Chance Emergence of Catalytic Activity and Promiscuity in a Self-Replicator

Jim Ottelé[1,±], Andreas S. Hussain[1,±], Clemens Mayer[2,*], Sijbren Otto[1,*]

Affiliations:

[1] Centre for Systems Chemistry, Stratingh Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, Netherlands

[2] Biomolecular Chemistry & Catalysis, Stratingh Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, Netherlands

[±]These authors contributed equally

*Correspondence to: c.mayer@rug.nl, s.otto@rug.nl

ABSTRACT

How life can emerge from inanimate matter is one of the grand challenges in contemporary science. Regardless of the approach to this challenge[1], self-replicating systems[2–6] play a central role in the transition from chemistry to biology. While self-replication is a necessary condition for life, it is not a sufficient one. Functions beyond replication need to be acquired and assimilated in order for life to emerge[5,7]. The ability to catalyse reactions is one of the most essential of such functionalities and key to another central characteristic of life: metabolism. However, the mechanisms through which self-replicators can acquire catalytic and metabolic properties remain to be established[8,9]. Here we show how catalytic activity in a self-replicator emerges through co-option: features which are selected to benefit replication inadvertently result in an arrangement of chemical functionalities that is conducive to catalysis. Specifically, in a system of self-assembly driven self-replication, the assembly process generates the combination of a substrate-binding pocket and a catalytically active lysine residue. This configuration does not only enable the catalysis of a retro-aldol reaction with activities comparable to the best designer enzymes, but also the cleavage of FMOC groups with high efficiencies. Notably, the latter transformation liberates an alkene, which promotes the formation of molecules that replicators use for replication, thereby exerting a positive feedback on replication. Such chance invention of new function at the molecular level is essential for open-ended evolution and marks a pivotal step in the process by which replicators can acquire metabolic activity. The

catalytic promiscuity exhibited by this system further augments the evolutionary significance of these findings.

INTRODUCTION

It is generally accepted that life requires the integration of replication and metabolism and that metabolism relies heavily on catalysis. Hence, catalytically active replicators are an important, but until now largely missing, stepping stone in the origin and de-novo synthesis of life. It is commonly thought that during its emergence, life went through a stage where the roles of replication and catalysis were shared by a single type of molecules. For example, one of the main arguments in support of the RNA world hypothesis is that RNA can act as catalyst and can replicate. Indeed, RNA molecules have been developed that catalyse reactions and RNA molecules have been made that can self-replicate, but no synthetic RNA molecules exist yet that can do both.[10] The central problem appears to be that catalysis requires folding, while replication requires an unfolded strand. This raises the question how catalytically active replicators can spontaneously arise. After more than 30 years of research on self-replicating molecules, the mechanisms that allow the integration of catalysis with self-replication without human intervention remain obscure. In search for a mechanism that allows for the autonomous emergence of catalytic abilities in self-replicators, we noted that the self-assembly of molecules can, in select cases, endow these with new or improved chemical reactivities[11–15]. By analogy, we surmised that a class of self-assembly driven self-replicators that we developed previously may well exhibit catalytic capabilities. Specifically, we investigated the replication system made from building block **1** (Fig. 1a), which features a small peptide chain appended to a benzene-1,3-dithiol core[16]. In brief: in aqueous solution and in the presence of oxygen from the air, this building block oxidizes to form a mixture of different sized disulfide macrocycles, dominated by trimers $\mathbf{1_3}$ and tetramers $\mathbf{1_4}$, that interconvert through disulfide exchange reactions (Fig. 1b). After an initial lag phase, the hexamer rings ($\mathbf{1_6}$) assemble into stacks through a nucleation-growth mechanism. Assembly of hexamer rings depletes the exchange pool of these rings, which get replenished[17]. Thus, the increasing the number of ends from which the stacks grow, enabling exponential replication[18]. We hypothesized that stacking of $\mathbf{1_6}$ would bring the lysine side chains appended to these rings into close proximity, thereby lowering the pK_a of some of the ε-amines

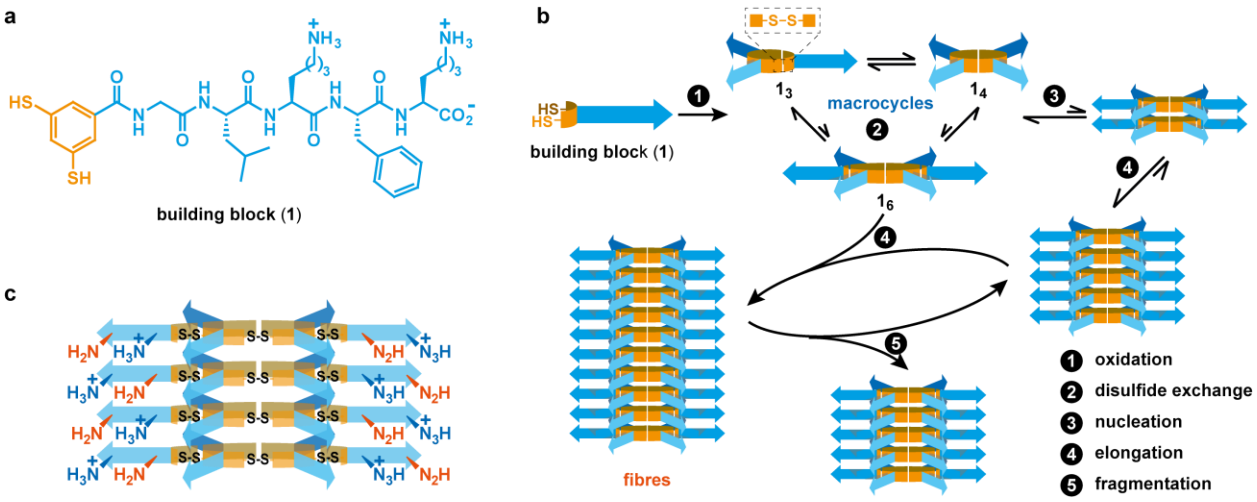


Fig. 1 | Fibre formation from building block 1 creates microenvironments that are conducive to catalysis. a, Molecular structure of building block **1**. The benzene-1,3-dithiol core is displayed in orange and the peptide chain in blue. **b**, Mechanism of self-replication driven by self-assembly of hexamer macrocycles $\mathbf{1}_6$. A series of oxidation (1) and disulfide exchange reactions (2) gives rise to $\mathbf{1}_6$. Following nucleation (3) stacking of additional $\mathbf{1}_6$ results in fibre growth (4) with mechanical energy leading to fragmentation (5), thereby increasing the number of available growing stack ends. **c**, Fibre assembly brings amines from different building blocks into close proximity (for clarity only amines on two of the six stacks of amines are shown). Coulombic repulsion perturbs the pK_a of these amines, thus increasing their nucleophilicity.

and enhancing their nucleophilicity and Brønsted basicity at neutral pH (Fig. 1c). Notably, analogous configurations are often found in the active sites of evolved and designed enzymes[11,19,20], where ionisable groups are positioned in proximity to catalytic residues to increase their reactivity. We further noted that the ε-amines in the replicator fibres are adjacent to hydrophobic leucine and phenylalanine residues. It is well established that placing amines in hydrophobic microenvironments can enhance their catalytic activity[21].

RESULTS AND DISCUSSION

To probe the catalytic potential of replicator fibres we first selected the retro-aldol reaction of methodol (**2**, Fig. 2a) and the cleavage of FMOC-glycine (**4**, Fig. 3a) as model transformations that may be accelerated by organocatalysis and Brønsted-base catalysis, respectively.

Retro-aldol catalysis.

The retro-aldol reaction of methodol may be catalysed by amines, through iminium ion formation, which activates the C-C bond towards cleavage (Fig. 2a)[22]. While the resulting

enamine undergoes tautomerization and hydrolysis to regenerate the catalyst, the cleavage also yields 6-methoxy-2-naphtaldehyde (**3**), which can be readily quantified by UV/Vis spectroscopy or reversed phase UPLC methods. Indeed, when incubating fibres of $\mathbf{1_6}$, (50 μM with respect to building block **1**) with methodol (200 μM) in borate buffer (50 mM, pH 8.12) at 25 °C, we observed the rapid formation of retro-aldol product **3** (Fig. 2b).

Critically, catalysis is an emergent property of the replicator assemblies, as evident from Fig. 2b, which shows that the activity of $\mathbf{1_6}$ fibres substantially exceeds that of **1** and a mixture of $\mathbf{1_3}$/$\mathbf{1_4}$ at equimolar concentrations of building block (see also Extended Data Figs. 1a-b). We attribute this difference to the altered microenvironment and higher local positive charge density that is provided by fibres of $\mathbf{1_6}$, inducing a larger shift in the lysine pK_as (vide infra) as compared to the building block or the non-assembled macrocycles. Furthermore, fibres are also more likely than the non-assembled molecules to provide a suitable microenvironment and binding pocket for catalysis.

To determine the catalytic efficiency of $\mathbf{1_6}$ replicator fibres, we first performed saturation kinetic studies and observed an initial burst phase, followed by a slower, steady-state rate for methodol consumption (Extended Data Fig. 1c). Such behaviour has previously been observed in designer enzymes and ascribed to slow release of the enamine intermediate or product rebinding to active sites[21]. When varying the concentrations of methodol, both regimes displayed saturation (Fig. 2c), indicative of the presence of a substrate-binding pocket. The kinetic analysis is complicated by the fact that the number of catalytically-active lysines (n) is a-priory unknown. To tackle this problem, we followed an established approach, comparing analyses performed using an excess of substrate to catalyst (Fig. 2c) to those using an excess of catalyst to substrate (Fig. 2d)[23,24]. Fitting data obtained using excess substrate to a Michaelis-Menten model yields a value for K_M and a composite value for k_{cat} and n, while data at excess catalyst provides k_{cat} and a composite term for K_M and n (Figs. 2c-d). Hence, a comparison between the kinetic parameters obtained under these conditions provides an estimate of the number of active lysines, and the catalytic parameters of a single active site (full details are provided in the Supplementary Information). Based on this comparison, we conservatively estimate that less than 10% of all lysines are catalytically active. This result is in line with recent observations indicating that $\mathbf{1_6}$ fibres are relatively poorly ordered[25] and, therefore, likely to provide a spectrum of microenvironments, of

