Organic & Biomolecular Chemistry

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

A self-replicating peptide nucleic acid†

Tobias A. Plöger* and Günter von Kiedrowski

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
DOI: 10.1039/b000000x

While the non-enzymatic ligation and template-directed synthesis of peptide nucleic acids (PNA) were reported since 1995 a case of self-replication of PNA has not been achieved yet. Here, we present evidence for autocatalytic feedback in a template directed synthesis of a self-complementary hexa-PNA from two trimeric building blocks. The course of the reaction was monitored in the presence of increasing initial concentrations of product by RP-HPLC. Kinetic modeling with the *SimFit* program revealed parabolic growth characteristics. The observed template effect, as well as the rate of the ligation, was significantly influenced by nucleophilic catalysts, pH value, and uncharged co-solvents. Systematic optimization of the reaction conditions allowed us to increase the autocatalytic efficiency of the system by two orders of magnitude. Our findings contribute to the hypothesis that PNA may have served as a primordial genetic molecule and was involved in a potential precursor of a RNA world.

Introduction

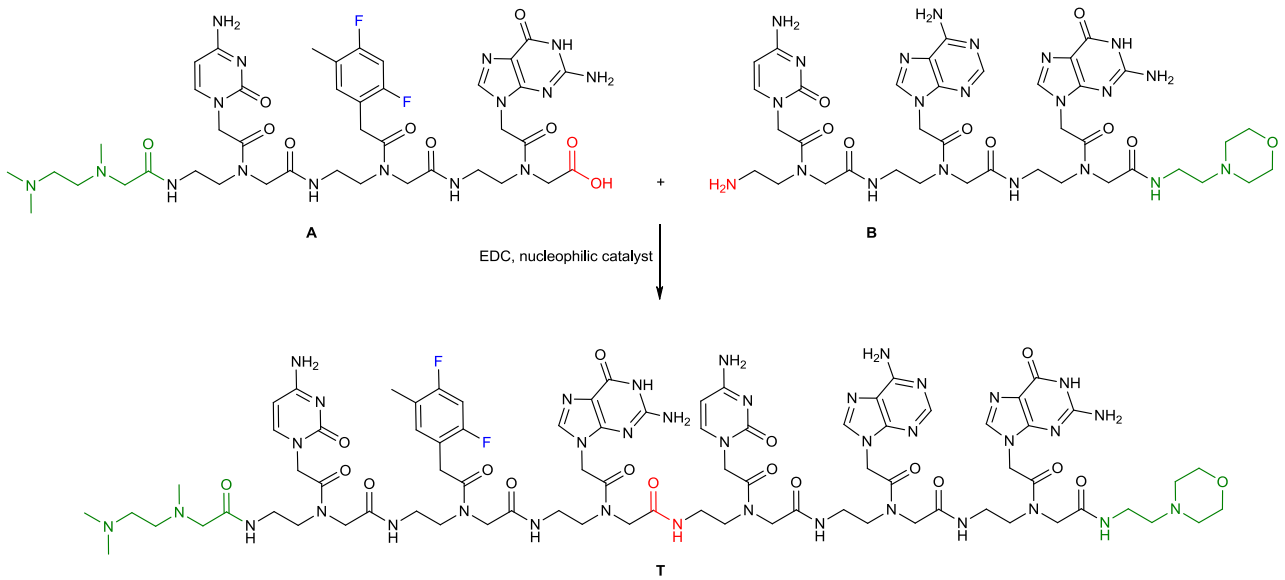

Fig. 1 Building blocks for a self-replicating system based on PNA. Trimeric building blocks **A** and **B** give the self-complementary hexa-PNA **T** upon condensation. Green: solubility enhancer; red: ligation site; blue: fluorine label.

PNA (1), a DNA or RNA mimic based on a non-charged, achiral, and pseudopeptidic backbone, as well as related peptide nucleic acid hybrids have been discussed as candidates for a primordial genetic material[1-3] preceding RNA[4-8]. PNA was recently proposed for the design of 'protocells' that are not based on chemistry occurring in today's biosphere.[9-11] Building blocks for PNA and related peptide nucleic acids were identified in prebiotic model experiments[12] and amongst the organic components of the *Murchison* meteorite[13]. Furthermore, when ligated to RNA, the achiral backbone offers a path for a gradual transformation to a homochiral polymer while avoiding enantiomeric cross-inhibition.[14] Finally, chemical ligation and the transfer of sequence information from PNA to RNA and from DNA to PNA have been demonstrated as well as the formation of DNA-PNA-chimeras on PNA- or DNA-templates, respectively.[15-25] Cases for autocatalytic and/or cross-catalytic PNA (self-)replication as a

prerequisite for natural selection and evolution have, however, not been identified so far.

