UNIVERSITY OF GRONINGEN

Catalytic sites on self-replicating fibres

Master Thesis Chemistry

J.E. Bos

26-10-2019

Supervised by:
J. Ottelé
Prof. dr. Sijbren Otto

Abstract

Creating life *de-novo* is one of the largest challenges that science is facing at the moment. It requires a system in which the characteristics of life are incorporated. Three such characteristics are essential for any living system: replication, metabolism and compartmentalisation. At the basis of metabolism lies the property that a system should be able to facilitate chemical transformations. Recently the group of Sijbren Otto discovered a peptide based self-replicator capable of catalysing both the retro-aldol reaction and Fmoc deprotection reaction, an emergent property of the system.

In this thesis we designed several replicator building blocks aimed at perturbing the active sites of this self-replicating system. Using several different techniques the properties of these self-replicators were investigated. After briefly investigating if the set of reactions catalysed could be expanded, the activity of the new replicators in the existing reactions was tested. The work further focusses on linking the discovered properties of the replicators to their activity as a catalyst. Mixing building blocks is utilised in order to further perturb the active sites in the system and the effects of this are studied.

Contents

Introduction

The origins of life

Perhaps one of the most amazing feats in chemistry is that it allows life to arise from inanimate matter. Yet even though our very existence is proof that this is possible, the process through which such abiogenesis happens remains largely unknown. The search for the origin of life is one of the largest unsolved problems of science and there are many groups with different backgrounds working on tackling this problem today.[1]

Even though the living things that we encounter in our everyday life can be intuitively classified as being alive, a common definition of it has so far remained elusive. Some even argue that formulating such a definition would either be impossible or useless.[2] Nevertheless we can still recognise a shared set of commonalities that all living beings possess. Even in the simplest form of life we find metabolism, compartmentalisation and reproduction.[3] These three very important characteristics are often covered under a variety of different names in textbooks and research literature (see figure 1).[4] The meaning is however generally the same: a living being needs to be separate from its environment and through the utilisation of chemical energy be able to make copies of itself.[3]

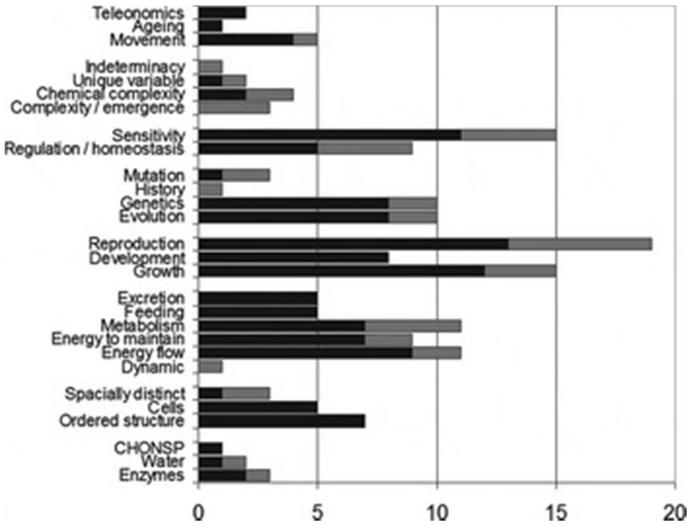


Figure 1: Characteristics of life from definitions found in textbooks (black bars) as well as monographs and research literature (grey bars). Taken from Bains.[4]

How life as we know it acquired these properties is the focus of origins of life research. This field is dominated primarily by the two opposing views of a replication first versus a metabolism first scenario.[5,6] In the replication first scenario information carrying molecules capable of self-replication are at this origin,[6] whereas in the metabolism first scenario there are simple molecular reaction networks that increase in complexity giving rise to living beings.[7]

In 1989 the Nobel prize in chemistry was awarded to Altman and Cech for the discovery of catalytic properties of RNA.[8] This observation that RNA can both encode information as well as catalyse reactions, acting as a RNA enzyme or ribozyme, gave rise to idea of the RNA world.[9,10] This is a moment in life's early history where RNA is the main genetic system as well as being responsible for metabolism with RNA based organisms.[11]

This has great implications for the replicator first scenario. If different RNA oligomers can be formed under prebiotic conditions, some of these oligonucleotides might be ribozymes that can catalyse their own formation. These self-replicators could then evolve through a process akin to natural selection, acquiring traits that increase chances of survival in order to not die out and leading to the RNA world.[12,13] This hypothesis assumes that there are prebiotic conditions in which a mixture of activated nucleotides are formed which are then capable of assembling into a library of different oligonucleotides.[11]

The field of pre-biotic chemistry focusses on finding such reactions and conditions. The most famous experiment that yielded a mixture of biologically relevant molecules was performed by Miller and Urey. They utilised a mixture of gasses likely to be found on the early earth, and an electrical discharge that simulated lightning.[14] In this mixture several biologically relevant molecules were created, such as simple amino acids.[15] Since then the field has had many successes in creating other relevant biomolecules trough prebiotic pathways that might eventually lead to the aforementioned oligonucleotide mixture.[16–18]

However the reactions that lead to these molecules usually require rather specific conditions and high concentrations of the right reactants, which decreases the possibility of finding such conditions outside of the lab. For example the synthesis of activated ribonucleotides through a pyrimidine pathway reported by Powner *et al.* requires cycles of heating and cooling, and hydration and dehydration.[19] Moreover, some consider RNA too complex a molecule and its catalytic repertoire too limited to have arisen prebiotically.[20]

On the basis of this, some scientist reject this replicator first hypothesis.[21] The hypothesis that is put forward instead is that of metabolism first scenario. Here it is recognised that life is a complex set of many reactions forming networks. Starting from a simple self-sustaining network of reactions, and giving that they can evolve, such a system could generate the compounds necessary to create a replicating informational polymer.[7,22]

Also this hypothesis is not without criticism. It is argued that metabolic cycles with nonenzymatic catalyst would lack in specificity and heredity to evolve into the cycles envisioned by their proponents.[5,23] Moreover there remains little experimental work that supports this hypothesis.[24]

Finding direct archaeological evidence to support one theory over the other is very difficult, if not impossible. Another approach that might give insight into the workings of abiogenesis is creating synthetic life, or life *de novo*. We can recognise that all life that we know is chemical in origin, and it should therefore be possible for a chemist to create a system that can transform from being not alive to one that is living.[25] Like the Miller-Urey experiment such an example could give experimental backing to theories regarding our own origins.

Life *de-novo*

Creating a living chemical system is a very ambitious goal, and has so far not been successful.[26] As we have seen before three characteristic appear to be used in most definitions in various forms, and we consider them essential in a life-like system. These characteristics are: metabolism, compartmentalisation and replication.[4,27] Of these three, metabolism has so far not been incorporated successfully into replicating systems and our current work focusses on developing this.

Metabolism can be defined as the set of all chemical reactions that sustain a living being.[27] To make this happen it needs to be capable of building all the necessary components of the organism, by creating large molecules from smaller parts. Moreover, it needs to supply the energy to create and utilise these components as well as breaking down the components that are no longer usefull.[27] An example of a metabolic process would be the creation of an RNA polymer due to the degradation of a high energy molecule in a cell.[11] At the very basis of this lies the observation that a living system should be able to execute chemical transformations, and to create favourable kinetics it has to employs catalysts. In nature this role is usually performed by enzymes, although as we have seen ribozymes can also do this.[28]

By shaping an active site within an enzyme through natural selection, nature has managed to create highly specific and active catalysts.[29] The catalytic properties of such enzymes are highly advanced and can hardly be matched by designed synthetic catalysts.[30] Because enzymes have a sequence that allows them to fold, a larger structure is created in which the functional groups of the active-site amino acids are perturbed.[31] This perturbation is what causes the catalytic role of many enzymes.[32,33] For example, bringing two charged amino acids into close proximity can influence the pK_a of the functional groups on these residues, so that the charge is alleviated. This enables them to act as a nucleophile, and work as a catalyst. This demonstrates that large (supramolecular) structures can become catalyst, and it is this supramolecular structure that enables this catalysis.

Self-replication

A self-replicating system is any system that can catalyse its own formation, and as discussed above one of the properties a living system has.[34] A molecule that can arrange its precursors so as to transfer its own structure, will create an exact copy of itself.[27] In the group of Sijbren Otto such a self-replicator has been developed.[35] This self replicator is based on an aromatic dithiol with a peptidic tail (see figure 2). The thiols can oxidise, forming disulphide bonds with other thiols, resulting in a mixture of macrocycles (figure 2, step 1).

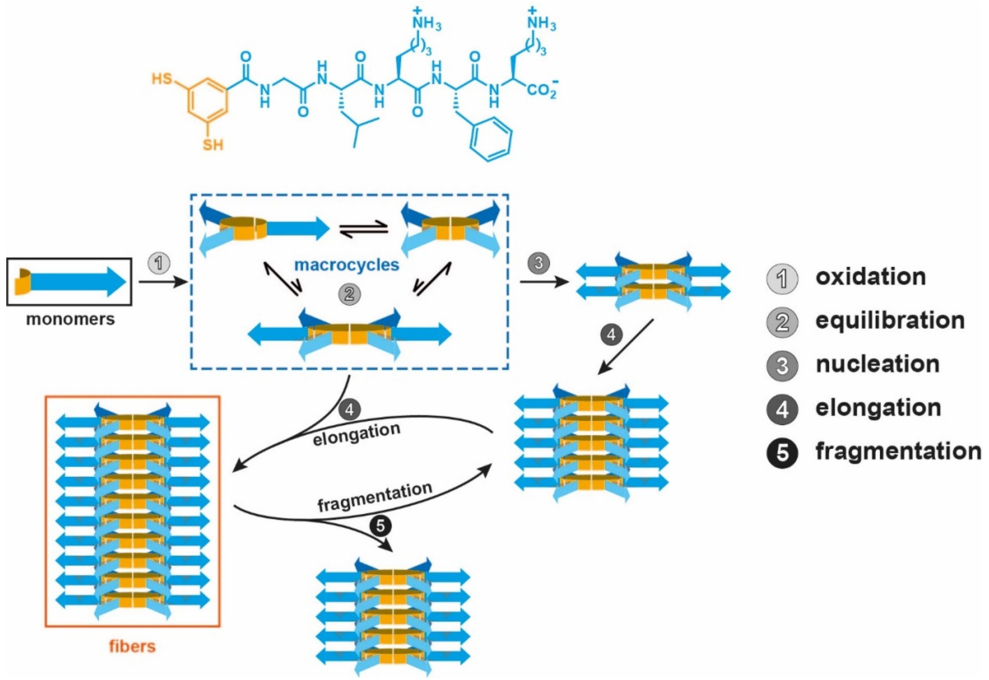
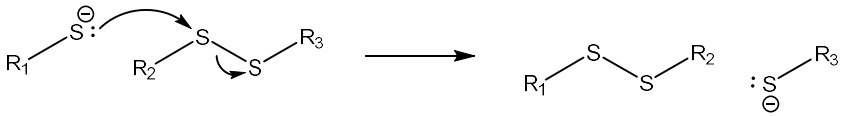


Figure 2: Mechanism of fibre formation in the XGLKFK system with step 1: oxidation of the monomers into a mixture of various macrocycles; step 2: equilibration of the macrocycles into a dynamic combinatorial library; step 3: nucleation of a fibre by stacking of the first macrocycles; step 4: elongation of the fibre by further stacking; step 5: fragmentation of the fibre by mechanical energy doubling the fibre ends. Adapted from Ottelé et al.[38]

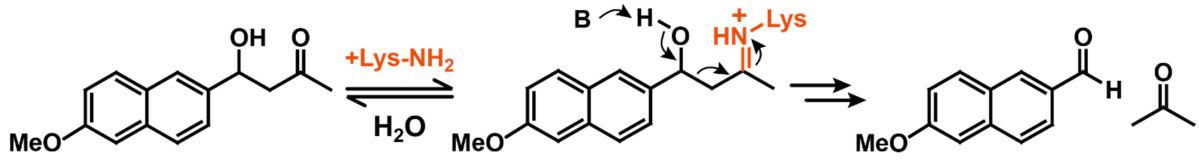
As long as there are free thiols left, the disulphide bond is dynamic and can exchange. A thiolate anion can attack a disulphide bond which will form a new disulphide bond and release a new thiolate anion (see scheme 1). This will equilibrate the system into a dynamic combinatorial library with macrocycles of various sizes.[36] In some cases a macrocycle size possesses the ability to stack, by self-assembling trough π-π interactions and β-sheet formation. The β-sheet formation is facilitated by the peptide tail which offers alternating hydrophobic and hydrophilic residues. We call the stacking of the first few macrocycles a nucleation event. After nucleation macrocycles of the same size are either templated on its end or sequestered from the macrocylce equilibrium, elongating the stack. The molecules assemble in this way into a supramolecular fibre (figure 2, step 3 and 4). By input of mechanical energy, fibres can be broken which doubles the amount of fibre ends. From these ends the fibres can again elongate, and thus pushes the system into a regime of exponential self-replication (figure 2, step 5).[37]



Scheme 1: Mechanism of disulphide exchange, showing a thiolate anion attacking a disulphide bond, displacing one sulphur atom and forming a new disulphide bond with the other.

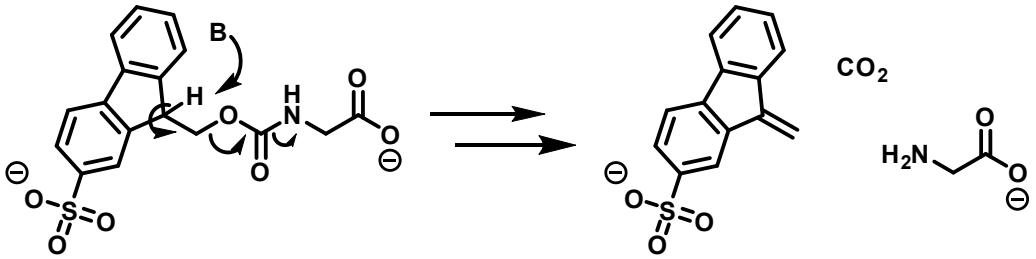
These self-replicating properties form a good starting point to create a life-like chemical system, since it already possesses one of the three important properties of life, *i.e.* replication. Developing catalytic functionality in this system forms the first step to also acquire a second property: metabolism. The way the fibres are self-assembled also offers some opportunities with respect to catalysis, as the β-sheets offer hydrophobic grooves, where several charged residues are brought into close proximity to one another.[39]

These properties were explored previously and led to the discovery that upon assembly the replicator can catalyse two types of reactions, the retro-aldol reaction and Fmoc deprotection reaction. The retro-aldol reaction of methodol is shown in scheme 2. In the reaction an iminium ion is formed from the aldol carbonyl and an amine from a lysine on the supramolecular fibre.[38] Which of the two lysines present in the structure creates the iminium ion was before this study unknown. In the supramolecular fibre the lysines are brought into close proximity of each other which lowers the pK_a of the amines to alleviate unfavourable electrostatic interactions. This decreased pK_a means that the residue is no longer protonated at a slightly basic pH and this allows it to act as a nucleophile in the reaction. Since this property is only present in the supramolecular fibre it is an emergent property of the system. The creation of the iminium ion facilitates the aldol cleavage, obtaining the aldehyde and acetone.[40] If the backwards reaction, *i.e.* the aldol formation, can also be achieved the set could be expanded to perform both anabolic and catabolic roles in the system.



Scheme 2: Proposed mechanism for the retro-aldol reaction of methodol. Showing the formation of the iminium ion intermediate which upon deprotonation breaks the aldol bond, forming acetone and 6-methoxy-2-naphthaldehyde. Adapted from Ottelé et al.[38] .

Besides this reaction also the deprotection reaction of Fmoc can be catalysed by the replicator fibres. When this was discovered, it became necessary to create a more water soluble variant of this reaction to be more useful in this system. A permanent negative charge on the Fmoc protecting group achieves this increased solubility while it is simultaneously attracted to the positively charged fibres. This negative charge was obtained by sulfonating the Fmoc group. The sulfonated Fmoc group is abbreviated as Sulfmoc. Here we have studied the Sulfmoc deprotection of glycine (see scheme 3). The reaction is catalysed by the same perturbed lysines on the replicator fibre, now acting as a base. This lysine deprotonates the fluorenyl hydrogen, which will cleave the molecule. This generates dibenzofulvene, CO_2 and the deprotected glycine.



Scheme 3: Mechanism for the Sulfmoc deprotection of glycine. Showing the deprotonation of the fluorenyl hydrogen which cleaves the molecule forming dibenzofulvene, glycine and CO_2.

The Fmoc deprotection reaction has the added benefit of also having a positive feedback on the replication of the system. Dibenzofulvene enhances the speed of sulphide oxidation, and thereby increases the speed of replicator formation. The exact mechanism of how this operates is however largely unknown and is beyond the scope of this project.

Aim and outline

If the concept of metabolism is to be explored in the context of *de-novo* life, it is paramount that there exists a thorough understanding of the catalytic properties of the system under study. We set out to further explore the active sites on the self-replicating system and we attempt to improve upon it. We designed several new building blocks with permutations from the original replicator that are relevant to the catalysis. These building blocks were fully characterised and their catalytic activities studied. Simultaneously we investigated if there are other reactions that can be catalysed by these self-replicating structures.

In the first chapter we take a look at these potentially self-replicating structures and we explore the characteristics of the structures that are formed. First a rational is given for the choice of these specific structures, which are then investigated with a myriad of techniques and under various conditions relevant to the catalysis.

These structures are then taken in the second chapter where they are studied for their catalytic activity for both the retro-aldol and the Sulfmoc deprotection reaction. Some other potential reactions are briefly discussed. The activities are linked to the properties found in the previous chapter.

In the final chapter several of the building blocks are combined to create mixed libraries, and the subsequent effect on catalysis is studied. The question is if the sum of both replicators is greater than its parts, and if further insight into the active site can be gained from these structures.

Characterisation of Replicator Sequences

In order to further develop catalytic functionality in replicating systems, a better understanding is needed of the way the self-replicating fibres become catalytically active. One way of going about this is by taking a reference structure and studying the change in reaction kinetics after a minor permutation in the replicator structure. Since the previous work by Ottelé *et al.* had focused on the behaviour of the XGLKFK replicator, it was chosen to take this compound as a reference.[38] As described in the introduction, in the proposed mechanism it was postulated that an amine from lysine is perturbed and this allows it to form an iminium ion, driving the reaction. Because of this, permutations were chosen that modify both the lysines in the structure or that influence the position immediately next to it. The rest of this chapter will be devoted to describing the properties of these structures.

Although our group has found a variety of self-replicating systems over the years, there exists some generality in their structure. First of all, they are all based on a very small peptide of varying lengths, with the pentapeptide being the most studied.[41,42] An aromatic dithiol core is linked through the N-terminus of the pentapeptide. This core moiety allows for the reversible covalent bonds that are central in dynamic combinatorial chemistry and is here abbreviated as X. When describing replicator building blocks they are referred to their one letter amino acid abbreviation, and counted from their N-terminal side, *e.g.* XGLKFK. In a pentapeptide there are over 3 million structures that can be generated from the canonical amino acids alone. It is therefore apparent that structures need to be rationalised and carefully chosen before synthesis.

Parameters of the active sites on a fibre

There are many parameters that could influence the active sites on the fibre (see figure 3). First there is the catalytic amine in the active site. Changing its electronic properties, *e.g.* in aromatic or secondary amines, will influence how well it performs. Secondly there is the local environment of the active site. It is this local environment that is responsible for the perturbation of the pK_a of the amines. To further

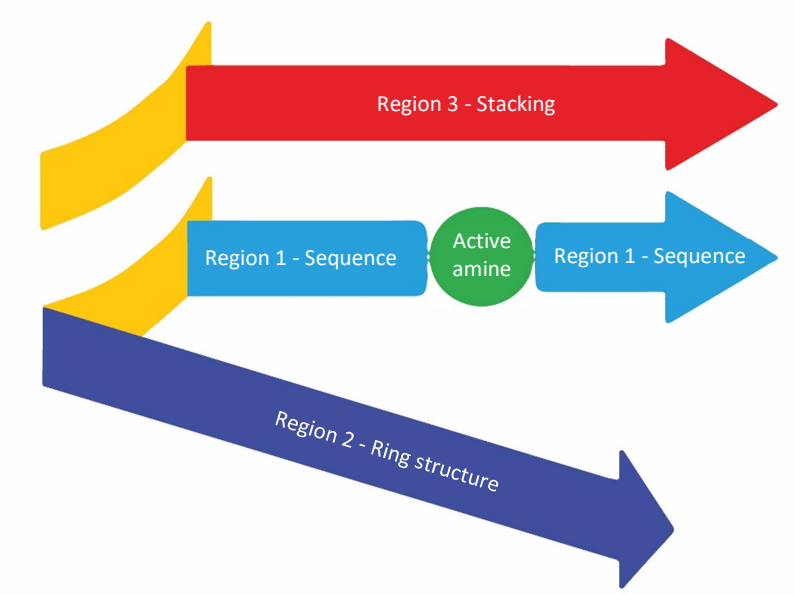


Figure 3: Schematic representation of the active site, and the corresponding local environment in a replicating fibre. Showing the various regions which make up this local environment and the parameters that influence it. The active amine is shown as a green circle; region 1 is located on the same peptide tail as the active amine; region 2 is formed by a peptide tail on the same macrocycle; and region 3 is created by the peptide tails of a stacking macrocycle.

illustrate how this can be influenced the local environment is divided into three regions. In the first region we find the neighbouring amino acids from the building block sequence. Changes in this sequence can directly influence the active site. Region two is made up of other building blocks that are covalently attached via the ring structure. Since the residues in the peptide tail of this building block can come into close proximity of the catalytic amine they can perturb the active site. Yet in order for that to happen, the ring needs to have the right size and shape. Changes in this shape can influence the active amine. Finally we have region three, which is similar to region two. This time however the residues come from the peptide tails of another ring, that is non covalently attached. Again the right shape is necessary for the residues to perturb the active site. Changes in the stacking macrocycles will affect the influence on the active site from this region.

Rationalisation of chosen structures

For this thesis eleven building block structures were chosen for study. As mentioned previously XGLKFK was taken to be a reference compound. Since the performance as catalyst was our prime interest modification mainly involved attempts to influence the active site through the parameters outlined above. Although it was proposed in the mechanism for the retro-aldol reaction that it was the amine of a lysine that was responsible for the catalysis, it was previously unknown which of the two lysines was the main contributor. Therefore the initial structures created replaced each lysine for an arginine group. We postulated that it would have a minor impact on the supramolecular structure, both groups mainly being charged at the working pH of 8.1. However, arginine has a higher pK_a, and is known to retain its protons even in hydrophobic environments, thus it cannot act as a nuclophile.[43] We initially created three structures: *i.e.* XGLKFR, XGLRFK, and XGLRFR as a control (figure 4, top row). The details of the structures formed will be described in detail further on in the text, however it is important to mention here that they allowed us to conclude that it is the inner lysine, on the third amino acid position that is mainly involved in the catalysis.

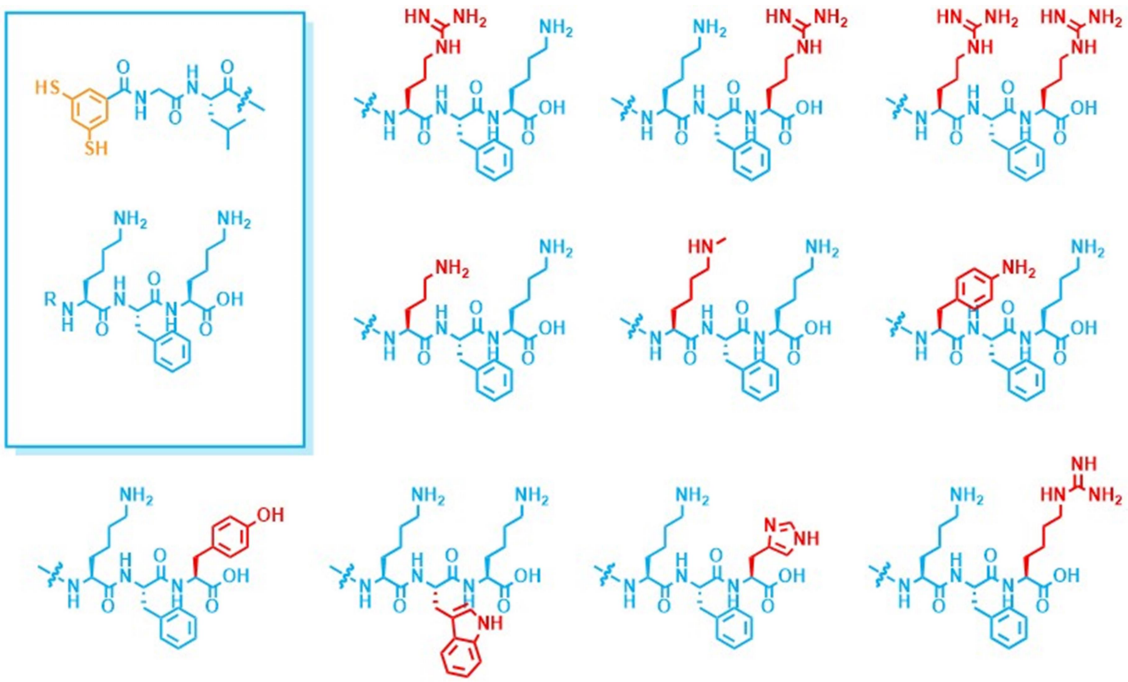


Figure 4: The eleven building block structures under study, the reference compound is shown in the square. The XGL-unit is general for all structures and only shown in the square.

Based on this result seven further structures were conceived that might be able to improve or disrupt the working as a catalyst. Focusing first on the catalytically active amine, three more structures were designed. The third amino acid position was modified using the (uncommon) amino acids ornithine, methyl-lysine, and amino-phenylalanine (see figure 4, middle row). These allow assessing the influence of the distance from the backbone for ornithine, the difference between a primary and secondary amine for methyl-lysine, and the difference between an aliphatic and aromatic amine in amino-phenylalanine.

The other four building blocks made changes to residues in the local environment (see figure 3, bottom row). Since the proposed mechanism for the retro-aldol reaction involves the movement of several protons in the creation of the iminium ion, three further structures were formulated in which hydrogen bonding amino acids could facilitate this proton movement. On the fourth amino acid position the phenylalanine was replaced with a tryptophan, which is simultaneously a hydrophobic residue and capable of hydrogen bonding. The fifth amino acid position was exchanged for both a tyrosine and a histidine. This position was of interest since in the β-sheets of the fibre the residues alternatingly lie next to each other. Therefore, this position might be in close proximity to the active amine. Both residues are also capable of hydrogen bonding.

The final structure created was devised to influence the position of a charged structure neighbouring the catalytically active lysine. Since it was postulated that the pK_a of the lysine amine was perturbed by a charged structure nearby which lowers the amine pK_a, it was chosen to replace the fifth amino acid for a homo-arginine amino acid (see figure 4, bottom right). We chose this amino acid position because it is the only charged residue besides the catalytically active amine. Homo-arginine has a guanidinium group on the epsilon carbon, and with a pK_a higher than a lysine residue, it might cause a larger perturbation in the catalytically active lysine residue

Analytical techniques

Several analytical methods were used to probe the structures that arose from the different building blocks. The composition of macrocycles in the various libraries was followed over time by UPLC in combination with UPLC/MS to confirm ring size. After the libraries were fully oxidised they were probed with Transmission Electron Microscopy, Circular Dichroism spectroscopy and a Thioflavin T assay. With the TEM images the supramolecular structures can be visualised and their dimensions and relation to each other can be explored. In contrast the CD-spectra and ThT assay will give an indication of the ordering within each structure. CD-spectra will give an indication of secondary structures present, such as β-sheet and α-helices. A β-sheet would have a maximum around 194 nm and a minimum around 214 nm (see figure 5).[44,45] A ThT assay explores the presence of β-sheet interactions. By following the increase in fluorescence that ThT exhibits upon binding β-sheets.[46]

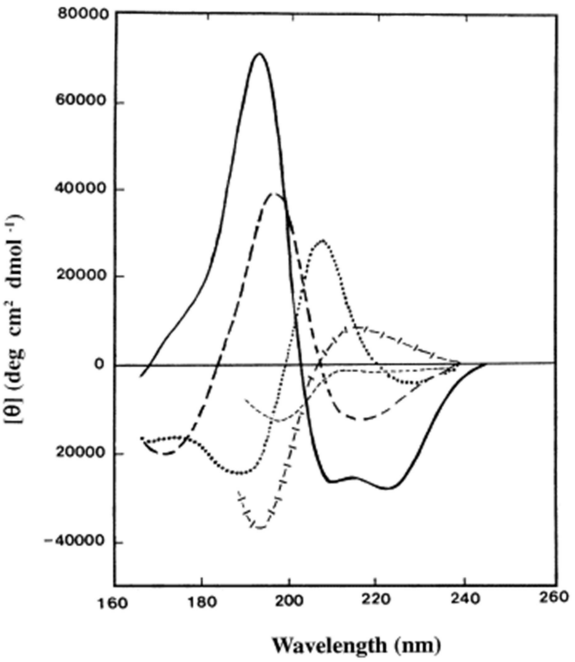


Figure 5: Secondary structures in CD spectra with an α-helices (solid line), anti-parallel ß-sheet (long dash), type-I ß-turn (dotted line), Poly (pro) II helix (cross-dash), irregular structure (short dash). Taken from Kelly et al. [45]

Arginine replicators

For reasons outlined above three structures were created to discover which lysine is catalytically active, by replacing them with an arginine amino acid. The structures created were XGLKFR, XGLRFK and XGLRFR. Their properties will be explored in this section.

XGLKFR

When substituting the C-terminus lysine for an arginine we obtain the structure XGLKFR. This is a self-replicating structure that creates supramolecular fibres consisting of hexamer macrocycles as confirmed by seeding experiments (Figure S18). The replicator macrocycle starts to emerge within 24 hours, yet a small amount of trimer appears to persist for a longer time (figure S17). On TEM images fibres can clearly be observed and they appear to be laterally associating (figure 6B). This is reflected in the width measurements based on TEM micrographs, which yielded an average length of 21 ± 8 nm (table S1). Earlier research indicated that single fibres of XGLKFK have a width of approximately 3 nm.[47] The ThT assay clearly indicates a large amount of β-sheet formed as there is a clear increase of fluorescence intensity compared to both the oxidised monomers and the monomers before assembly (figure 6D). The intensity is rather large compared to the other building blocks. A similar result is found in the CD spectra (figure 6C). When the building block is left *unstirred*, only trimers emerge. On TEM it can be seen that these also form fibres (Figure S22).

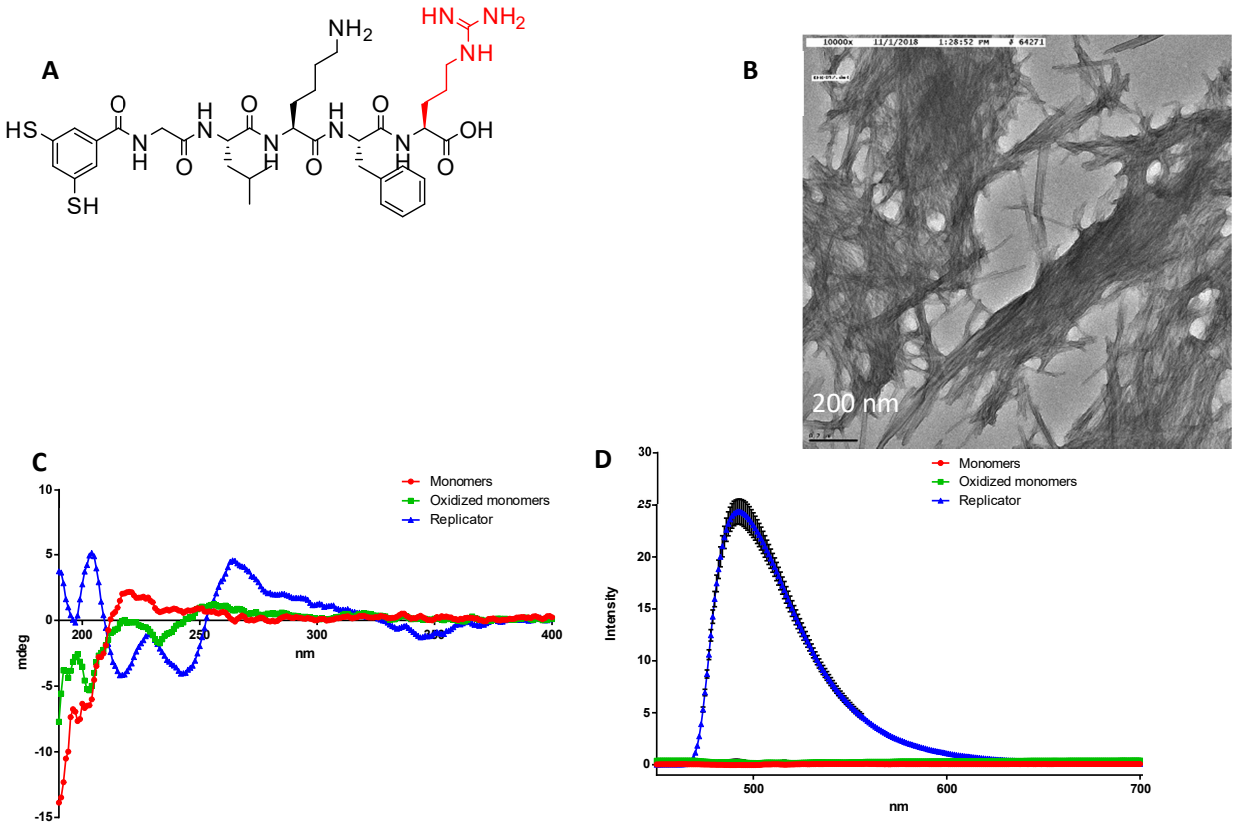


*Figure 6: **A)** Structure of XGLKFR **B)** Representative negative stain TEM image of a library of XGLKFR after approximately 2 weeks, showing fibres have formed. **C)** CD spectra and **D)** UV spectrum of a ThT assay of XGLKFR (in triplicates) showing samples with monomers as red circles, oxidised monomers as green squares and fibres as blue triangles; error bars are indicated in black. Both spectra indicate that the fibres are based on β-sheets. Samples were prepared as described in the supporting information.*

XGLRFK

Interestingly when making the other substitution with arginine the resulting molecule oxidises much slower, with 5% of monomer still present after 100 hours. After 60 hours the hexamer macrocycles emerges that are assembled into fibres (see figure S32). On TEM images we can clearly see these fibres and they appear to be similar to fibres of XGLKFR, with an average of 19 ± 6 nm (figure 7B, table S2). The intensity of the ThT assay is also much lower than for XGLKFR, yet still clearly amplified over the monomer and oxidised monomers . The β-sheets, might therefore not be as abundant as those found in XGLKFR. In the CD spectrum we find quite a strong signal in the β-sheet region. If this building block is *not* stirred, the building block produces trimers and pentamers. A fibre is still formed (figure S36).

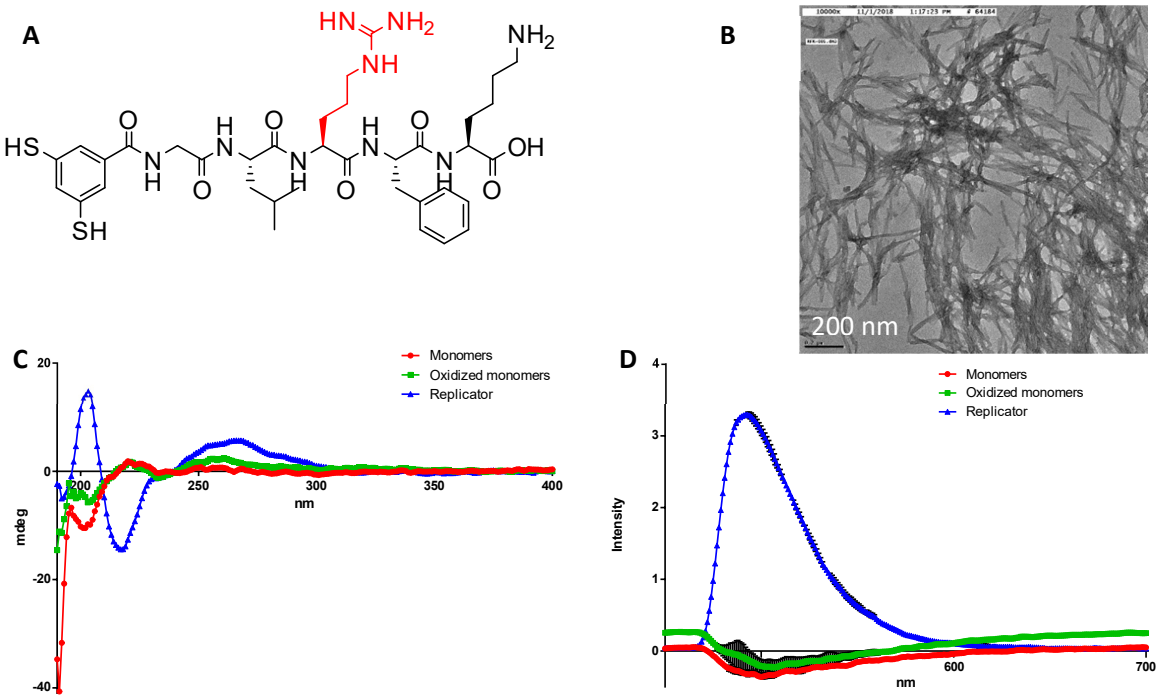


*Figure 7: **A)** Structure of XGLRFK **B)** Representative negative stain TEM image of a library of XGLRFK after approximately 2 weeks, showing fibres have formed. **C)** CD spectra and **D)** UV spectra of a ThT assay of XGLRFK (in triplicates) showing samples with monomers as red circles, oxidised monomers as green squares and fibres as blue triangles; error bars are indicated in black. Both spectra indicate that the fibres are based on β-sheets. Samples were prepared as described in the supporting information.*

XGLRFR

The building block XGLRFR was designed as a control experiment to show that the arginine residues are indeed not capable of participating in the catalysis. Unfortunately this structure is hardly soluble. This prevented proper purification after solid phase peptide synthesis, and only a very small amount was obtained. This was used to create a library, however upon oxidation a precipitate was formed (figure S107). Considering that this building block contains two permanent positive charges this was an unexpected result. Proper analysis could not be performed and the investigation into this compound was not pursued any further.

Permutations on the catalytically active lysine

The structures in the previous section allowed us to conclude that the inner lysine is mainly involved in the catalysis, as will be further described in chapter 2. The following three building blocks modified the N-terminus lysine residue to study the effects of differently spaced, secondary and aromatic amines. Their properties are investigated in the next section.

XGLOFK

Ornithine is a non-canonical amino acid that is nearly identical to lysine apart from the length of the aliphatic side chain. Where in lysine there are four carbon atoms between the backbone and sidechain amine, there are three in ornithine. This building block also assembles into a fibre consisting of hexamer macrocycles (figure S46). The emergence of this fibre is very rapid compared to the previous replicators, seemingly starting from the moment the library is created, and self-replication was confirmed by seeding experiments (figures S47). The TEM images show the fibrous structure formed, and these appear in smaller bundles with an average width of 16 ± 4 nm (figure 8B, table S3). Again it is seen on both the CD spectrum and the ThT assay that there is a large amount of. The CD spectrum however shows many more peaks outside of the β-sheet region, which might indicate other forms in which the fibre is ordered. The structure that is created also seems to be unaffected by stirring, since an unstirred library produces similar hexamer fibres (figure S51).

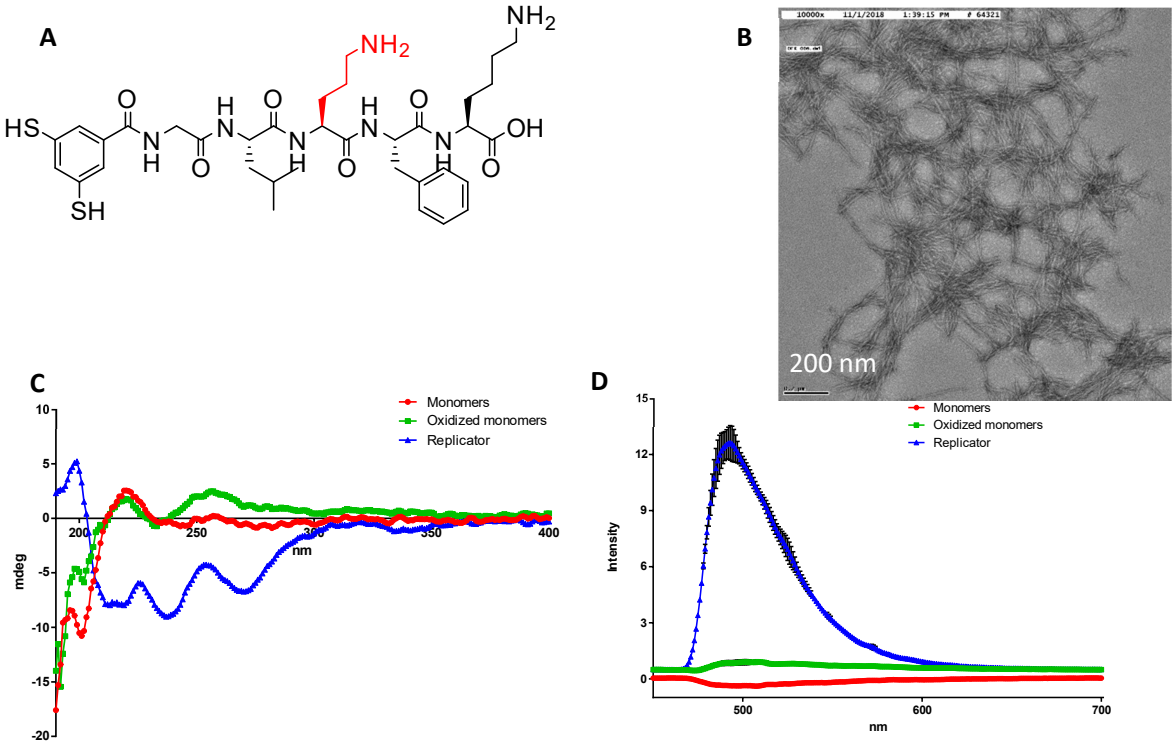


*Figure 8: **A)** Structure of XGLOFK **B)** Representative negative stain TEM image of a library of XGLOFK after approximately 2 weeks, showing fibres have formed. **C)** CD spectra and **D)** UV spectra of a ThT assay of XGLOFK (in triplicates) showing samples with monomers as red circles, oxidised monomers as green squares and fibres as blue triangles; error bars are indicated in black. Both spectra indicate that the fibres are based on β-sheets. Samples were prepared as described in the supporting information.*

XGLmKFK

The field of organocatalysis, that focusses on creating catalysts from organic molecules, has had many successes with the use of secondary amines.[48,49] This due to the fact that an iminium ion created from a secondary amine is more stable than that of a primary amine.[48] In order to see this effect in our system a methylated lysine residue was incorporated at the catalytic site. Interestingly this small permutation prevented the system from assembling, generating a mixture of macrocycles of varying sizes ranging from trimers to large oligomers up to a ring size of 23. In TEM images no supramolecular structures were found (figure S60). That no assembly occurred was further confirmed by CD-spectra and a ThT assay, which gave no indication of β-sheets (see figures S61).

XGLaFFK

Lastly, we wanted to investigate the effect of using aromatic amines, and the catalytic lysine residue was exchanged for an amino-phenylalanine residue. This non-canonical amino acid, has a p-phenylamine as sidechain. This aromatic amine was initially envisioned to be able to catalyse different reactions, such as an oxime formation reaction, in a similar fashion to an artificial enzyme recently created by Drienovská *et al.* as described in the following chapter.[50] When the library was completely oxidised there are still several ring sizes in the library; trimers, pentamers and hexamers (figure S71). Approximately 60% is trimer. On TEM images we can clearly see that at least one of these rings has the possibility to form fibrous structures (figure 10A). They have an average width of 18 ± 8 nm, indicating that they are laterally associating. However, we cannot determine which macrocylce is responsible for this structure as there are several present. Moreover, in seeding experiments it is shown that these structures are not self-replicating as formation of one macrocylce is not enhanced by addition of the fibre (figure S72). When the building block is oxidised, the library turns turbid which indicates that some large aggregates were formed. On a UPLC column this aggregate only comes of in the wash and appears to be smeared over the chromatogram (figure S64). In both the CD and ThT assay we can identify β-sheets formation for both the fibres and oxidised monomers (figure 9 C and D). However, the composition of the fibres still remains unknown. In *unstirred* samples this library mainly yielded unassembled trimers and tetramers (figure S76).

