Kinetics and regioselectivity of peptide-to-heterocycle conversions by microcin B17 synthetase

Peter J Belshaw, Ranabir Sinha Roy, Neil L Kelleher and Christopher T Walsh

Background: The *Escherichia coli* peptide antibiotic microcin B17 (MccB17) contains four oxazole and four thiazole rings introduced post-translationally in the 69 amino acid McbA gene product, an MccB17 precursor, by the microcin B,C,D enzyme complex. Both monocyclic and 4,2-bis-heterocyclic moieties are generated. The enzymatic cyclization involves 14 of the last 43 amino acids of McbA and requires the presence of the first 26 amino acids that function as a specificity-conferring propeptide.

Results: We have constructed maltose-binding protein (MBP)–$McbA_{1-46}$ fusion proteins and have mutagenized the Gly39–Ser40–Cys41 (GSC) wild-type sequence to assess the regioselectivity and chemoselectivity of MccB17-synthetase-mediated heterocycle formation at the first two loci, residues 40 and 41 of McbA. Four single-site and four double-site substrates showed substantial differences in turnover as assessed by western assays, UV–visible spectroscopy and mass spectrometry. Cysteine-derived thiazoles form at a greater rate than serine-derived oxazoles. Formation of bis-heterocycles is sensitive both to composition and sequence context.

Conclusions: The *E. coli* McbB,C,D MccB17 synthetase is the first peptide heterocyclization enzyme to be characterized. This study reveals substantial regioselectivity and chemoselectivity (thiazole > oxazole) at the most amino-terminal bis-heterocyclization site of McbA. The heterocyclization of GSS and GCC mutants of $McbA_{1-46}$ by MccB17 synthetase demonstrates that the complex can efficiently generate tandem bis-oxazoles and bis-thiazoles, moieties not found in MccB17 but present in natural products such as hennoxazole and bleomycin. The observations suggest a common enzymatic mechanism for the formation of peptide-derived heterocyclic natural products.

Address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA.

Correspondence: Christopher T Walsh
E-mail: walsh@walsh.med.harvard.edu

Key words: antibiotics, microcin B17, oxazole, post-translational modification, thiazole

Received: **20 April 1998**
Accepted: **18 May 1998**

Published: **24 June 1998**

Chemistry & Biology July 1998, **5**:373–384
http://biomednet.com/elecref/1074552100500373

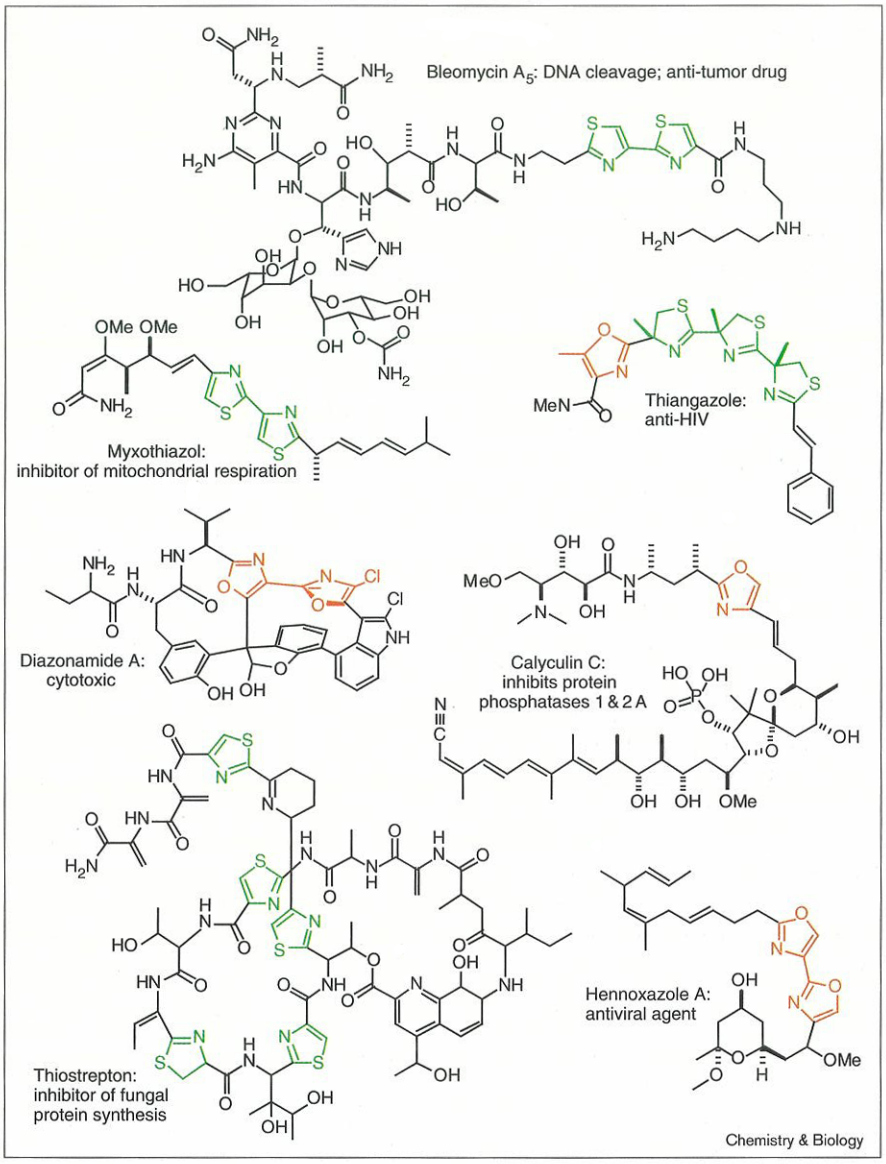
Introduction

A wide variety of natural products contain oxazole and thiazole moieties that are presumably derived from precursor peptides containing Xaa–Ser/Cys/Thr sequences. These natural products exhibit a broad spectrum of activities and many have medicinal properties (Figure 1).

The peptide antibiotic microcin B17 (MccB17) is produced by strains of *Escherichia coli* harboring the plasmid pMccB17 [1]. The plasmid encodes seven genes, *mcbA–G*, that together allow production of MccB17 [2,3]. The recent structural elucidation of MccB17 [4,5] revealed the presence of oxazole, thiazole, 4,2-bis-oxazole–thiazole and 4,2-bis-thiazole–oxazole rings in MccB17, spurring renewed interest in this antibiotic. The ribosomally synthesized peptide product of *mcbA*, pre-MccB17 (hereafter termed McbA, 69 amino acids), is post-translationally modified to convert six glycine, four serine and four cysteine residues into four oxazoles and four thiazoles in pro-MccB17 by MccB17 synthetase, a complex of enzymes encoded by the genes *mcbB,C,D* [6]. Pro-MccB17 is exported from the cell by an ATP-binding-cassette (ABC)-type transporter encoded by the products of *mcbE,F* [2]. The amino-terminal 26 residues of pro-MccB17 are ultimately removed by an as yet unknown protease to produce the mature antibiotic MccB17 [7] (Figure 2). Another gene, *mcbG*, cooperates with *mcbE,F* to provide immunity for producer strains to the effects of MccB17 [2]. MccB17 causes double-stranded DNA breaks in a DNA gyrase-dependent fashion, much like the chemotherapeutic quinolones and coumarins; MccB17 forms its own class of topoisomerase inhibitor, however, as mutations in DNA gyrase that confer resistance to MccB17 do not impart resistance to the quinolones or coumarins and *vice versa* [4].