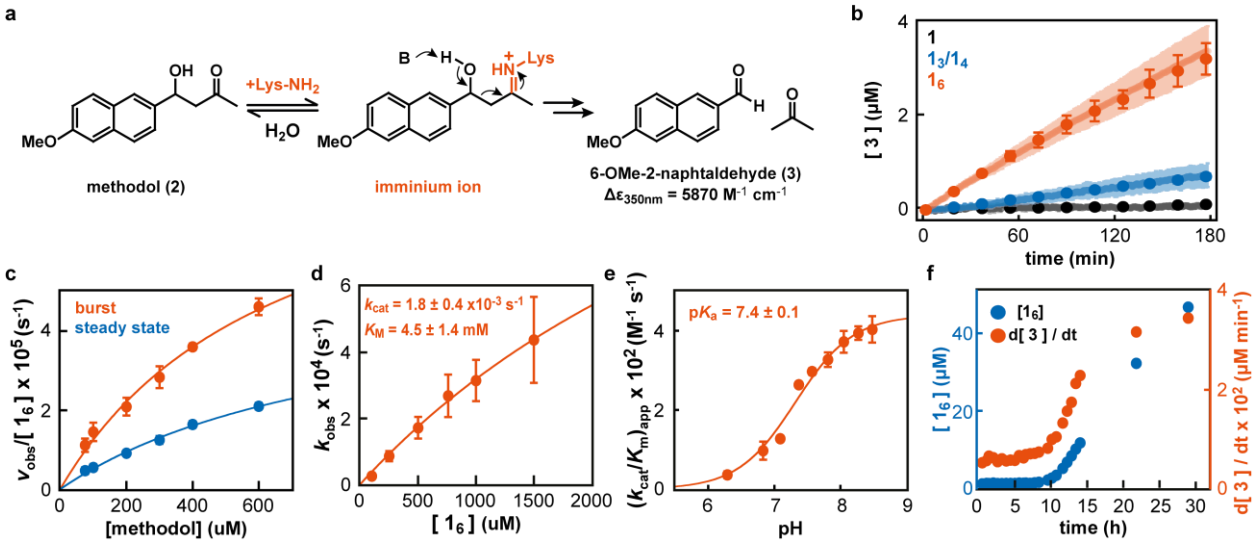


Fig. 2 | Catalytic activity of $\mathbf{1_6}$ in a model retro-aldol reaction emerges due to the combination of a substrate-binding pocket and nucleophilic lysine residues. a, The retro-aldol cleavage of methodol is accelerated in presence of a nucleophilic amine by the formation of an iminium ion intermediate. The C-C bond cleavage yields **3** and an enamine intermediate, which tautomerizes and hydrolyses in subsequent steps (not shown) to give acetone and regenerate the catalyst. **b,** At equimolar concentrations (50 μM with respect to **1**), fibres of $\mathbf{1_6}$ (red) produce **3** more rapidly than a mixture of $\mathbf{1_3}$/$\mathbf{1_4}$ (blue) or building block **1** (black). Data were obtained by either following the reaction by UPLC (filled circles and standard deviations) or spectrophotometrically (lines, standard deviations shown as shaded areas) and are the result of at least two independent measurements. **c,** Saturation kinetics for the burst phase (red) and the steady state phase (blue) for catalysis by $\mathbf{1_6}$ (rates and standard deviations of at least two independent measurements are shown). Obtained velocities for both regimes were fitted to the Michaelis-Menten equation (solid lines). **d,** Kinetic characterization for a large excess of $\mathbf{1_6}$ fibres (100 – 1500 μM in **1**) to **2** (10 μM). Obtained pseudo-first order rate constants were fitted to a Michaelis-Menten model and the obtained k_{cat} and K_M values are displayed. Data points are the average of at least two independent measurements with standard deviations shown. **e,** pH rate profile obtained for catalysis by $\mathbf{1_6}$ (40 μM in **1**) of the retro-aldol reaction of methodol (200 μM). **e,** Emergence of catalytic activity for the retro-aldol reaction in a stirred sample of **1** correlates with fibre formation as shown by a comparison of the change in the concentration of $\mathbf{1_6}$ (blue, left axis) and rate of formation of **3** (red, right axis) with time. All error bars shown correspond to the standard deviation.

which only some appear catalytically active. For single active sites on $\mathbf{1_6}$ fibres, K_M values of 670 ± 140 μM (burst phase) and 930 ± 160 μM (steady state) attest on an appropriate substrate-binding pocket. Furthermore, we obtained apparent second order rate constants for the burst (b) and steady state (ss) phases of $k_{cat}/K_{M,b}$ = 0.72 ± 0.07 M^{-1} s^{-1} and $k_{cat}/K_{M,ss}$ = 0.29 ± 0.02 M^{-1} s^{-1}. As such, individual active sites on $\mathbf{1_6}$ fibres outperform non-assembling macrocycles $\mathbf{1_3}$/$\mathbf{1_4}$ by a factor of 106 (burst) and 42 (steady state) and provide chemical proficiencies ($1/K_{TS}$ = $(k_{cat}/K_M)/k_{uncat}$) of 1.91 x 10^7 M^{-1} and 0.76 x 10^7 M^{-1} for the burst and steady-state phases, respectively. Notably, the k_{cat}/K_M values exhibited by the replicator assemblies are comparable to those of the best computationally designed enzymes for the same reaction (Extended Data Table

1)[26–28], which is remarkable, given that the present system emerged spontaneously, essentially unaided by human design or evolutionary optimization.

Next, we probed whether stacking of $\mathbf{1_6}$ units boosts the catalytic prowess of the lysine groups. The pH-rate profile of the catalysed reaction shows a single inflection point at pH 7.4 ± 0.1 (Fig. 2e), demonstrating a perturbation of the pK_a of the catalytically active lysine groups by approximately three units compared to the pK_a of lysine in solution. When performing pH titrations of $\mathbf{1_6}$ fibres we were unable to identify a significant change in ionisation at pH 7.4 (Extended Data Fig. 1d), confirming that the fraction of the lysines on the $\mathbf{1_6}$ fibres that are catalytically active is small.

Next, we monitored the correlation between the emergence of the replicator and the emergence of catalytic activity *in situ* starting from building block **1**. We dissolved **1** (154 μM) and methodol (200 μM) in borate buffer (pH 8.12) at 40 °C applying mechanical agitation. In parallel, we ran control samples that either lacked **1** or agitation; the latter preventing fibre breakage, thus suppressing replicator growth. We monitored the concentrations of building block **1**, non-assembling tetramer macrocycle $\mathbf{1_4}$ and assembling hexamer macrocycles $\mathbf{1_6}$, as well as retro-aldol product **3** over time using reversed-phase UPLC analysis (Fig. 2f and Extended Data Fig. 2; see Method section for details)[29]. Notably, the increase in the rate of the retro-aldol reaction correlates directly with the emergence of the replicator in agitated samples (Fig. 2f).

In the unstirred control replicator emergence is hampered and virtually no acceleration of the retro-aldol reaction was observed (see Extended Fig. 2 for a full analysis).[18] Finally, to confirm hexamer formation entails fibre assembly, we also subjected samples from the experiments shown in Fig. 2f periodically to TEM analysis, which confirmed the appearance of fibres as soon as the concentration of $\mathbf{1_6}$ started to increase (Extended Data Fig. 3). Together, these results demonstrate how chance invention at the molecular level can endow self-replicators with catalytic abilities. Moreover, they constitute the first *de novo* example of a molecule with the concurrent abilities to replicate and to catalyse a chemical reaction that is of a very different nature than the replication reaction.

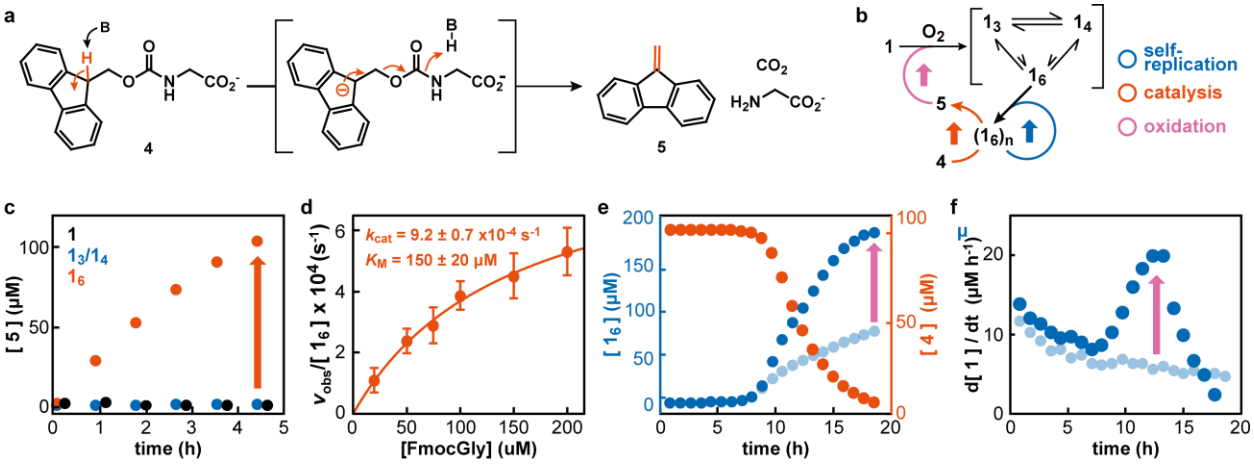


Fig. 3| Emergence of a protometabolism in response to the cleavage of FMOC-glycine by replicator fibres. a, Mechanism of the Brønsted-base-catalysed cleavage of FMOC-glycine. **b**, Concept of a positive feedback loop, in which the conversion of **4** by the self-replicator $(\mathbf{1_6})_n$ yields **5**, which speeds replication by accelerating the oxidation of building block **1** (mechanism of replication shown in Fig. 1b). **c,** At equimolar concentrations (200 μM with respect to **1**), replicator fibres $\mathbf{1_6}$ (red) accelerate FMOC-glycine cleavage (200 μM), while neither building block **1** nor non-assembled macrocycles $\mathbf{1_3}/\mathbf{1_4}$ provide a detectable activity when compared to the uncatalyzed reaction (not shown). **d**, Saturation kinetics for the cleavage of **4** by $\mathbf{1_6}$ fibres (observed rates and standard deviations of at least 3 independent measurements are shown). Kinetic parameters were obtained by fitting the data to the Michaelis-Menten equation (solid lines). **e**, In an agitated sample prepared from **1** (200 μM) and FMOC-glycine (100 μM) in 5% aqueous acetonitrile in borate buffer pH 8.0, the emergence of $\mathbf{1_6}$ (dark blue circles) coincides with the onset of FMOC-glycine cleavage (red circles). Notably, in a sample not containing FMOC-glycine (light blue circles) replicator $\mathbf{1_6}$ emerges at the same time, but grew significantly slower. The increase in replication efficiency in the presence of FMOC-glycine is consistent with a positive feedback loop (*cf.* panel b). **f**, A comparison of oxidation rates in the absence (light blue) and presence (dark blue) of FMOC-glycine show a 3.5 fold increase (purple arrow) once **5** is liberated, providing further evidence for the positive feedback loop. All error bars shown correspond to the standard deviation.

