Spontaneity in the patellamide biosynthetic pathway

Bruce F. Milne,[a] Paul F. Long,[b] Antonio Starcevic,[c] Daslav Hranueli[c] and Marcel Jaspars*[d]

Received 9th November 2005, Accepted 9th December 2005
First published as an Advance Article on the web 9th January 2006
DOI: 10.1039/b515938e

Post-translationally modified ribosomal peptides are unusual natural products and many have potent biological activity. The biosynthetic processes involved in their formation have been delineated for some, but the patellamides represent a unique group of these metabolites with a combination of a macrocycle, small heterocycles and D-stereocentres. The genes encoding for the patellamides show very low homology to known biosynthetic genes and there appear to be no explicit genes for the macrocyclisation and epimerisation steps. Using a combination of literature data and large-scale molecular dynamics calculations with explicit solvent, we propose that the macrocyclisation and epimerisation steps are spontaneous and interdependent and a feature of the structure of the linear peptide. Our study suggests the steps in the biosynthetic route are heterocyclisation, macrocyclisation, followed by epimerisation and finally dehydrogenation. This study is presented as testable hypothesis based on literature and theoretical data to be verified by future detailed experimental investigations.

Introduction

Marine natural products show promise as candidates in many therapeutic areas, but the issue of a viable economic supply has stalled the development of many.[1] Chemical synthesis is a solution in some cases, but biotechnological approaches may be advocated in others. The patellamides (Fig. 1) are a family of highly conserved thiazole and oxazoline containing cyclic octapeptides isolated from the Indo-Pacific seasquirt *Lissoclinum patella*. The patellamides display a variety of biological activities including cytotoxicity[2] and as selective antagonists for reversing the multidrug resistant CEM/VLB100 human leukemic cell line towards vinblastine, colchicine and adriamycin treatment.[3] The patellamides are part of a family of bioactive thiazole, thiazoline, oxazole and oxazoline containing natural products including bleomycin, thiostrepton and diazonamide A.[4] The biological activities of this family of alkaloids can be attributed to the conformational constraints imposed by the heterocycles and their ability to bind metals or intercalate into DNA.[4] *L. patella* contains an as yet uncultured primitive photosynthetic prokaryote *Prochloron didemni*, which is closely related to the cyanobacteria, but differs in the fact that it possesses both chlorophylls *a* and *b*, lacks phycobilins and has plant-like thylakoids.[5]

Recent work has indicated that the patellamides are unprecedented examples of post-translationally modified ribosomal peptides,[6] produced by the *Prochloron didemni* symbiont.[7] The patellamides are the most complex examples of this unusual family of peptides which includes the lantibiotics[8] and microcins,[4] as they contain a unique combination of heterocycles (*cf* microcin B17),[4] a macrocycle (*cf* microcin J25)[9] and D stereocentres. In addition, the same gene encodes two different products. The gene cluster shows extremely low sequence homology to genes of known function,[6] indicating that the pathway might contain novel enzymes. In addition, the heterocyclisation of dipeptides (XC, XT, XS) to give thiazole and oxazoline rings appears to be more tolerant to changes at X[10] than that observed in microcin B17, indicating that this process might be exploitable for biotechnological applications such as the combinatorial biosynthesis of novel bioactive products and semi-synthetics.

In brief, the biosynthesis of such post-translationally modified ribosomal peptides occurs by the tailoring of the pre-propeptide coding sequence by the post-translational modification machinery, which recognises the leader sequence and start/stop sequences (Scheme 1). The gene cluster sequenced by Schmidt *et al.* (Scheme 1)[6] allows ascribing of function to some of the genes, but raises many questions. In particular, *patB*, *patC* and *patF* show low or no similarity to proteins of known function. Some of the biosynthetic steps can be rationalised using the putative function of the other genes (Scheme 1). It is suggested that *patD2* is involved in the heterocyclisation to form the oxazoline and thiazoline rings, the latter of which is oxidised to the thiazole by *patG1*. Cleavage of the mature *patE* could occur under the influence of *patG2* or *patA*, followed by adenylation by *patD1* and macrocyclisation. One step that is unclear is the epimerisation of the stereocentres adjacent to the thiazole rings, although synthetic studies have shown that this type of system can self-organise into the correct epimer at the thiazoline stage.[11] Macrocyclisation of the mature peptide could occur under the influence of one of the genes of unassigned function, or the linear mature octapeptide may self-organise to allow macrocyclisation with simple activation of the C-terminus. In this paper we present literature and theoretical data that suggests that the epimerisation and macrocyclisation processes may be spontaneous and interdependent.

