single band of ~45 kDa. The difference is attributed to the pET-28b polylinker extension of hexahistidine and the post-translational changes of the protein that had occurred, such as glycosylation. The activity of rSCKP towards synthetic substrates confirmed its classification as a serine peptidase.

All the observed results indicate that the rSCKP protease possesses high thermostability under alkaline conditions. As detailed before, an enzyme appropriate for detergent applications should exhibit activity within a pH range of 8 to 12 [48]. Therefore, rSCKP shows strong potential for use as a bio additive in both solid and liquid laundry detergents. Additionally, rSCKP exhibits superior biochemical properties compared to many reported commercial basophilic proteases from *Streptomyces* and related extremophilic microorganisms, particularly in terms of its high optimal temperature, alkaline pH preference (pH 11), and remarkable thermal stability, highlighting its strong potential for industrial applications under harsh processing conditions (Supplementary data, Table S6). The rSCKP enzyme's theoretical physico-chemical parameters are particularly advantageous, as high solubility is a key factor for industrial applications, especially in enzymatic formulations where protein precipitation can be a major obstacle. Moreover, the significant similarities with proteases of the S9 family, particularly those from the genus *Streptomyces,* suggest that rSCKP shares functional and structural characteristics with other enzymes of this family, known for their robust catalytic activity and wide range of industrial applications. The instability index indicates that rSCKP is stable, while the GRAVY score suggests predominant hydrophilicity. The EMBOSS pepstats of rSCKP indicated good potential solubility. The predominance of β-strands suggests that rSCKP adopts a folding pattern characteristic of subtilases, with a central β-sheet core, which is consistent with known crystallographic structures of this enzyme family.

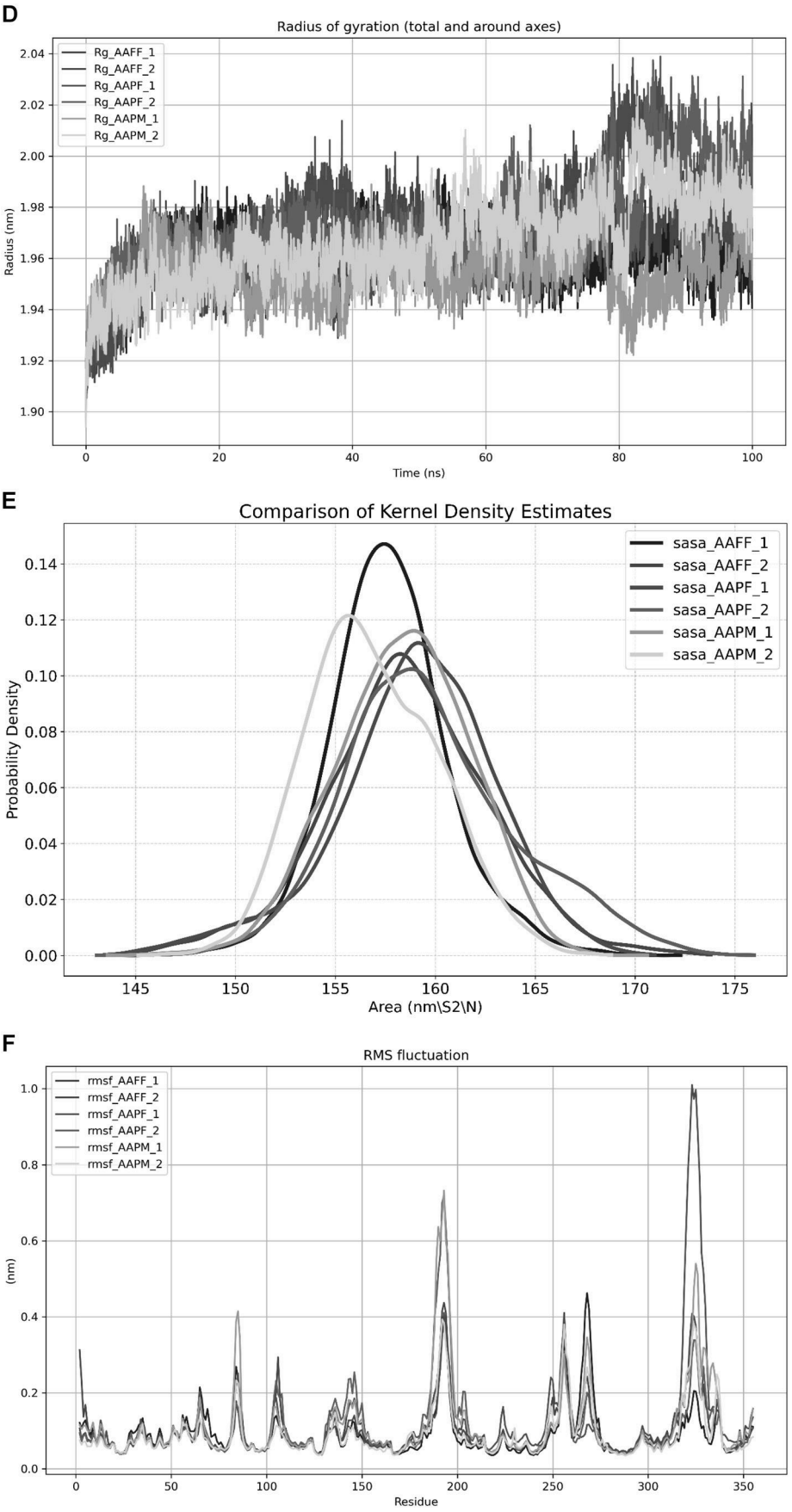

Fig. 11. (*continued*).

