Peptide Self-Replication via Template-Directed Ligation

Kay Severin, David H. Lee, Jose A. Martinez, and M. Reza Ghadiri*

Abstract: A 32-residue α-helical peptide with a sequence similiar to that of the GCN4 leucine zipper region is shown to catalyze its own formation by accelerating the amide bond formation of a 17-residue peptide, preactivated as a thiobenzyl ester, and a 15-residue peptide with a *N*-terminal cysteine. The self-replication process displays parabolic growth characteristics as revealed by a detailed kinetic analysis. Control reactions with single-mutant peptides strongly support a mechanism in which a ternary and/or quaternary complex of the product with both peptide fragments act(s) as the catalytically active intermediate(s). Furthermore, these experiments reveal a remarkable sequence selectivity, as evidenced by the loss of autocatalytic activity as a result of a single replacement of leucine or valine residues with an alanine at the recognition interface.

Keywords
autocatalysis · coiled coil · kinetics · peptides · self-replication

Introduction

In a distant future, it may be possible to construct complex molecular systems through *self-instructed nonlinear* chemical processes. Such endeavors would first require the ability to rationally design *informational self-reproducing and self-organizing* molecular systems. The design and study of self-replicating molecular systems is viewed as the first step toward these long-term goals and is expected to provide the vehicle for exploring such uncharted chemistry frontiers. Here we describe the design principles, the kinetic profile, and the molecular information transfer properties of the first self-replicating peptide structure.

Molecular self-replication, in its most basic manifestation, is a reaction in which the product(s) functions as a specific catalyst for the reaction. Thus, self-replicating molecular systems are necessarily nonlinear chemical processes. They can exhibit varying degrees of nonlinearity depending on the molecular nature and the complexity of the overall system.[1] In recent years a few intriguing examples of self-replicating molecular structures have been described.[2] These primarily operate based on the well-defined pattern of hydrogen-bond donors and acceptors of nucleic acid base-pairing interactions. In contrast, the self-replicating molecular structure described here is the first example of a peptide-based replicator that utilizes a multitude of noncovalent chemical interactions.[3]

[*] M. R. Ghadiri, K. Severin, D. H. Lee, J. A. Martinez
Departments of Chemistry and Molecular Biology and
The Skaggs Institute for Chemical Biology
The Scripps Research Institute, La Jolla, CA 92037 (USA)
Fax: Int. code +(619)784-2798
e-mail: ghadiri@scripps.edu

Results and Discussion

Design Principles: As opposed to nucleic acids, polypeptides are not intrinsically self-complementary molecular structures. Expression of the required complementary surfaces in self-replicating peptides depends not only on the primary sequence, but also on its folded three-dimensional structure. Accordingly, our design was based on a simple protein folding motif, namely, the α-helical coiled coil, for which the factors contributing to the thermodynamic and kinetic stability of the helical ensembles are reasonably well-understood.[4, 5] The most simple coiled-coil motif is made up of an identical pair of parallel α-helical peptides that wrap around one another with a slightly left-handed superhelical twist. The primary sequence of a coiled-coil peptide exhibits a heptad repeat (*abcdefg*; see Figure 4) in which the residues at *a* and *d* positions constitute the primary interhelical recognition motif forming a specific (knobs into holes) hydrophobic interface (Figure 1). Electrostatic interactions between the residues at the *e* and *g* positions are thought to constitute a secondary level of molecular recognition.[5] Residues at the *b*, *c*, and *f* positions are exposed to the solvent and do not participate directly in the self-assembly of the coiled-coil structure and thus

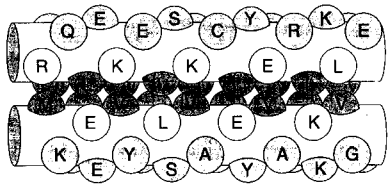

Figure 1. Schematic representation of the coiled-coil peptide structure employed in this study. Amino acid side chains are represented as balls. The hydrophobic core consists of alternating valine and leucine residues (darker gray) which show a knobs-into-holes type packing.

can tolerate a large variety of amino acid substitutions. The thermodynamic and kinetic stability, the aggregation state, and the relative orientation of the helical strands in such peptides have been shown to depend largely on the identity of the amino acid residues at positions *a* and *d*. Certain amino acid substitutions at these positions can transform the two-stranded coiled-coil structure into three- or four-helix bundles.[4f]

Monomeric peptide subunits of a coiled coil in aqueous solutions are typically random coils, but almost completely α-helical in the aggregated state. Therefore, conceptually a given peptide subunit of a coiled-coil structure may be viewed as the template that cooperatively directs the self-assembly and self-organization of the other strand(s). If so, it seemed reasonable that, given sufficient thermodynamic driving force, a coiled-coiled peptide subunit may also act as a template to organize two shorter peptide fragments onto itself (Figure 2). Furthermore, we

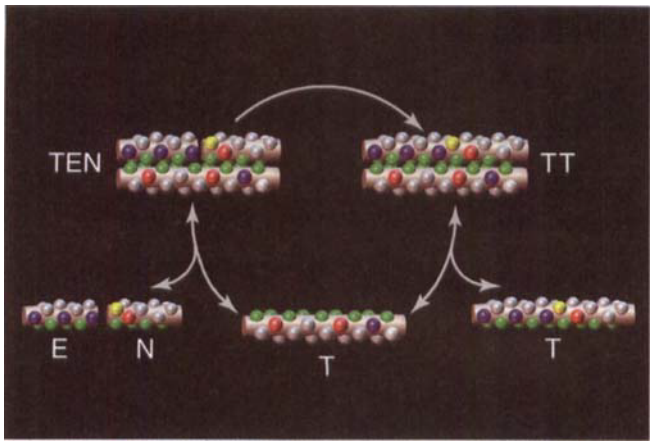

Figure 2. Schematic representation of the minimal autocatalytic reaction cycle for the self-replicating α-helical peptide. The electrophilic and nucleophilic peptide fragments E and N are preorganized on the template T through interhelical hydrophobic interactions forming the ternary complex TEN. Subsequent amide bond formation gives an identical copy of the template itself, which then becomes part of the autocatalytic cycle to promote further peptide ligation. The following color code is used: hydrophobic residues are shown in green, positively charged residues are blue, while negatively charged residues are red. The cysteine residue, critical for the ligation, is shown in yellow.

hypothesized that peptide fragments that are equipped with appropriate reactive end groups should enjoy significantly higher coupling rates, due to their close proximity and higher effective molarity when preorganized on the template strand.[6] We envisioned that, if the reacting peptides are constituent fragments of the template sequence, the newly formed product will be an identical copy of the template itself, which sets up a new catalytic cycle that further catalyzes the formation of more templates. Therefore, the template-directed catalysis (autocatalysis) creates a positive feedback loop, which establishes the nonlinear growth profile of self-replicating molecular systems.

