Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Structural-based engineering expands the substrate scope of a cyclodipeptide synthase

Emmajay Sutherland,[a] Christopher John Harding [a] and Clarissa Melo Czekster[a]

Cyclodipeptide synthases (CDPSs) are a growing family of enzymes capable of producing a large variety of cyclodipetide products using aminoacylated tRNA. Histidine-containing cyclic dipeptides have important biological activities as anticancer and neuroprotective molecules. Out of the 120 experimentally validated CDPS members, only two are known to accept histidine as a substrate. Here, we studied the activities of both *Para*-CDPS from *Parabacteroides sp. 20_3* and *Parcu*-CDPS from *Parcubacteria bacterium RAAC4_OD1_1* which synthesise cyclo(His-Phe) and cyclo(His-Pro) respectively. Both enzymes accepted canonical and non-canonical amino acids as substrates to generate a library of novel molecules. In order to understand the substrate selectivity of these CDPSs, the crystal structure of *Parcu*-CDPS was solved (alongside a number of mutants) and the role of residues important for catalysis and histidine recognition were probed using mutagenesis. Three successive generations of mutants containing both single and double residue substitutions were generated leading to a change in substrate selectivity from histidine to phenylalanine and leucine. The research detailed herein is the first instance of successful engineering of a CDPS to yield different products, paving the way to direct the promiscuity of these enzymes to produce molecules of our choosing.

Introduction

Cyclodipeptide synthases (CDPSs) are an interesting family of enzymes with remarkable potential.[1] CDPSs use aminoacylated tRNA (aatRNA) substrates to form peptide bonds between two amino acids yielding a cyclic dipeptide product (CDP).[2] CDPs contain an integral 6-membered ring, allowing derivatisation at up to 6 positions leading to a plethora of molecules with structurally similar backbones.[3] These molecules possess privileged scaffolds and display remarkable properties such as proteolytic resistance, blood-brain barrier permeability and the ability to mimic functional pharmacophores.[4, 5] CDPSs can exist alone or as part of gene clusters containing tailoring enzymes, adding significant complexity to the types of natural products that can be produced.[6] Computational prediction of the specificity of CDPSs has had limited sucess[7], and the most reliable strategy for determination of substrates and products of each enzyme requires experimental testing, which has been performed for over 120 enzyme variants, in a time consuming and low throughput process.[8-10] Therefore, it is imperative that we better understand how CDPSs are selecting their substrates, as this would enhance our capacity to predict and modulate their activity to synthesise a wide variety of CDPs with valuable biological activities.

Given the extensive research into CDPSs, there are now more than 1000 predicted members of the family (Uniprot IPR038622). These enzymes are typically small and monomeric in nature but display variable degrees of sequence conservation.[11] Of the 120 empirically determined CDPS activities, all but one originate from 3 main phyla of bacteria: Actinobacteria; Firmicutes and Proteobacteria.[12, 13] Cyclodipeptide synthases are known to be promiscuous enzymes as most are capable of synthesising more than one cyclic dipeptide product.[14] Typically this includes one major product and one/several minor products usually containing a constant and a variable amino acid. This substrate selectivity is due to two separate solvent accessible pockets (P1 and P2) which interact with specific aminoacylated tRNA (aatRNA) substrates. It has been hypothesised that P1 has more stringent binding restrictions for the first amino acid compared to the larger and shallower P2 pocket.[15]

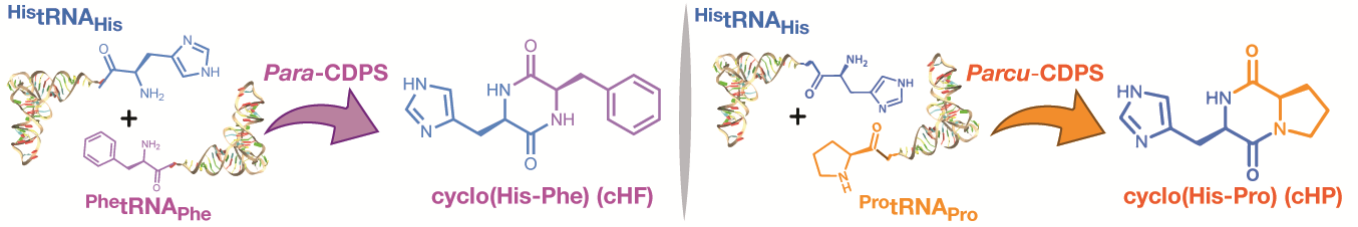
The catalytic mechanism of these enzymes was first elucidated by Sauguet *et al.* and since then details of the reaction they catalyse have been explored using experimental and computational methods.[16, 17] Briefly, the first aatRNA substrate binds and the amino acid moiety (situated in the P1 pocket) is covalently transferred to the catalytic serine forming an aminoacyl-enzyme intermediate. The second substrate then binds (aa moiety situated in the P2 pocket) and amino acid is transferred to the intermediate, generating a dipeptidyl-enzyme complex. This subsequently undergoes intramolecular cyclisation (facilitated by an essential tyrosine residue- Y202 in the case of AlbC) leading to the formation of a cyclic dipeptide.[11, 16, 18-20] For this reaction to proceed, there are four essential residues needed for substrate interaction and positioning: S37 and Y202 which are strictly conserved and Y178 and E182 which are highly conserved.[17]

Additionally, crystal structures from both CDPS families have been solved: AlbC;[16] RV2275;[19] YvmC[17], and BtCDPS from the NYH sub-family,[15] and Nbra-CDPS; Rgry-CDPS; and Fdum-CDPS from XYP.[21] These structures all share a common architecture featuring a Rossmann-fold domain. Structural similarity relate to the class of Ic aminoacyl tRNA synthetases (aaRSs), specifically TyrRS and TrpRS, suggesting CDPSs are

[a] *School of Biology, Biomedical Sciences Research Complex, University of St Andrews, St Andrews, Fife, KY16 9ST, UK.*

†

Electronic Supplementary Information (ESI) available: Supplementary information including materials and methods are available online. Coordinates of the proteins mentioned within have been deposited in the Protein Data Bank: *Parcu*-CDPS (7QB8); Y55F (7QAY); D58N (7QAU); E171Q (7QAX); E174A (7QAQ); E174L (7QAT) AND y189f (7QAW). See DOI: 10.1039/x0xx00000x



Scheme 1: Reactions catalysed by *Para*-CDPS and *Parcu*-CDPS.

derived from these Class-I aaRS.[16] However, CDPSs differ as they can function as monomers with the presence of a second pocket not found in the aaRSs and they lack an ATP-binding motif necessary for the function of aaRSs.[22] Within the two CDPS sub-groups, the structural differences are most evident in the first half of the protein i.e. the first pocket. Regardless of this, the conserved residues play parallel roles in positioning and stabilising the catalytic serine and tyrosine in the active site. Therefore, it has been hypothesised that all CDPSs will share a similar catalytic mechanism[21]