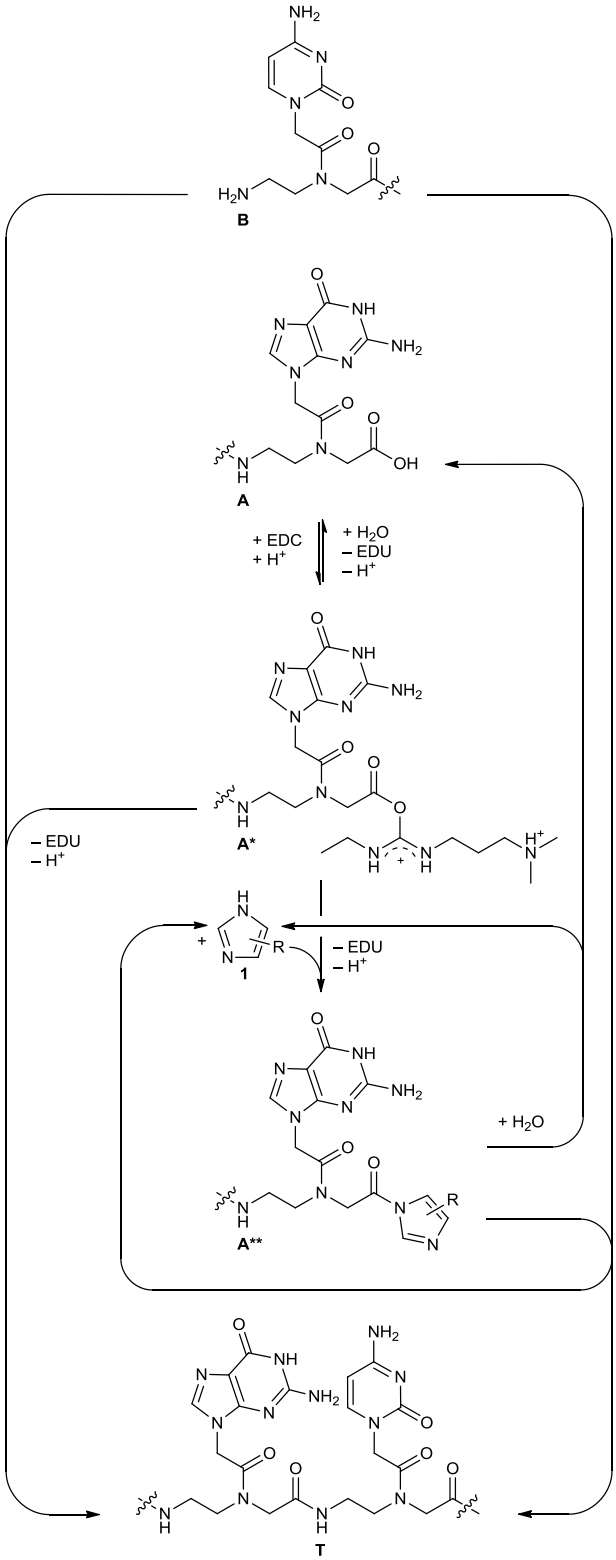

Fig. 2 Reaction network proposed for EDC mediated and imidazole catalyzed PNA ligation. The new amide bond is formed from the reaction of the amine **B** with the EDC activated carboxylic acid **A*** and/or with the imidazolide **A**** derived from the latter. Reactions of this type can occur for all carboxylic acid molecules, regardless whether they are involved in complexes or not.

Recently, we designed a system to evaluate potential PNA self-replication and presented an efficient large-scale synthesis of the building blocks needed.[26] The system consists of two PNA trimers **A** and **B** leading to the self-complementary hexameric PNA **T** upon ligation (Fig. 1), thus enabling a comparison with earlier systems from our laboratory.[27-31] Natural thymine was replaced by 2,4-difluorotoluene[32, 33] to introduce a ^{19}F-NMR probe which should allow kinetic NMR titration[34] at a later stage. Furthermore, the C- and N-terminal ends were modified with backbone extending solubility enhancers which are protonated and thus positively charged at physiological pH while avoiding a bias towards homochiral helicity caused by the influence of stereogenic centers.[35]

Results

Design of the study

The typical situation in a study of molecular replication is that in the beginning the experimenter has only little information under which conditions self-replication might show up. While the theory of minimal replicators makes predictions how the replication optimum depends on the temperature,[36] the concentration of building blocks, as well as the thermodynamic properties of complexes involved in the course of the reaction, there are dependencies which are barely predictable, especially if data are missing. In the case of our self-complementary hexamer, we even did not know the melting temperature of the duplex as the UV melting curve did not reveal a sigmoidal profile which could be analyzed by standard means (see Fig. S1 in ESI†). In such situations it is recommendable to build the kinetic analysis on a reaction model that is general enough to allow data extraction and parameter comparison for a wide range of conditions. If applicable, such macrokinetic screening allows the inclusion of experiments, in which the effect of various additives such as activating agents, nucleophilic catalysts, salts or crowding reagents become the subject of a quantitative comparison. For the analysis, we selected a reaction model which consists of three apparent reaction channels:

$$\mathbf{A}^* + \mathbf{B} \xrightarrow{k_1} \mathbf{T} \quad (1)$$

$$\mathbf{A}^* + \mathbf{B} + 0.5\,\mathbf{T} \xrightarrow{k_2} 1.5\,\mathbf{T} \quad (2)$$

$$\mathbf{A}^* \xrightarrow{k_3} \mathbf{A} \quad (3)$$

The model assumes that template molecules **T** are synthesized from an activated form of building block **A** (**A*** or **A****) and from its counterpart **B** via a bimolecular reaction (k_1) and a template-directed pathway (k_2), the latter fulfilling the square-root law of autocatalysis thus fixing parabolic growth dynamics. In addition, a first order decay of activated species **A*** has been considered (k_3) accounting for hydrolysis. Note that a full kinetic description of events preceding the formation of template molecules (Fig. 2) would include the activation of the carboxylic acid **A** by water-

