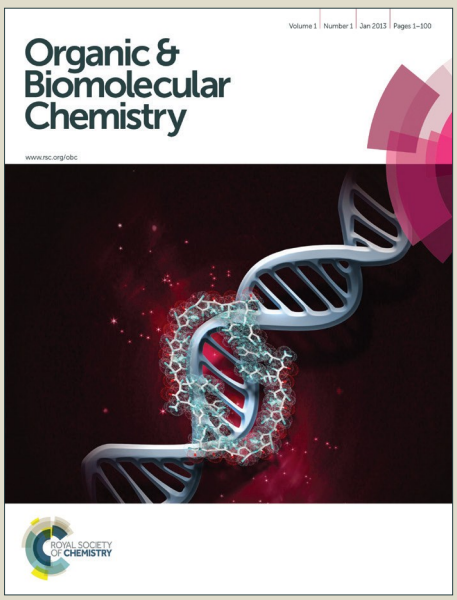
Organic & Biomolecular Chemistry

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc

Self-Replicating Systems

Gregory Clixby, Lance Twyman

Over the past 25 years, there has been a surge of development in research towards self-replication and self-replicating systems. The interest in these systems relates to one of the most fundamental questions posed in all fields of science - How did life on Earth begin? *Investigating how the self-replication process evolved may hold the key to understanding the emergence and evolution of living systems and, ultimately, gain a clear insight to the origin of life on Earth. This introductory review aims to highlight the fundamental prerequisites of self-replication along with the important research that has been conducted over the past few decades.*

1. Introduction

The ability of a system to form a perfect copy of itself - the process of replication - is central to the survival of living organisms. As self-replication is a key component in the evolution of biological life, it is envisaged that understanding more about how the mechanism works will lead to a greater understanding of how self-replicating systems have arisen and, therefore, how the generation of living systems on primitive Earth began. Furthermore, the study of self-replicating systems is of keen interest to the synthetic chemist, with the goal of creating the ultimate synthetic machine, capable of making perfect copies of itself from starting reagents, being an exciting prospect. The amount of research being conducted in this field and related areas is rapidly expanding and a vast number of papers and reviews have been published over recent years.

1.1. Why study self-replicating systems?

It is believed that the occurrence of living organisms could be due to self-replication. The theory of Darwinian evolution describes an evolutionary system as being able to metabolise, self-replicate and undergo mutations. Therefore, self-replication is one of the three criteria which distinguish living from non-living systems.[1]

Nucleic acids are candidates to be the first reproducing molecules.[1] The nucleic acids, DNA and RNA, play the cellular role of information transfer and storage. The DNA molecule is made up by the association of two nucleic acid strands in a helix whereas RNA is generally a single strand.[2] Several arguments suggest RNA may have predated DNA in evolution,[2] leading to the argument that DNA is a modified RNA molecule, altered in order to fit efficient storage and genetic information. In the 'RNA world' hypothesis for the origin of life,[3] RNAs are assumed to be the central macromolecules able to self-replicate by the processes of base-pairing, conserving information, and catalysing reactions necessary for a primitive metabolism.

When nucleic acids such as RNA and DNA replicate, the nucleic acid strands separate and, by complementarity, each one of them serves to regenerate the missing strand. In the cell, enzymes carry out this task.[2] It is unlikely that reproduction during the early phases in the development of life already featured these advanced biomolecules, therefore, under primitive conditions, we must assume that replication occurred without the intervention of enzymes, and by the simplest possible template directed synthesis.[4] However, there is a large knowledge gap between the emergence of a prebiotic Earth and the time the first organisms lived and died, and therefore there is no real clue as to the nature and origin of these original replicating systems and their properties. Therefore, chemists must create model systems to demonstrate the required principles.[5]

In the field of prebiotic chemistry, carefully designed systems serve as models for processes implicated in the origin of life. The study of such systems sheds light on prebiotic chemical evolution.[6] It is therefore hoped that by exploring different self-replicating systems of biological and synthetic origins that more light may be shed on the true identity of living systems.

Furthermore, the development of self-replicating systems represents the ultimate synthetic machine, capable of templating the production of a large number of perfect copies of itself from a single original molecule.[7] In order to develop synthetic machinery capable of directing its own synthesis and cooperating with other systems in an organised way, it is essential to develop the fundamental understanding of the recognition mediated processes which enable molecules to operate as highly efficient templates for the formation of themselves (autocatalysis) and in tandem with other molecules (cross-catalysis).[5]

1.2. What are self-replicating systems?

In order to define a self-replicating system, it is important to distinguish between the mechanism of self-replication and simple autocatalysis, which has been investigated extensively.[8] In any autocatalytic reaction, the product of a reaction is itself a catalyst for the reaction.[5] The autocatalytic cycle returns more catalyst to the reaction each time the cycle is completed and the rate of reaction accelerates until starting material runs out – the supply of precursor is the limiting factor and the reaction rate decreases when concentrations become low. Therefore, a sigmoidal rate profile of concentration vs. time is observed for an autocatalytic mechanism – a feature of exponential growth of product and decay of starting material. Conversely, a standard catalytic reaction is defined by a simple exponential rate profile (Figure 1).

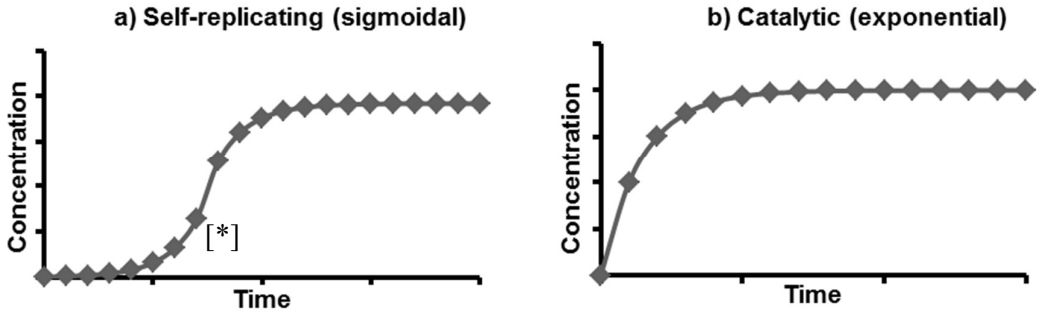
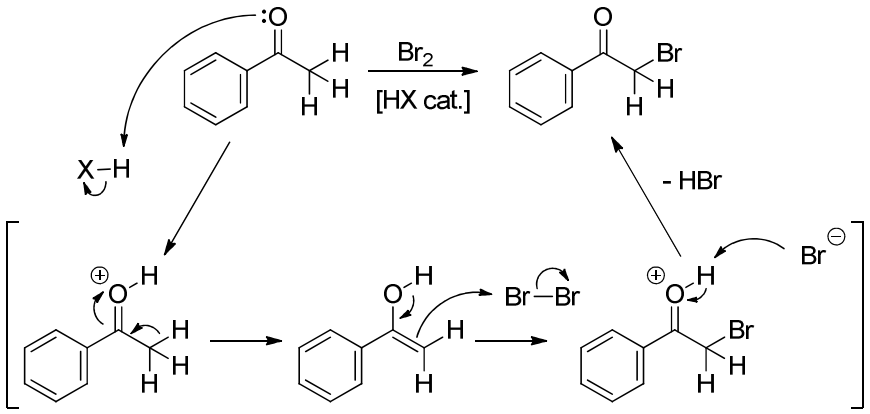


Figure 1 a) The rate profile of a self-replicating mechanism demonstrates a delay in the onset of exponential product growth [*] as catalyst is formed by the reaction; b) A standard catalytic mechanism displays exponential growth from the start of the reaction due to a constant concentration of catalyst

In order for a system to be defined as 'autocatalytic', the reaction cycle must generate and return at least some catalyst to the reaction mixture. Furthermore, general autocatalytic reactions are non-specific. For example, the α-bromination of acetophenone with bromine yields 2-bromo-1-phenylethanone along with hydrogen bromide by-product, which catalyses the reaction (Scheme 1). Though autocatalytic, the reaction may be catalysed by any Brønsted acid, and the hydrogen bromide produced will react with any Brønsted base, demonstrating a lack of specificity for this type of autocatalysis.



Scheme 1 Non-specific autocatalytic formation of α-bromoacetophenone through general HX catalyst (source of H^+)

In contrast, self-replicating reactions are a subset of autocatalytic reactions where specificity is essential – a molecule capable of self-replication will only catalyse its own formation, and may be referred to as a 'selfish catalyst'.[5] As self-replicating systems are a branch of autocatalytic systems, they still operate under a sigmoidal rate profile. A standard autocatalytic system will not operate efficiently if unwanted side products are simultaneously formed alongside the target molecule. A self-replicating molecule, however, should be able to reduce any unwanted reactions due to its specific recognition properties, leading to three possible reaction channels (Figure 2):

Channel 1 - Uncatalysed channel: The uncatalysed bimolecular reaction of **A** and **B** building blocks generates complementary **T**

Channel 2 - Autocatalytic channel: Template **T** simultaneously co-ordinates the **A** and **B** building blocks through specific molecular recognition to generate ternary complex **[A.B.T]**. The close proximity of reactive sites and higher concentration of building blocks comprising the ternary complex accelerates the reaction of **A** and **B**, which yields an additional template molecule within the product duplex **[T.T]**. The dissociation of the duplex to produce two catalytic templates ends the autocatalytic cycle. The formation of more catalyst leads to exponential template growth until all the building block molecules are consumed.

Channel 3 - [**A.B**] complex channel: Association of **A** and **B** due to the complementary nature of their recognition sites leads to the reversible formation of an [**A.B**] complex. If the

reactive sites of the building blocks are oriented correctly, the reaction of **A** and **B** could be accelerated through this pathway to form template **T'**.