Table 6
Binding free energy components and thermodynamic parameters of peptide–ligand complexes calculated by MM/PBSA analysis.

Peptide	ΔH ± SD	-TΔS ± SD	ΔG ± SD	E_ligand ± SD	nAA	VDW ± SD	EEL ± SD	EPB ± SD	ENPOLAR ± SD	GGAS ± SD	GSOLV ± SD
AAPF_1	−17.07 ± 8.58	32.50 ± 0.05	15.43 ± 8.58	−8.14 ± 6.10	4	−25.39 ± 7.31	72.63 ± 27.86	−61.19 ± 25.36	−3.12 ± 0.69	47.24 ± 25.93	−64.31 ± 25.17
AAPF_2	−15.43 ± 9.02	42.22 ± 0.05	26.79 ± 9.02	−6.83 ± 6.44	6	−29.23 ± 5.59	79.55 ± 51.36	−61.98 ± 52.49	−3.78 ± 0.58	50.32 ± 49.38	−65.76 ± 52.23
AAFF_1	−16.21 ± 6.60	58.80 ± 0.05	42.59 ± 6.60	−5.59 ± 3.77	6	−29.77 ± 8.11	59.26 ± 27.02	−41.99 ± 30.13	−3.70 ± 1.00	29.49 ± 32.07	−45.70 ± 29.39
AAFF_2	−16.45 ± 6.47	23.11 ± 0.05	6.66 ± 6.47	−2.36 ± 4.52	9	−37.74 ± 3.90	49.32 ± 15.66	−23.39 ± 14.44	−4.64 ± 0.34	11.57 ± 14.95	−28.02 ± 14.40
AAPM_1	−12.86 ± 6.77	32.95 ± 0.05	20.09 ± 6.78	−8.36 ± 2.18	8	−33.22 ± 6.84	77.07 ± 30.27	−52.68 ± 28.83	−4.03 ± 0.77	43.85 ± 30.80	−56.71 ± 28.52
AAPM_2	−11.66 ± 6.28	28.91 ± 0.96	17.25 ± 6.35	−6.53 ± 3.25	4	−25.23 ± 6.96	80.76 ± 38.06	−64.22 ± 35.18	−2.97 ± 0.80	55.53 ± 40.44	−67.19 ± 34.77

The MolProbity analysis of the 3D model confirms the validity and stability of the predicted model. Additionally, the agreement with reference subtilase structures reinforces the model's reliability and supports further studies on molecular dynamics and enzyme-substrate interactions.

The PLIP results confirmed the COACH-predicted binding site and highlighted the critical roles of Val215, Thr217, Val237, and Asp239 in forming a stable calcium-binding pocket through mixed main chain and side chain interactions. The analysis of binding energy scores from molecular docking reveals a strong affinity between the rSCKP protease and the tested peptides, with values ranging from −7.0 to −8.4 kcal/mol. The COACH tool and binding energy analysis scores were consistent with the experimental observations, where peptides with the best binding energies also exhibited the highest enzymatic activity in vitro.

Examination of molecular interactions highlights several key aspects explaining the study findings. The AAPF peptide, which exhibited the most favorable binding energy (−8.4 kcal/mol) and the highest relative activity (99 ± 2.5 %), forms multiple stabilizing interactions with rSCKP. It establishes conventional hydrogen bonds with S131, N174, and S278, hydrophobic interactions with L138 and Y133, as well as π-alkyl contacts with K274 and A275. Furthermore, this peptide interacts with three water molecules (WAT1, WAT2, and WAT3), enhancing complex stability via hydrogen bonding involving D105, H70, and S68. The COFACTOR tool predicted that the active site residues are located at D27, H70, N174, and S278, highlighting the importance of N174 in enzymatic catalysis. N174 is suggested to play a key role in stabilizing the tetrahedral intermediate of the enzymatic reaction, thereby promoting substrate hydrolysis. This information was used to optimize ligand positioning during molecular docking.

The AAFF peptide, with a slightly lower binding energy (−8.1 kcal/mol) and enzymatic activity of 98 ± 2.4 %, also forms stabilizing hydrogen interactions with N174 and S278, and π-alkyl interactions with K274 and A275. However, it exhibits weaker interaction with the hydration environment, which may explain its slightly reduced affinity.

Conversely, the AAPM peptide, with a binding energy of −7.0 kcal/mol and enzymatic activity of 95 ± 2.3 %, displays stabilizing interactions with Y133, D135, and S278, along with a π-π stacked interaction with Y230. However, it interacts with only one water molecule (WAT3), which might limit its stabilization within the active site.

The specific role of N174 in catalysis is particularly intriguing. Its involvement in multiple hydrogen interactions with peptides suggests that it may play a crucial role in substrate positioning and transition state stabilization. Additionally, its proximity to the catalytic triad could facilitate the nucleophilic activation of S278, thereby enhancing enzymatic hydrolysis efficiency. These observations align with the COFACTOR predictions, which identified N174 as a key residue of the active site.

The involvement of water molecules in rSCKP-substrate interactions, particularly through hydrogen bonds with Ser68 and His70, suggests a potential role of solvation in the catalytic mechanism since these are peptide-hydrolysis reactions.

