

Supporting Information

Thermodynamics and Kinetics of Early Stages of Carbon Dots Formation: A Case of Citric Acid and Ethylenediamine Reaction

Martin Pykal, Jela Nociarová, David Řeha, Juraj Filo, Marek Šebela, Petr Zajiček, Markéta Paloncýová, Chiara Olla, Francesca Mocci, Antonio Cappai, Carlo Maria Carbonaro, Lukáš Zdražil, Radek Zbořil, Andrey L. Rogach, Miroslav Medved’, Michal Otyepka**

E-mail: miroslav.medved@upol.cz, michal.otyepka@upol.cz

Content

1	General experimental details	2
2	NMR of standards used for comparison with studied samples	4
3	¹ H NMR of samples prepared by heating of CA/EDA mixtures for 1 h at 150–240 °C	14
4	Detailed NMR analysis of sample CD-150-1 – structure determination of 12	17
5	MALDI-TOF analysis of samples CD-150-1, CD-180-1, CD-210-1 and CD-240-1	20
6	¹ H NMR of samples prepared by heating of CA/EDA mixtures (10 h/150–240 °C)	21
7	Detailed analysis of the sample CD-180-10	23
8	The transformation of 12 to IPCA	29
9	Calculated and experimental ¹³ C and ¹⁵ N NMR shifts for intermediates 7, 8 and 12	31
10	XPS measurements	32
11	Computational details	35
12	Computational analysis of the reaction mechanism	40
13	Heuristic molecular dynamics	55
14	Classical MD simulations	58

1. General experimental details

Citric acid (CA) monohydrate, ethylenediamine (EDA) and all other chemicals were purchased from Fluorochem Ltd and used as received. Acetonitrile was dried using molecular sieves (4Å, Thermo Fisher Scientific).

Hydrothermal syntheses were conducted in 25 mL teflon- and PPE-lined autoclaves (Cambridge Energy Solutions, Ltd.) in the muffle furnace (Nabetherm B180) or laboratory oven (Binder ED53). Exact temperatures (150–240 °C) and times (1–10h) are designated below. The furnace was set to reach the designated temperature in 30 minutes. If not stated otherwise, all experiments were conducted at 5.00 mmol scale, using 1.05 g of CA, 0.334 mL of EDA and deionized water (10 mL). After cooling to rt, the reaction mixtures were evaporated to dryness using a rotary evaporator (heating bath set to 65 °C).

NMR spectra were recorded with Bruker Ascend™ 400 (400 MHz for ^1H , 101 MHz for ^{13}C) or Varian NMR System 600 (600 MHz for ^1H and 151 MHz for ^{13}C) instruments. Chemical shifts (δ) are given relative to tetramethylsilane (for ^1H and ^{13}C) or nitromethane (for ^{15}N). Samples were prepared in deuterated water (^1H δ 4.79 ppm) and DMSO- d_6 (^1H δ 2.50, ^{13}C δ 39.5 ppm).

UV-vis spectra were measured on Agilent 8453 absorption spectrophotometer.

Size-exclusion chromatography was performed on Sephadex G-10 (Cytiva) as the stationary phase using deionized water as eluent (approx. column diameter \times height: 3 \times 25 cm). Thin layer chromatography (TLC) was performed with Merck plates (Silica Gel 60, F-254). Compounds were detected at 254 nm and 365 nm, respectively.

All MS measurements were conducted using a Microflex LRF20 MALDI-TOF mass spectrometer (Bruker Daltonik, Bremen, Germany) equipped with a 60-Hz nitrogen laser ($\lambda_{\text{max}} = 337$ nm). The instrument was operated in the reflectron positive ion mode and controlled using flexControl 3.4 spectra acquisition software (Bruker Daltonik). The instrument parameters were as follows: IS1 voltage of 19 kV, IS2 voltage of 16.2 kV, lens voltage of 8.9 kV, reflectron voltage of 20 kV, and detector voltage of 1650 V. Delayed extraction was employed with the pulsed ion extraction time set to 150 ns. The matrix used was 0.1 mol l $^{-1}$ α -cyano-4-hydroxycinnamic acid (Bruker Daltonik) in acetone: water (4:1, v/v) containing 0.1 mmol l $^{-1}$ cetyltrimethylammonium bromide.^[1] Initially, 1- μ l aliquots of the matrix were deposited onto sample positions on an MSP BigAnchor 96 BC target plate (Bruker Daltonik) and allowed to dry at ambient temperature. The analyzed samples (2 mg/ml) were then applied (0.5 μ l) onto the crystallized matrix and allowed to dry. Mass spectra were acquired from 2000

single-pulse shots. Matrix peaks at m/z 190, 284 and 379 were utilized for external calibration. The acquired mass spectra were exported as TXT formatted files for further evaluation.

High resolution mass spectrometry (HRMS) was performed on Thermo Scientific Orbitrap Velos Pro mass spectrometer, using ion source HESI (heated electrospray), capillary temperature 350 °C, source heater temperature 300 °C, mass range 80-600 m/z .

XPS was carried out using a Nexsa G2 (Thermo Fisher Scientific) spectrometer equipped with an Al $K\alpha$ source. (12 kV, spot size is 200 μm in diameter). The survey spectra were acquired using pass energy 150 eV, step size 1 eV, the HR-XPS spectra were acquired using pass energy 50 eV, step size 0.1 eV. The obtained data were evaluated and deconvoluted with the Advantage software package (Thermo Fisher Scientific).

Steady-state PL measurements were performed on an FLS980 fluorescence spectrometer (Edinburgh Instruments) equipped with a 450 W xenon arc lamp. The PL properties of individual extracted fractions were investigated in diluted aqueous solutions at room temperature.

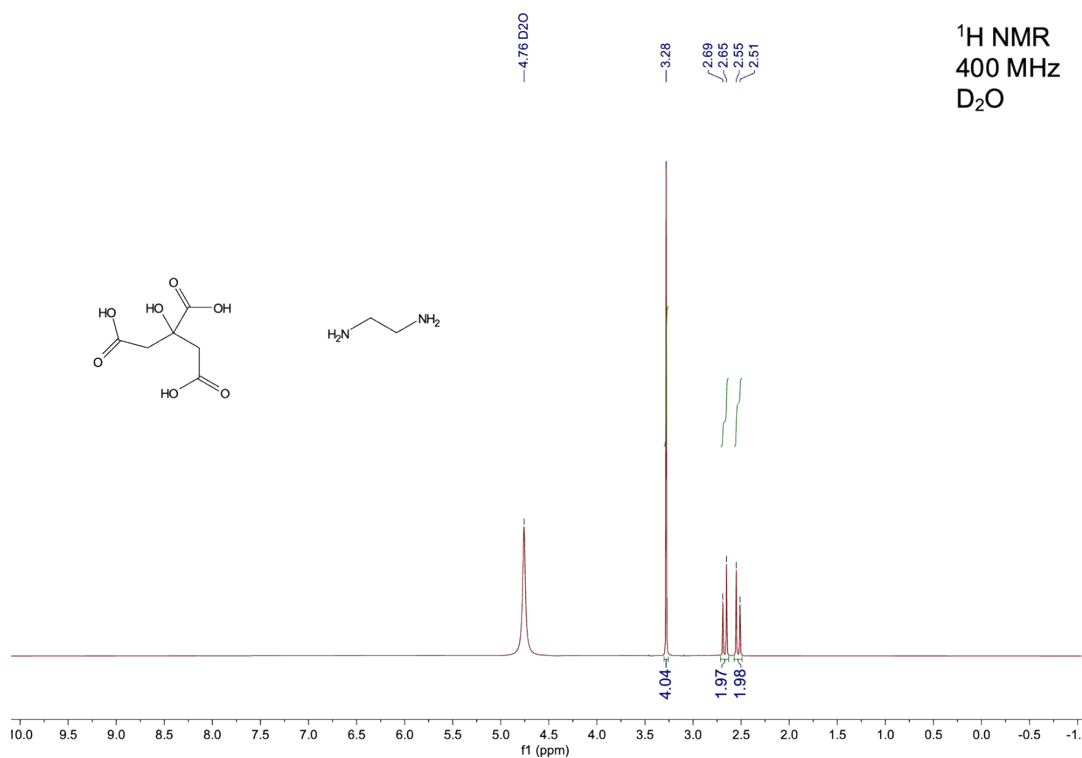
2. NMR of standards used for comparison with studied samples