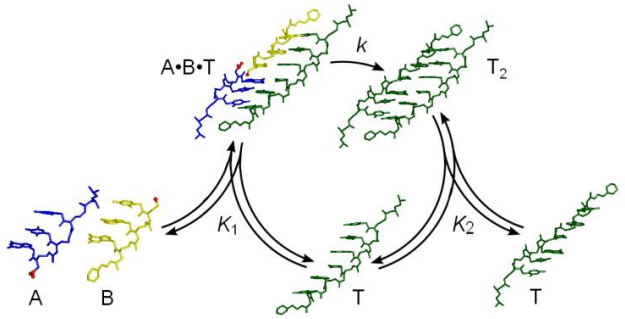

Fig. 3 Minimal model for self-replication. Template strand **T** catalyzes the formation of an identical copy from building blocks **A** and **B** in an autocatalytic ligation reaction (*k*). Ligation implies that **A** is in its activated form (viz. either **A*** or **A****).

soluble 1-ethyl-3-[3-(dimethylamino)propyl]carbodiimide (EDC), the attack of the transient *O*-acylisouronium form **A*** by nucleophilic catalysts, as well as individual hydrolysis reactions.[36] Neither of these is considered in our model, in which the parameters k_1-k_3 reflect all variations. This treatment is permitted if the activation by EDC as well as the transformation into activated conjugates **A**** are always fast and quasi-reversible reactions (due to hydrolysis) establishing the stationary concentrations of all activated forms within some minutes. Under these circumstances parameters k_1 and k_2 particularly depend on the concentration of the carbodiimide as well as the nucleophilic catalyst, while k_3 reflects the hydrolysis of the carbodiimide mediated by carboxylic acids. The modeling as a pseudo-first order process is permitted so long as the degree of activation is small and the consumption of building block **A** is negligible. The latter assumption will not hold in all cases, but from a pragmatic view it is better to account for hydrolysis approximated as a first order process than to overlook hydrolysis completely. In any case, while a three parameter model is a justifiable simplification a full modeling of all events would definitely surcharge our experimental data. The square-root law implicit in equation (2) is expected if self-replication proceeds as outlined in Fig. 3. Here, template **T** undergoes complexation (K_1) with its precursors **A** (or **A***/**A****) and **B** to yield a termolecular complex [**A•B•T**] which is then ligated in a pseudo-unimolecular fashion (*k*) leading to template duplex [**T•T**]. Cycle completion requires reversible dissociation (K_2) generating two template molecules **T**, which may both enter another round of replication. Conditions for a square-root law and thus parabolic growth are fulfilled, if

$$\sqrt{2K_2c} >> K_1ab >> 1 \qquad (4)$$

in which *a*, *b*, and *c* hold the total concentrations of precursors **A** and **B** as well as product **T**, respectively. For the special case of complete activation, k_2 can be rationalized in terms of elementary rate and equilibrium constants:

$$k_2 = \frac{kK_1}{\sqrt{2K_2}} \qquad (5)$$

The model Eq. 1-3 is then equivalent with the following differential equation:

$$\frac{\mathrm{d}c}{\mathrm{d}t} = (a_0 - c)(b_0 - c)[k_1 + k_2(c + c_\mathrm{T})^p]\exp(-k_3 t) \qquad (6)$$

Table 1 Kinetic data for the optimization of PNA self-replication at 5 mM tri-PNAs