[a]*Centro de Estudos de Química Orgânica, Fitoquímica e Farmacologia, Faculdade de Farmácia, Universidade do Porto, Rua Aníbal Cunha 164, 4050-047, Porto, Portugal*
[b]*School of Pharmacy, University of London, 29/39 Brunswick Square, London, UK WC1N 1AX*
[c]*Faculty of Food Technology and Biotechnology, University of Zagreb, Pierottijeva 6, 10000, Zagreb, Croatia*
[d]*Marine Natural Products Laboratory, Department of Chemistry, University of Aberdeen, Old Aberdeen, Scotland, UK AB24 3UE. E-mail: m.jaspars@abdn.ac.uk; Fax: +44 1224 272 921*

		R_1	R_2	R_3	R_4	Sequence
1	Pat A	$CHMeEt$	$CHMe_2$	$CHMeEt$	$CHMe_2$ 1-no Me	ITVCISVC
2	Pat B	CH_2CHMe_2	$CHMeEt$	$CHMeEt$	CH_2Ph	LTICITFC
3	Pat C	$CHMe_2$	Me	$CHMeEt$	CH_2Ph	VTACITFC
4	Pat D	$CHMeEt$	Me	$CHMeEt$	CH_2Ph	ITACITFC
5	Pat E	$CHMe_2$	$CHMe_2$	$CHMeEt$	CH_2Ph	VTVCITFC
6	Pat F	$CHMe_2$	$CHMe_2$	$CHMe_2$	CH_2Ph	VTVCVTFC
7	Pat G	$CHMeEt$	Me	CH_2CHMe_2	CH_2Ph	ITACLTFC
8	Asc	$CHMeEt$	$CHMe_2$	$CHMeEt$	$CHMe_2$	ITVCITVC

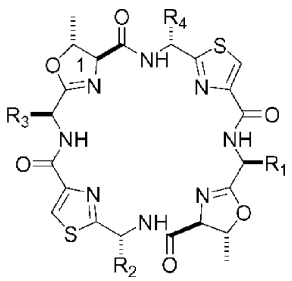
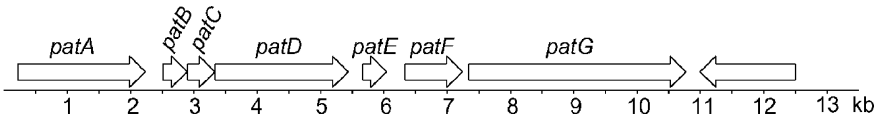


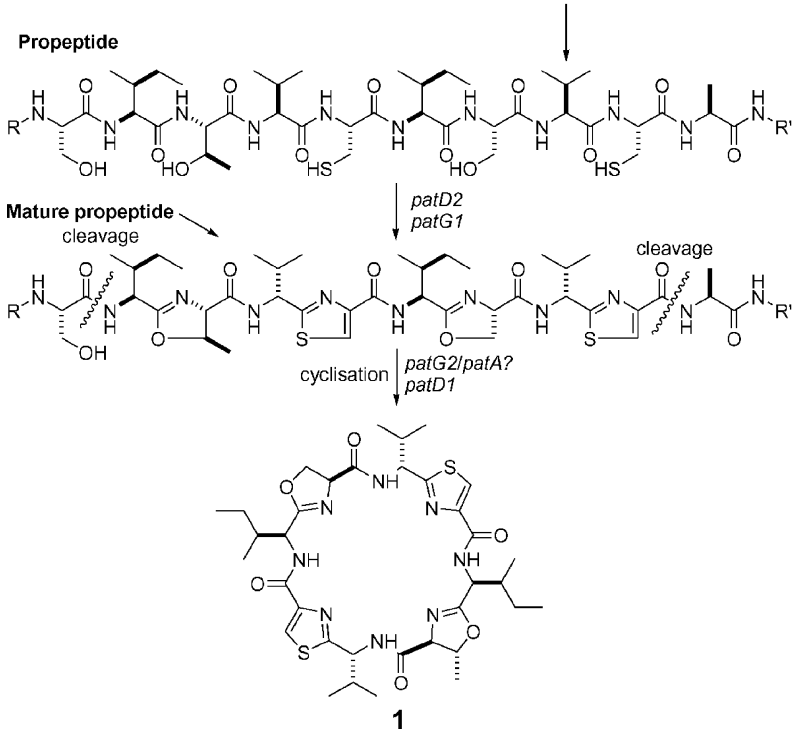
Fig. 1 Structures of the known patellamides from *Lissoclinum patella* and their peptide sequences.



PatE gene product (pre-propeptide, single 71 aa peptide)

*MNKKNILPQQGQPVIRLTAGQLSSQLAELSEEALG*DA**GLEAS**VTACITFC**AYD**

GVEPSITVCISVC**AYDGE**



Scheme 1 The *pat* gene cluster, the *patE* sequence encoding for **3** (upper) and **1** (lower) and suggested biosynthetic pathway for **1**. Italic = leader sequence. Bold = start, stop and start/stop cyclisation sequences. Underline = patellamide coding sequences. 'Start' and 'stop' peptide sequences in the prepeptide indicate the limits between which the tailoring enzymes must act.

Results and discussion

Repeating essentially the autoannotation of Schmidt *et al.*,[6] a search of the whole Pfam HMM library with each protein sequence of seven *Prochloron didemni* genes did not show the presence of additional adenylation, epimerisation, racemisation or cyclisation domains other than those shown in the sequence description. The highest scores and most relevant results were obtained with peptidase S8 from the subtilase family found in *patA* subtilisin-like protein and as a part of *patG* thiazoline oxidase/subtilisin-like protease and the YcaO from YcaO-like family found in the *patD* adenylation/heterocyclisation protein (identified by match to protein families HMM PF00082 and HMM PF02624). Other genes (beside *patA*, *patD*, and *patG*) showed no hits at all or scores just above the threshold that cannot be considered reliable. For example, *patF* shows very weak homology to propeptide C1 and SpoVT/AbrB like domain, and there might be a nitroreductase in *patG*, but also with a very weak signal (identified by match to protein families HMM PF08127, HMM PF04014 and HMM PF00881, respectively). *PatB*, *patC*

