Collectivism in a Replication Network of Minimal Nucleobase Sequences

Sonia Vela-Gallego
Universidad Autonoma de Madrid
Zulay Pardo-Botero
IMDEA Nanociencia https://orcid.org/0000-0002-1472-4849
Cristian Moya-Álamo
Universidad Autonoma de Madrid https://orcid.org/0000-0001-5107-7751
Andrés de la Escosura (andres.delaescosura@uam.es)
Universidad Autónoma de Madrid https://orcid.org/0000-0002-0928-8317

Article

Keywords: replication, different, pathways, network, amino, cysteine, kinetic

Posted Date: November 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-927381/v1

Collectivism in a Replication Network of Minimal Nucleobase Sequences

Sonia Vela-Gallego,*[,a] Zulay Pardo-Botero,[b] Cristian Moya-Álamo,[a] Andrés de la Escosura*[,a,c]

[a] *Department of Organic Chemistry, Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid, Spain.*

[b] *IMDEA Nanociencia, C/ Faraday 7, 28049 Madrid, Spain.*

[c] *Institute for Advanced Research in Chemistry (IAdChem), Cantoblanco, 28049 Madrid, Spain.*

** Corresponding authors: S.V.G: sonia.vela@uam.es; A.d.l.E: andres.delaescosura@uam.es.*

Abstract. A major challenge for understanding the origins of life is to explore how replication networks can engage in an evolutionary process. Herein, we shed light on this problem by implementing a network constituted by two different types of extremely simple biological components: the amino acid cysteine and the canonical nucleobases adenine and thymine, connected through amide bonds to the cysteine amino group and oxidation of its thiol into three possible disulfides. Supramolecular and kinetic analyses revealed that both self- and mutual interactions between such dinucleobase compounds drive their assembly and replication pathways. Those pathways involving sequence complementarity led to enhanced replication rates, suggesting a potential bias for selection. The interplay of synergistic dynamics and competition between replicators was then simulated in an open reactor with experimental kinetic data, showing the selective amplification of different species depending on the initial mixture composition. Overall, this network configuration can favor a collective adaptability to changes in the availability of feedstock molecules, with disulfide exchange reactions serving as 'wires' that connect the different individual auto- and cross-catalytic pathways.

Research on life's origins constitutes a major multidisciplinary effort to unravel the physicochemical means by which living systems could emerge from non-living matter. Many questions remain open in the field, with implications that are both historical (how and where life originated), synthetic (how life can be synthesized from its basic molecular constituents), and conceptual (what essential features of living organisms allow characterization of their aliveness).[1-5] Systems chemistry is proving to be useful in this respect, as it adopts a holistic view for the study of complex chemical systems, wherein dynamic out-of-equilibrium reaction and self-assembly processes govern the system´s emergent behaviors.[6-8] An important line in this area involves the development of chimeric systems that combine the properties of distinct biological building blocks, as a step towards replication, protometabolic networks and protocellular assemblies.[9-13]

In the endeavor to mimic DNA's capacity for replication or, more generally, the capacity of living cells to self-reproduce, different forms of replication have been developed with both synthetic and biological molecules.[14-18] The literature is rich in processes that display *autocatalysis*, either through the product's catalysis of its own formation,[19] *cyclic autocatalysis*,[20,21] or in oscillatory reactions.[22] Most of these autocatalytic transformations cannot be considered self-replication, since they lack the specificity required for information transfer at the molecular level.[18] In search for such specificity, *template replication* has been proven with different kinds of biopolymers and oligomers, including DNA,[23,24] RNA[25,26] and oligopeptides,[27,28] as well as with synthetic molecules not present in extant biology.[29] However, this type of mechanism tends to halt the replication process due to an excessively strong binding (and therefore inhibition) of the template and product molecules, which handicaps efforts to achieve exponential product growth. *Network autocatalysis* has been proposed as an alternative,[30,31] with both theoretical and experimental models based on lipids,[32,33] peptides,[34,35] nucleic acids,[36] and synthetic molecules.[37] Autocatalysis in these networks is normally associated with self-assembly of the replicating species, most commonly into hybridized strands, fibers, micelles or vesicles, the latter being relevant to the formation of self-reproducing

compartments.[32,38] This type of replication was likely widespread in prebiotic scenarios, but in order to trigger subsequent evolution, replicators would have to acquire additional capacities including:[39] (i) catalysis, to establish a supportive metabolism;[40] (ii) performance out of equilibrium;[35,41,42] (iii) compartmentalization, to avoid parasitic reactions and dilution effects;[43] or (iv) variability control with nucleobase sequences.[44] Important open questions with respect to nucleobase sequences are how simple can be their constituent monomers, and what is the minimal sequence length that can drive the emergence of replication networks.

To shed light on those issues, herein we describe a new family of very simple exponential replicators emerging from monomers that display adenine or thymine (**A** and **T**), connected through amide bonds to the amino groups of cysteine (Figure 1). The role of the amino acid in these molecules is to link the nucleobases in a sequence, through oxidation of its reactive thiol into dynamic disulfide bonds. Despite the short length of these nucleobase sequences, supramolecular studies showed that they are able to control the self-assembly of the three formed species (**AA**, **TT** and **AT**), thus determining their replication efficiency. In-depth kinetic experiments and simulations were used to study how the resulting aggregates affect the irreversible auto- and cross-catalytic oxidation pathways of **A** and **T**, and the concomitant reversible disulfide exchange reactions. In spite of the low complexity of the studied replicators, both in terms of the monomers structure (much simpler than that of ribonucleotides) and of the sequence length (dimers), complementarity of nucleobases enhanced the replication rate for both the auto- and cross-catalytic pathways (**AT** and **AA/TT**, respectively), suggesting an adaptive potential that involves the interplay of different collective and competitive dynamic interactions between them.