Entry	Buffer	pH	Catalyst (M)	Salt	EDC (M)	PEG	T(°C)	RMS (%)	ε ($M^{-1/2}$)	k_1 (M^{-1} s^{-1})
1	1-MeIm	7.5	buffer	-	0.2	-	0	1.21	0.1±0.6	(2.64±0.07)×10^{-4}
2	Im	7.5	buffer	-	0.2	-	0	0.896	11±2	(3.1±0.2)×10^{-4}
3	1-MeIm	7.5	buffer	-	0.2	-	10	0.722	4.3±0.6	(5.4±0.1)×10^{-4}
4	Im	7.5	buffer	-	0.2	-	10	1.25	25±3	(3.6±0.2)×10^{-4}
5	MOPS	7.2	-	NaCl	0.2	-	10	0.826	(8±1)×10^{1}	(1.8±0.1)×10^{-4}
6	MOPS	7.2	0.1 Im	NaCl	0.2	-	10	1.46	38±3	(2.6±0.2)×10^{-4}
7	MOPS	7.2	0.1 1-MeIm	NaCl	0.2	-	10	1.24	0.2±3.2	(5.3±0.3)×10^{-4}
8	MOPS	7.2	0.1 Py	NaCl	0.2	-	10	0.606	15±2	(3.8±0.1)×10^{-4}
9	MOPS	7.2	0.1 HOAt	NaCl	0.2	-	10	7.94	0.09±4.50	(3.3±0.6)×10^{-2}
10	MOPS	7.6	0.1 Im	-	0.2	-	10	0.650	60±5	(1.83±0.08)×10^{-4}
11	MOPS	7.6	0.1 Im	NaCl	0.2	-	10	0.993	52±6	(2.3±0.1)×10^{-4}
12	MOPS	7.6	0.1 Im	Na_2SO_4	0.2	-	10	1.04	(7±1)×10^{1}	(1.3±0.1)×10^{-4}
13	MOPS	7.6	0.1 Im	NaI	0.2	-	10	0.821	79±9	(2.0±0.1)×10^{-4}
14	MOPS	7.6	0.1 Im	NaCl	0.4	-	10	0.667	52±4	(5.2±0.2)×10^{-4}
15	MOPS	6.6	0.1 Im	NaCl	0.2	-	10	2.79	(1.7±0.6)×10^{2}	(3.5±0.8)×10^{-4}
16	MOPS	6.6	0.1 Im	-	0.4	-	10	2.34	72±9	(1.3±0.1)×10^{-3}
17	MOPS	7.2	0.1 Im	NaI	0.2	-	10	0.850	45±5	(6.9±0.3)×10^{-4}
18	MOPS	7.2	0.2 Im	NaI	0.2	-	10	1.16	44±6	(5.4±0.3)×10^{-4}
19	MOPS	7.6	0.1 Im	NaCl	0.2	-	-19	0.978	11±4	(1.70±0.09)×10^{-4}
20	MOPS	7.6	0.1 Im	NaCl	0.2	400	10	0.592	(2.6±0.4)×10^{2}	(5.3±0.5)×10^{-5}
21	MOPS	7.6	0.1 Im	NaCl	0.2	3350	10	1.38	(1.2±0.2)×10^{2}	(2.1±0.2)×10^{-4}

[a] Concentrations: 0.4 M imidazole buffer, 0.2 M MOPS buffer, 0.2 M Salt, and 20 wt% PEG. The agreement of the experimental data with the simplified reaction model is measured by root mean square (RMS) which can be understood as the average deviation of experimental and theoretical values in percent. Error margins reflect the mathematical error during the simulation.

Here, a_0, b_0, and c_T, are the initial concentrations of **A**, **B**, and **T**, while *c* is the concentration of synthesized ligation product **T** at time *t*. For a square-root law, the autocatalytic reaction order *p* is fixed to ½. This empirical rate equation was applied for data analysis in a previously reported self-replicating system based on oligonucleotides.[29] We followed the time course of the ligation

of building blocks **A** and **B** in aqueous buffered solutions by RP-HPLC. The reaction was monitored in the presence of different amounts of initially added ligation product **T**. Concentrations were derived from HPLC integrals after calibration. The resulting data points were then simultaneously approximated by non-linear fitting using our *SimFit* program.[37] Occurrence and efficiency of potential self-replication under different reaction conditions has been determined by the autocatalytic efficiency ε (Eq. 7), which provides the information how much faster the autocatalytic production of templates proceeds compared to the second order ligation if the template were at 1 M concentration:

$$\varepsilon = \frac{k_2}{k_1} \qquad (7)$$

Note that ε should never be used to compare systems which operate in different ranges of template concentration. Experiments on ribozyme self-replication in µM concentration will always result in high ε values compared to small organic replicators operating in mM concentrations regardless which one is more efficient. In addition *SimFit* was instructed to optimize the initial template concentration c_i over a range of maximally ±20 % of the actual value. The rationale behind is the compensation of possible errors during pipetting, sampling, and HPLC integration which are expected to exhibit a large influence for the calculation of the theoretical template concentration. This is partly bound to a so-called *Cauchy* problem as the theoretical concentrations are derived from the initially measured values. A statistical deviation of the initial value will otherwise lead to an incorrect approximation of all later values. Typically, the computed concentrations differed by not more than 5 % from the setpoint value (see ESI† for a comparison of experimental and optimized values). The influence of this approach on the reliability of the fitting as a whole is discussed below.