In this paper, we focus on two CDPS enzymes which synthesise a range of histidine-containing cyclic dipeptides (Scheme 1). *Para*-CDPS from *Parabacteroides sp. 20_3* (GenBank: EFK64745.1) produces cyclo(His-Phe) whilst *Parcu*-CDPS from *Parcubacteria bacterium RAAC4_OD1_1* (GenBank: ETB63777.1) can synthesise both cyclo(His-Glu) and cyclo(His-Pro) (Figure 1A). These CDPSs are the only members of this family known to use histidinyl-tRNA as a substrate to generate bioactive molecules. For example, cyclo(His-Phe) features the simple backbone structure of the anti-tumour compound plinabulin, the reduced form of phenylahistin – a natural product isolated from *Aspergillus ustus.*[23-25] On the other hand, cyclo(His-Pro) is endogenous to the human body as a by-product of the thyrotropin-releasing hormone.[26, 27] It has been proposed as a neuroprotective peptide against Parkinson disease and amyotrophic lateral sclerosis,[28, 29] as well as a molecule involved in the gut-brain-axis crosstalk,[30] with effects on glucose metabolism.[31] Thus, it would be advantageous to expand the chemical space of analogues we could produce by using enzymes such as CDPSs.

Here we describe a facile strategy for aminoacylated tRNA production, as these are substrates for CDPSs, with superior yield and decreased cost/labour. Our work reveals that both CDPS enzymes can use non-canonical amino acids as substrates to generate a diverse library of potentially bioactive compounds. We then used activated small molecules as substates, more specifically a dinitrobenzyl ester coupled to an amino acid, showing that these are substrates as well as chemical probes to determine if substrates are occupying pockets P1 or P2. To investigate histidine recognition by Parcu-CDPS, we solved the crystal structure of this protein and performed mutagenesis on key residues participating in interactions with substrates bound to P1. We rationally engineered P1 to become more hydrophobic and deeper, steering the substrate specificity away from histidine and towards more hydrophobic amino acids. These mutants displayed a remarkable shift in substrate specificity from the previously accepted histidine to two new substrates – leucine and phenylalanine. Therefore, our CDPS variants highlighted residues in P1 which are key for the recognition of histidine, unveiling important characteristics of how CDPSs select polar substrates. This has wide implications in our capacity to predict function as well as engineer CDPS enzymes to produce molecules of our choosing.

Results and discussion

Investigation of different tRNA sources

It has been well documented that CDPS enzymes use aminoacylated tRNA molecules as substrates for cyclic dipeptide production. They hijack the already aminoacylated tRNA within cells for use in cyclodipeptide synthesis. Therefore, it is vital to have methods of producing high quality tRNA for the *in vitro* reactions where we supply purified components to yield a CDP. Purified tRNA was previously synthesised by us and others using a time consuming *in vitro* transcription reaction which includes an initial PCR of the desired tRNA sequence to amplify the DNA template encoding the desired tRNA sequence, followed by *in vitro* transcription using a mutant T7 RNA polymerase (Δ172-173) to ensure homogeneous 3'-end in the tRNA, finishing with a phenol-chloroform extraction to yield purified tRNA.[32] This method, whilst reliable, has some caveats including cost, the duration of the experiment which spans three days and the rather extensive list of materials required to obtain tRNA. Searching for an easier alternative to this method, a procedure by Mechulam *et al.* was adapted to produce a 'pool' of tRNA naturally found in *E. coli.*[33] Using this protocol, a highly concentrated stock of all the tRNAs required for our experiments was purified and used with any CDPS/aaRSs combination of choice. More specifically, *Para*-CDPS and *Parcu*-CDPS could use this tRNA pool to generate their respective products – cHF and cHP – in relatively high yield (Figure 1B). This method is significantly cheaper, easier, and faster than previously described methods and has become a staple in our research with CDPSs.

Aiming to bypass the individual purification of the amino acid tRNA synthetases, we tested a bacterial lysate (S30 extract) isolated from *E. coli* containing amino acids, tRNA and amino acid tRNA synthetases.[34] Although easy to produce as no additional proteins/tRNAs need to be separately generated, this S30 extract was not as efficient as the previously described tRNA pool (Figure 1B). The extract gave a poor signal/noise baseline in LC-MS assays and quantification using a standard curve revealed low reaction yields. Furthermore, any use of non-canonical amino acids with the S30 was unsuccessful due to the high levels of endogenous