The chemo- and regioselectivity in peptide fragment coupling is of paramount importance, especially when free peptides having reactive side chains are employed in aqueous solutions. Although a number of peptide fragment coupling strategies may be applicable in the self-replicating process, in the present study we chose to employ the method of Kent et al. in which the *N*-terminal peptide fragment is preactivated as a thiobenzyl ester (electrophilic fragment **E**) and the *C*-terminal fragment is equipped with a free cysteine residue at its *N*-terminus (nucleophilic fragment **N**).[7] Such a coupling strategy circumvents the need for addition of external coupling reagents to the reaction mixture. In addition, the coupling reaction becomes highly chemo- and regioselective in producing the desired amide bond at the intended coupling site (Figure 3), because of the appropri-

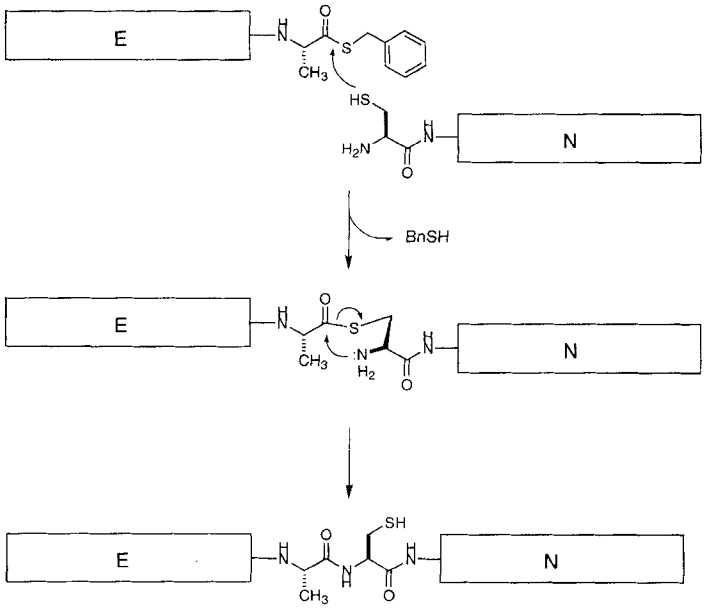

Figure 3. The first step of the ligation is a transthioesterification between the activated *C*-terminus of the electrophilic peptide fragment **E** and the *N*-terminal cysteine side chain functionality of the nucleophilic peptide fragment **N**. The intermediate thioester rapidly rearranges to give the native amide bond (the intermediate thioester has never been observed during the course of the self-replication process).

ate stability of the thiobenzyl ester in neutral aqueous solutions,as well as the greater nucleophilicity of the sulfhydryl functionality of the cysteine residue at neutral pH as compared to all other side chain moieties.

A 32-residue polypeptide, similar in sequence to the coiled-coil α-helical domain of the yeast transcription factor GCN4, was chosen as the putative catalyst (template **T**) in the self-replication process.[4i] The sequence employed in this study differs from the natural sequence in six positions (Figure 4). The wild-type GCN4 sequence contains a single neutral hydrophilic residue (Asn16) in the hydrophobic core structure, which has been shown to exert a critical influence in limiting the aggregation state of the peptide to a parallel homodimer, while at the same time considerably reducing the thermodynamic stability of the coiled-coil structure.[4i] We felt that, unlike the full-length peptide, a short fragment would not be able to bury such a polar residue in the hydrophobic core and thus may not favor the formation of a productive template–peptide fragment complex. Therefore, the asparagine residue at the position 16 of the natural sequence was replaced with valine. The consequence of this mutation is that the template can equilibrate between two- and three-stranded structures.[4f] At first glance the occurrence of this additional species may seem to be a complication, but it may also be a productive situation since the template dimer (**TT**) can potentially act as a catalyst in the ligation reaction (Figure 7).

0947-6539/97/0307-1018 $ 17.50+.50/0 Chem. Eur. J. 1997, 3, No. 7

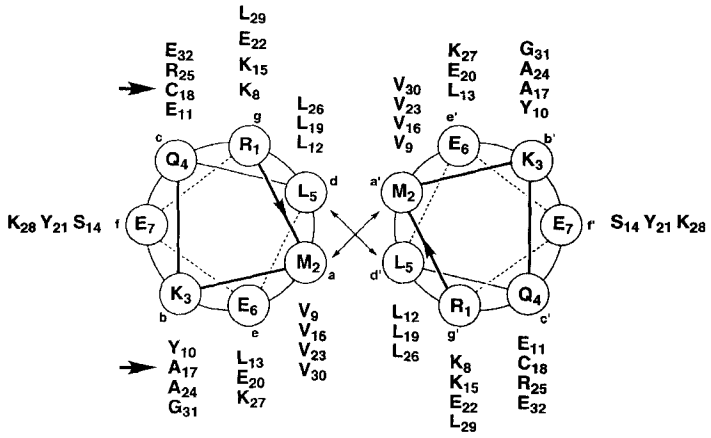

Figure 4. Helical-wheel diagram of the template peptide in the dimeric α-helical coiled-coil configuration emphasizing the heptad repeat motif (top). The interhelical recognition surface is dominated by hydrophobic packing interactions (positions *a* and *d*) and electrostatic interactions (positions *e* and *g*). Amino acids at positions *b*, *c*, and *f* lie on the solvent-exposed surface of the helical structure and do not participate in the molecular recognition processes. Arrows indicate the ligation site between a cysteine residue (*N*-terminus of the nucleophilic peptide fragment) and an alanine residue (as thiobenzyl ester activated *C*-terminus of the electrophilic peptide fragment). The following peptide sequences were employed in this study:
Ar-RMKQLEEKVYELLSKVACLEYEVARLKKLVGE-CONH$_2$ (**T**),
Ar-RMKQLEEK*E*YELLSKVACLEYEVARLKKLVGE-CONH$_2$ ($\mathbf{T}_{9E}$),
Ar-RMKQLEEKVYELLSKVACLEYEVAR*E*KKLVGE-CONH$_2$ ($\mathbf{T}_{26E}$),
Ar-RMKQLEEK*A*YELLSKVACLEYEVARLKKLVGE-CONH$_2$ ($\mathbf{T}_{9A}$),
Ar-RMKQLEEKVYELLSKVACLEYEVAR*A*KKLVGE-CONH$_2$ ($\mathbf{T}_{26A}$),
Ar-RMKQLEEKVYELLSKVA-COSBn (**E**),
Ar-RMKQLEEK*E*YELLSKVA-COSBn ($\mathbf{E}_{9E}$),
H_2N-CLEYEVARLKKLVGE-CONH$_2$ (**N**),
H_2N-CLEYEVAR*E*KKLVGE-CONH$_2$ ($\mathbf{N}_{26E}$).
For consistency, residue numbering of the smaller fragments are the same as the template. The templates and the electrophilic peptide fragments were acylated at the *N*-termini with 4-acetamidobenzoic acid (Ar) to allow sensitive monitoring of product formation by HPLC.