McbA, the 69 amino acid peptide precursor to MccB17, is composed of several regions including a leader peptide (residues 1–26), a glycine-rich linker (residues 27–38) and a region containing residues that are processed to heterocycles (residues 39–69). MccB17 synthetase (McbB,C,D), the enzyme complex responsible for introducing the eight

Figure 1

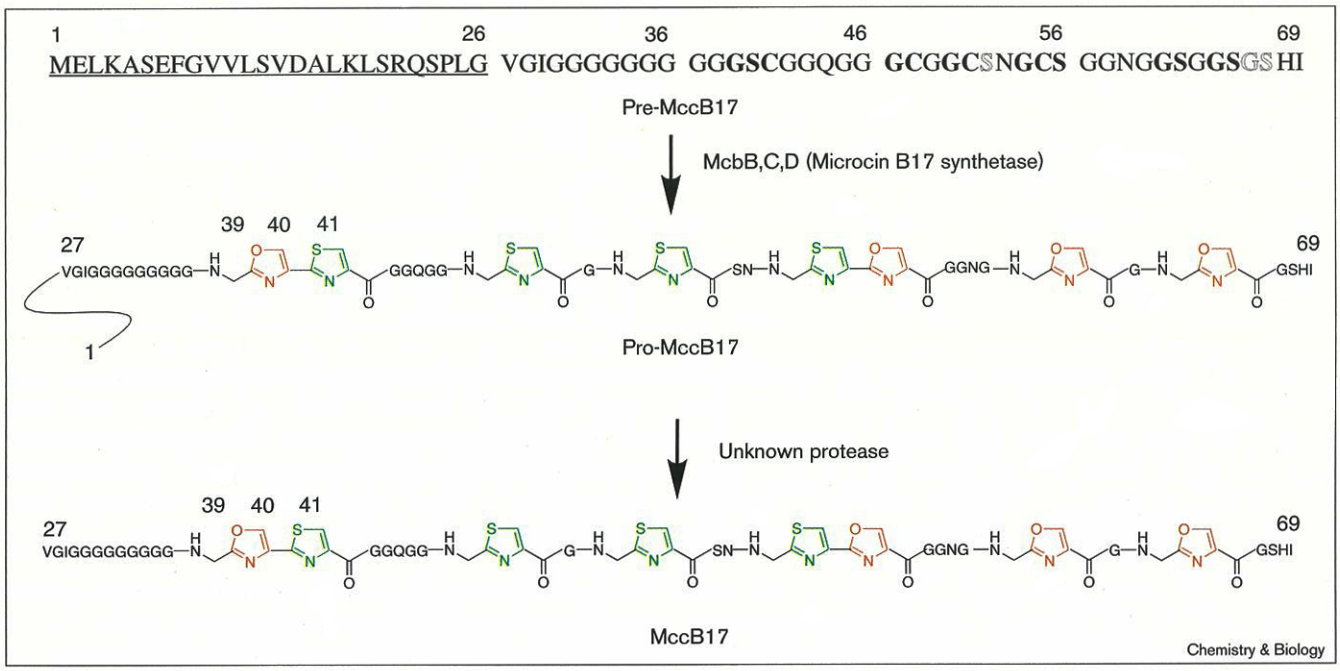


Natural products containing heterocycles derived from serine and cysteine residues. Oxazoles and oxazolines are highlighted in red. Thiazoles and thiazolines are highlighted in green.

heterocycles during post-translational processing of McbA, has recently been cloned, expressed, purified, and shown to possess heterocyclization activity towards substrate peptides [6]. There is an absolute dependence upon the leader peptide for processing activity [8] and investigations of the polyglycine linker have revealed a minimal linker length comprised of five glycine residues to be required for processing [9]. Investigations of the sequence determinants flanking the cyclized cysteine and serine residues revealed an absolute requirement for an upstream glycine residue and a lesser, yet still restrictive, requirement for small hydrophobic residues following the heterocycle [9].

As MccB17 synthetase produces both mono-heterocycles and bis-heterocycles, we sought to further ascertain its ability to generate alternate heterocycles from peptide precursors. Towards this end, we have produced a variety of potential substrate proteins containing residues 1–46 of McbA fused downstream of maltose-binding protein (MBP). Each substrate analog was incubated with MccB17 synthetase and the products were characterized by western immunoblot assay, UV–visible spectroscopy and electrospray ionization fourier-transform mass spectrometry (ESI–FTMS) [10]. The kinetics of processing eight substrate analogs representing the possible mono-heterocycle (GGC, GCG, GGS, GSG) and bis-heterocycle (GSC, GCS, GSS, GCC) pairs are presented. We correlate the formation of mono- and bis-heterocycles with changes in absorbance at 254 nm and 280 nm respectively and mass losses of 20 and 40 Da respectively, as

Figure 2



Post-translational modifications of microcin B17.

determined by high-resolution mass spectrometry. We present a mechanistic model for the processing of bis-heterocycle substrates by MccB17 synthetase.

Results

Substrate preparation

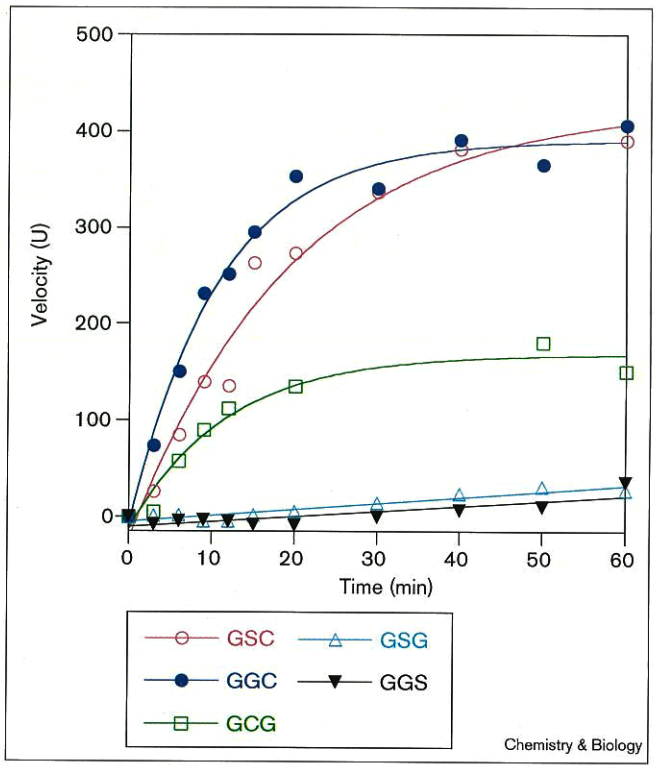
Prior studies on the recognition and processing of heterocycles in McbA by the McbB,C,D complex have indicated the obligate requirement for the propeptide (residues 1–26) of McbA as a specificity-determining element [8]. Thus, even though the first cyclization site is Gly39–Ser40–Cys41, no amino-terminally truncated McbA substrates were processed, either alone or in *trans* with added propeptide, requiring all structure–activity studies to be carried out using minimal substrates containing the propeptide and a linker. We have reported previously on $McbA_{1–46}$ constructs, demonstrating their efficient overproduction and purification as MBP–$McbA_{1-46}$ fusion proteins and the activity of the chimeric fusions as substrates for heterocyclization by the McbB,C,D synthetase complex [9].

MBP–$McbA_{1–46}$ fusion proteins were generated by polymerase chain reaction (PCR) and unique site elimination (USE) mutagenesis [11], expressed, purified in 10–20 mg quantities, and assayed as heterocyclization substrates. Although western assays using a polyclonal antibody raised against mature MccB17 [7] served as an initial assay to detect heterocyclic product in the MBP–$McbA_{1–46}$ fusions, the next set of assays developed for product characterization were matrix-assisted laser desorption ionization time-of-flight (MALDI–TOF) or ESI–FTMS to verify the expected loss of 20 mass units for each oxazole or thiazole moiety formed. To increase accuracy in mass analysis, the MBP tag (43 kDa) was removed via thrombin proteolysis following processing at an engineered site to release the $McbA_{1–46}$ fragments (~4.3 kDa) with a Gly–Ser dipeptide appended to the amino terminus. These product fragments were purified by high performance liquid chromatography (HPLC) [9] and analyzed by mass spectrometry. In all the assays reported here, the affinity-purified MccB17 synthetase with a calmodulin-binding domain fused in frame at the amino terminus of McbB was utilized as the heterocyclizing enzyme complex [9].