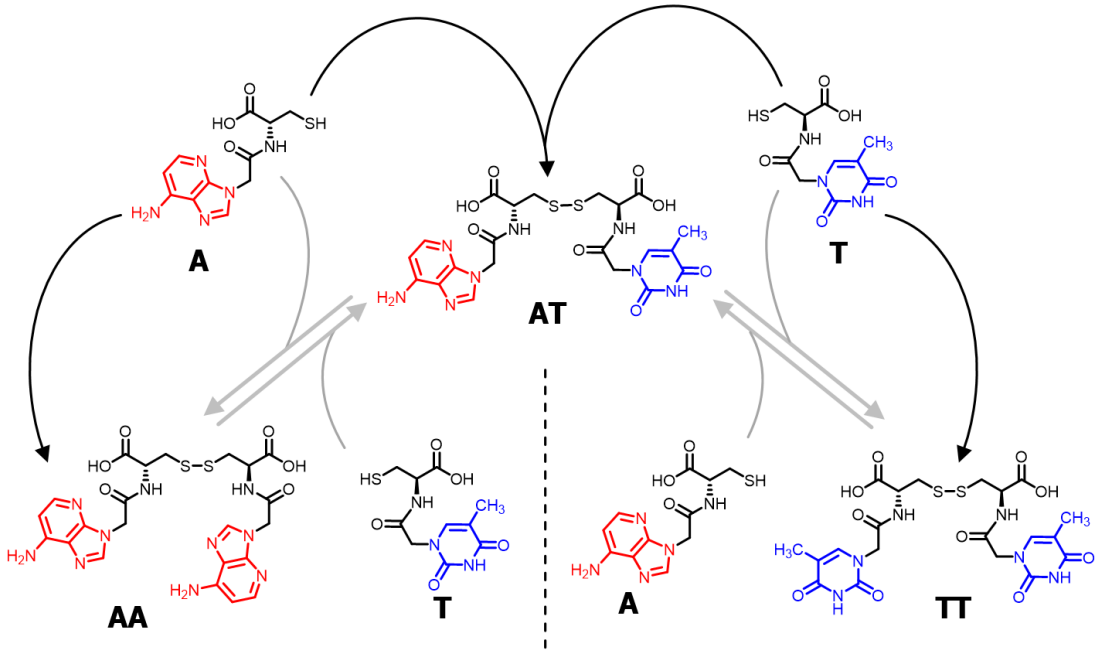


Figure 1. *Set of building blocks and transformations that constitute the reaction network under study.* ***A*** *and* ***T*** *are thiol monomers with a single nucleobase, and therefore lack the potential to self-assemble. Black arrows represent the oxidation of* ***A*** *and* ***T*** *into the disulfide dimers* ***AA***, ***TT*** *and* ***AT****, which can occur in a non-catalyzed manner or alternatively via autocatalysis provided that they are in sufficient concentration to form 'catalytic' aggregates. Grey arrows represent disulfide exchange reactions.* ***AA*** *and* ***TT*** *can also replicate through cross-catalysis based on the complementarity of their nucleobase sequences.*

Results

Synthesis and self-assembly of the network components. The network building blocks (**A** and **T**) and the corresponding disulfide homodimers were synthesized and characterized as described in the supporting information (SI, Scheme S1). ^{1}H-NMR revealed a slow oxidation into disulfides in DMSO (Figure S1), completely stopped in acidic water (Figure S2 bottom), and fast in aqueous basic medium (pH 11; see Figure S2, top). One of the requisites to establish base pairing interactions between adenine and thymine, however, is that they stay non-ionized.

pH titrations, monitored by ^{1}H-NMR, confirmed pK_a values of 4.14 ± 0.02 for the protonated adenine moiety in **AA** (Figure S3A), and of 9.62 ± 0.03 for the dissociation of the thymine imide NH in **TT** (Figure S3B). It can be assumed that similar pK_a values are applicable to **AT**, which could not be separated from the homodimers when generated in a mixture of **A** and **T**. On these bases, and considering that a slightly basic pH is optimum for disulfide formation/exchange,[34] all the subsequent self-assembly and replication experiments were run in 50 mM borate buffer at pH 8.2.

The existence of aggregation under such conditions was demonstrated through diffusion-ordered spectroscopy (DOSY) for samples containing **AA**, **TT** or an equimolar mixture of **AA/TT** at different concentrations (from 0.1 to 4 mM). The diffusion coefficient (*D*) was calculated for all samples, through monitoring of the monoexponential attenuation of NMR signals during a pulsed field gradient experiment, followed by plotting the obtained *D* values versus concentration (Figure 2A-C) to determine the critical aggregation concentration (*cac*). Datasets with two clearly differentiated linear regions were obtained for **AA** and **TT**, with the one above the *cac* showing a larger and constant *D* value, indicative of the presence of higher-order aggregates. The intersection between the two lines pointed to specific *cac* values of 0.51 and 0.9 mM, respectively. For the **AA/TT** mixture, a different data distribution was obtained, with a sharp transition between the aggregated and non-aggregated states. A *cac* value of 0.47 mM was obtained by fitting the data with a Boltzmann equation (R^2 = 0.9993) to detect the slope changing point, suggesting that nucleobase sequence complementarity induces a stronger and cooperative self-assembly.

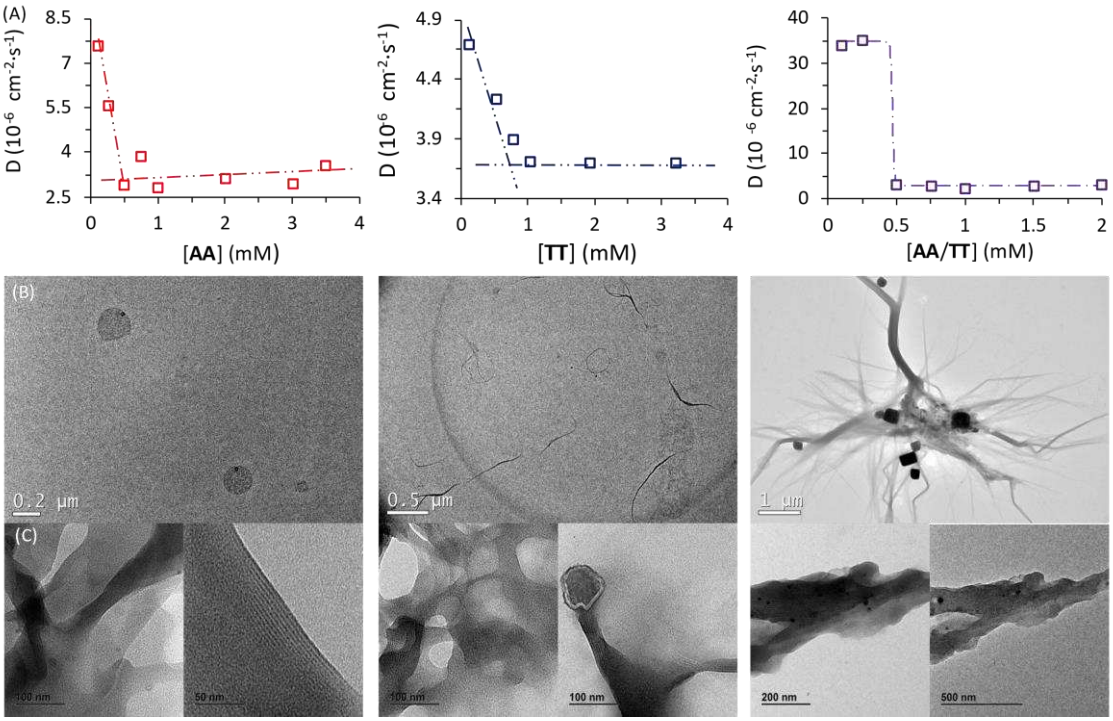


Figure 2. *Supramolecular studies of compounds* ***AA***, ***TT***, *and an equimolar mixture of* ***AA/TT***. *(A) Plot of the diffusion coefficient (D) obtained by DOSY experiments versus the concentration of* ***AA*** *(left),* ***TT*** *(middle) and* ***AA/TT*** *(right). Critical aggregation concentration (cac) values were calculated as the intersection of two straight lines in the plots for* ***AA*** *and* ***TT***, *and through fitting to a Boltzmann-type equation for* ***AA/TT***. *(B, C) TEM micrographs of* ***AA*** *(left),* ***TT*** *(middle) and* ***AA/TT*** *(right) at two different concentrations: 1 mM (B) and 2 mM (C) in 50 mM borate buffer (pH 8.2).*