Catalysis of carbamate hydrolysis.

A caveat of the retro-aldol reaction is that none of the liberated products are beneficial to the self-replicator, a necessity for the emergence of metabolism. To examine the possibility of $\mathbf{1_6}$ replicator fibres to spontaneously form a *protometabolism*, we assessed its ability to catalyse the cleavage of FMOC-glycine (**4**). This reaction proceeds through Brønsted-base-catalysed deprotonation, followed by fragmentation into glycine, CO_2 and dibenzofulvene (**5**, Fig. 3a). Attractive features of this reaction are that it is catalysed by basic amines (*e.g.* activated lysine chains in $\mathbf{1_6}$) and (2) and that it liberates an alkene, **5**, which is known to promote thiol oxidation[30,31]. Therefore, the cleavage of FMOC-glycine by replicator fibres could lead to the accelerated formation of the non-assembled macrocycles $\mathbf{1_3}/\mathbf{1_4}$. As they are the precursors from

which the replicator grows (Fig. 1b), accelerating their formation should, in turn speed up replication (Fig. 3b).

Indeed, when incubating a solution of FMOC-glycine (200 μM in borate buffer pH 8.1 containing 5% acetonitrile) with $\mathbf{1_6}$ fibres (200 μM in **1**) at 25 °C, rapid cleavage of **4** took place, while neither **1** or a mixture of $\mathbf{1}_3$ and $\mathbf{1}_4$ showed significant activity under otherwise identical conditions (Fig. 3c). As for the retro-aldol reaction, $\mathbf{1_6}$ fibres displayed saturation when varying the concentration of **4** (Fig. 3d) and the pH-rate profile of the catalysed reaction showed a single inflection point at pH 7.1 ± 0.1 (Extended Fig. 4a), which closely matches the one observed for methodol cleavage (pH 7.4 ± 0.1). While the poor solubility and stability of **5** prevented us from performing a more detailed kinetic analysis akin to that for the retro-aldol reaction, the absence of a recognizable plateau corresponding to a pH of 7.1 in the pH titration of $\mathbf{1_6}$ fibres (Extended Fig. 1d) indicates that, again, only a small fraction of all lysines is catalytically active. Notably, $\mathbf{1_6}$ fibres (k_{cat}/K_m = 6.2 ± 0.4 M^{-1} s^{-1})[32] outperform a mixture of $\mathbf{1_3}$ and $\mathbf{1_4}$ ($k_{app,3/4} = 6.4 \pm 0.6 \times 10^{-3}$ M^{-1} s^{-1}; see Methods) by almost 3 orders of magnitude. Replicators provide a chemical proficiency ($1/K_{TS} = (k_{cat}/K_M)/k_{uncat}$) of at least[32] 2.1 x 10^8 M^{-1}.

To test whether formation of **5** leads to the envisioned positive feedback on replication, we incubated FMOC-glycine (100 μM) with monomer **1** (200 μM) in borate buffer (pH 8.1, 5% acetonitrile) at 40 °C and simultaneously monitored the emergence of $\mathbf{1_6}$ replicator, its effect on FMOC-glycine cleavage, and the reciprocal effect of the cleavage products on the replication of $\mathbf{1_6}$ (Fig. 3e). As for the retro-aldol reaction, the emergence of replicator after approximately 10 h coincides perfectly with the onset of FMOC-glycine cleavage, consistent with only $\mathbf{1_6}$ fibres, but none of the precursors, being catalysts for the reaction. Strikingly, in a sample lacking **4** replicators emerged at the same time, but only grew at a rate that is less than half of that in presence of FMOC-glycine (Fig. 3e). Similarly, when monitoring the oxidation of **1** in both samples, the FMOC-containing one gave rise to a 3.5-times higher oxidation rate in response to formation of **5** (Fig. 3f), confirming the envisioned positive feedback loop (Fig. 3b; see Extended Data Fig. 4b-c for an analogous experiment containing 200 μM FMOC-glycine). Control experiments with isolated dibenzofulvene confirmed that this effect is due to this compound promoting the oxidation of **1** (Extended Figure 4d).

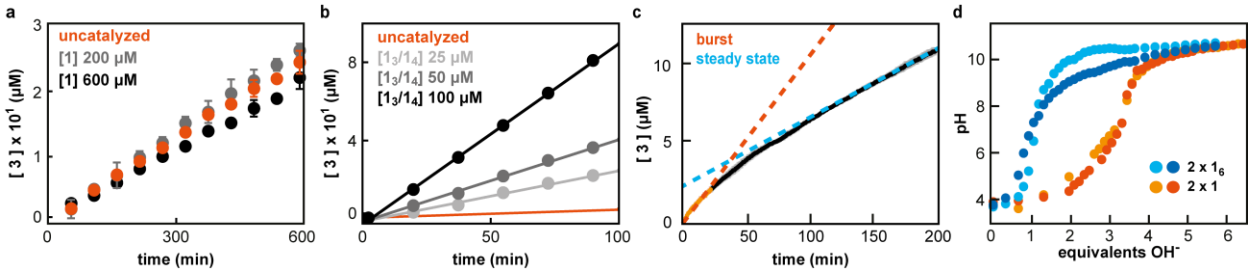
CONCLUSIONS

In conclusion, a self-assembling self-replicator was found to be able to catalyse not only its own formation, but also two other chemical transformations, that are each very different in nature from the replication reaction; i.e. the systems shows catalytic promiscuity. Catalytic function emerges at the level of the replicator assembly and cannot be ascribed to the starting materials or intermediates in the process of replicator formation. While the assemblies emerge spontaneously and are selected on the basis of their ability to replicate, our results show that such selection for replication can inadvertently also produce an arrangement conducive to catalysis, featuring a substrate-binding pocket and lysine residues that have enhanced catalytic activity, imposed by the microenvironment generated through the assembly into fibres. As shown for catalysis of FMOC-glycine cleavage, the replicators are able to generate molecules that promote the formation of the building blocks needed for replication, thereby accelerating the replication process. Such behaviour can be regarded as the onset of metabolism. Thus, these results reveal how replication and metabolism, two key characteristics of life, can autonomously become integrated and mark an important advance towards the de-novo creation of a minimal form of life, which now 'merely' requires the functional integration of a compartmentalisation system. The present system of self-assembly driven self-replicators has the important advantage over other replicators in that sites for catalytic activity (the sides of the fibres) are distinct from the sites of self-replication (occurring at the fibre ends). The catalytic promiscuity exhibited by the system is a highly desirable feature in early evolution, as it allows catalysing different reactions while only having to replicate a single entity. Catalytic promiscuity is also an important pathway for evolutionary inventions, as it enables a system that evolved for one purpose to be co-opted for another purpose.[33] Finally catalytic promiscuity has also been identified as a desirable evolutionary property in the metabolically coupled replicator system model,[34] which provides a solution to overcome Eigen's paradox (accurate replication needs a complex machinery, yet obtaining such complex self-replicators through evolution requires sufficiently accurate replication). Experimentally no efforts have yet been made to address Eigen's paradox, but with protometabolic replicators showing catalytic promiscuity, a path toward clearing this important hurdle in the development of life is starting to be unveiled.

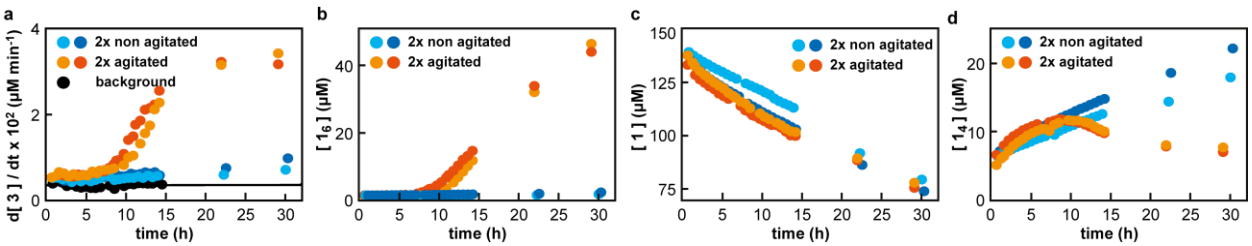
Extended Data Table 1: Summary of kinetic parameters for catalysis of the retro-aldol reaction of methodol (**2**) by replicators $\mathbf{1_6}$ as well as a comparison of catalytic activities with a number of computationally-designed retro-aldolases.

name	reference	k_{cat} x 10^4 (s^{-1})	K_M (μM)	k_{cat}/K_M x 10^1 (M^{-1} s^{-1})	$1/K_{TS}$ x 10^{-7} (M^{-1})
1_6 (b)*		**4.80 ± 0.61**	**670 ±140**	**7.17 ± 0.66**	**1.91**
1_6 (ss)*		**2.66 ± 0.32**	**930 ± 160**	**2.86 ± 0.18**	**0.76**
RA22	26	0.08	450	0.18	0.28
RA34	26	0.10	600	0.16	0.25
RA45	26	0.38	430	0.91	1.40
RA46	26	1.03	290	3.70	5.69
RA60	26	1.55	510	3.00	4.62
RA61	26	1.50	210	7.40	11.4
RA95	27	0.33	540	0.53	0.82
RA110	27	0.83	1600	0.48	0.74
RA114	28	0.03	34	0.95	1.46
RA115	28	0.15	1740	0.087	0.13
RA116	28	0.14	1570	0.088	0.14
RA117	28	0.28	470	0.58	0.89
RA120	28	0.04	880	0.041	0.06