Analysis of single heterocycle formation

Wild-type McbA has the sequence Gly39–Ser40–Cys41 at the first bis-cyclization site. To address potential interplay of regioselectivity and chemoselectivity for cyclization at positions 40 and 41, we sought first to deconvolute the problem by constructing the Gly39–Ser40–Gly41 and Gly39–Gly40–Cys41 versions of MBP–$McbA_{1–46}$ to address whether each could still be cyclized and at what rates. To complete the regioselectivity analyses, we also constructed the Gly39–Gly40–Ser41 and Gly39–Cys40–Gly41 isomers to enable direct comparison of the kinetics for cyclization of both serine and cysteine at positions 40 and 41. Figure 3 shows a time course for the western immunoblot assay of the four mutants and the wild-type MBP–$McbA_{1-46}$ (GSC) proteins. We note that the anti-MccB17 antibody is capable

Figure 3



Kinetics of processing of MBP–MccA fusion protein substrates by MccB17 synthetase. Data are shown for the four single-site substrates and the wild-type substrate as determined by the western immunoblot assay. U, arbitrary units.

of detecting the single thiazole products, a clear validation that the western assay in fact detected both mono- and 4,2-fused bis-heterocycles. The initial velocity of the GGC mutant is slightly greater than wild-type GSC and is some twofold faster than the regioisomer GCG. It is not clear why the GCG substrate gives only about half the amount of product as the GGC form, although similar results were observed previously with varying linker lengths ($[Gly]_{10}$–Ser–Cys versus $[Gly]_{11}$–Ser–Cys) [9]. By comparison, cyclization of a single serine to oxazole is much slower than thiazole formation for both GSG (> 100-fold) and GGS (> 1000-fold) substrates as compared with their respective cysteine-containing homologs. The signal in the western blot assay for the two mono-serine substrates, although low, is noticeably and reproducibly above background.

To further characterize products and obtain substrate-to-product ratios, samples from comparable incubations with MccB17 synthetase were purified using HPLC after 40 h of reaction and were subjected to ESI–FTMS. The wild-type GSC sequence yields product 40 Da lower in mass (M-40) after enzymatic processing by MccB,C,D as expected and previously reported [6] for loss of H_2O and 2H during the cyclization, dehydration and dehydrogenation in each heterocycle formation. The two cysteine mutants (GGC and GCG) produce the M-20 product (100% M-20) as do the two serine mutants (GGS, 0.1% M-20; GSG, 1% M-20), although to a minimal extent as anticipated by the western assay. Analysis of the UV–visible spectrum of unprocessed and processed samples (see below) showed general correlations between formation of single heterocycles with an increase in absorbance at 254 nm (A_{254}), and formation of bis-heterocycles with an increase in absorbance at 280 nm (A_{280}), presumably as a result of the extended conjugation in bis-heterocycles [12]. The UV–visible profiles of the assays from single-heterocycle-forming substrates agreed both with the results from the western assay and ESI–FTMS in that the single cysteine substrates showed large increases in A_{254} and no increase in A_{280}, whereas the incubations for single serine substrates had insignificant increases in either A_{254} or A_{280} (data not shown).

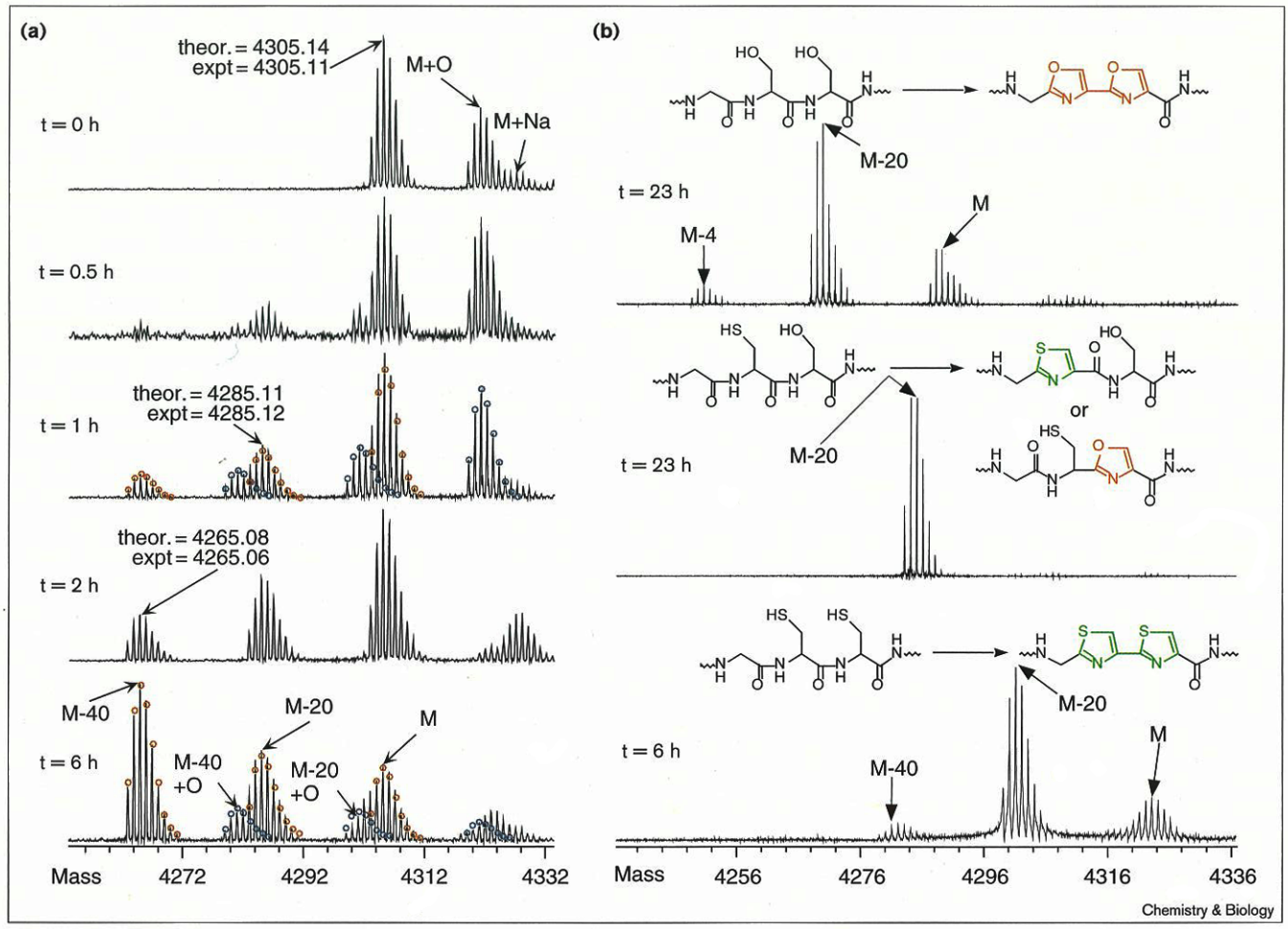
Distributive kinetics of bis-heterocycle formation at the Gly39–Ser40–Cys41 site in microcin B17

The 'wild-type' MBP–$MccA_{1-46}$ fusion protein with the native sequence Gly39–Ser40–Cys41 at the first site of (bis-)heterocyclization was monitored for kinetics of synthetase-mediated ring formation by ESI–FTMS as a baseline for interpretation of the GCS, GCC and GSS mutants noted below. Prior to exposure to the MccB,C,D complex, the GSC substrate yields a molecular weight (M_r) of 4305.11 Da, verifying the predicted substrate composition (theoretical M_r = 4305.14). When sample aliquots were analyzed by ESI–FTMS at various times after addition of enzyme (Figure 4a), a decrease in the abundance of the 4305 species (M) paralleled an increase first in the isotopic distribution for the 4285 species corresponding to the loss of 20.03 Da expected for one heterocycle ring in the intermediate. At later times, a 4265 species was observed (M-40) as the bis-heterocyclic product was produced. The 0.5 hour time point in Figure 4 shows formation of the M-20 intermediate, yet little M-40 product. At 6 h, the M-40 species accumulates and the isotopic peaks corresponding to M and M-20 decrease. Species corresponding to an addition of 16 Da are present in the spectra to varying degrees. We have localized the position of this modification to the leader peptide by MS–MS (data not shown) and thus it is most likely to be caused by air oxidation of an amino-terminal methionine residue to the sulfoxide.