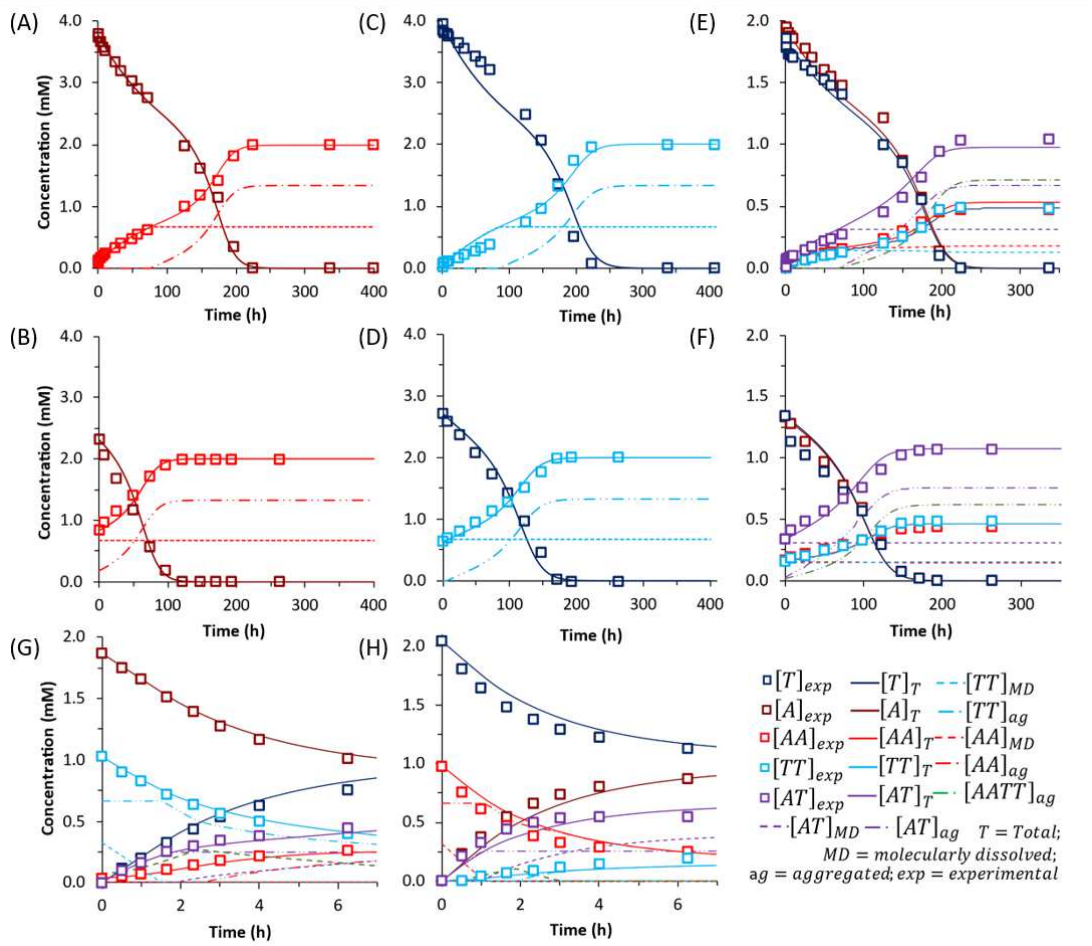
To assess the morphology of those aggregates, TEM studies were conducted at two different disulfide concentrations: 1 and 2 mM (Figure 2B and 2C, respectively). At 1 mM, the low density of objects over the grid indicated minor aggregation of **AA** and **TT**, showing spherical assemblies for the former (Figure 2B left / S5A), yet they were very exceptional and may not be representative of the main self-assembly pathway, and isolated fibrils for the latter (Figure 2B middle / S6A). Despite their low abundance, the presence of sulfur detected by energy dispersive x-ray spectroscopy (EXS) confirmed that such structures were formed by the corresponding disulfide compounds. For the **AA/TT** mixture, intertwined fiber assemblies were observed over the whole grid (Figure 2B right / S7A), in a significantly higher abundance than

for **AA** or **TT** alone. This points to the importance of complementary nucleobase pairing to induce aggregation in the mixture. At the high concentration, aggregation was much more prominent (Figure 2C) and resulted in lamellar structures for all cases, probably as a result of the hierarchical assembly of fibers into sheets. The distance between adjacent sheets could be clearly determined in areas where the lamellar arrangement was perpendicular with respect to the grid surface, yielding similar values (~4 nm) for **AA** and **TT** (Figure S5B/S6B bottom). For **AA/TT**, the stronger aggregation gave rise to thicker stacks of sheets (Figure 2B right / S7B) that did not permit estimation of the interlamellar distance. In any case, the formation of different assemblies for the three systems, including fibers and lamellar structures depending on concentration, points to a complex assembly landscape, with contribution from hydrogen bonding interactions, nucleobase π-π stacking and hydrophobic effects. The involved self-assembly mechanisms will be examined in-depth in subsequent studies but, since previous work has demonstrated the capacity of fibrillar and sheet assemblies to facilitate replication processes,[34,45] we assume these to be the catalytically active ones also in the present case.

Replication experiments. The initial replication experiments were performed with only one monomer, either **A** or **T**, at 4 mM concentration. The reaction kinetics were monitored through reverse phase high-performance liquid chromatography (HPLC, see experimental section and Figures S8-S11), coupled to electrospray ionization mass spectrometry (ESI-MS) for identification of species. For the non-templated oxidation reactions, Figures S8/S9 show a gradual decay of monomer and growth of dimer elution peak areas over the course of the experiment. Calibration curves were performed for **A**, **T**, **AA** and **TT** in order to quantify their molar concentrations in each aliquot (Figures S12-S15). The kinetic profiles, plotted from the obtained concentrations at different reaction times, showed typical features of autocatalysis for both **AA** and **TT** (Figure 3A, B): an initial slow growth of product (induction period) followed by a phase of faster growth (autocatalysis) until full conversion of monomer into dimer. Importantly, the change in curve slope was observed at a product concentration that approximately matches its *cac* value, as determined by DOSY, proving that there is no

autocatalysis in the absence of aggregates (below the *cac*). To confirm the products' autocatalytic nature, seeded experiments with 30% of **AA** or **TT** were conducted, maintaining the total concentration of starting materials in the same range as in the non-templated reactions. In both cases, a shortening of the induction period and an overall decrease in the reaction time was observed (Figure 3C, D; while Figure S16A and S16B shows the direct comparison of seeded and non-seeded experiments), indicating that they actually contribute to increase the reaction rate. This effect was less prominent for **TT**, probably due to its lower tendency to aggregate (higher *cac*).

Similar results were obtained when conducting the reactions from a mixture of **A** and **T** (2 mM each). Figure 3E depicts the kinetic profiles obtained from HPLC monitoring over time (Figure S10), revealing the characteristic features of replication kinetics for both **AA**, **TT** and **AT**. Experiments seeded with 20% of a previously finished reaction resulted in a shortening of the induction period for the three replicating species (Figure 3F, S16C and S16D). However, the scenario becomes significantly more complex when the two nucleobases are present, as disulfide exchange reactions can also occur. To study the role of these exchange processes in the global network kinetics, two different reactions were performed (4 mM **T** + 2 mM **AA** and 4 mM **A** + 2 mM **TT**) with HPLC monitoring (Figure S11). In the obtained kinetic curves, two stages could be distinguished, the first one corresponding to a preeminent role of disulfide exchange during the first 7 h of the reaction (Figure 3G and 3H).