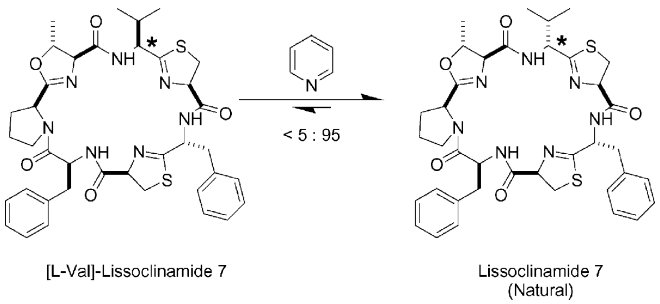
and *patE* did not score above the threshold in Pfam. In addition, when profiles of NRPS adenylation, cyclisation and epimerisation domains were generated[12] for a more detailed annotation of individual *Prochloron didemni* genes,[6] no significant hits were detected.

Further analysis indicates that *patA* contains another unknown domain. The peptidase S8 is localised in the first half of the peptide (amino acids 1–254), whereas the second half is still uncharacterised. *PatD*, which is postulated to have adenylation/heterocyclisation activity, shows a well-defined YcaO-like domain in the second half of the protein leaving space for another domain. Therefore it is likely that *patD* also contains two domains, one of which is totally unknown with no homology to anything in the database. For *patF* one can conclude that it contains two domains because the hits (identified by matches to protein families HMM PF08127 and HMM PF04014) do not overlap and expand throughout the peptide. One can only speculate about the nature of these domains because the *E*-values are simply too weak. *PatG* is the largest polypeptide in the cluster and probably contains four domains. The only one that can be certainly identified is the peptidase (peptidase S8 from the subtilase family) as the third domain (amino acids 522–837). The second domain might be a dehydrogenase localised around amino acids 276–480, while the first and fourth domains are localised around amino acids 1–270 and 840–1100.

We will analyse the biosynthetic pathway (Scheme 1) in detail. There are two possible routes by which heterocyclisation may occur, cyclodehydration followed by oxidation (dehydrogenation) or oxidation followed by cyclodehydration. The former has been observed for the microcin B17 system, whereas the latter is relatively rare.[4] In the case of the patellamides heterocycles at different oxidation levels, oxazolines and thiazoles, are incorporated suggesting that cyclodehydration occurs first to give the oxazoline and thiazoline and that the latter is then oxidised to the thiazole by *patG2* and/or *patA*.[6] However, in the microcin B17 system,[4] the same enzymes (*McbB*, *C*, *D*) process GC and GS to thiazole and oxazole respectively. Thiazolines are inherently more reactive than oxazolines suggesting that the oxidation of thiazoline to thiazole in the patellamide biosynthesis may occur spontaneously, but two other scenarios can be proposed for the lack of further oxidation of the oxazoline to oxazole. In the first, kinetic release occurs before the oxazoline is oxidised to the oxazole, and in the second it is proposed that the biosynthetic enzymes lack terminal dehydrogenase activity. The presence of a putative thiazoline oxidase in *patG1* suggests that the kinetic release scenario may be most likely. It is uncertain at which stage of the biosynthesis the oxidation of thiazoline to thiazole occurs.

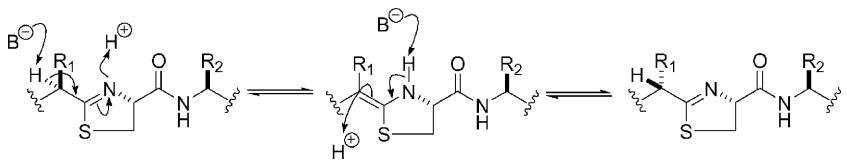
The initial sequence analysis[6] indicates that in the absence of an epimerase in the *pat* gene cluster this process may either be under the control of a novel enzyme or be spontaneous. It is unlikely to be catalysed by a separate racemase producing D-amino acids for incorporation into the patellamides, as the gene cluster suggests that L-amino acids only are incorporated into the patellamides, such that epimerisation must occur at a later stage. Epimerases in non-ribosomal peptide synthetases act by de- and re-protonating the α-carbon to give an equilibrium mixture of both epimers.[13] The downstream domain then selects the D-epimer for condensation to form a peptide bond. In the patellamide biosynthesis the linear peptide is formed first followed subsequently by tailoring steps. If a non-ribosomal peptide synthetase epimerase domain were present in the patellamide gene cluster this would result in a mixture of stereoisomers as no subsequent selectivity step is possible. This phenomenon is not observed suggesting that a different epimerisation process operates.

The spontaneous epimerisation hypothesis is proposed on the basis of experiments performed on related *L. patella* compounds, the lissoclinamides, which contain thiazolines as part of a cyclic heptapeptide.[11] In these, the α-carbon preceding a thiazoline is extremely labile to epimerisation due to the adjacent imine bond (Scheme 2). In lissoclinamide 7, it was observed that if the incorrect (L) stereochemistry was incorporated into the compound at the Val-(thiazoline) α-carbon *via* chemical synthesis, that under treatment with pyridine this centre would epimerise to give the naturally observed epimer of lissoclinamide 7 (Scheme 3).

