



Cloning, expression, purification and biophysical analysis of two putative halogenases from the glycopeptide A47,934 gene cluster of *Streptomyces toyocaensis*

Tabata P. Cardoso^{a, b, 1}, Larissa A. de Sá^{c, 1}, Priscila dos S. Bury^c, Sair M. Chavez-Pacheco^c, Marcio V.B. Dias^{a, b, c, *}

^a Laboratório Nacional de Biotecnologia, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, Brazil

^b Programa de Pós-Graduação em Microbiologia, Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, SP, Brazil

^c Instituto de Ciências Biomédicas, Universidade de São Paulo-SP, São Paulo, Brazil

ARTICLE INFO

Article history:

Received 6 September 2016

Received in revised form

22 December 2016

Accepted 4 January 2017

Available online 5 January 2017

Keywords:

Halogenase

Glycopeptide biosynthesis

Streptomyces toyocaensis

ABSTRACT

Glycopeptides are an important class of antibiotics used in the treatment of several infections, including those caused by methicillin resistant *Staphylococcus aureus*. Glycopeptides are biosynthesized by a Non Ribosomal Peptide Synthase (NRPS) and the resulting peptide precursors are decorated by several tailoring enzymes, such as halogenases and glycosyltransferases. These enzymes are important targets of protein engineering to produce new derivatives of known antibiotics. Herein we show the production of two putative halogenases, denominated Stal and StaK, involved in the biosynthesis of the glycopeptide A47,934 in *Streptomyces toyocaensis* NRRL 15,009. This antibiotic together with the compound UK-68,597 are the unique glycopeptides which have two putative halogenases identified in their gene clusters and three chloride substituent atoms attached to their aglycones. Stal and StaK were successfully produced in *E. coli* in the soluble fraction with high purity using the wild type gene for Stal and a synthetic codon optimized gene for StaK. We have purified both enzymes by two chromatographic steps and a good yield was obtained. These putative halogenases were co-purified with the co-factor FAD, which are differently reduced by the enzyme SsuE *in vitro*. We have further confirmed that these putative halogenases are monomeric using a calibrated gel filtration column and through circular dichroism, we confirmed that both enzymes are folded with a predominance of α -helices. Molecular models for Stal and StaK were generated and together with sequence and phylogenetic analysis, we could infer some structural insights of Stal and StaK from the biosynthesis of compound A47,934.

© 2017 Elsevier Inc. All rights reserved.

Glycopeptide antibiotics, such as vancomycin, are front line antibiotics for the treatment of severe infections caused by gram positive bacteria, including the troublesome methicillin resistant *Staphylococcus aureus* (MRSA) [1]. However resistance to vancomycin has also been reported, including strains which exhibit high-level of resistance (VRSA) [2]. Consequently the discovery of new strategies to treat MRSA and VRSA strains are needed, including those approaches to obtain new glycopeptide derivatives such as the semi-synthetic antibiotic telavacin and dalbavancin, which

demonstrated pharmacological improvements despite their low effect against VRSA [3]. Most glycopeptides are large and rigid molecules which have a central heptapeptide core with at least two aromatic non-proteinogenic amino acids decorated by glycosylations, halogenations, methylations and cross-links of their side chains [4]. The mechanism of action of glycopeptides generally involves the formation of a non-covalent complex with the acyl-D-alanyl-D-alanine portion of the pentapeptide of the peptidoglycan and consequently they are inhibitors of cell wall biosynthesis [4]. The activity of vancomycin and teicoplanin (another clinically important glycopeptide) is strongly dependent on the sugar and halogen substituents attached to its central core [5]. The chlorine substituents at the ring of the amino acids 2 and 6 of the heptapeptide core are found in most of the glycopeptides [4]. These halogens

* Corresponding author. Av Lineu Prestes, 1374, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo-SP, CEP 05508-000, Brazil.

E-mail address: mvbdi@usp.br (M.V.B. Dias).

¹ These authors contributed equally for this work.

attached to the glycopeptide have been proved to increase the stability of dimer formation and consequently contribute for the effect and specificity of these antibiotics [5,6]. The enzymes responsible for the halogenation of the heptapeptide core of glycopeptides have been identified to be flavin dependent halogenases, although the exact timing of this reaction during the biosynthesis of these antibiotics has not yet been determined [7]. However there are several evidences that the chlorination step occurs when the growing peptide chain substrate is bound to NRPS (non ribosomal peptide synthase) through a linkage with the PCP (peptidyl carrier protein) domain [8]. *In vitro* activity of a halogenase from glycopeptide antibiotic has only recently been proved to use a hexapeptidyl-PCP conjugate and VhaA halogenase from the vancomycin gene cluster [7]. Most of the glycopeptide biosynthetic gene clusters have a single halogenase, which catalyzes the introduction of both chlorine substituents to the heptapeptide aglicone [9–12]. Some glycopeptide gene clusters, such as ristocetin, do not have any putative halogenase [13] and consequently these glycopeptides are not halogenated. On the other hand, in the biosynthetic gene clusters of the glycopeptide A47,934 and UK-68,597, two putative halogenases were identified (Stal and StaK; and Auk21 and Auk23, respectively) and these compounds have three chlorinated aromatic amino acids at position 2, 5 and 6 [14,15]. Stal has higher similarity with other halogenases from glycopeptide gene clusters and should catalyze the chlorination at position 2 and 6 of heptapeptide. However, there are neither studies showing the requirement of StaK to perform the chlorination at position 5 of A47,934 nor evidence whether both Stal and StaK are essential to produce the glycopeptide A47,934.

Halogenases have recently become a target of several programs of biotechnological engineering and combinatorial biosynthesis due to the biological effect and importance of the incorporation of halogens into natural products [16,17]. Various research groups have had success with the incorporation of halogens through the manipulation of natural products biosynthetic pathways [18–22]. Halogenases are grouped in several classes based on its mechanism of catalysis and structures [23]. In this context, flavin dependent halogenases are one of the best characterized class and can be further grouped into two different subclasses, one that acts on free substrates and another that catalyzes the halogenation of substrates linked to the NRPS, such as halogenases or the putative halogenases found in the gene clusters of glycopeptides. However, although these halogenases may be interesting tools for the application of synthetic biology, there is a lack of studies about them, including structural and biophysical analysis. In this work, we show the cloning, expression, purification, preliminary biophysical analysis and molecular modelling of the two putative halogenases identified in the biosynthetic gene cluster of the glycopeptide A47,934 from *Streptomyces toyocaensis* NRRL 15,009. We have performed the expression of these enzymes in *E. coli* and analyzed the integrity of their structures by circular dichroism. To provide insights into the different functions of these two putative halogenases, we have made a sequence and phylogenetic analysis as well as molecular modelling to predict their tridimensional structure. The results indicate interesting differences that might be attributed to the divergent functions of Stal and StaK in the biosynthesis of compound A47,934.

1. Materials and methods

1.1. Cloning, expression and purification of Stal and StaK

Streptomyces toyocaensis NRRL 15,009 was obtained from Agricultural Research Service (ARS) culture collection from USA and the freeze-drying cells were inoculated in 50 mL of TSB medium (BD™)

(17 g tryptone, 3 g soytone, 2.5 g glucose, 5 g NaCl, and 2.5 HK₂PO₄) and grown for 4 days at a temperature of 30 °C with a rotation of 200 rpm. The cells were harvested by centrifugation at 4000 rpm (1860 × g) and the genomic DNA was extracted using the Promega extraction kit following the manufacturer instructions.