Although molecular docking initially suggested a stronger binding affinity for AAPF compared to AAFF, the MD/MM-PBSA analysis revealed a lower binding free energy for AAFF. This apparent discrepancy reflects the complementary nature of the two approaches: docking captures the static enthalpic fit of the ligand in the binding site, while molecular dynamics accounts for conformational flexibility, solvation, and entropic contributions. Notably, the experimentally measured enzymatic activities for AAPF (99 %) and AAFF (98 %) were nearly identical, reinforcing the conclusion that both peptides are efficiently recognized by rSCKP. The combination of these computational and biochemical findings supports the robustness of the predicted substrate preferences and highlights the dynamic stability of AAFF as a key factor in its favorable binding profile. In summary, considering binding affinity, energetic stability, and reproducibility across simulations, the ranking AAFF > AAPF > AAPM appears to be the most robust. These

Table 7
Molecular dynamics parameters of peptide complexes with the rSCKP protease from *Streptomyces cyaneofuscatus* strain CTM50504, including protein RMSD, ligand RMSD, radius of gyration (Rg), solvent-accessible surface area (SASA), and residue-wise RMSF.

Peptide	RMSD Prot (nm)	RMSD Lig (nm)	Rg (nm)	SASA (nm^2)	RMSF (nm)
AAFF_1	0.211 ± 0.027	0.454 ± 0.082	1.961 ± 0.011	157.699 ± 2.962	0.092 ± 0.064
AAFF_2	0.230 ± 0.033	0.805 ± 0.261	1.966 ± 0.014	158.580 ± 4.156	0.103 ± 0.071
AAPF_1	0.240 ± 0.079	0.995 ± 0.257	1.968 ± 0.021	158.976 ± 4.016	0.117 ± 0.136
AAPF_2	0.232 ± 0.058	0.762 ± 0.252	1.973 ± 0.018	159.726 ± 4.368	0.106 ± 0.092
AAPM_1	0.278 ± 0.058	1.173 ± 0.160	1.954 ± 0.011	158.357 ± 3.372	0.116 ± 0.107
AAPM_2	0.198 ± 0.026	1.283 ± 0.306	1.966 ± 0.017	156.856 ± 3.276	0.091 ± 0.066

findings provide a strong foundation for future rational design efforts aiming to optimize peptide–enzyme interactions through targeted residue-level modifications.

5. Conclusions

Whole genome sequencing and bioinformatic analyses of *S. cyaneofuscatus* strain CTM50504 revealed the presence of numerous hydrolase-encoding genes. Overall, this is the first report on the cloning of the *sckp* gene from CTM50504, which encodes a serine alkaline protease (rSCKP) possessing several properties highly valued in biotechnology. The docking simulation data provides insight into the involvement of key amino acids in substrate binding and helps explain the substrate selectivity of rSCKP. The integration of molecular dynamics and MM/PBSA analysis confirmed the structural and energetic consistency of the modeled rSCKP–peptide complexes. Accordingly, further studies, some currently underway in our laboratories, are needed to synthesize or clone the genes encoding other interesting industrially relevant enzymes. Site-directed mutagenesis (SDM) (e.g., N174A, K274A, and A275G) is scheduled for the near future to further validate the predicted functional residues and confirm the substrate binding mechanism at the atomic level.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2025.146696.

GenBank/ENA/DDBJ

PP346394 and PRJNA1065629.

CRediT authorship contribution statement

Sondes Mechri: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Séverine Croze:** Methodology, Investigation, Conceptualization. **Imen Rekik:** Writing – original draft, Software, Formal analysis, Conceptualization. **Fawzi Allala:** Writing – original draft, Software, Methodology, Conceptualization. **Fakher Frikha:** Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Alexandre Noiriel:** Writing – review & editing, Formal analysis, Conceptualization. **Marilize Le Roes-Hill:** Writing – review & editing, Formal analysis, Conceptualization. **Abdelkarim Abousalham:** Writing – review & editing, Formal analysis, Conceptualization. **Slim Tounsi:** Writing – review & editing, Supervision, Conceptualization. **Joel Lachuer:** Writing – original draft, Supervision, Funding acquisition, Conceptualization. **Bassem Jaouadi:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Funding

This work was conducted within the European Union's Horizon Europe Research and Innovation Program, under the *HORIZON-Coordination and Support Action (CSA)*, *HORIZON-WIDERA-2021-ACCESS-03-01*, as part of the Twinning Project NGS-4-ECOPROD (Grant Agreement ID: 101079425) 2022–2026, entitled: "FROM NEXT GENERATION SEQUENCING MICROORGANISMS TOWARDS ECOFRIENDLY BIOTECH BASED PRODUCTS". The project is coordinated by Prof. Slim Tounsi (Centre of Biotechnology of Sfax, CBS), in collaboration with Université Claude Bernard Lyon 1 (UCBL) (Prof. Joel Lachuer) and Georg-August-Universität Göttingen Stiftung Öffentlichen Rechts (UGOE) (Dr. Heiko Liesegang). Additionally, we acknowledge Dr. Anca-Adriana Cucu, Project Advisor for HE-Widening Participation, Research Executive Agency (Brussels Area, Belgium), for her valuable contributions. Furthermore, part of this work was supported by the French Government Scholarship Programme « *France Excellence* » under the short-term Scholarship 2023 « *Séjours Scientifiques de Haut Niveau (SSHN)* ».

Declaration of competing interest

The authors have declared no competing interests.