Enzymatic formation of alternate tandem heterocycles

To evaluate the relative efficiency of mono- and bis-heterocyclization of all four paired combinations of cysteine and serine at residues 40 and 41 of MccA, the three mutants potentially capable of bis-heterocyclization, Gly39–Cys40–Ser41, Gly39–Ser40–Ser41 and Gly39–Cys40–Cys41, were also prepared as MBP–$MccA_{1-46}$ fusion proteins. Figure 5 shows data from time course

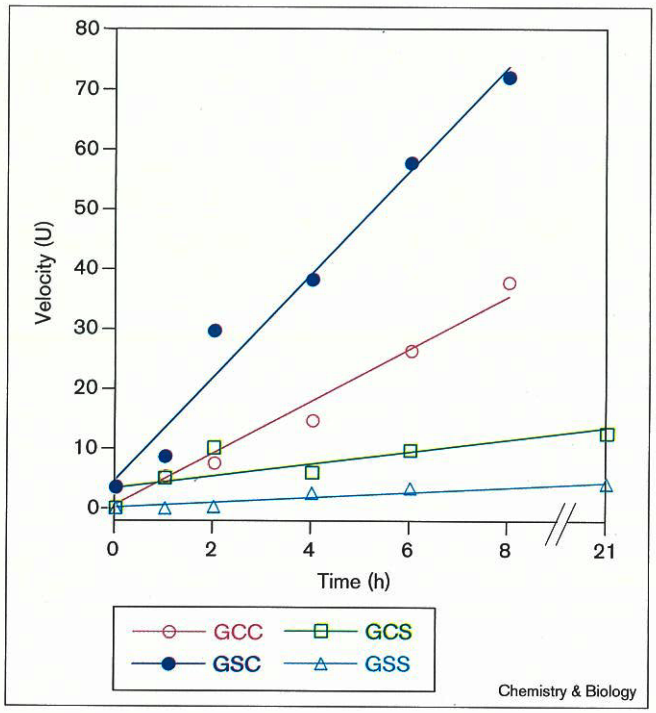
Figure 4



High-resolution mass spectral anaysis of heterocycle formation. **(a)** FTMS time series spectra for $McbA_{1-46}$ with the GSC cyclizable sequence. Theoretical isotopic distributions are indicated for spectra that were corrected to account for partially overlapping species: red circles, M, M-20 and M-40; blue circles, air oxidation products (+16 Da). **(b)** ESI–FTMS spectra for GSS (23 h), GCS (23 h) and GCC (6 h).

western blots for the four substrates. Each substrate showed a time-dependent increase in the antibody signal with the order of reactivity in the series: GSC > GCC > GCS > GSS. Samples of each of the four two-site substrates were treated with MccB17 synthetase, quenched with urea at time points between 0 and 23 h, and cleaved with thrombin to release the $McbA_{1-46}$ fragment, which was subsequently purified using HPLC. As we have noted previously [9], all processed and unprocessed samples containing the leader peptide have nearly identical retention times on C18 reverse-phase columns and thus could not be separated and quantified using HPLC alone. The UV–visible absorbance properties of each McbA substrate fragment was monitored with a diode array detector during elution from the column to determine the relative absorbances at 220, 254 and 280 nm (Figure 6). Each sample peak collected was also analyzed by ESI–FTMS to determine the relative abundances of each species present in the peak. The results are presented in Figure 7 where the percentages of species corresponding to M, M-20 and M-40 are plotted along with the A_{254}/A_{220} and A_{280}/A_{220} ratios. A direct correspondence is observed between heterocycle formation and an increase in A_{254}. Furthermore, formation of 4,2-bis-heterocycles leads to an increase in A_{280}. The differential processing of the four substrates by MccB17 synthetase is remarkable in that GCS is rapidly processed to an M-20 species yet does not go further to the M-40 species (Figure 4b, middle panel). In contrast, GSS is initially processed slowly yet forms a significant amount of M-40 species in the later time points. GCC is processed at an intermediate rate to give M-20 and M-40 species. The absorbance ratios in each case mirror the ESI–FTMS data showing large increases in A_{254}/A_{220} for GCS with little or no increase in A_{280}/A_{220}, indicating the formation of only unconjugated heterocycles. In contrast, for the

Figure 5



Kinetics of processing of MBP–McbA fusion protein substrates by MccB17 synthetase. Data are shown for the three two-site substrates and wild-type substrate as determined by western assay.

other substrates both the A_{254}/A_{220} and A_{280}/A_{220} ratios increase in proportion with the appearance of M-20 and M-40 species, respectively.

Discussion

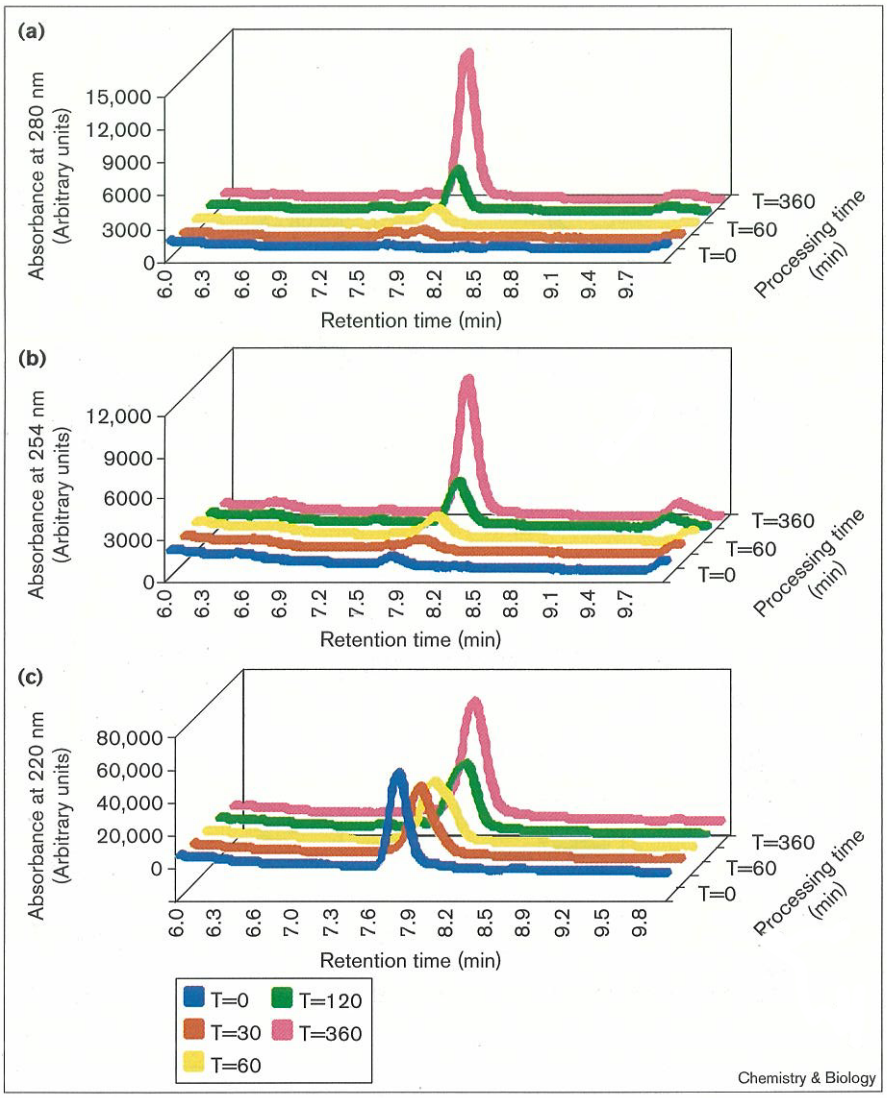
The studies reported here begin to address the catalytic capacity of the McbB,C,D enzyme complex for regioselectivity and chemoselectivity in single- and multiple-ring formations at the first locus of thiazole and oxazole formation, residues 40 and 41, in the *E. coli* antibacterial peptide microcin B17. Prior observations that the synthetase complex has an absolute requirement for the 26 residue propeptide [6,8,13] and a distance dependence in the glycine linker [9] (> 10 glycines required for maximal activity) have constrained our structure–activity investigations to using McbA substrate fragments of at least 46 residues ($McbA_{1-46}$) for each variant. This prerequisite has led us to a mutagenesis strategy rather than a solid-phase synthesis route to prepare and test the serine and cysteine mutants studied here. In turn, we have utilized in-frame maltose-binding protein fusions for ease and scale of purification. The MBP–$McbA_{1-46}$ chimeras are as active as $McbA_{1-46}$ alone, so kinetics were conducted on the fusion proteins.

