Electronic Supplementary Material (ESI) for Chemical Science.

Collective Adaptability in a Replication Network of Minimal Nucleobase Sequences

Sonia Vela-Gallego, Zulay Pardo-Botero, Cristian Moya, Andrés de la Escosura

Table of contents:

1. General methods

Characterization of compounds. NMR Spectra were recorded on a Bruker Advance III HD 400 MHz spectrometer equipped with 5 mm BBFO probe at room temperature. D_2O or DMSO-d_6 were used as deuterated solvents, and chemical shifts are given in ppm with the residual solvent peak as reference. IR spectra were recorded on a FT-IR Cary 630 (Agilent Technologies) using ATR as technique, correcting the intensity by ATR algorithm. ESI-MS was performed using an ultra-high-resolution QTOF instrument (MAXIS II, Bruker).

Conditions for 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_2O (100 ml), or D_2O 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_2O-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, δ) was 2.8 ms and the diffusion delay (big delta, Δ) 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 3 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, with a magnetic bar of 5 mm, 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, with a magnetic bar of 5 mm, followed by HPLC monitoring. Each experiment was repeated twice.

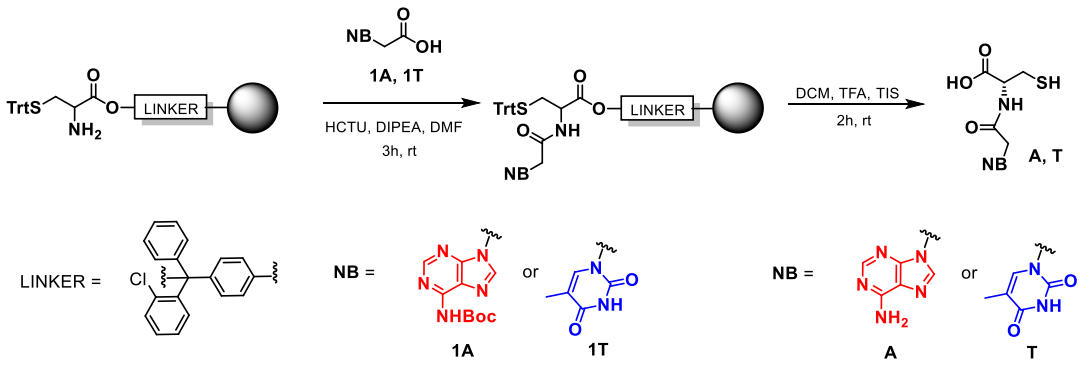
Disulfide exchange reactions. A solution of **A** (4 mM) in borate buffer (50 mM) was prepared as described above, and mixed with a finalized oxidation reaction of **T** (containing 100 mol% of **TT**) in a 2:1 molar ratio of **A**/**TT**. The protocol for the opposite reaction (**T** + **AA**) was identical 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, with a magnetic bar of 5 mm, 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).

2. Synthesis and characterization of the network components

Synthesis of building blocks A and T. The network building blocks **A** and **T** were prepared making use of a solid-phase synthesis (SPS) strategy (Scheme S1).[1,2] In particular, an amide coupling reaction was performed between 9-carboxymethyl-adenine (**1A**)[3] or the commercially available 1-carboxymethyl-thymine (**1T**) with L-cysteine, which was protected with S-trityl and N-Boc groups, and anchored in the commercial resin H-Cys(Trt)-2-chlorotrityl resin (100-200 mesh, 0.50-0.90 mmol/g). The coupling reaction was performed at room temperature, with 1,1,3,3-tetramethyluronium hexafluorophosphate (HCTU) as coupling reagent, N,N-diisopropylethylamine (DIPEA) as base and dimethylformamide (DMF) as solvent, followed by a cleavage/deprotection step with trifluoroacetic acid (TFA) and triisopropylsilane (TIS) in water for **T**, and with TFA, TIS and dithiothreitol (DTT) in dichloromethane (DCM) for **A**. The resulting compounds were obtained in 95 and 89% yield, respectively, and characterised by nuclear magnetic resonance (NMR), infrared (IR) spectroscopy, high-resolution mass spectrometry (HR-MS) and HPLC.



Scheme S1. Synthetic route for the preparation of building blocks **A** and **T**.

Experimental procedure for solid phase synthesis (SPS) of building blocks A and T. First, 150 mg of the resin H-Cys(Trt)-2-chlorotrityl resin (100-200 mesh, 0.50-0.90 mmol/g), purchased

from Bachem, was washed with DCM (2 mL x 0.5 min) and DMF (2 mL x 0.5 min). A solution of 9-carboxymethyl-adenine (**1A**)[3] or the commercially available 1-carboxymethyl-thymine (**1T**) (1.3 eq), HCTU (1.3 eq) and DIPEA (3 eq) in DMF (1.5 mL) was added. The mixture was left under orbital stirring for 3h. The resin was filtered under vacuum, and washed again with DCM and DMF (2 mL x 0.5 min each). This process was performed twice.

Cleavage and deprotection. A mixture of TFA/DCM/TIS (0.65:0.65:0.2) for **T**, or TFA/DCM/TIS/DTT (ratios) for **A**, was added to the resin, and the mixture was left under orbital stirring for 2h. The supernatant was collected and deposited into cold ether. A white precipitate appeared, which was filtered and washed with DCM, diethyl ether and acetonitrile. Yields were 95% for **T** and 89% for **A**.

*Characterization of compounds **A** and **T**.*

A; ^{1}H-NMR (300 MHz, DMSO-d_6): δ (ppm) = 8.84 (d, *J = 7.8 Hz*, CON**H**), 8.39 (s, 1H, N=C**H**-N), 8.32 (s, 1H, N=C**H**-N), 5.05 (s, 2H, CO-C**H$_2$**-N), 4.50-4.43 (m, 1H, C$_\alpha$**H**), 2.95-2.76 (m, 2H, C**H$_2$**-S). **^{13}C-NMR** (75 MHz, DMSO-d_6): δ (ppm) = 171.1, 166.1, 152.4, 149.1, 148, 143.6, 117.9, 54.5, 45.1, 25.6. **IR:** ν (cm^{-1}) = 3291, 3101, 1703, 1669, 1569, 1524. **HRMS** (ESI+, m/z): $[M]^+$ cald. for $C_{10}H_{13}N_6O_3S$, 297.07644; found 297.0759.

T; ^{1}H-NMR (300 MHz, DMSO-d_6): δ (ppm) = 11.24 (s, 1H, CON**H**CO), 8.53 (d, *J = 7.9 Hz*, 1H, CON**H**), 7.41 (d, *J = 1.1 Hz*, 1H, C**H**=C), 4.46-4.35 (m, 3H, C**H$_2$**N, C$_\alpha$**H**), 2.90-2.71 (m, 2H, C**H$_2$**-S), 2.42 (t, *J = 8.5 Hz*, 1H, S**H**), 1.74 (d, *J = 1.1 Hz*, 3H, C**H$_3$**). **^{13}C-NMR** (75 MHz, DMSO-d_6): δ (ppm) = 174, 169.2, 167.1, 152.3, 143.2, 111, 55.2, 54.4, 29.6, 11.3. **IR:** ν (cm^{-1}) = 3537, 3317, 3257, 1699, 1654, 1636, 1543. **HRMS** (ESI+, m/z): $[M+Na]^+$ cald. for $C_{10}H_{13}N_3NaO_5S$, 310.0474; found 310.0461.

Oxidation of thiols A and T into disulfide dimers. With the building blocks **A** and **T** in hand, an initial assessment of the oxidation of their thiol group into the disulfides **AA** and **TT**, respectively, was carried out by ^{1}H-NMR. The reactions were initially carried in DMSO-d_6, within the NMR tube, and resulted to be extremely slow, with incomplete conversion of the starting thiol after

30 days of reaction (Figure S1). In turn, the reaction progress in water was very much dependent on pH. Dissolving **A** or **T** in D_2O (2 mM), for example, yielded acidic solutions of pH 3.2 and 4, respectively (calculated from the measured pD values with the equation pH = 0.989 * pD + 0.472), which reduces the thiol nucleophilicity of both monomers and completely prevents their oxidation (Figure S2B, bottom spectra). At the other extreme of the tested pH range, formation of the disulfide compound **AA** or **TT** was immediate when the reaction pH was adjusted to 11 through addition of NaOD (Figure S2B, top spectra). Given this result, both oxidation reactions were performed in a larger scale in basic conditions (see experimental procedure in the next paragraph). The final cystine derivatives were isolated after 72 h of reaction, by acidification with hydrochloric acid down to pH 4 and filtration of the resulting white solid, obtaining **AA** and **TT** in 94 and 90% yield, respectively.

Experimental procedure for the synthesis of dimers AA and TT. A solution of **A** or **T** was brought to pH 11 (NaOH, 1M) and stirred for 16h at rt. The solution was acidified to pH 4 with HCl 1 M. A white precipitate was observed. The precipitate was filtered and washed with water, acetonitrile, and diethyl ether. The yields for both compounds are quantitative***.***

*Characterization of compounds **AA** and **TT**.*

AA; ^{1}H-NMR (300 MHz, DMSO-d_6): δ (ppm) = 8.95 (d, *J = 7.8 Hz*, 1H, CON**H**), 8.36 (s, 1H, N=C**H**-N), 8.33 (s, 1H, N=C**H**-N), 5.02 (s, 2H, CO-C**H$_2$**-N), 4.61-4.52 (m, 1H, C$_\alpha$**H**), 3.20 (dd, *J = 13.8*, *4.7 Hz*, 1H, C**H$_2$**-S), 3.00 (dd, *J = 13.8, 8.7 Hz*, 1H, C**H$_2$**-S). **^{13}C-NMR** (75 MHz, DMSO-d_6): δ (ppm) = 171.0, 166.5, 155.6, 152.4, 150.7, 149.6, 141.9, 54.4, 44.8, 24.4. **IR:** ν (cm^{-1}) = 3302, 1699, 1617, 1558, 1535. **HRMS** (ESI+, m/z) = $[M]^+$ cald. for $C_{20}H_{22}N_{12}NaO_6S$, 613.1124; found 613.1119.

TT; ^{1}H-NMR (300 MHz, DMSO-d_6): δ (ppm) = 11.24 (s, 1H, CON**H**CO), 8.53 (d, *J = 7.8 Hz*, 1H, CON**H**), 7.41 (d, *J = 1.1 Hz*, 1H, C**H**=C), 4.55-4.48 (m, 1H, C$_\alpha$**H**), 4.35 (s, 2H, C**H$_2$**N), 3.15 (dd, *J = 13.8, 4.0 Hz*, 1H, C**H$_2$**-S), 2.94 (dd, *J = 13.8, 8.6 Hz*, 1H, C**H$_2$**-S), 1.74 (d, *J = 1.1 Hz*, 3H, C**H$_3$**). **^{13}C-NMR** (75 MHz, DMSO-d_6): δ (ppm) = 174, 169.2, 167.1, 152.3, 143.2, 111, 55.2, 54.4, 29.6, 11.3.

IR: ν (cm^{-1}): 3291, 1707, 1703, 1654, 1643, 1550. **HRMS** (ESI+, m/z): $[M+Na]^+$ cald. for $C_{20}H_{24}N_6NaO_{10}S_2$, 596.0893; found 595.0888.