All other amino acid substitutions to the wild-type GCN4 sequence were made to residues at the solvent-exposed surface of the helical structure and are not implicated in the molecular information transfer process. The amino acids at positions 17 and 18 were chosen as the site for breaking the full sequence to form the electrophilic and nucleophilic fragments and were replaced with alanine and cysteine residues, respectively. Furthermore, in order to facilitate spectroscopic analysis of the reaction mixture, Glu 10 and Asn 21 were replaced with tyrosine residues, and 4-acetamidobenzoic acid was coupled to the *N*-termini of the electrophile and the template. Aspartic acid residue at position 7 was changed to Glu to avoid possible side reactions during solid-phase peptide synthesis. The 15-residue nucleophilic **N** and 17-residue electrophilic **E** peptide fragments were derived accordingly from the template sequence (Figure 4).

Peptide Self-Replication: Autocatalytic amide bond formation can be clearly established when reaction mixtures differing only in the initial concentration of the template are studied (Figure 5). Increasing the initial template concentrations in otherwise identical reaction mixtures was shown to lead to a significant increase of the initial rates of product formation. Three sets of reactions were carried out starting with initial peptide concentrations of [**E**] = 185 μM and [**N**] = 179 μM and in the presence of various amounts of template (0, 8, 28, and 53 μM).

The present study complements our recent preliminary report in a number of ways.[3] The self-replication studies were performed at higher initial peptide concentrations and monitored for a longer reaction time-course in order to improve the data quality and assist in more detailed kinetic analyses. It is interesting to note that the initial concentration of the peptide fragments significantly affects the reaction growth profile.[8] For example, in one of the reactions described above, in the presence of 30% initial template (53 μM), a 540% enhancement for the initial rate of template formation was observed as compared with the rate of product formation in the absence of added template. This value exceeds the 350% rate enhancement previously reported for similar experiments (in the presence 40% initial template) performed at twofold lower peptide fragment concentrations.[9]

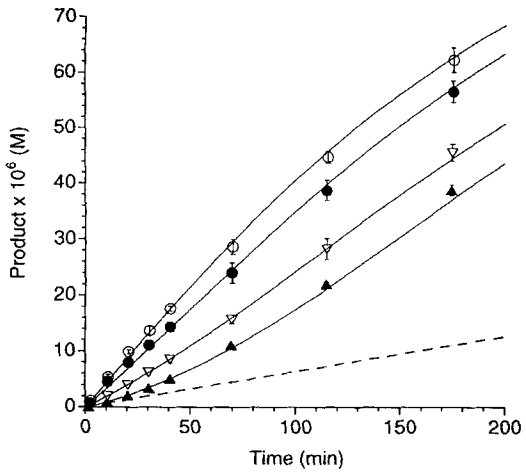

Figure 5. Template formation as a function of time for reaction mixtures with initial concentrations [**E**] = 185 μM, [**N**] = 179 μM, and [**T**] = 0 (▲), 8 (△), 28 (●), and 53 μM (○). Error bars reflect standard deviations of three independent runs. Curves were generated by nonlinear least-squares fit. The simulations are based on the reaction model described in the text. The dashed line represents the calculated production of template in the absence of autocatalysis.

The ligation reaction was found to be remarkably regio- and chemoselective. After 170 minutes less than 15% side products were observed (compared to the total amount of template produced) (Figure 6). Much of the side product was hydrolyzed thioester (**Hyd**) (≈9%). The second major side product was a "branched" peptide (**BR**) formed in a transesterification reaction between the thioester **E** and the cysteine side chain of the template **T** (≈6%). Both species were isolated and characterized by electrospray mass spectrometry, and their physical properties were compared with authentic samples. In order to establish whether the **BR**

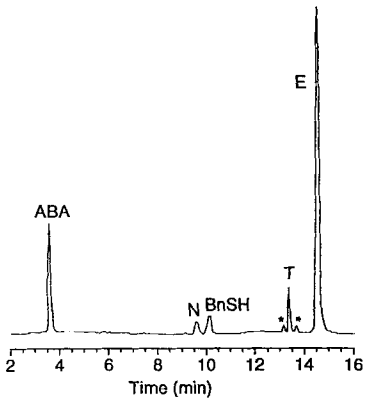

Figure 6. Reverse-phase HPLC reaction profile 70 min after initiation of the reaction. ABA denotes the internal standard 4-acetamidobenzoic acid, **N** the nucleophilic peptide fragment, T the template, **E** the electrophilic peptide fragment, and BnSH is benzyl mercaptan. The two minor side products—hydrolyzed electrophile and a ("branched") thioester formed from **T** and **E**—are marked with an asterisk.

species can act directly as a catalyst for the reaction, despite its very low concentration during the course of the reaction, we investigated the reaction of the **BR** peptide with the nucleophilic peptide fragment **N**. These studies indicated that product was only formed in the presence of benzyl mercaptan suggesting the equilibrium shown in reaction (4) of Scheme 1, thus negating any significant catalytic role of **BR** in the self-replication process. Other minor side products (<1%) that were identified by liquid chromatography and mass spectrometry were the mixed disulfides formed between the template and benzyl mercaptan and between the template and the nucleophile.

There are at least six template-dependent pathways that might possibly contribute to the autocatalytic process. The crucial intermediates of those pathways are schematically shown in Figure 7. Ligation via the ternary or quaternary complexes **TEN**

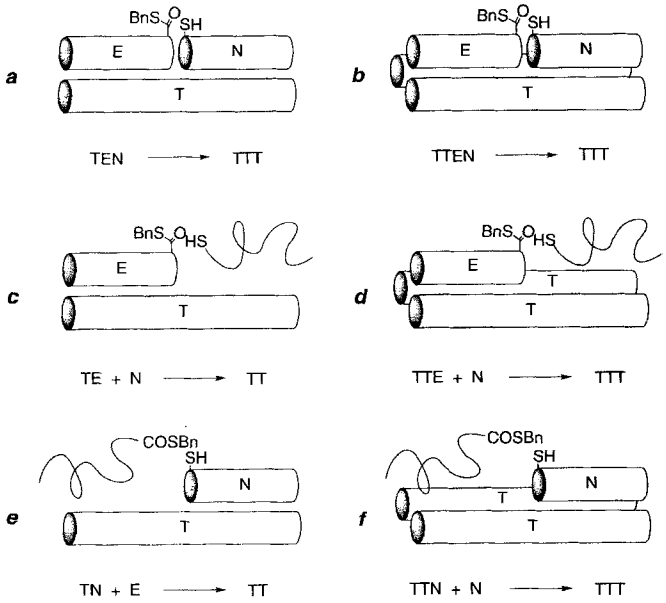

Figure 7. Schematic representation of plausible catalytically active intermediates which can contribute to the autocatalytic channel of template formation. Results of a variety of experiments suggest that only pathways a and b are active (see text).

or **TTEN** (Figure 7a and b) does involve molecular recognition of *both* fragments. Therefore, these aggregates can be considered as the "true" intermediates of the self-replication process. For the bimolecular, simply autocatalytic reactions proceeding through intermediates c–f, only one of the peptide fragments is structurally organized on the template.