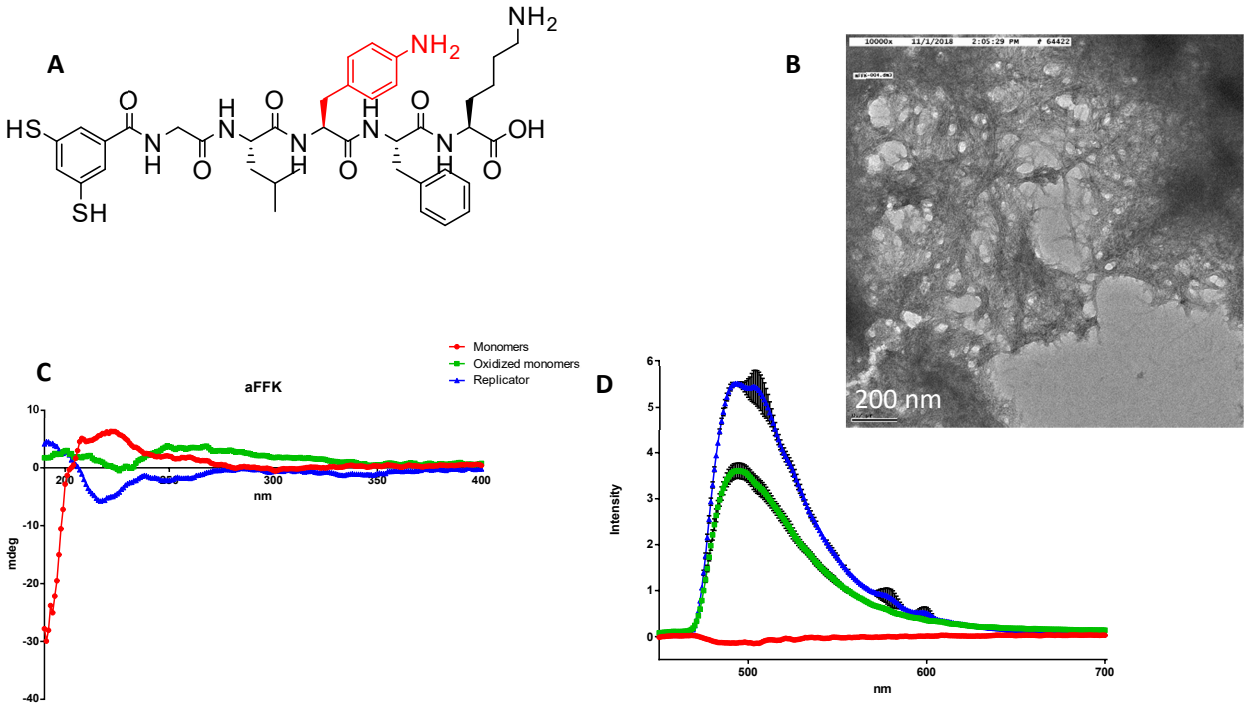


*Figure 9: **A)** Structure of XGLaFFK **B)** Representative negative stain TEM image of a library of XGLaFFK after approximately 2 weeks, showing fibres have formed. **C)** CD spectra and **D)** UV spectra of a ThT assay of XGLaFFK (in triplicates) showing samples with monomers as red circles, oxidised monomers as green squares and fibres as blue triangles; error bars are indicated in black. Both spectra indicate that the fibres are based on β-sheets. Samples were prepared as described in the supporting information.*

Permutations within the local environment

The final permutations tried to introduce groups that would change the local environment of the catalytic lysine, by changing possibilities for hydrogen bonding as well as charge interactions. This was done by changing the C-terminus lysine with tyrosine, histidine and homo-arginine. Besides this phenylalanine residue was changed for tryptophan.

XGLKFY

The tyrosine residue was chosen for its capabilities as a hydrogen bond donor. It is however also quite a hydrophobic amino acid, and not correctly placed in the alternating hydrophobic-hydrophilic residues. After a couple of days the molecules in the library start to form a hydrophobic aggregate. This is apparent from the UPLC chromatogram which shows the molecules only coming off in the acetonitrile wash (figure S103). On TEM pictures these aggregates can be seen quite clearly (figure 10B). Since no supramolecular was seen and the compound appeared to stick on the column, it was decided to stop following this library.

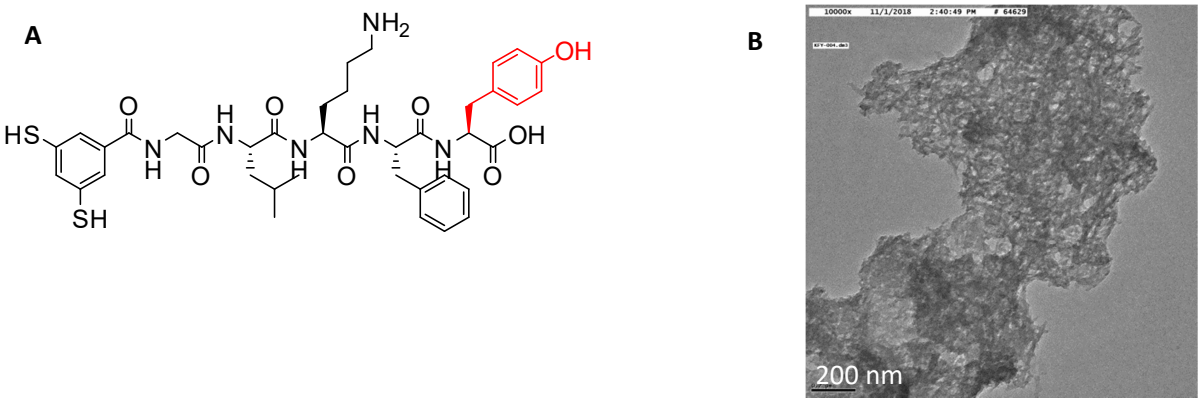


Figure 10: ***A)*** *Structure of XGLKFY* ***B)*** *TEM images of the aggregate formed by XGLKFY after 2 weeks at 200 nm*

XGLKFH

Also the permutation with the histidine residue was chosen for its hydrogen bonding capabilities. The initial results seemed promising however, when the library was followed for a couple of hours on the UPLC the compound started to precipitate and the signal was lost (figure S105 and S106). This prevented further characterisation and was therefore discontinued as well.

XGLKFhR

Homoarginine is a natural occurring, yet non-canonical amino acid, that can be found in the human body.[51] The homo prefix indicates the isomeric compound with an extra methylene group. In this case it describes arginine with a side chain that is one carbon atom longer. After slow oxidation by air, libraries form fibres of hexameric rings (figure S84). On TEM images fibres can be observed with a width of 14 ± 2 nm (table S5). After seeding experiments it was confirmed that this is indeed a self-replicator (figure S85). Investigation into the β-sheet structure by a ThT-assay and CD-spectroscopy showed that these fibres have relatively low amount of β-sheets compared to the other building blocks (figure 11A and B). The signal is only marginally increased over the monomer and fast-oxidised molecules. Not stirring this building block mainly makes trimers and tetramers, some other ring sizes can also be found in the base line. On TEM, a very small amount of fibres is visible, perhaps coming from one of the smaller species in the baseline (figure S89).

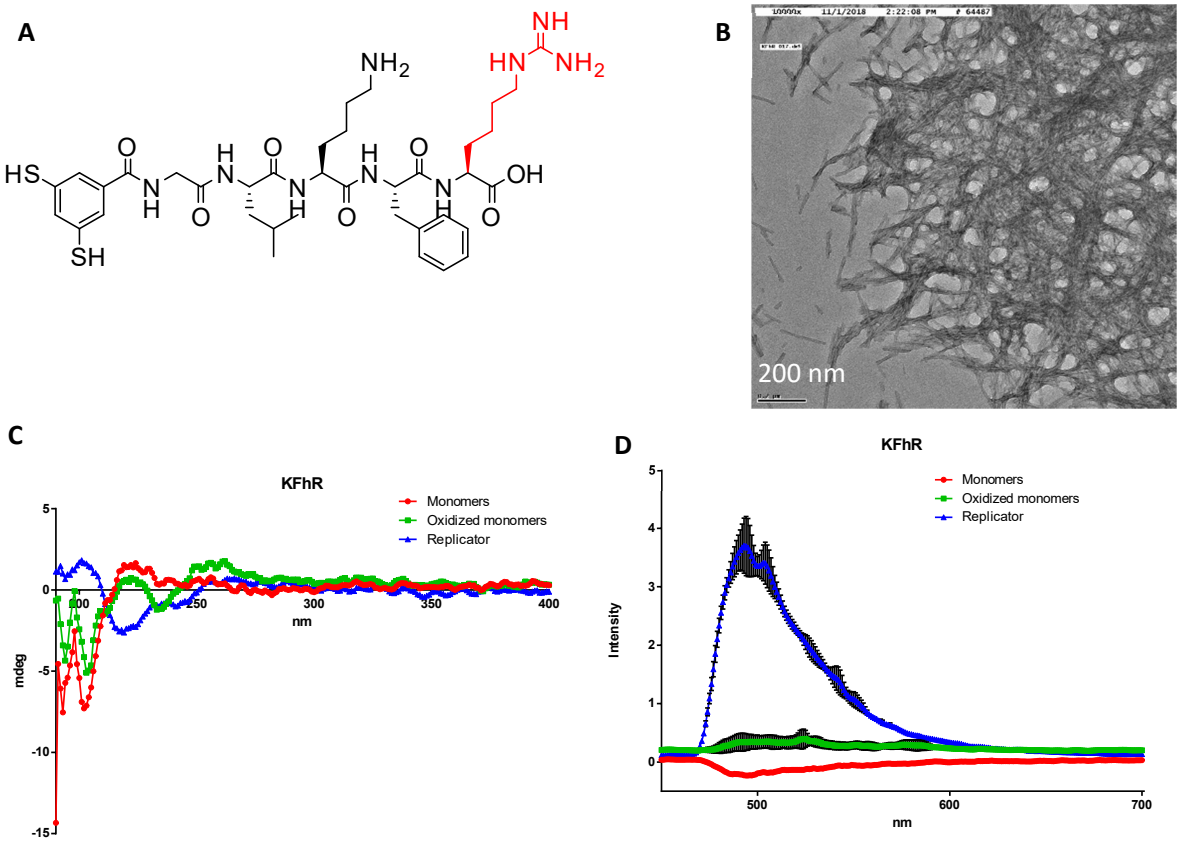


*Figure 11: **A)** Structure of XGLKFhR **B)** Representative negative stain TEM image of a library of XGLKFhR after approximately 2 weeks, showing fibres have formed. **C)** CD spectra and **D)** UV spectra of a ThT assay of XGLKFhR (in triplicates) showing samples with monomers as red circles, oxidised monomers as green squares and fibres as blue triangles; error bars are indicated in black. Both spectra indicate that the fibres are based on β-sheets. Samples were prepared as described in the supporting information.*

XGLKWK

The final structure discussed is the only one where the penultimate residue is exchanged. For this permutation tryptophan was chosen as it is both a hydrogen bond donating residue and relatively hydrophobic.[52] When slowly oxidised this building block forms hexameric rings that stack into fibres (figure S97). Interestingly after performing several seeding experiments, we were unable to see any autocatalytic behaviour common to self-replicators (see figure S98). Upon addition of a seed of fibres its own formation is not amplified. Nevertheless we continued our investigation into its characteristics. In ThT and CD experiments it can be seen that there are again quite some β-sheets present in these fibres (figure 12 C and D). Curiously on TEM-images the fibres appear to all align in one direction, however this is likely just an artefact of the TEM sample preparation (figure 12B). If the library is not stirred, it forms unassembled trimers and tetramers (figure S102).

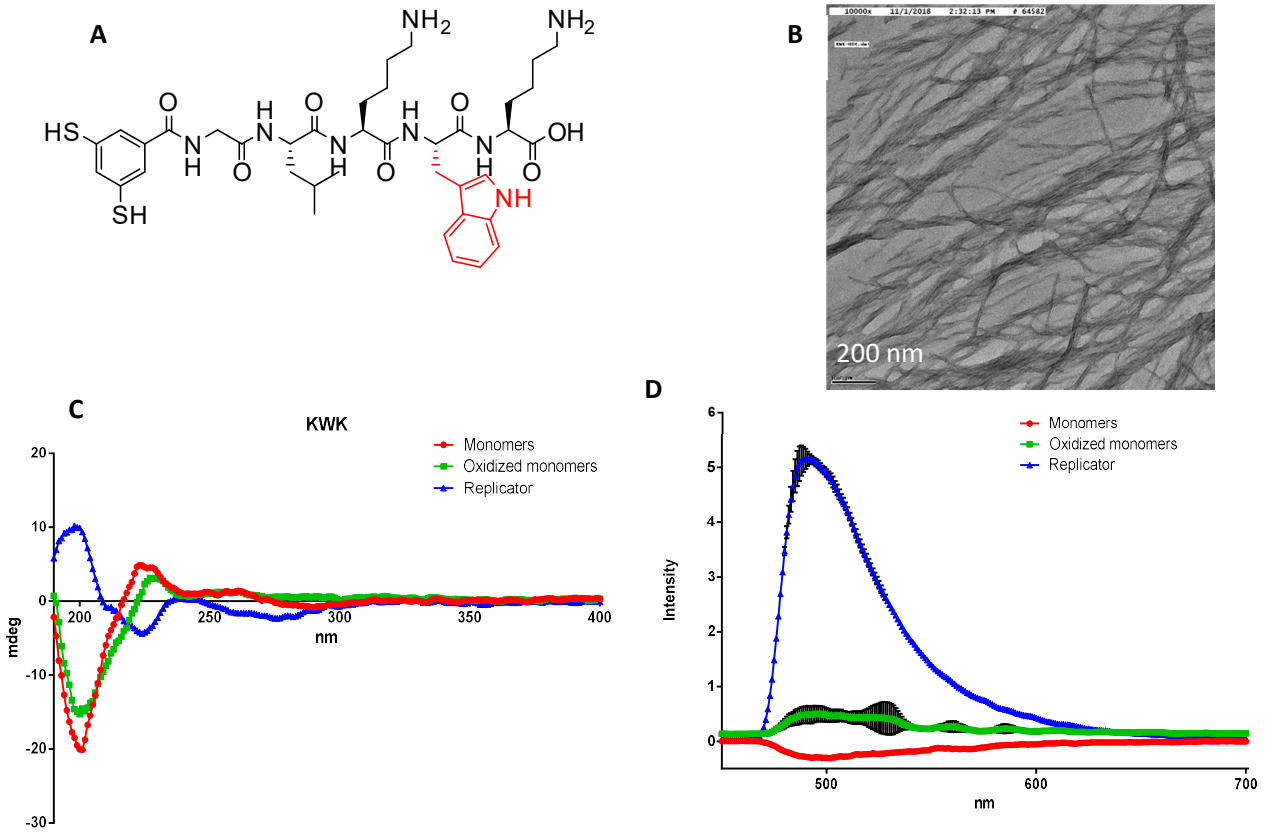


*Figure 12: **A)** Structure of XGLKWK **B)** Representative negative stain TEM image of a library of XGLKWK after approximately 2 weeks, showing fibres have formed. **C)** CD spectra and **D)** UV spectra of a ThT assay of XGLKWK (in triplicates) showing samples with monomers as red circles, oxidised monomers as green squares and fibres as blue triangles; error bars are indicated in black. Both spectra indicate that the fibres are based on β-sheets. Samples were prepared as described in the supporting information.*

General remarks on replicating structures

Having created several replicating structures, some generalities in their structures appear to exist. It is first of all rather interesting to observe that all the supramolecular assemblies are formed by hexamer macrocycles. Indicating that for this specific type of sequence, *i.e.* core-linker-hydrophobic-charged-hydrophobic-charged, the hexamer ring size is the most optimal.

For all building blocks, another set of libraries was created that were not stirred. In general for most libraries this meant that fibres could not emerge and the library mainly makes trimer and tetramers. However, in the case of the ornithine replicator a hexameric fibre still emerged. Besides, the two arginine replicators XGLKFR and XGLRFK also made fibres when left unstirred. However upon closer inspection these fibres are not made from hexamers but instead of trimers, and in the case of RFK both trimers and pentamers are observed. We did not focus our attention on the kinetics with which these fibres emerged, but this could be the focus of a further study, since the mechanism dictates that the input of mechanical energy is necessary for exponential growth.

Libraries were also followed at room temperature, but only the XGLOFK building block created fibres. Even when these building blocks were seeded with fibres of the sequence at room temperature no structures appeared. This indicates that the formation of these self-replicators is a process with a rather high energy barrier. It requires higher temperatures to be overcome. It is interesting to note that had these libraries only been investigated at this lower temperature as used to be the norm in the group, almost no replicating structure would have been found. It might be worthwhile to return to previous disregarded building blocks to investigate these at a higher temperature.

Chapter 2 Catalysing reactions

Fluorescence plate reader experiments on aldol reactions

Having characterised the structures of the self-replicators, we proceeded to investigate their ability to catalyse various chemical reactions. To be able to measure such reactions properly, the reactants or products need to be followed by time dependent spectroscopic techniques. Ultra-Performance Liquid Chromatography (UPLC) was used in the characterisation of the replicators, and seemed an obvious candidate. However this technique is rather time consuming and in order to screen reactions a more efficient technique was warranted.

Another method that allows screening of many reactions at once was by utilising a fluorescence plate reader (FPR). The fluorescence plate reader is an instrument that can excite and follow the fluorescent signal of up to 96 different samples. In this manner it can be used to efficiently screen possible reactions catalysed by replicators, provided that either one of the reactants or the product is fluorescently active. The retro-aldol cleavage of methodol that was discussed previously will yield 6-methoxy-2-naphthaldehyde as product. This compound is fluorescently active, which means this technique can be used for screening purposes.[40]

Previously our group had no experience with this technique and thus we needed to investigate if it gives reliable results. From its use in other fields we knew that photobleaching of the fluorescent compound would decrease the signal over time,[53] so we started out by investigating this effect on 6-methoxy-2-naphthaldehyde as to be able to correct for this later. 6-methoxy-2-naphthaldehyde was added to a well plate in various concentrations (20 µM to 0.16 µM, halving the concentration 7 times) and its fluorescence measured over one hour. The incident light had a wavelength of 330nm and fluorescence was measured at 452 nm. If we extrapolate the data of multiple one hour experiments over time, we can see how this bleaching would progress over time in a longer experiment (see figure 13).

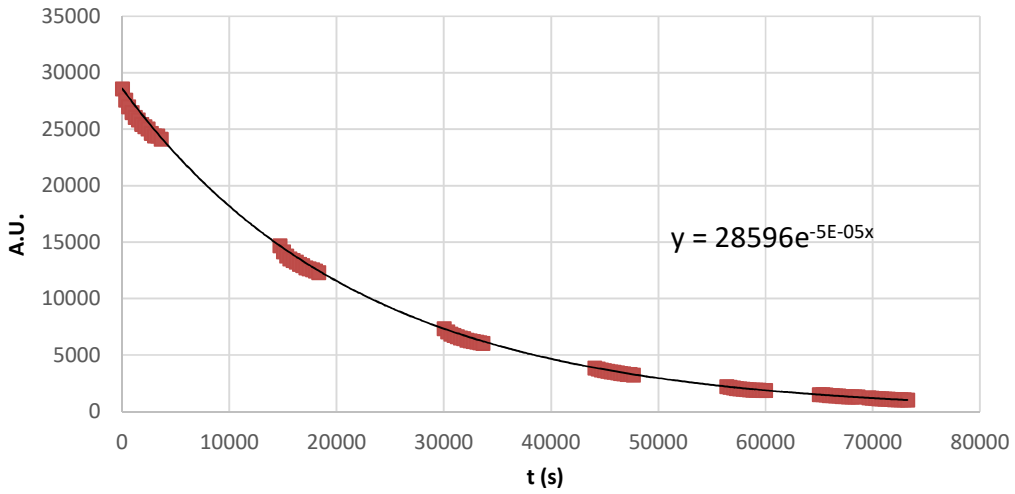


Figure 13: Photobleaching of 6-methoxy-2-naphthaldehyde (20 µM to 0.16 µM, halving the concentration each row) extrapolated over time, red squares indicate actual data from 6 repeats, and the black line demonstrates the fit. The excitation wavelength was set to 330 nm and fluorescence was measured at 452 nm.

From this fit we can calculate the how many molecules are bleached per time unit of excitement and correct for this in the actual experiment. In this experiment we followed the breakdown of methodol (400 µM) by fibres of XGLKFK (25 µM) over time with both the FPR and UPLC. The product formation observed in both these experiments is not the same, with the slope of the product formation differing by approximately 20% (see figure 14).The difference is likely a result from the likelihood for experimental error in filling 96 individual wells, as it requires several more pipetting steps. Nevertheless we were satisfied that this technique could be used to efficiently screen possible reactions involving 6-methoxy-2-naphthaldehyde, yet more accurate data should be obtained by UPLC experiments.

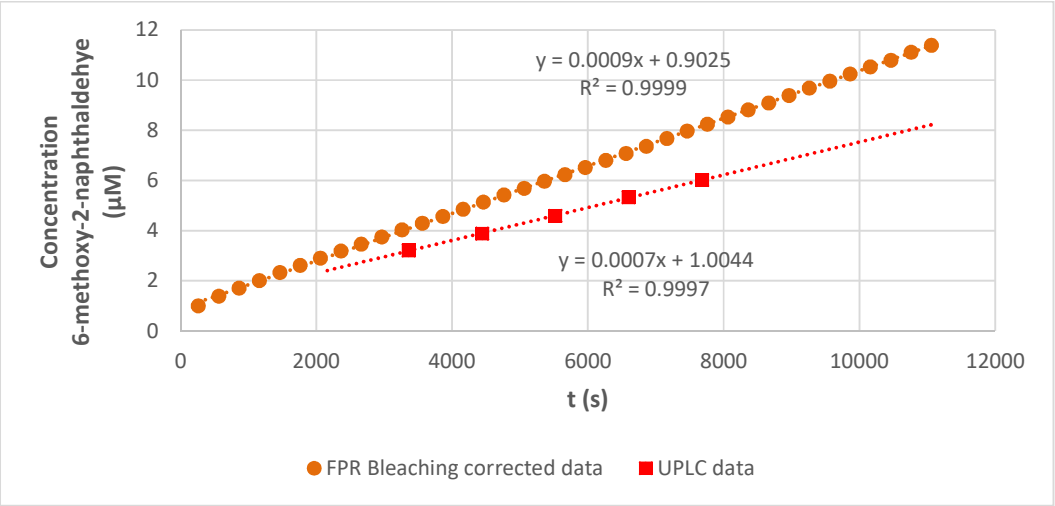


Figure 14: Formation of 6-methoxy-2-naphthaldehyde in µM from the breakdown of methodol (400 µM) by fibres of XGLKFK (25 µM) as followed by a fluorescence plate reader (orange circles) and as followed by UPLC (red squares). Samples contained 2% acetonitrile.

After having established that the FPR analysis could be performed reliably we chose several ketones and aldehydes that might be able to participate in a bond forming aldol reaction. The 23 ketones and aldehydes were chosen based on their availability and structural differences (see figure 15). We mixed them in several ratios in a 96 well plate and ran the fluorescence experiment as described above. With this method we hoped to see that the reaction kinetics would become more favourable, and small amounts of the aldol product would be formed. This would then prove a good starting point to expand upon. However, even though the FPR analysis appeared promising for some carbonyls, they turned out to not yield any significant amounts of product when subsequently studied by UPLC. We therefore abandoned this brute force method of finding suitable aldol reactants, and instead focused our attention on finding different reactions.

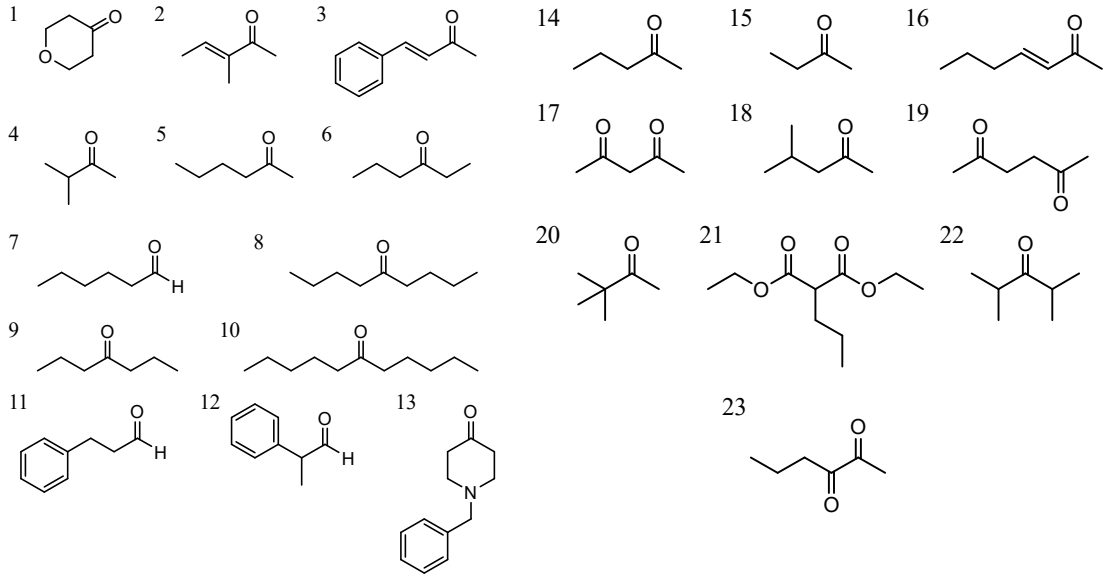


Figure 15: Scheme of the 23 carbonyl reactant chosen for the FPR analysis

Kemp elimination

One of the reactions that was investigated was the Kemp elimination of 5-nitrobenzisoxazole. In this reaction a general base is used to deprotonate the isoxazole hydrogen, which opens up the nitrogen oxygen bond forming a cyanide and an alcohol (figure 16A).[54,55] In the field of protein engineering this reaction had shown great potential in the rational design of enzymes that catalyse this reaction. [56] Similar to the mechanism for the FMOC deprotection reaction, We envisioned that the replicator fibres would be able to act as the base. Moreover we hypothesised that through neighbouring hydrogen bonding groups the transition state might be stabilised, as described in the literature.[56] To test whether the supramolecular fibres can catalyse this reaction an experiment was performed samples with 25 µM of XGLKFK fibres and 100 µM 5-nitrobenzisoxazole were made, and the product formation was followed by UV spectroscopy at 380 nm (see figure 16B). After following the reaction by UV we found that XGLKFK fibres did not catalyse the reaction, showing no enhanced rate over the background reaction in buffer. Moreover, it appears as though the reaction is slowed down, which is likely an artefact from turbidity created by the fibres.

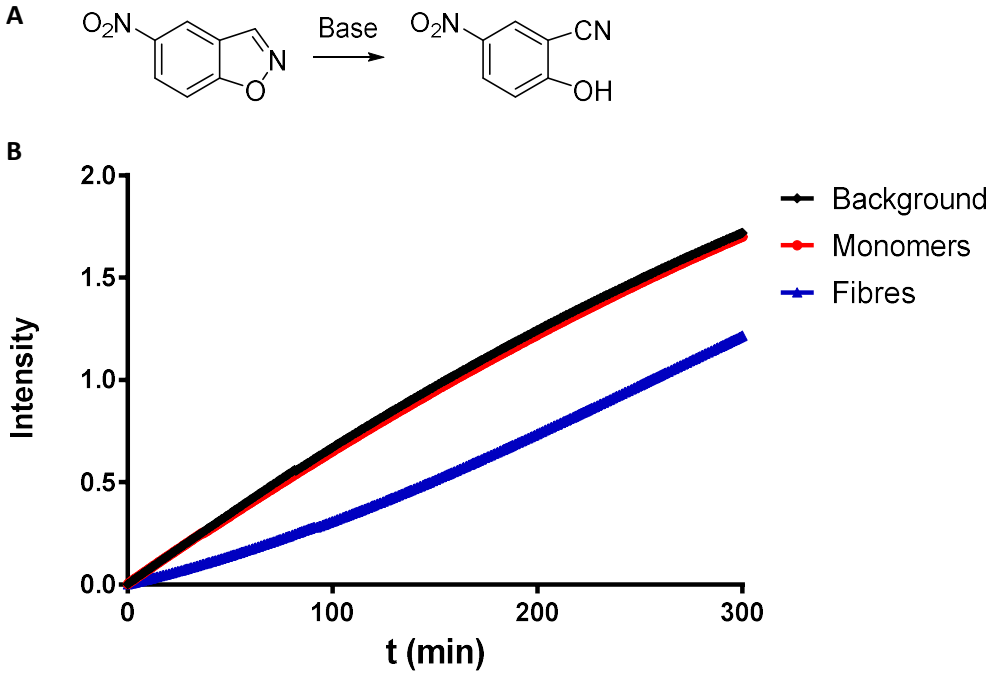


*Figure 16: **A)** The reaction scheme for the Kemp elimination of 5-nitrobenzisoxazole **B)** Graph following the product formation of the reaction at 380 nm, at pH 7.4. Starting with 100 µM of 5-nitrobenzisoxazole and 25 µM fibres of XGLKFK. With the background rate in black (top), the monomers in red (middle) and fibres in blue (bottom).*

Hydrazone formation

Another reaction that recently yielded great results in the design of enzymes, is the hydrazone formation reaction shown in figure 17A below.[50] Drienovská *et al.* used the unnatural aniline residue, aminophenylalanine, to create a designed enzyme capable of catalysing this reaction. This is the same aniline residue that we used in one of our sequences. Using the fibres with this residue and following the reaction by UV showed that the reaction rate is enhanced very slightly over the background reaction (figure 17B). Also the oxidized monomers can slightly catalyse the reaction. Over time however, some aggregates formed in these samples which causes the light to scatter. This makes the lines more jagged and the intensity decreases. Since the samples are then no longer homogeneous they also diverge. When the sample with the monomers was followed over time, it appears that the reaction rate is larger than for the other assemblies. This might mean that the monomers are capable of catalysing the reaction or that a side reaction is taking place. Given that XGLaFFK is not a replicator, and the increase in reaction rate is only minor we did not investigate this reaction any further.

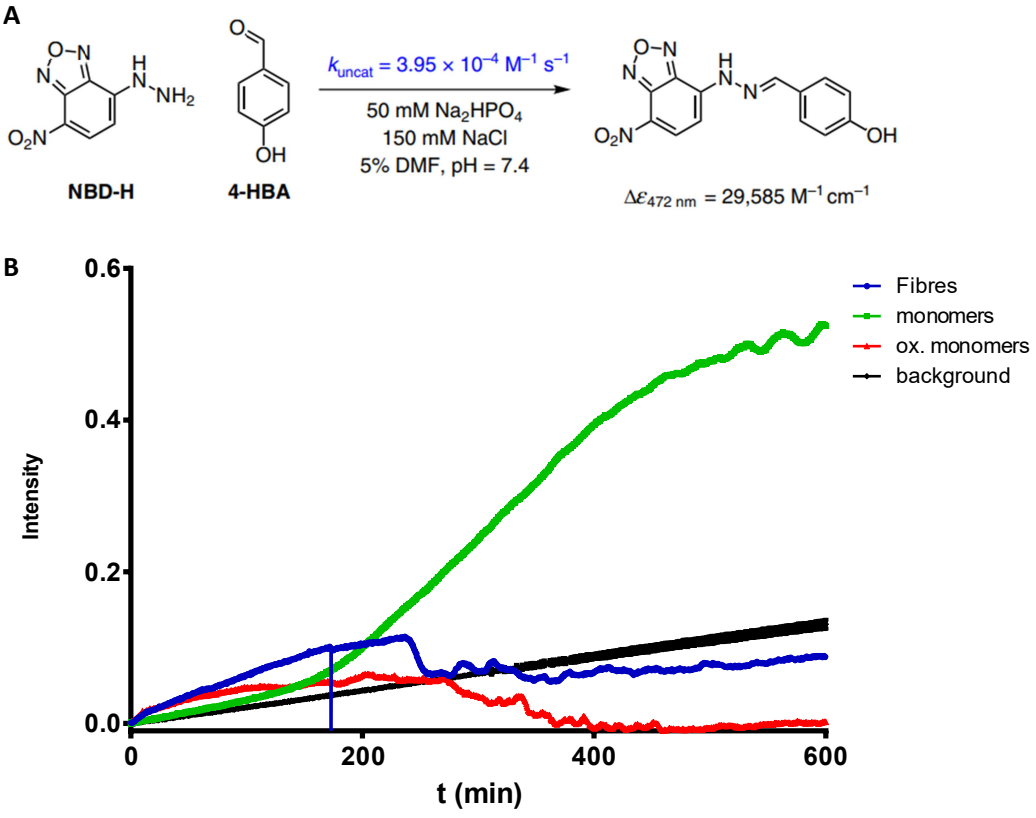


*Figure 17: **A)** Reaction scheme of the hydrazine formation. Taken from Drienovská et al[50]. **B)** Graph following product formation of the reaction between 4-HBA (5 mM) and NBD-H (200 µM) by the XGLaFFK building block (25 µM) at 472 nm at pH 7.4 in phosphate buffer with 5% DMF. Samples with monomers are green (top), oxidised monomers are red (bottom) and fibres are shown in blue (second from bottom). The background rate is shown in black (middle)*

Retro-aldol cleavage

The main focus of this thesis has been to explore the parameters that govern the retro-aldol reaction of methodol. As described in the introduction it is proposed that it is the amine from a lysine in the structure that makes an imminium ion, which facilitates the retro-aldol cleavage of methodol, yielding 6-methoxy-2-naphthaldehyde and acetone (figure 18A). In previous work it was already established that XGLKFK can catalyse the reaction only when in assembled form, and that the lysines from this sequence are in a local environment that perturbs their pK_a. Previous kinetic measurements for the XGLKFK fibres demonstrated that the amount of active sites was relatively low, with less than 10% of the lysines in the fibre being active.[38] It was however unknown which lysine makes the iminium ion and how it is perturbed. The sequences described in the previous chapter were designed to probe these characteristics specifically, as described there.

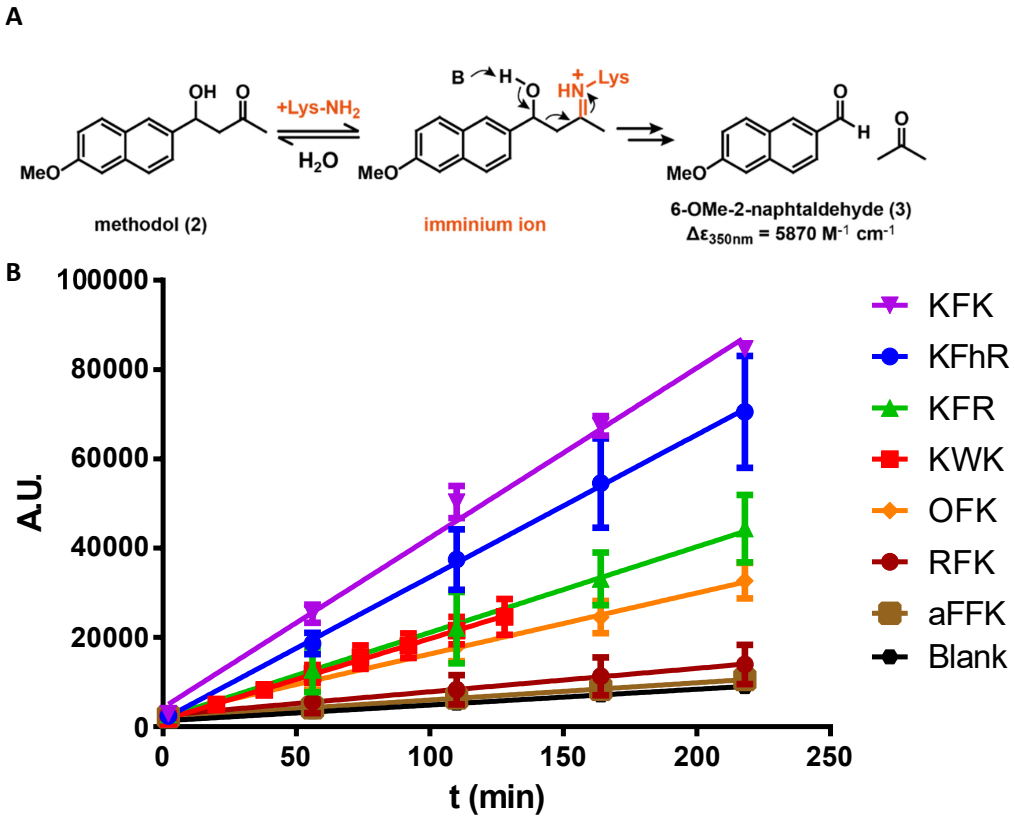


*Figure 18: **A)** Proposed mechanism of the retro-aldol reaction of methodol being broken down (adapted from Ottelé et al.)[38] **B)** Formation of 6-methoxy-2-naphthaldehyde by breakdown of methodol (400 µM) over time in arbitrary units from UPLC data at 313 nm using pre-formed fibres (25 µM) of the various building blocks at pH 8.1. Results are based upon three repeats. For clarity the prefix XGL- is omitted for all. Data of the fit is shown in figure S2.*

We set out to probe the catalytic prowess of the various building blocks. Fibres of the various building blocks were mixed with methodol, and the formation of 6-methoxy-2-naphthaldehyde was followed by UPLC (see figure 18B). In this experiment 400 µM Methodol was mixed with 25 µM of fibre in borate buffer with a pH of 8.1. To allow methodol to dissolve 2% acetonitrile is present. The reaction was also monitored with the building blocks in the monomeric and fully oxidised forms. These do not catalyse the reaction; their results can be found in the supporting information (figures S1).

It is helpful to frame these results with the help of the schematic representation of the active site from the previous chapter, and the figure is shown again here. Only upon formation of the fibre is the active amine created, as only fibres are active as a catalyst. What is more, only the sequences that have an amine on the inner position can catalyse the reaction. The difference between the rates of those various sequences is relatively small, spanning less than a factor of three. If we assume that structures that are created by the various building blocks are the same we can infer that region one only plays a small role in the active site, given that an amine is present at the correct position.

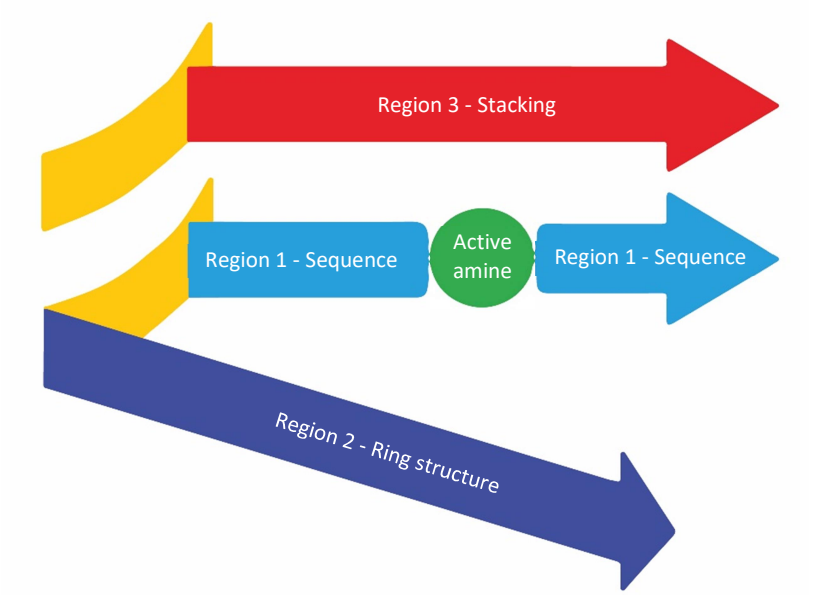


Figure 19: Schematic representation of the active site, and the parameters that influence it

Instead it must then be the supramolecular structure that is mainly responsible for forming the active site. In previous work it was already established that the fibres poses a certain heterogeneity.[44] Comparing this fact with the observation that only a small number of lysines is catalytically active and that catalysis seems less dependent on the amino acid sequence of the building block, suggests that it is at these disordered sites where the amines become perturbed. This means that the active site is dependent on the building blocks in region two and three to be in the correct position in these disordered sites. This is because the building block in these regions needs to be in the right spatial configuration, and at disordered sites different spatial configurations might be possible.

This observation is in contrast with the enzymes found in nature, where the structures are highly specific and deviations in the sequence usually have a major impact.[27,40] In the context of the origin of life this is interesting as it shows that crude structures with approximately similar sequences can already give rise to chemical transformation. Early and less complex systems might utilise this in their early evolution, without the necessity of an accurate replication system supporting it.

From the catalytic efficiencies shown in figure 18 we can divide the fibres into three distinct groups. At the bottom of the graph we find those replicators that have no aliphatic amine on the third position. These sequences barely catalyse the reaction; at most their rate is enhanced 1.5 times over the background. From this we can conclude that it is the lysine closest to the fibre core that is responsible for the catalysis.

The three sequences that occupy the middle of the graph show approximately a four to five fold increase over the background. These sequences all contain an aliphatic amine on the third position. Since their rates are so close together, it must be the local environment or the amount of active sites that brings about the differences, as discussed above. Compared to the XGLKFK fibres they perform 2.5 times worse.

The one sequence that performs similar to the XGLKFK standard, has a homo-arginine residue on the fifth position. When we compare this sequence to the structure with a regular arginine on the same position, we can explain its activity in one of two ways. The length and flexibility of the homo-arginine side chain might be ideally suited to perturb a neighbouring lysine, either within the same sequence or on another ring. In a similar fashion the length of the ornithine side chain might move it away from this ideal position and decrease its activity. This explanation assumes that the changes that are introduced do not influence the structure of the fibre, which might also be an explanation.

When looking at the ThT and CD spectroscopy data we can see that these fibres are not structured in exactly the same manner. Even though there are subtle differences they are largely the same. These subtle differences might however correlate well with their activity as a catalyst, since the β-sheets offer a hydrophobic space for the reaction. In the graph below the maximum intensity for the ThT analysis (see previous chapter) is plotted versus the slope of the formation of 6-methoxy-2-naphthaldehyde (See Figure 20). This graph shows that a small amount of β-sheets is present along the entire range of activity. What is more, there appears to be no relation between the different points, and we can therefore conclude that the catalysis cannot be explained by this single parameter.

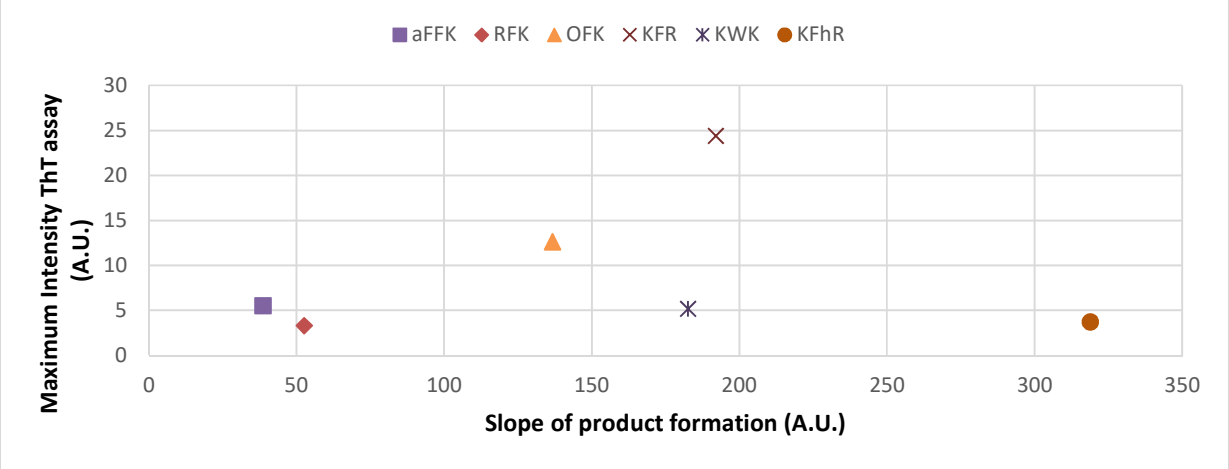


Figure 20: The activity of the different fibres plotted against their maximum ThT intensity for the retro-aldol reaction extracted from the ThT assays and the fit of the experiment of figure 18.

Sulfmoc deprotection

The other reaction catalysed by XGLKFK fibres is the Fmoc deprotection. To make this more applicable in our system Fmoc was sulfonated to gain more solubility, making Sulfmoc. The deprotection of Sulfmoc containing compounds by the supramolecular assemblies uses the same perturbed lysines as a base to deprotonate the fluorenyl hydrogen. The deprotonation leads to the formation of CO_2, dibenzofulvene and the previously protected group (figure 21A). Since the same amines in the sequences would perform the deprotonation, we expected that the performance in the retro-aldol reaction and Sulfmoc deprotection would be the same.