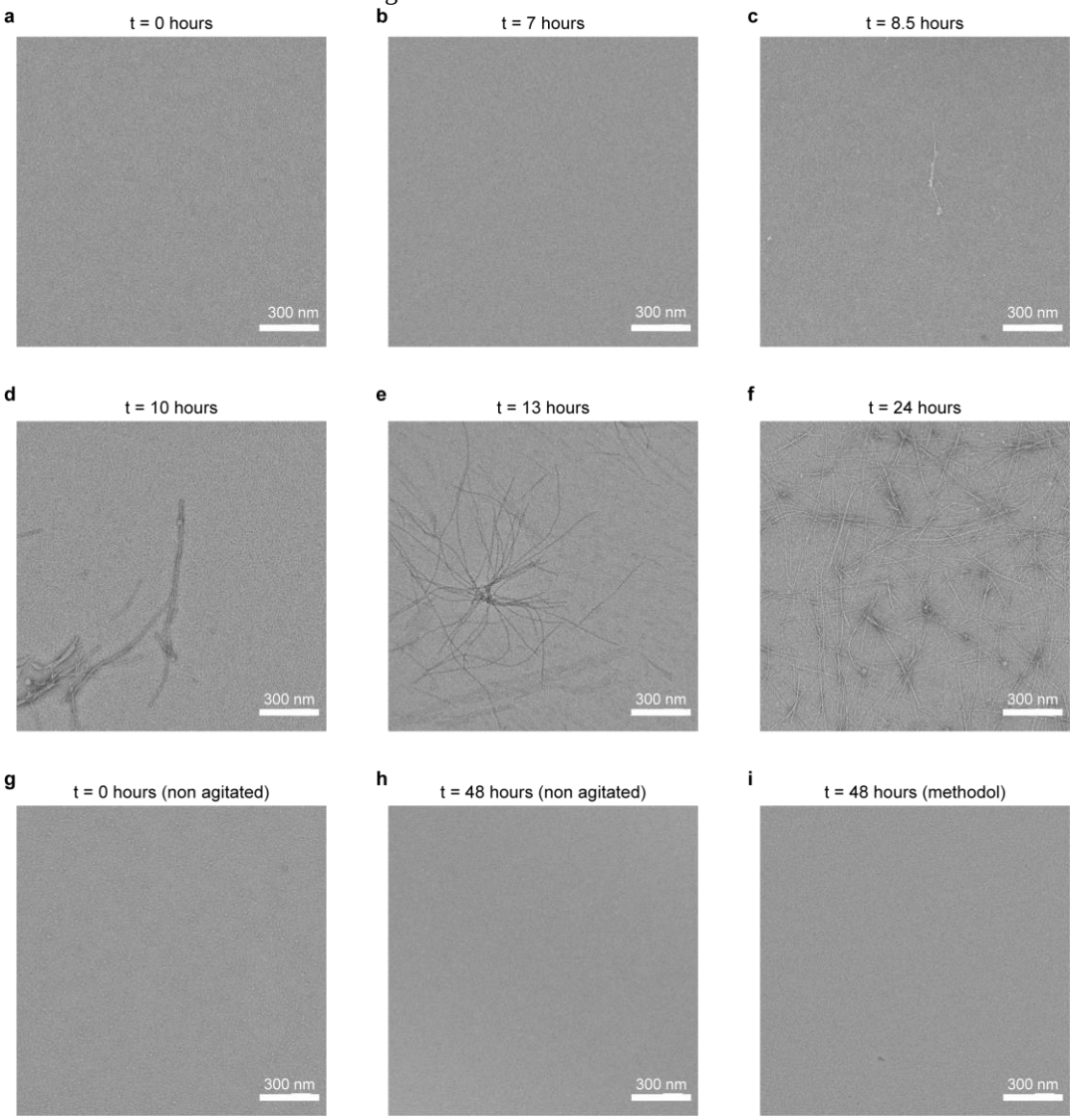
* k_{cat}, k_{cat}/K_M and $1/K_{TS}$ values are obtained from kinetic experiments in which we used an excess of substrate over catalyst, assuming that only 10% of all lysines are catalytically active.



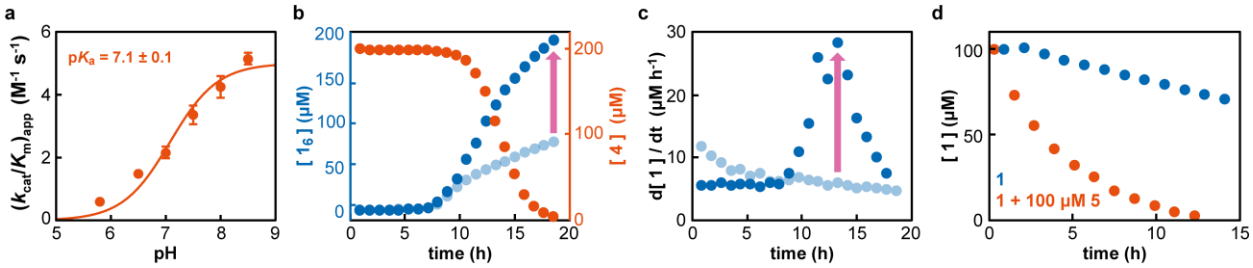
Extended Data Fig. 1: a-b: Comparison of the reaction progress between the uncatalyzed retro-aldol reaction (red in **a** and **b**) and that in the presence of different concentration of **1** (**a**) and varying concentrations of a mixture of $\mathbf{1_3}$/$\mathbf{1_4}$ (**b**). **c:** Reaction progress monitored spectrophotometrically with $\mathbf{1_6}$ (25 μM) and methodol (600 μM, in triplicates with standard deviation shown as shaded area). Red and blue dotted lines are linear fits for the observed burst (first 20 minutes) and steady state rate (100 – 200 minutes). **d:** pH titration for **1** and $\mathbf{1_6}$ fibres in water (both in duplicates). The buffering capacity of monomers at a pH between 5 and 7 can be attributed to the presence of aromatic thiols in **1**. In contrast, $\mathbf{1_6}$ fibres do not reveal appreciable levels of an ionizable group in the range from 5 – 8. The absence of a recognisable plateau at the pK_a of the lysines active in the retro-aldol (pK_a = 7.4; Fig. 2e) or FMOC-glycine cleavage reactions (pK_a = 7.1; Extended data Fig. 4a) indicates that, for each of these reactions, the fraction of lysine residues in the fibres that are catalytically active is low. All error bars shown correspond to the standard deviation.



Extended Data Fig. 2: Emergence of catalytic activity of the retro-aldol reaction in a stirred sample of 1 correlates with fibre formation. a: Rate of formation of retro-aldol product **3** over the course of the experiments. Solid line represents a linear fit to the uncatalyzed reaction. **b-d**: Concentration of $\mathbf{1_6}$ (**b**), **1** (**c**), and $\mathbf{1_4}$ (**d**) followed over time. Colour code as indicated in the legend.



Extended Data Fig. 3: Time-dependent TEM experiments. For agitated samples (**a-f**) fibres become detectable after 8.5 hours, which correlates with the increase in the concentration of $\mathbf{1_6}$ in the experiment shown in **Fig. 2**. Neither the non-agitated sample (**g-h**) nor the sample lacking **1** (**i**) showed detectable levels of fibres after 48 hours.



Extended Data Fig. 4: a, pH rate profile for the cleavage of FMOC-glycine (100 µM) catalysed by replicator $\mathbf{1}_6$ (40 µM in **1**). **b**, Data for experiments as shown in Fig. 3b, but using 200 µM FMOC-glycine. The emergence of $\mathbf{1}_6$ (dark blue circles) coincides with the onset of FMOC-glycine cleavage (red circles) and $\mathbf{1}_6$ replicates faster when compared to a sample that does not contain FMOC-glycine (light blue circles). **c**, A comparison of oxidation rates in the same experiment showed an up to 4.7 fold increase (purple arrow) once **5** is liberated, providing further evidence for the positive feedback loop shown in Fig. 3b. **d,** A comparison of the oxidation of **1** (100 µM) in presence (red circles) and absence (blue circles) of **5** (100 µM) confirmed that this FMOC-cleavage product is responsible for enhancing the rate of oxidation of **1**. Experiments were performed in borate buffer pH 8.1, containing 5% acetonitrile at 40 °C. All error bars shown correspond to the standard deviation.

METHODS

General procedures

All reagents, solvents and buffer salts were obtained from commercial suppliers and used without further purification, unless otherwise noted. Building block **1** was obtained from Cambridge Peptides Ltd (Birmingham, UK). Borate buffer concentrations are reported in concentration of boron atoms. UPLC analysis was performed on a Waters Acquity UPLC-H class system equipped with a PDA detector. All analyses were performed using a reversed-phase UPLC column (Aeris Peptide 1.7 µm XB-C18 x 2.10 mm, Phenomenex). The column temperature was kept at 35 °C, and the sample plate was kept at 25 °C unless otherwise specified. UV absorbance was monitored at 254 and 313 nm. Samples were separated with a gradient from 28-40% acetonitrile (0.1% TFA) in water (0.1% TFA) at a flow rate of 0.3 mL/min (see Supplementary Table S1 for the detailed method used). To heat the samples beyond 40 °C, a Thermo Fisher compact dry bath with a custom milled aluminium block to accommodate HPLC vials (with a dimension of 12 x 32 mm) was used. UPLC-MS experiments were performed on a Waters Acquity UPLC H-class system coupled to a Waters Xevo-G2 TOF. The mass spectrometer was operated in positive electrospray ionization mode. Capillary, sampling cone and extraction cone voltages were kept at 2.5 kV, 30 V and 4V, respectively. Source and desolvation temperatures were set at 140 °C and 500 °C. Nitrogen was used as both cone (5 L/h) and desolvation gas (500 L/h). Reaction mixtures were analysed by direct injection of the sample (5 µL). Transmission electron microscopy was performed on a Philips CM120 electron microscope operating at 120 kV. Images were recorded on a slow scan CCD camera (Gatan). Spectrophotometry measurements were performed on a Jasco V-660 spectrophotometer at 25 ºC unless otherwise noted. Methodol was synthesized using a previously published method[35]. All concentrations involving $\mathbf{1_3}$, $\mathbf{1_4}$ and $\mathbf{1_6}$ are given with respect to building block **1**, unless stated otherwise. A mixture of $\mathbf{1_3}$/$\mathbf{1_4}$ was obtained by oxidizing **1** with sodium perborate (0.95 eq.) in borate buffer (50 mM, pH 8.12). The resulting mixture, after oxidation, was analysed by UPLC and mass spectrometry (Supplementary Figs. 1-2). FMOC-glycine **4** (Fmoc-Gly-OH, 97%) was obtained from Sigma-Aldrich. Dibenzofulvene **5** was synthesised according to a known literature procedure[36].