Systematic optimization of the reaction conditions

The reaction conditions were optimized by varying temperature, buffer, pH, salt, EDC concentration, addition of uncharged co-solvents, and the nature and concentration of a nucleophilic catalyst. The latter one may be identical with the buffer substance as in the case of imidazole (Im) buffers, or is added to a non-nucleophilic buffer (3-(*N*-morpholino)propanesulfonic acid (MOPS)). While the influence of nucleophilic heterocycles on the efficiency of template-directed reactions were studied for the case of RNA and DNA as ligators[38-40] little is known about nucleophilic catalysis of PNA ligation[15, 25]. Table 1 reveals imidazole to be the most capable catalysts in terms of autocatalytic efficiency, while 1-methyl-imidazole (1-MeIm) and 1-hydroxy-7-azabenzotriazole (HOAt) favoured hexamer formation via the non-instructed reaction channel. The latter, as well as pyridine (Py), could rather be identified as efficient catalysts for PNA ligation at room temperature (see Fig. S7 in ESI†) and therefore offer an alternative to methods which use native chemical ligation or the related native chemical *i*Cys-ligation.[19, 21, 22] In the absence of any added catalysts, replication was clearly observable (entry 5), albeit significant amounts of side products were detected (see Table S7 in ESI†). These could be suppressed effectively by adding any of the heterocyclic catalysts studied. Interestingly, the interplay between ligation rate and autocatalytic efficiency is more complex than in the case of DNA replicators where the autocatalytic efficiency was found to be enhanced by faster ligation chemistries.[27-29] From theoretical considerations,[36] one can expect a parabolic replicator to work at its maximum rate if the temperature is adjusted between the melting temperatures of the termolecular complex and the product duplex because, both, the rate constant *k*, and the equilibrium concentration of the termolecular complex [**A•B•T**] depend on the temperature: The rate constant of the ligation is expected to increase with the temperature, while the equilibrium concentration of [**A•B•T**] is expected to decrease due to the weakening of forces which stabilize the complex. As mentioned above, the melting temperature of [**T•T**] could not be determined (see Fig. S1 in ESI†) due to a non-cooperative melting process which presumably results from the disturbing effect of the fluoroaromatic isostere.[41] However, a temperature of 10 °C obviously ensured a sufficient population of the productive termolecular complex, while experiments at lower temperatures (entries 1, 2, and 19) or room temperature (see e.g. Table S5 in ESI†) proved less efficient. Although PNA duplexes are typically independent to salt concentration owing to the non-charged backbone,[42] we found a slight influence of different anions on the replication system (entries 10-13) which can be attributed to the positively charged terminal modifications. Entries 11 and 14 show that the rate of the ligation is limited by the hydrolysis of the condensing agent: Increasing the EDC concentration from 0.2 to 0.4 M doubles the rate of hexa-PNA formation, while not biasing the autocatalytic efficiency ε. Furthermore, rate and yield of the ligation increase with decreasing pH, as it can be expected for an EDC mediated peptide coupling (entries 6, 11 and 15). In terms of autocatalytic efficiency, pH 6.6 was the best of the three pH values tested (entry 15). Subsequently, the combination of pH 6.6 and 0.4 M EDC allowed to establish reaction conditions optimized for fast replication and high conversion (entry 16, Fig. S11 in ESI†). Worth mentioning, the hydrolysis of the coupling reagent EDC is effectively catalyzed by carboxylic acids whose concentration is not uniform in all experiments: While the concentration of PNA carboxylic acid **A** is kept constant at 5 mM in all experiments, the quantity of trifluoroacetic acid (TFA) correlates with the initial template concentration because all PNAs were used as trifluoroacetates owing to HPLC purification. In fact, the model accounts for EDC hydrolysis in general (Eq. 3) but does not consider its dependency on TFA concentration. Increasing the imidazole concentration from 0.1 to 0.2 M gave no significant change in the autocatalytic efficiency ε, while slightly reducing the rate of the ligation (entries 17-18). In the following, we explored the self-replicating system in the eutectic phase (entry 19) as this is known to favor aggregates and organized structures for the case of RNA.[43] Here, hydrolysis is repressed due to a lower activity of water. In our case, eutectic freezing led to even a decrease of the autocatalytic synthesis when compared to the analogous experiment at 10 °C (entry 11). Finally, the influence of uncharged co-solvents on the self-replicating system was studied because they were considered to have an impact on the structures and stabilities of the complexes involved (entries 20-21).[44, 45] In fact, both, the addition of polyethylene glycol

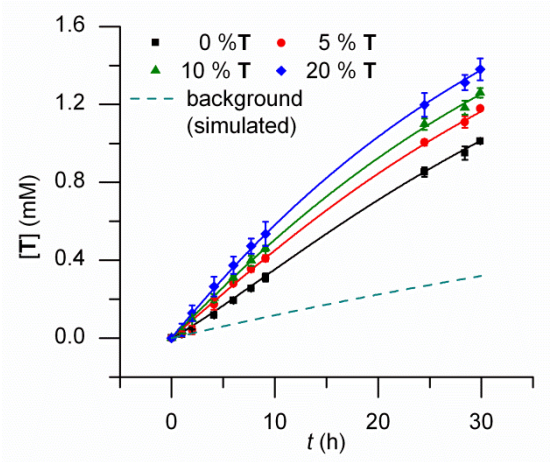

Fig. 4. Effect of increasing initial template addition on the kinetics of the ligation reaction. Experimental data points and theoretical curves derived from kinetic fitting (ε = 51±2 $M^{-1/2}$, 0.650 % RMS) according to equations (1)-(3). The approximated initial template concentrations were subtracted from the experimental values to enable comparison of template production. Reactions were carried out in the presence of the amount of template **T** shown in the legend using the reaction conditions of entry 11 in Table 1. Error bars indicate the standard deviation due to experimental inaccuracies and HPLC integration over three independent series of experiments. The resulting rate constants and error estimates are k_1 = (2.76±0.06)×10^{-4} M^{-1} s^{-1}, k_2 = (1.40±0.03)×10^{-2} $M^{-3/2}$ s^{-1}, and k_3 = (3.2±0.2)×10^{-6} s^{-1}. Initial template concentrations were optimized to c_1 = (0.262±0.004) mM, c_2 = (0.514±0.004) mM, and c_3 = (0.972±0.005) mM.