To amplify *stal* and *staK* we used the *Streptomyces toyocaensis* NRRL 15,009 genomic DNA and two set of primers for each gene: *stal_Foward* (ATCGATCGCA-TATGGACGGCTTGAGCGGCGATCCGTTCG) and *stal_Reverse* (ATCGCTCGAGTCATCCACGGTAGGGAAGCCACTTCATCCC) and *staK_Foward* (ATCGATCGCATATGTCCGCAGAGACGTTCCGACGTGGTTCG) and *staK_Reverse* (TCGCTCGAGTCATCGGGCGTCCTCTCGTTCCTTCGACGC), respectively for *stal* and *staK* (the endonuclease restriction sites for *Nde*I and *Xho*I are shown in bold for the forward and reverse primers, respectively). The corresponding amplified DNA fragments with the size of *stal* (1536 bp) and *staK* (1305 bp) were cloned into the cloning vector pGEM-T (Promega) and sub-cloned into the expression vector pET-28a(+) (Novagen). In addition, we also purchased a synthetic gene (GenScript Company) which produces StaK and had a codon optimization for expression in *E. coli*. This gene was cloned in pUC57 vector with the endonuclease restriction sites for *Nde*I and *Xho*I for 5' and 3' positions respectively. After the amplification of this plasmid using DH5 α competent cells, it was digested using the set of restriction endonucleases mentioned above and the corresponding fragment with the size of *staK* was subcloned into pET-28a(+) to create a fusion protein with a 6-His-tag at the N-terminal followed by a thrombin cleavage site.

To confirm the correct cloning site and the integrity of *stal* and *staK*, we have performed a sequencing of the plasmids containing the genes of interest using the Sanger method in the facility of Laboratório Nacional de Biociências – CNPEM, Campinas, Brazil.

To analyse the expression of both Stal and StaK, expression vectors containing *stal* and *staK* were introduced into BL21(DE3) competent cells by heat shock transformation. A single colony of BL21(DE3) containing the plasmid of interest was randomly chosen and inoculated into 10 mL of Luria-Bertani (LB) medium with further addition of 50 μ g/mL kanamycin. The culture was initially incubated at 37 °C overnight under a shaking of 200 rpm. For protein expression, 1 mL of this culture was inoculated into 250 mL of LB medium at the same conditions of the pre-inoculum. When the optical density at 600 nm (OD₆₀₀) of the culture reached 0.6 to 0.8, isopropyl β -D-1-thiogalactopyranoside (IPTG; Sigma) was added to a final concentration of 0.2–1.0 mM and the culture was further grown for 4–20 h at 37–18 °C under shaking of 200 rpm. A negative control was prepared using the same protocol of expression with the exception that IPTG was not added to the culture. To check the level of expression, each culture sample was centrifuged at 4000 rpm (1860 × g) and the corresponding cell pellet was lysed by an ultrasonic sonicator or using BugBuster lyse solution (Millipore) following the manufacturer instructions. The soluble and insoluble fractions were separated by centrifugation at 10,000 rpm (11,627 × g) for 10 min and 5 μ L of each fraction was separated and checked on a SDS-PAGE 10–15% (w/v) which was stained using Coomassie Brilliant Blue (Sigma). For larger scale protein production, we chose the best expression condition identified in the expression test and in order to produce a satisfactory yield of protein for biophysical experiments, we usually inoculated 200 mL of pre-inoculum into 4 L culture. After the expression protocol as defined above, the cells was pelleted by centrifugation at 4000 rpm (1860 × g) for 10 min at 4 °C and the resulting pellet was stored at –20 °C until purification.

For the purification, cell pellets (approximately 10 g of cells) were re-suspended in 50–75 mL of buffer (generally 50 mM Potassium Phosphate and 10% glycerol pH 7.0 or 50 mM Tris-HCl,

100 mM NaCl, pH 7.8). In this cell suspension, 100 μ L of Phenylmethylsulfonyl fluoride (PMSF) (Sigma-Aldrich) 10 mM, 100 μ L lysozyme (Sigma-Aldrich) 10 mg/mL and 100 μ L deoxyribonuclease I (Sigma-Aldrich) 1 mg/mL were further added. The cells were homogenized using an ultrasonic sonicator (Branson) continuously for 15 min with pulses of 5 s–5 s intervals at a frequency of 50% under ice bath. The cell debris were pelleted by centrifugation at 10,000 rpm (11,627 \times g) for 30 min at 4 °C and the supernatant was used for protein purification.

For the protein purification, an Akta Purifier (General Electric) system (FPLC- Fast Protein Liquid Chromatography) was used. The soluble lysate was loaded to a HisTrap IMAC HP 5 mL column charged with nickel and pre-equilibrated with buffer A (50 mM Tris-HCl, 100 mM NaCl, pH 7.8) and for the elution of binding proteins, buffer B (buffer A plus 500 mM imidazole) was used. Before protein elution, the column was washed with 100 mL of buffer A plus 25 mM of imidazole. The binding protein was eluted using a linear gradient of buffer B to 100% with a flow of 2 mL min⁻¹ during 30 min. The fraction containing the eluted protein was collected and concentrated to a final volume of 2 mL, which was loaded to a Superdex 200 (16/60) column (GE Healthcare) previously equilibrated with buffer A. The fractions containing the eluted protein were collected and concentrated to 10 mg/mL, flash frozen in liquid nitrogen, and stored at –80 °C until further biophysical experiments.

To check the purity of the eluted proteins of both affinity and molecular size exclusion purifications, the fractions corresponding to the peaks were analyzed on SDS-PAGE 15%. For the analysis of the affinity purification step, we also added as controls on the SDS-PAGE, samples of dissolved pellets in loading buffer, supernatants, non-binding protein solution and the wash solution with 25 mM of imidazole. All gels were stained with Coomassie Brilliant Blue.

1.2. Estimative of hydrodynamic mass

The hydrodynamic masses of Stal and StaK were calculated by a calibration of the Superdex 200 (16/60) column. For the calibration of the column, we used the HMW and LMW kits (GE Healthcare) containing carbonic anhydrase (29,000,000 M_r), ovalbumin (43,000,000 M_r), conalbumin (75,000,000 M_r), aldolase (158,000,000 M_r), ferritin (440,000,000 M_r) and blue dextran (>2,000,000 M_r). Blue dextran was separately used to estimate the void volume of the Superdex 200 (16/60). The calibration was effected in the same buffer used for the purification of Stal and StaK. The samples were injected using a loop of 200 μ L at concentration of 3–10 mg ml⁻¹ and with a flow rate of one mL.min⁻¹. The analysis of Stal, StaK and the calibration curves, along with the linear regression, were performed using the program Origin 2016.

1.3. Circular dichroism

Circular dichroism measurements were performed on a JASCO J810 spectropolarimeter (Tokyo, Japan) equipped with a temperature controller (Peltier Type Control System PFD 4255, JASCO). A circular quartz cuvette of 0.05 cm optical path or a rectangular quartz cuvette with 1 cm optical path was used, and samples of both Stal and StaK were at a concentration of 10 μ mol L⁻¹ in a buffer constituted of 50 mM Tris-HCl, 100 mM of NaCl, pH 7.8. Spectra were obtained over a wavelength between 190 and 260 nm with a scanning speed of 50 nm min⁻¹. The curves were plotted using the software Excel or Origin 2016. The deconvolution of the data were carried out using the online server K2D3 [24].

Thermal denaturation was monitored at 222 nm as function of temperature. The protein concentration and the used buffer are the same as described above, and the temperature was increased in a

rate of 1.0 °C.min⁻¹ from 10 to 100 °C. The melting temperature (T_m) was measured using the Boltzman fitting equation on Origin 2016 and identifying the point at which 50% of the samples were unfolded.