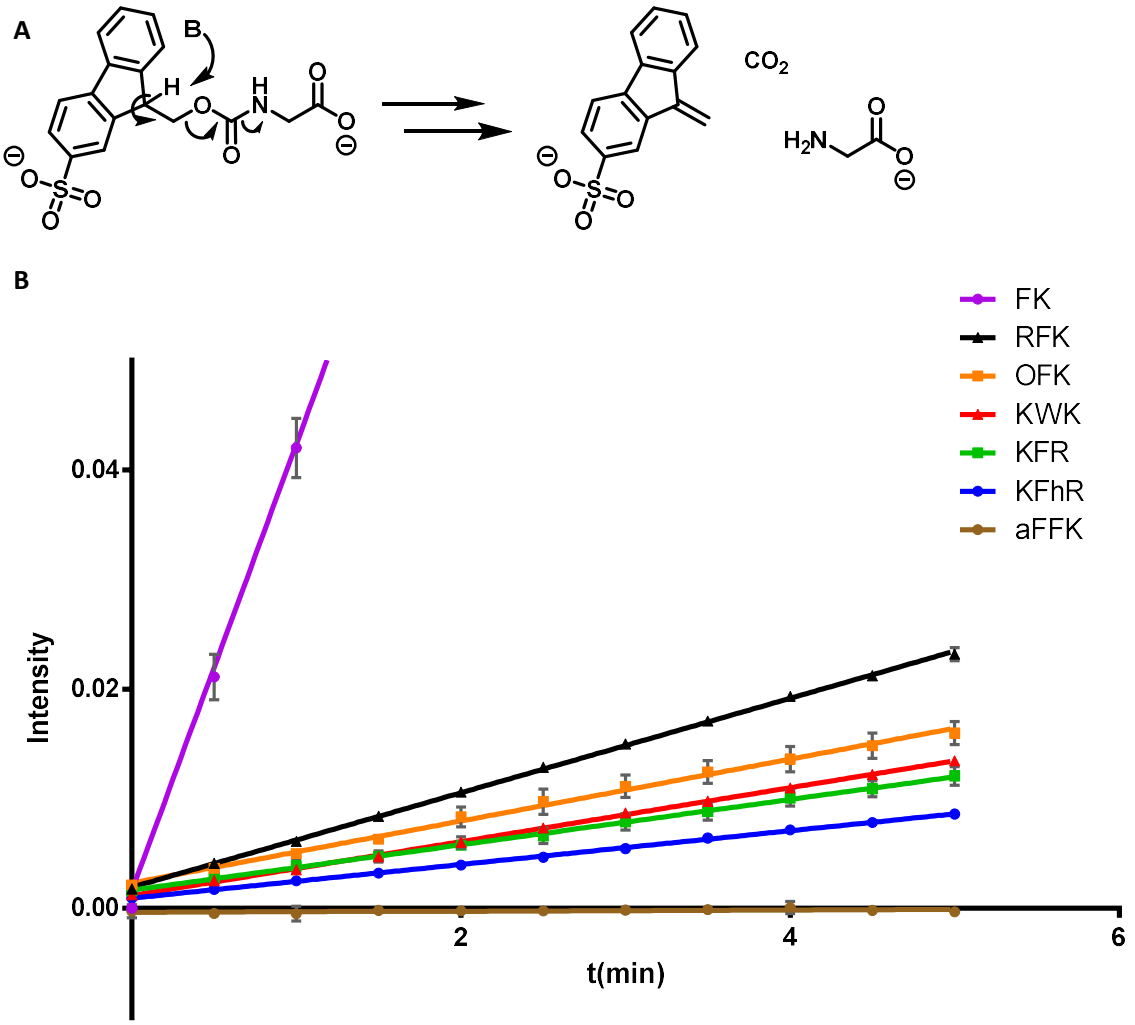


*Figure 21: **A)** Mechanism of the Sulfmoc deprotection reaction of protected glycine **B)** Formation of dibenzofulvene from the deprotection of Sulfmoc-glycine (100 µM) over time in arbitrary units from UV data at 308 nm using fibres (40 µM) of the indicated sequences at pH 8.1. Results are based upon three repeats. For clarity the prefix XGL- is omitted in the legend.*

For these experiments we mixed pre-formed fibres (40 µM) with Sulfmoc protected glycine (100 µM). The initial linear rates were of interest and are shown in figure 21B, where the UV intensity from dibenzofulvene is followed over time at 308 nm. Like before, the monomers and oxidised monomers do not catalyse the reaction, with one exception. The oxidised XGLKFhR building block, which seemingly makes no structure, also shows a small amount of activity for the reaction (see figure S4). Over time this activity is minor compared to the fibre.

What is striking about these results is that all sequences perform approximately 10 times worse compared to XGLKFK. Moreover, their activity seems to correlate with the activity in the retro aldol reaction, however in an inverse manner; a good retro-aldol catalyst is a relatively poor Sulfmoc deprotection catalyst and vice versa. This suggests that active sites for the deprotection reaction are different and have opposite characteristics to those from the retro-aldol reaction. It is now the outer position in which the lysine is most active. Although for this reactions both positions are capable of catalysing the reaction.

Similar as for the retro-aldol reaction, we tried to explain the activities as a function of the amount of β-sheets. Again we plotted the slope of the product formation versus the maximum intensity of the β-sheets (see figure 22). A direct correlation between those characteristics is however still not possible as seen in the graph below, indicating that this reaction is also influenced by several parameters.

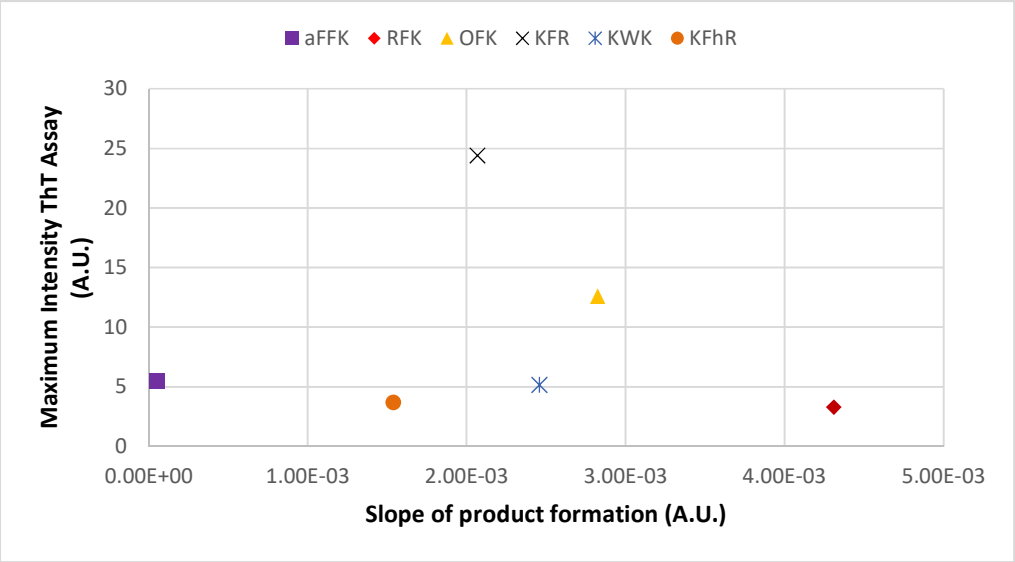


Figure 22: The activity of the different fibres plotted against their maximum ThT intensity for the retro-aldol reaction extracted from the ThT assays and the fit of the experiment of figure 21.

Catalysis by mixed libraries

In an attempt to further perturb the active sites in the structure, libraries were made where two building blocks were mixed in a one to one ratio. In mixed fibres the sequence as present in region one from figure 19 does not change. Instead it is the other two regions that *can* house a different building block and influence the active site by presenting a different residue close to the active amine. This can influence the active site by bringing about a change in hydrophobicity or acidity of the active amine.

Of the many combinations possible, six mixed libraries where chosen. XGLKFK was mixed with five other building blocks, since it was the best performing catalyst. We opted for the combination with XGLKWK, XGLKFR, XGLRFK, XGLKFhR and XGLOFK. Also the combination of just the two arginine replicators was of interest as their sequences are complementary to each other. All building blocks were mixed in a 1:1 ratio and they all made fibres with a statistical distribution of the possible hexamer macrocycles. Unfortunately these fibres caused all the libraries to become gel-like, which complicated making homogeneous samples. This was especially the case for the fibres made by combining XGLKFK with either XGLKFR or XGLKFhR. A lot of peak area was lost on the UPLC chromatogram (figures S104 and S106), due to the difficulty in sample preparation by pipetting gels. Their activities as catalysts are included but they are not representative of an accurate measurement.

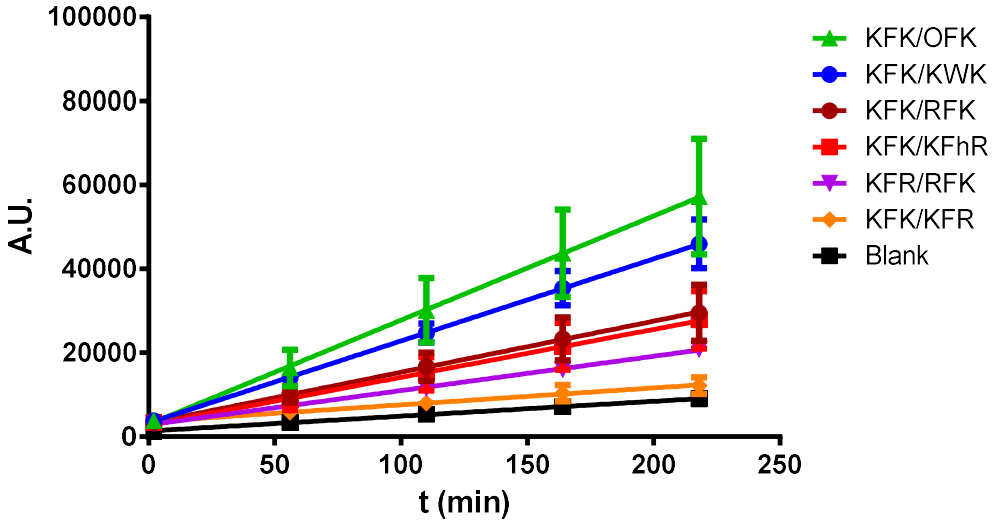


Figure 23: Formation of 6-methoxy-2-naphthaldehyde by breakdown of methodol (400 µM) over time in arbitrary units from UPLC data at 313 nm using pre-formed fibres (25 µM) of the various building blocks at pH 8.1. Results are based upon three repeats. For clarity the prefix XGL- is omitted for all.

For these mixed libraries we investigated their activity towards both the retro-aldol reaction and Sulfmoc deprotection as in the previous chapter. First we looked at the methodol catalysis (See Figure 23). When mixing building blocks, the results are best framed with respect to their individual counterparts. We were interested in seeing if introducing different building blocks in region two and three of the active site could perturb the active lysine more. If a different building block would have no effect, the mixed fibre will give the average of the two rates of the individual fibres. Any deviation from this average means that these building blocks perturb the active site in a different way.

All the mixed libraries are plotted together with their individual counter parts in the graphs below. The mixed library is represented by the grey line with the squares in each instance. As mentioned before, the KFK/KFR and KFK/KFhR fibres could hardly be pipetted which likely plays an important role in the lack of

catalytic activity as seen in the experiment. The inhomogeneity of the libraries of the other fibres also caused problems in the experiment. Even though the libraries were homogenised by vigorous pipetting, it was difficult to consistently add the same amount of fibre. This makes the data less reliable than for the single building block fibres.

Nevertheless, most of the mixed fibres perform worse than average of the two rates. Although the error for the KFK/OFK fibres is too large to be certain. We can therefore carefully conclude that it must indeed be the interaction from a neighbouring building block in region two or three that is responsible for perturbing the lysine.

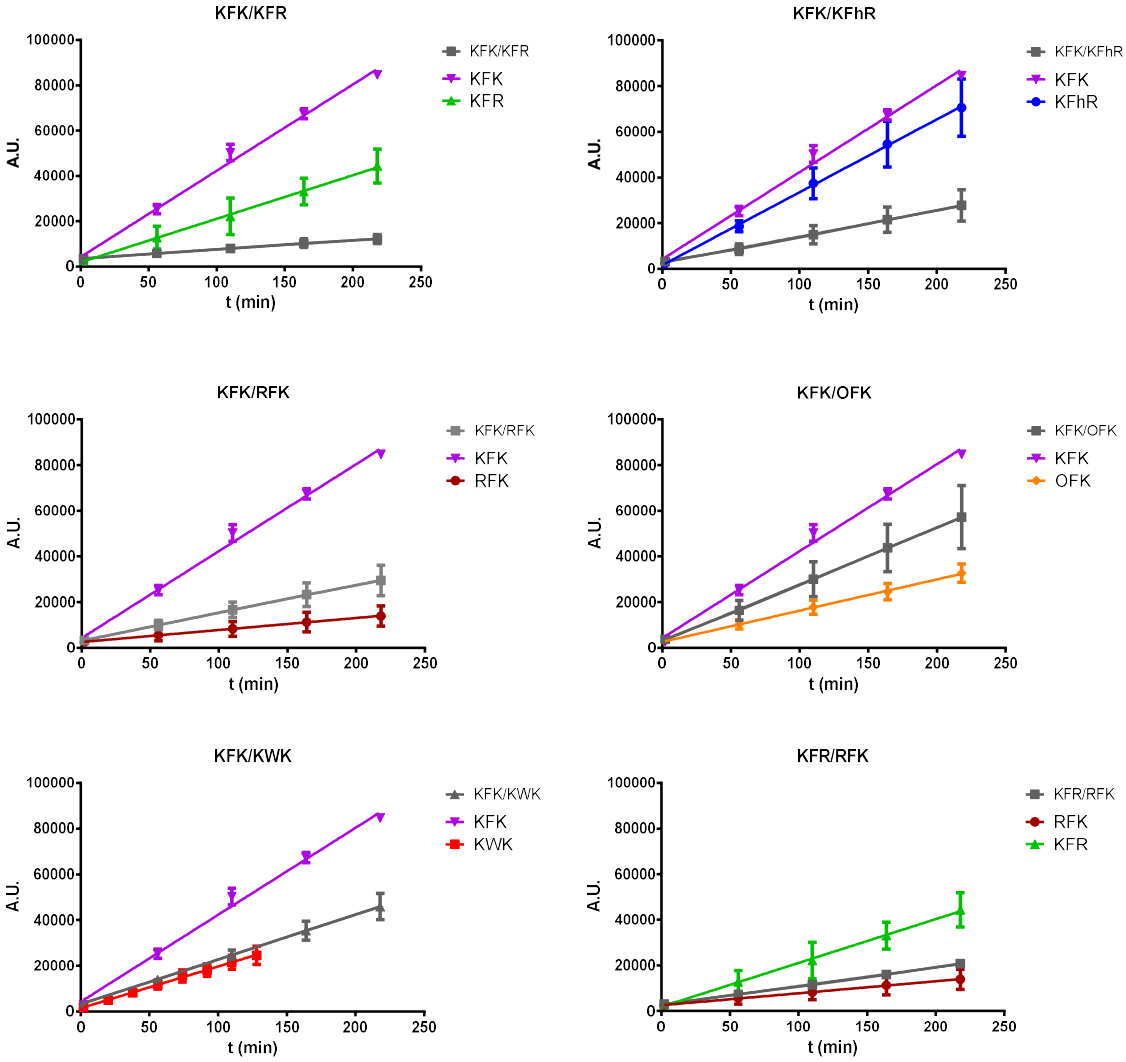


Figure 24: Results of the retro aldol reaction performed by mixed fibres as seen in figure 24 plotted together with their individual counter parts.

In the case of the Sulfmoc deprotection we observe a similar scenario. Again the interactions between the different building blocks cause the mixed fibre to differ from the average (figure 25). This time however there are is also a mixed fibres that show a positive interaction. The KFR/RFK fibres perform as good as its best counterpart. The activity is however still very low compared to the fibres of XGLKFK.

Like for the individual fibres, the results for both reactions are not similar. This again shows that the active sites of the two reactions are not the same. Both of them, however, can be influenced by changing the neighbouring building blocks present in region two and three of the active site.

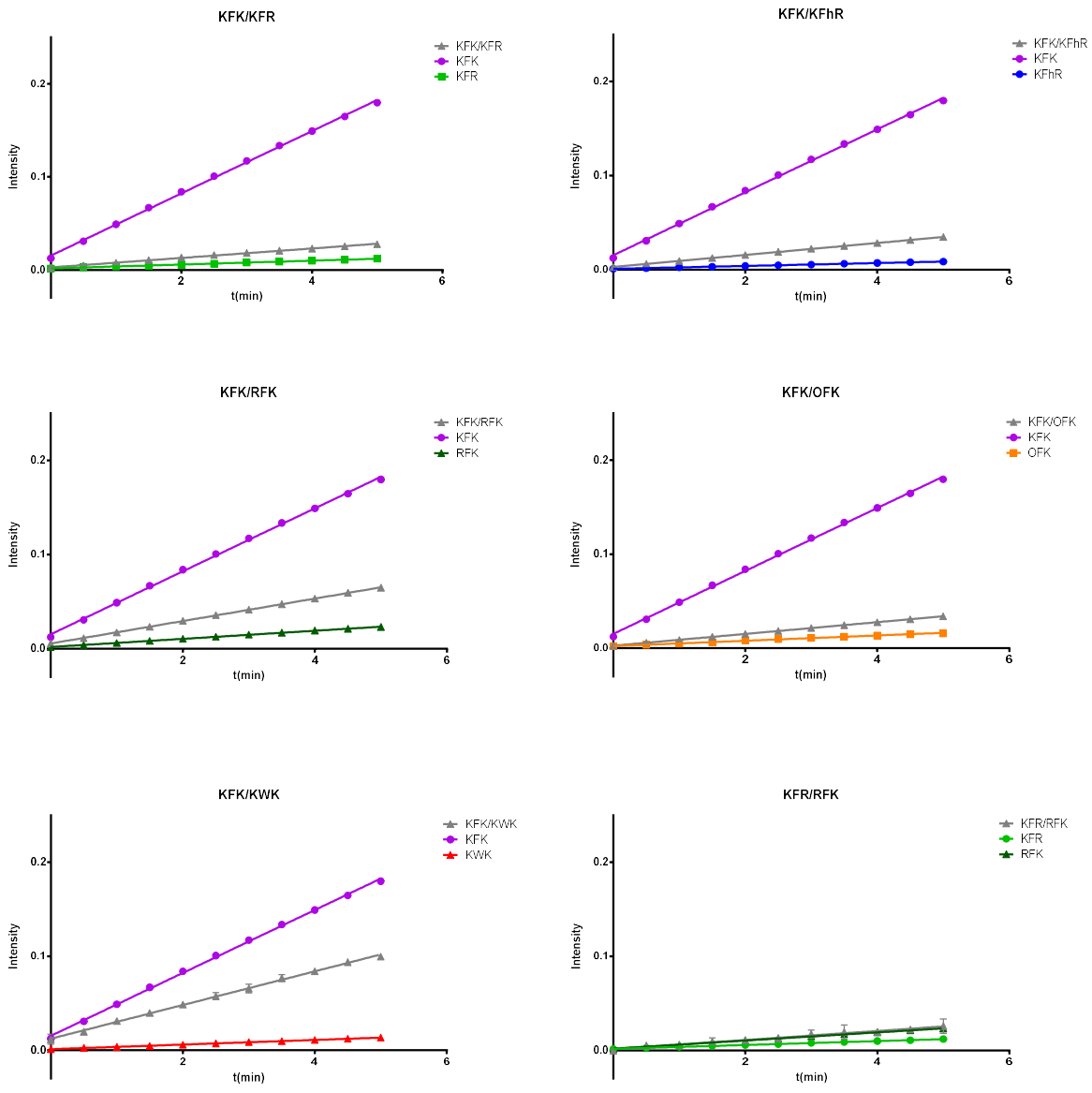


Figure 25: Results of the Sulfmoc deprotection reaction performed by mixed fibres plotted together with their individual counter parts.

Conclusions

For this thesis we set out to further develop the pathway towards metabolic capabilities in self-replicating systems. Within the context of de-novo life, self-replicating structures that are capable of chemical transformations are an essential step in the creation of a life-like chemical system. Ten new building blocks were investigated for their ability to form self-replicating structures as well as the ability to act as a catalyst. These are shown in figure 26. Of these ten building blocks, six formed supramolecular fibres. Four of those showed self-replicating behaviour. These six fibres were further characterised by several techniques to probe their structure and properties.

Several interesting observations were made during the characterisation of these building blocks. All self-replicators created fibres made-up of hexamer macrocycles. This shows that for sequences of this type, this specific ring size is the most favourable. Another interesting observation was made when the libraries were followed at a lower temperature. All but one replicator failed to assemble into fibres. We concluded that the self-replication process has a rather high energy barrier. If the building blocks were not followed at an elevated temperature, their self-replicating properties might have been overlooked. To find more self replicators, future experiments should return to previous building blocks that failed to replicate at room temperature, and these should be investigated at an elevated temperature as well.

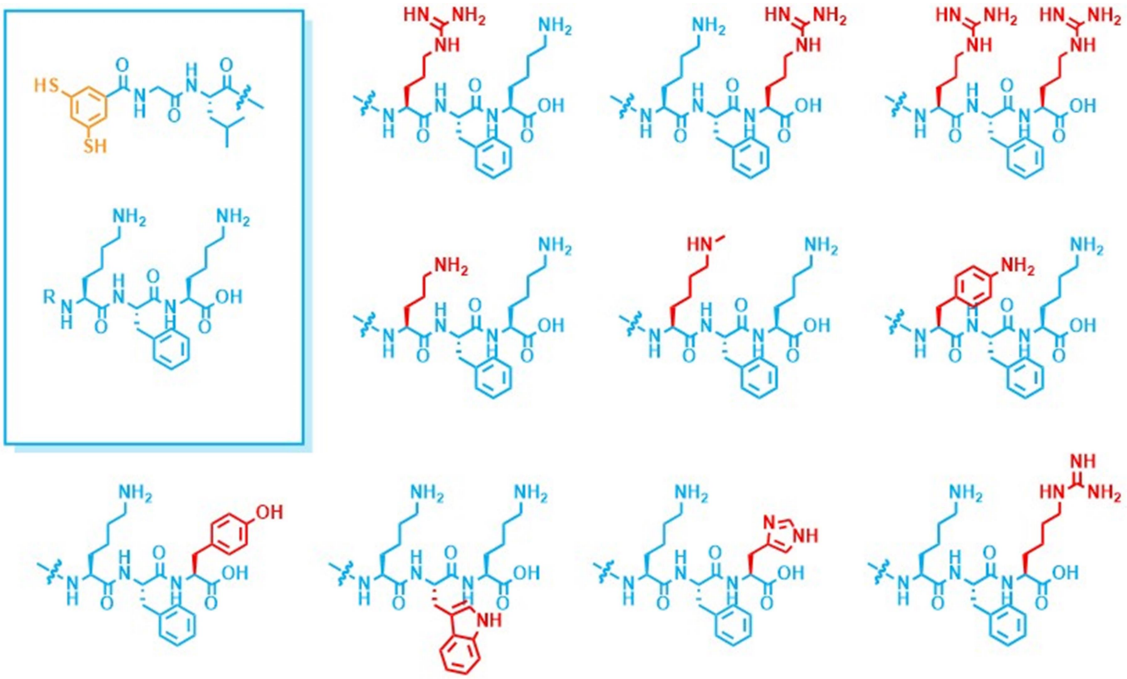


Figure 26: The eleven building block structures under study, the reference compound is shown in the square.

Furthermore, all building blocks were also left unstirred. According to the proposed mechanism of the fibre formation, it is the mechanical energy supplied by stirring that causes fibres to break and which gives rise to exponential replication. Without mechanical energy a fibre might still emerge, just not at an exponential rate. However, for most building blocks we investigated this mechanical energy is necessary to cause a fibre to emerge. We still found three building blocks were a fibre could also emerge without mechanical energy. Interestingly not all of these fibres consisted of the same macrocycles as their stirred counterparts. In future experiments these unstirred libraries need to be characterised further to see how much its properties differ from the stirred replicators.

An attempt was made to broaden the scope of reactions which can be catalysed by these assemblies. First the aldol reaction was investigated by a newly developed technique. Using a fluorescent plate reader, several carbonyl reactants could be investigated at the same time. Even though the thermodynamics of the reaction in the bond forming direction are unfavourable, we expected to find reactants where a small amount of product would be formed. No such reactants were found however, and our attention shifted to other reactions. Subsequently we investigated both the kemp-elimination reaction and a hydrazone formation reaction. Both of these were not significantly accelerated over the background reaction by the supramolecular assemblies. Ultimately no new reactions were found in which these building blocks are catalytically active.

The main focus of this thesis has been to explore the active sites that are present on the fibres created by these building blocks. The reactions for which it was known that they could be catalysed by these fibres are the retro-aldol and Fmoc and Sulfmoc deprotection reaction. As model reactions we chose the breakdown of methodol and Sulfmoc protected glycine, for the retro-aldol and FMOC deprotection respectively. The activities of the various fibres in these reactions were compared with a previously investigated building block.

The retro-aldol reaction formed our starting point. Although it was known that the lysines in the fibre were perturbed it was unknown which lysine was responsible for the catalysis. From the results of the arginine replicators, we gathered that it is the inner position where the lysine is catalytically active. When the lysine is placed on the outer position its rate is 1.5 times that of the background.

Furthermore we concluded that the changes in the sequence only impacted the activity to a minor extent, as long as the lysine was on the inner position. This is a striking observation as this is in direct contrast to the enzymes found in nature, where the sequence and therefore structure is very well defined. This demonstrates that the structure formed by a self-replicator is much more important than the specifics of the sequence. This implies that when life originated ill-defined supramolecular structures could be responsible for the creation of function in self-replicating chemical networks. This would result in an evolutionary advantage and could eventually lead to metabolism.

When our focus shifted to the Sulfmoc deprotection reaction, surprisingly the relative activity did not match the retro-aldol reaction. We had hypothesised that the active site would be the same in both reactions. However, all fibres performed remarkably worse than the standard building block, and their relative order seemed to correlate with the retro-aldol reaction in an inverse manner. This indicates that both reactions have a different active site, but with opposing characteristics.

Several fibres were created with building blocks mixed in a one to one ratio. Unfortunately these libraries were prone to forming gels, causing them to become inhomogeneous to the point where they could not be accurately pipetted. This resulted in less accurate data. Even so, it appears as though the activity was not simply the average of the individual counterparts. This means that the perturbation in the active site does not come from the residues within the same building block, but from a building block further removed, *i.e.* on the same macrocycle or the next attached macrocycle. The results from the Sulfmoc reaction showed a similar trend, and further confirmed that a different active site is responsible for this reaction.

This thesis demonstrates that metabolism still proves to be a challenging trait to develop in self-replicating systems. There are a myriad of parameters that influence the active sites, which makes it difficult to predict how well a self-replicator is capable of catalysing certain chemical transformations. Moreover there are multiple different active sites on these fibres that can catalyse different reactions. Further efforts towards this goal should focus on further shedding light on these various parameters.

Since it is suspected that the active site is located on the disordered places in the fibre, a method needs to be found to directly influence the formation of these places. If the interactions that hold the fibre together can be weakened, it might introduce more of these disordered sites. Using denaturing agents for example might weaken the β-sheets and could introduce more disorder during the fibre formation.

Building blocks can also be designed to further disrupt the system. A method to achieve this might be to increase the bulk of the residues in the first part of the sequence, *e.g.* by exchanging the leucine for isoleucine. This will also give further insights into the parameters that govern the self-replicating process, similar to the building blocks in this thesis.

Other efforts could focus on probing the active site more directly. Dyes that can bind the fibre trough the active amine could provide a direct method to image where these active sites are. This opens the door to using super resolution imaging which showed a large potential in the field of molecular biology.[57] In a similar fashion, heavier atoms that also bind the active amine can be imaged by electron microscopy. For this the halogen atoms could be used and can also provide an image of the active site.

When more parameters of the active sites are known attention should be refocused on the chemical transformations itself. With this knowledge other reactions should be easier to find. The goal of this project should be to allow the fibres to create molecules which are useful for the assembly. This brings metabolism a step closer, and forms the next step towards de-novo life.

Although this self-replicating system still has a long way to go before it is capable of metabolism, here we have investigated its first catalytic properties; a small step on the road to de-novo life. Even though this road might seem very long, it wouldn't be de-novo if we did not start from nothing.

Acknowledgements

First of all I would like to thank professor Otto for his supervision of this research and allowing me in his group. With the many opportunities available I have developed myself in many new areas. Thanks for all the things I could learn in the group.

Secondly I would like to thank Jim Ottelé for supervising my project. I really appreciated the many ideas that we discussed during the drinking of endless amounts of coffee really. You created a very pleasant environment to work in. Thanks to Jelmer Visser for being an excellent student and for all the work that we did on the fluorescent plate reader together. I would also like to thank Clemens Mayer for his help with the fluorescent plate reader and the hydrazone formation reaction.

I would like to thank Charalampos Pappas for introducing me into the world of peptide synthesis. Over our many conversations you taught me a lot about the world of academia. I also thank Bin Liu for the UPLC-MS measurements, and being a great help if I didn't understand how it is operated. Thank you for teaching me how the flash column is used. Thanks also to Ivanna Marić for reminding Jim and me how to do synthesis.

Lastly, I would like to thank the rest of the group. The pleasant environment meant I have had an amazing time here.

Bibliography

1. Scharf, C. *et al.* A Strategy for Origins of Life Research. *Astrobiology* **15,** 1031–42 (2015).
2. Machery, E. Why I stopped worrying about the definition of life... and why you should as well. *Synthese* **185,** 145–164 (2012).
3. Gánti, T. *The principles of life*. (Oxford University Press, 2003).
4. Bains, W. What do we think life is? A simple illustration and its consequences. *Int. J. Astrobiol.* **13,** 101–111 (2014).
5. Anet, F. AL. The place of metabolism in the origin of life. *Curr. Opin. Chem. Biol.* **8,** 654–659 (2004).
6. Leslie E., O. Prebiotic Chemistry and the Origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* **39,** 99–123 (2004).
7. Kauffman, S. A. *The origins of order : self-organization and selection in evolution*. (Oxford University Press, 1993).
8. The Nobel Prize in Chemistry 1989 - NobelPrize.org. Available at: https://www.nobelprize.org/prizes/chemistry/1989/summary/. (Accessed: 16th October 2019)
9. Gilbert, W. Origin of life: The RNA world. *Nature* **319,** 618–618 (1986).
10. Kruger, K. *et al.* Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31,** 147–57 (1982).
11. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418,** 214–221 (2002).
12. Lal, A. K. Origin of Life. *Astrophys Sp. Sci* **317,** 267–278 (2008).
13. Robertson, M. P. & Joyce, G. F. The Origins of the RNA World. *Cold Spring Harb. Perspect. Biol.* **4,** a003608–a003608 (2012).
14. Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science (80-.).* **117,** 528–529 (1953).
15. Bada, J. L. New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. *Chem. Soc. Rev.* **42,** 2186 (2013).
16. Saladino, R., Di Mauro, E. & García-Ruiz, J. M. A Universal Geochemical Scenario for Formamide Condensation and Prebiotic Chemistry. *Chem. - A Eur. J.* **25,** 3181–3189 (2019).
17. McCollom, T. M. Miller-Urey and Beyond: What Have We Learned About Prebiotic Organic Synthesis Reactions in the Past 60 Years? *Annu. Rev. Earth Planet. Sci.* **41,** 207–229 (2013).
18. Neveu, M., Kim, H.-J. & Benner, S. A. The 'Strong' RNA World Hypothesis: Fifty Years Old. *Astrobiology* **13,** 391–403 (2013).
19. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459,** 239–242 (2009).
20. Bernhardt, H. S. The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)(a). *Biol. Direct* **7,** 23 (2012).
21. Shapiro, R. A Replicator Was Not Involved in the Origin of Life. *IUBMB Life* **49,** 173–176 (2000).

22. Vasas, V., Fernando, C., Santos, M., Kauffman, S. & Szathmáry, E. Evolution before genes. *Biol. Direct* **7,** 1 (2012).

23. 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.* **107,** 1470–5 (2010).

24. Orgel, L. E. The Implausibility of Metabolic Cycles on the Prebiotic Earth. *PLoS Biol.* **6,** e18 (2008).

25. Duim, H. & Otto, S. Towards open-ended evolution in self-replicating molecular systems. *Beilstein J. Org. Chem.* **13,** 1189–1203 (2017).

26. Pascal, R., Pross, A. & Sutherland, J. D. Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics. *Open Biol.* **3,** 130156 (2013).

27. Danchin, A. From chemical metabolism to life: the origin of the genetic coding process. *Beilstein J. Org. Chem.* **13,** 1119–1135 (2017).

28. Symons, R. H. & Uhlenbeck, O. C. Ribozymes. *CRC. Crit. Rev. Plant Sci.* **10,** 189–234 (1991).

29. Tawfik, O. K. and D. S. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annu. Rev. Biochem.* **79,** 471–505 (2010).

30. Jacobsen, E. N. & Finney, N. S. Synthetic and biological catalysts in chemical synthesis: how to assess practical utility. *Chem. Biol.* **1,** 85–90 (1994).

31. Raynal, M., Ballester, P., Vidal-Ferran, A. & Leeuwen, P. W. N. M. van. Supramolecular catalysis. Part 2: artificial enzyme mimics. *Chem. Soc. Rev.* **43,** 1734–1787 (2014).

32. Cook, S. A. & Borovik, A. S. Molecular Designs for Controlling the Local Environments around Metal Ions. *Acc. Chem. Res.* **48,** 2407–2414 (2015).

33. Raynal, M., Ballester, P., Vidal-Ferran, A. & van Leeuwen, P. W. N. M. Supramolecular catalysis. Part 1: non-covalent interactions as a tool for building and modifying homogeneous catalysts. *Chem. Soc. Rev.* **43,** 1660–1733 (2014).

34. Vidonne, A. & Philp, D. Making Molecules Make Themselves - the Chemistry of Artificial Replicators. *European J. Org. Chem.* **2009,** 593–610 (2009).

35. Carnall, J. M. A. *et al.* Mechanosensitive Self-Replication Driven by Self-Organization. *Science (80-.).* **327,** 1502–1506 (2010).

36. Otto, S., Furlan, R. L. E. & Sanders, J. K. M. Dynamic combinatorial libraries of macrocyclic disulfides in water [35]. *Journal of the American Chemical Society* **122,** 12063–12064 (2000).

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

38. Ottelé, J., Mayer, C. & Otto, S. Chance Emergence of Catalysis by a Self-Replicator. *Man. in Prep.* (2019).

39. J Ottelé. No Title. *Man. in Prep.*

40. Jiang, L. *et al.* De Novo Computational Design of Retro-Aldol Enzymes. *Science (80-.).* **319,** 1387–1391 (2008).

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

42. 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.* **139,** 13612–13615 (2017).

43. Fitch, C. A., Platzer, G., Okon, M., Garcia-Moreno, B. E. & McIntosh, L. P. Arginine: Its pKa value revisited. *Protein Sci.* **24,** 752–761 (2015).

44. Frederix, P. W. J. M. *et al.* Structural and Spectroscopic Properties of Assemblies of Self-Replicating Peptide Macrocycles. *ACS Nano* **11,** 7858–7868 (2017).

45. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics* **1751,** 119–139 (2005).

46. Hudson, S. A., Ecroyd, H., Kee, T. W. & Carver, J. A. The thioflavin T fluorescence assay for amyloid fibril detection can be biased by the presence of exogenous compounds. *FEBS J.* **276,** 5960–5972 (2009).

47. J M Frederix, P. W. *et al.* Structural and Spectroscopic Properties of Assemblies of Self-Replicating Peptide Macrocycles. *ACS Nano* **11,** 33 (2017).

48. Yu, X. & Wang, W. Organocatalysis: Asymmetric cascade reactions catalysed by chiral secondary amines. *Org. Biomol. Chem.* **6,** 2037–2046 (2008).

49. Bertelsen, S. & Jørgensen, K. A. Organocatalysis - After the gold rush. *Chemical Society Reviews* **38,** 2178–2189 (2009).

50. Drienovská, I., Mayer, C., Dulson, C. & Roelfes, G. A designer enzyme for hydrazone and oxime formation featuring an unnatural catalytic aniline residue. *Nat. Chem.* **10,** 946–952 (2018).

51. Tsikas, D. & Wu, G. Homoarginine, arginine, and relatives: analysis, metabolism, transport, physiology, and pathology. *Amino Acids* **47,** 1697–1702 (2015).

52. Wimley, W. C. & White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Mol. Biol.* **3,** 842–848 (1996).

53. Boger, D. L., Fink, B. E., Brunette, S. R., Tse, W. C. & Hedrick, M. P. A Simple, High-Resolution Method for Establishing DNA Binding Affinity and Sequence Selectivity. *J. Am. Chem. Soc.* **123,** 5878–5891 (2001).

54. Kemp, D. S. & Casey, M. L. Physical organic chemistry of benzisoxazoles. II. Linearity of the Broensted free energy relation for the base-catalyzed decomposition of benzisoxazoles. *J. Am. Chem. Soc.* **95,** 6670–6680 (1973).

55. Casey, M. L., Kemp, D. S., Paul, K. G. & Cox, D. D. Physical organic chemistry of benzisoxazoles. I. Mechanism of the base-catalyzed decomposition of benzisoxazoles. *J. Org. Chem.* **38,** 2294–2301 (1973).

56. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453,** 190–195 (2008).

57. Fernández-Suárez, M. & Ting, A. Y. Fluorescent probes for super-resolution imaging in living cells. *Nat. Rev. Mol. Cell Biol.* **9,** 929–943 (2008).

Supporting information

General procedures

The chemicals used were generally purchased from Sigma-Aldrich and used as received. Unless stated building block peptides were obtained from GenScript. XGLKFK was obtained from Cambridge Peptides. Acetonitrile (ULC-MS grade) and Water (ULC-MS grade) were purchased from Biosolve BV.

Peptide synthesis

XGLKFH and XGLRFR were prepared by solid state peptide synthesis. A syringe with filter was loaded with the appropriate FMOC protected amino acids on wang resin, to get 0.1 mmol active carboxylate sites. The resin was swelled with DMF, after which the amino acid was washed three times with 25% piperidine in DMF. Subsequent amino acids were coupled by adding 1M Oxyma in DMF (0.3 ml, 0.3 mmol, 3 eq) and 0.5M (0.6 ml, 0.6 mmol, 6 eq) *N*,*N*'-diisopropylcarbodiimide in DMF to the FMOC protected amino acids (0.3 mmol, 3 eq). This coupling cocktail was added to the resin after rinsing with DMF and was left up to 2 hours. This procedure was repeated until the sequence was synthesised. Peptides were deprotected with a cocktail of trifluoroacetic acid (TFA) (2850 µL), doubly distilled H_2O (60 µL), 1,2-ethanedithiol (60 µL) and triisopropylsilane (30 µL) for up to 4 hours. After filtration the supernatant was added to diethylether (20ml) from which it precipitates. The resulting suspension was centrifuged and washed with diethyl ether several times. The ether was decanted and the pellet was redissolved in doubly distilled H_2O, and subsequently freeze dried. The impure peptides were purified with flash chromatography on a reverse phased C18 Silica column with acetonitrile and water with 0.1% TFA.

Library preparation

All libraries were made at a concentration of 2mM with respect to monomer building block in borate buffer (pH 8.1, 200mM in boron atoms, or 50mM in buffering ions). Libraries were kept at 45°C and stirred at 1200rpm unless otherwise stated. Before analysis, care was taken to homogenise the libraries by vigorous pipetting after which the desired amount was taken. After libraries were completely oxidised for several days, they were kept at room temperature.

For measurements that required monomers or oxidised monomers; we prepared the samples fresh before measuring. Oxidation was done with a 20mM sodium perborate tetrahydrate solution in buffer (either to 80% or to 100%).

Fluorescence plate reader measurements

FPR experiments were performed on a Biotek Synergy HTX plate reader with 96 well plates (8 x 12). Wells were filled up to 200 µL with reactants in various concentrations. For this a multichannel pipet and corresponding reservoir was utilised. After addition of the final reactant the plate was immediately inserted in the machine, and reactions were followed for several hours at constant temperature (25°C) while shaking the plate. The incident light was set to 330nm to excite 6-methoxy-2-naphthaldehyde and emitted light was measured at 452 nm.

UPLC measurements

All UPLC analyses were performed on a Waters Acquity UPLC I-class equipped with a PDA UV/Vis detector and reversed-phase Aeris Peptide 1.7 µm column (XB-C18 130 x 2.10 mm, purchased from Phenomenex). Libraries were diluted 10 times with UPLC grade H2O and 10 µL was injected. Analysis was performed with ULC-MS grade water (eluent A) and ULC-MS grade acetonitrile (eluent B), both containing 0.1 V/V % TFA. UV absorbance was monitored at 254 nm, 313 nm and 330 nm, and column temperature was kept at 35°C, with a flow rate of 0.3 mL/min. The method was as follows:

Time (minutes)	%A	%B
Initial	90	10
1.00	90	10
1.30	75	25
3.00	72	28
11.00	60	40
11.50	5	95
12.00	5	95
12.50	90	10
17.00	90	10

Mass spectrometry

All LC-MS analyses were performed on a Waters Acquity UPLC H-class coupled to a Waters Xevo G2 TOF with a diode array UV/Vis set to 245 nm. The UPLC operated under the same conditions as described above. All analyses were done in positive electrospray ionization mode. Samples were diluted 10 times and 5 µL was injected.

UV measurements

All UV measurements were performed in a Jasco V-650 UV-Vis spectrometer with temperature control set to 25°C. Samples were prepared in quartz cuvettes cleaned with nitric acid and gently shaken. After the UV active reactant was added the samples were followed for 10 minutes to ensure their homogeneity. After addition of the catalyst fibres the samples were shaken and immediately followed for up to 15 hours.

ThT measurements

A 2.2 mM thioflavin T stock solution in borate buffer (pH 8.1) was prepared and stored in the dark until used in the measurements. The stock solution was diluted a 100 times to generate a working solution of 22 µM. 900 µL was added to a plastic cuvette, and the fluorescence intensity was measured on a JASCO FP6200 Fluorimeter. An excitation wavelength of 440 nm was used with a 480 nm cut-off filter; the readout was set to 482 nm. After this blank 176 µL buffer and 4 µL 2mM library solution were added, and after 2 minutes the fluorescence was measured again with 5 accumulations.

CD measurements

CD spectra of the peptides were obtained on a JASCO 715 CD instrument. Samples were prepared in a 300 µL quartz cuvette with a path length of 1 mm. Concentration of samples always was 0.33 mM (in building block). Samples were diluted with UPLC grade water. Spectra were taken from 300nm to 190nm although for monomer samples this lead to high voltages below 200nm so these results are less reliable.

TEM measurements

A small drop (5 µL) of sample was deposited on a 400 mesh copper grid covered with a thin carbon film (Agar Scientific). After 30 s, the droplet was blotted on filter paper. The sample was then stained twice (4 µL each time) with a solution of 2% uranyl acetate deposited on the grid and blotted on the filter paper after 30 s each time. The grids were observed in a Philips CM120 cryo-electron microscope operating at 120 kV. Images were recorded on a slow scan CCD camera. Using image processing software (ImageJ) the width of fibres was measured using the measuring tool. Each micrograph was scaled using the magnification parameters of the figure. For every building block, 25 non overlapping fibres were measured across multiple pictures.

Kemp elimination

The kemp elimination reaction was followed in a UV spectrometer at 380 nm. A solution of 10 mM 5-nitro-1,2-benzisoxazole in acetonitrile was prepared. 10 µL of this stock solution was added to 977.5 µL borax buffer (pH 7.4) and 12.5 µL 2 mM replicator building block was added. This created samples of 100 µM 5-nitro-1,2-benzisoxazole and 25 µM of the XGLKFK building block. The reaction was followed with monomers, pre-formed fibres and a background with H_2O.

Hydrazone formation

Hydrazone formation was followed at 472 nm in a sample containing 5 mM 4-HBA, 200 µM NBD-H and 25 µM XGLaFFK building block. The reaction was carried out in phosphate buffer (50 mM Na_2HPO_4, 150 mM NaCl, pH 7.4) and contained 5% DMF. The reaction was followed with monomers, oxidised monomers, pre-formed fibres and a background with H_2O.

Methodol kinetics

For the retro-aldol reaction of methodol a 30mM methodol solution in acetonitrile was prepared. This stock solution can be kept in the fridge for up to a week. A working solution was then prepared of 2668 µM methodol with 13.5% V/V acetonitrile in buffer by diluting 20 µL 30 mM methodol in 173 µL buffer and adding 7 µL acetonitrile. In an UPLC vial 12.5 µL replicator solution was diluted in 837 µL borate buffer, and at exactly 16 minutes into the previous injection 150 µL of the working solution was added. This creates a reaction mixture that is 400 µM in methodol, 25 µM in replicator building blocks in buffer with 2% V/V acetonitrile at pH 8.1. The sample was analysed by UPLC starting from exactly 2 minutes after addition of the catalyst. The UV trace was subsequently followed at 313 nm. Figure S1 shows the breakdown of methodol by monomers and oxidised monomers of all the building blocks. Due to time constraints these were not followed by UPLC but by UV directly in a UV-Vis spectrometer.

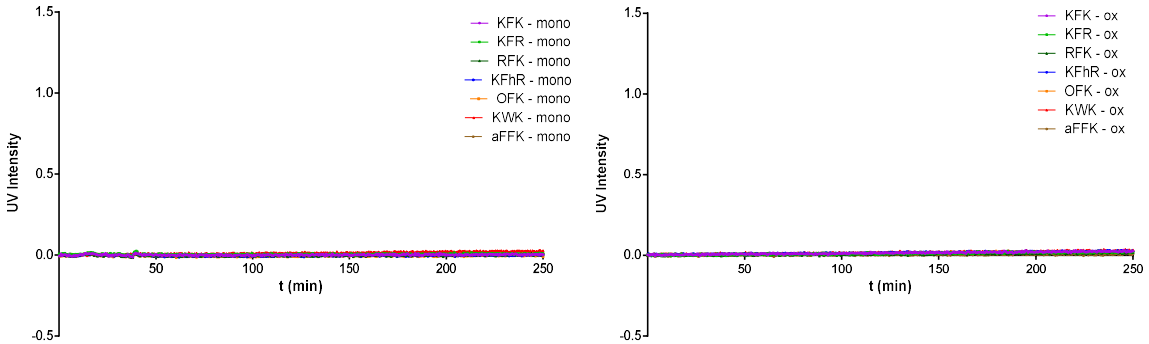
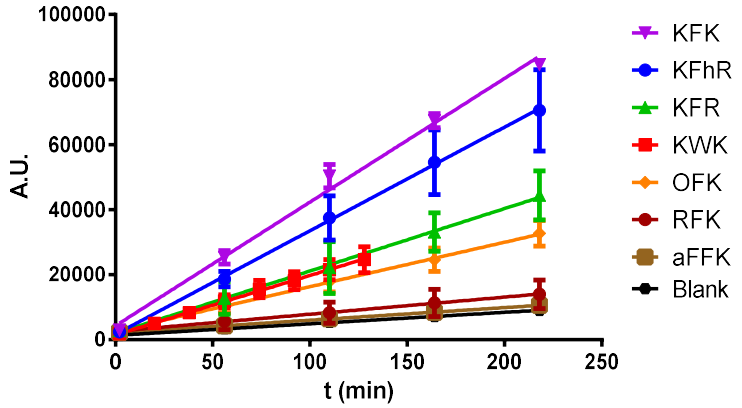
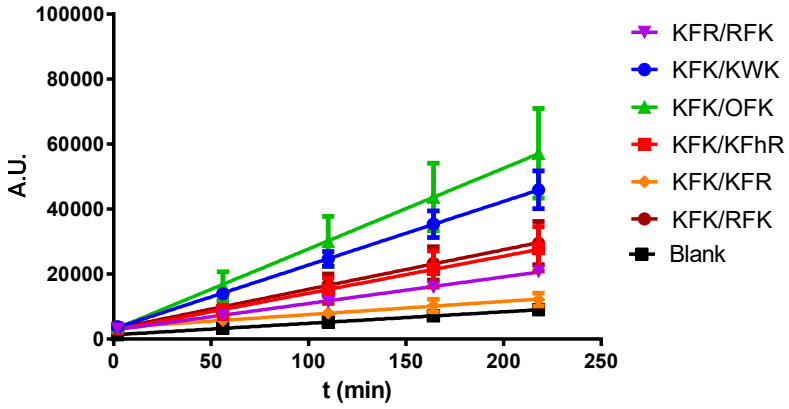


Figure S1: Breakdown of methodol by the monomers (left) and oxidised monomers (right) of the various building blocks. The above passage describes the protocol.



