Active site remodelling of a cyclodipeptide synthase redefines substrate scope

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

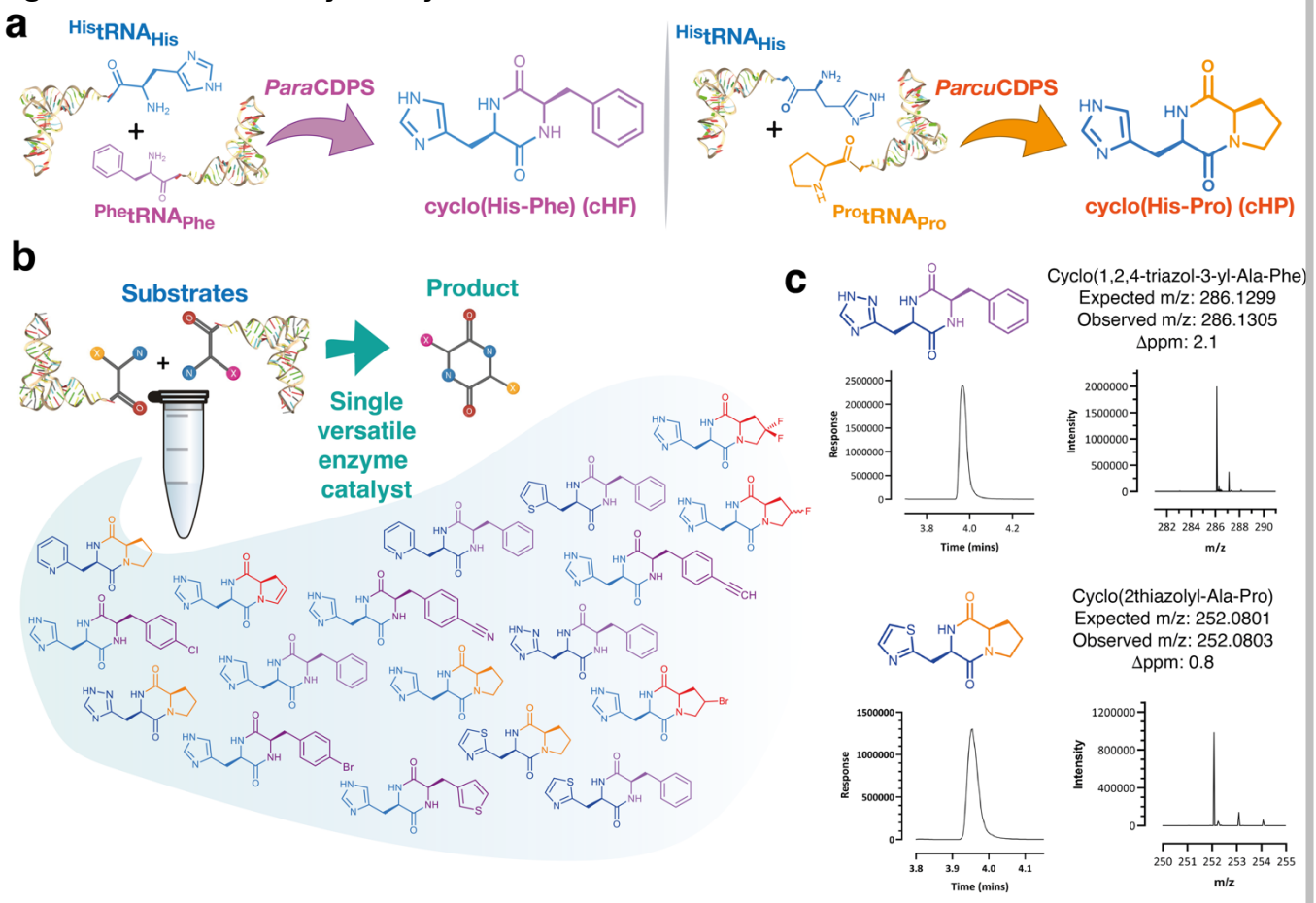
Cyclodipeptide synthases (CDPSs) generate a wide range of cyclic dipeptides using aminoacylated tRNAs as substrates. Histidine-containing cyclic dipeptides have important biological activities as anticancer and neuroprotective molecules. Out of the 120 experimentally validated CDPS members, only two are known to accept histidine as a substrate yielding cyclo(His-Phe) and cyclo(His-Pro) as products. It is not fully understood how CDPSs select their substrates, and we must rely on bioprospecting to find new enzymes and novel bioactive cyclic dipeptides. Here, we generated an extensive library of molecules using canonical and non-canonical amino acids as substrates, expanding the chemical space of histidine-containing cyclic dipeptide analogues. To investigate substrate selection we determined the structure of a cyclo(His-Pro)-producing CDPS. Three consecutive generations harbouring single, double and triple residue substitutions elucidated the histidine selection mechanism. Moreover, substrate selection was redefined, yielding enzyme variants that became capable of utilising phenylalanine and leucine. Our work pioneers the successful engineering 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) use aminoacylated tRNA (aatRNA) substrates to form peptide bonds between two amino acids yielding a cyclic dipeptide product (CDP).[1] CDPs contain a diketopiperazine ring, a scaffold that has been coined as privileged due to its remarkable properties such as proteolytic resistance, blood-brain barrier permeability and the ability to mimic functional pharmacophores. [2-4] CDPSs act in combination with tailoring enzymes, adding significant complexity to the types of natural products that can be produced.[5,6] Because computational prediction of the specificity of CDPSs is challenging[7], the determination of substrates and products of each enzyme requires experimental testing, in a time consuming and low throughput process.[8-10] Prior to this work there were no reports of successfully engineering CDPS enzymes to direct substrate selection.