3350 (PEG3350) and 400 (PEG400) improved the template effect: While 20 % PEG3350 increased the autocatalytic efficiency ε from 52 (entry 11) to 120 $M^{-1/2}$, 20 % PEG400 lead to the overall optimum value of ε for the experiments presented (ε = 260 $M^{-1/2}$). Both PEGs led to an enhanced production of side-products however (see Table S22 in ESI†). Following the screening, we decided to elaborate an illustrative replication assay based on the conditions presented in entry 11 which, while not exhibiting the highest value for ε, served as a reference for many of the variations presented above (Fig. 4). To clearly indicate the relative contributions of both pathways to the observed product formation we also simulated the concentration of template that would be produced in the absence of any autocatalysis.

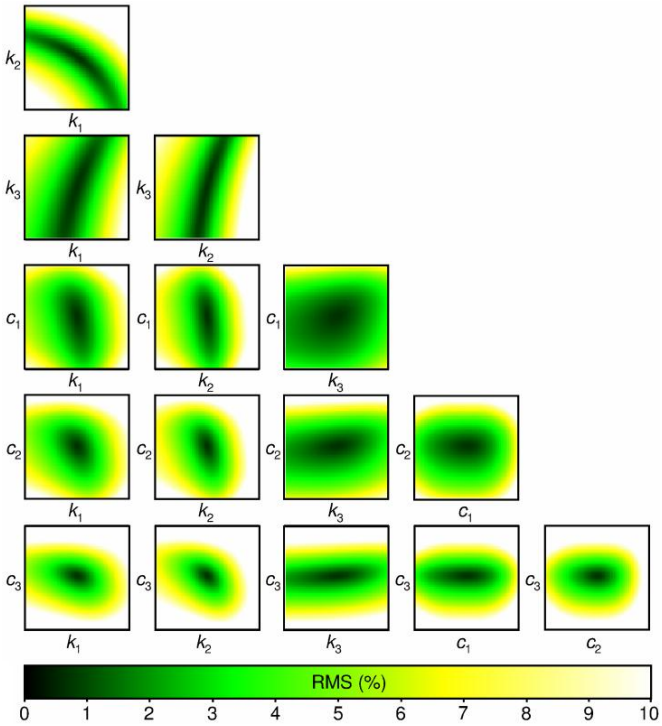

Fig. 5 Visualization of the error hypersurface of the 6 parameter fitting by its 15 possible projection planes. Details are described in the text.

We expanded our *SimFit* program by a method (called 'scan') allowing visualization of the error hypersurface based on the mapping of RMS to all possible projection planes spanned by two parameters (Fig. 5). Generally, for N parameters there are N×(N-1)/2 of such planes. Logarithmic scaling on each plane's axis was applied to view RMS as a function of rate parameters scaling linearly with the corresponding free energies of activation. Similarly, initial concentrations of template **T** (c_1-c_3) were also varied logarithmically (like in the case of pH values). Each axis ranges from the 0.5 fold (1.03^{-25}) to the 2.09 fold (1.03^{+25}) of the respective pair of parameters varied to calculate RMS. Each plane shows RMS for optimized parameters in the very centre. Three planes illustrate three typical cases: The k_1,k_2-plane exemplifies a negative slope of covariation. Here, increasing one parameter necessitates decreasing the other to stay close to the minimum. The opposite is revealed by the k_1,k_3-plane where both parameters need to be increased or decreased to keep low RMS. The c_2,c_3-plane indicates that both initial template concentrations are determined rather independently from each other as a change in any direction leads to an increase of RMS. The latter one basically also applies for most mixed planes of the k_1,c_i- and k_2,c_i-type. This is in a way counter-intuitive as the increase of the number of fit parameters is usually associated with the risk to surcharge the experimental data and thereby to create a moment of arbitrariness reflected by elevated covariation. Here, error compensation by additional optimization of initial template concentrations c_1-c_3 is even more reliable in this term than the extraction of independent rate parameters k_1-k_3. In fact, further addition of three more fit parameters a_1-a_3 accounting for the initial concentration of building block **A** did not change the reliability of the fitting and the conclusions derived from it (see Table 2 and Fig. S15 in ESI†). Therefore, we tend to believe for

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Table 2 Kinetic data for three different types of fittings

Parameters	RMS (%)	k_1 (M^{-1} s^{-1})	k_2 ($M^{-3/2}$ s^{-1})	k_3 (s^{-1})	ε ($M^{-1/2}$)
3: k_1-k_3	0.800	$(2.7\pm0.1)\times10^{-4}$	$(1.38\pm0.07)\times10^{-2}$	$(2.5\pm0.4)\times10^{-6}$	51±5
6: k_1-k_3 and c_1-c_3	0.650	$(2.76\pm0.06)\times10^{-4}$	$(1.40\pm0.03)\times10^{-2}$	$(3.2\pm0.2)\times10^{-6}$	51±2
9: k_1-k_3, c_1-c_3, and a_1-a_3	0.520	$(2.68\pm0.05)\times10^{-4}$	$(1.49\pm0.04)\times10^{-2}$	$(3.6\pm0.3)\times10^{-6}$	56±3

[a] 5 mM tri-PNAs **A** and **B**, 0.2 M EDC, 0.2 M MOPS pH 7.6, 0.2 M NaCl, 0.1 M imidazole, 16 µl scale.

such systems that the arbitrariness of parameter extraction in general is not increased by co-optimization of initial concentrations.