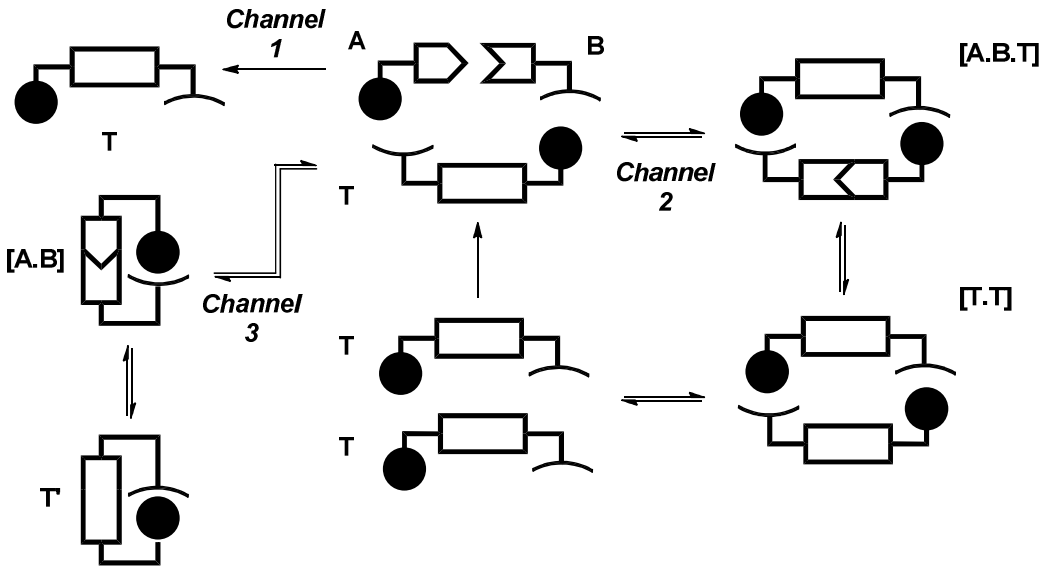


Figure 2 A minimal replicating system, demonstrating the potential pathways of formation for template **T** via uncatalysted **Channel 1**; autocatalytic **Channel 2**; or [**A.B**] Complex **Channel 3**

Binary complex formation is difficult to overcome and it has the potential of quenching the autocatalytic cycle. This is particularly the case as [**A.B**] complex formation requires only one association between two components, whereas the [**A.B.T**] ternary complex in the autocatalytic cycle requires two associations between three components.[5] The significance of the autocatalytic cycle to the overall reaction rate depends on the magnitude of the kinetics and rates within **Channel 1** compared with **Channel 2** and **Channel 3**. This leads to the definition of either a parabolic or exponential growth in template production (Equation 1).

a) *Parabolic*: $p = 1/2$; $c_t = (c_o^{1/2} + \alpha(t/2))^2$

b) *Exponential*: $p = 1$; $c_t = c_o e^{\alpha t}$

Equation 1 'p', the autocatalytic reaction order, defines the type of autocatalytic growth as parabolic, where autocatalysis is one aspect of a complex reaction mixture, or exponential, describing a system operating efficiently through the autocatalytic channel

Equation 1a) represents the parabolic growth of template **T**, whereby the rate of autocatalytic template formation is proportional to the square-root of template concentration (√[**T**]), the square-root law of autocatalysis. Equation 1b) represents the exponential growth of template **T**.[9] As stated by von Kiedrowski, *"true Darwinian selection necessitates exponential, and not parabolic, growth"*.[1] Therefore, in order to maximise the autocatalytic function of a self-replicating system, the binary complex formation should be minimised or, ideally, removed.

The additional problem of the autocatalytic cycle is that in order for template to be released back into the reaction, the [**T.T**] product duplex must dissociate. This will only occur if the [**T.T**] template duplex is less stable than the [**A.B.T**] ternary complex.[5] A system with a highly stable product duplex will not be able to complete the autocatalytic cycle, as it does not release template back into the reaction. This means that whilst the system could be considered to be self-replicating as it successfully generates a copy of itself, it would *not* be autocatalytic as it is does not deliver catalyst back to the cycle.

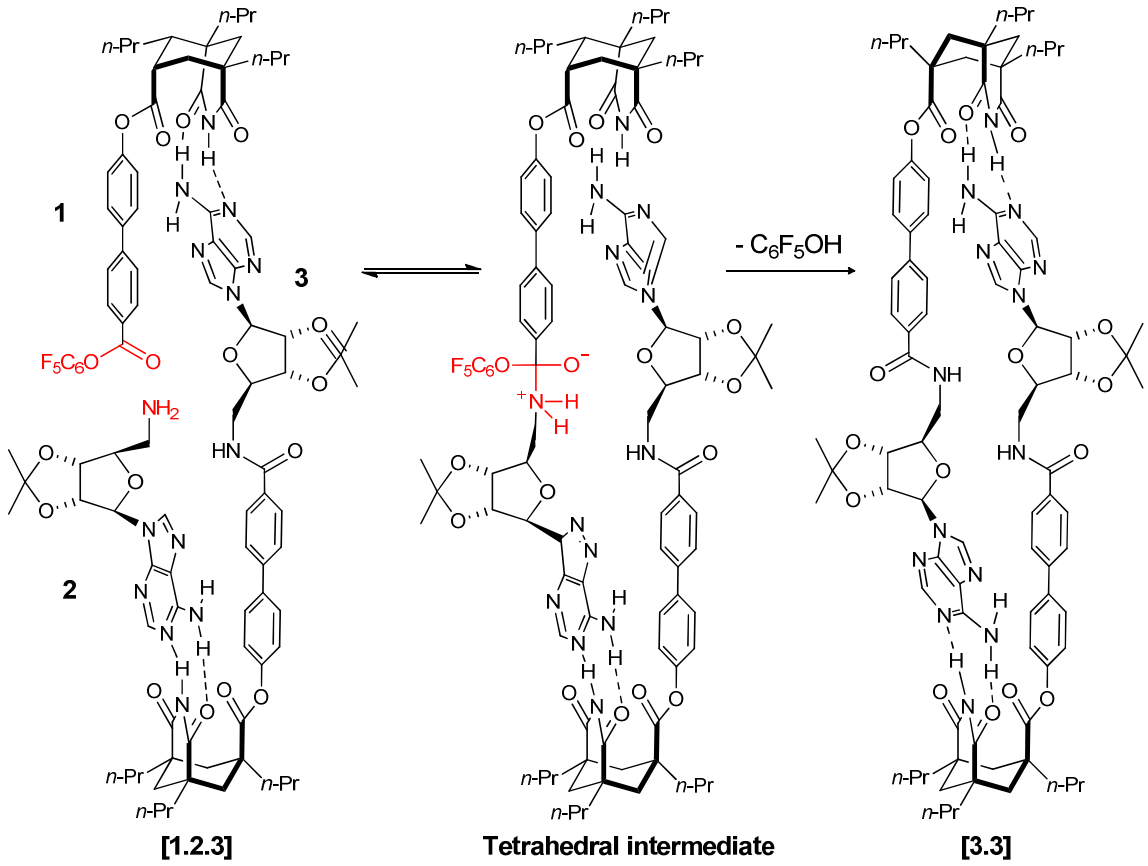
2. Synthetic self-replicating systems

Early studies of replicating systems have identified the necessity for minimising the activity of the reactive **[A.B]** complex (Figure 2 above).[5] The use of synthetic rather than biologically based systems provides the opportunity to eliminate undesirable aspects of a self-replicating cycle which can reduce the efficiency of the system - the ease of altering the design of synthetic systems being a particular advantage.[5,10] Therefore, viewing nature's designs as an example of what is possible, rather than a blueprint, offers a minimalist approach in designing simple synthetic molecules capable of self-replication.[10]

2.1. Early designs of synthetic self-replicators

The first attempt to design a synthetic self-replicating molecule was by reported by Rebek in 1990.[11,12] The system was comprised of complementary adenine and imide recognition sites in building blocks **1** and **2**. Aminolysis served as the bond forming reaction to generate self-complementary template **3**, capable of assembling its own building blocks. The initial designs incorporated a phenyl or naphthyl spacer in the ester backbone, which allowed the [**A.B**] complex reaction pathway to template formation to dominate and hence quench the

autocatalytic pathway, although self-replication was observed using the naphthyl spacer.[11] Modification of the backbone by replacing these spacer groups with a longer biphenyl spacer facilitated a self-replication pathway (Scheme 2).[12] This adaptation separated the reactive centres within the [**A.B**] complex and rendered it less effective in the intramolecular sense. This modification produced a rate curve with clear sigmoidal character.



Scheme 2 Rebek's first attempt at a self-replicating system, where the tetrahedral intermediate transition state is stabilised by H-bonding recognition and aromatic stacking

The legitimacy of whether Rebek's system was operating through an autocatalytic mechanism was heavily discussed in the early 1990's, and it was questioned whether amide catalysis could be causing the rate enhancement displayed.[13] Through heavy exploration into the system using various models and modifications, it was discovered that several modes of catalysis were operating within the Rebek system.[13-15] It was established that self-replication, as defined by Rebek, does operate in this system, however other pathways obscure the simple picture of a ternary complex as the *only* complex that leads to the rate enhancement, and one of those bimolecular pathways is amide catalysis. This argument led to demonstrate the

complexities that operate in a minimal self-replicating system, and that designing a system capable of achieving exponential turnover would be a challenging prospect.

2.2. Developments in synthetic self-replication

Rebek and co-workers continued their studies by developing several other minimal replicators, and were able to demonstrate the necessity of self-complementarity and molecular recognition in achieving an efficient self-replicating system.[16] Wang and Sutherland applied these synthetic design principles constructed by Rebek to develop a template directed system with a [4+2] Diels-Alder cycloaddition as the template forming step.[17] The system demonstrates the first example of near-exponential growth in a completely synthetic system ($\alpha = 0.8$), which was later optimised by von Kiedrowski to achieve a rate of autocatalytic template formation of 0.89.[18]

Philp further explored cycloaddition mediated self-replication by developing a system based on a 1,3-dipolar cycloaddition between nitrone **A1** and maleimide **B1** to form a diastereomeric mixture of isoxazolidines **T1** and **T1'**.[19] The corresponding reaction between **A1** and **B2**, the ester of **B1**, to yield **T2** and **T2'** was also conducted as control for monitoring any catalytic effect (Figure 3). During the reaction, nitrone **A1** and maleimide **B1** are capable of associating via complementary H-bonding interactions with templates **T1** and **T1'** to form the potentially catalytic complexes [**A1.B1.T1**] and [**A1.B1.T1'**]. The introduction of the recognition motif during the reaction significantly increased the rate of reaction, and a clear sigmoidal-shaped curve was particularly evident for major *trans* isomer **T1**. Addition of **T1** at $t = 0$ led to the enhanced formation of **T1** and not **T1'**, resulting in an improved ratio of 9:1 compared to the reaction of the corresponding esters at a 4:1 ratio. This therefore demonstrates that the reaction of **A1** with **B1** generates template **T1** which is capable of self-replication. The autocatalytic reaction order was determined to be 0.9 due to the destabilisation of the [**T1.T1**] template duplex, leading to efficient turnover in the autocatalytic cycle.

