The reversible macrocyclization of Tyrocidine A aldehyde: a hemiaminal reminiscent of the tetrahedral intermediate of macrolactamization†

Sebastian Enck, Florian Kopp,‡ Mohamed A. Marahiel* and Armin Geyer*

Received 25th August 2009, Accepted 30th October 2009
First published as an Advance Article on the web 27th November 2009
DOI: 10.1039/b917549k

In spite of the important role of peptide macrocyclizations for the generation of conformationally constrained biological ligands, our knowledge of factors that determine the inclination of a substrate to cyclize is low. Therefore, methods that give access to the thermodynamic characterization of these processes are desirable. In this work, we present the isosteric substitution of the amide ligation site of a cyclopeptide by an imine. Applied to the decapeptide antibiotic Tyrocidine A (TycA), the reversible cyclization of the linear aldehyde TycA–CHO resulted in the unexpectedly stable hemiaminal Ψ[CH(OH)NH]–TycA, which is equivalent to the tetrahedral intermediate of macrolactamization, and which is observed for the first time in a peptidic structure. On the basis of NMR spectroscopy and molecular modeling, we discuss the observed high stereoselectivity of hemiaminal formation, as well as its reluctance to be dehydrated to the imine. As required for thermodynamic analysis, it is possible to establish a pH- and temperature-dependent cyclization equilibrium, which allows determination of the entropy loss of the peptide chain, and quantification of the extent of preorientation of the cyclization precursor.

Introduction

Macrocyclization induces conformational restriction, which can increase biological activity and selectivity of protein ligands.[1] This is demonstrated by the multitude of macrocyclic natural products synthesized by complex enzyme machineries,[2] and by the great effort of medicinal chemists to establish suitable methods for the cyclization of long molecular chains.[3,4] However, concerning the propensity of a substrate to ring closure, one must rely on a mainly qualitative understanding of the biological processes and on the empirical approach to synthetic macrocyclizations. A prerequisite for a better understanding is thermodynamic analysis, which could give the free energy difference between linear and cyclic species, characterizing the inclination of a linear precursor for cyclization independently of the activation method. Thermodynamic data can then be used to validate computational predictions of the cyclization processes and can help to systematically improve synthetic macrocyclizations, which often suffer from low yields and thus become the bottleneck of total syntheses. Thermodynamic studies, however, are hampered by the irreversibility of the macrocyclization processes. The nostocyclopeptides A1 and A2 produced by the terrestrial cyanobacterium *Nostoc* sp. ATCC53789 stand out from all other known cyclopeptides due to their unique backbone imino linkage.[5] Recently, we demonstrated that this structural feature gives the required reversible cyclization behavior and allows the estimation of the entropy balance of macrocyclic ring closure, which can be quantitatively explained by the amount of released water and the number of restricted rotational degrees of freedom in the peptide backbone.[6] Considering the great number of macrocyclic compounds of natural and synthetic origin, it is desirable to expand these possibilities given by reversible cyclization processes.

In this study, we present a strategy to access the thermodynamics of peptide macrocyclizations. The amide bond at the ligation site of a cyclopeptide is isosterically substituted by a reversibly closing imino functionality (Fig. 1). By variation of temperature and pH, we control the ring-chain equilibrium of the 30-membered macrocycle of an accordingly modified Tyrocidine A (TycA), which is significantly larger than the 21-membered rings formed by the nostocyclopeptides. As the entropy balance of the cyclization process can be estimated, this strategy allows us to quantify the extent of preorientation and to identify determinants that play a role in the inclination of a substrate to cyclize.