Conclusions

We have presented evidence that a short self-complementary PNA sequence is capable to promote its autocatalytic self-replication. The synthesis in the autocatalytic channel was optimized by a kinetic screening of reaction conditions leading to the finding that macromolecular crowding using PEG proved as a suitable means to increase self-replication. While the length of the PNA-hexamer is subcritical to determine thermodynamic data on complex formation, we believe that the effect of crowding can be understood as increasing the population of productive complexes. The kinetics are consistent with the square-root law of autocatalysis expected for parabolic growth which was detected in DNA systems of the same length.[27-29] In an independent study *Singhal* and *Nielsen* report on cross-catalytic systems of various template lengths providing evidence that systems using tetrameric building blocks which react to yield octameric templates are more efficient than systems based on penta- or hexameric building blocks.[46] Kinetic results from a previously studied DNA system in which cross-catalysis and autocatalysis could be observed simultaneously revealed no differences in the efficiencies of autocatalytic and cross-catalytic pathways (30, 37). Furthermore, minimal replicator theory predicts a replication rate optimum which is a function of the thermodynamic properties of template associated complexes.[36] Taken together, these findings mean that our hexameric system is most likely slightly below optimal thermodynamic stability while *Nielsen's* system based on an octameric template is slightly above. Both studies are nevertheless complementary supporting the potential prebiotic role of PNA as a primordial replicator. Future studies may ask whether or not structural modifications on the PNA skeleton are conceivable that could allow to combine the principle of self-replication with organocatalysis[47] and the emergence of biomolecular homochirality[14, 35, 48]. In this context, it will be worthwhile to explore different ligation chemistries that are for instance based on reductive amination.[20, 49-51]

Acknowledgements

This work was supported by the EU projects PACE (FP6-IST-FET-002035), ECCell (FP7-ICT-222422), and MatchIT (FP7-ChemIT-249032), as well as the EU networks ReAd (Marie Curie) and Systems Chemistry (COST CM0703).

Notes and references

Ruhr-Universität Bochum, Chair of Organic Chemistry I - Bioorganic Chemistry, 44780 Bochum, Germany. Fax: +49 234 32 14355; Tel: +49 234 32 28218; E-mail: tobias@oc1.rub.de

† Electronic Supplementary Information (ESI) available: Characterization of hexa-PNA **T,** full description of experimental methods, detailed kinetic data for all replication assays and an exemplary *SimFit* command file. See DOI: 10.1039/b000000x/

1 P. E. Nielsen, *Origins Life Evol. Biosphere*, 1993, **23**, 323-327.
2 A. Eschenmoser, *Tetrahedron*, 2007, **63**, 12821-12843.
3 C. Dose, S. Ficht and O. Seitz, *Angew. Chem. Int. Ed.*, 2006, **45**, 5369-5373.
4 U. Diederichsen, D. Weicherding and N. Diezemann, *Org. Biomol. Chem.*, 2005, **3**, 1058-1066.
5 G. F. Joyce, *Nature*, 2002, **418**, 214-221.
6 T. A. Lincoln and G. F. Joyce, *Science*, 2009, **323**, 1229-1232.
7 X. Chen, N. Li and A. D. Ellington, *Chem. Biodiversity*, 2007, **4**, 633-655.
8 M. Levy and A. D. Ellington, *J. Mol. Evol.*, 2003, **56**, 607-615.
9 J. Attwater, S. Tagami, M. Kimoto, K. Butler, E. T. Kool, J. Wengel, P. Herdewijn, I. Hirao and P. Holliger, *Chem. Sci.*, 2013, **4**, 2804-2814.
10 S. Rasmussen, L. Chen, D. Deamer, D. C. Krakauer, N. H. Packard, P. F. Stadler and M. A. Bedau, *Science*, 2004, **303**, 963-965.
11 T. Rocheleau, S. Rasmussen, P. E. Nielsen, M. N. Jacobi and H. Ziock, *Phil. Trans. R. Soc. B*, 2007, **362**, 1841-1845.
12 *Protocells - Bridging Nonliving and Living Matter*, MIT Press, Cambridge, London, **2009**.
13 K. E. Nelson, M. Levy and S. L. Miller, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 3868-3871.
14 U. J. Meierhenrich, G. M. M. Caro, J. H. Bredehoft, E. K. Jessberger and W. H. P. Thiemann, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 9182-9186.
15 P. E. Nielsen, *Origins Life Evol. Biosphere*, 2007, **37**, 323-328.
16 C. Bohler, P. E. Nielsen and L. E. Orgel, *Nature*, 1995, **376**, 578-581.
17 J. G. Schmidt, P. E. Nielsen and L. E. Orgel, *J. Am. Chem. Soc.*, 1997, **119**, 1494-1495.
18 J. G. Schmidt, P. E. Nielsen and L. E. Orgel, *Nucleic Acids Res.*, 1997, **25**, 4797-4802.
19 M. Koppitz, P. E. Nielsen and L. E. Orgel, *J. Am. Chem. Soc.*, 1998, **120**, 4563-4569.
20 A. Mattes and O. Seitz, *Chem. Commun.*, 2001, 2050-2051.
21 D. M. Rosenbaum and D. R. Liu, *J. Am. Chem. Soc.*, 2003, **125**, 13924-13925.
22 C. Dose and O. Seitz, *Org. Lett.*, 2005, **7**, 4365-4368.
23 J. M. Heemstra and D. R. Liu, *J. Am. Chem. Soc.*, 2009, **131**, 11347-11349.
24 Y. Ura, J. M. Beierle, L. J. Leman, L. E. Orgel and M. R. Ghadiri, *Science*, 2009, **325**, 73-77.
25 M. Gorlero, R. Wieczorek, K. Adamala, A. Giorgi, M. E. Schininà, P. Stano and P. L. Luisi, *FEBS Letters*, 2009, **583**, 153-156.