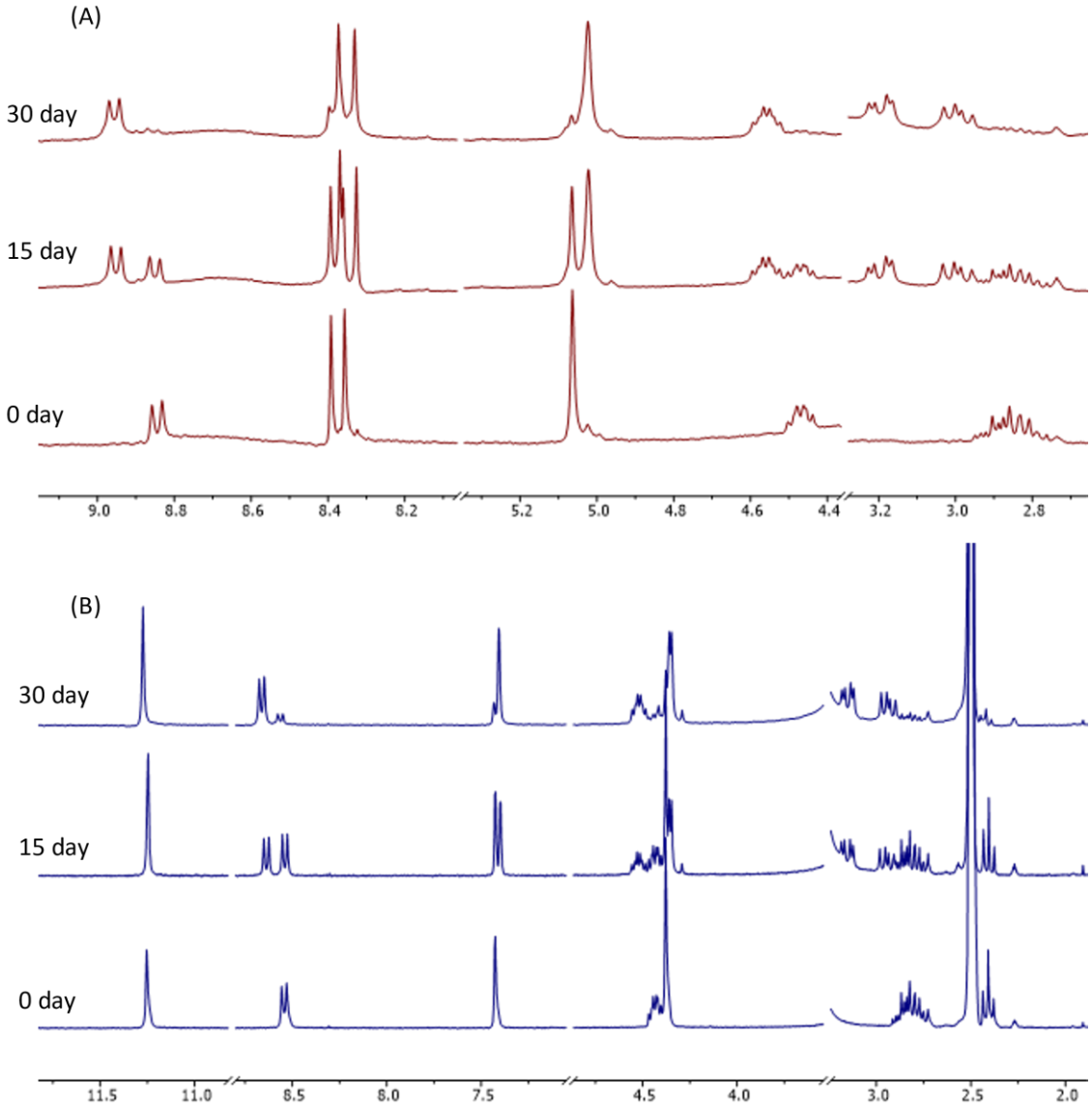


Figure S1. ^{1}H-NMR monitoring of the oxidation of **A** and **T** into the dimeric disulfides **AA** (A) and **TT** (B) in DMSO-d_6.

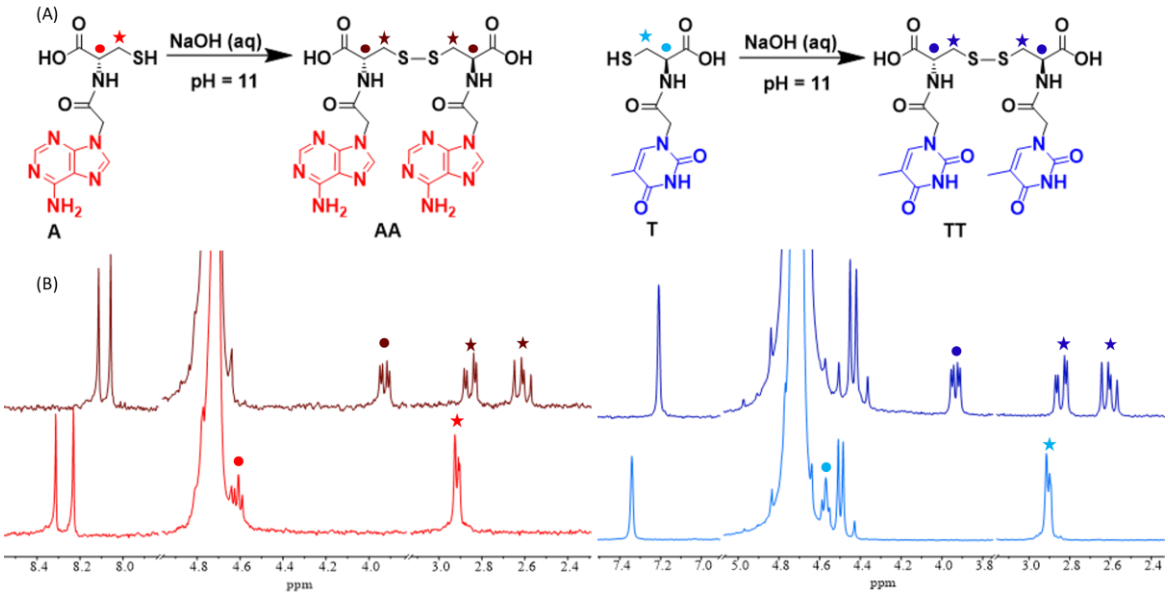


Figure S2. (A) Oxidation reactions of **A** and **T** into **AA** and **TT**, respectively. (B) ^{1}H-NMR monitoring of the oxidation of **A** and **T** into the dimeric disulfides **AA** (left) and **TT** (right) at acidic (bottom) and basic (top) conditions.

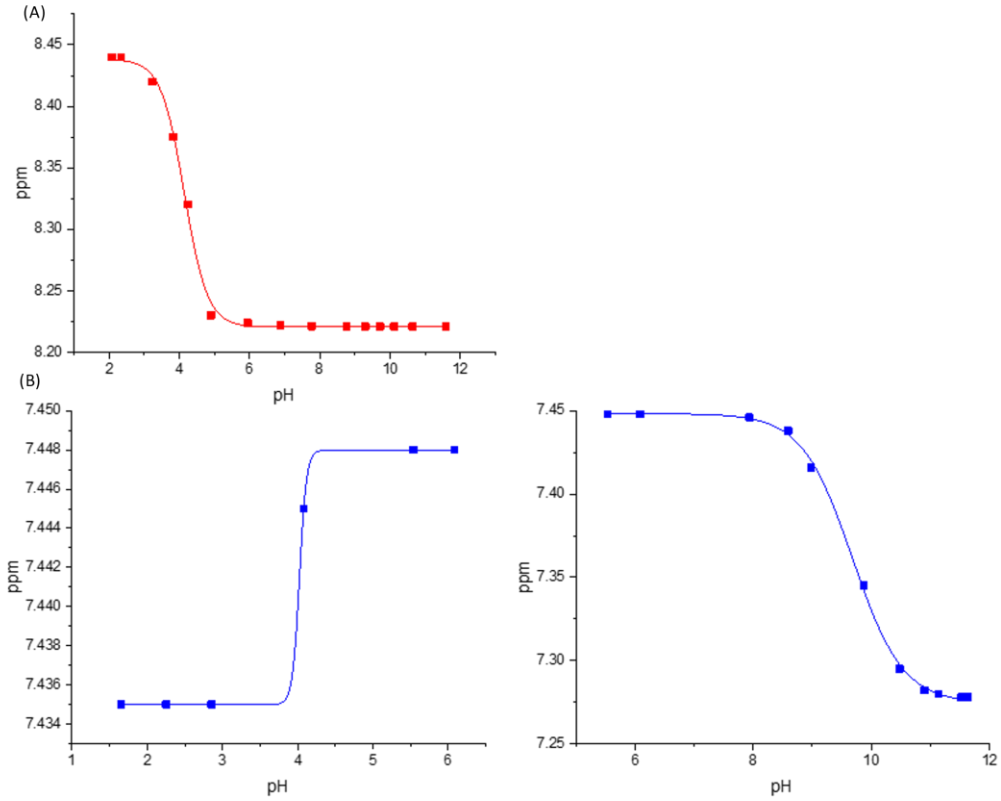


Figure S3. pH titrations monitored by ^{1}H-NMR to calculate the pK_a of **AA** (A), following the chemical shift of the most downfield aromatic proton of the adenine moiety, and **TT** (B), following the aromatic thymine proton. In the latter case, the first process (left) corresponds to the carboxylic acid-based equilibrium, and the second one (right) to the dissociation of the thymine imide NH. The obtained data were fitted with a Boltzmann equation for an accurate determination of the pK_a values.

3. Supramolecular studies

Calculation of the diffusion coefficient (D) by DOSY NMR. The DOSY measurements were carried out by monitoring the attenuation of NMR signals during a pulsed field gradient experiment. The degree of attenuation is a function of the magnetic gradient pulse amplitude and occurs at a rate proportional to the diffusion coefficient (*D*). Assuming that a line at a given (fixed) chemical shift *f* belongs to a single sample *X* component with a diffusion constant *D*, it can be calculated according to the following equation:

$$\boldsymbol{S(f,z) = S_x(f)^{-DZ}}$$

Where ***S(f,z)*** is integral, $\boldsymbol{S_X(f)}$ is the spectral intensity of component *X* in zero gradient ('normal' spectrum of *X*), $\boldsymbol{D_x}$ is its diffusion coefficient, and ***Z*** encodes the different gradient amplitudes used in the experiment.

