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  
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

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
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). 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. 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.

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. The literature is rich in processes that display
autocatalysis, either through the product’s catalysis of its own formation, cyclic
autocatalysis, or in oscillatory reactions. Most of these autocatalytic transformations
cannot be considered self-replication, since they lack the specificity required for information
transfer at the molecular level. In search for such specificity, template replication has been
proven with different kinds of biopolymers and oligomers, including DNA, RNA and
oligopeptides, as well as with synthetic molecules not present in extant biology. 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, with both theoretical and experimental models based on lipids, peptides,
nucleic acids, and synthetic molecules. 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
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: (i) catalysis, to establish a supportive metabolism; (ii) performance out of equilibrium; (iii) compartmentalization, to avoid parasitic reactions and dilution effects; or (iv) variability control with nucleobase sequences. 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, 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,\textsuperscript{34,45} we assume these to be the catalytically active ones also in the present case.

\textbf{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).
Figure 3. Kinetic profiles corresponding to: (first row) the autocatalytic formation of AA (A), TT (C), and the mixture of possible replicators AA/TT/AT (E); (second row) seeded experiments for the autocatalytic formation of AA (B), TT (D), and AA/TT/AT (F); (third row) disulfide exchange reactions (up to 7 h) between AA and T (G) or TT and A (H). Each panel shows the evolution of all involved species over time through experimental data (square data points) and fit curves, while the set of kinetic equations used for fitting are depicted in Table 1. In the templated reactions (A, C, E), the total concentration of starting materials was maintained in the same range as in the non-templated ones (B, D, F), with minor experimental deviations that are difficult to avoid but are considered in the mathematical analysis. The non-seeded experiments were repeated three times, while seeded experiments and disulfide exchange reactions were repeated twice (Figures S16-S19). The treatment of fitting errors with such 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 \( \text{R-A1, R-T1} \) and the autocatalytic oxidation \( \text{R-A2, R-T2} \) of the corresponding monomeric thiol (Table 1, boxes 1 and 2). In the reaction schemes, \( \text{AA}_{ag} \) and \( \text{TT}_{ag} \) refer to the self-assembled products, and emerge from the aggregation processes \( \text{R-A3} \) and \( \text{R-T3} \). For calculation of kinetic constants, the concentration of aggregated replicator is needed at any given time. According to Chen et al., the concentrations of \( \text{AA}_{ag} \) and \( \text{TT}_{ag} \) are given by equation \( \text{Eq-S1} \) (Section 5 in the SI), in which the total concentration of aggregating compound \( (C_T) \) can be expressed as a function of \( C_I \) (the concentration of molecularly dissolved compound), \( \rho \) (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 \( \text{Eq-S1} \) for different values of \( \rho \) are compared to the proposed simplification, the outcome differences are negligible.

The rate equations for non-catalyzed (\( \text{Eq-A1 and Eq-T1} \)) and autocatalytic regimes (\( \text{Eq-A2 and Eq-T2} \)) 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 \( \text{AA}_{ag} \) (or \( \text{TT}_{ag} \)) were evaluated through fitting of the experimental data. The equilibrium constants of aggregate formation were considered through equations \( \text{Eq-A3} \) and \( \text{Eq-T3} \). 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 (\( \text{Eq-A2 and Eq-T2} \)) 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.\textsuperscript{34,47} The orders obtained in the rate equations
\textbf{Eq-A2} and \textbf{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 ($K_{AAag} = 1.95 \text{ mM}^{-1}$ and $K_{TTag} = 1.6 \text{ mM}^{-1}$) 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 (\textbf{R-E1} to \textbf{R-E4} / \textbf{Eq-E1} to \textbf{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 \textit{AA} and \textit{TT}, two other aggregate types
must be considered, resulting from either complementary interaction between \textit{AA} and \textit{TT}
(\textit{AATT}\textsubscript{ag}) or from self-pairing of \textit{AT} (i.e., \textit{AT}\textsubscript{ag}). These aggregates enable auto- and cross-
catalytic reactions, as \textit{AATT}\textsubscript{ag} can aid in reactions producing \textit{AA} and \textit{TT} (\textbf{R-A4} and \textbf{R-T4},
respectively -- Table 1, box 4), and \textit{AT}\textsubscript{ag} can assist in its own formation (\textbf{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 \textit{TT} and \textit{AA}, the disulfide exchange reactions, the
aggregation constants of \textit{AA} and \textit{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.

