University of Groningen

Understanding emergent functional properties in self replicating systems

Ottel, Johannes

DOI:
[10.33612/diss.182238948](https://doi.org/10.33612/diss.182238948)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Ottel, J. (2021). *Understanding emergent functional properties in self replicating systems*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.182238948>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Understanding emergent functional properties in self-replicating systems

Jim Ottelé

university of
 groningen

Understanding emergent functional properties in self replicating systems

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Friday 5 November 2021 at 11.00 hours

by

Johannes Ottelé
born on 27 August 1989
in Groningen

Supervisors

Prof. S. Otto

Prof. J.G. Roelfes

Assessment Committee

Prof. A.J. Minnaard

Prof. D. Lynn

Prof. B. Escuder

*Ter nagedachtenis aan Alida (Ali) Geertruida Bakker, die me leerde
wat verlies was
en Johannes (Joop) Kesselaar, die me leerde
hoe je daarmee om kan gaan.*

TABLE OF CONTENTS

Preface: Molecular replicators as a starting point for the de novo

synthesis of life	1
Outline and structure of this dissertation	1
The importance of outreach	2

Chapter 1: Introduction: where are we now and where are we going? 4

1.1 The dynasty of living things	4
1.2 The odds of our unlikely existence	6
1.3 Defining Life	7
1.4 Functional properties of living systems	9
1.5 Self-replication	10
1.6 Dynamic Combinatorial Chemistry	12
1.7 Metabolism	15
1.8 Integration of metabolism	21
1.9 Compartmentalisation	23
1.10 The integration of compartmentalisation	26
Acknowledgements	28
Bibliography	29

Chapter 2: Mechanistic studies of supramolecular self-replication 39

2.1 Introduction	39
2.2 Results and discussion	46
2.3 Conclusions and outlook	69
Methods	71
Acknowledgements	78
Bibliography	79
Supplementary Figures	83
Supplementary Tables	91

Chapter 3: Mechanistic insight into supramolecular polymerization-driven self-replication from real-time visualization 92

3.1 Introduction	92
3.2 Results and discussion	94
3.3 Conclusions and outlook	103
Methods	104
Acknowledgements	111
Bibliography	112
Supplementary Information	117

Supplementary Video descriptions	117
Supplementary Figures	119
Supplementary Tables	130

Chapter 4: Chance emergence of catalytic activity and promiscuity in a self-replicator	132
4.1 Introduction	132
4.2 Results and discussion	135
4.3 Conclusions	142
Methods	143
Extended Figures	154
Acknowledgements	159
Bibliography	160
Supplementary Information	163
 Chapter 5: Diversification through multi building block replicators	174
5.1 Introduction	174
5.2 Results and discussion	179
5.4 Conclusions and outlook	194
Methods	194
Acknowledgements	197
Bibliography	198
Supplementary Figures	201
 Chapter 6: Conclusions and future perspectives	228
The future and beginning of synthetic life	232
Bibliography	233
 Summary	234
Samenvatting	237
Acknowledgements	240
About the Author	250
List of publications	251

PREFACE

Molecular replicators as a starting point for the *de novo* synthesis of life

The question of what exactly separates living things from inanimate matter has been on the minds of scientists, philosophers, artists, theologists and many others since humans started to record history (and presumably before that). Truth be told, I have been asking myself this question – albeit phrased much simpler – since I was around ten years old, when the loss of a loved one opened my eyes to (some of) the harsh realities of this world. Let me immediately get this out of the way: you will obviously not find an answer to such questions in this dissertation. Instead, we will look at how molecules can self-assemble into supramolecular structures and in the process become more life-like. It might offer a glimpse in how life could emerge – what types of machineries would be required and how they might come to be.

However, one should understand that the work described in this dissertation was not aimed at discovering how life might have originated on our Earth. There are many fascinating publications that describe how certain biologically relevant molecules could have formed under prebiotic conditions – but here we will rather investigate the requirements for the synthesis of *de novo* life. We are not trying to re-invent the wheel, we are trying to develop a different manner of transportation altogether.

OUTLINE AND STRUCTURE OF THIS DISSERTATION

Chapter one will give a general introduction that aims to inform the reader about the path between lifeless and living systems. We will discuss the many definitions of life, why these are important and why we might do without – at least for a while. Instead, we look at which functional systems are incorporated in living systems and how they might originate (in the lab or otherwise). Furthermore, the self-replicating system that is central to this dissertation is introduced. **Chapter two** focuses on mechanistic studies done to identify and quantify the parameters involved in the mechanism of self-replication, where-

as **chapter three** describes the discovery of a whole new mechanism altogether. **Chapter four** describes the integration of catalytic activity in self-replicators resulting in proto-metabolic activity of the self-replicating system. **Chapter five** introduces a new approach to screen mixtures of building blocks for self-replicating behaviour as a tool to probe catalytic activity and self-sorting behaviour. Finally, **chapter six** aims to give a glimpse of the future: the prospect of combining proto-metabolic activity with compartmentalisation.

Chapters two and five contain preliminary results: these are based on experiments that are not repeated as often but still contain information that gives additional insight.

THE IMPORTANCE OF OUTREACH

Before going any further, I would like to ask your attention for an issue that has started to concern me during the course of my research. I believe that as scientists, we have the responsibility to inform the public of our findings. Currently, many (most¹) publications are behind paywalls – inaccessible to anyone without a subscription. Even though more journals are opening up possibilities of open-access articles, scientists are often (indirectly) discouraged to use them as they charge the authors exorbitant amounts of money while having much lower impact factors. To me it makes little sense that the people who pay for the research must also pay for their work to be published (although this is almost always the case, even for articles published behind paywalls!). Furthermore, publications are often written in a way that is incomprehensible to the general audience. I believe that we, as scientists, have the responsibility to deliver our findings to the general public through *understandable* and *affordable* knowledge transfer. They are the ones paying for it, after all (for universities, research is mostly funded through taxpayers money).

During my PhD, I have tried to spread our findings as much as I could, through plenary lectures and masterclasses at conferences, talks at local festivals, visiting high- and elementary schools and finally with the initiation of a local outreach programme called *Noorderlab* aimed to inform and engage local citizens about the research being done in their own city. None of these efforts would have been possible without the tremendous support from my supervisor, professor Sijbren Otto, who has always encouraged these initiatives.

The *Noorderlab* initiative would not have been possible without the tremendous support from Gaël Schaeffer, Ingeborg Veldman, Kas Jansma, Amanda Brouwers, Renske de Jonge, Annelies van Santen and - last but not least - the Koninklijk Natuurkundig Genootschap (KNG), who funded the initiative.

CHAPTER 1

Introduction: where are we now and where are we going?

“LIVING MATTER EVADES THE DECAY TO EQUILIBRIUM” – ERWIN SCHRÖDINGER, 1944

In this chapter, we will explore the transition from inanimate matter into gradually more life-like systems with the purpose of – ultimately – discovering entities that we could consider to be alive. We will begin with a discussion about our own ancestry and continue back in time to discuss how life could have emerged: is the origin of life the result of a single random event or is there some force that acts upon nature to produce ever more complex systems? We progress towards current and past definitions of life and conclude with an alternative: instead of defining life we should look at which functional subsystems are integrated in known living things: self-replication, metabolism and compartmentalisation. We will discuss each of these subsystems, their importance from an origin of life perspective but also ways in which they can be approached synthetically. Dynamic combinatorial chemistry is introduced and used to develop the self-replicators that are used in the subsequent chapters of this thesis. Finally, we discuss the integration of metabolism and compartmentalisation in self-replicating systems: how this could potentially be achieved and ways in which it already has. Indeed, de novo life might not be as elusive as it once was thought to be.

1.1 THE DYNASTY OF LIVING THINGS

Living things are all around us. Who is not familiar with the irritating mosquito that buzzes around one's head at night, the spider that catches it and the human who screams in fear for his arachnid housemates? Earth, the planet all known living things call their home, hosts vast quantities of tremendously ingenious and diverse creatures: from chemolithoautotrophic bacteria living near hydrothermal vents¹ to the endangered Rüppell's Griffon – flying so high it is known to have collided with commercial airplanes.²

With some animals we often feel a strong sense of familiarity and recognition. Primates in particular exhibit behaviour not unlike that of our own.³ When the bonobo *Kuni* discovered a dazed sterling that had flown against the glass wall of his enclosure, he took it up to a

high treetop, spread its wings and released it.⁴ With others – such as the arachnids – most humans feel a sense of fear or repulsion.⁵ It is curious that we often value other living entities based solely on their likeness or usefulness to us, even though all known living things should be considered to be far family. Indeed, all known living things share a single ancestral species known as LUCA: the last universal common ancestor (see **Figure 1**).^{6–10}

Although no fossil record of LUCA exists, its existence can be inferred from the comparison of the genomes of modern organisms.⁹ For us to be descendants of such an organism, LUCA must have contained a rudimentary set of functional machineries allowing it to un-

Figure 1: A two-domain tree of life. The last universal common ancestor (LUCA, bottom) was a population of organisms from which all known living things are descendant. Initially, it was thought that archaea, eukaryotes and bacteria all directly descended from LUCA but recent evidence suggests that eukaryotes descend from prokaryotes.^{8,11}

dergo Darwinian evolution. But even if LUCA is indeed the ancestor to all known living things on earth, it is not necessarily the *first* living

being. LUCA undoubtedly originated from something far simpler. If we walk down the path of the ancestry of LUCA itself, at what point would one consider the chemical systems we find no longer to be *alive*? At what point does biology transition back into chemistry? The ancient question of what Life really is and what it means to be *alive* has fascinated the scientific world since recorded history. Indeed, the origin of life on Earth and the requirements for life to be able to emerge are still one of the biggest challenges in modern science.

1.2 THE ODDS OF OUR UNLIKELY EXISTENCE

It is extraordinarily improbable – even bordering on the impossible – for a single event to have occurred that has led to the sudden emergence of life on Earth. Imagine a short ribonucleic acid (RNA) strand that consists of the genetic monomeric building blocks common in all living things: nucleotides. In RNA, there are four different nucleotides that can be incorporated in such a strand (nucleic acid). Various RNA self-replicators have been developed, so let us take one as an example that consists of 61 nucleotides.¹² The possibility of such a sequence being randomly generated from a mixture of four different nucleotides is:

$$\frac{1}{4^{61}} = 1.9 \times 10^{-37} \quad (\text{Eq 1.1})$$

Even if one hundred million different combinations would be attempted every *microsecond*, it would still take 10^{11} billion years to synthesize one of each possible combination. This is roughly 7 billion times the age of the universe.¹³ And even then, we only have a self-replicating RNA strand – something no one would consider to be truly “alive”. Alternatively, life could have emerged not by a single leap but through the progression of intermediate stages of gradually becoming more life-like (see **Figure 2**).^{14,15}

In such a scenario, the various evolving systems at each intermediate stage must have a sufficiently long lifetime to be able to develop themselves further and increase their life-likeness. These intermediate stages cannot be populated statistically: that would result in the same probability as life emerging in a single leap. Instead, they must be populated in an out-of-equilibrium fashion that must constantly

Figure 2: The emergence of life as a single event (left) and as a progressive, step-wise process that gradually increases the life-likeness of a chemical system (right).

be maintained.¹⁶ Self-replicating systems are particularly suited for this purpose. Even though individual replicators might have a short lifetime, their ability to produce new generations allow for an overall long lifetime of an entire *population* while enabling evolvability of the system. Furthermore, replicating systems can only maintain a stable population when their formation and decay are balanced. This requires a certain stability that affects the *population* of replicators, and not the individual constituents, known as dynamic kinetic stability (DKS).¹⁷ Systems of self-replicating molecules could progress through the intermediate stages by increasing their complexity, acquiring additional functionalities and slowly evolving into something that could be considered *alive*.¹⁸ But when do we consider something to be alive?

1.3 DEFINING LIFE

Since the sixties, mankind has started to look for extra-terrestrial life using (radio) telescopes, laser detectors and space probes.¹⁹ But how would we – or rather astrobiologists – be able to recognize it? What exactly should they be looking for? Scientists, philosophers and other creative spirits have attempted to postulate a theory of life for

a very long time indeed. Aristotle once wrote “*By life I mean both nutrition through itself and growth and decline*” in his treatise about the soul.²⁰ Although pointing in the right direction, this definition would mean that a forest fire should be considered alive: it consumes wood, grows and declines (when it runs out of fuel).

More recently, a committee called the “Exobiology Discipline Working Group” led by Dick Young formulated the famous NASA working definition of life, often accredited solely to Gerald Joyce.^{21,22} It states that life is a “*self-sustaining chemical system capable of Darwinian evolution*”, similar to what Horowitz and Miller proposed in 1962.²³ Including *Darwinian evolution* infers various machineries life should contain: replication, mutation and selection. A further important aspect is the use of the words “system” and “capable”. This avoids the frequent counterexample that a single individual would not be considered alive since it would not be able to reproduce (with few exceptions). For example, Lonesome George (the last known Pinta Island tortoise), would not be considered alive since there is no way he would have been able to reproduce. However, this definition applies to a population, not an individual. This results in a distinction between what *life* is and whether a single entity is *alive*. The working definition is still being adjusted as we learn more about what that distinction truly means.²⁴

Another interesting take on living systems is the notion of *autonomous agents*. In his book *Investigations*, Kauffman proposed that all living systems act on their own behalf in a certain environment.²⁵ Instead of moving towards a thermodynamic sink, living systems will act according to their own agenda (in biology, this behaviour is known as teleonomy). In my opinion, this is not *necessarily* contrary to previously established definitions, but could be considered an additional requirement – albeit one that seems rather vague and quite ungraspable when considering the laws of physics and chemistry.

Still, any definitions and all-encompassing theories on life we come up with is related to the extraordinarily small sample set ($n=1$) of known living things, since they are all related to one another. Who knows what forms of evolution could have existed on the early earth?²⁶ In the metabolism-first scenario – one of the scientific theories on the origin of life on Earth – non-Darwinian evolution must have played a key role in forming the first molecular replicators.²⁷ Indeed, some believe an all-encompassing theory of life and an accompanying definition is meaningless – at least until the discovery of extra-terrestrial

life or the synthesis of life *de novo*.^{15,26} So how do we move forward from here? It looks like it has become an unwitting race between astrobiologists and systems chemists: either we discover life that has originated elsewhere, or we synthesize our own. The remainder of this dissertation will discuss the latter: from the particular properties chemical systems should have for them to be considered life-like, ways in which such individual sub-systems can be obtained and eventually combined.

1.4 FUNCTIONAL PROPERTIES OF LIVING SYSTEMS

For mankind to be able to synthesize life in a lab, we must have an idea of which functional machineries to implement in chemical systems for them to become more and more life-like. For this, a rather operational approach is required in which we establish the machineries essential for a minimal form of life to adhere to NASA's working definition.^{28,29} All known living systems are capable of reproduction, are separated from their environment and display metabolic activity (see **Figure 3A**). Living systems reproduce by making copies of themselves (in one way or another). Small mistakes can be made during this copying process, that enable Darwinian evolution through mutation and selection. The separation between an organism and its environment is essential to keep the organism's components together. Furthermore, it enables higher reaction rates by generating local concentrations of reactants and can lead to energy and material transport through the boundary, allowing for out-of-equilibrium growth.³⁰ Finally, metabolic activity enables the harvesting of energy. Living systems are dissipative: they consume energy to maintain their out-of-equilibrium steady state (dynamic kinetic stability). A metabolism further enables the synthesis of required components to empower other machineries. The true challenge will be the integration of all of these machineries in a complex chemical system that subsequently can exhibit *open ended* Darwinian evolution (see **Figure 3B**). In this sense, open ended means that we, as creators and observers, obtain a fully integrated complex chemical system that increases its complexity through evolution while allowing for ongoing adaptation to its environment.³¹

In the next sections we will discuss these individual machineries or *sub-systems* in further detail. We will examine ways in which they can be made – in the lab or otherwise – and how they might be integrated

in a single chemical system.

Figure 3: Essential functional machineries and their implementation in self-replicating systems. A) Replication (1), metabolism (2) and compartmentalization (3) must be integrated in a single system. B) Open ended evolution where replicators can evolve, adapt to their environment and remain in an out-of-equilibrium state.

1.5 SELF-REPLICATION

Reproduction is exhibited by all known living systems and can take curious shapes and forms. Sexual reproduction relies on two organisms combining their genetic material to produce offspring (common in mammals) while asexual reproduction needs only a single parent organism. Hence, asexual reproduction leads to a faster growth than sexual reproduction. Bacteria often use binary fission to propagate themselves through time in which the bacterial cell elongates and splits in two. A great many machineries are needed for such an – intuitively – simple process.³² In this process, bacteria create an identical copy of themselves (replication) while still allowing for random mu-

tations. In chemical systems, we will look at replication rather than reproduction: this is by far the simplest approach and could still allow for mutations by the incorporation of multiple, different building blocks (see **Chapter 5**).

All molecular self-replicators are autocatalytic: they catalyse the formation of themselves. However, not all autocatalytic systems are self-replicators. For a system to self-replicate, some form of information transfer between the original molecule and the replicate is required.^{33a,b–35} Self-replicating systems have been notoriously difficult to synthesize. In 1984, a pentanucleotide consisting of guanine and cytosine was shown to be able to act as a template for the complementary strand. However, no autocatalytic cycle was obtained, since the strands would need to separate to enable further rounds of replication.³⁶ Two years later, Von Kiedrowski discovered the first autocatalytic cycle based on DNA in which a hexameric template directs the condensation of two trinucleotides.³⁷ Other nucleotide-based replicating systems quickly followed^{38,39} and was soon expanded to peptide-based systems⁴⁰ and completely synthetic systems.^{41,42}

The order (the relationship between the reaction rate and the concentration of the species involved in it) of Von Kiedrowski's first replicator was assumed to be $\frac{1}{2}$ in replicator (but definitely <1), resulting in parabolic growth of the replicating species. The biggest hurdle to overcome in duplex-forming replicating systems is always the dissociation between template and copy (see **Figure 4**). Many other autocatalytic systems reported orders close to 1 in the following years, but the first replicator exhibiting exponential growth was not described until 1998,⁴³ when Von Kiedrowski used surface immobilisation to avoid the pitfall of product inhibition. Indeed, the functional integration of molecular self-replication in life-like systems usually requires exponential growth.

Sub-exponential growth in a system with multiple replicators typically leads to co-existence.⁴⁴ Even though this could lead to a larger variability in the diversity of the replicators (if they would incorporate the same building blocks) it would also mean that Darwinian evolution would remain out of reach: only if species propagate exponentially can useful hereditary traits be attained by the whole population. Furthermore, the extinction of *weaker* replicating systems could also never be achieved during sub-exponential growth.^{34,45} Competition in an evolutionary scenario can only be achieved when all replicators *can* grow exponentially and only selection pressures can curb that

growth.

Exponential growth in autocatalytic systems can be obtained if product inhibition is avoided. The previously mentioned surface immobilisation by Von Kiedrowski allowed for a separation of the duplex by immobilizing the template molecule (see **Figure 5A**). Other approaches utilize weak binding between the template and substrate, often resulting in *non-rate limiting* product inhibition¹² and cross catalysis⁴⁶ or avoid the dissociation step altogether by utilizing supramolecular polymerisation (see **Figure 5B**).^{47–49}

Figure 4: A minimal self-replicator. The template (1) binds the building blocks (2) who are subsequently able to react with one another due to the close proximity of their reactive groups (3), generating another template molecule. The dissociation stage of the template duplex (4) is often the rate limiting step. Templates that have recognition sites on both sides could result in oligomer formation and prevent product inhibition by duplex formation.

1.6 DYNAMIC COMBINATORIAL CHEMISTRY

There are many ways in which individual subsystems known to living organisms have been synthesized in the lab or even in one's household: dish soap will readily form self-assembled compartments when added to water capable of storing hydrophobic material. The challenge lies in the facilitation of an active communication between the individual subsystems. One way of approaching this problem is by linking the dynamicity of a system to its properties.

In our group, we have used *dynamic combinatorial chemistry* (DCC) as a tool to develop self-replicators that exhibit exponential growth.

Figure 5: Two examples of exponential growth in self-replicators. A) Surface-immobilisation leading to exponential replication in DNA analogues (inspired by reference 43). A single strand is irreversibly attached to a solid support surface (1). Complementary fragments in solution are bound to the template and generate the complementary strand through chemical ligation (2). The template copy dissociates and irreversibly binds to the solid support at a different location. B) Self-replication through an elongation-breakage mechanism.⁵⁰ Building blocks form hexamer macrocycles that are either sequestered from solution or produced through templated reactions at fibre ends, elongating the fibre nuclei (1). If mechanical agitation is applied to the system, long fibres break and double the amount of growing fibre ends, resulting in exponential growth of the hexamer macrocycle (if continuous mechanical agitation is applied).

In DCC, building blocks can interact with one another reversibly (either covalently or non-covalently) resulting in a *library* that contains complex mixtures of interconverting molecules. These libraries are under *thermodynamic control*: the interconversion of molecules allows for the product distribution to move towards the thermodynamic minimum of the system.⁵¹ DCC has many potential applications such as drug discovery in the pharmaceutical industry and the synthesis of otherwise inaccessible molecules.⁵² Since the concentrations of the library constituents depend on their stability, external stimuli such as binding sites on target molecules can lead to the amplification of a particular library constituent – the one that binds to the target.^{53–55} Since DCC is built on the formation of complex networks of interconverting molecules, its use in synthetic biology and systems chemistry is not surprising.⁵⁶ Since such a diversified product distribution is generated, unforeseen molecular structures can be obtained that can exhibit properties that further complexify the library (see **Chapter 4**).

Many reversible chemistries exist^{51,57} that can enable the use of DCC: from covalent hydrazone exchange to non-covalent metal-ligand coordination or hydrogen bond exchange.

We utilize disulfide chemistry to implement reversible reactions in dynamic combinatorial libraries (DCLs). Building blocks functionalized with thiol groups can bind to one other upon oxidation by forming disulfide bonds. Free thiolate anions enable building blocks to exchange with one another (see **Figure 6A**). When building blocks are functionalised with two thiol groups, a variety of macrocyclic and linear structures can be formed (see **Figure 6B**).

We have developed a range of pentapeptide functionalised building blocks that contain a benzene-1,3-dithiol core moiety.^{48,58} The pentapeptide alternates the incorporation hydrophobic and hydrophilic amino acids to promote β -sheet formation. Upon oxidation of the building blocks, a mixture of interconverting macrocyclic species

Figure 6: Reversible covalent chemistry utilizing disulfide exchange. A) Different building blocks functionalized with thiol groups readily exchange with one another (other possible products not shown). B) Building blocks functionalized with two thiol groups form an interconverting mixture of macrocycles upon oxidation (linear- and other possible species not shown).

emerge that are initially dominated by small macrocycles like trimers and tetramers. As these DCLs are under thermodynamic control, the relative stability of the macrocyclic species determines library composition: eventually a larger macrocyclic species appears (we generally observe hexamer, heptamer or octamers) that can stabilise themselves by stacking on top of one another. Both β -sheet formation from the pentapeptides and π - π stacking of the aromatic core moieties are the foundation of the stacking stabilisation and result in the

formation of supramolecular polymers. Subsequently, the small fibre nuclei can elongate by promoting the formation of the macrocyclic species that constitute the fibre (either by sequestration or templation). When mechanical agitation is applied to such a system, elongated fibres can break, doubling the number of growing fibre ends. This elongation-breakage mechanism enables exponential growth in these self-replicating systems (see **Figure 7**).⁵⁰ An attribute of these self-replicating systems is that upon addition of the replicating species to a mixture of small macrocycles the self-replication process immediately begins. This process is called *seeding* and is extensively used in mechanistic studies (see **Chapter 2** and **Figure 8A**). The supramolecular structures can be visualised using conventional transmission electron microscopy (TEM) and generally have lengths of hundreds of nanometres with widths of a few nanometres (see **Figure 8B**).

We have recently discovered that the self-replication mechanism of these systems is far more complex than described here but the general ideas remain the same (for a more detailed analysis, see **Chapter 2** and **Chapter 3**). The peptide-based self-replicators described above are the main ingredient and starting point of this dissertation. The ultimate goal is to move the complex chemical systems generated through DCC further up the steps shown in **Figure 2** – slowly increasing the life-likeness of these synthetic systems.

We have come a long way since Von Kiedrowski's first replicator. A large variety of self-replicating systems composed of different types of building blocks exist (biological, synthetic and anywhere in between) that can now even exhibit exponential growth. But life is far more complex than simple self-replication. Cells consist of incredibly complex and inventive reaction networks in which chemical reactions are catalysed by enzymes *whilst being able to replicate*.

The next step towards life *de novo* is the complexification of individual molecules able to replicate themselves. These self-replicating systems must acquire additional functionalities beyond replication and progress towards the domain of complex chemical reaction networks.

1.7 METABOLISM

A metabolism is generally described as all of the chemical reactions needed to sustain a living system. This typically includes the

Figure 7: The (simplified) mechanism of self-replication of the pentapeptide replicators developed in our group. An aromatic core moiety functionalised with two thiol groups and a pentapeptide forms small macrocycles when oxidized (1). These macrocycles are under thermodynamic control and interconvert into one another readily (2). Eventually, a larger macrocyclic species emerges that can stabilise itself by stacking on top of one another (3). The fibre ends elongate (4) and when mechanical agitation is applied to the system, the fragmentation of fibres results in a doubling of the growing fibre ends (5), enabling exponential growth in these systems.

consumption of food, the subsequent generation – and storing – of energy and building blocks and the elimination of waste products. Metabolic processes are separated in two distinct classes: anabolic and catabolic reactions. In anabolic reactions, molecules are built up from small building blocks, typically requiring the consumption of an energy source like adenine triphosphate (ATP) or sunlight (see **Figure**

Figure 8: A) When adding a small amount of pre-formed replicator fibres to a mixture of small macrocycles, the self-replication process begins immediately. Alternatively, the emergence of the self-replicator without seeding progresses much slower. Both seeded and unseeded libraries were monitored by an ultra-performance liquid chromatography (UPLC) system. Only the hexamer species is shown for clarity. B) Typical negative stain TEM micrograph of the supramolecular structures obtained through self-replication. In this particular case, the fibres are built up out of hexamer macrocycles. The different widths observed are due to the lateral association of multiple fibres.

9A). In contrast, catabolic reactions break down molecules into smaller constituents to produce energy and building blocks (see **Figure 9B**).

These processes allow for a living system to harvest and consume energy, keeping it in an out-of-equilibrium state, while the in-house production of building blocks enables growth and reproduction. The metabolic processes of known living systems are tremendously complex and involve thousands of interconnecting reactions (see **Figure 10**).^{59,60}

Suffice to say that all biological systems rely on the consumption of energy to function: they are dissipative. Recently, examples have arisen of dissipative systems capable of autocatalytic self-assembly^{61,62} and self-replication⁶³. Such systems can be maintained in an out-of-equilibrium fashion by the consumption of a chemical fuel.

Even though these are significant steps in the direction of out-of-equilibrium systems, they lack the metabolic aspect required

Figure 9: The two classes of metabolic processes. A) Photosynthesis is an anabolic process. Glucose and oxygen are formed from CO_2 , H_2O and requires (sun)light as an energy source. B) Cell respiration is a catabolic process. Here, CO_2 , H_2O and energy (in the form of ATP) are formed while glucose and oxygen are consumed.

in living systems: the replicators are not responsible for the catalytic formation of their own building blocks or the storage of energy.

In this section we will focus on the minimal requirements for a life-like system to exhibit (proto) metabolic processes.

From an *origin of life* perspective, the emergence of complex inter-connecting molecular networks before the development of replicators is one of the theories on the origin of life itself known as the *metabolism first scenario* (as briefly mentioned in the section **1.3 Defining Life**). In this situation, small molecules are spontaneously formed from chemicals presumed to be available on the early Earth

Figure 10: Some 500 interconnecting reactions occurring in a part of the metabolic pathways in an emblematic cell. Each molecule is represented by a filled circle and the enzyme catalysed reactions by the lines between them. Image credit: Zlir'a (public domain, but surely heavily inspired by reference 59).

under thermodynamically favorable conditions.^{64,65} These molecules could subsequently react in various ways, increasing the complexity of the system and resulting in the formation of a complex molecular

network. Eventually, some networks would produce autocatalytic sets of reactions resulting in the specific amplification of the molecules involved.^{66,67} This scenario would then solve the typical question “where did the first replicator come from?”.²⁷ However, metabolism-first theories are often criticized as not being chemically plausible^{68,69} or to have an inherent lack of evolvability.⁷⁰

In any case, the emergence of metabolic processes in life-like systems must have been a crucial stage in the transition from inanimate to animate matter. Molecular replication is typically governed by thermodynamically down-hill reactions, while cellular replication is made possible only by the consumption of an external energy source (i.e. metabolism). It could even be argued (by some) that the development of a metabolism would directly lead to teleonomic (*purposeful*) behavior by converting the life-like system from thermodynamically driven reaction to kinetically driven ones.⁷¹ Moreover, if one metabolic replicator would compete with a non-metabolic replicator for (sparse or fluctuating amounts of) building blocks, metabolic replicators would drive the non-metabolic replicators to extinction. This implies that the invention of metabolic capabilities could have enhanced the kinetic stability of early replicating systems.⁷²

However, we will approach metabolic processes from a life *de novo* perspective in which self-replicators have already been developed. In living organisms, enzymes typically enable metabolic pathways by acting as catalysts and lowering the activation energy of reactions.

By extension, the first milestone on the path towards the invention of a full-fledged metabolism in life-like systems would be the development of catalytic functionalities in systems of self-replicators. Indeed, if self-replicating systems would be able to benefit from their newly developed catalytic capabilities (e.g. by increasing their replication rate), one could argue that proto-metabolic pathways are being developed.³³ Furthermore, catalysis could also lead to anabolic reactions in which the replicator building blocks could be synthesized from smaller precursors. Alternatively, spontaneous replicator formation from its building blocks might be inhibited by the introduction of protecting groups. Catalytically active replicators could then catalyse the deprotection of these building blocks, enabling the catabolic generation of building blocks (see **Chapter 6: Conclusions**).

1.8 INTEGRATION OF METABOLISM

The first system in which a known (synthetic) self-replicator could exhibit catalytic activity was discovered by Rebek et al. in 2010, where a synthetic autocatalytic molecule was able to catalyse hydrogenation or nucleophilic addition reactions.⁷³ However, in this system replication and catalysis could not be coupled to one another: different solvents were required to enable either self-replication or catalysis. Another interesting approach has used known RNA replicases⁷⁴ derived from the *Azoarcus* bacterium to implement a rudimentary catabolism. Here, the inactive RNA substrates *WXY* and *Z* (see **Figure 11B**) can form the covalent RNA catalyst *WXYZ* that can catalyse its own formation. Both *WXY* and *Z* can be modified with a foreign sequence (consisting of 27 nucleotides) on the 3' end that contain a short trinucleotide spacer required for a transesterification reaction.^{74,75} Surprisingly, the unmodified catalyst *WXYZ* is formed even if only modified building blocks are fed to the system. Moreover, the modified *WXY* and *Z* can be transformed into usable building blocks by the *WXYZ* catalyst (catabolism) that can subsequently utilise these building blocks to drive its own formation (anabolism, see **Figure 11A**).⁷⁶ However, no exponential amplification of the *WXYZ* catalyst is observed in these systems, limiting their evolvability. Furthermore, the system is far from synthetic, as they are built up out of nucleotides in a sequence that is the product of evolution.

Our group has recently discovered that the previously described (see section **1.6 Dynamic Combinatorial Chemistry**) self-replicators are able to display two distinct emergent catalytic capabilities: the generation of *cofactors* to enable photooxidation and the ability to catalyse chemical reactions through active site generation.

The formation of fibrous supramolecular structures is partially stabilised by β -sheet formation. Photosensitizers like Thioflavin T (ThT) are known to bind to β -sheets and are often used to investigate protein aggregation due to the changes in their fluorescence signal upon binding.⁷⁷ Indeed, such photosensitizers are able to bind to the families of replicators our group has developed⁵⁸ and are used to assess fibre formation (see **Chapter 5**). Similarly, the photosensitizer Rose Bengal was found to be able to bind to the fibrous structures. Rose Bengal can generate singlet oxygen ($^1\text{O}_2$) upon irradiation which, in turn, is able to enhance thiol oxidation. The recruitment of such a *cofactor* by the self-replicators enables a positive feedback mechanism

Figure 11: Coupled metabolic processes to replication. A) Precursor material is initially not usable due to a modification with a 27-nucleotide foreign sequence (in red). The *WXYZ* catalyst can cleave the foreign sequence from the raw material, resulting in precursor building blocks (*WXY* and *Z*) that – through an anabolic process – can be transformed into the catalyst. B) The autocatalytic generation of the covalent RNA catalyst *WXYZ* catalyst from inactive RNA substrates *WXY* and *Z*. Reprinted with permission from reference 76. Copyright©, 2018, OUP.

when irradiated with an appropriate wavelength. Interestingly, the photocatalysis can be tuned to only occur when the replicating species is present since the excitation wavelength of Rose Bengal red-shifts upon binding to the fibrous structures.⁷⁸

The emergence of active sites along the fibre is a product of the self-assembly mechanism. The non-covalent stacking of macrocycles inadvertently brings the peptide residues into close proximity with one another. If these peptide residues are protonated, electrostatic repulsion could lead to a perturbation of their pK_a , resulting in partial deprotonation. More specifically, we have discovered that lysine residues in the pentapeptide sequence become potent nucleo-

philes when deprotonated. Furthermore, they are positioned in relative hydrophobic environments along the fibre backbone. Indeed, we have discovered *two* chemical reactions that can be catalyzed by our fibres: the retro-aldol reaction of methanol and Fmoc deprotection (see **Chapter 4**). The latter results in the formation of dibenzofulvene (DBF), which is capable of promoting thiol oxidation⁷⁹, enabling a positive feedback mechanism.⁸⁰

While the photooxidation reaction requires the input of energy, the catalytic Fmoc deprotection does not. Yet both reactions have the same outcome: an increased rate of thiol oxidation. However, this reaction is spontaneous in nature (i.e. it is thermodynamically downhill). Thus, an essential part of the road towards life *de novo*, the consumption of energy to maintain an out-of-equilibrium state while generating building blocks has yet to be fully implemented.

1.9 COMPARTMENTALISATION

A compartment is a useful boundary that enables us to determine where an organism ends and its environment begins. For single-celled organisms like bacteria such a boundary is easily established, but for multicellular organisms it is more difficult to determine where this border exactly is (are dead skin cells part of your body? What about hair and fingernails?). The boundaries between organism (multicellular or otherwise) and environment have many functions: they tend to keep our important components together;⁸¹ they form a protective barrier against outside attack and can sometimes function as sensory systems. Spatial self-organisation is an essential attribute in all living systems – something unattainable without the formation of compartments.

When approaching compartments from a life *de novo* or even origins of life perspective, the importance of compartments becomes apparent quickly.^{82,83} Without compartments to hold essential molecules together, diffusion would quickly eliminate any possibility of self-organisation, let alone self-replication. Indeed, compartmentalisation is essential for protocell formation as it can encapsulate the various subsystems required.

The integration of compartments in life-like systems can enable a local high concentration of reactants, speeding up single reactions or entire reaction networks.^{84–86} Reactants in a compartmentalised reaction network will run out if not replenished, and waste will accumu-

late if not depleted. Indeed, an excellent control over which molecules are able to permeate the boundary of life-like systems and which do not is an essential regulatory machinery. Nowadays, intricate pore forming proteins (such as porins) enable cells to regulate the transport of matter. It is, however, unlikely for such complex membrane transport machineries to have been present in very minimal forms of life and by extension useful in the quest towards life *de novo*. It would be more likely that the compartments themselves allow for the transmission of certain materials through the boundary.

When thinking about how life could originate – or how we could synthesize our own – one quickly runs into the issue of parasitism.^{82,87} Recently, the spontaneous emergence of parasitic RNA strands in a translation-coupled RNA replication system was discovered.⁸⁸ In this system, the parasitic RNA strands do not contain the encoding of the RNA replicase, drastically reducing their size and thus increasing their replication rate. Due to their much higher replication rate, the formation of the parasitic species inhibits the formation of the RNA host. Such behaviour can be expected in any complex self-replication system in which a certain product (P) is produced that can promote replication. Less complex, parasitic replicators can benefit from the formation of P while selecting for the highest replication rate, quickly dominating the system (see **Figure 12A**).

The introduction of compartments can help to address the issue of parasitism. Compartmentalisation of the previously mentioned RNA translation-coupled replication system resulted in oscillating population dynamics.⁸⁹ In this case, the host RNA population did not collapse, but instead coexisted with the parasitic RNA. In another experiment,⁹⁰ the transient compartmentalization of RNA replicators by droplets generated in a microfluidic channel prevents the extinction of the replicator by parasitic RNAs. Indeed, the physical separation of host and parasite limits parasite propagation and prevents collapse of the host population. When we extend this to the previous discussion about parasitism in self-replicating systems, compartments can also prevent the exploitation of one replicator by another (see **Figure 12B**). In this simplified scenario, the blue replicator cannot utilize the positive feedback loop generated by the catalytic abilities of the red replicator, while in the un-compartmentalised situation (as shown in **Figure 12A**), the blue replicator can utilize the same positive feedback loop without contributing to the growth of the red replicator.

Figure 12: Open versus compartmentalised systems of replicators. A) The red replicator is able to catalyse the formation of a certain molecule (denoted with the star symbol) that is able to promote the production of building blocks for both replicators from precursors. The blue replicator does not exhibit any catalytic activity but is able to exploit the catalytic prowess of the red replicator. B) If both replicators are compartmentalized (note that this is a strongly simplified representation) the blue replicator cannot reproduce as fast as it is spatially separated from the catalytically active replicator. In this scenario, the red replicator will quickly dominate the system.

Darwinian evolution relies on *selection*. In a system where different replicators compete for the same food, the replicator that is able to reproduce more rapidly will be able to propagate their hereditary information while those that are slower or produce fewer offspring will eventually die out. However, if the replicators as shown in **Figure 12** exist freely in solution, there can be no selection based on beneficial mutation.^{28,82,91} Compartmentalisation allows for a population of closely related replicators to remain in close proximity to one another and benefit from any advantageous mutation – resulting in a higher rate of replication.^{91–93} So not only does compartmentalisation solve

a variety of technical problems, it is essential to enable selection in the ultimate objective: open ended Darwinian evolution in a synthetic system.

1.10 THE INTEGRATION OF COMPARTMENTALISATION

Even though it is generally accepted that amphiphilic molecules capable of forming primitive membranes were around on the early Earth^{82,94}, there are many other ways in which complex chemical networks can be compartmentalised – even without the use of (organic) membrane forming molecules. For instance, inorganic compartments that consist of FeS can be formed near hydrothermal vents⁹⁵, ice eutectic phases can lead to compartment formation with a high local concentration of reactants enclosed by ice crystals^{96,97} and even atmospheric aerosol droplets have been shown to supply spatially separated microreactors.⁹⁸ Indeed, there exist many different ways in which compartmentalisation can be achieved. As mentioned before, compartment formation allows for selection on a group level – but this requires the compartments to eventually reproduce and divide to create a next generation (see **Figure 13**)

As the compartmentalised self-replicators produce copies of themselves, their environment would become more and more crowded

Figure 13: Two replicators, red (R) and purple (P) reproduce and lead to the formation of a new green (G) replicator through mutation. The compartment grows to accommodate the additional material and to enable division, resulting in a generation of two new compartments that contain parts of the material encompassed by the parent compartment. Selection is enabled if one of the offspring compartments can replicate faster than the other.

– if the compartments do not grow with them. Compartments must have some self-replicating properties *themselves*. Not only to prevent overcrowding: it is also essential to enable division of compartments for if they were not able to self-replicate, they would become smaller and smaller as new generations are produced.

Since the nineties, a variety of biologically relevant self-reproducing compartments have been discovered⁹⁹: vesicles^{100–102}, micelles¹⁰³, reverse micelles¹⁰⁴ and even water-oil droplets¹⁰⁵ have been shown to self-reproduce. However, self-reproduction alone is not enough. The biggest hurdle to overcome is division. Previously, compartment growth and division were coupled by the addition of building blocks to a vesicle and applying mild mechanical agitation. Interestingly, only trace amounts of leakage from the compartments was observed.¹⁰⁶ More recently, the replication of DNA was coupled to the growth and division of giant vesicles^{107,108}. However, as Luisi pointed out, in these systems not all components are synthesized in situ: the addition of a DNA polymerase and a catalyst is required to continue the formation of next generations.¹⁰⁹ Another recent example was able to couple the division of compartments to its replication: amphiphilic imines in chloroform can catalyse the formation of themselves by bringing hydrophobic and hydrophilic precursors in close proximity. The autocatalyst lowers the interfacial tension between chloroform and an aqueous phase, resulting in division of the droplets.¹¹⁰

Another interesting approach towards the implementation of compartmentalisation is the use of coacervates. Coacervates are membrane-less compartments formed through liquid-liquid phase separation. They are generally formed in aqueous solution by the association of macromolecules with one another (e.g. by mixing oppositely charged polymers), resulting in polymer-rich coacervate droplets and a dilute solution. Alternatively, non-associative phase separation can be achieved by using macromolecules that do not group together readily. In that case, two different phases are obtained, each rich in one of the two macromolecules.¹¹¹ The fact that these phase-separated compartments have no membranes can solve one of the major issues in protocell development: the difficulties associated with the transportation of material through the membrane. As mentioned before, protocells should not incorporate complex transportation machineries as they would be improbable to have been developed at such an early stage of life-like systems. Not having a membrane at all would avoid this issue altogether. Furthermore, coacervates have

been shown to sequester a range of materials from solution, further encouraging their use in an origin of life scenario.^{112,113} Such behaviour leads to a local high concentration of reactants, increasing reaction rates while also *keeping important components together*.^{114,115}

Of course, there are still plenty of challenges to overcome. First and foremost, coacervates tend to coalesce – something that must be prevented if compartment division is to be realised. Furthermore, replicating coacervates have yet to be realised, let alone the coupling of replication between compartments and replicators – although these seem more of technical problems than conceptual ones. The use of complex coacervation as a tool to develop protocells finds its origin in Oparin's *The Origin of Life*¹¹⁶, it has only recently gained its well-deserved momentum in the field of synthetic biology.¹¹⁷

Finally, I would like to point out an alternative scenario in which replication and compartmentalisation could be coupled. If the replicators *themselves* promote the production of compartment forming material (e.g. by catalysing its formation) – emerging replicators would lead to the formation of compartments that would grow as replicators do. Moreover: if the product of such a reaction would promote replicator growth, we could very well be on our way towards the integration of a proto-metabolism at the same time (see **Chapter 6: Conclusions**).