Calculation of the critical aggregation concentration (cac) for AA/TT. Once we had determined the diffusion coefficient for each system at different concentrations, it was necessary to determine the cac values. The calculation of cac for **AA**/**TT** could not be performed as for the pure compounds, **AA** and **TT**, since the data did not show two straight lines crossing at a given point. Therefore, a mathematical fitting of the experimental data was made into a Boltzmann equation provided by Origin 2019 (see below), and the result is the curve shown with a dotted line in Figure 2A (right image). The inflection point of this curve actually corresponds to cac value given for **AA**/**TT**.

$$Y = \frac{A_1 - A_2}{\left(1 + e^{\frac{x - x_0}{dx}}\right)} + A_2$$

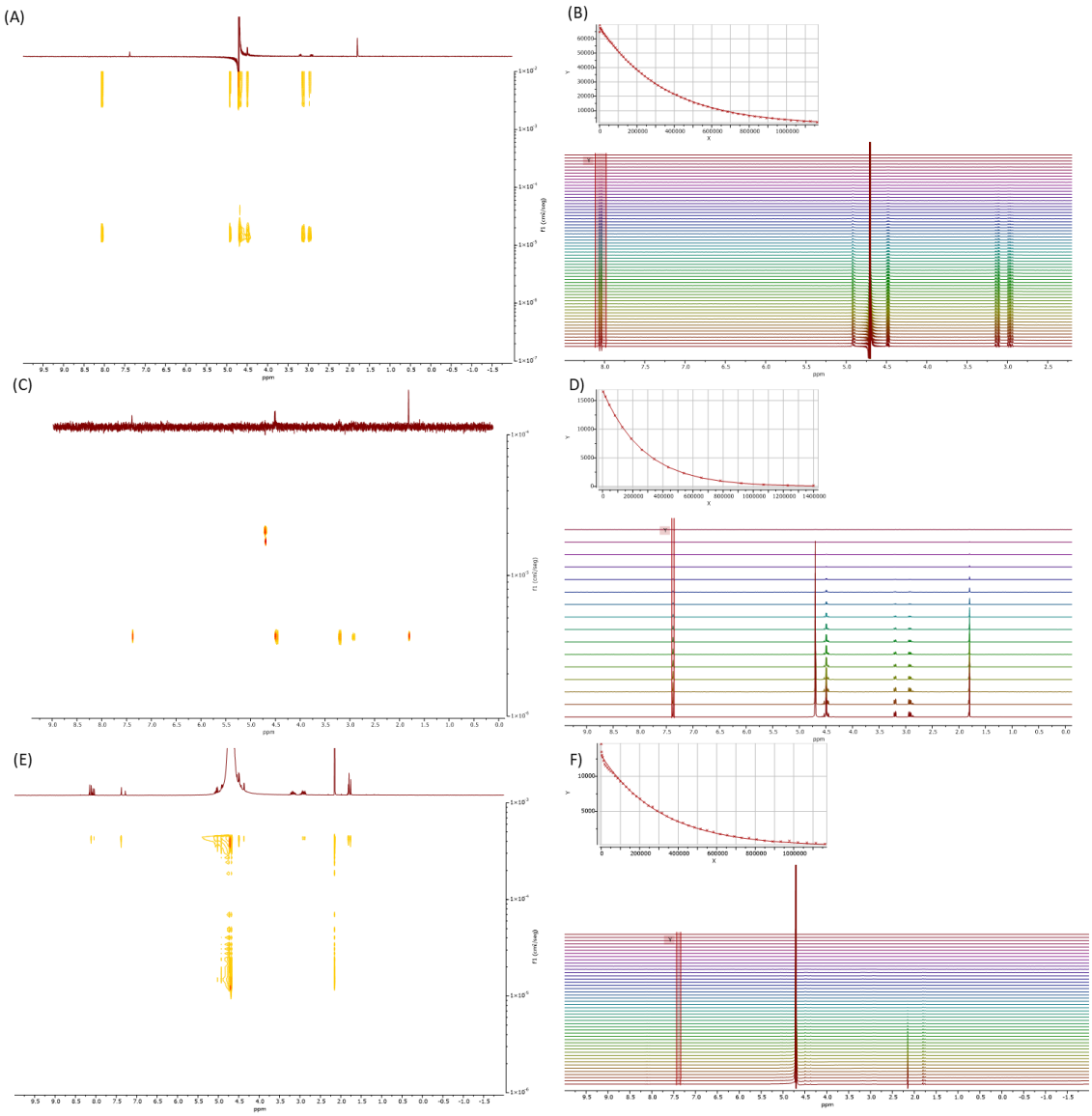


Figure S4. Left column graphs: representative DOSY spectra of **AA** (A), **TT** (C) and **AA**/**TT** (E) at 2mM. This type of spectra was recorded for the three systems at different concentrations from 0.1 to 4 mM. **Right column graphs:** Examples of NMR datasets that were used for calculation of the diffusion constant (from the degree of attenuation of the NMR signals) for the **AA** (B), **TT** (D) and **AA**/**TT** (F) systems at 2mM.

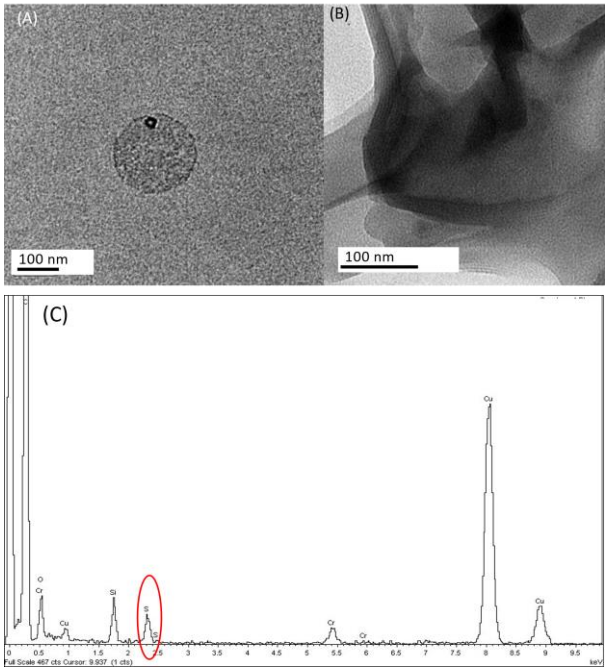


Figure S5. TEM micrographs of **AA** at 1 mM (A) and 2 mM (B). (C) EDX plots for the observed objects, in which the signal corresponding to sulfur is highlighted with a red mark.

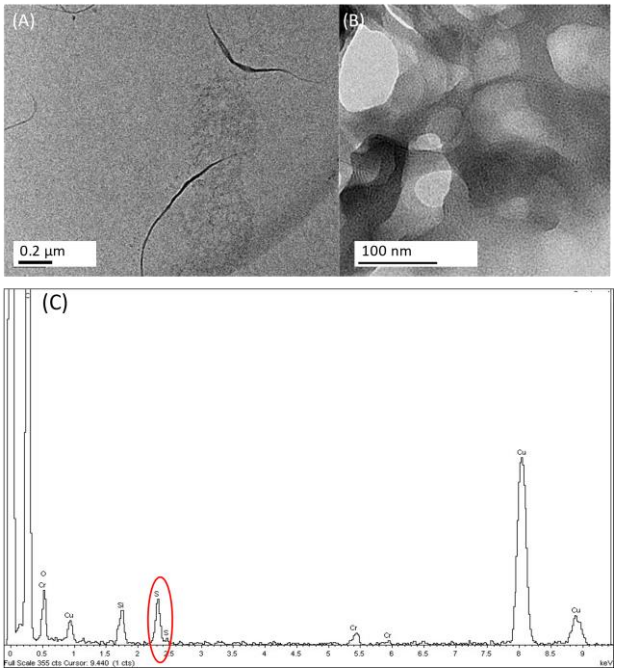


Figure S6. TEM micrographs of **TT** at 1 mM (A) and 2 mM (B). (C) EDX plots for the observed objects, in which the signal corresponding to sulfur is highlighted with a red mark.

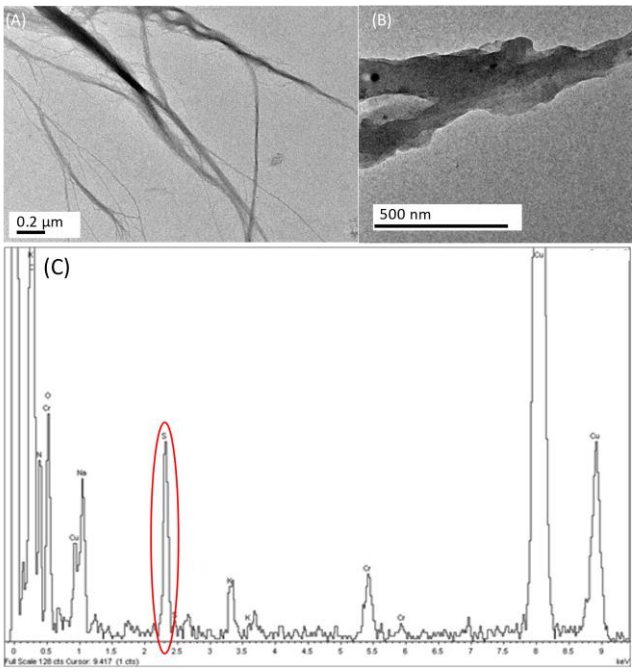


Figure S7. TEM micrographs of **AA**/**TT** at 1 mM (A) and 2 mM (B). (C) EDX plots for the observed objects, in which the signal corresponding to sulfur is highlighted with a red mark.

4. HPLC monitoring of replication kinetics

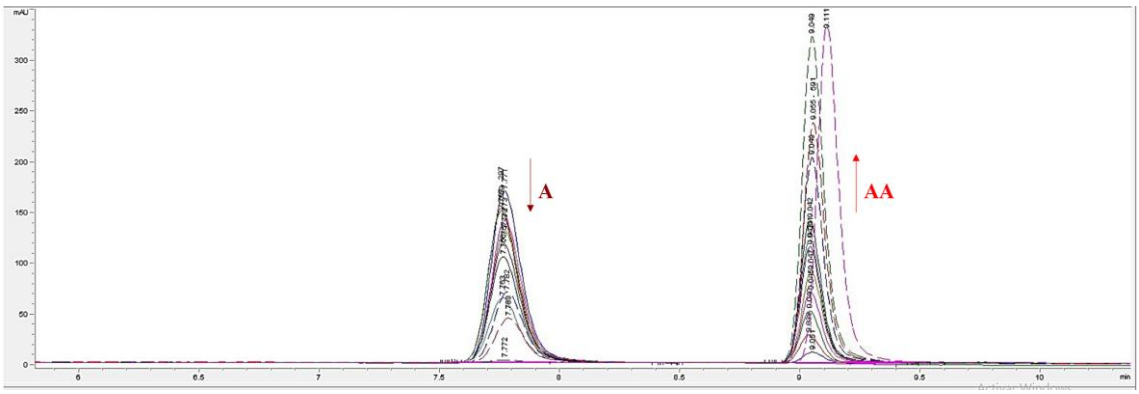


Figure S8. Chromatograms of a replication experiment with **A**, from 0 to 400 h.

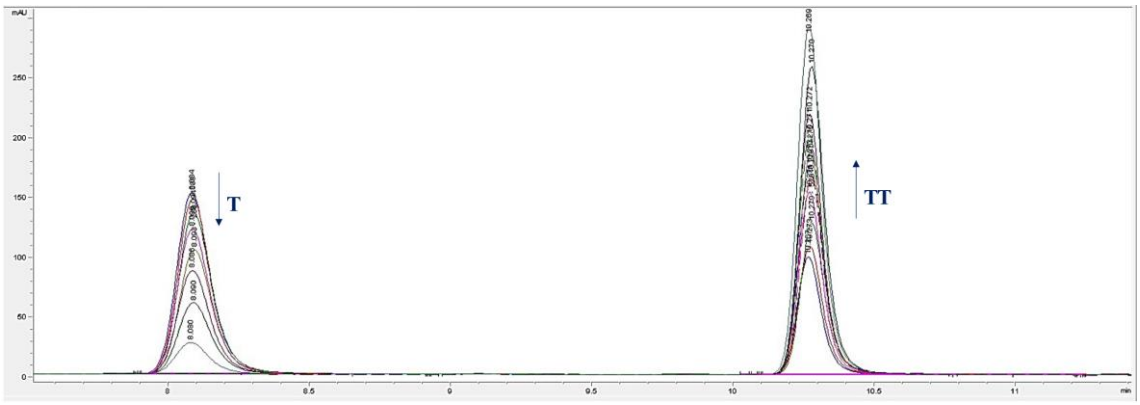


Figure S9. Chromatograms of a replication experiment with **T**, from 0 to 400 h.