Preparation of $\mathbf{1_6}$

A stock solution of $\mathbf{1_6}$ fibres was typically prepared from building block **1** as follows. In brief, building block **1** (2.28 mg, 2.31 μmol) was added to an HPLC vial (12 x 32 mm) containing a Teflon coated magnetic stirring bar (5 x 2 mm, VWR). The sample was dissolved in 1.5 mL borate buffer (50 mM, pH 8.12) to give a final concentration of 1.54 mM **1**, and the vial was subsequently closed with a Teflon septum screw cap. Then, the mixture was stirred at 1,200 rpm at 40 °C for >72 hours. The composition of the reaction mixture was periodically analysed by UPLC (10 μL injection) and stirred until $\mathbf{1_6}$ accounted for >90% of the mixture (see Supplementary Figures S1 and S2 for a representative UPLC chromatogram and mass spectra of the final mixture). The resulting stock solution was kept at 25 ºC while stirring (1,200 rpm) and could be used for at least 12 weeks without observing any significant changes in the sample composition.

Initial assessment of the activity of $\mathbf{1_6}$ in catalysing the retro-aldol reaction

Acetonitrile (10 μL) and a stock solution of methodol (10 μL, 20 mM in acetonitrile) were added to 948 μL of borate buffer (50 mM, pH 8.12) and the resulting mixture was incubated for 10 minutes at 25 °C. At this point, 32.5 μL of a stock solution of $\mathbf{1_6}$ (1.54 mmol in the same borate buffer) was added to yield a final concentration of 50 μM $\mathbf{1_6}$ (with respect to building block **1**), 200 μM methodol and 2% acetonitrile. Reaction progress was either monitored spectrophotometrically (ε_{350} = 5870 M^{-1} cm^{-1}) at 25 °C or through periodic analysis of the reaction mixture by UPLC at the same temperature (10 μL injection of the mixture, 18 minute intervals). For the latter, the concentration of **3** was quantified by comparison to a calibration curve that was obtained by injecting known quantities of **3** in the reaction buffer (λ_{max} = 313 nm, Supplementary Figure S3). As control, the experiment was carried out in presence of **1** as well as a mixture of $\mathbf{1_3}$/$\mathbf{1_4}$ under otherwise identical conditions. The data presented in Fig. 2b represents the average of at least two independent experiments.

Effect of 1 and a mixture containing 1_3 and 1_4 on the retro-aldol reaction

Reaction mixtures containing methodol (final concentration 200 μM) and acetonitrile (2%) were prepared and incubated for 10 minutes at 25 °C as described above. To determine the kinetic parameters for **1**, either 66.7 μL or 200 μL of a stock solution of **1** (3.08 mM in borate buffer) was added to give a final concentration of 200 μM or 600 μM **1**, respectively. In order to avoid oxidation of **1** to small, non-assembling macrocycles 10 μL of a stock solution of dithiothreitol (200 mM, 2 mM final concentration) was added. The resulting reaction mixtures were incubated at 25 ºC and periodically analysed by UPLC (10 μL injections, 54 minute intervals for 10 hours). The analysed retro-aldol reaction was measured in parallel by replacing the stock solution of **1** with borate buffer (pH 8.12). The concentration of **3** was quantified by comparing the peak area at 313 nm to the calibration curve shown in Supplementary Figure S3. The reaction progress for each reaction is displayed in Extended Data Fig. 1 and demonstrates that at these concentrations **1** does not lead to appreciable levels of acceleration of this retro-aldol reaction.

Kinetic characterization of a freshly prepared mixture of **1_3/1_4** was performed in a similar fashion. To a sample containing methodol (200 μM) and acetonitrile (2%) in borate buffer (pH 8.12), 16.2, 32.5 or 65 μL of a stock solution of **1_3/1_4** (1.54 mM with respect to building block **1** in borate buffer pH 8.12) was added, giving total a total concentration of 25, 50 or 100 μM (with respect to building block **1**). Reaction mixtures were monitored for 90 minutes by periodic analysis by UPLC (10 μL injections, 18 minute intervals) at 25 ºC. The concentration of **3** was quantified as specified above. Observed second order rate constants at all three concentrations were measured by determining the initial rate of the reaction and were comparable for the **1_3/1_4** mixture at all concentrations tested (8.25×10^{-3} $M^{-1}s^{-1}$ (25 μM), 7.41×10^{-3} $M^{-1}s^{-1}$ (50 μM) and 8.29×10^{-3} $M^{-1}s^{-1}$ (100 μM)), indicative of the reaction rate being first order with respect to the catalyst. An apparent second order rate constant was determined by averaging the obtained values.

Kinetic characterization of the effect of the 1_6 fibres on the retro-aldol reaction

Excess substrate to catalyst: This kinetic characterization of **1_6** fibres was performed by measuring reaction progress spectrophotometrically (ε_{350nm} = 5,870 M^{-1} cm^{-1}) at varying concentration of **2** (75, 100, 200, 300, 400, and 600 μM) at 25 °C. Samples were prepared by

adding appropriate amounts of acetonitrile and a stock solution of **2** (20 mM in acetonitrile) into 1 cm path length cuvettes containing 948 µL borate buffer (50 mM, pH 8.12). Reaction mixtures were incubated for 10 minutes at 25 °C, at which point 16.2 µL of a stock solution of $\mathbf{1_6}$ (1.54 mM with respect to building block **1** in borate buffer, final concentration 25 µM with respect to building block **1**) was added. Reaction progress was followed for 250 minutes at 25 °C. Observed rates for the burst phase (0 – 20 minutes) and the steady state (100-200 minutes) were extracted by determining the slope of product formation over time for the indicated time frames. The measured rates were corrected for the background rate and the resulting values, which represent the average of at least two independent measurements, were fitted (*RStudio*) to the following equation, in which [**1**] = 2.5 x 10^{-5} M:

$$\frac{v_0}{[\mathbf{1}]} = \frac{k_{cat}\,[\mathbf{2}]}{K_M + [\mathbf{2}]}$$

Excess catalyst to substrate: A stock solution of **2** in acetonitrile (0.50 mM) was prepared and kept at -21 ºC. The solution was used no longer than one week after preparation. Stock solutions of $\mathbf{1_6}$ fibres (2.0 mM with respect to **1**) were prepared as described before and used within two weeks after the library composition was >90% $\mathbf{1_6}$. For the kinetic characterization, varying volumes of $\mathbf{1_6}$ fibres (30 – 450 µL), were added to HPLC vials (12 x 32 mm) containing borate buffer (50 mM, pH 8.12, 138 – 558 µL) and 12 µL of the stock solution of **2**, giving a total volume of 600 µL. Final concentrations for $\mathbf{1_6}$ ranged from 100-1500 µM with respect to building block **1** with a constant concentration of **2** (10 µM). The formation of **3** was monitored by periodic analysis of the reaction mixture by UPLC (10 µL injections, 18 minute intervals for 128 minutes). From the obtained results, pseudo-first order rate constants were calculated for different concentrations of $\mathbf{1_6}$. Averages of values from least two independent measurements were subsequently fitted (*Rstudio*) to the following equation, in which k_{obs} are the measured pseudo-first order rate constants and [**1**] varies from 1 x 10^{-4} to 1.5 x 10^{-3} M:

$$k_{obs} = \frac{k_{cat}\,[\mathbf{1}]}{K_M + [\mathbf{1}]}$$

pH rate profile the effect of $\mathbf{1_6}$ on the retro-aldol reaction

Buffer solutions (50 mM) were prepared spanning the pH range from 6.0 – 8.75 (phosphate buffer pH 6.3-7.1; borate buffer pH 7.4-8.5). Samples were prepared by adding acetonitrile (10 µL) and 10 µL of a stock solution of methodol (20 mM in acetonitrile) to 1 cm cuvettes containing 954 µL of the specific buffer solution. The resulting mixtures were incubated for 10 minutes at 25 °C before addition of a stock solution of $\mathbf{1_6}$ (26 µL, 1.54 mM with respect to building block **1** in borate buffer, pH 8.12) to obtain a final concentration of 40 µM $\mathbf{1_6}$ (with respect to building block **1**), 200 µM methodol and 2% acetonitrile. The reaction progress was monitored spectrophotometrically (ε_{350} = 5,870 M^{-1} cm^{-1}) at 25 °C for 250 minutes. Observed rates for the steady state (100 – 250 minutes) were extracted by analysing the data for the product formation over time. The measured rates were corrected for the background and $(k_{cat}/K_M)_{app}$ values were determined by dividing the corrected rates by the concentrations for **1** (=4 x 10^{-5} M) and substrate (=2 x 10^{-4} M). Averages of values from least two independent measurements were subsequently fitted (*Rstudio*) to the following equation:

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

pH titration of 1 and $\mathbf{1_6}$ fibres

To prepare a stock solution of **1** and $\mathbf{1_6}$ fibres in H_2O, building block **1** (3.21 mg, 3.25 µmol in **1**) was added to a glass vial (15 x 45 mm) containing a Teflon coated magnetic stirring bar (5 x 2 mm, VWR). UPLC grade H_2O (2.1 mL) was added to give a final concentration of 1.54 mM. To enable disulfide formation and exchange, which is necessary for $\mathbf{1_6}$ formation, 1 M NaOH (4 µL) was added to one of the samples, resulting in a pH of 8.2. This vial was closed with a screw cap and the mixture was stirred at 1,200 rpm at 45 °C for 7 days. The library composition was periodically analysed by UPLC (10 µL injections). At the time, $\mathbf{1_6}$ accounted for >90% of the mixture, the stock solution was cooled to 25 °C and continued to be stirred at 1200 rpm. This stock solution could be used for up to 12 weeks without observing a significant change in the sample composition, while the stock solution containing **1** was prepared freshly before the pH titration.

For the pH titration, stock solutions for **1** and $\mathbf{1_6}$ fibres (1 mL each) were diluted with UPLC grade H_2O (1 mL) to obtain a solution containing 1.54 µmol of **1** and $\mathbf{1_6}$. To the latter 1 M HCl (2.5 µL) was added to lower to pH to ~3.6, similar to the pH of the freshly prepared solution of **1**. Then, small aliquots (1-5 µL) of a 0.1 or 0.5 M NaOH solution were added to the samples and the pH monitored using a Hanna HI-4221 Research Grade Bench pH & ORP Meter. Measurements were performed in duplicates and are summarized in Extended Data Fig. 1.