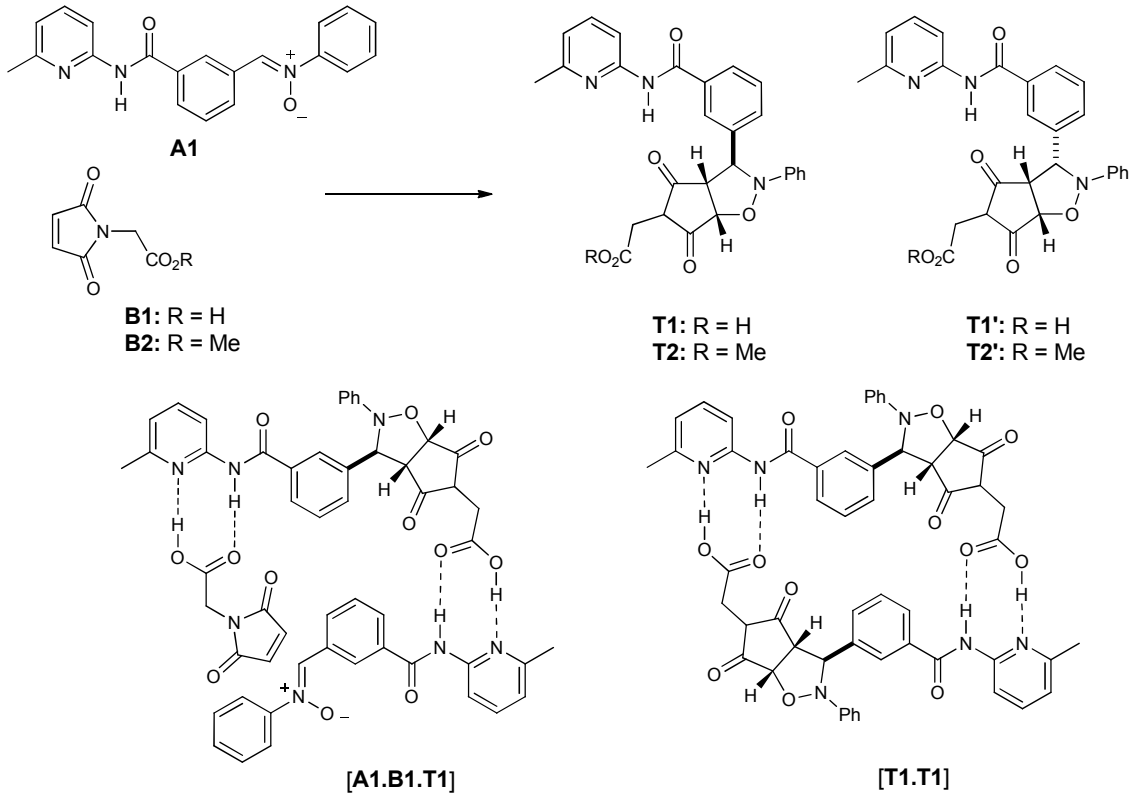
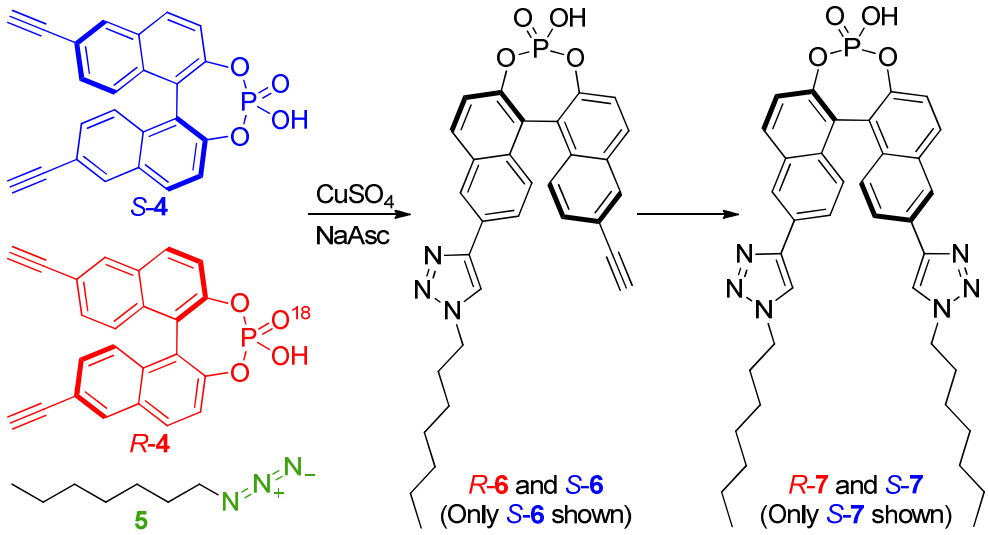


Figure 3 1,3-dipolar cycloaddition between nitrone **A1** and maleimide **B1** achieves efficient self-replication in a synthetic system

By continuing studies on cycloaddition mediated self-replicating systems, Philp and co-workers were able to demonstrate how minor alterations of building blocks can significantly impact the efficiency of the replicating system.[20,21]This thereby makes the design of a self-replicating system challenging, as it is unclear whether a system will operate effectively until it has been constructed. The group also led to demonstrate replication within a system of templates which cooperate through a series of autocatalytic and cross-catalytic pathways.[22]

Rodionov described a system in which the self-replication of micellar aggregates result in the spontaneous amplification of chirality in the reaction product. Compounds **4** and **5** were selected based on their hydrophilic and hydrophobic properties, respectively, meaning that by setting up a biphasic reaction medium, the ensuing reaction would require a phase transfer mediated pathway to encourage formation of amphiphilic mono- and ditriazoles **6** and **7** (Scheme 3). An ^{18}O label was incorporated into the *R*-enantiomer of **4** in order to monitor the reaction of both enantiomers independently. In a 50% *S*-enriched *RS*-**7** mixture, the formation of micelles proceeded via an uncatalysed pathway, with a slow phase transfer of azide **5** to

the aqueous reaction media followed by fast and unselective reaction with either *R*-**7** or *S*-**7**, yielding a disordered aggregate of *R*- and *S*- amphiphiles. However, when pre-seeded with racemic **7**, the reaction was accelerated and the *S*-**7** enantiomer amplified. Upon seeding with enantiomerically pure *S*-**7**, an even higher rate of formation was observed. It was postulated that catalysis resulted from the ditriazole **7** forming predominantly homochiral aggregates of the *S*-enantiomer, which is capable of catalysing the phase transfer of azide **5** into the aqueous layer. Therefore, within this system, the intermediate amphiphiles are capable of phase transfer catalysis, information transfer and self-assembly, which Rodionov claims to presents "*a plausible model for prenucleic acid 'lipid world' entities*".[23] These results encourage the development of more complex recognition mediated replication networks, which could have the potential to assist in bridging the gap between synthetic and prebiotic chemistry.



Scheme 3 Racemic mixture of dialkyne **4** couples with azide **5** to form mono- and ditriazine amphiphiles **6** and **7** respectively, of which *S*-**7** is capable of selectively amplifying its own formation through a self-replicating pathway

3. Natural self-replicating systems

RNA has the capability of acting as a catalyst and as a carrier of genetic information. In addition, RNA is a likely candidate for the first prebiotic self-replicating molecule.[24] This has generated a spark of interest to devise nucleotides that replicate without enzymes. Nucleic acid replicators based on polycondensations of activated mononucleotides directed by non-

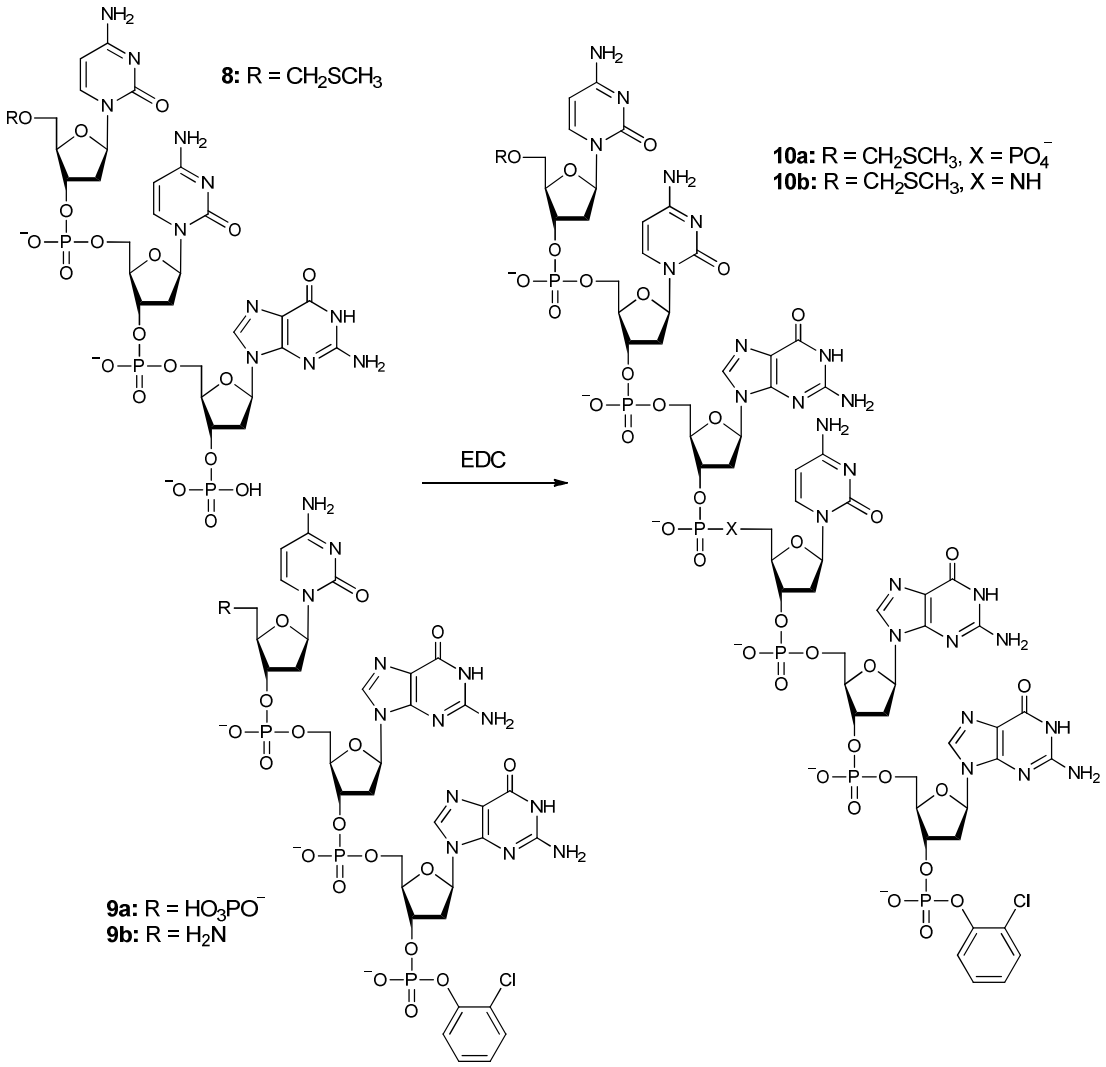
enzymatic templates have been studied by Orgel extensively.[6] Orgel's group were able to demonstrate the non-enzymatic synthesis of fully complementary products, however, no complete reaction cycle could be achieved using mononucleotides as precursors.[1] These initial findings generated a surge of interest into the possibility of accessing a non-enzymatic 'natural' self-replicating system.