Results and discussion

Cyclization studies

The well-studied decapeptide antibiotic TycA appeared to be a promising candidate for the substitution survey because of its high degree of preorganization, and its ability to cyclize spontaneously and without the need of an enzyme or an activation reagent in aqueous solution.[7] In order to introduce reversibility into the irreversible process of native TycA cyclization, we synthesized an analog of the linear TycA cyclization precursor with an aldehyde instead of a carboxylic acid function at the C-terminus (TycA–CHO). While in TycA biosynthesis the macrocyclic amide is closed between D-Phe1 and Leu10, cyclization of the aldehyde precursor should result in reversible formation

Fachbereich Chemie, Philipps-Universität Marburg, Hans-Meerwein-Strasse, D-35032 Marburg, Germany. E-mail: geyer@staff.uni-marburg.de

† Electronic supplementary information (ESI) available: Peptide synthesis and purification, characterization of all compounds by HR-ESI-MS, detailed description of NMR measurements, spectral data, details of pH/temperature dependence of the macrocyclization equilibrium, as well as molecular modeling data.

‡ Present address: *The Scripps Research Institute, Cravatt Laboratory, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.*

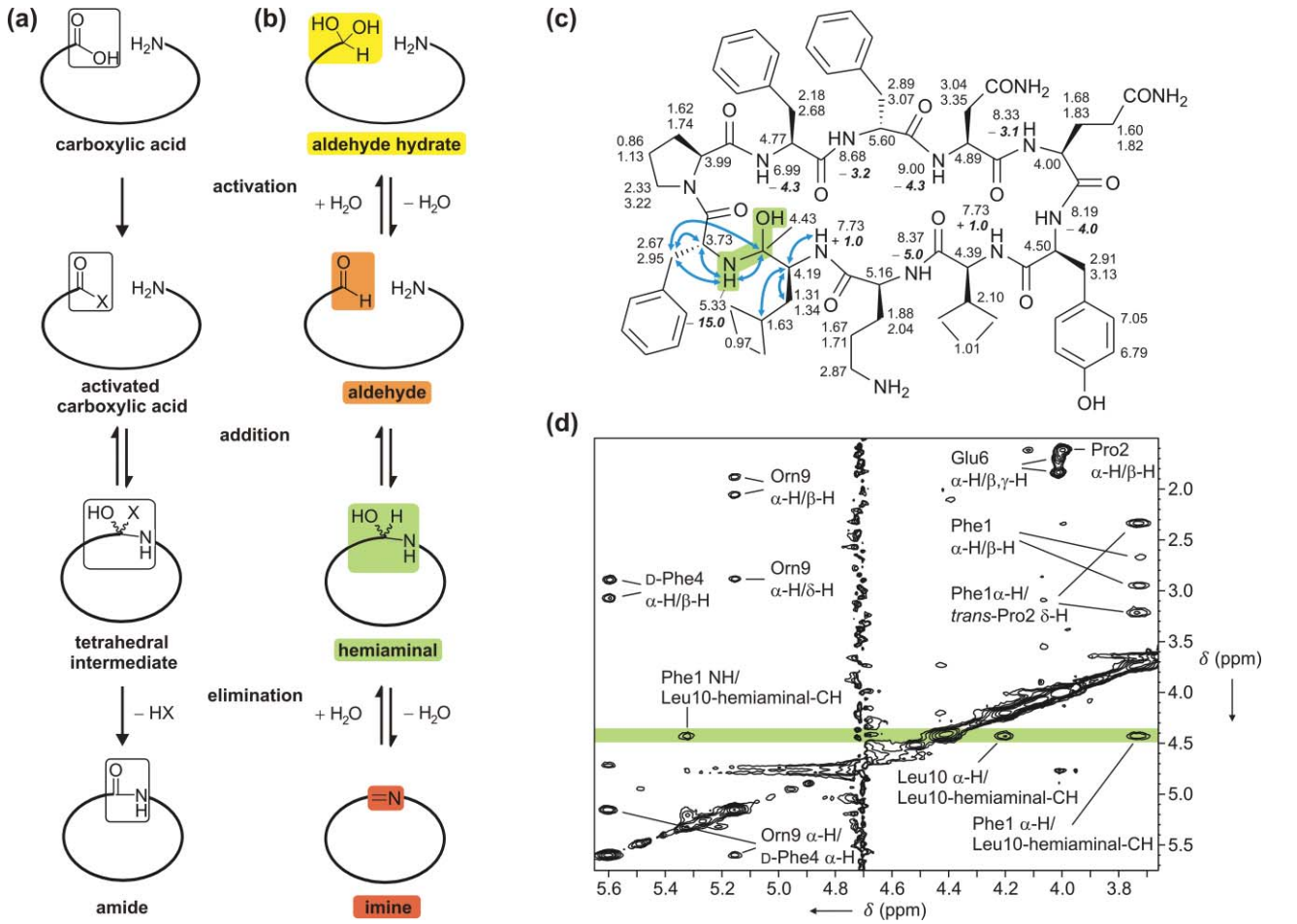

Fig. 1 Amides condense irreversibly in aqueous solution (a) while all steps of macrocyclic imine generation are reversible (b) and thus allow for thermodynamic analysis. Both pathways have in common the same sequence of intermediates, as the substrate first has to be activated to react with an amine to form a tetrahedral species, which subsequently eliminates to yield the amide or the imine. While the amide is irreversibly generated from a tetrahedral intermediate, the imine (red) is in equilibrium with a hemiaminal (green). (c) One main species out of four possible diastereomers is formed in the reversible cyclization of TycA–CHO. All assigned ^{1}H chemical shifts are written next to their