1.4. Molecular modelling

Molecular models for both Stal and StaK were constructed using the program Modeller [25]. The best templates were obtained by a search using the online server BlastP [26] using the sequences of Stal and StaK against the Protein Data Bank. The best structural template for both enzymes were CndH (3E1T PDB code), a halogenase from the chondrocloren biosynthesis in Myxobacteria [27]. The identity of CndH to Stal and StaK were 56 and 53%, respectively. A total of ten models were generated for both Stal and StaK and the best models were selected based on the stereochemistry quality, which was obtained by the program SFcheck [28]. Visual inspection of the models was performed using the program COOT [29] and PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.), which was also used to calculate the electrostatic surfaces.

1.5. Alignment analysis and phylogeny

The protein sequences of halogenases or putative halogenases identified in biosynthetic gene clusters of different glycopeptides were obtained in the database Uniprot (<http://www.uniprot.org/>) and the alignments were done using the online programs Clustal Omega [30], MultAlin [31] or Muscle [32].

The phylogenetic tree of 9 different sequences of halogenases or putative ones from glycopeptide gene clusters was constructed by the program MEGA6 [33]. The evolutionary histories of these enzymes were inferred by using the Maximum Likelihood method based on the Whelan and Goldman [34] with a bootstrap of 1000 replicates. The tree with the highest log likelihood was chosen and

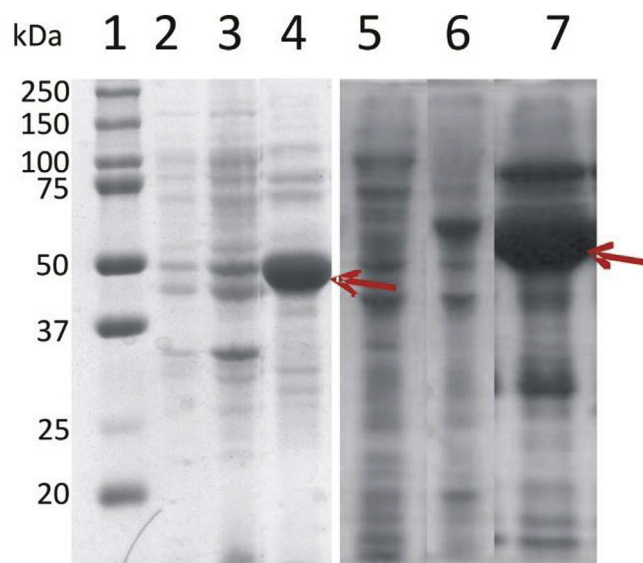


Fig. 1. SDS PAGE analysis of the superexpression experiment and for the first step purification using a Ni-NTA column chromatography for both putative halogenases. Lane 1: LMW-SDS; lane 2: insoluble fraction of StaK. Lane 3: soluble fraction of StaK, lane 4: elution of StaK of the nickel column using 130 mM of imidazole; lane 5: insoluble fraction of Stal; lane 6: soluble fraction of Stal; lane 7: elution of Stal of the nickel column using 60 mM of imidazole. To prepare the gel, it was used 5 μ L of sample and 5 μ L of loading buffer which was boiled for 5 min. The gels were stained using Coomassie Brilliant Blue. The halogenases are indicated using arrows, which are in agreement with the expected molecular weight (53 kDa and 63 kDa for StaK and Stal, respectively).

it was drawn to scale, with branch lengths measured in the number of substitutions per site.

2. Results

stal and *staK* from *Streptomyces toyocaensis* NRRL15,009 were

cloned into pET28a(+) vector and the expression resulted in a recombinant protein fused with a 6 × His-tag on the N-terminal (His-Stal and His-StaK). The best expression condition for His-Stal was obtained with the induction of 0.2 μM of IPTG for 16 h at 18° C using BL21(DE3) cells. Using this condition, we have obtained about 10 g of cells containing both plasmid constructs carrying *stal* and *staK*.

