Cyclase library phylogenetic tree  
A phylogenetic tree of the cyclase library produced by Katja

Phylogenetic analysis methods

Amino acid sequences were aligned with MAFFT v7.490 and ran with the --auto flag (1). A maximum-likelihood phylogenetic tree was produced with iqtree2 using ModelFinder Plus (MFP) and 1000 ultrafast bootstrap replicates (2). The phylogenetic tree was visualised and annotated in FigTree v1.4.4 (<https://github.com/rambaut/figtree>).

As a starting point for library construction, sequences encoding 13 previously characterized spirotetronate cyclases were selected for inclusion (Table S2). These were complimented with the sequence of Tsn15, which possesses a characteristic spirotetronate cyclase 8 stranded β-barrel core fold, but which performs a pericyclic rearrangement of a six-membered ring in tetronasin biosynthesis (Scheme S1F).(3) To expand the library yet further, 286 unique amino acid sequences were identified in public databases and assembled into a sequence similarity network (SSN, Figure 1). Distinct clustering was observed for the known cyclases KijU, LobD1, LonU2, TcaU4, and VstJ, and for ChlL and PyrI4. From this SSN analysis 17 putative cyclase sequences were selected (Table S4) and homology models generated for these polypeptides. Three methods were used, (i) RosettaCM(4) as implemented in the Cyrus-CAD bench application, (ii) a template based approach using YASARA,(5) employing the reported crystal structures of AbmU, AbyU, PyrI4, Tsn15, and (iii) AlphaFold 2.0.(6) Models which predicted 8 stranded β-barrel folds, consistent with those of known spirotetronate cyclases, were taken as an indication of the likely function of these enzymes. As a consequence of these analyses 12 of the 17 putative cyclase sequences were selected for inclusion in the library, with five candidates excluded.

With the release of AlphaFold 2.0 (AF2), it was possible to perform a full structural comparison of the 26 selected known and predicted cyclases (Figure S2). While AF2 could predict the core structure of the candidate cyclases with high confidence, in 60% of the cases the terminal regions were not predicted to be structured. Furthermore, the candidate cyclases Cyc06, Cyc12, and Cyc13 were predicted to adopt non-β-barrel folds, with mn bhjvFor the 23 structures which exhibit known or predicted β-barrel folds, each houses a predominantly hydrophobic binding pocket within the barrel lumen, populated with a high proportion of aromatic residues. In many instances the enzyme binding pocket contains a pair of aromatic residues, which are resident on opposing sides of the barrel, whose side chains form π-π stacking interactions, e.g. Trp124 and Phe41 in AbyU (Figure S1). Notably, such interactions are not observed in Tmn8 and Tsn15, which could account for the distinctive pericyclic rearrangements reported for these enzymes.(3, 7) The predicted structures of these polypeptides are also distinct in that they each lack a salt-bridge at the barrel ‘head’. In the majority of the 23 structures a glutamic acid and arginine pair fulfil this role, however, in PyrI4, ChlL, Cyc03, and Cyc04 the arginine residue is replaced by histidine. PyrI4, ChlL and Cyc03 reside in the same cluster in our SSN analysis and share an extended yet structured N-terminal region, analogous to that reported in PyrI4.(8) With the exception of Cyc05, each of our models possesses a 6-15 residue capping loop, which regulates access to the enzyme active site and may play a role in bringing the substrate into a reactive conformation.(8, 9)

Current models of catalysis in spirotetronate cyclases are derived predominantly from studies of AbmU(10), AbyU(9, 11, 12)and PyrI4.(8, 13) It is proposed that the substrate binding cavity is essential for forming an environment which brings the diene and dienophile into close proximity, thus facilitating the [4+2] cycloaddition reaction. There are no explicitly conserved active site residues shared by these three enzymes, and there is a general lack of mechanistic understanding with respect to the broader family of cyclases.(14, 15, 16) To establish if this plasticity impacts on substrate selectivity in the spirotetronate cyclases, we further expanded the scope of our cyclase library through inclusion of 19 additional candidate sequences from the literature (Table S5), (more details?) to yield a 45-member library. We performed a phylogenetic analysis of the 45 cyclase library to explore the evolutionary relationships and potential functional differences among the cyclases (Figure S2). The spirotetronate-forming cyclases claded clearly together. The three abyssomicin diels-alderases (AbmU, AbyU and AbsU) formed a clade with 5 putative cyclases, including Cyc17, which is likely involved in the production of a recently discovered class of spirotetronates the Wychimicins (ref). The phylogenetic analysis potentially explained the adoption of non-β-barrel folds in Cyc06, Cyc12, and Cyc13, with all three showing stronger evolutionary relationships with non-spirotetranate forming cyclases. To enable activity screening of library members each enzyme was recombinantly over-expressed in *Escherichia coli* and was soluble purified to homogeneity. To expedite this process we established a semi-automated protein-production workflow which enabled us to prepare 12 purified proteins per week and involved expression in auto-induction media, followed by affinity and size exclusion chromatography. Protein identity and homogeneity were evaluated by SDS-PAGE analysis, melting temperature (TM) studies and peptide mapping (Table S6). Of our target 45 proteins 31 were successfully produced using this workflow, including 12 spirotetronate cyclases previously described in the literature. The only exceptions were the individual N- and C-terminal domains of QmnH which proved recalcitrant to production using this approach. Of the 12 selected sequences of putative spirotetronate cyclases, 7 could be produced solubly and in high yields, with poor expression observed for Cyc01, Cyc02, Cyc06, Cyc16, and Cyc17 under our standard conditions. An additional 12 cyclases identified in the literature, but with no previously reported spirotetronate cyclase activity, were also solubly expressed and purified, while five cyclases from fungal sources (CcsF, EupF, gNR600, mAsR5, Sol5) and two bacterial enzymes (PyrE3, SpnF) were not expressed in sufficient quantities to enable further characterisation. Five proteins exhibited melting temperatures above 70°C (AbyU, LonU2, PyrI4, Cyc03, and Cyc15). Although there was variation in purified protein yields, it was possible to obtain at least 1 mg of protein for 29 of the cyclases.