3.1. Nucleic acid based systems

The first biomimetic replicating systems based on minimal hexadeoxynucleotide templates were achieved by von Kiedrowski in 1986,[25] and Zielinski and Orgel in 1987.[26] Both systems demonstrated the ability of template directed synthesis to increase the initial rate of feedstock reactions, however experiments did not produce a sigmoidal time curve. Template addition did not increase the rate of autocatalytic template formation linearly, and it was instead proportional to √[**T**], leading to a reaction order, p, of 0.5 rather than 1. This firmly fits the square-root law of autocatalysis, reflecting the influence of both limited autocatalytic ability and product inhibition.[1,9] This was due to the non-autocatalytic predominance of the systems and presence of unwanted side reactions, meaning that the ideal sigmoidal rate cure was not observed.

Following these first systems developed by von Kiedrowski and Orgel, it was evident that product inhibition was a significant factor in preventing autocatalytic self-replication. Therefore, von Kiedrowski attempted to develop a system which would increase template induced autocatalytic product synthesis whilst keeping non-instructed synthesis as low as possible. For nucleic acid based systems, autocatalytic synthesis benefits from enhanced nucleophilicity at the attacking 5'-terminus, therefore, trimer **9a** was used in its 5'-phosphorylated form instead of 5'-hydroxy form, used in von Kiedrowski's initial system. Carbodiimide dependant condensation with 5'-methylthiomethyl (MTM) trideoxynucleotide **8** yielded hexamer template **10a** with a central 3'-5'-pyrophosphate linkage (Scheme 4), and resulted in a rate of template formation two orders of magnitude greater than von Kiedrowski's original system. The system was further developed through the replacement of the 5'-phosphate with a 5'-amino group **9b**, leading to the formation of a 3'-5'-phosphoramidate bond **10b** and resulting in a rate enhancement of almost four orders of magnitude compared to the phosphodiester.[27] This was the first observation of a sigmoidal increase in template concentration and furthermore demonstrated that 'faster' replicators are

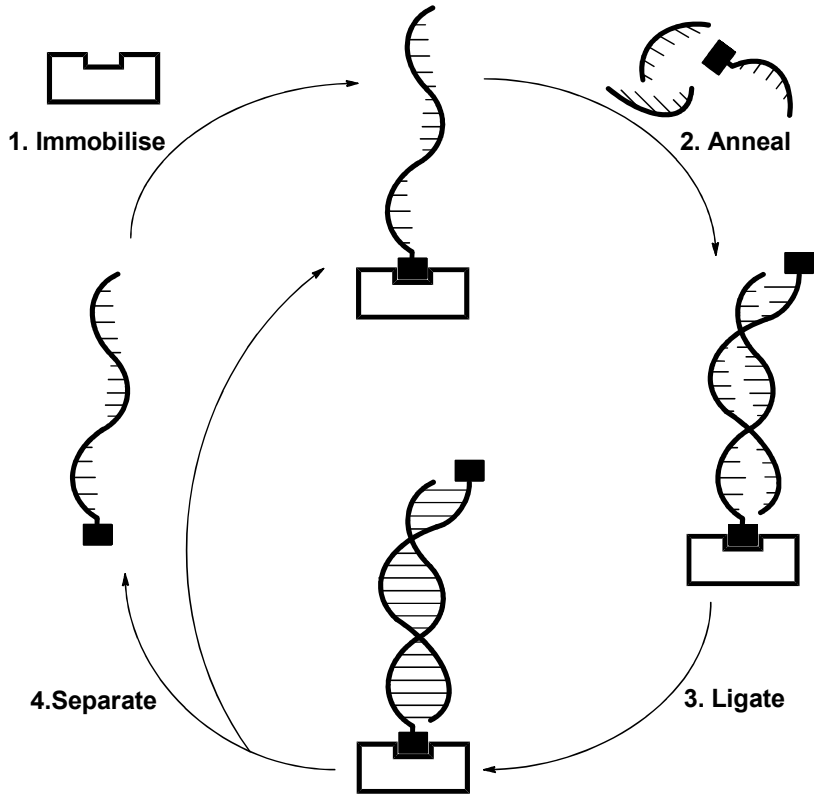
more selective. Despite these advances, the system still follows the square-root law, displaying parabolic rather than exponential growth, although displays a large rate enhancement relative to previously reported systems.



Scheme 4 First demonstration of a sigmoidal rate of product formation in a nucleic acid based replicating system displaying parabolic growth of product

The first successful demonstration of exponential growth in a self-replicating system was achieved by von Kiedrowski in 1998 through the use of a solid support in the replication mechanism, a technique called SPREAD (Surface Promoted Replication and Exponential Amplification of DNA Analogues).[28] By immobilising oligonucleotide template molecules on the surface of the support, stable product duplexes were generated through template directed condensations with complementary oligonucleotide building blocks. The duplexes inhibit the

release of template molecules whilst they are immobilised on the support, however, once cleaved from the support, individual template strands may be separated, thereby generating further material for another cycle (Scheme 5). This approach towards obtaining highly efficient self-replicating systems seems very appealing, as the inhibition of template production may eventually be bypassed by removing the feedstock periodically, flushing off product from the solid support and recycling template for the next addition of feedstock.[5] Despite this obvious advantage, it may be argued that conducting reactions in such manner does not give a true representation of a self-replicating system in the Darwinian sense, as product inhibition is circumvented rather than prevented. However, it has been postulated that early replicators may have utilised minerals and rocks by spreading over their surface, thereby acting as a natural solid support system.[28]



Scheme 5 Highly efficient approach towards product replication achieves near exponential growth through surface templated catalysis

In addition to pursuing exponential growth in self-replicating systems, von Kiedrowski and co-workers expanded the studies of self-replication to more complex projects, furthering their studies to investigate the conditions for prebiotic self-replication, and the design of a cross-

catalytic replicating system.[1] While both systems were only able to demonstrate parabolic rather than exponential growth, these intricate systems have paved a way for the future designs and ideas, expanding the possible studies that could be conducted with replicating systems.

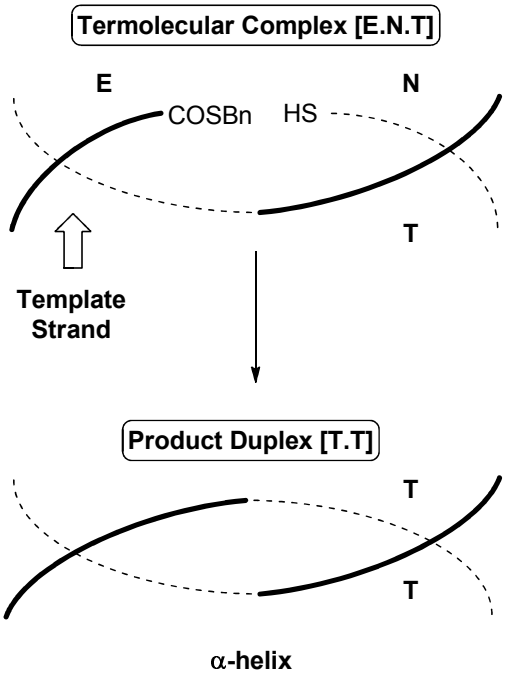
The Joyce group have studied the replication function of RNA extensively and recently constructed an RNA enzyme entirely of $_{L}$-ribonucleotides, which undergoes ligand-dependant, self-sustained replication with exponential growth. It was demonstrated that efficient amplification was dependent on the presence of ligand, and exponential growth rate depends on ligand concentration. Unlike biological enzymes, this system is capable of being generalised to any ligand. Enzymes composed from biologically known materials were shown to degrade, or were incapable of achieving efficient amplification in biological samples, due to rapid degradation by ribonucleases. For example, the enantiomerically pure $_{D}$-RNA enzyme, capable of being accessed enzymatically by transcription of a DNA template, demonstrates efficient amplification outside the realms of biological systems but rapidly degrades when subjected to 10% human serum. Conversely, $_{L}$-RNA has no interaction with biological macromolecules and displays exponential growth in the presence of ribonuclease. Therefore the Joyce group successfully achieved an efficient non-biological self-sustaining RNA enzyme, capable of efficient amplification in the presence of a wide variety of ligands, leading to the determination of ligand concentration in unknown samples.[29] This is an interesting example of a system operating under self-replication that not only requires a template to achieve autocatalytic growth, but the presence of ligand. The $_{L}$-RNA enzyme was further demonstrated to cross-replicate with a partner enzyme, resulting in their mutual exponential growth and *"enabling self-sustained Darwinian evolution*".[30,31]

3.2. Peptide based systems

Experiments conducted under prebiotic conditions have often been found to yield short peptides as products. If such peptides were to self-assemble and have a route towards information-transfer through self-replication, the consideration of their potential role in the molecular origins of life would be greatly enhanced.[32]

In 1996, Ghadiri and co-workers set out to achieve the first self-replicating peptide using a 32-residue α-helical peptide based on leucine-zipper sequence to induce recognition.[32] The

leucine-zipper motif readily forms an α-helix structure through a single stranded peptide forming a helix with a second strand, creating a coiled-coil. The 32-residue α-helical peptide template **T** acts as a recognition unit for 15- and 17- residue peptide fragments, placing the reactive sites in close proximity. Ligation between a thiol nucleophile **N** and thiobenzyl ester electrophile **E** generates a copy of the original template strand **T** (Scheme 6).[33]



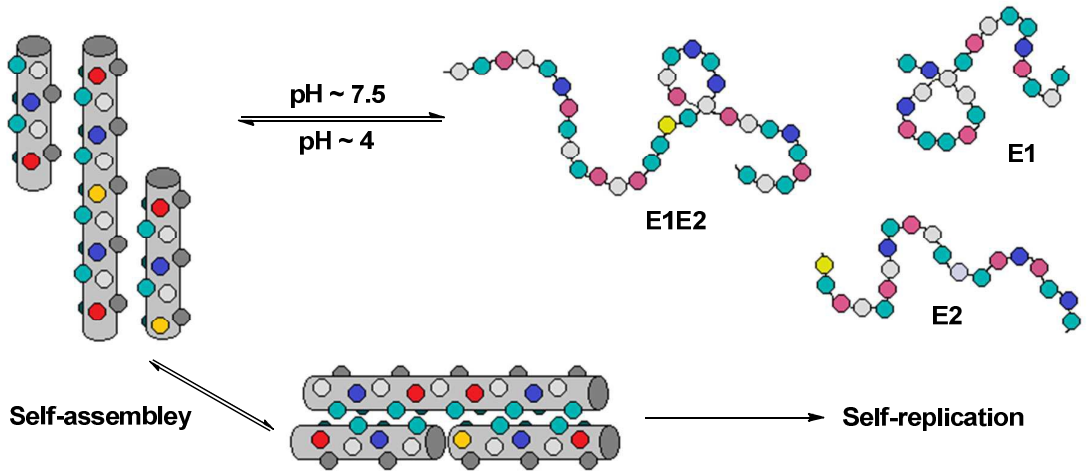
Scheme 6 Peptide α –helix formation is utilised as a self-replicating template through peptide template **T** with thiol nucleophile **N** and thiobenzyl ester electrophile **E**

Template formation can be accessed via several possible pathways, 6 of which lie in the autocatalytic channel, however only two of these are self-replicating with the ability to form a ternary complex.[5,34] By carefully considering the ability of all potential pathways to template formation, all bimolecular reactions were discredited by designing 'crippled' templates, capable of recognising one feedstock strand, but not the other. Conducting reactions in the presence of crippled template observed no increase in initial rates of product formation. Therefore, it was concluded that autocatalysis does not proceed through these pathways and template turnover goes through termolecular ternary complexes. Autocatalysis in template production is observed in Ghadiri's reactions, and sigmoidal growth of template is noticeable without the addition of prefabricated template, however, the increase in product formation correlates with the square-root law of initial template concentration, which reflects product inhibition. It is unlikely that inhibition is due to a highly stable template duplex, as, although

reactions were not taken to complete conversion, there was no premature reduction in the rate of template production. This suggests that the leucine-zipper recognition motif does allow dissociation of the template duplex, as template product is added and returned to the reaction mixture.