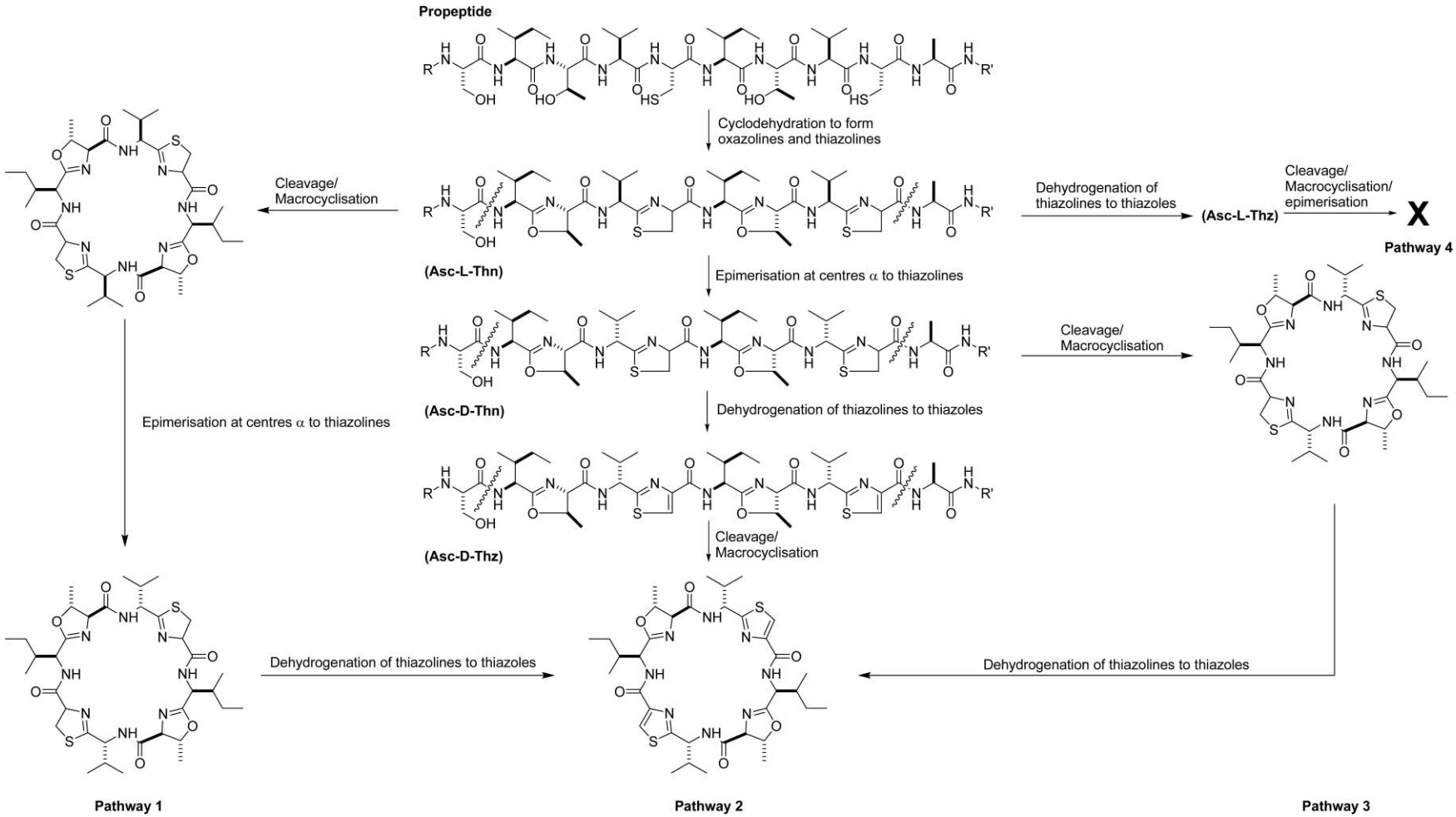


Scheme 3 Equilibration of unnatural epimer of lissoclinamide 7 to the natural epimer at the centre marked with a * occurs in the presence of 5–10 equivalents pyridine at 60 °C.

In the case of the patellamides, we suggest that this process could occur at the thiazoline stage whilst still incorporated into the 71 aa pre-propeptide, prior to oxidation to the thiazole, which fixes the stereochemistry and subsequent macrocyclisation (Scheme 4, Pathway 2). An alternative scenario is that the macrocyclised patellamides are biosynthesised containing L-thiazolines, which then epimerise into the observed (D) geometry followed by enzymic or spontaneous oxidation (Scheme 4, Pathway 1). Enzymic oxidation of the thiazolines to thiazoles would occur under the influence of *patG1*, although spontaneous oxidation of thiazolines to thiazoles has often been proposed for the lissoclinamides, in which identical compounds are observed differing only in oxidation states.[14] An



Scheme 2 Proposed mechanism by which epimerisation at the α-carbon adjacent to the thiazole occurs.