	KFhR	KWK	KFR	KFK	OFK	RFK	aFFK	Blank
Slope	318.8 ± 23.43	182.6 ± 10.15	192.0 ± 17.99	380.4 ± 10.99	136.7 ± 8.683	52.57 ± 9.987	38.68 ± 0.9074	35.39 ± 2.456
Y-intercept when X=0.0	1646 ± 3137	1462 ± 781.2	1906 ± 2410	4263 ± 1471	2610 ± 1163	2558 ± 1337	2109 ± 121.5	1310 ± 328.9
X-intercept when Y=0.0	-5.165	-8.007	-9.926	-11.21	-19.09	-48.65	-54.52	-37.02
R square	0.9344	0.9364	0.8975	0.9893	0.9502	0.6807	0.9929	0.9411

Figure S2: Full experiment of the breakdown of methodol by all the fibres. The same graphs are seen in figure 17. The data of the fit is summarised in the table.



	KFR/RFK	KFK/KWK	KFK/OFK	KFK/KFhR	KFK/KFR	KFK/RFK	Blank
Slope	81.78 ± 2.958	195.6 ± 10.96	248.3 ± 28.41	114.2 ± 13.31	40.37 ± 5.005	121.8 ± 12.38	35.39 ± 2.456
Y-intercept when X=0.0	2775 ± 396.2	3221 ± 1468	2917 ± 3804	2700 ± 1782	3508 ± 670.2	3144 ± 1658	1310 ± 328.9
X-intercept when Y=0.0	-33.93	-16.47	-11.75	-23.65	-86.9	-25.82	-37.02
R square	0.9896	0.9755	0.9052	0.8499	0.8335	0.8815	0.9411

Figure S3: Full experiment of the breakdown of methodol by all the mixed fibres. The same graphs are seen in figure 22. The data of the fit is summarised in the table.

Sulfmoc kinetics

A stock solution of 10 mM Sulfmoc-Gly in buffer was prepared. This stock solution can be kept for several days in the fridge. Samples were prepared by diluting 10 µL 10 mM stock solution in 970 µL borate buffer of pH. After stabilising 20 µL 2 mM building block solution was added, generating a reaction mixture of 0.1 mM in Sulfmoc-Gly and 40 µM in building blocks and pH 8.1. The reaction was subsequently monitored using a UV spectrometer at 308 nm. Figure S2 shows the deprotection of Sulfmoc glycine by monomers and oxidised monomers.

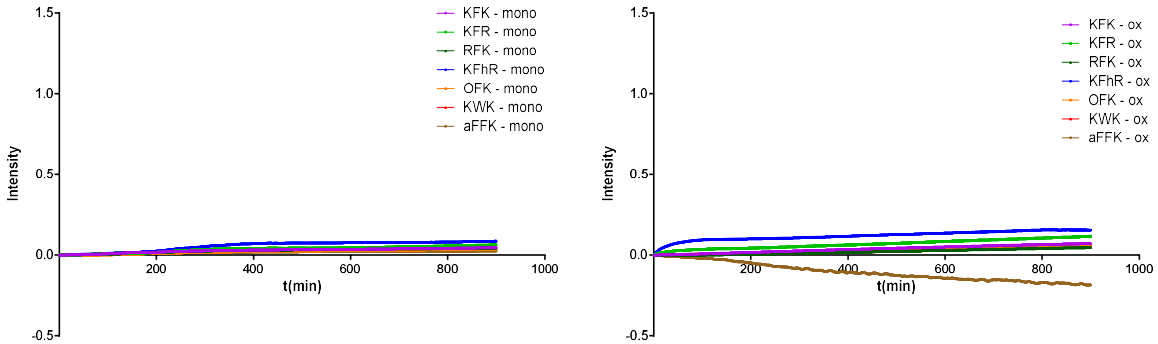


Figure S4: Breakdown of Sulfmoc glycine by the monomers (left) and oxidised monomers (right) of the various building blocks. The above passage describes the protocol.

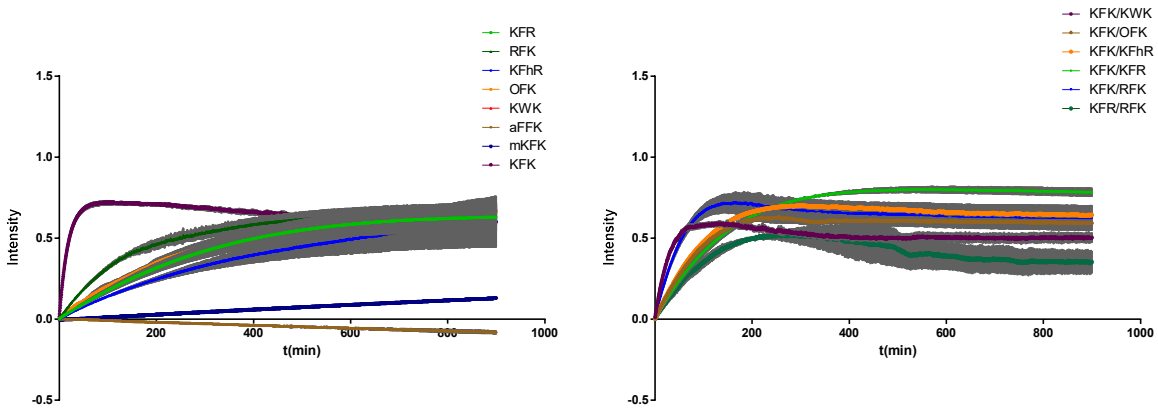
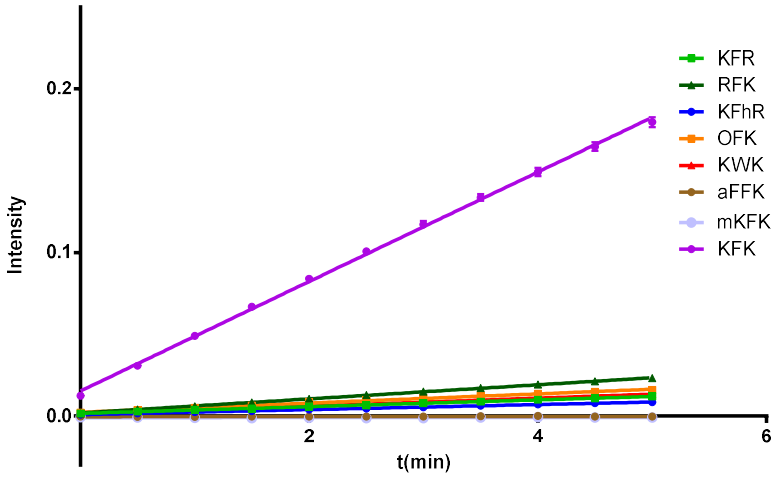
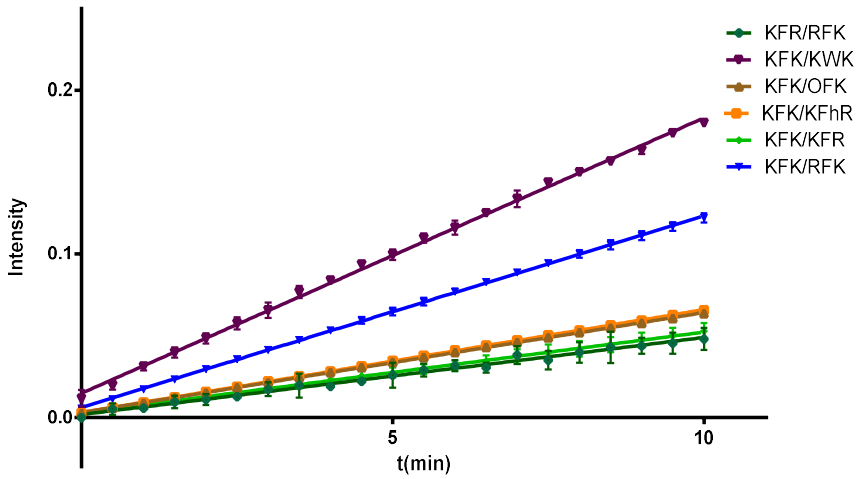


Figure S5: Full experiment of the deprotection of Sulfmoc glycine by all the fibres (left) and all mixed libraries (right). The same graphs are seen with a shorter time axis in figure 20 (left) and figure 24 (right)



	KFhR	KWK	KFR	KFK	OFK	RFK	aFFK	mKFK
Slope	0.00154 ± 2.55e-005	0.00246 ± 3.08e-005	0.00207 ± 6.06e-005	0.0335 ± 0.000262	0.00282 ± 8.61e-005	0.00431 ± 2.86e-005	5.17e-005 ± 4.12e-005	1.35e-005 ± 0.000106
Y-intercept when X=0.0	0.000908 ± 7.54e-005	0.00115 ± 9.10e-005	0.00164 ± 0.000179	0.0153 ± 0.000775	0.00230 ± 0.000255	0.00191 ± 8.45e-005	-0.000383 ± 0.000122	-0.000974 ± 0.000312
X-intercept when Y=0.0	-0.59	-0.469	-0.7948	-0.4576	-0.8157	-0.4426	7.408	71.98
R square	0.9916	0.9952	0.974	0.9981	0.9719	0.9986	0.04847	0.0005304

Figure S6: Initial minutes of the deprotection of Sulfmoc glycine by all the fibres. The data of the fit is summarised in the table.

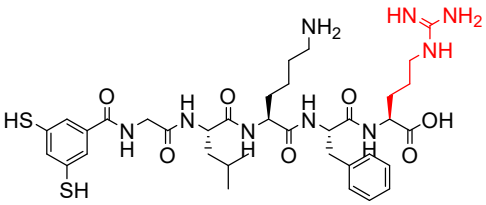


	KFR/RFK	KFK/KWK	KFK/OFK	KFK/KFhR	KFK/KFR	KFK/RFK
Slope	0.004719 ± 0.0001817	0.01684 ± 0.0001324	0.006107 ± 4.138e-005	0.006288 ± 2.194e-005	0.004935 ± 0.0001179	0.01172 ± 7.024e-005
Y-intercept when X=0.0	0.001790 ± 0.001062	0.01483 ± 0.0007738	0.003081 ± 0.0002419	0.003120 ± 0.0001282	0.002904 ± 0.0006894	0.006127 ± 0.0004106
X-intercept when Y=0.0	-0.3792	-0.8806	-0.5044	-0.4963	-0.5884	-0.5229
R square	0.9171	0.9962	0.9972	0.9993	0.9663	0.9978

Figure S7: Initial minutes of the deprotection of Sulfmoc glycine by all mixed fibres. The data of the fit is summarised in the table.

Characterisation of compounds

XGLKFR "**1**"



UPLC traces of 1

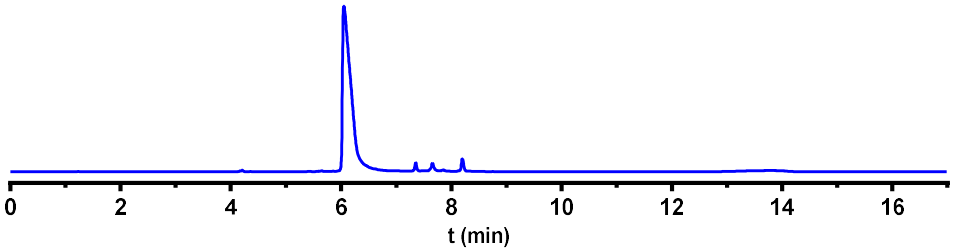


*Figure S8: UPLC chromatogram of **1** after library creation*

Peak	Retention time	Percentage
1 Monomer	6.05	94.69
2 Tetramer	7.35	1.05
3 Linear dimer	7.66	1.68
4 Trimer	8.20	2.05

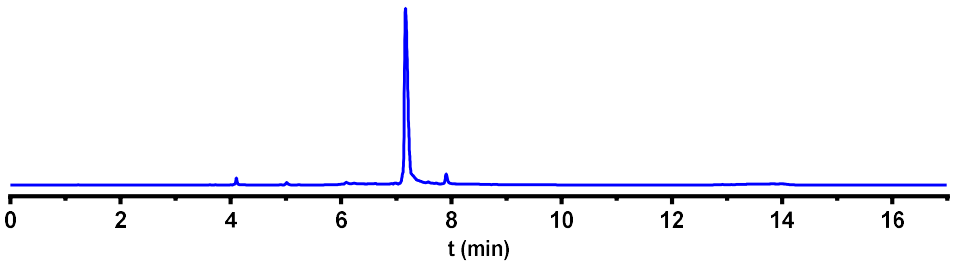


*Figure S9: UPLC chromatogram of **1** after 3 days*

Peak	Retention time (min)	Percentage of total
1 Over oxidised species	4.10	1.23
2 Monomer	6.10	1.28
3 Hexamer	7.17	87.22
4 Linear trimer	7.58	1.33
5 Trimer	7.91	4.25

Mass Spectra of 1

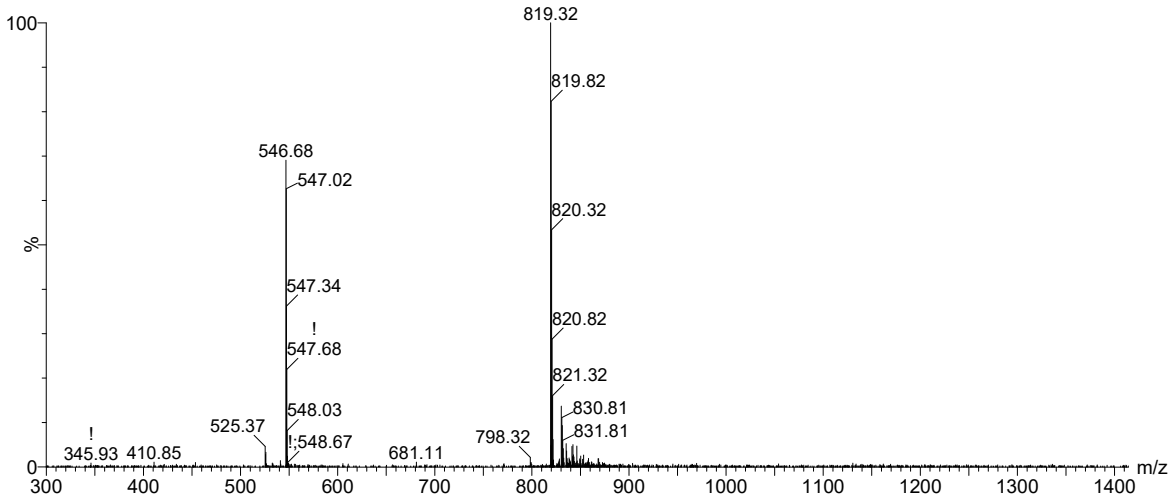


*Figure S10: Mass spectrum of over oxidised (**1**)$_2$ (retention time 4.10) from the LC-MS analysis of the building block after 5 days*
m/z observed: 819.32 $[M+4O+2H]^{2+}$; 546.68 $[M+4O+3H]^{3+}$
m/z calculated: 819.34 $[M+4O+2H]^{2+}$; 546.56 $[M+4O+3H]^{3+}$

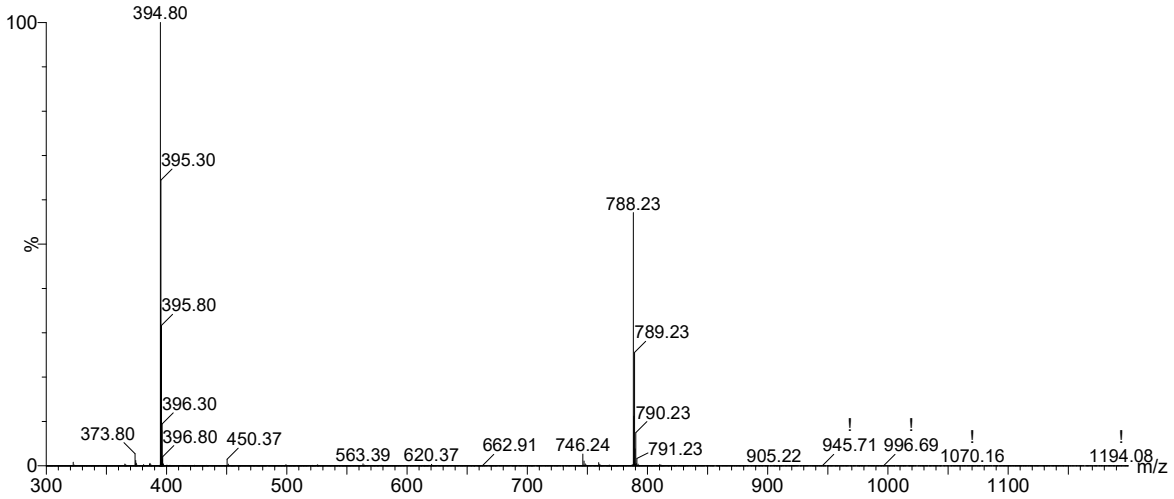


*Figure S11: Mass spectrum of (**1**)$_1$ (retention time 6.1) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 788.23 $[M+1H]^{1+}$; 394.80 $[M+2H]^{2+}$
m/z calculated: 788.36 $[M+1H]^{1+}$; 394.68 $[M+2H]^{2+}$

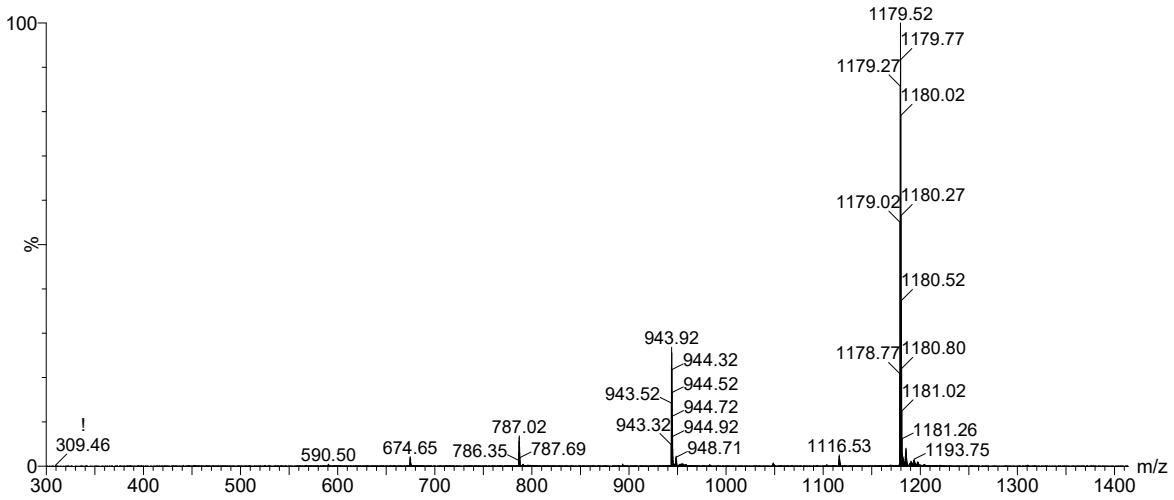


*Figure S12 Mass spectrum of (**1**)$_6$ (retention time 7.2) from the LC-MS analysis of the building block after 5 days.*
m/z observed: 1179.52 $[M+4H]^{4+}$; 943.92 $[M+5H]^{5+}$
m/z calculated: 1179.51 $[M+4H]^{4+}$; 943.81 $[M+5H]^{5+}$

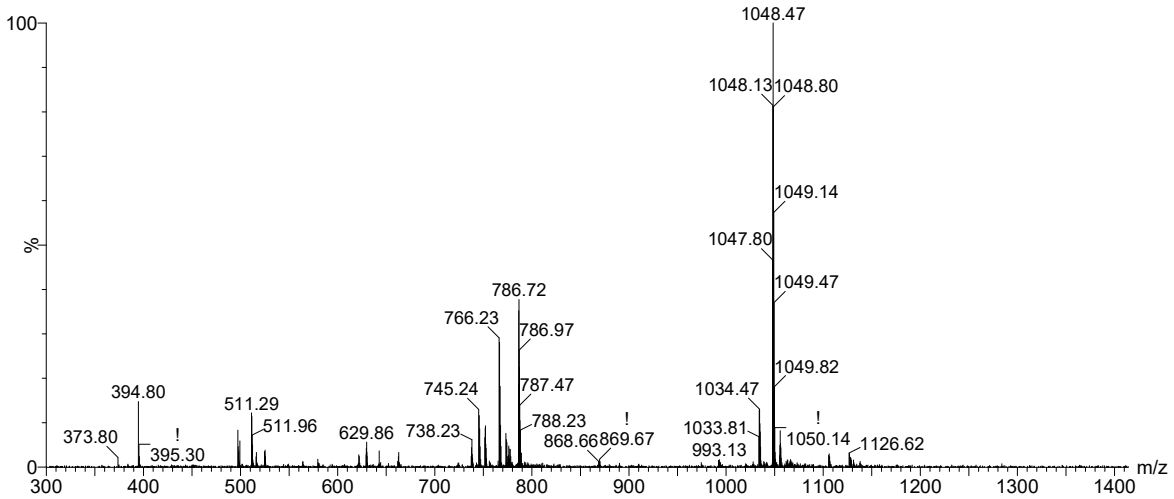


*Figure S13 Mass spectrum of (**1**)$_4$ (retention time 7.4) from the LC-MS analysis of the building block after library preparation. NB: The mass sample was accidently contaminated with a small amount of another building block.*
m/z observed: 1048.47 $[M+3H]^{3+}$; 786.72 $[M+4H]^{4+}$
m/z calculated: 1048.46 $[M+3H]^{3+}$; 786.59 $[M+4H]^{4+}$

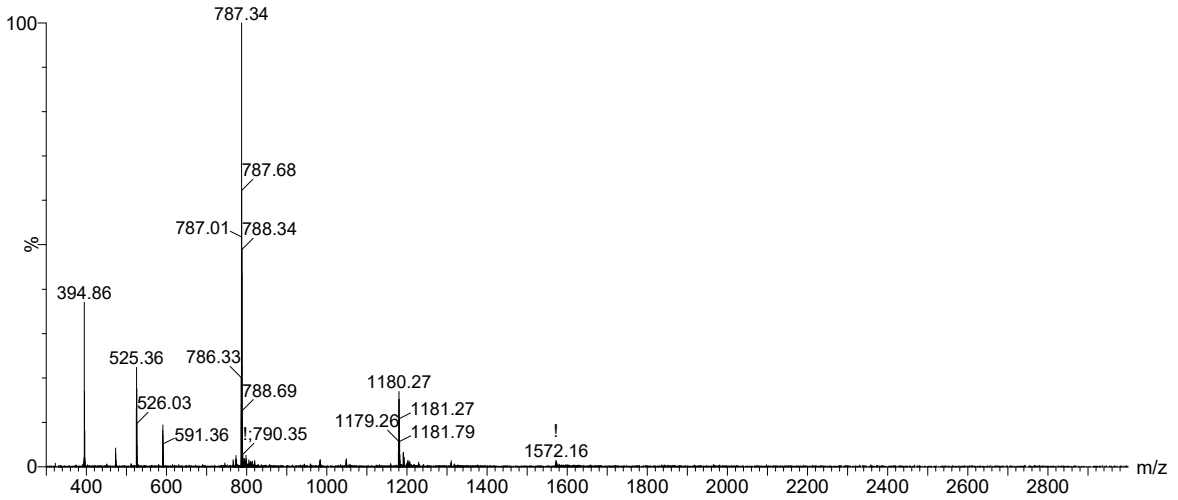


*Figure S14 Mass spectrum of linear (**1**)$_3$ (retention time 7.6) from the LC-MS analysis of the building block after 1 days. NB. Other macrocycles still visible from background*
m/z observed: 1180.27 $[M+2H]^{3+}$; 787.34 $[M+3H]^{3+}$
m/z calculated: 1180.52 $[M+2H]^{2+}$; 787.35 $[M+3H]^{3+}$

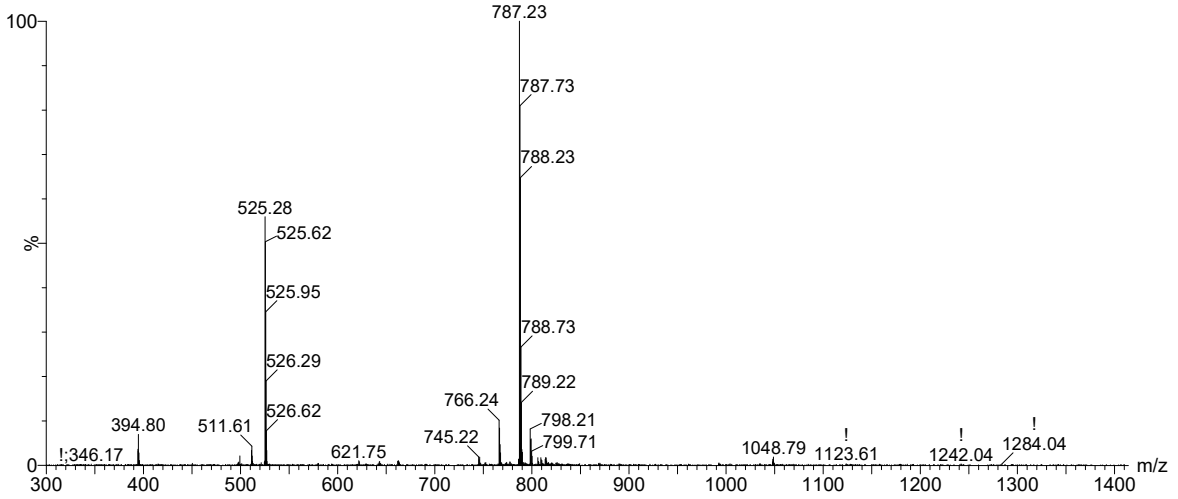


*Figure S15: Mass spectrum of (**1**)$_2$ (retention time 7.7) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 787.23 $[M+2H]^{2+}$; 525.28 $[M+3H]^{3+}$
m/z calculated: 787.35 $[M+2H]^{2+}$; 525.24 $[M+3H]^{3+}$

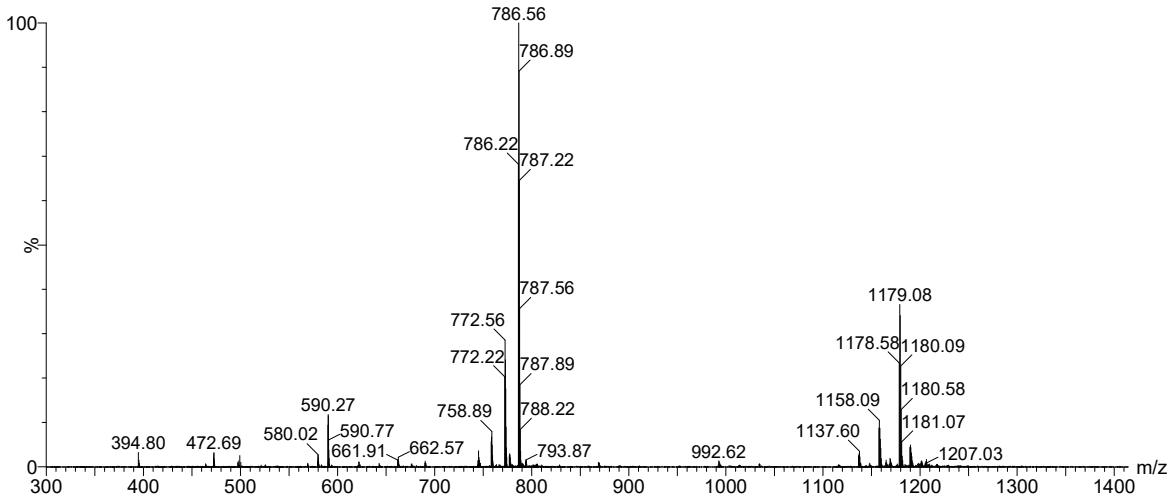


*Figure S16: Mass spectrum of (**1**)$_3$ (retention time 8.2) from the LC-MS analysis of the building block after library preparation. NB: The mass sample was accidently contaminated with a small amount of another building block.*
m/z observed: 1179.08 $[M+2H]^{2+}$; 786.72 $[M+3H]^{3+}$
m/z calculated: 1179.51 $[M+2H]^{2+}$; 786.68 $[M+3H]^{3+}$

Kinetic plots

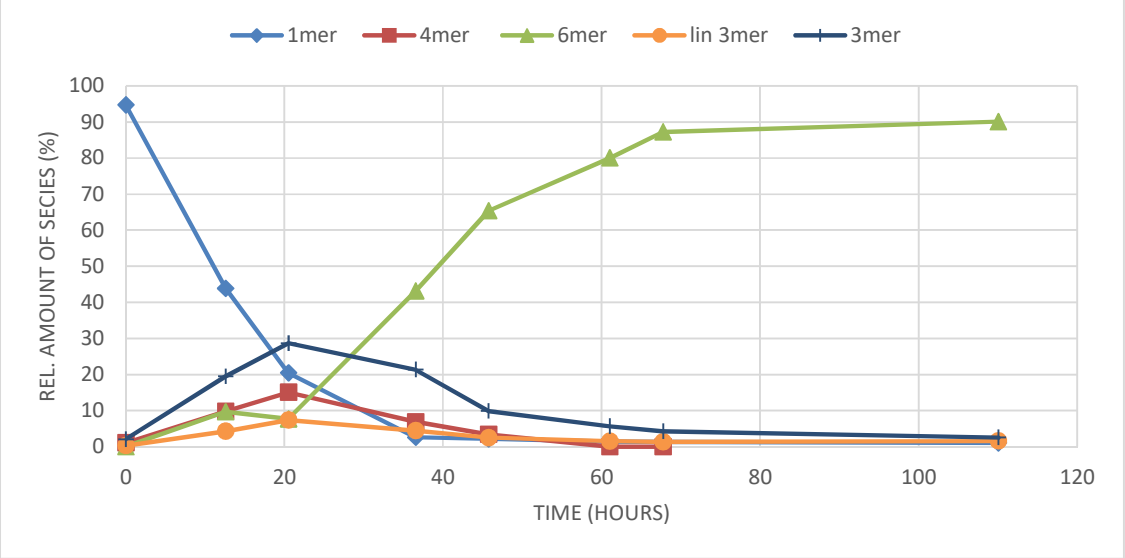


*Figure S17: Kinetic profile of a library made from **1** in borate buffer (pH 8.1) under continuous stirring at 45°C*

Seeding

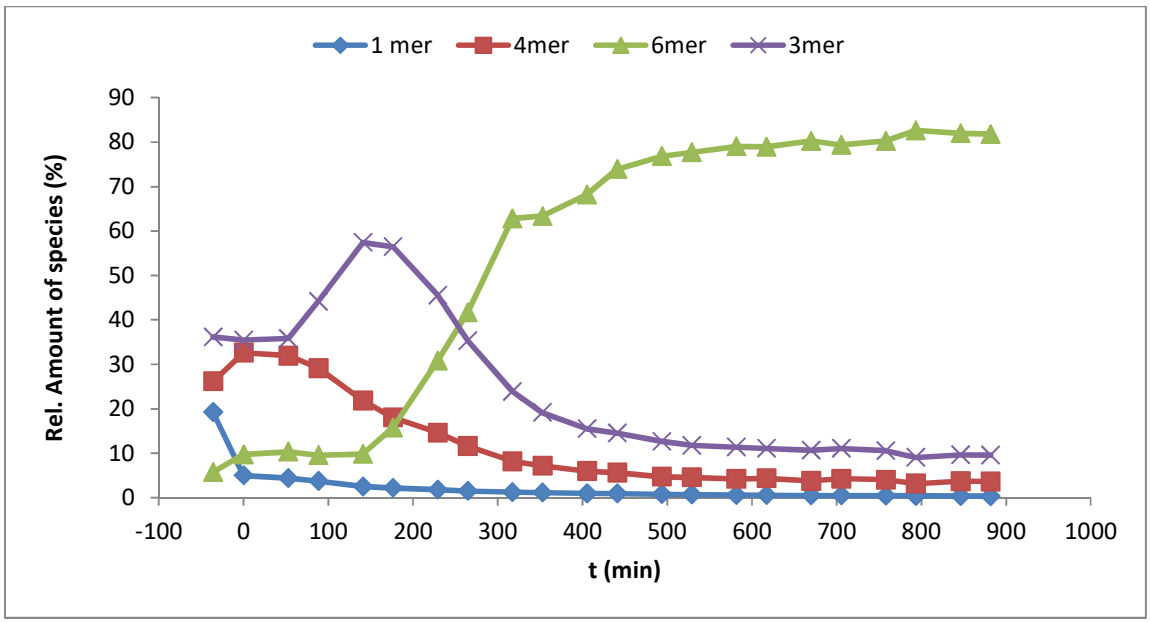


*Figure S18: Kinetic profile of a library made from an 80% oxidised library of **1** in borate buffer (pH 8.1) under continuous stirring at 40°C, upon addition of a seed of fibres of **1** at t = 0*

CD and ThT data

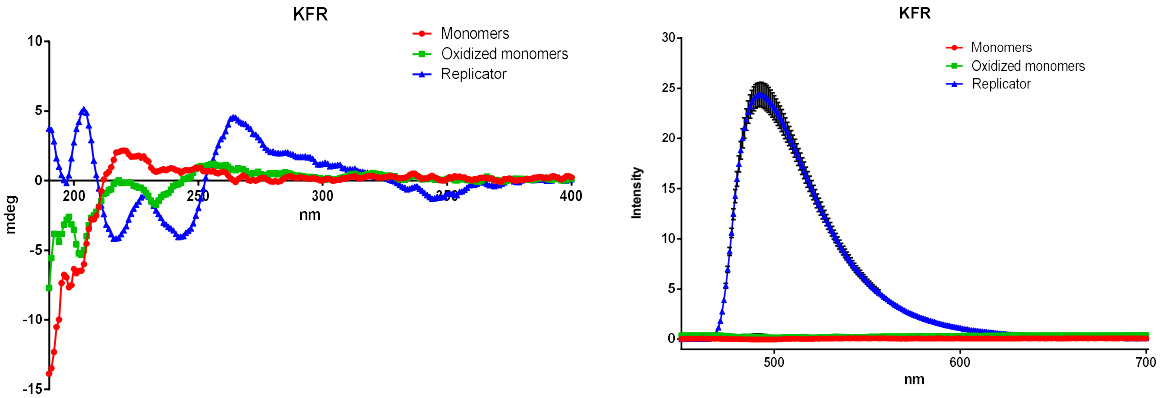


Figure S19: CD spectra (left) and UV spectra (right) of a ThT assay of XGLKFR (in triplicates) showing samples with monomers as red circles, oxidised monomers as green squares and fibres as blue triangles; error bars are indicated in black. Samples were prepared as described in the previous section

TEM images

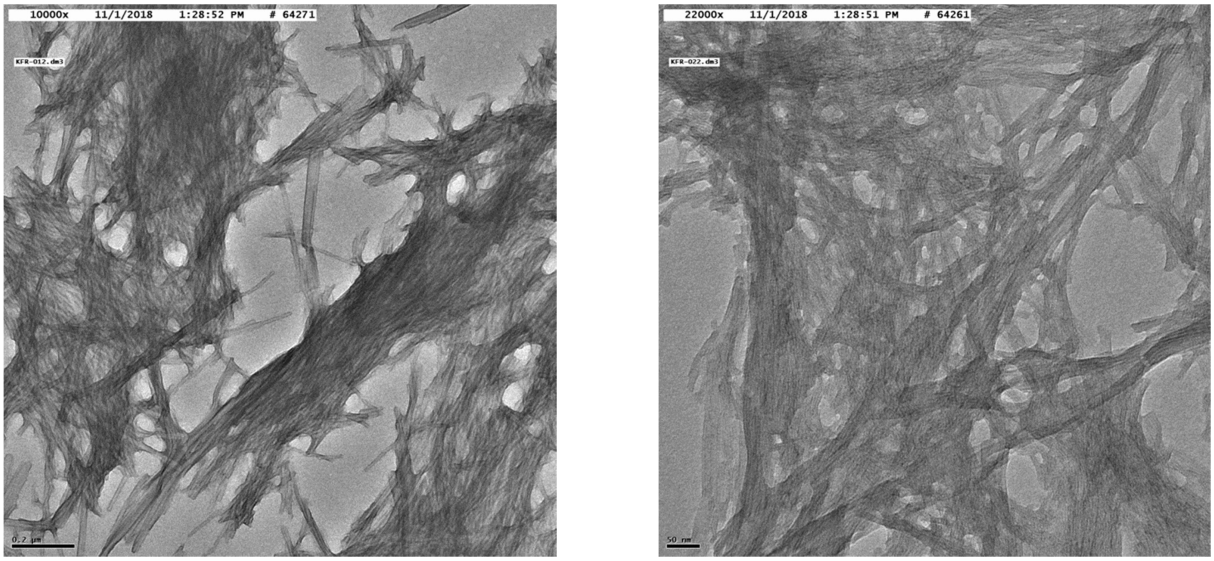


*Figure S20: Representative negative stain TEM images of fibres of **1** at 200 nm (left) and 50 nm (right), after approximately 2 weeks.*

Measurement	Width (nm)	Measurement	Width (nm)
1	38.034	**16**	29.626
2	39.931	**17**	18.667
3	28.567	**18**	21.845
4	25.314	**19**	15.189
5	24.447	**20**	15.612
6	27.015	**21**	15.743
7	21.353	**22**	21.907
8	22.72	**23**	12.411
9	21.576	**24**	19.927
10	22.508	**25**	19.668
11	18.502		
12	12.217	Mean	21.015
13	10.788	SD	7.692
14	11.847	Min	9.963
15	9.963	Max	39.931

*Table S1: Width of fibres of **1** measured across various TEM images*

Catalysis

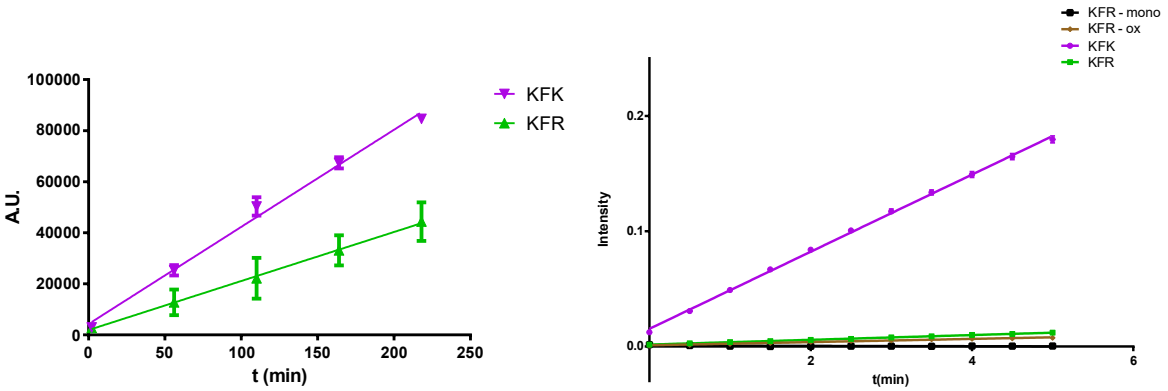
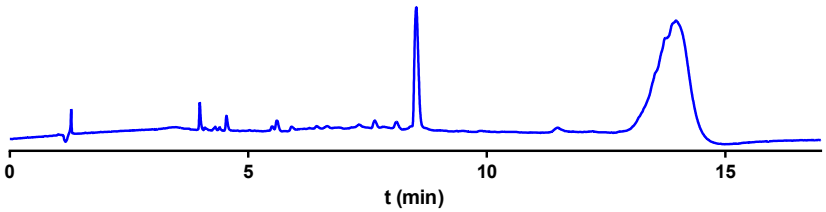


*Figure S21: Methodol breakdown (left) and Sulfmoc glycine deprotection (right) performed by **1** as discussed in the main body of the text*

Unstirred library



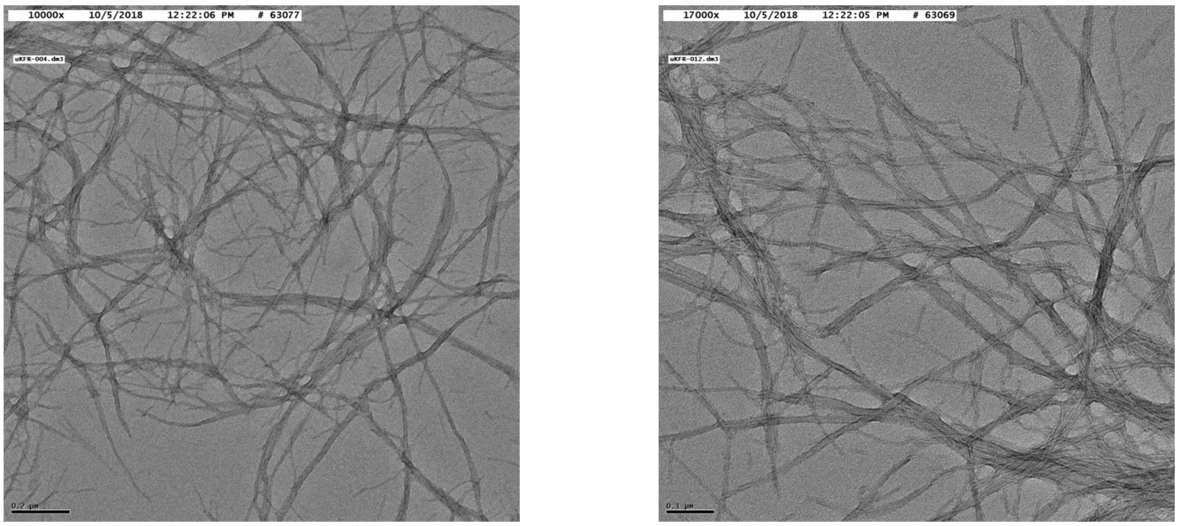
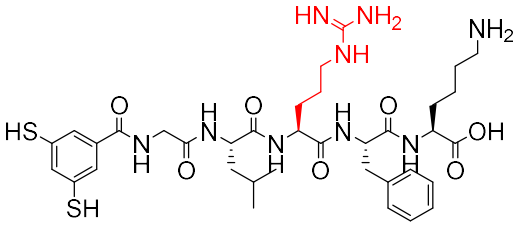


*Figure S22: Top: UPLC trace of a unstirred library of **1** after 2 weeks, showing a trimer peak. Bottom: Representative negative stain TEM images of an unstirred library of **1** at 200 nm (left) and 50 nm (right) after approximately 2 weeks.*

XGLRFK "2"



UPLC traces of 2

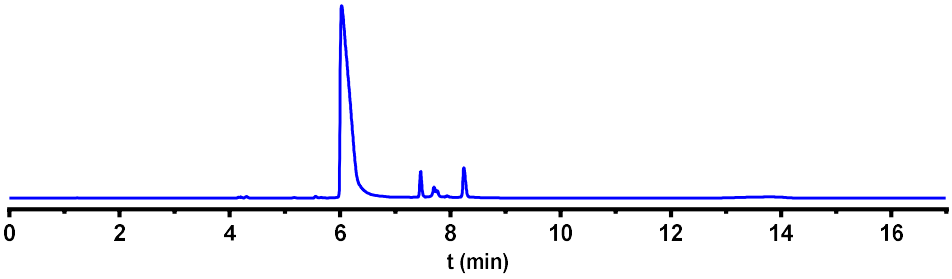


*Figure S23: UPLC chromatogram of **2** after library preparation*

Peak	Retention time	Percentage
1 Monomer	6.03	90.47
2 Tetramer	7.46	2.62
3 Linear dimer	7.70	1.30
4 Trimer	8.25	4.07

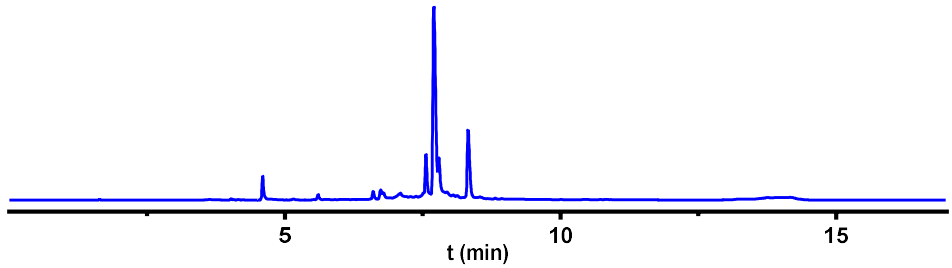


*Figure S24: UPLC chromatogram of **2** after approximately 1 month*

Peak	Retention time	Percentage
1 Over oxidised species	4.59	3.51
2 Tetramer	7.56	7.63
3 Hexamer	7.70	42.71
4 Pentamer	7.79	8.87
5 Trimer	8.32	15.33