Mixture of citric acid (CA, 1) and ethylenediamine (EDA, 2) in 1:1 molar ratio at RT

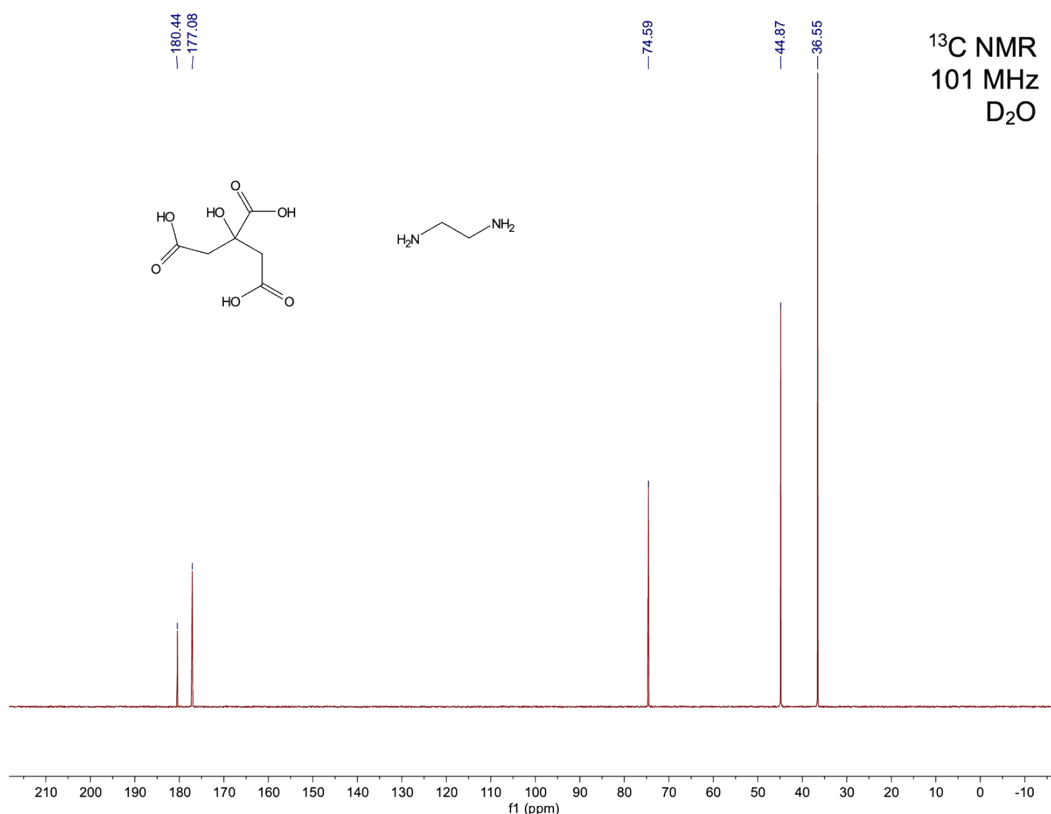
Citric acid (1): ^1H NMR (400 MHz, D_2O) δ 2.67 (d, $J = 15.3$ Hz, 2H), 2.53 (d, $J = 15.3$ Hz, 2H); ^{13}C NMR (101 MHz, D_2O) δ 180.4, 177.1, 74.6, 36.6.

Ethylenediamine (2): ^1H NMR (400 MHz, D_2O) δ 3.28 (s, 4H); ^{13}C NMR (101 MHz, D_2O) δ 44.9.

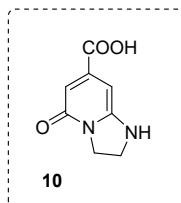
^1H NMR of CA/EDA mixture in 1:1 molar ratio at RT (400 MHz, D_2O)



^{13}C NMR of **CA/EDA mixture** in 1:1 molar ratio at RT (101 MHz, D_2O)



5-oxo-1,2,3,5-tetrahydroimidazo[1,2-*a*]pyridine-7-carboxylic acid (IPCA, 10)



Pure IPCA was prepared by hydrothermal synthesis. The mixture of CA (1.05 g, 5.00 mmol), EDA (0.334 mL, 5.00 mmol, 1.00 eq.) and deionized water (10 mL) was heated to 180 °C for 10 h in a 25 mL-hydrothermal autoclave reactor. After cooling to approx. 50 °C, the solution was transferred to a glass vial and left to crystallize for 3 days at RT. The product was isolated by filtration to obtain pure IPCA (0.325 g, 36%) as a brownish solid.

^1H NMR (400 MHz, D_2O) δ 5.99 (d, $J = 1.2$ Hz), 5.94 (d, $J = 1.3$ Hz), 4.12 (t, $J = 8.9$ Hz, 2H), 3.69 (t, $J = 8.9$ Hz, 2H).

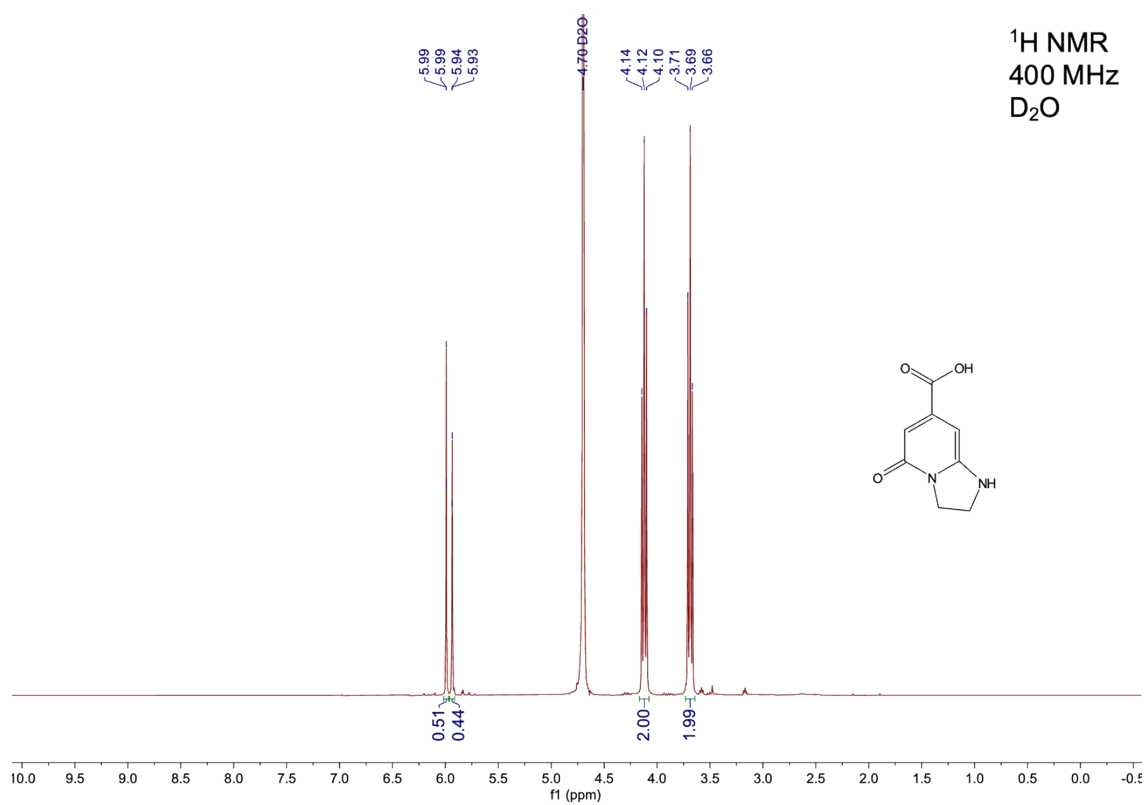
Notes: Due to hydrogen-deuterium exchange, the intensity of signals at 5.99 and 5.94 ppm is lower than expected (the signals completely vanish over time or upon heating).