The comprehensive analysis of single-site and double-site mutants shows that the western assay is suitable for detection of single heterocycles, both thiazole and oxazole, as well as the bis-heterocycles. For mass spectrometric analysis, however, the size of the MBP–$McbA_{1-46}$ fusions (47 kDa) was inconvenient for quantitation of loss of 20 or 40 Da to confirm progressive heterocycle formation. All fusion protein substrates were therefore constructed with a thrombin site between MBP and $McbA_{1-46}$ (see the Materials and methods section) to allow post-incubation proteolysis and HPLC purification of the ~4.3 kDa $McbA_{1-46}$ product fragments for ESI–FTMS analysis. UV analysis of HPLC peaks containing mixtures of partially processed substrates has led to the utilization of gain in absorbance at 254 nm as a marker for formation of mono-thiazole or mono-oxazole and gain of absorbance at 280 nm as an indicator of the increased conjugation that accompanies formation of the fused 4,2-bis-heterocyclic products. A literature search on the UV–visible properties of compounds with 4,2-fused bis-oxazoles/thiazoles conjugated to an amide supported this hypothesis, providing several examples with λ_{max} generally between 270 nm and 290 nm [12], whereas compounds with a single oxazole/thiazole had λ_{max} between 230 and 250 nm [14]. Finally, although MALDI–TOF MS was suitable for qualitative detection of M, M-20, and M-40 species, we often observed differential quantitation of M, M-20 and M-40 peaks, suggesting selective ionization, destruction or retention of material on the matrix, complicating kinetic analysis. These problems were ameliorated by ESI for the kinetic analyses shown in Figures 4 and 6, coupled with high-resolution FTMS, which allows deconvolution of mixtures arising from nonenzymatic sulfoxidation (+16 Da) at Met1 of the McbA substrates or from tightly bound sodium ions (+22 Da).

The results for the two pairs of mono-cysteine- and mono-serine-containing MBP–$McbA_{1-46}$ substrates show both chemoselectivity and regioselectivity by the McbB,C,D synthetase complex as it processes residues 40 and 41 in McbA. The cysteine-containing substrates are rapidly processed by MccB17 synthetase whereas the serine-containing substrates are only slowly turned over. There also appears to be regioselectivity within the cysteine substrates in which the GGC substrate is processed more rapidly and to a fuller extent than GCG. This result might reflect a preference by MccB17 synthetase regarding the exact register of a cyclizable residue relative to the propeptide or sensitivity to residues within the local sequence context (e.g. Gln44) of these substrates. The large differences in rate and product yield between cysteine- and serine-containing substrates probably arises from a requirement for attack of the sidechain nucleophile (e.g. cysteine or serine at residue 40) on the upstream peptide carbonyl to yield a cyclic hemithioaminal (from cysteine) or hemiaminal (from serine) that is subsequently dehydrated and dehydrogenated (by the McbC subunit) to the thiazole or oxazole (Figure 8). The pKa difference

Figure 6

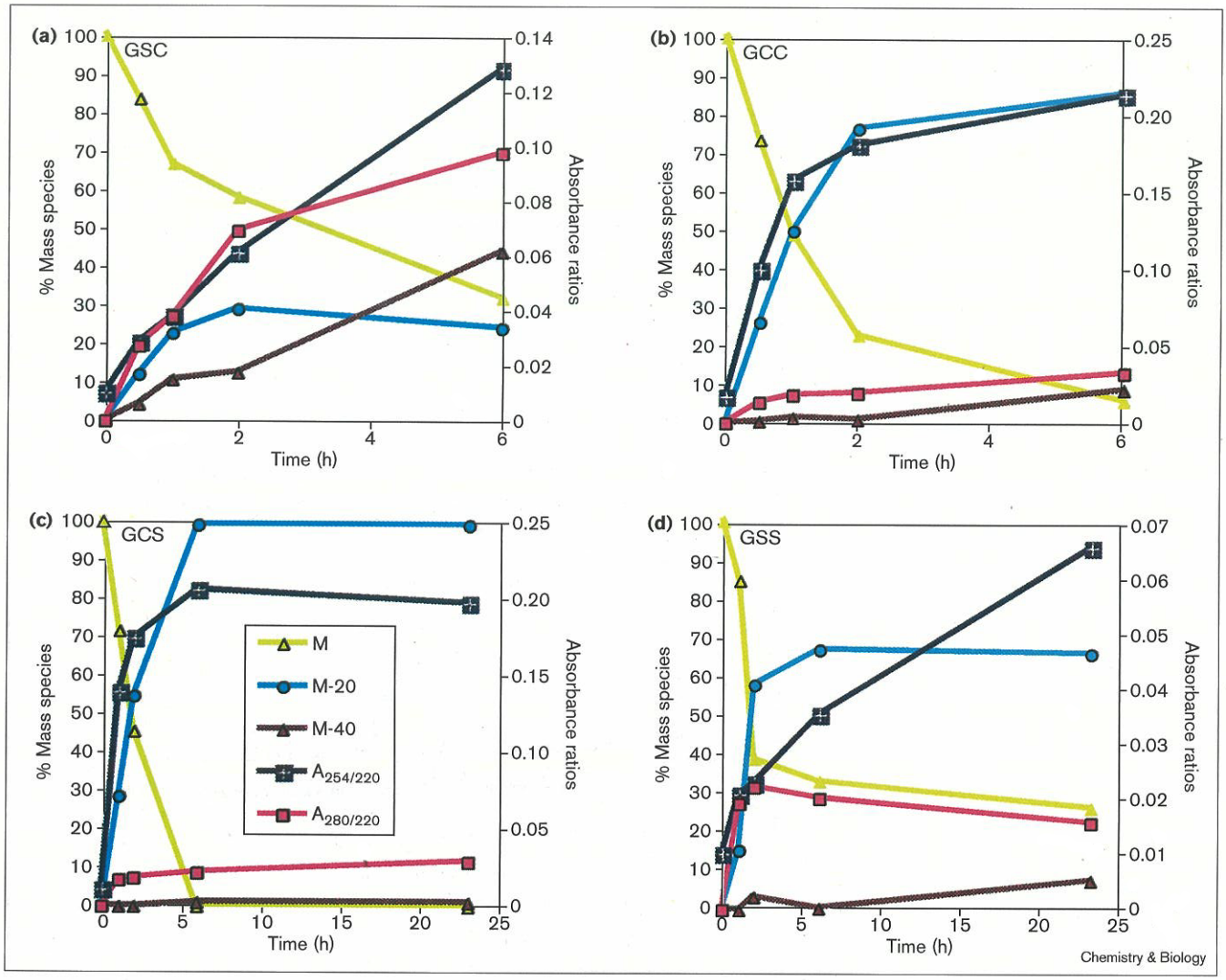
Time series processing of MBP–$McbA_{1-46}$ (GSC) by MccB17 synthetase. Stacked plots of HPLC spectra at **(a)** 280 nm, **(b)** 254 nm and **(c)** 220 nm.