Here, we focus on the only two known CDPS enzymes which synthesise bioactive histidine-containing cyclic dipeptides (Scheme 1). [11-15] Cyclo(His-Phe) (cHF) encompasses the backbone structure of the anti-tumour compound plinabulin, while cyclo(His-Pro) (cHP) is endogenous to the human body and proposed as a neuroprotective peptide against Parkinson's disease and amyotrophic lateral sclerosis,[16,17] as well as a

Fig. 1: Reactions catalysed by *Para*CDPS and *Parcu*CDPS



a, Enzymes characterised here, both use aminoacylated tRNAs as substrates to produce cyclic dipeptide. *Para*CDPS from *Parabacteroides sp. 20_3* (GenBank: EFK64745.1) produces cyclo(His-Phe), and *Parcu*CDPS from *Parcubacteria bacterium RAAC4_OD1_1* (GenBank: ETB63777.1) can synthesise both cyclo(His-Glu) – not shown – and cyclo(His-Pro). **b**, strategy envisioned here, using a single enzyme catalyst to produce novel cyclodipeptide products; **c**, representative LC-HRMS spectra showing production of new molecules. Spectra for all cyclodipeptides described here are on Supplementary Figure 6 and Supplementary Figure 7.

molecule involved in the gut-brain-axis crosstalk,[18] with effects on glucose metabolism.[19] Thus, it would be advantageous to expand the chemical space of histidine-containing cyclic dipeptide analogues we could produce by using CDPS enzymes.

We describe a facile strategy for cyclic dipeptide production with superior yield and decreased cost/labour. Using canonical and non-canonical amino acids as substrates we generated a diverse library of unprecedented compounds (Supplementary Table 5). We then used small molecule substrates as chemical probes to characterise substrate binding pockets P1, which is more stringent and deeper, and P2, which is shallower and more solvent exposed.[20] To investigate histidine recognition on P1, we solved crystal structures of the cyclo(His-Pro)-producing enzyme, as well as of several mutants. 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

Enzymes that use histidinyl-tRNA as substrates

*Para*CDPS from *Parabacteroides sp. 20_3* (GenBank: EFK64745.1) produces cyclo(His-Phe)[8] whilst *Parcu*CDPS from *Parcubacteria bacterium RAAC4_OD1_1* (GenBank: ETB63777.1) can synthesise both cyclo(His-Glu) and cyclo(His-Pro)[21]. We produced both enzymes in high yield and purity and confirmed their activity and products using purified components *in vitro* (Supplementary Figure 1).

Facile production of tRNA simplifies CDP production

CDPS enzymes use aatRNA molecules as substrates, hijacking the already aminoacylated tRNA within cells for cyclodipeptide synthesis.[10] 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.[22] This method, whilst reliable, is costly, lengthy and requires specialised materials.

Aiming to bypass the individual purification of the amino acid tRNA synthetases as well as the individual tRNAs, we tested a bacterial lysate (S30 extract) isolated from *E. coli* containing amino acids, tRNA and amino acid tRNA synthetases.[23] Although initially easy to produce, this S30 extract was not as efficient at yielding CDPs as expected (Fig. 2b). Furthermore, use of non-canonical amino acids with the S30 extract was unsuccessful due to the high levels of endogenous aminoacylated tRNA present.

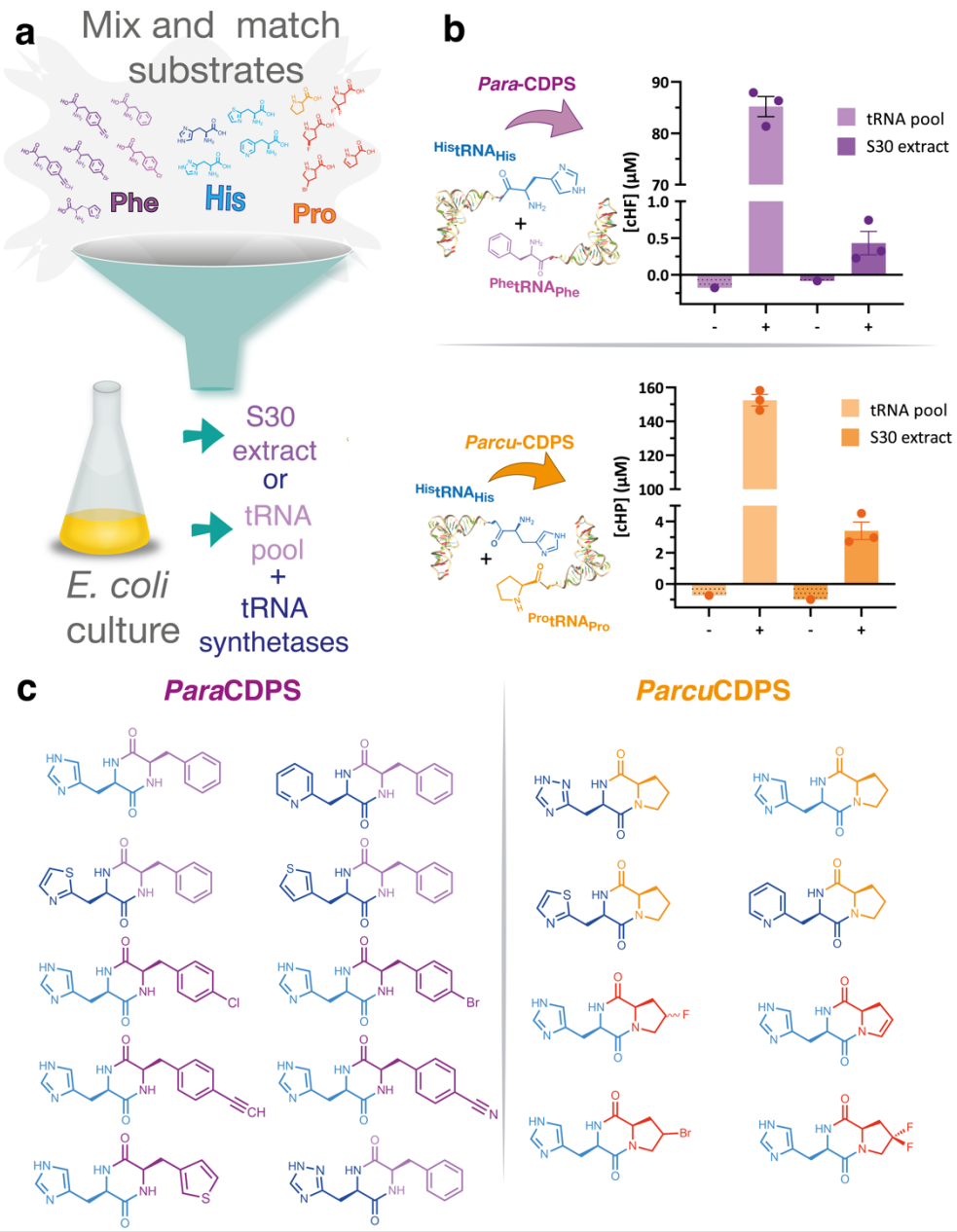
We then adapted a procedure by Mechulam *et al.* to produce a 'pool' of tRNA naturally found in *E. coli.*[24] Using this method, a highly concentrated stock of all the tRNAs required for our experiments was purified and used with any CDPS/aaRSs (aminoacyl-tRNA synthetase) combination of choice. More specifically, *ParaCDPS* and *Parcu*CDPS could use this tRNA pool to generate their respective products – cHF and cHP – in relatively high yield (Fig. 2b). In comparison, we also performed reactions using *in vitro* transcribed tRNA to confirm that the same cyclodipeptides were produced, however the tRNA pool method remains significantly cheaper, easier, and faster than previously described methods and has become a staple in our research with CDPSs. This method could be used for other enzymes that use aatRNAs as substrates, such as Fem transferases and LanB-like dehydratases.[25] 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 some CDPSs accept these as substrates to form CDPs.[26,27] Consequently, we explored the capability of *Para*CDPS and *Parcu*CDPS in accepting amino acid analogues as substrates for cyclic dipeptide production. Using a range of commercially available non-canonical analogues with the tRNA pool (Fig. 2a), a library of diverse CDPs was produced from just two CDPSs (Fig. 2c). *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. Since all analogues employed were previously shown to be substrates for histidinyl tRNA synthetase (HisRS)[26], the CDPS enzyme is posing an additional sieve for substrate selection. Prior to our work, no information was available about histidine recognition by CDPSs. The unnatural substrate utilisation emphasises 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 observation that a nitrogen may be required in close proximity to the alpha carbon of the amino acid. Supplementary Table 1 and Supplementary Table 2 summarize all analogues tested.