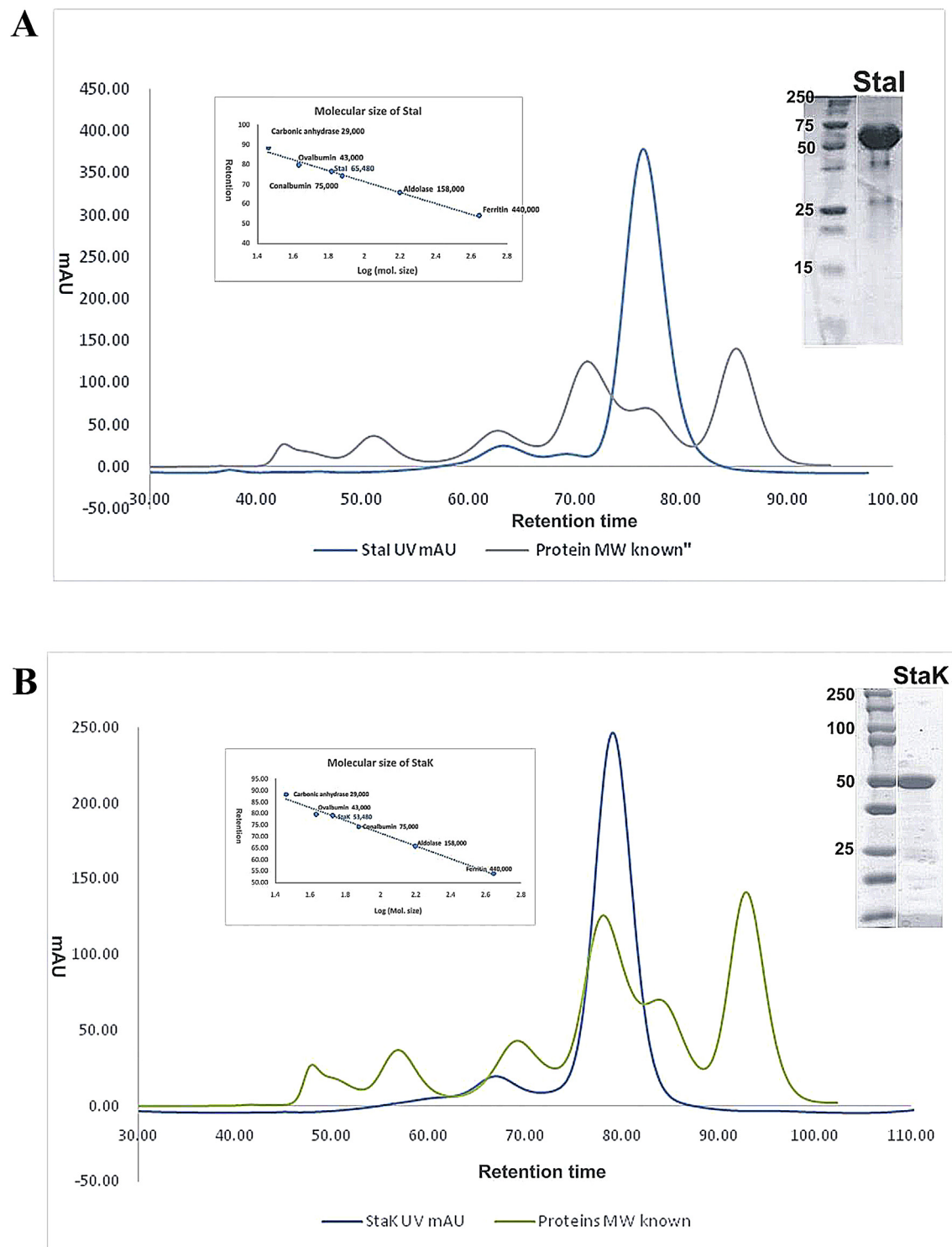


Fig. 2. Elution profile of (A) Stal and (B) StaK. Gray and green lines show the peaks of the proteins with known molecular weight used to calibrate a Superdex 200 (16/60) column (carbonic anhydrase, ovalbumin, conalbumin, aldolase and ferritin). The blue line represents the retention time for (A) Stal and (B) StaK. To obtain the molecular weight of both enzymes we have used a linear regression by Origin or Excel software (left inset). The right inset, shows the result of the molecular exclusion purification step and the purity of the sample used for biophysical analysis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Following this condition, approximately 80% of His-Stal was found in the soluble fraction by analysis of SDS-PAGE stained using Coomassie Brilliant Blue (Fig. 1). On the other hand, various attempts to obtain StaK were performed, including using different IPTG concentrations, temperatures, mediums and different expression systems, as BL21(DE3), Rosetta(DE3), BL21-pLys(DE3) and co-expression with the chaperone system pGRO7 (GroEL/GroES) (Takara). However, overexpression of the protein was not observed when the soluble and insoluble fractions were analyzed by SDS-PAGE. Nevertheless, in an analysis of Western blot using an anti-His antibody, it was possible to observe a very low expression of this enzyme in both soluble and insoluble fraction (Figure not shown). This encouraged us to obtain a synthetic gene with codon optimization for expression in *E. coli* to produce StaK. The gene was cloned in pET28a(+) vector, which also resulted in a $6 \times$ His-Tag product and the expression was carried out under the same conditions of His-Stal. About 25% of the protein was found in the soluble fraction when an aliquot of the result of superexpression was analyzed in SDS-PAGE (Fig. 1).

Both enzymes were purified in two chromatographic steps, a Ni-NTA His-tag affinity (Fig. 1) followed by molecular size exclusion using a Superdex 200 (16/60) gel filtration column (Fig. 2). His-Stal binds weakly to the affinity column and elutes in a low concentration of imidazole and, consequently the sample had a high concentration of impurity. Both enzymes were yellowish indicating the co-purification with FAD, as expected. A scanning using the wavelength between 300 and 800 nm confirmed the co-purification with FAD in both proteins (Fig. 3). However, after the

molecular exclusion step, the protein was satisfactorily pure for further biophysical analysis. His-StaK eluted using a higher concentration of imidazole and the sample was practically pure after the affinity column step. However, to increase the purity of this enzyme, a second step using size exclusion chromatography was performed. Analysis of the molecular weight of both proteins using a calibrated Superdex 200 (16/60) indicated that Stal and StaK eluted as a monomer with a hydrodynamic mass of 56 kDa and 46 kDa, respectively (Fig. 2), which is in agreement with the theoric molecular weight values obtained in the Uniprot Database. About 40% and 15% of Stal and StaK, respectively, eluted in the void volume of the gel filtration column, indicating that there was the formation of large oligomers or aggregation, and these fractions were not used for further experiments. Approximately 10 and 5 mg of pure protein was obtained per liter of culture for Stal and StaK, respectively. To check the ability of a flavin reductase to reduce the bound flavin at Stal and StaK, we have cloned, expressed and purified the enzyme SsuE from *E. coli* and performed an assay as described in Schartz et al., 2014 [8]. Briefly, Stal and StaK at 20 μ M were incubated in the presence of the of SsuE 10 μ M and 1 mM of NADPH for 2 min. The yellowish color of the FAD could be rapidly lost because of its reduction to FADH₂ and the oxidation of NADPH to NADP. However, the FADH₂ is again rapidly re-oxidized to FAD because of the aerobic condition used in the assay. Interestingly this experiment indicates that the FAD bound to Stal and StaK have different availability to SsuE (Fig. 3). While the FAD bound to Stal is partially reduced by SsuE, the FAD bound to StaK is not completely available to SsuE, since there is no significant decreasing of

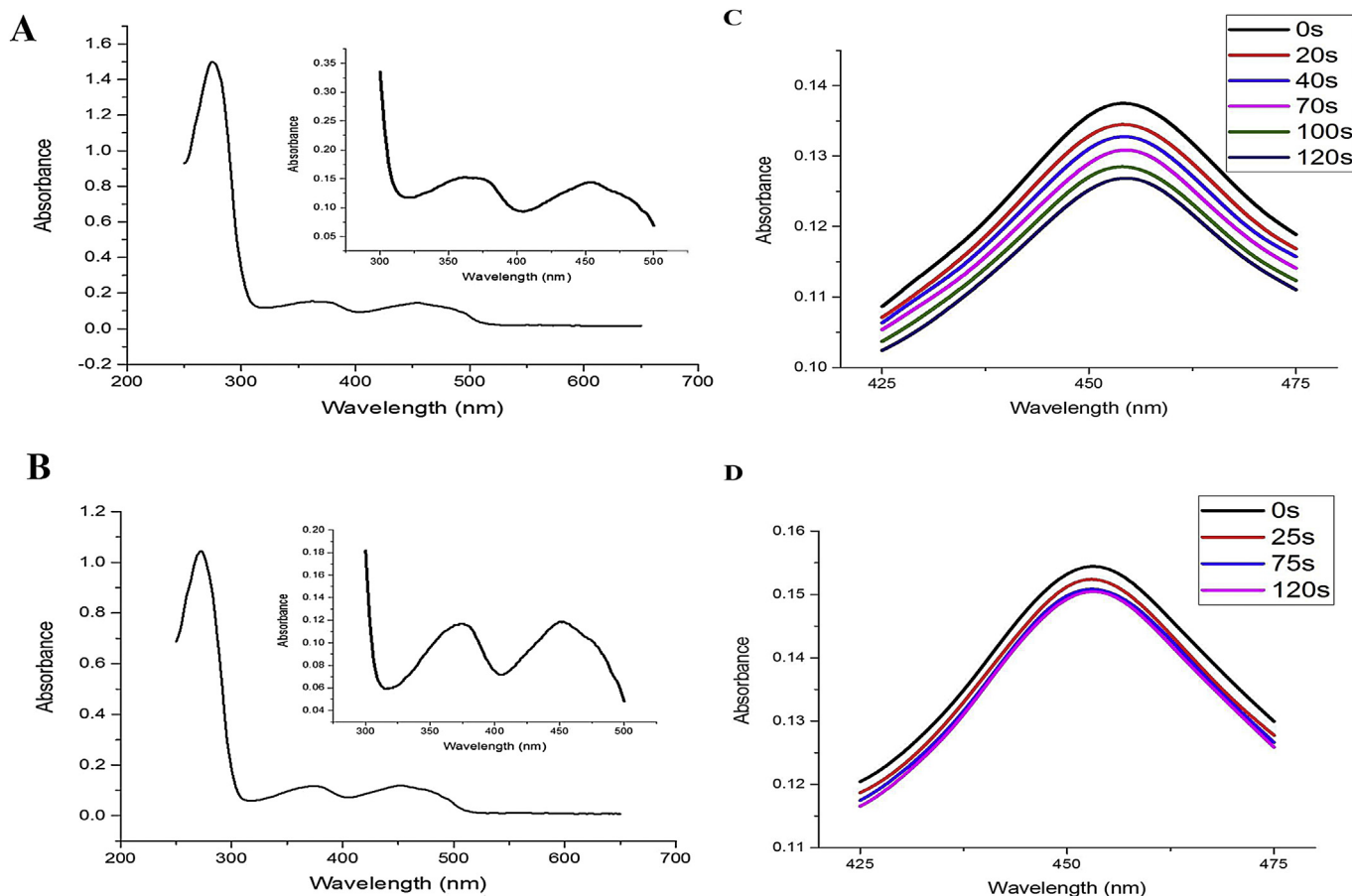


Fig. 3. Analysis of the UV-Vis spectra of Stal and StaK. A) Spectra of Stal and B) StaK indicating the co-purification with FAD. C) Analysis of the reduction of the FAD bound to Stal by SsuE. D) Analysis of the reduction of the FAD bound to StaK by SsuE.

absorbance at 452 nm indicating that it is more buried within the active site of the protein. Thus, this assay demonstrated that alike VhaA [7], Stal could also be reduced by a flavin reductase as expected considering these two enzymes are more similar (84% of identity), while StaK is more divergent (61% of identity) and might have a less accessible active site.