172

173 ***Figure 3.*** *Kinetic profiles corresponding to: (first row) the autocatalytic formation of* ***AA*** *(A),*
174 ***TT*** *(C), and the mixture of possible replicators* ***AA/TT/AT*** *(E); (second row) seeded*
175 *experiments for the autocatalytic formation of* ***AA*** *(B),* ***TT*** *(D), and* ***AA/TT/AT*** *(F); (third row)*
176 *disulfide exchange reactions (up to 7 h) between* ***AA*** *and* ***T*** *(G) or* ***TT*** *and* ***A*** *(H). Each panel*
177 *shows the evolution of all involved species over time through experimental data (square data*
178 *points) and fit curves, while the set of kinetic equations used for fitting are depicted in Table 1.*
179 *In the templated reactions (A, C, E), the total concentration of starting materials was*
180 *maintained in the same range as in the non-templated ones (B, D, F), with minor experimental*
181 *deviations that are difficult to avoid but are considered in the mathematical analysis. The non-*
182 *seeded experiments were repeated three times, while seeded experiments and disulfide exchange*
183 *reactions were repeated twice (Figures S16-S19). The treatment of fitting errors with such*
184 *repeats is shown in section 6 of the SI.*

Kinetic analyses. The above replication data were used to analyze the contribution of all involved processes in the global network kinetics. Two different reactions were initially considered for the formation of each disulfide homodimer: the non-catalyzed (**R-A1**, **R-T1**) and the autocatalytic oxidation (**R-A2**, **R-T2**) of the corresponding monomeric thiol (Table 1, boxes 1 and 2). In the reaction schemes, $\mathbf{AA_{ag}}$ and $\mathbf{TT_{ag}}$ refer to the self-assembled products, and emerge from the aggregation processes **R-A3** and **R-T3**. For calculation of kinetic constants, the concentration of aggregated replicator is needed at any given time. According to Chen *et al.*,[46] the concentrations of $\mathbf{AA_{ag}}$ and $\mathbf{TT_{ag}}$ are given by equation **Eq-S1** (Section 5 in the SI), in which the total concentration of aggregating compound (C_T) can be expressed as a function of C_1 (the concentration of molecularly dissolved compound), ρ (a parameter related to the reaction mechanism) and K_{eq} (the equilibrium constant of the aggregation process). This equation can become complex to solve, but the problem is approachable considering that when C_T is below $1/K_{eq}$, most molecules are in the monomeric form, whereas molecules aggregate rapidly if C_T exceeds that value. An equivalence can therefore be assumed between $1/K_{eq}$ and the *cac,* an assumption that is valid for any possible aggregation mechanism. In Figure S20, for example, when the analytical solutions of **Eq-S1** for different values of ρ are compared to the proposed simplification, the outcome differences are negligible.

The rate equations for non-catalyzed (**Eq-A1** and **Eq-T1**) and autocatalytic regimes (**Eq-A2** and **Eq-T2**) were defined in a MATLAB program (Table 1, boxes 1 and 2), establishing that they operate below and above the *cac*, respectively. For autocatalytic processes, the mechanism of catalysis was not known and so equations with different orders with respect to both **A** (or **T**) and $\mathbf{AA_{ag}}$ (or $\mathbf{TT_{ag}}$) were evaluated through fitting of the experimental data. The equilibrium constants of aggregate formation were considered through equations **Eq-A3** and **Eq-T3**. The best fitting curves can be seen in Figures 3A-D (with R^2 above 0.99 in all cases), and correspond to a global order of three for the autocatalytic stage (**Eq-A2** and **Eq-T2**) and two with respect to the aggregated replicator. For a complete statistical treatment of fitting errors, see section 6 of the SI, Tables S1-S3 and Figures S21 and S22. The fact that the best fittings

were obtained for a reaction order of two with respect to the replicating species (therefore higher than one) points to an exponential growth.[34,47] The orders obtained in the rate equations **Eq-A2** and **Eq-T2** actually imply that in the 'catalytic' hybrid assemblies of monomeric thiol and disulfide dimer, the required ratio between both for the monomer to get activated towards oxidation is of 1:2. Further studies will however be devoted to propose a solid mechanistic scheme of this replication process. In any case, it is worth mentioning that the kinetic constants for the catalyzed reactions were about one order of magnitude greater than the non-catalyzed ones. The calculated equilibrium constants (K_{AAag} = 1.95 mM^{-1} and K_{TTag} = 1.6 mM^{-1}) in turn led to *cac* values of 0.51 and 0.63 mM, respectively, which are close to those obtained from DOSY experiments.

A similar procedure was applied for calculation of disulfide exchange kinetic constants, considering four possible reactions/equations (**R-E1** to **R-E4 / Eq-E1** to **Eq-E4** -- Table 1, box 3; fitting curves in Figure 3G, H) and all previously calculated constants (k_A, k_T, k_{Ac}, k_{Tc}, K_{AAag} and K_{TTag}). The constants resulting upon fitting (k_{e1}, k_{e2}, k_{e3}, k_{e4}) are approximately one order of magnitude larger than those for the autocatalytic oxidation of monomeric thiols, which makes their reaction rates comparable. Concerning the analysis of replication from mixtures of both nucleobase monomers, the landscape of non-covalent assembly pathways is more complex than for single replicators. In addition to the self-assembly of **AA** and **TT**, two other aggregate types must be considered, resulting from either complementary interaction between **AA** and **TT** (**AATT**$_{ag}$) or from self-pairing of **AT** (i.e., **AT**$_{ag}$). These aggregates enable auto- and cross-catalytic reactions, as **AATT**$_{ag}$ can aid in reactions producing **AA** and **TT** (**R-A4** and **R-T4**, respectively – Table 1, box 4), and **AT**$_{ag}$ can assist in its own formation (**R-AT2**) (Table 1, box 5). For mathematical fitting of these processes (Figure 3E, F), all constants concerning the non-catalyzed and autocatalytic formation of **TT** and **AA**, the disulfide exchange reactions, the aggregation constants of **AA** and **TT** and the order of those reactions with respect to aggregates in the autocatalytic regime were used as previously calculated.