To further expand our diketopiperazine library we exploited the increased promiscuity of CDPSs to produce cHF analogues using a phenylalanyl tRNA synthetase mutant (PheRS-A294G) which has a wider binding pocket allowing interactions with a larger range of related Phe analogues.[28] All but two of the known non-canonical amino acids accepted by PheRS-A294G were incorporated into the ring by *Para*CDPS (Fig. 2c). All halogen para-substitutions in the phenyl ring were easily accepted by the CDPS, but not para substitutions introducing polar groups such as amine, nitro and azido (Supplementary Table 1). To produce cHP variants, several halogenated proline analogues reported as substrates for ProRS were also utilised by *Parcu*CDPS. 4-Bromo-proline was not previously shown to be a substrate for ProRS but it was

Fig. 2: Production of novel cyclodipeptides using *Para*CDPS and *Parcu*CDPS.



a, cartoon schematic depicting possible substrates for each CDPS studied herein. The tRNA structure of the aminoacylated tRNA is from the PDB 1EHZ. **b**, quantification of CDP production by LC-MS using a standard calibration curve. The presence (+) and absence (-) of CDPS is displayed for each source 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.

hypothesised to display similar chemistry as 4-fluoro-proline.[29] *Parcu*CDPS can tolerate conservative derivatisations of proline however functional groups including hydroxyl and amines on the ring were not accepted. This may be due to positioning in the pocket causing steric clashes and forcing the amino acid into unfavourable conformations for cyclisation. Although *Parcu*CDPS also produces cHE (Supplementary Fig. 8), it did not accept any glutamate analogues, demonstrating surprisingly narrow specificity.