Circular dichroism (CD) analysis showed that both halogenases are folded and have similar spectra with negative bands around 208 nm and 220 nm and positive bands around 190 nm, which is an indication of a high content of α -helices (Fig. 4). We performed the deconvolution of the data using the web server K2D3 [24] to estimate the percentage of secondary elements of both samples. Stal has an estimated secondary structure content of about 27% of α -helices and 17% of β -sheets, while StaK has approximately 32% of α -helices and 20% of β -sheets. The CD thermal denaturation of both proteins showed that Stal and StaK have a melting temperature about of 64 °C and 70 °C, respectively, indicating that these proteins are quite stable (Fig. 4).

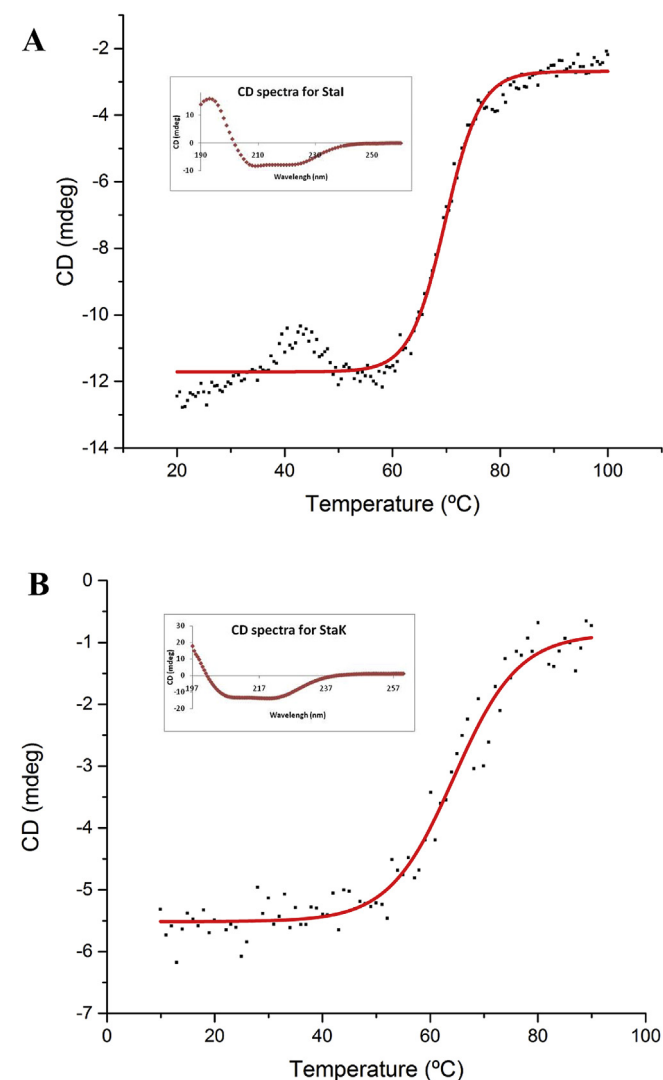


Fig. 4. Circular dichroism denaturation curves for (A) Stal and (B) StaK. Fraction of unfolded protein was monitored as a function of temperature by CD signal at 222 nm (■). 10 μ M mL⁻¹ of protein was heated at a rate of 1 °C.min⁻¹ starting from 0 or 10 °C to 90–100 °C. The T_m was considered when half of the protein samples were denatured. A Boltzmann fitting using the Origin Software is also shown and it was used to calculate the T_m. Inset: CD spectra of (A) Stal and (B) StaK. To calculate the percentage of secondary elements content, we used the program K2D3 [24].

2.1. Molecular modeling

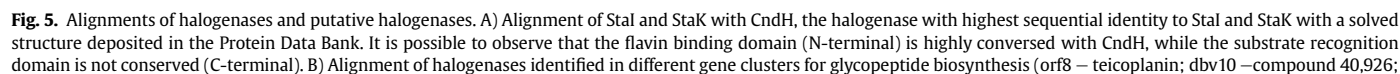
To gain insights into the possible structures of Stal and StaK, we used molecular modeling to construct models for both putative halogenases. We have used the amino acid residues sequence of Stal and StaK to perform a search on the program BlastP against the Protein Data Bank (PDB). The protein which matched the best score and the highest similarity was CndH, a halogenase from *Chondromyces crocatus* involved in the chondrochloren biosynthesis [27]. Stal and StaK have a sequential identity of 56% and 53% with CndH, respectively (Fig. 5a). CndH is a halogenase that belongs to the glutathione reductase superfamily and has a characteristic folding of the FAD binding proteins. The halogenases of this superfamily or the FADH₂ dependent halogenases are described to have a folding constituted of a single domain with two modules or subdomains: a box subdomain in the N Terminal region and a triangular pyramid subdomain in the C terminal region [35]. The box subdomain is the flavin-binding module, which is highly conserved in this family and characterized to have two large β -sheets, which correspond to the “ceiling” and “floor” of the box and four α -helices which are said to be the lateral walls of the box. The triangular pyramid subdomain is the substrate recognition module, which generally has three α -helices forming a pyramidal shape and is not conserved in different halogenases (Figure not shown) [35]. However, the C Terminal portion of CndH is lacking its tridimensional structure and it was attributed to this region being disordered or denatured during the crystal packing formation [27]. Thus, due to the absence of the C Terminal subdomain in CndH and the low similarity of this region with any other proteins with tridimensional structures, we have not modeled the substrate recognition domain for Stal and StaK. Based on our molecular models, both Stal and StaK have the flavin binding subdomain very similar to CndH.

Taking the C Terminal domain as disordered, and calculating the percentage of secondary elements, we have an theoretic estimative for our models of: 18.7% of β -sheet and 27.2% of α -helices for Stal, and 31.3% of β -sheet and 21.8% α -helices for StaK. These percentages are near of the experimental estimative obtained by circular dichroism (17% of β -sheet and 27% of α -helices for Stal and 32% of β -sheet and 20% α -helices for StaK). The agreement between these values indicates that our models have a high probability to be correct and corroborates the possibility of the C terminal domain of the putative halogenases Stal and StaK to be disordered.

In addition, according to the mechanism of other halogenases, we have further evidence that C Terminal subdomains are disordered in our structures in the absence of heptaptidyl-PCP substrate. The structure of PrnA, one of the best characterized halogenase from flavin halogenase family, the substrate binds between the FAD binding subdomain and the substrate recognition subdomain [35] and this event seems to trigger a movement engaging the substrate binding subdomain closer to the flavin binding subdomain for the activated chloride transfer [35].