(~13 for serine O–H deprotonation, ~7 for cysteine S–H deprotonation) favors thiolate over alkoxide formation by $> 10^6$ for nucleophile generation. It is unclear to what extent any of the chemical steps of Figure 8 limit enzymatic action of MccB17 synthetase but, even at equal concentrations, thiolates (from cysteine) are better nucleophiles than alkoxides (from serine) so both pKa and nucleophilicity considerations would strongly suggest initial cyclization to be more facile from cysteine than from serine. Notwithstanding the predicted and observed rate differences between cysteine and serine cyclization at residues 40 and 41 in McbA, single serine residues at residues 62 and 65 are also cyclized to oxazoles in mature MccB17 (Figure 2).

Analysis of the four two-site substrates reveals some interesting features. The order of reactivity for formation of bis-heterocycles decreases in the series GSC > GCC > GSS > GCS as determined by the rate of accumulation of M-40 species. In contrast, the order of reactivity in producing the first heterocycle decreases in the order GCC > GCS > GSC > GSS, as measured by disappearance of M species. The observation that the GCS-containing substrate is cyclized rapidly to the M-20 species and subsequently stalls without progressing to the M-40 species is surprising in light of the fact that the other three bis-heterocycle substrates (GSS, GSC, GCC) produce significant amounts of M-40 even though they may produce the M-20 species more slowly. This raises some additional issues about electronic and/or geometric requirements specific to the second heterocycle formed in the tandem 4,2-bis-heterocycle: if the first ring is produced from a Gly–Ser dipeptide in wild-type McbA where there is no interfering R group on the upstream glycine residue, the second ring would have to form in the context of an upstream oxazole.

Figure 7



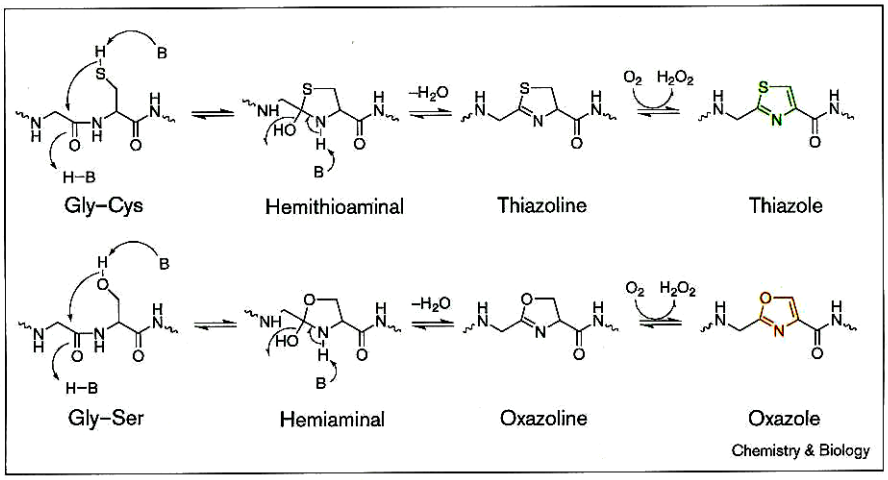
Plot of kinetics of processing of MBP–McbA fusion protein substrates by MccB17 synthetase. The species present in the samples as determined by FTMS are plotted as percentages on the left axis (% M, M-20, M-40). The absorbance ratios (A_{254}/A_{220} and A_{280}/A_{220}) are plotted on the right y axis. **(a)** Gly39–Ser40–Cys41–MBP–$McbA_{1-46}$, **(b)** Gly39–Cys40–Cys41–MBP–$McbA_{1-46}$, **(c)** Gly39–Cys40–Ser41–MBP–$McbA_{1-46}$, **(d)** Gly39–Ser40–Ser41–MBP–$McbA_{1-46}$.

If the nucleophilic attack of a sidechain on the preceding amide carbonyl group is the rate-determining step in the formation of the heterocycle, the order of heterocycle formation in a bis-heterocycle could have a significant electronic effect because the electrophilicity (reactivity towards attack by a sidechain) of the amide carbonyl group would differ in the two cases in whether the carbonyl group is conjugated to a heterocycle or not (Figure 9). Thus the reactivity of bis-heterocyclic substrates may be governed by the electronic properties of not only the nucleophile (cysteine is more reactive than serine) but also the electrophile (conjugated versus unconjugated carbonyl groups). Steric differences in the substrates also have significant effects as clearly evidenced by the absolute requirement for a glycine immediately upstream of the heterocycle-forming sites [9]. Full resolution of this issue will require the synthesis of mono-heterocyclic $McbA_{1-46}$ substrates (e.g., Gly39–oxazole40–Cys41) and initial velocity kinetic analyses.

A closer look at the differing fates of GSC→bis-heterocycle and GCS→mono-heterocycle at residues 39–41 in the context of other residues cyclized in full-length McbA raises additional issues of selectivity. A scan of the McbA sequence (Figure 2) reveals three sites of potential bis-heterocyclization: Gly39–Ser40–Cys41, Gly50–Cys51–Ser52

Figure 8

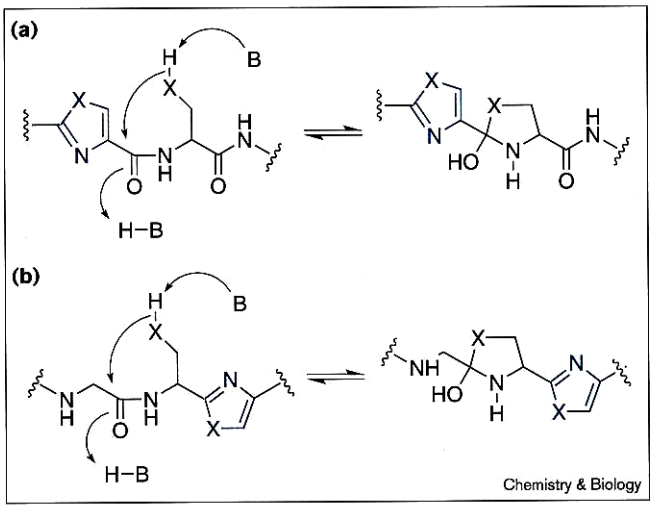
Proposed mechanism for the formation of oxazoles and thiazoles from Gly–Ser and Gly–Cys sequences.



and Gly54–Cys55–Ser56. Two bis-heterocycles are in fact realized: the 4,2-oxazole–thiazole from the 39–41 tripeptide and the 4,2-thiazole–oxazole from the 54–56 sequence. But the Gly50–Cys51–Ser52 sequence, identical to the adjacent Gly54–Cys55–Ser56, is converted only to the thiazole and Ser52 is not cyclized. This fate of Gly50–Cys51–Ser52 mimics the reverse regioisomer Gly39–Cys40–Ser41 studied here and argues for controls on cyclization by determinants beyond the immediate tripeptide. One issue still to be tested for those sites is whether the immediate downstream residue modulates heterocyclization catalytic efficiency. It is Gly42 and Gly66 in the two sequences that bis-heterocyclize but Asn53 where Ser52 fails to cyclize. In our earlier study on the sequence requirements surrounding Gly39–Ser40–Cys41, we found that the mutations Gly42→Asn or Val blocked processing at the mono-heterocycle, whereas Gly42→Ala was processed to the bis-heterocycle [9]. As our substrates all have glycine residues following the bis-heterocyclization site, yet GCSG does not form M-40 species, the local sequence beyond the immediate upstream and downstream residues must also influence the processing of bis-heterocycles. The context surrounding the Gly54–Cys55–Ser56 sequence in McbA is SNG*CSG*GNGG rather than GGG*CSG*GQGG for our substrate analogs. The difference in reactivity observed could also be a result of a change in mechanism in which the local sequence context controls the order of cyclization. For example, Ser56 could cyclize before Cys55 in MccB17 and then continue to the bis-heterocycle, whereas, in the GCS substrate analog, Cys40 might cyclize before Ser41, preventing further progress to the bis-heterocycle (Figure 10). If 4,2-fused bis-heterocycles must be processed in a carboxyl→amino direction to allow formation of the second heterocycle, the strange behavior of the GCS (processing stops at M-20 species) and GCC (fast rate of M-20 formation, slow and incomplete formation of M-40) substrates might be reconciled. By this logic, the M-20 species formed from GCC and GSS could be mixtures of mono-heterocycles at positions 40 and 41, with only the heterocycle at position 41 being further processed to bis-heterocycle. Correspondingly, the M-20 species of GCS and GSC would be thiazoles. These questions can be addressed with MS–MS fragmentation of the M-20 intermediates for the GCS and GSC substrates, which should identify the location of the first heterocycle formed in each substrate.