Acknowledgments

The authors extend their sincere gratitude to Prof. Catherine Legras-Lachuer from the Viroscan 3D VIROSCAN3D, 11 Allée des Acacias 01600, Trévoux, France, for her invaluable support with the 5200 Fragment Analyzer System analyses. Additionally, we wish to express our appreciation to Mr. Taha Bilel Chalbi and Miss Marwa Lazreg (LMEBB-CBS) for their kind technical assistance. We are also deeply grateful to Dr. Stefani Maria Díaz Valerio and Dr. Heiko Liesegang (UGOE, Institut für Mikrobiologie und Genetik, Laboratorium für Genomanalyse, Georg-August Universität, Grisebachstr. 8, Göttingen, D-37077, Germany), for their invaluable contributions and insightful discussions, which greatly enriched the preparation of this manuscript. Special thanks to Dr. Hela Gargouri, Project Manager of NGS-4-ECOPROD, and Ms. Sawssen Kchaou, from the Financial Department at the CBS, for their valuable support and assistance. The authors warmly express their sincere gratitude to Prof. Stefan Grünert, PhD in Biochemistry from University of Cambridge, United Kingdom (co-founder of biolution – Project Development, Training, & Coaching – Medical University of Vienna, Austria), Mrs. Maggy Pézeril (Chargée de mission, Pôle Universitaire Européen de Montpellier et du Languedoc-Roussillon, Université de Montpellier), and the dedicated members of UPGO-HE, MHESR, Tunisia — namely Prof. Helmi Mardassi (General Director of UPGO-HE) and Mrs. Amani Mahjoubi Charrad (COST National Correspondent, MHESR – DG R&I EU Programmes) — for their unwavering encouragement, inspiring training, and highly valuable coaching sessions that greatly contributed to the success of the NGS-4-ECOPROD twinning project (https://cordis.europa.eu/project/id/101079425) and, ultimately, to the accomplishment of this work.

Data availability

Data will be made available on request.

References

[1] P.S. Murthy, M. Vedashree, H. Sneha, I. Prakash, Extremophiles as a source of biotechnological products, in: I.R. Management Association (Ed.), Research