CDP formation using a minimal substrate

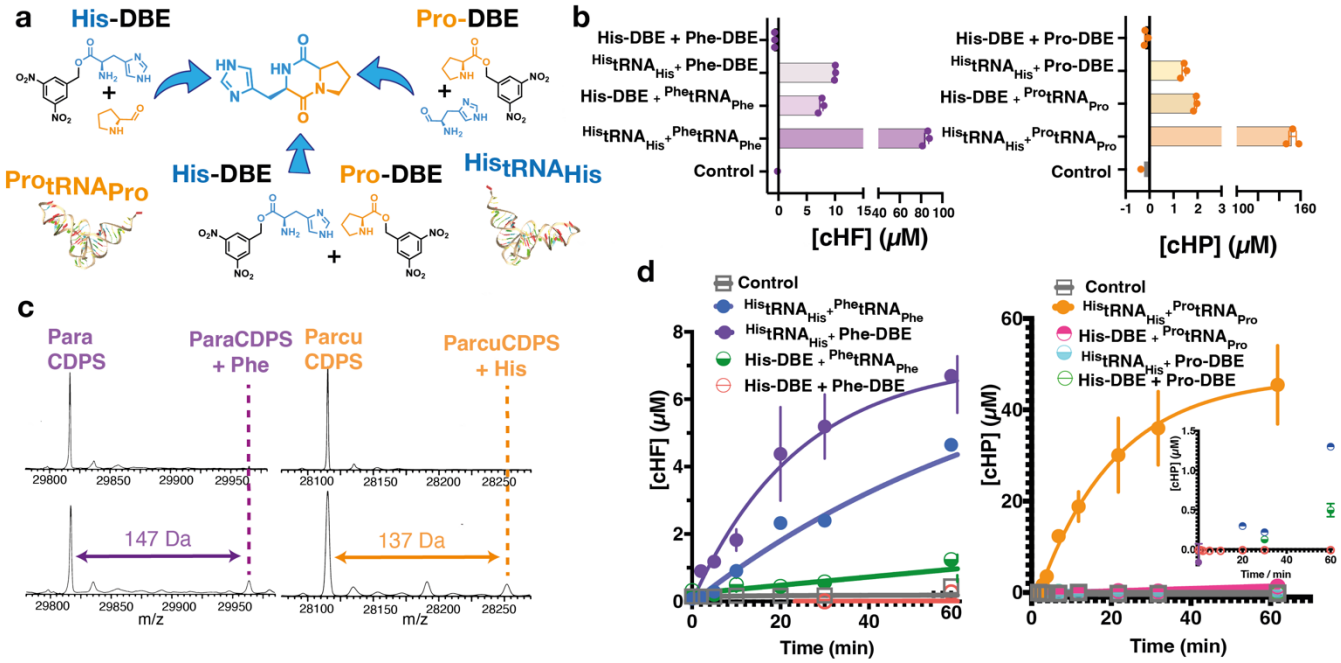
Previous research from our group demonstrated the use of small molecules containing a dinitrobenzyl ester coupled to an amino acid (aa-DBE) as substrates for CDPSs.[20] These minimal substrates are useful tools to investigate substrate specificity, and binding to P1/P2, while also presenting an alternative to using aatRNA as substrates. Prior to this work, hybrid reaction conditions using a combination of aa-DBE and aatRNA as substrates were not performed.

To test if hybrid reaction conditions would yield product, we synthesised His-DBE; Pro-DBE, and Phe-DBE. Reactions containing only minimal aa-DBE substrates were highly unfavourable and yielded very little CDP. In contrast, hybrid reactions using all available combinations for aa-DBE and aatRNA substrates (Fig. 3a) revealed that more product was formed using a combination of the aa-tRNA/aa-DBE substrates, albeit with lower overall yield when compared to the natural tRNA substrates (Fig. 3b). It is important to note that aa-DBE has a limited half-life, which also impacts reaction yields.[20] In these hybrid reaction conditions, we hypothesised that the small aa-DBE substrate would be turned over to generate a cyclic dipeptide product when occupying the deeper narrower P1 pocket, while P2 was occupied by an aatRNA, and not the other way around since aa-DBE occupying P2 would result in unproductive binding conformations, leading to aa-DBE hydrolysis and not cyclic dipeptide formed.

Our results show that *Para*CDPS generated more product using $^{His}tRNA_{His}$ + Phe-DBE, in disagreement to what would be expected if histidine was occupying P1 if the hypothesis above was correct. 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 (Fig. 3c). This therefore corroborates our hypothesis, and most likely DBE-aa occupying P2 can sample several unproductive conformations leading to DBE hydrolysis before successful product formation.

Minimal aa-DBE substrates are also useful tools to investigate kinetics of the first half reaction (acylation) and full reaction (cyclisation), as using aa-DBE is expected to increase the energy barrier for the reaction it participates in, being the first acylation step or the second half reaction to generate a cyclic peptide product. Clear differences between *Para*CDPS and *Parcu*CDPS were observed when we analysed progress curves monitoring cyclic peptide formation in hybrid reactions with aa-DBE and aa-tRNA (Fig. 3d). *Para*CDPS displayed a small difference in the rate

Fig. 3: Use of minimal substrates to yield CDPs.



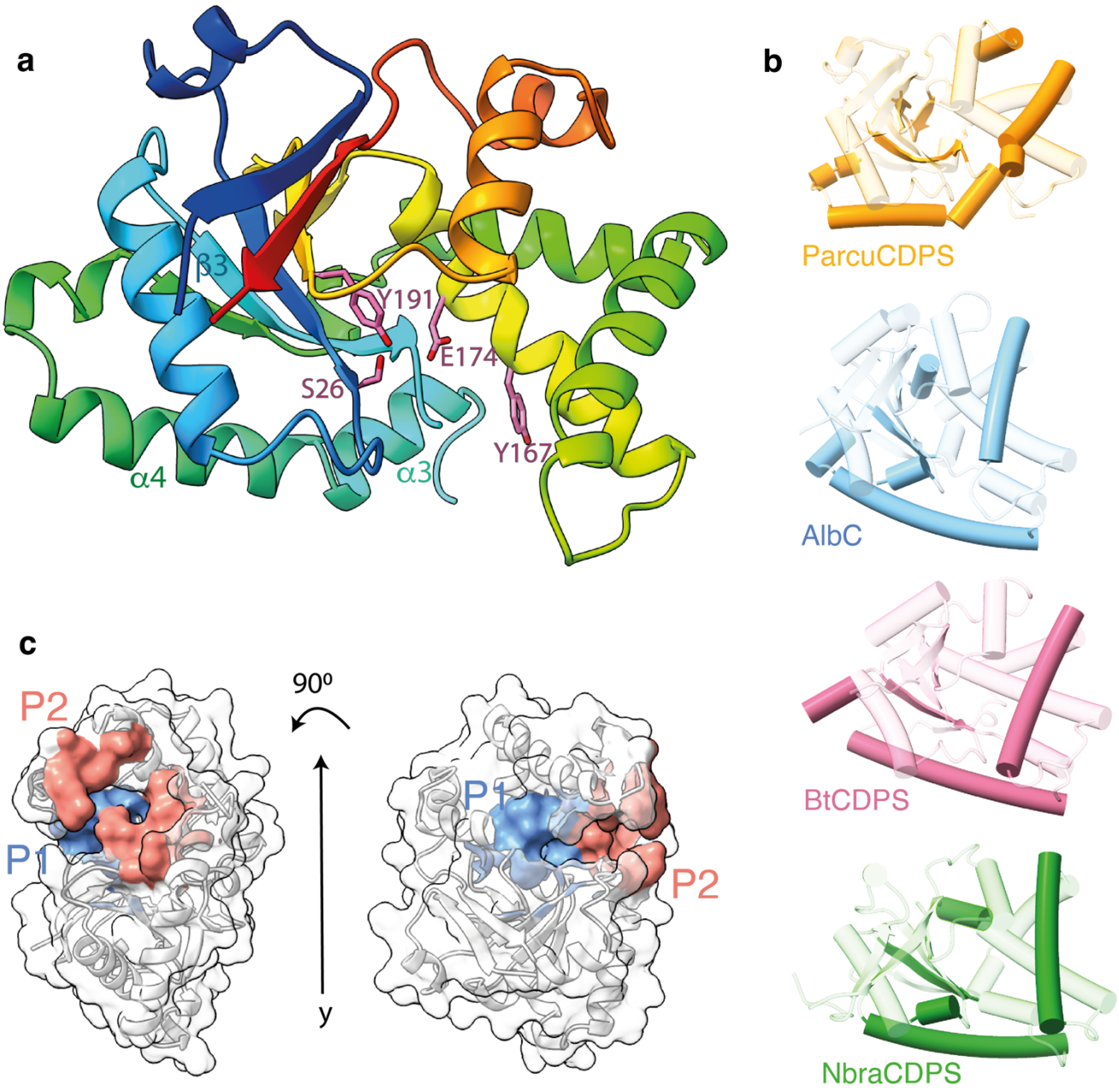
a, reaction scheme highlighting the 3 possible combinations when using DBE substrates in conjunction with aatRNA. **b,** quantification of product yield from aa-DBE and aatRNA reactions with each CDPS. The use of two aatRNA substrates continues to give the highest concentration of product. **c,** intact 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. Lines are fits to an exponential equation to yield an apparent reaction rate.