respective positions. The temperature gradients of the NH proton signals (in ppb K^{-1}) are assigned with bold numbers, and the ^{1}H–^{1}H-TOCSY correlations in the hemiaminal region are marked by blue arrows. (d) ROESY region (600 MHz, 300 K, KH_2PO_4/H_3PO_4 buffer, H_2O/D_2O 5 : 1 with 0.19 mol L^{-1} SDS-d_{25}, pH 9.0) of cyclized TycA–CHO showing correlations in the hemiaminal region (green bar), the *trans*-Pro2 configuration and the β sheet.

of the TycA imino analog cyclo(Asn5–Gln6–Tyr7–Val8–Orn9–Ψ[C=N]Leu10–D–Phe1–Pro2–Phe3–D–Phe4).[8]

An aqueous solution of the precursor TycA–CHO was subjected to various temperature and pH conditions, enabling us to study the cyclization equilibrium. Despite the potential presence of several linear aldehydes and aldehyde hydrates together with diastereomeric macrocyclic imines, as well as dimers and oligomers, the quantities of the dominant contributing species can be clearly identified from the ^{1}H NMR. Unexpectedly, at pH 9.0 no cyclic imine is observed, but the hemiaminal (Ψ[CH(OH)NH]–TycA) dominates the cyclization equilibrium (Fig. 1c). The cyclization product of TycA–CHO is the first example of a macrocyclic peptide with a hemiaminal, which was expected to be unstable in aqueous solution.

The ten amino acid sequence ($4n$ + 2 residues) of TycA has a high antiparallel amphipathic β sheet content[9–11] and is inclined to self-aggregate in aqueous solution, a fact which hampers its handling. The combination of perdeuterated sodium dodecyl sulfate (SDS-d_{25}) micelles and aqueous phosphate buffer prevented precipitation when the pH was increased from 3.0 to 9.0 (see ESI† for details). As in the case of the nostocyclopeptides, at pH 3.0 only linear species are observed and the C-terminal aldehyde is mostly hydrated (amount of aldehyde <10%).[6] The appearance of several low-intensity sets of signals may also result from the epimerization of the aldehyde, as it was observed in the case of the nostocyclopeptides, as well as from the Pro2 *cis*/*trans* isomerism.