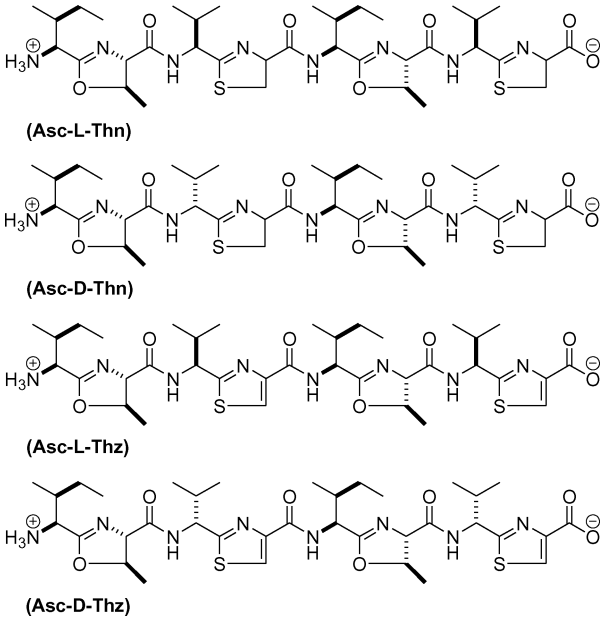
Scheme 4 The three pathways (1–3) considered for the spontaneous macrocyclisation/epimerisation hypothesis and subjected to molecular dynamics simulations. Pathway 4 is deemed unlikely on biosynthetic grounds but has been included in the simulations for completeness.

intermediate route is shown by pathway 3 in Scheme 4, in which epimerisation occurs first to generate D-thiazolines followed by macrocyclisation and oxidation to the thiazoles. A fourth pathway can be proposed (pathway 4 in Scheme 4), in which the epimerisation occurs after the formation of the thiazoles, with macrocyclisation occurring prior or subsequent to this step. Although this pathway has been included in our molecular dynamics simulations, it is felt that this is biosynthetically extremely unlikely. The initial suggestion[6] was that epimerisation occurs during the maturation of the pre-propeptide and prior to proteolysis and macrocyclisation (Scheme 4, Pathway 2).

Schmidt *et al.* suggest that the mature propeptide containing the oxazoline and thiazole rings and two D-stereocentres is cleaved from the propeptide under the influence of *patG2* or *patA* (Scheme 1).[6] *PatD1* could adenylate the linear peptide followed by macrocyclisation, which could be spontaneous, or under enzymic control. Spontaneous macrocyclisation has also been proposed as a possibility for the lantibiotics, related post-translationally modified ribosomal peptides.[8] Studies on several lantibiotics using synthetic precursors have shown that spontaneous macrocyclisation is possible, and that in some cases the naturally observed stereochemistry is obtained.[8] We propose here that this process in the patellamides is spontaneous as the linear octapeptide containing heterocycles is capable of pre-organising itself without outside influence into a conformation suitable for macrocyclisation. This effect is due to the conformational constraints induced by the thiazoline or thiazole and oxazoline rings. This effect of pre-organising of cyclic peptides by including conformational constraints such as heterocycles has been observed in the synthesis of cyclic peptides by Jolliffe and co-workers, and was found to increase macrocyclisation yields up to 99%, even at high substrate concentrations.[15] Chemical syntheses of the patellamides do not utilise this effect because the oxazoline rings are formed last.[16]

For these reasons we decided to simulate the behaviour of the putative linear peptides represented in Scheme 4 pathways 1–4 using molecular dynamics simulations to determine which of these pathways is the most probable. We have investigated the conformational preferences of the patellamides previously under different solvent conditions using a mixture of circular dichroism, NOE-restrained molecular dynamics and unrestrained molecular dynamics calculations.[17]

In this study the four possible linear peptides encompassing the thiazole and thiazoline oxidation states and L and D stereochemistries at the α-centre adjacent to the thiazole/thiazoline were constructed for molecular dynamics in explicit water (Asc L-Thn, Asc D-Thn, Asc L-Thz, Asc D-Thz). Ascidiacyclamide (**8**) sidechains were chosen for simplicity and the macrocycle junction was created where indicated by the *patE* coding sequences. Zwitterionic forms present at the expected physiological pH were used, as well as relevant concentrations of NaCl. Adenylate groups were not included, as the eventual distance between the N and C termini (Fig. 2) would determine whether cyclisation was possible, and not the nature of the activating group. After generating starting conformations using simulated annealing followed by energy minimisation, six simulations of 1000 ps with a 2 fs timestep were executed at 300 K for each of the four linear peptides. The distance between the N and C termini were extracted from the trajectories for each of these simulations and plotted in Fig. 2. For clarification, the actual distance measured during these simulations is represented in Fig. 3.



The simulations for the expected route *via* pathway 2 (Asc D-Thz in Fig. 2) in which epimerisation is followed by dehydrogenation and macrocyclisation show that the N and C termini never get within reasonable bonding distance. Therefore, if this is the actual biosynthetic route, although epimerisation may occur spontaneously, the macrocyclisation must be under enzymic control. This process would then differ from macrocyclisation in non-ribosomal peptide synthetases, in which an enzyme-bound peptide chain is cleaved by a terminal thioesterase and cyclised under control of the same enzyme.[18] In pathway 2 a free fully tailored Asc D-Thz chain would be expected to complex to an enzyme which would activate it and enforce a conformation in which macrocyclisation is possible.