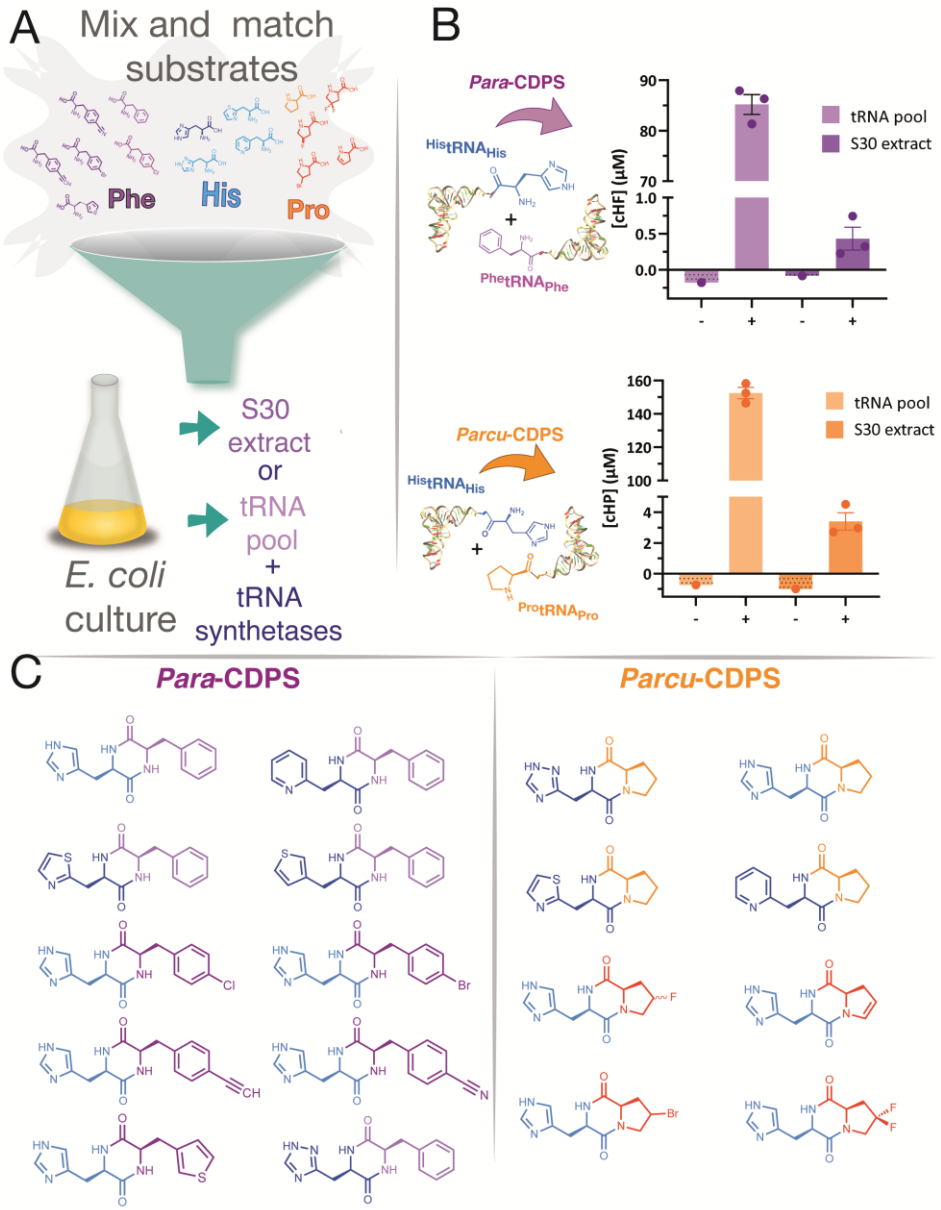


Figure 1. Cyclic dipeptide production using *Para*-CDPS and *Parcu*-CDPS. (A) Cartoon schematic depicting possible substrates and products from each CDPS studied herein. The tRNA structure of the aminoacylated tRNA was taken from the PDB (1EHZ). The distinctive 6-membered CDP ring is shown in green, and the common histidine is shown in blue. (B) Quantification of CDP production using a standard calibration curve. The presence (+) and absence (-) of CDPS is displayed for each method of tRNA; tRNA pool and S30 Extract. It is evident that the tRNA pool is a far superior method of producing CDP for both CDPSs. (C) Generation of a cyclic dipeptide library using *Para*-CDPS and *Parcu*-CDPS using non-canonical amino acids: histidine derivatives are shown in dark blue; phenylalanine derivatives in dark purple and proline derivatives in red.

aminoacylated tRNA present. We added a deacylation step to the S30 production, but that decreased overall yield and was not further pursued. Overall, the tRNA pool was a far superior alternative to produce high quality tRNA for use and therefore was employed in subsequent reactions.

Incorporation of non-canonical amino acids into cyclic dipeptides

Hartman *et al.* previously determined the ability of aaRSs to accept non-canonical amino acids and literature has shown that CDPSs can incorporate these into the ring.[35, 36] Consequently we aimed to explore the capability of our two active enzymes in accepting other amino acids as substrates for cyclic dipeptide production. Using a range of commercially available non-canonical analogues with the previously described tRNA pool, a library of diverse CDPs was produced from just two CDPSs (Figure 1C). *Para*-CDPS and *Parcu*-CDPS both accepted the same histidine analogues - H-β-(2-Thiazolyl)-alanine and 3-(2-pyridyl)-L-alanine – with *Parcu*-CDPS also able to use β-(1,2,4-Triazol-3-yl)-DL-alanine. It has been demonstrated that these amino acid analogues are accepted by HisRS[35] so the only factor influencing product synthesis was the CDPS. Prior to our work, no information was available about histidine recognition by these enzymes, nevertheless the unnatural substrates highlight that the nitrogen on position three of the imidazole ring is important for substrate recognition. The incorporation of the pyridyl ring indicates that this enzyme can accept larger ring structures however only one isomer (2,3) was found to be introduced into a CDP. Moreover, the rejection of isomers 3-(3-) and 3-(4-pyridyl)-L-alanine supports the argument that a nitrogen may be required in close proximity to the alpha carbon of the amino acid.

In addition, we used PheRS A294G which has a wider binding pocket allowing interactions with a larger range of related Phe analogues.[37] All but two of the known non-canonical amino acids accepted by this synthetase were incorporated into the ring by *Para*-CDPS. The para-substituted phenyl ring is easily accepted by the CDPS due to the additional functional group being positioned on the outskirts of the pocket reducing potential steric clashes with pocket residues. On the other hand, several halogenated proline analogues were reported as accepted substrates for ProRS and were seen as part of the CDPs produced by *Parcu*-CDPS. 4-Bromo-proline was not previously shown to be a substrate for PheRS-A294G but it was hypothesised to display similar chemistry as 4-fluoro-proline.[38] This CDPS can tolerate conservative derivatisations of proline however functional groups including hydroxyl and amines on the ring were not accepted. We hypothesise this may be due to their positioning in the pocket causing steric clashes and forcing the amino acid into unfavourable conformations for cyclisation.

Nevertheless, the production of a diverse library from just two CDPSs highlights the comprehensive scope these enzymes possess to accept both canonical and non-canonical amino acids as substrates. Importantly, most of the synthesised molecules have a functional group which could lead to further applications. For example,

halogenated amino acids are valuable tools as NMR spectroscopic probes whilst certain groups such as nitriles and terminal alkynes can participate in click chemistry reactions with fluorophores for use as fluorescent probe as well as other conjugation reactions. This list is by no means exhaustive of the potential applications for these molecules and the opportunities provided by non-canonical amino acids has been recently reviewed.[38] However, a range of non-canonical amino acids that are known substrates of aaRSs were not accepted which highlights that the CDPS enzyme is selective in which substrates it will accept regardless of aminoacylation. Having explored the substrate landscape using commercially available amino acids, our focus turned to a reduced form composed of a dinitrobenzyl ester (DBE) coupled to an amino acid known to be accepted by CDPSs. Previous research from our group highlighted the use of these molecules to produce CDPs however that particular enzyme – BtCDPS – only required one aa-DBE to give a product.[15] Herein we explore the use of a combination of both DBE and tRNA to yield cyclic products.

CDP formation using a minimal substrate

We synthesised three activated amino esters harbouring dinitrobenzyl ester as a leaving group: His-DBE; Pro-DBE, and Phe-DBE (Synthesis detailed in ESI). Considering that the CDPSs require two DBE substrates to produce a final product, it was hypothesised that this reaction would be highly unfavourable and yield very little CDP, due to the fact that both first and second half reactions are likely to be impaired by the use of a minimal substrate. This hypothesis holds true and to maximise the efficiency of the reaction we tested the use of one aa-DBE and one aa-tRNA with the enzymes in question (Figure 2A). Interestingly, more product was formed using a combination of the aa-tRNA/aa-DBE substrates however the overall yield was lower in comparison to the natural tRNA substrates (Figure 2B). The yield of product varies proportionally with the concentration of enzyme used however using DBE substrates overall slows down the enzymatic turnover which consequently affects CDP synthesis. aa-DBE has a limited half-life, which also impacts reaction yields.[15]