Mass Spectra of 2

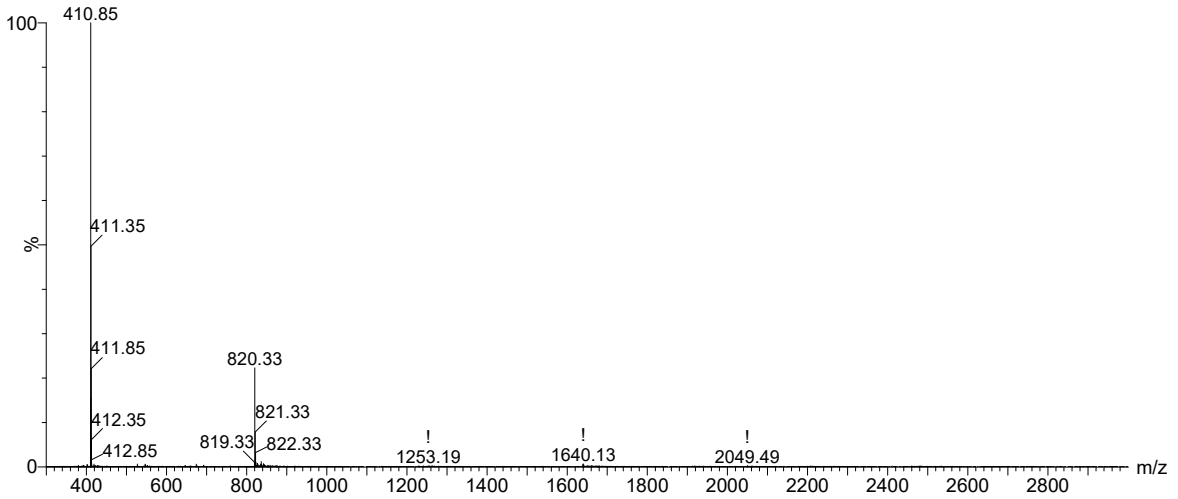


*Figure S25: Mass spectrum of over oxidised (**2**)$_1$ (retention time 4.2) from the LC-MS analysis of the building block after 5 days*
m/z observed: 820.33 $[M+2O+1H]^{1+}$; 410.85 $[M+2O+2H]^{2+}$
m/z calculated: 820.35 $[M+2O+1H]^{1+}$; 410.68 $[M+2O+2H]^{2+}$

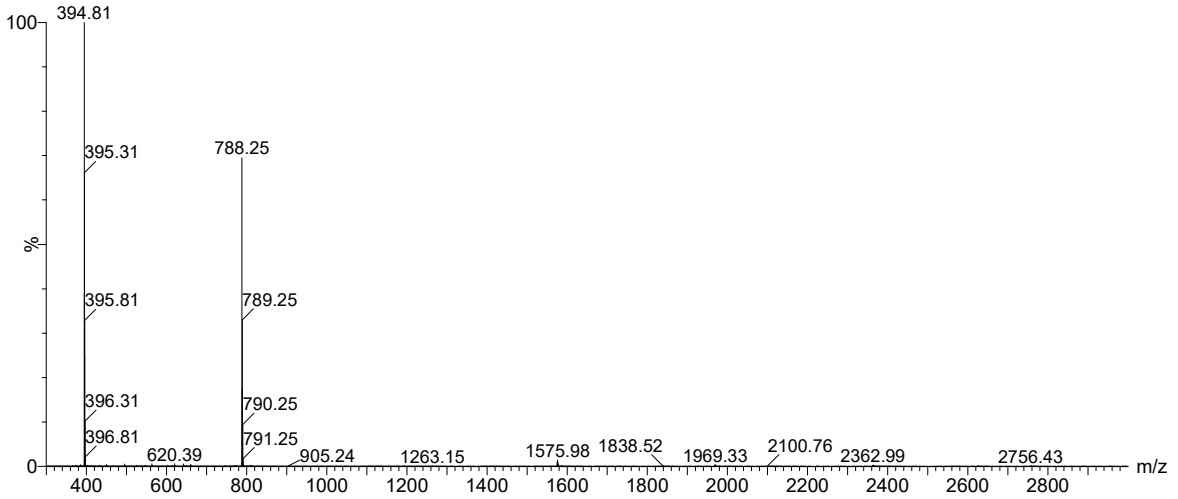


*Figure S26: Mass spectrum of (**2**)$_1$ (retention time 6.0) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 788.25 $[M+1H]^{1+}$; 394.81 $[M+2H]^{2+}$
m/z calculated: 788.36 $[M+1H]^{1+}$; 394.68 $[M+2H]^{2+}$

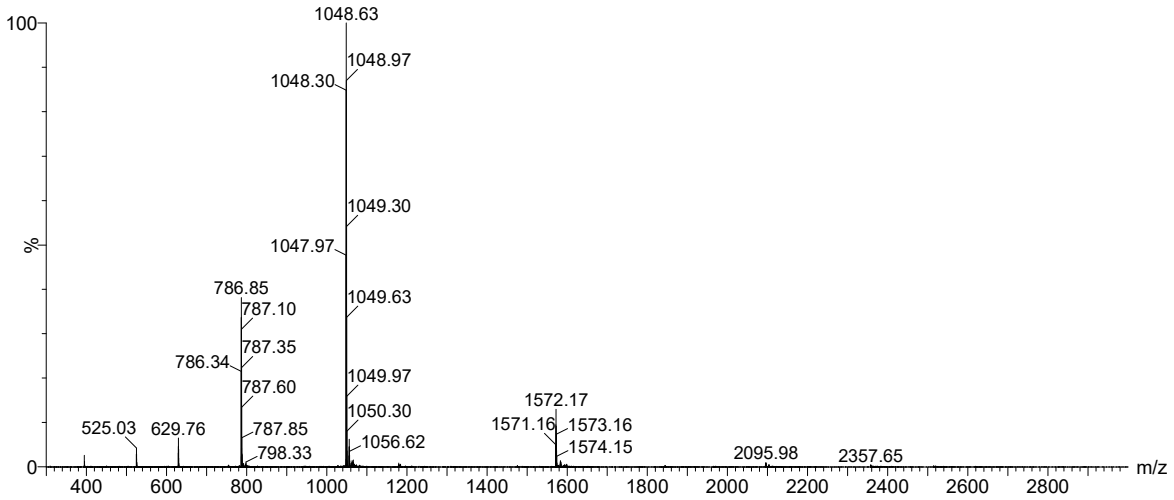


*Figure S27 Mass spectrum of (**2**)$_4$ (retention time 7.6) from the LC-MS analysis of the building block after 5 days.*
m/z observed: 1048.63 $[M+3H]^{3+}$; 786.85 $[M+4H]^{4+}$
m/z calculated: 1048.46 $[M+3H]^{3+}$; 786.59 $[M+4H]^{4+}$

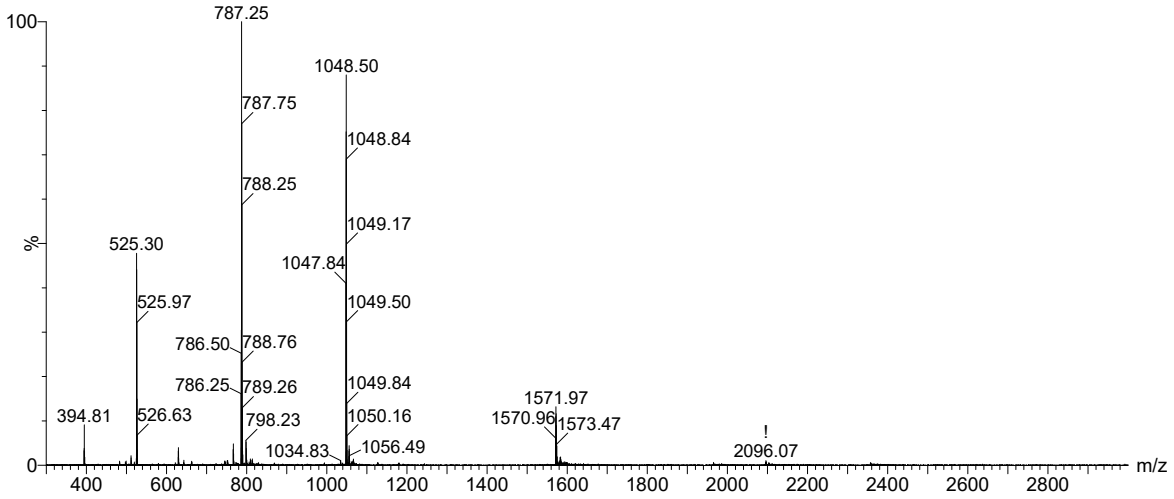


*Figure S28: Mass spectrum of (**2**)$_2$ (retention time 7.7) from the LC-MS analysis of the building block after library preparation. NB: Tetramer co-elutes.*
m/z observed: 787.25 $[M+2H]^{2+}$; 525.30 $[M+3H]^{3+}$
m/z calculated: 787.35 $[M+2H]^{2+}$; 525.24 $[M+3H]^{3+}$

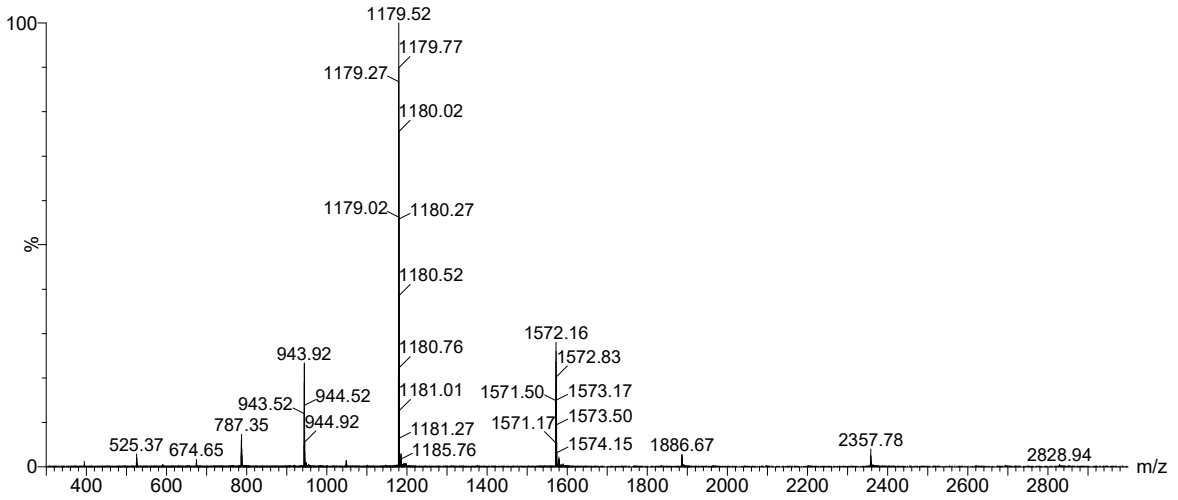


*Figure S29 Mass spectrum of (**2**)$_6$ (retention time 7.7) from the LC-MS analysis of the building block after 5 days. NB. Pentamer co-elutes with this peak*

m/z observed: 1179.52 $[M+4H]^{4+}$; 943.92 $[M+5H]^{5+}$

m/z calculated: 1179.51 $[M+4H]^{4+}$; 943.81 $[M+5H]^{5+}$

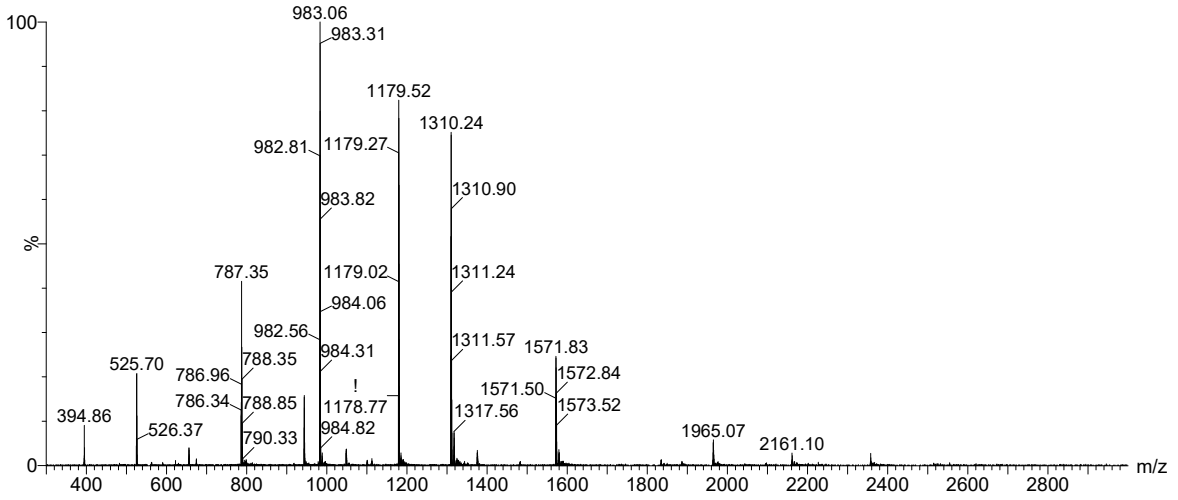


*Figure S30 Mass spectrum of (**2**)$_5$ (retention time 7.8) from the LC-MS analysis of the building block after 5 days. NB. Hexamer co-elutes with this peak*

m/z observed: 1310.24 $[M+3H]^{3+}$; 983.06 $[M+4H]^{4+}$

m/z calculated: 1310.23 $[M+3H]^{3+}$; 982.93 $[M+4H]^{4+}$

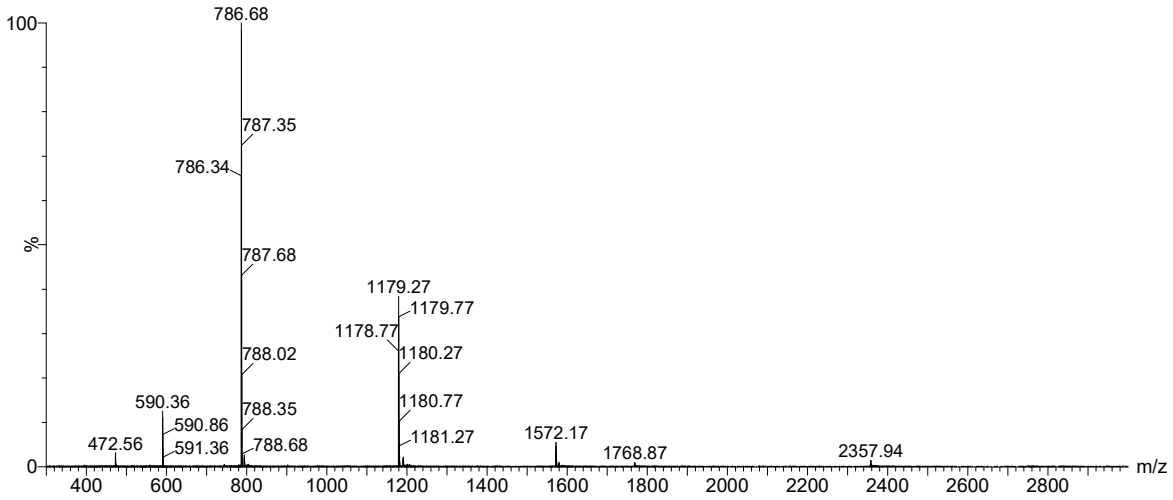


*Figure S31: Mass spectrum of (**2**)$_3$ (retention time 8.3) from the LC-MS analysis of the building block after 5 days.*
m/z observed: 1179.27 $[M+2H]^{2+}$; 786.68 $[M+3H]^{3+}$
m/z calculated: 1179.51 $[M+2H]^{2+}$; 786.68 $[M+3H]^{3+}$

Kinetic plots

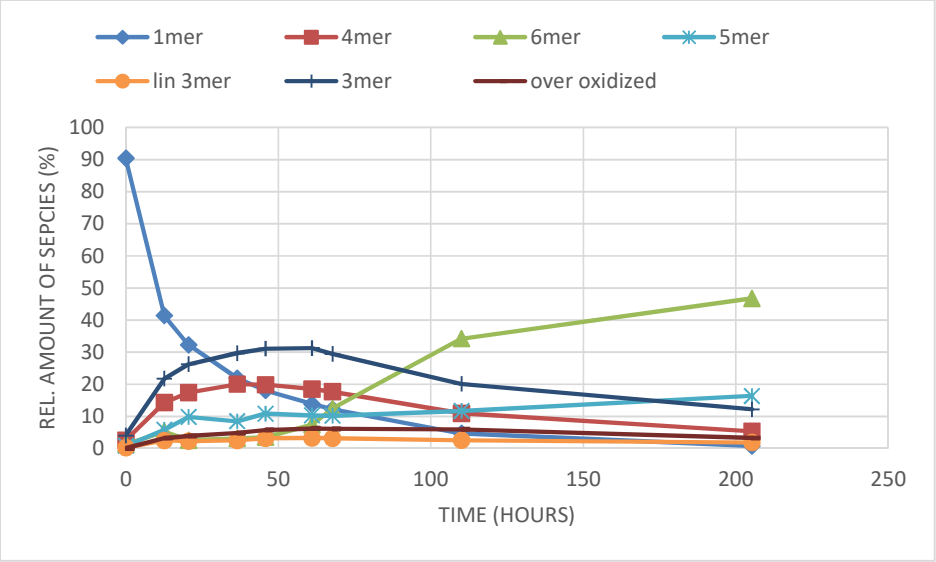


*Figure S32: Kinetic profile of a library made from **2** in borate buffer (pH 8.1) under continuous stirring at 45°C NB. The hexamer will keep growing, but was not followed for longer.*

Seeding

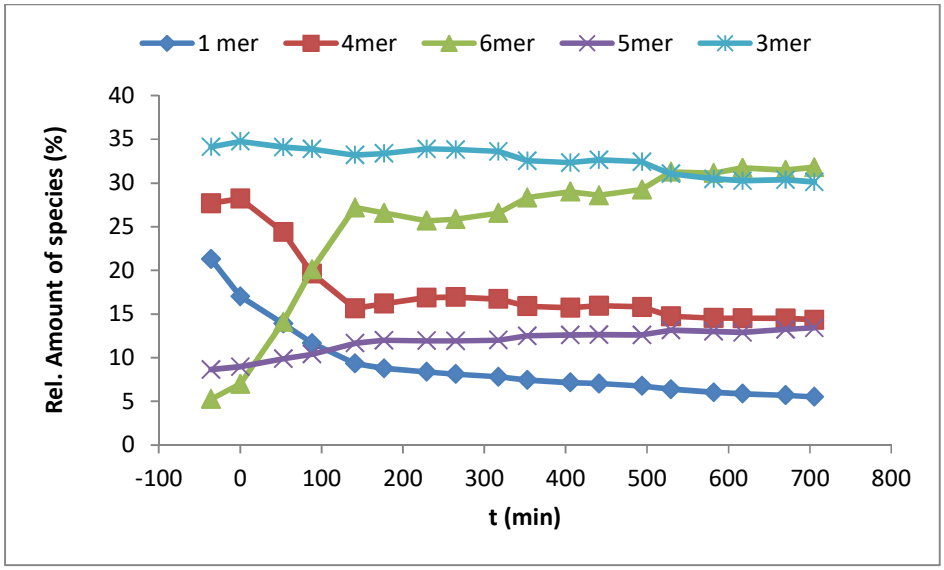


*Figure S33: Kinetic profile of a library made from an 80% oxidised library of **2** in borate buffer (pH 8.1) under continuous stirring at 40°C, upon addition of a seed of fibres of **2** at t = 0*

CD and ThT data

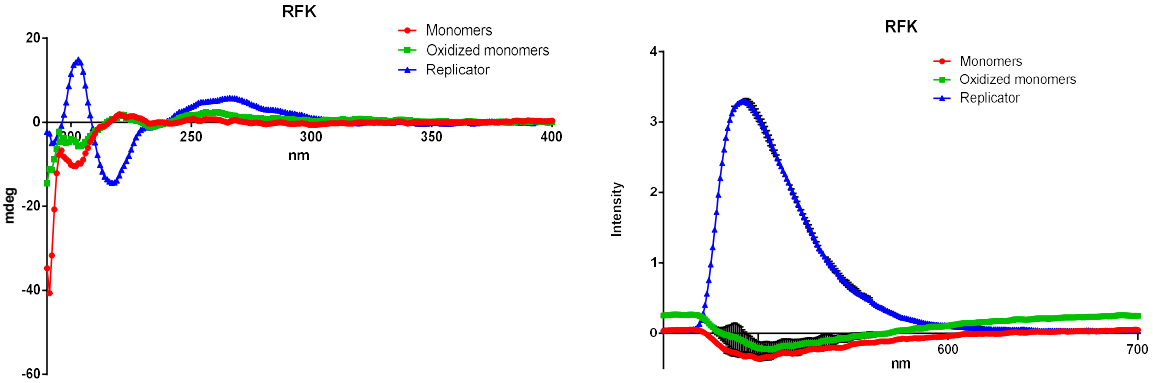


Figure S34: CD spectra (left) and UV spectra (right) of a ThT assay of XGLRFK (in triplicates) showing samples with monomers as red circles, oxidised monomers as green squares and fibres as blue triangles; error bars are indicated in black. Samples were prepared as described in the previous section

TEM images

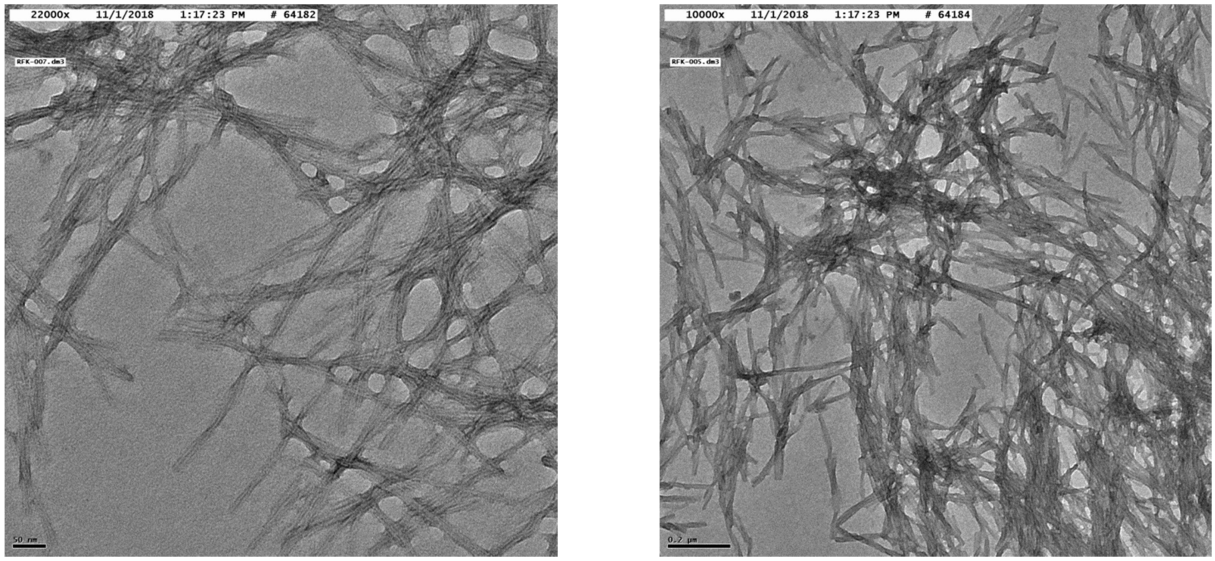


*Figure S35: Representative negative stain TEM images of fibres of **2** at 200 nm (left) and 50 nm (right) after approximately 2 weeks.*

Measurement	Width (nm)	Measurement	Width (nm)
1	15.257	**16**	16.066
2	19.668	**17**	17.277
3	22.508	**18**	15.894
4	16.424	**19**	15.916
5	14.917	**20**	20.686
6	14.801	**21**	17.552
7	18.278	**22**	16.256
8	32.227	**23**	12.411
9	12.951	**24**	21.144
10	12.047	**25**	14.09
11	15.894	**26**	27.015
12	29.982	Mean	19.11
13	22.356	SD	5.765
14	31.713	Min	12.047
15	23.52	Max	32.227

*Table S2: Width of fibres of **2** measured across various TEM images*

Unstirred library

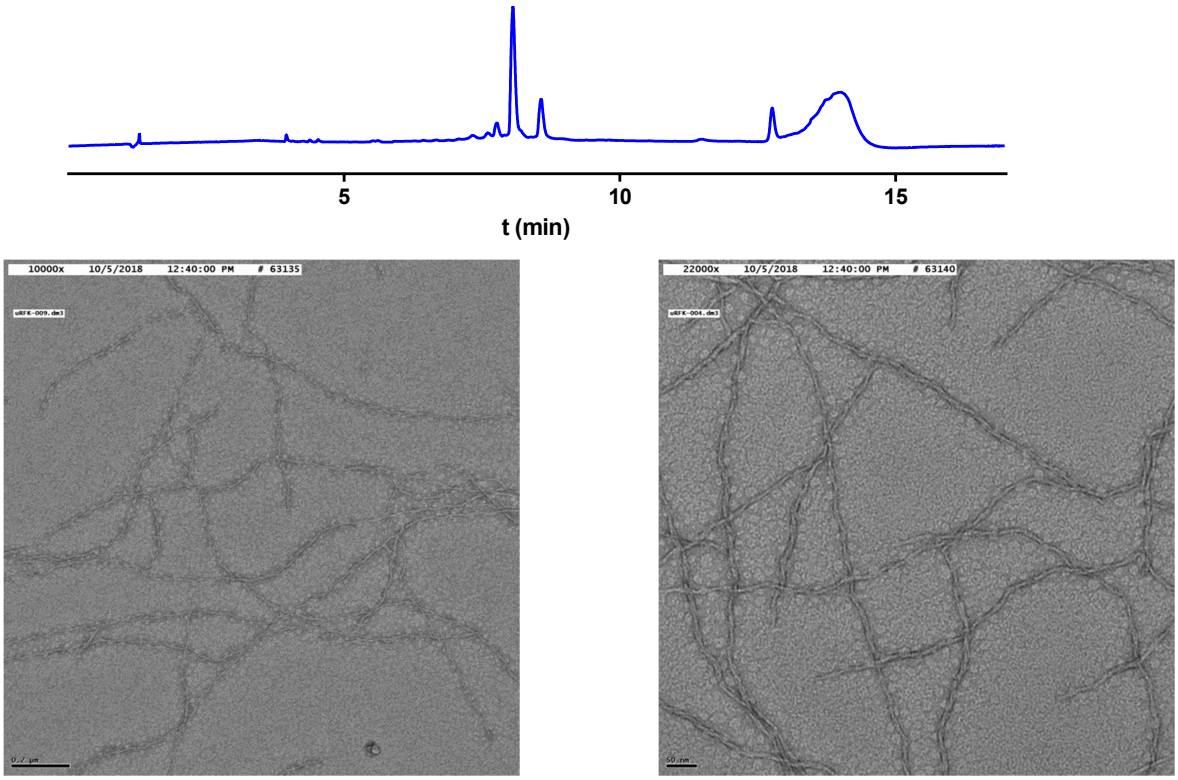


*Figure S36: Top: UPLC trace of an unstirred library of **2** after 2 weeks showing a pentamer and trimer peak. Bottom: Representative negative stain TEM images of an unstirred library of **2** at 200 nm (left) and 50 nm (right) after approximately 2 weeks.*

Catalysis

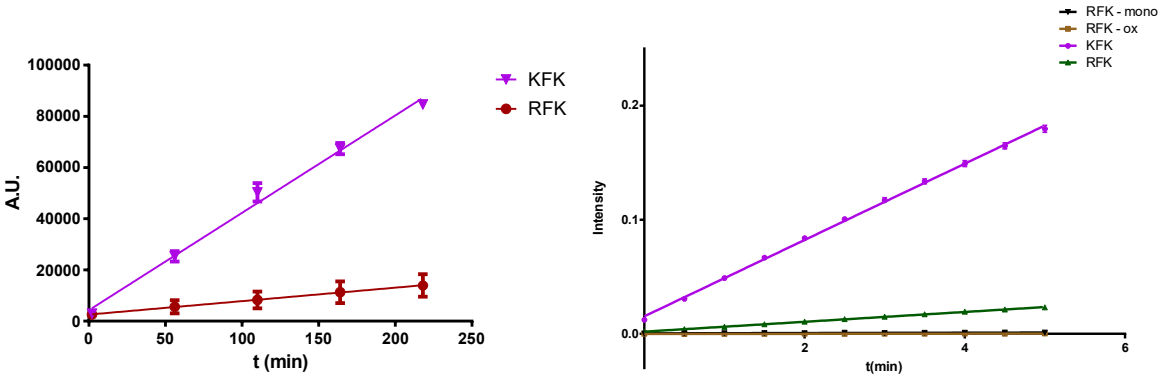
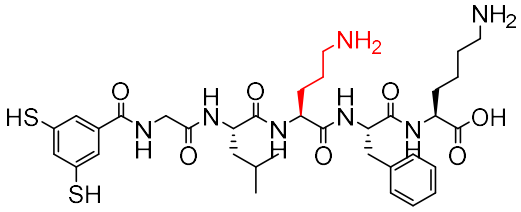


*Figure S37: Methodol breakdown (left) and Sulfmoc glycine deprotection (right) performed by **2** as discussed in the main body of the text.*

XGLOFK "**3**"



UPLC traces of 3

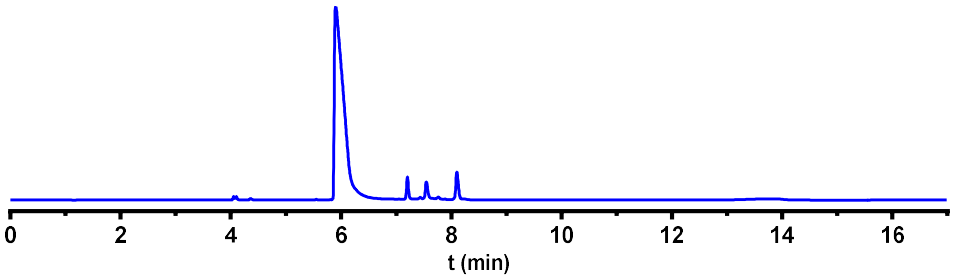


*Figure S38: UPLC chromatogram of **3** after library preparation*

Peak	Retention time	Percentage
1 Monomer	5.90	89.36
2 Tetramer	7.20	2.41
3 Linear Dimer	7.55	2.85
4 Trimer	8.10	3.75

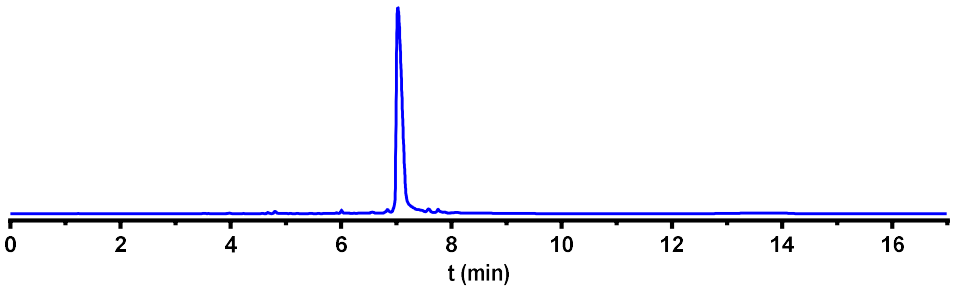


*Figure S39: UPLC chromatogram of **3** after 3 days*

Peak	Retention time	Percentage
1 Hexamer	7.03	93.45
2 Hexamer - H_2O	7.59	1.44
3 trimer	7.56	1.25

Mass spectra of 3

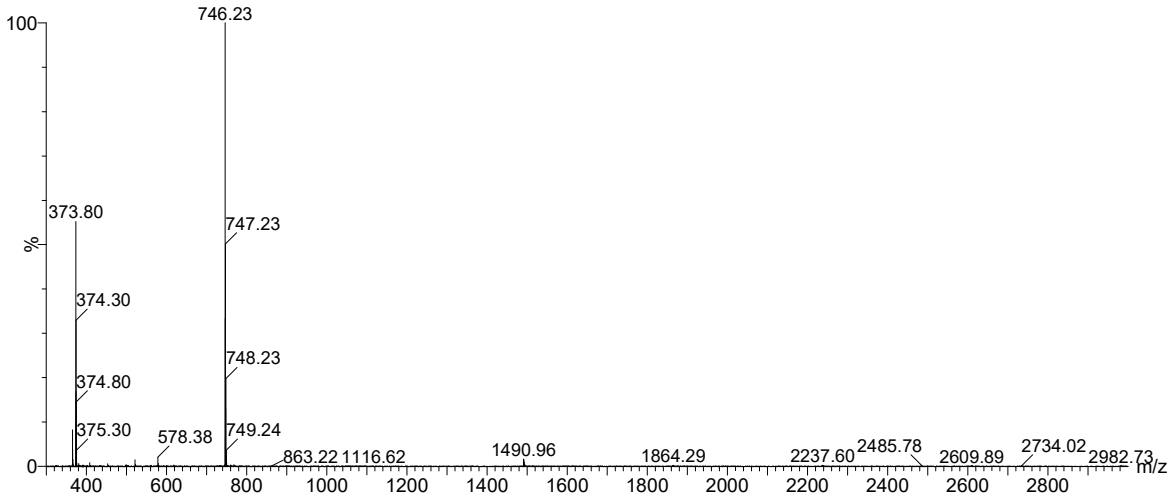


*Figure S40: Mass spectrum of (**3**)$_1$ (retention time 5.9) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 746.23 $[M+1H]^{1+}$; 373.80 $[M+2H]^{2+}$
m/z calculated: 746.34 $[M+1H]^{1+}$; 373.67 $[M+2H]^{2+}$

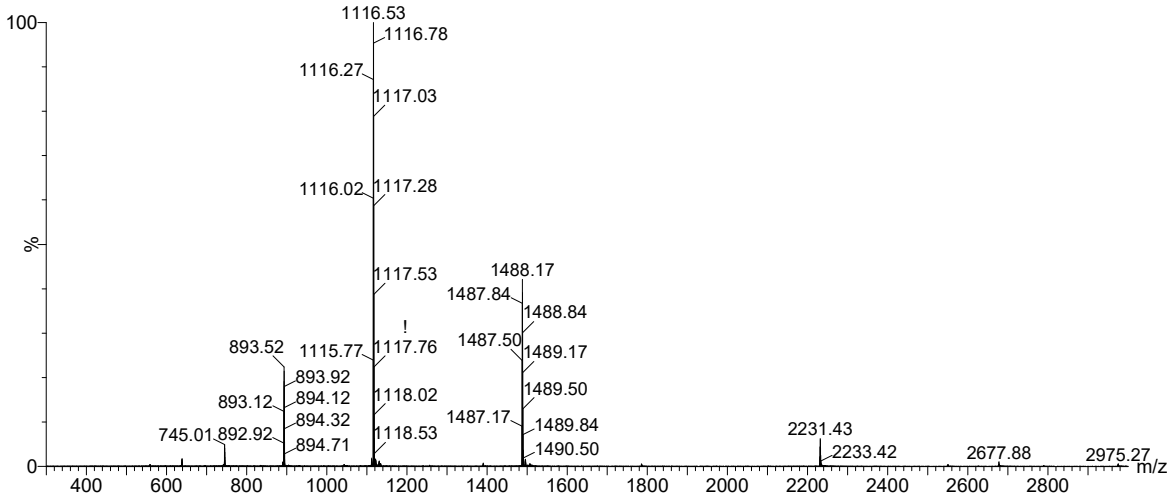


*Figure S41: Mass spectrum of (**3**)$_6$ (retention time 7.0) from the LC-MS analysis of the building block after 5 days.*
m/z observed: 1488.17 $[M+3H]^{3+}$; 1116.53 $[M+4H]^{4+}$
m/z calculated: 1488.30 $[M+3H]^{3+}$; 1116.48 $[M+4H]^{4+}$

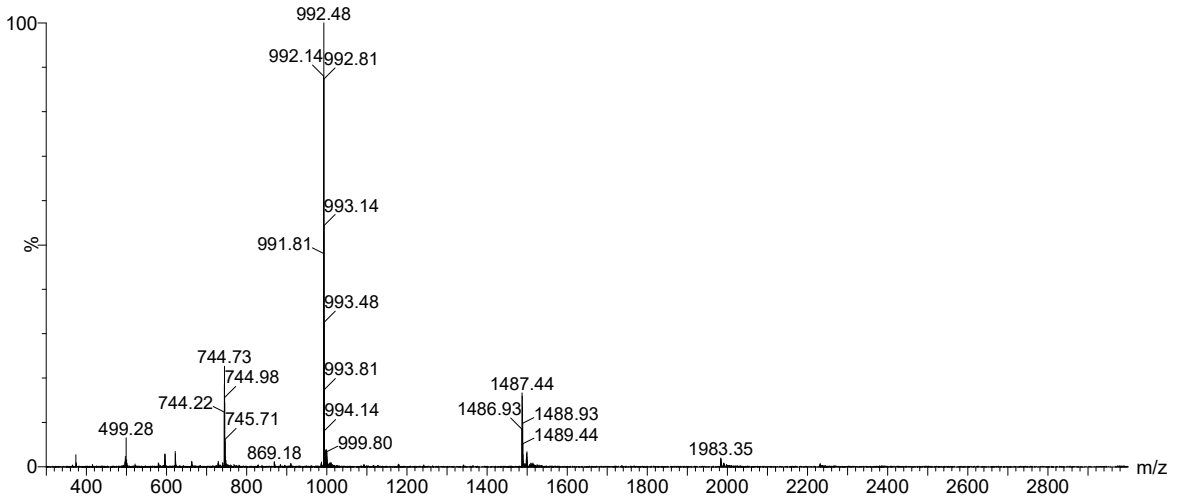


*Figure S42: Mass spectrum of (**3**)$_4$ (retention time 7.2) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 992.48 $[M+3H]^{3+}$; 744.73 $[M+4H]^{4+}$
m/z calculated: 992.43 $[M+3H]^{3+}$; 744.57 $[M+4H]^{4+}$

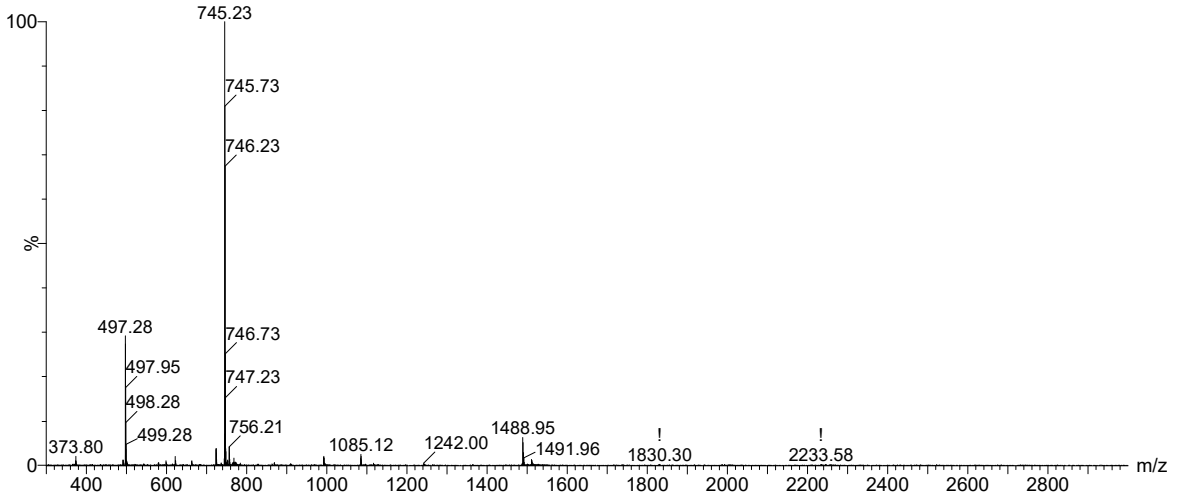


*Figure S43: Mass spectrum of (**3**)$_2$ (retention time 7.6) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 745.23 $[M+2H]^{2+}$; 497.28 $[M+3H]^{3+}$
m/z calculated: 745.33 $[M+2H]^{2+}$; 497.22 $[M+3H]^{3+}$

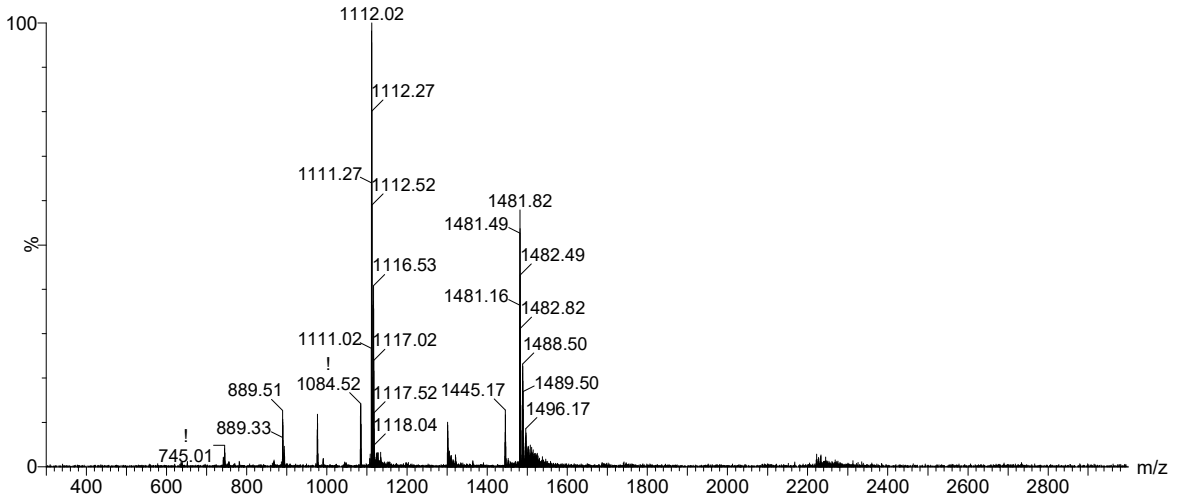


*Figure S44: Mass spectrum of (**3**)$_6$ - H_2O(retention time 7.6) from the LC-MS analysis of the building block after 5 days.*
m/z observed: 1481.82 $[M+3H]^{3+}$; 1112.02 $[M+4H]^{4+}$
m/z calculated: 1482.30 $[M+3H]^{3+}$; 1111.98 $[M+4H]^{4+}$

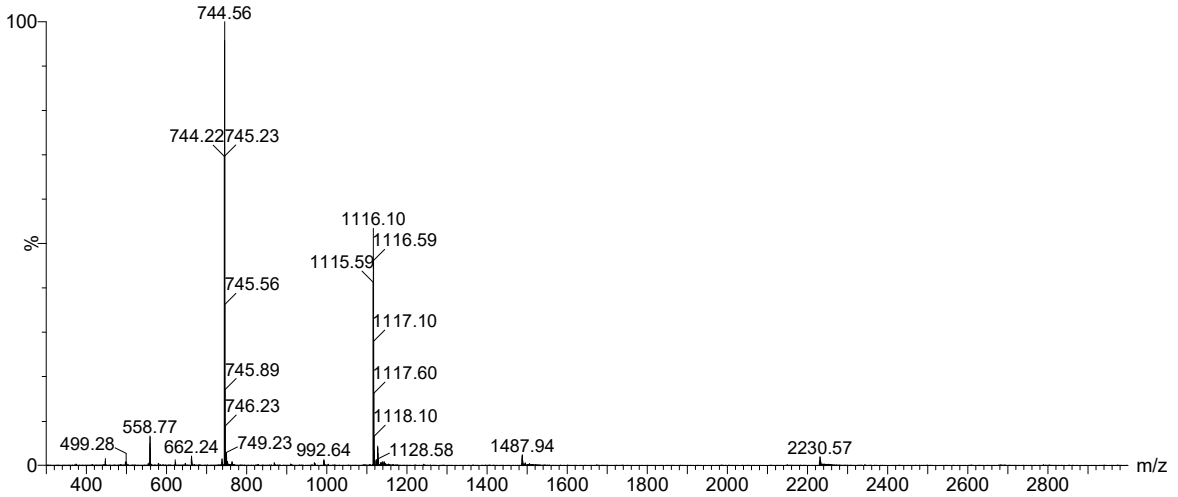


*Figure S45: Mass spectrum of (**3**)$_3$ (retention time 8.1) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 1116.10 $[M+2H]^{2+}$; 744.56 $[M+3H]^{3+}$
m/z calculated: 1116.48 $[M+2H]^{2+}$; 744.66 $[M+3H]^{3+}$

Kinetic plots

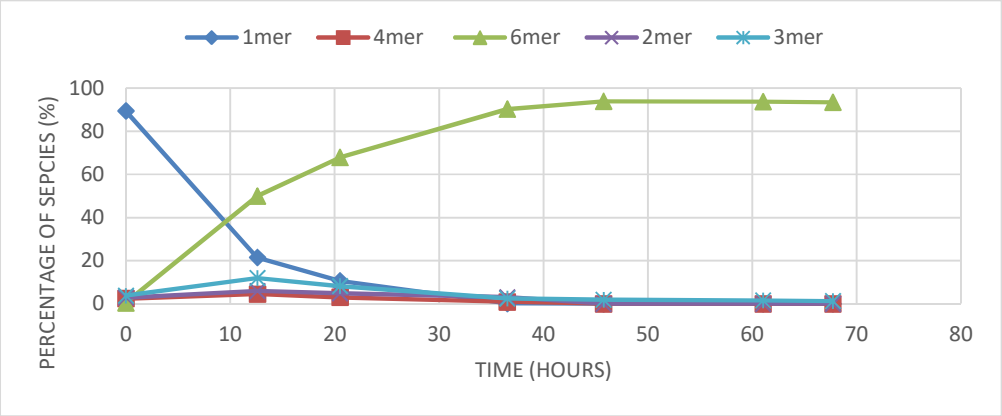


*Figure S46: Kinetic profile of a library made from **3** in borate buffer (pH 8.1) under continuous stirring at 45°C*