A ^{1}H NMR spectrum recorded 20 min after raising the pH to 9.0 indicated the formation of one main product and only traces (<5%) of the linear species. This fast response of TycA–CHO to pH increase, as well as the high signal dispersion of the new set of signals, showed that cyclization had occurred. The temperature gradients of the NH protons are characteristic for the Tyc macrocycles, and the strong long-range D-Phe4 α-H/Orn9 α-H NOE indicates the formation of the β sheet. A covalent linkage of D-Phe1 and Leu10 is proven by NOESY and TOCSY spectra, but no imine signals of N=CH, which are expected at

chemical shifts of approx. 7 ppm (^{1}H) and 165 ppm (^{13}C), were detectable in the ^{1}H and 2D NMR spectra. In fact, the spectral information identifies a hemiaminal linkage (Ψ[CH(OH)NH]–TycA), with the hemiaminal-CH signal at 4.43 ppm and the NH signal at 5.33 ppm, respectively (Fig. 1d).

As the α position of the Leu10 aldehyde is subjected to racemization and a new stereogenic center is generated by formation of Ψ[CH(OH)NH]–TycA, four possible diastereomers can emerge from the cyclization reaction. The presence of only one cyclic main species, however, demonstrates that cyclization proceeds along a single reaction path, and that the hemiaminal is incorporated into a structurally optimized cyclic structure, which can be dominated by a distinct main stereoconfiguration. This remarkable behavior is reminiscent of the nostocyclopeptides, which also deracemize upon cyclization, yielding stereopure macrocyclic *E*-configured imines from a mixture of *S*- and *R*-configured aldehydes.[6]

Stability of the hemiaminal

An important question concerns the origin of the unprecedented stability of the hemiaminal Ψ[CH(OH)NH]–TycA. The main difference to the corresponding imine Ψ[C=N]–TycA is the sp^3 hybridization and the presence of a hydroxyl function, which is stabilized as part of the hydrogen bonding network of the antiparallel TycA β sheet, thereby suppressing imine formation. If the hemiaminal oxygen acts as an electron donor for Phe3-NH (temperature gradient: –4.3 ppb K^{-1}), a hydrogen bond analogous to the Phe3-NH/Leu10-CO interaction in native TycA is present, though the hemiaminal oxygen is expected to be weaker in its electron-donating ability than the amide oxygen. Of course, the conditions applied in this case are different, as TycA is investigated in a membrane mimetic environment in water, while DMSO and MeOH have been used in earlier works.[12] However, the similar pattern of temperature gradients $\Delta\delta/\Delta T$ of the NH protons (Fig. 1c), which give information on the antiparallel β sheet, indicates that, in spite of the modifications of TycA and of its environment, the β sheet with the four hydrogen bonds is maintained.[13]

The sp^3 hybridized carbon of the hemiaminal resembles the geometrically similar tetrahedral intermediate of amide formation (Fig. 1). Hence, Ψ[CH(OH)NH]–TycA, for the first time, allows the observation of the carbonyl analog of this species formed during a macrocyclic amide formation. In previous studies concerning the cyclization process of native TycA, it was postulated that the tetrahedral intermediate may be involved in a stabilizing hydrogen bridge to Phe3-NH, which in turn is a result of the pronounced precursor preorganization.[14] The unexpected hemiaminal stability fits into this picture as it shows that the preorganization of the linear chain brings together both termini and thus facilitates ring closure.

The remarkable selective formation of only one out of four possible diastereomers upon cyclization raises the question of the absolute stereochemistry of the cyclization site. The NOE data of Ψ[CH(OH)NH]–TycA are consistent with the NH temperature gradients and support the β sheet formation under the NMR conditions. As they are similar compared to native TycA, they suggest the *S*-configuration of Leu10, like in the linear precursor. The NOEs were used as interproton restraints for molecular dynamics simulations of Ψ[*R*-CH(OH)NH]–TycA and Ψ[*S*-CH(OH)NH]–TycA. However, none of the two hemiaminal epimers could be unambiguously identified as the more stable one, because similar geometries are obtained in both cases, and the hemiaminal is capable of hydrogen bonding with Phe3-NH (Fig. 2).