There exists no synthetic system that fully integrates all previously described sub-systems, but the scientific community has made enormous advancements over the last decades. Life *de novo* – although tremendously complex – seems to be an achievable aspiration.

ACKNOWLEDGEMENTS

I have attempted to write this chapter in a way that it would be understandable for people who are not experts on this topic. For that purpose, I would like to extend my thanks to N. C. Soer for proofreading this chapter. Additionally, I would like to extend my thanks to C. Pappas for making sure that experts might want to read it as well.

BIBLIOGRAPHY

1. Stetter, K. O. Hyperthermophilic procaryotes. *FEMS Microbiol. Rev.* **1996**, 18, 149–158.
2. Laybourne, R. C. Collision between a vulture and an aircraft at an altitude of 37,000 feet. *Wilson Bull.* **1974**, 86, 461–462.
3. de Waal, F. Een tijd voor empathie. (Olympus, **2015**).
4. de Waal, F. The empathic ape. *New Sci.* 188, 52–54 (**2005**).
5. Davey, G. C. L. Characteristics of individuals with fear of spiders. *Anxiety Res.* **1991**, 4, 299–314.
6. Theobald, D. L. A formal test of the theory of universal common ancestry. *Nature*, **2010**, 465, 219–222.
7. Kim, K. M., Caetano-Anollés, G. The proteomic complexity and rise of the primordial ancestor of diversified life. *BMC Evol. Biol.* **2011**, 11.
8. Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S. & Martin, W. F. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **2016**, 1, 1–8.
9. Koonin, E. V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* **2003**, 1, 127–136.
10. Darwin, C. R. On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life. (John Murray, 1859).
11. Williams, T. A., Foster, P. G., Cox, C. J., Embley, T. M. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature*, **2013**, 504, 231–236.
12. Paul, N., Joyce, G. F. A self-replicating ligase ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99, 12733–12740.
13. Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Bal-

lardini, M., Banday, A. J., et al. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.* (**2020**).

14. Bruylants, G., Bartik, K., Reisse, J. Prebiotic chemistry: a fuzzy field. *Comptes Rendus Chim.* **2011**, 14, 388–391.

15. Bruylants, G., Bartik, K., Reisse, J. Is it useful to have a clear-cut definition of life? On the use of fuzzy logic in prebiotic chemistry. *Orig. Life Evol. Biosph.* **2010**, 40, 137–143.

16. Pascal, R., Pross, A., Sutherland, J. D. Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics. *Open Biol.* **2013**, 3, 1–9.

17. Wagner, N., Pross, A. The nature of stability in replicating systems. *Entropy*, **2011**, 13, 518–527.

18. De Duve, C. The onset of selection. *Nature*, **2005**, 433, 581–582.

19. SETI at 50. *Nature*, **2009**, 461, 316–316.

20. Polansky, R. Aristotle's de anima. (Cambridge University Press, **2007**).

21. Deamer, D. W. & Fleischaker, G. R. Origins of life: the central concepts. (Jones and Bartlett, **1994**).

22. Benner, S. A. Defining life. *Astrobiology*, **2010**, 10, 1021–1030.

23. Horowitz, N. H., Miller, S. L. Current theories on the origin of life. in Progress in the Chemistry of Organic Natural Products vol. 10 423–459 (Springer Vienna, **1962**).

24. Vitas, M. & Dobovišek, A. Towards a general definition of life. *Orig. Life Evol. Biosph.* **2019**, 49, 77–88.

25. Kauffman, S. A. Investigations. (Oxford University Press, **2000**).

26. Cleland, C. E., Chyba, C. F. Defining 'life'. *Orig. Life Evol. Biosph.* **2002**, 32, 387–393.

27. Shapiro, R. A replicator was not involved in the origin of life. *IUBMB Life*, **2000**, 49, 173–176.
28. Balaram, P. Synthesizing life. *Curr. Sci.* **2003**, 85, 1509–1510.
29. Tsokolov, S. A. Why is the definition of life so elusive? Epistemological considerations. *Astrobiology*, **2009**, 9, 401–412.
30. Ashkenasy, G., Hermans, T. M., Otto, S., Taylor, A. F. Systems chemistry. *Chem. Soc. Rev.* **2017**, 46, 2543–2554.
31. Packard, N., Bedau, M. A., Channon, A., Ikegami, T., Rasmussen, S., Stanley, K., Taylor, T. Open-ended evolution and open-endedness: editorial introduction to the open-ended evolution I special issue. *Artif. Life*, **2019**, 25, 1–3.
32. Angert, E. R. Alternatives to binary fission in bacteria. *Nat. Rev. Microbiol.* **2005**, 3, 214–224.
- 33a. Adamski, P., Eleveld, M., Sood, A., Kun, Á., Szilágyi, A., Czárán, T., Szathmáry, E., Otto, S. From self-replication to replicator systems en route to de novo life. *Nat. Rev. Chem.* **2020**, 4, 386–403.
- 33b. Chen, C., Tan, J., Hsieh, M.-C., Pan, T., Goodwin, J. T., Mehta, A. K., Grover, M. A., Lynn, D. G. Design of multi-phase dynamic chemical networks. *Nat. Chem.* **2017**, 9, 799–804.
34. von Kiedrowski, G. Minimal replicator theory I: parabolic versus exponential growth. in *Bioorganic Chemistry Frontiers* (eds. Dugas, H. & Schmidtchen, F. P.) vol. 3 113–146 (Springer, Berlin, Heidelberg, **1993**).
35. Eigen, M., Schuster, P. The hypercycle: a principle of natural self-organization. (Springer-Verlag Berlin Heidelberg, **1979**).
36. Inoue, T., Joyce, G. F., Grzeskowiak, K., Orgel, L. E., Brown, J. M., Reese, C. B. Template-directed synthesis on the pentanucleotide CpCpGpCpC. *J. Mol. Biol.* **1984**, 178, 669–676.
37. Von Kiedrowski, G. A self-replicating hexadeoxynucleotide. *Angew. Chem. Int. Ed.* **1986**, 25, 932–935.

38. Zielinski, W. S., Orgel, L. E. Autocatalytic synthesis of a tetranucleotide analogue. *Nature*, **1987**, 327, 346–347.
39. von Kiedrowski, G., Wlotzka, B., Helbing, J. Sequence dependence of template-directed syntheses of hexadeoxynucleotide derivatives with 3'–5' pyrophosphate linkage. *Angew. Chemie Int. Ed.* 1989, 28, 1235–1237.
40. Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K., Ghadiri, M. R. A self-replicating peptide. *Nature*, **1996**, 382, 525–528.
41. Tjivikua, T., Ballester, P., Rebek, J. A self-replicating system. *J. Am. Chem. Soc.* **1990**, 112, 1249–1250.
42. Terfort, A. & von Kiedrowski, G. Self-replication by condensation of 3-aminobenzamidines and 2-formylphenoxyacetic acids. *Angew. Chemie Int. Ed.* **1992**, 31, 654–656.
43. Luther, A., Brandsch, R., Von Kiedrowski, G. Surface-promoted replication and exponential amplification of DNA analogues. *Nature*, **1998**, 396, 245–248.
44. Szathmáry, E., Gladkih, I. Sub-exponential growth and coexistence of non-enzymatically replicating templates. *J. Theor. Biol.* **1989**, 138, 55–58.
45. Wills, P. R., Kauffman, S. A., Stadler, B. M. R., Stadler, P. F. Selection dynamics in autocatalytic systems: Templates replicating through binary ligation. *Bull. Math. Biol.* **1998**, 60, 1073–1098.
46. Lincoln, T. A., Joyce, G. F. Self-sustained replication of an RNA enzyme. *Science*, **2009**, 323, 1229–1232.
47. Rubinov, B., Wagner, N., Rapaport, H., Ashkenasy, G. Self-replicating amphiphilic β -sheet peptides. *Angew. Chem. Int. Ed.* **2009**, 48, 6683–6686.
48. Carnall, J. M. A., Waudby, C. A., Belenguer, A. M., Stuart, M. C. A., Peyralans, J. J.-P., Otto, S. Mechanosensitive self-replication driven by self-organization. *Science*, **2010**, 327, 1502–1506.
49. Nowak, P., Colomb-Delsuc, M., Otto, S., Li, J. Template-triggered emergence of a self-replicator from a dynamic combinatorial library. *J. Am. Chem.*

Soc. **2015**, 137, 10965–10969.

50. Colomb-Delsuc, M., Mattia, E., Sadownik, J. W., Otto, S. Exponential self-replication enabled through a fibre elongation/breakage mechanism. *Nat. Commun.* **2015**, 6, 7427.

51. Reek, J. N. H., Otto, S. Dynamic combinatorial chemistry. (Wiley, **2010**).

52. Liu, B., Pappas, C. G., Zangrando, E., Demitri, N., Chmielewski, P. J., Otto, S. Complex molecules that fold like proteins can emerge spontaneously. *J. Am. Chem. Soc.* **2019**, 141, 1685–1689.

53. Lehn, J. M., Eliseev, A. V. Dynamic combinatorial chemistry. *Science*, **2001**, 291, 2331–2332.

54. Otto, S., Furlan, R. L., Sanders, J. K. Dynamic combinatorial chemistry. *Drug Discov. Today*, **2002**, 7, 117–125.

55. Hartman, A. M., Elgaher, W. A. M., Hertrich, N., Andrei, S. A., Ottmann, C., Hirsch, A. K. H. Discovery of small-molecule stabilizers of 14-3-3 protein-protein interactions via dynamic combinatorial chemistry. *ACS Med. Chem. Lett.* **2020**, 11, 1041–1046.

56. Ludlow, R. F., Otto, S. Systems chemistry. *Chem. Soc. Rev.* **2008**, 37, 101–108.

57. Corbett, P. T., Leclaire, J., Vial, L., West, K. R., Wietor, J.-L., Sanders, J. K. M., Otto, S. Dynamic combinatorial chemistry. *Chem. Rev.* **2006**, 106, 3652–3711.

58. Malakoutikhah, M., Peyralans, J. J. P., Colomb-Delsuc, M., Fanlo-Virgós, H., Stuart, M. C. A., Otto, S. Uncovering the selection criteria for the emergence of multi-building-block replicators from dynamic combinatorial libraries. *J. Am. Chem. Soc.* **2013**, 135, 18406–18417.

59. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. Molecular biology of the cell. (Garland Science, **2002**).

60. The Molecular Origins of Life. (Cambridge University Press, **1998**).

61. Colomer, I., Morrow, S. M., Fletcher, S. P. A transient self-assembling self-replicator. *Nat. Commun.* **2018**, 9, 1–6.
62. Morrow, S. M., Colomer, I., Fletcher, S. P. A chemically fuelled self-replicator. *Nat. Commun.* **2019**, 10, 1–9.
63. Yang, S., Schaeffer, G., Mattia, E., Markovitch, O., Liu, K., Andreas, S., Otelé, J., Sood, A., Otto, S. Chemical fueling enables molecular complexification of assembly-driven self-replicators. *Angew. Chem. Int. Ed.* **2020**, 60, 2–8.
64. Wachtershauser, G. Evolution of the first metabolic cycles. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, 87, 200–204.
65. Ruiz-Mirazo, K., Briones, C., De La Escosura, A. Prebiotic systems chemistry: new perspectives for the origins of life. *Chem. Rev.* **2014**, 114, 285–366.
66. Vasas, V., Fernando, C., Santos, M., Kauffman, S. & Szathmáry, E. Evolution before genes. *Biol. Direct*, **2012**, 7, 1–14.
67. Lancet, D., Zidovetzki, R. & Markovitch, O. Systems protobiology: origin of life in lipid catalytic networks. *J. R. Soc. Interface*, **2018**, 15, 20180159.
68. Orgel, L. E. The implausibility of metabolic cycles on the prebiotic earth. *PLoS Biol.* **2008**, 6, 0005–0013.
69. Pross, A. Causation and the origin of life. Metabolism or replication first? *Orig. Life Evol. Biosph.* **2004**, 34, 307–321.
70. Vasas, V., Szathmáry, E., Santos, M. Lack of evolvability in self-sustaining autocatalytic networks constraints metabolism-first scenarios for the origin of life. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107, 1470–5.
71. Pross, A. How can a chemical system act purposefully? Bridging between life and non-life. *J. Phys. Org. Chem.* **2008**, 21, 724–730.
72. Wagner, N., Pross, A., Tannenbaum, E. Selection advantage of metabolic over non-metabolic replicators: A kinetic analysis. *BioSystems*, **2010**, 99, 126–129.
73. Kamioka, S., Ajami, D., Rebek, J. Autocatalysis and organocatalysis with

synthetic structures. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107, 541–544.

74. Draper, W. E., Hayden, E. J., Lehman, N. Mechanisms of covalent self-assembly of the Azoarcus ribozyme from four fragment oligonucleotides. *Nucleic Acids Res.* **2008**, 36, 520–531.

75. Vaidya, N., Manapat, M. L., Chen, I. A., Xulvi-Brunet, R., Hayden, E. J. & Lehman, N. Spontaneous network formation among cooperative RNA replicators. *Nature*, **2012**, 491, 72–77.

76. Nghe, P., Griffiths, A. D., Arsène, S., Ameta, S., Lehman, N. Coupled catabolism and anabolism in autocatalytic RNA sets. *Nucleic Acids Res.* **2018**, 46, 9660–9666.

77. Biancalana, M., Koide, S. Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochim. Biophys. Acta - Proteins Proteomics*, **2010**, 1804, 1405–1412.

78. Monreal Santiago, G., Liu, K., Browne, W. R., Otto, S. Emergence of light-driven protometabolism on recruitment of a photocatalytic cofactor by a self-replicator. *Nat. Chem.* **2020**, 12, 603–607.

79. Dénès, F., Pichowicz, M., Povie, G., Renaud, P. Thiyl radicals in organic synthesis. *Chem. Rev.* **2014**, 114, 2587–2693.

80. Ottelé, J., Hussain, A. S., Mayer, C., Otto, S. Chance emergence of catalytic activity and promiscuity in a self-replicator. *Nat. Catal.* **2020**, 3, 547–553.

81. Berg, J. M., Tymoczko, J. L., Gatto, G. J., Stryder, L. Biochemistry. (W. H. Freeman, **2015**).

82. Silva, E., Lazcano, A. Membranes and prebiotic evolution: compartments, spatial isolation and the origin of life. in *Molecules in Time and Space* 13–25 (Kluwer Academic Publishers, **2006**).

83. Eigen, M., Winkler-Oswatitsch, R. Steps towards life: a perspective on evolution. (Oxford University Press, **1992**).

84. Sunami, T., Ichihashi, N., Nishikawa, T., Kazuta, Y., Yomo, T. Effect of liposome size on internal RNA replication coupled with replicase translation.

ChemBioChem, **2016**, 0871, 1282–1289.

85. Jia, T. Z., Kuruma, Y. Recent advances in origins of life research by biophysicists in Japan. *Challenges*, **2019**, 10, 28.

86. Adamala, K., Szostak, J. W. Nonenzymatic template-directed RNA synthesis inside model protocells. *Science*, **2013**, 342, 1098–1100.

87. Kun, Á., Szilágyi, A., Könnyu, B., Boza, G., Zachar, I., Szathmáry, E. The dynamics of the RNA world: insights and challenges. *Ann. N. Y. Acad. Sci.* **2015**, 1341, 75–95.

88. Bansho, Y., Ichihashi, N., Kazuta, Y., Matsuura, T., Suzuki, H., Yomo, T. Importance of parasite RNA species repression for prolonged translation-coupled RNA self-replication. *Chem. Biol.* **2012**, 19, 478–487.

89. Bansho, Y., Furubayashi, T., Ichihashi, N., Yomo, T. Host-parasite oscillation dynamics and evolution in a compartmentalized RNA replication system. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, 113, 4045–4050.

90. Matsumura, S., Kun, A., Ryckelynck, M., Coldren, F., Szilagyi, A., Jossinet, F., Rick, C., et al. Transient compartmentalization of RNA replicators prevents extinction due to parasites. *Science*, **2016**, 354, 1293–1296.

91. Szathmáry, E., Demeter, L. Group selection of early replicators and the origin of life. *J. Theor. Biol.* **1987**, 128, 463–486.

92. Grey, D., Hutson, V., Szathmary, E. A re-examination of the stochastic corrector model. *Proc. R. Soc. B Biol. Sci.* **1995**, 262, 29–35.

93. Joyce, G. F., Szostak, J. W. Protocells and RNA self-replication. *Cold Spring Harb. Perspect. Biol.* **2018**, 10,.

94. Deamer, D. The role of lipid membranes in life's origin. *Life*, **2017**, 7.

95. Koonin, E. V., Martin, W. On the origin of genomes and cells within inorganic compartments. *Trends Genet.* **2005**, 21, 647–654 .

96. Khorana, H. G. Nucleic acid synthesis. *Pure Appl. Chem.* **1968**, 17, 349–382.

97. Monnard, P. A., Ziock, H. Eutectic phase in water-ice: a self-assembled environment conducive to metal-catalyzed non-enzymatic RNA polymerization. *Chem. Biodivers.* **2008**, 5, 1521–1539.
98. Dobson, C. M., Ellison, G. B., Tuck, A. F., Vaida, V. Atmospheric aerosols as prebiotic chemical reactors. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, 97, 11864–11868.
99. Stano, P., Luisi, P. L. Achievements and open questions in the self-reproduction of vesicles and synthetic minimal cells. *Chem. Commun.* **2010**, 46, 3639–3653.
100. Schmidli, P. K., Schurtenberger, P., Luisi, P. L. Liposome-mediated enzymatic synthesis of phosphatidylcholine as an approach to self-replicating liposomes. *J. Am. Chem. Soc.* **1991**, 113, 8127–8130.
101. Walde, P. Self-reproducing vesicles. in *Self-Production of Supramolecular Structures* 209–216 (Springer Netherlands, **1994**).
102. Chen, I. A., Walde, P. From self-assembled vesicles to protocells. *Cold Spring Harb. Perspect. Biol.* **2010**, 2, 1–13.
103. Bachmann, P. A., Luisi, P. L., Lang, J. Autocatalytic self-replicating micelles as models for prebiotic structures. *Nature*, **1992**, 357, 57–59.
104. Bachmann, P. A., Walde, P., Luisi, P. L., Lang, J. Self-replicating reverse micelles and chemical autopoiesis. *J. Am. Chem. Soc.* **1990**, 112, 8200–8201.
105. Fiordemondo, D., Stano, P. Lecithin-based water-in-oil compartments as dividing bioreactors. *ChemBioChem*, **2007**, 8, 1965–1973.
106. Zhu, T. F., Szostak, J. W. Coupled growth and division of model protocell membranes. *J. Am. Chem. Soc.* **2009**, 131, 5705–5713.
107. Kurihara, K., Tamura, M., Shohda, K. I., Toyota, T., Suzuki, K., Sugawara, T. Self-reproduction of supramolecular giant vesicles combined with the amplification of encapsulated DNA. *Nat. Chem.* **2011**, 3, 775–781.
108. Matsuo, M., Kan, Y., Kurihara, K., Jimbo, T., Imai, M., Toyota, T., Hirata, Y., Suzuki, K., Sugawara, T. DNA length-dependent division of a giant vesi-

cle-based model protocell. *Sci. Rep.* **2019**, 9, 1–11.

109. Luisi, P. L., Stano, P. Minimal cell mimicry. *Nat. Chem.* **2011**, 3, 755–756.

110. Taylor, J. W., Eghesadi, S. A., Points, L. J., Liu, T., Cronin, L. Autonomous model protocell division driven by molecular replication. *Nat. Commun.* **2017**, 8, 4–9.

111. Poudyal, R. R., Pir Cakmak, F., Keating, C. D., Bevilacqua, P. C. Physical principles and extant biology reveal roles for RNA-containing membraneless compartments in origins of life chemistry. *Biochemistry*, **2018**, 57, 2509–2519.

112. Marianelli, A. M., Miller, B. M., Keating, C. D. Impact of macromolecular crowding on RNA/spermine complex coacervation and oligonucleotide compartmentalization. *Soft Matter*, **2018**, 14, 368–378.

113. Koga, S., Williams, D. S., Perriman, A. W., Mann, S. Peptide-nucleotide microdroplets as a step towards a membrane-free protocell model. *Nat. Chem.* **2011**, 3, 720–724.

114. Sokolova, E., Spruijt, E., Hansen, M. M. K., Dubuc, E., Groen, J., Chokalingam, V., Piruska, A., Heus, H. A., Huck, W. T. S. Enhanced transcription rates in membrane-free protocells formed by coacervation of cell lysate. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, 110, 11692–11697.

115. Mason, A. F., Yewdall, N. A., Welzen, P. L. W., Shao, J., Stevendaal, M. Van, Hest, J. C. M. V., Williams, D. S., Abdelmohsen, L. K. E. A. Mimicking cellular compartmentalization in a hierarchical protocell through spontaneous spatial organization. *ACS Cent. Sci.* **2019**, 5, 1360–1365.

116. Oparin, A. I. *The Origin Of Life.* (1952).

117. Mason, A. F., van Hest, J. C. M. Multifaceted cell mimicry in coacervate-based synthetic cells. *Emerg. Top. Life Sci.* **2019**, 3, 567–571.

CHAPTER 2

Mechanistic studies of supramolecular self-replication

Parts of this chapter are published: Markovitch, O., Ottel, J., Veldman, O., Otto, S. Automated device for continuous stirring while sampling in liquid chromatography systems. Commun. Chem. 3, 2020

ABSTRACT

The ability for certain molecules to be able to catalyse the formation of copies of themselves is thought to be an essential component for the emergence of life – be it either life as we know it or life de novo. Design rules for such molecules have historically been difficult to device, but nowadays scientists have produced a large variety of synthetic self-replicators. However, the exact mechanism with which these molecules are able to self-replicate often remains elusive. In this chapter we design new methods to perform in-depth mechanistic studies of a self-replicator previously developed in our group. The ability to exert strict control over various variables has allowed for hitherto unobtainable reproducible kinetic data. The development of a 3D printed stirring device capable of replacing the sample holder of liquid chromatography systems has further improved upon these methods while enabling the possibility for high-throughput kinetic experiments. Rate constants obtained can be used in the development of a computer model which should facilitate exploring new behaviour in complex mixtures of self-replicators.

2.1 INTRODUCTION

Systems capable of self-replication are of profound interest to the scientific community. Not only are they thought to be essential components in the emergence of life on earth^{1–6}, they also play vital roles in a variety of research topics such as materials science^{7–9}, molecular electronics^{10–12} and even Alzheimer’s disease^{13–15}. However, the mechanism by which self-replicators create copies of themselves is often difficult to probe experimentally. Computer models could supply us with the tools needed to investigate the inner workings of self-replication.

We have previously^{16–19} developed a range of peptide-based self-replicators that form supramolecular structures and can, through an elongation breakage mechanism¹⁷, exhibit exponential growth. Here we will focus on building block **1** that consists of a short peptide chain attached to a benzene-1,3-dithiol core moiety (see **Figure 1A**). In aqueous solution and in the presence of atmospheric oxygen, building block **1** oxidises to form a mixture of small disulfide macrocycles initially dominated by non-assembled cyclic trimers (**1**₃) and tetramers (**1**₄). After an initial lag phase that is typical of sigmoidal growth kinetics, the hexamer (**1**₆)* self-replicators emerge. This species is able to nucleate and self-assemble into supramolecular fibres. These fibres then catalyse the formation of more hexamer, leading to the elongation of the short stacks. We found that mechanical agitation plays an essential role in achieving significant formation of the self-replicating species within a reasonable timeframe – it leads to fibre breakage which doubles the amount of growing ends and pushes the formation of the hexamers into an exponential regime¹⁶ (see **Figure 1B** and **1C**). Mechanically induced fragmentation solves the common issue of product inhibition described in **Chapter 1**. The fibre ends play a crucial role in the process of hexamer formation, but the exact chemical processes taking place at the fibre ends are largely unknown.

Recently we discovered that the formation of supramolecular structures in these systems inadvertently lead to self-replicators acquiring additional functions, such as the recruitment of cofactors²⁰ or the ability to catalyse chemical reactions.²¹ Such interesting behaviour is dictated by the pentapeptide chain attached to the core moiety. When we constrain ourselves to the twenty-two proteinogenic amino acids that could be incorporated into this pentapeptide chain, there are already more than five million possible combinations one could synthesise – far too many to study experimentally, especially within the timeframe of a typical PhD considering that experiments performed with our peptide-based replicators usually take a few weeks to months.

In this largely unexplored peptide space, there is a considerable potential for new discoveries – it could harbour building blocks with peptide sequences that exhibit essential behaviour for the next step in the road towards life de novo: self-sorting and the coexistence of various replicators in a single system. Due to the sheer number of possible peptide combinations, such interesting behaviours cannot easily be tackled experimentally: a computer model that can predict

Figure 1: A) Structure of building block **1**, where the aromatic core unit is depicted in yellow and the pentapeptide in turquoise. B) A typical sigmoidal growth curve characteristic of self-replication. C) Proposed mechanism of self-replication. The yellow semicircle represents the aromatic core unit and the blue rounded rectangle symbolises the pentapeptide. Initially, building block **1** oxidises by oxygen in the atmosphere (1) and forms small disulphide macrocycles that can exchange with one another (2). After an initial lag phase, hexamers are formed that can nucleate by assembling

to form a short stack (3). These stacks elongate (4) and promote the formation of more hexamers. When applying mechanical agitation, the elongated fibres can fragment (5), doubling the number of fibre ends and resulting in exponential growth kinetics.

interesting behaviour is required to properly explore this vast peptide sequence space.

The use of a variety of building blocks in single systems further establishes the need for such a model. Mixed systems are an essential next step on the road towards *de novo* life: they can exhibit mutations by stacking mismatches, change catalytic behaviour and lead to competition between different replicators. However, a system that only encompasses *two* building blocks is already tremendously difficult to adequately describe. Imagine the two building blocks **A** and **B** that are capable of mixing with one another to form hexamer self replicators. One can already construct *thirteen* distinct hexamers from **A** and **B** (see **Figure 2**).

Figure 2: The thirteen different hexamers constructed from building blocks **A** (in turquoise) and **B** (in red).

Now, let us consider the stacking of these hexamers. Since the building blocks and thus the macrocycles are chiral, there are 13^2 ways in which a dimer of hexamers can form – if we do not take rotations into account. However, since the relative orientation of the hexamers changes the overall stacking, we must find a way to account for all possible rotations. Consider the following argument: for the first

choice of hexamer there are only thirteen possibilities (rotating them does not result in a new unique orientation). For the second hexamer in the stack, the rotational symmetry matters. For example, the three possible combinations of A_4B_2 generate fifteen possible rotational isomers (see **Figure 3**).

Figure 3: The fifteen different rotational isomers of the three A4B2 combinations.

Taking all of these isomers into account, the total number of different hexamer dimer stacks becomes:

$$13 \times (1 + 6 + (3 + 6 + 6) + (2 + 6 + 6) + (3 + 6 + 6) + 6 + 1) = 754 \quad (\text{Eq 2.1})$$

For n number of hexamers in a stack, the total number of possibilities can be described as follows:

$$13 \times (1 + 6 + (3 + 6 + 6) + (2 + 6 + 6) + (3 + 6 + 6) + 6 + 1)^{n-1} = 13 \times 58^{n-1} \quad (\text{Eq 2.2})$$

Under normal circumstances (see Methods: *preparation of libraries*) the fibres in our systems can grow to lengths in the order of hundreds of nanometers to a few micrometers. Since one hexamer constructed from building block **1** is about half a nanometer in height,²² the number of different possibilities in a stack is comparable to the number of atoms in the observable universe when a fibre has grown to about 23 nm (~46 hexamer macrocycles). It is clear we cannot investigate the stacking behaviour experimentally.

Thus, the reasons for constructing a computer model to describe the mechanism of replication in our systems are twofold:

To obtain more insight in the actual chemistries involved in the mechanism of self-replication.

To explore peptide space in order to discover interesting behaviour.

Both the large number of stacking possibilities and the variety in peptide sequence introduce a large amount of *randomness* to dynamic combinatorial libraries. Deterministic models would be prohibitively complicated and computationally demanding – a stochastic model based on the probabilities of interactions between molecules could describe these systems with inherent randomness more appropriately, cheaper and faster.

Stochastic models should stay as close as possible to the experimental system in order to obtain insight into the latter. The construction of such a model requires the determination of as many experimentally accessible parameters as possible. Parameters that are problematic to capture experimentally can then subsequently be resolved by fitting them to the model.

Various parameters can immediately be identified when considering **Figure 1**. The thiol oxidation of the building blocks to form disulfide bonds (**1**), the exchange reactions between free thiolate anions and disulfide bonds (**2**), nucleation events that lead to hexamer dimers and the formation of a short fibre (**3**), the elongation (**4**) and the subsequent fragmentation (**5**) of such a fibre. Previous work has focused on the rates of fibre elongation¹⁷ and the effect of stirring rates or other means of agitation on replication and fragmentation.¹⁸ However, at the times these experiments were performed no temperature control was used (subjecting the libraries temperature changes at night,

during the weekends and with the change of seasons) nor was there any control over the amount of fibre ends present. This resulted in a large variability in elongation rates and less than optimal linear regression fits. To obtain as many data points as possible during these experiments, libraries were acidified using trifluoroacetic acid (TFA) to reduce the speed of thiol exchange mechanisms (which require free thiolate anions). This required extensive pipetting and introduced the possibility of human error. In this chapter we aim to minimize the introduction of errors by performing strictly controlled kinetic experiments. For this purpose, a 3D printed stirring device has been designed and constructed that is able to operate inside the sample holder of an Ultra Performance Liquid Chromatography (UPLC) system, eliminating most human interactions with the libraries. We further investigate the rates associated with the mechanism of self-replication for the purpose of an eventual implementation in a computer model that can give us mechanistic insight and can be used as a tool to discover interesting behaviour in peptide space.

2.2 RESULTS AND DISCUSSION

2.2.1 DESIGN AND DEVELOPMENT OF A 3D PRINTED STIRRING DEVICE

Chromatography is at the heart of analytical chemistry. Nowadays, automated apparatuses have been developed to make the detection and analysis of complex mixtures possible. UPLC in particular is an important tool to distinguish different species in complex mixtures,²³ making it particularly suitable for the analysis of dynamic combinatorial libraries. In brief, small aliquots of solutions are injected from the sample holder and separated with the use of a column and eluents. A column typically interacts with the sample so that compounds can be separated based on their physical properties such as ionic strength, hydrophobicity or size.

The stirring of samples is often a crucial part of an experiment as it promotes homogeneity and can even be essential for product formation.¹⁷ To automate sampling and analysis even further, we have developed a UPLC stirring device that can replace a standard sample holder and enables battery-powered magnetic stirring. Any experimental setup that requires continuous stirring can now be run inside a UPLC system, allowing for multiple measurements at any time

without the need for sample preparation and human intervention. This eliminates human errors, makes measurements more reproducible and gives access to standard deviations (quantifying variability in identical experiments).

To obtain maximum versatility in design and manufacturing, as well as to allow easy user-specific modifications, we used 3D printing to develop the stirring device. User requests and stirring readouts are handled through an Arduino Pro Mini 3.3V 8Mhz microcontroller. A standard motor is used for the rotation of a plate that contains magnets. Control over the exact rotational speed is accomplished through the use of a magnetic sensor and can be varied between 100 and 1200 rpm with a step size of 100 rpm. The control module that contains the Arduino and batteries is separated from the stirring module (see **Figure 4**). Battery levels and rotational speed can be observed through an OLED screen on the control module.

We tested this device by monitoring the evolution of a library constructed from building block **1**. Since mechanical agitation is essential for the exponential growth of the hexamer replicator, it is an ideal system to determine whether or not the device can produce reproducible kinetic data. At room temperature the rate of replication is rather slow – it can take up to a few weeks for significant replicator formation. To solve the inconvenience of continuous measurements for multiple weeks, we performed our experiments at elevated temperatures (40 °C), where it should only take a few days to convert most of the material to replicator.

Four libraries were monitored over the course of 110 hours where two samples contained Teflon coated stirring bars (see **Figure 5A** and **Supplementary Figure 1A**) and two did not (see **Figure 5B** and **Supplementary Figure 1B**). Each sample was injected consecutively and column washes were performed when necessary. All species were identified using previously performed mass spectrometry measurements and peak retention times (see **Supplementary Figure 2**).

The use of the device has resulted in reproducible data with many data points. A clear difference is observed between stirred and unstirred samples: at around 60 hours roughly 50% of material is converted to the replicator in the stirred sample, where in the unstirred sample <5% of replicator is formed.

To test the device even further, we looked at the influence of rotational speed of the magnets on the replication rate. Previous work²⁴ has already investigated such effects and had found that a larger

Figure 4: Overview of the 3D printed stirring device. The electric motor takes up the space of 16 sample holder positions, leaving 32 positions still accessible for experiments.

amount of mechanical agitation (i.e. faster stirring rate) leads to an increase in replication rate. Since the experiments performed with the 3D printed stirring device occur at elevated temperatures (due to time constraints), the rate of replication becomes limited by the availability of oxidized monomers and by extension the speed of oxidation. To circumvent this problem, pre-oxidized libraries are used to investigate the effect of the stirring rate on replication rate (see **Figure 6**).

Indeed, samples stirred at 1000 rpm show faster growth than those stirred at 200 rpm, in line with previous observations (albeit at higher temperature). Unstirred samples containing the same pre-oxidized building blocks show no significant formation of replicating species (see **Supplementary Figure 3A**).

Finally, to ensure that the sampling quality remains consistent throughout the experiment we quantify the total amount of peak area for each measurement (see **Supplementary Figure 3B**).

Figure 5: Library progression when A) stirred at 1000 RPM and when B) unstirred. Graphs are plotted based on two repeats each. Standard deviations are shown as shaded areas between the dotted lines. All libraries were made from of 500 μ M building block **1** in borate buffer (50 mM, pH 8.12) and were kept at 40° C during the experiment.

Figure 6: Relative amount (in percentage) of A) monomers and hexamers and B) of trimers and tetramers when stirred at 200 and 1000 rpm. At 1000 rpm, library composition reaches 50% hexamer at 8.2 ± 0.9 hours while at 200 rpm this takes 17.2 ± 2.2 hours. Graphs are plotted based on four repeats. Standard deviations are shown as shaded areas between the dotted lines. All libraries were made from of 500 μM building block **1** in borate buffer (50 mM, pH 8.12) and were pre-oxidized by the addition of 0.5 equivalents of sodium perborate. All libraries were kept at 40° C during the experiment.

The total peak area appears to be well preserved with only a standard deviation of <2 %. The visual inspection of positions A4, A5, C2, D2, C7, D7, F4 and F5 (see **Figure 4**) confirmed that there was equal rotation of the stirring bar, while positions in columns one and eight exhibited less uniform stirring behaviour as they were further away from the magnetic field. The battery level is typically reduced to 35 % after approximately 72 hours of continuous stirring at 1000 rpm.

The development and construction of this device has allowed us to perform experiments otherwise inaccessible (due to time-constraints or human interference) inside the sample holder of a UPLC system. This allows for further automation of experiments and high frequency and reproducibility of UPLC injections. This device has been used in the investigation of various parts of the mechanism of self-replication of systems comprised of building block **1**.

2.2.2 THE KINETICS OF SELF-REPLICATION

OXIDATION OF THIOLS

The first reaction we encounter when looking at the proposed mechanism of self-replication is thiol oxidation that results in disulfide formation (see **Scheme 1**). To properly construct the rate laws that describe this part of the mechanism we must identify the reaction order with which these reactions occur in the presence of atmospheric oxygen.

Scheme 1: The oxidation of building block **1** to form a linear dimer.

The rate of a chemical reaction is generally defined as the ratio between the change in concentration of a substrate or reactant and the change in time. A *rate law* describes how the concentration of a reactant influences the rate of a chemical reaction.

For a two-reactant reaction, a general rate law can be written as:

$$Rate = k \times [S_1]^x \times [S_2]^y \quad (\text{Eq 2.3})$$

Where k is the rate constant, S_1 and S_2 are the substrate concentrations and x and y the exponents that determine the influence of the concentration of these substrates. For thiol oxidation, the rate law can be written as:

$$Rate = k \times [SH]^x \times [O_2]^y \quad (\text{Eq 2.4})$$

In order to investigate the order of the thiols in this reaction (i.e. to determine exponent x) the influence of the concentration of free thiols on the reaction rate must be determined. We can do so by utilising the newly developed 3D printed stirring device and measuring the amount of free thiol present as a function of time. However, there are many variables involved in these complex mixtures which require us to make the following assumptions:

1: The extinction coefficients of all macrocycles involved in the mechanism are a scalar multiple of the extinction coefficient of building block 1.

The extinction coefficients are determined by multiplying the amount of building block **1** in any given macrocycle by the extinction coefficient of building block **1**, e.g. a tetramer contains four monomers thus has four times the extinction coefficient of the monomer. The libraries are monitored with a UPLC equipped with PDA detector. More specifically, a wavelength of 254 nm is used as it has been found that there the monomers and hexamers have the same extinction coefficient.²⁵ Indeed, when performing UPLC analysis on our libraries we observe that the total peak area remains – more or less – the same, independent on library composition (see **Supplementary Figure 3B**).

2: The amount of oxygen in the libraries remains constant during the experiments.

Since the reaction vessels are not airtight and the mixtures are stirred, we can safely assume the concentration of oxygen ranges between 250 and 500 μM during the course of an experiment.^{26,27}

3: The amount of free thiol is described by:

$$[SH] = 2 \times ([monomer] + [dimer]) \quad (\text{Eq 2.5})$$

Other linear species that could be present (linear trimer, linear tetramer) were not detected when performing these experiments. This does not mean they are absent, nor that we believe them not to be formed: we strongly believe that those species play an important role in the mechanism of self-replication. However, to have avoided detection they must be present at such low concentrations that we can assume the total concentration of free thiols is not heavily influenced by their omission from Eq 2.5.

4: Measurements are taken in a regime where there is no influence on the mechanism of oxidation by the replicating species.

We have found a few examples where it seems that the replicating species speeds up the oxidation of its building blocks. Even though those effects are still under scrutiny and we have found no conclusive evidence of this behaviour in libraries constructed from building block **1**, we measure the oxidation of free thiols without a significant presence of the replicating species (<2%) to minimize any possible interactions.

We set up four different libraries with concentrations of either 250 μM or 500 μM in building block **1** (duplicates), stirred at 1000 rpm and kept at 40 °C. We monitored library progression for 78 hours and used the first six hours to determine the initial rates of thiol oxidation (see **Figure 7**).

When the concentration of thiols is doubled (**Figure 7C** vs **Figure 7D**), we see that the rate of thiol oxidation increases with a factor of $1,81 \pm 0,28$. It follows that the exponent describing the influence of the concentration of thiols in the rate law is $0,86 \pm 0,22$ (by solving $2^x=1,81$). As will become clear in **Chapter 4**, we believe thiol oxidation to occur through a one-electron radical oxidation mechanism. It has

Figure 7: Rates of thiol oxidation. A) Library composition when $[1] = 250 \mu\text{M}$ and B) when $[1] = 500 \mu\text{M}$. C) Initial rate of thiol oxidation when $[1] = 250 \mu\text{M}$ and D) when $[1] = 500 \mu\text{M}$. Rates of thiol oxidation. Two repeats were performed for both 250 μM and 500 μM in building block **1**. Standard deviations are shown in shaded areas between the dotted lines. Linear regression fits are shown by black lines and the black dots represent the standard deviation of those fits.

been previously shown^{28,29} that such mechanisms are described by first order kinetics, which is in line with our observations. When fitting the experimental data to a first order rate law, we find an apparent rate constant for thiol oxidation of:

$$k = 5,5 \times 10^{-6} \pm 6,0 \times 10^{-7} \text{ M s}^{-1}.$$

EXCHANGE REACTIONS

Yang et al. recently²⁵ investigated the rate of thiol-disulfide exchange using a stopped-flow setup with a UV detector. These reactions were found to be overall second order; first order in thiolate anion and in disulfide.³⁰ The two monothiol model compounds **2** and **3** were used

(see **Figure 8A**) to prepare two samples: one that contains only the thiols of building block **2** and one that contains only the disulfides of building block **3**.

The initial kinetics of the reaction shown in **Figure 8B** were studied so that there is no significant influence of other (subsequent) exchange reactions. The extinction coefficient of compounds **2**, **3**, **3₂** and **2-3** were determined by UV or UPLC which allowed for rate constant determination by measuring the change in compound **3** over time and fitting the results to a second order rate law:

$$k = 1.08 \times 10^4 \pm 1 \times 10^2 \text{ M s}^{-1}.$$

Figure 8: Building blocks and reaction used to measure the rate of disulfide exchange. A) Chemical structure of building blocks **2** and **3** and their schematic representations. B) Disulfide exchange reaction monitored in a stopped-flow setup.

ELONGATION

The rate at which self-replicating fibres elongate has previously been probed by performing a variety of seeding experiments. In such an experiment, a certain amount of pre-formed fibres are added to a library that consists of – predominantly – monomers, trimers and tetramers. Other species might be present at low concentrations (<10%) but no significant amount of replicating species should be present at the start of the experiment. Usually these “young” libraries are prepared by the rapid oxidation of monomers using an oxidant such as sodium perborate or hydrogen peroxide. After the rapid oxidation pre-formed fibres are added to the mixture and the library is subsequently monitored through periodic UPLC analysis. Small aliquots of the library are diluted with doubly distilled water and injected in the UPLC roughly every fifteen to eighteen minutes (depending on the measurement time of an experiment).

In order to acquire data points at a higher resolution it would be convenient to halt the process of self-replication. Since thiol exchange is an integral part of the mechanism without which no hexamer can be formed, slowing down or even halting the exchange should lead to a library of which the composition remains – more or less – the same. In practice, this means taking an aliquot of the library and diluting it with doubly distilled water that contains TFA to acidify the mixture. Since the rate of thiol-disulfide exchange depends on the amount of free thiolate anions present, lowering the pH should significantly slow down the rate at which hexamers can be formed. Using this technique to “freeze” the dynamic combinatorial libraries, one can take aliquots at any time and determine their composition at leisure.

There is little information about the pre-formed fibres with which seeding experiments are performed. For the sake of properly calculating the rate of elongation, the concentration of fibre ends needs to be determined. The fibres grow from their ends, so increasing the number of fibre ends increases the rate of hexamer production. The amount of fibre ends can be estimated if the average length of fibres and the height of a hexamer macrocycle are known. When libraries are stirred using conventional magnetic stirrers and Teflon coated stirring bars, the fibres can grow as long as a few hundred nanometres to a few micrometres. Such fibres are typically not completely straight and their exact length – and by extension the amount of fibre ends – are difficult to probe (see **Figure 9A**).