Mass spectrometry of BtCDPS indicated that the enzyme could form the first acyl-enzyme intermediate when a DBE substrate was used.[15] Docking simulations suggested the decreased product formation when aa-DBE is a substrate occurs due to the second substrate binding and unproductive positioning in P2, lacking essential residue interactions for the reaction to proceed. Initially we assumed *Para*-CDPS and *Parcu*-CDPS both accepted histidine in P1 as they are the only members of the CDPS family to display affinity for this amino acid. However, *Para*-CDPS is able to make more product using $^{His}tRNA_{His}$ + Phe-DBE, in disagreement to what would be expected if histidine was binding on P1. Consequently, we investigated both enzymes using a trapped acyl-enzyme intermediate experiment developed in-house, which exploits the minimal aa-DBE substrate as a chemical probe to verify substrate binding order. Here we saw that *Para*-CDPS accepts Phe in P1 instead of His whereas *Parcu*-CDPS accepts His in P1 (Figure 2C). This is agreement with our original hypothesis that using a DBE substrate in P1 is more favourable for CDPSs and allows for a higher yield of product by restricting the aa-DBE substrate role to the first half reaction (acylation), ultimately leading to more productive turnover cycles. It is likely that the DBE substrate occupying P2 can sample different conformations, most of which are unproductive and given the instability of this substrate hydrolysis occurs before the CDPS is able to turnover.

Interested in the formation of the acyl-enzyme intermediate and the rate limiting nature of the cyclization reaction, we then performed a discontinuous time course assay monitoring the rate of cyclic dipeptide formation from initial (0 minutes) to the theorised endpoint (overnight). The experiment was carried out for both *Para*-CDPS and *Parcu*-CDPS using the previously described possible

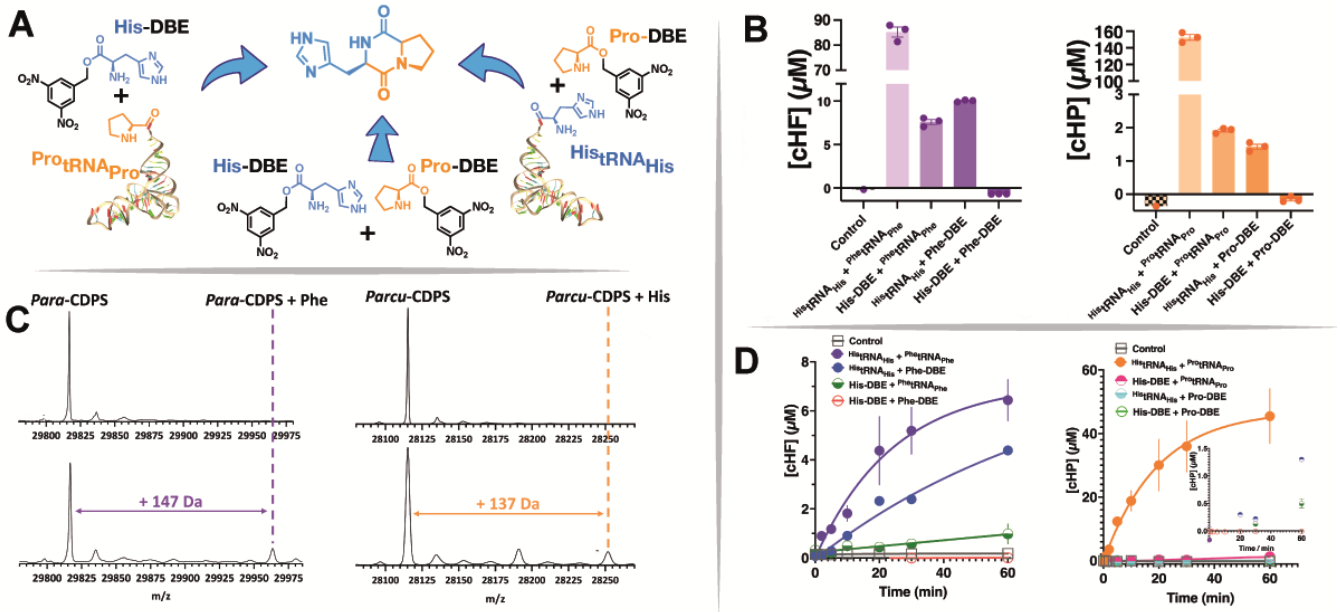


Figure 2. Use of minimal substrates to yield CDPs. (A) Reaction scheme highlighting the 3 possible combinations when using DBE substrates in conjunction with aatRNA. (B) Quantification of product yield from aa-DBE and aatRNA reactions with each CDPS. The use of two aatRNA substrates continues to give the highest concentration of product. (C) Intact protein mass spectrometry of trapped acyl-enzyme intermediates for *Para*-CDPS and *Parcu*-CDPS. Contrary to our original hypothesis, *Para*-CDPS binds phenylalanine in P1 as shown by the mass relating to the *Para*-CDPS+Phe. *Parcu*-CDPS however does bind histidine in the first pocket. (D) Time course assay for product formation of cHP and cHF highlighting the rate-limiting capability of the enzyme acylation step. In the first hour of reaction *Para*-CDPS catalyses the reaction with similar rates when either aatRNA or aa-DBE bind to P1, whereas *Parcu*-CDPS generates very little product when histidine binds to P1 regardless of which co-substrate is provided.

combinations of substrates (Figure 2D). Our data show that *Para*-CDPS displayed a small difference in the rate of product formed in the first hour of reaction when using $^{Phe}tRNA_{Phe}$ or Phe-DBE as substrate, indicating that either the rate of formation for the *Para*-CDPS-Phe acyl-enzyme intermediate is similar for both substrates or that they possess similar rate limiting steps for the first half reaction regardless which substrate is used. This points towards a less significant role of aatRNA in substrate positioning for the first half reaction, in contrast to what is observed in *Parcu*-CDPS. Furthermore, the overnight reaction yield and reaction rate when aa-DBE binds on P2 supports the hypothesis that aa-DBE compounds in P2 are less productively positioned, which combined with the instability of aa-DBE substrates results in a lower yield of product overall. The slowest step of the reaction catalysed by *Parcu*-CDPS is likely to be in the second half reaction, after substrate is productively bound to P1, and the first acyl-enzyme intermediate is formed. This is because in the reaction catalysed by *Parcu*-CDPS having a single DBE substrate in either pocket significantly reduces the yield of cHP. Following these experiments we focused on understanding Histidine selection, and more specifically on P1, as it was predicted to possess a narrower binding pocket. To do this, we solved the crystal structure of *Parcu*-CDPS and explored the residues determining substrate selection.