<table>
<thead>
<tr>
<th>Box 1</th>
<th>Reaction</th>
<th>Equation</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-A1</td>
<td>A + A $\rightarrow$ AA</td>
<td>Eq-A1</td>
<td>$k_d[A]^2$</td>
</tr>
<tr>
<td>R-A2</td>
<td>A + A $\rightarrow$ AA $\rightarrow$ AA$_{ag}$</td>
<td>Eq-A2</td>
<td>$k_{Ac}[A][AA_{ag}]^2$</td>
</tr>
<tr>
<td>R-A3</td>
<td>AA + AA $\rightarrow$ AA$_{ag}$</td>
<td>Eq-A3</td>
<td>$K_{eqAA_{ag}} = 1/[AA]_{cac}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Box 2</th>
<th>Reaction</th>
<th>Equation</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-T1</td>
<td>T + T $\rightarrow$ TT</td>
<td>Eq-T1</td>
<td>$k_T[T]^2$</td>
</tr>
<tr>
<td>R-T2</td>
<td>T + T $\rightarrow$ TT</td>
<td>Eq-T2</td>
<td>$k_{TC}[T][TT_{ag}]^2$</td>
</tr>
<tr>
<td>R-T3</td>
<td>TT + TT $\rightarrow$ TT$_{ag}$</td>
<td>Eq-T3</td>
<td>$K_{eqTT_{ag}} = 1/[TT]_{cac}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Box 3</th>
<th>Reaction</th>
<th>Equation</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-E1</td>
<td>A + TT$<em>d$ $k</em>{c_e}[A][TT]_{MD}$</td>
<td>Eq-E1</td>
<td>$k_{c_e}[A][TT]_{MD}$</td>
</tr>
<tr>
<td>R-E2</td>
<td>T + AA$<em>d$ $k</em>{c_e}[T][AA]_{MD}$</td>
<td>Eq-E2</td>
<td>$k_{c_e}[T][AA]_{MD}$</td>
</tr>
<tr>
<td>R-E3</td>
<td>A + AT$<em>d$ $k</em>{c_e}[A][AT]_{MD}$</td>
<td>Eq-E3</td>
<td>$k_{c_e}[A][AT]_{MD}$</td>
</tr>
<tr>
<td>R-E4</td>
<td>T + AT$<em>d$ $k</em>{c_e}[T][AT]_{MD}$</td>
<td>Eq-E4</td>
<td>$k_{c_e}[T][AT]_{MD}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Box 4</th>
<th>Reaction</th>
<th>Equation</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-A4</td>
<td>A + A $\rightarrow$ AA</td>
<td>Eq-A4</td>
<td>$k_{AA_{TT}1}[A][AA_{TT}]_{ag}$</td>
</tr>
<tr>
<td>R-T4</td>
<td>T + T $\rightarrow$ TT</td>
<td>Eq-T4</td>
<td>$k_{AA_{TT}2}[T][AA_{TT}]_{ag}$</td>
</tr>
<tr>
<td>R-AATT</td>
<td>AA + TT $\rightarrow$ AA$_{ag}$</td>
<td>Eq-AATT</td>
<td>$K_{eqAA_{TT}<em>{ag}} = 1/[AA</em>{TT}]_{cac}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Box 5</th>
<th>Reaction</th>
<th>Equation</th>
<th>Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-AT1</td>
<td>A + T $\rightarrow$ AT</td>
<td>Eq-AT1</td>
<td>$k_{AT}[A][T]$</td>
</tr>
<tr>
<td>R-AT2</td>
<td>A + AT $\rightarrow$ AT</td>
<td>Eq-AT2</td>
<td>$k_{AT}[A][AT]_{ag}$</td>
</tr>
<tr>
<td>R-AT3</td>
<td>AT + AT $\rightarrow$ AT$_{ag}$</td>
<td>Eq-AT3</td>
<td>$K_{eqAT_{ag}} = 1/[AT]_{cac}$</td>
</tr>
</tbody>
</table>
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 $AATT_{ag}$ and $AT_{ag}$ (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 ([replicators]$_0$ = 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
increasing initial AT concentrations (Figure 4B). SS\textsubscript{AT} was also obtained when the reactor was filled with AA/TT below a threshold concentration (see below). When [AA/TT]\textsubscript{o} was above that threshold, however, a new steady state appeared (SS\textsubscript{AA/TT}) where AA/TT were the dominant species ([AA/TT] = 1.77 mM) and AT decreased in concentration ([AT] = 0.19 mM) (Figure 4C and S23). A range of initial proportions of AA, TT and AT, and of total initial disulfide concentration (from 0 to 4 mM), was then tested to determine which steady state would be reached in each case. The result was a 3D surface (Figure 4D) that marks the boundary between initial conditions that favor SS\textsubscript{AT} (below the surface) or SS\textsubscript{AA/TT} (above the surface). Interestingly, the threshold below which AT always gets amplified was ~1 mM; above that threshold, SS\textsubscript{AA/TT} was favored except when the initial molar fraction of AT was higher than 0.8.
Figure 4. (A) Replication network topology. (B, C) Simulations of the network evolution when fed with two input streams ($Q_{in} = 1 \mu L/min$) of $A$ and $T$, in the presence of replicator $AT$ ($0 – 2 mM$) (B) or $AA/TT$ ($0 – 2 mM$) (C). (D) 3D surface representing the reached steady state (SS$_{AT}$ below the surface; SS$_{AA/TT}$ above) for a range of initial replicator proportions (horizontal axes on the graph) and total replicator concentrations (vertical axis). The surface marks the boundary between initial conditions that favor SS$_{AT}$ (below the surface) or SS$_{AA/TT}$ (above). (E, F) Simulations of the network evolution when decreasing the exchange constants by six orders of magnitude, with the same input streams and conditions as in (B) and (C).