Two types of control experiments were designed to reveal the nature of the template-catalyzed process. In the first set of control experiments two reactions, one with 13% initial template and the other without, were performed in the presence of 2.5 M guanidinium hydrochloride in the reaction buffer. The chaotropic agent was expected to denature the thermodynamically least stable template–peptide fragment(s) species, such as the ternary complex **TEN**, thereby abolishing any template-directed effect without affecting any functional groups present in the peptides that might somehow catalyze the ligation reaction. Figure 8 shows the results of this experiment at an initial peptide concentration of 190 mM. As in the control experiment previ-

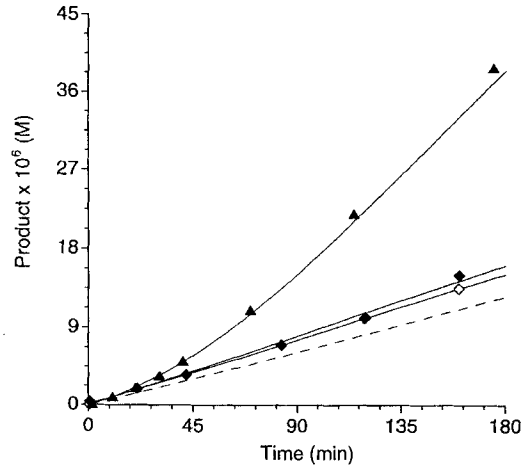

Figure 8. Template formation as a function of time in the presence (◆) and absence (◇) of 25 μM initial template. The reactions were carried out in a buffered solution of 2.5 M guanidinium hydrochloride. Production of template in the absence of guanidinium hydrochloride without any added template is shown for reference (▲). The branched peptide **BR** was included as part of the template **T**.

ously reported at 90 mM,[3] there is very little difference in rate between the reaction with template compared to that without template– practically none when one accounts for the minute production of the branched species. Furthermore, the rates of the reactions in guanidinium hydrochloride are similiar to the calculated rate of the template-independent background reaction (see kinetic analysis). It is also noteworthy to point out that the sigmoidal growth (a characteristic signature of autocatalytic processes) that was so prominent in the experiments lacking guanidinium hydrochloride was lost. Therefore, it can be unambiguously concluded that loss of template-assisted structural preorganization in the reaction mixture abolishes catalysis and that no unusual or fortuitous off-template catalysis is provided by the peptide side-chain functionalities in the primary step of the fragment condensation process.

The second set of control experiments was performed in order to determine whether the template-dependent channels c–f (Figure 7) contribute to the production of the full-length peptide. Two "crippled" peptide sequences were designed for use as templates in these reactions. The crippled peptide sequences differ from the original template sequence only in one amino acid position at the recognition interface: a glutamic acid residue in place of Val 9 ($\mathbf{T_{9E}}$) and Leu 26 ($\mathbf{T_{26E}}$) in the native sequence, respectively. The locations of these amino acid substitutions were chosen to correspond to the putative binding regions of **E** and **N** on the native template strand. Placement of a hydrophilic residue inside the hydrophobic recognition interface of coiled coils is known to disrupt interhelical associations.[4j] Therefore, in the presence of a given crippled template the reaction can initially proceed only through pathways c–f, and thus their contributions to the overall self-replicating process can be directly assessed. As shown in Figure 9 reactions initially containing 17% mutant templates $\mathbf{T_{9E}}$ and $\mathbf{T_{26E}}$ had approximately the same rate of product formation as the reaction with no initial template at all. Thus, any autocatalysis observed was derived from the normal template that was produced in situ during the course of the reaction, and not from the mutant template that was initially present in the reaction mixture. These

0947-6539/97/0307-1020 $ 17.50+.50/0 Chem. Eur. J. 1997, 3, No. 7

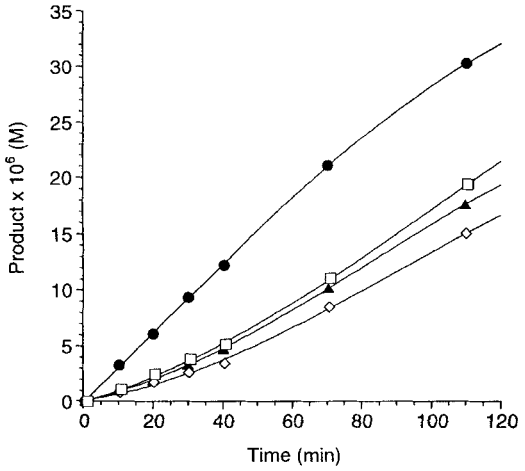

Figure 9. Template formation as a function of time for reaction mixtures initially containing 29 μM native template **T** (●), 30 μM crippled template $\mathbf{T_{9E}}$ (◇), 28 μM crippled template $\mathbf{T_{26E}}$ (□), or no template at all (▲). In the presence of templates with Glu mutations in the hydrophobic core, no significant rate-enhancement is observed.

experiments indicate that intermediates c–f (Figure 7) do not contribute significantly to the production of **T** and thus the self-replication process must indeed proceed through the intermediate(s) (such as a and b) in which both fragments are interacting with the template through specific hydrophobic packing interactions.

Sequence Selectivity in Peptide Self-Replication: The above studies in the presence of crippled templates also suggest a high degree of sequence specificity in the self-replication process. In order to more fairly probe the sequence fidelity of the self-replication process—the degree and preciseness of the information transfer process—we prepared an additional electrophilic ($\mathbf{E_{9A}}$) and nucleophilic ($\mathbf{N_{26A}}$) peptide fragment for our studies. Each of these fragments differs from its corresponding native counterpart by a single, very "conservative" mutation in the hydrophobic core (alanine in place of Val9 and Leu26, respectively; see Figure 4). Two sets of reactions were employed to assess the sequence selectivity issues in the self-replication process. In the first set of reactions, the mutant fragment $\mathbf{E_{9A}}$ and the native (nonmutated) fragment **N** were allowed to react in the presence of 15% $\mathbf{T_{9A}}$, 19% native template **T**, and in the absence of any added initial template. In the second set of reactions, the fragments **E** and $\mathbf{N_{26A}}$ were allowed to react in the presence of 41% $\mathbf{T_{26A}}$, 19% native template **T**, and in the absence of any added initial template. Interestingly, none of these reactions displayed any significant template-assisted catalysis. The above studies clearly establish a remarkable sequence selectivity in the formation of the (auto)catalytically productive intermediate complexes and thus strongly suggest a high sequence fidelity in the peptide self-replication process.