Structure of wild type *Parcu*-CDPS

Parcu-CDPS belongs to the XYP sub-group of the CDPS family, characterised by the presence of 3 residues: X40 where X is a non-conserved residue, Y202, and P203 (numbering respective to AlbC).[14] Structures of three previous members from XYP have been solved by Bourgeois *et al.* – *Rgry*-CDPS; *Fdum*-CDPS and *Nbra*-CDPS.[21] However, these enzymes all use relatively hydrophobic and non-polar amino acids such as glycine, alanine and leucine. By solving the structure of *Parcu*-CDPS, we aimed to uncover unique characteristics allowing histidine to act as a substrate. The crystal structure of *Parcu*-CDPS was solved at a resolution of 1.90 Å where 9 residues out of the total 230 are not traceable (Figure 3A). This small loop (residues 61 - 69) lacked sufficient electron density for a model to be built with certainty, likely due to high flexibility. It is important to note the structure was solved by Iodide SAD phasing after extensive trials of molecular replacement failed, suggesting significant deviation from previously determined CDPS structures.

At a first glance, *Parcu*-CDPS displays the stereotypical Rossman-fold known to be common throughout the CDPS family. The active site includes the four conserved residues previously identified: S26, Y167, E171 and Y191 (Figure 3A). The identity of the residue known as X40 is an asparagine which is found in 25% of the recognised XYP CDPSs.[21] D58 was previously not identified as an active site residue, but our modelled structure shows it is in hydrogen bond distance from the catalytic serine (S26), and this interaction could be important for S26 to act as a nucleophile. A Ser/Asp dyad was observed in Phospholipase A2,[39] and future work could be directed towards better understanding the catalytic mechanism of *Parcu*-CDPS, and whether similarities are present. Moreover, the side chain of the predicted catalytic important Tyr167 points away from the active site, removing the conventional hydrogen bonding network seen in other CDPSs. Y167 plays the same role as Y178 in AlbC which is hypothesised to stabilise the aminoacyl moiety formed from the binding of the first substrate to P1. Upon further comparison with other CDPSs, RMSD values ranging from 2.764 Å over 174 aligned residues for BtCDPS (PDB: 6ZTU) to 2.984 Å over 171 aligned residues for AlbC (PDB: 4Q24) and 2.764 Å 2.901 Å over 177 residues for *Nbra*-CDPS (PDB: 5MLQ) were found. Structural comparison of *Parcu*-CDPS with these three CDPSs highlighted a divergence in secondary structure. CDPS enzymes typically share a highly conserved secondary structure, upon which appendages exist in individually solved structures. Figure 3B depicts the core fold in transparent colour and the divergent regions shown in solid colour. *Parcu*-CDPS structure diverges from the common fold in helix α-3 and beta-strand β3, where these regions change direction. In the common CDPS fold helix α-3 and α-4 exist as a single continuous helix whereas, in *Parcu*-CDPS a glycine residue (G84) provides a significant bend and change of direction, splitting the helix into two. The direction of β3 is another important deviation seen in *Parcu*-CDPS' structure, as it changes direction (compared to common CDPS fold) at I56 to divide the active site pocket. This directional change of β3 facilitates the placement of D58 into H-bonding distance of the active site S26. These two motifs differentiate *Parcu*-CDPS from the previously solved CDPS structures which are fairly conserved with respect to each other.

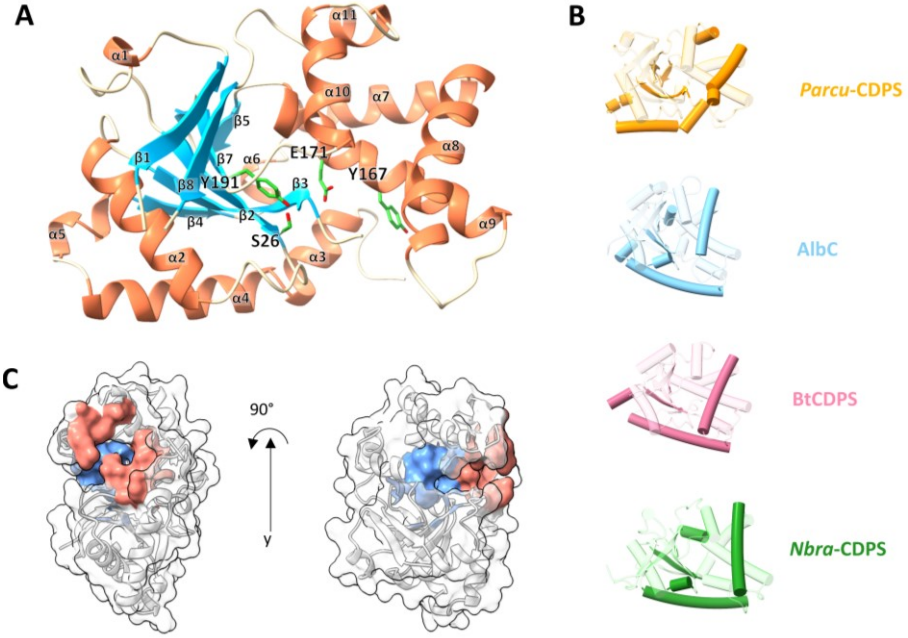


Figure 3. Structure of *Parcu*-CDPS. (A) Full structure of *Parcu*-CDPS depicting the alpha helices (orange), beta-sheets (blue) and active site residues (green). (B) Secondary structure comparison of *Parcu*-CDPS to three other CDPSs: AlbC; BtCDPS and *Nbra*-CDPS. The common CDPS core is shown as transparent colour whilst the major differences are 100% opaque. (C) Pocket volumes of both P1 (blue) and P2 (red) in *Parcu*-CDPS are displayed here as calculated by CASTp.[40]

CASTp[40] was used to investigate the pocket volume of *Parcu*-CDPS and highlighted pocket residues which are in contrast to the predicted pocket residues from Gondry *et al.*[12] This finding emphasises the importance of characterising individual members of the CDPS family which can display divergences

from the well-known archetype AlbC. P1 (shown in blue in Figure 3C) is found deeper within the enzyme and is smaller and more restricted in size than P2 which sits at the solvent-accessible edge of the CDPS (red on Figure 3C). The large size of P2 further explains poor positioning of Pro-DBE as the substrate can potentially sample several conformations, most of which are likely unproductive. We used PROPKA[41] to calculate theoretical pKa values for residues in our model at pH 7.0 and generate electrostatic potential maps for each protein variant (calculated data found in Table S4, ESI). From these calculations, P1 is predicted to be mostly neutral apart from Tyr55 and Glu174 which form a negatively charged microenvironment, while P2 is predicted to be mostly lined by positively charged residues. The more positively charged section of P2 could be facilitating the production of cHE – which is also a product of *Parcu*-CDPS (Figure S5, ESI). Glutamate is a large flexible amino acid, likely negatively charged at reaction pH.[42] Thus participation in electrostatic interactions of the $^{Glu}tRNA_{Glu}$ substrate with positively charged residues in P2 is plausible, while specific interaction with $^{Pro}tRNA_{Pro}$ are less obvious.