The low solubility of IPCA in water prevents acquisition of well-resolved ^{13}C NMR spectrum.

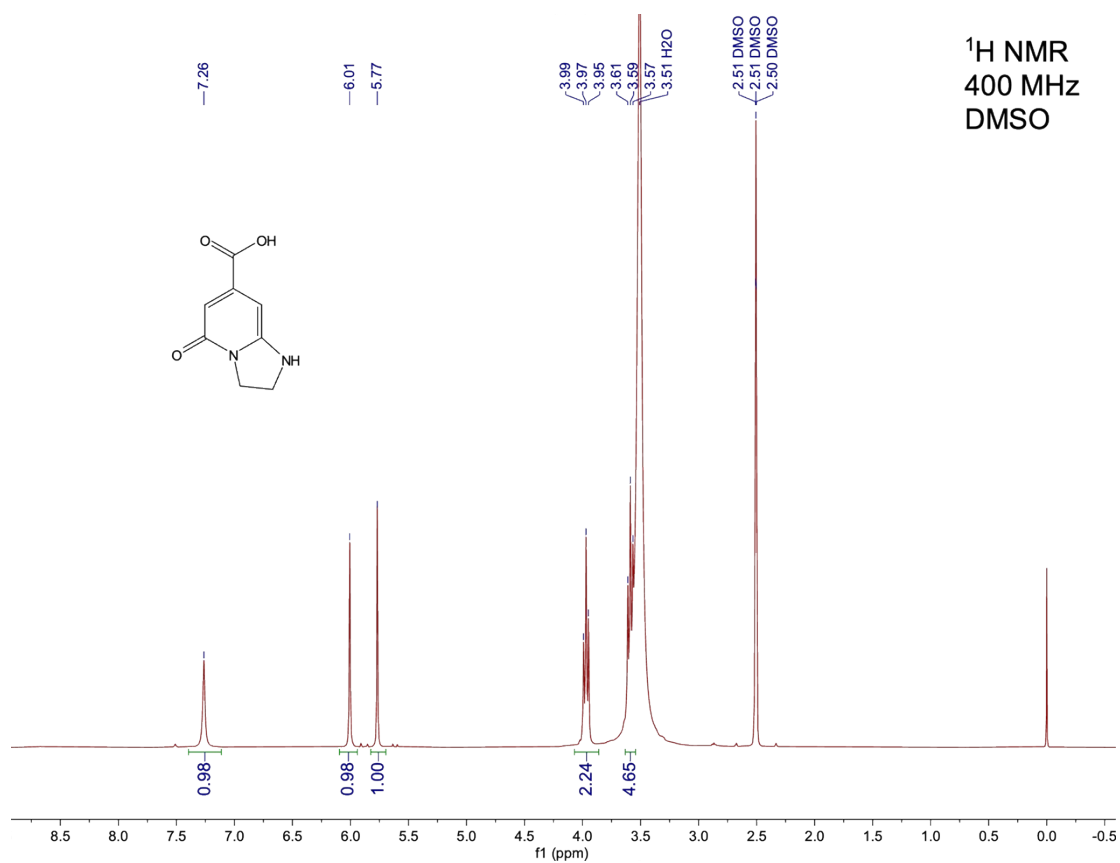
^1H NMR (400 MHz, DMSO) δ 7.26 (s, 1H), 6.01 (s, 1H), 5.77 (s, 1H), 3.97 (t, $J = 8.7$ Hz, 2H), 3.59 (t, $J = 8.9$ Hz, 2H).

^{13}C NMR (101 MHz, DMSO) δ 170.1, 160.8, 153.5, 150.7, 103.9, 82.6, 43.9, 41.9.

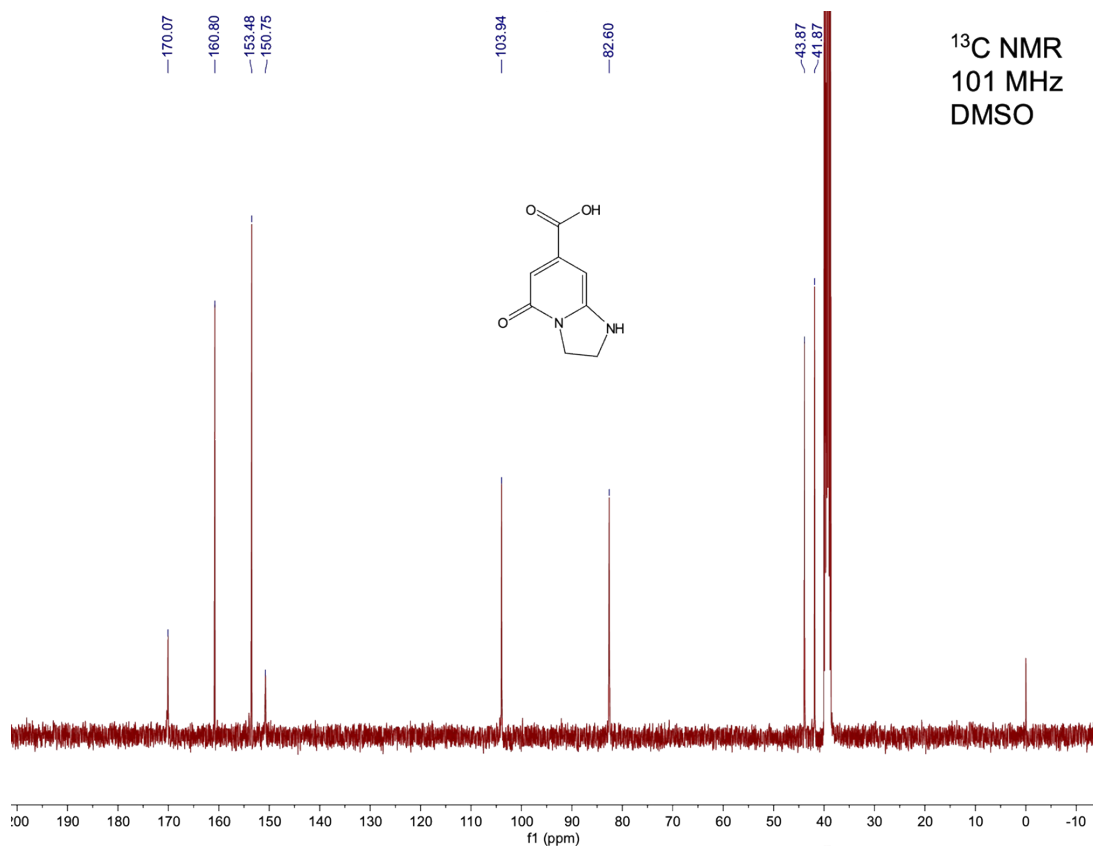
^1H NMR of **IPCA** (400 MHz, D_2O)



^1H NMR of **IPCA** (400 MHz, DMSO)



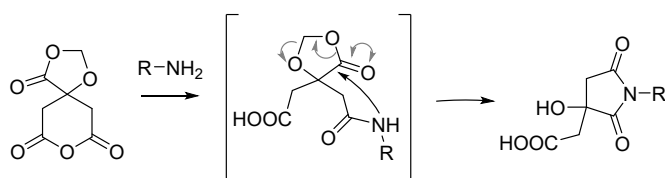
^{13}C NMR of **IPCA** (101 MHz, DMSO)



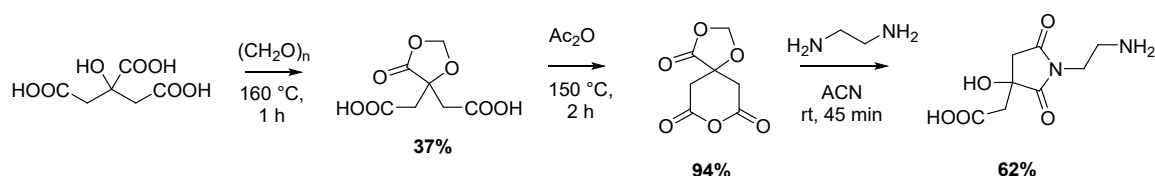
An alternative preparation of compound 12