Experiment corresponding to Fig. 2f and Extended Data Fig. 2

Borate buffer (880 µL, 50 mM, pH 8.12) was added to an HPLC vial (12 x 32 mm) containing a Teflon coated magnetic stirring bar (5 x 2 mm, VWR). The sample was heated to 40 °C and acetonitrile (10 µL) and a stock solution of methodol (10 µL, 20 mM in acetonitrile) were added to the preheated buffer. The sample was incubated at 40 °C for 15 minutes. Then, 100 µL of a freshly prepared solution of **1** (1.54 mM in 50 mM borate buffer, pH 8.12) was added to the mixture to give a total concentration of 154 µM **1**, 200 µM methodol and 2% acetonitrile. The reaction mixture was kept at 40 °C and was stirred at 500 rpm for the entire duration of the experiment (30 hours). The reaction was monitored by periodic analysis of the reaction mixture by UPLC (10 µL injection). The concentrations of **1**, $\mathbf{1_4}$, and $\mathbf{1_6}$ were quantified by comparison to a calibration curve that was obtained by injecting known quantities of **1** in the reaction buffer (Supplementary Fig. S3). The concentration of **3** was quantified by comparison to a calibration curve, as described above. Control experiments were performed by either not applying any mechanical agitation or by omitting **1** from the sample. Representative chromatograms recorded near the start and the end of the experiment are depicted in Supplementary Fig. S4.

Time dependent TEM

Samples were prepared as described before for the experiments corresponding to Fig. 2. Fibre formation was followed by periodically preparing TEM samples from the reaction mixtures (*e.g.* for the agitated samples at t = 0, 1, 6, 7.5, 8.5, 10, 13, 24, and 48 hours). For this, a small drop (5 µL) of the reaction mixtures was deposited on a 400 mesh copper grid covered with a thin carbon film (Agar Scientific). After one minute, the grid was blotted on filter paper. The sample was subsequently stained twice with uranyl acetate by depositing 5 µL of a saturated uranyl acetate solution on the grid and blotting it on filter paper after incubation for 20 seconds. Samples obtained from this procedure were then analysed by TEM (Extended Data Fig. 3).

Neither the control sample for which **1** was omitted nor the one for which no mechanical agitation was applied showed detectable levels of fibre formation after prolonged times (48 hours).

Initial assessment of the activity of $\mathbf{1_6}$ in catalysing the FMOC-glycine cleavage

A stock solution of FMOC-glycine **4** (20 μL, 10 mM in acetonitrile) was added to 900 μL of borate buffer (50 mM, pH 8.12) with 30 μL of acetonitrile and the resulting mixture was incubated for 5 minutes at 25 °C. At this point, 50 μL of a stock solution of $\mathbf{1_6}$ (4.0 mmol in the same borate buffer) was added to yield a final concentration of 200 μM $\mathbf{1_6}$ (with respect to building block **1**), 200 μM FMOC-glycine and 5% acetonitrile. Reaction progress was monitored through periodic analysis of the reaction mixture by UPLC at 25 °C (10 μL injection of the mixture, 18 minute intervals). As control, the experiment was carried out in presence of **1** as well as a mixture of $\mathbf{1_3/1_4}$ under otherwise identical conditions. For the latter, the concentration of **4** was quantified by comparison to a calibration curve that was obtained by injecting known quantities of **4** in the reaction buffer (Supplementary Fig. S3c). As control, the experiment was carried out in presence of **1** as well as a mixture of $\mathbf{1_3/1_4}$ under otherwise identical conditions.

Determination of uncatalysed rate and the apparent second order rate constant for a mixture of $\mathbf{1_3}$ and $\mathbf{1_4}$

Since the uncatalysed reaction was observed to be very slow at 25 °C and 40 °C, the reaction mixture was heated to different temperatures using a block heater. Two samples with a total volume of 1000 μL with **4** (40 μM) and acetonitrile (5%) in borate buffer (pH 8.1) were prepared in HPLC vials (12 x 32 mm) and incubated for 5 minutes. The samples were put to three different temperatures in doublets (T = 60, 70 and 80 °C, Supplementary Fig. S6a). The vials were alternatively taken out of the heater to sample for UPLC analysis (10 μL injections, 36 minute intervals for 4.2 h) at 40 ºC. The resulting decrease of **4** was quantified by comparing the peak area at 254 nm to the calibration curve shown in Supplementary Figure S1c. Observed first order rate constants were measured by determining the initial rate of the reaction via the Arrhenius equation, yielding the following data: 2.37×10^{-5} s^{-1} (80 °C), 8.18×10^{-6} s^{-1} (70 °C) and 2.67×10^{-6} s^{-1} (60 °C). These rate constants were used to extrapolate the first order rate

constant at room temperature (2.86 x 10^{-8} s^{-1} at 25 °C). The Arrhenius plot is shown in Supplementary Figure S5b.

To test the catalytic proficiency of the **1**, 75 µL of a stock solution of **4** (1.0 mM in acetonitrile) in a concentration of 40 µM in **4** with an overall acetonitrile content of 5 % were prepared under inert atmosphere and incubated for 5 minutes. To that, 75 µL of a stock solution of **1** (4.0 mM in borate buffer) was added under inert atmosphere to avoid oxidation of **1** into small, non-assembling macrocycles to give a final concentration of **1** of 300 µM. The sample and a control sample without **1** were heated to 60 °C and were periodically analysed by UPLC (50 µL, daily for 3d). The progress of the reaction is displayed in Supplementary Figure S5a and demonstrates that at these concentrations **1** does not lead to significant acceleration of the FMOC-glycine cleavage reaction.

Kinetic characterization of a freshly prepared mixture of $\mathbf{1_3}$/$\mathbf{1_4}$ was performed by adding to a sample containing FMOC-glycine (40 µM) and acetonitrile (5%) in borate buffer (pH 8.1), 12.5, 37.5 or 75 µL of a stock solution of $\mathbf{1_3}$/$\mathbf{1_4}$ (4.0 mM with respect to building block **1** in borate buffer pH 8.1), giving total a total concentration of 50, 150 or 300 µM (with respect to building block **1**). The reaction mixtures were incubated for 10 minutes and then monitored for 6 days by UPLC (10 µL injection of the mixture, daily for 6 days). Observed second order rate constants at all three concentrations were measured by determining the initial rate of the reaction and were comparable for the $\mathbf{1_3}$/$\mathbf{1_4}$ mixture at all concentrations tested (6.31 x 10^{-3} $M^{-1}s^{-1}$ (50 µM), 5.89 x 10-3 $M^{-1}s^{-1}$ (150 µM), 7.07 x 10^{-3} $M^{-1}s^{-1}$ (300 µM)) at room temperature, suggesting first order in catalyst concentration. By averaging the obtained values an apparent second order rate constant was determined (6.43 x 10^{-3} $M^{-1}s^{-1}$). Plots depicting the reaction of **4** in the presence of $\mathbf{1_3}$/$\mathbf{1_4}$ are shown at Supplementary Figure S5c.

Kinetic characterization of the effect of the $\mathbf{1_6}$ fibres for the FMOC-glycine cleavage

Excess substrate to catalyst: This kinetic characterization of $\mathbf{1_6}$ fibres was performed by measuring reaction progress spectrophotometrically (ε_{305nm} = 3,088 M^{-1} cm^{-1}) at varying concentration of **4** (20, 50, 75, 100, 150, and 200 µM) at 25 °C. Samples were prepared by adding appropriate amounts of acetonitrile and a stock solution of **4** (10 mM in acetonitrile) into 1 cm path length cuvettes containing 937 µL borate buffer (50 mM, pH 8.1). Reaction mixtures

were incubated for 20 minutes at 25 °C, at which point 13 µL of a stock solution of $\mathbf{1_6}$ (1.54 mM with respect to building block **1** in borate buffer, final concentration 20 µM with respect to building block **1**) was added. Reaction progress was followed for 60 minutes at 25 °C. Observed rates for the initial phase (0 – 20 minutes) were extracted by determining the slope of product formation over time. The measured rates were corrected for the background rate and the resulting values, which represent the average of at least two independent measurements, were fitted (*RStudio*) to the following equation, in which [**1**] = 2.0 x 10^{-5} M:

$$\frac{v_0}{[\mathbf{1}]} = \frac{k_{cat}\,[\mathbf{2}]}{K_M + [\mathbf{2}]}$$

pH rate profile for the effect of $\mathbf{1_6}$ on the FMOC-glycine cleavage

Different buffer solutions (50 mM) were prepared and adjusted to the pH range from 5.8 – 8.5 (MOPS buffer pH 5.8-7.0; borate buffer pH 7.5-8.5). Samples were prepared by adding acetonitrile (40 µL) and 10 µL of a stock solution of **4** (10 mM in acetonitrile) to 1 cm cuvettes containing 924 µL of the specific buffer solution. The resulting mixtures were incubated for 10 minutes at 25 °C before addition of a stock solution of $\mathbf{1_6}$ (26 µL, 1.54 mM with respect to building block **1** in borate buffer, pH 8.12) to obtain a final concentration of 40 µM $\mathbf{1_6}$ (with respect to building block **1**), 100 µM **4** and 5% acetonitrile. The reaction progress was monitored photometrically (ε_{305nm} = 3,088 M^{-1} cm^{-1}) at 25 °C for 60 minutes. The measured rates were corrected for the background and $(k_{cat}/K_M)_{app}$ values were determined by dividing the corrected rates by the concentrations of **1** (= 4 x 10^{-5} M) and substrate (= 1 x 10^{-4} M). Averages of values from least two independent measurements were subsequently fitted (Rstudio) to the following equation:

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

Experiment corresponding to Fig. 3e-f and Extended Data Fig. 4b-c

A fresh solution of **4** (10 µL or 20 µL, 10 mM in acetonitrile) was added to borate buffer (900 µL, 50 mM, pH 8.1) and acetonitrile (40 µL or 30 µL) in an HPLC vial (12 x 32 mm) equipped

with a Teflon coated magnetic stirring bar (5 x 2 mm, VWR), to give a resulting concentration of 100 µM or 200 µM, respectively, with an acetonitrile content of 5%. The mixture was preheated to 40 °C for 10 minutes inside the UPLC system. To that, 50 µL of a freshly prepared solution of **1** (4.0 mM in borate buffer, pH 8.1) were added while the temperature was kept at 40 °C and the reaction was stirred at 600 rpm for the entire duration of the experiment (27 hours). The reaction mixture was monitored periodically via UPLC (10 µL injection every 54 minutes). The concentrations of **1**, $\mathbf{1_4}$, and $\mathbf{1_6}$ were quantified by comparing to a calibration curve which was obtained by injecting known quantities of **1** in the buffer (Supplementary Fig. S3). The consumption of **4** was quantified by comparing the peak area to the calibration curve shown in Supplementary Fig. S3. Representative chromatograms recorded for the experiment are depicted in Supplementary Fig. S6.