Anthology on Bioinformatics, Genomics, and Computational Biology, IGI Global, Hershey, PA, USA, 2024, pp. 606–632.
[2] M.A. Dahmani, N. Boulenouar, O. Alami, B. Saoudi, K. Bouacem, S. Mechri, F. Allala, O. Messaoudi, M. Yousfi, M. Le Roes-Hill, B. Jaouadi, Purification and biochemical characterization of a new thermostable and detergent-stable serine protease from a novel thermo-halotolerant bacterial strain, *Laceyella sacchari* strain BK-TM, Biotechnol. Appl. Biochem. (2025), https://doi.org/10.1002/bab.70015. Epub ahead of print. PMID: 40551601.
[3] A. Zakrzewska, G. Van Eikenhorst, J.E. Burggraaff, D.J. Vis, H. Hoefsloot, D. Delneri, S.G. Oliver, S. Brul, G.J. Smits, Genome-wide analysis of yeast stress survival and tolerance acquisition to analyze the central trade-off between growth rate and cellular robustness, Mol. Biol. Cell 22 (22) (2011) 4435–4446.
[4] R.R. Salwan, V. Sharma, Genomics of prokaryotic extremophiles to unfold the mystery of survival in extreme environments, Microbiol. Res. 264 (2022) 127156.
[5] A.D. Alma'abadi, T. Gojobori, K. Mineta, Marine metagenome as a resource for novel enzymes, Genom. Proteom. Bioinform. 13 (5) (2015) 290–295.
[6] I. Kohli, N.C. Joshi, S. Mohapatra, A. Varma, Extremophile–an adaptive strategy for extreme conditions and applications, Curr. Genomics 21 (2020) 96–110.
[7] A. Vaishnav, J. Sahu, H.B. Singh, Genomics of extremophiles for sustainable agriculture and biotechnological applications (part I), Curr. Genomics 21 (2) (2020) 78.
[8] A. Purohit, S.K. Yadav, Genome sequencing of a novel *Microbacterium camelliasinensis* CIAB417 identified potential mannan hydrolysing enzymes, Int. J. Biol. Macromol. 208 (2022) 219–229.
[9] J.A. John, E. Selvarajan, Genomic analysis of lignocellulolytic enzyme producing novel *Streptomyces* sp.MS2A for the bioethanol applications, Int. J. Biol. Macromol. 250 (2023) 126138.
[10] K. Alam, A. Mazumder, S. Sikdar, Y.M. Zhao, J. Hao, C. Song, Y. Wang, R. Sarkar, S. Islam, Y. Zhang, A. Li, *Streptomyces*: the biofactory of secondary metabolites, Front. Microbiol. 13 (2022) 968053.
[11] P. Khushboo, K.K. Kumar, Z. Dubey, M. Usmani, V.K. Sharma, Gupta, bioproducts, biorefining, biotechnological and industrial applications of Streptomyces metabolites, Biofuels Bioprod. Biorefin. 16 (1) (2022) 244–264.
[12] K. Alam, A. Mazumder, S. Sikdar, Y.-M. Zhao, J. Hao, C. Song, Y. Zhang, R. Wang, S. Sarkar, Y. Islam, A. Li, Streptomyces: The biofactory of secondary metabolites, Front. Microbiol. 13 (2022) 968053.
[13] C. Chen, Y. Ye, R. Wang, Y. Zhang, C. Wu, S.C. Debnath, Z. Ma, J. Wang, M. Wu, *Streptomyces nigra* sp. nov. is a novel actinobacterium isolated from mangrove soil and exerts a potent antitumor activity in vitro, Front. Microbiol. 9 (2018) 1587.
[14] Y. Ohnishi, J. Ishikawa, H. Hara, H. Suzuki, M. Ikenoya, H. Ikeda, A. Yamashita, M. Hattori, S. Horinouchi, Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350, J. Bacteriol. 190 (2008) 4050–4060.
[15] S.D. Bentley, K.F. Chater, A.M. Cerdeño-Tárraga, G.L. Challis, N. Thomson, K. D. James, D.E. Harris, M.A. Quail, H. Kieser, D. Harper, Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3, Nature 417 (2002) 141–147.
[16] H. Ikeda, J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori, S. Ōmura, Complete geQnome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*, Nat. Biotechnol. 21 (2003) 526–531.
[17] Y.H. Chung, H. Kim, C.H. Ji, H.W. Je, D. Lee, S.H. Shim, H.S. Joo, H.S. Kang, Comparative genomics reveals a remarkable biosynthetic potential of the *Streptomyces* phylogenetic lineage associated with rugose-ornamented spores, mSystems 6 (2021) e0048921.
[18] V. Napolioni, L. Cimarelli, A. Miano, A. La Teana, R. Çapuni, A.M. Giuliodori, A. Fabbretti, R. Spurio, F.J. Stewart, Draft genome sequence of *Streptomyces* sp. strain AM-2504, identified by 16S rRNA comparative analysis as a *Streptomyces kasugaensis* strain, Microbiol. Resour. Announc. 8 (2019) e00966–19.
[19] N. Lee, S. Hwang, J. Kim, S. Cho, B. Palsson, B.-K. Cho, Mini review: genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in *Streptomyces*, Comput. Struct. Biotechnol. J. 18 (2020) 1548–1556.
[20] E.B. Shirling, D. Gottlieb, Methods for characterization of *Streptomyces* species, Int. J. Syst. Evol. Microbiol. 16 (1966) 313–340.
[21] A. Wilmotte, Molecular Evolution and Taxonomy of the cyanobacteria, The molecular biology of cyanobacteria, Springer, 1994, p. 125.
[22] S. Mechri, M. Ben Elhoul Berrouina, M. Omrane Benmrad, N. Zaraî Jaouadi, H. Rekik, E. Moujehed, A. Chebbi, S. Sayadi, M. Chamkha, S. Bejar, B. Jaouadi, Characterization of a novel protease from *Aeribacillus pallidus* strain VP3 with potential biotechnological interest, Int. J. Biol. Macromol. 94 (2017) 221–232.
[23] G. Kouker, K.E. Jaeger, Specific and sensitive plate assay for bacterial lipases, Appl. Environ. Microbiol. 53 (1987) 211–213.
[24] A. Dab, I. Hasnaoui, S. Mechri, F. Allala, K. Bouacem, A. Noiriel, A. Bouanane-Darenfed, E. Saalaoui, A. Asehraou, F. Wang, A. Abousalham, B. Jaouadi, Biochemical characterization of an alkaline and detergent-stable lipase from *fusarium annulatum* Bugnicourt strain CBS associated with olive tree dieback, PLoS One 18 (2023) e0286091.
[25] I. Hasnaoui, A. Dab, S. Mechri, H. Abouloifa, E. Saalaoui, B. Jaouadi, A. Noiriel, A. Asehraou, A. Abousalham, Purification, biochemical and kinetic characterization of a novel alkaline sn-1,3-Regioselective triacylglycerol lipase from *Penicillium crustosum* Thom strain P22 isolated from moroccan olive mill wastewater, Int. J. Mol. Sci. 23 (2022) 11920.
[26] R. Rahier, H. Abla, Y. Arhab, A. Noiriel, A. Abousalham, Direct and continuous measurement of phospholipase d activities using the chelation-enhanced fluorescence property of 8-Hydroxyquinoline, in: G. Sandoval (Ed.), Lipases and Phospholipases: Methods and Protocols, Springer, New York, 2018, pp. 129–138.