(2-(1-(2-aminoethyl)-3-hydroxy-2,5-dioxopyrrolidin-3-yl)acetic acid)

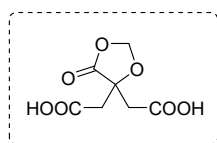
R. Weaver and I. H. Gilbert^[2] showed that upon treatment of anhydromethylenecitric acid anhydride (1,3-dioxo-8-azaspiro[4.5]decane-4,7,9-trione) with *N*-containing nucleophiles, after initial formation of the corresponding amide, 1,3-dioxolan-4-one ring is cleaved to obtain a cyclic amide:



We used this strategy to prepare the intermediate **12**:

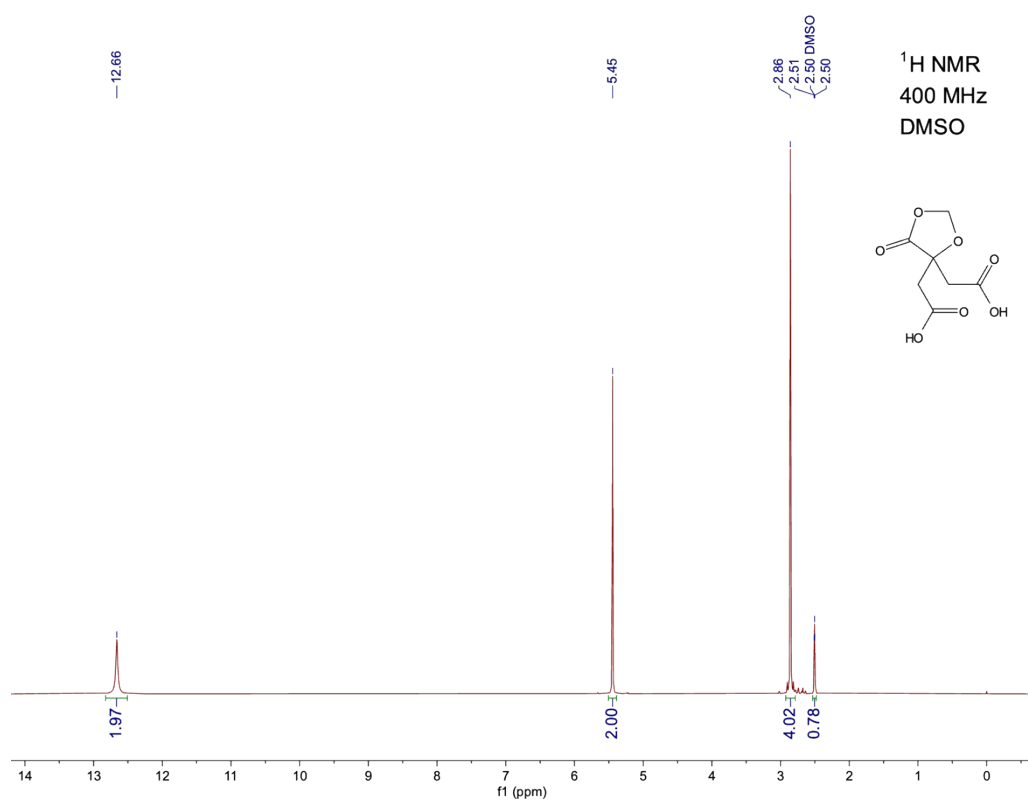


Anhydromethylenecitric acid (2,2'-(5-oxo-1,3-dioxolane-4,4-diyl)diacetic acid)

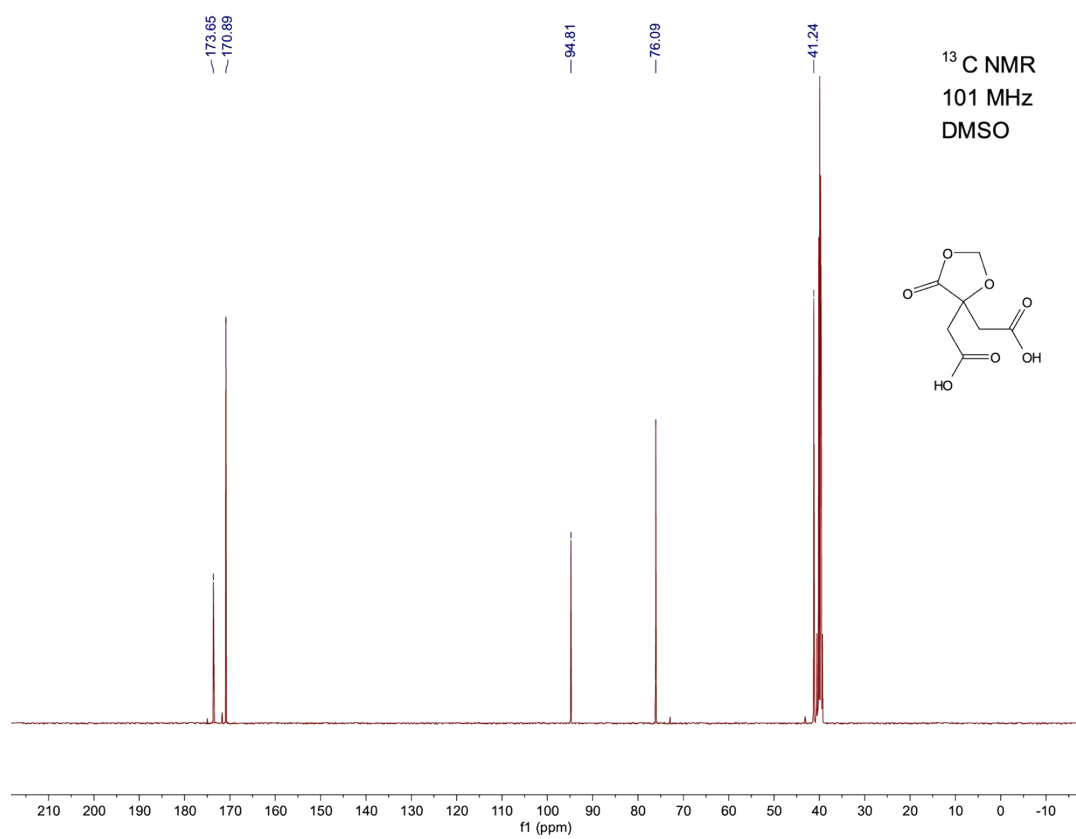


In a well-ventilated fumehood, a round-bottom flask containing anhydrous citric acid (15 g, 78 mmol) and paraformaldehyde (4.7 g, 2 eq.), closed with an argon-filled balloon, was heated to 160 °C for 60 min. During this time, the solid mixture first melts and then again solidifies. After cooling to rt, the mixture was recrystallized from water (approx. 50 mL) to obtain the product as a white solid (5.9 g, 37%). ¹H NMR (400 MHz, DMSO) δ 12.66 (s, 2H), 5.45 (s, 2H), 2.86 (s, 4H); ¹³C NMR (101 MHz, DMSO) δ 173.6, 170.9, 94.8, 76.1, 41.2.