Table 1. *Kinetic analysis of the replication network. Boxes 1 and 2 concern the irreversible reactions (either non-catalyzed or autocatalytic) of oxidation of* ***A*** *and* ***T*** *into* ***AA*** *and* ***TT****, respectively, and the equilibrium of aggregation of the latter species, as the resulting aggregates are involved in the autocatalysis. Box 3 includes the four possible disulfide exchange reaction steps. Boxes 4 and 5 refer to the cross-catalysis of* ***AA/TT*** *and the autocatalysis of* ***AT****, respectively. In all boxes, the table details each of the involved processes, the equations that govern them and the values of the resulting rate/equilibrium constants. The order of the reactions, with respect to the monomers and the catalytic aggregates, in equations* ***Eq-A2****,* ***Eq-T2****,* ***Eq-A4****,* ***Eq-T4*** *and* ***Eq-AT2*** *was determined through fitting of the experimental data into rate equations with different orders (see Tables S1-S3), selecting those that led to the lowest mean absolute percentage errors and* R^2 *values > 0.99. Dispersion graphs for the different fits (Figures S21 and S22) helped confirming the selected rate equations.*

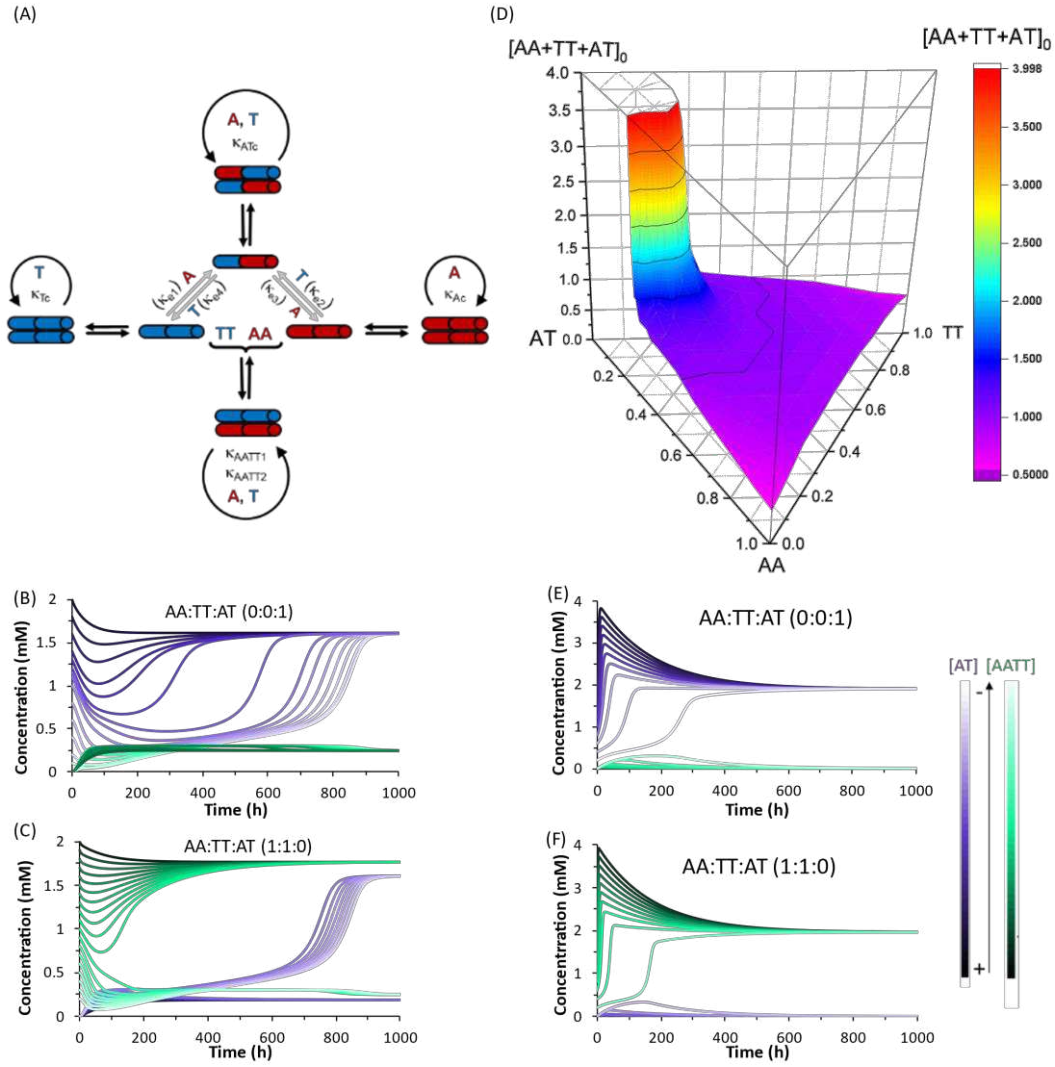
		Reaction		Equation	Constant
Box 1	R-A1	$A + A \rightarrow AA$	Eq-A1	$k_A[A]^2$	$k_A = 6.62 \cdot 10^{-4}$ mM^{-1} h^{-1}
	R-A2	$A + A \xrightarrow{AA_{ag}} AA$	Eq-A2	$k_{Ac}[A]\left[AA_{ag}\right]^2$	$k_{Ac} = 1.38 \cdot 10^{-2}$ mM^{-2} h^{-1}
	R-A3	$AA + AA \rightleftarrows AA_{ag}$	Eq-A3	$K_{eq_{AAag}} = 1/[AA]_{cac}$	$K_{eq_{AAag}}$ =1.95 mM^{-1}
Box 2	R-T1	$T + T \rightarrow TT$	Eq-T1	$k_T[T]^2$	$k_T = 6.89 \cdot 10^{-4}$ mM^{-1} h^{-1}
	R-T2	$T + T \xrightarrow{TT_{ag}} TT$	Eq-T2	$k_{Tc}[T]\left[TT_{ag}\right]^2$	$k_{Tc} = 1.73 \cdot 10^{-2}$ mM^{-2} h^{-1}
	R-T3	$TT + TT \rightleftarrows TT_{ag}$	Eq-T3	$K_{eq_{TTag}} = 1/[TT]_{cac}$	$K_{eq_{TTag}}$ =1.6 mM^{-1}
Box 3	R-E1	$A + TT_d \xrightarrow{K_{e_1}} T + AT_d$	Eq-E1	$k_{e_1}[A][TT]_{MD}$	$k_{e_1} = 0.15$ mM^{-1} h^{-1}
	R-E2	$T + AA_d \xrightarrow{K_{e_2}} A + AT_d$	Eq-E2	$k_{e_2}[T][AA]_{MD}$	$k_{e_2} = 0.21$ mM^{-1} h^{-1}
	R-E3	$A + AT_d \xrightarrow{K_{e_3}} T + AA_d$	Eq-E3	$k_{e_3}[A][AT]_{MD}$	$k_{e_3} = 0.11$ mM^{-1} h^{-1}
	R-E4	$T + AT_d \xrightarrow{K_{e_4}} A + TT_d$	Eq-E4	$k_{e_4}[T][AT]_{MD}$	$k_{e_4} = 0.07$ mM^{-1} h^{-1}
Box 4	R-A4	$A + A \xrightarrow{AATT_{ag}} AA$	Eq-A4	$k_{AATT1}[A][AATT]_{ag}^2$	$k_{AATT} = 5.86 \cdot 10^{-2}$ mM^{-2} h^{-2}
	R-T4	$T + T \xrightarrow{AATT_{ag}} TT$	Eq-T4	$k_{AATT2}[T][AATT]_{ag}^2$	$k_{AATTc} = 6.02 \cdot 10^{-2}$ mM^{-2} h^{-2}
	R-AATT	$AA + \mathrm{TT} \rightleftarrows AA\mathrm{TT}_{ag}$	Eq-AATT	$K_{eq_{AATTag}} = 1/[AATT]_{cac}$	$K_{eq_{AATTag}}$ =3.27 mM^{-1}
Box 5	R-AT1	$A + T \rightarrow AT$	Eq-AT1	$k_{AT}[A][T]$	$k_{AT} = 1.54 \cdot 10^{-3}$mM^{-1} h^{-1}
	R-AT2	$A + T \xrightarrow{AT_{ag}} AT$	Eq-AT2	$k_{AT_c}[A]^{0.5}[T]^{0.5}[AT]_{ag}^2$	$k_{AT_c} = 4.69 \cdot 10^{-2}$ mM^{-2} h^{-2}
	R-AT3	$AT + \mathrm{AT} \rightleftarrows AT_{ag}$	Eq-AT3	$K_{eq_{ATag}} = 1/[AT]_{cac}$	$K_{eq_{ATag}}$ =3.25 mM^{-1}