Similarly, pathway 3 (Asc D-Thn in Fig. 2) in which epimerisation is followed by macrocyclisation and dehydrogenation also suggests that macrocycle closure will rarely occur. This then suggests that the epimerisation step must occur as a later step in the biosynthetic process, as both pathways 1 and 4 (Asc L-Thn and Asc L-Thz in Fig. 2) show trajectories in which the C and N termini are frequently within bonding distance. Pathway 4 (Asc L-Thz in Fig. 2) was ruled out on biosynthetic grounds above as this would require an epimerisation of a non-labile α-centre adjacent to a thiazole. In the absence of an epimerase domain and evidence suggesting that spontaneous epimerisation at the α-centre adjacent to thiazoline is facile this suggests that pathway 4 is extremely unlikely. The remaining route, pathway 1, (Asc L-Thn in Fig. 2) in which macrocyclisation occurs on the linear peptide with thiazolines and L stereochemistry at the α-centres adjacent to the thiazolines, followed by spontaneous epimerisation and finally enzymic or spontaneous oxidation of the thiazolines to thiazoles is therefore deemed the most likely. The trajectories from the dynamics simulations indicate that the N and C termini are within bonding distance for a greater proportion of the time. If adenylation of the linear Asc L-Thn occurs under the influence of *patD1*, then macrocyclisation will occur readily. This pathway is in keeping with the macrocyclisation experiments by Jolliffe

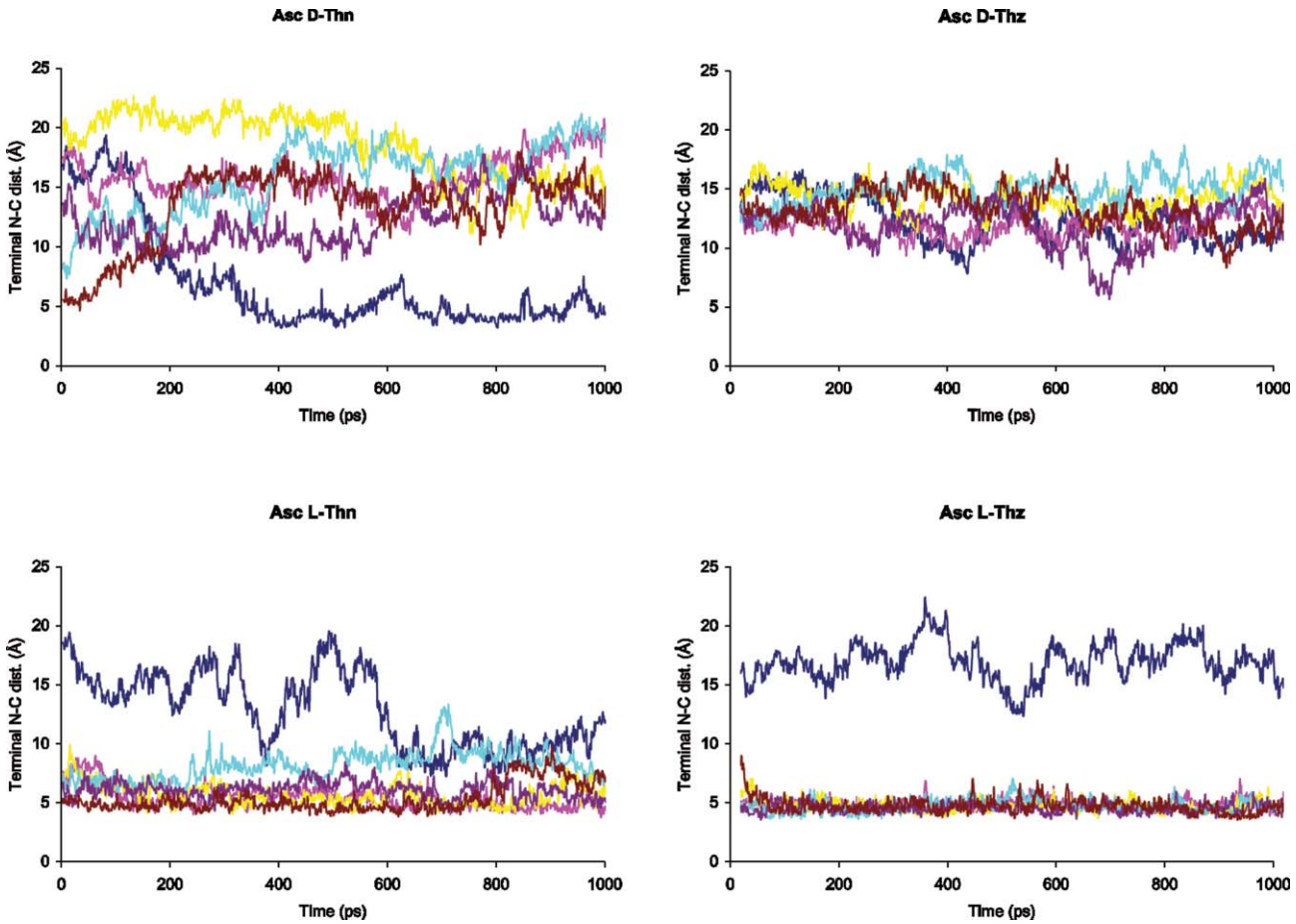


Fig. 2 Terminal N–C distances monitored during a total of six 1000 ps dynamics simulations for all geometries of ascidiacyclamide considered in this study.