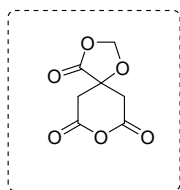
¹H NMR of anhydromethylenecitric acid (400 MHz, DMSO)



¹³C NMR of anhydromethylenecitric acid (101 MHz, DMSO)

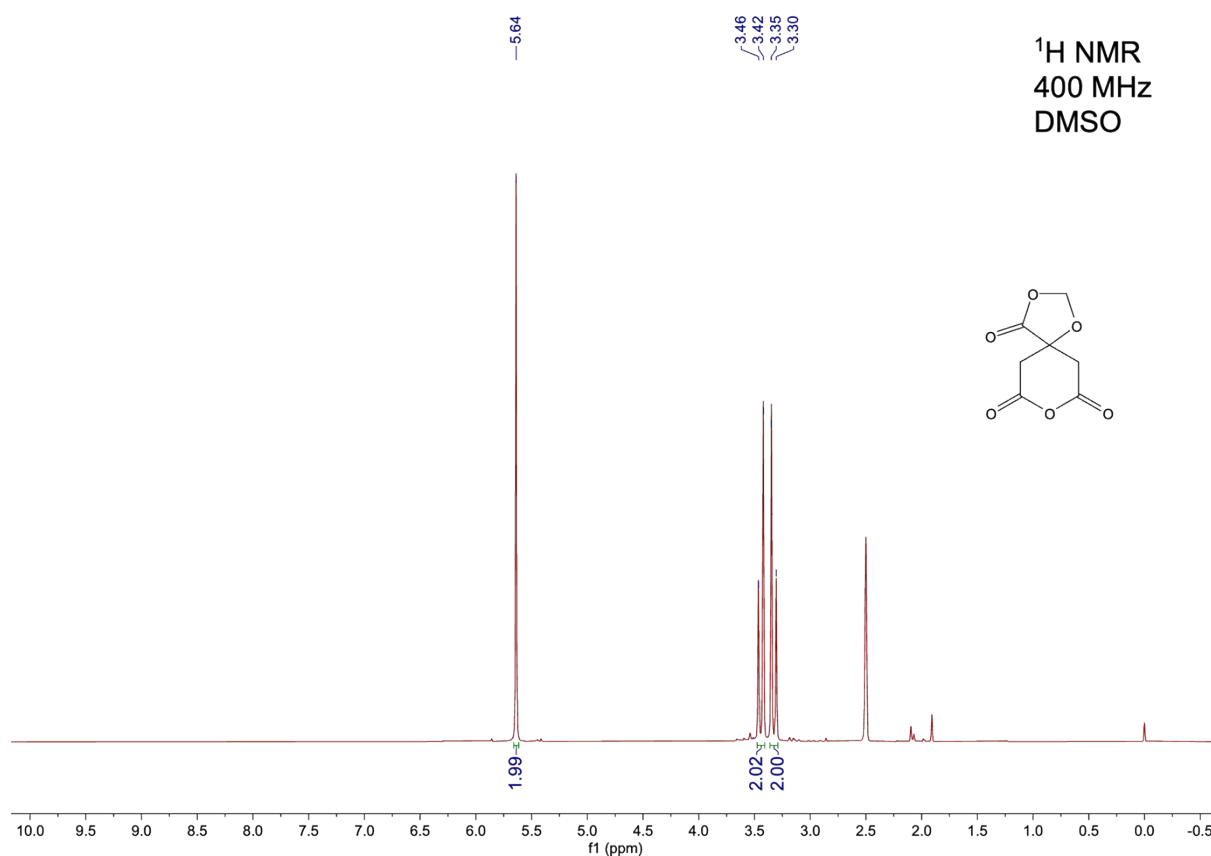


Anhydride of anhydromethylenecitric acid (1,3-dioxaspiro[4.5]decane-4,7,9-trione)

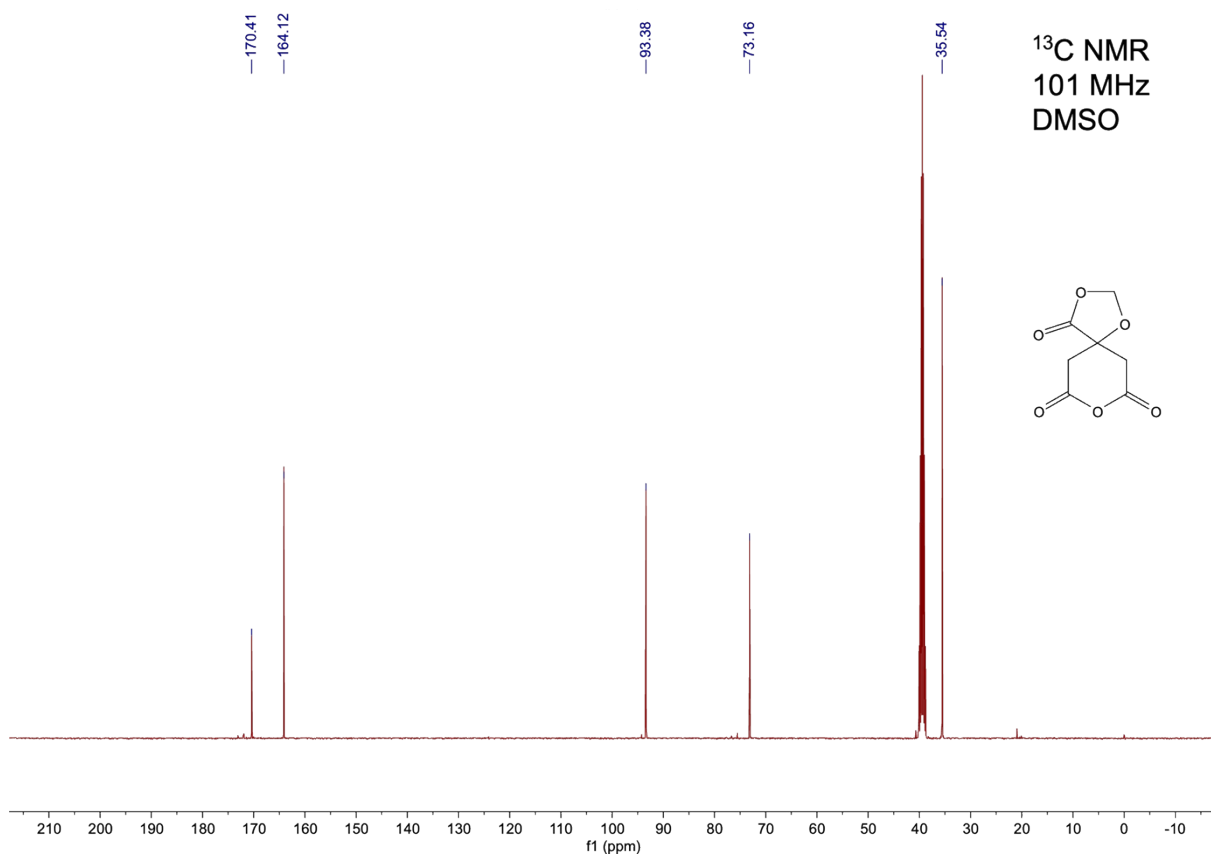


Anhydromethylenecitric acid (2.00 g, 9.8 mmol) and acetic anhydride (5 mL) were heated to 150 °C for 2 h. After this time, the volatiles were removed by vacuum distillation (at 70 °C/100 mbar). The residue (1.72 g, 94 %) was used in the next step without further purification. ^1H NMR (400 MHz, DMSO) δ 5.64 (s, 2H), 3.44 (d, J = 16.8 Hz, 2H), 3.33 (d, J = 17.1 Hz, 2H); ^{13}C NMR (101 MHz, DMSO) δ 170.4, 164.1, 93.4, 73.2, 35.5.

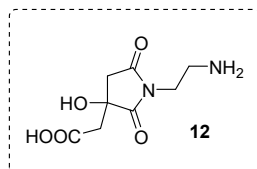
^1H NMR of anhydromethylenecitric acid anhydride (400 MHz, DMSO)



¹³C NMR of anhydromethylenecitric acid anhydride (101 MHz, DMSO)