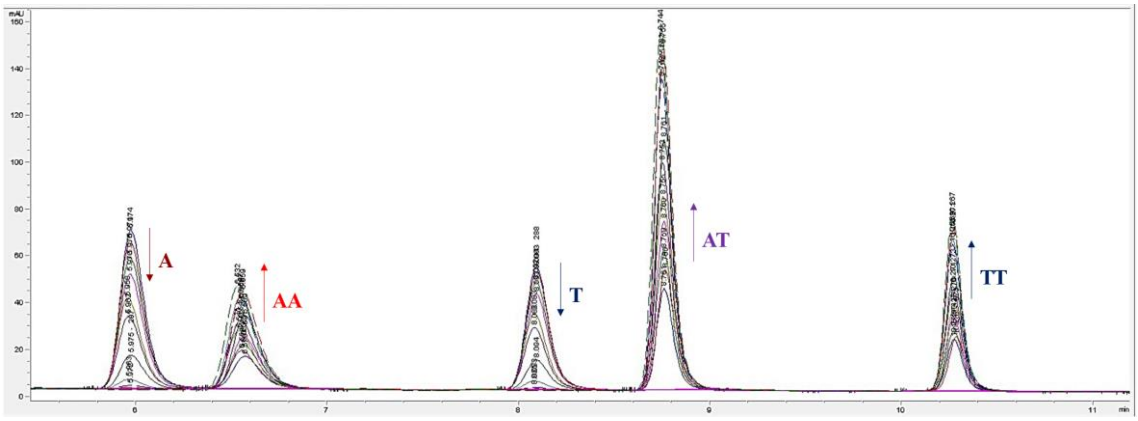


Figure S10. Chromatograms of a replication experiment with **T** and **A**, from 0 to 400 h.

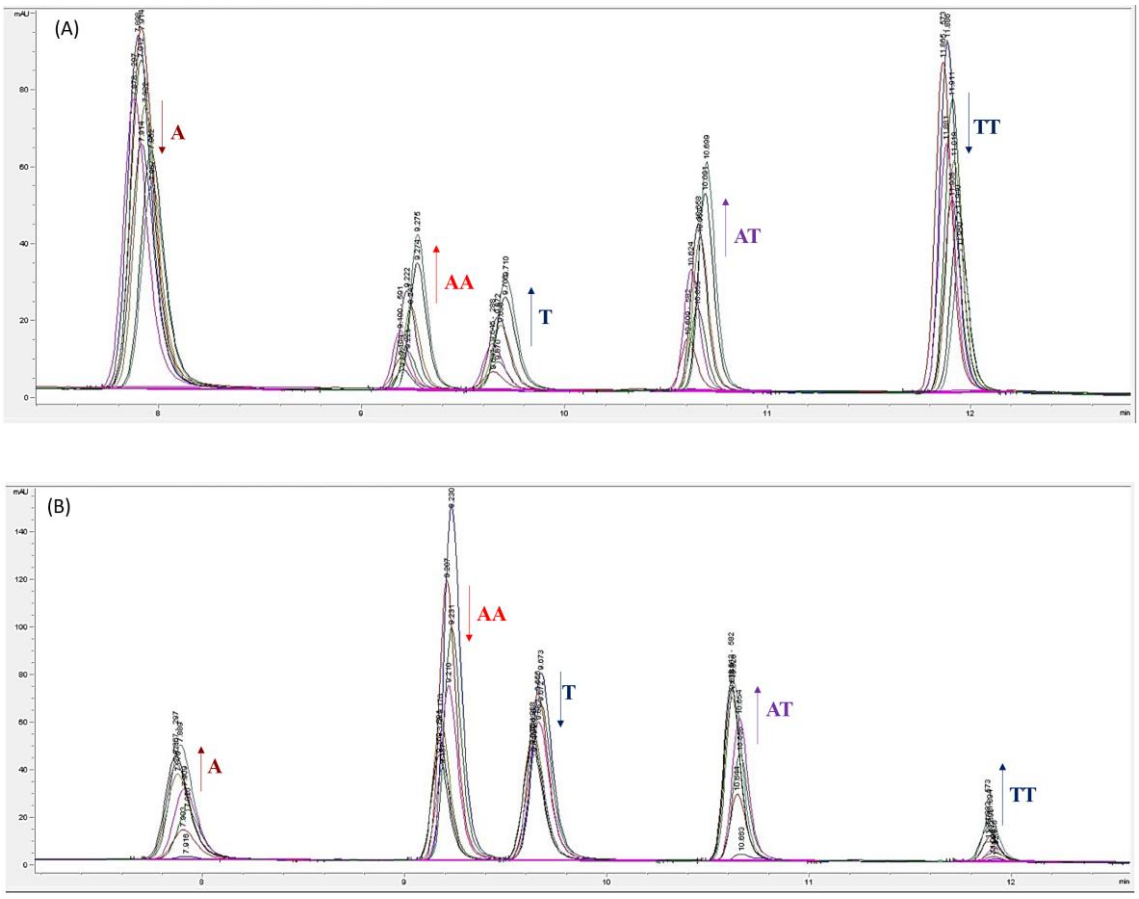


Figure S11. Chromatograms of the exchange reactions between **A** and **TT** (A) or **AA** *and* **T** (B), from 0 to 7 h.

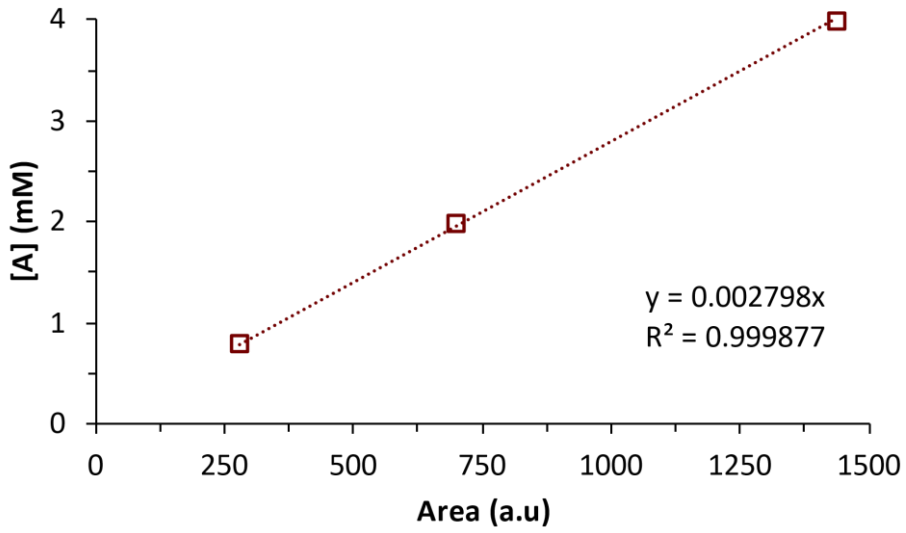


Figure S12. HPLC calibration curve for **A**.

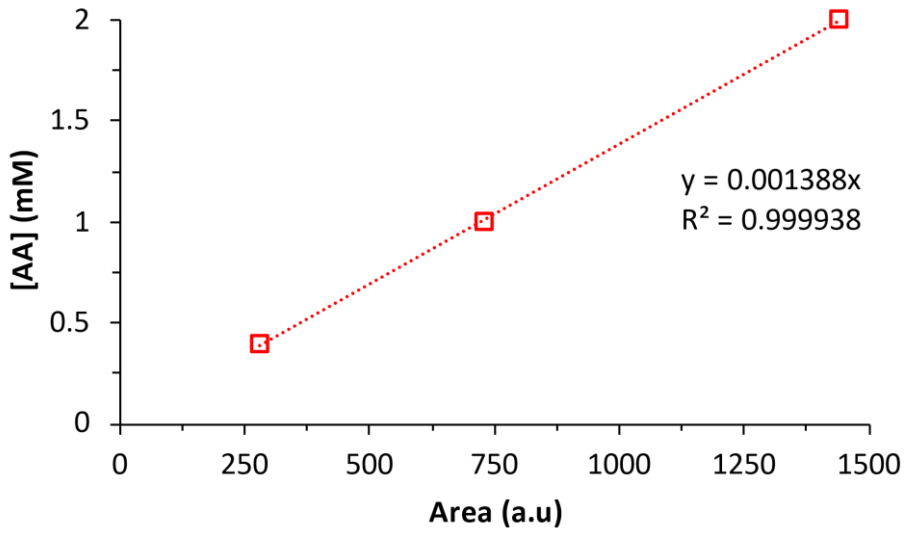


Figure S13. HPLC calibration curve for **AA**.

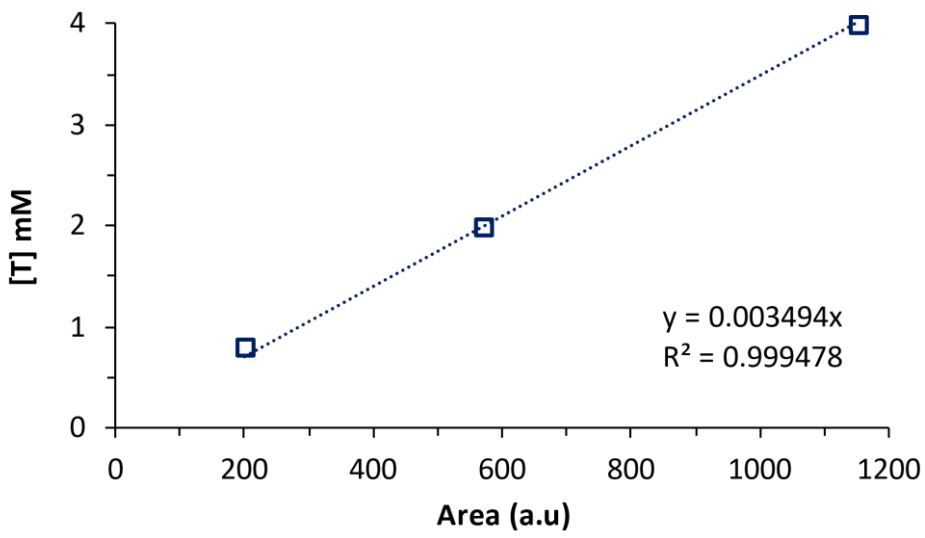


Figure S14. HPLC calibration curve for **T**.

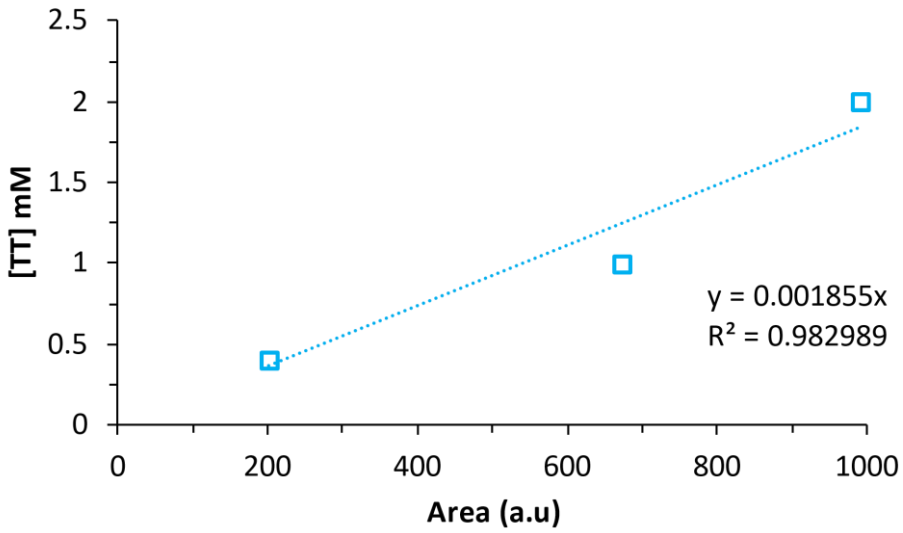


Figure S15. HPLC calibration curve for **TT**.