[27] F. Allala, K. Bouacem, N. Boucherba, Z. Azzouz, S. Mechri, M. Sahnoun, S. Benallaoua, H. Hacene, B. Jaouadi, A. Bouanane-Darenfed, Purification, biochemical, and molecular characterization of a novel extracellular thermostable and alkaline α-amylase from *Tepidimonas fonticaldi* strain HB23, Int. J. Biol. Macromol. 132 (2019) 558–574.
[28] K. Mohamed, S. Bouacem, N.A. Mechri, H. Addou, M.L. Laribi-Habchi, B. Fardeau, A. Jaouadi, H. Bouanane-Darenfed, Hacène, Purification and biochemical characterization of a novel acido-halotolerant and thermostable endochitinase from *Melghiribacillus thermohalophilus* strain Nari2A^{T}, Carbohydr. Res. 473 (2019) 46–56.
[29] J. Cao, L. Zheng, S. Chen, Screening of pectinase producer from alkalophilic bacteria and study on its potential application in degumming of ramie, Enzym. Microb. Technol. 14 (1992) 1013–1016.
[30] Z.E. Elsababty, S.H. Abdel-Aziz, A.M. Ibrahim, A.A. Guirgis, G.E. Dawwam, Purification, biochemical characterization, and molecular cloning of cellulase from *Bacillus licheniformis* strain Z9 isolated from soil, J. Genet. Eng. Biotechnol. 20 (2022) 34.
[31] J.E. Stach, L.A. Maldonado, A.C. Ward, M. Goodfellow, A.T. Bull, New primers for the class Actinobacteria: application to marine and terrestrial environments, Environ. Microbiol. 5 (2003) 828–841.
[32] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics (Oxford, England) 30 (15) (2014) 2114–2120.
[33] J.R. Grant, E. Enns, E. Marinier, A. Mandal, E.K. Herman, C.Y. Chen, M. Graham, G. Van Domselaar, P. Stothard, Proksee: in-depth characterization and visualization of bacterial genomes, Nucleic Acids Res. 51 (W1) (2023) W484–W492.
[34] J.P. Meier-Kolthoff, M. Göker, TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy, Nat. Commun. 10 (2019) 2182.
[35] M. Richter, R. Rosselló-Móra, F. Oliver Glöckner, J. Peplies, JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison, Bioinformatics 32 (6) (2016) 929–931.
[36] A.L. Delcher, D. Harmon, S. Kasif, O. White, S.L. Salzberg, Improved microbial gene identification with GLIMMER, Nucleic Acids Res. 27 (23) (1999) 4636–4641.
[37] K. Blin, S. Shaw, H.E. Augustijn, Z.L. Reitz, F. Biermann, M. Alanjary, A. Fetter, B. R. Terlouw, W.W. Metcalf, E.J.N. Helfrich, G.P. van Wezel, M.H. Medema, T. Weber, antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation, Nucleic Acids Res. 51 (W1) (2023) W46–W50.
[38] M. Madhaiyan, V. Saravanan, W.S. See-Too, Genome-based analyses reveal the presence of 12 heterotypic synonyms in the genus *Streptomyces* and emended descriptions of *Streptomyces bottropensis*, *Streptomyces celluloflavus*, *Streptomyces fulvissimus*, *Streptomyces glaucescens*, *Streptomyces murinus*, and *Streptomyces variegatus*, Int. J. Syst. Evol. Microbiol. 70 (2020) 3924–3929.
[39] J. Sun, F. Lu, Y. Luo, L. Bie, L. Xu, Y. Wang, OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes, Nucleic Acids Res. 51 (W1) (2023) W397–W403.
[40] S. Mechri, N. Zaraî Jaouadi, K. Bouacem, F. Allala, A. Bouraoui, C. Ferard, H. Rekik, A. Noiriel, A. Abousalham, A. Bouanane-Darenfed, H. Hacène, F. Lederer, L. Baciou, B. Jaouadi, Cloning and heterologous expression of subtilisin SAPN, a serine alkaline protease from *Melghiribacillus thermohalophilus* Nari2A^{T} in *Escherichia coli* and *Pichia pastoris*, Process Biochem. 105 (2021) 27–41.
[41] Z. Wu, S. Violot, A. Abousalham, A. Noiriel, A new bacterial phospholipase D with specificity for phosphatidylethanolamine over phosphatidylcholine, Int. J. Biol. Macromol. 304 (2025) 140578.
[42] S. Mechri, K. Bouacem, T.-B. Chalbi, M. Khaled, F. Allala, A. Bouanane-Darenfed, H. Hacene, B. Jaouadi, A Taguchi design approach for the enhancement of a detergent-biocompatible alkaline thermostable protease production by *Streptomyces mutabilis* strain TN-X30, J. Surfactant Deterg. 25 (4) (2022) 487–504.
[43] M.S. Valdés-Tresanco, M.E. Valdés-Tresanco, P.A. Valiente, E. Moreno, gmx_MMPBSA: A new tool to perform end-state free energy calculations with GROMACS, J. Chem. Theory Comput. 17 (10) (2021) 6281–6291.
[44] S.C. Heinsch, S.Y. Hsu, L. Otto-Hanson, L. Kinkel, M.J. Smanski, Complete genome sequences of *Streptomyces* spp. isolated from disease-suppressive soils, BMC Genomics 20 (2019) 994.
[45] H. Otani, D.W. Udwary, N.J. Mouncey, Comparative and pangenomic analysis of the genus *Streptomyces*, Sci. Rep. 12 (1) (2022) 18909.
[46] J.R. Grant, P. Stothard, The CGView server: a comparative genomics tool for circular genomes, Nucleic Acids Res. 36 (2008) W181–W184 (Web Server issue).
[47] K. Blin, S. Shaw, A.M. Kloosterman, Z. Charlop-Powers, G.P. Van Wezel, M. H. Medema, T. Weber, antiSMASH 6.0: improving cluster detection and comparison capabilities, Nucleic Acids Res. 49 (W1) (2021) W29–W35.
[48] S. Mechri, K. Bouacem, N. Jaouadi, H. Rekik, M. Elhoul, M. Benmrad, H. Hacene, S. Bejar, A. Bouanane-Darenfed, B. Jaouadi, Identification of a novel protease from the thermophilic *Anoxybacillus kamchatkensis* M1V and its application as laundry detergent additive, Extremophiles 23 (2019) 687–706.
[49] K. Knapik, M. Becerra, M.I. González-Siso, Microbial diversity analysis and screening for novel xylanase enzymes from the sediment of the Lobios hot spring in Spain, Sci. Rep. 9 (2019) 11195.
[50] H. Sahay, A.N. Yadav, A.K. Singh, S. Singh, R. Kaushik, A.K. Saxena, Hot springs of Indian Himalayas: potential sources of microbial diversity and thermostable hydrolytic enzymes, 3 Biotech. 7 (2017) 118.
[51] S. Mechri, F. Allala, K. Bouacem, I. Hasnaoui, H. Gwaithan, T.B. Chalbi, E. Saalaoui, A. Asehraou, A. Noiriel, A. Abousalham, H. Hacene, A. Bouanane-Darenfed, M. Le Roes-Hill, B. Jaouadi, Preparation, characterization, immobilization, and molecular docking analysis of a novel detergent-stable