Ghadiri extended his work in peptide self-replication by investigating the effect of adding mutant fragments to the initial peptide mixture of **E**, **N** and **T**, in order to simulate the spontaneous formation of errors during the self-replication process.[35] It was demonstrated that Ghadiri's mutant templates could catalyse the formation of native template **T** from the **E** and **N** building blocks, whilst mutant templates could not catalyse their own formation from mixtures of their corresponding feedstock. The catalytic rate efficiency here for native template formation is roughly 75% of that produced when using the native template itself, which seems to indicate a successful error correcting network of catalytic cycles.

Peptide self-replication was further studied by exploring the effect of environmental control over the reaction course. An initial system was designed by Chmielewski,[36] containing electrophilic thioester **E1** and nucleophilic peptide **E2** fragments, which generated **E1E2** template containing negatively charged Glu residues on the basis of the thioester-promoted peptide bond formation. Under neutral conditions (~ pH 7.5), **E1E2** formation proceeded by a non-autocatalytic pathway due to the structure existing as a non-templating random coil, as a result of the repulsion of negatively charged fragments. When the pH was reduced to achieve acidic conditions, a large rate increase was observed upon reaching pH 4, attributed to the coiled-coil templating ability of **E1E2** at this pH. Acidic conditions enable protonation of the negatively charged Glu residues, preventing electrostatic repulsion and enabling favourable coiled-coil formation. This allowed the coupling of **E1** and **E2** to proceed by an autocatalytic pathway, with sigmoidal growth in **E1E2** being observed at pH 4, and the initial rate of product formation proportional to the square root of the template concentration (Scheme 7).

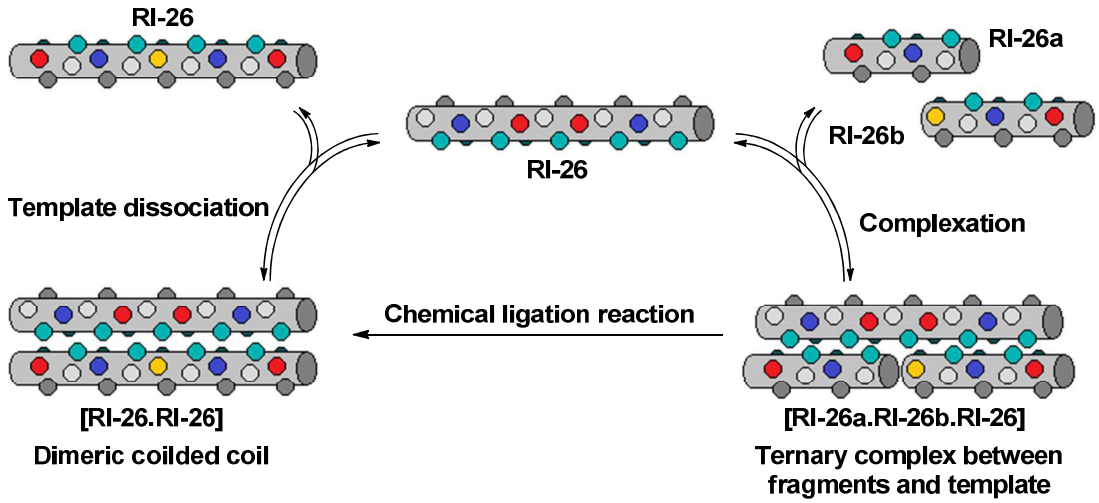


Scheme 7 A self-replicating peptide which utilises environmental control over the reaction course, forming a coiled-coil conformation only under acidic conditions.

Chmielewski further developed the **K1K2** peptide, containing positively charged Lys residues, which was designed to operate under ionic control and demonstrated effective autocatalysis through the addition of negatively charged counterions.[37] The group continued its studies by using both of the previous two peptides, **E1E2** and **K1K2**, in a cross catalytic replication cycle.[38] It was demonstrated that a cross-catalytic system was able to exhibit product selectivity derived from the relative stabilities of the coiled-coils formed in the reaction mixture, with additional amplification achieved by adjusting the pH. However, any increase in the initial rate stemmed from relatively high concentration of additional templates, due to product inhibition. This is not surprising based on the charged peptide fragments forming a highly stable coiled-coil duplex.

The Chmielewski group's attention subsequently focused on attempting to improve the catalytic efficiency of the self-replicating peptide **E1E2** by destabilising the coiled-coil structure. The first system consisted of peptide fragments **RI-26a** and **RI-26b** forming template **RI-26**, with three full heptad repeats within the coiled-coil, one shorter than the original **E1E2** peptide.[39] The shorter chain length reduced the hydrophobic interactions between chains, therefore slowing any uninstructed non-catalysed or background reactions as a result of association between the two fragments (Scheme 8). A second system was also based on the **E1E2** system, where either hydrophobic or hydrophilic regions were modified by replacing them with a proline kink to give **XL-1** and **XL-2** templates, respectively.[40]

The **XL-1** path showed little product from the feedstock after 24 hours and no template acceleration with added template. This was explained by the proline kink creating a break in the hydrophobic surface of the coiled-coil, with the two fragments being directed away from each other. Conversely, proline replacement in the hydrophilic face of **XL-2** template creates a bent, but continuous, hydrophobic surface which promotes ligation between bound fragments. Once ligation is complete, the kink acts to destabilise the template duplex, reducing template inhibition significantly. For both **RI-26** and **XL-2** templates, a significantly higher reaction order ($p \sim 0.91$) was observed, classifying the replicating systems as weakly exponential ($0.75 < p > 1$), a catalytic rate enhancement close to known enzymes.



Scheme 8 Formation of template **RI-26** approaching exponential rates due to minimised non-catalysed and background reactions

More recently the Ashkenasy group have explored replication of short chain peptides which are capable of forming β-sheets in water through alternating hydrophobic and hydrophilic interactions.[41] Peptide electrophile **αE** and nucleophile **αN** fragments readily associate to the template β-sheet strand $\boldsymbol{\beta}_n$, catalysing the ligation of the monomers (Figure 4). Studies into the mode of catalysis revealed that the peptide molecules formed spherical micelles after a short time, which rearranged into fibril helical structures followed by hollow nanotubes.[42] The fibrils were demonstrated to serve as catalysts for replication, leading to exponential product growth, while other supramolecular aggregates do not enhance ligation.[43]

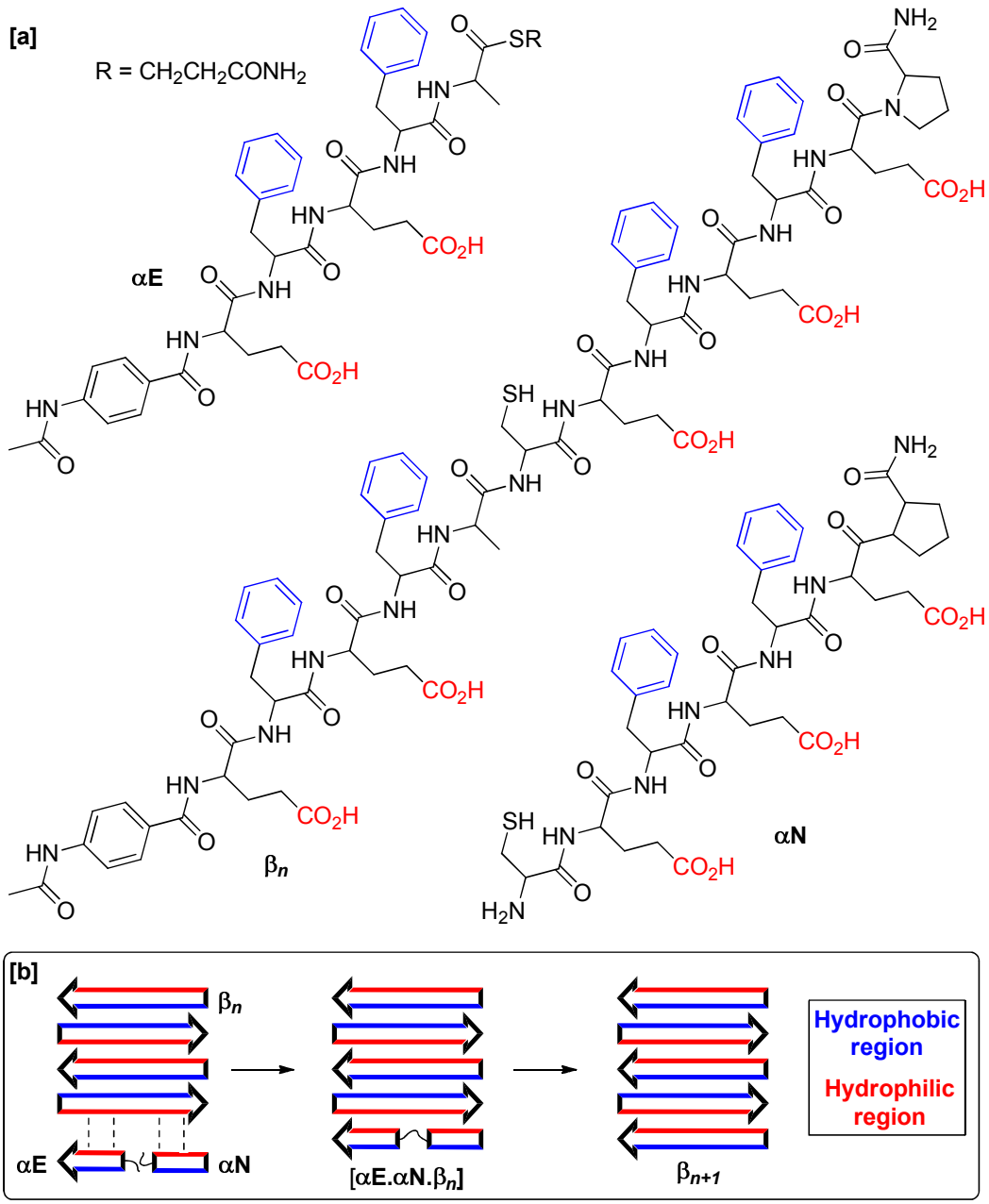


Figure 4 [a] Peptide fragments, **E** and **N**, and single strand template $\mathbf{1}_n$ ($_{n = 1}$); [b] Beta sheet $\mathbf{1}_n$ serves as template for the association of fragments **E** and **N**, leading to enhanced ligation resulting in exponential growth of larger template aggregate $\mathbf{1}_{n+1}$.