This global analysis of the network kinetics revealed interesting aspects of its behavior. As expected, for statistical reasons related to the number of pathways to its formation, **AT** is produced twice as fast as **AA** or **TT**. More importantly, the auto-/cross-catalytic reactions were ~4x faster, and the aggregation constants 2x higher when there is complementarity between nucleobase sequences (i.e. for **AT** and **AA/TT**) than when there is not (i.e., for **AA** or **TT**). Indeed, according to the calculated K_{ag} values for both $\mathbf{AATT}_{ag}$ and $\mathbf{AT}_{ag}$ (3.3 mM^{-1}), their *cac* would correspond to 0.3 mM (total concentration of disulfides), in agreement with the DOSY data. These minimal nucleobase sequences therefore seem to enhance their replication through complementary base pairing.

Interplay between synergistic/self-replication pathways. According to the above kinetic analysis, the network topology presents two dominant competing pathways: the autocatalytic replication of **AT** (Figure 4A, top cycle); and the synergistic assembly of **AA/TT**, which results in cross-catalysis towards their common formation (Figure 4A, bottom cycle). In a closed reactor, fast disulfide exchange contributes to balance both pathways, leading to a statistical mixture of the three replicating species. In contrast, in an open reactor, the asymmetry between both competing pathways, together with their possible interconversion through exchange reactions, may lead to adaptive behaviors.[48] To predict the evolutionary potential of this network topology in open environments, a continuous perfect mixing reactor (section 5 in the SI) was modeled in excel software using the previously determined kinetic data. In the studied configuration, the 20 mL reactor could be initially loaded with any possible combination of the three species and continuously fed with monomers **A** and **T** (4 mM) through two different input streams of 1 µL/min, while maintaining a constant reactor volume through the extraction of an equivalent output current.

When starting from an empty reactor ($[\text{replicators}]_o = 0$), the system evolved as expected into a steady state dominated by **AT** (SS_{AT}), where [**AT**] = 1.61 mM and [**AA/TT**] = 0.24 mM, the latter meaning total disulfide concentration (Figure 4B, lightest color curves). A significant decrease in the time needed to reach SS_{AT} was observed when loading the reactor with

278 increasing initial **AT** concentrations (Figure 4B). SS_{AT} was also obtained when the reactor was
279 filled with **AA/TT** below a threshold concentration (see below). When $[\mathbf{AA/TT}]_o$ was above
280 that threshold, however, a new steady state appeared ($SS_{AA/TT}$) where **AA/TT** were the dominant
281 species ([**AA/TT**] = 1.77 mM) and **AT** decreased in concentration ([**AT**] = 0.19 mM) (Figure
282 4C and S23). A range of initial proportions of **AA**, **TT** and **AT**, and of total initial disulfide
283 concentration (from 0 to 4 mM), was then tested to determine which steady state would be
284 reached in each case. The result was a 3D surface (Figure 4D) that marks the boundary between
285 initial conditions that favor SS_{AT} (below the surface) or $SS_{AA/TT}$ (above the surface).
286 Interestingly, the threshold below which **AT** always gets amplified was ~1 mM; above that
287 threshold, $SS_{AA/TT}$ was favored except when the initial molar fraction of **AT** was higher than 0.8.
288



289

290 ***Figure 4***. *(A) Replication network topology. (B, C) Simulations of the network evolution when*
291 *fed with two input streams (Q_{in} = 1 µL/min) of* ***A*** *and* ***T****, in the presence of replicator* ***AT*** *(0 – 2*
292 *mM) (B) or* ***AA/TT*** *(0 – 2 mM) (C). (D) 3D surface representing the reached steady state (SS_{AT}*
293 *below the surface; $SS_{AA/TT}$ above) for a range of initial replicator proportions (horizontal axes*
294 *on the graph) and total replicator concentrations (vertical axis). The surface marks the*
295 *boundary between initial conditions that favor SS_{AT} (below the surface) or $SS_{AA/TT}$ (above). (E,*
296 *F) Simulations of the network evolution when decreasing the exchange constants by six orders*
297 *of magnitude, with the same input streams and conditions as in (B) and (C).*

298 The capacity of a dynamic system to reach two different steady states depending on the
299 initial conditions is called bistability, and in the present case it seems to be related to their
300 possible interconversion through disulfide exchange. To test this possibility, the network

behavior was simulated in a hypothetical scenario where the exchange reactions were kinetically frozen, artificially reducing the values of their kinetic constants by six orders of magnitude compared to the experimental data (see Table S4). In that situation, the dominant species in the reached steady state correlated exclusively with the replicating system (**AT** or **AA/TT**) initially present (Figure 4E and 4F). In addition, the concentration of the 'losing' replicative system drops almost to extinction. The reason for this is that, in the absence of exchange reactions, the dominant catalytic species gets amplified sufficiently quickly to consume all the substrates fed into the reactor. This simulation thus proves the importance of disulfide exchange as wiring reactions that connect the different auto- and cross-catalytic pathways, endowing the whole replication network with a collectively better adaptability, as it can switch from one replicating system to the other if the conditions are favorable.

Conclusions

The results of this work underscore the likelihood of replication networks emerging in conditions and from building blocks with reasonable prebiotic plausibility. The replicating species are built from the amino acid cysteine and two canonical nucleobases, which have been reported in Miller-type experiments,[49] Strecker-derived chemistry[50] and HCN/cyanoacetylene oligomerization reactions.[51] Although for practical reasons our synthesis of **A** and **T** was performed following standard organic synthesis techniques, the chemistry of amide and disulfide bond formation/exchange has been extensively studied in prebiotic contexts.[2] More importantly, the molecular complexity of the replicators **AA**, **TT** and **AT** is significantly less than that of other different replicator families reported to date, suggesting that the structural requirements for chemical evolution to step into replicating species was probably not so high. The present cysteine-based derivatives do not need an oligopeptide or lipid chain to drive their self-assembly and replication processes. On the other hand, the smallest nucleic acid template replicators previously described required a minimum sequence of 6 nucleotides.[15] As a merge of both approaches, the self-assembly of **AA**, **TT** and **AT** must be mostly promoted by H-bond

328 interactions between nucleobases, together with π-π stacking and hydrophobic effects. Despite
329 this makes their self-assembly weaker and the replication rates slower compared to previous
330 peptide- or lipid-based replicators, they gain a rudimentary sequence-based control of the
331 replication process. Building on such a capacity, the network described herein presents a
332 collective behaviour that can provide significant adaptability between the individual synergistic
333 and 'selfish' replication pathways, aided by exchange reactions that allow interconversion
334 between the different replicator species.