Seeding

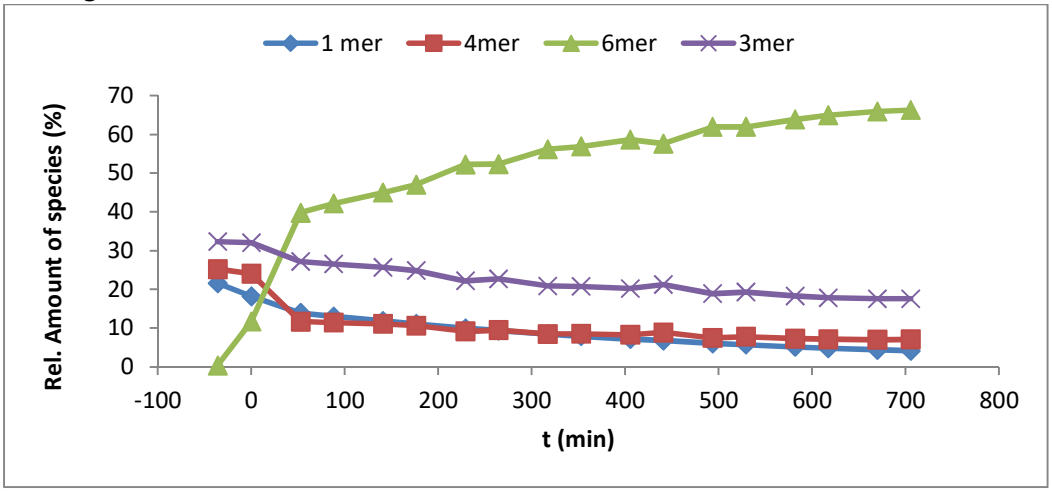


*Figure S47: Kinetic profile of a library made from an 80% oxidised library of **3** in borate buffer (pH 8.1) under continuous stirring at 40°C, upon addition of a seed of fibres of **3** at t = 0*

CD and ThT data

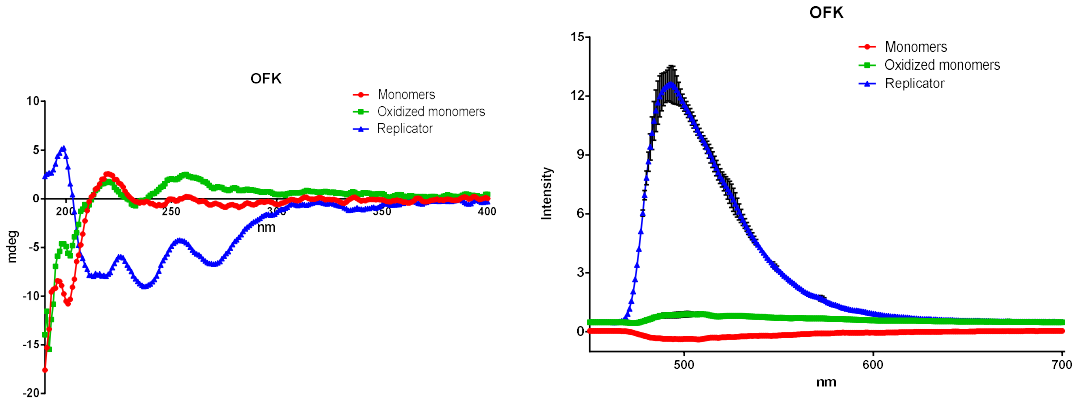


Figure S48: CD spectra (left) and UV spectra (right) of a ThT assay of XGLOFK (in triplicates) showing samples with monomers as red circles, oxidised monomers as green squares and fibres as blue triangles; error bars are indicated in black. Samples were prepared as described in the previous section

TEM images

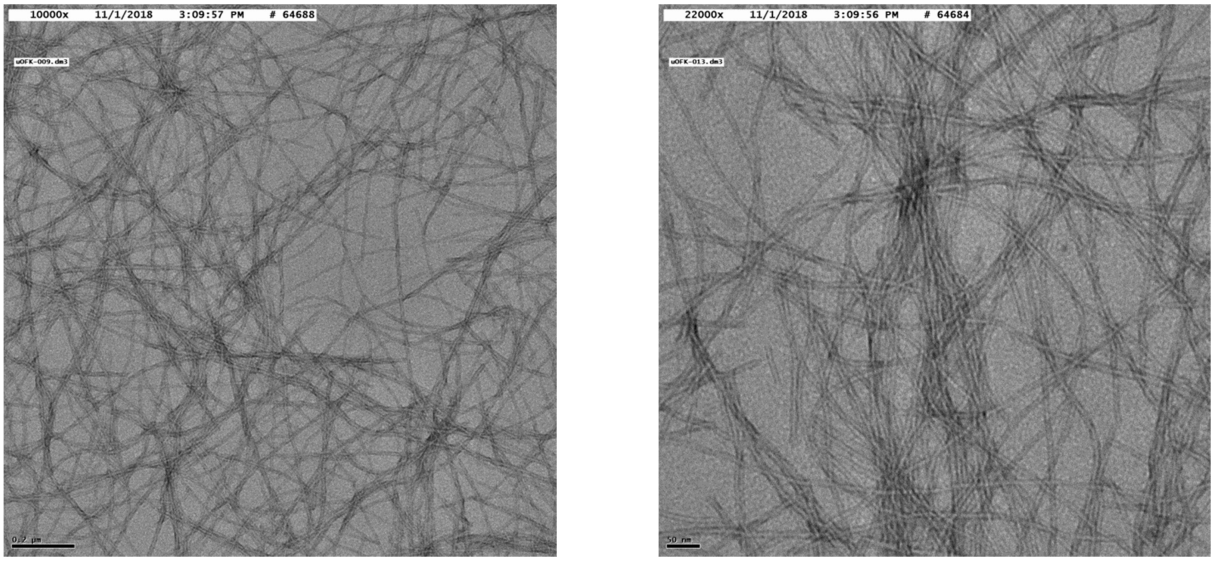


*Figure S49: Representative negative stain TEM images of fibres of **3** at 200 nm (left) and 50 nm (right) after approximately 2 weeks*

Measurement	Width (nm)	Measurement	Width (nm)
1	13.87	**16**	12.047
2	10.883	**17**	15.808
3	18.372	**18**	13.469
4	15.721	**19**	17.218
5	10.788	**20**	15.257
6	26.684	**21**	20.132
7	13.771	**22**	24.166
8	15.189	**23**	20.03
9	18.147		
10	14.09		
11	17.862		
12	16.256	Mean	16.194
13	13.469	SD	3.884
14	15.894	Min	10.788
15	13.342	Max	26.684

*Table S3: Width of fibres of **3** measured across various TEM images*

Catalysis

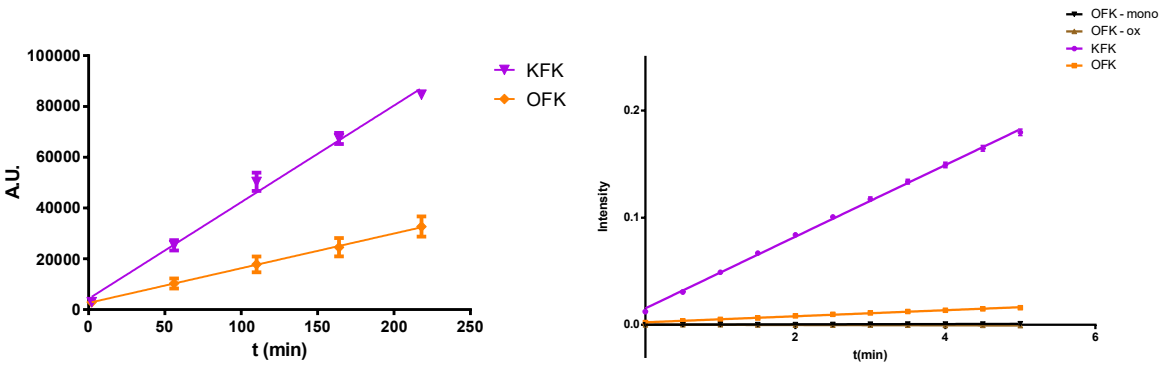


*Figure S50: Methodol breakdown (left) and Sulfmoc glycine deprotection (right) performed by **3** as discussed in the main body of the text.*

Unstirred library

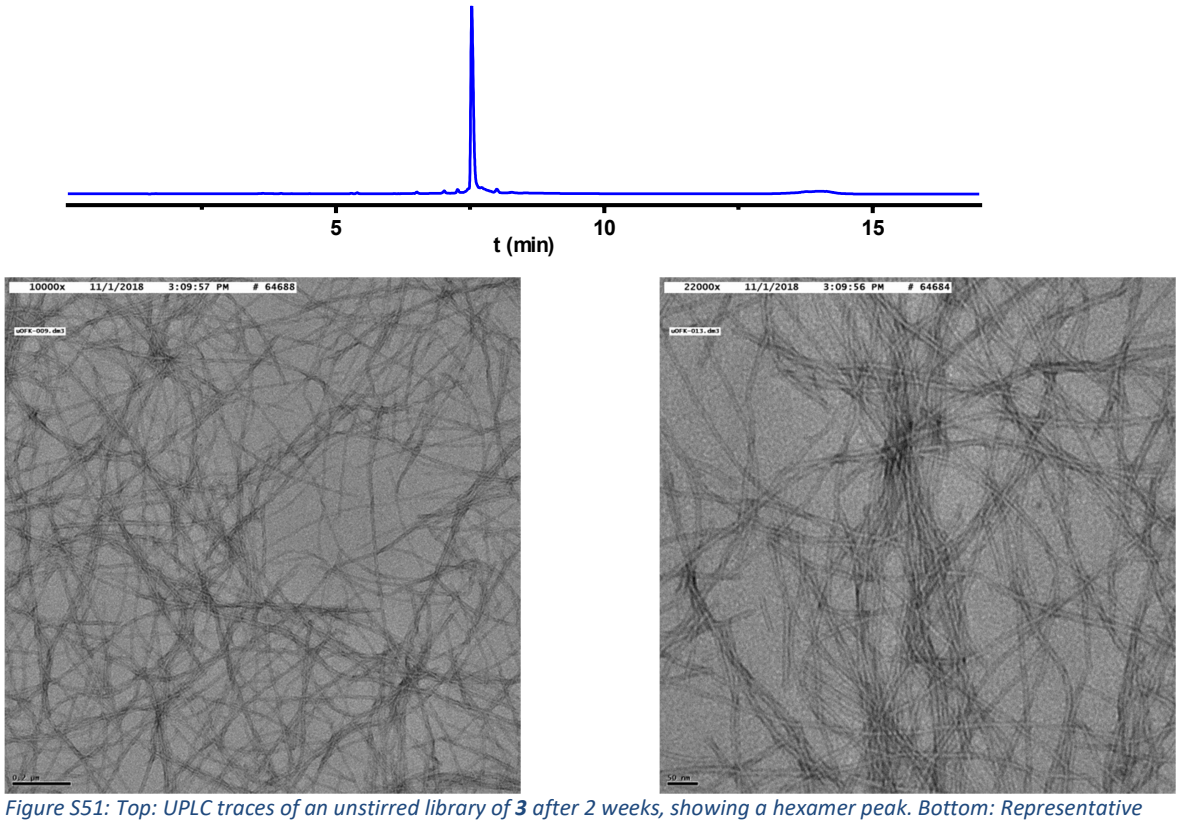
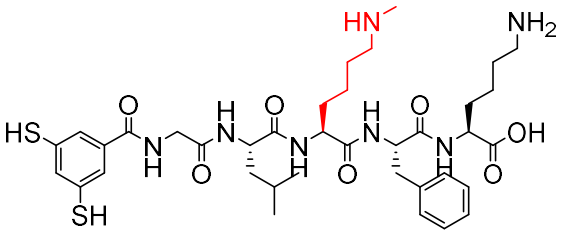


*Figure S51: Top: UPLC traces of an unstirred library of **3** after 2 weeks, showing a hexamer peak. Bottom: Representative negative stain TEM images of an unstirred library of **3** at 200 nm (left) and 50 nm (right) after approximately 2 weeks.*

XGLmKFK "**4**"



UPLC traces of 4

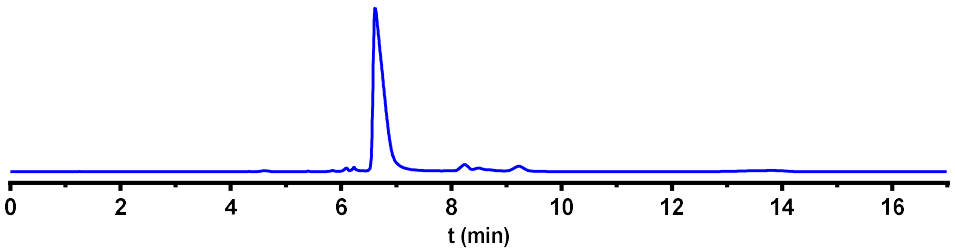


*Figure S52: UPLC chromatogram of **4** after library preparation*

Peak	Retention time	Percentage
1 Monomer	6.62	89.72
2 Tetramer	8.24	3.14
3 Linear Dimer	8.50	2.42
4 Trimer	9.22	2.91

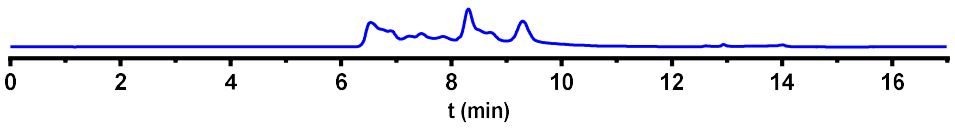


*Figure S53: UPLC chromatogram of **4** after 3 days*

Peak	Retention time	Percentage
1 Large macrocycles (3-20+)	-	-

Mass spectra of 4

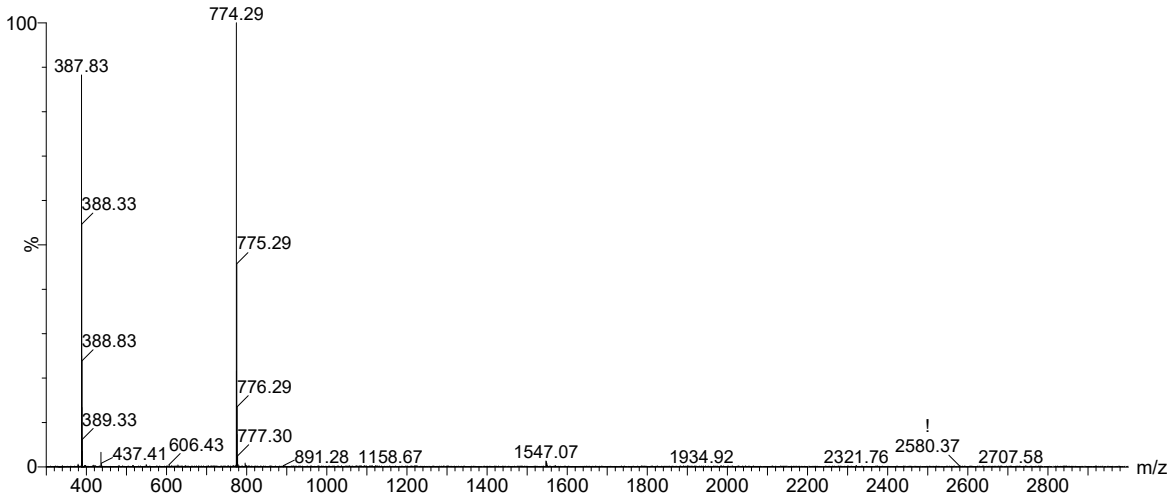


*Figure S54: Mass spectrum of (**4**)$_1$ (retention time 6.6) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 774.29 $[M+1H]^{1+}$; 387.83 $[M+2H]^{2+}$
m/z calculated: 774.37 $[M+1H]^{1+}$; 387.69 $[M+2H]^{2+}$

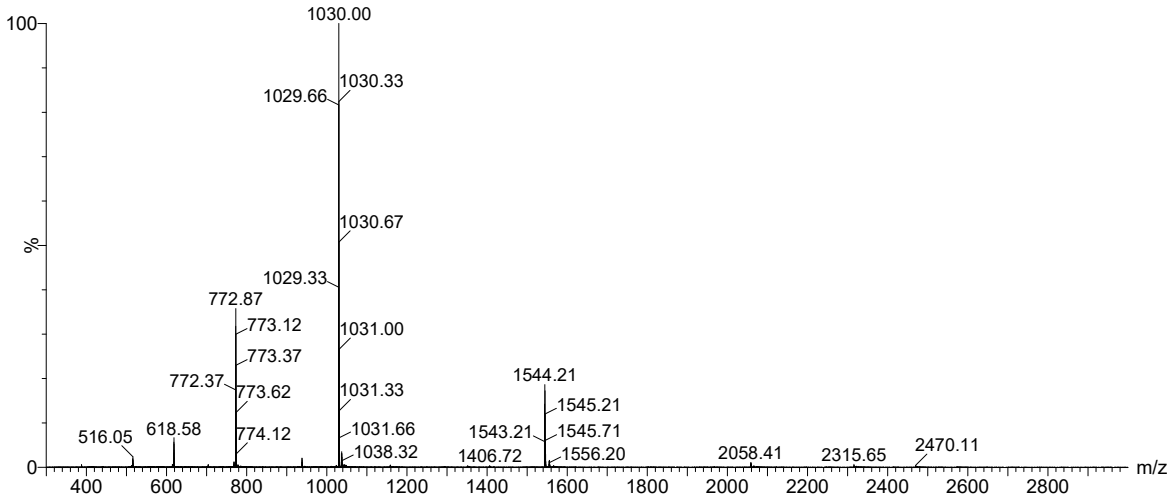


*Figure S55: Mass spectrum of (**4**)$_4$ (retention time 8.2) from the LC-MS analysis of the building block after 1 day.*
m/z observed: 1030.00 $[M+3H]^{3+}$; 772.87 $[M+4H]^{4+}$
m/z calculated: 1029.80 $[M+3H]^{3+}$; 772.60 $[M+4H]^{4+}$

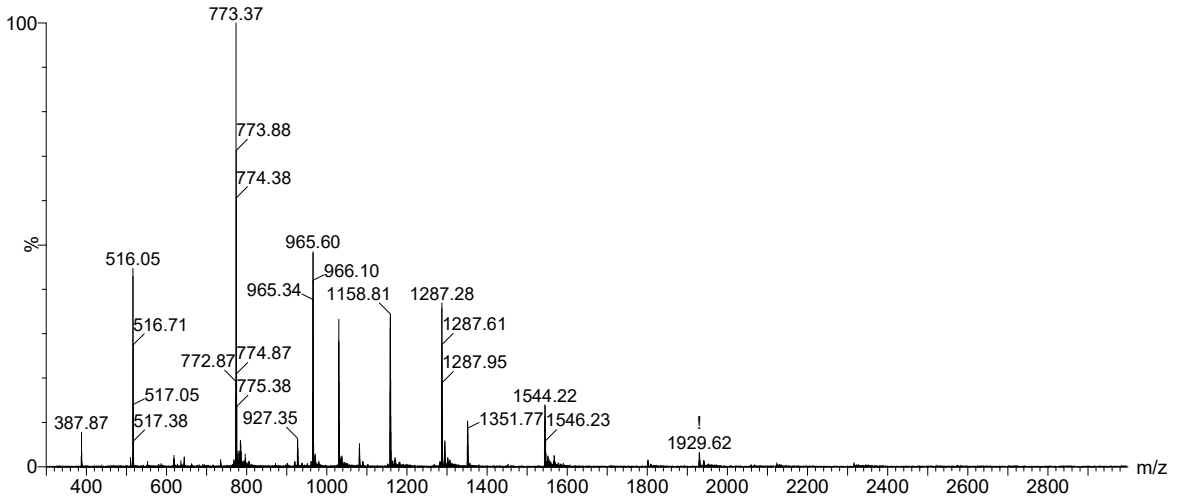


*Figure S56: Mass spectrum of (**4**)$_2$ (retention time 8.5) from the LC-MS analysis of the building block after 1 day.*
m/z observed: 773.37 $[M+2H]^{2+}$; 516.05 $[M+3H]^{3+}$
m/z calculated: 773.36 $[M+2H]^{2+}$; 515.91 $[M+3H]^{3+}$

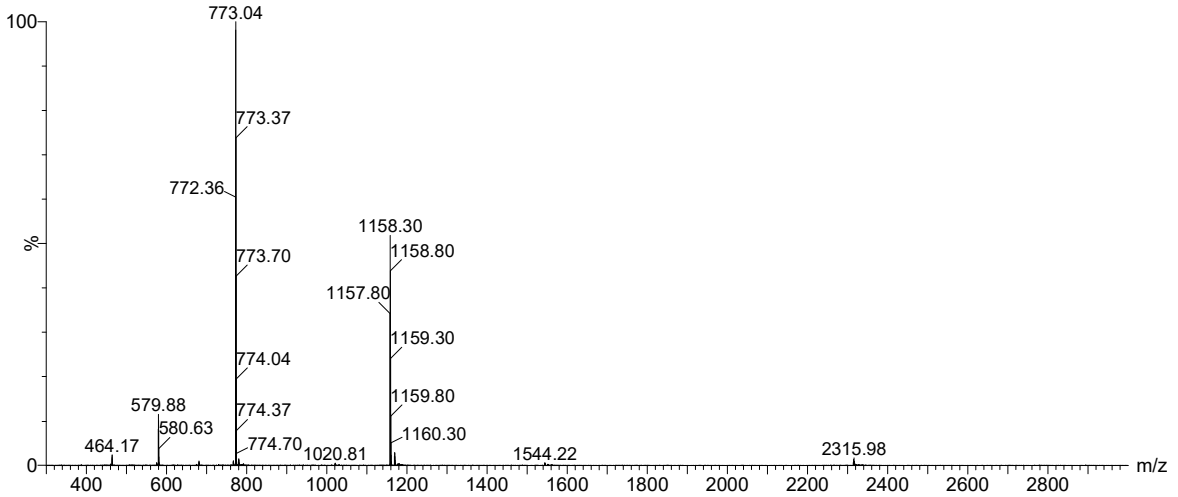


*Figure S57: Mass spectrum of (**4**)$_3$ (retention time 9.2) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 1158.30 $[M+2H]^{2+}$; 773.04 $[M+3H]^{3+}$
m/z calculated: 1158.53 $[M+2H]^{2+}$; 772.69 $[M+3H]^{3+}$

Kinetic plots

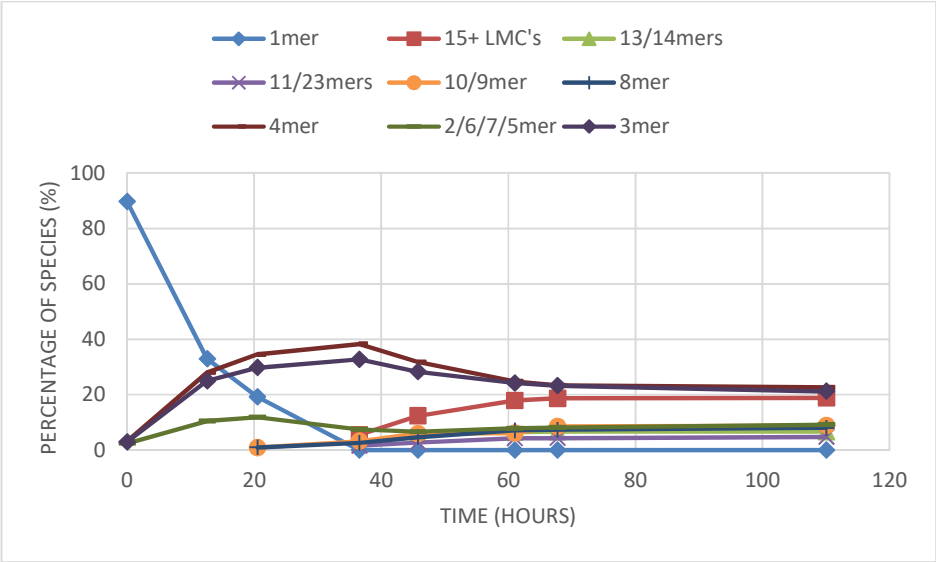


*Figure S58: Kinetic profile of a library made from **4** in borate buffer (pH 8.1) under continuous stirring at 45°C*

Seeding

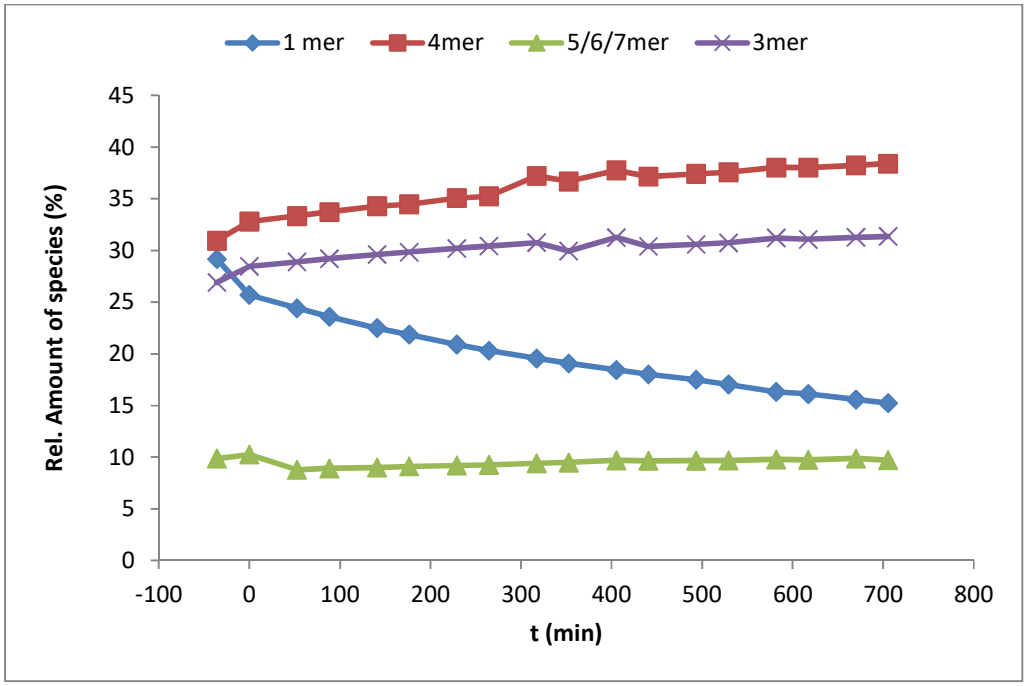


*Figure S59: Kinetic profile of a library made from an 80% oxidised library of **4** in borate buffer (pH 8.1) under continuous stirring at 40°C, upon addition of a fully oxidised library of **4** at t = 0*

TEM Images

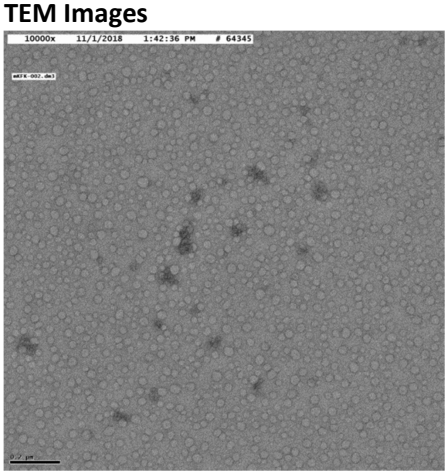


*Figure S60: Representative negative stain TEM images of a fully oxidised library of **4** at 200 nm after approximately 2 weeks*

CD and ThT spectra

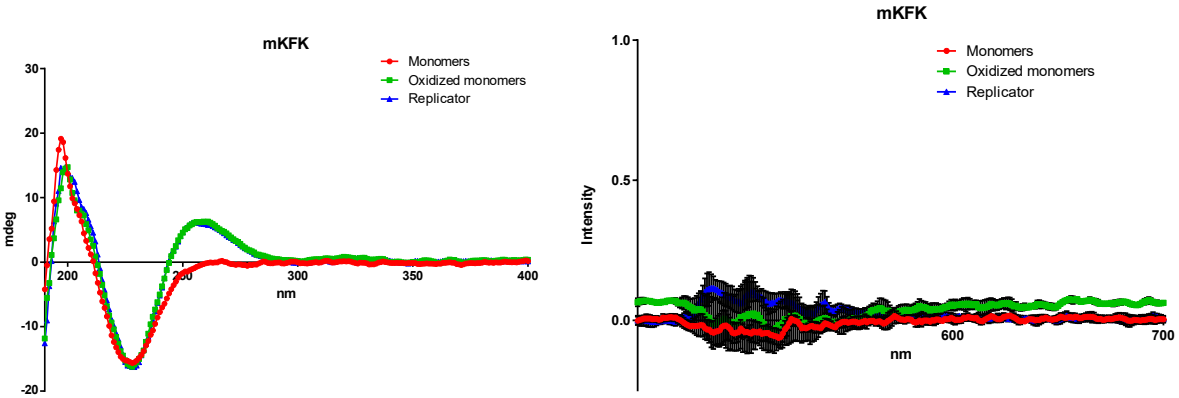
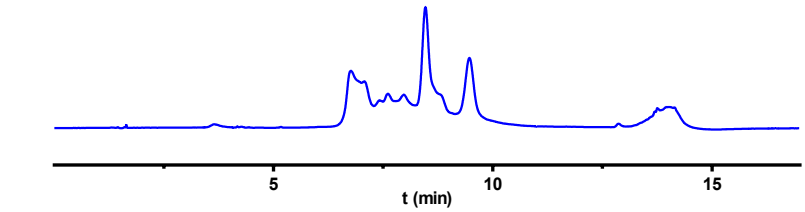


Figure S61: CD spectra (left) and UV spectra (right) of a ThT assay of XGLmKFK (in triplicates) showing samples with monomers as red circles, oxidised monomers as green squares and fibres as blue triangles; error bars are indicated in black. Samples were prepared as described in the previous section

Unstirred library



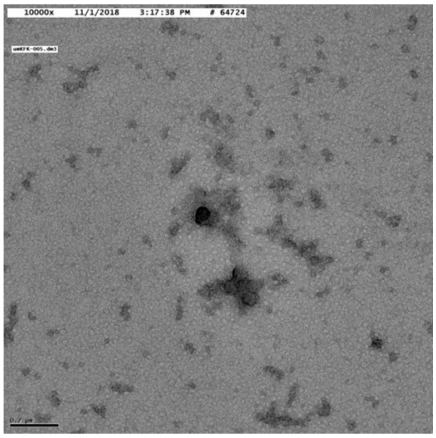
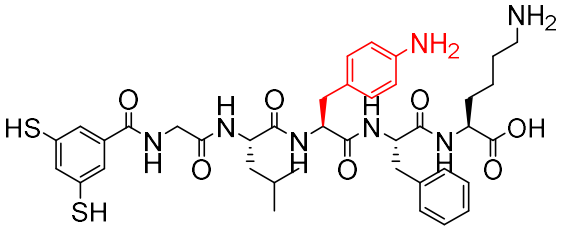


*Figure S62: Top: UPLC traces of an unstirred library of **4** after 2 weeks, showing peaks of various macrocycle sizes. Bottom: Representative negative stain TEM images of an unstirred library of **4** at 200 nm (left) after approximately 2 weeks*

XGLaFFK "**5**"



UPLC traces of 5

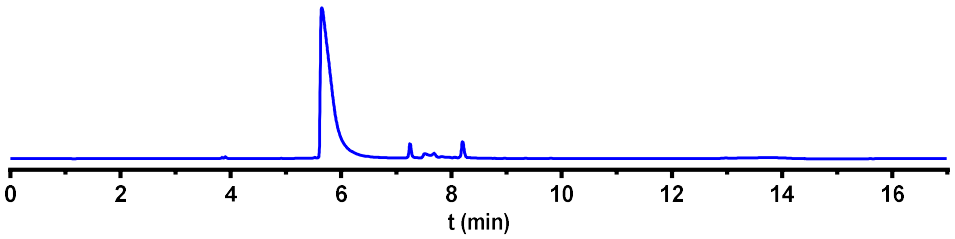


*Figure S63: UPLC chromatogram of **5** after library preparation*

Peak	Retention time	Percentage
1 Monomer	5.65	91.08
2 Tetramer	7.25	2.22
3 Pentamer or Hexamer	7.52	1.46
4 Pentamer or Hexamer	7.69	1.22
5 Trimer	8.204	2.58

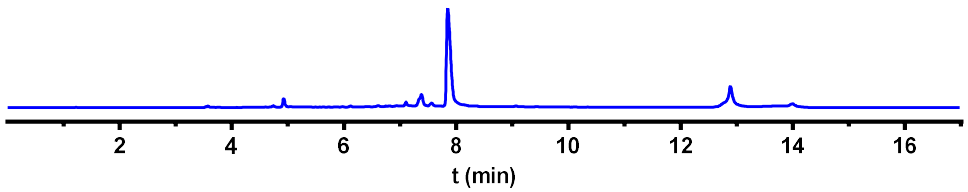


*Figure S64: UPLC chromatogram of **5** after 8.5 days*

Peak	Retention time	Percentage
1 Over oxidised species	4.92	2.7
2 Tetramer	7.11	1.72
3 Pentamer/Hexamer	7.38	7.78
4 Trimer	7.85	59.77
5 Unknown in wash	12.89/14.00	18.66

Mass analysis of 5

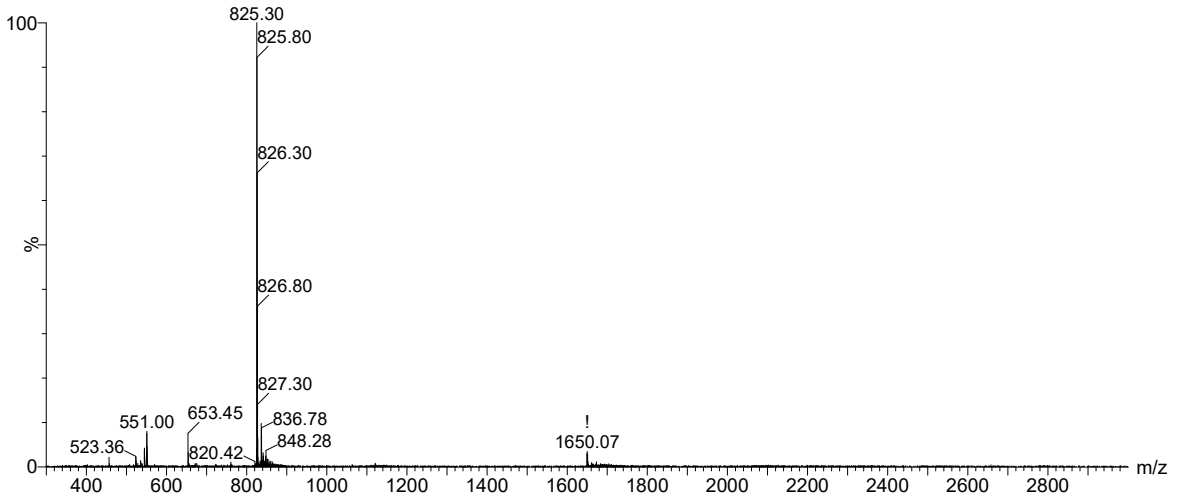


*Figure S65: Mass spectrum of over oxidised (**5**)$_2$ (retention time 4.9) from the LC-MS analysis of the building block after 5 days*
m/z observed: 825.30 $[M+4O+2H]^{2+}$
m/z calculated: 825.32 $[M+4O+2H]^{2+}$

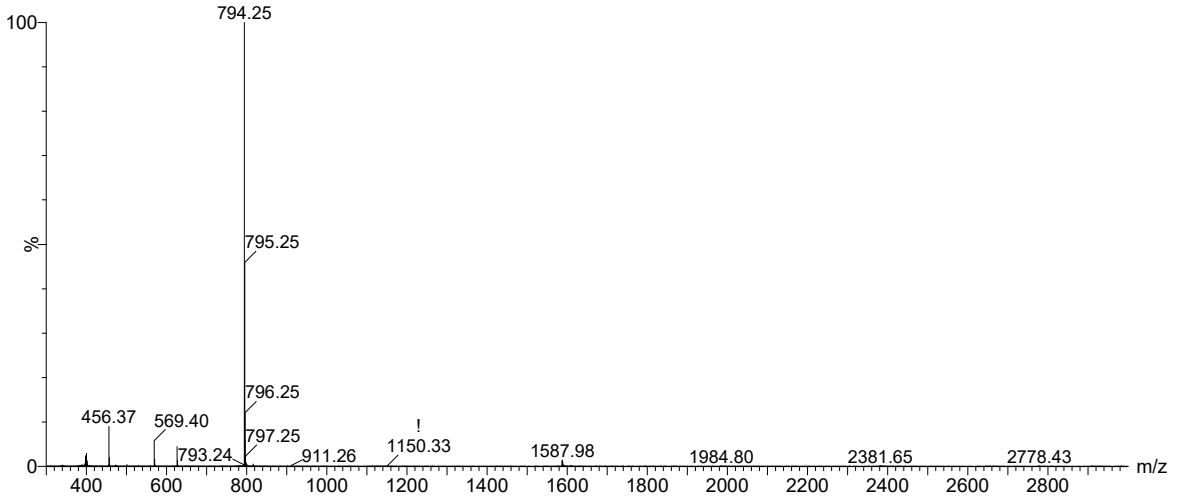


*Figure S66: Mass spectrum of (**5**)$_1$ (retention time 5.7) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 794.25 $[M+1H]^{1+}$
m/z calculated: 794.34 $[M+1H]^{1+}$

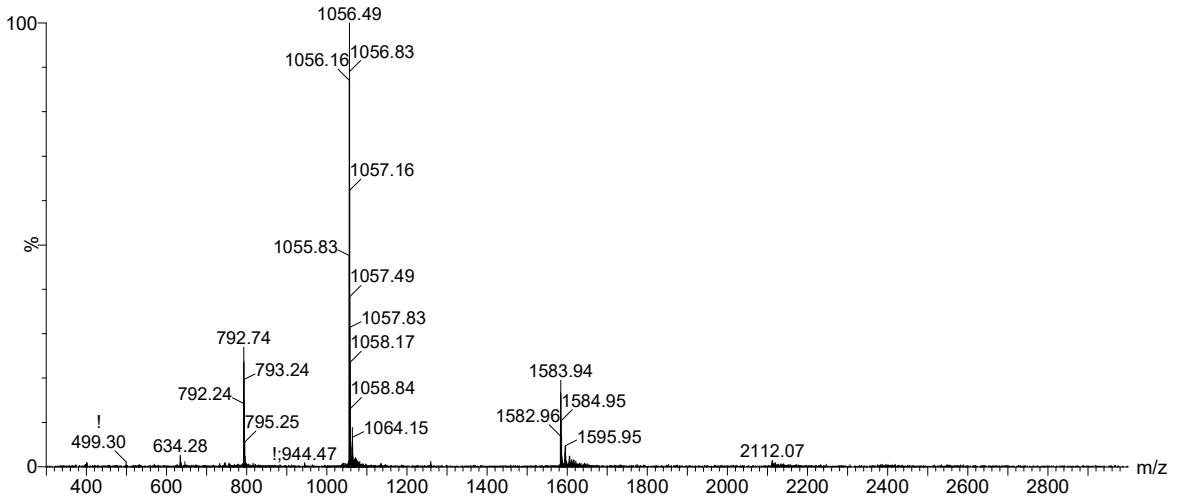


*Figure S67: Mass spectrum of (**5**)$_4$ (retention time 7.2) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 1056.49 $[M+3H]^{3+}$; 792.74 $[M+4H]^{4+}$
m/z calculated: 1056.53 $[M+3H]^{3+}$; 792.57 $[M+4H]^{4+}$

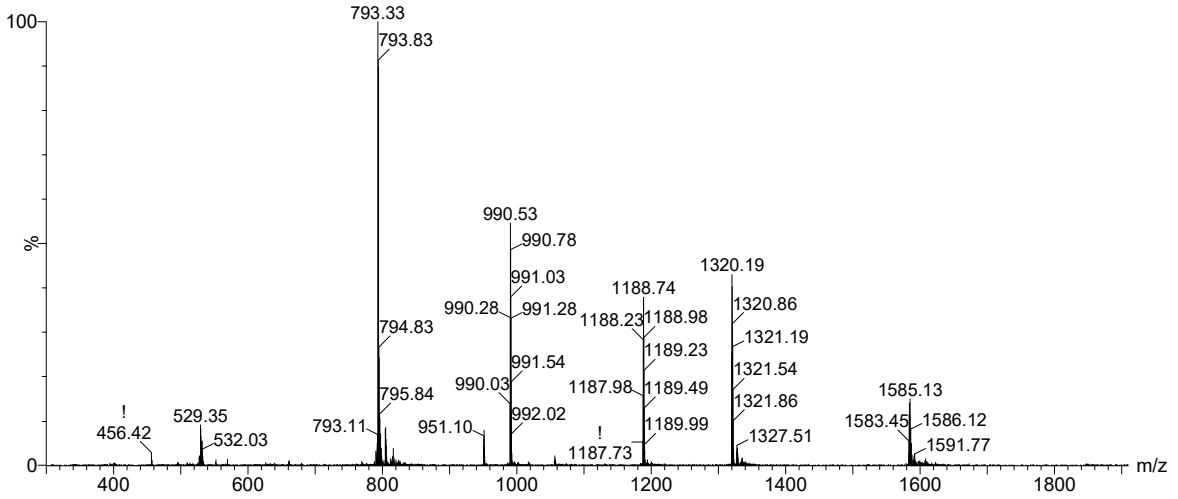


*Figure S68: Mass spectrum of (**5**)$_5$ and (**5**)$_6$ (retention time 7.5-7.7) from the LC-MS analysis of the building block after 1 day. NB. The 793 mass is shared by differently charged species*
Pentamer:
m/z observed: 1320.19 $[M+3H]^{3+}$; 990.53 $[M+4H]^{4+}$
m/z calculated: 1320.53 $[M+3H]^{3+}$; 990.65 $[M+4H]^{4+}$
Hexamer:
m/z observed: 1585.13 $[M+3H]^{3+}$; 1188.74 $[M+4H]^{4+}$
m/z calculated: 1584.30 $[M+3H]^{3+}$; 1188.48 $[M+4H]^{4+}$

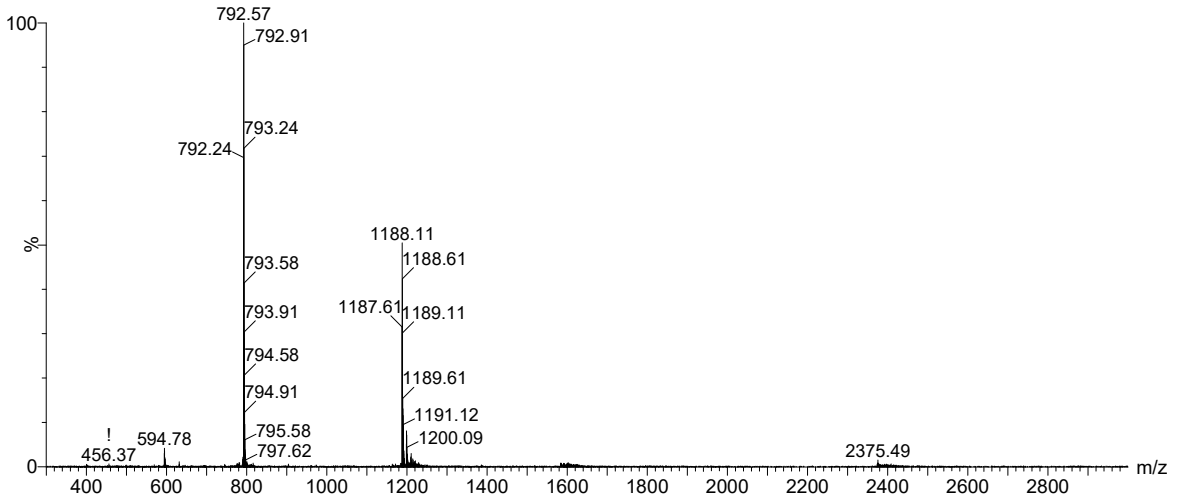


*Figure S69: Mass spectrum of (**5**)$_3$ (retention time 7.9) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 1188.11 $[M+2H]^{2+}$; 792.57 $[M+3H]^{3+}$
m/z calculated: 1188.48 $[M+2H]^{2+}$; 792.66 $[M+3H]^{3+}$

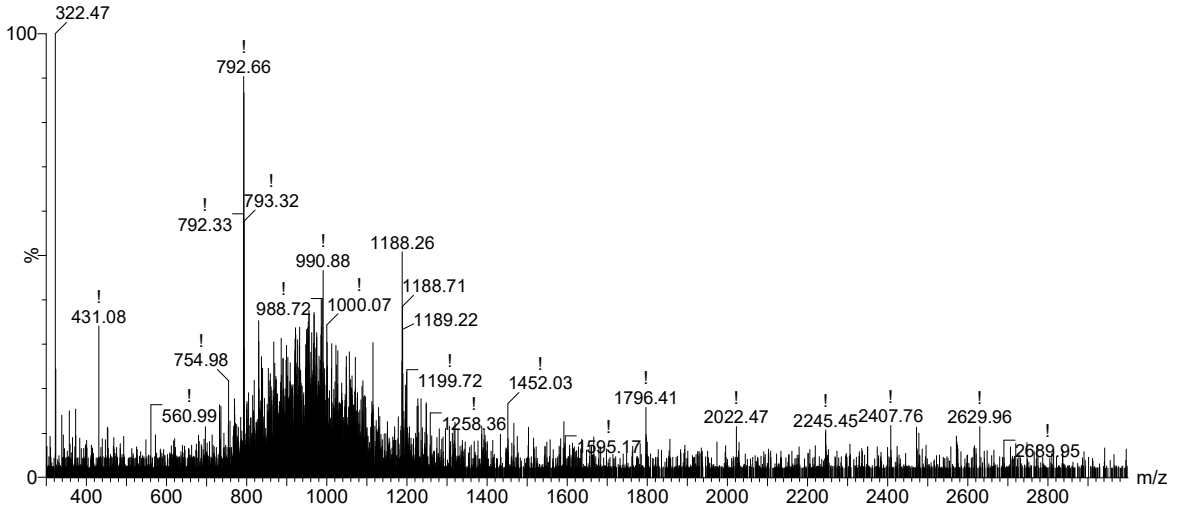


Figure S70: Mass spectrum of the unknown compound in the wash (retention time 12-14) from the LC-MS analysis of the building block after 5 days. Notice that a large amount of trimer has accumulated here.