|  |
| --- |
|  |
| Figure 1. Unrooted maximum likelihood phylogenetic analysis of protein sequences. IQ-tree2 MFP selected WAG+F+R4 as the best model and identified 152 constant sites (15.46%) across the 45 sequences, Branches coloured according to proven or predicted spirotetronate activity (blue), non-spirotetronate cyclases (green) and putative cyclases from this study (red). |

|  |
| --- |
|  |
|  |

Figure 2. Alternative with proposed enzyme clades highlighted.



Now published as:

Wychimicins - <https://www.nature.com/articles/s41429-022-00560-4>

IQ-tree-2 data

**Input data:** 45 sequences with 983 amino-acid sites

**Number of constant sites:** 152 (= 15.4629% of all sites)

**Number of parsimony informative sites:** 485

**Number of distinct site patterns:** 934



Reviewing the paper for Hmm work:

Two well studied families of **[4+2]** cycloadditions – spirotetronate cyclases and and decalin forming cyclases

To date, five enzymes catalysing spirotetronate formation have been characterised in detail, PyrI4,[7](https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/cbic.202300382#cbic202300382-bib-0037) AbyU,[8](https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/cbic.202300382#cbic202300382-bib-0038) AbmU,[9](https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/cbic.202300382#cbic202300382-bib-0042) AbnU[10](https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/cbic.202300382#cbic202300382-bib-0043) and PloI4.[11](https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/cbic.202300382#cbic202300382-bib-0044)

13 previously characterised spirotetronate cyclases were selected for inclusion (Table S2)

Tsn15 performs a weird cyclization

286 unique amino acid sequences were identified

A cluster of 100 sequences was identified and excluded, since the 43 annotated sequences belonged to different protein-fold families exhibiting cofactor binding motives like the two FAD-dependent decalin forming cyclases ChlE3 and LobP3

So we have the 14 spirotetronate cyclases

The 17 putative cyclases

Plus 19 additional candidate sequences from the litature

Do we have confirmation of DA activity for any of the other cyclases?

Cyclase15B?

Two models?

1. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution. 2013;30(4):772-80.

2. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37(5):1530-4.

3. Little R, Paiva FCR, Jenkins R, Hong H, Sun Y, Demydchuk Y, et al. Unexpected enzyme-catalysed [4+2] cycloaddition and rearrangement in polyether antibiotic biosynthesis. Nature Catalysis. 2019;2(11):1045-54.

4. Song Y, DiMaio F, Wang RY, Kim D, Miles C, Brunette T, et al. High-resolution comparative modeling with RosettaCM. Structure. 2013;21(10):1735-42.

5. Venselaar H, Joosten RP, Vroling B, Baakman CA, Hekkelman ML, Krieger E, et al. Homology modelling and spectroscopy, a never-ending love story. Eur Biophys J. 2010;39(4):551-63.

6. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-9.

7. Little RF, Samborskyy M, Leadlay PF. The biosynthetic pathway to tetromadurin (SF2487/A80577), a polyether tetronate antibiotic. PLoS One. 2020;15(9):e0239054.

8. Zheng Q, Guo Y, Yang L, Zhao Z, Wu Z, Zhang H, et al. Enzyme-Dependent [4 + 2] Cycloaddition Depends on Lid-like Interaction of the N-Terminal Sequence with the Catalytic Core in PyrI4. Cell Chem Biol. 2016;23(3):352-60.

9. Byrne MJ, Lees NR, Han LC, van der Kamp MW, Mulholland AJ, Stach JE, et al. The Catalytic Mechanism of a Natural Diels-Alderase Revealed in Molecular Detail. J Am Chem Soc. 2016;138(19):6095-8.

10. Li Q, Ding W, Tu J, Chi C, Huang H, Ji X, et al. Nonspecific Heme-Binding Cyclase, AbmU, Catalyzes [4 + 2] Cycloaddition during Neoabyssomicin Biosynthesis. ACS Omega. 2020;5(32):20548-57.

11. Marsh CO, Lees NR, Han LC, Byrne MJ, Mbatha SZ, Maschio L, et al. A Natural Diels‐Alder Biocatalyst Enables Efficient [4+2] Cycloaddition Under Harsh Reaction Conditions. ChemCatChem. 2019;11(20):5027-31.

12. Ding W, Chi C, Wei X, Sun C, Tu J, Ma M, et al. Enzymatic Synthesis of a Diastereomer of Neoabyssomicin Derivative Using the

Diels‐Alderase AbyU. Chinese Journal of Chemistry. 2021;39(7):1871-7.

13. Zou Y, Yang S, Sanders JN, Li W, Yu P, Wang H, et al. Computational Investigation of the Mechanism of Diels-Alderase PyrI4. J Am Chem Soc. 2020;142(47):20232-9.

14. Xu G, Yang S. Diverse evolutionary origins of microbial [4 + 2]-cyclases in natural product biosynthesis. Int J Biol Macromol. 2021;182:154-61.

15. Watanabe K. Discovery and investigation of natural Diels-Alderases. J Nat Med. 2021;75(3):434-47.

16. Minami A, Oikawa H. Recent advances of Diels-Alderases involved in natural product biosynthesis. J Antibiot (Tokyo). 2016;69(7):500-6.