Rationally altering substrate specificity of *Parcu*-CDPS

Previous research by us and others has shown that mutations in active site residues seriously reduces product formation in CDPS enzymes. However, attempts at changing the substrate scope of a CDPS by mutating select residues has been unsuccessful until now.[10] Having solved the structure of *Parcu*-CDPS, we aimed to investigate which residues were crucial for histidine recognition by using site-directed mutagenesis. We first designed P1 mutants referred to as Generation 1 (seven variants: Y55F; Y55V; E174A; E174H; E174L; Y189F and Y189L), and these variants were cloned, expressed in *E.coli*, purified and used for activity assays. The presence of the mutation was confirmed using intact protein mass spectrometry (Figure S8) and the enzymatic activity was confirmed using LC-MS as previously described. The active site mutants – S26A; S26C; D58A; D58N; Y167A; Y167F; E171A; and E171Q - were included to confirm the loss of activity upon removing known catalytic residues. (Figure 4A).

Initially, the mutants were tested for cHP production using the same assay performed on the WT. Figure 4B shows that only S26C and Y167F from the active site variants could still produce cHP albeit with a lower yield. This result is akin to the trend in activity published by Bourgeois *et al.* who also mutated the catalytic Tyr in three different XYP CDPSs which still produced their respective products.[21] This demonstrates that the phenyl ring is vital in substrate binding rather than the hydrogen bonding interactions from the phenolic hydroxyl.[19, 20] When the H-bond between S26 and D58 is disrupted by mutating D58 to either an alanine or an asparagine, the enzyme is unable to produce any cyclic dipeptide. This highlights the essential role that D58 has in potentially polarizing and positioning the serine in the active site for substrate binding and acyl enzyme formation. By mutating the P1 residues, the variants were still capable of accepting histidine with only one mutant – E174H – displaying no activity. E174H was designed to reverse the charge of the residue and was predicted to repel the incoming histidine from P1. Further investigation into the changes imposed by these mutants was performed by solving the crystal structures of a select few (Figure S6, ESI). The mutant structures confirmed that protein remained intact and properly folded resembling a stereotypical CDPS. Therefore, the disruption of activity is caused exclusively by the change in these few residues which appear essential for the enzyme to produce a CDP.

Following on from these results, we designed Generation 2 of double mutants from P1 residues (three variants: Y55F+Y189F, Y55F+E174A, Y189F+E174A,), aiming to severely perturb the activity of the enzyme towards histidine. When we investigated the formation of cHP using Generation 2 variants, all mutants synthesised significantly less cHP than the single Gen 1 mutants

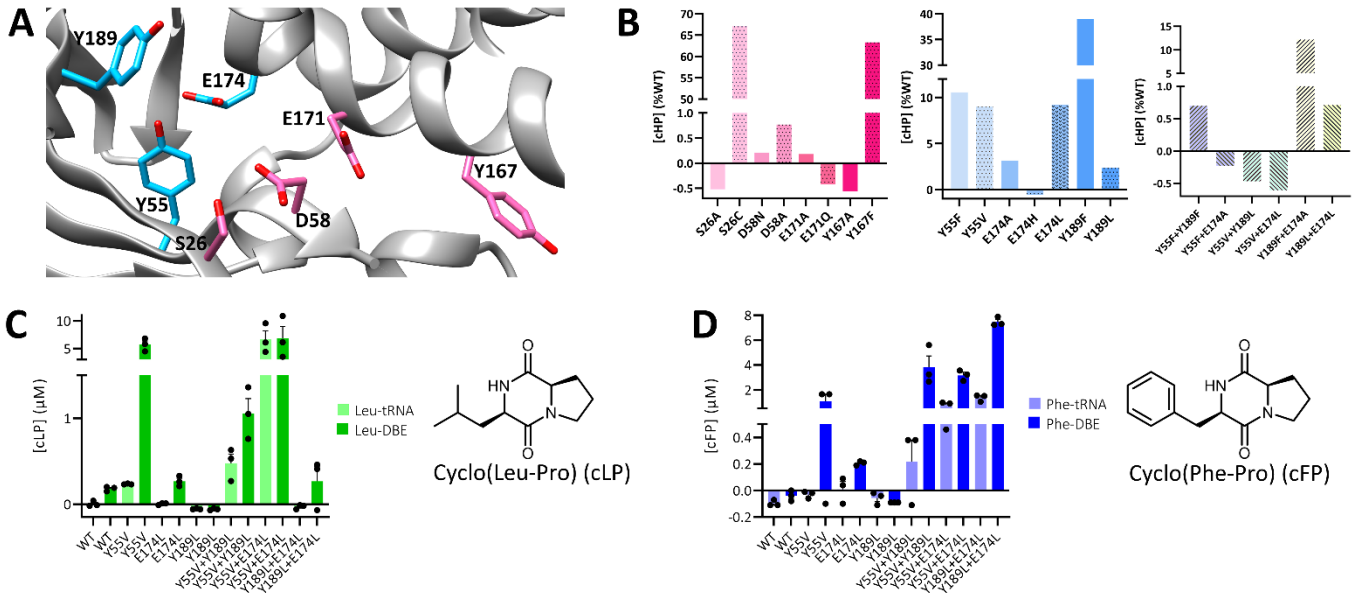


Figure 4. Site directed mutagenesis of *Parcu*-CDPS. (A) Enhanced image of the WT structure highlighting the two sets of residues targeted for mutagenesis – active site (pink) and pocket 1 (turquoise). (B) cHP activity assay for each set of mutants – the activity is shown as a percentage of the wild-type *Parcu*-CDPS activity. Each residue is shown in a different colour and the patterned bar represents a second mutation of the same residue. The overall trend shown is a decrease in the capability of the mutants to produce cHP with S26C displaying the highest yield. (C) Quantification of cyclo(Leu-Pro) production using *Parcu*-CDPS variants. Pro-tRNA was used in combination with Leu-tRNA and Leu-DBE to investigate the use of different substrates on product yield. (D) Quantification of cyclo(Phe-Pro) using the same mutants as tested for cLP.

alone (Figure 4B). This suggests that binding to the P1 pocket is not facilitated by a single residue alone, rather by a combination of interactions. The impact of mutating Y55 carries more weight than Y189, suggesting it may directly interact with the histidine residue (via a H-bond) rather than adding to the polar surface.