Kinetic Plots

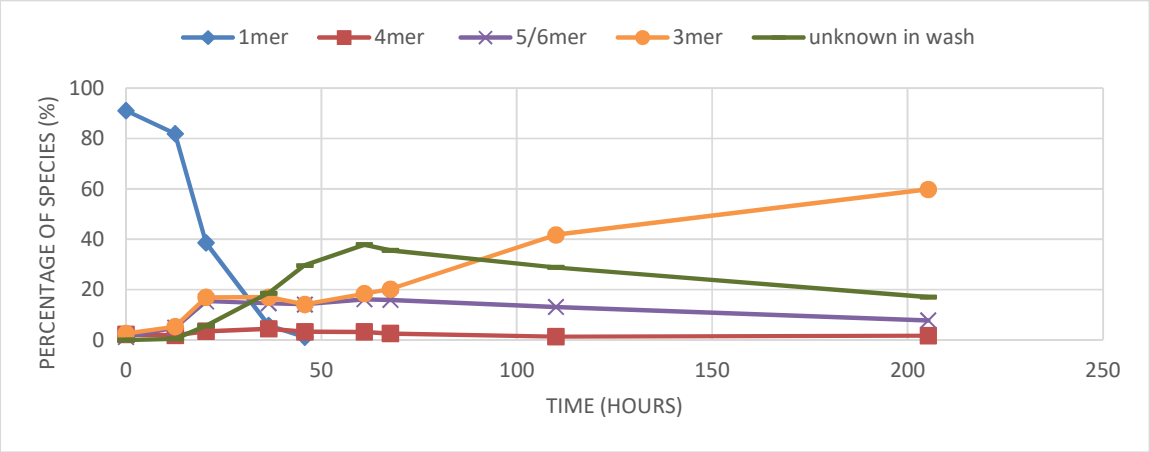


*Figure S71: Kinetic profile of a library made from **5** in borate buffer (pH 8.1) under continuous stirring at 45°C*

Seeding

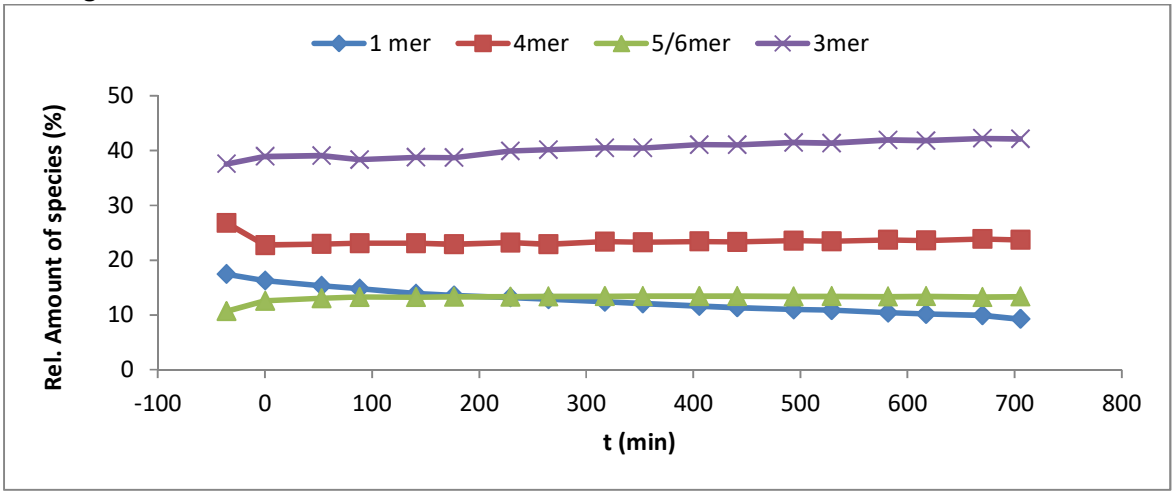


*Figure S72: Kinetic profile of a library made from an 80% oxidised library of **5** in borate buffer (pH 8.1) under continuous stirring at 40°C, upon addition of a seed of fibres of **5** at t = 0*

CD and ThT data

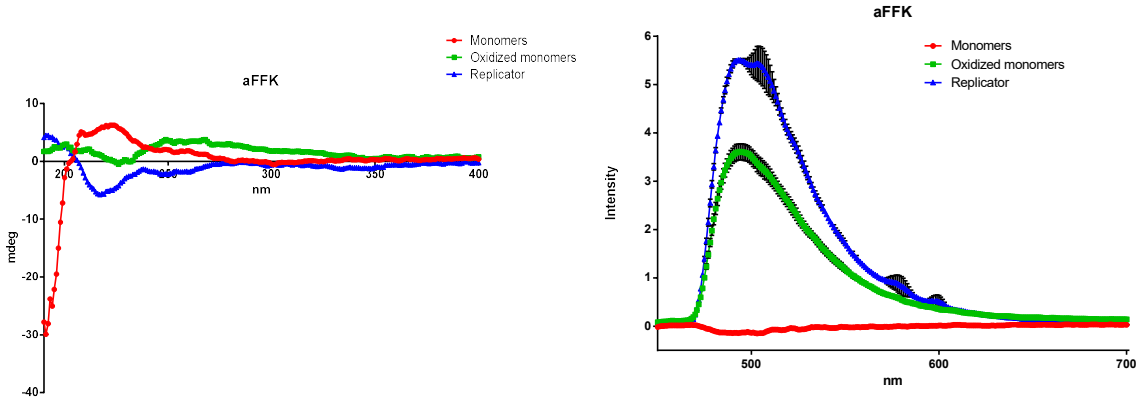


Figure S73: CD spectra (left) and UV spectra (right) of a ThT assay of XGLaFFK (in triplicates) showing samples with monomers as red circles, oxidised monomers as green squares and fibres as blue triangles; error bars are indicated in black. Samples were prepared as described in the previous section

TEM Images

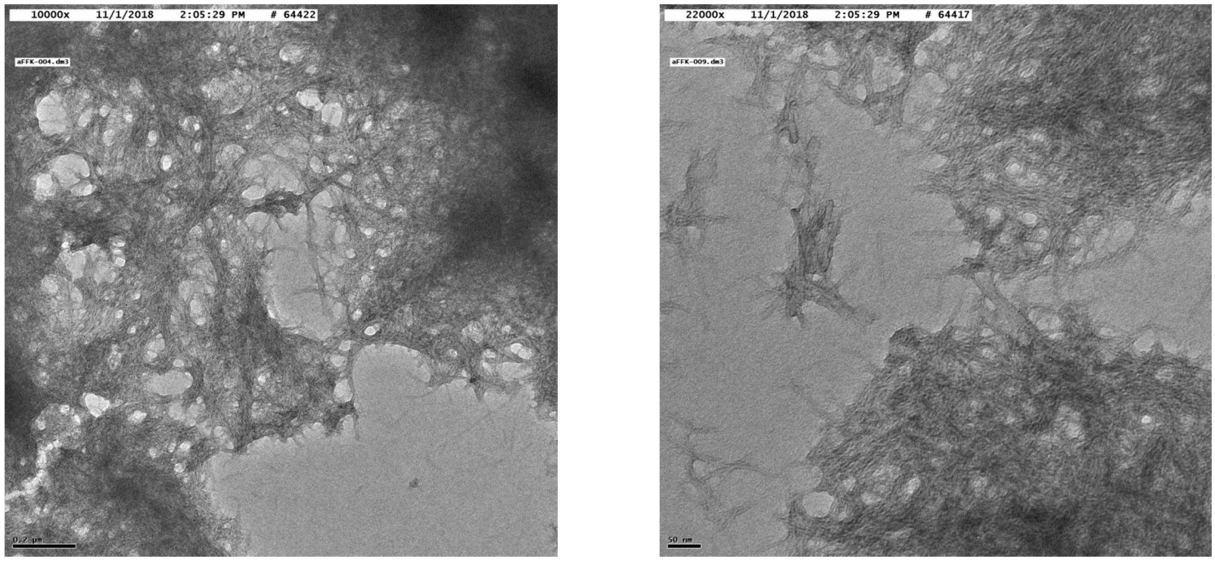


*Figure S74: Representative negative stain TEM images of fibres of **5** at 200 nm (left) and 50 nm (right) after approximately 2 weeks*

Measurement	Width (nm)	Measurement	Width (nm)
1	8.912	**16**	17.396
2	13.083	**17**	10.788
3	11.254	**18**	13.771
4	26.684	**19**	17.038
5	11.039		
6	12.845		
7	24.123		
8	15.257		
9	16.569		
10	29.347		
11	17.552		
12	40.102	Mean	18.175
13	20.2	SD	7.646
14	21.095	Min	8.912
15	18.278	Max	40.102

*Table S4: Width of fibres of **5** measured across various TEM images*

Catalysis

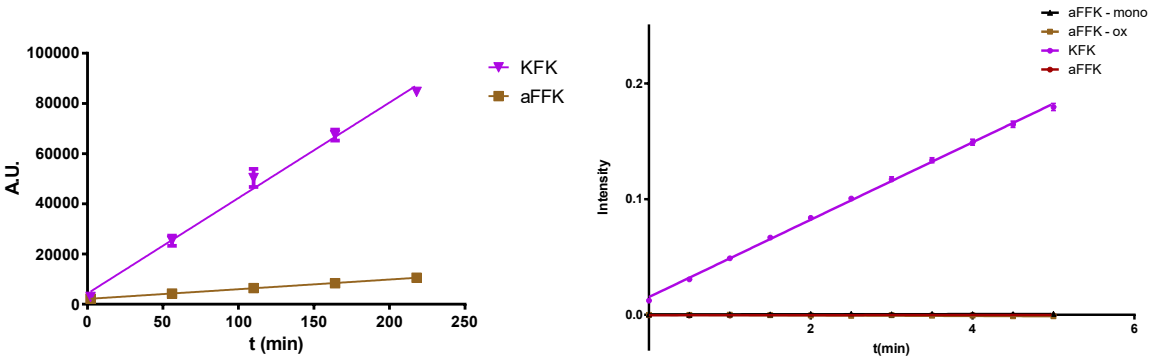
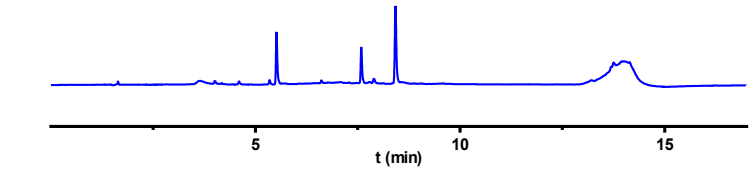


*Figure S75: Methodol breakdown (left) and Sulfmoc glycine deprotection (right) performed by **5** as discussed in the main body of the text.*

Unstirred library



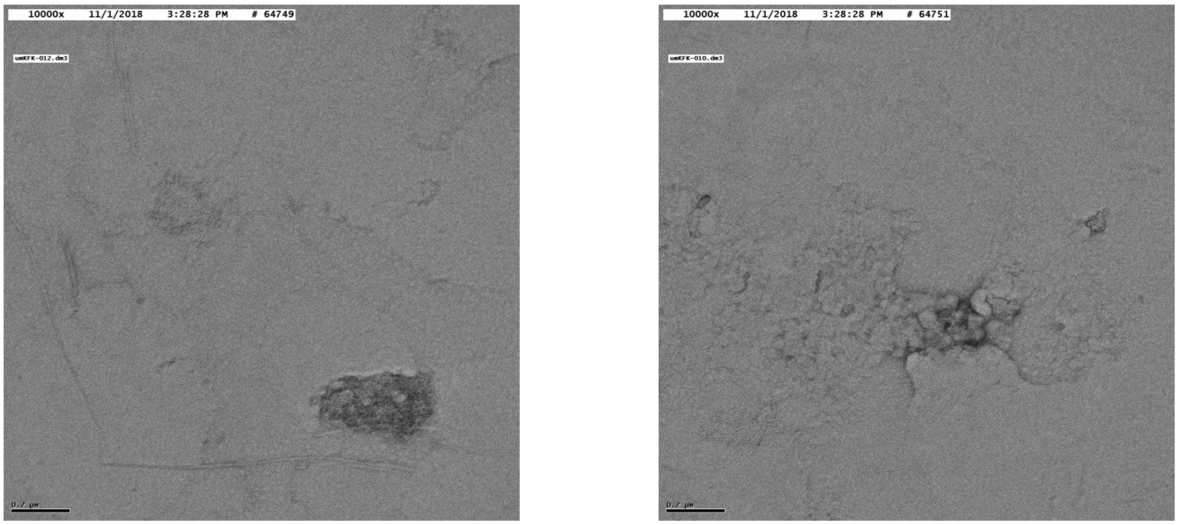
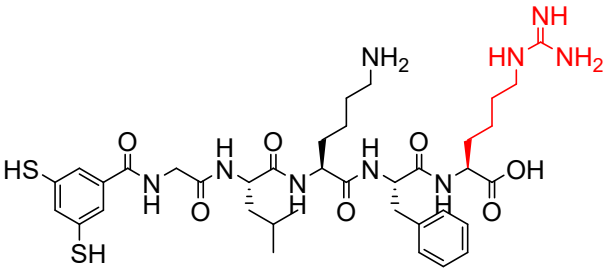


*Figure S76: Top: UPLC traces of an unstirred library of **5** after two weeks, showing a peak for over oxidised monomer, tetramer and trimer. Bottom: Representative negative stain TEM images of an unstirred library of **5** at 200 nm (left) and 50 nm (right) after approximately 2 weeks*

XGLKFhR "**6**"



UPLC traces of 6

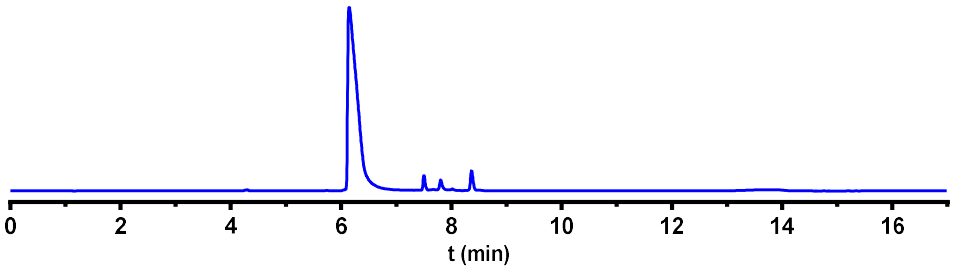


*Figure S77: UPLC chromatogram of **6** after library preparation*

Peak	Retention time	Percentage
1 Monomer	6.15	92.34
2 Tetramer	7.50	1.84
3 Linear Dimer	7.81	1.89
4 Trimer	8.37	2.88

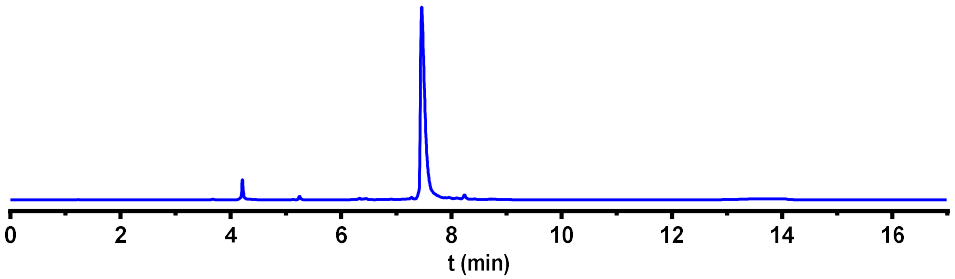


*Figure S78: UPLC chromatogram of **6** after 3 days*

Peak	Retention time	Percentage
1 Over oxidised species	4.21	3.14
2 Hexamer	7.46	91.44
3 Linear Trimer	7.95	1.01
4 Trimer	8.24	1.57

Mass analysis of 6

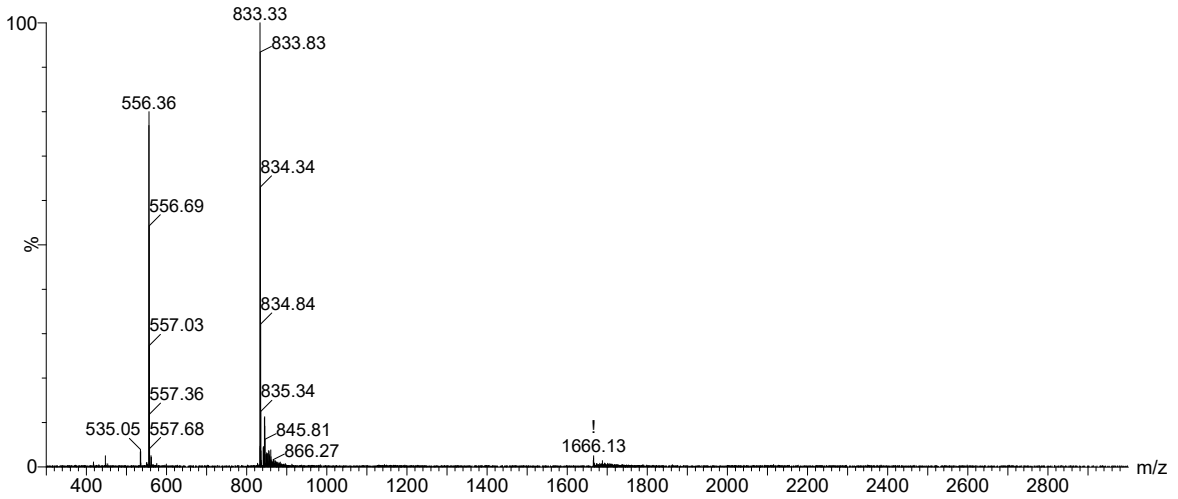


*Figure S79: Mass spectrum of over oxidised (**6**)$_2$ (retention time 4.2) from the LC-MS analysis of the building block after 5 days*
m/z observed: 833.33 $[M+4O+2H]^{2+}$; 556.36 $[M+4O+3H]^{3+}$
m/z calculated: 833.36 $[M+4O+2H]^{2+}$; 555.90 $[M+4O+3H]^{3+}$

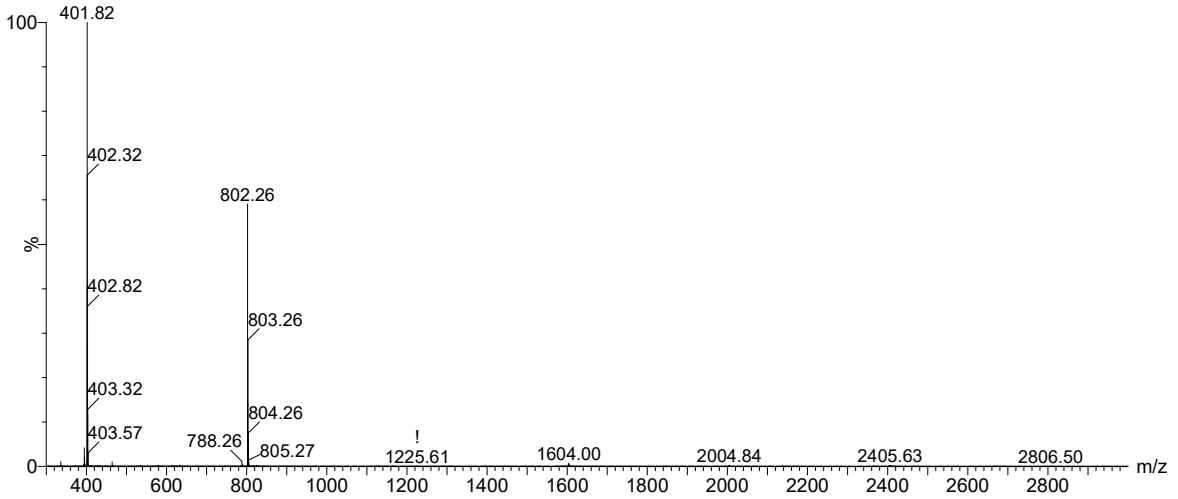


*Figure S80: Mass spectrum of (**6**)$_1$ (retention time 6.2) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 802.26 $[M+1H]^{1+}$; 401.82 $[M+2H]^{2+}$
m/z calculated: 802.37 $[M+1H]^{1+}$; 401.69 $[M+2H]^{2+}$

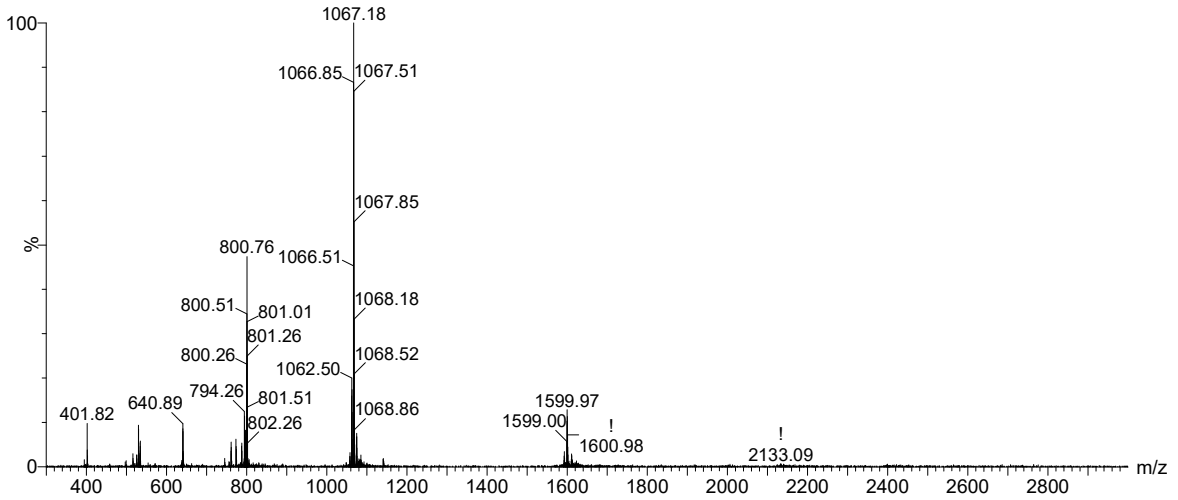


*Figure S81: Mass spectrum of (**6**)$_4$ (retention time 7.5) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 1067.18 $[M+3H]^{3+}$; 800.76 $[M+4H]^{4+}$
m/z calculated: 1067.14 $[M+3H]^{3+}$; 800.61 $[M+4H]^{4+}$

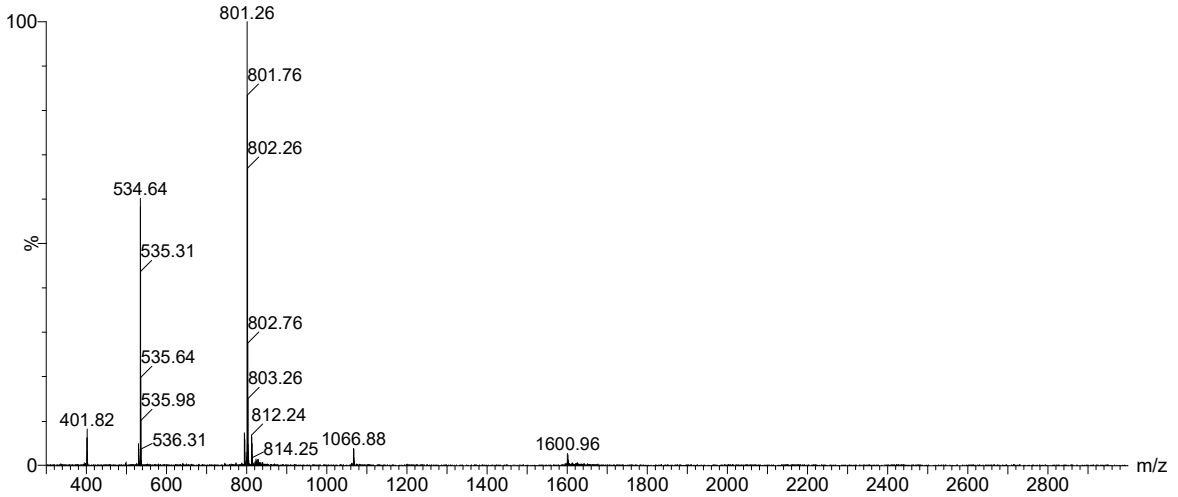


*Figure S82: Mass spectrum of (**6**)$_2$ (retention time 7.8) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 801.26 $[M+2H]^{2+}$; 534.64 $[M+3H]^{3+}$
m/z calculated: 801.37 $[M+2H]^{2+}$; 534.58 $[M+3H]^{3+}$

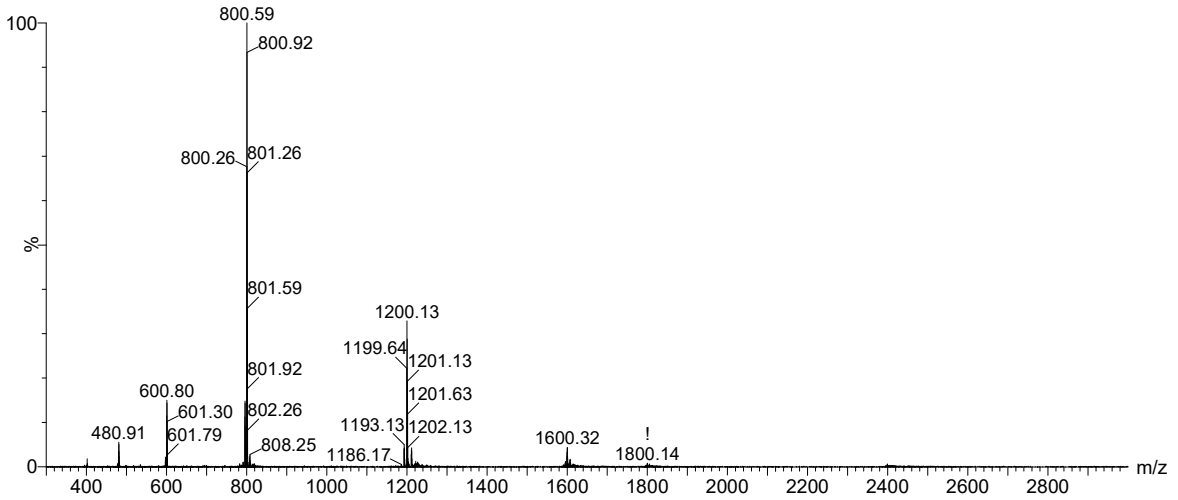


*Figure S83: Mass spectrum of (**6**)$_3$ (retention time 8.0) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 1200.13 $[M+2H]^{2+}$; 800.59 $[M+3H]^{3+}$
m/z calculated: 1200.54 $[M+2H]^{2+}$; 800.69 $[M+3H]^{3+}$

Kinetic Plots

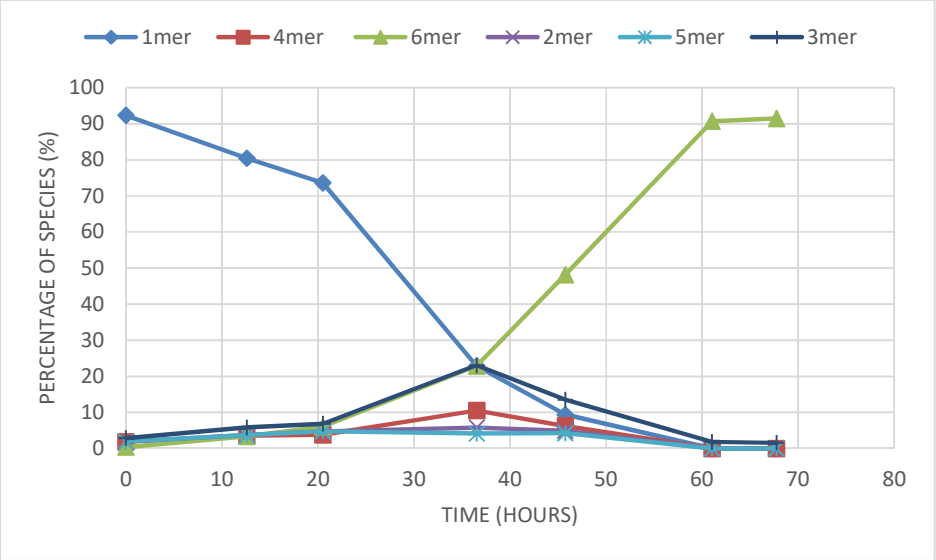


*Figure S84: Kinetic profile of a library made from **6** in borate buffer (pH 8.1) under continuous stirring at 45°C*

Seeding

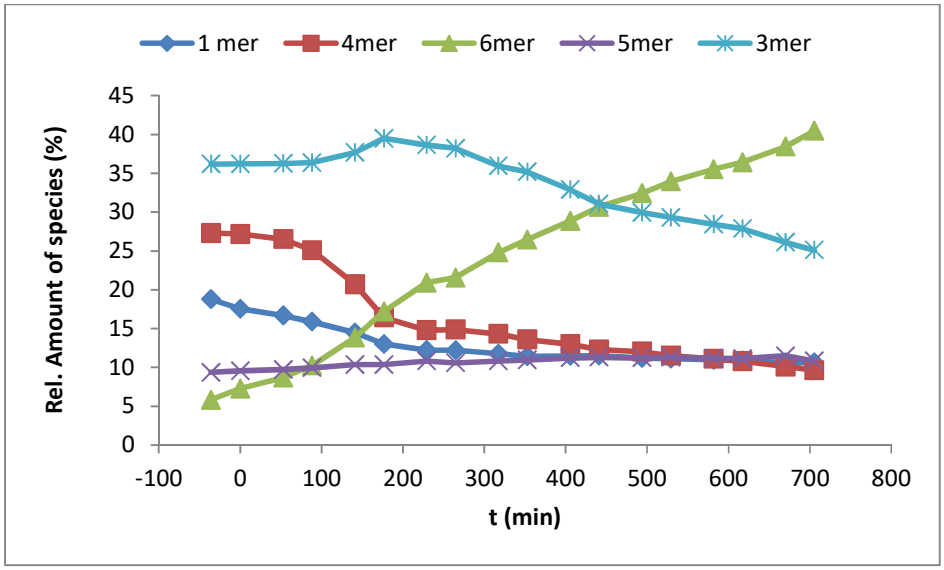


*Figure S85: Kinetic profile of a library made from an 80% oxidised library of **6** in borate buffer (pH 8.1) under continuous stirring at 40°C, upon addition of a seed of fibres of **6** at t = 0*

CD and ThT data

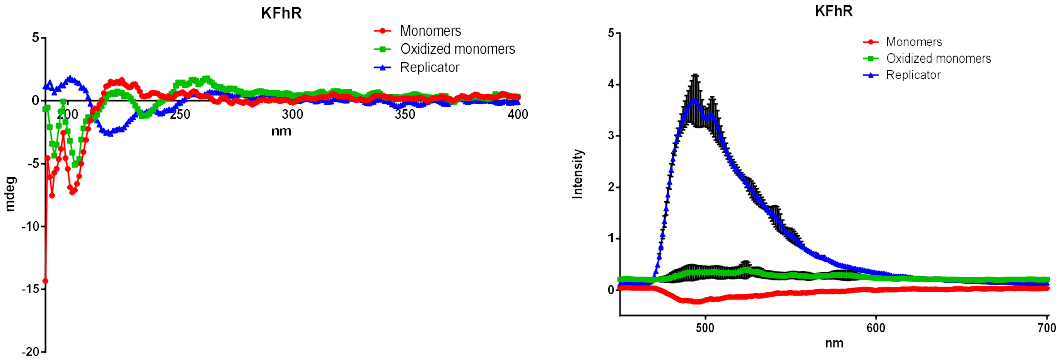


Figure S86: CD spectra (left) and UV spectra (right) of a ThT assay of XGLKFhR (in triplicates) showing samples with monomers as red circles, oxidised monomers as green squares and fibres as blue triangles; error bars are indicated in black. Samples were prepared as described in the previous section

TEM Images

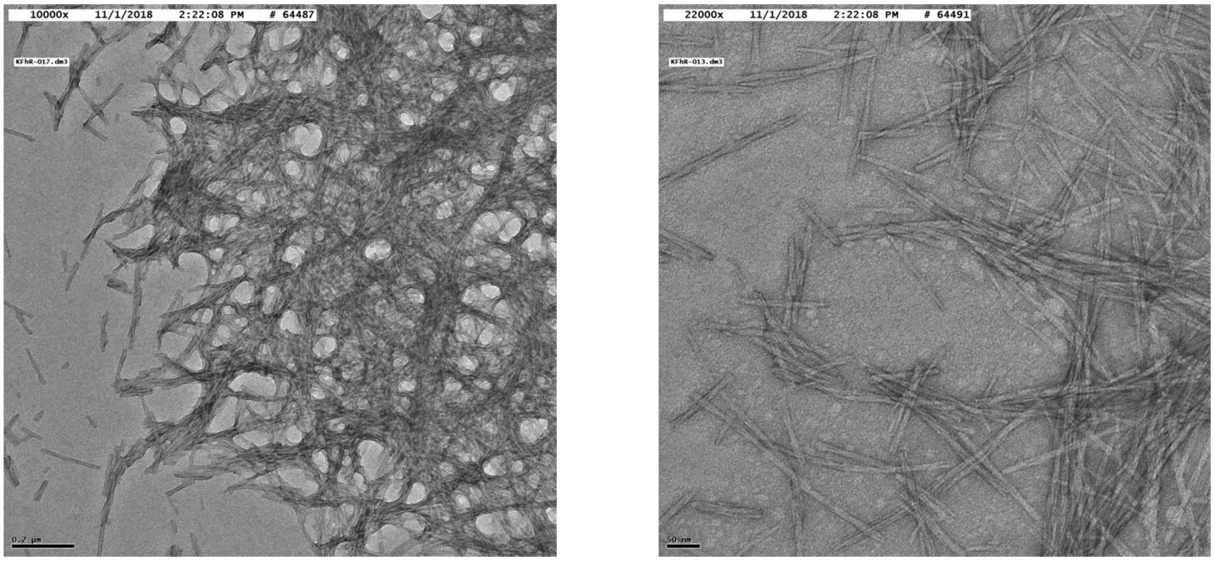


*Figure S87: Representative negative stain TEM images of fibres of **6** at 200 nm (left) and 50 nm (right) after approximately 2 weeks*

Measurement	Width (nm)	Measurement	Width (nm)
1	14.474	**16**	15.391
2	12.3	**17**	13.696
3	13.646	**18**	15.808
4	15.98	**19**	15.808
5	13.083	**20**	16.424
6	14.917	**21**	13.469
7	9.963	**22**	16.066
8	14.042	**23**	15.546
9	15.212	**24**	13.646
10	12.217	**25**	13.265
11	11.847		
12	12.951	Mean	13.801
13	13.083	SD	1.934
14	8.315	Min	8.315
15	13.87	Max	16.424

*Table S5: Width of fibres of **6** measured across various TEM images*

Catalysis

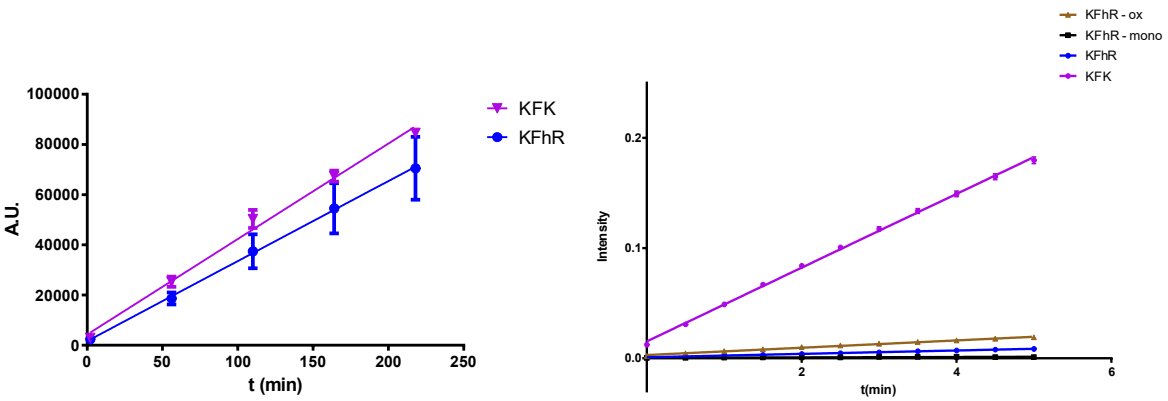


*Figure S88: Methodol breakdown (left) and Sulfmoc glycine deprotection (right) performed by **6** as discussed in the main body of the text. Notice how the oxidised monomers show more activity than the fibres.*

Unstirred library

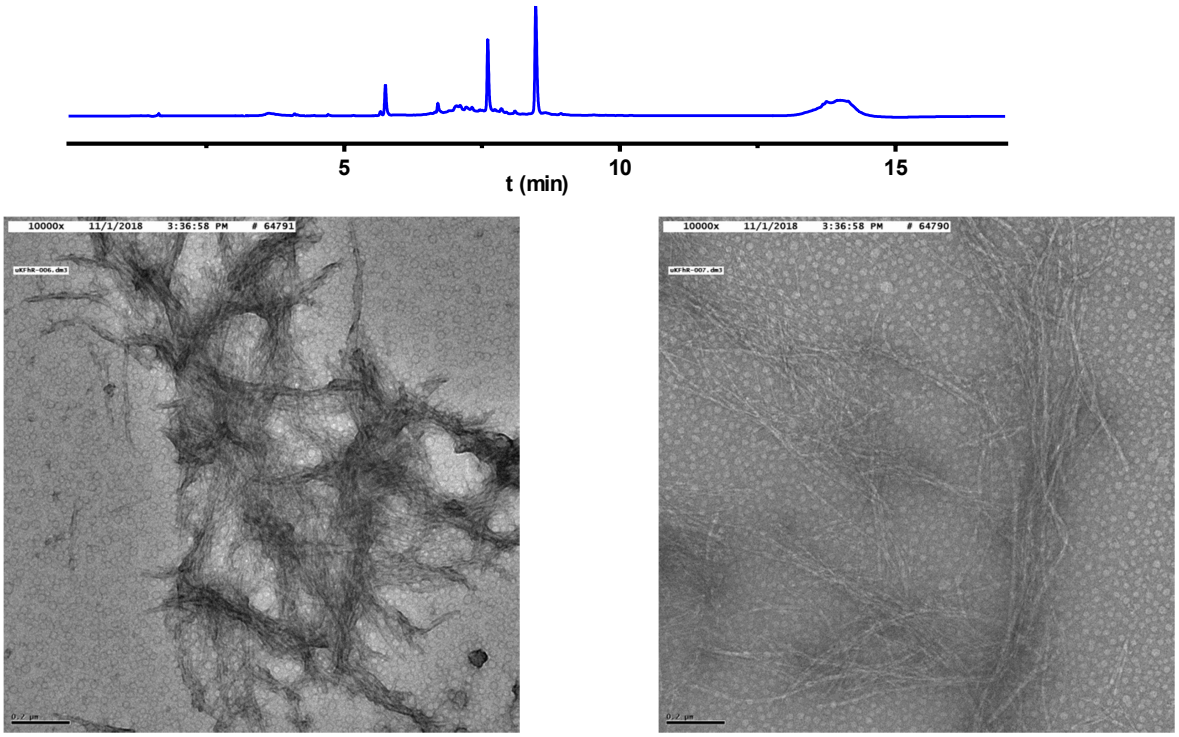
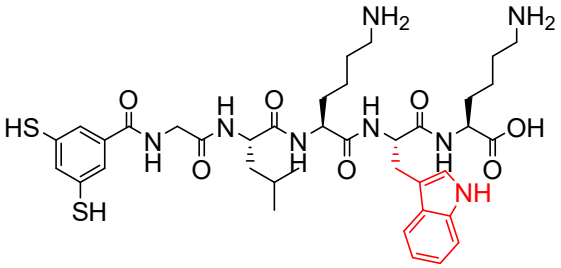


*Figure S89: Top: UPLC traces of an unstirred library of **6** after 2 weeks, showing a peak for over oxidised monomer, tetramer and trimers. Bottom: Negative stain TEM images of an unstirred library of **6** at 200 nm (left) and 50 nm (right) after approximately 2 weeks. NB. Only a very small amount of fibres was observed.*

XGLKWK "**7**"



UPLC traces of 7

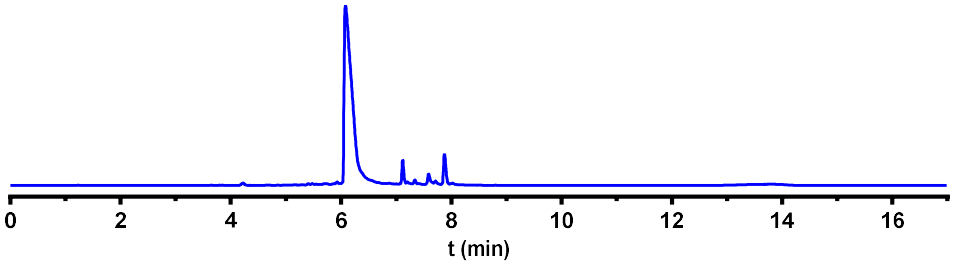


*Figure S90: UPLC chromatogram of **7** after library preparation*

Peak	Retention time	Percentage
1 Monomer	6.08	85.00
2 Tetramer	7.12	3.04
3 Linear Dimer	7.59	1.94
4 Trimer	7.88	4.51

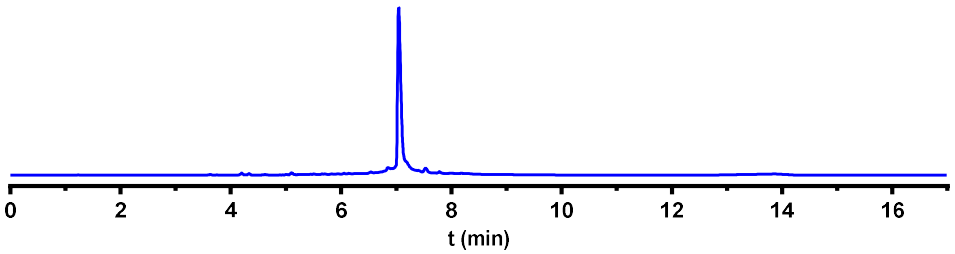


*Figure S91: UPLC chromatogram of **7** after 3 days*

Peak	Retention time	Percentage
1 Hexamer	7.05	82.94
3 Linear Dimer	7.53	5.03
4 Trimer	7.78	1.83

Mass analysis of 7

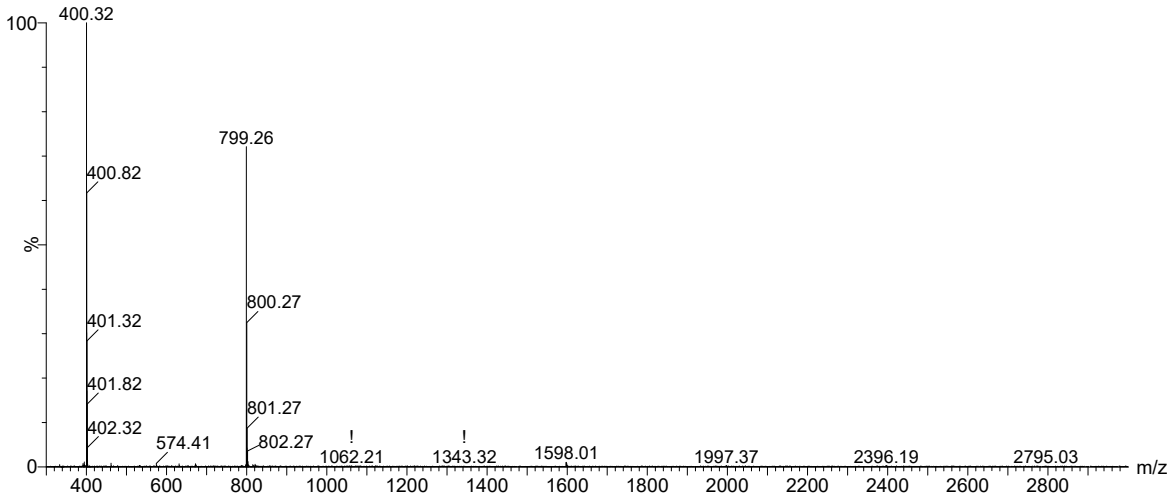


*Figure S92: Mass spectrum of (**7**)$_1$ (retention time 6.1) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 799.26 $[M+1H]^{1+}$; 400.32 $[M+2H]^{2+}$
m/z calculated: 799.36 $[M+1H]^{1+}$; 400.19 $[M+2H]^{2+}$