of product formed in the first hour of reaction when using $^{Phe}tRNA_{Phe}$ or Phe-DBE as substrate, suggesting the rate for the first half reaction is similar for both substrates and that cyclisation is rate determining. This points towards a less significant role of aatRNA in substrate positioning for the first half reaction. Conversely, 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.

Structure of wild type *Parcu*CDPS

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.

Fig. 4: Structure of *Parcu*CDPS.



a, structure of *Parcu*CDPS coloured rainbow (blue N terminus to red C terminus) and active site residues (pink). **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]

*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).[30] Structures of three previous members from XYP are available[31], but these enzymes use relatively hydrophobic and non-polar amino acids such as glycine, alanine and leucine. We solved the crystal structure of *Parcu*CDPS at a resolution of 1.90 Å (9 residues out of the total 230 were not traceable), uncovering unique characteristics of histidine substrate selection (Fig. 4a). The structure was solved by Iodide SAD phasing after extensive failed trials of molecular replacement, suggesting significant deviation from previously determined CDPS structures.

*Parcu*CDPS displays a Rossman-fold common throughout the CDPS family. The active site includes the four conserved residues previously identified: S26, Y167, E171 and Y191 (Fig. 4a).[31] Additionally, we hypothesised D58 was acting as a potential active site residue, within hydrogen bond distance from the catalytic serine (S26), which could be important for S26 to act as a nucleophile. Indeed, when the hydrogen bond between S26 and D58 is disrupted by mutating D58 to either an alanine or an asparagine, *Parcu*CDPS loses over 90% of activity, while preserving its structure (Fig.5b, Supplementary Figure 9). A Ser/Asp dyad was observed in Phospholipase A2,[32] and future work could be directed towards better understanding the catalytic mechanism of *Parcu*CDPS. Moreover, the side chain of the predicted catalytically important Y167 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. Further comparison with other CDPSs from the XYP family reveals RMSD values ranging from 2.7-2.9 Å. This structural comparison highlighted a divergence in secondary structure. Fig. 4b 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 α3 and beta-strand β3, where these regions change direction. In the common CDPS fold helix α3 and α4 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 features differentiate *Parcu*CDPS from the previously solved CDPS structures which are fairly conserved with respect to each other.