Using an electrostatic surface analysis for both Stal and StaK, we could infer the hypothetical region which forms a large tunnel in the opposite side of the FAD binding site as described in the structure of PrnA [35]. In addition, as described in the structure of CndH [27], it is possible to observe in both structures a large cleft where the peptidyl-PCP substrate should be accommodated for the catalysis (Fig. 6a–b).

A detailed analysis of the non-conserved aromatic residues of StaK observed in the alignment (Fig. 5b) shows that several of these residues are not involved in the formation of the halogen transfer tunnel, which has the conserved amino acid residue Lys76, an essential residue for the active halogen transference to the substrate. However, most of these residues are clustered in a region,



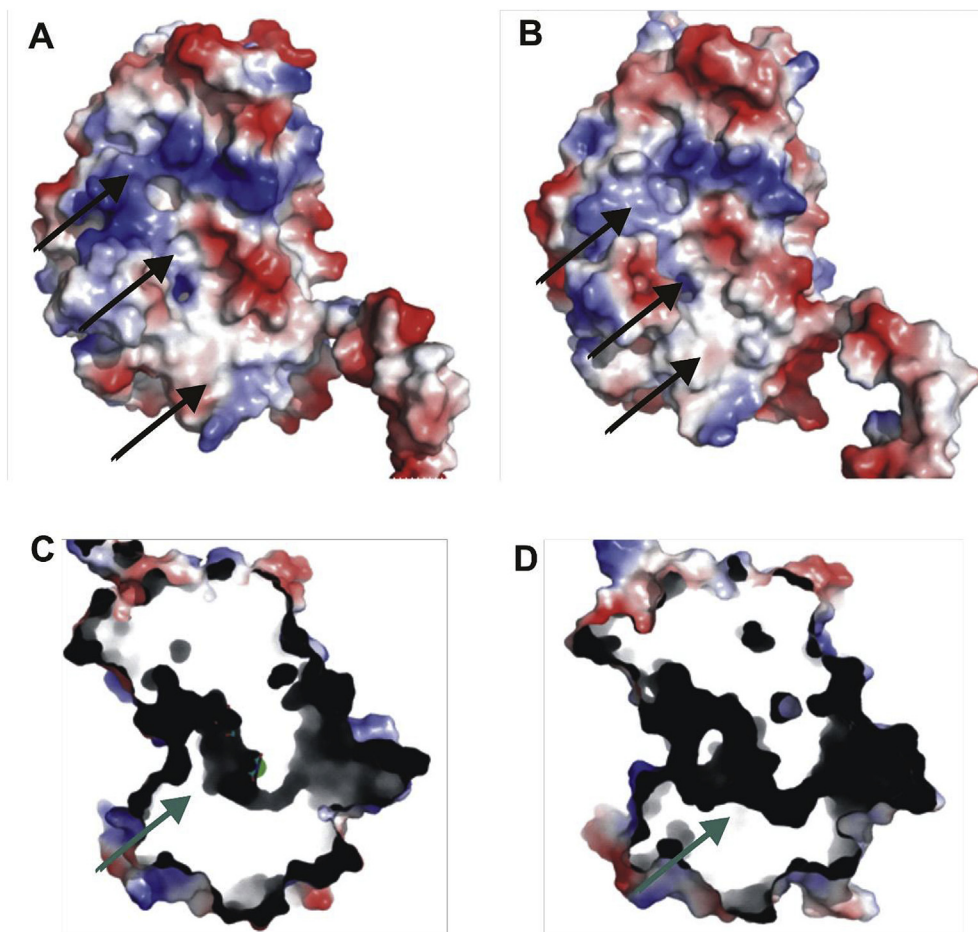


Fig. 6. Electrostatic surfaces for (A) StaI and (B) StaK models. The arrows indicate the region that the heptapeptidyl-PCP substrate should anchor to the putative halogenases. In (C) and (D) it is possible to observe the characteristic halogen transfer tunnel, as indicated by the arrows of flavin dependent halogenases.

which probably interacts with the flavin-binding subdomain near of the cleft where the peptidyl-PCP should bind (Fig. 7). The position of these residues in this interface might influence the binding mode of the substrate, which allows the correct positioning for catalysis and regiospecificity of both halogenases.

2.2. Phylogenetic analysis and sequence analysis

To gain insights concerning the divergence and evolution of halogenases involved in the biosynthesis of glycopeptide antibiotics, we have also performed an alignment of all putative halogenases sequences identified in different glycopeptide biosynthetic gene clusters and also constructed a phylogenetic tree. The alignment of these enzymes reveals that they have, in general, high sequential identity to the N terminal subdomain (Fig. 5b). However, the putative halogenases StaK and Auk23 have a deletion of about 40 amino acid residues in the C Terminal subdomain while all the other sequences have this segment very conserved. In addition, StaK and Auk23 have other similarities in the sequence that are divergent from the other glycopeptide biosynthesis halogenases (Fig. 5b). Thus, the phylogenetic analysis

reveals, as expected, that StaK and Auk23 are separated in a different branch (Fig. 8b) supporting that enzymes has evolved independently from the other halogenases and eventually could have gained a slightly different function. However to confirm this hypothesis, experiment of gene disruption and complementation in *Streptomyces toyocaensis* and further analysis of the produced metabolites are needed.

3. Discussion

The FADH halogenases or the putative ones from the biosynthesis of glycopeptides have a biotechnological potential for the use of combinatorial biosynthesis. However, these enzymes are not sufficiently studied and there is no tridimensional structure for them. In these studies, we have performed a preliminary characterization of the two putative halogenases from the biosynthesis of the A47,934, one of the two known glycopeptides which have two putative halogenases identified in their biosynthetic gene clusters. The function of the StaI and StaK in the biosynthesis of this glycopeptide is not well defined since this glycopeptide has a third halogenation at residue 5 of heptapeptide which differs from the

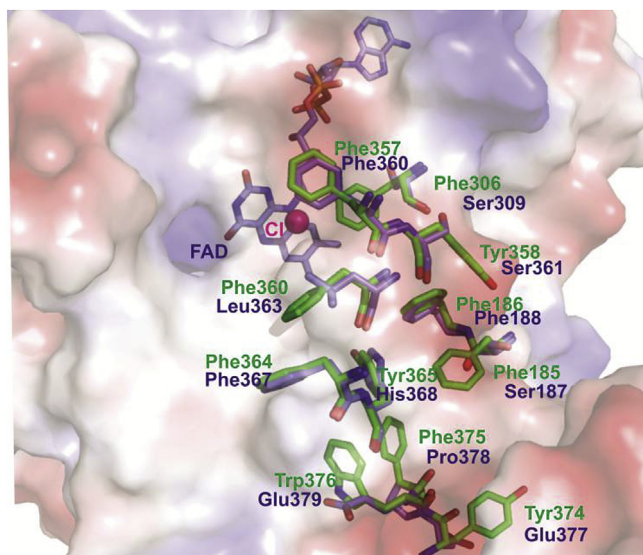


Fig. 7. Superposition of Stal and StaK molecular models. The superposition shows the region of the flavin binding domain near to the substrate recognition domain with a predominance of substitutions of aromatic or other large side chain residues (Stal) for shorter side chain residues (StaK).