Kinetic Analysis: Simulations of the experimental data were carried out by means of the program SimFit (nonlinear fitting by dynamic simulations).[2g] This program derives rate constants together with their errors by nonlinear curve-fitting. Following a general theoretical analysis of autocatalysis in self-complementary systems,[1b] we focused on the minimal reaction model

$$\mathbf{E} + \mathbf{N} \xrightarrow{k_1} \mathbf{T} \quad (1)$$

$$\mathbf{E} + \mathbf{N} + p\mathbf{T} \xrightarrow{k_2} (1+p)\mathbf{T} \quad (2)$$

$$\mathbf{E} \xrightarrow{k_3} \mathbf{Hyd} \quad (3)$$

$$\mathbf{E} + \mathbf{T} \underset{k_5}{\overset{k_4}{\rightleftharpoons}} \mathbf{BR} + \mathrm{BnSH} \quad (4)$$

Scheme 1. Minimal reaction model for autocatalysis in self-complementary systems.

depicted in Scheme 1. Reaction (1) reflects the template-independent background with the rate constant k_1, assuming a simple bimolecular ligation. Reaction (2) reflects the autocatalytic channel with the apparent rate constant k_2. To account for the major side products, a pseudo first-order hydrolysis of **E** and a slow equilibrium between **E**, **T**, and **BR** were included (reactions (3) and (4)). Approximations for the rate constants k_3 to k_5 were determined in separate simulations including side products and fixed to the following values: $k_3 = 2.4 \times 10^{-6}\,\mathrm{s}^{-1}$, $k_4 = 7.4 \times 10^{-2}\,\mathrm{M}^{-1}\mathrm{s}^{-1}$, and $k_5 = 2.6 \times 10^{-2}\,\mathrm{M}^{-1}\mathrm{s}^{-1}$. The initial benzyl mercaptan concentration was set to 1.6 mM.

Of special interest is the reaction order p with respect to the template. If we assume that autocatalysis proceeds through the ternary complex **TEN**, that the concentration of **TEN** is low in comparison to the concentration of the template dimer **TT** ("product inhibition"), and that the rate-limiting step is the irreversible reaction of **TEN** to form **TT**, then a reaction order of $p = 0.50$ is expected, giving rise to parabolic growth of template ("square-root law").[1b] If, however, higher order aggregates are involved (e.g. autocatalysis proceeds through the two-stranded α-helical template) or significantly higher concentrations of **TEN** are present (little or no product inhibition), p may have a higher value: $0.5 < p \leq 1$ (see Appendix). A priori it was hypothesized that the unique feature of peptide self-replication based on coiled-coil sequences may allow higher than parabolic growth of template. Figure 10 shows a plot of the reaction order against the goodness of fit.[10] The best fit is obtained for $p = 0.63$. This value is in accordance with the mechanism shown in Figure 2, but also suggests a possible contribution from **TTEN** to the autocatalytic pathway.

As outlined above, the reaction order as well as the autocatalytic efficiency is strongly dependent on the relative thermody-

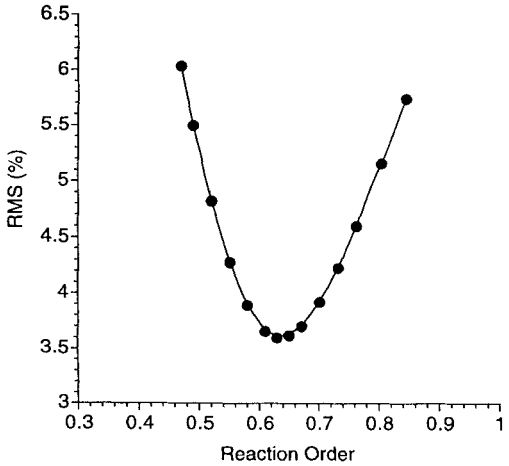

Figure 10. RMS value calculated with SimFit as a function of the reaction order p with respect to the template. The minimum RMS value is obtained for $p = 0.63$.

namic stability of the ternary complex **TEN** with respect to the template dimer **TT**. Temperature is one of the parameters which is expected to influence the stability of these aggregates. Therefore another set of experiments was carried out at 5 °C (similiar concentrations as in previous reactions). At this temperature the peptide was still able to replicate (Figure 11). However, in comparison to the experiments carried out at 21 °C the rate enhancement was slightly diminished (290% at 5 °C instead of 460% at 21 °C for an initial template concentration of 55 μM).

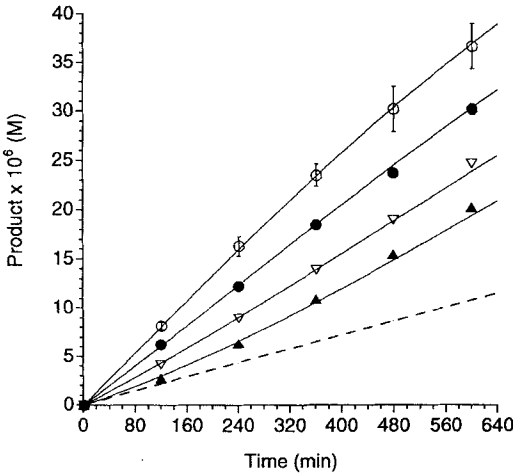

Figure 11. Template formation as a function of time for reaction mixtures initially containing 0 μM (▲), 9 μM (△), 27 μM (●), and 55 μM (○) of template at 5 °C. Error bars reflect standard deviations of two independent experiments. Curves were generated by nonlinear least-squares fit. The simulations are based on the reaction model described in the text. The dashed line represents the calculated production of template in the absence of autocatalysis.

Simulations based on the reaction model depicted in Scheme 1 with the fixed rate constants $k_3 = 6.2 \times 10^{-7}\,s^{-1}$, $k_4 = 1.8 \times 10^{-2}\,M^{-1}s^{-1}$, and $k_5 = 1.0 \times 10^{-2}\,M^{-1}s^{-1}$ give also an excellent fit for $p = 0.63$, suggesting no significant change in the reaction order p within our experimental error (Table 1).

Table 1. Results obtained from numerical fittings of the experimental data according to the reaction model described in the text.

T (°C)	p	k_1 ($10^{-2}\,M^{-1}s^{-1}$)	k_2 ($10^{1}\,M^{-(1+p)}s^{-1}$)	RMS (%)
21	0.63	3.2 (±0.1)	11.9 (±0.1)	3.60
5	0.63	1.1 (±0.1)	1.9 (±0.1)	2.12

Despite the remarkably good fit of the theoretical curves to the experimental data, it must be pointed out that the reaction models used are simpified representations of a potentially very complex kinetic system. Full modeling of this system should include a large number of equilibria between monomers and possible aggregates such as **TT**, **TTT**, **TEN**, and **TTEN** complexed with peptides in parallel and antiparallel orientations. Furthermore, a full model should also account for side reactions, nonspecific association, and the fact that the ligation between the thioester and the nucleophilic peptide fragment is a two-step process. However, as pointed out by von Kiedrowski,[1b] the advantage of empirical reaction models is that the derived apparent rate constants are more reliable than the often difficult and inaccurate experimentally derived parameters for various plausible pathways/intermediates.

Summary and Outlook

This study describes the design and the synthesis of a self-replicating peptide together with a detailed kinetic analysis of the autocatalytic process at two different temperatures. Control reactions in guanidinium hydrochloride solutions and with peptides having single amino acid mutations strongly support a mechanism in which a ternary and/or quaternary complex of the template with *both* peptide fragments act(s) as the catalytically active intermediate(s). Furthermore these experiments reveal a remarkable sequence selectivity in the self-replication process: single alanine mutations at the recognition interface are sufficient to abolish autocatalysis.