26 T. A. Plöger and G. von Kiedrowski, *Helv. Chim. Acta*, 2011, **94**, 1952-1980.
27 G. von Kiedrowski, *Angew. Chem. Int. Ed.*, 1986, **25**, 932-935.
28 G. von Kiedrowski, B. Wlotzka and J. Helbing, *Angew. Chem. Int. Ed.*, 1989, **28**, 1235-1237.
29 G. von Kiedrowski, B. Wlotzka, J. Helbing, M. Matzen and S. Jordan, *Angew. Chem. Int. Ed.*, 1991, **30**, 423-426.
30 D. Sievers and G. von Kiedrowski, *Nature*, 1994, **369**, 221-224.
31 H. Schöneborn, J. Bülle and G. von Kiedrowski, *ChemBioChem*, 2001, **2**, 922-927.
32 B. A. Schweitzer and E. T. Kool, *J. Org. Chem.*, 1994, **59**, 7238-7242.
33 E. T. Kool and H. O. Sintim, *Chem. Commun.*, 2006, 3665-3675.
34 I. Stahl and G. von Kiedrowski, *J. Am. Chem. Soc.*, 2006, **128**, 14014-14015.
35 I. A. Kozlov, L. E. Orgel and P. E. Nielsen, *Angew. Chem. Int. Ed.*, 2000, **39**, 4292-4295.
36 G. von Kiedrowski, *Bioorg. Chem. Front.*, 1993, **3**, 113-146.
37 D. Sievers and G. von Kiedrowski, *Chem. Eur. J.*, 1998, **4**, 629-641.
38 T. Inoue and L. E. Orgel, *J. Am. Chem. Soc.*, 1981, **103**, 7666-7667.
39 M. Röthlingshöfer, E. Kervio, T. Lommel, U. Plutowski, A. Hochgesand and C. Richert, *Angew. Chem. Int. Ed.*, 2008, **47**, 6065-6068.
40 M. Röthlingshöfer and C. Richert, *J. Org. Chem.*, 2010, **75**, 3945-3952.
41 K. A. Frey and S. A. Woski, *Chem. Commun.*, 2002, 2206-2207.
42 M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden and P. E. Nielsen, *Nature*, 1993, **365**, 566-568.
43 P. A. Monnard and H. Ziock, *Chem. Biodiversity*, 2008, **5**, 1521-1539.
44 S.-i. Nakano, H. T. Karimata, Y. Kitagawa and N. Sugimoto, *J. Am. Chem. Soc.*, 2009, **131**, 16881-16888.
45 M. Z. Markarian and J. B. Schlenoff, *J. Phys. Chem. B*, 2010, **114**, 10620-10627.
46 P. E. Nielsen and A. Singhal, *Org. Biomol. Chem.*, 2014,
47 S. Kamioka, D. Ajami and J. Rebek, *Proc. Natl. Acad. Sci. USA*, 2010, **107**, 541-544.
48 M. Froeyen, F. Morvan, J.-J. Vasseur, P. Nielsen, A. Van Aerschot, H. Rosemeyer and P. Herdewijn, *Chem. Biodiversity*, 2007, **4**, 803-817.
49 J. T. Goodwin and D. G. Lynn, *J. Am. Chem. Soc.*, 1992, **114**, 9197-9198.
50 Z. Y. J. Zhan and D. G. Lynn, *J. Am. Chem. Soc.*, 1997, **119**, 12420-12421.
51 Z. Y. Li, Z. Y. J. Zhang, R. Knipe and D. G. Lynn, *J. Am. Chem. Soc.*, 2002, **124**, 746-747.