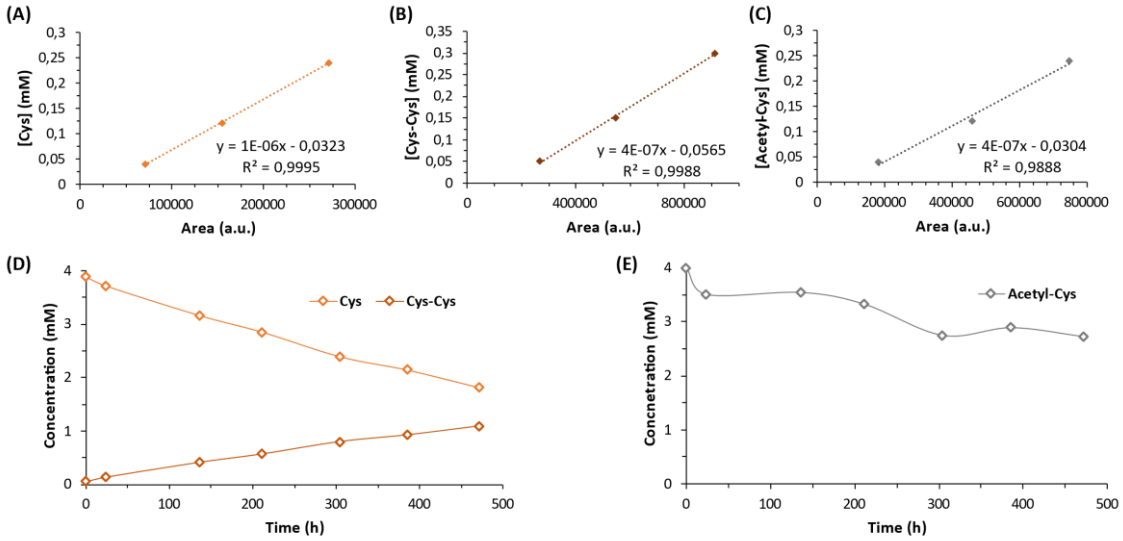


Figure S16. HPLC calibration curves for cysteine (**Cys**) (A)**,** cystine (**Cys-Cys**) (B) and N-acetylcysteine. Kinetic profiles of the oxidation reaction of cysteine (4 mM) into cystine (D) and of N-acetylcysteine (E), in borate buffer (50 mM) at pH 8.2 and 20ºC. The curves show no autocatalytic effect when this oxidation is run in absence of nucleobases as side moieties in the amino acid derivative.

5. Replication experiments

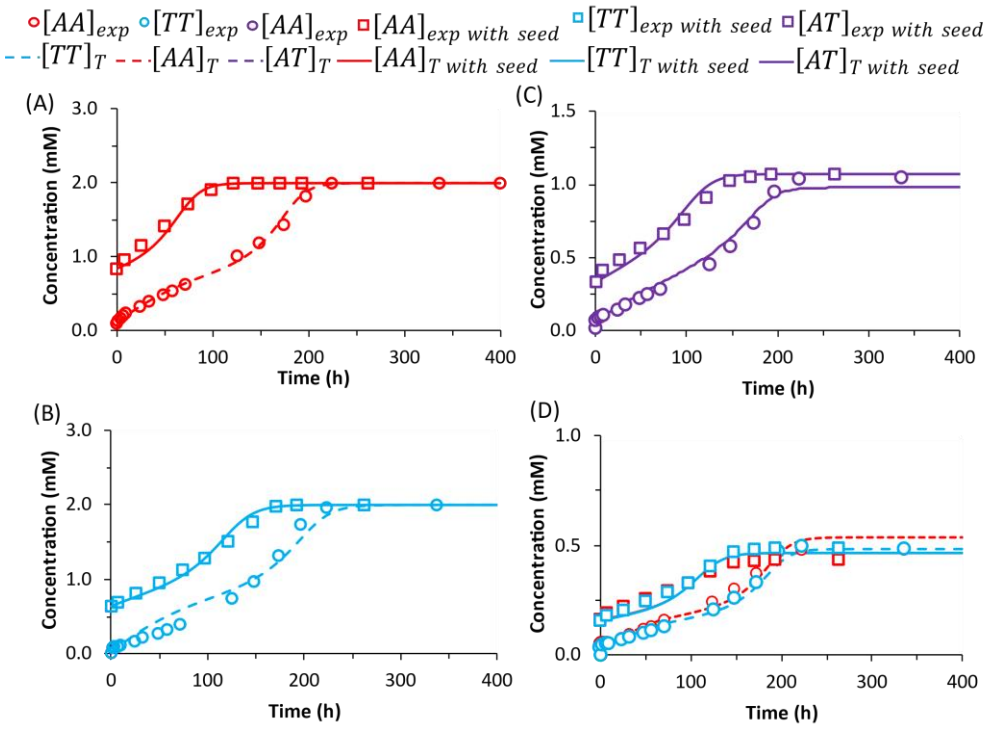


Figure S17. Direct comparison of kinetics between templated and non-templated reactions for the replication of **AA** (A), **TT** (B), **AT** (C) and **AA**/**TT** (D). Panel A merges Figure 3A and C; Panel B merges Figure 3B and 3D; Panels C and D merge curves corresponding to the growth of **AT** or **AA**/**TT** from Figure 3E and 3F. Circles correspond to data from non-templated reactions; squares correspond to data from templated reactions. In all cases, the graphs only display the evolution of the relevant species.

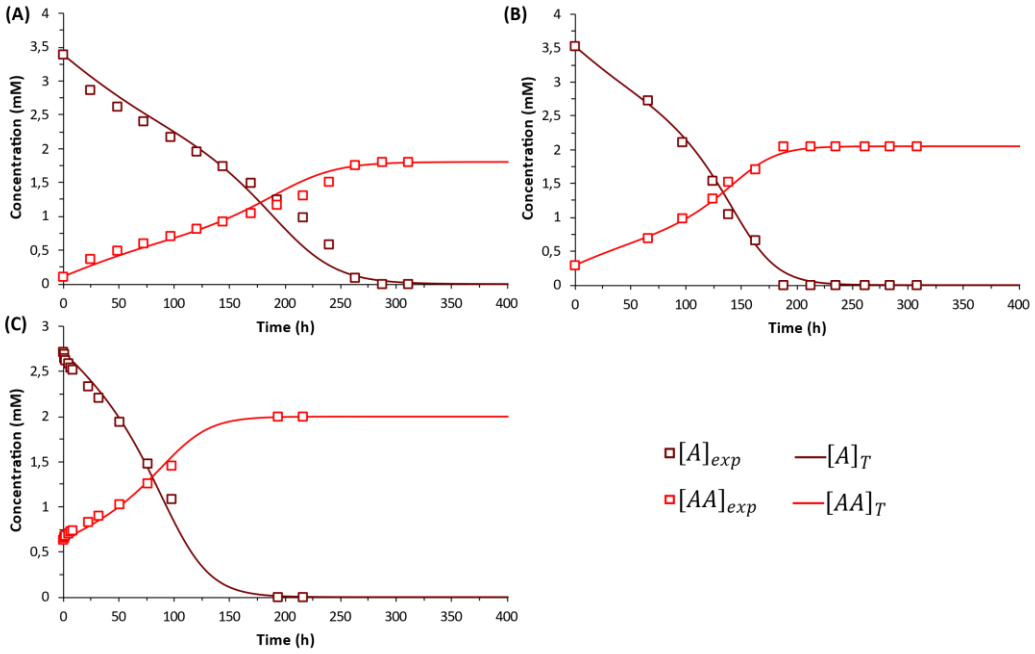


Figure S18. Kinetic profiles corresponding to repetitions of the autocatalytic formation of **AA** from **A** in non-seeded (A, B) and seeded (C) experiments.

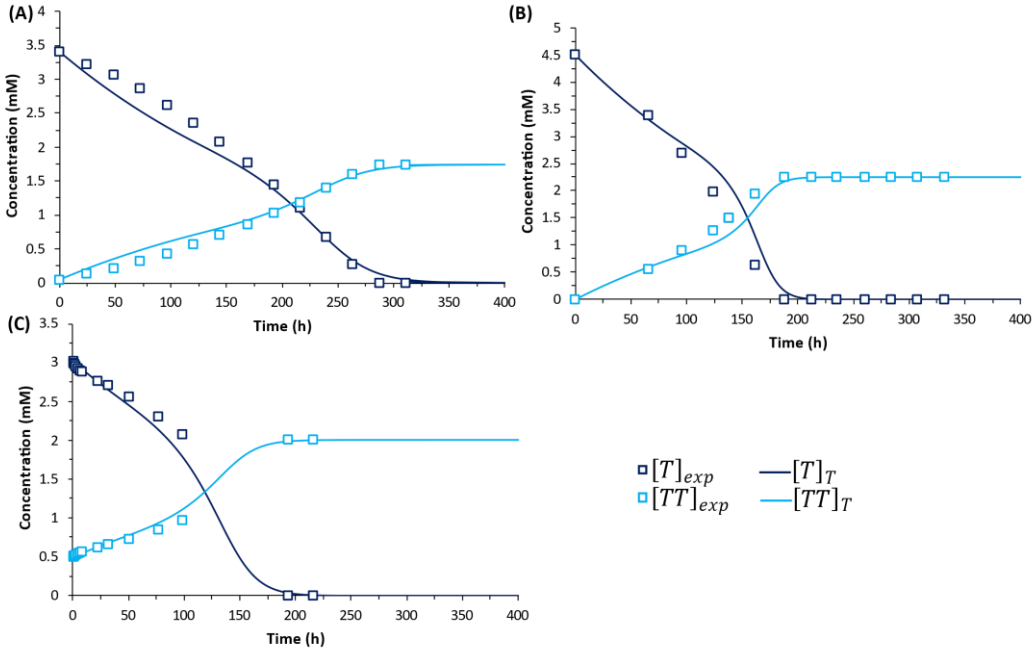


Figure S19. Kinetic profiles corresponding to repetitions of the autocatalytic formation of **TT** from **T** in non-seeded (A, B) and seeded (C) experiments.