Unique characteristics of binding pockets P1 and P2

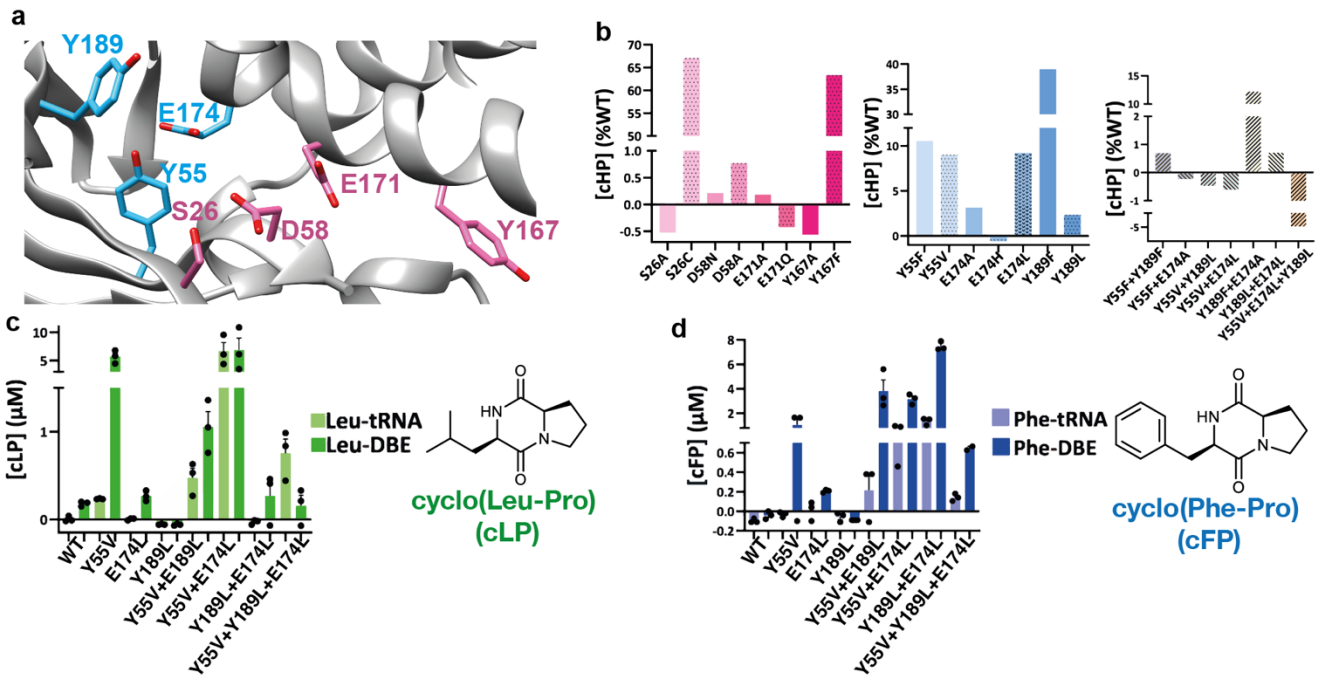
CASTp[33] was used to investigate the pocket volume of *Parcu*CDPS and highlighted pocket residues which are unique.[21] P1 (shown in blue in Fig. 4c) is found deeper within the enzyme and is smaller and more restricted in size than P2 which sits at the solvent-accessible edge of the CDPS (red on Fig. 4c). 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[34] 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 Supplementary Table 4). 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 (Supplementary Figure 8). Glutamate is a large flexible amino acid, likely negatively charged at reaction pH.[35] Thus participation in electrostatic interactions of the $^{Glu}tRNA_{Glu}$ substrate with positively charged residues in P2 is plausible, while specific interactions with $^{Pro}tRNA_{Pro}$ are less obvious.

Rationally altering substrate selection by *Parcu*CDPS

Previous research by us and others has shown that mutations in active site residues seriously reduce product formation in CDPS enzymes. However, attempts at changing the substrate scope of a CDPS by mutating select residues have been largely unsuccessful until now.[10] Inspired by HisRS and histidine recognition more generally (Supplementary Figure 11) we systematically altered three unique residues in *Parcu*CDPS P1, which we hypothesised to be participating in crucial interactions with the polar side chain of histidine. We first designed Generation 1 containing seven P1 mutants (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 (Supplementary Figure 2) 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 also generated to confirm the loss of activity upon removing catalytic residues. (Fig. 5a).

All mutants were probed for cHP production using the same activity assay performed on the WT as well as the trapped acyl-enzyme intermediate assay (Supplementary Figure 5). Fig. 5b 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.[31] This demonstrates that the phenyl ring is vital in substrate binding rather than the hydrogen bonding interactions from the phenolic hydroxyl.[36,37] When the H-bond between S26 and D58 is disrupted by mutating D58 to either an alanine or an asparagine, the enzyme is inactive. 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

Fig 5: Rational engineering of *Parcu*CDPS P1



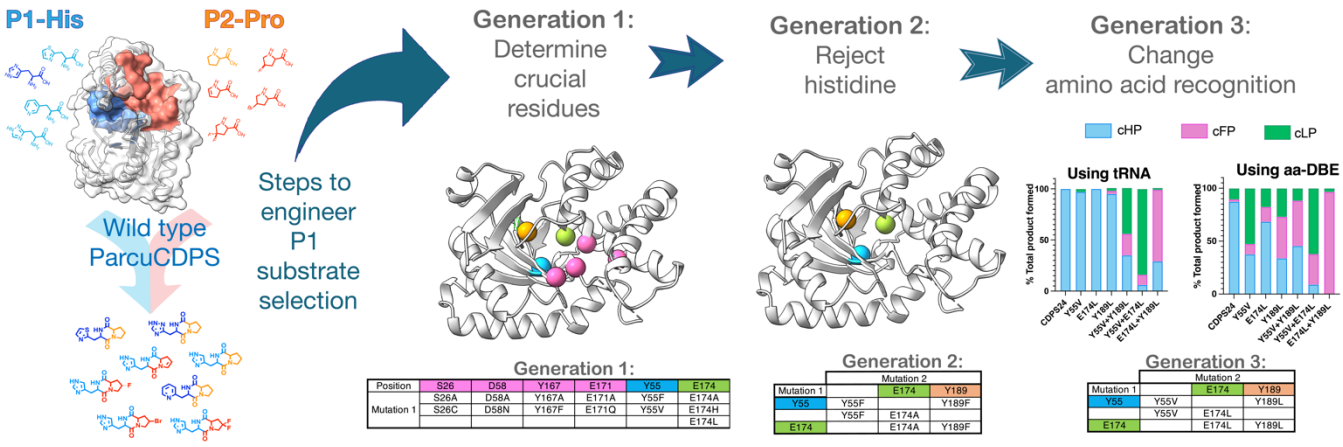
a, enhanced image of the WT structure highlighting the two sets of residues targeted for mutagenesis – active site (pink) and pocket 1 (turquoise). **b**, cHP activity assay for each set of mutants – the activity is shown as a percentage of the wild-type *Parcu*CDPS activity. Each residue is shown in 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.

solving the crystal structures of a select few (Supplementary Figure 9, Supplementary Table 3). The mutant structures retained an equivalent fold to WT enzyme. Therefore, the disruption of activity is caused exclusively by the change in these few residues which appear important for the enzyme to produce a CDP.