other glycopeptides, with exception of UK-68,597 [15]. In the biosynthesis of A47,934, Stal and StaK share high identity (65%), even though StaK has a smaller substrate-binding domain and does not have several conserved aromatic amino acids. Although both enzymes are similar, the wild type gene showed very little expression in *E. coli* and to obtain a satisfactory yield of protein, a synthetic gene for StaK with codon optimization for this host was necessary. Both proteins shown soluble and are yellowish in color,

which is an indication of the tiny binding of FAD, which is differently accessible to the reduction by flavin reductases, such as SsuE. Stal has a massive predominance of expression in the soluble fraction while StaK had a small portion of the expression in the insoluble fraction. On the other hand, Stal has a higher tendency to aggregate as it had a large amount of protein in the void volume of the size exclusion chromatography. Analysis of CD indicate the integrity of both proteins, its folding and a high predominance of α -helices in agreement with our molecular models. Results of analytical gel filtration suggest that both Stal and StaK are monomeric enzymes. Using a phylogenetic analysis, we could observe that StaK and Auk23 are more divergent from the other halogenases and are grouped in a different branch of a phylogenetic tree. We have also constructed molecular models for Stal and StaK and we could observe that both models are similar between them and very similar to CndH, a halogenase involved in the chondrochloren biosynthesis [27]. However, StaK has several substitutions and a gap of 40 residues in the C Terminal subdomain which are divergent to the other putative halogenases, with exception of Auk23. These differences might be involved in a slightly different function of StaK as the halogenation at the position 5 of the heptapeptidyl-PCP. An *in vivo* experiment with the *staK* deletion and analysis of the produced metabolites might confirm this hypothesis.

Acknowledgements

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo under grant 2010/15971-3 and fellowships 2012/23427-7 for TPC and 2012/16631-7 for LAS. SMCP also thanks CAPES for his fellowship. We would like to thank the facilities of the Laboratory of Spectrometry and Calorimetry (LEC) and the National Laboratory of Bioscience for assistance and Dr. Juliana Fatori for helpful discussion and assistance during the experiments of circular dichroism.

References

- [1] Y. Keynan, E. Rubinstein, Staphylococcus aureus bacteremia, risk factors, complications, and management, Crit. Care Clin. 29 (3) (2013) 547–562, <http://dx.doi.org/10.1016/j.ccc.2013.03.008>. PMID: 23830653.
- [2] S. Gardete, A. Tomasz, Mechanisms of vancomycin resistance in Staphylococcus aureus, J. Clin. Invest. 124 (7) (2014) 2836–2840, <http://dx.doi.org/10.1172/JCI68834>. PMID: 24983424 PMID: PMC4071404.
- [3] M. Bassetti, M. Merelli, C. Temperoni, A. Astilean, New antibiotics for bad bugs: where are we? Ann. Clin. Microbiol. Antimicrob. 12 (2013) 22, <http://dx.doi.org/10.1186/1476-0711-12-22>. PMID: 23984642 PMID: PMC3846448.
- [4] R. Nagarajan, Glycopeptide Antibiotics, Marcel Dekker, New York, 1994.
- [5] C. Harris, R. Kannan, H. Kopecka, T. Harris, The role of the chlorine substituents in the antibiotic vancomycin – preparation and characterization of monodechlorovancomycin and didechlorovancomycin, J. Am. Chem. Soc. 107 (23) (1985) 6652–6658, <http://dx.doi.org/10.1021/ja00309a038>. PMID: WOS: A1985AUA7900038.
- [6] U. Gerhard, J. Mackay, R. Maplestone, D. Williams, The Role of the sugar and chlorine substituents in the dimerization of vancomycin antibiotics, J. Am. Chem. Soc. 115 (1) (1993) 232–237, <http://dx.doi.org/10.1021/ja00054a033>. PMID: WOS: A1993KG95400033.
- [7] P.C. Schmartz, K. Zerbe, K. Abou-Hadeed, J.A. Robinson, Bis-chlorination of a hexapeptide-PCP conjugate by the halogenase involved in vancomycin biosynthesis, Org. Biomol. Chem. 12 (30) (2014) 5574–5577, <http://dx.doi.org/10.1039/c4ob00474d>. PMID: 24756572.
- [8] O. Puk, P. Huber, D. Bischoff, J. Recktenwald, G. Jung, R.D. Süßmuth, K.H. van Pée, W. Wohlleben, S. Pelzer, Glycopeptide biosynthesis in Amycolatopsis mediterranei DSM5908: function of a halogenase and a haloperoxidase/perhydrolase, Chem. Biol. 9 (2) (2002) 225–235. PMID: 11880037.
- [9] A.M. van Wageningen, P.N. Kirkpatrick, D.H. Williams, B.R. Harris, J.K. Kershaw, N.J. Lennard, M. Jones, S.J. Jones, P.J. Solenberg, Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic, Chem. Biol. 5 (3) (1998) 155–162. PMID: 9545426.
- [10] S. Pelzer, R. Süßmuth, D. Heckmann, J. Recktenwald, P. Huber, G. Jung, W. Wohlleben, Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in Amycolatopsis mediterranei DSM5908, Antimicrob. Agents Chemother. 43 (7) (1999)

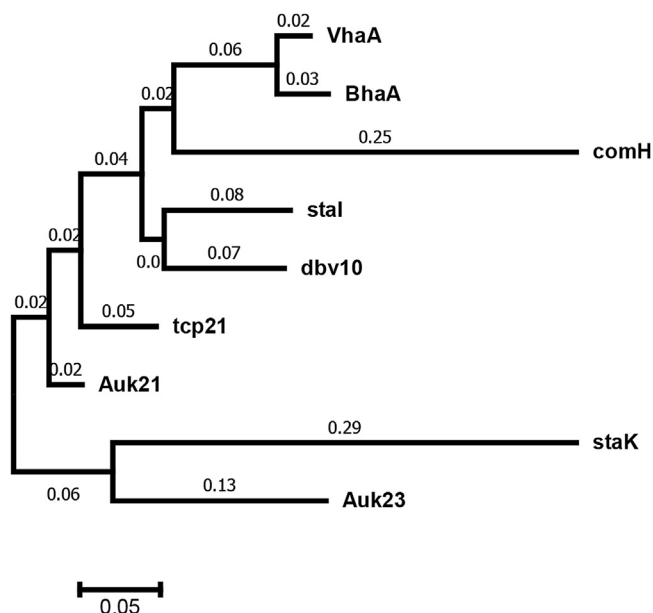


Fig. 8. Molecular Phylogenetic analysis by Maximum Likelihood method of the different halogenases and putative ones identified in glycopeptide gene clusters. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site (next to the branches). The name of proteins are from following glycopeptide gene cluster: VhaA (vancomycin); BhaA (balhimycin); ComH (Complestatin); Stal and StaK (Compound A47,934); Dbv10 (compound A40,926); Tcp21 (Teicoplanin); Auk21 and Auk23 (compound UK68,597).