Having hindered the capacity of *Parcu*-CDPS to recognize histidine, we then focussed on Generation 3 composed of three double mutants (Y55V+E174L; Y55V+Y189L; E174L+Y189L), aiming to switch the substrate specificity to a less polar amino acid. This was because Generation 1 and 2 variants essentially incorporate hydrophobic and non-polar amino acids potentially altering the overall environment and electrostatics of P1. Therefore, we hypothesised that whilst these enzymes were uncapable of cHP production, they would tolerate a different, less charged substrate in P1. Previously we showed that CDPSs can use both tRNA and DBE substrates to yield a CDP and so using a combination of these, the production of cLP and cFP by *Parcu*-CDPS mutants was investigated (Figure 4C/D). Interestingly, reactions using $^{Pro}tRNA_{Pro}$ and Leu-DBE/Phe-DBE gave a higher yield of product compared to $^{Pro}tRNA_{Pro}$ and $^{Leu}tRNA_{Leu}$/$^{Phe}tRNA_{Phe}$. This could indicate that the enzyme is still able to recognise and reject the tRNA body, which can be overcome by using a smaller DBE substrate. It is evident that whilst wild-type *Parcu*-CDPS is unable to use either leucine or phenylalanine as substrates, the new mutants can accept these amino acids. These mutants, however, did not accept other small hydrophobic amino acids such as valine and isoleucine thus demonstrating that the enzyme is still actively selecting its substrates. Overall, this is the first example of a CDPS displaying a change in substrate specificity using targeted enzyme engineering. The new products – cLP and cFP – produced by the *Parcu*-CDPS variants are biologically relevant molecules with known applications as anti-cancer drugs. Jinendiran *et al*. reported cell death of colorectal cancer cells (HT-29) in zebrafish xenograft model after dosing with either CDP.[43] This finding showcases the advantages of mutating a CDPS to produce interesting molecules with untapped potential.

Conclusions

We set out to investigate the cyclodipeptide-synthesising capability of two cyclodipeptide synthases which both accept histidine as a substrate. Our work uncovered that the use of a collective tRNA pool was sufficient for the CDPSs to yield their expected product in addition to accepting a variety of unnatural amino acids as substrates for CDP formation. This new method can easily be scaled up and has simplified the process of synthesising tRNA for multiple reactions, proving useful for both canonical and non-canonical amino acids. Thus, our method to produce a diverse library of molecules is a significant improvement allowing more compounds with biologically relevant capabilities to be produced in the future.

Additionally, structural characterisation of *Parcu*-CDPS revealed the interesting pocket topology of one out of the only two enzymes which accept histidine as a substrate. The acyl-enzyme intermediate was trapped using a minimal substrate composed of a dinitrobenzyl ester coupled to an amino acid. This confirmed that histidine was bound in P1 of *Parcu*-CDPS but on P2 for *Para*-CDPS, highlighting the difference between the two enzymes. Future work should focus on solving the structure for *Para*-CDPS to further uncover substrate selection on P2.

Finally, combining structural biology, enzymatic assays, and rational engineering, we provide a much clearer picture of how polar residues such as histidine are selected on P1, as well as how this selectivity can be manipulated. pKa calculations using experimentally determined structures reveal that several residues in proximity are likely altering electrostatics of the binding pocket and therefore influencing substrate selection. Although product yield by a CDPS protein has been improved engineering P1,[44] we are unaware of successful attempts to alter the accepted substrates of a CDPS enzyme. Therefore, this is a pivotal finding which could lead to a wide array of CDPs from a single CDPS and its engineered mutants.

Author Contributions

ES performed experiments, interpreted data, and wrote the manuscript; CJH and CMC interpreted and discussed data and revised the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank the BSRC Mass Spectrometry and Proteomics Facility, University of St. Andrews for collecting data on intact protein mass. We also thank Alison Dickson for valuable discussions about mass spectrometry and routine upkeep of the instrument.

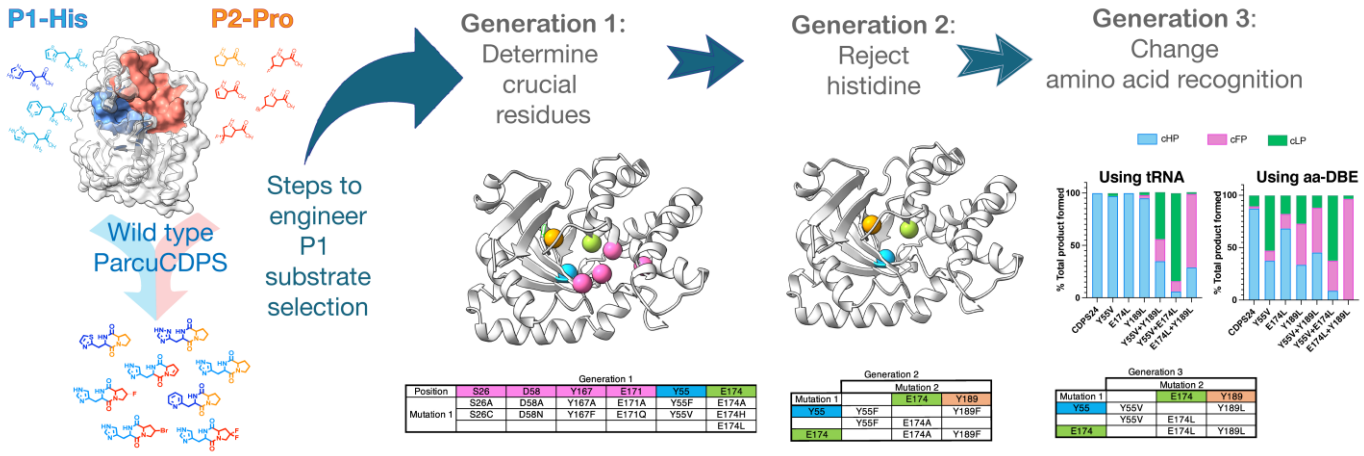


Figure 5: Swapping the substrate selection on P2 by *Parcu*-CDPS. Based on P1 residues we produced a series of rationally designed mutants to 1) determine residues crucial for histidine recognition; 2) reject histidine as a substrate and 3) select a different amino acid on P1 to produce cyclic dipeptides that no longer contain histidine.