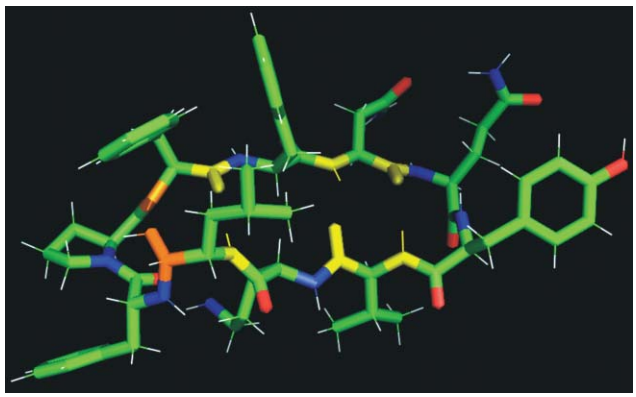

Fig. 2 NOE-based energy-minimized average structure of Ψ[*R*-CH(OH)-NH]–TycA in SDS-d_{25} environment in H_2O at pH 9.0. The structure is similar to the antiparallel β sheet structure of native TycA. Carbons are shown in green, oxygens in red, and nitrogens in blue. The three amide hydrogen bonds are marked in yellow, and the hemiaminal hydrogen bond is marked in orange.

The presence of a single stereoconfiguration results from the preference for one of the two half-spaces of the aldehyde by the amine nucleophile. This apparently arises from the steric shielding by the surrounding side chains, in combination with a precoordination of the aldehyde to the Phe3-NH (Fig. 3a). As visible from the structure shown in Fig. 2, this should result in an orientation of the hemiaminal-NH to the solvent, which is indicated by the large temperature coefficient of –15.0 ppb K^{-1}. 30 min after cyclization, a low-intensity second set of signals (<10%) is visible, which grows during the following hours, finally reaching approx. 25% after 38 h (Fig. S6 and S7 in the ESI†). As the expected dehydration to the imine does not occur, two explanations are conceivable. Either this secondary set of signals results from the epimeric hemiaminal, slowly formed *via* an imine intermediate by dehydration and subsequent addition of water, which occurs with less stereocontrol than the addition of the amine to the aldehyde, or alternatively, a slow epimerization of Leu10 may also be the reason for the emerging second set of signals.[6] The formation of dimers in principle is also possible, but was never observed in cyclization studies of native TycA.[7]

Thermodynamics of TycA macrocyclization

In order to elucidate the entropy balance of the cyclization process, we investigated the temperature dependence of the cyclization equilibrium. In spite of high signal density (two main sets of decapeptide signals) and the line broadening caused by the SDS-d_{25}, the Val8 and Leu10 methyl signals are well enough resolved to be integrated, and thus to give information about the composition of the equilibrating system (Fig. 3b). The thermodynamic parameters can be determined from the temperature dependence of the percentages of linear species (TycA–CH(OH)$_2$) and cyclic species (Ψ[CH(OH)NH]–TycA). As the composition of the linear/cyclic peptide mixture at pH 6.6 does not change significantly between 290 and 330 K (variations of <5% are within measuring