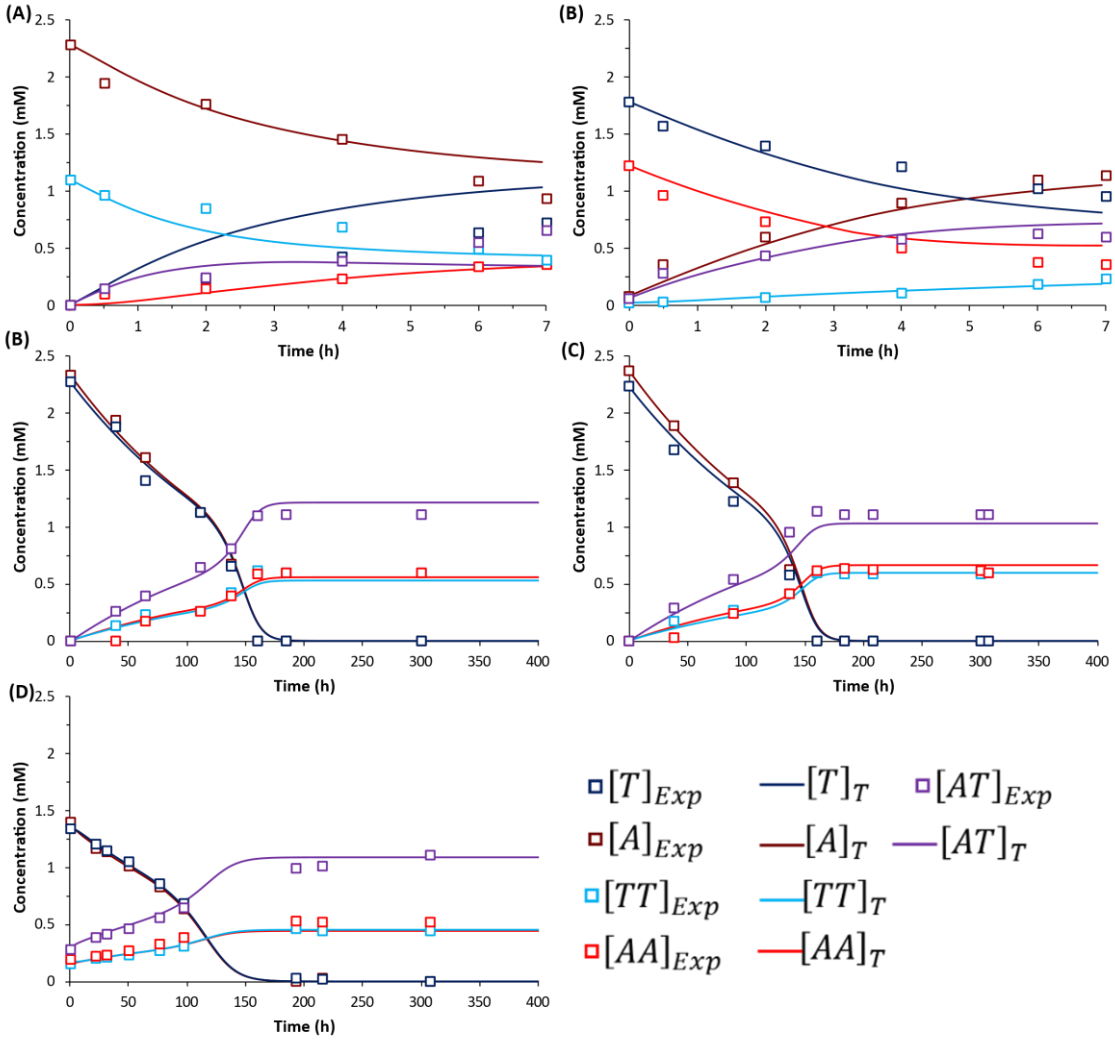
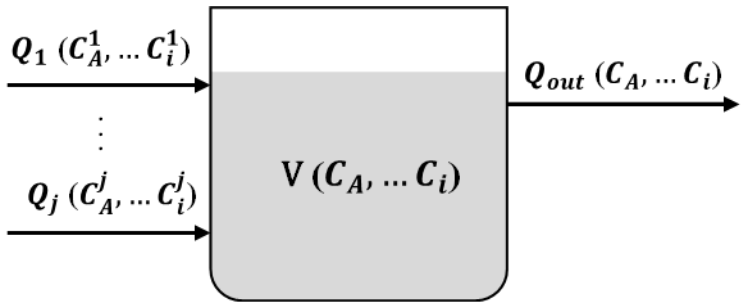


Figure S20. Kinetic profiles corresponding to repetitions of the experiments: **AA + T** (A), **TT + A** (B), **A** + **T** (C, D) and **A** + **T** + 20% seed (E).

6. Mathematical model for kinetic analysis

In this section, it is described the mathematical model employed to reproduce the replication experiments, to fit the experimental data and to determine the corresponding rate constants. The system is considered as a batch reactor with homogeneous concentration of volume **V**:



Into the reactor can enter one or more inlet streams with a volumetric flow of $\boldsymbol{Q_j}$ and a concentration of each component $\boldsymbol{C_i^j}$. From the reactor, there is an outlet stream with a volumetric flow of $\boldsymbol{Q_{out}}$ and a concentration equal to the one inside the reaction, $\boldsymbol{C_i}$.

First, we consider the total mass balance inside the reactor:

$$\frac{\partial V \cdot \rho}{\partial t} = \sum_{j=1}^{j} Q_j \cdot \rho - Q_{out} \cdot \rho$$

Where the variation in the reactor volume can be calculated from the difference between the inlet and outlet flows. As the modelled experiments were with highly diluted aqueous solutions, it can be assumed an equal density in all the system. Once accounted for the variation in volume, the molar balance for each component can be written with the following equation:

$$\frac{\partial N_i}{\partial t} = \sum_{j=1}^{j} Q_j \cdot C_i^j - Q_{out} \cdot C_i + G_i \cdot V$$

Were the variation in the total number of moles for each species is equal to the sum of the inlet minus the outlet, plus a generation term that represents the amount appearing or disappearing through the different chemical reactions. This term is calculated as the summatory of the

products of the stoichiometric coefficient of the species $\boldsymbol{\mu_i}$ and the global reaction rate $\boldsymbol{r_i}$ for each of the involved reactions (refer to Table 1 for all the reactions and their corresponding $\boldsymbol{r_i}$)

$$\boldsymbol{G_i} = \sum \mu_i \cdot r_i$$

In the model, it is assumed that only the dissociated (non-aggregated) species take part in the reactions, and so the formation and disruption of aggregated species also had to be addressed. These supramolecular processes are considered to take place instantaneously and, for simplicity, a critical aggregation concentration (*cac*) is defined. Below the *cac*, all the species are in their monomeric form, while above that concentration they are considered to form aggregates.

$$cac = \frac{1}{Keq} \therefore \begin{cases} C_i = C_T \\ C_{i,g} = 0 \end{cases}, \; C_T < \frac{1}{Keq}$$

$$\begin{cases} C_i = \frac{1}{Keq} \\ C_{i,g} = C_T - \frac{1}{Keq} \end{cases}, \; C_T > \frac{1}{Keq}$$

The convenience of this simplification is discussed in section 6.1 of this supplementary material.

Finally, the resulting system of differential equations is solved using the Euler method, iteratively from the initial conditions until the final time considered.

6.1 Modelling of the aggregation processes

In order to model the kinetics of the catalyzed oxidation reactions, it is first necessary to know the concentration of aggregates in the medium. This concentration is usually defined by equation **Eq-S1**.[4]

$$C_T = (1-p) * C_i + \rho * C_i / 1 * (1 - K_{eq} * C_i)^2 \qquad \textbf{Eq-S1}$$

For this type of compounds, it can be assumed that the aggregation mechanism is of the cooperative or isodesmic type, so that ρ can be said to be less than or equal to 1. This allows simplifying the equation into **Eq- A3** and **Eq-T3**, that is, the equilibrium constant is equal to the

inverse of the critical aggregation concentration. To check if this simplification for calculating the concentration of aggregated compound at each point is adequate, **Eq-S1** was solved analytically, giving different values to ρ and K_{eq}. This allowed representing the obtained concentrations of aggregated species ($\mathbf{C_{ag}}$) versus the total concentration of compound ($\mathbf{C_T}$) for different values of ρ (therefore assuming different mechanisms of the supramolecular polymerization process), and compare them with the same kind of plot in which the simplified expression of the equilibrium constant (equal to 1/*cac*) was considered. Figure S21 shows that the differences between all plots are negligible. Therefore, this simplification will be used to calculate the aggregation constant and the concentration of aggregated compound.

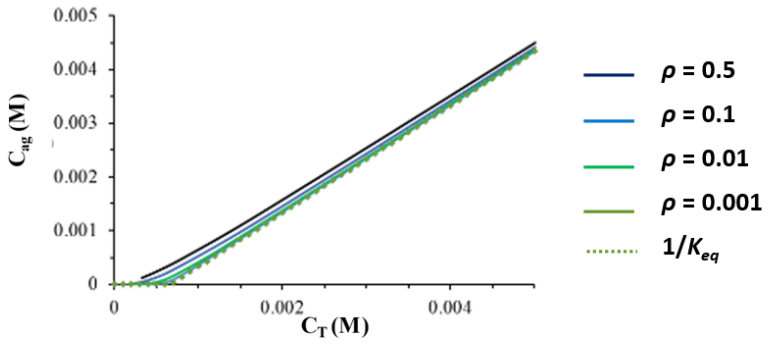


Figure S21. Representation of the concentration of aggregated compound versus the total concentration for different ρ values and the same K_{eq}.

7. Calculation of the different kinetic and equilibrium constants

Calculation of the rate constants for the different oxidation and exchange reactions, and of the equilibrium constants for aggregation processes, were carried out through fitting of the experimental data using a program designed in Matlab.

7.1 Determination of the reaction order of the catalyzed reactions (Eq-A2 and Eq-T2)

The reactions to be studied are described in Table 1 of the main text. Since the mechanism of the reactions was not known (both in its uncatalyzed and autocatalyzed versions), and neither the reaction orders, eight different rate equations with different orders with respect to **A** and $\mathbf{AA}_{ag}$ (Table S1), or **T** and $\mathbf{TT}_{ag}$ (Table S2), were initially proposed. In those equations, reaction orders vary between 0.5 and 1 for **A** or **T** and, 1 and 2 for $\mathbf{AA}_{ag}$ or $\mathbf{TT}_{ag}$. The eight different equations were used to fit five different experiments with distinct initial conditions.

In the reaction from **A** to **AA**, both the error and R^2 values show a better fit for the equation in Box 8 of Table S1. These data were also corroborated with dispersion graphs (Figures S22), where a lower dispersion of data is clearly observed when the overall reaction order is one in the uncatalyzed reaction and three for the autocatalytic one (Figure S22, panel H). In particular, the order is one with respect to the monomer and two with respect to the catalytic aggregate. Similar results were obtained for the oxidation of **T** into **TT** (the best dispersion of data correspond to the fit of Box 8 in Table S2, panel H in Figure S23).

Table S1. Rate equations with different orders for the autocatalytic reaction of **A**, used to fit data from five different experiments, and the respective obtained constant values, mean absolute percentage error and R^2.

	Equation	Constant	Experiment	Error (%)*	R^2
Box 1	$k_A[A]^{0.5}$	$3.80 \cdot 10^{-3}$ $mM^{-0.5}$ h^{-1}	Figure 3A	10	0.99
			Figure 3C	22	0.96
			Figure S18A	22	0.92
			Figure S18B	12	0.99
			Figure S18C	3	0.95
	$k_{Ac}[A]^{0.5}[AA_{ag}]$	$5.34 \cdot 10^{-3}$ $mM^{-0.5}$ h^{-1}			
Box 2	$k_A[A]^{0.5}$	$3.91 \cdot 10^{-3}$ $mM^{-0.5}$ h^{-1}	Figure 3A	8	0.99
			Figure 3C	15	0.97
			Figure S18A	22	0.92
			Figure S18B	14	0.94
			Figure S18C	3	0.99
	$k_{Ac}[A]^{0.5}[AA_{ag}]^2$	$7.63 \cdot 10^{-3}$ $mM^{-1.5}$ h^{-1}			
Box 3	$k_A[A]^{0.5}$	$3.49 \cdot 10^{-3}$ $mM^{-0.5}$ h^{-1}	Figure 3A	11	0.99
			Figure 3C	16	0.92
			Figure S18A	18	0.93
			Figure S18B	10	0.96
			Figure S18C	4	0.99
	$k_{Ac}[A][AA_{ag}]$	$5.20 \cdot 10^{-3}$ mM^{-1} h^{-1}			
Box 4	$k_A[A]^{0.5}$	$3.75 \cdot 10^{-3}$ $mM^{-0.5}$ h^{-1}	Figure 3A	11	0.99
			Figure 3C	16	0.96
			Figure S18A	18	0.92
			Figure S18B	10	0.96
			Figure S18C	4	0.99
	$k_{Ac}[A][AA_{ag}]^2$	$759 \cdot 10^{-3}$ mM^{-2} h^{-1}			
Box 5	$k_A[A]$	$2.05 \cdot 10^{-3}$ h^{-1}	Figure 3A	9	0.99
			Figure 3C	10	0.94
			Figure S18A	17	0.95
			Figure S18B	10	0.96
			Figure S18C	4	0.99
	$k_{Ac}[A]^{0.5}[AA_{ag}]$	$7.38 \cdot 10^{-3}$ $mM^{-0.5}$ h^{-1}			

*Mean absolute percentage error

Table S1. Continuation.