References

1. S. Lautru, M. Gondry, R. Genet and J.-L. Pernodet, *Chem. Biol.*, 2002, **9**, 1355-1364.
2. M. Gondry, L. Sauguet, P. Belin, R. Thai, R. Amouroux, C. Tellier, K. Tuphile, M. Jacquet, S. Braud, M. Courçon, C. Masson, S. Dubois, S. Lautru, A. Lecoq, S.-i. Hashimoto, R. Genet and J.-L. Pernodet, *Nat. Chem. Bio.*, 2009, **5**, 414-420.
3. A. D. Borthwick, *Chem. Rev.*, 2012, **112**, 3641-3716.
4. J. F. González, I. Ortín, E. de la Cuesta and J. C. Menéndez, *Chem. Soc. Rev.*, 2012, **41**, 6902-6915.
5. M. B. Martins and I. Carvalho, *Tetrahedron*, 2007, **63**, 9923-9932.
6. P. Borgman, R. D. Lopez and A. L. Lane, *Org Biomol Chem*, 2019, **17**, 2305-2314.
7. M. A. Skinnider, C. W. Johnston, N. J. Merwin, C. A. Dejong and N. A. Magarvey, *BMC Genomics*, 2018, **19**, 45.
8. I. B. Jacques, M. Moutiez, J. Witwinowski, E. Darbon, C. Martel, J. Seguin, E. Favry, R. Thai, A. Lecoq, S. Dubois, J. L. Pernodet, M. Gondry and P. Belin, *Nat Chem Biol*, 2015, **11**, 721-727.
9. M. Gondry, I. B. Jacques, R. Thai, M. Babin, N. Canu, J. Seguin, P. Belin, J. L. Pernodet and M. Moutiez, *Front Microbiol*, 2018, **9**, 46.
10. N. Canu, M. Moutiez, P. Belin and M. Gondry, *Nat Prod Rep*, 2020, **37**, 312-321.
11. N. Canu, M. Moutiez, P. Belin and M. Gondry, *Nat. Prod. Rep.*, 2019, DOI: 10.1039/C9NP00036D.
12. M. Gondry, I. B. Jacques, R. Thai, M. Babin, N. Canu, J. Seguin, P. Belin, J.-L. Pernodet and M. Moutiez, *Front. Microbiol.*, 2018, **9**.
13. M. Moutiez, P. Belin and M. Gondry, *Chem. Rev.*, 2017, **117**, 5578-5618.
14. I. B. Jacques, M. Moutiez, J. Witwinowski, E. Darbon, C. Martel, J. Seguin, E. Favry, R. Thai, A. Lecoq, S. Dubois, J.-L. Pernodet, M. Gondry and P. Belin, *Nat. Chem. Bio.*, 2015, **11**, 721.
15. C. J. Harding, E. Sutherland, J. G. Hanna, D. R. Houston and C. M. Czekster, *RSC Chem. Bio.*, 2021, **2**, 230-240.
16. L. Sauguet, M. Moutiez, Y. Li, P. Belin, J. Seguin, M.-H. Le Du, R. Thai, C. Masson, M. Fonvielle, J.-L. Pernodet, J.-B. Charbonnier and M. Gondry, *Nucleic Acids Res.*, 2011, **39**, 4475-4489.
17. E. Schmitt, G. Bourgeois, M. Gondry and A. Aleksandrov, *Sci. Rep.*, 2018, **8**, 7031.
18. L. Bonnefond, T. Arai, Y. Sakaguchi, T. Suzuki, R. Ishitani and O. Nureki, *PNAS*, 2011, **108**, 3912-3917.
19. M. W. Vetting, S. S. Hegde and J. S. Blanchard, *Nat. Chem. Bio.*, 2010, **6**, 797-799.
20. M. Moutiez, E. Schmitt, J. Seguin, R. Thai, E. Favry, P. Belin, Y. Mechulam and M. Gondry, *Nat. Commun.*, 2014, **5**, 5141.
21. G. Bourgeois, J. Seguin, M. Babin, P. Belin, M. Moutiez, Y. Mechulam, M. Gondry and E. Schmitt, *J. Struct. Biol.*, 2018, **203**, 17-26.
22. L. Aravind, R. F. de Souza and L. M. Iyer, *Biol. Direct*, 2010, **5**, 48.
23. K. Kanoh, S. Kohno, T. Asari, T. Harada, J. Katada, M. Muramatsu, H. Kawashima, H. Sekiya and I. Uno, *Bioorg. Med. Chem. Lett.*, 1997, **7**, 2847-2852.
24. B. Nicholson, G. K. Lloyd, B. R. Miller, M. A. Palladino, Y. Kiso, Y. Hayashi and S. T. C. Neuteboom, *Anti-Cancer Drugs*, 2006, **17**, 25-31.
25. P. J. Cimino, L. Huang, L. Du, Y. Wu, J. Bishop, J. Dalsing-Hernandez, K. Kotlarczyk, P. Gonzales, J. Carew, S. Nawrocki, M. A. Jordan, L. Wilson, G. K. Lloyd and H.-G. Wirsching, *Biomed Rep*, 2019, **10**, 218-224.
26. S. F. Grottelli, I.; Pietrini, G.; Peirce, M.J.; Minelli, A.; Bellezza, I., *Int. J. Mol. Sci.*, 2016, **17**, 1332.
27. C. Prasad, *Neurosci. Biobehav. Rev.*, 1988, **12**, 19-22.
28. I. Bellezza, M. J. Peirce and A. Minelli, *Trends Mol. Med.*, 2014, **20**, 551-558.
29. S. Grottelli, L. Mezzasoma, P. Scarpelli, I. Cacciatore, B. Cellini and I. Bellezza, *Mol. Cell. Neurosci.*, 2019, **94**, 23-31.
30. I. Bellezza, M. J. Peirce and A. Minelli, in *Quorum Sens.*, ed. G. Tommonaro, Academic Press, 2019, ch. 10, pp. 257-286.
31. M. K. Song, D. S. Bischoff, A. M. Song, K. Uyemura and D. T. Yamaguchi, *BBA Clinical*, 2017, **7**, 41-54.
32. B. M. Bertand Beckert, *RNA: Methods in molecular biology*, 2011, **703**, 29-41.
33. Y. Mechulam, L. Guillon, L. Yatime, S. Blanquet and E. Schmitt, in *Methods in Enzymology*, ed. J. Lorsch, Academic Press, 2007, vol. 430, pp. 265-281.
34. N. Krinsky, M. Kaduri, J. Shainsky-Roitman, M. Goldfeder, E. Ivanir, I. Benhar, Y. Shoham and A. Schroeder, *PLOS ONE*, 2016, **11**, e0165137.
35. M. C. T. Hartman, K. Josephson and J. W. Szostak, *Proc Natl Acad Sci U S A*, 2006, **103**, 4356-4361.
36. N. Canu, P. Belin, R. Thai, I. Correia, O. Lequin, J. Seguin, M. Moutiez and M. Gondry, *Angewandte Chemie International Edition*, 2018, **57**, 3118-3122.
37. P. Kast and H. Hennecke, *Journal of Molecular Biology*, 1991, **222**, 99-124.
38. M. C. T. Hartman, *ChemBioChem*, 2021, **22**, 1-17.
39. A. Dessen, J. Tang, H. Schmidt, M. Stahl, J. D. Clark, J. Seehra and W. S. Somers, *Cell*, 1999, **97**, 349-360.
40. W. Tian, C. Chen, X. Lei, J. Zhao and J. Liang, *Nucleic Acids Research*, 2018, **46**, W363-W367.
41. E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G. W. Wei, M. J. Holst, J. A. McCammon and N. A. Baker, *Protein Sci*, 2018, **27**, 112-128.
42. T. K. Harris and G. J. Turner, *IUBMB Life*, 2002, **53**, 85-98.
43. S. Jinendiran, W. Teng, H.-U. Dahms, W. Liu, V. K. Ponnusamy, C. C.-C. Chiu, B. S. D. Kumar and N. Sivakumar, *Scientific Reports*, 2020, **10**, 13721.
44. K. Brockmeyer and S. M. Li, *J Nat Prod*, 2017, **80**, 2917-2922.