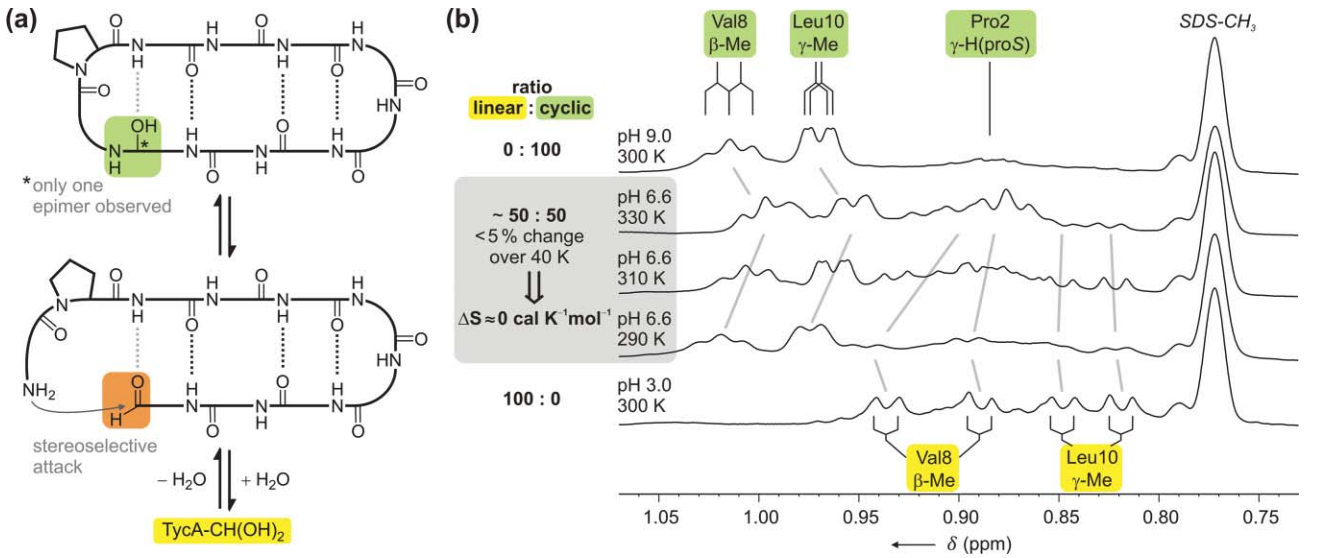

Fig. 3 (a) Schematic representation of backbone, Pro residue and hydrogen bonds before (Tyc–CHO) and after (Ψ[CH(OH)NH]–TycA) cyclization. As also assumed for native TycA, residues 3–9 in the linear peptide are preorganized for cyclization by the presence of three hydrogen bonds (black dotted lines: Phe3-CO/Leu10-NH, Val8-CO/Asn5-NH, Asn5-CO/Val8-NH; from left to right). This enables a precoordination of the aldehyde (orange) by another hydrogen bond (grey dotted line) and the stereoselective attack by the amine of the predisposed N-terminal segment. The hydrogen bond between the aminal oxygen and the Phe3-NH is analogous to the hydrogen bond formed by the tetrahedral intermediate in native TycA cyclization and to the Leu10-CO/Phe3-NH interaction resulting after amide formation. (b) Aliphatic proton range of ^{1}H NMR spectra recorded at different temperatures and pH values using water suppression by the Gradient-Tailored Excitation (Watergate) method (600 MHz, KH_2PO_4/H_3PO_4 buffer, $H_2O:D_2O$ 5 : 1 with 0.19 mol L^{-1} SDS-d_{25}). The spectra at pH 6.6, depicted in the middle, show the presence of an approx. 50 : 50 (linear : cyclic) mixture, which is temperature-independent of composition. For comparison, the spectrum of TycA–$CH(OH)_2$ at pH 3.0 and 300 K is shown at the bottom, while at the top, the respective section of the freshly cyclized macrocyclic hemiaminal at pH 9.0 and 300 K, showing only one main set of signals, is depicted.

uncertainty), the balance of cyclization enthalpy ΔH can be set to zero according to the van't Hoff equation. The amounts of linear and cyclic species are approx. 50% each, which indicates that the free energy difference ΔG can, in an approximation, also be set to zero. Consequently, the Gibbs–Helmholtz equation gives the entropy balance ΔS of the macrocyclization reaction, which is equal to zero. This value represents the total entropy change, which also includes the released water molecule.