Following on from these results, we designed Generation 2 of double mutants from P1 residues (variants Y55F+Y189F, Y55F+E174A, Y189F+E174A, E174L+Y189L), 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 alone (Fig. 5b). 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. A similar pattern of residues interacting with the imidazole from histidine is seen in HisRS, in which two tyrosine and a glutamate residue mediate interactions with the nitrogen groups of histidine (Supplementary Figure 11).

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) and a triple mutant (Y55V+E174L+Y189L), aimed at switching 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 incapable 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 (**Fig. 5c and 5d**). 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 indicates that the enzyme is still able to recognise and reject the tRNA body, which can be circumvented 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. All double mutants were capable of producing cLP and cFP albeit with varying yields: 84% of the total products formed by Y55V+E174L was cLP, whilst for E174L+Y189L, 70% of products formed was cFP. 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.

Fig. 6: Swapping the substrate selection on P1 by *Parcu*CDPS.



Based on P1 residues we produced a series of rationally designed mutants to Generation 1) determine residues crucial for histidine recognition; Generation 2) reject histidine as a substrate and Generation 3) select a different amino acid on P1 to produce cyclic dipeptides that no longer contain histidine.

Trapped acyl enzyme intermediate experiments confirmed this trend of switched amino acid selection (Supplementary Figure S5). Overall, this is the first example of a CDPS displaying a change in substrate specificity using targeted enzyme engineering. **Fig. 6** summarizes the steps required to achieve this shift. 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.[38] 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 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 proved useful for generating cyclic dipeptides containing both canonical and non-canonical amino acids.

Additionally, structural characterisation of *Parcu*CDPS revealed a so far unique pocket topology to accommodate histidine as a substrate. By trapping the acyl-enzyme intermediate we determined that histidine was bound in P1 of *Parcu*CDPS but on P2 for *Para*CDPS, stressing differences between the two enzymes, and more broadly on the rate limiting nature of different steps in the reactions they catalyse.

Finally, combining structural biology and activity assays, we provide a much clearer picture of how polar residues such as histidine are selected by cyclodipeptide synthases as substrates on P1, as well as how this selectivity can be manipulated by rational engineering to produce molecules of our choosing. 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 by engineering P1,[39] there are no published attempts to alter the substrate scope of a CDPS enzyme. Therefore, this is a pivotal finding which could lead to a wide array of CDPs from a single engineered enzyme, taking control and manipulating substrate selection by these enzymes.

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, and Professor Malcolm White for thoughtful discussions.

References

1 Gondry, M. *et al.* Cyclodipeptide synthases are a family of tRNA-dependent peptide bond?forming enzymes. *Nat. Chem. Bio.* **5**, 414-420, doi:10.1038/nchembio.175 (2009).

2 Borthwick, A. D. 2,5-Diketopiperazines: Synthesis, Reactions, Medicinal Chemistry, and Bioactive Natural Products. *Chem. Rev.* **112**, 3641-3716, doi:10.1021/cr200398y (2012).

3 González, J. F., Ortín, I., de la Cuesta, E. & Menéndez, J. C. Privileged scaffolds in synthesis: 2,5-piperazinediones as templates for the preparation of structurally diverse heterocycles. *Chem. Soc. Rev.* **41**, 6902-6915, doi:10.1039/C2CS35158G (2012).

4 Martins, M. B. & Carvalho, I. Dikeopiperazines: biological activity and synthesis. *Tetrahedron* **63**, 9923-9932 (2007).

5 Borgman, P., Lopez, R. D. & Lane, A. L. The expanding spectrum of diketopiperazine natural product biosynthetic pathways containing cyclodipeptide synthases. *Org Biomol Chem* **17**, 2305-2314, doi:10.1039/c8ob03063d (2019).

6 Giessen, T. W. & Marahiel, M. A. Rational and combinatorial tailoring of bioactive cyclic dipeptides. *Front Microbiol* **6**, 785, doi:10.3389/fmicb.2015.00785 (2015).

7 Skinnider, M. A., Johnston, C. W., Merwin, N. J., Dejong, C. A. & Magarvey, N. A. Global analysis of prokaryotic tRNA-derived cyclodipeptide biosynthesis. *BMC Genomics* **19**, 45, doi:10.1186/s12864-018-4435-1 (2018).

8 Jacques, I. B. *et al.* Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity. *Nat Chem Biol* **11**, 721-727, doi:10.1038/nchembio.1868 (2015).