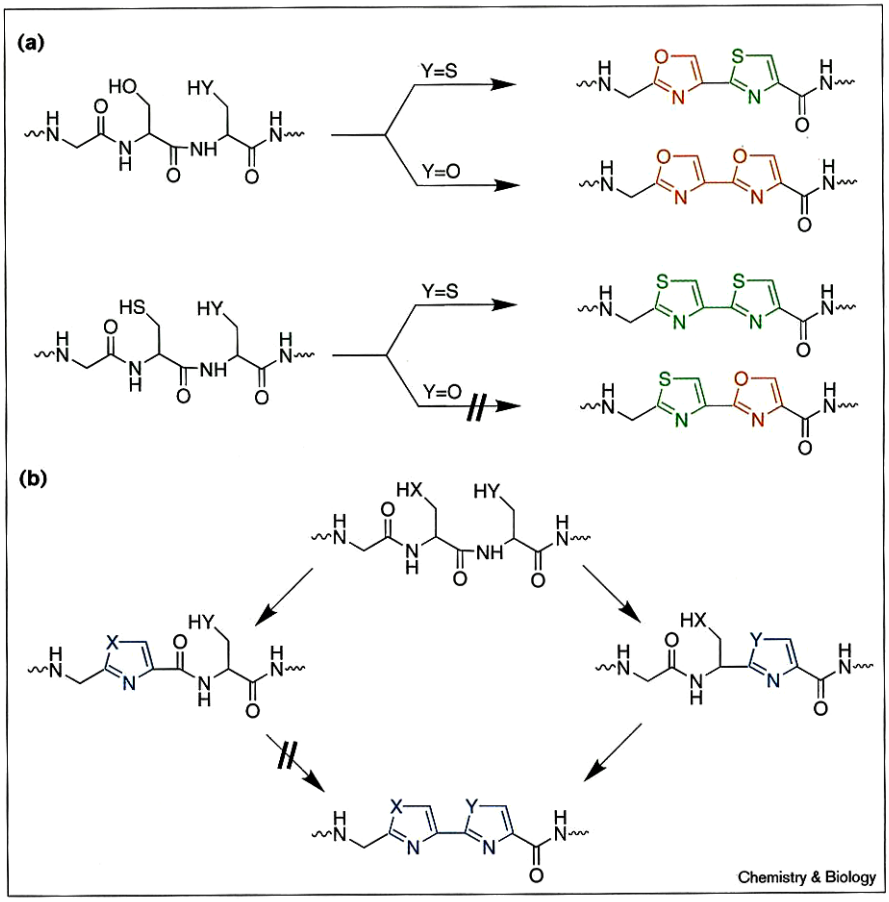
Although it remains to be proven in the –39,40,41– tripeptide locus context whether the first ring forms either from

Figure 9



Differing contexts for cyclization of the second heterocycle of bis-heterocycle pairs in microcin B17. **(a)** The carbonyl group being attacked by a sidechain nucleophile is conjugated with the previously formed heterocycle. **(b)** The carbonyl group under attack is unconjugated. X represents O or S, B represents a general base.

Figure 10



Mechanistic interpretation for differential heterocyclization of serine and cysteine. **(a)** Reactivity differences observed between serine and cysteine in first or second positions during formation of 4,2-fused bis-heterocycles in the context of the sequence Gly39–X40–Y41. **(b)** Formation of the second heterocycle is facilitated when the carbonyl involved in the cyclization reaction is unconjugated.

residues 39,40, or from 40,41, it is clear that the M-20 species is released freely from the enzyme. If it had been the case that the McbB,C,D synthetase made both rings in the conjugated 4,2-bis-heterocycle of McbA without release of monocyclic intermediate, the MS kinetics assay would have noted M and M-40 species but not M-20. Instead, Figure 4a shows an initial buildup of M-20 without detectable M-40, then a decline in M-20 as M-40 bis-heterocycle product forms, consistent with the discrete binding and enzymatic cyclization of both starting substrate and monocyclic intermediate as in Figure 8. For the bis-heterocyclization substrates GCC, GCS and GSS (Figure 4b), the single-ring intermediates (mono-thiazole and mono-oxazole) are produced as discreetly detectable M-20 species, again consistent with distributive processing of substrates and free release of monocyclic intermediates on the way to bis-heterocyclic products. Thus the heterocyclization is distributive rather than processive but it may still be strictly directional (e.g. proceeding from the amino to carboxyl terminus of McbA; residues 39–41, 47, 48, 50, 51, 54–56, 61, 62, 64 and 65 are modified). Among other issues, this distributive characteristic of heterocyclization might mean that one could utilize partially cyclized substrates both to study mechanism and for combinatorial biosynthesis of novel tandem heterocycles.

Lastly, we note that the MccB17 synthetase complex has now demonstrated the ability to make novel combinations of tandem heterocycles, most notably the 4,2-bis-thiazole from GCC and a 4,2-bis-oxazole from the GSS substrate. The bis-thiazole moiety is the signature sequence of the DNA-cleaving antitumor antibiotic bleomycin, widely used in cancer chemotherapy, whereas the 4,2-bis-oxazole is found in such natural products as hennoxazole (Figure 1). The bis-thiazole of bleomycin has been demonstrated to intercalate between base pairs of DNA and is undoubtedly the DNA-targeting portion of this anti-cancer drug [15] This suggests, *inter alia*, that the two 4,2-bis-heterocycles in the wild-type microcin B17 may be DNA-targeting groups in blocking the action of DNA gyrase [16] and that construction of the Ser40→Cys mutation in the full-length McbA gene would be worthwhile to evaluate any change in microcin B17 antibiotic activity. With the gain in knowledge of the local sequence requirements of substrates for MccB17

synthetase, libraries of peptides containing tandem heterocycles could be generated through combinatorial biosynthesis and screened for bioactivity.

Significance

Peptide-based heterocycles are constituents of a diverse array of biologically active natural products, in which the thiazole and oxazole moieties serve as recognition determinants. The post-translational enzymatic cyclization of peptide precursors alters backbone connectivity and imposes architectural scaffolding elements in the products that permit selective interactions with such macromolecular targets as DNA (bleomycin), rRNA (thiostrepton) and proteins (e.g., DNA gyrase for microcin B17). Analysis of the regioselectivity and chemoselectivity of microcin B17 synthetase to carry out both monocyclizations and tandem cyclizations gives insight into the chemical constraints and the molecular logic of the enzymatic machinery that might be generally applicable to this class of natural products. These results also have implications for efforts at combinatorial biosynthesis of heterocycle-containing compounds by microcin synthetase.

Materials and methods

Generation of MBP–McbA$_{1-46}$ fusion proteins

DNA constructs for overexpression of the MBP–McbA$_{1-46}$ mutants reported here were generated as described previously [9] by PCR and USE [11] mutagenesis. The proteins were overproduced in *E. coli* and purified by affinity chromatography on amylose beads (NEB). Primers for USE mutagenesis: 5′-CTGGAATGCTGTTTTACCAGGTATCGCAGTGGTG-3′ (*Sma*I removal); 5′-CCCATACAATCGGTAGATTGTCGCAC-3′ (*Cla*I removal); 5′-TGGAGGTAGCAGCGGTGGCC-3′ (GSS, plasmid pMSS15); 5′-TGGAGGTTGCTGCGGTGGCC-3′ (GCC, pMSS17); 5′-TGGTGGAGGTTGCAGCGGTGGCCAAGG-3′ (GCS, pMSS18). Primers for PCR mutagenesis: 5′-AAGCTTAGCTTTCTTGGCCACCGCAGCCACCTCC-3′ (GSG, pMSS36); 5′-AAGCTTAGCTTTCTTGGCCACCGCAGCCACCTCC-3′ (GGC, pMSS37); 5′-AAGCTTAGCTTTCTTGGCCACCGCCGCAACCTCCA-3′ (GCG, pMSS49); 5′-AAGCTTAGCTTTCTTGGCCACCGCTGCCACCT-3′ (GGS, pMSS48).