In order to estimate the entropy loss of the peptide chain, several contributions must be taken into account, as already discussed for the case of nostocyclopeptide macrocyclization.[6] The changes in the solvation sphere can, in a first approximation, be neglected in comparison to the water released upon condensation, which is supported by mostly slight changes of the chemical shifts (Table S2 in the ESI†). While two equivalents of water are released upon imine formation,[6] hemiaminal formation is accompanied by the loss of only one equivalent of water. Subtracting the entropy of one released water molecule (16.7 cal K^{-1} mol^{-1})[15] from the obtained entropy balance of approx. 0 cal K^{-1} mol^{-1} gives an entropy loss of the TycA peptide chain of approx. 17 cal K^{-1} mol^{-1}. The rule of thumb that half of the rotational degrees of freedom are lost upon cyclization proved to be a well applicable approximation.[6] Concerning the cyclization of the decapeptide TycA–CHO, possessing 18 rotational degrees of freedom in the backbone, we should expect an entropy loss of 27 (9 × 3) cal K^{-1} mol^{-1}, which exceeds the experimentally obtained value by 10 cal K^{-1} mol^{-1}. Hence, the overall entropy balance of a decapeptide cyclization with the release of one equivalent of water should be in the range of approx. –10 cal K^{-1} mol^{-1}, and thus entropically disfavored. This demonstrates that the linear precursors of the Tyrocidine peptides feature such an extraordinary pronounced preorientation that the cyclization entropy balance is shifted from a negative to an approx. neutral value. The 10 cal K^{-1} mol^{-1} reduction observed for TycA–CHO cyclization is equivalent to approx. three rotational degrees of freedom, and the cyclizing substrate acts more like a hepta- than a decapeptide. These results are in agreement with former qualitative investigations on native TycA, in which its outstanding ability to self-cyclize without reagent was demonstrated,[7] and now add a quantitative estimate of the TycA preorientation and cyclization tendency.

Conclusions

In summary, we use the isosteric substitution of the peptide bond against a reversibly closing carbonyl functionality for the structural and thermodynamic analysis of a peptide cyclization in order to quantify the extent of preorientation of the linear cyclization precursor TycA–CHO. This readily cyclizes in aqueous solution, and surprisingly yields a hemiaminal that is not further dehydrated to the expected imine. The unprecedented stability of the aminal results from its apparent incorporation into the hydrogen bonding network. Furthermore, the formation of a dominating stereoisomer suggests a stereoselective attack, which is only possible if the cyclization site is structurally well-defined. The isosteric substitution of an amide by a carbonyl analog has promising potential for the elucidation of macrocyclization thermodynamics. A systematic "carbonyl scan" of all amide bonds in a cyclopeptide could allow the ranking of different precursors

with respect to their inclination to cyclize, and to compare the native with non-native ligation sites. It also has potential for the determination of the native cyclization site if it is not known. As an analog of the tetrahedral intermediate in amide formation, the macrocyclic hemiaminal is of particular interest. Notably, a comparison of the TycA hemiaminal and the native amide with respect to biological activity will help the understanding of the TycA structure–activity relationship. We are confident that the examination of further lactams should reveal more details, which may help us to better understand the determinants of macrocyclization processes.

Experimental section

Peptide synthesis

The synthesis of Tyrocidine A aldehyde was carried out according to a protocol described previously,[16] starting from H–Thr–Gly–NovaSyn® TG resin (Novabiochem) preloaded with leucine amino aldehyde. The chain extension was carried out by automated Fmoc solid phase synthesis. After assembly of the target sequence and removal of protecting groups, the crude peptide aldehyde was cleaved from resin, precipitated in hexane, and purified by preparative HPLC. The identity of the product obtained after lyophilization was confirmed by analytical HPLC-ESI-MS. A sample of cyclized TycA in basic aqueous solution was also controlled *via* ESI-MS. Further experimental details, the HPLC trace of the purified peptide, as well as high-resolution ESI-MS spectra of all linear and cyclic compounds, are provided in the ESI.†

NMR analytics

Homo- and heteronuclear NMR measurements were carried out using a Bruker Avance DRX600 spectrometer operating at 600.13 MHz (^{1}H) and 150.90 MHz (^{13}C), respectively, with a 5 mm BBI probe head. All chemical shifts are given in parts per million (ppm). For water suppression, the double pulsed field-gradient spin echo (WATERGATE) sequence[17] was applied, which had also been appropriate in preceding studies.[6] All measurements described were carried out with the same NMR sample to prove the reversibility of the system. TycA–CHO was dissolved in a partially deuterated KH_2PO_4/H_3PO_4 buffer solution (H_2O/D_2O 5 : 1) at pH 3.0. The pH was set to the desired value by addition of solid Na_2CO_3 and aqueous H_3PO_4, respectively. For the temperature-dependent series of measurements, 10 K steps (290–330 K) were chosen and the sample was allowed to equilibrate in the spectrometer for 1 h. In the ESI,† further details, including assigned NMR spectra, as well as chemical shifts and temperature gradients, are given.