REFERENCES

1. Ruiz-Mirazo, K., Briones, C. & De La Escosura, A. Prebiotic systems chemistry: New perspectives for the origins of life. *Chem. Rev.* **114,** 285–366 (2014).
2. Patzke, V. & von Kiedrowski, G. Self replicating systems. *Arkivoc* **v,** 293–310 (2007).
3. Dadon, Z., Wagner, N. & Ashkenasy, G. The road to non-enzymatic molecular networks. *Angew. Chem. Int. Ed.* **47,** 6128–6136 (2008).
4. Vidonne, A. & Philp, D. Making molecules make themselves - The chemistry of artificial replicators. *Eur. J. Org. Chem.* **2009,** 593–610 (2009).
5. Duim, H. & Otto, S. Towards open-ended evolution in self-replicating molecular systems. *Beilstein J. Org. Chem.* **13,** 1189–1203 (2017).
6. Higgs, P. G. & Lehman, N. The RNA World: molecular cooperation at the origins of life. *Nat. Rev. Genet.* **16,** 7–17 (2014).
7. Moreno, A. & Ruiz-Mirazo, K. The problem of the emergence of functional diversity in prebiotic evolution. *Biol. Philos.* **24,** 585–605 (2009).
8. Eschenmoser, A. The search for the chemistry of life's origin. *Tetrahedron* **63,** 12821–12844 (2007).
9. Kamioka, S. Ajami, D. Rebek, J. Autocatalysis and organocatalysis with synthetic structures. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 541–544 (2010).
10. Le Vay, K., Weise, L. I., Libicher, K., Mascarenhas, J. & Mutschler, H. Templated Self-Replication in Biomimetic Systems. *Adv. Biosyst.* **3,** 1–16 (2019).
11. Müller, M. M., Windsor, M. A., Pomerantz, W. C., Gellman, S. H. & Hilvert, D. A rationally designed aldolase foldamer. *Angew. Chem. Int. Ed.* **48,** 922–5 (2009).
12. Rufo, C. M. *et al.* Short peptides self-assemble to produce catalytic amyloids. *Nat. Chem.* **6,** 303–309 (2014).
13. Weingarten, A. S. *et al.* Self-assembling hydrogel scaffolds for photocatalytic hydrogen production. *Nat. Chem.* **6,** 964–970 (2014).
14. Tena-Solsona, M. *et al.* Emergent catalytic behavior of self-assembled low molecular weight peptide-based aggregates and hydrogels. *Chem. Eur. J.* **22,** 6687–6694 (2016).
15. Omosun, T. O. *et al.* Catalytic diversity in self-propagating peptide assemblies. *Nat. Chem.* **9,** 805–809 (2017).
16. Malakoutikhah, M. *et al.* Uncovering the selection criteria for the emergence of multi-building-block replicators from dynamic combinatorial libraries. *J. Am. Chem. Soc.* **135,** 18406–18417 (2013).
17. We cannot exclude the possibility that the stack ends also play a catalytical role in the production of additional hexamer rings.
18. 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).
19. Westheimer, F. H. Coincidences, decarboxylation, and electrostatic effects. *Tetrahedron* **51,** 3–20 (1995).

20. Giger, L. *et al.* Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9,** 494–8 (2013).

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

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

23. Suh, J., Scarpa, I. S. & Klotz, I. M. Catalysis of decarboxylation of nitrobenzisoxazolecarboxylic acid and of cyanophenylacetic acid by modified polyethylenimines. *J. Am. Chem. Soc.* **98,** 7060–7064 (1976).

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

25. Frederix, P. W. J. M. *et al.* Structural and spectroscopic properties of assemblies of self-replicating peptide macrocycles. *ACS Nano* **11,** 7858–7868 (2017).

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

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

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

29. Methodol **2** and trimer macrocycle $\mathbf{1_3}$ co-eluted under our analytical conditions and could thus not be quantified.

30. Capozzi, G. & Modena, G. Oxidation of Thiols. in *The Chemistry of the Thiol Group* (ed. Patai, S.) 785–839 (Wiley: London, 1974).

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

32. For the kinetic analysis we conservatively estimate that each building block in the replicator fibres provides a single active site. From the fact that there is not recognizable plateau in the pH titration of the fibres (Extended Data Fig. 1d) at the p<i>.

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

34. Czárán, T., Könnyu, B. & Szathmáry, E. Metabolically Coupled Replicator Systems: Overview of an RNA-world model concept of prebiotic evolution on mineral surfaces. *J. Theor. Biol.* **381,** 39–54 (2015).

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

36. Nakano, T. & Yade, T. Synthesis, Structure, and Photophysical and Electrochemical Properties of a π-Stacked Polymer. *J. Am. Chem. Soc.* **125,** 15474–15484 (2003).

Data availability: The raw datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgements: We gratefully acknowledge a Marie Skłodowska Curie Individual Fellowship (project no. 751509) to C.M. and funding from Netherlands Organisation for Scientific Research (NWO, Veni grant 722.017.007 to C.M., Vici grant 724.012.002 to S.O.), as well as support from the ERC, COST (CM1304) and the Dutch Ministry of Education, Culture and Science (Gravitation program 024.001.035).

Author contributions: C.M and S.O. supervised the overall project. J.O. conceived, designed and performed the experiments on the retro-aldol reaction. A.S.H. conceived, designed and performed the experiments on the FMOC-glycine cleavage reaction. J.O, A.S.H, C.M. and S.O. co-wrote the paper. All authors discussed the results and commented on the manuscript.

Author information: Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.M. (c.mayer@rug.nl) or S.O. (s.otto@rug.nl).

Supplementary Materials for

Chance Emergence of Catalytic Activity in a Self-Replicator

Jim Ottelé, Andreas S. Hussain, Clemens Mayer*, Sijbren Otto*

*To whom correspondence should be addressed. E-mail: c.mayer@rug.nl, s.otto@rug.nl

This file includes:

SUPPLEMENTARY DISCUSSION

Estimating the fraction of catalytically active lysine residues for catalysis of the retro-aldol reaction by $\mathbf{1_6}$: Here we provide a comprehensive discussion of the results and the assumptions we made to obtain an estimate for the fraction of catalytically active lysine residues on a given fibre sample. First, for all kinetic characterizations of $\mathbf{1_6}$ fibres we assumed that: (1) active sites on the fibres are independent of each other, and (2) the measured activity can solely be attributed to lysine side chains in the active site. We define n ($0 < n < 2$) as the number of lysine residues per monomer (a maximum of two lysines per **1**) that contribute a catalytically active ε-amine.

In a first step to determining n, we performed a kinetic characterization, in which the concentration of **2** was varied (75 – 600 µM) in presence of 25 µM (with respect to building block **1**) $\mathbf{1_6}$ fibres (**Fig. 2C**). Under these conditions [**2**] >> [catalytic sites]. Performing a Michaelis-Menten analysis yields an apparent "k_{cat}" that corresponds to $k_{cat}*n$. Conversely, the obtained K_M value – which is not dependent on the concentration of catalyst, assuming a Michaelis Menten model – accurately reflects the K_M of individual active sites.

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

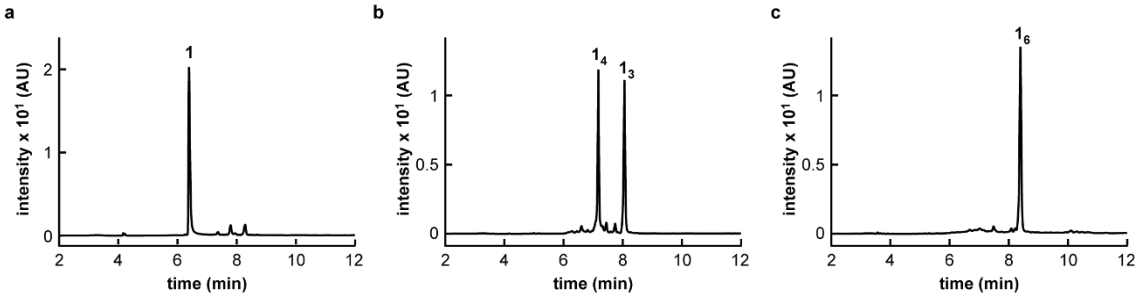
Conditions	$n*k_{cat}$ x 10^4 (s^{-1})	K_M (µM)	k_{cat} x 10^4 (s^{-1})	K_M / n (µM)	n
[2] >> [catalytic sites]	0.96*	670*			0.054 (from k_{cat})
[catalytic sites] >> [2]			17.8	4570	0.147 (from K_M)

* values obtained from the "burst" phase.

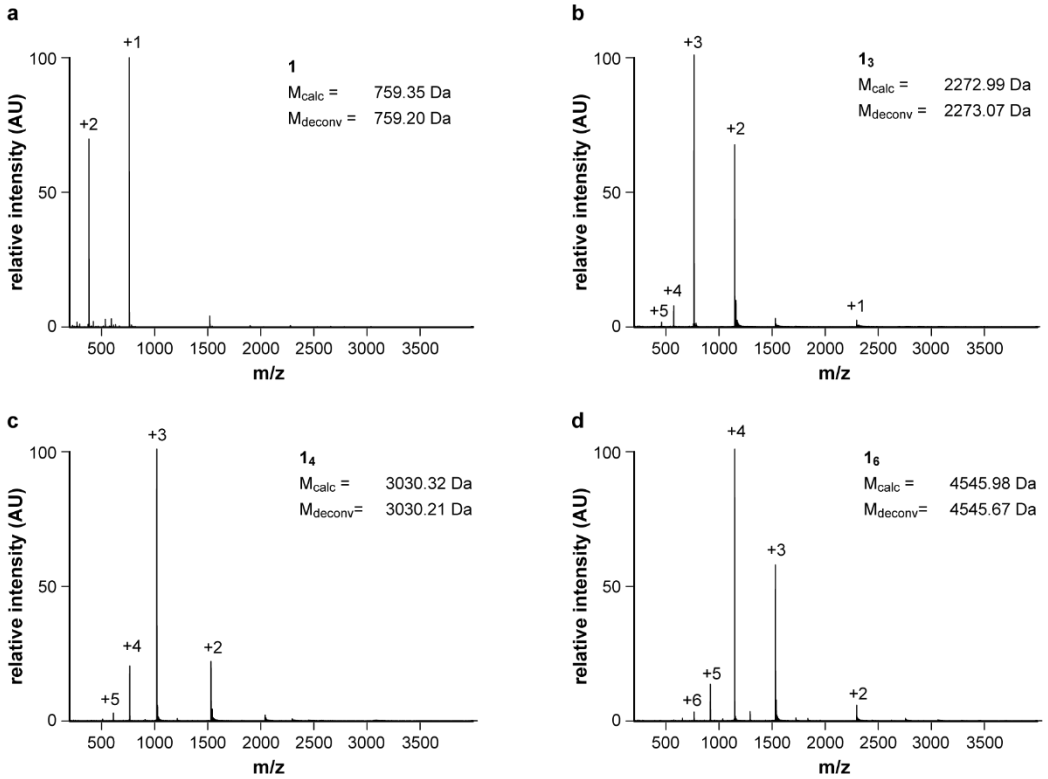
The kinetic parameters obtained from both experiments are summarized in the Table above. From these parameters, we can determine n, by either (1) $(n*k_{cat})/n$ or (2) $K_M/(K_M/n)$, yielding 0.054 and 0.147 respectively. These values correspond to 2.7% and 7.3% of all lysines or 5.4% and 14.7% of all building blocks on a typical fibre sample being catalytically active.