Figure 9: TEM micrographs of fibres consisting of **1**₆. A) Under standard conditions (25 °C, 1200 rpm for three weeks) long hexamer fibres are observed. B) Fibres exposed to high shear stress fragment into short stacks with a narrow length distribution.

A device capable of exerting variable shear stress to liquid samples has previously been developed in our group²⁴. This device, known as a Couette cell (named after the physicist Maurice Couette) consists of a spinning cylinder in a fixed cuvette separated by just 250 μm (see **Figure 10A**). The cylinder is connected to a motor using a small drive belt, and both cylinder and cuvette are made from titanium which was found to have negligible influence on library progression. The Couette cell is controlled by a computer that provides real time data of the measured versus the set rotational speed. However, the exact amount of shear stress generated during operation cannot be quantified; the rotational speed is such that the flow inside the Couette cell is turbulent. Because of this, the rotational frequency is used as an approximation of shear stress instead. Nonetheless, this device has allowed us to obtain short fibre fragments with a narrow length distribution (see **Figure 9B**).³¹ We performed extensive analysis of TEM micrographs to determine the length distribution and average length of the fibres constructed from building block **1** after being subjected to 66.7 Hz (or 4000 rpm) for thirty minutes. They were found to have an average length of 34.8 ± 15.2 nm based on the measurement of 994 fibres in nine TEM micrographs (see **Figure 10B** and **Supplementary Figure 4**)

The height of a single hexamer macrocycle ($C_a - C_a$) of building block **1** has been estimated to be 0.48 ± 0.1 nm from molecular dynamics (MD) simulations.²² If we look at the short fibre stacks generated with

Figure 10: Design of the Couette cell and resulting length distribution at 66.7 Hz. A) Top, side and isometric view of the Couette cell. The most essential components are indicated with various colours. B) Length distribution of fibres consisting of **1**, after being exposed to 66.7 Hz in the Couette cell for 30 minutes (average length of 34.8 ± 15.2 nm). Nine TEM micrographs were analysed using ImageJ software.

the Couette cell, we find that such fibres generally consist of 70 ± 20 hexamer macrocycles. Since the concentration of hexamer in a sample of pre-formed fibres is known, we can determine how many fibre ends are introduced when performing a seeding experiment.

Initially, short fibre stacks were prepared as described above and stored in a $-12\text{ }^{\circ}\text{C}$ freezer. We hoped that this approach would allow for the same pre-formed fibres to be used in each elongation experiment, reducing the potential variability introduced by the Couette cell's turbulent flow.

When performing elongation experiments using thawed fibre stacks, we observed highly irreproducible data and found no significant improvement compared to previously performed measurements (see **Figure 11A**). Remarkably, when the short fibre stacks were *not* frozen but were freshly prepared before each elongation experiment, we obtained very reproducible data (see **Figure 11B**). This led us to exclude the use of frozen samples, and to prepare freshly sheared short fibres before every seeding experiment.

The discrepancy between frozen and freshly prepared sheared fibre stacks can be explained by a large variability in the amount of introduced fibre ends during the seeding step. The slow freezing of samples leads to a high local concentration of solutes in areas where there is still liquid water (see **Figure 12B**).³² At these local high concentrations the fibre ends can then quickly recombine to form longer fibres, eliminating the narrow length distribution and control over the amount of fibre ends (see **Figure 12C**).

On the other hand, there is excellent control over fibre length in samples freshly prepared with the Couette cell. From the four repeats shown in **Figure 11B** we were able to determine that the initial growth rate (using the first three datapoints) is $2.7 \pm 0.7\text{ nm (fibre end)}^{-1}\text{ min}^{-1}$ at a concentration of $154\text{ }\mu\text{M}$ in building block **1**. It should be noted that the rate of elongation depends strongly on the availability of precursor material. Indeed, the oxidation of thiols often seem to be the rate limiting step during self-replication (see **Figure 5A**).

The “freezing” of libraries using TFA introduced a large experimental error (see **Supplementary Figure 5A**) and should therefore be rejected as an approach to obtain a higher resolution of data (i.e. more data points). Instead, based on the present work two alternative methods are suggested: the use of lower concentrations or lower temperature. Since library progression has historically been monitored on the lab benches – subject to substantial temperature changes as night falls,

Figure 11: Hexamer formation when pre formed 1_6 fibres were added to pre-oxidized libraries to ensure the availability of enough precursor material needed for elongation. All libraries were prepared in borate buffer (50 mM, pH 8.12). Short fibre stacks were generated by exposing samples (1.54 mM in building block **1**) consisting of >95% 1_6 to 66.7 Hz (or 4000 RPM) in a Couette cell for thirty minutes. Pre-oxidized libraries were prepared by the addition of 0.75 equivalents of sodium perborate to a solution containing 1.54 mM building block **1** and contained mostly 1_3 and 1_4 (denoted by “oxidized monomers”). A) Fibres of 1_6 were frozen and stored before use or B) used without freezing.

during the weekends when there is no heating and with the change of the seasons, we studied the effects of temperature on elongation rate (see **Supplementary Figure 5B**). There is an order of magnitude difference between the rates of elongation at 10 °C and 35 °C showing yet again the importance of temperature-controlled experiments, especially when studying reaction kinetics.

Figure 12: Recombination of short fibres through the formation of high local concentrations of fibre ends. A) Distribution obtained after exposing fibres to high shear stress. B) As ice crystals are formed, fibres are concentrating in water that is still liquid. C) In these areas the fibres recombine, resulting in a loss of control over the amount of fibre ends. D) TEM micrograph of sheared fibres before and E) after freezing and thawing.

2.2.3 PRELIMINARY RESULTS

Two important elements of the proposed mechanism of self-replication have not been discussed yet: nucleation and fragmentation. Chronologically, experiments related to these aspects have been per-

formed before the development of the 3D printed stirring device and have not been repeated yet. Even though new methods have been developed to study these features in more detail, the preliminary results obtained in this work shed light on the manner in which fibres are built and broken and will be discussed here. It should be noted that these experiments have fewer repeats than that I am comfortable with (<3) so their results should not be taken at face value – the preliminary results rather point to areas which show interesting behaviour and warrant further investigation.

NUCLEATION

The formation of a fibre requires an occurrence in which the first few macrocycles of the replicating species stack on top of one another: a nucleation event. Here, the foundations of the supramolecular structure are built – a critical and fragile step in the mechanism of self-replication. Since the binding energies of the macrocycles are cumulative (strong cooperative binding), the initial incentive of stacking through β -sheet formation and π - π stacking are less pronounced than they would be for a longer stack of macrocycles.²² This means that initially, macrocycles are more likely to disassociate from a short stack than they would from a long fibre, further increasing the duration of a lag phase in self-replication. To investigate at what point a short stack of macrocycles transitions into a fibre, i.e. when the rate of dissociation becomes very small, experimental data was fitted to a computer model.

MODEL FOR NUCLEATION

To investigate the nature of the early nucleation stages, a model was constructed that solely focusses on the self-assembly part of the mechanism. The model assumes that initially, all of building block **1** present consists of homogeneously dissolved non-aggregated hexamers and proceeds to find a stacking configuration that provides the best fit to the observed hexamer formation during the time course of an experiment. The model starts by taking all of the freely diffusing hexamers and tries to incorporate them into stacks of various lengths. In the most simplified model, freely diffusing hexamers (δ_{free}) are assumed to spontaneously assemble into a fibre directly (δ_{fibre} , see **Scheme 2**) which, in turn, are autocatalytic (the elongation-break-

age mechanism is implicit in the model). Alternatively, in the more detailed models there could be intermediate stages through the formation of nuclei of various sizes (see **Scheme 3**).

Scheme 2: Freely diffusing hexamers spontaneously assemble into supramolecular fibrils.

Scheme 3: Freely diffusing hexamers go through intermediate stages of fibre formation, each stage becoming more thermodynamically favourable. The number of nucleation stages is given by $n-1$, where n is the total of stacked hexamers.

An experimental setup was used in which the composition of a library (154 μM building block **1**, see Methods: *Nucleation*) was monitored every 18 minutes by manual injection (see **Figure 13A**). As this study is focused on the nucleation part, only the first eight hours of the experiment were monitored (13 % conversion to hexamer). Models containing various nucleation stages (zero to seven) were fitted to the experimental data using standard data fitting techniques (see **Figure 13C**, **13D** and Methods). A nucleation stage is an intermediate species that consists of various hexamer macrocycles that have stacked on top of one another (see **Scheme 3**). During these nucleation stages hexamer macrocycles can readily detach from the stacks – the backward reaction is still significant. The data fitting for each model was done by comparing the number of observed hexamers to the total amount of assembled hexamers. For example, if the model only had one nucleation stage, then the fit considered both 6_2 and 6_{fibre} and compared the sum of the two to the number of observed hexamers, where a smaller difference between experiment and model signifies a better fit. We found that a model with either two or four intermediate nucleation stages (see **Figure 13B**) shows the closest fit to

the experimental data since these have the lowest root mean square difference (RMSD) values. The values of the rate constants obtained through data-fitting are shown in **Table 1**.

Figure 13: Experimental and computational data related to the nucleation mechanism. A) Experimental library composition. Concentrations of the two constituents are plotted on left and right y-axes. B) Calculated RMSD values for various nucleation stages (n) shown in C) and D). The RMSD values represent how well a model with a certain amount of nucleation stages (n) fits the experimental detail. Generally speaking: the lower the RMSD, the better the fit. C) and D) show the hexamer concentration predicted by models with various nucleation stages.

VALIDATION OF THE MODEL AND RECOGNITION OF LIMITATIONS

This model is based on the assumption that all building block **1** is present as freely diffusing hexamers at the beginning of the experiment. We know, of course, that this is not the case, and that a large variety of macrocycles are present throughout the experiment. Furthermore, data fitting was only performed on a single experiment – something that should be repeated for many different experiments.

The algorithm that determines the values for the various parameters in the models has only been run once. This means that the parameter values found for the closest fit to the experimental data could be a local minimum rather than an absolute minimum. To obtain more certainty, such an algorithm should be run many times with different initial guesses for the parameters.

Finally, as more nucleation stages are introduced, the number of parameters increases. More parameters generally facilitate data fitting. To illustrate that fact, Johnny von Neumann famously stated: “With four parameters I can fit an elephant, and with five I can make him wiggle his trunk”.³³ Turns out that you actually can, if these parameters are complex numbers.³⁴

In light of the previous, we utilize the Akaike information criterion (AIC).³⁵ This relates the goodness of fit – in our case the RMSD – and the number of estimated parameters in the following way:

$$AIC = 2k - 2\ln(GoF) \quad (\text{Eq 2.6})$$

Here, k is the number of predictable parameters in the model and GoF is the goodness of fit. The AIC allows us to compare models to one another – on the one hand the AIC value is lowered by the GoF while on the other hand the AIC is increased by the number of parameters used. This allows us to decrease the bias obtained from having more predictable parameters in a model. The AIC steadily increases as the number of parameters are increased – we find that for our models, the number of parameters has a much stronger influence than the obtained RMSD values (see **Supplementary Table 1**). From a visual inspection it is clear that models for two and four nucleation stages show the best fit. The AIC predicts that of these two models, the one with two nucleation stages is preferable as it contains less predictable parameters. It should be noted that models that contain more parameters do not necessarily give the best fit – further confirming the need for algorithm optimisation.

It is clear that these preliminary results cannot be taken at face value. More experiments with higher resolution of data points are required, assumptions must be reassessed and the data fitting must be optimized. The observed trend shows that the presence of intermediate nucleation stages is likely – but needs further confirmation.

	NUCLEATION STAGES							
Rate constants	0	1	2	3	4	5	6	7
k_{auto} ($\mu\text{M min}^{-1}$)	3,48E-07	1,12E-02	7,56E-03	5,92E-03	8,94E-03	6,16E-03	5,47E-03	5,40E-03
k_1 ($\mu\text{M min}^{-1}$)	3,70E-05	3,50E-04	3,03E-04	3,47E-04	3,36E-04	3,51E-04	3,51E-04	3,51E-04
k_{-1} (min^{-1})	9,22E-02	1,32E-02	1,91E-02	1,71E-02	1,56E-02	2,44E-02	1,24E-02	2,73E-03
k_2 ($\mu\text{M min}^{-1}$)		8,30E-05	9,42E-04	2,79E-04	4,29E-04	6,04E-04	3,98E-04	3,99E-04
k_{-2} (min^{-1})		6,49E-03	1,23E-01	1,71E-03	7,66E-03	1,02E-02	5,88E-03	4,15E-04
k_3 ($\mu\text{M min}^{-1}$)			2,23E-04	1,10E-03	7,27E-05	2,06E-03	2,36E-03	1,28E-03
k_{-3} (min^{-1})			3,72E-06	5,87E-03	5,07E-03	2,86E-02	2,24E-03	1,91E-04
k_4 ($\mu\text{M min}^{-1}$)				1,08E-03	6,42E-04	2,36E-03	2,41E-03	1,71E-03
k_{-4} (min^{-1})				6,95E-05	1,31E-05	4,93E-03	1,98E-06	1,27E-04
k_5 ($\mu\text{M min}^{-1}$)					4,29E-01	1,81E-03	9,61E-04	1,71E-03
k_{-5} (min^{-1})					9,82E-04	7,06E-04	1,94E-03	7,14E-05
k_6 ($\mu\text{M min}^{-1}$)						3,46E-03	1,19E-02	3,13E-03
k_{-6} (min^{-1})						6,49E-07	2,12E-03	7,69E-05
k_7 ($\mu\text{M min}^{-1}$)							1,14E-03	1,45E-03
k_{-7} (min^{-1})							6,49E-07	1,02E-04
k_8 ($\mu\text{M min}^{-1}$)								1,91E-03
k_{-8} (min^{-1})								7,30E-05

Table 1: Rate constants in models with various nucleation stages obtained by data fitting.

FRAGMENTATION

The rate at which fibres fragment influences the amount of fibre ends at any given time. The rotational speed of the magnetic stirring bars impacts the shear stress applied to the fibres – the effect of which is largely unknown. Previous experiments¹⁷ have shown that the rotational speed influences the time at which the self-replicator enters the exponential growth regime. We envisioned a deeper insight into the effects of fragmentation could be obtained by performing seeding experiments where the rotational speed is varied in various experiments, resulting in the observation of hexamer formation as a function of rotational speed.

Libraries were constructed from building block **1**, oxidized and seeded with short fibre stacks (prepared as described in *section 2.2.2 Elongation*). Subsequently, these libraries were subjected to magnetic stirring at 0, 300, 600, 900 and 1200 rpm. Library composition was monitored every 18 minutes by UPLC injection (see **Figure 14A** and **Supplementary Figure 6**).

In the analysis of the effects of stirring rate on hexamer formation, we assume that each fibre end grows with the same speed. The initial rate of the unagitated sample (0 rpm) is used to estimate the rate of hexamer addition per fibre end. Since the amount of fibre ends can be calculated from a combination of TEM observations, MD simulations and known concentrations, we can estimate the rate of hexamer formation per fibre end. Then, based on the initial number of short stacks with which the libraries were seeded (estimated using the hexamer concentration and the average stack length in the seed) and using the initial rate of the unagitated sample, we can calculate a theoretical number of hexamers that should be produced if all samples were unagitated (see **Figure 14B** and Methods: *Fragmentation calculation* for how this is calculated). The difference in concentration of **1₆** between the observed and theoretical values (see **Figure 14C**) can then be attributed to the formation of additional fibre ends (see **Figure 14D**). These are calculated by dividing the excess production of hexamer by the rate of hexamer formation of the unagitated sample. We see that indeed, samples that have a higher stirring rate produce more hexamer than samples with a lower stirring rate.

Figure 14: Effects of stirring rate on hexamer production in seeded libraries. All samples have a concentration of 154 μM in building block 1. A) Experimental hexamer formation. B) Hexamer formation based on the initial number of short stacks with which the samples were seeded and using the initial rate of hexamer production in the unagitated (0 rpm) sample. C) Difference between experimental A) and theoretical B) values. D) Fibre ends required to explain the difference shown in C).

These results can then be extended to determine how many additional fibre ends must be produced every 18 minutes (see **Figure 15A**). Comparing these results to the total number of fibres, we can estimate the fraction of fibres breaking between injections (see **Figure 15B**). There is a high probability of breakage during the initial phase of the experiment which decreases as time progresses. Since we generally observe a fast initial growth rate during elongation measurements (see *section 2.2.2 Elongation*), such behaviour can be explained by a depletion of precursor material at later stages of the experiment.

Even though these results should not be taken at face value since it is based on only a single experiment, the trend is clear: higher mechanical agitation leads to a higher fragmentation rate and thus a

faster production of fibre ends, increasing the rate at which hexamers are produced. It would be interesting to see what the minimal length of a fibre should be before it is capable of fragmenting due to the relatively low shear stress generated by magnetic stirring bars. Combining this information with additional, strictly controlled experiments could lead to control over the fragmentation rate of these supramolecular structures.

Figure 15: Generated fibre ends and fraction of fibres that fragment during the experiment. A) Additional fibre ends generated between each UPLC injection (every 18 minutes) required to account for the additional hexamers produced during that period. B) The fraction of fibres that must break to obtain the required amount of fibre ends between each UPLC injection. Samples subjected to higher mechanical agitation show a larger amount of produced fibre ends, indicating that a larger fraction of fibres break during the experiment.

2.3 CONCLUSIONS AND OUTLOOK

In this chapter we have performed in-depth studies on microscopic mechanisms and have developed improved methods to study the kinetics of self-replication. Strict control over fibre ends and temperature while minimising human error has led to highly reproducible kinetic data. The design and development of a 3D printed stirring device has further improved on these parameters while also enabling a method of high-throughput kinetic experiments. We confirmed that thiol oxidation in these systems is first order in thiol which – at high building block concentrations – might be limited by the amount of oxygen present in solution. Through TEM analysis of sheared fibres produced by the Couette cell we found narrow length distributions which allowed us to calculate the elongation rate. As will be discussed in

the next chapter, these rates are strongly influenced by the concentration of precursor macrocycles (dominated by $\mathbf{1}_3$ and $\mathbf{1}_4$) – concentrations that we have tried to keep as consistent as possible. Fitting experimental data to a computer model that contains various stages of nucleation has shed some light on initial fibre formation. Current evidence suggests that there are various nucleation stages involved before the dissociation reaction of hexamers becomes insignificant. Optimisation of the model and its validation by fitting more experimental data could quantify the actual number of nucleation stages. We have also been able to confirm that the fragmentation rate of elongating fibres is influenced by the rotational speed of magnetic stirring bars. By comparing experimental and theoretical concentrations of $\mathbf{1}_6$, we could estimate the number of fibre ends generated to account for additional production of $\mathbf{1}_6$. Even though the various rates obtained are useful for the construction of a computer model, we still need to obtain more insight in the mechanism of self-replication. For this, we are currently investigating methods to identify all possible macrocyclic species present in a dynamic combinatorial library – even those that might be present in very small amounts.

The use of the 3D printed stirring device allows for the generation of high-resolution data and the automation of experiments that would otherwise be unfeasible or even impossible. With a newly obtained control over various essential variables that strongly influence the rates involved in the mechanism of self-replication, we now have the tools in hand to perform kinetic studies in a large variety of single and multi-building block replicators.

METHODS

General procedures

All reagents, solvents and buffer salts were obtained from commercial suppliers and used without further purification. Building block **1** was obtained from Cambridge Peptides Ltd (Birmingham, UK). UPLC analysis was performed on a Waters Acquity UPLC H-class system equipped with a PDA detector. Analysis were performed on reversed-phase UPLC columns: either an Aeris Peptide 1.7 μm XB-C18 x 2.10 mm column (Phenomenex) or an Acquity UPLC peptide 1.7 μm CSH C18 x 2.10 mm column (Waters). Column temperature was set at 35 °C and the sample holder was kept at 25 °C unless otherwise noted. The UV absorbance was monitored at 254 nm. UPLC-MS experiments were performed on a Waters Acquity UPLC H-class system coupled to a Waters Xevo-G2 TOF. The mass spectrometer was operated in positive electrospray ionization mode. Capillary, sampling cone and extraction cone voltages were kept at 2.5 kV, 30V and 4V respectively. Source and desolvation temperatures were set at 140 °C and 500 °C. Nitrogen was used as both cone (5 L/h) and desolvation gas (500 L/h).

Negative staining transmission electron microscopy

Transmission electron microscopy was performed on a Philips CM120 electron microscope operating at 120 kV. Images were recorded on a slow scan CCD camera (Gatan). Micrographs were analyzed using ImageJ software. Samples were prepared by depositing a small aliquot (5 μL) on a 400 mesh copper grid covered with a thin carbon film (van Loenen instruments). After one minute, the grid was blotted on filter paper. The sample was subsequently stained twice with uranyl acetate (2%) by depositing 5 μL of a saturated uranyl acetate solution on the grid and blotting it on filter paper after incubation for 30 seconds.

Samples used to measure the effect on a freeze-thaw cycle on fibre recombination were observed on a FEI Tecnai T20 electron microscope operated at 200 keV. Sample preparation was performed as described above.

UPLC separation

The gradient used for the UPLC separation of dynamic combinatorial libraries is shown below. The same gradient was used for both columns (Phenomenex and Waters). A = UPLC grade H₂O containing 0.1 v% TFA, B = UPLC grade MeCN containing 0.1 v% TFA. The flow rate was kept at 0.300 mL/minute throughout the analysis.

Borate buffer preparation

Concentrations of borate buffers are reported in concentration of

TIME (MINUTES)	%A	%B
0.0	90	10
1.0	90	10
1.3	75	25
3.0	72	28
11.0	60	40
11.5	5	95
12.0	5	95
12.5	90	10
17.0	90	10

boron atoms. Borate buffer is prepared by dissolving anhydrous borax (Na₂B₄O₇) in doubly distilled water to a concentration of 250 mM. The pH is subsequently adjusted to 8.2 by addition of 1M HCl. The solution is subsequently diluted to a final concentration of either 50 or 200 mM.

Preparation of libraries

Concentrations of library constituents are reported with respect to their building blocks, unless otherwise specified. In general, libraries were prepared by adding building block **1** (2.1 mg, 2.12 μmol) to a UPLC vial (12 x 32 mm) containing a Teflon coated magnetic stirring

bar (5 x 2 mm, VWR). The sample was dissolved in 1.06 mL borate buffer (either 50 or 200 mM) to give a final concentration of 2.0 mM building block **1** (unless otherwise specified). If required, samples were oxidized by the addition of sodium perborate (20 mM). The vial was closed with a Teflon septum screw cap and stirred at 1200 rpm at 25 °C unless otherwise specified. The temperature of the reaction mixtures was controlled using a Thermo Fisher compact dry bath with a custom milled aluminium block to accommodate HPLC vials (with dimensions 12 x 32 mm).

Preparation of **1**₆ seed

Stock solutions of **1**₆ were prepared by stirring a library (see *preparation of libraries*) until >90 % of building block **1** was converted to the replicating species. The composition of the library was analyzed by periodic UPLC injection. Typically this procedure takes up to two weeks, but can be sped up to roughly five days by stirring (1200 rpm) at 40 °C.

Experiments related to the 3D printed stirring device

Stirred versus unstirred samples (corresponding to **Figure 5** and **Supplementary Figure 1**)

Each of the four samples were prepared by diluting freshly built libraries (see *preparation of libraries*) with borate buffer (200 mM, pH 8.12) to 500 μM in a UPLC vial (12 x 32 mm). The volume of each vial was 1 mL to prevent significant loss of volume during injections. Two samples contained Teflon coated stirring bars (5 x 2 mm). The samples were placed in the 3D printed stirring device that had replaced the standard sample holder of the UPLC. The temperature of the UPLC sample holder was kept at 40 (± 0.1) °C. Samples were stirred at 1000 rpm and injected (10 μL, Aeris Peptide 1.7 μm XB-C18 x 2.10 mm column) every 72 minutes for five days. This experiment was repeated at a different time to obtain four repeats of both stirred and unstirred samples.

Effect of stirring rate on pre-oxidized libraries (corresponding to **Figure 6** and **Supplementary Figure 3**)

Each sample was prepared by oxidizing a freshly prepared library (see *preparation of libraries*) with sodium perborate so that 50% of building block **1** was oxidized. Subsequently, libraries were diluted to 500 μM with borate buffer (200 mM, pH 8.12) in a UPLC vial (12 x 32 mm). Each sample contained a Teflon coated stirring bar (5 x 2 mm), except for the four samples that were unstirred. All vials were placed in the 3D printed stirring device that had replaced the standard sample holder of the UPLC. The temperature of the UPLC sample holder was kept at 40 ($\pm$ 0.1) $^{\circ}\text{C}$. Samples were stirred at 200 rpm, 1000 rpm or left unstirred (each experiment contained four samples) and injected (10 μL , Aeris Peptide 1.7 μm XB-C18 x 2.10 mm column) every 90 minutes for 25 hours.

Thiol oxidation (corresponding to Figure 7)

Four libraries were prepared by diluting freshly built libraries (from building block **1**, see *preparation of libraries*) to 250 and 500 μM with borate buffer (200 mM, pH 8.12) in a UPLC vial (12 x 32 mm). Each sample contained a Teflon coated stirring bar (5 x 2 mm). All vials were placed in the 3D printed stirring device that had replaced the standard sample holder of the UPLC. The temperature of the UPLC sample holder was kept at 40 ($\pm$ 0.1) $^{\circ}\text{C}$. All samples were stirred at 1000 rpm and were injected every 90 minutes for 78 hours.

Elongation experiments (corresponding to Figure 11)

Libraries were constructed from building block **1** (see *preparation of libraries*) and oxidized to ~75% by the addition of sodium perborate. Subsequently, libraries were diluted to a final concentration of 154 μM . Short stacks of **1**₆ fibres were prepared (see Couette cell) and either stored in a freezer (**Figure 11A**) or used within 30 hours. The short stacks were diluted to 154 mM (either freshly prepared or thawed) and used to seed the libraries (15-22 %). Samples were placed inside the UPLC sample holder with a fixed temperature (25 ($\pm$ 0.1) $^{\circ}\text{C}$ for **Figure 11** or a range of temperatures for **Supplementary Figure 5B**) where the library composition was analysed every 18 minutes. Experiments were repeated four times.

Couette cell

Short stacks of **1₆** fibres were obtained by the following procedure: First, 150 µL of a library containing hexamers (see *preparation of 1₆ seed*) was pipetted in the cuvette. The cylinder was placed on top and the apparatus was screwed shut. The Couette cell operates by periodically increasing the rotational speed until the desired rpm is reached. To operate the machine at 4000 rpm, the rotational speed is increased by 500 rpm every five minutes (see **Supplementary Figure 7** for the full operational procedure).

Nucleation

A library was prepared (see *preparation of libraries*) and diluted to 154 µM in a UPLC vial (12 x 32 mm) with borate buffer (200 mM, pH 8.12) containing 200 µM methodol and a Teflon coated stirring bar (5 x 2 mm). The sample was stirred at 500 rpm and the temperature of the library (45 °C) was controlled using a Thermo Fisher compact dry bath with a custom milled aluminium block to accommodate HPLC vials (with dimensions 12 x 32 mm). The library composition was monitored by manual UPLC injection every 18 minutes.

Computer model

The models are given as a set of ordinary differential equations (ODEs). As an example, a model that contains a single nucleation stage is shown. This model consists of δ_{free} , δ_2 , and δ_{fiber} .

$$1) \frac{d[\delta_{free}]}{dt} = -2 k_1 [\delta_{free}]^2 + 2 k_{-1} [\delta_2] + k_{-2} [\delta_{fiber}] - k_2 [\delta_{free}] [\delta_2] - k_{auto} [\delta_{free}] [\delta_{fiber}]$$

$$2) \frac{d[\delta_2]}{dt} = 2 k_1 [\delta_{free}]^2 - 2 k_{-1} [\delta_2] + 2 k_{-2} [\delta_{fiber}] - 2 k_2 [\delta_{free}] [\delta_2]$$

$$3) \frac{d[\delta_{fiber}]}{dt} = 3 k_2 [\delta_{free}] [\delta_2] - 3 k_{-2} [\delta_{fiber}] + k_{auto} [\delta_{fiber}] [\delta_{free}]$$

Concentrations of species are given as the amount of hexamer in that particular species, which ensures mass conservation. The numbers before rate constants show the stoichiometry of the reaction. For example, the first term in 2) is $2 k_1 [\delta_{free}]^2$, which implies that two stoichiometric amounts of δ_{free} are used to generate two stoichiomet-

ric amounts of hexamer in δ_2 .

For a given set of rate constants, the model is simulated by numerical integration to give the kinetic plots (using MATLAB's `ode15s` function). Data fitting is used in order to locate a set that minimizes the difference between the simulated values and the experimentally measured ones (using `lsqcurvefit` function). Experimental data is compared with the sum of all the assembled states, by adding the concentration of all nucleation stages and resulting fiber (e.g. δ_2 and δ_{fibre}).

Fragmentation

Libraries were constructed from building block **1** (see *preparation of libraries*) and oxidized to ~85% by the addition of sodium perborate. Subsequently, libraries were diluted to a final concentration of 154 μM . Short stacks of **1**₆ fibres were prepared (see *Couette cell*) and were used within 20 hours. Libraries were seeded (between 9 and 13% **16**), stirred at 0, 300, 600, 900 and 1200 rpm while their temperature was kept at 25 °C using a Thermo Fisher compact dry bath with a custom milled aluminium block to accommodate HPLC vials (with dimensions 12 x 32 mm). The library composition was monitored by manual UPLC injection every 18 minutes.

Fragmentation calculations

For clarity, calculations for two rotational speeds are shown below.

Experimental results

From the concentrations shown in the table on the next page, we can determine the number of hexamer molecules in solution at any time point. To calculate the number of fibre ends, we assume that at $t=2$ all observed hexamers are solely from the added seed. We assume that each fibre is 34.8 nm long (average, based on TEM measurements) and each hexamer macrocycle has a height of 0.48 nm (average, based on MD simulations.)

Fibre ends 0 rpm

$$= 5,24 \times 10^{16}$$

Fibre ends 1200 rpm

$$= 4,26 \times 10^{16}$$

Time (minutes)	Hexamer in library (0 rpm, mM)	Hexamer in library (1200 RPM, mM)
2	0.00315	0.00256
20	0.00426	0.00423
38	0.00562	0.00595
56	0.00705	0.00777
74	0.00857	0.00962
92	0.01008	0.01132
110	0.01156	0.01287
128	0.01261	0.01410
146	0.01402	0.01476

Then, we calculate the initial rate of hexamer formation by taking the first five datapoints of the unagitated sample and using a linear regression fit. From this, we obtain a rate of $7,57 \times 10^{-5}$ mM hexamer minute^{-1} . Dividing this by the number of calculated fibre ends in the unagitated sample results in a rate of $1,44 \times 10^{-21}$ mM hexamer fibre end $^{-1}$ minute^{-1} .

Since all samples are seeded with different amounts of hexamer (experimental error) we must then multiply this rate by the number of fibre ends introduced at the start of the experiment ($t=2$) to obtain the rate of hexamer formation if these samples would be unagitated.

This results in rates of:

$$\begin{aligned} \text{Rate } 0 \text{ rpm} &= 7,57 \times 10^{-5} \text{ mM hexamer minute}^{-1} \\ \text{Rate } 1200 \text{ rpm (unagitated)} &= 6,15 \times 10^{-5} \text{ mM hexamer minute}^{-1} \end{aligned}$$

This means that if these samples would be unagitated, we would expect to see a faster rate of hexamer formation in the 0 rpm sample since it contained more seed at the beginning of the experiment.

These rates allow us to calculate a theoretical amount of hexamer formation if all samples would be unagitated and taking into account the initial amount of hexamer with which the libraries were seeded (see **Figure 14 B**). Subsequently, we subtract these theoretical concentrations from the observed concentrations to obtain a residual amount of hexamer that must have been formed due to the generation of more fibre ends through mechanical agitation (see **Figure 14 C**). Since we know the rate of hexamer formation per fibre end, we can then calculate the total number of fibre ends needed to explain the difference (see **Figure 14 D**).

ACKNOWLEDGEMENTS

J. Ottelé and S. Otto conceived the project. O. Markovitch programmed and developed the computer model. O. Markovitch, O. Veldman, J. Ottelé and A. Hussain developed the 3D stirring device. J. Ottelé performed all experiments unless otherwise noted. G. P. R. Schaeffer, H. Potgieser and O. Markovitch are thanked for the proofreading of this chapter and the ensuing discussions. L. Swier and R. Sneep are gratefully acknowledged for the mass spectrometry measurements.

BIBLIOGRAPHY

1. Dawkins, R. *The Selfish Gene*. (Oxford University Press, **1976**).
2. Orgel, L. E. Molecular replication. *Nature*, **1992**, 358, 203–209.
3. Lee, D. H., Severin, K., Ghadiri, M. R. Autocatalytic networks: The transition from molecular self-replication to molecular ecosystems. *Curr. Opin. Chem. Biol.* **1997**, 1, 491–496.
4. Paul, N., Joyce, G. F. Minimal self-replicating systems. *Curr. Opin. Chem. Biol.* **2004**, 8, 634–639.
5. Patzke, V., Von Kiedrowski, G. Self replicating systems. *Arkivoc*, **2007**, 293–310.
6. Ruiz-Mirazo, K., Briones, C., De La Escosura, A. Prebiotic systems chemistry: New perspectives for the origins of life. *Chem. Rev.* **2014**, 114, 285–366.
7. Wang, T. et al. Self-replication of information-bearing nanoscale patterns. *Nature*, **2011**, 478, 225–228.
8. Zhang, R., Walker, D. A., Grzybowski, B. A., Olvera De La Cruz, M. Accelerated self-replication under non-equilibrium, periodic energy delivery. *Angew. Chem. Int. Ed.* **2014**, 53, 173–177.
9. Dempster, J. M., Zhang, R., Olvera De La Cruz, M. Self-replication with magnetic dipolar colloids. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2015**, 92, 1–11.
10. Mange, D., Sipper, M., Stauffer, A., Tempesti, G. Toward self-repairing and self-replicating hardware: The embryonics approach. *NASA/DoD Conf. Evolvable Hardw.* **2000**, 205–214.
11. Reinhoudt, D. N., Crego-Calama, M. Synthesis beyond the molecule. *Science*, **2002**, 295, 2403–2407.
12. Jacobson, J. M., Griffith, S., Goldwater, D. Self-replication from random parts. *Nature*, **2005**, 437.

13. Kumar, A., Pate, K. M., Moss, M. A., Dean, D. N., Rangachari, V. Self-propagative replication of A β oligomers suggests potential transmissibility in Alzheimer disease. *PLoS One*, **2014**, 9, 1–11.
14. Xiao, Y. et al. A β (1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nat. Struct. Mol. Biol.* **2015**, 22, 499–505.
15. Scheidt, T. et al. Secondary nucleation and elongation occur at different sites on Alzheimer's amyloid-b aggregates. *Sci. Adv.* **2019**, 5.
16. Carnall, J. M. A. et al. Mechanosensitive Self-Replication Driven by Self-Organization. *Science*, **2010**, 327, 1502–1506.
17. Colomb-Delsuc, M., Mattia, E., Sadownik, J. W., Otto, S. Exponential self-replication enabled through a fibre elongation/breakage mechanism. *Nat. Commun.* **2015**, 6, 7427.
18. Malakoutikhah, M. et al. Uncovering the selection criteria for the emergence of multi-building-block replicators from dynamic combinatorial libraries. *J. Am. Chem. Soc.* **2013**, 135, 18406–18417.
19. Altay, M., Altay, Y., Otto, S. Parasitic Behavior of Self-Replicating Molecules. *Angew. Chem. Int. Ed.* 57, **2018**, 10564–10568.
20. Monreal Santiago, G., Liu, K., Browne, W. R., Otto, S. Emergence of light-driven protometabolism on recruitment of a photocatalytic cofactor by a self-replicator. *Nat. Chem.* **2020**, 12, 603–607.
21. Ottel, J., Hussain, A. S., Mayer, C., Otto, S. Chance emergence of catalytic activity and promiscuity in a self-replicator. *Nat. Catal.* **2020**, 3, 547–553.
22. Frederix, P. W. J. M. et al. Structural and spectroscopic properties of assemblies of self-replicating peptide macrocycles. *ACS Nano*, **2017**, 11, 7858–7868.
23. Swartz, M. E. UPLCTM: An introduction and review. *J. Liq. Chromatogr. Relat. Technol.* **2005**, 28, 1253–1263.
24. Colomb-Delsuc, M. Mechanistic Studies on a Peptide-Based Self-Repli-

cating System. (University of Groningen, **2015**).

25. Yang, S. et al. Chemical Fueling Enables Molecular Complexification of Self - Replicators. *Angew. Chem. Int. Ed.* **2020**, 60, 2-8.

26. Xing, W. et al. Oxygen Solubility, Diffusion Coefficient, and Solution Viscosity. Rotating Electrode Methods and Oxygen Reduction Electrocatalysts (Elsevier, **2014**).

27. Vendruscolo, F., Rossi, M. J., Schmidell, W., Ninow, J. L. Determination of Oxygen Solubility in Liquid Media. *ISRN Chem. Eng.* **2012**, 1–5.

28. Zhao, R., Lind, J., Merényi, G., Eriksen, T. E. Kinetics of one-electron oxidation of thiols and hydrogen abstraction by thiyl radicals from α -amino C-H bonds. *J. Am. Chem. Soc.* **1994**, 116, 12010–12015.

29. Nagy, P. Kinetics and mechanisms of thiol-disulfide exchange covering direct substitution and thiol oxidation-mediated pathways. *Antioxidants Redox Signal.* **2013**, 18, 1623–1641.

30. Singh, R., Whitesides, G. M. Thiol-disulfide interchange. *Chem. Sulphur-Containing Funct. Groups*, **2010**, 633–658

31. Pal, A. et al. Controlling the Structure and Length of Self-Synthesizing Supramolecular Polymers through Nucleated Growth and Disassembly. *Angew. Chem. Int. Ed.* **2015**, 54, 7852–7856.

32. Monnard, P. A., Ziock, H. Eutectic phase in water-ice: A self-assembled environment conducive to metal-catalyzed non-enzymatic RNA polymerization. *Chem. Biodivers.* **2008**, 5, 1521–1539.

33. Dyson, F. A meeting with Enrico Fermi How one intuitive physicist rescued a team from fruitless research. *Nature*, **2004**, 427, 8540.

34. Mayer, J., Khairy, K., Howard, J. Drawing an elephant with four complex parameters. *Am. J. Phys.* **2010**, 78, 648–649.

35. Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. in *Selected Papers of Hirotugu Akaike*. Springer Series in Statistics (Perspectives in Statistics). (ed. Parzen E., Tanabe K., K. G.) 199–

213 (Springer, **1998**).

SUPPLEMENTARY FIGURES

Supplementary Figure 1: Chromatograms at various timepoints during an experiment inside the sample holder of the UPLC. The scale in all chromatograms is the same. For each chromatogram small offsets on the x-axis (0.5 minutes) and y axis (0.15 a.u.) are used for clarification purposes. A) Samples that are stirred at 1000 rpm show significant conversion to the hexamer replicator. B) Unstirred samples exhibit only little hexamer formation, which is only present in a minimal amount >5 %.

Observed mass	Calculated mass	Macrocycle
760.37	760.34	$[1+1H]^{1+}$
1010.80	1010.79	$[1_4+1H]^{3+}$
1137.03	1137.01	$[1_6+1H]^{4+}$
1263.25	1263.23	$[1_5+1H]^{3+}$
759.37	759.34	linear $[1_2+1H]^{2+}$
759.03	759.01	linear $[1_3+1H]^{3+}$
1137.03	1137.00	$[1_3+1H]^{2+}$

Supplementary Figure 2: Chromatogram and mass spectrum obtained by LC-MS of a library constructed from building block **1** that has not converted all material to the self-replicating species yet (although the replicator is present). A) Chromatogram that shows the retention time of each species. B-H) Mass spectrum identifying the formation of monomer, tetramer, hexamer, pentamer, linear dimer, linear trimer and trimer respectively.

Supplementary Figure 3: A) Library composition of unstirred samples under the same conditions as the experiments shown in **Figure 6**. B) Total integrated peak area during the experiment shown in **Figure 6A**. Each line represents a single sample. We observe standard deviations of less than 2%.

Supplementary Figure 4: Length distribution of fibres exposed to high shear stress (corresponding to **Figure 10B**) were based on the analysis of these TEM micrographs. The libraries consisted of >95% **1₆** and were subjected to 66.7 Hz (or 4000 rpm) as described under General procedures.

Supplementary Figure 5: A) Fluctuations in library composition of samples “frozen” using TFA (by adding 10 μ L of a 1.54 mM library to 90 μ L UPLC grade H₂O containing 0.6 v% TFA). B) Temperature effects on elongation rate (experiments performed as described in section 2.2.2 Elongation).

Supplementary Figure 6: Library composition when subjected to different stirring rates. Libraries were constructed from building block **1**, oxidized to ~85% using sodium perborate and seeded with short fibre stacks (~10%, prepared as described in section 2.2.2 Elongation). The libraries were subjected to various degrees of mechanical agitation (stirred at 0, 300, 600, 900 and 1200 rpm using a magnetic stirring bar) while library composition was periodically monitored using UPLC injection. For clarity, oxidized monomers refer to any species that is not monomer or hexamer (mostly trimers and tetramers).

Supplementary Figure 7: Couette cell operational procedure when preparing short stacks of 1_{ϕ} .

SUPPLEMENTARY TABLES

Nuclei	AIC	Number of parameters	RMSD
0	7.586943	3	0.452272
1	10.70463	5	0.70306
2	18.80225	7	0.090616
3	21.01914	9	0.221005
4	27.36916	11	0.06825
5	29.624	13	0.163327
6	32.41325	15	0.299205
7	34.95687	17	0.619754

Supplementary Table 1: Calculated Akaike information criterion for models with various nuclei. It should be noted that the number of parameters influences the AIC much more than the differences in the RMSD.