subtilisin-like serine protease from *Streptomyces mutabilis* strain TN-X30, Int. J. Biol. Macromol. 222 (2022) 1326–1342.
[52] S.H. Yoon, S.M. Ha, J. Lim, S. Kwon, J. Chun, A large-scale evaluation of algorithms to calculate average nucleotide identity, Antonie Van Leeuwenhoek 110 (2017) 1281–1286.
[53] H. Liu, S. Yin, L. An, G. Zhang, H. Cheng, Y. Xi, G. Cui, F. Zhang, L. Zhang, Complete genome sequence of *Bacillus subtilis* BSD-2, a microbial germicide isolated from cultivated cotton, J. Biotechnol. 230 (2016) 26–27.
[54] Y. Zhou, Q. Li, Z. Peng, J. Zhang, J.J.J.o.A. Li, F., Chemistry, biocontrol effect of *Bacillus subtilis* YPS-32 on potato common scab and its complete genome sequence analysis, J. Agric. Food Chem. 70 (2022) 5339–5348.
[55] H. Wu, W. Liu, L. Shi, K. Si, T. Liu, D. Dong, T. Zhang, J. Zhao, D. Liu, Z.J.S.R. Tian, Comparative genomic and regulatory analyses of natamycin production of *Streptomyces lydicus* A02, Sci. Rep. 7 (2017) 9114.
[56] T.G. Pridham, C.W. Hesseltine, R.G. Benedict, A guide for the classification of *Streptomycetes* according to selected groups; placement of strains in morphological sections, Appl. Microbiol. 6 (1958) 52–79.
[57] V. Sharma, A. Sharma, A.B. Malannavar, R. Salwan, Molecular aspects of biocontrol species of *Streptomyces* in agricultural crops, Molecular aspects of plant beneficial microbes in agriculture, Elsevier, 2020, pp. 89–109.
[58] J. Jiang, X. He, D.E. Cane, Biosynthesis of the earthy odorant geosmin by a bifunctional *Streptomyces coelicolor* enzyme, Nat. Chem. Biol. 3 (2007) 711–715.
[59] K. Tahlan, H.U. Park, S.E. Jensen, Three unlinked gene clusters are involved in clavam metabolite biosynthesis in *Streptomyces clavuligerus*, Can. J. Microbiol. 50 (2004) 803–810.
[60] H. Aldemir, S.V. Kohlhepp, T. Gulder, T.A. Gulder, Structure of a putative fluorinated natural product from *Streptomyces* sp. TC1, J. Nat. Prod. 77 (2014) 2331–2334.
[61] L. Tang, Y.X. Zhang, C.R. Hutchinson, Amino acid catabolism and antibiotic synthesis: valine is a source of precursors for macrolide biosynthesis in *Streptomyces ambofaciens* and *Streptomyces fradiae*, J. Bacteriol. 176 (1994) 6107–6119.
[62] L.T.-H. Tan, K.-G. Chan, L.-H. Lee, B.-H. Goh, *Streptomyces* bacteria as potential probiotics in aquaculture, Front. Microbiol. 7 (2016) 79.
[63] S.T. Williams, M. Goodfellow, G. Alderson, E.M. Wellington, P.H. Sneath, M. J. Sackin, Numerical classification of *Streptomyces* and related genera, J. Gen. Microbiol. 129 (6) (1983) 1743–1813.
[64] T. Narancic, R. Davis, J. Nikodinovic-Runic, K.E. O'connor, Recent developments in biocatalysis beyond the laboratory, Biotechnol. Lett. 37 (5) (2015) 943–954.
[65] N.J. Turner, M.D. Truppo, Biocatalysis enters a new era, Curr. Opin. Chem. Biol. 17 (2) (2013) 212–214.
[66] A. Bhalla, N. Bansal, S. Kumar, K.M. Bischoff, R.K. Sani, Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes, Bioresour. Technol. 128 (2013) 751–759.
[67] L.C. Smith, F. Faustinella, L. Chan, Lipases: three-dimensional structure and mechanism of action, Curr. Opin. Struct. Biol. 2 (1992) 490–496.
[68] B. Jaouadi, B. Abdelmalek, D. Fodil, F.Z. Ferradji, H. Rekik, N. Zaraî, S. Bejar, Purification and characterization of a thermostable keratinolytic serine alkaline proteinase from *Streptomyces* sp. strain AB1 with high stability in organic solvents, Bioresour. Technol. 101 (2010) 8361–8369.
[69] A. Mohy Eldin, N. Hossam, Microbial surfactants: characteristics, production and broader application prospects in environment and industry, Prep. Biochem. Biotechnol. 53 (2023) 1013–1042.
[70] F. Hasan, A.A. Shah, A. Hameed, Industrial applications of microbial lipases, Enzym. Microb. Technol. 39 (2006) 235–251.
[71] R. Gupta, Q.K. Beg, P. Lorenz, Bacterial alkaline proteases: molecular approaches and industrial applications, Appl. Microbiol. Biotechnol. 59 (1) (2002) 15–32.
[72] M. Kapoor, D. Panwar, G. Kaira, Bioprocesses for enzyme production using agro-industrial wastes: technical challenges and commercialization potential, agro-industrial wastes as feedstock for enzyme production, Elsevier, 2016, pp. 61–93.
[73] T. Kantyka, N.D. Rawlings, J. Potempa, Prokaryote-derived protein inhibitors of peptidases: a sketchy occurrence and mostly unknown function, Biochimie 92 (11) (2010) 1644–1656.