Both the limited solubility of **2** under the assay conditions and the limited availability of **1** prohibit performing reactions under conditions in which all available catalytic sites are saturated. Being unable to achieve saturation results in significant uncertainty for the parameters computed. Given these constraints we conservatively estimate <10% of all lysines to be catalytically active. The fact that we cannot identify an ionizable group in the pH titration of the fibres (**Fig. S3**) with a pK_a that corresponds to the one observed in the pH rate profile (**Fig. 2E**) also suggests that the fraction of lysines that are catalytically active is small.

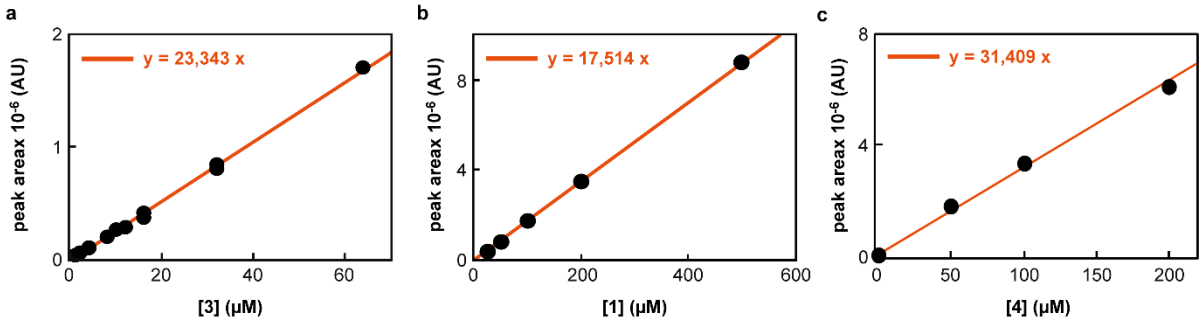
SUPPLEMENTARY FIGURES



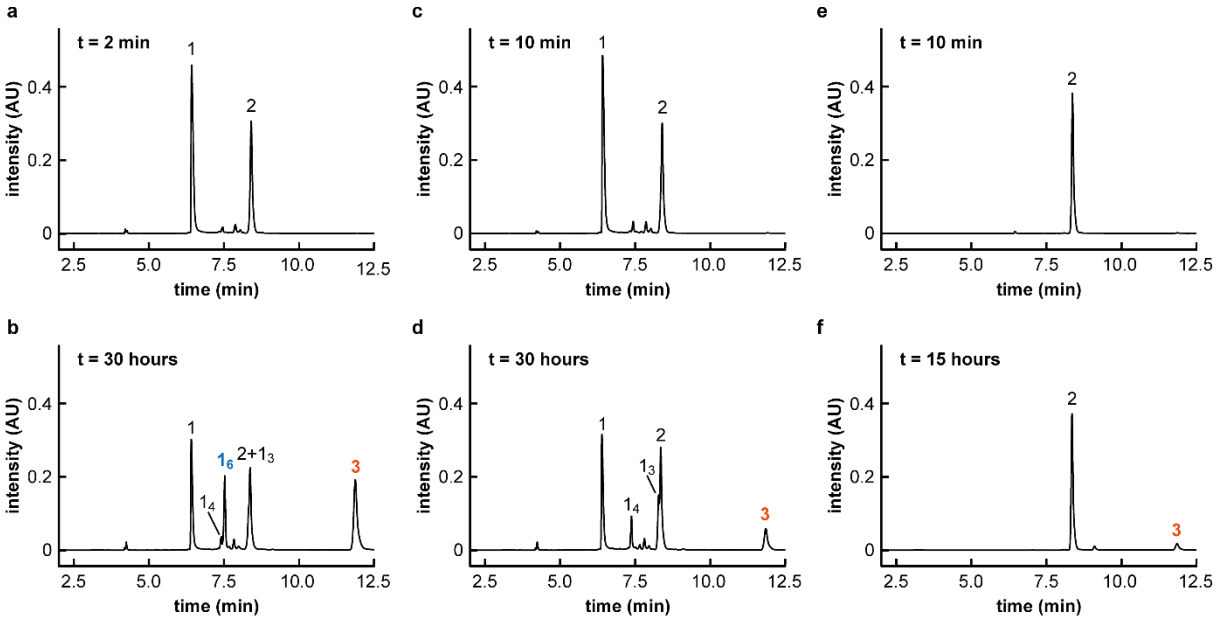
Supplementary Fig. S1: Representative UPLC chromatograms of **1** (**a**), a solution of **1** that was oxidized using sodium perborate containing a mixture of $\mathbf{1_3}$/$\mathbf{1_4}$ (**b**) and $\mathbf{1_6}$ (**c**) prepared as described in the Materials and Methods section.



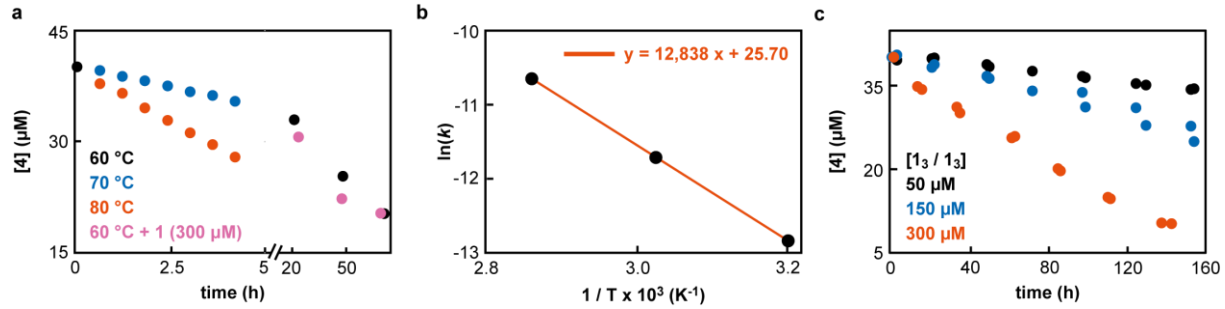
Supplementary Fig. S2: Representative mass spectra for **1** (**a**), $\mathbf{1_3}$ (**b**) $\mathbf{1_4}$ (**c**), and $\mathbf{1_6}$ (**d**) obtained for the samples shown in **Fig. S1**. Relevant charge states and the resulting deconvoluted masses (M_{deconv}) are indicated in the corresponding spectra.



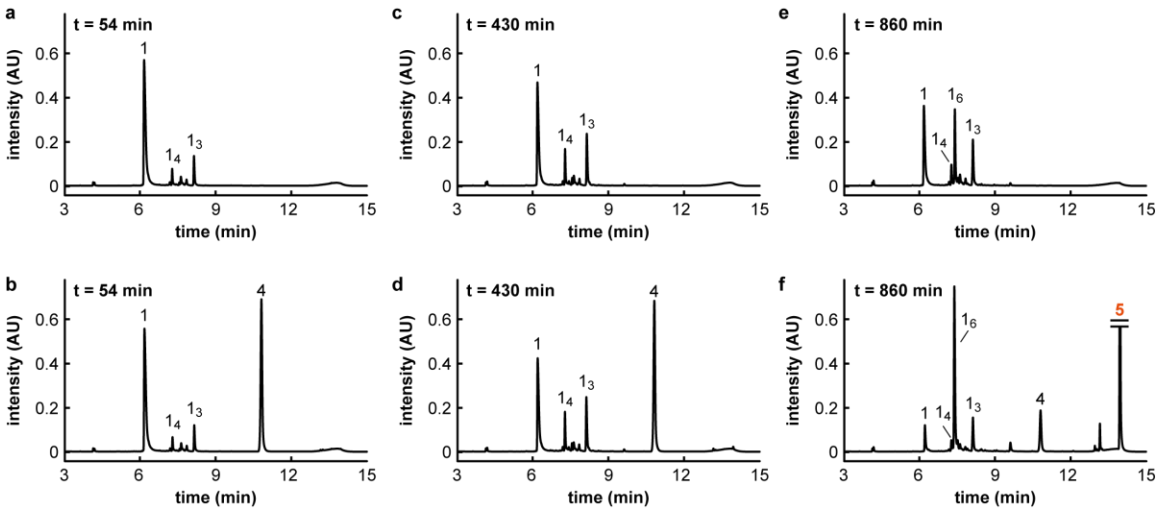
Supplementary Fig. S3: Calibration curves for **3** (**a**) and **1** (**b**), and **4** (**c**) obtained from subjecting samples with known quantities of these compounds to UPLC analysis. Linear fit and corresponding equation are shown in red.



Supplementary Fig. S4: Representative UPLC chromatograms near the start and end of the experiments shown in Fig. 2f and Extended Figure 2. From left to right: agitated sample (**a-b**), non-agitated control (**c-d**) and background (**e-f**, the experiment was stopped after 15 hours).



Supplementary Fig. S5: **(a)** Comparison of the reaction progress between the uncatalyzed FMOC-glycine cleavage at different temperatures and in presence of **1** (300 μM). **(b)** Arrhenius plot of the uncatalysed cleavage reaction that was used to extrapolate the rate constant of the reaction at 25 °C. Linear fit and corresponding equation highlighted in red. **(c)** Reaction progress in presence of $\mathbf{1_3}$/$\mathbf{1_4}$ (50, 150, 300 μM) at 25 °C.



Supplementary Fig. S6: Representative UPLC chromatograms near the start (54 min), near the emergence (430 min) and after emergence of $\mathbf{1_6}$ (860 min) for the experiment shown in Fig. 3f in absence (**a-c**) and presence (**d-f**) of **4.**

SUPPLEMENTARY TABLE

Table S1: Gradient used for UPLC separation of retro-aldol reaction mixtures (A=water containing 0.1% TFA, B=acetonitrile containing 0.1% TFA).

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