CHAPTER 3

Mechanistic insight into supramolecular polymerization-driven self-replication from real-time visualization

This chapter has been published: Maity, S., Ottel, J., Santiago, G. M., Frederix, P. W. J. M., Kroon, P., Markovitch, O., Stuart, M. C. A., Marrink, S. J., Otto, S., Roos, W. Caught in the Act: Mechanistic Insight into Supramolecular Polymerization-Driven Self-Replication from Real-Time Visualization. J. Am. Chem. Soc. 141, 3, 1685-1689, 2020

ABSTRACT

Self-assembly features prominently in fields ranging from materials science to biophysical chemistry. Assembly pathways, often passing through transient intermediates, can control the outcome of assembly processes. Yet, the mechanisms of self-assembly remain largely obscure due to a lack of experimental tools for probing these pathways at the molecular level. Here, the self-assembly of self-replicators into fibers is visualized in real-time by high-speed atomic force microscopy (HS-AFM). Fiber growth requires the conversion of precursor molecules into six-membered macrocycles, which constitute the fibers. HS-AFM experiments, supported by molecular dynamics simulations, revealed that aggregates of precursor molecules accumulate at the sides of the fibers, which then diffuse to the fiber ends where growth takes place. This mechanism of precursor reservoir formation, followed by one-dimensional diffusion, which guides the precursor molecules to the sites of growth, reduces the entropic penalty associated with colocalizing precursors and growth sites and constitutes a new mechanism for supramolecular polymerization.

3.1 INTRODUCTION

The focus of research on supramolecular self-assembly is broadening from exclusively thermodynamically controlled structures to out-of-equilibrium systems.¹⁻⁵ While for the former the final structure

is not influenced by the route it takes, the outcome of out-of-equilibrium self-assembly is dictated by the assembly pathway,^{6–8} which spurs efforts to unravel assembly mechanisms. Insight is needed into how labile nanoscale assemblies change with time, for which only a few techniques are available. Current methods for real-time visualization of such systems use confocal laser scanning microscopy⁹ or stochastic optical reconstruction microscopy.¹⁰ These methods provide for resolutions down to 80 and 20 nm, respectively, and require the use of fluorescent probes. Recent advances in the field of atomic force microscopy (AFM) have enabled the studying of dynamic processes of (bio)molecular systems using high-speed AFM (HS-AFM)¹¹ at even smaller length scales, including the configurational dynamics of proteinaceous structures,^{12–14} the assembly of amyloid-like fibrils,^{15,16} and the movement of synthetic molecular transporters¹⁷ with unprecedented spatiotemporal resolution.

By employing HS-AFM, we have now been able to elucidate the molecular mechanism of the recently discovered systems of self-assembly-driven self-replication.¹⁸ A prominent feature of such systems is that, under mechanical agitation (shaking, stirring), the assembly processes take place in a mixture of interconverting molecules leading to the autocatalytic sequestration of the assembling molecules and causing their exponential self-replication.^{18,19} The spontaneous emergence of self-replicators out of such systems appears general, has been observed for different compound classes,^{20–22} and is relevant in the context of the origin and the *de novo* synthesis of life.^{23,24}

We focused our mechanistic investigation on self-assembling replicators that are formed from the monomeric building block **1**, which features two thiols that are readily oxidized to form disulfide bonds, and initially produces a mixture of differently sized macrocycles that interconvert through thiol–disulfide exchange²⁵ (**Figure 1A**). When investigated under mechanical agitation (stirring), subsequent to a lag phase in which trimers and tetramers (**1**₃ and **1**₄) are the dominant products, a selfreplicator (cyclic hexamer **1**₆) emerged, following a nucleation–growth mechanism, during which **1**₃ and **1**₄ are converted into **1**₆ (**Figure 1B**).

To obtain the molecular details of the supramolecular polymerization-driven replication process, we used HS-AFM, supported by chemical analysis and molecular dynamics (MD) simulations. The results (presented below) revealed an unexpected assembly/replication mechanism summarized in **Figure 1C**. Monomer **1** jointly with

the small macrocycles $\mathbf{1}_3$ and $\mathbf{1}_4$ (together termed the “precursors”) form small off-pathway aggregates in solution that are in equilibrium with the nonassembled precursors. Importantly, the free precursors can also bind and accumulate as aggregates onto the fiber surface, to form a reservoir from where they diffuse toward the end of the fiber, which results in fiber elongation. This new mechanism of reservoir formation followed by one-dimensional diffusion is essential for efficient fiber growth as it directs the precursors to the fiber ends that are present at very low concentration.

3.2 RESULTS AND DISCUSSION

3.2.1 CAPTURING THE ASSEMBLY PATHWAY OF A SELF-REPLICATOR AT THE SINGLE PARTICLE LEVEL

We monitored fiber growth by HS-AFM. Fibers were absorbed onto a mica surface covered with a slightly negatively charged lipid bilayer to ensure binding while still allowing for fiber growth (see *Methods*). Upon addition of a solution containing precursors, we observed aggregates of these precursors attached to the sides of the fibers (**Figure 2A**). We distinguished between instances where (i) precursors are attached near the end of the fiber, (ii) precursors were attached near the middle of the fiber, and (iii) no precursor was attached (**Figure 2A,B**). Analyzing AFM images revealed that precursor attachment near the end occurs at a higher frequency than attachment close to the middle (**Figure S1**). Interestingly, the majority of the fibers with precursors attached near their ends were growing substantially over time (**Figure 2C**, see supplementary note for a statistical analysis), while the fibers without bound precursor, or those where the precursors stayed attached near the middle, showed limited or no growth during the same time interval (**Figure 2D,E**). These observations suggest that both the formation of aggregates of precursors and their proximity to the fiber end are necessary conditions for extension of the fiber.

Next, we performed dynamic studies using HS-AFM imaging at 0.5 frame/s, to focus on the fibers that have precursors bound near their end (**Figure S2A–F**). As can be seen in **Figure S2A–F** and **Video S1**, these fibers grow at an average rate of ~ 5 nm/min ($N = 40$), which corresponds to the attachment of ~ 11 units of $\mathbf{1}_6$ per minute.

Figure 1 (left): The fiber self-assembly pathway. A) Fiber formation from building block **1**. Upon oxidation (1) the monomer forms a mixture of macrocycles that can interchange building blocks with one another through disulfide exchange reactions (2). Following slow nucleation (3), **1₆** can elongate by the stacking of additional hexamer macrocycles (4). B) Representative kinetic analysis of relative molecular concentrations over time, performed using ultra-performance-liquid-chromatography (UPLC) under stirred conditions. Concentration of monomer (**1**) (black squares) diminishes by reaction with oxygen to form small, soluble macrocycles (cyan triangles). After an initial lag phase (roughly 500 min), the concentration of hexamer (**1₆**) (orange circles) increases as fibers are formed. Insets show the coarse-grained models of monomer, trimer (**1₃**), tetramer (**1₄**), and stacks of hexamers (fibers), and a high-resolution AFM image of a single fiber on a lipid bilayer (top right, scale bar 10 nm). Amino acid side chains are not shown in the coarse-grained models. C) Model representation of the self-assembly pathway summarizing the main findings of the present work (see main text).

Note that, as the fibers grow, the relative volume of the attached precursors shrinks (**Figure S2D,E**). On the other hand, for a fiber with precursors attached in the middle, both the length and the relative volume of attached precursors remain unchanged over time (**Figure S2G–I** and **Video S2**). These observations of decrease in precursor volume, only on a growing fiber, suggest that the aggregates serve as a reservoir supplying material for fiber extension. Further dynamic studies confirm this hypothesis by capturing the complete process of fiber selfassembly. **Video S3** and **Figure 3A** show a growing fiber for which five stages were identified. These distinct states are clearly visible when a kymograph is made from a line section over the fiber for the full period of the recording (**Figure 3B** and **C**). An initial growth phase from 0 to 340 s at ~8.5 nm/min (phase 1) is followed by a slower growth of ~2 nm/min as the reservoir becomes more distant from the growing fiber end (phase 2 [until ~880 s]). Later, a stagnant phase is observed when the reservoir is presumably too far away from the growing end of the fiber (phase 3 [until ~1330 s]). New precursor accumulation from solution takes place in between ~930–1330 s along with diffusion along the fiber. Next, slow growth occurs from ~1330–1800 s (phase 4), while a new reservoir is gradually forming on the fiber. From ~1800 s on, the fiber grows rapidly again at ~8 nm/min, while the reservoir starts depleting (phase 5).

Figure 2: Precursor attachment near the fiber end is essential for successful growth. A) Example of an AFM image of fibers grown on a membrane surface in the presence of precursors. Three different fiber-precursor interactions are indicated as green circles (precursors attached near the end of the fiber), a blue square (precursors attached in the middle of the fiber), and gray arrows (fibers without any precursor attached). B) Schematic representation of these three fiber-precursor arrangements. C) Histogram of the length of fibers having precursors attached near the end and imaged at three different times in the presence of a 2.31 mM precursor solution. An increase in fiber length can be observed going from 0 min (in purple, $N = 70$), to 30 min (in cyan, $N = 139$), and 60 min (in orange, $N = 108$). D) Same as C) but for the fibers having precursors attached in the middle of the fibers. No significant growth was observed after 60 min. $N = 20, 17$, and 39 for imaging after 0, 30, and 60 min, respectively. Inset shows the same data enlarged. E) Same as C) but for fibers having no precursor attached on the fiber. No significant growth was observed after 60 min. $N = 120, 77$, and 90 for imaging after 0, 30, and 60 min, respectively

Interestingly, while the gradually emerging precursor aggregates show an early stage diffusivity toward the fiber end (phase 3 in **Figure 3C** and **Video S3**), the aggregate eventually stabilizes on the fiber surface (phase 5 in **Figure 3C**) and participates in the growth process. This observation is consistent with what is shown in **Video S1**, where the already formed precursor aggregate does not move while the fiber still grows. These observations lead us to speculate that the precursor attachment followed by accumulation into aggregates occurs randomly, while the rate of diffusion of aggregates reduces with increasing aggregate mass. This would also explain the low probability of finding the precursors attached near the middle of the fiber (**Figure 2**).

Precursors may initially have attached anywhere along the fiber, but both individual molecules as well as precursor aggregates seem to be prone to diffuse along the fiber toward its end. The small fraction of fibers that were observed to contain the precursors roughly in the middle (**Figure 2A,D** and **Figure S1**) is likely the result of inefficient diffusion due to a surface attachment-induced blockage. For the growing fibers, typically a clear asymmetry in growth was observed as only one direction of the fiber elongated. At the used frame rate of 0.5 frame/s and an image size of 300×300 pixels, the pixel to pixel acquisition time is 22 μs /pixel, which seems to allow for capturing the presence of precursors on top of the fiber without smearing their position out. When a movie of a growing fiber is analyzed using image segmentation by masking the height of the fiber (~ 3.5 nm) and anything above this height is monitored over time, the dynamics on top of the fiber surface can be observed (**Figure S3** and **Video S4**). Following this dynamics indicates that the diffusion dominates in the direction of fiber growth, as visualized in **Figure S3**.

3.2.3 ELONGATION OF THE FIBERS SCRUTINIZED IN BULK SOLUTION

We also monitored the rate of fiber elongation without any mechanical agitation, and in bulk solution containing 2.31 mM of precursors and 4.3% (in terms of building block **1**) of **1₆** replicator by UPLC analysis, to reveal a growth rate of ~ 4 nm per minute per fiber end, or ~ 8 units of **1₆** per minute (**Figure 3D**), consistent with the AFM results obtained from the fibers attached and grown on a membrane surface (**Figure S2F**). When we repeated the same experiment using a 15-

Figure 3: Precursor accumulation, diffusion, and fiber growth as observed in real-time. A) Snapshots of AFM images of fibers growing on a membrane surface at different times. The cyan arrows indicate the first active precursor aggregate's position, the pink arrows indicate the growth site of the fiber, and the green arrows indicate the second precursor aggregate's accumulation and position. B) Representative image of the growing fiber from panel A showing the line (in cyan) selected to construct the time-resolved intensity kymograph in panel C. Scale bars in A) and B) are 20 nm. C) Kymograph along the line section in B) over 1930 s. Dashed lines mark the different stages of growth as described in the text. The immobile fiber that is encountered by the growing fiber in phase 2 was removed from the kymograph for clarity. D) Fiber elongation rate determination by UPLC. Shortened fibers were used as a seed in the presence of preoxidized precursors to measure the elongation speed in solution. The results are consistent with the values obtained from AFM experiments.

fold less concentrated precursor solution, we found that the growth rate decreased only by a factor of ~ 3 (**Figure S4**). A systematic investigation revealed that the growth rate levels off at high precursor concentrations (**Figure S5**). This behavior cannot be explained by saturation of the fibers with precursors, because AFM analysis shows a significant number of fibers that are devoid of precursors in the same concentration range. Instead, these results suggest the existence of an off-pathway assembled state of the precursors that does not contribute directly to the growth of the fibers.

To investigate the nature of this off-pathway state, we studied the precursor solutions by cryo-transmission electron microscopy, by dynamic light scattering, and by analyzing the fluorescence of a solvatochromic probe. The resulting data confirm that the precursors form aggregates in solution (**Figure S6**). These aggregates are present already from very low concentrations of precursors (critical aggregation concentration of $23 \pm 5 \mu\text{M}$ in building block **1**), and because we do not observe them directly attaching to the fibers by AFM (fiber-bound aggregates appear to grow gradually), we infer that they do not contribute to fiber growth. However, they do have an active role by releasing free precursors into the solution, which replenishes those that participated in fiber growth.

To prove that the postulated fiber growth mechanism is also occurring for fibers free in solution and to rule out surface artifacts, we allowed fibers to grow in bulk solution, while periodically taking samples and imaging these using AFM. The resulting data (**Figures 4A** and **S7**) reveal a distribution of precursor attachment to fibers similar to that observed in the on-surface experiments. For the first 10 min, the precursors stay attached while the fiber length increases. Interestingly, after 25–35 min, the precursors appear to have migrated to the ends of the fibers, while the fibers had continued to grow. After 60 min, the aggregates have almost disappeared from the fibers and presumably also from the solution, while the fibers have grown to their full extent. The spreading of the aggregates with time is accompanied by a reduction in their height (**Figure S7**). From the change of fiber length over time (**Figure S7**), a growth rate of around $\sim 5 \text{ nm/min}$ ($N > 20$ for each interval) was obtained, in good agreement with our previous estimation from **Figure S2F** (fiber grown on a membrane surface) and **Figure 3D** (estimated in bulk by UPLC from fibers grown free in solution).

Figure 4: Diffusion of precursors along the fiber toward the end. A) Representative AFM images of fibers immobilized at different times on a surface after growth free in solution. Scale bar 50 nm. The green insets show the cross-section side view of the fiber along the dotted green line in the corresponding figures. The arrows indicate the position of the precursor aggregates. B) Simulated coarse-grained (CG) fiber structure (16 hexamers, 8 nm) with one molecule of trimer showing the diffusion at different times over a 500 ns simulation. Different colors (from red to blue) indicate the position of the trimer at different times (from 0 to 500 ns). The fiber is represented as a gray surface. (C–E) CG MD simulation of fiber (outlined in black) and single trimer molecules over time. Relative density of trimer added to a preassembled fiber is averaged over 400 simulations at the start of the simulation (0 ns), after 60 ns, and after 500 ns, respectively. The color bar representing the density plots in C)–E) is normalized from 0 to 1.

A summary of estimated growth rates from different experimental approaches can be found in **Table S1** and **supplementary discussion 1**. Also, as observed in **Figure 4A** (and **Figure S7**), in the course of the first 35 min, the precursors have spread out from the reservoir and moved toward the fiber ends. This suggests that the rate of diffusion of the precursors along the fiber may be significantly faster than the rate of growth, which causes precursors to accumulate near the fiber ends rather than being inserted into the fiber. It also suggests that the diffusion along the fibers occurs unobstructed in solution, while surface artifacts inhibit efficient diffusion when the precursors are

attached in the middle (**Figures 2A,D and S1**). However, the observed rates of fiber growth in both conditions are similar to each other (**Table S1**). These observations suggest that the diffusivity of the precursors along the fiber is not the growth-limiting factor in this assembly mechanism. To support this hypothesis, we have compared the rate of volume gained by the fiber itself during an elongation to the rate of volume loss by the corresponding precursor aggregate (**Figure S8 and Video S5**). It can be observed that the precursors diffuse out of the aggregate faster than the growth of the fiber occurs, which fits with our other observations.

3.2.4 MECHANISTIC INSIGHTS INTO PRECURSOR DIFFUSION AND FIBER ELONGATION FROM MD SIMULATIONS AND MASS-KINETIC MODELS

Further support for the postulated assembly and replication mechanism comes from MD simulations. Recent improvements in soft- and hardware, force fields, and enhanced sampling techniques have created new possibilities for the study of complex assembly processes.^{26–30} Specifically, we employed MD simulations to confirm the hypothesis of the diffusion of precursors along the fiber. A short fiber composed of 16 **1₆** macrocycles was simulated as previously,³⁰ and a single **1₃** macrocycle was added at a random location in the surrounding explicit solvent at a distance of >2 nm from the stack. The macrocycle was allowed to diffuse and bind to the fiber. Repeated atomistic simulations (N = 200, **Video S6**) show that binding is typically rapid (<20 ns), but exhibits no preference for any specific location along the fiber axis, except for a tendency to bind to the hydrophobic core of the fiber (**Figure 4B**). As time progressed, diffusion of **1₃** occurred along the fiber toward the fiber ends, in agreement with the experimental observations. To probe this process over longer time scales, using larger fibers and with better statistics, a coarse-grained (CG) model for the self-replicating macrocycles was developed using the Martini force field.³¹ This model correctly reproduces the binding free energy of hexamer macrocycles to the fiber ends, the chiral pitch observed using cryo-TEM and AFM (**Figures 1A, S9, and S10**), and other structural parameters (**Figure S11**). CG binding simulations, performed in the same way as for the atomistic simulations, confirmed the diffusion of the macrocycles along the fibers in the course

of 500 ns ($N = 400$) after binding to the fiber (**Figure 4C–E** and **Video S7**). The indications that precursor diffusion follows the chiral pitch of the fiber could upon strong fiber attachment to the surface lead to precursor blockage by the surface as shown by the data of **Figure 2D**. Finally, mass-action kinetic models of the assembly and replication processes were constructed (**supplementary discussion 2**). In the simplest model, the fibers grow directly from nonassembled **1_g** that is sequestered from bulk solution, where it is a minor constituent of the dynamic macrocycle mixture.

The more elaborate model includes a role of the fibers in converting precursors into **1_g**, which thus captures the role of the fibers in assimilating the precursors and directing them to the fiber ends. Attempts to fit the experimental kinetic data using the two models only yielded an acceptable fit for the more elaborate model, which provided further support for the validity of the postulated mechanism, as shown in **Figure 1B**.

3.3 CONCLUSION

We have been able to directly visualize molecular selfreplication in real-time with unprecedented resolution and obtained detailed and unexpected insights into the mechanism of the self-assembly-driven self-replication process. HS-AFM revealed a mechanism of supramolecular polymerization, where accumulation of precursor reservoirs occurs along the sides of the existing assemblies. While this mechanism bears some resemblance to the previously proposed secondary nucleation model for amyloids,³² it is distinctly different from this model, as this new aggregate does not itself elongate, but instead promotes growth of the fiber after diffusion of the precursors from the reservoirs toward the fiber end. Results from atomistic and CG MD simulations provide support for molecular diffusion along the fiber as an important step in fiber growth. This diffusion of precursors reduces their search for a growing fiber end from a 3D to 1D problem, which lowers the entropic barrier of supramolecular polymerization.

HS-AFM imaging of surfaces can be performed at a frequency of 10 frames per second.¹¹ However, height fluctuations of a single line can be studied 100 times faster,³³ which thereby allows for single-millisecond temporal resolution. Next, for studies of man-made systems, the presented HS-AFM approach can likely also shed light on the mechanism in which secondary nucleation and elongation oc-

curs for amyloid fibrils, as the exact mechanism of this process remains unclear.³⁴ To summarize, our results not only establish a new self-assembly mechanism that might well extend to other biological/synthetic systems, but also establish HS-AFM as a powerful tool to unravel self-assembly processes.

METHODS

Fiber Formation

A stock solution of **1₆** was prepared by adding building block **1** [(3,5-dimercaptobenzoyl)glycyl-L-leucyl-L-lysyl-L-phenylalanyl-L-lysine]) to a 1 mL HPLC vial (12 × 32 mm) containing a Teflon-coated magnetic stirring bar (5 × 2 mm, VWR). The building block was dissolved in borate buffer prepared from 25 mM B₂O₃ and adjusted to a pH of 8.1 to a final concentration of 1.54 mM and was kept at elevated temperatures while mechanical agitation was applied (1200 rpm, 45 °C). The sample was subjected to periodic UPLC analysis and kept at the conditions described above until the sample contained >90% **16**. The **1₆** fibers were then stored at room temperature while stirring and could be used up to 8 weeks after preparation without observing any significant changes in sample composition.

Seed Preparation by Mechanical Shearing

From the **1₆** stock solution, a 150 μL aliquot was placed in a Couette cell ($R_{\text{cup}} = 20.25$ mm, $R_{\text{bob}} = 20$ mm, average radius (R) = 20.125 mm). The sample was subjected to mechanical shearing by rapid rotation of the inner cylinder. The rotational frequency used was 4000 rpm for 30 min, corresponding to a shear rate ($\dot{\gamma}$) of 33 702 s⁻¹. The resulting seeds were stored at room temperature and used within 2 days of preparation.

Fiber End Estimation

The average fiber length of the sheared seeds was analyzed using transmission electron microscopy. Using ImageJ software, we measured the length of 994 sheared seeds, which resulted in an average length of 34.8 ± 15.2 nm. The average height of a single **1₆** macrocycle³⁰ is 0.485 nm; therefore, we find an average of 71.8 ± 31.3 macro-

cycles in a sheared seed fiber.

Precursor Solution Preparation

A stock solution of a mixture of **1**, **1**₃, and **1**₄ was prepared by adding building block **1** to a HPLC vial (12 × 32 mm) and transferring it to a glovebox. Building block **1** then was oxidized using sodium perborate (0.80–0.85 equiv) in borate buffer to obtain final concentrations of 1.54 and 2.31 mM. The resulting mixture was analyzed by UPLC and could be used up to 3 days if no mechanical agitation was supplied.

Ultra-Performance-Liquid-Chromatography (UPLC) Elongation Experiments

Low Concentration.

In a 1 mL HPLC vial was diluted 85 μ L of an oxidized precursor solution (92–94% oxidation, 1.54 mM in building block in 50 mM borate buffer, pH 8.12) with 900 μ L of UPLC grade H₂O. Sheared seeds were added (15 μ L) and the mixture was thoroughly mixed, which resulted in a final concentration of 0.154 mM in building block (precursors). This mixture was kept without any mechanical agitation at a constant temperature of 25 °C, and the composition of the sample was monitored by UPLC every 18 min for >800 min. The elongation experiment was repeated four times.

High Concentration.

A glass insert was placed in a 1 mL HPLC vial. Of an oxidized precursor solution (88% oxidation, 2.31 mM in building block in 50 mM borate buffer, pH 8.12), 95 μ L was added to the insert. Two minutes before the UPLC injection, 5 μ L of sheared seeds (2.0 mM in building block) was added and the sample was mixed thoroughly, which resulted in a final concentration of 2.29 mM. The mixture was kept without any mechanical agitation at a constant temperature of 25 °C, during which the library composition was monitored by UPLC every 18 min for 128 min. The elongation experiment was repeated four times.

Elongation Experiments Monitored by Fluorescence

Samples containing sheared **1**₆ seeds (60 μ M in building block),

thioflavin T (500 μM), and increasing concentrations of precursors (0–2.3 mM in building block, 86% oxidation) were prepared in a 96-well plate using borate buffer as a solvent. The samples were shaken (orbital shaking for 30 s) at a controlled temperature of 25 $^{\circ}\text{C}$, and the fluorescence of thioflavin T ($\lambda_{\text{exc}} = 440 \text{ nm}$, $\lambda_{\text{em}} = 500 \text{ nm}$) was measured every 5 min using a Synergy|H1 microplate reader (BioTek, U.S.). Simultaneously, samples containing thioflavin T (500 μM) and increasing concentrations of **1₆** seeds or precursors were monitored in the same way. These samples were used as calibration to correlate the fluorescence signal with concentrations of both precursors and **1₆**, and to monitor the photobleaching of thioflavin T (which remained always lower than 5% of the initial signal). The fluorescence intensity at every time point was converted to the concentration of **1₆** after subtracting the signal coming from precursors, and the initial growth rate was calculated by linear regression of the first five points of each sample. This experiment was repeated three times.

Surface Preparation for AFM Studies

To immobilize the fibers on the surface and to provide at the same time freedom for the fiber to grow, we have used a lipid bilayer deposited on top of freshly cleaved mica. The lipid bilayer was formed by absorption of large unilamellar vesicles (LUVs) onto a freshly cleaved mica surface. LUVs were prepared using a lipid mixture composed of 60% dioleoylphosphatidylcholine (DOPC) and 40% dioleoylphosphatidyl-serine (DOPS) (mol:mol) from Avanti Polar. The lipid mixture containing 1 mg/mL of total lipids was mixed in 200 μL of chloroform in a small glass vial. Next, chloroform was evaporated using argon gas while the vial was slowly rotated to produce a lipid film on the glass wall. The film was kept in a vacuum desiccator for 30–45 min. After the lipid film was dried, 200 μL of a buffer composed of 10 mM HEPES, pH 7.4, 100 mM NaCl, and 50 mM sucrose was added and vortexed for 30 s. The mixture was freeze–thawed three times using liquid nitrogen. The LUVs were stored at -20°C for further use within 1 month. For deposition on a mica surface, we have used 0.2 mg/mL concentration of the stock preparation (diluted in the same buffer) and incubated on top of freshly cleaved mica (HS-AFM sample holder) for 15–30 min. The surface was then cleaned 3–5 times with 50 mM borate buffer, pH 8.1.

HS-AFM Experiments

All of the AFM studies were done using HS-AFM (RIBM, Japan) in amplitude modulation tapping mode in liquid.^{17,35–37} Short cantilevers (USC-F1.2-k0.15, NanoWorld, Switzerland) with a spring constant of 0.15 N/m, resonance frequency around 0.6 MHz, and a quality factor of ~2 in buffer were used. The cantilever free amplitude was set to 1 nm, and the set-point amplitude for the cantilever oscillation was set around 0.9 nm. Images were taken at 0.2–0.5 frame/s depending on the size of the image. A mica surface of diameter 1.5 mm glued on top of a 5 mm high glass rod was used as the AFM sample stage. The glass rod was then attached to the scanner Z-piezo using a small amount of wax. After formation of the lipid bilayer (as mentioned above), the short preassembled fibers (seeds) were incubated for 30 s and then cleaned with borate buffer. The scanner head was then put upside down into a small liquid chamber containing the cantilever and filled with 120 μ L of the recording solution. All on-surface growth experiments were performed in the presence of 2.31 mM precursor in 50 mM borate buffer, pH 8.12. The HS-AFM works as a sample scanning system, and a minimum imaging force (<100 pN) was applied throughout all experiments.

All AFM measurements were done in solution, and we performed AFM imaging experiments on fibers that were grown on a membrane surface and on fibers that were grown free in solution (in a glass bottle) and later deposited on a mica surface to image the fiber. For long-term (> ~10 min) characterization of the growth process, small seeds were immobilized on a lipid surface, and the chamber was filled with 2.31 mM precursor solution. AFM images were taken at different time points. Because of mechanical drift, imaging the exact place after several minutes was not possible; therefore, we have estimated the elongation by measuring the length of at least 10 fibers on the surface for each point in time. For dynamic studies, we used a similar approach, but after localizing a seed with precursors attached near its end, we zoomed in and imaged it continuously at typically 0.5 frames/s. Because of the low growth rate, mechanical drift during imaging, and the small piezo limit (900 nm $\times$ 900 nm), we were only able to follow a fiber for typically ~10 min. Finally, we also performed AFM experiments on fibers that were not grown on a surface, but free in solution. To do this, we incubated the seeds and the precursors at a 5%:95% molar ratio in a sealed glass bottle. For every predecided

time, we then took a small amount of the mixture, which we added onto a freshly cleaved mica surface and left to incubate for 30 s. Next, the surface was rinsed with borate buffer, and the AFM imaging was done immediately afterward in borate buffer.

AFM Data Analysis

For AFM data analysis, we have used Igorpro software with built-in script from RIBM (Japan) and ImageJ software with additional home-written plugins. The HS-AFM images/movies were only processed minimally, through tilt correction, drift correction, and brightness correction. The kymographs were obtained from the cross-section at a fixed scale (marked for each image) over the entire movie. It represents the height distribution (in terms of intensity) along the line cross section as a function of time. For all different experimental conditions, we obtained and reported the results from several days of experiments. Rounding of growth rates and other values was performed on final numbers. Statistical tests on the relevant AFM data sets are reported in the supplementary notes (Tables S2 and S3).

Cryo-Transmission Electron Microscopy

An aliquot (3 μ L) of solutions containing **1**₆, precursors, or both (4 mM, prepared in borate buffer) was deposited on holey carbon-coated grids (3.5/1 Quantifoil Micro Tools, Jena, Germany) that were previously glowdischarged for 15 s. After the excess liquid was blotted for 4 s, the grids were vitrified in liquid ethane using a Vitrobot (FEI, Eindhoven, The Netherlands) and transferred to a FEI Tecnai T20 electron microscope equipped with a Gatan model 626 cryo-stage operating at 200 keV. Micrographs were recorded under low-dose conditions with a slow-scan CCD camera.

Dynamic Light Scattering

Dynamic light scattering measurements were performed on a NanoBrook 90Plus PALS Particle Size Analyzer (Brookhaven, NY), using a 659 nm laser at a 90° detection angle. Samples were prepared in borate buffer and filtered through a 0.2 μ m pore size filter. The refractive index used for the particles was 1.5, but no significant differences were observed when changing it from 1.4 to 1.6. A set of 10 repeats

were recorded for each sample.

Fluorescent Probe Measurements

A borate buffer solution containing Nile Red (15 μM) was titrated with a concentrated solution of precursors (4 mM in building block, 85% oxidized). After each addition of precursors, the sample was homogenized by immersing it in an ultrasound bath for 1 min, and its fluorescence spectra were recorded using a JASCO FP6200 fluorimeter ($\lambda_{\text{exc}} = 553 \text{ nm}$). The titration was repeated three times, and we measured in each of them the point when the fluorescence band started blue-shifting and increasing in intensity.³⁸

Atomistic Molecular Dynamics Simulations

Atomistic simulations were performed using the GROMOS 54a8 united atom force field,^{39,40} as described in detail in ref 30. A fully equilibrated fiber of 12 stacked hexameric macrocycles was created by simulating the fiber in water for 50 ns while using harmonic distance restraints between neighboring C_α atoms (force constant 1000 kJ/nm^2 , equilibrium distance 0.48 nm) and further simulation for 50 ns without these restraints. 200 binding simulations were performed using the following procedure: A single macrocycle configuration was randomly extracted from a separate 50 ns simulation of a single macrocycle in excess aqueous solvent. The selected macrocycle was then inserted in the simulation box containing the equilibrated fiber, in a randomized rotation at a distance of approximately 3.5 nm away from the surface of the fiber. The system was solvated, neutralized using chloride ions, and energy-minimized for 5000 steepest descent steps. The single macrocycle was allowed to diffuse and/or bind to the fiber during a 60.2 ns simulation with 2.0 fs time-steps while the fiber was maintained at its original position along the z-axis by means of rototranslational center of mass motion removal (software extension developed in-house⁴¹). No restraints were applied to keep the structure stable during production runs. 0.2 ns of simulation time was discarded as the equilibration period after which the density of all macrocycle atoms was averaged in blocks of 2.5 ns/25 frames to generate time-dependent density plots. Atomistic simulations were run using GROMACS 4.6.7.⁴² Explicit aromatic or polar hydrogens were converted to virtual sites, and all bonds were constrained in production runs

using the LINCS algorithm,⁴³ except for the SPC water,⁴⁴ which was constrained using the efficient SETTLE algorithm.⁴⁵ Center-of-mass motion was removed every 100 timesteps. The production runs were performed in the NPT ensemble with the velocity-rescaling thermostat⁴⁶ ($\tau_T = 1.0$ ps, separate coupling for solute and water+ions) and the Berendsen barostat⁴⁷ ($\tau_p = 1.5$ ps) while the temperature was kept at 298 K and the pressure at 1.0 bar, respectively. A Barker–Watts reaction field ($\epsilon_{RF} = 62$) was used to treat electrostatic interactions with Coulomb and van der Waals forces cut off at 1.4 nm.

Coarse-Grained Molecular Dynamics Simulations

CG molecular dynamics simulations were performed using the Martini force field v. 2.2.^{48,49} Parameters for the dithiobenzene group were derived from the atomistic simulations by matching bond, angle, dihedral, and nonbonded distributions. Previous work has demonstrated a random coil secondary peptide structure for macrocycles in solution, while fibers exhibit high β -sheet content.³⁰ As such, separate parameters were used for the peptide parts of the macrocycles in the fiber and in solution. In the fiber, the parameters were taken as β -sheet parameters with extended dihedrals from the standard Martini protein parameters,³¹ while for the single macrocycle the coil parameters were used. A fiber of 16 stacked hexamer macrocycles was constructed. The structure was solvated in a box of $10.8 \times 11.7 \times 14.5$ nm, and counterions (96 Na^+ , 192 Cl^- ; 261 mM) were added. 10% of the water beads were replaced with Martini “anti-freeze” particles to avoid possible freezing of the water in the confined geometry of the simulation box. The system was equilibrated for 85 ns with 0.52 nm distance restraints with a force constant of 100 kJ/nm₂ between backbone beads of neighboring peptides. Afterward, the fiber was simulated for 1 μs without distance restraints. Separately, a single trimeric macrocycle was solvated in a box of $6.4 \times 5.1 \times 5.8$ nm, together with counterions (3 Na^+ , 6 Cl^- ; 79 mM) and 10% “antifreeze” particles. The system was equilibrated for 75 ns, before a 1 μs production simulation. Four hundred binding simulations were performed. They were set up by taking a random frame from the 1 μs fiber simulation and a random frame from the 1 μs macrocycle simulation. The macrocycle was inserted in the box of the fiber at a random place in the XY plane at the middle for the fiber at 2–2.5 nm from the fiber surface. The original solvent was removed, the resulting structure was

resolvated, and counterions were added (99 Na⁺, 198 Cl⁻; 269 mM) together with 10% “anti-freeze” particles. The system was equilibrated for 5 ns with 100 kJ/nm₂ position restraints on the backbone beads of both the fiber and the macrocycle. The system was then simulated for 500 ns without restraints. To generate time-dependent density plots, the density of all macrocycle atoms was averaged in blocks of 2.5 ns/25 frames. Coarse-grained simulations were performed using GROMACS versions 5.1 and 2018 (ref 48). In all cases, the fiber was maintained at its original orientation and position along the z-axis by means of rototranslational center of mass motion removal as for the atomistic simulations. A time step of 10 fs was used. The production runs were performed in the NPT ensemble with the velocity-rescaling thermostat⁴⁵ ($\tau_T = 1.0$ ps) and the Parrinello–Rahman barostat⁵⁰ ($\tau_p = 36$ ps) keeping the temperature at 298 K and the pressure at 1.0 bar, respectively. Other simulation parameters used are described by De Jong et al.⁵¹

ACKNOWLEDGEMENTS

S. Maity performed the AFM experiments and data analysis on samples prepared by J. Ottel. G. M. Santiago performed DLS and Nile Red measurements. J. Ottel and G. M. Santiago performed UPLC and fluorescent kinetic analysis. G. M. Santiago and M. C. A. Stuart performed electron microscopy measurements. P. Kroon performed coarse-grain MD simulations. P. W. J. M. Frederix performed atomistic MD simulations. O. Markovitch performed kinetic modelling and data-fitting. The first draft was written by J. Ottel and S. Maity, which was subsequently edited by all other authors. S. Maity, J. Ottel, S. Otto and W. Roos conceived the project.

BIBLIOGRAPHY

1. Aida, T. Meijer, E. W. Stupp, S. I. Functional supramolecular polymers. *Science*, **2012**, 335, 813–817.
2. Zhang, S. Emerging biological materials through molecular self-assembly. *Biotechnol. Adv.* **2002**, 20, 321–339.
3. Grzelczak, M. Vermant, J. Furst, E. M. Liz-Marzan, L. M. Directed self-assembly of nanoparticles. *ACS Nano*, **2010**, 4, 3591–3605.
4. Mattia, E. Otto, S. Supramolecular systems chemistry. *Nat. Nanotechnol.* **2015**, 10, 111–119.
5. Van Esch, J. H. Klajn, R. Otto, S. Chemical systems out of equilibrium. *Chem. Soc. Rev.* **2017**, 46, 5474–5475.
6. Kumar, M. Ing, N. L. Narang, V. Wijerathne, N. K. Hochbaum, A. I. Ulijn, R. V. Amino-acid-encoded biocatalytic selfassembly enables the formation of transient conducting nanostructures. *Nat. Chem.* **2018**, 10, 696–703.
7. Ter Huurne, G. M. De Windt, L. N. J. Liu, Y. Meijer, E. W. Voets, I. K. Palmans, A. R. A. Improving the folding of supramolecular copolymers by controlling the assembly pathway complexity. *Macromolecules*, **2017**, 50, 8562–8569.
8. Fukui, T. Kawai, S. Fujinuma, S. Matsushita, Y. Yasuda, T. Sakurai, T. Seki, S. Takeuchi, M. Sugiyasu, K. Control over differentiation of a metastable supramolecular assembly in one and two dimensions. *Nat. Chem.* **2017**, 9, 493–499.
9. Onogi, S. Shigemitsu, H. Yoshii, T. Tanida, T. Ikeda, M. Kubota, R. Hamachi, I. In situ real-time imaging of self-sorted supramolecular nanofibres. *Nat. Chem.* **2016**, 8, 743–752.
10. Da Silva, R. M. P. Van Der Zwaag, D. Albertazzi, L. Lee, S. S. Meijer, E. W. Stupp, S. I. Super-resolution microscopy reveals structural diversity in molecular exchange among peptide amphiphile nanofibres. *Nat. Commun.* **2016**, 7, 11561.
11. Ando, T. Uchihashi, T. Scheuring, S. Filming biomolecular processes by

high-speed atomic force microscopy. *Chem. Rev.* **2014**, 114, 3120–3188.

12. Kodera, N. Daisuke, Y. Ryoki, I. Ando, T. Video imaging of walking myosin V by high-speed atomic force microscopy. *Nature*, **2010**, 486, 72–76.

13. Colom, A. Redondo-Morata, L. Chiaruttini, N. Roux, A. Scheuring, S. Dynamic remodeling of the dynamin helix during membrane constriction. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, 114, 5449–5454.

14. Maity, S. Caillat, C. Miguët, N. Sulbaran, G. Effantin, G. Schoehn, G. Roos, W. H. Weissenhorn, W. VPS4 triggers constriction and cleavage of ESCRT-III helical filaments. *Sci. Adv.* **2019**, 5, No. eaau7198.

15. Fukui, T. Uchihashi, T. Sasaki, N. Watanabe, H. Takeuchi, M. Sugiyasu, K. Direct observation and manipulation of supramolecular polymerization by high-speed atomic force microscopy. *Angew. Chem., Int. Ed.* **2018**, 57, 15465–15470.

16. Milhiet, P. E. Yamamoto, D. Berthoumieu, O. Dosset, P. le Grimmellec, C. Verdier, J. M. Marchal, S. Ando, T. Deciphering the structure, growth and assembly of amyloid-like fibrils using high-speed atomic force microscopy. *PLoS One*, **2010**, 5, No. e13240.

17. Katsuda, Y. Wickham, S. F. J. Bath, J. Hidaka, K. Sugiyama, H. Turberfield, A. J. Endo, M. Direct observation of stepwise movement of a synthetic molecular transporter. *Nat. Nanotechnol.* **2011**, 6, 166–169.

18. Carnall, J. M. A. Waudby, C. A. Belenguer, A. M. Stuart, M. C. A. Peyralans, J. J.-P. Otto, S. Mechanosensitive self-replication. *Science* **2010**, 327, 1502–1507.

19. Colomb-Delsuc, M. Mattia, E. Sadownik, J. W. Otto, S. Exponential self-replication enabled through a fibre elongation/breakage mechanism. *Nat. Commun.* **2015**, 6, 7427.

20. Rubinov, B. Wagner, N. Rapaport, H. Ashkenasy, G. Selfreplicating amphiphilic β -sheet peptides. *Angew. Chem., Int. Ed.* **2009**, 48, 6683–6686.

21. Nowak, P. Colomb-Delsuc, M. Otto, S. Li, J. TemplateTriggered Emergence of a Self-Replicator from a Dynamic Combinatorial Library. *J. Am. Chem. Soc.*

2015, 137, 10965–10969.

22. Komaromy, D. Tezcan, M. Schaeffer, G. Marić, I. Otto, S. Effector-triggered self-replication in coupled subsystems. *Angew. Chem. Int. Ed.* **2017**, 56, 14658–14662.

23. Ruiz-Mirazo, K. Briones, C. De La Escosura, A. Prebiotic systems chemistry: New perspectives for the origins of life. *Chem. Rev.* **2014**, 114, 285–366.

24. Lancet, D. Zidovetzki, R. Markovitch, O. Systems protobiology: Origin of life in lipid catalytic networks. *J. R. Soc., Interface*, **2018**, 15, 20180159.

25. Otto, S. Furlan, R. L. E. Sanders, J. K. M. Dynamic combinatorial libraries of macrocyclic disulfides in water. *J. Am. Chem. Soc.* **2000**, 122, 12063–12064.

26. Frederix, P. W. J. M. Patmanidis, I. Marrink, S. J. Molecular simulations of self-assembling bio-inspired supramolecular systems and their connection to experiments. *Chem. Soc. Rev.* **2018**, 47, 3470–3489.

27. Bochicchio, D. Pavan, G. M. Molecular modelling of supramolecular polymers. *Adv. Phys. X*, **2018**, 3, 1436408.

28. Bochicchio, D. Salvalaglio, M. Pavan, G. M. Into the dynamics of a supramolecular polymer at submolecular resolution. *Nat. Commun.* **2017**, 8, 147.

29. Brotzakis, Z. F. Gehre, M. Voets, I. K. Bolhuis, P. G. Stability and growth mechanism of self-assembling putative antifreeze cyclic peptides. *Phys. Chem. Chem. Phys.* **2017**, 19, 19032–19042.

30. Frederix, P. W. J. M. Ide, J. Altay, Y. Schaeffer, G. Surin, M. Beljonne, D. Bondarenko, A. S. Jansen, T. L. C. Otto, S. Marrink, S. Structural and spectroscopic properties of assemblies of selfreplicating peptide macrocycles. *ACS Nano*, **2017**, 11, 7858–7868.