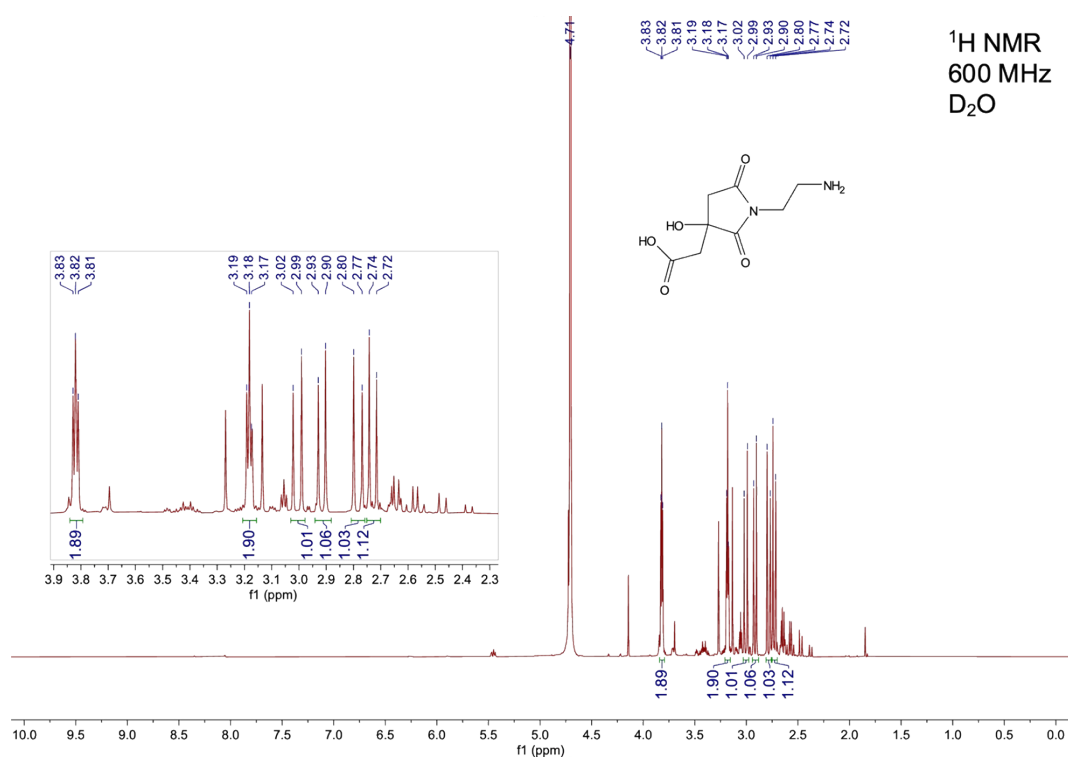


2-(1-(2-aminoethyl)-3-hydroxy-2,5-dioxopyrrolidin-3-yl)acetic acid (12**)**

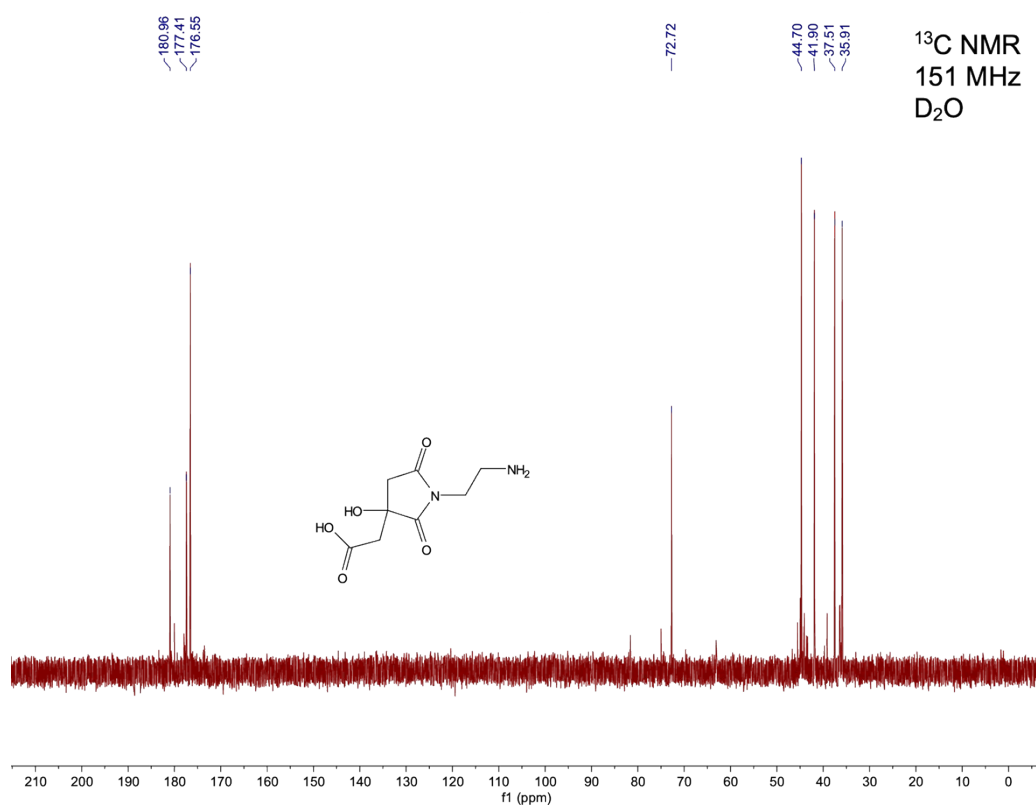


To a solution of anhydromethylenecitric acid anhydride (290 mg, 1.56 mmol) in dry acetonitrile (15 mL) under argon, ethylenediamine (0.104 mL, 93 mg, 1.56 mmol, 1.0 eq.) was added in one portion, which resulted in an immediate formation of greyish precipitate. The reaction mixture was further stirred at rt for 45 min and the precipitate was then isolated by filtration to obtain **12** as a greyish sticky solid, which was purified by size-exclusion chromatography on Sephadex G-10 using deionized water as eluent (column diameter × height: 3×25 cm). The product (yellowish solid, 210 mg, 62%) was isolated by evaporation of water at the rotary evaporator from two colorless fractions, which followed immediately after the first brown band. ¹H NMR (600 MHz, D₂O) δ 3.82 (t, *J* = 5.5 Hz, 2H), 3.18 (t, *J* = 5.4 Hz, 2H), 3.01 (d, *J* = 18.6 Hz, 1H), 2.92 (d, *J* = 16.1 Hz, 1H), 2.78 (d, *J* = 18.6 Hz, 1H), 2.73 (d, *J* = 16.1 Hz, 1H); ¹³C NMR (151 MHz, D₂O) δ 181.0, 177.4, 176.6, 72.7, 44.7, 41.9, 37.5, 35.9; HRMS, calcd for C₈H₁₃N₂O₅ [M + H]⁺, 217.0819; found, 217.0824.

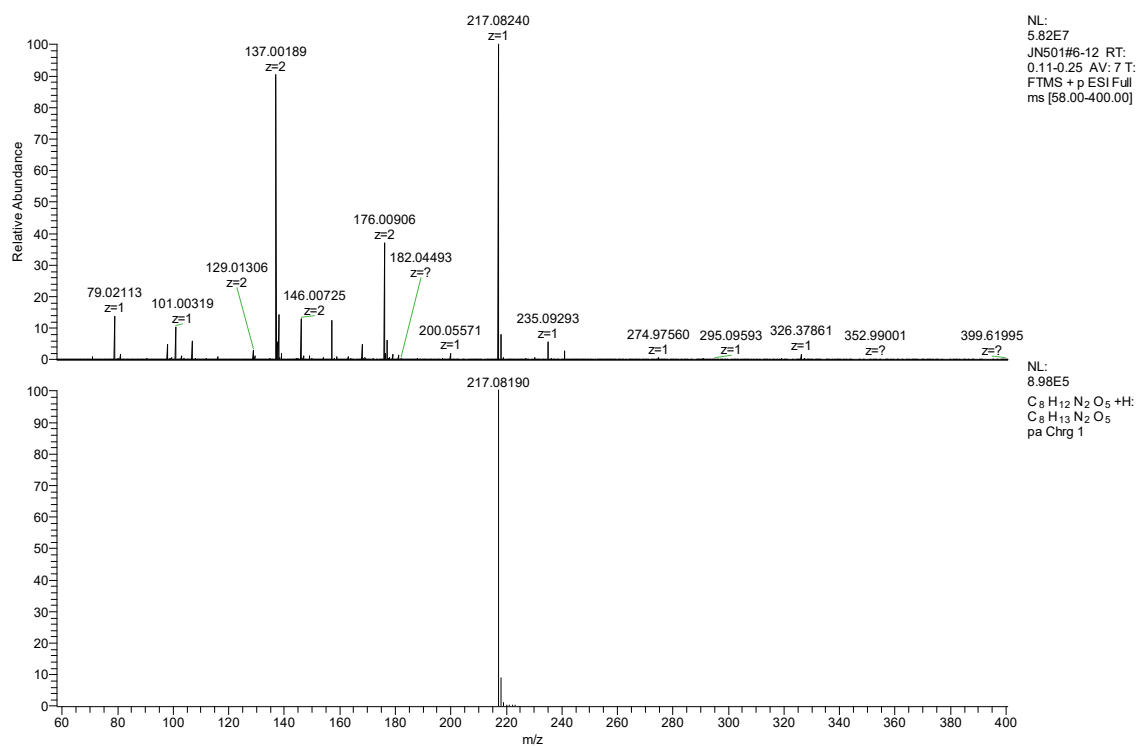
^1H NMR of compound **12** (600 MHz, D_2O)