Characterization of the first self-replicating peptide sequence marks only the beginning of a research program aimed at the discovery and better understanding of self-organized nonlinear chemical systems. In the near future, the scope and limitations of the self-replication process as well as the behavior of multicomponent autocatalytic systems need to be explicitly addressed. The results of several ongoing studies along these lines will be reported shortly.

Appendix

If the rate-determining step of the autocatlytic cycle is described by the transition of the quarternary complex TTEN to the template trimer TTT, the reactions (5) to (8) comprise the minimal reaction model.

$$3\mathbf{T} \rightleftharpoons \mathbf{TTT} \qquad K_1 \tag{5}$$

$$2\mathbf{T} \rightleftharpoons \mathbf{TT} \qquad K_2 \tag{6}$$

$$\mathbf{TT} + \mathbf{E} + \mathbf{N} \rightleftharpoons \mathbf{TTEN} \qquad K_3 \tag{7}$$

$$\mathbf{TTEN} \longrightarrow \mathbf{TTT} \qquad k \tag{8}$$

Under the assumption that the the equilibria (5) to (7) are fast in comparison to the irreversible reaction, the initial rate of template formation can be described by Equation (9). Using the explicit equations for the equilibria (5)

$$d[\mathbf{T}]/dt = k[\mathbf{TTEN}] \tag{9}$$

to (7) the concentration of TTEN can be expressed by Equation (10). If we

$$[\mathbf{TTEN}] = K_1^{-2/3} K_2 K_3 [\mathbf{E}][\mathbf{N}][\mathbf{TTT}]^{2/3} \tag{10}$$

assume that $[\mathbf{E}] \approx [\mathbf{E}]_{init}$, $[\mathbf{N}] \approx [\mathbf{N}]_{init}$, and $[\mathbf{TTT}] \approx 1/3[\mathbf{T}]_{init}$ (template trimer is the dominant **T**-containing species in solution and is the product-inhibited catalyst), the initial rate of template formation is given by Equation (11), leading to a reaction order of $p = 2/3$ with respect to the template.

$$\begin{aligned} d[\mathbf{T}]/dt &= (1/3)^{2/3} k K_1^{-2/3} K_2 K_3 [\mathbf{E}]_{init}[\mathbf{N}]_{init}[\mathbf{T}]_{init}^{2/3} \\ &= k^*[\mathbf{E}]_{init}[\mathbf{N}]_{init}[\mathbf{T}]_{init}^{2/3} \end{aligned} \tag{11}$$

Experimental Section

General: Dichloromethane (optima grade), dimethylformamide (sequencing grade), and diisopropylethylamine (peptide synthesis grade) were purchased from Fisher. Dichloromethane and dimethylformamide were dried over

0947-6539/97/0307-1022 $ 17.50+.50/0 Chem. Eur. J. 1997, 3, No. 7

4 Å molecular sieves and stored under nitrogen; diisopropylethylamine was either used from a newly opened bottle or distilled prior to use. The following chemicals were used as purchased: acetonitrile (Fisher), 4-acetamidobenzoic acid (ABA) (Aldrich), trifluoroacetic acid (New Jersey Halo Carbon), 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) (Richelieau Biotechnologies), 3-(*N*-morpholino)propanesulfonic acid (MOPS) (Fisher), benzyl mercaptan (BnSH) (Aldrich), 2-(*N*-morpholino)ethanesulfonic acid (MES) (Sigma), guanidinium hydrochloride (Gnd·HCl) (Fisher), *N,N*-diisopropylethylamine (DIEA) (Fisher), thiophenol (Aldrich), 3-mercaptopropionic acid (Fluka), dicyclohexylamine (DCHA) (Fisher). All Boc amino acids were purchased from Bachem California. Side chains of serine and glutamic acid were protected with o-benzyl, tyrosine with dichlorobenzyloxycarbonyl, lysine with *o*-chlorobenzyloxycarbonyl, glutamine with xanthyl, arginine with *p*-tosyl, and cysteine with *p*-methylbenzyl group. *p*-Methylbenzhydrylamine (MBHA, substitution = 0.87–1.05 meq/g) resin was purchased from Novabiochem. Electrospray mass spectrometry (ES-MS) data were collected on a Sciex API 3 and Sciex API 100 mass analyzers by direct infusion at 4 $\mu L\,min^{-1}$. Liquid chromatography-mass spectrometry (LC-MS) experiments were carried out by coupling the Sciex API 100 with a Hewlett-Packard 1090 liquid chromatograph. ^1H NMR spectra were collected on Bruker AM-300/AMX-500 spectrometers.

Synthesis of BOC-Ala-$SCH_2CH_2CO_2H$:[11] 3-Mercaptopropionic acid (3.49 mL, 40 mmol) and DIEA (13.0 mL, 75 mmol) were added to a solution of BOC-Ala-OSu[12] (10 g, 35 mmol) in CH_2Cl_2 (120 mL). The resulting mixture was stirred at room temperature for 19 h. After evaporation of the solvent under reduced pressure, the crude product was dissolved in ethyl acetate (100 mL). The ethyl acetate layer was washed with 0.5 M citric acid (2 × 30 mL) and brine (1 × 30 mL), dried over sodium sulfate, and filtered. The solution was concentrated, and the resulting oil was dissolved in diethyl ether (70 mL) and crystallized upon addition of DCHA (3.42 mL, 28 mmol) and hexane. After recrystallization from hot ethyl acetate, the salt (13.2 g) was suspended in ethyl acetate (100 mL) and DCHA was extracted with 0.5 M citric acid (3 × 30 mL). The organic layer was washed with brine, dried over sodium sulfate, and filtered. After concentration and freezing, BOC-Ala-$SCH_2CH_2CO_2H$ was obtained as a white solid (7 g, 72%; m.p. = 73–75 °C). ^1H NMR ($CDCl_3$); δ = 1.28 (d, 3H), 1.36 (s, 9H), 2.57 (m, 2H), 3.02 (m, 2H), 4.07, 4.30 (2m, 1H), 5.29 and 6.64 (2m, 1H). Enantiomeric purity was confirmed by mild hydrolysis (NH_4OH, 3 h, 20 °C) and Marfey's test.[13]

Peptide Synthesis: All peptides were synthesized on a 0.5–1.0 mmol scale with MBHA resin using Boc chemistry and the optimized protocol of Kent,[14] except that DMF flow washes were replaced with 5 × 30 s shake washes. BOC-Ala-$SCH_2CH_2CO_2H$ was coupled to the resin with either DCC or the Kent protocol. 4-Acetamidobenzoic acid was coupled to the *N*-terminus of the electrophilic peptide using the same procedure. After HF cleavage (10% anisole/HF, 1 h, 0 °C) the crude peptides were washed with ether, dissolved in water, lyophilized, and then purified by C_{18} reverse-phase HPLC with 99% water/acetonitrile/0.1% TFA (A) and 90% acetonitrile/water/0.07% TFA (B) binary gradient. Purity was confirmed by analytical HPLC and mass spectrometry (electrospray or MALDI).