31. De Jong, D. H. Singh, G. Bennett, W. F. D. Arnarez, C. Wassenaar, T. A. Schafer, L. V. Periole, X. Tieleman, D. P. Marrink, S. J. Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theory Comput.* **2013**, 9, 687–697.

32. Törnquist, M. Michaels, T. C. T. Sanagavarapu, K. Yang, X. Meisl, G. Cohen,

S. I. A. Knowles, T. P. J. Linse, S. Secondary nucleation in amyloid formation. *Chem. Commun.* **2018**, 54, 8667–8684.

33. Heath, G. R. Scheuring, S. High-speed AFM height spectroscopy reveals μ s-dynamics of unlabeled biomolecules. *Nat. Commun.* **2018**, 9, 4983.

34. Scheidt, T. Łapinska, U. Kumita, J. R. Whiten, D. R. Klenerman, D. Wilson, M. R. Cohen, S. I. A. Linse, S. Vendruscolo, M. Dobson, C. M. Knowles, T. P. J. Arosio, P. Secondary nucleation and elongation occur at different sites on Alzheimer's amyloid- β aggregates. *Sci. Adv.* **2019**, 5, No. eaau3112.

35. Uchihashi, T. Kodera, N. Ando, T. Guide to video recording of structure dynamics and dynamic processes of proteins by highspeed atomic force microscopy. *Nat. Protoc.* **2012**, 7, 1193–1206.

36. Ando, T. Kodera, N. Naito, Y. Kinoshita, T. Furuta, K. Toyoshima, Y. Y. A High-speed Atomic Force Microscope for Studying Biological Macromolecules in Action. *ChemPhysChem*, **2003**, 4, 1196–1202.

37. Bertin, A. de Franceschi, N. de la Mora, E. Maity, S. Alqabandi, M. Miguet, N. di Cicco, A. Roos, W. H. Mangenot, S. Weissenhorn, W. Bassereau, P. Human ESCRT-III polymers assemble on positively curved membranes and induce helical membrane tube formation. *Nat. Commun.* **2020**, 11, 2663.

38. Stuart, M. C. A. Van De Pas, J. C. Engberts, J. B. F. N. The use of Nile Red to monitor the aggregation behavior in ternary surfactant water-organic solvent systems. *J. Phys. Org. Chem.* **2005**, 18, 929–934.

39. Reif, M. M. Hünenberger, P. H. Oostenbrink, C. New interaction parameters for charged amino acid side chains in the GROMOS force field. *J. Chem. Theory Comput.* **2012**, 8, 3705–3723.

40. Reif, M. M. Winger, M. Oostenbrink, C. Testing of the GROMOS force-field parameter set 54A8: Structural properties of electrolyte solutions, lipid bilayers, and proteins. *J. Chem. Theory Comput.* **2013**, 9, 1247–1264.

41. GROMACS RTC code at <https://github.com/Tsjerk/gromacs/tree/rtc>.

42. Hess, B. Kutzner, C. Van Der Spoel, D. Lindahl, E. GRGMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J.*

Chem. Theory Comput. **2008**, 4, 435–447.

43. Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **2008**, 4, 116–122.

44. Berendsen, H. J. C. Postma, J. P. M. van Gunsteren, W. F. Hermans, J. Interaction models for water in relation to protein hydration. In *Intermolecular Forces* Pullman, B., Ed. Springer: Dordrecht, **1981** Vol. 14, pp 331–342.

45. Miyamoto, S. Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, 13, 952–962.

46. Bussi, G. Donadio, D. Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, 126, 014101.

47. Berendsen, H. J. C. Postma, J. P. M. Van Gunsteren, W. F. Dinola, A. Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, 81, 3684–3690.

48. Marrink, S. J. Risselada, H. J. Yefimov, S. Tieleman, D. P. De Vries, A. H. The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, 111, 7812–7824.

49. Abraham, M. J. Murtola, T. Schulz, R. Pall, S. Smith, J. C. Hess, B. Lindah, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1, **2015**, 2, 19–25.

50. Parrinello, M. Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, 52, 7182–7190.

51. De Jong, D. H. Baoukina, S. Ingolfsson, H. I. Marrink, S. J. Martini straight: Boosting performance using a shorter cutoff and GPUs. *Comput. Phys. Commun.* **2016**, 199, 1–7.

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY VIDEO DESCRIPTIONS:

Supplementary Video 1. Growth of a fibre captured by HS-AFM.

The video shows the growth of a fibre weakly immobilized on a membrane bilayer. Imaging rate is 0.5 frame/second. For detailed description see **Figure S2 A-F** and the main text. The video can be accessed through this link: tinyurl.com/3hypryms

Supplementary Video 2. No observable growth of fibre with precursors attached in the middle as captured by HS-AFM.

The video captured a fibre immobilized on a membrane bilayer with precursors attached in the middle. Imaging rate is 2 frame/second. For detailed description see **Figure S2 G-I** and the main text. The video can be accessed through this link: tinyurl.com/yy4zy967

Supplementary Video 3. Growth of a fibre, accumulation and diffusion of precursor aggregate captured by HS-AFM.

The video shows the pathway of fibre growth capture for a long period of time (1930 s). Imaging rate 0.5 frame/second. For detailed description see **Figure 3** and the main text. The video can be accessed through this link: tinyurl.com/5348utw3

Supplementary Video 4. Precursors diffusivity on a fibre surface captured by HS-AFM.

The video shows growth of a fibre without mask (left), and with mask (right) used for the height of a fibre (3.5 nm). The black dots represent all protrusions above 3.5 nm. Imaging rate 0.5 frame/second. For more detailed description see **Figure S3**. The video can be accessed through this link: tinyurl.com/d8h2uk8w

Supplementary Video 5. Growth of a fibre captured by HS-AFM at an extended time.

The video shows the growth of a fibre weakly immobilized on a

membrane bilayer at a longer time frame (after 32 minutes). Imaging rate is 0.5 frame/second. For detailed description see **Figure S8** and the main text. The video can be accessed through this link: tinyurl.com/t8wm8sb4

Supplementary Video 6. GROMOS all-atom simulation results of the diffusion of a precursor on a fibre.

The video shows the normalized density profile of a **1₃** precursor averaged over 200 independent 60 ns simulations. The fibre outline is represented by the black line. The video can be accessed through this link: tinyurl.com/49fkyep6

Supplementary Video 7. Martini coarse grain simulation results of the diffusion of a precursor on a fibre.

The video shows the normalized density profile of a **13** precursor averaged over 400 independent 500 ns simulations. The fibre outline is represented by the black line. The video can be accessed through this link: tinyurl.com/dwvy2zsf

SUPPLEMENTARY DISCUSSION

Supplementary Discussion 1.

The growth rate values measured by fluorescence and UPLC should only be taken as an approximation because of the assumptions made in the conversion from $\mu\text{M}/\text{minute}$ to $\text{nm}/\text{minute}/\text{fibre end}$ (i.e. the calculated number of fibre ends per mM or the width of each hexamer). The experiments shown in **Supplementary Figure 5**, however, were performed simultaneously and under exactly the same conditions, and the above mentioned conversion factor is the same for all results.

The **Supplementary Discussion 2**, which includes mass-action kinetic modelling, data-fitting, **Supplementary Figures 12-18** and **Supplementary Tables 4-6** can be found here: tinyurl.com/uatds74t

SUPPLEMENTARY FIGURES

Supplementary Figure 1: Fibres having precursors attached near the fibre end have the highest frequency of occurrence and grow the most. A) Example of an AFM image of fibres immobilized on a membrane surface at 0 min, in presence of 2.31 mM precursors. B) Histogram of fibre length at 0 min. N= 210 (fibres) from 10 different images. C) Relative frequencies of fibres having precursors attached near the end (green), in the middle (blue) or no precursor attachment (grey). D) Same as A, but after 30 min of incubation on the membrane surface with 2.31 mM precursors in solution. (E, F) Same as B and C, respectively, but after 30 min. N= 234 (fibres) from 10 different images. (G, H, I) Same as A, B and C, respectively, but after 60 min of incubation on the membrane surface with precursors in solution. N= 238 (fibres) from 10 different images. The error bar in panels C, F and I are representing the standard deviation of the distribution. Statistical analysis using the two-sample KolmogorovSmirnov test confirmed that at t=60 min the lengths of fibres with precursors attached to their ends are significantly different from those with precursors attached to their middle (highest pvalue= 4.39×10^{-7} , Supplementary Table 2). Evaluating the elongation of fibres between t=0, 30 and 60 min confirmed that only the group of fibres with precursors attached at their ends grew significantly (highest p-value= 2.41×10^{-7} , Supplementary Table 3), while the group of fibres with precursors attached at the middle and that without precursors did not significantly grow in length (lowest p-value= 5.22×10^{-2} and 7.07×10^{-2} , respectively).

Supplementary Figure 2: The attached precursors participate actively in the growth process. A) Snapshots of fibre growth on a membrane surface at different times. Scale bar 20 nm. B) Image processing for the calculation of relative volume. The first frame (left) represents an example of a raw image. In the second frame (middle) the background is masked. In the third frame (right) the black line indicates the area occupied by the precursors. The selected area was fixed throughout image acquisition. C) Example of a height distribution histogram (bin size of 0.08 nm) of the selected area in panel B (right). D) Kymograph of relative height (in terms of intensity) along the green dotted line in panel A for the duration of the imaging (266 s). E) Change in precursor volume with time from the experiment in panel A. F) Histogram of fibre growth rate for all the growing fibres, calculated every 60 s (N=40). G) Snapshot of AFM images of a fibre with precursors attached in the middle. Scale bar 20 nm. H) Kymograph of relative height along the blue dashed line in panel G. I) Measured relative volume of the attached precursors over time from panel G.

Supplementary Figure 3: Precursor diffusion on a fibre surface favouring directionality in a fibre growth. A) Snapshots of AFM images of a growing fibre on a membrane surface taken at initial time, and after 292 seconds. The left image shows the unmasked fibre, and right image after masking for the fibre height of 3.5 nm. The black dots on the fibre surface represent any protrusion above the fibre surface (>3.5 nm). B) kymograph along the dotted green line in panel A for the period of 292 s. C) Superimposed masked images of the fibre for every 40 s intervals, showing overall diffusion of precursors in every 40 seconds. The red arrow indicates the direction of growth. It can be observed that there is a higher number of black dots (i.e. precursors) at the growing end. Scale bars 20 nm.

Supplementary Figure 4: UPLC analysis of library composition over time with 0.154 mM precursors in presence of 5 % seed, and without mechanical agitation. The data (monomers as circles, trimers and tetramers as squares, and hexamers as triangles) shown is the average of four separate experiments. The calculated growth rate for hexamers from the initial linear part of purple curve is ~1 nm/min.

Supplementary Figure 5: Initial growth rate of fibres in presence of different concentrations of precursors, determined by fluorescence spectroscopy (See Methods). The black squares represent the average of three experiments and the error bars their standard deviation, with the red circles representing each of the individual experiments. The fact that the fibre growth rate levels off at higher precursor concentrations suggests that precursor aggregates form which are off-pathway: i.e. these aggregates do not contribute to fibre growth.

Supplementary Figure 6: Characterization of the aggregates formed by precursors in solution. Cryo-TEM micrographs of samples containing A) precursors and B) 16 fibres and precursors. To aid interpretation, arrows have been added indicating the location of aggregates (red), fibres (orange), and ice contamination (blue).¹⁻³ Scale bars are 20 nm. The insert in A highlights (red circles) typical examples of precursors aggregates. C) A comparative histogram of measured aggregate diameter by TEM, and aggregate height by AFM. The diameter of 100 different aggregates was measured by TEM in different micrographs, giving an average value of 2.6 ± 0.6 nm. By AFM, the aggregates were only visible when they were attached on a fibre surface. The measured height of 63 different aggregates is 6 ± 2 nm, comparatively higher than that measured by TEM. D) Chromatogram representation of

the sample from A. The results depict the major composition of molecules present in the precursors solution. E) Size distribution of a precursor sample, as measured by DLS. The accuracy of DLS is relatively low for the size range measured, so 10 individual measurements were recorded (from 10 repeats each) and averaged, obtaining a diameter of 3 ± 2 nm. F) Determination of the critical aggregation concentration of precursors using Nile Red as a fluorescent probe. The blue shift and increase in intensity of the fluorescence band of Nile Red above a certain concentration corresponds to the formation of a supramolecular structure with a hydrophobic environment.^{4,5}

Supplementary Figure 7: Growth of fibres free in solution in presence of 2.31 mM precursors as periodically observed by AFM. (A-E) Example of AFM images of fibres immobilized on mica after 0, 10, 35, and 60 min of incubation, respectively. The relative position of the precursors on the fibres changes with time. In the 0-10 min period, the aggregates do not show significant differences in position; while after 20 min, a spread in the precursor density can be observed. After 30 min, an accumulation of the precursors towards the fibre end can be seen. After 60 min, the precursors had disappeared as the fibres grew to the full extent. Scale bars are 100 nm. F) Plots of aggregates height vs fibre length as observed at different times. The aggregates height decreases as the

fibre elongates and the precursors spread over the fibre. G) Fibre growth upon incubation with 2.31 mM (final 2.19 mM) precursor solution. Growth rate is ~ 5 nm/min. This number is likely an underestimate as the longer fibres tend to break in smaller pieces upon attachment to the surface.

Supplementary Figure 8: Correlation of diffusivity with the fibre growth. A) Snapshots of HS-AFM images of a growing fibre on a membrane surface after 32 min incubation period. Scale bar 20 nm. B) kymograph of the growth along the dotted green line in panel A. An average growth rate of 3.6 nm/min is estimated from this kymograph. C) Comparative measurements of relative volume over time for the precursors aggregate (reservoir), and the fibre. The slope estimated for the reservoir indicates a decrease in volume at least at a rate of 5 relative volume units/s. It is important to mention that this rate may be an underestimation, as free precursors in the solution can still be absorbed into the existing aggregates during this time period. The slope estimated for the fibre shows an increase in volume at a rate of 1.5 relative volume units/s.

Supplementary Figure 9: Free energy profiles of binding comparing the atomistic and coarse-grained models. Both profiles were obtained using umbrella sampling by pulling one hexamer from a fibre of 8 hexamers. The reaction coordinate for the atomistic simulation was defined as the center of mass (com) distance between neighbouring C-terminal lysines of hexamers 1 and 2, simulating for 10 ns per window. The reaction coordinate for the coarsegrained simulation was defined as the com distance between one of the C-terminal lysines of hexamer 1 and the com of full hexamer 2 (hence the horizontal shift of the curve), simulating for 6 ns per window.

Supplementary Figure 10: Angle of the hexamer as function of position in the fibre in a 1 μ s coarse-grained simulation. To obtain the period of the twist of the fibre a line was fit (orange line) to the found angles (blue dots), excluding the first and last hexamers. This resulted in a slope of -1.75 ± 0.01 $^{\circ}/\text{nm}$ ($R^2=0.9991$). The fit was performed using SymFit.⁶ This amounts to 205 ± 3 nm for a full twist, somewhat overestimating the experimental and atomistic findings.⁷

Supplementary Figure 11: Distributions of angles and (improper) dihedrals in the dimercaptobenzene headgroup. Evidence of CG model accuracy vs atomistic data. Distributions determined based on simulations of a single hexamer in solution. Atomistic reference data is plotted as dots, the corresponding coarse-grained data is plotted as solid lines. Bead notation as in insets, S/SS denote either S1 or S2. The observable mismatch between atomistic and CG model in the BB-SS-SS-BB dihedral distribution at angles below -100° and above 100° is considered acceptable as the main mode is still captured, especially since the macroscopic CG fibre behaviour matches the AA fibre behaviour.

SUPPLEMENTARY TABLES

AFM MEASUREMENTS		UPLC MEASUREMENTS		MEASURED BY FLUORESCENCE SPECTROSCOPY	
Fibres grown on a surface (2.31 mM pre-cursors)	Fibres grown free in solution (2.31 mM pre-cursors)	In presence of 2.31 mM precursors	In presence of 0.154 mM precursors	In presence of 2.31 mM precursors	In presence of 0.100 mM precursors
~5 nm/min	~5 nm/min	~4 nm/min	~1 nm/min	~9 nm/min	~1 nm/min

Supplementary Table 1. Summary of experimentally measured elongation rate of the fibre using different techniques and sampling.

	0 MIN	30 MIN	60 MIN
END VS. MIDDLE	4.54×10^{-1}	3.22×10^{-1}	4.39×10^{-7}
END VS. NONE	3.16×10^{-8}	1.14×10^{-30}	5.71×10^{-37}
MIDDLE VS. NONE	7.41×10^{-4}	5.08×10^{-10}	2.73×10^{-10}

Supplementary Table 2: p-values of statistical tests comparing the observed fibre lengths at specific times. The two-sample Kolmogorov-Smirnov test was used, with null hypothesis that two datasets are from the same continuous distribution and the alternative hypothesis is that they are from different continuous distributions.

	END	MIDDLE	NONE
0 VS. 30 MIN	2.21×10^{-13}	5.22×10^{-2}	2.03×10^{-1}
0 VS. 60 MIN	5.05×10^{-26}	5.04×10^{-1}	7.07×10^{-2}
30 VS. 60 MIN	2.41×10^{-7}	3.16×10^{-1}	9.41×10^{-1}

Supplementary Table 3: p-values when comparing the same group of fibres at different times. The analysis is identical to that in **Supplementary Table 2**.

CHAPTER 4

Chance emergence of catalytic activity and promiscuity in a self-replicator

This chapter has been published: Ottel, J., Hussain, A. S., Mayer, C., Otto, S. Caught in the Act: Chance emergence of catalytic activity and promiscuity in a self-replicator. Nat. Catal. 3, 547-553, 2020

ABSTRACT

How life can emerge from inanimate matter is one of the grand questions in science. Self-replicating molecules are necessary for the transition from chemistry to biology, but they need to acquire additional functions for life to emerge. Catalysis is one of the most essential of such functionalities, but mechanisms through which self-replicators can acquire catalytic and, in extension, metabolic properties have remained elusive. Here we show how catalytic activity and promiscuity in a self-replicator emerges through co-option: features that are selected to benefit replication inadvertently result in an arrangement of chemical functionalities that is conducive to catalysis. Specifically, we report self-assembly driven self-replicators that promote both a model retro-aldol reaction and the cleavage of fluorenylmethoxycarbonyl groups, with the latter transformation exerting a positive feedback on replication (protometabolism). Such chance invention of new function at the molecular level marks a pivotal step toward the de novo synthesis of life.

4.1 INTRODUCTION

It is generally accepted that life requires the integration of replication and metabolism and that the latter relies heavily on catalysis.¹⁻⁴ Hence, catalytically active replicators are an important but, until now, largely missing stepping stone in the origin and de novo synthesis of life.⁵⁻⁷ It is commonly thought that during its emergence, life went through a stage in which the roles of replication and catalysis were shared by a single type of molecule.^{8,9} For example, one of the main arguments in support of the RNA world hypothesis is that RNA can act as catalyst and can replicate. RNA molecules that can perform each of these tasks have been developed, but no synthetic RNA molecules

exist yet that can do both.¹⁰ The central problem seems to be that catalysis requires folding, whereas replication requires an unfolded strand. This raises the question of how catalytically active replicators can spontaneously arise? After more than thirty years of research on self-replicating molecules, the mechanisms that allow the integration of catalysis with self-replication without human intervention remain obscure. In search for a mechanism that allows for the autonomous emergence of catalytic abilities in self-replicators, we noted that the self-assembly of molecules can, in select cases, endow these molecules with new or improved chemical reactivities.^{11–15} By analogy, we surmised that a class of self-assembly driven self-replicators that we developed previously may well exhibit catalytic capabilities. Specifically, we investigated the replication system made from building block **1** (**Figure 1A**), which features a small peptide chain (GLKFK) appended to a benzene-1,3-dithiol core.¹⁶ In summary, in aqueous solution and in the presence of oxygen from the air, this building block oxidizes to form a mixture of uniquely sized disulfide macrocycles—dominated by trimers **1**₃ and tetramers **1**₄—that interconvert through disulfide exchange reactions (**Figure 1B**). After an initial lag phase, the hexamer rings (**1**₆) assemble into stacks through a nucleation-growth mechanism. Assembly of hexamer rings depletes the exchange pool of these rings, which get replenished (although we cannot exclude the possibility that the stack ends also play a catalytic role in the production of additional hexamer rings). Thus, the assembly process drives the synthesis of the very molecules that assemble, amounting to their self-replication. Upon supplying mechanical energy, long stacks are broken into smaller ones, thereby increasing the number of ends from which the stacks grow, enabling exponential replication.¹⁷ We hypothesized that stacking of **1**₆ would bring the lysine side chains appended to these rings into close proximity, thereby lowering the apparent pK_a of some of the ϵ -amines and enhancing their nucleophilicity and Brønsted basicity at neutral pH (**Figure 1C**). Notably, analogous configurations are often found in the active sites of evolved and designed enzymes^{11,18,19}, where ionizable groups are positioned in proximity to catalytic residues to increase their reactivity. We further noted that the ϵ -amines in the replicator fibres are adjacent to hydrophobic leucine and phenylalanine residues. It is well established that placing amines in hydrophobic microenvironments can enhance their catalytic activity.²⁰

Figure 1: Fibre formation from building block **1** creates microenvironments that are conducive to catalysis. A) The molecular structure of building block **1**. The benzene-1,3-dithiol core and the peptide chain are coloured orange and blue, respectively. B) A mechanism of self-replication that is driven by self-assembly of hexamer macrocycles **1₆**. A series of oxidation (1) and disulfide exchange reactions (2) gives rise to **1₆**. Following nucleation (3), stacking of additional **1₆** results in fibre elongation (4), with mechanical energy leading to fragmentation (5), thereby increasing the number of available growing stack ends. C) Fibre assembly brings amines from different building blocks into close proximity (only amines on two of the six stacks of peptides are shown for clarity). Coulombic repulsion perturbs the apparent pK_a of some of these amines (in red), thus increasing their nucleophilicity.

4.2 RESULTS AND DISCUSSION

4.2.1 RETRO-ALDOL CATALYSIS

To probe the catalytic potential of replicator fibres, we first selected the retro-aldol reaction of methodol (**2**, **Figure 2A**) and the cleavage of fluorenylmethoxycarbonyl (Fmoc)-glycine (**4**, **Figure 3A**) as model transformations that may be accelerated by organocatalysis and Brønsted-base catalysis, respectively. The retro-aldol reaction of methodol is catalysed by amines through iminium ion formation, which activates the C–C bond towards cleavage (**Figure 2A**).²¹ The resulting enamine undergoes tautomerization and hydrolysis to regenerate the catalyst, and the cleavage also yields 6-methoxy-2-naphthaldehyde (**3**), which can be readily quantified by ultraviolet–visible spectroscopy or reversed-phase ultra-performance liquid chromatography (UPLC) methods. We observed the rapid formation of retro-aldol product **3** when incubating fibres of **1₆** (50 μM with respect to building block **1**) with methodol (200 μM) in borate buffer (50 mM, pH 8.12) at 25 °C (**Figure 2B**).

Critically, catalysis is an emergent property of the replicator assemblies, as is evident from **Figure 2B**, which shows that the activity of the **1₆** fibres substantially exceeds that of **1** and a mixture of **1₃**/**1₄** at equimolar concentrations of building block (see also **Extended Data Figure 1A,B**). We attribute this difference to the altered microenvironment and higher local positive charge density that is provided by fibres of **1₆**, which induce a larger shift in the apparent lysine pK_as (vide infra) in comparison with the building block or the non-assembled macrocycles. Furthermore, fibres are also more likely to provide a suitable microenvironment and binding pocket for catalysis than the non-assembled molecules.

Figure 2: The catalytic activity of **1₆** in a model retro-aldol reaction emerges due to the combination of a substrate-binding pocket and nucleophilic lysine residues. A) The retro-aldol cleavage of methodol is accelerated in the presence of a nucleophilic amine by the formation of an iminium ion intermediate. The C–C bond cleavage yields **3** and an enamine intermediate, which tautomerizes and hydrolyses in subsequent steps (not shown) to give acetone and regenerate the catalyst. B) At equimolar concentrations (50 μM with respect to **1**), fibres of **1₆** (red) produce **3** more rapidly than a mixture of **1₃/1₄** (blue) or building block **1** (black). Data were obtained by either following the reaction by UPLC (filled circles and single standard deviations) or spectrophotometrically (lines, single standard deviations shown as shaded areas) and are the result of at least two independent

measurements. C) Saturation kinetics for the burst phase (red) and the steady-state phase (blue) for catalysis by **1₆**. Initial rates (v_{obs}) are plotted as the average of three independent experiments with error values displayed as single standard deviation. The velocities that were obtained for both regimes were fitted to the Michaelis–Menten equation (solid lines). D) Kinetic characterization in the case of a large excess of **1₆** fibres (100–1500 μM in **1**) compared with **2** (10 μM). The obtained pseudo-first-order rate constants were fitted to a Michaelis–Menten model and the k_{cat} and K_{M} values are displayed. Data points are the average of at least two independent measurements with single standard deviations shown. E) A pH rate profile obtained for catalysis by **1₆** (40 μM in **1**) of the retro-aldol reaction of methodol (200 μM). F) The emergence of catalytic activity for the retro-aldol reaction in a stirred sample of **1** correlates with fibre formation, as shown by a comparison of the change in the concentration of **1₆** (blue, left axis) with the rate of formation of **3** (red, right axis) with time. All of the error bars shown correspond to a single standard deviation.

To determine the catalytic efficiency of **1₆** replicator fibres, we first performed saturation kinetic studies and observed an initial burst phase, followed by a slower, steady-state rate for methodol consumption (**Extended Data Figure 1C**). Such behaviour has previously been observed in designer enzymes and ascribed to slow release of the enamine intermediate or product rebinding to active sites.²⁰ When varying the concentrations of methodol, both regimes displayed saturation (**Figure 2C**), which is indicative of the presence of a substrate-binding pocket. The kinetic analysis is complicated by the fact that the number of catalytically active lysines (n) is a priori unknown. We followed an established approach to tackle this problem, comparing analyses performed using an excess of substrate to catalyst (**Figure 2C**) to those using an excess of catalyst to substrate (**Figure 2D**).^{22,23} Fitting the data obtained using excess substrate to a Michaelis–Menten model yields a value for the Michaelis constant (K_{M}) and a composite value for the catalytic rate constant (k_{cat}) and n , whereas fitting data at excess catalyst provides k_{cat} and composite terms for K_{M} and n (**Figure 2C,D**). Hence, a comparison between the kinetic parameters obtained under these conditions provides an estimate of the number of active lysines and the catalytic parameters of a single active site (see Supplementary Discussion). On the basis of this comparison, we conservatively estimate that less than 10% of all lysines are catalytically active. This result is in line with recent observations indicating that fibres of **1₆** are relatively poorly ordered²⁴ and are therefore likely to provide a spectrum of microenvironments, of which only some seem catalytically active. For single active sites on **1₆** fibres, K_{M} values of $670 \pm 140 \mu\text{M}$ (burst phase) and 930 ± 160

μM (steady state) attest to an appropriate substrate-binding pocket. Furthermore, we obtained apparent second-order rate constants for the burst (b) and steady-state (ss) phases of

$$k_{\text{cat}}/K_M^b = 0.72 \pm 0.07 \text{ M}^{-1} \text{ s}^{-1} \text{ and } k_{\text{cat}}/K_M^{\text{ss}} = 0.29 \pm 0.02 \text{ M}^{-1} \text{ s}^{-1}.$$

As such, individual active sites on **1**₆ fibres outperform non-assembling macrocycles **1**₃/**1**₄ by a factor of 106 (burst) and 42 (steady state) and provide chemical proficiencies ($1/K_{\text{TS}} = (k_{\text{cat}}/K_M)/k_{\text{uncat}}$) of $1.91 \times 10^7 \text{ M}^{-1}$ and $0.76 \times 10^7 \text{ M}^{-1}$ for the burst and steady-state phases, respectively. Notably, the k_{cat}/K_M values exhibited by the replicator assemblies are comparable with those of the best computationally designed enzymes for the same reaction (**Supplementary Table 1**)^{25–27}, which is remarkable given that the present system emerged spontaneously, essentially unaided by human design or evolutionary optimization.

We next probed whether stacking of **1**₆ units boosts the catalytic prowess of the lysine groups. The pH rate profile of the catalysed reaction shows a single inflection point at $\text{pH } 7.4 \pm 0.1$ (**Figure 2E**), demonstrating a perturbation of the apparent $\text{p}K_{\text{a}}$ of the catalytically active lysine groups by approximately three units in comparison with the $\text{p}K_{\text{a}}$ of lysine in solution. When performing pH titrations of **1**₆ fibres, we were unable to identify a notable change in ionization at $\text{pH } 7.4$ (**Extended Data Figure 2**), which confirms that the fraction of the lysines on the **1**₆ fibres that are catalytically active is small. We synthesized building blocks featuring lysine-to-arginine mutations on either the inner (XGLR_{FK}) or outer (XGLK_{FR}) lysine to further pinpoint the catalytic roles of the two lysine side chains in **1** (**Figure 1A**). Replacing the inner lysine with a positively charged but catalytically inactive arginine side chain resulted in an approximately fivefold loss in activity of the hexamer replicators (which is consistent with the inner lysine being more buried), whereas changing the outer lysine reduced the activity of the corresponding replicator by about 2.5-fold (**Extended Data Figure 1D**). Combined, our kinetic characterization identified that replicator **1**₆ features configurations that are conducive for catalysis, in which catalytically active lysine residues are rare and predominantly buried. We next monitored the correlation between the emergence of the replicator and the emergence of catalytic activity in situ starting from building block **1**. We dissolved **1** ($154 \mu\text{M}$) and **2** ($200 \mu\text{M}$) in borate buffer ($\text{pH } 8.12$) at 40°C , applying mechanical agitation. In parallel, we ran control samples that either lacked **1** or

agitation; the latter preventing fibre breakage, thus suppressing replicator growth. We monitored the concentrations of building block **1**, non-assembling tetramer macrocycle **1₄** and assembling hexamer macrocycles **1₆**, as well as retro-aldol product **3** over time using reversed-phase UPLC analysis (**Figure 2f** and **Extended Data Figure 3**; see *Methods* for details). The increase in the rate of the retro-aldol reaction directly correlates with the emergence of the replicator in agitated samples (**Figure 2F**). In the unstirred control replicator emergence is hampered and virtually no acceleration of the retro-aldol reaction was observed (see **Extended Data Figure 3** for a full analysis).¹⁷ Finally, to confirm hexamer formation entails fibre assembly, we also subjected samples from the experiments shown in **Figure 2F** periodically to transmission electron microscopy (TEM) analysis, which confirmed the appearance of fibres as soon as the concentration of **1₆** started to increase (**Extended Data Figure 4**). These results together demonstrate how chance invention at the molecular level can endow self-replicators with catalytic abilities. Moreover, they constitute a de novo example of a molecule with the concurrent abilities to replicate and to catalyse a chemical reaction that is of a very different nature than the replication reaction.

4.2.2 CATALYSIS OF FMOC CLEAVAGE

A caveat of the retro-aldol reaction is that none of the liberated products are beneficial to the self-replicator, which is a necessity for the emergence of metabolism. To examine the possibility of **1₆** replicator fibres spontaneously forming a protometabolism, we assessed its ability to catalyse the cleavage of **4**. This reaction proceeds through Brønsted-base-catalysed deprotonation, followed by fragmentation into glycine, CO₂ and dibenzofulvene (**5**, **Figure 3A**). Attractive features of this reaction are that it is catalysed by basic amines (for example, activated lysine chains in **1₆**) and that it liberates an alkene (**5**) that is known to promote thiol oxidation.^{28,29} The cleavage of FMOC-glycine by replicator fibres could therefore lead to the accelerated formation of the non-assembled macrocycles **1₃/1₄**. As they are the precursors from which the replicator grows (**Figure 1B**), accelerating their formation should in turn speed up replication (**Figure 3B**). When incubating a solution of **4** (200 μM in borate buffer pH 8.1 containing 5% acetonitrile) with **1₆** fibres (200 μM in **1**) at 25 °C, rapid cleavage of **4** took place, whereas neither **1** nor a mixture of **1₃**

and **1₄** showed notable activity under otherwise identical conditions (**Figure 3C**). As for the retro-aldol reaction, **1₆** fibres displayed saturation when varying the concentration of **4** (Fig. 3d) and the pH rate profile of the catalysed reaction showed a single inflection point at pH 7.1 ± 0.1 (**Extended Data Figure 5A**), which closely matches the pH rate profile observed for methodol cleavage (pH 7.4 ± 0.1). Although poor solubility and stability of **5** prevented us from performing a more detailed kinetic analysis akin to that for the retro-aldol reaction, the absence of a recognizable plateau corresponding to a pH of 7.1 in the pH titration of **1₆** fibres (**Extended Data Figure 2A**) again indicates that only a small fraction of all lysines is catalytically active. Notably, **1₆** fibres ($k_{\text{cat}}/K_{\text{M}} = 6.2 \pm 0.4 \text{ M}^{-1} \text{ s}^{-1}$) outperform a mixture of **1₃** and **1₄** ($k_{\text{app}, 3/4} = 6.4 \pm 0.6 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$; see *Methods*) by almost three orders of magnitude (for the kinetic analysis, we conservatively estimate that each building block in the replicator fibres provides a single active site, with 50% of all lysines therefore being active). Moreover, replicators provide a chemical proficiency ($1/K_{\text{TS}} = (k_{\text{cat}}/K_{\text{M}})/k_{\text{uncat}}$) of at least $2.1 \times 10^8 \text{ M}^{-1}$, which is one order of magnitude higher than that observed for the methodol cleavage.

Consistent with **1₆** fibres featuring a more effective active site architecture that results in the increased transition-state stabilization for the Fmoc cleavage, lysine-to-arginine mutations proved more deleterious than for the retro-aldol reaction (**Extended Data Figure 5B**). Replacing the inner lysine in replicators proved crippling and led to a >300-fold decrease in activity. Installation of an arginine at the outer position yielded a replicator that only retained ~2% of the catalytic efficiency of **1₆**. To test whether formation of **5** leads to the envisioned positive feedback on replication, we incubated Fmoc-glycine (100 μM) with monomer **1** (200 μM) in borate buffer (pH 8.1, 5% acetonitrile) at 40 °C and simultaneously monitored the emergence of **1₆** replicator, its effect on Fmoc-glycine cleavage and the reciprocal effect of the cleavage products on the replication of **1₆** (**Figure 3E**). As for the retro-aldol reaction, the emergence of replicator after approximately 10h coincides perfectly with the onset of Fmoc-glycine cleavage, which is consistent with only **1₆** fibres (but none of the precursors) being catalysts for the reaction.

Figure 3: The emergence of a protometabolism in response to the cleavage of FMOC-glycine by replicator fibres. A) The mechanism of the Brønsted-base-catalysed cleavage of FMOC-glycine. B) The concept of a positive feedback loop, in which the catalysed conversion of **4** (red arrow) by the self-replicator $(\mathbf{1}_6)_n$ yields **5**, which speeds replication (blue arrow) by accelerating the oxidation (purple arrow) of building block **1** (mechanism of replication shown in **Figure 1B**). C) At equimolar concentrations (200 μM with respect to **1**), replicator fibres $\mathbf{1}_6$ (red) accelerate FMOC-glycine cleavage (200 μM), whereas neither building block **1** nor non-assembled macrocycles $\mathbf{1}_3/\mathbf{1}_4$ provide a detectable activity when compared with the uncatalysed reaction (not shown). D) Saturation kinetics for the cleavage of **4** by $\mathbf{1}_6$ fibres (observed rates and single standard deviations of at least three independent measurements are shown). Kinetic parameters were obtained by fitting the data to the Michaelis–Menten equation (solid lines). E) In an agitated sample prepared from **1** (200 μM)

and Fmoc-glycine (100 μ M) in 5% aqueous acetonitrile in borate buffer pH 8.0, the emergence of **1₆** (dark blue circles) coincides with the onset of Fmoc-glycine cleavage (red circles). Notably, in a sample not containing Fmoc-glycine (light blue circles), replicator **1₆** emerges at the same time, but grew considerably slower. The increase in replication efficiency in the presence of Fmoc-glycine is consistent with a positive feedback loop (shown in panel B). F) A comparison of oxidation rates in the absence (light blue) and presence (dark blue) of Fmoc-glycine show a 3.5-fold increase (purple arrow) once **5** is liberated, providing further evidence for the positive feedback loop. All of the error bars shown correspond to a single standard deviation.

Strikingly, a sample that lacked Fmoc-glycine replicators emerged at the same time, but only grew at a rate that is less than half of that in presence of Fmoc-glycine (**Figure 3E**). Similarly, when monitoring the oxidation of **1** in both samples, the Fmoc-containing sample gave rise to a 3.5-times higher oxidation rate in response to formation of **5** (**Figure 3F**), confirming the envisioned positive feedback loop (**Figure 3B**; see **Extended Data Figure 5C,D** for an analogous experiment containing 200 μ M Fmoc-glycine). Control experiments with isolated dibenzofulvene confirmed that this effect is due to this compound promoting the oxidation of **1** (**Extended Data Figure 5C**).

4.3 CONCLUSIONS

In conclusion, a self-assembling self-replicator was found to be able to catalyse not only its own formation, but also two other chemical transformations that are each different in nature from the replication reaction; that is, the systems show catalytic promiscuity. Catalytic function emerges at the level of the replicator assembly and cannot be ascribed to the starting materials nor intermediates in the process of replicator formation. Although the assemblies emerge spontaneously and are selected on the basis of their ability to replicate, our results show that such selection for replication can also inadvertently produce an arrangement that is conducive to catalysis. Specifically, replicators feature a substrate-binding pocket and lysine residues that have enhanced catalytic activity, imposed by the microenvironment generated through the assembly into fibres. As shown for the catalysis of Fmoc-glycine cleavage, the replicators are able to generate molecules that promote the formation of the building blocks needed for replication, thereby accelerating the replication process. Such behaviour can be regarded as the onset of metabolism. These results thus reveal how replication and metabolism, two key charac-

teristics of life, can autonomously become integrated and mark an important advance towards the de novo creation of a minimal form of life.

The present system of self-assembly driven self-replicators has an important advantage over other replicators in that sites for catalytic activity (the sides of the fibres) are distinct from the sites of self-replication (which occurs at the fibre ends). The catalytic promiscuity exhibited by the system is a highly desirable feature in early evolution as it allows for catalysis of different reactions while only having to replicate a single entity. Catalytic promiscuity is also an important pathway for evolutionary inventions³⁰ as it enables a system that evolved for one purpose to be co-opted for others.³¹ Finally, catalytic promiscuity has also been identified as a desirable evolutionary property in the metabolically coupled replicator system model³², which provides a solution to overcome Eigen's paradox (which states that accurate replication needs complex machinery, yet, obtaining such complex self-replicators through evolution requires sufficiently accurate replication). No efforts have yet been made experimentally to address Eigen's paradox, but with protometabolic replicators showing catalytic promiscuity, a path towards overcoming this important hurdle in the development of life is starting to be unveiled.

METHODS

General procedures

All reagents, solvents and buffer salts were obtained from commercial suppliers and used without further purification, unless otherwise noted. Building block **1**, XGLKFR and XGLRFK were obtained from Cambridge Peptides. Borate buffer concentrations are reported as the concentration of boron atoms. UPLC analysis was performed on a Waters Acquity UPLC-H class system equipped with a photodiode array detector. All analyses were performed using a reversed-phase UPLC column (Aeris Peptide 1.7 μm XB-C18 $\times$ 2.10mm, Phenomenex). The column temperature was kept at 35 °C and the sample plate was kept at 25 °C unless otherwise specified. Ultraviolet absorbance was monitored at 254 and 313 nm. Samples were separated with a gradient from 28–40% acetonitrile (0.1% trifluoroacetic acid) in water (0.1% trifluoroacetic acid) at a flow rate of 0.3 mL min⁻¹ (see **Supplementary Table 2** for details on the method used). A Thermo-Fisher com-

compact dry bath with a custom-milled aluminium block to accommodate for high-performance liquid chromatography (HPLC) vials (with a dimension of $12 \times 32\text{mm}$) was used to heat the samples beyond 40°C . Ultra-performance liquid chromatography–mass spectrometry experiments were performed on a Waters Acquity UPLC-H-class system coupled with a Waters Xevo-G2 TOF. The mass spectrometer was operated in positive electrospray ionization mode. Capillary, sampling cone and extraction cone voltages were kept at 2.5 kV, 30 V and 4 V, respectively. Source and desolvation temperatures were set at 140°C and 500°C . Nitrogen was used as both a cone (5 L h^{-1}) and desolvation gas (500 L h^{-1}). Reaction mixtures were analysed by direct injection of the sample ($5\text{ }\mu\text{L}$). Transmission electron microscopy was performed on a Philips CM120 electron microscope operating at 120 kV. Images were recorded on a slow-scan charge-coupled device camera (Gatan). Spectrophotometry measurements were performed on a Jasco V-660 spectrophotometer at 25°C unless otherwise noted. Methodol was synthesized using a previously published method.³³ All concentrations involving **1**₃, **1**₄ and **1**₆ are given with respect to building block **1**, unless stated otherwise. A mixture of **1**₃/**1**₄ was obtained by oxidizing **1** with sodium perborate (0.95 equiv.) in borate buffer (50 mM, pH 8.12). The resulting mixture after oxidation was analysed by UPLC and mass spectrometry (**Supplementary Figures. 1 and 2**). Fmoc-glycine (**4**, Fmoc-Gly-OH, 97%) was obtained from Sigma-Aldrich. Dibenzofulvene (**5**) was synthesized according to a known literature procedure.³⁴

Preparation of **1**₆

A stock solution of **1**₆ fibres was typically prepared from building block **1** as follows. In brief, building block **1** (2.28 mg, $2.31\text{ }\mu\text{mol}$) was added to an HPLC vial ($12 \times 32\text{mm}$) containing a Teflon-coated magnetic stirring bar ($5 \times 2\text{mm}$, VWR). The sample was dissolved in 1.5 mL borate buffer (50 mM, pH 8.12) to give a final concentration of 1.54 mM **1** and the vial was subsequently closed with a Teflon septum screw cap. The mixture was then stirred at 1200 rpm at 40°C for >72h. The composition of the reaction mixture was periodically analysed by UPLC (10 μL injection) and stirred until **1**₆ accounted for >90% of the mixture (see **Supplementary Figures 1 and 2** for a representative UPLC chromatogram and mass spectra of the final mixture). The resulting stock solution was kept at 25°C while stirring (1200 rpm) and could be used for at least twelve weeks without observing

any notable changes in the sample composition. Hexamer replicators for XGLKFR and XGLRFK were obtained following the same protocol.