4. Systems chemistry approach to self-replication

Systems chemistry is a young field that brings together aspects of supramolecular and dynamic combinatorial chemistry (DCC), by attempting to capture the complexities of

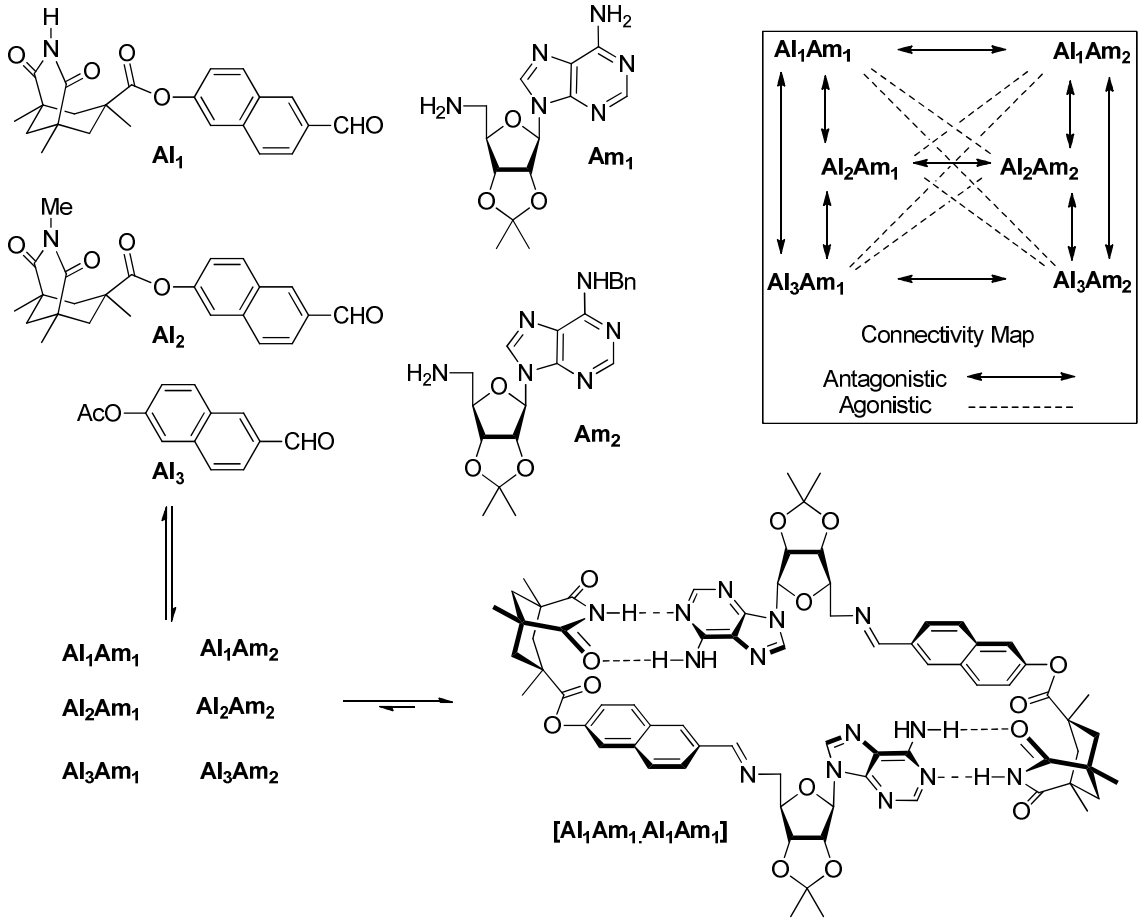
biological systems within a wholly synthetic chemical framework.[44] The question of how complex natural structures may have emerged in prebiotic times is starting to be addressed in recent work on replication in the context of dynamic combinatorial chemistry.[45] The use of dynamic control within a system enables the expression of a dominant chemical species from a library of synthetic compounds through molecular recognition between library members, which are designed to interact and react with each other in programmed ways.[44] The overexpression of a compound is achieved by making copies of itself from a pool of reshuffling components in a series of competing equilibria.[4,44] Thus, the selected structures can be considered to be self-replicating in a thermodynamically controlled fashion.

The early studies of replicating systems have mainly focused on generating the thermodynamic reaction products in order to achieve an efficient self-replicating molecule. In these cases, the product of the reaction is more stable than its corresponding building blocks and furthermore, replication is usually irreversible. However, a replicating reaction with a reversible step would bring us closer to achieving dynamic kinetic stability, a significant factor in the operation of living organisms.[46] The investigation into synthetic systems aiming to model these properties aids to bridge the gap towards discovering the possible origin of life, and bring minimal self-replicating systems closer to realisation using DCC.[47]

4.1. Development of dynamic self-replicating systems

A significant step towards a dynamic replicating system was demonstrated by the group of Giuseppone, who reported a proof of principle study featuring a library member exhibiting self-complementary recognition through hydrogen bonding,[47] inspired by the Rebek replicator.[11,12] The group developed an 11-membered library comprising of several building blocks capable of reversible covalent association and displaying, or not displaying, complementary supramolecular units in order to produce, or not produce, a template with self-recognition properties. The key molecule was $\mathbf{Al_1Am_1}$, synthesised by condensation of aldehyde $\mathbf{Al_1}$ and adenosine amine $\mathbf{Am_1}$. This self-complementary dynamic compound is able to strongly associate with itself into complex [$\mathbf{Al_1Am_1.Al_1Am_1}$]. The system further comprised of two other aldehydes, $\mathbf{Al_2}$, the methyl protected variant of $\mathbf{Al_1}$, and $\mathbf{Al_3}$, which contains no recognition features, and an additional amine, $\mathbf{Am_2}$, with a sterically restricted imide recognition site (Scheme 9).

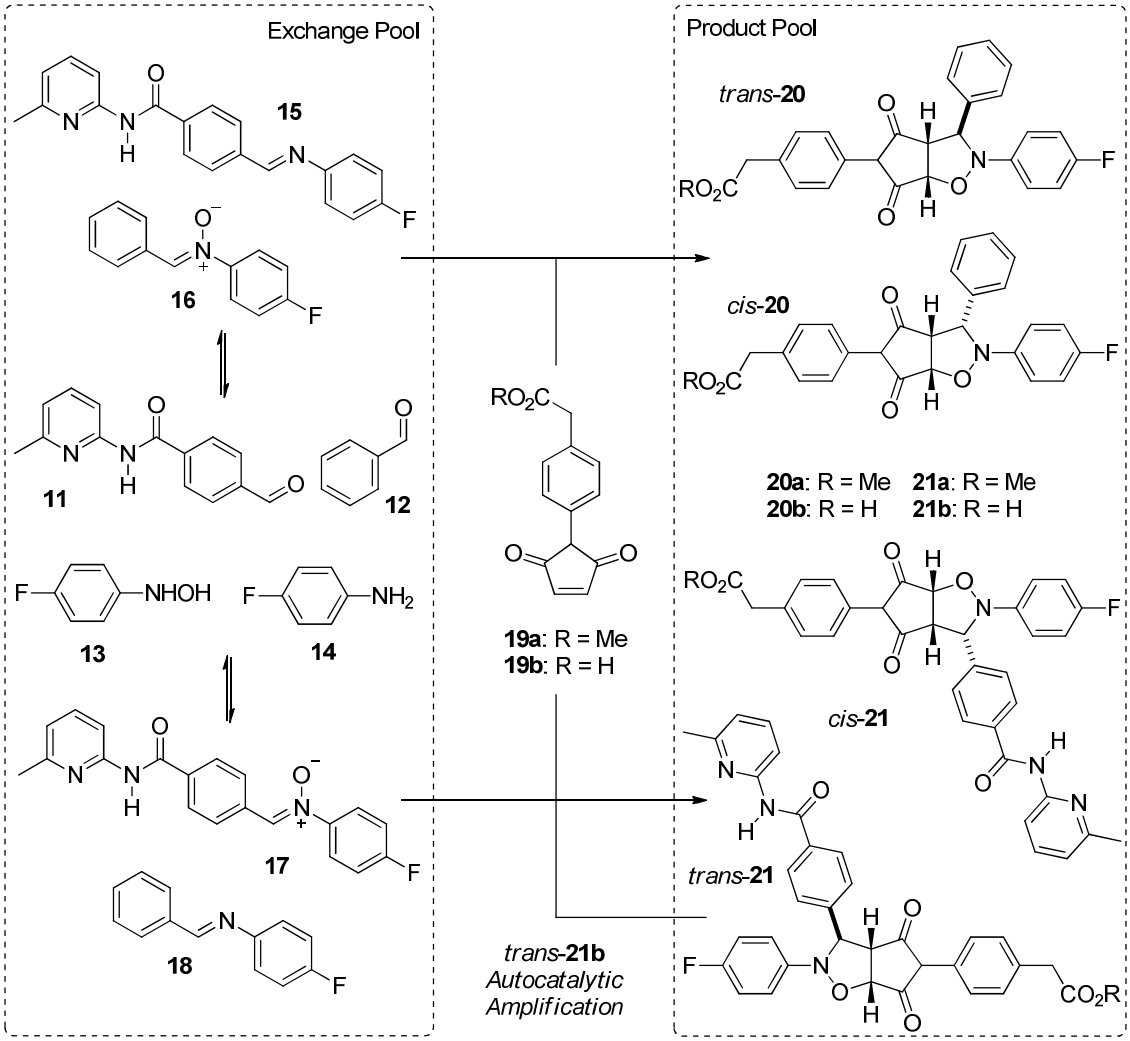
A series of equilibrium experiments demonstrated that **Al_1Am_1** was overexpressed above the other potential imine products when the building block components were added in equimolar proportions. The group therefore demonstrated that it was possible to self-amplify one product in a dynamic combinatorial library (DCL), namely the one that can self-complementarily direct its own formation. The expression of compounds in the library evolves along both kinetic and thermodynamic bases, that both lead to the amplification of the best duplicator. From a Darwinian point of view, such a system illustrates the selection of the most efficient self-duplicator by the destruction of the entities that are not able, or are less able in the case of **Al_1Am_2**, to duplicate themselves.