Synthesis of E: Benzyl mercaptan (180 mL, 1.53 mmol) was added to a solution of the propionamide thioester peptide [E($SCH_2CH_2CONH_2$), 17 μmol] in degassed 4 M Gnd·HCl (50 mM MES buffer, pH 7, 11 mL) and acetonitrile (0.9 mL). The resulting mixture was stirred at room temperature under argon for 18 h. After acidification with neat TFA, excess benzyl mercaptan was removed with diethyl ether (4 × 5 mL). The crude mixture was purified on C_{18} RP-HPLC to afford E(SCH_2Ph) (11 μmol, 65%). The thiobenzyl esters of the mutant peptides $\mathbf{E_{9A}}$ and $\mathbf{E_{9E}}$ were synthesized in an analogous manner.

Synthesis of T: The peptides E($SCH_2CH_2CONH_2$) (5.2 μmol) and **N** (5.1 μmol) were dissolved in a mixture of 5 mL degassed Gnd·HCl solution (4 M, 50 mM MES buffer, pH 7) and 0.5 mL acetonitrile. To this solution thiophenol (400 mmol) was added four times during 30 h, once 15 min before acidification with neat TFA. Excess thiophenol was removed with diethyl ether (5 ×) and the crude peptide was purified on C_4 RP-HPLC (4.0 μmol, 78%). The mutant peptides $\mathbf{T_{9A}}$, $\mathbf{T_{26A}}$, $\mathbf{T_{9E}}$, and $\mathbf{T_{26E}}$ were synthesized in an analogous fashion using the respective mutant peptide fragments.

Self-replication reactions: The reactions were performed in 0.6 mL Eppendorf tubes. Temperature was maintained at 21.0 (±0.2) °C. All components were dissolved in degassed H_2O at acidic pH to inhibit initiation of the reaction. Peptide concentrations were determined by amino acid analysis and adjusted relative to the internal standard 4-acetamidobenzoic acid. Reactions were initiated by adding argon-purged, benzyl mercaptan saturated MOPS buffer to give a total volume of 300 μL (final average concentrations: [**E**] = 185 μM, [**N**] = 179 μM, [BnSH] = ≈1.6 mM, [MOPS] = 157 mM (pH 7.50), [ABA] = 37 μM; [**T**] = 0, 8, 28, and 53 μM, respectively). All experiments were repeated three times. Samples (33 μL) were removed from the reaction vessel at various time points, immediately quenched with 2% trifluoroacetic acid (70 μL), and stored at −78 °C prior to HPLC analysis.

Peptide fragment condensations in the presence of Gnd·HCl were performed in the same way as the reactions described above, except that degassed, benzyl mercaptan saturated MOPS buffer containing Gnd·HCl was used. Final concentrations: [Gnd·HCl] = 2.5 M, [BnSH] ≈ 1.6 mM, [MOPS] = 84 mM (pH 7.50), [ABA] = 235 μM, [**E**] = 195 μM, [**N**] = 187 μM; [**T**] = 0 and 25 μM, respectively.

Peptide fragment condensations in the presence of crippled templates $\mathbf{T_{9E}}$ and $\mathbf{T_{26E}}$ were performed similiarly to the reactions described above. Final concentrations: [BnSH] ≈ 1.6 mM, [MOPS] = 139 mM (pH 7.50), [ABA] = 90 μM, [**E**] = 170 μM, [**N**] = 173 μM; [**T**] = 0 μM, [**T**] = 29 μM, [$\mathbf{T_{9E}}$] = 30 μM, and [$\mathbf{T_{26E}}$] = 28 μM, respectively.

Peptide condensations of alanine mutant fragments in the presence of nonmutated as well as single-mutant templates were performed similiarly to the reactions described above. Final concentrations: [$\mathbf{E_{9A}}$] = 163 μM, [**N**] = 167 μM, [$\mathbf{T_{9A}}$] = 24 μM, [**T**] = 0 μM and 31 μM; [**E**] = 160 μM, [$\mathbf{N_{26A}}$] = 178 μM, [$\mathbf{T_{26A}}$] = 63 μM, [**T**] = 0 μM and 31 μM; in all experiments, [ABA] = 90–150 μM, [MOPS] = 100–150 mM (pH = 7.5), [BnSH] ≅ 1.6 mM.

Peptide fragment condensations at low temperature were performed similiarly to the reactions described above, except that the temperature was maintained at 5 °C. Final concentrations: [BnSH] ≈ 1.6 mM, [MOPS] = 139 mM (pH 7.50), [ABA] = 93 μM, [**E**] = 166 μM, [**N**] = 168 μM; [**T**] = 0, 9, 27, and 55 μM, respectively.

HPLC Analysis: RP-HPLC analysis was carried out using a Zorbax C-8 300SB column connected to a Hitachi D-7000 diode array HPLC system. The initial buffer (A) consisted of acetonitrile/water 1:99 containing 0.1% trifluoroacetic acid; the final buffer (B) of acetonitrile/water 90:10 contained 0.07% trifluoroacetic acid. The following gradient program was used: 10 → 25% B within 4 min, 25 → 30% B within 5 min, 30 → 50% B within 2 min, 50% B for 4 min, 50 → 100% B within 2 min, 100% B for 3 min, and 100 → 10% B within 2 min. The flow rate was 1.5 $mL\,min^{-1}$. The injection volume was 100 μL. HPLC peaks were detected by monitoring the UV absorbance at λ = 270 nm. Known amounts of peptide (concentration measured by standardized amino acid hydrolysis) were calibrated against known amounts of the internal standard 4-acetamidobenzoic acid. Peptide concentrations during the course of the reaction were determined relative to the internal standard.

Simulations: The experimental data were analyzed according to the empirical reaction models described in the text using the program SimFit. Nonlinear curve fitting by least squares was achieved using the Simplex algorithm followed by the Newton-Raphson algorithm. For screening the reaction order only the Simplex algorithm was used.

Acknowledgments: We thank the Medical Research Council of Canada for a predoctoral fellowship to DHL, Ministry of Education and Science of Spain for a postdoctoral fellowship to JAM, and Deutsche Forschungsgemeinschaft for a postdoctoral fellowship to KS.

Received: February 17, 1997 [F617]

[1] a) M. Eigen, P. Schuster, *The Hypercycle. A Principle of Natural Self-Organization*, Springer, Berlin, **1979**; b) G. von Kiedrowski, *Bioorg. Chem. Front.* **1993**, *3*, 113–146; c) S. A. Kauffman, *The Origins of Order*, Oxford Univ. Press., New York, **1993**.

[2] For recent reviews see ref. [2a–c]. a) B. G. Bag, G. von Kiedrowski, *Pure Appl. Chem.* **1996**, *68*, 2145–2152; b) E. A. Wintner, J. Rebek, Jr., *Acta Chim. Scand.* **1996**, *50*, 469–485; c) L. E. Orgel, *Nature* **1992**, *358*, 203–209; d) S. Pitsch, R. Krishnamurthy, M. Bolli, S. Wendeborn, A. Holzner, M. Minton, C.