Initial assessment of the activity of **1**₆ in catalysing the retro-aldol reaction.

Acetonitrile (10 μ l) and a stock solution of methodol (10 μ l, 20 mM in acetonitrile) were added to 948 μ l of borate buffer (50 mM, pH 8.12) and the resulting mixture was incubated for 10 min at 25 °C. At this point, 32.5 μ l of a stock solution of **1**₆ (1.54 mmol in the same borate buffer) was added to yield a final concentration of 50 μ M **1**₆ (with respect to building block **1**), 200 μ M methodol and 2% acetonitrile. The progress of the reaction was either monitored spectrophotometrically ($\epsilon_{350} = 5,870 \text{ M}^{-1} \text{ cm}^{-1}$) at 25 °C or through periodic analysis of the reaction mixture by UPLC at the same temperature (10 μ l injection of the mixture, 18 min intervals). For the latter, the concentration of **3** was quantified by comparison with a calibration curve that was obtained by injecting known quantities of **3** in the reaction buffer ($\lambda_{\text{max}} = 313 \text{ nm}$, **Supplementary Figure 3**). As control, the experiment was carried out in presence of **1** as well as a mixture of **1**₃/**1**₄ under otherwise identical conditions. The data presented in **Figure 2b** represents the average of at least two independent experiments.

Effect of **1** and a mixture containing **1**₃ and **1**₄ on the retro-aldol reaction

Reaction mixtures containing methodol (final concentration 200 μ M) and acetonitrile (2%) were prepared and incubated for 10 min at 25 °C as described above. To determine the kinetic parameters for **1**, either 66.7 μ l or 200 μ l of a stock solution of **1** (3.08 mM in borate buffer) was added to give a final concentration of 200 μ M or 600 μ M **1**, respectively; 10 μ l of a stock solution of dithiothreitol (200 mM, 2 mM final concentration) was added to avoid oxidation of **1** to small, non-assembling macrocycles. The resulting reaction mixtures were incubated at 25 °C and periodically analysed by UPLC (10 μ l injections, 54 min intervals for 10 h). The analysed retro-aldol reaction was measured in parallel by replacing the stock solution of **1** with borate buffer (pH 8.12). The concentration of **3** was quantified by comparing the peak area at 313 nm to the calibration curve shown in **Supplementary Figure 3**. The reaction progress for each reaction is dis-

played in **Extended Data Figure 1** and demonstrates that **1** does not lead to appreciable levels of acceleration of this retro-aldol reaction at these concentrations. Kinetic characterization of a freshly prepared mixture of **1**₃/**1**₄ was performed in a similar fashion; 16.2, 32.5 or 65 μL of a stock solution of **1**₃/**1**₄ (1.54 mM with respect to building block **1** in borate buffer pH 8.12) was added to a sample containing methanol (200 μM) and acetonitrile (2%) in borate buffer (pH 8.12), giving total a total concentration of 25, 50 or 100 μM (with respect to building block **1**). Reaction mixtures were monitored for 90 min through periodic analysis by UPLC (10 μL injections, 18min intervals) at 25 °C. The concentration of **3** was quantified as specified above. Observed second-order rate constants at all three concentrations were measured by determining the initial rate of the reaction and were comparable for the **1**₃/**1**₄ mixture at all concentrations tested ($8.25 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ (25 μM), $7.41 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ (50 μM) and $8.29 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ (100 μM)), which is indicative of the reaction rate being first order with respect to the catalyst. An apparent second-order rate constant was determined by averaging the obtained values.

Kinetic characterization of the effect of the **1**₆ fibres on the retro-aldol reaction

Excess substrate to catalyst. This kinetic characterization of **1**₆ fibres was performed by measuring reaction progress spectrophotometrically ($\epsilon_{350} \text{ nm} = 5,870 \text{ M}^{-1} \text{ cm}^{-1}$) at varying concentration of **2** (75, 100, 200, 300, 400 and 600 μM) at 25 °C. Samples were prepared by adding appropriate amounts of acetonitrile and a stock solution of **2** (20 mM in acetonitrile) into 1 cm path-length cuvettes containing 948 μL borate buffer (50 mM, pH 8.12). Reaction mixtures were incubated for 10min at 25 °C, at which point 16.2 μL of a stock solution of **1**₆ (1.54mM with respect to building block **1** in borate bufer, final concentration 25 μM with respect to building block **1**) was added. Reaction progress was followed for 250 min at 25 °C. Observed rates for the burst phase (0 – 20 min) and the steady state (100 – 200 min) were extracted by determining the slope of product formation over time for the indicated time frames. Te measured initial rates (v_0) were corrected for the background rate and the resulting values (which represent the average of at least two independent measurements) were fitted (RStudio) to equation (1), in which $[\mathbf{1}] = 2.5 \times 10^{-5} \text{ M}$:

$$\frac{v_0}{[1]} = \frac{k_{cat}[2]}{K_M + [2]} \quad (\text{Eq 4.1})$$

Excess catalyst to substrate. A stock solution of **2** in acetonitrile (0.50 mM) was prepared and kept at -21°C . The solution was used no later than one week after preparation. Stock solutions of **1₆** fibres (2.0 mM with respect to **1**) were prepared as described before, and used within two weeks after the library composition was $>90\%$ **1₆**. For the kinetic characterization, varying volumes of **1₆** fibres (30 – 450 μl) were added to HPLC vials (12 $\times$ 32 mm) containing borate buffer (50 mM, pH 8.12, 138 – 558 μl) and 12 μl of the stock solution of **2**, giving a total volume of 600 μl . The final concentrations of **1₆** ranged from 100 – 1500 μM with respect to building block **1**, with a constant concentration of **2** (10 μM). The formation of **3** was monitored by periodic analysis of the reaction mixture by UPLC (10 μl injections, 18 min intervals for 128 min). From the obtained results, pseudo-first-order rate constants were calculated for different concentrations of **1₆**. Averages of values from least two independent measurements were subsequently fitted (RStudio) to equation (2), in which k_{obs} is the measured pseudo-first-order rate constant and $[1]$ varies from 1×10^{-4} to 1.5×10^{-3} M:

$$k_{obs} = \frac{k_{cat}[1]}{K_M + [1]} \quad (\text{Eq 4.2})$$

A pH rate profile of the effect of **1₆** on the retro-aldol reaction

Buffer solutions (50 mM) were prepared that spanned the pH 6.0 – 8.75 range (phosphate buffer pH 6.3 – 7.1; borate buffer pH 7.4 – 8.5). Samples were prepared by adding acetonitrile (10 μl) and 10 μl of a stock solution of methodol (20 mM in acetonitrile) into 1 cm

cuvettes containing 954 μl of the specific buffer solution. The resulting mixtures were incubated for 10 min at 25°C before addition of a stock solution of **1₆** (26 μl , 1.54 mM with respect to building block **1** in borate buffer, pH 8.12) to obtain a final concentration of 40 μM **1₆** (with respect to building block **1**), 200 μM methodol and 2% acetonitrile. The reaction progress was monitored spectrophotometrically ($\epsilon_{350} = 5,870 \text{ M}^{-1} \text{ cm}^{-1}$) at 25°C for 250 min. Observed rates for the steady state (100–250 min) were extracted by analysing the data for

the product formation over time. The measured rates were corrected for the background and $(k_{cat}/K_M)_{app}$ values (apparent catalytic efficiencies) were determined by dividing the corrected rates by the concentrations for **1** (4×10^{-5} M) and substrate (2×10^{-4} M). Averages of values from least two independent measurements were subsequently fitted (RStudio) to equation (3):

$$\left(\frac{k_{cat}}{K_M}\right)_{app} = \frac{\left(\frac{K_{cat}}{K_M}\right)_{max}}{1 + 10^{pK_a - pH}} \quad (\text{Eq 4.3})$$

pH titration of **1** and **1₆** fibres

Building block **1** (3.21 mg, 3.25 μ mol in **1**) was added to a glass vial (15 $\times$ 45 mm) containing a Teflon-coated magnetic stirring bar (5 $\times$ 2 mm, VWR) to prepare a stock solution of **1** and **1₆** fibres in H₂O. UP-LC-grade H₂O (2.1 mL) was added to give a final concentration of 1.54 mM. To enable disulfide formation and exchange, which is necessary for **1₆** formation, 1 M NaOH (4 μ L) was added to one of the samples, resulting in a pH of 8.2; this vial was closed with a screw cap and the mixture was stirred at 1200 rpm at 45 °C for 7 d. The library composition was periodically analysed by UPLC (10 μ L injections). At the time, **1₆** accounted for >90% of the mixture, the stock solution was cooled to 25 °C and continued to be stirred at 1200 rpm. This stock solution could be used for up to twelve weeks without observing a notable change in the sample composition, whereas the stock solution containing **1** was prepared freshly before the pH titration. For the pH titration, stock solutions for **1** and **1₆** fibres (1 mL each) were diluted with UPLC-grade H₂O (1 mL) to obtain a solution containing 1.54 μ mol of **1** and **1₆**; 1 M HCl (2.5 μ L) was added to the latter to lower the pH to ~3.6, similar to the pH of the freshly prepared solution of **1**. Small aliquots (1 – 5 μ L) of a 0.1 or 0.5 M NaOH solution were then added to the samples and the pH monitored using a Hanna HI-4221 research-grade bench pH and ORP meter. Measurements were performed in duplicates and are summarized in **Extended Data Figure 1**.

Experiments corresponding to Figure 2F and Extended Data Figure 3

Borate buffer (880 μ L, 50 mM, pH 8.12) was added to an HPLC vial

(12 × 32 mm) containing a Teflon-coated magnetic stirring bar (5 × 2 mm, VWR). The sample was heated to 40 °C; acetonitrile (10 µL) and a stock solution of methodol (10 µL, 20 mM in acetonitrile) were added to the preheated buffer. The sample was incubated at 40 °C for 15 min; 100 µL of a freshly prepared solution of **1** (1.54 mM in 50 mM borate buffer, pH 8.12) was subsequently added to the mixture to give a total concentration of 154 µM **1**, 200 µM methodol and 2% acetonitrile. The reaction mixture was kept at 40 °C and stirred at 500 rpm for the entire duration of the experiment (30 h). The reaction was monitored by periodic analysis of the reaction mixture by UPLC (10 µL injection). The concentrations of **1**, **1₄** and **1₆** were quantified by comparison to a calibration curve that was obtained by injecting known quantities of **1** in the reaction buffer (**Supplementary Figure 3**). The concentration of **3** was quantified by comparison to a calibration curve, as described above. Control experiments were performed by either not applying any mechanical agitation or by omitting **1** from the sample. Representative chromatograms recorded near the start and the end of the experiment are depicted in **Supplementary Figure 4**.

Time-dependent TEM

Samples were prepared as described before for the experiments corresponding to **Figure 2**. Fibre formation was followed by periodically preparing TEM samples from the reaction mixtures (for example, for the agitated samples at $t = 0, 1, 6, 7.5, 8.5, 10, 13, 24$ and 48 h). For this, a small drop (5 µL) of the reaction mixtures was deposited onto a 400 mesh copper grid covered with a thin carbon film (Agar Scientific). After 1 min, the grid was blotted on filter paper; the sample was subsequently stained twice with uranyl acetate by depositing 5 µL of a saturated uranyl acetate solution onto the grid and blotting it on filter paper after incubation for 20 s. Samples obtained from this procedure were then analysed by TEM (**Extended Data Figure 4**). Neither the control sample for which **1** was omitted, nor the control sample for which no mechanical agitation was applied, showed detectable levels of fibre formation after a prolonged time (48 h).

Initial assessment of the activity of **1₆** in catalysing the Fmoc-glycine cleavage

A stock solution of Fmoc-glycine **4** (20 µL, 10 mM in acetonitrile)

was added to 900 μL of borate buffer (50 mM, pH 8.12) with 30 μL of acetonitrile, the resulting mixture was incubated for 5 min at 25 $^{\circ}\text{C}$. At this point, 50 μL of a stock solution of **1**₆ (4.0 mmol in the same borate buffer) was added to yield a final concentration of 200 μM **1**₆ (with respect to building block **1**), 200 μM Fmoc-glycine and 5% acetonitrile. Reaction progress was monitored through periodic analysis of the reaction mixture by UPLC at 25 $^{\circ}\text{C}$ (10 μL injection of the mixture, 18 min intervals). As a control, the experiment was carried out in presence of **1** as well as a mixture of **1**₃/**1**₄ under otherwise identical conditions. For the latter, the concentration of **4** was quantified by comparison with a calibration curve that was obtained by injecting known quantities of **4** into the reaction buffer (**Supplementary Figure 3C**).

Determination of uncatalysed rate and the apparent second-order rate constant for a mixture of **1**₃ and **1**₄

As the uncatalysed reaction was observed to be very slow at 25 $^{\circ}\text{C}$ and 40 $^{\circ}\text{C}$, the reaction mixture was heated to different temperatures using a block heater. Two samples with a total volume of 1,000 μL with **4** (40 μM) and acetonitrile (5%) in borate buffer (pH 8.1) were prepared in HPLC vials (12 $\times$ 32 mm) and incubated for 5 min. The samples were heated placed under three different temperatures in doublets (at $T = 60, 70$ and 80 $^{\circ}\text{C}$, **Supplementary Figure 5A**). The vials were taken out of the heater alternately to be sampled for UPLC analysis (10 μL injections, 36 min intervals for 4.2 h) at 40 $^{\circ}\text{C}$. The resulting decrease of **4** was quantified by comparing the peak area at 254 nm with the calibration curve shown in **Supplementary Figure 3C**. Observed first-order rate constants were measured by determining the initial rate of the reaction via the Arrhenius equation, yielding the following data: $2.37 \times 10^{-5} \text{ s}^{-1}$ (80 $^{\circ}\text{C}$), $8.18 \times 10^{-6} \text{ s}^{-1}$ (70 $^{\circ}\text{C}$) and $2.67 \times 10^{-6} \text{ s}^{-1}$ (60 $^{\circ}\text{C}$). These rate constants were used to extrapolate the first-order rate constant at room temperature ($2.86 \times 10^{-8} \text{ s}^{-1}$ at 25 $^{\circ}\text{C}$). The Arrhenius plot is shown in **Supplementary Figure 5B**. To test the catalytic proficiency of **1**, a 75 μL stock solution of 40 μM **4** (1.0 mM in acetonitrile) was prepared under inert atmosphere and incubated for 5 min; 75 μL of a stock solution of **1** (4.0 mM in borate buffer) was then added under inert atmosphere to avoid oxidation of **1** into small, non-assembling macrocycles, giving a final concentration of 300 μM **1** and an overall acetonitrile content of 5%. The sample, and a control sample without **1**, were heated to 60 $^{\circ}\text{C}$ and periodically

analysed by UPLC (50 μL , daily for 3 d). The progress of the reaction is displayed in **Supplementary Figure 5A**, which demonstrates that **1** does not lead to a notable acceleration of the Fmoc-glycine cleavage reaction at these concentrations. Kinetic characterization of a freshly prepared mixture of **1**₃/**1**₄ was performed by adding 12.5, 37.5 or 75 μL of a stock solution of **1**₃/**1**₄ (4.0 mM with respect to building block **1** in borate buffer pH 8.1) to a sample containing Fmoc-glycine (40 μM) and acetonitrile (5%) in borate buffer (pH 8.1), giving total a total concentration of 50, 150 or 300 μM (with respect to building block **1**). The reaction mixtures were incubated for 10 min and then monitored for 6 d by UPLC (10 μL injection of the mixture daily for 6 d). Observed second-order rate constants at all three concentrations were measured by determining the initial rate of the reaction and were comparable for the **1**₃/**1**₄ mixture at all of the concentrations tested ($6.31 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ (50 μM), $5.89 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ (150 μM), $7.07 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$ (300 μM)) at room temperature, suggesting that the reaction is first order with respect to catalyst concentration. By averaging the obtained values an apparent second-order rate constant was determined ($6.43 \times 10^{-3} \text{ M}^{-1} \text{ s}^{-1}$). Plots depicting the reaction of **4** in the presence of **1**₃/**1**₄ are shown in **Supplementary Figure 5C**.

Kinetic characterization of the effect of the **1**₆ fibres for the Fmoc-glycine cleavage

Excess substrate to catalyst. This kinetic characterization of **1**₆ fibres was performed by measuring reaction progress spectrophotometrically ($\epsilon_{305 \text{ nm}} = 3,088 \text{ M}^{-1} \text{ cm}^{-1}$) at varying concentrations of **4** (20, 50, 75, 100, 150 and 200 μM) at 25 °C. Samples were prepared by adding appropriate amounts of acetonitrile and a stock solution of **4** (10 mM in acetonitrile) into 1 cm path-length cuvettes containing 937 μL borate buffer (50 mM, pH 8.1). The reaction mixtures were incubated for 20 min at 25 °C, at which point 13 μL of a stock solution of **1**₆ (1.54 mM with respect to building block **1** in borate buffer, final concentration 20 μM with respect to building block **1**) was added. The reaction progress was followed for 1 h at 25 °C. The observed rates for the initial phase (0–20 min) were extracted by determining the slope of product formation over time. The measured rates were corrected for the background rate and the resulting values — which represent the average of at least two independent measurements—were fitted (RStudio35) to equation (4), in which $[\mathbf{1}] = 2.0 \times 10^{-5} \text{ M}$:

$$\frac{v_0}{[1]} = \frac{k_{cat}[2]}{K_M + [2]} \quad (\text{Eq 4.4})$$

The pH rate profile for the effect of **1**₆ on the Fmoc-glycine cleavage

Different buffer solutions (50 mM) were prepared and adjusted to the pH 5.8 – 8.5 range (MOPS buffer pH 5.8 – 7.0; borate buffer pH 7.5 – 8.5). Samples were prepared by adding acetonitrile (40 µL) and 10 µL of a stock solution of **4** (10 mM in acetonitrile) into 1 cm cuvettes containing 924 µL of the specific buffer solution. The resulting mixtures were incubated for 10 min at 25 °C before addition of a stock solution of **1**₆ (26 µL, 1.54 mM with respect to building block **1** in borate buffer, pH 8.12) to obtain a final concentration of 40 µM **1**₆ (with respect to building block **1**), 100 µM **4** and 5% acetonitrile. The reaction progress was monitored photometrically ($\epsilon_{305} \text{ nm} = 3,088 \text{ M}^{-1} \text{ cm}^{-1}$) at 25 °C for 60 min. The measured rates were corrected for

the background and $(k_{cat}/K_M)_{app}$ values were determined by dividing the corrected rates by the concentrations of **1** ($4 \times 10^{-5} \text{ M}$) and substrate ($1 \times 10^{-4} \text{ M}$). Averages of values from least two independent measurements were subsequently fitted (RStudio) to equation (5).

$$\left(\frac{k_{cat}}{K_M} \right)_{app} = \frac{\left(\frac{K_{cat}}{K_M} \right)_{max}}{1 + 10^{pK_a - pH}} \quad (\text{Eq 4.5})$$

Experiment corresponding to Figure 3E,F and Extended Data Figure 5C,D

A fresh solution of **4** (10 µL or 20 µL, 10 mM in acetonitrile) was added to borate buffer (900 µL, 50 mM, pH 8.1) and acetonitrile (40 µL or 30 µL) in an HPLC vial (12 × 32 mm) equipped with a Teflon-coated magnetic stirring bar (5 × 2 mm, VWR), resulting in a concentration of 100 µM or 200 µM, respectively, with an acetonitrile content of 5%. The mixture was preheated to 40 °C for 10 min inside the UPLC system; 50 µL of a freshly prepared solution of **1** (4.0 mM in borate buffer, pH 8.1) was then added while the temperature was kept at 40 °C and the reaction was stirred at 600 rpm for the entire duration of the experiment (27 h). The reaction mixture was monitored periodically via UPLC (10 µL injection every 54 min). The concentrations of **1**, **1**₄

and **1₆** were quantified by comparing to a calibration curve that was obtained by injecting known quantities of **1** in the buffer (**Supplementary Figure 3**). The consumption of **4** was quantified by comparing the peak area to the calibration curve shown in **Supplementary Figure 3**. Representative chromatograms recorded for the experiment are depicted in **Supplementary Figure 6**.

EXTENDED FIGURES

Extended Data Figure 1: Kinetic characterization of **1**, **1₃/1₄** and replicators. Comparison of the reaction progress between the uncatalyzed retro-aldol reaction (red in A and B) and that in the presence of different concentration of **1** (A) and varying concentrations of a mixture of **1₃/1₄** (B). (C): Reaction progress monitored spectrophotometrically with **1₆** (25 μM) and methodol (600 μM , in triplicates with single standard deviation shown as shaded area). Red and blue dotted lines are linear fits for the observed burst (first 20 minutes) and steady state rate (100–200 minutes). (D): Apparent catalytic efficiencies for the cleavage of methodol (200 μM) of **1₆** and replicators featuring lysine-to-arginine mutations.

Extended Data Figure 2: pH titration and TEM studies of 1_6 replicators. (A) pH titration for 1 and 1_6 fibres in water (both in duplicates). The buffering capacity of monomers at a pH between 5 and 7 can be attributed to the presence of aromatic thiols in 1 . In contrast, 1_6 fibres do not reveal appreciable levels of an ionizable group in the range from 5–8. The absence of a recognisable plateau at the apparent $\text{p}K_a$ of the lysines active in the retro-aldol ($\text{p}K_{a,\text{app}} = 7.4$; **Figure 2E**) or Fmoc-glycine cleavage reactions ($\text{p}K_{a,\text{app}} = 7.1$; **Extended data Figure 5A**) indicates that, for each of these reactions, the fraction of lysine residues in the fibres that are catalytically active is low. All error bars shown correspond to a single standard deviation. (B–C) TEM images of 1_6 fibres at low and high pH regimes do not show a change in morphology.

Extended Data Figure 3: Emergence of catalytic activity of the retro-aldol reaction in a stirred sample of **1** correlates with fibre formation. (A): Rate of formation of retro-aldol product **3** over the course of the experiments. Solid line represents a linear fit to the uncatalyzed reaction. (B–D): Concentration of **1₆** (B), **1** (C), and **1₄** (D) followed over time. Colour code as indicated in the legend.

Extended Data Figure 4: Time-dependent TEM experiments. For agitated samples (A–F) fibres become detectable after 8.5 hours, which correlates with the increase in the concentration of **1₆** in the experiment shown in **Figure 2**. Neither the non-agitated sample (G, H) nor the sample lacking **1** (i) showed detectable levels of fibres after 48 hours.

Extended Data Figure 5: Emergence of catalytic activity of Fmoc cleavage in a stirred sample of **1** correlates with fibre formation and an increase in the oxidation rate. (A), pH rate profile for the cleavage of Fmoc-glycine (100 μM) catalysed by replicator **1**₆ (40 μM in **1**). (B), Apparent catalytic efficiencies for the cleavage of Fmoc glycine (200 μM) by **1**₆ and replicators featuring lysine-to-arginine mutations. (C), Data for experiments as shown in Figure 3B, but using 200 μM Fmoc-glycine. The emergence of **1**₆ (dark blue circles) coincides with the onset of Fmoc-glycine cleavage (red circles) and **1**₆ replicates faster when compared to a sample that does not contain Fmoc-glycine (light blue circles). (D), A comparison of oxidation rates in the same experiment showed an up to 4.7 fold increase (purple arrow) once **5** is liberated, providing further evidence for the positive feedback loop shown in Figure 3B. (E), A comparison of the oxidation of **1** (100 μM) in presence (red circles) and absence (blue circles) of **5** (100 μM) confirmed that this Fmoc-cleavage product is responsible for enhancing the rate of oxidation of **1**. Experiments were performed in borate buffer

pH 8.1, containing 5% acetonitrile at 40 °C. All error bars shown correspond to a single standard deviation.

ACKNOWLEDGEMENTS

J. Ottel, S. Otto and C. Mayer conceived the project. C. Mayer and S. Otto supervised the overall project. J. Ottel designed and performed the experiments related to the retro-aldol reaction and replicators XGLRFK and XGLKFR. A. S. Hussain conceived, designed and performed the experiments relating to the Fmoc-glycine cleavage reaction. J. Ottel performed electron microscopy measurements and pH titrations. J. Ottel, A. S. Hussain, C. Mayer and S. Otto co-wrote the paper. All authors discussed the results and commented on the manuscript.

BIBLIOGRAPHY

1. Ruiz-Mirazo, K., Briones, C., de la Escosura, A. Prebiotic systems chemistry: new perspectives for the origins of life. *Chem. Rev.* **2014**, 114, 285–366.
2. Patzke, V., von Kiedrowski, G. Self replicating systems. *Arkivoc*, **2007**, 293–310.
3. Dadon, Z., Wagner, N., Ashkenasy, G. The road to non-enzymatic molecular networks. *Angew. Chem. Int. Ed.* **47**, **2008**, 6128–6136.
4. Vidonne, A., Philp, D. Making molecules make themselves—the chemistry of artificial replicators. *Eur. J. Org. Chem.* **2009**, 593–610.
5. Duim, H., Otto, S. Towards open-ended evolution in self-replicating molecular systems. *Beilstein J. Org. Chem.* **2017**, 13, 1189–1203.
6. Moreno, A., Ruiz-Mirazo, K. The problem of the emergence of functional diversity in prebiotic evolution. *Biol. Philos.* **2009**, 24, 585–605.
7. Kamioka, S., Ajami, D., Rebek, J. Autocatalysis and organocatalysis with synthetic structures. *Proc. Natl Acad. Sci. USA.* **2010**, 107, 541–544.
8. Higgs, P. G., Lehman, N. The RNA world: molecular cooperation at the origins of life. *Nat. Rev. Genet.* **2014**, 16, 7–17.
9. Eschenmoser, A. The search for the chemistry of life's origin. *Tetrahedron.* **2007**, 63, 12821–12844.
10. Le Vay, K., Weise, L. I., Libicher, K., Mascarenhas, J., Mutschler, H. Templated self-replication in biomimetic systems. *Adv. Biosyst.* **2019**, 3, 1–16.
11. Müller, M. M., Windsor, M. A., Pomerantz, W. C., Gellman, S. H., Hilvert, D. A rationally designed aldolase foldamer. *Angew. Chem. Int. Ed.* **2009**, 48, 922–925.
12. Rufo, C. M. et al. Short peptides self-assemble to produce catalytic amyloids. *Nat. Chem.* **2014**, 6, 303–309.
13. Weingarten, A. S. et al. Self-assembling hydrogel scaffolds for photocat-

alytic hydrogen production. *Nat. Chem.* **2014**, 6, 964–970.

14. Tena-Solsona, M. et al. Emergent catalytic behavior of self-assembled low molecular weight peptide-based aggregates and hydrogels. *Chem. A Eur. J.* **2016**, 22, 6687–6694.

15. Omosun, T. O. et al. Catalytic diversity in self-propagating peptide assemblies. *Nat. Chem.* **2017**, 9, 805–809.

16. Malakoutikhah, M. et al. Uncovering the selection criteria for the emergence of multi-building-block replicators from dynamic combinatorial libraries. *J. Am. Chem. Soc.* **2013**, 135, 18406–18417.

17. Colomb-Delsuc, M., Mattia, E., Sadownik, J. W., Otto, S. Exponential self-replication enabled through a fibre elongation/breakage mechanism. *Nat. Commun.* **2015**, 6, 7427.

18. Westheimer, F. H. Coincidences, decarboxylation, and electrostatic effects. *Tetrahedron*, **1995**, 51, 3–20.

19. Giger, L. et al. Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **2013**, 9, 494–498.

20. Lassila, J. K., Baker, D., Herschlag, D. Origins of catalysis by computationally designed retroaldolase enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, 107, 4937–4942.

21. Tanaka, F., Fuller, R., Shim, H., Lerner, R. A. & Barbas, C. F. Evolution of aldolase antibodies in vitro: correlation of catalytic activity and reaction-based selection. *J. Mol. Biol.* **2004**, 335, 1007–1018.

22. Suh, J., Scarpa, I. S., Klotz, I. M. Catalysis of decarboxylation of nitrobenz-isoxazolecarboxylic acid and of cyanophenylacetic acid by modified polyethylenimines. *J. Am. Chem. Soc.* **1976**, 98, 7060–7064.

23. Hollfelder, F., Kirby, A. J., Tawfik, D. S. Efficient catalysis of proton transfer by synzymes. *J. Am. Chem. Soc.* **1997**, 119, 9578–9579.

24. Frederix, P. W. J. M. et al. Structural and spectroscopic properties of assemblies of self-replicating peptide macrocycles. *ACS Nano*, **2017**, 11,

7858–7868.

25. Jiang, L. et al. De novo computational design of retro-aldol enzymes. *Science*, **2008**, 319, 1387–1391.

26. Althoff, E. A. et al. Robust design and optimization of retroaldol enzymes. *Protein Sci.* **2012**, 21, 717–726.

27. Bjelic, S. et al. Exploration of alternate catalytic mechanisms and optimization strategies for retroaldolase design. *J. Mol. Biol.* **2014**, 426, 256–271.

28. Capozzi, G., Modena, G. in *The Chemistry of the Thiol Group* (ed. Patai, S.), Wiley, **1974**, 785–839.

29. Dénès, F., Pichowicz, M., Povie, G., Renaud, P. Thiyl radicals in organic synthesis. *Chem. Rev.* **2014**, 114, 2587–2693.

30. Leveson-Gower, R. B., Mayer, C., Roelfes, G. The importance of catalytic promiscuity for enzyme design and evolution. *Nat. Rev. Chem.* **2019**, 3, 687–705.

31. Carbonell, P., Lecointres, G. & Faulon, J. L. Origins of specificity and promiscuity in metabolic networks. *J. Biol. Chem.* **2011**, 286, 43994–44004.

32. Czárán, T., Könnyu, B., Szathmáry, E. Metabolically coupled replicator systems: overview of an RNA-world model concept of prebiotic evolution on mineral surfaces. *J. Theor. Biol.* **2015**, 381, 39–54.

33. Guo, H., Tanaka, F. A fluorogenic aldehyde bearing a 1,2,3-triazole moiety for monitoring the progress of aldol reactions. *J. Org. Chem.* **2009**, 74, 2417–2424.

34. Nakano, T., Yade, T. Synthesis, structure, and photophysical and electrochemical properties of a π -stacked polymer. *J. Am. Chem. Soc.* **2003**, 125, 15474–15484.

35. RStudio Team (RStudio Inc., 2015); <https://www.rstudio.com>

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY DISCUSSION

Estimating the fraction of catalytically active lysine residues for catalysis of the retro-aldol reaction by $\mathbf{1}_6$:

Here we provide a comprehensive discussion of the results and the assumptions we made to obtain an estimate for the fraction of catalytically active lysine residues on a given fibre sample. First, for all kinetic characterizations of $\mathbf{1}_6$ fibres we assumed that: (1) active sites on the fibres are independent of each other, and (2) the measured activity can solely be attributed to lysine side chains in the active site. We define n ($0 < n < 2$) as the number of lysine residues per monomer (a maximum of two lysines per $\mathbf{1}$) that contribute a catalytically active ϵ -amine. In a first step to determining n , we performed a kinetic characterization, in which the concentration of $\mathbf{2}$ was varied (75 – 600 μM) in presence of 25 μM (with respect to building block $\mathbf{1}$) $\mathbf{1}_6$ fibres (see **Figure 2C**). Under these conditions $[\mathbf{2}] \gg [\text{catalytic sites}]$. Performing a Michaelis-Menten analysis yields an apparent “ k_{cat} ” that corresponds to $k_{\text{cat}}^* n$. Conversely, the obtained K_{M} value – which is not dependent on the concentration of catalyst, assuming a Michaelis-Menten model – accurately reflects the K_{M} of individual active sites.

In a second experiment, we determined the pseudo-first order rate constants in presence of a large excess of $\mathbf{1}_6$ fibres (100 – 1,500 μM with respect to building block $\mathbf{1}$) when compared to $\mathbf{2}$ (10 μM , see **Figure 2D**). In these experiments $[\text{catalytic sites}] \gg [\mathbf{2}]$ and all $\mathbf{2}$ should eventually be bound to the catalyst. Fitting the obtained pseudo-first order rate constants at varying concentrations of $\mathbf{1}_6$ fibres to a Michaelis-Menten model gives k_{cat} and K_{M}/n . In this case, it is the K_{M} value that needs adjustment by n , as it depends on the number of active lysines.

Conditions	$n * k_{cat} * 10^4$ [s ⁻¹]	K_M [μM]	$k_{cat} * 10^4$ [s ⁻¹]	K_M [μM] / n	n
[2] >> [catalytic sites]	0.96*	670*			0.054 (from k_{cat})
[catalytic sites] >> [2]			17.8	4570	0.147 (from K_M)

* values obtained from the “burst” phase.

The kinetic parameters obtained from both experiments are summarized in the Table above. From these parameters, we can determine n , by either (1) $(n * k_{cat}) / n$ or (2) $K_M / (K_M / n)$, yielding 0.054 and 0.147 respectively. These values correspond to 2.7% and 7.3% of all lysines or 5.4% and 14.7% of all building blocks on a typical fibre sample being catalytically active. Both the limited solubility of **2** under the assay conditions and the limited availability of **1** prohibit performing reactions under conditions in which all available catalytic sites are saturated. Being unable to achieve saturation results in significant uncertainty for the parameters computed. Given these constraints we conservatively estimate <10% of all lysines to be catalytically active. The fact that we cannot identify an ionizable group in the pH titration of the fibres (see **Extended Data Figure 2A**) with an apparent pK_a that corresponds to the one observed in the pH rate profile (see **Figure 2E**) also suggests that the fraction of lysines that are catalytically active is small.

Supplementary Figure 1: Representative UPLC chromatograms of **1** (A), a solution of **1** that was oxidized using sodium perborate containing a mixture of **1₃**/**1₄** (B), **1₆** (C), (XKLKFR)₆ (D), and (XGLRFK)₆ (E) prepared as described in the Materials and Methods section.

Supplementary Figure 2: Representative mass spectra for **1** (A), **1₃** (B), **1₄** (C), and **1₆** (D) obtained for the samples shown in **Supplementary Figure 1**. Relevant charge states and the resulting deconvoluted masses (M_{deconv}) are indicated in the corresponding spectra.

Supplementary Figure 3: Calibration curves for **3** (A) and **1** (B), and **4** (C) obtained from subjecting samples with known quantities of these compounds to UPLC analysis. Linear fit and corresponding equation are shown in red.

Supplementary Figure 4: Representative UPLC chromatograms near the start and end of the experiments shown in **Figure 2F** and **Extended Figure 2**. From left to right: agitated sample (A-B), non-agitated control (C-D) and background (E-F, the experiment was stopped after 15 hours).

Supplementary Figure 5: A) Comparison of the reaction progress between the uncatalyzed Fmoc-glycine cleavage at different temperatures and in presence of **1** (300 μM). B) Arrhenius plot of the uncatalysed cleavage reaction that was used to extrapolate the rate constant of the reaction at 25 °C. Linear fit and corresponding equation highlighted in red. (c) Reaction progress in presence of **1**₃/**1**₄ (50, 150, 300 μM) at 25 °C.

Supplementary Figure 6: Representative UPLC chromatograms near the start (54 min), near the emergence (430 min) and after emergence of **1₆** (860 min) for the experiment shown in **Figure 3F** in absence (A-C) and presence (D-F) of **4**.

SUPPLEMENTARY FIGURES

SUPPLEMENTARY TABLES

Name	Reference	$k_{\text{cat}} * 10^4 \text{ (s}^{-1}\text{)}$	$K_M (\mu\text{M})$	$k_{\text{cat}}/K_M * 10^1 \text{ (M}^{-1}\text{s}^{-1}\text{)}$	$1/K_{\text{TS}} * 10^1 \text{ (M}^{-1}\text{)}$
1_b(b)*		4.80 ± 0.61	670 ± 140	7.17 ± 0.66	1.91
1_b(ss)*		2.66 ± 0.32	930 ± 160	2.86 ± 0.18	0.76
RA22	26	0.08	450	0.18	0.28
RA34	26	0.10	600	0.16	0.25
RA45	26	0.38	430	0.91	1.40
RA46	26	1.03	290	3.70	5.96
RA60	26	1.55	510	3.00	4.65
RA61	26	1.50	210	7.40	11.4
RA95	27	0.33	540	0.53	0.82
RA110	27	0.83	1600	0.48	0.74
RA114	28	0.03	34	0.95	1.46
RA115	28	0.15	1740	0.087	0.13
RA116	28	0.14	1570	0.088	0.14
RA117	28	1.08	470	0.58	0.89
RA120	28	0.04	880	0.041	0.06

Supplementary Table 1: Summary of kinetic parameters for catalysis of the retro-aldol reaction of methodol (2) by replicators **1_b**, as well as a comparison of catalytic activities with a number of computationally-designed retroaldolases.

<i>Time (minutes)</i>	<i>%A</i>	<i>%B</i>
<i>0.0</i>	<i>90</i>	<i>10</i>
<i>1.0</i>	<i>90</i>	<i>10</i>
<i>1.3</i>	<i>75</i>	<i>25</i>
<i>3.0</i>	<i>72</i>	<i>28</i>
<i>11.0</i>	<i>60</i>	<i>40</i>
<i>11.5</i>	<i>5</i>	<i>95</i>
<i>12.0</i>	<i>5</i>	<i>95</i>
<i>12.5</i>	<i>90</i>	<i>10</i>
<i>17.0</i>	<i>90</i>	<i>10</i>

Supplementary Table 2: Gradient used for UPLC separation of retro-aldol reaction mixtures (A=water containing 0.1% TFA, B=acetonitrile containing 0.1% TFA).

CHAPTER 5

Diversification through multi building block replicators

ABSTRACT

In this chapter, we discuss the introduction of mutations into self-replicating systems. One way to attain diversification in self-replicators is the use of multi building-block libraries. Here, we investigate a new analytical tool with which large numbers of such libraries can be screened. Using a fluorescent probe, we were able to indirectly monitor a key stabilising factor of the replicating supramolecular structures: β -sheet formation. Simultaneously, the addition of a Fmoc derived substrate enabled a positive feedback loop in these mixed libraries that can be activated through catalytic deprotection. The steepness of the curve describing β -sheet formation could allow us to indirectly probe which mixtures of building-blocks are able to utilize the feedback loop and which are not. We had hoped to discover self-sorting behaviour combined with catalytic activity, but we were unable to definitively prove either objective. However, the analytical method investigated has shown to be a useful new tool in the analysis of many libraries simultaneously: be it for the screening of catalytic activity or the ability to form supramolecular structures.

5.1 INTRODUCTION

Living systems are – amongst other things – characterized by their ability to undergo Darwinian evolution.^{1–3} Small, congenital disparities in an individual can lead to a larger probability to reproduce, allowing for the propagation of useful hereditary traits. To enable Darwinian evolution in synthetic life-like systems, we must introduce machineries that permit mutations to occur.^{4,5} Furthermore, a selection pressure is required so that beneficial mutations can propagate through future generations, and detrimental mutations lead to extinction.

When Darwinian evolution is applied to self-replicating systems, selection may be based on replication rate.^{6,7} If multiple competing replicators exist in a single system, the one that replicates fastest will quickly outgrow the others. Mutations within replicators that increase the efficiency of their autocatalysis will thus result in the selection of that particular replicator.

In previous chapters, we have extensively examined various systems of self-replicating molecules. All of these systems were based on the use of a single building block that, upon oxidation, can form a variety of macrocycles. Some of these macrocycles are capable of minimizing their free energy by stacking on top of one another to create supramolecular fibrillar structures held together by non-covalent forces. Mechanical agitation leads to the fragmentation of these structures, increasing the number of growing fibre ends and pushing the formation of the replicating macrocycle into an exponential regime.⁸

To facilitate mutation machineries in self-replicating systems, one can make use of the mechanism with which the supramolecular structures are formed. The non-covalent interactions stabilising the formation of these fibres are induced by π - π stacking of the aromatic core units and β -sheet formation from the peptide appendages.^{9,10} A change in the peptide sequence can lead to the disruption or strengthening of the β -sheet, which can lead to an array of different properties such as changes in replicating macrocycle size, replication rate and catalytic activity through active site modification.^{11–13} The latter can be further expanded on by the use of multiple building blocks, which can lead to stacking mismatches and thus the generation of a variety of microenvironments alongside the supramolecular structures (see **Figure 1 A,B**) that can result in changes in their catalytic activity (see **Figure 1C**).

Most importantly, the use of different building blocks is essential to obtain a coexistence or intermingling of various replicators in a single system: exceptionally strong, non-disruptive interactions between different building blocks can lead to social self-sorting of replicators (see **Figure 2A**). Here, social refers to the preference for building blocks to mix with one another to form one particular macrocycle (e.g. a hexamer that consists of alternating red and turquoise building blocks). On the other hand, if there are strong non-disruptive interactions between identical building blocks but disruptive interactions between different building blocks, a system of mixed replicators might give rise to narcissistic self-sorting (see **Figure 2B**). Narcissistic implies the formation of macrocycles that do not mix with one another at all: they consist of only one of the available building blocks (e.g. hexamers that consist only of red building blocks). Finally, when such interactions are neither strongly disruptive or stabilizing, a statistical

distribution of the incorporation of building blocks is expected.¹²

Figure 1: A) Hexamer fibre constructed from two building blocks. B) Random incorporation of either building block leads to stacking mismatches which can lead to C) active site modification, influencing catalytic activity.

However, in practice, multi-building block systems do not necessarily lead to replicator formation: often, mixing various building blocks leads to ill-defined aggregates or freely diffusing macrocycles without the construction of supramolecular structures such as fibres or ribbons. Depending on the choice of building blocks, the resulting macrocycles might become too hydrophobic and aggregate without the formation of replicating species. On the other hand, building blocks that are too hydrophilic will lead to the formation of small, solvated macrocycles. A balance of the two is required: hydrophobic enough to be able to minimise free energy by stacking yet hydrophilic enough to not immediately aggregate.

The stacking of the macrocycles leads to the formation of supramolecular structures and is stabilised by the π - π stacking of aromatic core units and β -sheet formation between the peptide tails. Both of these interactions can be modified to tailor the interactions required for self-sorting phenomena. Core unit modification can not only change π - π stacking interactions, but also influence the flexibility of the thiol moieties, thus potentially leading to variations in preferred macrocycle size which can promote self-sorting behaviour. Chang-

ing the peptide sequence influences β -sheet interactions and hydrophilicity, determines the self-sorting propensity (if any) and is essential for the incorporation of emergent functions such as catalysis (see **Chapter 4**).