Processing and analysis of substrates by MccB17 synthetase

Purified substrates (50 μM) were incubated in 100 μl reactions containing MccB17 synthetase (0.2 mg/ml), Tris-Cl (50 mM, pH 7.5), NaCl (150 mM), $MgCl_2$ (2 mM) and ATP (2 mM) and DTT (10 mM) at 37°C. After varying times, the reactions were quenched with urea (6 M, 20 μl) and analyzed by western immunoblotting with anti-microcin B17 antibodies as previously described [9]. For the western assays of the single-site mutants presented in Figure 3, the substrate concentration was 20 μM, hence the rate of reaction appears greater for the GSC assay results in Figure 3 as compared with Figure 5, due to substrate inhibition in the latter.

HPLC purifications

Quenched reactions were cleaved with thrombin (NEB, 3 units) and purified by reverse-phase chromatography (Vydac, C18, 5μ, 300Å, 4.5 × 250 mm) using a Beckman System Gold HPLC, eluting with a H_2O/acetonitrile (ACN) gradient (30-42% ACN over 12 min at 1.5 ml/min). Peaks were detected on line with a Beckman 168 diode array spectrophotometer.

Mass spectrometry

HPLC fractions were dried under vacuum and redissolved in 10–25 μl of 50:49:1 (H_2O/ACN/acetic acid). Samples (3 μl) were loaded into a nanoelectrospray [17] tip (New Objective, Cambridge, MA) with a 2 or 4 μM orifice; a potential difference of 600 to 1500 V between the tip and the mass spectrometer inlet established electrospray ionization at a 20–100 nl/min flow rate. The resulting ions were guided through four stages of differential pumping into the Infinity Cell (< 10^{-9} Torr) of a 4.7 Tesla APEX II Fourier-transform MS (Bruker, Billerica, MA) for excitation and detection. FTMS instrumentation and operation are described in detail elsewhere [18,19]. Spectra were stored as 256 or 512 K data sets and were calibrated externally using bovine ubiquitin. Theoretical isotopic distributions were generated using Bruker's XMASS software. All relative molecular weight (M_r) values reported are for the peptide's most abundant isotopic peak; for the 48 residue substrates presented here, this peak is 2 Da heavier than the monoisotopic value (e.g., two ^{13}C atoms in the molecule). Species of similar mass (e.g., the +16 oxidation peak of a −20 species, net $\Delta Mr = -4$ Da from the precursor) have partially overlapping isotopic distributions. Component ratios in such cases were determined by correcting isotopic peak abundances of the heavier component by amounts determined from extrapolation of a scaled theoretical distribution of the lighter component. These ratios should be accurate to within 10%, assuming no difference in ionization efficiency for heterocycle-containing peptides.

Acknowledgements

We thank Jill Milne for helpful discussions. We thank Frank Laukien, Gary Kruppa, and Paul Speir of Bruker Daltonics for providing access to their FTMS instrument and for assistance with data collection. This research was supported by NIH Grant GM 20011 to C.T.W. P.J.B. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. R.S.R. is a Parke-Davis Fellow of the Life Sciences Research Foundation. N.L.K is a NIH Fellow and has also been supported by the Armenise Foundation. This investigation has been aided by a grant from the Jane Coffin Childs Memorial Fund for Medical Research.

References

1. San Millan, J.L., Hernandez-Chico, C., Pereda, P. & Moreno, F. (1985). Cloning and mapping of the genetic determinants for microcin B17 production and immunity. *J. Bacteriol.* **163**, 275-281.
2. Garrido, M.C., Herrero, M., Kolter, R. & Moreno, F. (1988). The export of the DNA replication inhibitor microcin B17 provides immunity for the host cell. *EMBO J.* **7**, 1853-1862.
3. San Millan, J.L., Kolter, R. & Moreno, F. (1985). Plasmid genes required for microcin B17 production. *J. Bacteriol.* **163**, 1016-1020.
4. Yorgey, P., *et al.*, & Kolter, R. (1994). Posttranslational modifications in microcin B17 define an additonal class of DNA gyrase inhibitor. *Proc. Natl Acad. Sci. USA* **91**, 4519-4523.
5. Bayer, A., Freund, S. & Jung, G. (1995). Post-translational heterocyclic backbone modifications in the 43-peptide antibiotic microcin B17. Structure elucidation and NMR study of a 13C,15N-labelled gyrase inhibitor. *Eur. J. Biochem.* **234**, 414-426.
6. Li, Y.M., Milne, J.C., Madison, L.L., Kolter, R. & Walsh, C.T. (1996). From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. *Science* **274**, 1188-1193.
7. Yorgey, P., Davagnino, J. & Kolter, R. (1993). The maturation pathway of microcin B17, A peptide inhibitor of DNA gyrase. *Mol. Microbiol.* **9**, 897-905.
8. Madison, L.L., Vivas, E.I., Li, Y.M., Walsh, C.T. & Kolter, R. (1997). The leader peptide is essential for the post-translational modification of the DNA-gyrase inhibitor microcin B17. *Mol. Microbiol.* **23**, 161-168.
9. Sinha Roy, R., Belshaw, P.J. & Walsh, C.T. (1998). Mutational analysis of posttranslational heterocycle biosynthesis in the gyrase inhibitor microcin B17: distance dependence from propeptide and tolerance for substitution in a GSCG cyclizable sequence. *Biochemistry* **37**, 4125-4136.
10. McLafferty, F.W. (1994). High-resolution tandem FT mass spectrometry above 10 kDa. *Accounts Chem. Res.* **27**, 379-386.
11. Deng, W.P. & Nickoloff, J.A. (1992). Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. *Anal. Biochem.* **200**, 81-88.
12. Brookes, P., Clark, R.J., Fuller, A.T., Mijovic, M.P.V. & Walker, J. (1960). Chemistry of Micrococcin P. Part III. *J. Chem. Soc.* 916-925.
13. Sinha Roy, R., Kim, S., Baleja, J.D. & Walsh, C.T. (1998). Role of the microcin B17 propeptide in substrate recognition: solution structure and mutational analysis of McbA$_{1-26}$. *Chem. Biol.* **5**, 217-228.

14. Kovacs, L., Herczegh, P., Batta, G. & Farkas, I. (1987). Two acyclic analogues of 2-β-D-ribofuranosyl thiazole-4-carboxamide (Tiazofuran). *Heterocycles* **26**, 947-960.
15. Kane, S.A., Natrajan, A. & Hecht, S.M. (1994). On the role of the bithiazole moiety in sequence-selective DNA cleavage by Fe-bleomycin. *J. Biol. Chem.* **269**, 10899-10904.
16. Vizan, J.L., Hernandez-Chico, C., del Castillo, I. & Moreno, F. (1991). The peptide antibiotic microcin B17 induces double-strand cleavage of DNA mediated by *E. coli* DNA gyrase. *EMBO J.* **10**, 467-476.
17. Wilm, M. & Mann, M. (1996). Analytical properties of the nanoelectrospray ion source. *Anal. Chem.* **68**, 1-8.
18. Cassady, C.J., Wronka, J., Kruppa, G.H. & Laukien, F.H. (1994). Deprotonation of multiply protonated ubiquitin ions. *Rapid Commun. Mass Spectrom.* **8**, 394-400.
19. Kofel, P., Allemann, M., Kellerhals, H.P. & Wanczek, K.P. (1989). External trapped ion-source for ion cyclotron resonance spectrometry. *Int. J. Mass Spectrom. Ion Processes* **87**, 237-247.

Because *Chemistry & Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published via the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cmb – for further information, see the explanation on the contents pages.