Scheme 9 DCL of imines **$Al_{(1\text{-}3)}Am_{1,2}$**, cooperate to favour the formation of dimer [**$Al_1Am_1.Al_1Am_1$**]

Sadownik and Philp later reported a kinetically controlled self-replicating system within a DCL, which overexpresses the autocatalytic reaction product over the remaining product mixture.[44] The DCL was constructed from aldehyde **11**, bearing an amidopyridine recognition site, and benzaldehyde **12**. The presence of 4-fluoroaniline **14** generates two

unreactive imines (**15** and **18**) and 4-fluorophenylhydroxyamine **13**, in turn, permits the formation of reactive nitrones **16** and **17** (Scheme 10). At equilibrium, the DCL exchange pool contains two imines and two nitrones along with their respective precursors. Materials can be transferred irreversibly from the exchange pool to the product pool through the reaction of either nitrone **16** or **17** with maleimide **19a** or the recognition based variant, **19b**.



Scheme 10 Dynamic exchange pool of building blocks drives the formations of corresponding diastereomeric pairs of cycloadducts through reactions with maleimides **19a** and **19b**

The dipolar cycloaddition reactions create a group of products containing two pairs of diastereomeric cycloadducts, *cis*- and *trans*-**20** and *cis*- and *trans*-**21**. Only cycloadduct *trans*-**21b** is capable of self-replication. *trans*-**21b** acts as a template for its own formation through

the recognition and binding of nitrone **17** and maleimide **19b** to generate catalytic ternary complex [**17.19b.***trans*-**21b**]. This accelerates the cycloaddition reaction between **17** and **19b** by more than 100 times to give product duplex [*trans*-**21b.***trans*-**21b**] which dissociates to add more template to the reaction (Figure 5). Autocatalysis by *trans*-**21b** causes the cycloadduct of **17** to dominate over that of **16**, resulting in the selective consumption of building blocks **11** and **13**.

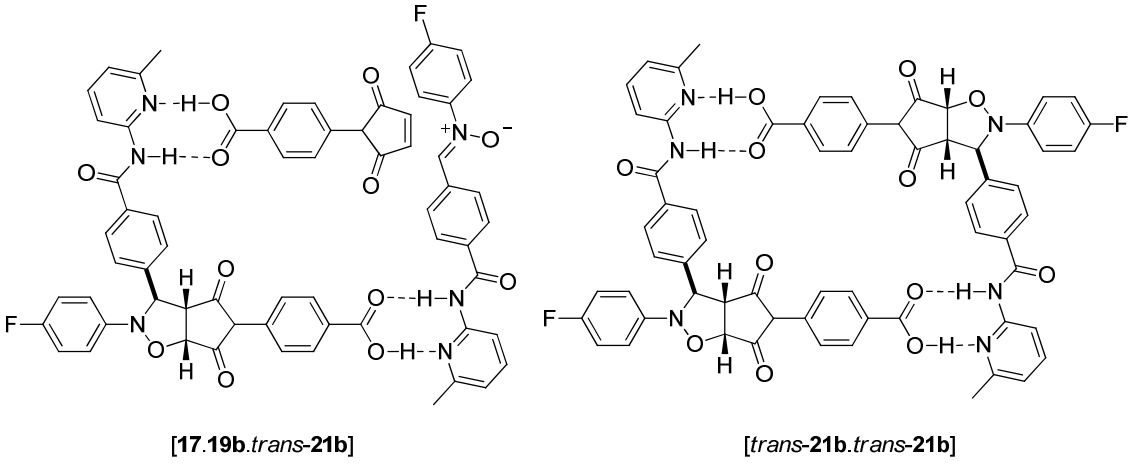


Figure 5 Ternary complex [**17.19b.***trans*-**21b**] and template duplex [*trans*-**21b.***trans*-**21b**]

A significant result was achieved by Otto through template initiated replication. As natural systems do not emerge spontaneously, as has previously been investigated within the realms of synthetic replication, a necessary step in the evolution of synthetic systems is to model the triggered response functions observed in many biological processes.[48,49] A system was developed whereby oxidation of a simple aromatic dithiol, **22**, strongly favoured the formation of a series of isomeric catananes, **23**, held together by reversible disulfide linkages. Introduction of a suitable guest template, **24**, amplified four distinct isomeric tetramers, **I-IV**, within the library (Figure 6).[50,51] In depth studies of the system showed that at low concentrations of template, all tetramers displayed near-linear dependence of the concentrations of template **24**. However, increasing the template concentration led to a sudden amplification of the minor and weakest binding tetramer, isomer **IV**, which displayed unexpected self-replication. Without the presence, or at low concentrations of template, the concentration of isomer **IV** is below the critical aggregation concentration (CAC) required for self-replication. By adding a high enough concentration of template, the concentration of the ensuing complex **IV-24** is raised above the CAC, leading to the onset of replication. Once the concentration of isomer **IV** alone is increased beyond its CAC, the tetramer is capable of

promoting further replication in the absence of template.[52] This is the first known example of a self-replicator in a DCL that is dependent on a triggered initiation by a template molecule, developing a further connection between artificial replicating systems and examples seen in nature.

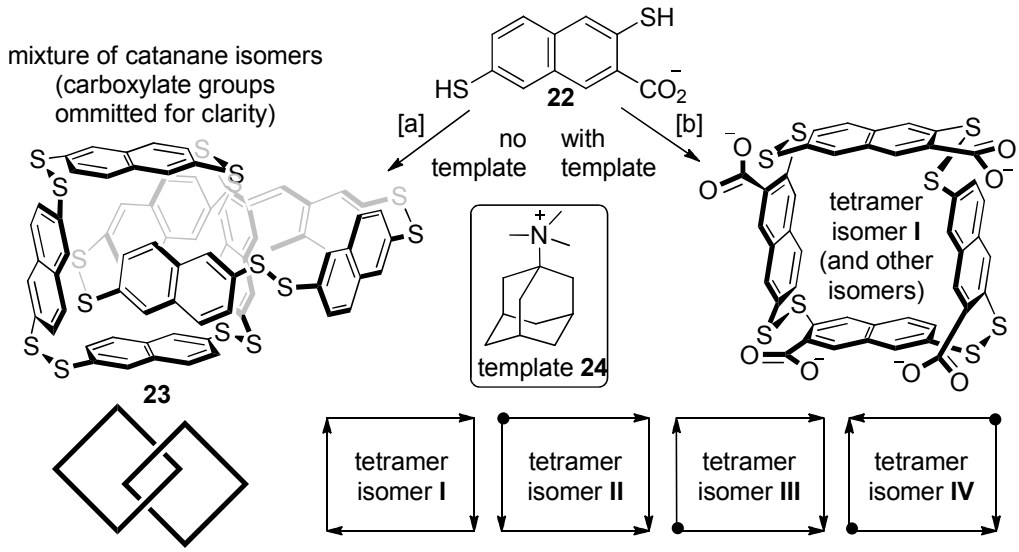


Figure 6 [a] dithiol **22** preferentially forms catanane **23** isomers; [b] addition of template **24** favours the production of tetramers, of which isomer **IV** is capable of self-replication

The Otto group have also achieved interesting results through amphiphilic peptide β-sheet replication in a dynamic library of macrocycles. By stirring a simple dithiol in borate buffer solution, thiol oxidation resulted in the formation of a dynamic library of disulfides of varying sizes. With no agitation, the entropically favourable trimer and tetramer species were the dominant library members, however stirring resulted in the emergence of cyclic heptamer while shaking gave cyclic hexamer.[53] The larger macrocycles formed β-sheets through the greater number of peptide chains promoting self-assembly, leading to the growth of fibrilar aggregates. The breaking of the fibrils generates new fibre ends that promote the formation of more assembling macrocycle, leading to exponential replication. The group further demonstrated that by reducing the hydrophobicity in the peptide sequence, the size of the emerging macrocycle increased, with larger self-assembled macrocycles displaying autocatalytic rates of formation.[54] These replicating macrocycles were initially investigated as separate entities, with only one peptide chain group forming a DCL. However, this study evolved by investigating the effect of combining two different peptide initiators, **X** and **Y**, in the same DCL, with fascinating results. In separate DCLs, **X** formed hexamer $(\mathbf{X})_6$ and **Y** formed the octamer $(\mathbf{Y})_8$, respectively. However when combined in a single DCL, only

hexamers formed, ranging through all possible combinations of peptides in two distinct sets: *Set I*, comprising of $(\mathbf{X})_6$, $(\mathbf{X})_5(\mathbf{Y})_1$, $(\mathbf{X})_4(\mathbf{Y})_2$, $(\mathbf{X})_3(\mathbf{Y})_3$, formed after 3 days and *Set II*, including $(\mathbf{X})_2(\mathbf{Y})_4$, $(\mathbf{X})_1(\mathbf{Y})_5$, $(\mathbf{Y})_6$, was observed after 7 days. It is evident that peptide **X** is the prevalent species within *Set I*, resulting in the dominant formation of hexamer macrocyles. Therefore, the fibrils formed through *Set I* can act as a template and transfer information about the macrocylce size to *Set II* (Figure 7).