Lesueur, I. Schlönvogt, B. Jaun, A. Eschenmoser, *Helv. Chim. Acta* **1995**, *78*, 1621–1635; e) E. A. Wintner, B. Tsao, J. Rebek, Jr., *J. Org. Chem.* **1995**, *60*, 7997–8001; f) F. M. Menger, A. V. Eliseev, N. A. Khanjin, M. J. Sherrod, *ibid.* **1995**, *60*, 2870–2878; g) D. Sievers, G. von Kiedrowski, *Nature* **1994**, *369*, 221–224; h) T. Li, K. C. Nicolaou, *ibid.* **1994**, *369*, 218–221; i) F. M. Menger, A. V. Eliseev, N. A. Khanjin, *J. Am. Chem. Soc.* **1994**, *116*, 3613–3614; j) E. A. Wintner, M. M. Conn, J. Rebek Jr., *ibid.* **1994**, *116*, 8877–8884; k) J. S. Nowick, Q. Feng, T. Tjivikua, P. Ballester, J. Rebek, Jr., *ibid.* **1991**, *113*, 8831–8839; l) G. von Kiedrowski, B. Wlotzka, J. Helbing, M. Matzen, S. Jordan, *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 423–426; m) T. Tjivikua, P. Ballester, J. Rebek, Jr., *J. Am. Chem. Soc.* **1990**, *112*, 1249–1250; n) G.von Kiedrowski, *Angew. Chem. Int. Ed. Engl.* **1986**, *25*, 932–935.

[3] D. H. Lee, J. R. Granja, J. A. Martinez, K. Severin, M. R. Ghadiri, *Nature* **1996**, *382*, 525–528.

[4] a) F. K. Junius, S. I. O'Donoghue, M. Nilges, A. S. Weiss, G. F. King, *J. Biol. Chem.* **1996**, *271*, 13663–13667; b) O. D. Monera, N. E. Zhou, P. Lavigne, C. M. Kay, R. S. Hodges, *J. Biol. Chem.* **1996**, *271*, 3995–4001; c) D. N. Woolfson, T. Alber, *Protein Sci.* **1995**, *4*, 1596–1607; d) H. Wendt, C. Berger, A. Baici, R. M. Thoms, H. R. Bosshard, *Biochemistry* **1995**, *34*, 4097–4107; e) P. B. Harbury, P. S. Kim, T. Alber, *Nature* **1994**, *371*, 80–83; f) P. B. Harbury, T. Zhang, P. S. Kim, T. Alber, *Science* **1993**, *262*, 1401–1407; g) B. Lovejoy, S. Choe, D. Cascio, D. K. McRorie, W. F. DeGrado, D. Eisenberg, *ibid.* **1993**, *259*, 1288–1293; h) N. E. Zhou, C. M. Kay, R. S. Hodges, *Biochemistry* **1992**, *31*, 5739–5746; i) E. K. O'Shea, J. D. Klemm, P. S. Kim, T. Alber, *Science* **1991**, *254*, 539–544; j) J. C. Hu, E. K. O'Shea, P. S. Kim, R. T. Sauer, *ibid.* **1990**, *250*, 1400–1403.

[5] a) Y. Yu, O. D. Monera, R. S. Hodges, P. L. Privalov, *Biophys. Chem.* **1996**, *59*, 299–314; b) P. Lavigne, L. H. Kondejewski, M. E. Houston, Jr., F. D. Sönnichsen, B. Lix, B. D. Sykes, R. S. Hodges, C. M. Kay, *J. Mol. Biol.* **1995**, *254*, 505–520; c) K. J. Lumb, P. S. Kim, *Science* **1995**, *268*, 436–439; d) O. D. Monera, C. M. Kay, R. S. Hodges, *Biochemistry* **1994**, *33*, 3862–3871; e) N. E. Zhou, C. M. Kay, R. S. Hodges, *J. Mol. Biol.* **1994**, *237*, 500–512; f) E. K. O'Shea, K. J. Lumb, P. S. Kim, *Curr. Biol.* **1993**, *3*, 658–667; g) T. J. Graddis, D. G. Myszka, I. M. Chaiken, *Biochemistry* **1993**, *32*, 12664–12671; h) E. K. O'Shea, R. Rutkowski, W. F. Stafford III, P. S. Kim, *Science* **1989**, *245*, 646–648.

[6] For reviews on template effects in organic chemistry see "Templating, Self-Assembly, and Self-Organization" in *Comprehensive Supramolecular Chemistry*, *Vol. 9* (Eds.: J. Atwood, J. D. Davies, D. D. Macnicol, F. Vögtle), Pergamon, Oxford, New York, Tokyo, **1996**, and the following selected manuscripts: a) R. K. Bruick, P. E. Dawson, S. B. H. Kent, N. Usman, G. F. Joyce, *Chem. Biol.* **1996**, *3*, 49–56; b) C. A. Hunter, *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 1079–1081; c) R. Hoss, F. Vögtle, *ibid.* **1994**, *33*, 375–384; d) S. Anderson, H. L. Anderson, J. K. M. Sanders, *Acc. Chem. Res.* **1993**, *26*, 469–475; e) T. R. Kelly, G. J. Bridger, C. Zhao, *J. Am. Chem. Soc.* **1990**, *112*, 8024–8034.

[7] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776–779.

[8] The reaction order p of the autocatalytic pathway as well as the observed rate enhancement are expected to depend on the initial peptide concentrations. For a detailed discussion see ref. [1b] and: D. N. Reinhoudt, D. M. Rudkevich, F. de Jong, *J. Am. Chem. Soc.* **1996**, *118*, 6880–6889.

[9] The calculated rate enhancements are based on linear fits of the theoretical curves obtained by SimFit for the first 10 min of the reaction. This method includes all data points and is therefore more accurate then a linear fit of the initial data points in each experiment.

[10] The goodness of fit for simulations made with SimFit is characterized by the RMS (root mean square) factor. An RMS factor below 2% is considered an excellent fit, below 3% a good fit, and below 5% a reasonable fit.

[11] H. Hojo, S. Aimoto, *Bull. Chem. Soc. Jpn.* **1991**, *64*, 111–117.

[12] G. W. Anderson, J. E. Zimmerman, F. M. Callahan, *J. Am. Chem. Soc.* **1964**, *86*, 1839–1842.

[13] Under Marfey's conditions ($NaHCO_3$, 1 h, 30–40 °C) the thioester gave 5–10% epimerization; P. Marfey, *Carlsberg Res. Commun.* **1984**, *49*, 591–596.

[14] M. Schnolzer, P. Alewood, A. Jones, D. Alewood, S. B. H. Kent, *Int. J. Pept. Protein Res.* **1992**, *40*, 180–193.

0947-6539/97/0307-1024 $ 17.50+.50/0