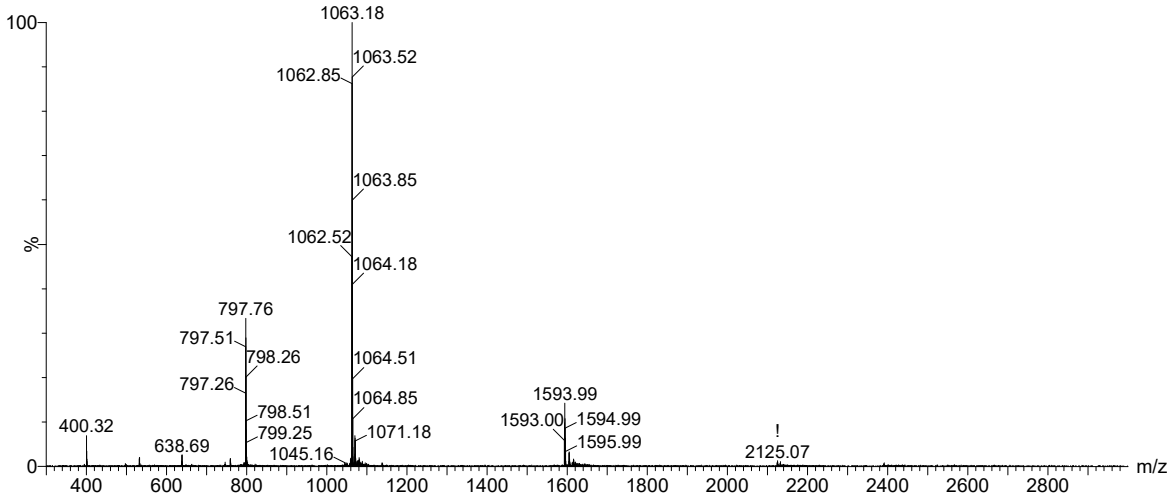


*Figure S93: Mass spectrum of (**7**)$_4$ (retention time 7.1) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 1063.18 $[M+3H]^{3+}$; 797.76 $[M+4H]^{4+}$
m/z calculated: 1063.13 $[M+3H]^{3+}$; 797.60 $[M+4H]^{4+}$

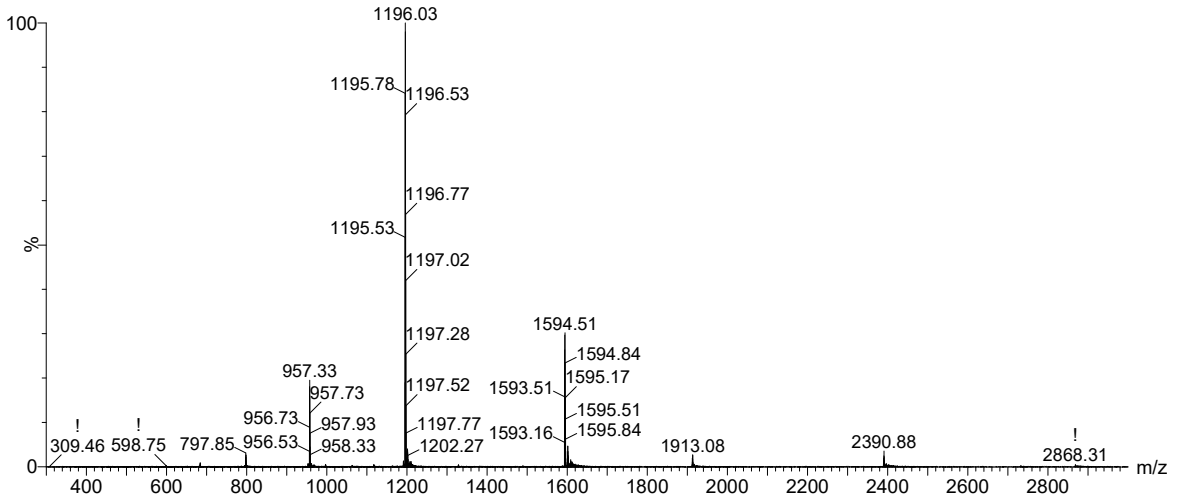


*Figure S94: Mass spectrum of (**7**)$_6$ (retention time 7.1) from the LC-MS analysis of the building block after 5 days.*
m/z observed: 1594.51 $[M+3H]^{3+}$; 1196.03 $[M+4H]^{4+}$
m/z calculated: 1594.36 $[M+3H]^{3+}$; 1196.02 $[M+4H]^{4+}$

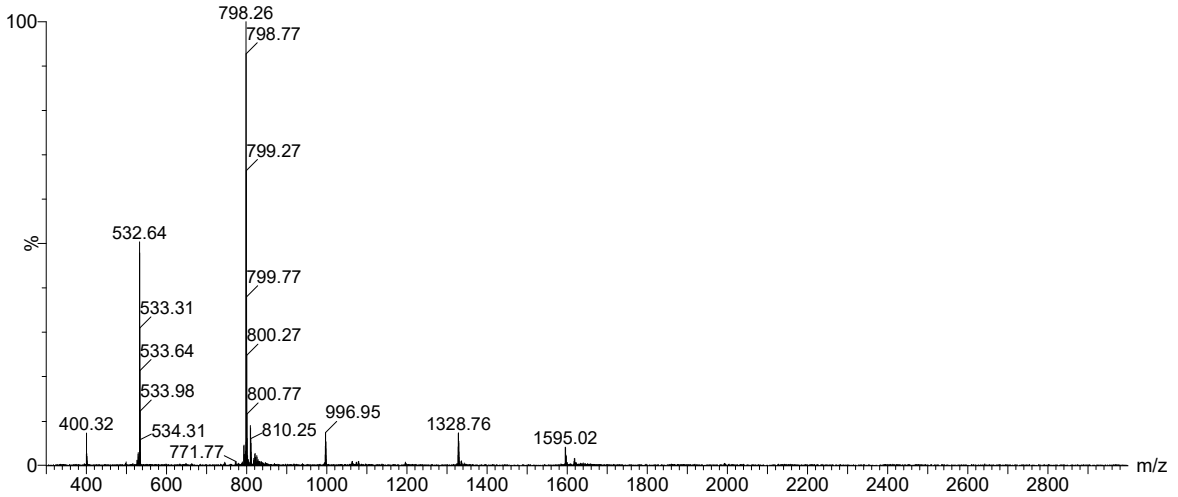


*Figure S95: Mass spectrum of (**7**)$_2$ (retention time 7.6) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 798.26 $[M+2H]^{2+}$; 532.64 $[M+3H]^{3+}$
m/z calculated: 798.35 $[M+2H]^{2+}$; 532.57 $[M+3H]^{3+}$

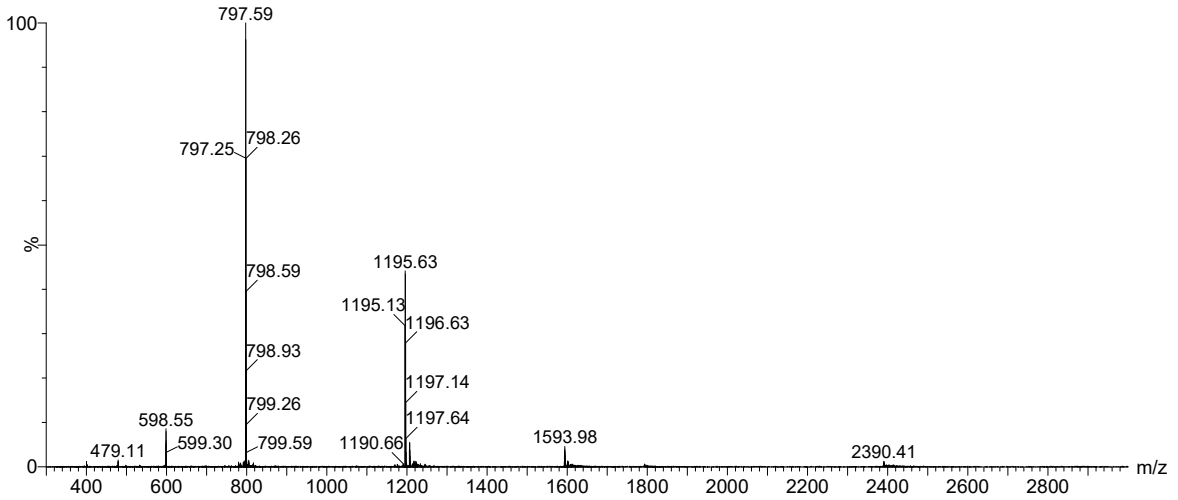


*Figure S96: Mass spectrum of (**7**)$_3$ (retention time 7.8) from the LC-MS analysis of the building block after library preparation.*
m/z observed: 1195.63 $[M+2H]^{2+}$; 797.59 $[M+3H]^{3+}$
m/z calculated: 1196.02 $[M+2H]^{2+}$; 797.68 $[M+3H]^{3+}$

Kinetic Plots

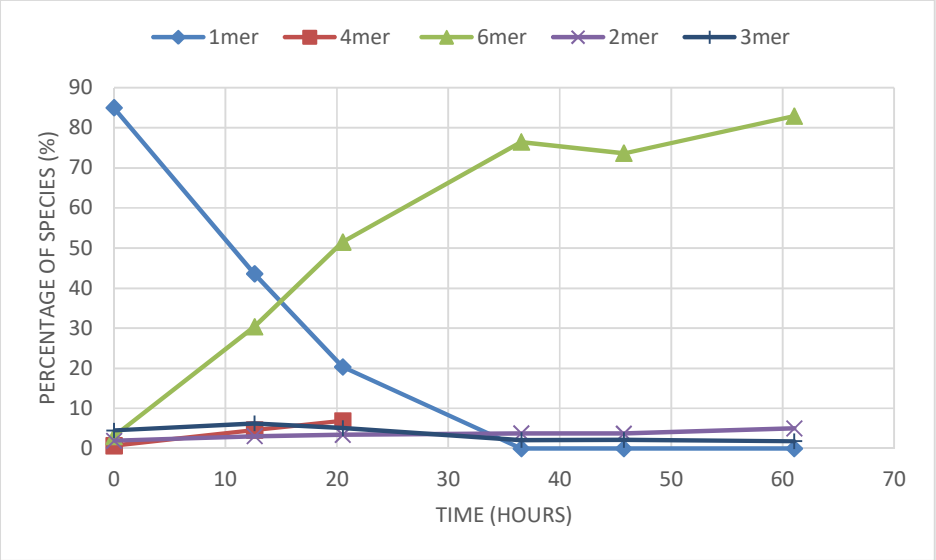


*Figure S97: Kinetic profile of a library made from **7** in borate buffer (pH 8.1) under continuous stirring at 45°C. NB. Tetramer co-elutes with hexamer.*

Seeding

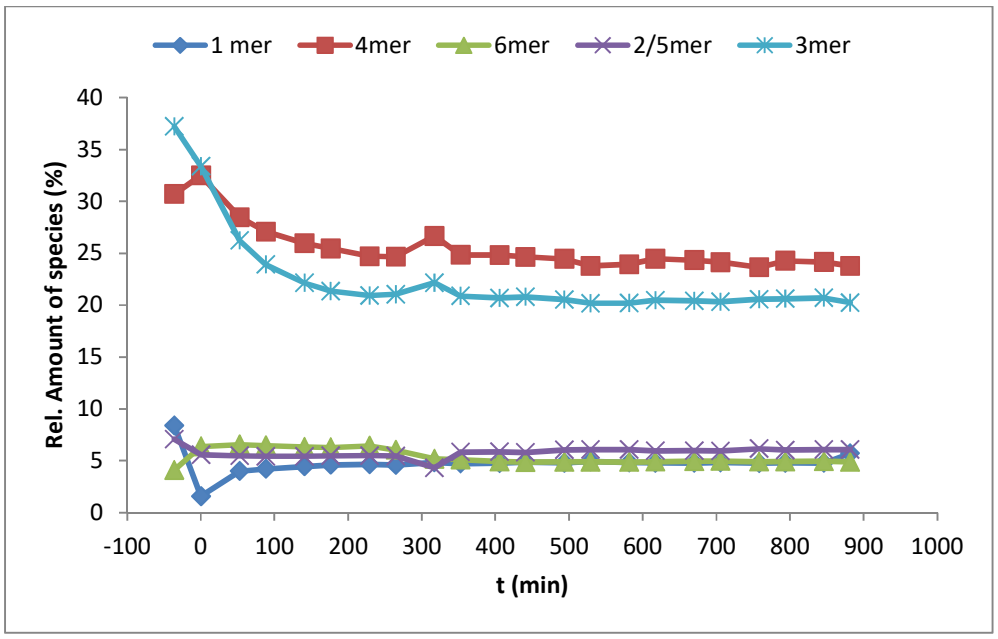


*Figure S98: Kinetic profile of a library made from an 80% oxidised library of **7** in borate buffer (pH 8.1) under continuous stirring at 40°C, upon addition of a seed of fibres of **7** at t = 0*

CD and ThT data

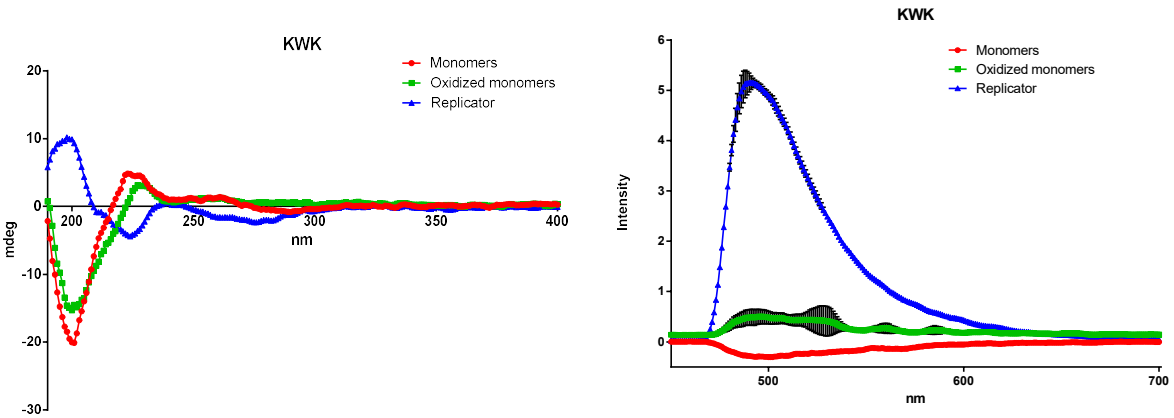


Figure S99: CD spectra (left) and UV spectra (right) of a ThT assay of XGLKWK (in triplicates) showing samples with monomers as red circles, oxidised monomers as green squares and fibres as blue triangles; error bars are indicated in black. Samples were prepared as described in the previous section

TEM Images

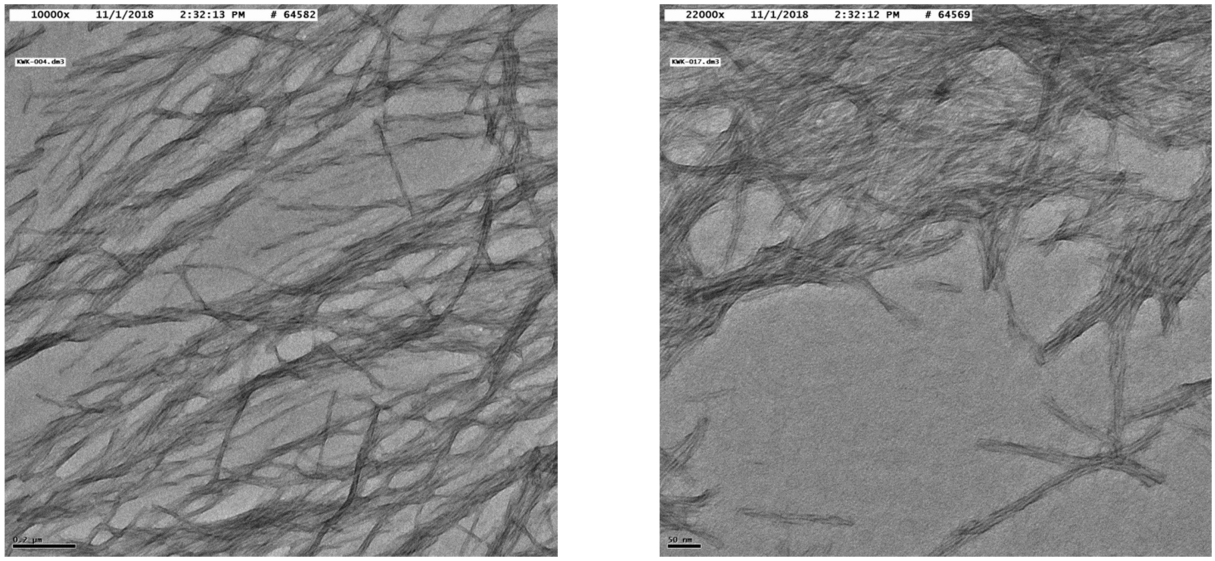


*Figure S100: Representative negative stain TEM images of fibres of **7** at 200 nm (left) and 50 nm (right) after approximately 2 weeks*

Measurement	width (nm)	Measurement	width (nm)
1	18.868	**17**	16.631
2	24.795	**18**	19.102
3	15.98	**19**	16.734
4	27.405	**20**	16.548
5	40.704	**21**	14.09
6	19.528	**22**	19.102
7	19.927	**23**	16.256
8	15.457	**24**	21.719
9	11.76	**25**	17.396
10	13.239	**26**	22.356
11	10.596	**27**	18.594
12	15.391	**28**	26.387
13	25.968	Mean	19.349
14	16.548	SD	5.965
15	17.688	Min	10.596
16	22.99	Max	40.704

*Table S6: Width of fibres of **7** measured across various TEM images*

Catalysis

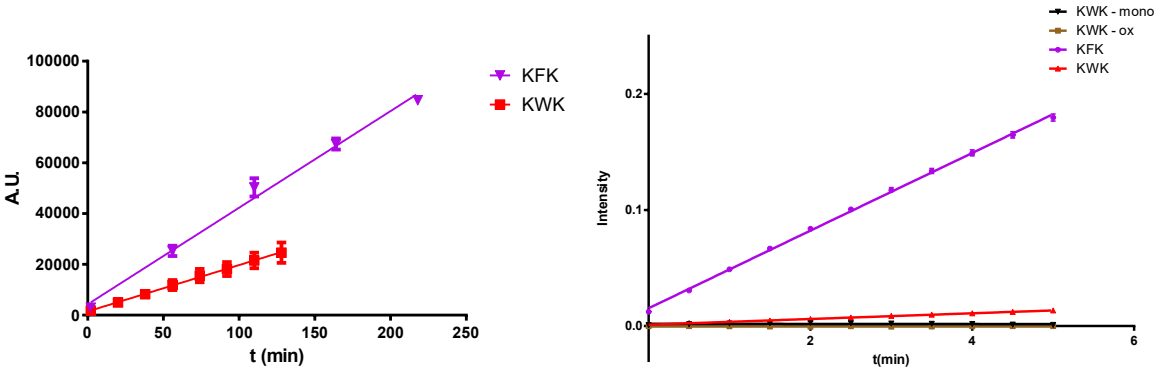


*Figure S101: Methodol breakdown (left) and Sulfmoc glycine deprotection (right) performed by **7** as discussed in the main body of the text*

Unstirred library

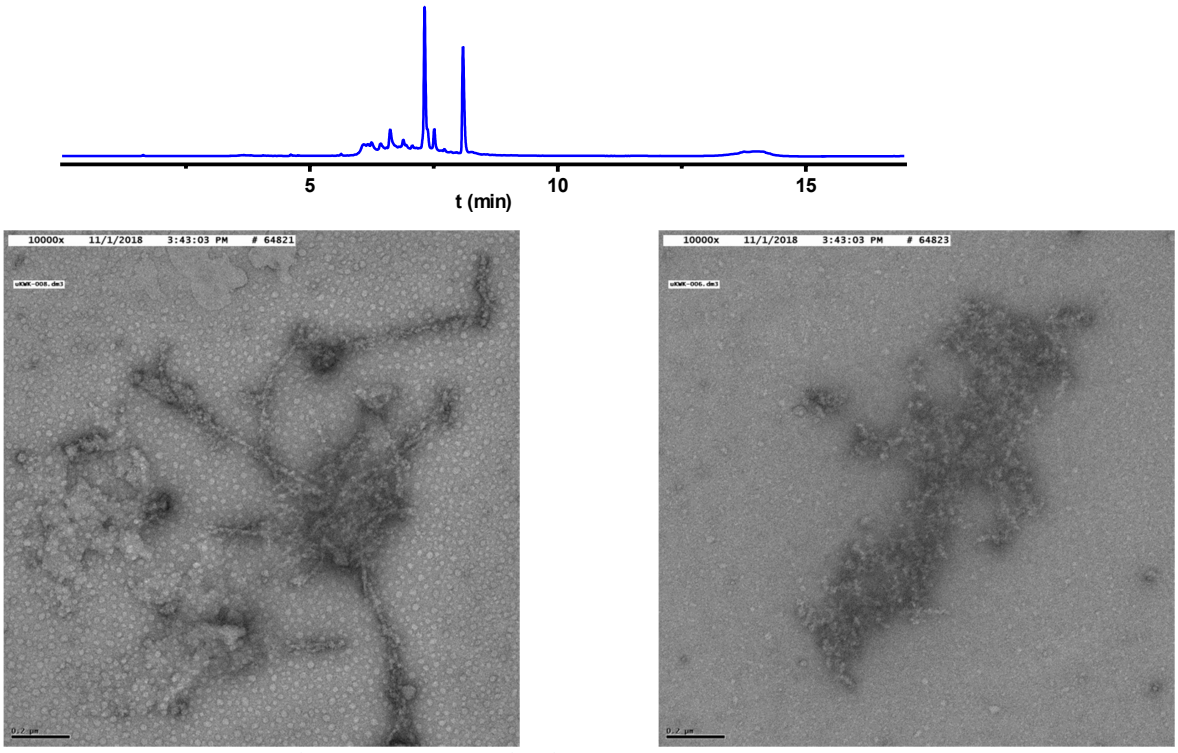
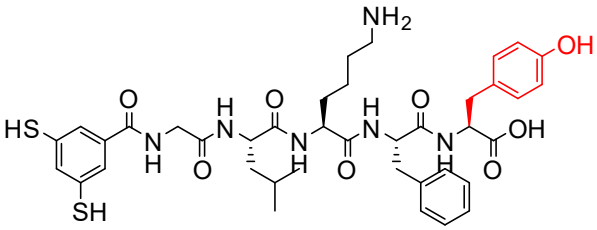


*Figure S102: Top: UPLC trace of an unstirred library of **7** after 2 weeks, showing a tetramer and trimer peak. Bottom: Representative negative stain TEM images of an unstirred library of **7** at 200 nm (left) and 50 nm (right) after approximately 2 weeks*

.

XGLKFY "**8**"



UPLC Trace

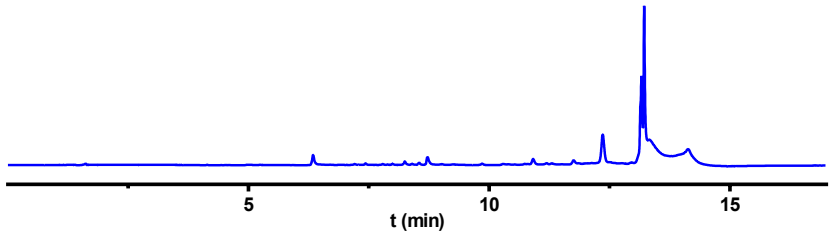


*Figure S103: UPLC chromatogram of **7** after 3 days*

TEM Images

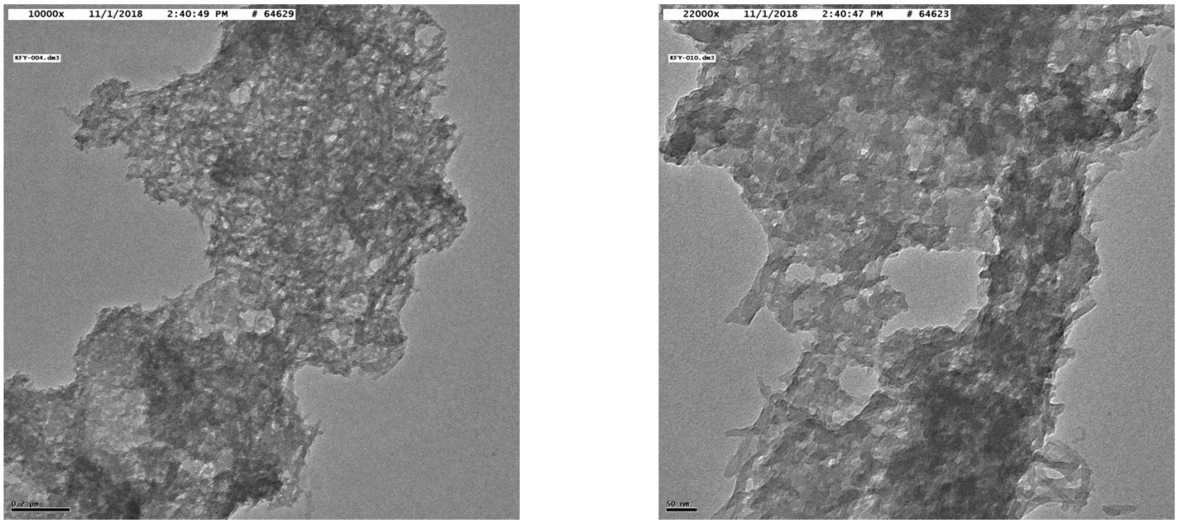
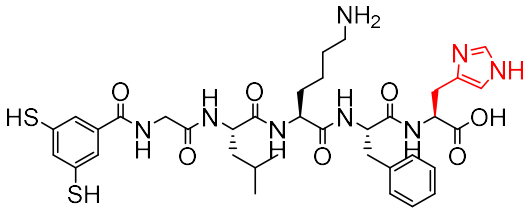


*Figure S104: Representative negative stain TEM images of the aggregate formed by **8** at 200 nm (left) and 50 nm (right) after approximately 2 weeks*

XGLKFH "**9**"



UPLC Traces

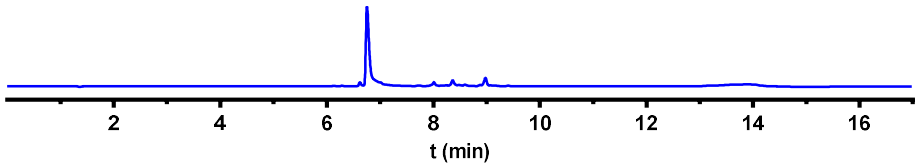


*Figure S105: UPLC chromatogram of **9** after library preparation*

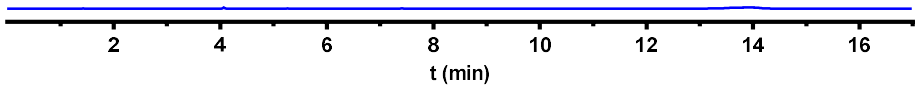
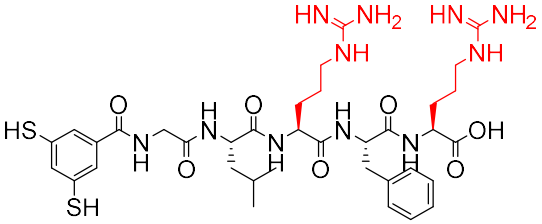


*Figure S106: UPLC chromatogram of **9** after 30 hours. All peak area is lost.*

XGLRFR "**10**"



UPLC Trace

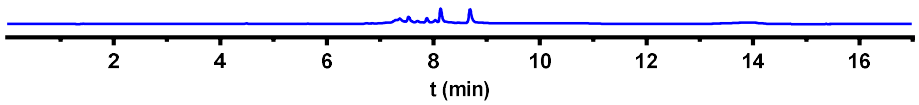


*Figure S107: UPLC chromatogram of **10** after 3 days, the chromatogram has a low total peak area, and is scaled the same as the other chromatograms.*

XGLKFK/XGLKFR "**11**"

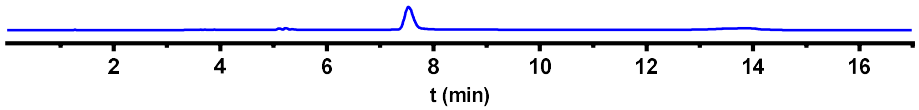


*Figure S108: UPLC chromatogram of **11** after ~2 months*

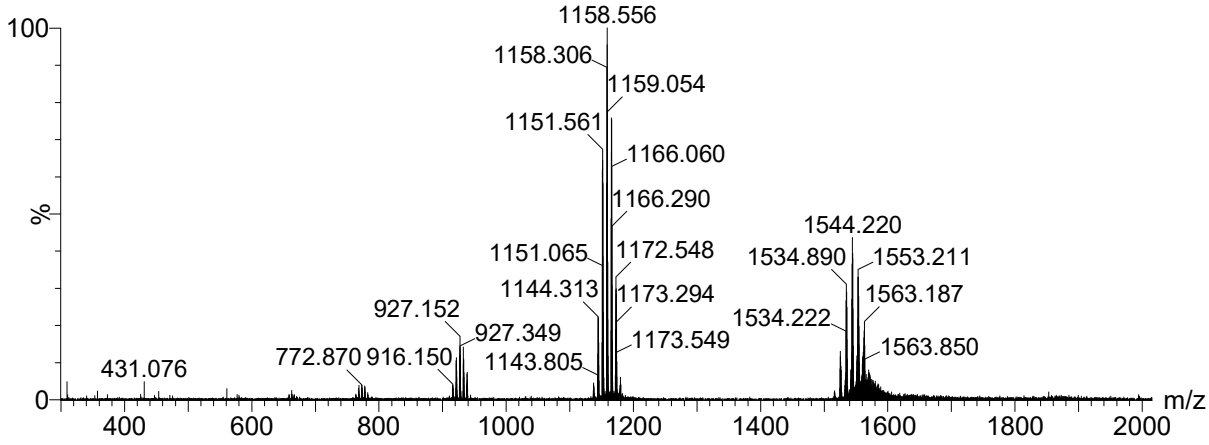


*Figure S109: Mass spectrum of (**11**)$_6$ (retention time 7.5) from the LC-MS analysis of the building block after 2 months.*
m/z observed: 1137.31 $[M+4H]^{4+}$ to 1179.55$[M+4H]^{4+}$
m/z calculated: 1137.50 $[M+4H]^{4+}$ to 1179.51 $[M+4H]^{4+}$

XGLKFK/XGLKFhR "**12**"

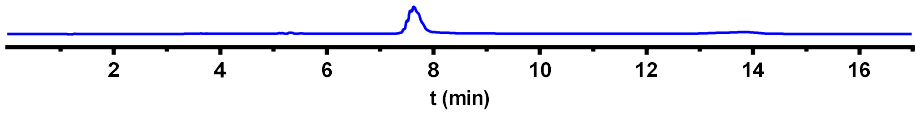


*Figure S110: UPLC chromatogram of **12** after ~2 months*

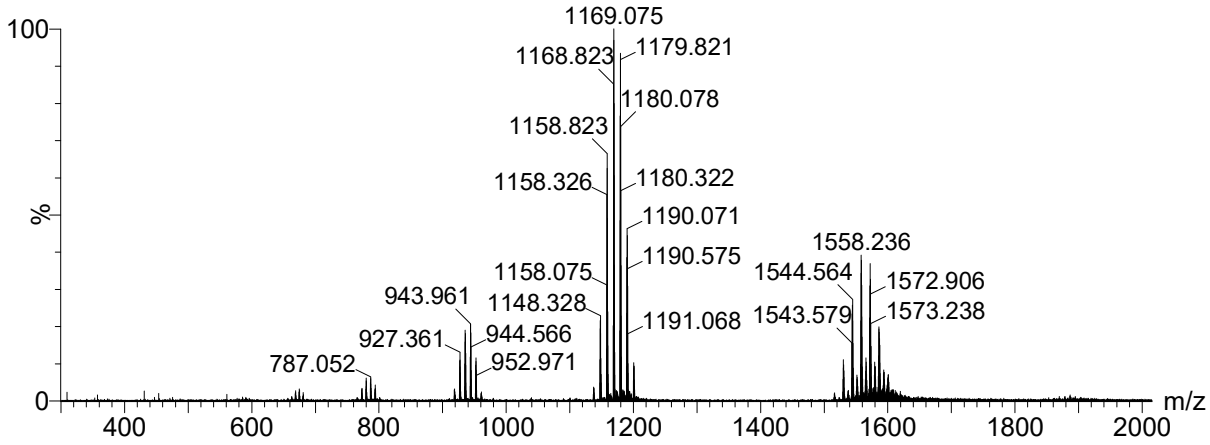


*Figure S111: Mass spectrum of (**12**)$_6$ (retention time 7.6) from the LC-MS analysis of the building block after 2 months.*
m/z observed: 1137.57 $[M+4H]^{4+}$ to 1200.57$[M+4H]^{4+}$
m/z calculated: 1137.50 $[M+4H]^{4+}$ to 1200.54 $[M+4H]^{4+}$

XGLKFK/XGLRFK "**13**"

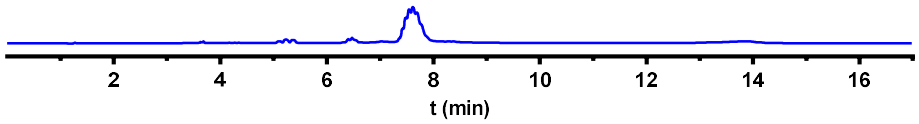


*Figure S112: UPLC chromatogram of **13** after ~2 months*

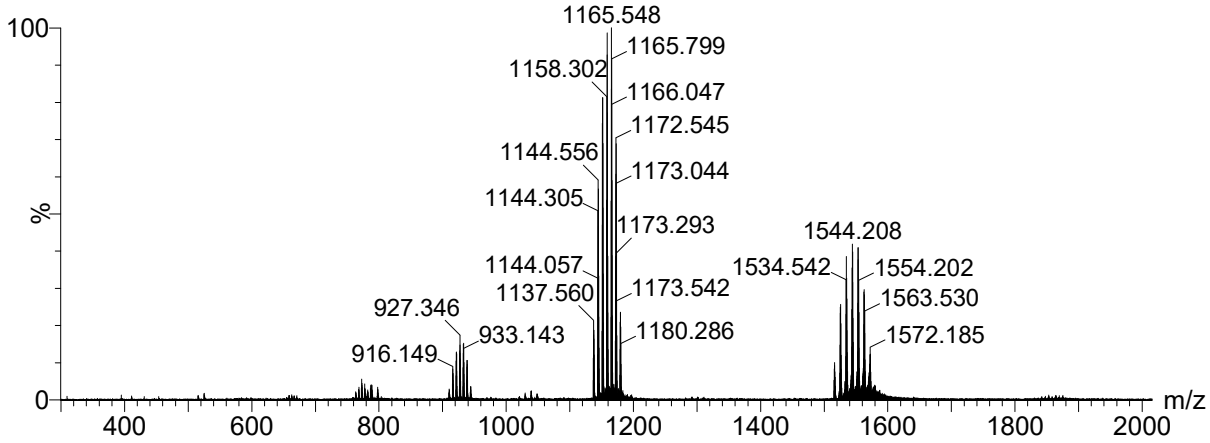


Figure S113: Mass spectrum of $(\mathbf{13})_6$ (retention time 7.5) from the LC-MS analysis of the building block after 2 months.
m/z observed: 1137.56 $[M+4H]^{4+}$ to 1179.79 $[M+4H]^{4+}$
m/z calculated: 1137.50 $[M+4H]^{4+}$ to 1179.51 $[M+4H]^{4+}$

XGLKFK/XGLOFK "**14**"

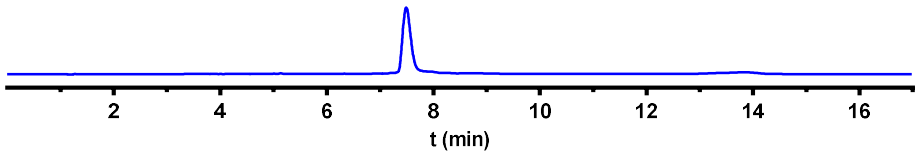


*Figure S114: UPLC chromatogram of **14** after ~2 months*

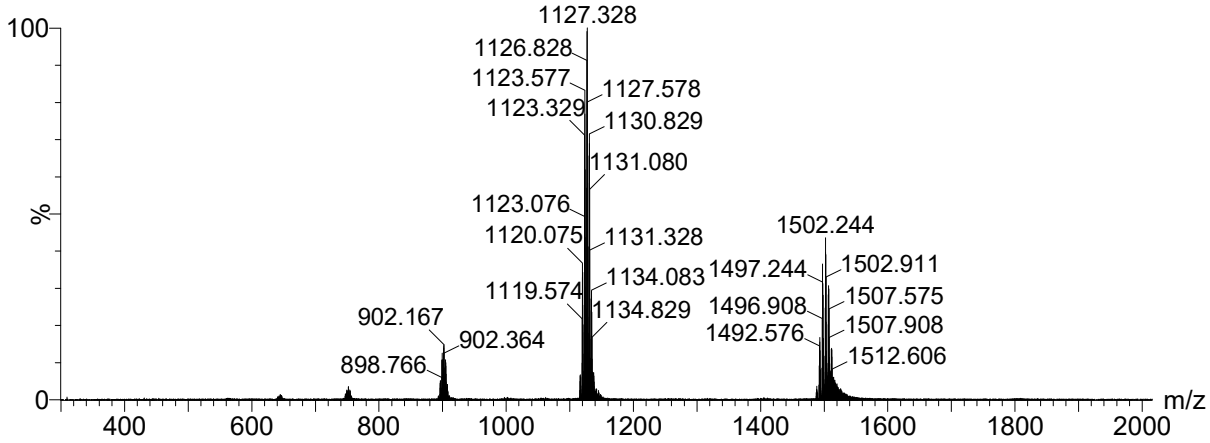


Figure S115: Mass spectrum of $(\mathbf{14})_6$ (retention time 7.46) from the LC-MS analysis of the building block after 2 months.
m/z observed: 1116.32 $[M+4H]^{4+}$ to 1137.58 $[M+4H]^{4+}$
m/z calculated: 1116.48 $[M+4H]^{4+}$ to 1137.50 $[M+4H]^{4+}$

XGLKFK/XGLKWK "**15**"

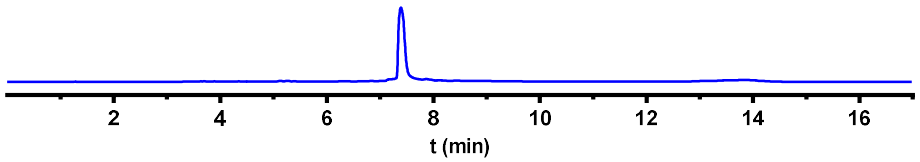


*Figure S116: UPLC chromatogram of **15** after ~2 months*

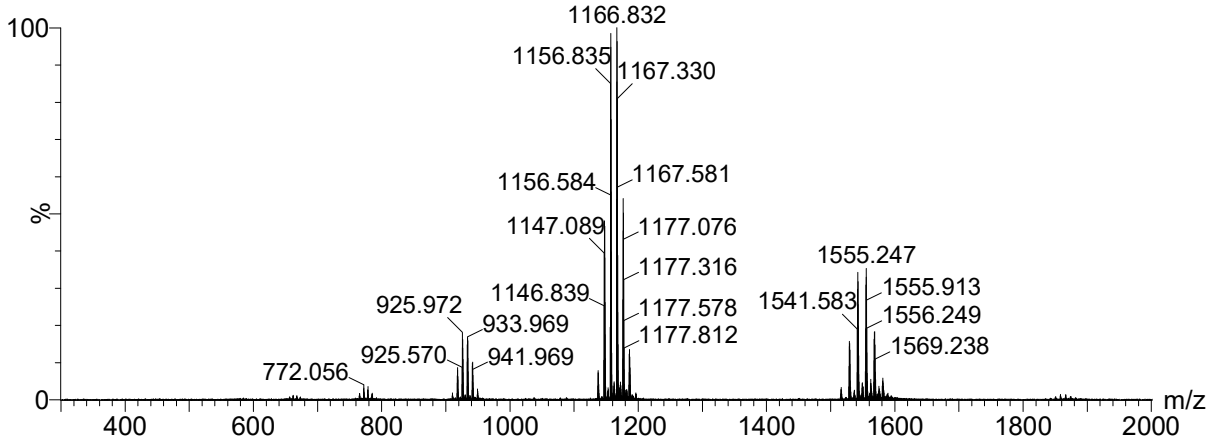


Figure S117: Mass spectrum of $(\mathbf{15})_6$ (retention time 7.36) from the LC-MS analysis of the building block after 2 months.
m/z observed: 1137.59 $[M+4H]^{4+}$ to 1196.32 $[M+4H]^{4+}$
m/z calculated: 1137.50 $[M+4H]^{4+}$ to 1196.02 $[M+4H]^{4+}$

XGLRFK/XGLKFR "**16**"

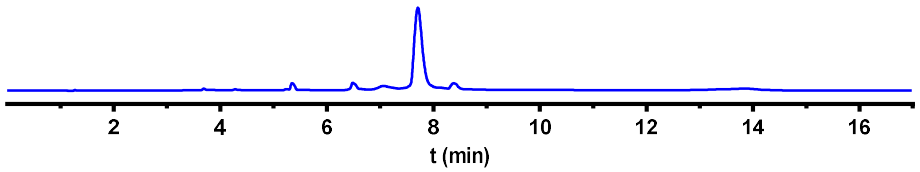


*Figure S118: UPLC chromatogram of **16** after ~2 months*

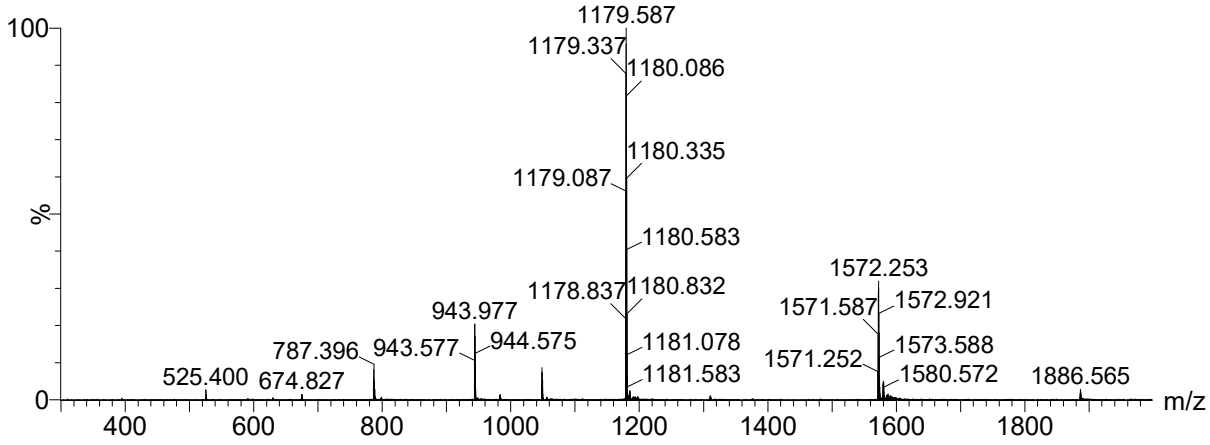


Figure S119: Mass spectrum of $(\mathbf{16})_6$ (retention time 7.68) from the LC-MS analysis of the building block after 2 months.
m/z observed: 1179.58 $[M+4H]^{4+}$
m/z calculated: 1179.51 $[M+4H]^{4+}$

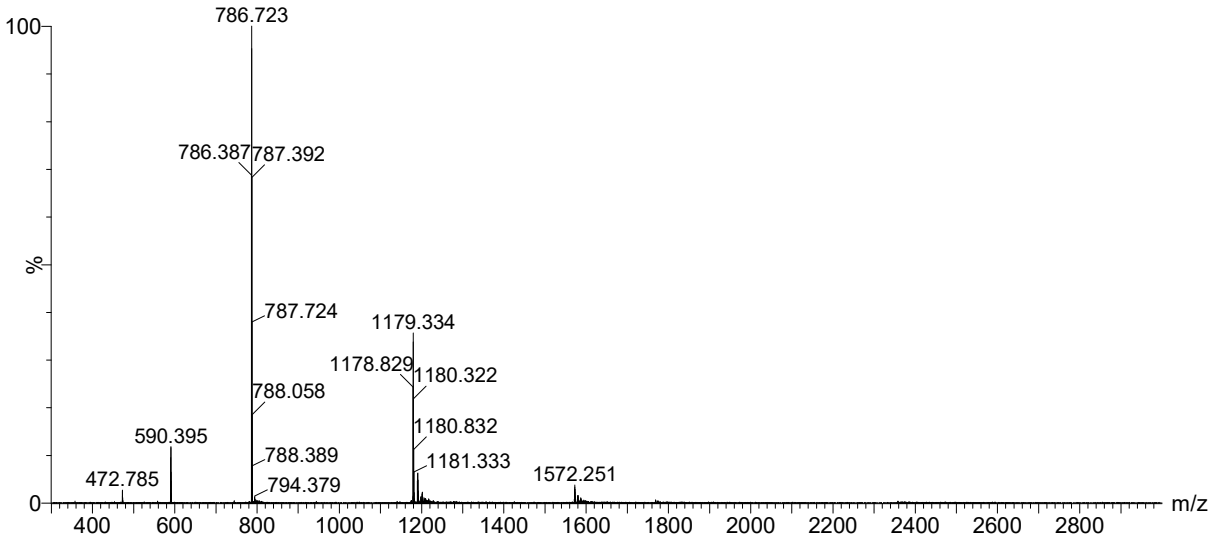


Figure S120: Mass spectrum of $(\mathbf{16})_3$ (retention time 8.33) from the LC-MS analysis of the building block after 2 months.
m/z observed: 786.72 $[M+3H]^{3+}$
m/z calculated: 786.68 $[M+3H]^{3+}$