- 1565–1573. PMID: 10390204 PMCID: PMC89325].
- [11] R.M. Shawky, O. Puk, A. Wietzorrek, S. Pelzer, E. Takano, W. Wohlleben, E. Stegmann, The border sequence of the balhimycin biosynthesis gene cluster from *Amycolatopsis balhimycina* contains bbr, encoding a StrR-like pathway-specific regulator, *J. Mol. Microbiol. Biotechnol.* 13 (1–3) (2007) 76–88, <http://dx.doi.org/10.1159/000103599>. PMID: 17693715.
 - [12] T.L. Li, F. Huang, S.F. Haydock, T. Mironenko, P.F. Leadlay, J.B. Spencer, Biosynthetic gene cluster of the glycopeptide antibiotic teicoplanin: characterization of two glycosyltransferases and the key acyltransferase, *Chem. Biol.* 11 (1) (2004) 107–119, <http://dx.doi.org/10.1016/j.chembiol.2004.01.001>. PMID: 15113000.
 - [13] A.W. Truman, M.J. Kwun, J. Cheng, S.H. Yang, J.W. Suh, H.J. Hong, Antibiotic resistance mechanisms inform discovery: identification and characterization of a novel amycolatopsis strain producing ristocetin, *Antimicrob. Agents Chemother.* 58 (10) (2014) 5687–5695, <http://dx.doi.org/10.1128/AAC.03349-14>. PMID: 25022591 PMCID: PMC4187901.
 - [14] J. Pootoolal, M.G. Thomas, C.G. Marshall, J.M. Neu, B.K. Hubbard, C.T. Walsh, G.D. Wright, Assembling the glycopeptide antibiotic scaffold: the biosynthesis of A47934 from *Streptomyces toyocaensis* NRRL15009, *Proc. Natl. Acad. Sci. U. S. A.* 99 (13) (2002) 8962–8967, <http://dx.doi.org/10.1073/pnas.102285099>. PMID: 12060705 PMCID: PMC124406.
 - [15] G. Yim, L. Kalan, K. Koteva, M.N. Thaker, N. Wagglechner, I. Tang, G.D. Wright, Harnessing the synthetic capabilities of glycopeptide antibiotic tailoring enzymes: characterization of the UK-68,597 biosynthetic cluster, *Chembiochem* 15 (17) (2014) 2613–2623, <http://dx.doi.org/10.1002/cbic.201402179>. PMID: 25255985.
 - [16] S. Brown, S.E. O'Connor, Halogenase engineering for the generation of new natural product analogues, *Chembiochem* 16 (15) (2015) 2129–2135, <http://dx.doi.org/10.1002/cbic.201500338>. PMID: 26256103.
 - [17] D.R. Smith, S. Grischow, R.J. Goss, Scope and potential of halogenases in biosynthetic applications, *Curr. Opin. Chem. Biol.* 17 (2) (2013) 276–283, <http://dx.doi.org/10.1016/j.cbpa.2013.01.018>. PMID: 23433955.
 - [18] C. Sánchez, L. Zhu, A.F. Braña, A.P. Salas, J. Rohr, C. Méndez, J.A. Salas, Combinatorial biosynthesis of antitumor indolocarbazole compounds, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2) (2005) 461–466, <http://dx.doi.org/10.1073/pnas.0407809102>. PMID: 15625109 PMCID: PMC544307.
 - [19] E. Yeh, S. Garneau, C.T. Walsh, Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis, *Proc. Natl. Acad. Sci. U. S. A.* 102 (11) (2005) 3960–3965, <http://dx.doi.org/10.1073/pnas.0500755102>. PMID: 15743914 PMCID: PMC554827.
 - [20] S. Zehner, A. Kotzsch, B. Bister, R.D. Stüssmuth, C. Méndez, J.A. Salas, K.H. van Pée, A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005, *Chem. Biol.* 12 (4) (2005) 445–452, <http://dx.doi.org/10.1016/j.chembiol.2005.02.005>. PMID: 15850981.
 - [21] A.S. Eustáquio, B. Gust, S.M. Li, S. Pelzer, W. Wohlleben, K.F. Chater, L. Heide, Production of 8'-halogenated and 8'-unsubstituted novobiocin derivatives in genetically engineered streptomyces coelicolor strains, *Chem. Biol.* 11 (11) (2004) 1561–1572, <http://dx.doi.org/10.1016/j.chembiol.2004.09.009>. PMID: 15556007.
 - [22] A. Deb Roy, S. Grischow, N. Cairns, R.J. Goss, Gene expression enabling synthetic diversification of natural products: chemogenetic generation of pacidamycin analogs, *J. Am. Chem. Soc.* 132 (35) (2010) 12243–12245, <http://dx.doi.org/10.1021/ja1060406>. PMID: 20712319.
 - [23] L.C. Blasiak, C.L. Drennan, Structural perspective on enzymatic halogenation, *Acc. Chem. Res.* 42 (1) (2009) 147–155, <http://dx.doi.org/10.1021/ar800088r>. PMID: 18774824 PMCID: PMC3980734.
 - [24] C. Louis-Jeune, M.A. Andrade-Navarro, C. Perez-Iratxeta, Prediction of protein secondary structure from circular dichroism using theoretically derived spectra, *Proteins* 80 (2) (2012) 374–381, <http://dx.doi.org/10.1002/prot.23188>. PMID: 22095872.
 - [25] A. Sali, J.P. Overington, Derivation of rules for comparative protein modeling from a database of protein structure alignments, *Protein Sci.* 3 (9) (1994) 1582–1596, <http://dx.doi.org/10.1002/pro.5560030923>. PMID: 7833817 PMCID: PMC2142932.
 - [26] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: architecture and applications, *BMC Bioinforma.* 10 (2009) 421, <http://dx.doi.org/10.1186/1471-2105-10-421>. PMID: 20003500 PMCID: PMC2803857.
 - [27] S. Buedenbender, S. Rachid, R. Müller, G.E. Schulz, Structure and action of the myxobacterial chondrochloren halogenase CndH: a new variant of FAD-dependent halogenases, *J. Mol. Biol.* 385 (2) (2009) 520–530, <http://dx.doi.org/10.1016/j.jmb.2008.10.057>. PMID: 19000696.
 - [28] A.A. Vaguine, J. Richelle, S.J. Wodak, SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model, *Acta Crystallogr. D. Biol. Crystallogr.* 55 (Pt 1) (1999) 191–205, <http://dx.doi.org/10.1107/S0907444998006684>. PMID: 10089410.
 - [29] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, *Acta Crystallogr. D. Biol. Crystallogr.* 60 (Pt 12 Pt 1) (2004) 2126–2132, <http://dx.doi.org/10.1107/S0907444904019158>. PMID: 15572765.
 - [30] H. McWilliam, W. Li, M. Uludag, S. Squizzato, Y.M. Park, N. Buso, A.P. Cowley, R. Lopez, Analysis tool web services from the EMBL-EBI, *Nucleic Acids Res.* 41 (Web Server issue) (2013) W597–W600, <http://dx.doi.org/10.1093/nar/gkt376>. PMID: 23671338 PMCID: PMC3692137.
 - [31] F. Corpet, Multiple sequence alignment with hierarchical clustering, *Nucleic Acids Res.* 16 (22) (1988) 10881–10890. PMID: 2849754 PMCID: PMC338945.
 - [32] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.* 32 (5) (2004) 1792–1797, <http://dx.doi.org/10.1093/nar/gkh340>. PMID: 15034147 PMCID: PMC339037.
 - [33] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, *Mol. Biol. Evol.* 30 (12) (2013) 2725–2729, <http://dx.doi.org/10.1093/molbev/mst197>. PMID: 24132122 PMCID: PMC3840312.
 - [34] S. Whelan, N. Goldman, A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach, *Mol. Biol. Evol.* 18 (5) (2001) 691–699. PMID: 11319253].
 - [35] C. Dong, S. Flecks, S. Unversucht, C. Haupt, K.H. van Pée, J.H. Naismith, Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination, *Science* 309 (5744) (2005) 2216–2219, <http://dx.doi.org/10.1126/science.1116510>. PMID: 16195462 PMCID: PMC3315827.
 - [36] X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, *Nucleic Acids Res.* 42 (Web Server issue) (2014) W320–W324, <http://dx.doi.org/10.1093/nar/gku316>. PMID: 24753421 PMCID: PMC4086106.