The β -sheet formation is a crucial part of the replication process. We envisaged that this could be used to indirectly probe the formation of supramolecular structures – be it fibres, ribbons or otherwise. There exist a variety of fluorescent probes such as Congo Red (CR) and Thioflavin T (ThT) that display enhanced fluorescence upon binding to β -sheets. ThT is a particularly suitable candidate for β -sheet detection as it can be used in aqueous solution.^{14,15}

Figure 2: Self-sorting properties of replicators in mixed building block systems depend on the interactions between the building blocks. A) Red and turquoise building blocks have strong, non-disruptive interactions which promotes social self-sorting. B) Red and blue building blocks experience strong disruptive interactions between one another, promoting narcissistic self-sorting. Note that other configurations could also be called narcissistic or social self-sorting.

Combining fluorescent probes with the high-throughput sampling of a (fluorescent) plate reader allows us to screen large numbers of mixed building block systems for the formation of supramolecular structures (see **Figure 3**).

As mentioned before, attaining additional functions over time is essential to enable Darwinian evolution in systems of self-replicators. In **Chapter 4** the catalytic capabilities of various replicators were examined and compared. Replicators were found to be able to catalyse

Figure 3: High-throughput screening of replicating species using ThT and plate readers. 1) Building blocks are mixed with one another to create a large variety of dynamic combinatorial libraries. 2) A fluorescent probe (ThT) is added to the libraries, which are mechanically agitated through orbital shaking in the plate reader. 3) Fluorescent probe is excited (440 nm) and the emission (482 nm) is recorded for each individual library. Libraries in which the fluorescent signal increases over time is indicative of β -sheet formation, which hints at the presence of self-replicators.

two bond-breaking reactions using lysine residues with a perturbed pK_a – the direct result of the formation of supramolecular structures. One of the reactions, fluorenylmethyloxycarbonyl (FMOC) deprotection, was found to introduce a positive feedback loop in the self-replicating system: dibenzofulvene (DBF) – the product of the reaction – is capable of enhancing thiol oxidation.¹⁶ We discovered an increase in replication rate during emergence experiments containing catalytically active replicators and FMOC glycine.¹³

Since the positive feedback loop obtained through DBF formation increases replication rate, we can screen for potential catalytically active replicators by addition of DBF forming substrates. Combinations of building blocks that are catalytically active should then be

able to access the positive feedback loop, leading to an increased rate of replicator formation and by extension to a stronger ThT signal in the plate reader (see **Figure 4**).

Finally, screening large combinations of building blocks might reveal particular combinations that show self-sorting behaviour. The fluorescent probe will identify which mixtures form β -sheets and subsequent liquid chromatography–mass spectrometry (LC-MS) analysis will identify which macrocyclic species are present. Finding coexisting replicators is a crucial step towards the implementation of competition scenarios to eventually enable open ended evolution. In this chapter, we introduce a new approach to screen large numbers of potential replicators constructed from mixed building blocks while simultaneously probing their catalytic activity through the availability of a catalytically induced positive feedback loop.

5.2 RESULTS AND DISCUSSION

5.2.1 BUILDING BLOCK SELECTION

Throughout the last decade, our group has gathered a large collection of diverse thiol-functionalized building blocks. To still somewhat limit the number of possible combinations, twenty-two pentapeptide based building blocks were selected (see **Figure 5** and **Figure 6**). From this selection, 231 libraries can be constructed that consist of two different building blocks (see Eq 5.1)

$$\text{Combinations} = \binom{n}{m} = \frac{n!}{m!(n-m)!} \quad (\text{Eq 5.1})$$

Where n is the number of different building blocks and m the amount of building blocks in any one library. In this case, $n = 22$ and $m = 2$. Out of the twenty-two selected building blocks, fourteen have a modification in the pentapeptide sequence compared to building block **1** (see **Figure 5**, building blocks **2-15**). These building blocks were loosely selected based on changes in their hydrophobicity or lysine modifications – both expected to influence the common disruptive or stabilising interactions and their ability to catalyse chemical reactions.

Figure 4: Multi-building block libraries are constructed in microwell plates as described in the Supplementary Information. A) Some building block combinations can result in catalytically active replicators (X), while other combinations (Y) do not. Those that are catalytically active enable a positive feedback mechanism through DBF mediated thiol oxidation, increasing replication rate. B) The fluorescent signal obtained through β -sheet binding of ThT would then increase more rapidly for the catalytically active replicators.

1 = XGLKFK

2 = XGLOFK

3 = XGLmKFK

4 = XGLaFFK

5 = XGLKFhR

6 = XGLKWK

7 = XGLKFR

8 = XGLRFK

9 = XGLKSK

10 = XGLKNapthK

11 = XGLKFY

12 = XGLKAK

Figure 5 (left and continued above): Building block design based on the modification of the peptide tail compared to building block 1 (shown in red). Building blocks were – in general – designed based on lysine modification (building blocks **2-5**, **7** and **8**) and changes in hydrophobicity (building blocks **6**, **9-15**).

Building blocks **1** and **9-14** have been under investigation before.^{11,17-19} Building blocks **2-8** were developed for the investigation of catalytic capabilities by active site modification and have also been studied extensively.²⁰ The library progression of these individual building blocks often reveals the formation of a single library constituent and the formation of supramolecular structures (see **Supplementary Figures 1-4**).

Four building blocks are equipped with an inverse peptide sequence: the sequence of hydrophobic and hydrophilic amino acids in the peptide is reversed with respect to building block **1** (see **Figure 6**, building blocks **16-19**). Finally, three building blocks have a variation in the aromatic core unit (see **Figure 6**, building blocks **20-22**) but have the same peptide sequence as building block **1**. To go into much detail about the properties of these individual building blocks is beyond the scope of this chapter and will most likely be extensively discussed in the dissertations of my colleagues Marcel Eleveld and Paul Adamski, who designed and synthesized them.

Figure 6: Building block design in which hydrophobic and hydrophilic amino acids are reversed in the sequence with respect to building block **1**, (shown in red, building blocks **16-19**) or where the aromatic core moiety is modified (shown in red, building blocks **20-22**).

However, it should be noted that not all of these building blocks are known to give rise to disulfide macrocycles that spontaneously self-assemble (let alone form self-replicators) – or do so under specific conditions that are very different from the conditions in these experiments. Regardless, their behaviour in mixed systems can be very different depending on the counterpart building block present. The table on the next page summarizes the known behaviour of all building blocks used.

BUILDING BLOCK	DOMINANT RING SIZE	REPLICATOR (Y/N)	STRUCTURE TYPE
1	Hexamer	Y	Fibres
2	Hexamer	Y	Fibres
3	No particular species dominates the library	N	Aggregates
4	Trimer	N (no seeding effects observed)	Fibres
5	Hexamer	Y	Fibres
6	Hexamer	N (no seeding effects observed)	Fibres
7	Hexamer	Y	Fibres
8	Hexamer	Y	Fibres
9	Octamer	Y	Fibres
10	Hexamer	Y	Fibres
11	No particular species dominates the library	N	Aggregates
12	Octamer	Y	Fibres
13	Hexamer	Y	Fibres
14	Octamer	Y	Fibres
15	Hexamer	Unknown	Fibres
16	Hexamer	Y	Fibres
17	Hexamer	Y	Fibres
18	Hexamer	Y	Fibres
19	Trimer	Unknown	Unknown
20	Dimer	Unknown	Ribbons
21	Dimer	Unknown	Aggregates
22	Dimer	Unknown	Fibres

Table 1: Results of single building-block library progression, taken from references **11**, **17–19** and **20**. Most of the individual building blocks form replicating species, but not all. Similarly, not all species that form supramolecular structures are shown to replicate or require the formation of larger macrocycles.

The implementation of large-scale screening of mixed building block systems requires addressing various technical issues. First of all, experiments must be run at elevated temperatures to keep them within a reasonable timeframe. Since the 96 (micro)well plates used in plate readers can only contain small volumes ($< 500 \mu\text{L}$), evaporation of the aqueous media must be circumvented: increasing the concentration of building blocks can lead to the formation of otherwise inaccessible products.²¹ Luckily, this is easily circumvented by using a water-tight microplate sealing film (see Methods: *Plate reader experiments*). Furthermore, the manner in which these plates are agitated is different from typical experiments, in which they are stirred using magnetic stirring bars. In the plate reader, agitation is achieved through orbital shaking. Changes in agitation can lead to the formation of different products.²² To confirm which replicating species are present after screening, promising libraries are submitted to UPLC analysis and mass spectrometry. Subsequently, fresh libraries will be monitored while mechanical agitation is applied through conventional magnetic stirring bars. Finally, the Fmoc glycine substrate used in catalysis produces the highly insoluble DBF. Over time, libraries that contain Fmoc glycine and a catalytically active replicator show significant precipitation. Such precipitates will disrupt plate reader measurements, so a water-soluble analogue is required.

5.2.2 DEVELOPMENT OF A WATER-SOLUBLE Fmoc ANALOGUE

Experiments in which DBF was generated led to significant precipitation in aqueous solution. To be able to utilize the positive feedback loop generated by DBF through the enhancement of thiol oxidation, precipitation must be avoided. The introduction of Darwinian evolution in self-replicating systems requires the extinction of those that do not replicate fast enough – something that can be attained with the use of flow systems²³ or serial dilutions.²⁴ Positive feedback loops would be a crucial implementation in such systems, since replicators able to use them could increase their replication rate and, by extension, their survivability. Precipitation of material is detrimental to such experiments, as it leads to a build-up of hydrophobic material and causes a range of technical problems (clogging of tubing, problems with analysis and interference with mechanical agitation).

In the seventies, the compound 9-(2-sulfo)fluorenylmethyloxycarbonyl (SULFmoc) chloride was developed as a reagent in peptide pu-

rification.²⁵ This compound closely resembles the structure of Fmoc (see **Figure 7**), making it an excellent candidate for substitution. The sulfonic acid group has a reported pK_a of around -1.3, resulting in an almost permanent negative charge on the substrate – especially at the mildly basic conditions employed to facilitate disulfide exchange (see **Chapter 2**). This not only drastically enhances solubility; it could also play a role in substrate binding as the supramolecular structures generally carry many positive charges. Furthermore, the product obtained after SULFMOC deprotection retains the sulfonic acid functional group (and is thus dubbed SULFDBF), eliminating any precipitation issues related to DBF formation.

Figure 7: Fmoc chloride and its sulfonated analogue SULFMOC chloride. The sulfonic acid functional group is shown in red.

The deprotection of SULFMOC glycine was tested using spectroscopic methods. Hexamers constructed from building block **1** were able to catalyse the deprotection reaction, while monomers and oxidized monomers were not (see **Figure 8A**). Furthermore, an emergence experiment was set up, similar to those performed in **Chapter 4**. Here, SULFMOC glycine was mixed with building block **1** in an agitated aqueous solution. The samples were kept at 40 °C and were analysed with UPLC. Indeed, samples that contained the SULFMOC glycine substrate showed enhanced thiol oxidation (see **Figure 8B**), similar to those observed when using Fmoc glycine as substrate. A more quantified approach to SULFMOC synthesis, characterisation and effects as substrate for catalysis was studied elsewhere.²⁶

SULFMOC glycine was deemed an apt candidate for plate reader experiments: it is soluble in aqueous media, can be deprotected by replicators constructed from building block **1** and promotes thiol oxida-

tion. However, it was found that high concentrations of SULFDBF thiol oxidation proceeded so rapidly that the dynamic nature of the library was compromised. Therefore, we employed only a small amount of substrate relative to building block concentrations (1:20) in the plate reader experiments.

Figure 8: SULFMOc glycine as a substrate for catalysis. A) Deprotection of SULFMOc glycine (74 μM) by libraries constructed from building block 1. Monomers (40 μM), oxidized monomers (40 μM), hexamers (20 μM) and the background reaction are compared using UV spectroscopy. Only the self-replicating hexamer is able to catalyse the deprotection reaction. Three repeats were performed, where the measurement error is shown by the shaded areas within the dotted lines. B) Amount of building block 1 present during the course of an emergence experiment (200 μM building block 1) monitored with UPLC. Samples that contain SULFMOc glycine (200 μM) – and by extension SULFDBF

when catalysis is performed – have an increased rate of thiol oxidation. All experiments were performed in borate buffer (200 mM, pH 8.2).

5.2.3 PLATE READER SCREENING

Each 250 μ L well in the 96 well plates contained two different building blocks; a fluorescent probe (ThT) and SULFMOC glycine as a substrate for catalysis. Additionally, each 96 well plate contained three libraries that consisted solely of building block **1** as an internal reference. All 231 combinations of building blocks were run in triplicate, resulting in the screening of 693 libraries (plus an additional 24 libraries due to the internal reference). Each plate was sealed with microplate sealing films to prevent evaporation of aqueous media. When placed in the plate reader, a constant temperature of 40 °C was maintained while agitation of libraries was achieved through orbital shaking. All libraries were monitored for approximately three days. The choice for this time period was based on three factors: libraries constructed from individual building blocks usually form supramolecular structures within two days (see **Supplementary Figures 1-4**). Furthermore, our initial interest went out to replicators that would be able to utilise the positive feedback loop mechanism implemented by catalytic SULFDBF generation, further speeding up replicator formation. A final time constraint was the number of libraries to be screened – a total of eight 96 well plates were required to perform these experiments. Promising candidates (see **Figure 9**) were selected based on the ThT response: it should be increasing over time as this would signify the continuous formation of β -sheets.

Out of the 231 combinations, 21 were considered to produce significant ThT signals over time (see **Figure 10 A-D**). These multi-building block libraries were prepared anew and monitored by UPLC and mass spectrometry to determine the library composition.

Out of these 21 selected candidates, ten libraries lost significant portions of their peak area (up to 100%) when monitored with UPLC (see **Figure 11D** and **Supplementary Figures 6(D, F), 7(D, E), 8(B, E, F)** and **9(A, B, C)**). Remarkably, all six libraries containing building block **11** showed this behaviour (see **Supplementary Figures 6D, 7D, 8(E, F)** and **9(A, B)**). Building block **11** contains a tyrosine residue located at the terminus of the peptide sequence, next to a phenyl alanine residue. This interferes with the β -sheet promoting alternating hydrophobic-hydrophilic amino-acid sequence and results in the formation

of aggregates as evident from visual inspection: samples quickly turn cloudy and material precipitates. Strong interaction with the column results in the observed peak area loss – only at high acetonitrile concentrations can the aggregates be disassembled. This made it impossible to properly quantify which species were present, but does not exclude the possibility of self-sorting behaviour (be it social or narcissistic self-sorting). However, no further effort was put into the use of different experimental conditions that might allow for a proper quantification of systems with significant peak area loss at this time.

Figure 9: Example of plate reader data output. Each reaction mixture (200 mM borate buffer, pH 8.12) consists of 200 μM building blocks (100 μM per building block), 250 μM ThT and 10 μM SULFMOC glycine. The fluorescent signal from ThT (measured at 482 nm) of the first seven mixed libraries is shown over time. From these libraries, two potential candidates are selected due to their strong ThT response (marked with a star). Note that library **1+4** is not selected: there is no strong indication of rapidly increasing β -sheet formation. All graphs are plotted based on three repeats. Standard deviations are shown as shaded areas between the dotted lines.

The remaining eleven libraries all revealed the emergence of mixed hexamer macrocycles that eventually dominated the library (see **Supplementary Figures 6(A-C, E), 7(A-C, F) and 8(A, C, D)**). The main products of the libraries were identified using LC-MS (see **Supplementary Figures 10-20**). We focused on the analysis of the most

abundant peak(s). For each compound multiple charged states were detected. For clarity, only the part of the spectrum corresponding to a single charged state is shown.

Figure 10: Fluorescent signal obtained from all 21 potential candidates from the plate reader screening experiment. Libraries consisting of building block **1** are shown as a reference in each graph. Experimental conditions are the same as described in **Figure 9**. All graphs are plotted based on three repeats. Standard deviations are not shown here for clarity and are shown in **Supplementary Figure 5(A-G)**.

Remarkably, of the 231 screened libraries only a single one showed clear signs of self-sorting behaviour: the library constructed from building blocks **1** and **15**. Here, a preference for the **1₅15₁** macrocycle is observed (see **Supplementary Figures 7A** and **14**). However, when looking at the total peak area from the UPLC chromatograms, we detect a loss of 46% (see **Supplementary Figure 21**). Due to the more hydrophobic nature of building block **15**, the seemingly self-sorting behaviour can be explained by a tendency for **15**-rich macrocycles to precipitate, resulting in a skewed representation of the most predominant species. Of course, this is not necessarily the case: it might very

well be that this combination does show self-sorting behavior, but it is impossible to conclude so from the current experimental data.

Figure 11: Library progression monitored by UPLC of a selection of four building block combinations (100 μ M per building block) that showed a large fluorescent signal during plate reader screening. Mixed hexamers are the dominant products in A–C while D shows 100% peak area loss. Red lines mark the library composition at day zero (dominated by the two monomeric building blocks), green lines at day four and purple lines at day 14. Chromatograms are plotted with an offset (on both the X and Y axis) for clarity. All experiments were performed in borate buffer (200 mM, pH 8.12), were kept at 40 $^{\circ}$ C and stirred at 1200 rpm.

Even though no definitive proof of self-sorting could be discovered during these experiments – it does not mean it cannot occur within this building-block space. The biggest limitation is the short duration of the plate reader experiments. It might very well be the case that three days are not enough for all potential replicators to form. This does not diminish the inherent value of such experiments: it is clear that very large numbers of combinations can be screened for the formation of β -sheet stabilized supramolecular structures with relative ease – the duration of which can easily be adjusted.

5.3 PRELIMINARY RESULTS

The implementation of a positive feedback loop generated through SULFMOC glycine deprotection remains elusive. From the current results, it is unclear whether mixed building block systems are catalytically active and whether they can utilize SULFDBF mediated thiol oxidation. To investigate whether the substrate has any effect on the emergence and replication speed, another plate reader experiment was performed where the eleven most promising candidates were monitored without the addition of the substrate.

Remarkably, we found that the emergence of the main species almost always occurs faster without the addition of SULFMOC glycine. Perhaps this can be explained by a perturbation of the nucleation stage: the interactions between building blocks and the charged substrate might inhibit the formation of the replicating nucleus. As briefly mentioned before, the SULFDBF mediated oxidation can proceed very rapidly. This has been shown to disrupt the dynamic nature of libraries which halts replicator formation. SULFDBF might need to be scavenged during such experiments to ensure the positive feedback loop does not overshoot its purpose.

The large differences in emergence time of the β -sheets, the relatively large errors observed and the curious shapes of these curves (for example, **Figure 12D** and **E**) indicate that the influence of the substrate warrants further investigation. Furthermore, the use of a fluorescent probe to correlate β -sheet formation to the emergence of replicating species should be quantified. Still, the use of these large-scale screening setups can help us to identify self-assembling or replicating species in the vast peptide space generated by multi-building block libraries and are a viable option for assessing catalytic capabilities of their assemblies.

Figure 12: Fluorescent signal obtained from the eleven most promising candidates, with and without the addition of the SULFMOG glycine substrate. Each reaction mixture (200 mM borate buffer, pH 8.12) consists of 200 μ M building blocks (100 μ M per building block), 250 μ M ThT and either 10 μ M SULFMOG glycine or no SULFMOG glycine. The fluorescent signal from ThT (measured at 482 nm) of each mixed library is shown over time. Experiments performed with the addition of SULFMOG glycine were performed in triplicate, while the experiments performed without the substrate were performed in duplicate. Standard deviations are shown as shaded areas between the dotted lines.

5.4 CONCLUSIONS AND OUTLOOK

This chapter reveals an easy, high-throughput method of screening for potential replicators in mixed building block systems. A total of 22 building blocks were mixed to generate 231 unique building block combinations. From this building-block space, a total of eleven mixed hexamer assemblies were identified through plate reader experiments using a fluorescent probe sensitive to β -sheet formation. To discover self-sorting behaviour the duration of plate reader experiments should be lengthened and should be disconnected from the introduction of a positive feedback loop in the system. Indeed, the effects of the SULFMOC glycine substrate on oxidation and replication rate should be investigated further and the reproducibility of the plate reader technique must be verified.

That being said, this method opens up new pathways for the directed screening of mixed building block systems (whether for the capability to replicate or to catalyse chemical reactions) – especially when it is combined with predictive computer modelling (**Chapter 2**). Indeed, the use of fluorescence spectroscopy allows for a more shotgun style approach to discover other chemical reactions that assemblies might catalyse. The only requirement for such experiments is the use of a fluorophore, either as a reactant or product. The retro-aldol reaction discussed in **Chapter 3** could have been discovered in this way, as the 6-methoxy-2-naphthalene product is fluorescently active. Finally, this approach can lead to a form of directed evolution within self-replicating systems: amino acids in the peptide backbone can be modified, altering the resulting active site in the supramolecular structures. The use of high-throughput screening techniques is required to select mutations that are beneficial, after which further rounds of mutations can be applied.

METHODS

General procedures

All reagents, solvents and buffer salts were obtained from commercial suppliers and used without further purification. Building blocks were obtained from Cambridge Peptides Ltd (Birmingham, UK) or GenScript (Piscataway, USA). Building blocks **20–22** were synthesized in-house by Paul Adamski. UPLC analysis was performed on a

Waters Acquity UPLC H-class system equipped with a PDA detector. Analysis were performed on a reversed-phase UPLC column (Aeris Peptide 1.7 μm XB-C18 x 2.10 mm), purchased from Phenomenex. Column temperature was set at 35 °C and the sample holder was kept at 25 °C unless otherwise noted. The UV absorbance was monitored at 254 nm. UPLC-MS experiments were performed on a Waters Acquity UPLC H-class system coupled to a Waters Xevo-G2 TOF. The mass spectrometer was operated in positive electrospray ionization mode. Capillary, sampling cone and extraction cone voltages were kept at 2.5 kV, 30V and 4V respectively. Source and desolvation temperatures were set at 140 °C and 500 °C. Nitrogen was used as both cone (5 L/h) and desolvation gas (500 L/h). The mass range selected was 300-3000 m/z.

UPLC separation

The gradient used for the UPLC separation of dynamic combinatorial libraries is shown below. The same gradient was used for all samples. A = UPLC grade H₂O containing 0.1 v% TFA, B = UPLC grade MeCN containing 0.1 v% TFA. The flow rate was kept at 0.300 mL/minute throughout the analysis.

Negative staining transmission electron microscopy

Transmission electron microscopy was performed on a Philips CM120 electron microscope operating at 120 kV. Images were recorded on a slow scan CCD camera (Gatan). Micrographs were analyzed using ImageJ software. Samples were prepared by depositing a small aliquot (5 μL) on a 400 mesh copper grid covered with a thin carbon film (van Loenen instruments). After one minute, the grid was blotted on filter paper. The sample was subsequently stained twice with uranyl acetate (2%) by depositing 5 μL of a saturated uranyl acetate solution on the grid and blotting it on filter paper after incubation for 30 seconds.

Borate buffer preparation

Concentrations of borate buffers are reported in concentration of boron atoms. Borate buffer is prepared by dissolving anhydrous borax (Na₂B₄O₇) in doubly distilled water to a concentration of 250 mM.

The pH is subsequently adjusted to 8.2 by addition of 1M HCl. The solution is subsequently diluted to a final concentration of either 50 or 200 mM.

Preparation of libraries

Concentrations of library constituents are reported with respect to their building blocks, unless otherwise specified. In general, mixed libraries were prepared by mixing stock solutions of individual libraries. Stock solutions were always prepared by adding a building block (1-3 mg, 1-3 μ mol) to a UPLC vial (12 x 32 mm). The building block was dissolved in borate buffer (200 mM, 1-2 mL) to give a final concentration of 2.08 mM in building block.

Subsequently, equal volumes of stock solutions (300 μ L) were mixed in a UPLC vial (12 x 32 mm) containing a Teflon coated magnetic stirring bar (5 x 2 mm, VWR). The vial was closed with a Teflon septum screw cap and stirred at 1200 rpm at 40 °C unless otherwise specified. The temperature of the reaction mixtures was controlled using a Thermo Fisher compact dry bath with a custom milled aluminium block to accommodate HPLC vials (with dimensions 12 x 32 mm). Libraries constituting of single building blocks were prepared in the same way, where a volume of 600 μ L from the stock solution was used instead.

Plate reader experiments

All plate reader experiments were performed on a Biotek Synergy HTX plate reader. Thermo Scientific MicroWell 96-well plates (8 x 12) with optical bottom (purchased from Thermo Fischer) were used for fluorescence measurements.

Each well had a total volume of 250 μ L. Each well contained 200 μ M in building block(s), 250 μ M Thioflavin T and either 10 μ M or no SULF-MOC glycine. All experiments were performed in 200 mM borate buffer (pH 8.12). Stock solutions of building blocks were prepared as described by Methods: *Preparation of libraries*. Mixed libraries inside the 96 microwell plates were constructed using a multichannel pipette and various reservoirs (containing building block stock solutions, ThT and SULFMOC glycine). After combining all reagents in each microwell, the plates were sealed using water-tight microplate sealing film.

Subsequently, microwell plates were immediately inserted in the plate reader.

Sample temperature was kept at 40 °C and samples were agitated by continuous orbital shaking (282 cycles per minute). An excitation wavelength of 440 nm was used and emission was monitored at 482 nm. All mixed libraries were monitored for approximately three days.

ACKNOWLEDGEMENTS

J. Ottel, C. Mayer and S. Otto conceived the project. J. Ottel performed all experiments unless otherwise specified. J. Bos was involved with the experiments related to the characterisation of building blocks **2-8**. K. van Esterik synthesized and characterized all compounds related to the SULFMOc protecting group, and tested the catalytic activity of building block **1** towards its deprotection. M. J. Eleveld designed building blocks **15-19**. P. R. Adamski designed and synthesized building blocks **20-22**. G. P. R. Schaeffer and M. J. Eleveld are thanked for the proofreading of this chapter and the ensuing discussions. B. Liu and M. J. Eleveld are gratefully acknowledged for mass spectrometry measurements.

BIBLIOGRAPHY

1. Darwin, C. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. *Br. Foreign Med. Chir. Rev.* **1860**, 25, 367–404.
2. Bedau, M. A., Cleland, C. E. A universal definition of life: autonomy and open-ended evolution. in *The Nature of Life* 310–325 (Cambridge University Press, **2010**).
3. Adamski, P., Eleveld, M., Sood, A., Kun, Á., Szilágyi, A., Czárán, T., Szathmáry, E., Otto, S. From self-replication to replicator systems en route to de novo life. *Nat. Rev. Chem.* **2020**, 4, 386–403.
4. Duim, H., Otto, S. Towards open-ended evolution in self-replicating molecular systems. *Beilstein J. Org. Chem.* **2017**, 13, 1189–1203.
5. Moreno, A., Ruiz-Mirazo, K. The problem of the emergence of functional diversity in prebiotic evolution. *Biol. Philos.* **2009**, 24, 585–605.
6. Szathmáry, E., Gladkih, I. Sub-exponential growth and coexistence of non-enzymatically replicating templates. *J. Theor. Biol.* **1989**, 138, 55–58.
7. Lifson, S., Lifson, H. Coexistence and Darwinian selection among replicators: Response to the preceding paper by Scheuring and Szathmary. *J. Theor. Biol.* **2001**, 212, 107–109.
8. Colomb-Delsuc, M., Mattia, E., Sadownik, J. W., Otto, S. Exponential self-replication enabled through a fibre elongation/breakage mechanism. *Nat. Commun.* **2015**, 6, 7427.
9. Fleming, S., Ulijn, R. V. Design of nanostructures based on aromatic peptide amphiphiles. *Chem. Soc. Rev.* **2014**, 43, 8150–8177.
10. Frederix, P. W. J. M., Idé, J., Altay, Y., Schaeffer, G., Surin, M., Beljonne, D., Bondarenko, A. S., et al. Structural and spectroscopic properties of assemblies of self-replicating peptide macrocycles. *ACS Nano*, **2017**, 11, 7858–7868.
11. Malakoutikhah, M., Peyralans, J. J.-P., Colomb-Delsuc, M., Fanlo-Virgós,

H., Stuart, M. C. A., Otto, S. Uncovering the Selection Criteria for the Emergence of Multi-Building-Block Replicators from Dynamic Combinatorial Libraries. *J. Am. Chem. Soc.* **2013**, 135, 18406–18417.

12. Sadownik, J. W., Mattia, E., Nowak, P., Otto, S. Diversification of self-replicating molecules. *Nat. Chem.* **2016**, 8, 264–269.

13. Ottel, J., Hussain, A. S., Mayer, C., Otto, S. Chance emergence of catalytic activity and promiscuity in a self-replicator. *Nat. Catal.* **2020**, 3, 547–553.

14. Groenning, M. Binding mode of Thioflavin T and other molecular probes in the context of amyloid fibrils-current status. *J. Chem. Biol.* **2010**, 3, 1–18.

15. Gade Malmos, K., Blancas-Mejia, L. M., Weber, B., Buchner, J., Ramirez-Alvarado, M., Naiki, H., Otzen, D. ThT 101: a primer on the use of thioflavin T to investigate amyloid formation. *Amyloid*, **2017**, 24, 1–16.

16. Dns, F., Pichowicz, M., Povie, G., Renaud, P. Thiyl Radicals in Organic Synthesis. *Chem. Rev.* **2014**, 114, 2587–2693.

17. Altay, M., Altay, Y., Otto, S. Parasitic Behavior of Self-Replicating Molecules. *Angew. Chem. Int. Ed.* **2018**, 57, 10564–10568.

18. Altay, Y., Tezcan, M., Otto, S. Emergence of a new self-replicator from a dynamic combinatorial library requires a specific pre-existing replicator. *J. Am. Chem. Soc.* **2017**, 139, 13612–13615.

19. Altay, Y., Altay, M., Otto, S. Existing Self-Replicators Can Direct the Emergence of New Ones. *Chem. Eur. J.* **2018**, 24, 11911–11915.

20. Bos, J. E. Catalytic sites on self-replicating fibres. (Rijksuniversiteit Groningen, **2019**).

21. Liu, B., Pappas, C. G., Zangrando, E., Demitri, N., Chmielewski, P. J., Otto, S. Complex Molecules That Fold Like Proteins Can Emerge Spontaneously. *J. Am. Chem. Soc.* **2019**, 141, 1685–1689 .

22. Carnall, J. M. A., Waudby, C. A., Belenguer, A. M., Stuart, M. C. A., Peyralans, J. J.-P., Otto, S. Mechanosensitive Self-Replication Driven by Self-Organization. *Science*, **2010**, 327, 1502–1506.

23. Mattia, E., Otto, S. Supramolecular systems chemistry. *Nat. Nanotechnol.* **2015**, 10, 111–119.

24. Hordijk, W., Vaidya, N., Lehman, N. Serial transfer can aid the evolution of autocatalytic sets. *J. Syst. Chem.* **2014**, 5, 1–8.

25. Merrifield, R. B., Bach, A. E. 9-(2-Sulfo)fluorenylmethyloxycarbonyl Chloride, a New Reagent for the Purification of Synthetic Peptides. *J. Org. Chem.* **1978**, 43, 4808–4816.

26. van Esterik, K. (Sul)Fmoc deprotection as a central strategy to incorporate proto-metabolism and compartmentalisation in a system of synthetic replicators. Master thesis - Rijksuniversiteit Groningen (Rijksuniversiteit

SUPPLEMENTARY FIGURES

Supplementary Figure 1: Library progression of building blocks 2-5 during the first eight days. All libraries were stirred, prepared and operated as written in Methods: Preparation of libraries. The coloured lines represent different stages of library progression. Chromatograms are plotted with an offset (on both the X and Y axis) for clarity.

Supplementary Figure 2: Library progression constructed from building blocks **6**, **7** or **8** during the first eight days. All libraries were stirred, prepared and operated as written in Methods: Preparation of libraries. The library constructed from building block **6** shows peak area loss – often observed by more hydrophobic building blocks. The coloured lines represent different stages of library progression. Chromatograms are plotted with an offset (on both the X and Y axis) for clarity.

Supplementary Figure 3: Supramolecular structures observed during negative stain TEM after 14 days of library progression. Only libraries that are constructed from building block **3** show amorphous aggregates. All other libraries show the formation of fibrillar supramolecular structures.

Supplementary Figure 4: Supramolecular structures observed in negative stain TEM after 14 days of library progression. All libraries show the formation of fibrillar supramolecular structures.

Supplementary Figure 5: Standard deviations (shown as shaded areas between the dotted lines) of the fluorescent signals shown in **Figure 10**.

Supplementary Figure 6: Library progression of building block combinations that showed a large fluorescent signal during plate reader screening. Mixed hexamers are the dominant product in the libraries in A, B, C, D and E, while large peak area losses are observed in D and F. Red lines mark the library composition at day zero, green lines at day four and purple lines at day 14. Chromatograms are plotted with an offset (on both the X and Y axis) for clarity. Experimental conditions are the same as described in **Figure 11**.

Supplementary Figure 7: Library progression of building block combinations that showed a large fluorescent signal during plate reader screening. Mixed hexamers are the dominant product in the libraries in A, B, C and F, while large peak area losses are observed in D and E. A non-statistical distribution of mixed hexamers is observed in library A, where the **1₅15₁** hexamer is predominantly formed. Red lines mark the library composition at day zero, green lines at day four and purple lines at day 14. Chromatograms are plotted with an offset (on both the X and Y axis) for clarity. Experimental conditions are the same as described in **Figure 11**.

Supplementary Figure 8: Library progression of building block combinations that showed a large fluorescent signal during plate reader screening. Mixed hexamers are the dominant product in the libraries in A, C and D, while large peak area losses are observed in B, E and F. Red lines mark the library composition at day zero, green lines at day four and purple lines at day 14. Chromatograms are plotted with an offset (on both the X and Y axis) for clarity. Experimental conditions are the same as described in **Figure 11**.

Supplementary Figure 9: Library progression of building block combinations that showed a large fluorescent signal during plate reader screening. Large peak area losses are observed for all combinations. Red lines mark the library composition at day zero, green lines at day four and purple lines at day 14. Chromatograms are plotted with an offset (on both the X and Y axis) for clarity. Experimental conditions are the same as described in **Figure 11**.

Observed mass	Calculated mass	Macrocycle
1115.89	1116.00	$[2_6+4H]^{4+}$
1119.62	1119.50	$[1_1 2_5+4H]^{4+}$
1123.12	1123.00	$[1_2 2_4+4H]^{4+}$
1127.37	1126.50	$[1_3 2_3+4H]^{4+}$
1130.13	1130.01	$[1_4 2_2+4H]^{4+}$
1133.63	1133.51	$[1_5 2_1+4H]^{4+}$
1137.12	1137.01	$[1_6+4H]^{4+}$

Supplementary Figure 10: Chromatogram and mass spectrum obtained by LC-MS of a library constructed from building blocks **1** and **2** after nine days. A) Chromatogram showing the main product of the library. B) Mass spectrum identifying a statistical distribution of hexamers. The table shows the observed mass, calculated mass and corresponding macrocycle.

Observed mass	Calculated mass	Macrocycle
1137.16	1137.01	$[1_6+4H]^{4+}$
1143.99	1144.01	$[1_5 8_1+4H]^{4+}$
1151.16	1151.02	$[1_4 8_2+4H]^{4+}$
1158.16	1158.02	$[1_3 8_3+4H]^{4+}$
1165.16	1165.02	$[1_2 8_4+4H]^{4+}$
1172.17	1172.02	$[1_1 8_5+4H]^{4+}$
1179.16	1179.03	$[8_6+4H]^{4+}$

Supplementary Figure 11: Chromatogram and mass spectrum obtained by LC-MS of a library constructed from building blocks **1** and **8** after nine days. A) Chromatogram showing the main product of the library. B-H) Mass spectrum identifying a statistical distribution of hexamers. The table shows the observed mass, calculated mass and corresponding macrocycle.

Observed mass	Calculated mass	Macrocycle
1137.17	1137.01	$[1_6+4H]^{4+}$
1149.67	1149.52	$[1_5 10_1+4H]^{4+}$
1162.18	1162.02	$[1_4 10_2+4H]^{4+}$
1174.68	1174.53	$[1_3 10_3+4H]^{4+}$
1187.19	1187.03	$[1_2 10_4+4H]^{4+}$
775.12	775.01	$[1_2 10_1+3H]^{3+}$
1199.69	1199.54	$[1_1 10_5+4H]^{4+}$
791.80	791.69	$[1_1 10_2+3H]^{3+}$
1212.20	1212.04	$[10_6+4H]^{4+}$

Supplementary Figure 12: Chromatogram and mass spectrum obtained by LC-MS of a library constructed from building blocks **1** and **10** after nine days. A) Chromatogram showing the main product of the library. B-J) Mass spectrum identifying a statistical distribution of hexamers and other products. The table shows the observed mass, calculated mass and corresponding macrocycle.

Observed mass	Calculated mass	Macrocycle
909.55	909.41	$[12_4+3H]^{3+}$
934.89	934.76	$[1_1 12_3+3H]^{3+}$
1061.14	1060.98	$[1_2 12_4+4H]^{4+}$
682.41	682.31	$[12_3+3H]^{3+}$
960.24	960.10	$[1_2 12_2+3H]^{3+}$
1080.15	1079.99	$[1_3 12_3+4H]^{4+}$
1099.16	1099.00	$[1_4 12_2+4H]^{4+}$
1061.14	1060.98	$[1_2 12_1+2H]^{2+}$
1118.17	1118.00	$[1_5 12_1+4H]^{4+}$
1137.18	1137.01	$[1_6+4H]^{4+}$
1137.18	1137.01	$[1_3+2H]^{2+}$

Supplementary Figure 13: Chromatogram and mass spectrum obtained by LC-MS of a library constructed from building blocks **1** and **12** after nine days. A) Chromatogram showing the main product of the library. B-L) Mass spectrum identifying the formation of hexamers and other products. The table shows the observed mass, calculated mass and corresponding macrocycle.

Observed mass	Calculated mass	Macrocycle
1137.19	1137.01	$[1_6+4H]^{4+}$
1145.69	1145.69	$[1_515_1+4H]^{4+}$
1154.19	1154.01	$[1_415_2+4H]^{4+}$
1162.69	1162.50	$[1_315_3+4H]^{4+}$
1171.19	1171.00	$[1_210_4+4H]^{4+}$
1179.69	1179.50	$[1_115_5+4H]^{4+}$
758.47	758.34	$[1_3+3H]^{3+}$
769.80	769.67	$[1_215_1+3H]^{3+}$
781.12	781.00	$[1_115_2+3H]^{3+}$

Supplementary Figure 14: Chromatogram and mass spectrum obtained by LC-MS of a library constructed from building blocks **1** and **15** after nine days. A) Chromatogram showing the main product of the library. B-J) Mass spectrum identifying the formation of hexamers and other products. The table shows the observed mass, calculated mass and corresponding macrocycle.

Here, we observe a non-statistical distribution of hexamers where the **1₅15₁** hexamer is preferred.

Observed mass	Calculated mass	Macrocycle
1116.17	1116.00	$[2_6+4H]^{4+}$
1126.69	1126.50	$[2_57_1+4H]^{4+}$
1137.20	1137.01	$[2_47_2+4H]^{4+}$
1147.71	1147.51	$[2_37_3+4H]^{4+}$
1158.46	1158.02	$[2_27_4+4H]^{4+}$
1168.72	1168.52	$[2_17_5+4H]^{4+}$
1179.23	1179.03	$[7_6+4H]^{4+}$

Supplementary Figure 15: Chromatogram and mass spectrum obtained by LC-MS of a library constructed from building blocks **2** and **7** after nine days. A) Chromatogram showing the main product of the library. B) Mass spectrum identifying a statistical distribution of hexamers. The table shows the observed mass, calculated mass and corresponding macrocycle.

Observed mass	Calculated mass	Macrocycle
1116.19	1116.00	$[2_6+4H]^{4+}$
1126.70	1126.70	$[2_5 8_1+4H]^{4+}$
1137.70	1137.70	$[2_4 8_2+4H]^{4+}$
1147.71	1147.71	$[2_3 8_3+4H]^{4+}$
1158.22	1158.22	$[2_2 8_4+4H]^{4+}$
1168.72	1168.72	$[2_1 8_5+4H]^{4+}$

Supplementary Figure 16: Chromatogram and mass spectrum obtained by LC-MS of a library constructed from building blocks **2** and **8** after nine days. A) Chromatogram showing the main product of the library. B) Mass spectrum identifying a statistical distribution of hexamers. The table shows the observed mass, calculated mass and corresponding macrocycle.

Observed mass	Calculated mass	Macrocycle
1182.25	1181.78	$[6_1 7_5+4H]^{4+}$
1184.75	1184.53	$[6_2 7_4+4H]^{4+}$
1188.50	1188.28	$[6_3 7_3+4H]^{4+}$
1190.25	1190.04	$[6_4 7_2+4H]^{4+}$
1193.25	1192.79	$[6_5 7_1+4H]^{4+}$

Supplementary Figure 17: Chromatogram and mass spectrum obtained by LC-MS of a library constructed from building blocks **6** and **7** after nine days. A) Chromatogram showing the main product of the library. B) Mass spectrum identifying a statistical distribution of hexamers. The table shows the observed mass, calculated mass and corresponding macrocycle.

Observed mass	Calculated mass	Macrocycle
810.53	810.36	$[10_1+1H]^{1+}$
1198.51	1198.29	$[6_110_5+4H]^{4+}$
1201.26	1201.04	$[6_210_4+4H]^{4+}$
1204.01	1203.79	$[6_310_3+4H]^{4+}$
1207.01	1206.54	$[6_410_2+4H]^{4+}$
801.17	801.03	$[6_210_1+3H]^{3+}$
1209.52	1209.29	$[6_510_1+4H]^{4+}$
804.84	804.69	$[6_110_2+4H]^{4+}$
808.51	808.36	$[10_3+3H]^{3+}$

Supplementary Figure 18: Chromatogram and mass spectrum obtained by LC-MS of a library constructed from building blocks **6** and **10** after nine days. A) Chromatogram showing the main product of the library. B-J) Mass spectrum identifying the formation of hexamers and other products. The table shows the observed mass, calculated mass and corresponding macrocycle.

Observed mass	Calculated mass	Macrocycle
1179.25	1179.03	$[7_x 8_{6-x} + 4H]^{4+}$

Supplementary Figure 19: Chromatogram and mass spectrum obtained by LC-MS of a library constructed from building blocks **7** and **8** after nine days. A) Chromatogram showing the main product of the library. B) Mass spectrum identifying only that there are hexamer species, as both **7** and **8** have the same exact mass. The table shows the observed mass, calculated mass and corresponding macrocycle.