^{13}C NMR of compound **12** (151 MHz, D_2O)

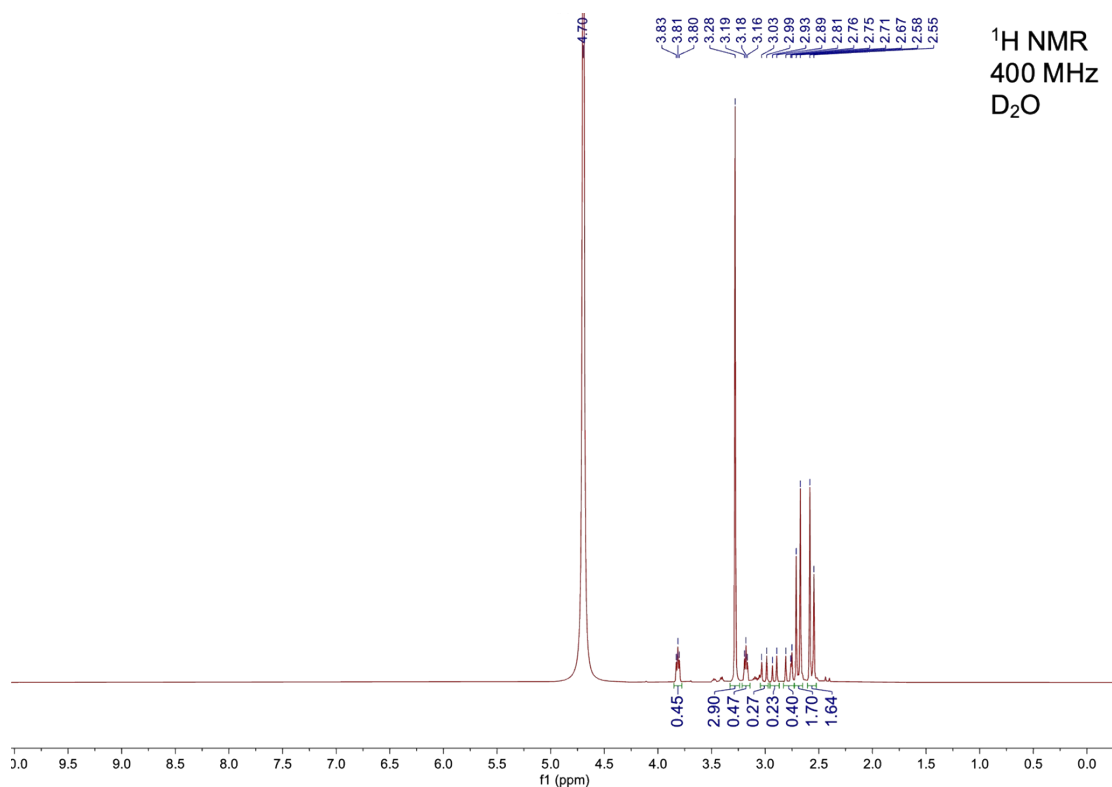


HRMS spectrum of compound **12**: measured (top), calculated (bottom)

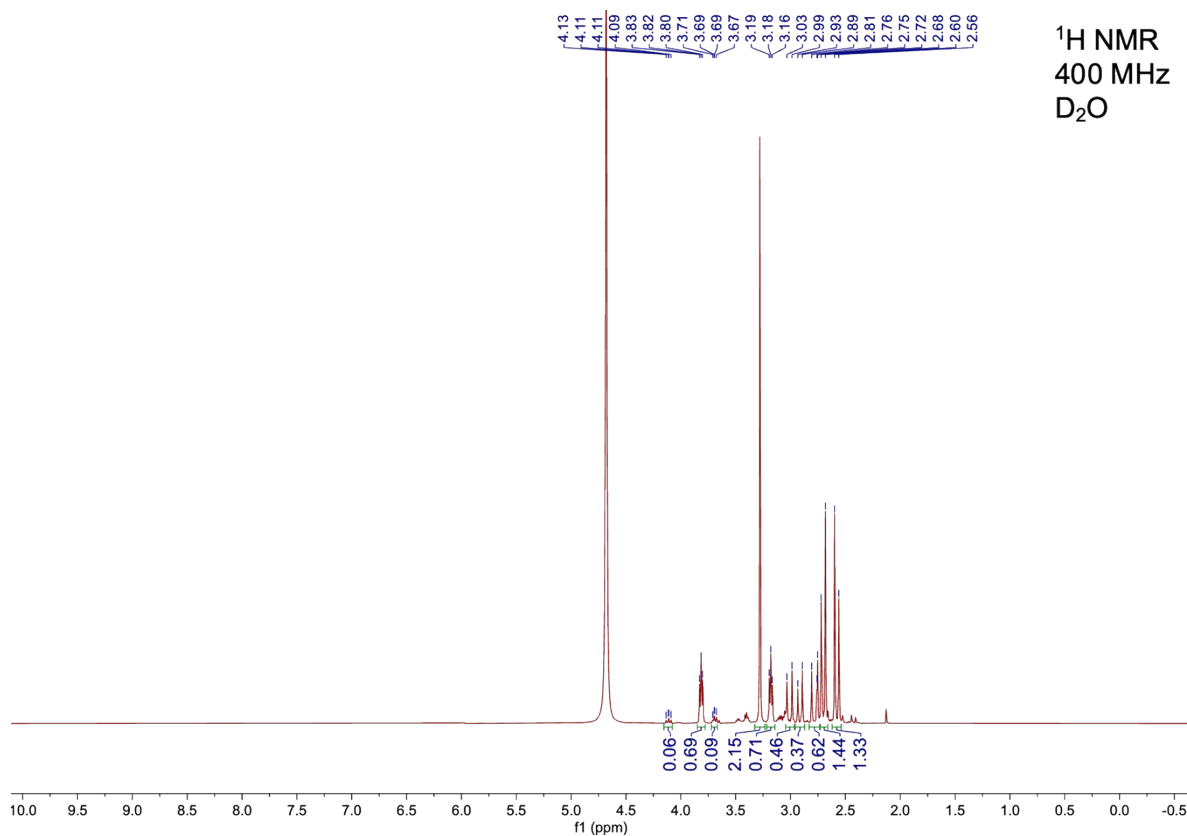


3. ^1H NMR of samples prepared by heating of CA/EDA mixtures for 1 h at 150–240 °C