	Equation	Constant	Experiment	Error (%)*	R^2
Box 6	$k_A[A]$	$2.29 \cdot 10^{-3}$ h^{-1}	Figure 3A	19	0.99
			Figure 3C	16	0.97
	$k_{Ac}[A]^{0.5}[AA_{ag}]^2$	$1.46 \cdot 10^{-2}$ $mM^{-1.5}$ h^{-1}	Figure S18A	25	0.93
			Figure S18B	19	0.96
			Figure S18C	4	0.97
Box 7	$k_A[A]$	$2.10 \cdot 10^{-3}$ h^{-1}	Figure 3A	26	0.99
			Figure 3C	27	0.96
	$k_{Ac}[A][AA_{ag}]$	$8.50 \cdot 10^{-3}$ mM^{-1} h^{-1}	Figure S18A	14	0.96
			Figure S18B	24	0.95
			Figure S18C	6	0.98
Box 8	$k_A[A]$	$2.22 \cdot 10^{-3}$ h^{-1}	Figure 3A	4	0.99
			Figure 3C	3	0.99
	$k_{Ac}[A][AA_{ag}]^2$	$9.68 \cdot 10^{-3}$ mM^{-2} h^{-1}	Figure S18A	6	0.99
			Figure S18B	2	0.99
			Figure S18C	2	0.99

*Mean absolute percentage error

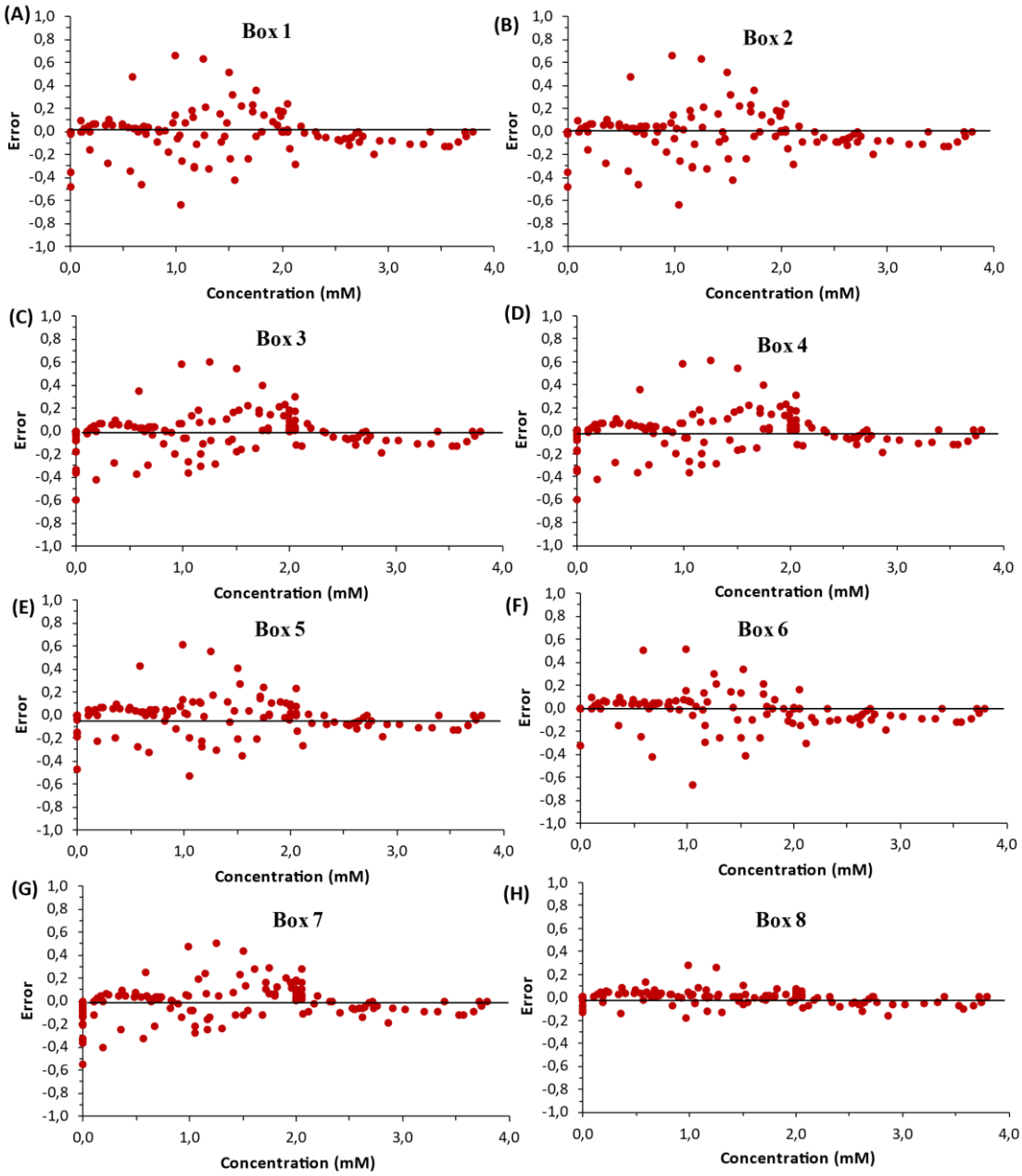


Figure S22. Dispersion graphs for the different box (1 to 8) parameters from Table S1. Concentration in the *x* axis refers to the monomer **A**.

To make the data in this Figure S22 and the fitting protocol cleared, we have added, as an example, two new graphs below that aim to make easier the comparison between different fits. In these graphs, the error associated to each datapoint is represented, for comparison, from kinetic experiments of the **AA** autocatalysis that were fitted to three possible sets of equations: $k_A[A]^{0.5}$, $k_{Ac}[A]^{0.5}[AA_{ag}]$ (blue data, (left graph)); $k_A[A], k_{Ac}[A][AA_{ag}]^2$ (green data (both

graphs)); and $k_A[A], k_{Ac}[A][AA_{ag}]$ (orange data (right graph)). Through these direct comparisons, it is actually possible to evaluate in which phase of the reaction we are getting the highest errors. At high **A** concentration, from 2.5 mM, only the monomer is present, and the reaction rate is mainly driven by the non-catalyzed oxidation reaction. In this phase the errors are lower, and the fits point to an order 1 with respect to monomer, which actually agrees with the literature (refs. 49 and 50 in the manuscript). At lower concentrations of **A** (between 2.5 and 0.5 mM), when there is already sufficient dinucleobase compound, the autocatalyzed reaction starts to gain a major weight. In that range, the errors (dispersion of data) for rate equations that do not match the real process kinetics get very amplified (notice the greater differences between the errors obtained for blue and orange data, compared to the green ones, in this region of both graphs). For the orange data, the high dispersion becomes even larger at the lowest concentrations of **A**, where the reactions should be much slower if the order with respect to replicator was one. This can be observed in the rest of the kinetics where we have put this order of reaction for replicators. Such careful error analysis of the different fits gives as a result that the correct rate equation corresponds to the green data.

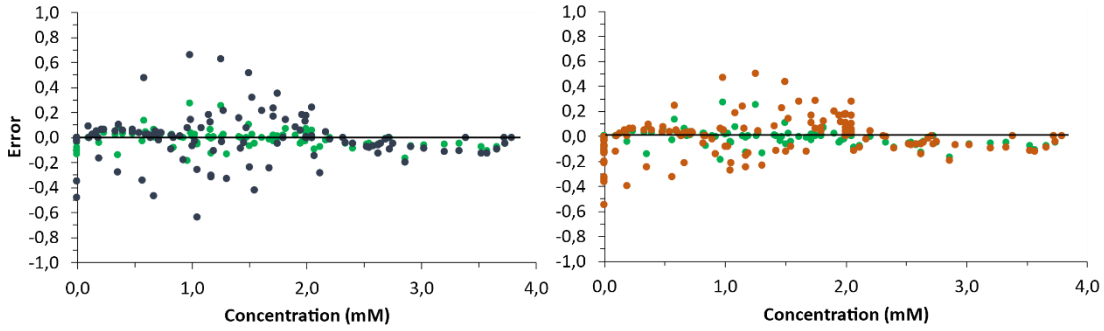


Table S2. Rate equations with different orders for the autocatalytic reaction of **T**, used to fit data from five different experiments, and the respective obtained constant values, mean absolute percentage error and R^2.

	Equation	Constant	Experiment	Error (%)*	R^2
Box 1	$k_T[T]^{0.5}$	$3.89 \cdot 10^{-3}$ $mM^{-0.5}$ h^{-1}	Figure 3B	21	0.99
			Figure 3D	15	0.96
	$k_{Tc}[T]^{0.5}[TT_{ag}]$	$6.33 \cdot 10^{-3}$ $mM^{-0.5}$ h^{-1}	Figure S19A	36	0.87
			Figure S19B	24	0.88
			Figure S19C	6.6	0.98
Box 2	$k_T[T]^{0.5}$	$4.10 \cdot 10^{-3}$ $mM^{-0.5}$ h^{-1}	Figure 3B	17	0.99
			Figure 3D	17	0.95
	$k_{Tc}[T]^{0.5}[TT_{ag}]^2$	$1.02 \cdot 10^{-2}$ $mM^{-1.5}$ h^{-1}	Figure S19A	35	0.88
			Figure S19B	25	0.89
			Figure S19C	5	0.99
Box 3	$k_T[T]^{0.5}$	$3.72 \cdot 10^{-3}$ $mM^{-0.5}$ h^{-1}	Figure 3B	25	0.99
			Figure 3D	22	0.91
	$k_{Tc}[T][TT_{ag}]$	$6.49 \cdot 10^{-2}$ mM^{-1} h^{-1}	Figure S19A	31	0.96
			Figure S19B	34	0.88
			Figure S19C	7	0.97
Box 4	$k_T[T]^{0.5}$	$3.94 \cdot 10^{-3}$ $mM^{-0.5}$ h^{-1}	Figure 3B	17	0.99
			Figure 3D	15	0.96
	$k_{Tc}[T][TT_{ag}]^2$	$1.06 \cdot 10^{-2}$ mM^{-2} h^{-1}	Figure S19A	40	0.90
			Figure S19B	32	0.90
			Figure S19C	5	0.99
Box 5	$k_T[T]$	$2.12 \cdot 10^{-3}$ h^{-1}	Figure 3B	21	0.99
			Figure 3D	14	0.97
	$k_{Tc}[T]^{0.5}[TT_{ag}]$	$9.10 \cdot 10^{-3}$ $mM^{-0.5}$ h^{-1}	Figure S19A	18	0.94
			Figure S19B	27	0.94
			Figure S19C	5	0.98

*Mean absolute percentage error

Table S2. Continuation.