NOE-based molecular modeling

The structure calculations of Ψ[*R*-CH(OH)NH]–TycA were carried out as described previously[6,18] using the program package HyperChem[19] with MM+ force field and without explicit water included. The NOE-derived distances that were used as restraints are listed in Table S4 in the ESI.† Ten snapshots from a 10 × 10 ps molecular dynamics simulation (step size 1 fs, 300 K) were averaged and the resulting structure was allowed to relax without restraints to verify the plausibility. The resulting structure is depicted in Fig. 2, and the acquired distances are listed in Table S4 in the ESI.†

Acknowledgements

We are grateful to Dr U. Linne for his assistance with HPLC and ESI-MS measurements. Financial support was provided by the Deutsche Forschungsgemeinschaft (DFG) and by the Studienstiftung des Deutschen Volkes (Ph.D. scholarship for S.E.).

References

1 E. M. Nolan and C. T. Walsh, *ChemBioChem*, 2009, **10**, 34–53.
2 F. Kopp and M. A. Marahiel, *Nat. Prod. Rep.*, 2007, **24**, 735–749.
3 E. M. Driggers, S. P. Hale, J. Lee and N. K. Terrett, *Nat. Rev. Drug Discovery*, 2008, **7**, 608–624.
4 J. Blankenstein and J. Zhu, *Eur. J. Org. Chem.*, 2005, 1949–1964.
5 T. Golakoti, W. Y. Yoshida, S. Chaganty and R. E. Moore, *J. Nat. Prod.*, 2001, **64**, 54–59.
6 S. Enck, F. Kopp, M. A. Marahiel and A. Geyer, *ChemBioChem*, 2008, **9**, 2597–2601.
7 X. Bu, X. Wu, G. Xie and Z. Guo, *Org. Lett.*, 2002, **4**, 2893–2895.
8 This modified TycA is abbreviated to Ψ[C=N]–TycA. The fact that no position for the modified tether is given indicates that the native cyclization position is retained.
9 M.-C. Kuo and W. A. Gibbons, *J. Biol. Chem.*, 1979, **254**, 6278–6287.
10 M.-C. Kuo and W. A. Gibbons, *Biophys. J.*, 1980, **32**, 807–836.
11 R. M. Kohli, J. W. Trauger, D. Schwarzer, M. A. Marahiel and C. T. Walsh, *Biochemistry*, 2001, **40**, 7099–7108.
12 A. P. Tonge, P. Murray-Rust, W. A. Gibbons and L. K. McLachlan, *J. Comput. Chem.*, 1988, **9**, 522–538.
13 N. Zhou, P. Mascagni, W. A. Gibbons, N. Niccolai, C. Rossi and H. Wyssbrod, *J. Chem. Soc., Perkin Trans. 2*, 1985, 581–587.
14 J. W. Trauger, R. M. Kohli and C. T. Walsh, *Biochemistry*, 2001, **40**, 7092–7098.
15 L. M. Amzel, *Proteins: Struct., Funct., Genet.*, 1997, **28**, 144–149.
16 F. Kopp, C. Mahlert, J. Grünewald and M. A. Marahiel, *J. Am. Chem. Soc.*, 2006, **128**, 16478–16479.
17 T.-L. Hwang and A. J. Shaka, *J. Magn. Reson., Ser. A*, 1995, **112**, 275–279.
18 M. Haack, S. Enck, H. Seger, A. Geyer and A. G. Beck-Sickinger, *J. Am. Chem. Soc.*, 2008, **130**, 8326–8336.
19 *HyperChem*, Hypercube, Inc., Gainesville, FL, 2000.