[74] I. Lee, C.K. Suzuki, Functional mechanics of the ATP-dependent Lon protease-lessons from endogenous protein and synthetic peptide substrates, Biochim. Biophys. Acta, Proteins Proteomics 1784 (5) (2008) 727–735.
[75] L.L. Lairson, B. Henrissat, G.J. Davies, S.G. Withers, Glycosyltransferases: structures, functions, and mechanisms, Annu. Rev. Biochem. 77 (2008) 521–555.
[76] B.B. He, X. Bai, Y. Tan, W. Xie, Y. Feng, G.Y. Yang, Glycosyltransferases: mining, engineering and applications in biosynthesis of glycosylated plant natural products, Synth. Syst. Biotechnol. 7 (2022) 602–620.
[77] K. Choure, S. Parsai, R. Kotoky, A. Srivastava, A. Tilwari, P.K. Rai, A. Sharma, P. Pandey, Comparative metagenomic analysis of two alkaline hot springs of Madhya Pradesh, India and deciphering the extremophiles for industrial enzymes, Front. Genet. 12 (2021) 643423.
[78] E.I. Petersen, G. Valinger, B. Sölkner, G. Stubenrauch, H. Schwab, A novel esterase from *Burkholderia gladioli* which shows high deacetylation activity on cephalosporins is related to β-lactamases and DD-peptidases, J. Biotechnol. 89 (1) (2001) 11–25.
[79] U.G. Wagner, E.I. Petersen, H. Schwab, C. Kratky, EstB from *Burkholderia gladioli*: a novel esterase with a β-lactamase fold reveals steric factors to discriminate between esterolytic and β-lactam cleaving activity, Protein Sci. 11 (3) (2002) 467–478.
[80] L.M. Ouyang, J.Y. Liu, M. Qiao, J.H. Xu, Isolation and biochemical characterization of two novel metagenome-derived esterases, Appl. Biochem. Biotechnol. 169 (2013) 15–28.
[81] M. Takehara, K. Kinoshita, M. Miyamoto, H. Hirohara, A novel alkaline esterase from *Sporosarcina* sp. nov. strain eSP04 catalyzing the hydrolysis of a wide variety of aryl-carboxylic acid esters, Biosci. Biotechnol. Biochem. 76 (2012) 1721–1727.
[82] Y.O. Kim, I.S. Park, B.H. Nam, D.G. Kim, Y.J. Jee, S.J. Lee, C.M. An, A novel esterase from *Paenibacillus* sp. PBS-2 is a new member of the β-lactamase belonging to the family VIII lipases/esterases, J. Microbiol. Biotechnol. 24 (2014) 1260–1268.
[83] G. Rodríguez-Alonso, J. Toledo-Marcos, L. Serrano-Aguirre, C. Rumayor, B. Pasero, A. Flores, A. Saborido, P. Hoyos, M.J. Hernáiz, I. de la Mata, M. Arroyo, A novel lipase from *Streptomyces exfoliatus* DSMZ 41693 for biotechnological applications, Int. J. Mol. Sci. 24 (23) (2023) 17071.
[84] P. Mander, S.S. Cho, J.R. Simkhada, Y.H. Choi, D.J. Park, J.W. Ha, J.C. Yoo, An organic solvent-tolerant alkaline lipase from *Streptomyces* sp. CS268 and its application in biodiesel production, Biotechnol. Bioprocess Eng. 17 (1) (2012) 67–75.
[85] G.M. Borrelli, D. Trono, Recombinant lipases and phospholipases and their use as biocatalysts for industrial applications, Int. J. Mol. Sci. 16 (9) (2015) 20774–20840.
[86] H. Tsujibo, N. Hatano, H. Endo, K. Miyamoto, Y. Inamori, Biotechnology, biochemistry, Purification and characterization of a thermostable chitinase from *Streptomyces thermoviolaceus* OPC-520 and cloning of the encoding gene, Biosci. Biotechnol. Biochem. 64 (1) (2000) 96–102.
[87] É.R.d. Santos, Z.N.S. Teles, N.M. Campos, D.A.J.d. Souza, A.S.d.R. Bispo, R.P. d. Nascimento, Production of α-amylase from *Streptomyces* sp. SLBA-08 strain using agro-industrial by-products, Braz. Arch. Biol. Technol. 55 (5) (2012) 793–800.
[88] S. Kar, R.C. Ray, U.B. Mohapatra, Purification, characterization and application of thermostable amylopullulanase from *Streptomyces erumpens* MTCC 7317 under submerged fermentation, Ann. Microbiol. 62 (3) (2012) 931–937.
[89] M.M. Bradford, A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254.
[90] N. Zaraî Jaouadi, H. Rekik, A. Badis, S. Trabelsi, M. Belhoul, A.B. Yahiaoui, H. B. Aicha, A. Toumi, S. Bejar, B. Jaouadi, Biochemical and molecular characterization of a serine keratinase from *Brevibacillus brevis* US575 with promising keratin-biodegradation and hide-dehairing activities, PLoS One 8 (10) (2013) e76722.
[91] V. Sangal, M. Goodfellow, A.L. Jones, E.C. Schwalbe, J. Blom, P.A. Hoskisson, I. C. Sutcliffe, Next-generation systematics: an innovative approach to resolve the structure of complex prokaryotic taxa, Sci. Rep. 6 (2016) 38392.