335

336 ## Methods

337 Complete procedures for the synthesis and characterization of the network components are
338 described in the SI.

339 ***Self-assembly and replication experiments.*** A borate-buffered solution (200 mM, pH 8.2) was
340 employed for all the self-assembly and replication experiments. Boric acid (20 mmol, 1.24 g)
341 was dissolved in H_2O (100 ml), or D_2O for the DOSY experiments, and the pH was adjusted to
342 8.2 with 1 M aq. NaOH (or NaOD). This buffer was used as a stock to prepare all samples,
343 adjusting the borate concentration to 50 mM in all cases.

344 ***Transmission electron microscopy and energy dispersive x-ray spectroscopy*** was performed in
345 a JEOL JEM-2100 electron microscope (JEOL Ltd., Tokyo, Japan) operated at 200 kV,
346 preparing samples as follows: 5 μL of sample solution were applied to glow discharged
347 formvar/carbon-coated grids. Images were acquired with a CCD ORIUS SC1000 camera.

348 ***DOSY NMR.*** Different solutions of **AA, TT** or equimolar mixtures of **AA/TT** (0.1, 0.25, 0.5,
349 0.75, 1, 2, 3 and 4 mM total concentration of disulfide) were prepared in D_2O-based borate
350 buffer, and the pH was readjusted to 8.2. The DOSY measurements were performed using the
351 longitudinal eddy current (LED) delay pulse sequence. The duration of the magnetic field pulse
352 gradient (small delta, δ) was 2.8 ms and the diffusion delay (big delta, Δ) was 100 ms in order to
353 obtain less than 3% residual signal with the maximum gradient strength. The number of

accumulated scans (ns) was set between 32 and 80 depending of the sample concentration. The pulse gradients were incremented in ns steps from 2% to 95% of the maximum gradient strength (53.5 G/cm) in a linear ramp. The Eddy Current delay (*te*) and the pulse separation (*ts*) were set at 5 and 0.2 ms, respectively, in all experiments. For details on the calculation of *D*, see section 2 in the SI.

Replication experiments. A solution of **A** or **T** (5.3 mM) in water (for the one-component autocatalytic reactions), or an equimolar mixture of **A** and **T** (2.7 mM each for the two-component replication processes) was vortexed for 1 min, followed by addition of borate buffer (200 mM) until dilution to 4 mM of monomer (total cysteine concentration) and 50 mM of buffer. The mixture was vortexed for 1 min, and the pH was readjusted to 8.2 with NaOH (1 mM). The reaction was stirred at 600 rpm and 20 °C, and monitored through HPLC. Each experiment was repeated at 3 times.

Seeded replication experiments. Monomer solutions, either with a single component or with an equimolar mixture of **A** and **T**, were prepared as described in the previous paragraph. Once prepared and while being stirred, the corresponding percentage of seed (20/30% of cysteine) from a finished reaction was added, and the reaction was kept stirring at 600 rpm and 20 °C, followed by HPLC monitoring. Each experiment was repeated twice.

Disulfide exchange reactions. A solution of **A** (4 mM) in borate buffer (50 mM) was prepared as described above, and mixed with a finalized oxidation reaction of **T** (containing 100 mol% of **TT**) in a 2:1 molar ratio of **A/TT**. The protocol for the opposite reaction (**T** + **AA**) was identical except for the switch of the nucleobases in the monomeric thiol and the disulfide derivative. The reaction was stirred at 600 rpm and 20 °C, and monitored through HPLC, repeating twice each of the experiments.

HPLC-MS. 50 µL aliquots from every experiment were collected at the indicated times and deposited into 1% aqueous TFA to quench the reaction. The samples were then frozen until analyzed in a Waters Symmetry® C18 5µm 250×4.6 mm column, eluting them with a linear

gradient of water to acetonitrile for 15 min. The different species were identified with a single quadrupole mass detector and quantified with a UV-Vis detector (λ = 260 nm).

Acknowledgements

This research was supported by the H2020 FET-Open (A.d.l.E and G.A.; CLASSY project, Grant Agreement Nº 862081) and the Spanish Ministry of Economy and Competitivity (A.d.l.E; MINECO: CTQ-2014-53673-P, CTQ-2017-89539-P, and EUIN2017-87022). We would like to acknowledge Dr. Ana Pizarro for her advice in the NMR determination of the nucleobase pK_a values in compounds **AA** and **TT**, and Dr. Francisco Tato for initial tests on HPLC separations. The professional editing service NB Revisions was used for technical preparation of the text prior to submission.

Author contributions

S.V.G performed the experiments, Z.P.B. contributed to the DOSY analyses, C.M.A. contributed to the computational kinetic analyses, A.d.l.E. conceived the project, S.V.G. and A.d.l.E. contributed to designing and analyzing the experiments, and writing and editing the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Cronin, L. & Walker, S. I. Beyond prebiotic chemistry. *Science* **352**, 1174–1175 (2016).
2. 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).
3. Szostak, J. W. The narrow road to the deep past: in search of the chemistry of the origin of life. *Angew. Chem. Int. Ed.* **56**, 11037–11043 (2017).

4. Sutherland, J. D. Studies on the origin of life—the end of the beginning. *Nat. Rev. Chem.* **1**, 1–7 (2017).
5. Krishnamurthy, R. & Hud, N. V. Introduction: chemical evolution and the origins of life. *Chem. Rev.* **120**, 4613–4615 (2020).
6. Ashkenasy, G., Hermans, T. M., Otto, S. & Taylor, A. F. Systems chemistry. *Chem. Soc. Rev.* **46**, 2543–2554 (2017).
7. Ura, Y., Beierle, J. M., Leman, L. J., L. E. Orgel, M. R. Ghadiri. Self-assembling sequence-adaptive peptide nucleic acids. *Science* **325**, 73–77 (2009).
8. Carbajo, D., Pérez, Y., Bujons, J. & Alfonso, I. Live-cell-templated dynamic combinatorial chemistry. *Angew. Chem.* **132**, 17355–17359 (2020).
9. Mann, S. Systems of creation: the emergence of life from nonliving matter. *Acc. Chem. Res.* **45**, 2131–2141 (2012).
10. Gibard, C., Bhowmik, S., Karki, M., Kim, E. K., & Krishnamurthy, R. Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nat. Chem.***10**, 212 (2018).
11. Bonfio, C., Godino, E., Corsini, M., de Biani, F. F., Guella, G. & Mansy, S. S. Prebiotic iron–sulfur peptide catalysts generate a pH gradient across model membranes of late protocells. *Nat. Catal.* **1**, 616–623 (2018).
12. Morales-Reina, S., Giri, C., Leclercq, M., Vela-Gallego, S., de la Torre, I., Castón, J. R., Surin, M. & de la Escosura, A. Programmed recognition between complementary dinucleolipids to control the self-assembly of lipidic amphiphiles. *Chem. Eur. J.* **26**, 1082–1090 (2020).
13. Frenkel-Pinter, M., Haynes, J. W., Mohyeldin, A. M., Martin, C., Sargon, A. B., Petrov, A. S., Krishnamurthy, R., Hud, N. V., Williams, L. D. & Leman, L. J. Mutually stabilizing interactions between proto-peptides and RNA. *Nat. Commun.* **11**, 1–14 (2020).
14. Paul, N. & Joyce, G. F. Minimal self-replicating systems. *Curr. Opin. Chem. Biol.* **8**, 634–639 (2004).
15. Patzke, V. & von Kiedrowski, G. Self-replicating systems. *Arkivoc* **5**, 293–310 (2007).