Observed mass	Calculated mass	Macrocycle
930.92	930.75	$[9_4+3H]^{3+}$
960.28	960.06	$[7_1 9_3+3H]^{3+}$
1047.16	1046.97	$[9_3+2H]^{2+}$
989.63	989.44	$[7_2 9_2+3H]^{3+}$
1113.21	1113.00	$[7_3 9_3+4H]^{4+}$
1091.19	1090.99	$[7_1 9_2+2H]^{2+}$
1135.22	1135.01	$[7_4 9_2+4H]^{4+}$
1157.24	1157.02	$[7_5 9_1+4H]^{4+}$
757.15	757.00	$[7_2 9_1+3H]^{3+}$
1179.25	1179.03	$[7_6+4H]^{4+}$
1179.26	1179.03	$[7_3+2H]^{2+}$

Supplementary Figure 21: Chromatogram and mass spectrum obtained by LC-MS of a library constructed from building blocks **7** and **9** after nine days. A) Chromatogram showing the main product of the library. B-L) Mass spectrum identifying the formation of hexamers and other products. The table shows the observed mass, calculated mass and corresponding macrocycle.

Supplementary Figure 21: The library constructed from building blocks **1** and **15** showed apparent self-sorting behaviour. However, since large peak area losses are observed during the monitoring of library composition by UPLC, the self-sorting behaviour could be explained by the precipitation of macrocycles incorporating the more hydrophobic building block **15**.

CHAPTER 6

Conclusions and future perspectives

This final chapter connects the main results of all previous chapters and in doing so reveals perspectives on future research directions. Attempts were made to withhold from repeating the conclusions from each individual chapter but it is insurmountable to achieve it whilst maintaining a coherent independent story. If this entire thesis would have to be summarized in one sentence, one could state that “the work in this dissertation has increased the life-likeness of the systems that we started out with and, as is common in the scientific community, has raised more questions than it has answered”.

The path between inanimate systems and truly living systems is long, winding and has an ill-defined end – and arguably an ever more elusive beginning. As our research group is mainly interested in the synthesis of *de novo* life, the lack of an all-encompassing definition and theory of life makes it difficult to determine when this goal would have been achieved. Instead of focusing on the current widely accepted definitions of life, we identify which functional machineries are essential to known living systems and attempt to implement these in our designed synthetic systems. In brief, these can be considered to be self-replication, metabolism and compartmentalization (see **chapter one**).

The starting point of this dissertations are the previously developed self-replicating systems that can be formed through dynamic combinatorial chemistry and are under thermodynamic control.^{1,2} The building-blocks of these replicators consist of a pentapeptide appended to a small benzene-1,3-dithiol core moiety functionalized with two thiol groups. Upon oxidation and equilibration, larger macrocyclic species emerge that form supramolecular polymers capable of catalysing their own formation – even allowing for exponential growth through an elongation-breakage mechanism.³ Remarkably, the exact chemical processes involved in the mechanism of self-replication are still unknown. After an introduction that allows the reader to familiarise him or herself with the concepts and systems presented in this dissertation, **chapter two** discusses detailed mechanistic studies that are aimed to uncover the many parameters involved in

the mechanism of self-replication which, in turn, can be used in the development of a computer model. Such models are essential tools for the future of this field due to the complexity of the issue at hand: to find replicators capable of coexisting with one another in a single system (see below). Since we have an enormous peptide space available to us, the use of computer modelling to predict which building-blocks might self-sort (or show other interesting behaviour) is indispensable – there are simply too many possible combinations to tackle this challenge experimentally.

Meanwhile, we have revealed that the mechanism of self-replication is much more complex than initially presumed. Using high-speed atomic force microscopy (HS-AFM), we were able to directly visualise the self-replication process on a surface. We found that replicator precursors can aggregate in solution and directly adsorb on the replicator supramolecular structure (see **chapter three**). This has two important implications. First of all, it turns a three-dimensional search problem into a one-dimensional one: fibre ends do not need to find precursor molecules; the whole fibre acts as a scaffold to which aggregates of precursor molecules can bind. Subsequently, these molecules diffuse along the fibre backbone towards the fibre ends, where they can be converted to replicator molecules. This explains the high growth rates observed in our systems – even at micromolar concentrations of building-blocks we observe the fast conversion of precursors to replicator. Furthermore, the existence of such reservoir aggregates suggests that there is already some form of compartmentalization taking place during the replication process. This is an especially promising development when combined with the results obtained in **chapter four**.

There, we have discovered that the construction of supramolecular polymers inadvertently leads to the generation of active sites along the fibre surface. Upon stacking of macrocycles, the lysine residues that are incorporated into the pentapeptide sequence of the building-blocks are brought into close proximity to one another, resulting in a perturbation of their pK_a by approximately three orders of magnitude. This gives rise to the formation of active sites all along the hydrophobic backbone of the fibres in which deprotonated primary amines act as catalytic residues. We found that the self-replicators were able to catalyse two chemical reactions: the retro-aldol reaction and fluorenylmethoxycarbonyl (Fmoc) deprotection. The retro-aldol reaction progresses through the formation of an iminium intermediate, while

Fmoc deprotection requires a hydrogen abstraction reaction – both provided by the deprotonated amines. Interestingly, these reactions only occur when the replicating macrocycle self-assembles: i.e. the precursor molecules show almost no catalytic efficiency towards either of these reactions, indicating that the pKa of their lysine residues are not sufficiently affected by potential intramolecular interactions between non-assembled macrocycles. While the retro-aldol reaction does not influence the process of self-replication, the Fmoc deprotection reaction certainly does. Dibenzofulvene (DBF), one of the products generated by the Fmoc cleavage surprisingly promotes thiol oxidation – probably through a radical stabilisation mechanism.⁴ This results in the implementation of a positive feedback loop in the mechanism of self-replication: as the self-replicating macrocycles start to stack and form supramolecular polymers, they become catalytically active and start to produce not only copies of themselves, but also start Fmoc cleavage, resulting in DBF formation. This in turn enhances the production of replicator precursors through its ability to promote thiol oxidation. With the discovery of this feedback loop, the first steps towards the implementation of proto-metabolic activity in self-replicating systems have been taken.

Combining these results with the precursor aggregate formation discovered in chapter three, we can already start to take advantage of the potential benefits derived from compartmentalisation. Since the precursor molecules attach to the fibre surface in aggregates and catalysis takes place along the fibre surface, one can envision that the products of catalysis will also coalesce in these aggregates. Specifically, the hydrophobic DBF would be encapsulated in these aggregates, where the local high concentration of precursor molecules would benefit most from the DBF mediated amplification of their rate of oxidation. More importantly, this might mean that replicators that are most catalytically active would also benefit from the resulting positive feedback loop the most. In other words, the replicators that develop the additional function of catalysis will also enjoy the fruits of their labor.

For this to work in a practical sense, efforts should be taken to develop systems of replicators that are able to coexist with one another, yet retain their differences. This could be achieved by mixed building-block systems containing replicators that show narcissistic self-sorting behavior and thus do not compete for the same building-blocks (or do so in a minimal sense). Otherwise, building-block

incorporation would quickly deteriorate into a statistical distribution, destroying any distinction between replicators that there might have been at the onset of the experiment. Furthermore, the aggregates on the fibre surface should preferably contain the building-blocks that the fibre can convert into replicator molecules – not the building-blocks used by other replicators in the same system. Solving this issue is not straight-forward: building blocks do not tend to easily detach from the fibre surfaces after adsorption. However, perhaps they do not need to. It is unclear what happens to unusable building-blocks attached to fibre surfaces. Do they inhibit further fibre elongation? Or do they simply stick to the surface? In the latter case, the replicators that do need those building-blocks to grow would simply go extinct faster. In the former case, perhaps synthetic adjustments to building-blocks are required to enable easier detachment from the fibre surface – for example through the use of electrostatic repulsion between charged residues in different building blocks.

We have developed a new method to screen multi building-block libraries for the ability to form supramolecular structures stabilized by β -sheet formation while indirectly probing catalytic activity (see chapter five). Using plate-reader technology, we were able to screen hundreds of libraries in a short timeframe. Through such a shotgun approach, we envisioned the discovery of self-sorting behaviour in mixed systems. Remarkably, no clear evidence of self-sorting was observed. This was most likely caused by the strict selection applied during the experiment: the libraries were only monitored for three days – mostly because we were interested in their ability to utilize the positive feedback loop through the use of catalysis. Current investigations have revealed signs of self-sorting behaviour in mixed building-block systems studied in this dissertation – but only after the prolonged monitoring of libraries at elevated temperatures.

Curiously, not a single combination of building blocks was shown to be able to utilize the positive feedback loop more efficiently than libraries constructed solely from building block **1**. Indeed, the positions and charge of the two lysine residues seem crucial to facilitate FMOC cleavage efficiently. Current efforts employ density functional theory (DFT) calculations to compare the free energy barriers of possible transitions states to investigate this hypothesis further. It is curious that the first attempt at the integration of self-replication and catalysis immediately yielded the best result – out of a great many possibilities with the building-blocks present in our lab.

THE FUTURE AND BEGINNING OF SYNTHETIC LIFE

So where do we go from here? With the integration of a proto-metabolism and self-replication, the next functional system is suddenly within reach. The use of FMOC protecting groups on surfactant (or co-acervate) forming molecules would couple replication, proto-metabolism and compartmentalisation in a single system. This would require the FMOC protected molecules to be inactive – only upon deprotection should they start to form compartments. Furthermore, the proto-metabolism can be expanded to generate the replicator building blocks in a similar fashion: using protecting groups to prevent them from forming replicating macrocycles.

However, the proto-metabolism must be expanded much further. So far, we have only achieved semi-catabolic reactions. As of yet there are no energy harvesting machineries in play. Additionally, bond forming reactions still remain elusive. Screening for other chemical transformations – such as the forward aldol reaction – are currently being conducted. The use of other amino acids in the pentapeptide (such as proline) might also lead to an expanded catalytic repertoire.

Finally, mutations within replicators can be achieved by using multi building-block systems and employing the different stacking possibilities to generate different types of active sites. For this purpose, replicator building blocks that are dissimilar to those we already employ should be developed to enable stronger self-sorting behaviour.

Even though we might not really understand what life truly is (yet), the *de novo* synthesis of life is gradually progressing towards more life-likeness in its systems – not dissimilar to how inanimate systems might have progressed into life-like systems billions of years ago.

Even though the concept of "life" is incredibly difficult to define, I think that in the end, we will know it when we see it.

BIBLIOGRAPHY

1. Carnall, J. M. A., Waudby, C. A., Belenguer, A. M., Stuart, M. C. A., Peyralans, J. J.-P. & Otto, S. Mechanosensitive Self-Replication Driven by Self-Organization. *Science*, **2010**, 327, 1502–1506.
2. Malakoutikhah, M., Peyralans, J. J. P., Colomb-Delsuc, M., Fanlo-Virg??s, H., Stuart, M. C. A. & Otto, S. Uncovering the selection criteria for the emergence of multi-building-block replicators from dynamic combinatorial libraries. *J. Am. Chem. Soc.* **2013**, 135, 18406–18417.
3. Colomb-Delsuc, M., Mattia, E., Sadownik, J. W. & Otto, S. Exponential self-replication enabled through a fibre elongation/breakage mechanism. *Nat. Commun.* **2015**, 6, 7427.
4. Dns, F., Pichowicz, M., Povie, G. & Renaud, P. Thiyl Radicals in Organic Synthesis. *Chem. Rev.* **2014**, 114, 2587–2693.

Summary

This dissertation is focused on two major investigations. The first one is about the mechanism with which a known self-replicator can create copies of itself. This study led to the first real-time visualization of a molecule capable of self-replication. The second is about the introduction of additional functionalities within self-replicating systems of molecules. Here, we were able to show emergent behavior: catalytic activity that would only start when the self-replication process takes place.

Chapter 1 gives a general introduction to the field, that starts with the transition between inanimate and animate matter. What separates the two? We briefly discuss the problems encountered when attempting to define life and quickly move towards an alternative: describing life by listing the required functional subsystems common to all known living entities. How can these subsystems (self-replication, metabolism and compartmentalization) be approached synthetically? We go on with an introduction to a powerful toolbox in the repertoire of a chemist: dynamic combinatorial chemistry. Finally, we discuss how multiple subsystems could be incorporated into a single synthetic system.

Chapter 2 further discusses an established peptide-based self-replicator that is central to this dissertation. We look at the importance of the amino acids chosen and their effects on replicator behavior. It becomes apparent that a computer model is required to be able to comprehensively investigate the tremendous structure space of these systems. We identify and quantify the parameters involved in the self-replication mechanism which requires a strict control over experimental variables. This led to the development of a 3D printed stirring device that can be used inside an ultra-performance liquid chromatography system. We conclude with the preliminary results of some experiments that probe the elusive elements of nucleation and fragmentation.

In **Chapter 3** we obtain new insights into the mechanism of the self-assembly driven self-replication process through the use of high-speed atomic force microscopy. We discovered that material

that acts as a precursor to the molecular self-replicator can form aggregates that can attach to the sides of the fibrillar self-replicator. Subsequently, this material is consumed as it diffuses towards the growing fibre ends. The state-of-the-art microscopy technique has revealed to us an entire new mechanism of supramolecular polymerization that transformed a three-dimensional search problem (how does a replicator fibre-end find material to replicate in solution?) into a one-dimensional search problem (precursor material can simply diffuse along the fibre towards the growing end).

Catalysis is one of the most essential functionalities in living systems. **Chapter 4** shows how catalytic activity and promiscuity in a self-replicator is introduced through co-option: certain features obtained through the selected amino-acids in the self-replicator that promote self-replication inadvertently led to a supramolecular arrangement that generated micro-environments suitable for catalysis. Importantly, the catalytic activity shown by the self-replicator does not occur with the building blocks it is built up out of. We discovered two chemical reactions that the self-replicator can catalyse: the retro-aldol cleavage of methodol and the cleavage of fluorenylmethoxycarbonyl groups. We have been able to show that the latter resulted in a positive feedback on the speed of replication through the generation of dibenzofulvene that can enhance thiol oxidation. This is a first step towards a full-fledged metabolism in synthetic systems.

Chapter 5 discusses the establishment of mutations and diversification in self-replicating systems. One approach to attain this is through the use of mixed building block systems that can – potentially – show self-sorting behavior. We investigate a new analytical tool that allows us to screen large numbers of such mixed systems by monitoring a key stabilizing factor of our self-replicators: β -sheet formation. By using a microplate photometer and a fluorescent probe, we can screen hundreds of combinations per day. Additionally, this tool can also be used to screen for catalytic reactions of known self-replicators by using fluorescently active substrates or products.

Finally, **Chapter 6** connects the main results of the previous chapters and shows a perspective on future research directives. Most importantly, future work should focus on the use of Fmoc derivatives on surfactant (or coacervate) forming molecules. This would couple

replication, a proto-metabolism and compartmentalization in a single system. Of course, the proto-metabolism should be expanded so that self-replicator mediated catalysis will result in its building blocks. Bond forming reactions would be a major step in the right direction.

With every experiment, we are moving towards systems that are gradually becoming more life-like, and I am confident that a system that contains all functional subsystems will not remain out of our reach for long.

Samenvatting

Dit proefschrift is gericht op twee primaire onderzoeken. Het eerste gaat over het mechanisme waarmee een bekende zelf-replicator kopieën van zichzelf kan maken. Dit onderzoek heeft geleid tot de eerste *real time* visualisatie van een molecuul dat in staat is tot zelf-replicatie. Het tweede onderzoek gaat over de introductie van extra functionaliteiten binnen zelf-replicerende systemen van moleculen. Hier waren we in staat om emergent gedrag te laten zien: katalytische activiteit die pas op gang komt wanneer het zelf-replicatie proces plaatsvindt.

Hoofdstuk 1 geeft een algemene inleiding tot het vakgebied, dat begint met de overgang tussen levenloze en levende materie. Wat onderscheidt de twee? We bespreken kort de problemen die zich voordoen bij pogingen om het leven te definiëren en gaan over op een alternatief: we beschrijven het leven door de vereiste functionele subsystemen die alle bekende levende entiteiten op onze planeet hebben op te sommen. Hoe kunnen deze subsystemen (zelf-replicatie, metabolisme en compartimentering) synthetisch worden benaderd? We gaan verder met de introductie van een belangrijke “toolbox” in het repertoire van een chemicus: dynamisch combinatoriële chemie. Tenslotte bespreken we hoe meerdere subsystemen in een enkel synthetisch systeem kunnen worden opgenomen.

Hoofdstuk 2 gaat verder met het beschrijven van een eerder ontdekte peptide-gebaseerde zelf-replicator die centraal staat in dit proefschrift. We kijken naar het belang van de gekozen aminozuren en hun effecten op het gedrag van de replicator. Het wordt duidelijk dat een computermodel nodig is om de enorme “structure space” van deze systemen uitgebreid te kunnen onderzoeken. We identificeren en kwantificeren de parameters die betrokken zijn bij het zelf-replicatie mechanisme, wat een strikte controle vereist over de experimentele variabelen. Dit leidt tot de ontwikkeling van een 3D-geprint roerapparaat dat kan worden gebruikt in conventionele ultra-performance vloeistofchromatografie systemen. We sluiten af met de voorlopige resultaten van enkele experimenten die zich richten op de lastig te benaderen processen van nucleatie en fragmentatie.

In **Hoofdstuk 3** verkrijgen we nieuwe inzichten in het mechanisme van het zelf-assemblage gedreven zelf-replicatieproces door het gebruik van “high-speed atomic force microscopy”. We komen tot de ontdekking dat het materiaal dat fungeert als voorloper van de moleculaire zelf-replicator aggregaten kan vormen die zich kunnen hechten aan de zijanten van de vezelachtige zelf-replicator. Vervolgens wordt dit materiaal verbruikt terwijl de zelf-replicator aan zijn uiteinden groeit. De state-of-the-art microscopie techniek onthult een geheel nieuw mechanisme van supramoleculaire polymerisatie dat een driedimensionaal zoekprobleem (hoe vindt een vezel-uiteinde van een replicator materiaal in oplossing dat het kan gebruiken om te repliceren?) in een eendimensionaal zoekprobleem transformeert (materiaal kan eenvoudigweg langs de vezelachtige structuur naar het groeiende uiteinde diffunderen).

Katalyse is een van de meest essentiële functionaliteiten in levende systemen. **Hoofdstuk 4** laat zien hoe katalytische activiteit en promiscuïteit in een zelf-replicator wordt geïntroduceerd door “co-option”: bepaalde eigenschappen die zijn verkregen door de geselecteerde aminozuren in de zelf-replicator die de replicatie bevorderen, leiden onbedoeld tot een supramoleculaire ordening die micro-omgevingen genereerde die bijzonder geschikt zijn voor katalyse. Belangrijk is dat de katalytische activiteit die de zelf-replicator vertoont, niet optreedt met enkel de bouwstenen waaruit de replicator is opgebouwd. We hebben twee chemische reacties ontdekt die de zelf-replicator kan katalyseren: de retro-aldolreactie van methodol en de splitsing van fluorenylmethoxycarbonyl groepen. We laten zien dat dit laatste resulteert in een positieve feedback op de snelheid van de replicatie door de generatie van dibenzofulvene, wat de snelheid van thiol oxidatie kan versnellen. Dit is een eerste stap naar een volwaardig metabolisme in een synthetisch systeem.

Hoofdstuk 5 bespreekt het tot stand brengen van mutaties en diversificatie in zelf-replicerende systemen. Eén benadering om structurele mutaties te verwezenlijken is door het gebruik van gemengde bouwsteen systemen die – potentieel – zelf-sorterend gedrag kunnen vertonen. We onderzoeken een analytisch instrument dat ons in staat stelt grote aantallen van zulke gemengde systemen te screenen door een belangrijke stabiliserende factor van onze zelf-replicerende systemen te monitoren: β -sheet vorming. Door gebruik te maken van

een microplaat fotometer en een fluorescente probe kunnen we honderden combinaties per dag screenen. Daarnaast kan dit instrument ook gebruikt worden om te screenen op katalytische reacties van zelf-replicatoren door gebruik te maken van fluorescente substraten of producten.

Tenslotte worden in **Hoofdstuk 6** de belangrijkste resultaten van de voorgaande hoofdstukken met elkaar in verband gebracht en wordt een perspectief op toekomstige onderzoeksrichtingen gegeven. Het belangrijkste daarvan is dat toekomstig werk zich zou moeten richten op het gebruik van FMOC derivaten op surfactant (of coacervaat) vormende moleculen. Dit zou replicatie, een proto-metabolisme en compartimentering in één enkel systeem kunnen verenigen. Natuurlijk zou het proto-metabolisme moeten worden uitgebreid zodat zelf-replicator gemedieerde katalyse kan resulteren in de bouwstenen van de zelf-replicator. Reacties die bindingen vormen in plaats van ze afbreken zouden een grote stap in de goede richting zijn.

Met ieder experiment komen we een stap dichterbij systemen die geleidelijk aan meer levensecht worden en ik ben ervan overtuigd dat een systeem dat alle functionele subsystemen bevat niet lang buiten ons bereik zal blijven.

Acknowledgements

I have mentioned before (see *Preface*) that this field of research interests me on an almost primordial level. The ability for me to directly work on something that has bothered me for the last twenty years has helped me put my questions somewhat at ease. It would not have been possible for me to do that research without the tremendous support (scientific or otherwise) from a great many people. I have looked forward to writing this final part of my dissertation since the day I began writing it, so forgive me for being long-winded. But since I fear that most people will only read the parts pertaining to them, I will make your names **bold**.

First and foremost I would like to thank my supervisor, professor **Sijbren Otto**. Of course, many of the kind words typically written at the end of a dissertation apply to you: the welcoming open-door policy that I have (very) frequently used, the encouragement, the patience, the mediation, the trust that you put in your students and the ability to give advice on any topic – scientific or otherwise. But above all else, I truly admire the way in which you have let me develop myself in various orthogonal directions. The closest I ever came to hearing a “no” from you was when you thought it would be unhealthy for me to put more things on my plate. Likewise, I greatly appreciate the freedom in the development of new tangible skills. For instance, even though many courses I took were of direct influence to my research – such as electron microscopy – others were perhaps not (Indesign, Illustrator, creative writing). And when I asked you if I could explore options in industry, you removed barriers instead of throwing them up. In my experience, you have not just supervised my research projects, you have supervised a large part of the development of the foundation on which I can build my future and for that I will always be grateful to you.

Secondly, allow me to express my gratitude to my second promotor **Gerard Roelfes**: even though we did not have many scientific interactions during the course of my PhD, I started to get to know you a bit better during the pandemic. In the evenings, when the halls were mostly empty, you would lighten the mood by playing Rock music (like Queens of the Stone Age) so loud I could hear it in my office.

When I started my PhD I could not have imaged a more distinguished reading committee: Professors **Beatriu Escuder**, **David Lynn** and

Adri Minnaard, thank you for taking the time out of your very busy schedules to read my dissertation and supply it with your valuable feedback.

Clemens, without your watchful eye during my talk at the NCCC conference, the whole Nature Catalysis paper might not have ever happened. For your supervision during that project I remain indebted. I wonder how your ability to provide.... constructive criticism has evolved after entering the tenure track at this university ;-) All jokes aside – I really appreciate the work you have put in this project and wish you all the best in your academic career.

To my (slightly unconventional) paranimfs **Hugo** and **Tim**: thank you for taking on this gruesome responsibility. I hope you will/have enjoyed it nonetheless! I know I will ;-) It was fun to notice you two making secret appointments and I appreciate the dedication you guys have shown from the start.

Tim, I have known you for most of my life – and all of yours! You have been there for me through thick and thin. We have started (various) companies together (remember Jims & Tims?), we have been colleagues, housemates, schemers and perpetrators of tremendous mischief (across from my parents' house). I always enjoy our discussions about (our) future prospects and above all else, I am happy to count you as a forever-friend.

Hugo, our relationship shares many similarities: we have also been colleagues, housemates and schemers (definitely). We also share another feature: permanent existential dread. Thank you for being a kindred spirit. It always helps to be able to talk to a like-minded person about these types of things. Furthermore, I really want to thank you for proofreading most of my thesis. That asked a lot from you and has benefited the readability of this dissertation a lot. I am really glad that you decided to move back to the cold north – for both our sakes!

Tom, thank you for your graphical support *throughout my entire PhD*. I remember doing revisions for the publication of Chapter three and you were talking me through issues with 3D renditions of fibres I made in Illustrator – at night. And yes, I know, it would have been easier and better to use a different program for these types of things.. Besides that, you have helped me tremendously with the lay-out of this thesis and I am justly proud that such a talented graphic designer was willing to design its cover. I have very fond memories of us starting our “adult” lives together with **Hessel** in 2008. We learned a lot of things through trial and error – like how it's a bad idea to paint your

house light blue because that would give it a “Caribbean style look”. It doesn’t, at least not in the Netherlands.

Gaël, thank you for always trying to help pretty much everyone. During your postdoc in the group (and after) you have always been there to give advice on projects, proofread, suggest new ideas and have been a tremendous support during outreach activities and even more so when we decided to set up our own. Besides all of the professional help, you are also a great friend with which I discovered the world of investment banking ;-). I’m really glad to be able to call you my friend.

Andreas, Boris and Guille, my brothers from various mothers: it is not the same without you in the group. **Andreas**, I still have your picture on my desk (heritage from **Babis**, I think) and when I leave here it is definitely coming with me. You were always the life of the party (and **Julia** as well) during your stay in Groningen. **Boro**, I have never seen anyone with such an endurance (in the gym and in a bar). Always the last to go home. Our common interest in the gym and gardening have always been... interesting topics of discussion. Too bad we cannot go to the ACLO forever. **Guille**, I have always been impressed by your willingness to try new things (from apps to partying with only Dutch people). I cannot begin to fathom the amount of different board games we have played together (I hope we will be able to play *klaverjassen* soon). I will never forget our elongation dilemma and I am happy we passed that test together.

Sourav, thank you for the days – or rather weeks or months – spent at the HSAFM to monitor the self-replication process of our fibres and the resulting discussions. These discoveries will be (are) the foundation of new research directions. I have sincerely enjoyed working with you. **Chris**, I’m very happy that you have inherited this project. At the time of writing this, there have already been interesting new results and I’m looking forward to the new discoveries.

Marcel, first of all a huge thank you for all your mass-spectrometry efforts – from measuring samples to helping identifying strange peaks and discovering the issue with the baseline in the diode array. You are full of good ideas and always ready to help with running projects (or starting new ones). We have many overlapping hobbies and I am happy to often participate in those together. Too bad the pandemic makes it difficult right now (by the time this is printed, probably much less so!), but I’m sure we’ll get back to it in no time. Besides that I want to thank you, **Sebas, Peter, Guille** and **Anne-Marie** for your many hours indulging my creative outlet during our homebrew cam-

paign. The same goes out to you, **Sebas** and **Hugo** for the many days (and nights) of strategic battles on my kitchen table (that is too small – I know, I know).

Babis, it has always been a great pleasure working on projects with you (and **Bin**!). I was very happy to share my fear of flying with someone else, although that someone else made the experience way worse on our trip to the United States when he was loudly proclaiming the elevation in small increments from 30.000 feet until landing. To you, I have lost more bets than I can remember (biggest lesson learned: never bet against a Greek guy, especially when that guy is called **Babis**). Thank you for all the discussions, advice and fun (on conferences or elsewhere) we have had over the past four years.

Ivana, thank you for all the dark humour, the fun we've had in the United states and your impressive organisation skills. Together with **Guille** you always organised fun activities for the group which always resulted in increased morale and new friendships. Thanks for pretending to be startled every time I said it was your turn for group meeting, and apologies for the times where you did not pretend ;-)

Ankush, I think it is safe to say that we are USVA course buddies. I was impressed by your story-writing skills. This is probably why **Guille's** movie will be such a great success (if the party ever comes). Thank you for not tiring of my many questions about India, its politics and social structure.

Armin, thank you for teaching me a lot about Iranian culture and food. You have (successfully!) taken over most of the tremendous task of performing electron microscopy for the group – for which I am very grateful indeed. But most of all, you are one of the most helpful and friendly people I have ever met and I am happy to share an office with you!

Paul, organic synthesis wizard, music enthusiast and probably one of the best climbers I know (not sure how you relate to **Jasper** and **Kayleigh** ;-)) It is always fun to work on projects with you. You organised the collaboration with the group of **Evan Spruijt** well and I think the weeks we're going to have to spend doing kinetics will be less tedious with you around!

Peter, thank you for developing the UPLC data parser, teaching me some Python and having very short group meeting presentations! Besides that you're always fun to have around – be it during board games, video games or discussing sci-fi.

Henning, although you've been here for only a short time, you've

already made a great impression! A German that knows how to cook (food or molecules) and with an interest in outreach! We've just started a few projects together and although everything is always broken and delayed, you always seem to approach things from the positive side. Together with **Kayleigh** you have really kept the group spirits up during social isolation.

Marc, thank you for being one of the best lab supervisors ever: you have great organisation skills resulting in a lab that has never looked better than it does now (or at least compared to the last six or so years). You take many (and fruitful!) initiatives but... need much more practice with Worms ;-)

Meagan, you took on the gruelling task of introducing catalytic activity in foldamers. Not an easy challenge indeed. If anyone can do it I'm sure it's you. So did the people who award Marie Curie fellowships! As I'm writing this we haven't had a bake-off yet, so looking forward to your best carrot cake!

Alex, even though we've mostly been online-buddies, I think we're buddies nonetheless. I always find your ideas about publications and the general field of origins research fascinating and have enjoyed reading the many megabytes of WhatsApp messages (and the Sjors stickers).

Juntian, you've really taught me a lot about Chinese culture and food! Thanks for showing me the only place in Groningen where they sell actual Chinese food. During the lockdown we've started playing some video games – hope we can keep that up!

Kirill, you're a friendly new addition to the lab! You immediately fit in very well, taught me how to say “eet smakelijk” in Russian (priyatnogo appetita) and often initiate social (online) gatherings. I think it's really great that you took the trouble to go and learn Dutch – really not an easy task while doing research and studying from home due to the pandemic. Thanks for helping keeping everyone sane!

Finally, a big thank you to all current and former group members: **Pim** (who I share a birthday with), **Bin**, **Kai**, **Joydev**, **Omer**, **Yannick**, **Ivica**, **Yulong**, **Falk**, **Yigit**, **Meniz**, **Piotr**, **Shuo**, **Dávid**, **Elio**, **Giulia**, **Jan**, **Asish**, **Andrea**, **Morteza**, **Mathieu**, **Hugo** and **Saleh**: some of you have fabricated the foundation my work is built upon, while others are still adding their parts to the entire structure that someday – without a doubt – will be considered a synthetic living system.

During my PhD I have supervised a reasonable number of students doing various types of projects. Above all else, I have really enjoyed

working with you all, hope I have taught you the things you needed to learn and know that I am very grateful for the things you have taught me.

Kayleigh, you were an amazing master student. After teaching you a chaotic practical course on supramolecular chemistry you (and Jasper for that matter) apparently thought: “yes, the guy that is running around like a crazy person should supervise my master project”. Well, I’m glad you did ;-). I enjoyed working with you and was happy (or lucky) you could endure the shenanigans from **Jasper** and myself. You developed a new substrate for catalysis: something that was desperately needed. The DBF resulting from the FMOC deprotection introduces severe solubility issues, and you managed to solve that issue completely. The calibration curve was an issue though.. the substrate was so hygroscopic it was almost impossible to weigh it in. After finishing your master project you did a short project in industry (“Paper Shaper” at BEBR consultancy) where I also jumped in to supervise you (for a bit). I think you did a great job in both cases and I’m very happy you decided to stick around at the Otto-group for a PhD!

Jasper, I don’t think I’ve laughed as much at work as I have during pretty much your entire master project. You very strangely collected and logged all of the coffee you drank during that time (or at least >99% of it) and stacked all of them into such an incredible masterpiece it still graces the corner you were sitting at. You looked into the effects of lysine modification in the XGLKFK sequence. We hoped to gain more insight in the mechanism with which the various catalytic reactions proceeded. The results were very promising, but when I repeated the measurements.... Well, that’s a story for another time. There were only two real things that were troublesome: the improbable deadlift achievement and the support of the monarchy ;-). It was a very sad day when you decided to move to Leiden.. But I’m sure you’re going to do great.

Gabriel, the French karate-master. It was fun having you around, doing UPLC together and discuss martial arts. You did a lot of catalysis and tried to find different substrates – something the Otto-lab members are still trying to continue today.

Edwin, you did an Honours College project with me in which we tried to synthesize a substrate for ester hydrolysis that would result in a decrease of negative charges in the individual hydrolysis products to stop product inhibition. The synthesis worked, but the monomers did the catalysis much better than the fibres.. Something that happens

more often than one would think...

Felien, you performed a lot of kinetic experiments during your bachelor project in which we tried to see which types of fibres (depending on the building block they were made up out of) grew the fastest. We also looked into the catalytic activity of mixed building block systems but, as always, XGLKFK was the fastest.. in all cases.

Mathijs, you looked into the effect of the peptide chain on the catalytic activity of the replicator – something Jasper continued in his master project. At the time, we did not look closely at lysine modifications but were more interested in modifying the phenylalanine. Additionally, we discussed a lot about investment banking.. something that I don't think you've started doing just yet ;-)

Jelmer, you also performed an Honours College project in our group in which you tried to optimize the analysis of catalytic reactions (by fibres) using fluorescence spectroscopy. Together with **Jasper**, you laid the foundation for Chapter five and introduced the use of this technique in our group (that is still being extensively used today! But now, we have our own multi-pipettes.. saves a lot of time).

Kas, Elianna, Vibuthi and **Martine**, thank you for signing up en masse for the outreach project I wrote. It started out as "Grunniger Science" but has evolved into Noorderlab (by the time you are reading this, the first podcasts should have already been released). You've helped tremendously with the development of the project and made explanation videos, various graphic designs, a cool puzzle and even recorded a try-out podcast.

Fleur and **Kbreab**, thank you for making a digital poster to explain the research being done at the Otto lab. The plan was to make a whole museum exposition, but the pandemic foiled our plans. Regardless, it was fun working with you!

Similarly, there were many people that have supported my research, an outreach project or my well-being in one way or another:

Annette, thank you for helping me with a lot of the day-to-day business. If it weren't for you, I wouldn't even know how to send a letter from the University. Sometimes it was hard to find my way through the bureaucracy, but somehow you always managed to know exactly what to do.

Renske, Amanda, Egbert, Ingeborg and **Maaïke**, thank you for your incredible support during the development of Noorderlab. **Renske, Ingeborg** and **Maaïke**: you are the ones that will keep the project alive after we – as Stratingh institute – pass on the baton to the next FSE

institute. **Amanda** and **Egbert**, thank you for your efforts in designing, recording and editing the podcasts and your valuable feedback on other aspects of the project.

Lou de Leij, Marianne Rots and the rest of **the board and members of the KNG**, thank you for giving us the opportunity to develop the Noorderlab project. Without your financial support, it would not have been possible to set this up. I hope that future releases will continue and be successful indefinitely!

Renze and **Monique**, you have saved me from missing a deadline more often than I can count. The most recent case was when reviewers requested additional experiments in a very short timeframe. All experiments had to be performed on the same machine (as calibration curves were made on that particular one) and everything broke down. The needle, the vacuum pump (and other things after). **Monique** was able to obtain a different type of needle very quickly and **Renze** switched out a vacuum pump from a different machine. We made the deadline with only around 12 hours to spare (the experiment took three days). Thank you for your never-ending support, explanation and most of all: hard work.

Johan, thank you for your help with finding creative solutions for bureaucratic problems.

Everyone at BEBR and **Marije** in particular, thank you for giving me the opportunity to broaden my horizon as a technical consultant. The projects were very interesting – I did not expect to learn as much as I did about 5G, potatoes, satellites and writing proposals.

Marco, ouwe vingerknipper (dat doe je vaak en is beter dan wat ik eerst had geschreven). De dag werd altijd een stukje leuker als we wat lol in de gang hadden. Het ga je goed!

BHV-ers, ik (en velen met mij) ben jullie dankbaar voor de tijd en moeite die jullie elke dag steken in de veiligheid van jullie collega's. Iedereen zou eens een weekje moeten meelopen hier en kijken wat er allemaal gebeurt en hoe het wordt opgelost. Ik denk dat dat een hoop ongelukken zou voorkomen. Voor mij was – en is – het een erg leerzame en dankbare functie.

Naomi, Tim, Sebas, Martine, Juanita en Laura (en alle (wisselende) aanhang), ik ben heel erg blij dat we elkaar ontmoet hebben. Er is veel te veel waar ik jullie dankbaar voor ben om hier over te schrijven. Dank voor alle nachten, ochtenden, het delen van plezier en verdriet, het redden van mijn leven (enigszins figuurlijk), het geven van ongezouten advies (ongevraagd of niet), de nieuwe ervaringen en de on-

voorwaardelijke support.

Thijs, ik ben blij dat ik je tot een echte vriend kan rekenen. Door dik en dun! Je vindt er geen doekjes om en vindt het niet erg om het ergens over oneens te zijn. Bedankt voor alle gedeelde ervaringen, de discussies en de goede (en slechte) tijden die we hebben meegeemaakt en nog zullen meemaken.

Jasper, de oudere-broer-die-eigenlijk-de-buurjongen-is. Wat een lol hebben wij gehad de laatste twee decennia. Op jou kan ik blindelings vertrouwen. Bedankt voor al het advies, de hulp en het introduceren van krachtsport aan een 15 jarig, dik kereltje. Daarvoor moet ik **Klaas** ook erkennen.

Hessel, dank voor alle gesprekken en inzichten. Het begon allemaal met een potje poolen in een voor mij erg moeilijke tijd en liep uit tot jaren samenwonen met **Tom**. Het was fantastisch om samen met jullie groot te worden.

Pieter, voor mij maakte onze vriendschap de middelbare school dragelijk. Ik kijk nog altijd met veel plezier terug naar de survival activiteiten van toen, en naar de wildkampeer activiteiten van de toekomst.

Mam, mam en pap. Zonder jullie eeuwige support was er van mij niet veel terecht gekomen. Toen ik de eerste existentiële crisis onderging toen ik elf was wist men eigenlijk al dat dat niet weg te rationaliseren is. Ik heb er mee om moeten leren gaan – iets wat in veel gevallen niet mogelijk is. Jullie vulden elkaar tijdens die periode goed aan, ook al voelde dat (voor jullie) op die momenten vaak niet zo. Voor alles wat jullie voor me hebben gedaan, wat jullie me hebben geleerd en wat jullie nog steeds aan me leren zal ik jullie voor altijd dankbaar blijven.

Lisanne, over jouw avonturen moet er ook nog een keer een boek verschijnen. Misschien kan **Stef** je dan ook helpen met de lay-out ;-). Lies, ik ben zo blij dat jij m'n zus bent. Je bent betrokken, staat altijd klaar om te helpen bij de echte problemen in het leven, bent een geweldige juf en een geweldige moeder. Je bent er altijd voor me geweest en ik zal er altijd voor jou (en **Dim**) zijn.

Naomi, mijn zon. Je maakt me een beter persoon en daagt me altijd uit om uit mijn comfort zone te komen. Bovendien ben ik soms veel te stellig en te ongenueanceerd en dan help je me de andere kant te zien. Bedankt voor alle nieuwe ontdekkingen (in het verleden en in de toekomst), je begrip, je humor en je liefde.

About the Author

Jim (Johannes) Ottelé was born in 1989 in Groningen, The Netherlands. He started studying physics at the University of Groningen in 2008 and decided to switch to chemistry after doing experimental chemistry courses. His interest lay in the interface between chemistry and physics, which motivated him to choose tunneling junctions of self-assembled monolayers (SAMs) with donor and acceptor moieties in the group of Ryan Chiechi as graduation topic for his Bachelor's degree. Subsequently, he enrolled in the Advanced Materials track – a Master's degree programme with both physics and chemistry courses. During his studies he worked as a student assistant at the Bètas-teunpunt, he developed experiments for use in practical courses of bachelor students, taught web classes about photovoltaics and was a bar manager at a night-club establishment. He obtained his Master's degree in the group of Sijbren Otto, where he studied the incorporation of catalytic functionalities in self-replicating systems, thereafter staying to pursue a PhD. During his PhD, he set up the outreach programme Noorderlab, participated at numerous scientific outreach activities, acted as a technical consultant and taught at a continued education program for high school chemistry teachers.

List of publications

1. Liu, B., Pappas, C. G., Ottel, J., Schaeffer, G., Jurissek, C., Pieters, P. F., Altay, M., Mari, I., Stuart, M. C. A., Otto, S. Spontaneous emergence of self-replicating molecules containing nucleobases and amino acids, *J. Am. Chem. Soc.*, **2020**, 142, 4181-4192.
2. Ottel, J., Hussain, A.S., Mayer, C., Otto, S. Chance Emergence of catalytic activity and promiscuity in a self-replicator, *Nat. Catal.*, **2020**, 3, 547-553.
3. Maity, S., Ottel, J., Monreal Santiago, G., Frederix, P. W. J. M., Kroon, P., Markovitch, O., Stuart, M. C. A., Marrink, S. J., Otto, S., Roos, W. H. Caught in the act: mechanistic insight into supramolecular polymerization-driven self-replication from real-time visualization, *J. Am. Chem. Soc.*, **2020**, 141, 1685-1689.
4. Markovitch, O., Ottel, J., Veldman, O., Otto, S. Automated device for continuous stirring while sampling in liquid chromatography systems, *Commun. Chem.*, **2020**, 180.
5. Yang, S., Schaeffer, G., Mattia, E., Markovitch, O., Liu, K., Hussain, A. S., Ottel, J., Sood, A., Otto, S. Chemical fueling enables molecular complexification of self-replicators, *Angew. Chem. Int. Ed.*, **2020**, 60, 2-8.
6. Pappas, C. G., Liu, B., Mari, I., Ottel, J., Kiani, A., van der Klok, M. L., Onck, P. R., Otto, S. Two sides of the same coin: emergence of foldamers and self-replicators from dynamic combinatorial libraries, *J. Am. Chem. Soc.*, **2021**, 143, 7388-7393.
7. Liu, B., Beatty, M. A., Pappas, C. G., Liu, K., Ottel, J., Otto, S. Self-sorting in dynamic combinatorial libraries leads to the co-existence of foldamers and self-replicators. *Angew. Chem. Int. Ed.*, **2021**, 60, 13569-13573
8. Komromy, D., Tiemersma-Wegman, T., Kemmink, J., Portale, G., Adamski, P. R., Blokhuis, A., Aalbers, F. S., Mari, I., Monreal Santiago, G., Ottel, J., Sood, A., Saggiomo, V., Liu, B., van der Meulen, P., Otto, S. Stoichiometry alone can steer supramolecular systems on complex free energy surfaces with high selectivity. *Chem.* **2021**, 7, 1933-1951.