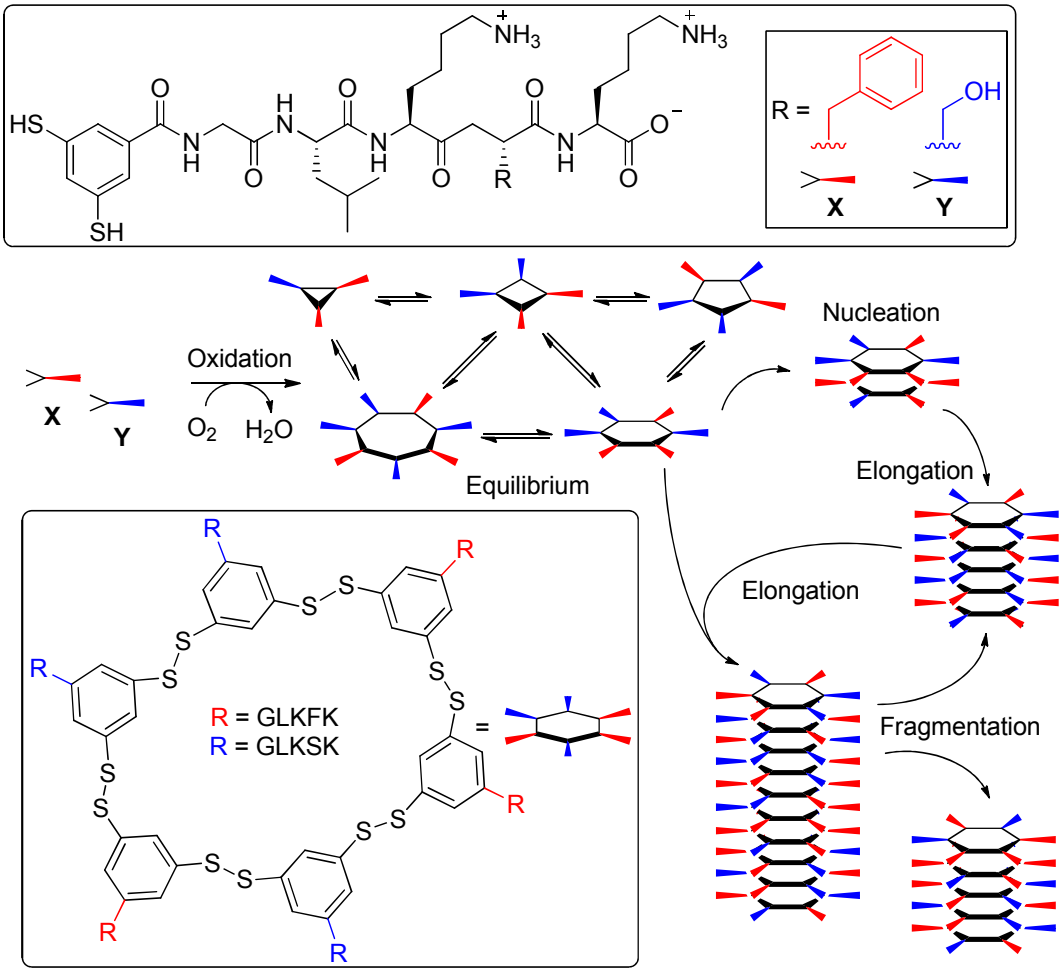


Figure 7 Oxidation of thiol building blocks with two different peptide sequences results in the formation of a DCL of macrocylic disulfides, which selectively stack as hexamers in a growing fibrilar strand through a self-replicating pathway

The observed information transfer within a simple replicating system clearly marks major progress within the realms of self-replication towards greater understanding the origin of biological species. Otto describes the results as *"an important step towards achieving Darwinian evolution with a system of fully synthetic molecules and the synthesis of life"*.[55]

5. Outlook for synthetic self-replicating systems

The existence of life on Earth relies on the process of replication. Over recent years, advances in developing synthetic models have achieved systems with the ability to perform the self-replication process, implicated in the origin of life, at near exponential rates. An extensive catalogue of synthetic replicators has provided information on the replication process, giving an insight into which of the system's features should be present or eliminated in order to demonstrate efficient replication. However, there is still much work to be done on understanding why some systems are more efficient replicators than others. The study of dynamic replication gains further insight into the true competitive nature of the emergence of living systems from the prebiotic soup. In order to expand these ideas and develop a more clear understanding of how these molecules evolved, additional factors, such as information transfer, have been successfully investigated and incorporated into several of these systems. With an ever growing interest in the field, and such a versatile array of approaches being explored, the opportunity for more features present within biological systems has the potential to be developed and integrated into self-replicating systems. A greater understanding of these systems will enable the development of replicating systems with synthetically useful applications, and ultimately gain an insight into the role of replication in the origin of life.

Bibliography

1. V. Patzke and G. von Kiedrowski, *ARKIVOC*, 2007, **5**, 293
2. M.-C. Maurel, *Tetrahedron*, 1999, **55**, 3141
3. W. Gilbert, *Nature*, 1986, **319**, 618
4. J. J. P. Peyralans and S. Otto, *Curr. Opin. Chem. Biol.*, 2009, **13**, 705
5. A. Robertson, A. J. Sinclair and D. Philp, *Chem. Soc. Rev*, 2000, **29**, 141
6. T. Inoue and L. E. Orgel, *Science*, 1983, **219**, 859
7. D. Philp and J. F. Stoddart, *Angew. Chem. Int. Ed. Engl.*, 1996, **35**, 1154
8. A. J. Bissette and S. P. Fletcher, *Angew. Chem. Int. Ed*, 2013, **52**, 12800
9. G. von Kiedrowski, *Bioorg. Chem. Front.*, 1993, **3**, 113
10. M. Famulok, J. S. Nowick and J. Rebek, Jr., *Acta. Chem. Scand.*, 1992, **46**, 315
11. T. Tjivikua, P. Ballester amd J. Rebek, Jr., *J. Am. Chem. Soc.*, 1990, **112**, 1249

12. V. Rotello, J.-I. Hong and J. Rebek, Jr., *J. Am. Chem. Soc.*, 1991, **113**, 9422
13. F. M. Menger, A. V. Eliseev, N. A. Khanjin and M. J. Sherrod, *J. Org. Chem.*, 1995, **60**, 2870
14. E. A. Wintner, B. Tsao and J. Rebek, Jr., *J. Org. Chem.*, 1995, **60**, 7997
15. D. N. Reinhoudt, D. M. Rudkevich and F. de Jong, *J. Am. Chem. Soc.*, 1996, **118**, 6880
16. Q. Feng, T. K. Park and J. Rebek, Jr., *Science*, 1992, **256**, 1179
17. B. Wang and I. O. Sutherland, *Chem. Commun.*, 1997, 1495
18. M. Kindermann, I. Stahl, M. Reimold, W. M. Pankau and G. von Kiedrowski, *Angew. Chem. Int. Ed.*, 2005, **44**, 6750
19. V. C. Allen, D. Philp and N. Spencer, *Org. Lett.*, 2001, **3**, 777
20. R. J. Pearson, E. Kassianidis, A. M. Z. Slawin and D. Philp, *Chem. Eur. J.*, 2006, **12**, 6829
21. E. Kassianidis, R. J. Pearson and D. Philp, *Chem. Eur. J.*, 2006, **12**, 8798
22. V. C. Allen, C. C. Robertson, S. M. Turega and D. Philp, *Org. Lett.*, 2010, **12**, 1920
23. K. V. Bukhryakov, S. Almahdali and V. O. Rodionov, *Langmuir*, 2015, **31**, 2931
24. P. A. Shaw, *Cell*, 1985, **42**, 397
25. G. von Kiedrowski, *Angew. Chem. Int. Ed. Engl.*, 1986, **25**, 932
26. W. S. Zielinski and L. E. Orgel, *Nature*, 1987, **327**, 346
27. G. von Kiedrowski, B. Wlotzka, J. Helbing, M. Matzen and S. Jordan, *Angew. Chem. Int. Ed. Engl.*, 1991, **30**, 423
28. A. Luther, R. Brandsch and G. von Kiedrowski, *Nature*, 1998, **396**, 245
29. C. Olea, Jr., D. P. Horning and G. F. Joyce, *J. Am. Chem. Soc.*, 2012, **134**, 8050
30. A. C. Ferretti and G. F. Joyce, *Biochemistry*, 2013, **52**, 1227
31. M. P. Robertson and G. F. Joyce, *Chemistry & Biology*, 2014, **21**, 238
32. D. H. Lee, J. R. Granja, J. A. Martinez, K. Severin and M. R. Ghadiri, *Nature*, 1996, **382**, 525
33. P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. H. Kent, *Science*, 1994, **266**, 776
34. K. Severin, D. H. Lee, J. A. Martinez and M. R. Ghadiri, *Chem. Eur. J.*, 1997, **3**, 1017
35. K. Severin, D. H. Lee, J. A. Martinez, M. Vieth and M. R. Ghadiri, *Angew. Chem. Int. Ed.*, 1998, **37**, 126
36. S. Yao, I. Ghosh, R. Zutshi and J. Chmielewski, *J. Am. Chem. Soc.*, 1997, **119**, 10559
37. S. Yao, I. Ghosh, R. Zutshi and J. Chmielewski, *Angew. Chem. Int. Ed.*, 1998, **37**, 478
38. S. Yao, I. Ghosh, R. Zutshi and J. Chmielewski, *Nature*, 1998 **396**, 447

39. R. Issacs and J. Chmielewski, *J. Am. Chem. Soc.*, 2002, **124**, 6808
40. X. Li and J. Chmielewski, *J. Am. Chem. Soc.*, 2003, **125**, 11820
41. B. Rubinov, N Wagner, H. Rapaport and G. Ashkenasy, *Angew. Chem. Int. Ed.*, 2009, **48**, 6683
42. B. Rubinov, N. Wagner, M. Matmor, O. Regev, N. Ashkenasy, G. Ashkenasy, *ACS Nano*, 2012, **6**, 7893
43. V. Bourbo, M. Matmor, E. Shtelman, B. Rubinov, N. Ashkenasy and G. Ashkenasy, *Orig. Life Evol. Biosph.*, 2011, **41**, 563
44. J. W. Sadownik and D. Philp, *Angew. Chem. Int. Ed.*, 2008, **47**, 9965
45. P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J.-L. Wietor, J. K. M. Sanders and S. Otto, *Chem. Rev.*, 2006, **106**, 3652
46. J. Li, P. Nowak and S. Otto, *J. Am. Chem. Soc.*, 2013, **135**, 9222
47. S. Xu and N. Giuseppone, *J. Am. Chem. Soc.*, 2008, **130**, 1826
48. Y. Zheng, M. L. Wong, B. Alberts, T. Mitchison, *Nature*, 1995, **378**, 578
49. J. M. Kollman, J. K. Polka, A. Zelter, T. N. Davis, D. A. Agard, *Nature*, 2010, **466**, 879
50. K. R. West, R. F. Ludlow, P.T. Corbett, P. Besenius, F. M. Mansfeld, P. A. G. Cormack, D. C. Sherrington, J. M. Goodman, M. C. A. Stuart and S. Otto, *J. Am. Chem. Soc.*, 2008, **130**, 10834
51. J. Li, P. Nowak, and S. Otto, *Angew. Chem. Int. Ed.*, 2015, **54**, 833
52. P. Nowak, M. Colomb-Delsuc, S. Otto, and J. Li, *J. Am. Chem. Soc.*, 2015, **137**, 10965
53. J. M. A. Carnall, C. A. Waudby, A. M. Belenguer, M C. A. Stuart, J. J.-P. Peyralans, S. Otto, *Science,* 2010*,* **327**, 1502
54. M. Malakoutikhah, J. J.-P. Peyralans, M. Colomb-Delsuc, H. Fanlo-Virgós, M. C. A. Stuart and S. Otto, *J. Am. Chem. Soc.*, 2013, **135**, 18406
55. J. W. Sadownik, E. Mattia, P. Nowak and S. Otto, *Nature Chemistry*, 2016, **8**, 264