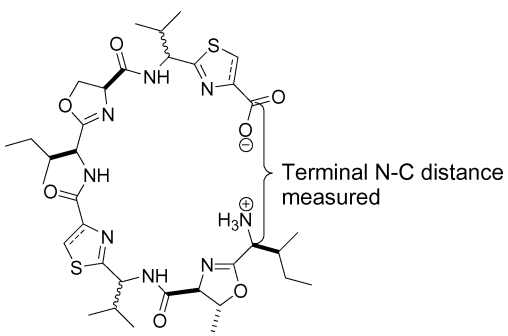


Fig. 3 Actual distance measured during the molecular dynamics simulations represented in Fig. 2.

mentioned above,[15] and Wipf's observation that α-centres adjacent to thiazolines in these macrocycles can spontaneously self organise into the lowest energy epimers (Scheme 3).[11] This then suggests that the oxidation of thiazoline to thiazole is the final step in the biosynthetic process, whether spontaneous or enzymic. The fact that the oxazolines are not oxidised to oxazoles in the patellamides may be due to the fact that this oxidation occurs at such a late stage in the biosynthesis. If the oxidation is spontaneous, then chemical reactivity differences may explain the difference in oxidation states between the two types of heterocycles present. If the process is enzymic at this late stage, then it may be that the oxazolines cannot gain access to the oxidase active site.

Conclusions

The inability of the preferred candidate, Asc D-Thz to close easily suggests that a macrocyclase may be involved, but in the absence of such an enzyme in the sequence analysis this indicates that the enzyme may be very novel, or that the process is spontaneous and occurs *via* a different pathway as suggested by the findings discussed above. Our study is suggestive that pathway 1 in Scheme 4 may be the route *via* which the patellamides are biosynthesised. A testable hypothesis is therefore proposed which can be verified by future detailed biosynthetic studies. The function of the genes in the patellamide gene cluster will only be fully understood after these investigations, but it is postulated that, in parallel with the lantibiotics[8] and microcins,[4] several of these might be involved in metabolite export and host immunity.

Speculation as to the biosynthetic origin of unusual metabolites has been used successfully to explain observed stereochemistries and product distributions. The most notable example is the

postulated electrocyclic reactions leading to the endiandric acids,[19] which were later confirmed by the synthesis of the putative precursor which gave identical product distribution and stereochemistry to that observed in nature.[20] Similarly, the biogenetic hypothesis towards the manzamine alkaloids[21] predicted the existence of the related xestocyclamine/madangamine alkaloids.[22] Speculation on the biosynthetic origin of alga-derived brominated terpenoids has recently been shown to be correct.[23] As can be seen, biogenetic hypotheses of the type in this paper has helped advance our understanding of biosynthetic processes in the past[24] and has assisted in the design of biomimetic syntheses.[25]

Experimental

Sequence analysis

The entire Pfam HMM library[26] was downloaded from the Pfam's FTP site. The library was searched locally using the latest HMMER hidden Markov model software.[27] The search was done using fastA files containing protein translations of all seven individual *Prochloron didemni* genes[6] taken from the GenBank's sequence AY986476.

To confirm the results from the search of the Pfam HMM library, profiles of all adenylation, cyclisation and epimerisation domains from NRPS's were prepared.[28] FastA protein sequences of domains were obtained from the NRPSDB database.[12] Multiple alignments of domains were made using ClustalW.[29] Three profiles were built from multiple alignments of domains by the preparation of a small library that was used in search.

Molecular dynamics simulations

Molecular dynamics (MD) calculations were carried out with the GROMACS v3.2.1 package.[30,31] The modified GROMOS-87 united-atom force field distributed with GROMACS was used in all simulations.[32–35] Aqueous solvation was modelled using the flexible form of the enhanced simple point charge (SPC/E) water model.[36,37] A 125 nm^3 cubic water box was used for solvation of each peptide, requiring ~4200 waters per peptide. In order to better simulate the physiological environment Na^+ and Cl^- ions were added with the *genion* program (supplied as part of GROMACS) so as to produce a concentration of 35 g l^{-1}, with minimum force calculations being used to determine which of the solvent molecules were to be replaced. The peptides were modelled in their zwitterionic forms and the PRODRG server at Dundee University (http://davapc1.bioch.dundee.ac.uk/programs/prodrg/) was used for generation of the required force field topology files.[38]

In order to generate starting configurations for the simulations each peptide–solvent system was subjected to 6 cycles of simulated annealing. The temperature was cycled from 50 K up to 500 K over 250 ps and back down to 50 K over 50 ps so as to allow reasonable sampling of conformational space followed by quenching of favoured conformations. Temperature coupling within each system was performed using the Berendsen method and electrostatics were treated using the particle–mesh Ewald (PME) approach. The time step employed was 2 fs. The resulting trajectories were manually sampled every 300 ps using the *ngmx* GROMACS trajectory viewer. The 6 samples for each derivative were then subjected to energy minimisation and subsequently used as starting configurations for the production runs. These runs were performed at 300 K and lasted for 1000 ps. A 2 fs timestep was used and Berendsen temperature coupling and PME electrostatics were employed. The evolution of the distance between the N-terminal nitrogen and the C-terminal carbon was extracted from the final trajectories using the gOpenMol software.[39]

Acknowledgements

MJ and PFL thank the Leverhulme Trust for funding (F/00 152/N). Bioinformatics work was supported by the grant 058008 (to DH) from the Ministry of Science, Education and Sports, Republic of Croatia. All molecular dynamics calculations were performed under openMosix on the SEERAD funded 64-node RRI/BioSS Beowulf cluster at the Rowett Research Institute, Aberdeen, Scotland with the help of Dr Tony Travis. BFM is the recipient of post-doctoral research fellowship SFRH/BPD/17830/2004 from the Portuguese Ministry for Science, Technology and Higher Education (Foundation for Science and Technology).