The capacity of a dynamic system to reach two different steady states depending on the initial conditions is called bistability, and in the present case it seems to be related to their 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,\textsuperscript{49} Strecker-derived chemistry\textsuperscript{50} and HCN/cyanoacetylene oligomerization reactions.\textsuperscript{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.\textsuperscript{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.\textsuperscript{15} As a merge of both approaches, the self-assembly of AA, TT and AT must be mostly promoted by H-bond
interactions between nucleobases, together with $\pi$-$\pi$ stacking and hydrophobic effects. Despite this makes their self-assembly weaker and the replication rates slower compared to previous peptide- or lipid-based replicators, they gain a rudimentary sequence-based control of the replication process. Building on such a capacity, the network described herein presents a collective behaviour that can provide significant adaptability between the individual synergistic and 'selfish' replication pathways, aided by exchange reactions that allow interconversion between the different replicator species.

**Methods**

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

**Self-assembly and replication experiments.** A borate-buffered solution (200 mM, pH 8.2) was employed for all the self-assembly and replication experiments. Boric acid (20 mmol, 1.24 g) was dissolved in H$_2$O (100 ml), or D$_2$O for the DOSY experiments, and the pH was adjusted to 8.2 with 1 M aq. NaOH (or NaOD). This buffer was used as a stock to prepare all samples, adjusting the borate concentration to 50 mM in all cases.

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

**DOSY NMR.** Different solutions of AA, TT or equimolar mixtures of AA/TT (0.1, 0.25, 0.5, 0.75, 1, 2, 3 and 4 mM total concentration of disulfide) were prepared in D$_2$O-based borate buffer, and the pH was readjusted to 8.2. The DOSY measurements were performed using the longitudinal eddy current (LED) delay pulse sequence. The duration of the magnetic field pulse gradient (small delta, $\delta$) was 2.8 ms and the diffusion delay (big delta, $\Delta$) was 100 ms in order to 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 ($\lambda = 260$ 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 $\text{pK}_{a}$ 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


Supplementary Files

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

- VelaetalSICommunChem281021.pdf