16. Bissette, A. J. & Fletcher, S. P. Mechanisms of autocatalysis. *Angew. Chem. Int. Ed.* **52**, 12800–12826 (2013).

17. Kosikova, T. & Philp, D. Exploring the emergence of complexity using synthetic replicators. *Chem. Soc. Rev.* **46**, 7274-7305 (2017).

18. Adamski, P., Eleveld, M., Sood, A., Kun, Á., Szilágyi, A., Czárán, T., Szathmáry, E. & Otto, S. From self-replication to replicator systems en route to de novo life. *Nat. Rev. Chem.* **4**, 386–403 (2020).

19. Soai, K., Shibata, T., Morioka, H. & Choji, K. Asymmetric autocatalysis and amplification of enantiomeric excess of a chiral molecule. *Nature* **378**, 767–768 (1995).

20. Breslow, R. On the mechanism of the formose reaction. *Tetrahedron Lett.* **1**, 22–26 (1959).

21. Hein, J. E. & Blackmond, D. G. On the origin of single chirality of amino acids and sugars in biogenesis. *Acc. Chem. Res.* **45**, 2045–2054 (2012).

22. Epstein, I. R. & Showalter, K. Nonlinear chemical dynamics: oscillations, patterns, and chaos. *J. Phys. Chem.* **100**, 13132–13147 (1996).

23. von Kiedrowski, G. A self-replicating hexadeoxynucleotide. *Angew. Chem. Int. Ed.* **25**, 932–935 (1986).

24. Kreysing, M., Keil, L., Lanzmich, S. & Braun, D. Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. *Nat. Chem.* **7**, 203–208 (2015).

25. Hayden, E. J. & Lehman, N. Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chem. Biol.* **13**, 909–918 (2006).

26. Lincoln, T. A. & Joyce, G. F. Self-sustained replication of an RNA enzyme. *Science* **323**, 1229–1232 (2009).

27. Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K. & Ghadiri, M. R. A self-replicating peptide. *Nature* **382**, 525–528 (1996).

28. Mukherjee, R., Cohen-Luria, R., Wagner, N. & Ashkenasy, G. A bistable switch in dynamic thiodepsipeptide folding and template-directed ligation. *Angew. Chem.* **127**, 12629–12633 (2015).

29. Tjivikua, T., Ballester, P. & Rebek Jr, J. A self-replicating system. *J. Am. Chem. Soc.* **112**, 1249–1250 (1990).

30. Kauffman, S. A. Autocatalytic sets of proteins. *J. Theor. Biol.* **119**, 1–24 (1986).

31. Vasas, V., Fernando, C., Santos, M., Kauffman, S., & Szathmáry, E. Evolution before genes. *Biol. Direct*, **7**, 1–14 (2012).

32. Segré, D., Ben-Eli, D., Deamer, D. W. & Lancet, D. The lipid world. *Orig. Life Evol. Biosph.* **31**, 119–145 (2001).

33. Colomer, I., Borissov, A. & Fletcher, S. P. Selection from a pool of self-assembling lipid replicators. *Nat. Commun.* **11**, 1–9 (2020).

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

35. Maity, I., Wagner, N., Mukherjee, R., Dev, D., Peacock-Lopez, E., Cohen-Luria, R. & Ashkenasy, G. A chemically fueled non-enzymatic bistable network. *Nat. Commun.* **10**, 1–9 (2019).

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

37. Semenov, S. N., Kraft, L. J., Ainla, A., Zhao, M., Baghbanzadeh, M., Campbell, V. E., Kang, K., Fox, J. M. & Whitesides, G. M. Autocatalytic, bistable, oscillatory networks of biologically relevant organic reactions. *Nature* **537**, 656–660 (2016).

38. Walde, P., Wick, R., Fresta, M., Mangone, A. & Luisi, P. L. Autopoietic self-reproduction of fatty acid vesicles. *J. Am. Chem. Soc.* **116**, 11649–11654 (1994).

39. Ruiz-Mirazo, K., Briones, C. & de la Escosura, A. Chemical roots of biological evolution: the origins of life as a process of development of autonomous functional systems. *Open Biol.* **7** 170050 (2017).

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

41. Yang, S., Schaeffer, G., Mattia, E., Markovitch, O., Liu, K., Hussain, A. S., Ottelé, J., Sood, A. & Otto, S. Chemical fueling enables molecular complexification of self-replicators. *Angew. Chem. Int. Ed.* **60**, 11344–11349 (2021).

42. Miao, X., Paikar, A., Lerner, B., Diskin-Posner, Y., Shmul, G., Semenov, S. N. Kinetic selection in the out-of-equilibrium autocatalytic reaction networks that produce macrocyclic peptides. *Angew. Chem. Int. Ed.* (2021). doi.org/10.1002/anie.202105790

43. Matsumura, S., Kun, Á., Ryckelynck, M., Coldren, F., Szilágyi, A., Jossinet, F., Rick, C., Nghe, P., Szathmáry, E. & Griffiths, A. D. Transient compartmentalization of RNA replicators prevents extinction due to parasites. *Science* **354**, 1293–1296 (2016).

44. Bandela, A. K., Wagner, N., Sadihov, H., Morales-Reina, S., Chotera-Ouda, A., Basu, K., Cohen-Luria, R., de la Escosura, A. & Ashkenasy, G. Primitive selection of the fittest emerging through functional synergy in nucleopeptide networks. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2015285118 (2021).

45 Rubinov, B., Wagner, N., Matmor, M., Regev, O., Ashkenasy, N. & Ashkenasy, G. Transient fibril structures facilitating non-enzymatic self-replication. *ACS Nano* **6**, 7893–7901 (2012).

46. Chen, Z., Lohr, A., Saha-Möller, C. R. & Würthner, F. Self-assembled π-stacks of functional dyes in solution: structural and thermodynamic features. *Chem. Soc. Rev.* **38**, 564–584 (2009).

47. von Kiedrowski, G. Minimal replicator theory I: parabolic versus exponential growth. *Bioorg. Chem. Front.* **3**, 113-146 (1993).

48. Wagner, N. & Ashkenasy, G. Rhythm before life. *Nat. Chem.* **11**, 681–683 (2019).

49. Parker, E. T., Cleaves, H. J., Dworkin, J. P., Glavin, D. P., Callahan, M., Aubrey, A., Lazcano, A. & Bada, J. L. Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark discharge experiment. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5526–5531 (2011).

50. Foden, C. S., Islam, S., Fernandez-Garcia, C., Maugeri, L., Sheppard, T. D. & Powner, M. W. Prebiotic synthesis of cysteine peptides that catalyze peptide ligation in neutral water. *Science* **370**, 865–869 (2020).

51. Islam, S. & Powner, M. W. Prebiotic systems chemistry: complexity overcoming clutter. *Chem* **2**, 470–501 (2017).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- VelaetalSICommunChem281021.pdf