References

1 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2004, **67**, 1216.
2 B. M. Degnan, C. J. Hawkins, M. F. Lavin, E. J. McCaffrey, D. L. Parry, A. L. v. d. Brenk and D. J. Watters, *J. Med. Chem.*, 1989, **32**, 1349.
3 A. B. Williams and R. S. Jacobs, *Cancer Lett.*, 1993, **71**, 97.
4 R. Sinha Roy, A. M. Gehring, J. C. Milne, P. J. Belshaw and C. T. Walsh, *Nat. Prod. Rep.*, 1999, **16**, 249.
5 R. Lewin and L. Chang, *Prochloron: A Microbial Enigma*, Chapman and Hall, New York, 1989.
6 E. Schmidt, J. Nelson, D. Rasko, S. Sudek, J. Eisen, M. Haygood and J. Ravel, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 7315.
7 P. Long, W. Dunlap, C. Battershill and M. Jaspars, *ChemBioChem*, 2005, **6**, 1760.
8 C. Chatterjee, M. Paul, L. Xie and W. A. van der Donk, *Chem. Rev.*, 2005, **105**, 633.
9 S. Rebuffat, A. Blond, D. Destoumiex-Garzon, C. Goulard and J. Peduzzi, *Curr. Protein Pept. Sci.*, 2004, **5**, 383.
10 R. S. Roy, P. J. Belshaw and C. T. Walsh, *Biochemistry*, 1998, **37**, 4125.
11 P. Wipf, P. C. Fritch, S. J. Geib and A. M. Sefler, *J. Am. Chem. Soc.*, 1998, **120**, 4105.
12 M. Z. Ansari, G. Yadav, R. S. Gokhale and D. Mohanty, *Nucleic Acids Res.*, 2004, **32**, W405.
13 D. B. Stein, U. Linne and M. A. Marahiel, *FEBS J.*, 2005, **272**, 4506.
14 L. A. Morris, B. F. Milne, M. Jaspars, J. J. Kettenes van den Bosch, K. Versluis, A. J. R. Heck, S. M. Kelly and N. C. Price, *Tetrahedron*, 2001, **57**, 3199.
15 D. Skropeta, K. A. Jolliffe and P. Turner, *J. Org. Chem.*, 2004, **69**, 8804.
16 Y. Hamada, M. Shibata and T. Shioiri, *Tetrahedron Lett.*, 1985, **26**, 6501.
17 B. F. Milne, L. A. Morris, M. Jaspars and G. S. Thompson, *J. Chem. Soc., Perkin Trans. 2*, 2002, 1076.
18 H. D. Mootz, D. Scharzer and H. A. Marahiel, *ChemBioChem*, 2002, **3**, 490.
19 W. Bandanarayake, J. Banfield and D. S. C. Black, *J. Chem. Soc., Chem. Commun.*, 1980, 902.
20 K. Nicolaou, N. Petasis and R. Zipkin, *J. Am. Chem. Soc.*, 1982, **104**, 5560.
21 J. E. Baldwin and R. C. Whitehead, *Tetrahedron Lett.*, 1992, **33**, 2059.
22 J. Rodriguez and P. Crews, *Tetrahedron Lett.*, 1994, **35**, 4719.
23 A. Butler and J. N. Carter-Franklin, *Nat. Prod. Rep.*, 2004, **21**, 180.
24 R. Thomas, *Nat. Prod. Rep.*, 2004, **21**, 224.
25 K. Nicolaou, T. Montagnon and S. Snyder, *Chem. Commun.*, 2003, 551.
26 A. Bateman, E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall and E. L. Sonnhammer, *Nucleic Acids Res.*, 2002, **30**, 276.

27 HMMER-Biological sequence analysis with profile hidden Markov models *Version 2.3.2*, HHMI/Washington University School of Medicine, 1992–2003.
28 S. R. Eddy, *Bioinformatics*, 1998, **14**, 755.
29 D. Higgins, J. Thompson, T. Gibson, J. D. Thompson, D. G. Higgins and T. J. Gibson, *Nucleic Acids Res.*, 1994, **22**, 4673.
30 H. J. C. Berendsen, D. van der Spoel and R. G. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43.
31 E. Lindahl, B. Hess and D. van der Spoel, *J. Mol. Model*, 2001, **7**, 306.
32 X. Daura, B. Oliva, E. Querol, F. X. Aviles and O. Tapia, *Proteins: Struct., Funct., Genet.*, 1997, **25**, 89.
33 A. R. van Buuren, S. J. Marrink and H. J. C. Berendsen, *J. Phys. Chem.*, 1996, **97**, 9206.
34 D. van der Spoel, A. R. van Buuren, D. P. Tieleman and H. J. C. Berendsen, *J. Biomol. NMR*, 1996, **8**, 229.
35 W. F. van Gunsteren and H. J. C. Berendsen, *GROMOS-87 Manual*, 1987.
36 H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, *J. Phys. Chem.*, 1987, **91**, 6269.
37 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren and J. Hermans, *Intermolecular Forces*, ed. B. Pullman, Reidel, Dordrecht, 1981.
38 A. W. Schuettelkopf and D. M. F. van Aalten, *Acta Crystallogr., Sect. D*, 2004, **60**, 1355.
39 D. L. Bergman, L. Laaksonen and A. Laaksonen, *J. Mol. Graphics*, 1997, **15**, 301.