9 Gondry, M. *et al.* A Comprehensive Overview of the Cyclodipeptide Synthase Family Enriched with the Characterization of 32 New Enzymes. *Front Microbiol* **9**, 46, doi:10.3389/fmicb.2018.00046 (2018).
10 Canu, N., Moutiez, M., Belin, P. & Gondry, M. Cyclodipeptide synthases: a promising biotechnological tool for the synthesis of diverse 2,5-diketopiperazines. *Nat Prod Rep* **37**, 312-321, doi:10.1039/c9np00036d (2020).
11 Kanoh, K. *et al.* (−)-Phenylahistin: A new mammalian cell cycle inhibitor produced by aspergillus ustus. *Bioorg. Med. Chem. Lett.* **7**, 2847-2852, doi:https://doi.org/10.1016/S0960-894X(97)10104-4 (1997).
12 Nicholson, B. *et al.* NPI-2358 is a tubulin-depolymerizing agent: in-vitro evidence for activity as a tumor vascular-disrupting agent. *Anti-Cancer Drugs* **17**, 25-31, doi:10.1097/01.cad.0000182745.01612.8a (2006).
13 Cimino, P. J. *et al.* Plinabulin, an inhibitor of tubulin polymerization, targets KRAS signaling through disruption of endosomal recycling. *Biomed Rep* **10**, 218-224, doi:10.3892/br.2019.1196 (2019).
14 Grottelli, S. F., I.; Pietrini, G.; Peirce, M.J.; Minelli, A.; Bellezza, I. The Role of Cyclo(His-Pro) in Neurodegeneration. *Int. J. Mol. Sci.* **17**, 1332 (2016).
15 Prasad, C. Cyclo(His-Pro): its distribution, origin and function in the human. *Neurosci. Biobehav. Rev.* **12**, 19-22, doi:10.1016/s0149-7634(88)80069-1 (1988).
16 Bellezza, I., Peirce, M. J. & Minelli, A. Cyclic dipeptides: from bugs to brain. *Trends Mol. Med.* **20**, 551-558, doi:https://doi.org/10.1016/j.molmed.2014.08.003 (2014).
17 Grottelli, S. *et al.* Cyclo(His-Pro) inhibits NLRP3 inflammasome cascade in ALS microglial cells. *Mol. Cell. Neurosci.* **94**, 23-31, doi:https://doi.org/10.1016/j.mcn.2018.11.002 (2019).
18 Bellezza, I., Peirce, M. J. & Minelli, A. in *Quorum Sens.* (ed Giuseppina Tommonaro) Ch. 10, 257-286 (Academic Press, 2019).
19 Song, M. K., Bischoff, D. S., Song, A. M., Uyemura, K. & Yamaguchi, D. T. Metabolic relationship between diabetes and Alzheimer's Disease affected by Cyclo(His-Pro) plus zinc treatment. *BBA Clinical* **7**, 41-54, doi:https://doi.org/10.1016/j.bbacli.2016.09.003 (2017).
20 Harding, C. J., Sutherland, E., Hanna, J. G., Houston, D. R. & Czekster, C. M. Bypassing the requirement for aminoacyl-tRNA by a cyclodipeptide synthase enzyme. *RSC Chem. Bio.* **2**, 230-240, doi:10.1039/D0CB00142B (2021).
21 Gondry, M. *et al.* A Comprehensive Overview of the Cyclodipeptide Synthase Family Enriched with the Characterization of 32 New Enzymes. *Front. Microbiol.* **9**, doi:10.3389/fmicb.2018.00046 (2018).
22 Bertand Beckert, B. M. Synthesis of RNA by In Vitro Transcription. *RNA: Methods in molecular biology* **703**, 29-41 (2011).
23 Krinsky, N. *et al.* A Simple and Rapid Method for Preparing a Cell-Free Bacterial Lysate for Protein Synthesis. *PLOS ONE* **11**, e0165137, doi:10.1371/journal.pone.0165137 (2016).
24 Mechulam, Y., Guillon, L., Yatime, L., Blanquet, S. & Schmitt, E. in *Methods in Enzymology* Vol. 430 (ed Jon Lorsch) 265-281 (Academic Press, 2007).
25 Moutiez, M., Belin, P. & Gondry, M. Aminoacyl-tRNA-Utilizing Enzymes in Natural Product Biosynthesis. *Chem Rev* **117**, 5578-5618, doi:10.1021/acs.chemrev.6b00523 (2017).
26 Hartman, M. C. T., Josephson, K. & Szostak, J. W. Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc Natl Acad Sci U S A* **103**, 4356-4361, doi:10.1073/pnas.0509219103 (2006).
27 Canu, N. *et al.* Incorporation of Non-canonical Amino Acids into 2,5-Diketopiperazines by Cyclodipeptide Synthases. *Angewandte Chemie International Edition* **57**, 3118-3122, doi:10.1002/anie.201712536 (2018).
28 Kast, P. & Hennecke, H. Amino acid substrate specificity of Escherichia coli phenylalanyl-tRNA synthetase altered by distinct mutations. *Journal of Molecular Biology* **222**, 99-124, doi:https://doi.org/10.1016/0022-2836(91)90740-W (1991).
29 Hartman, M. C. T. Non-canonical Amino Acid Substrates of E. coli Aminoacyl-tRNA Synthetases. *ChemBioChem* **22**, 1-17, doi:https://doi.org/10.1002/cbic.202100299 (2021).
30 Jacques, I. B. *et al.* Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity. *Nat. Chem. Bio.* **11**, 721, doi:10.1038/nchembio.1868 (2015).
31 Bourgeois, G. *et al.* Structural basis for partition of the cyclodipeptide synthases into two subfamilies. *J. Struct. Biol.* **203**, 17-26 (2018).
32 Dessen, A. *et al.* Crystal structure of human cytosolic phospholipase A2 reveals a novel topology and catalytic mechanism. *Cell* **97**, 349-360, doi:10.1016/s0092-8674(00)80744-8 (1999).
33 Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Research* **46**, W363-W367, doi:10.1093/nar/gky473 (2018).
34 Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Sci* **27**, 112-128, doi:10.1002/pro.3280 (2018).
35 Harris, T. K. & Turner, G. J. Structural Basis of Perturbed pKa Values of Catalytic Groups in Enzyme Active Sites. *IUBMB Life* **53**, 85-98, doi:https://doi.org/10.1080/15216540211468 (2002).
36 Vetting, M. W., Hegde, S. S. & Blanchard, J. S. The structure and mechanism of the Mycobacterium tuberculosis cyclodityrosine synthetase. *Nat. Chem. Bio.* **6**, 797-799, doi:10.1038/nchembio.440 (2010).
37 Moutiez, M. *et al.* Unravelling the mechanism of non-ribosomal peptide synthesis by cyclodipeptide synthases. *Nat. Commun.* **5**, 5141, doi:10.1038/ncomms6141 (2014).
38 Jinendiran, S. *et al.* Induction of mitochondria-mediated apoptosis and suppression of tumor growth in zebrafish xenograft model by cyclic dipeptides identified from Exiguobacterium acetylicum. *Scientific Reports* **10**, 13721, doi:10.1038/s41598-020-70516-x (2020).

39 Brockmeyer, K. & Li, S. M. Mutations of Residues in Pocket P1 of a Cyclodipeptide Synthase Strongly Increase Product Formation. *J Nat Prod* **80**, 2917-2922, doi:10.1021/acs.jnatprod.7b00430 (2017).