	Equation	Constant	Experiment	Error (%)*	R^2
Box 6	$k_T[T]$	$2.29 \cdot 10^{-3}$ h^{-1}	Figure 3B	19	0.99
			Figure 3D	16	0.97
	$k_{Tc}[T]^{0.5}[TT_{ag}]^2$	$1.46 \cdot 10^{-2}$ $mM^{-1.5}$ h^{-1}	Figure S19A	25	0.93
			Figure S19B	19	0.96
			Figure S19C	4	0.97
Box 7	$k_T[T]$	$2.10 \cdot 10^{-3}$ h^{-1}	Figure 3B	26	0.99
			Figure 3D	27	0.96
	$k_{Tc}[T][TT_{ag}]$	$8.50 \cdot 10^{-3}$ mM^{-1} h^{-1}	Figure S19A	14	0.96
			Figure S19B	24	0.95
			Figure S19C	6	0.98
Box 8	$k_T[T]$	$2.04 \cdot 10^{-3}$ h^{-1}	Figure 3B	7	0.99
			Figure 3D	4	0.99
	$k_{Tc}[T][TT_{ag}]^2$	$2.24 \cdot 10^{-2}$ mM^{-2} h^{-1}	Figure S19A	5	0.99
			Figure S19B	13	0.99
			Figure S19C	5	0.99

*Mean absolute percentage error

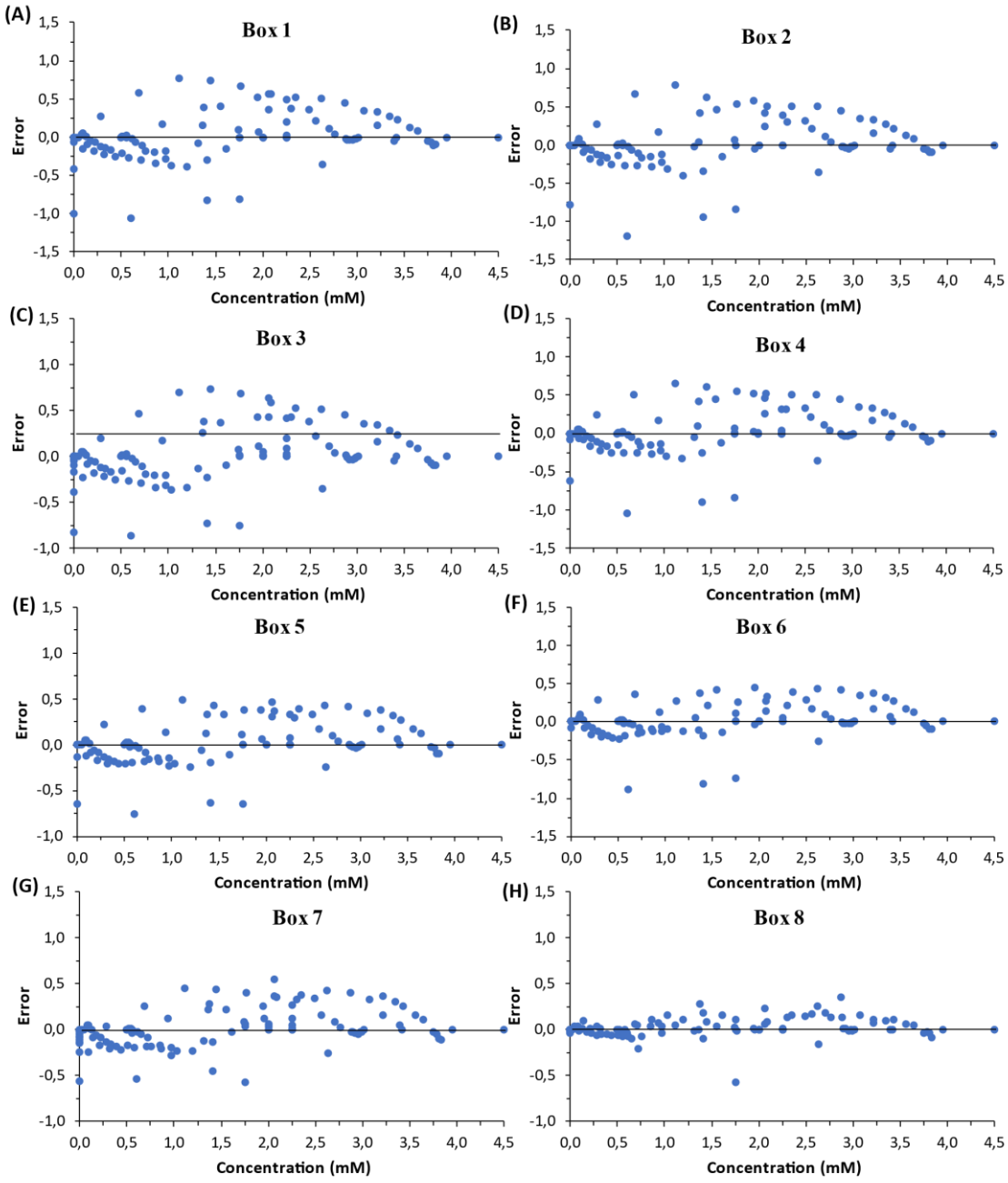


Figure S23. Dispersion graphs for the different box (1 to 8) parameters from Table S2. Concentration in the *x* axis refers to the monomer **T**.

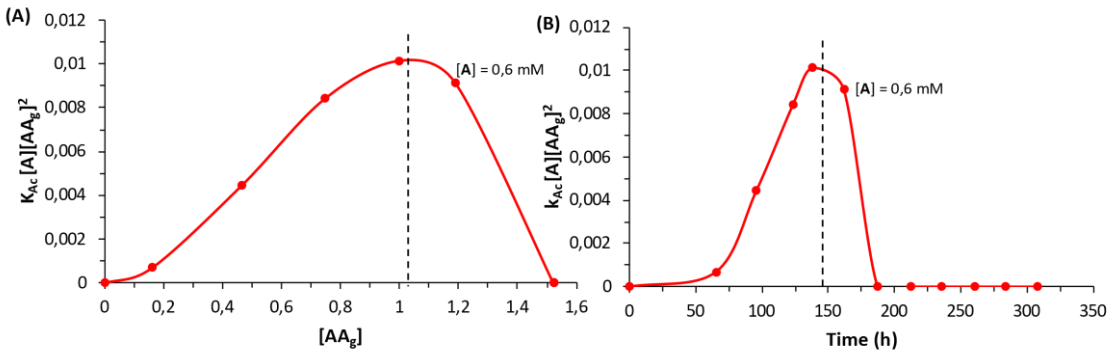


Figure S24. Representation of the **AA** production rate versus aggregated replicator concentration (mM) (A) or versus time (B) for experiment S18A. Both graphs are divided into two zones separated by a dotted line. The left zone shows an exponential growth of the reaction velocity, more clearly identified in the plot versus time (B). In the right zone, the reaction rate decreases rapidly, due to the drastic consumption (i.e., there is no middle point between that with [**A**] = 0.6 mM and the next one in which **A** has been fully consumed) of the starting material, which would of course not be observed in an open reactor with constant supply of the feedstock reagent.

7.2 Summary of errors estimation for all the experiments

Table S3. List of experiments and their replicas, with calculated mean absolute percentage error and R^2 values. Experiments marked in bold are included in the main text.

	Experiment	Figure	Error (%)*	R^2
Box 1	**A + A**	**Figure 3A**	**4**	**0.99**
	A + A + seed	**Figure 3C**	**3**	**0.99**
	A + A	Figure S18A	6	0.99
	A + A	Figure S18B	2	0.99
	A + A + seed	Figure S18C	2	0.99
Box 2	**T + T**	**Figure 3B**	**7**	**0.99**
	T + T + seed	**Figure 3D**	**4**	**0.99**
	T + T	Figure S19A	5	0.99
	T + T	Figure S19B	13	0.99
	T + T + seed	Figure S19C	5	0.99
Box 3	**AA + T**	**Figure 3G**		**0.99**
	TT + A	**Figure 3H**		**0.99**
	AA + T	Figure S20A		0.96
	TT + A	Figure S20B		0.97
Box 4	**A + T**	**Figure 3E**	**3**	**0.99**
	A + T + seed	**Figure 3F**	**3**	**0.99**
	A + T	Figure S20C	4	0.99
	A + T	Figure S20D	3	0.99
	A + T + seed	Figure S20E	3	0.99

*Mean absolute percentage error

8. Simulations in an open reactor

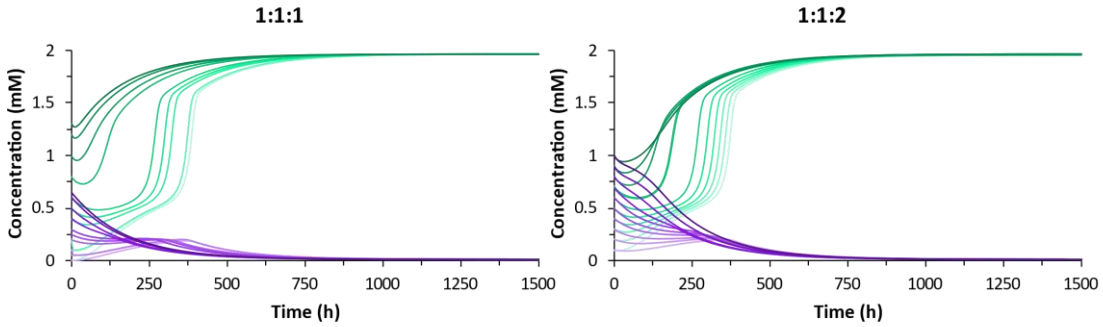


Figure S25. Simulations of the network evolution when fed with two input streams (Q_{in} = 1 µL/min) of **A** and **T**, in absence of replicators at time 0 (A), and in presence of replicators **AA**/**TT** + **AT** with an equimolar proportion (0 – 2 mM) (B) or **AA**/**TT** + **AT** with double molar concentration of **AA**/**TT** compared to **AT** (0 – 2 mM) (B).

Table S4. Rate equations and constants assumed for simulation of the network behavior in a hypothetical scenario with frozen disulfide exchange reactions.

Reaction	Equation	Constant
$A + TT \rightarrow T + AT$	$k_{e_1}[A][TT]$	$1.5 \cdot 10^{-7}$ mM^{-1} h^{-1}
$T + AA \rightarrow A + AT$	$k_{e_2}[T][AA]$	$2.3 \cdot 10^{-7}$mM^{-1} h^{-1}
$A + AT \rightarrow T + AA$	$k_{e_3}[A][AT]$	$1.3 \cdot 10^{-7}$mM^{-1} h^{-1}
$T + AT \rightarrow A + TT$	$k_{e_4}[T][AT]$	$7 \cdot 10^{-8}$mM^{-1} h^{-1}

9. References

1. K. Barlos, O. Chatzi, D. Gatos and G. Stavropoulos, *Int. J. Peptide Protein. Res.,* 1991, **37**, 513.
2. F. García-Martín, N. Bayó-Puxan, L. J. Cruz, J. C. Bohling and F. Albericio, *QSAR Comb. Sci*., 2007, **26**, 1027.
3. Y. C. Huang, C. Cao, X. L. Tan, X. Li and L. Liu, *Org. Chem. Front.* 2014, **1**, 1050.
4. Z. Chen, A. Lohr, C. R. Saha-Möller and F. Würthner, *Chem. Soc. Rev.,* 2009, **38**, 564.