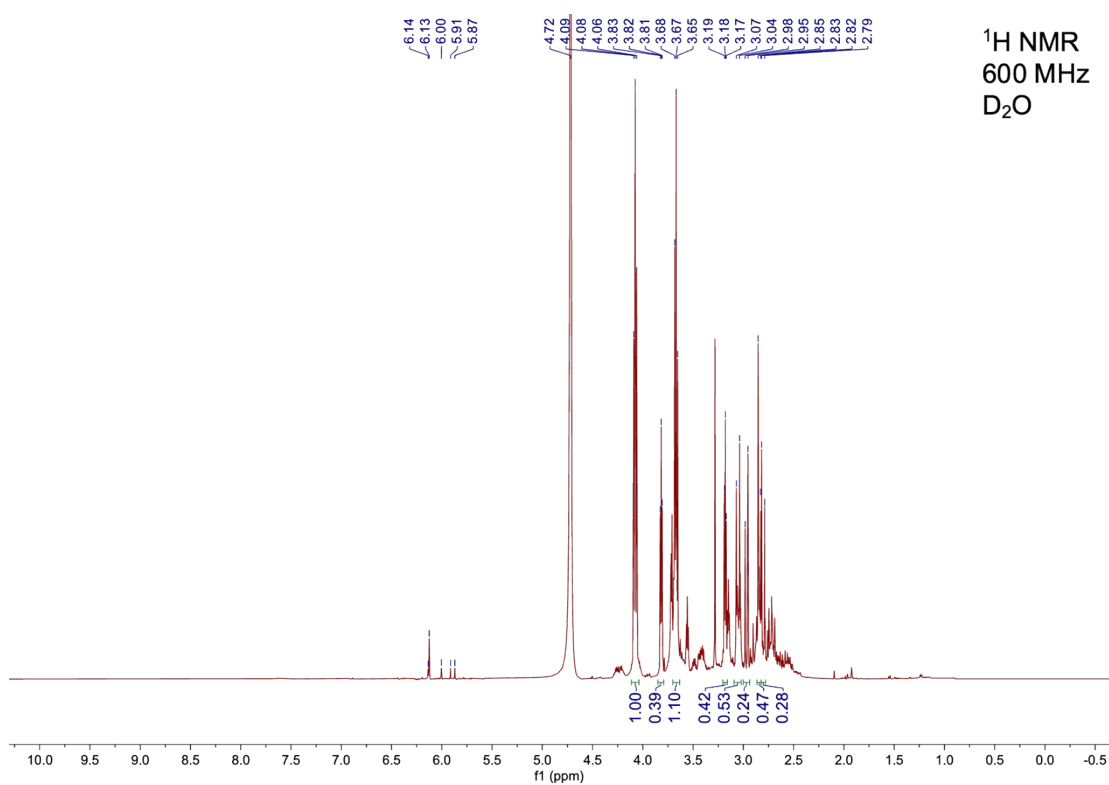
^1H NMR of the sample **CD-150-1** (CA+EDA 1:1 heated at 150 °C for 1h)



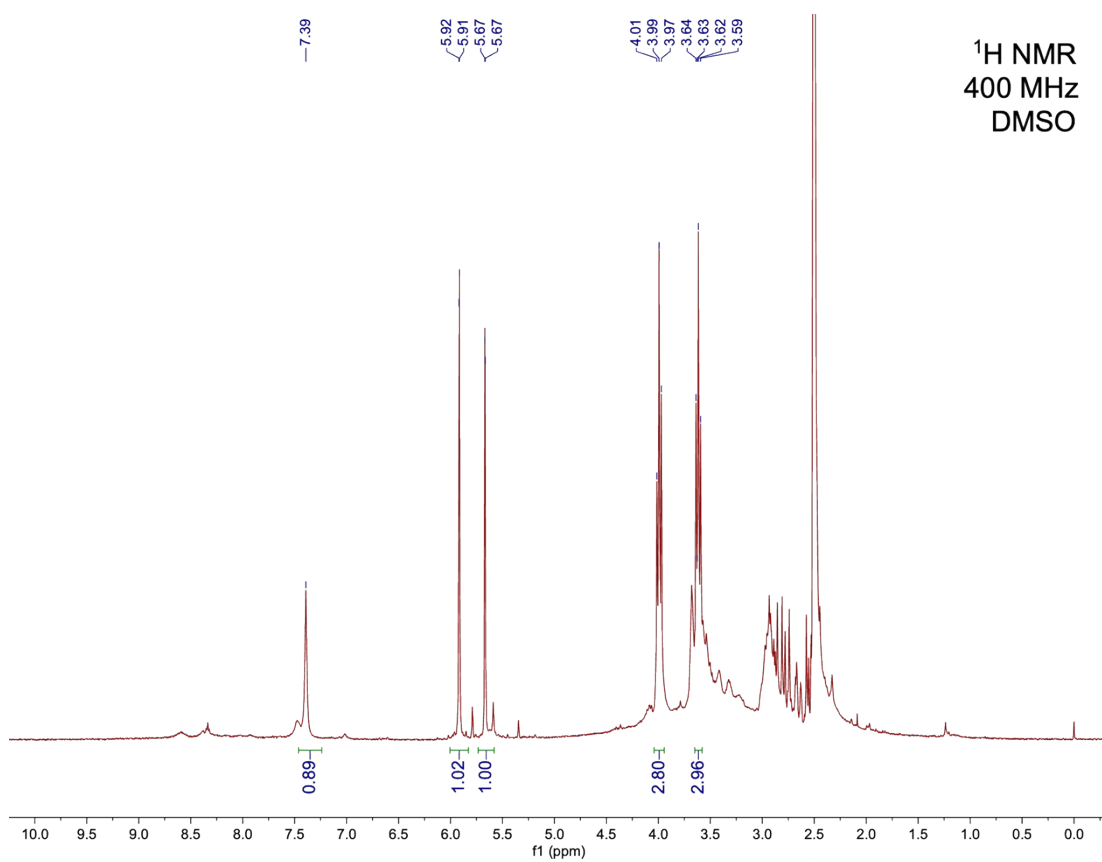
^1H NMR of the sample **CD-180-1** (CA+EDA 1:1 heated at 180°C for 1h)



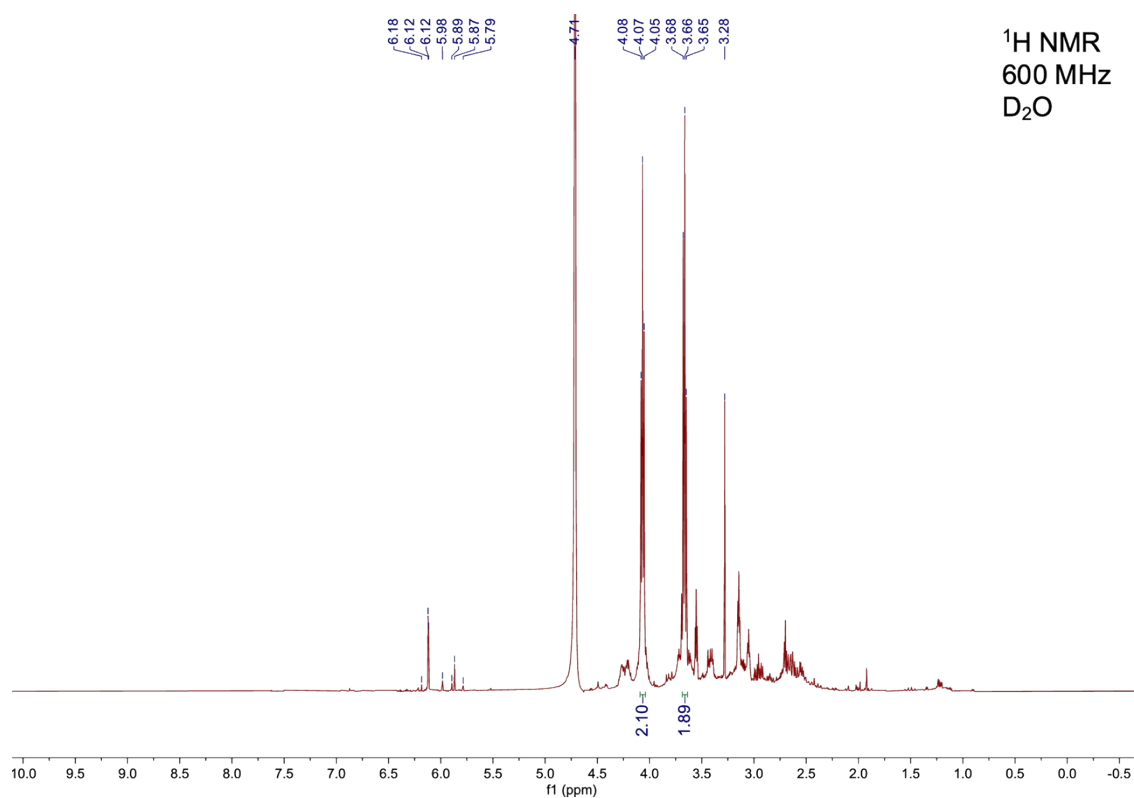
^1H NMR of the sample **CD-210-1** (CA+EDA 1:1 heated at 210°C for 1h)



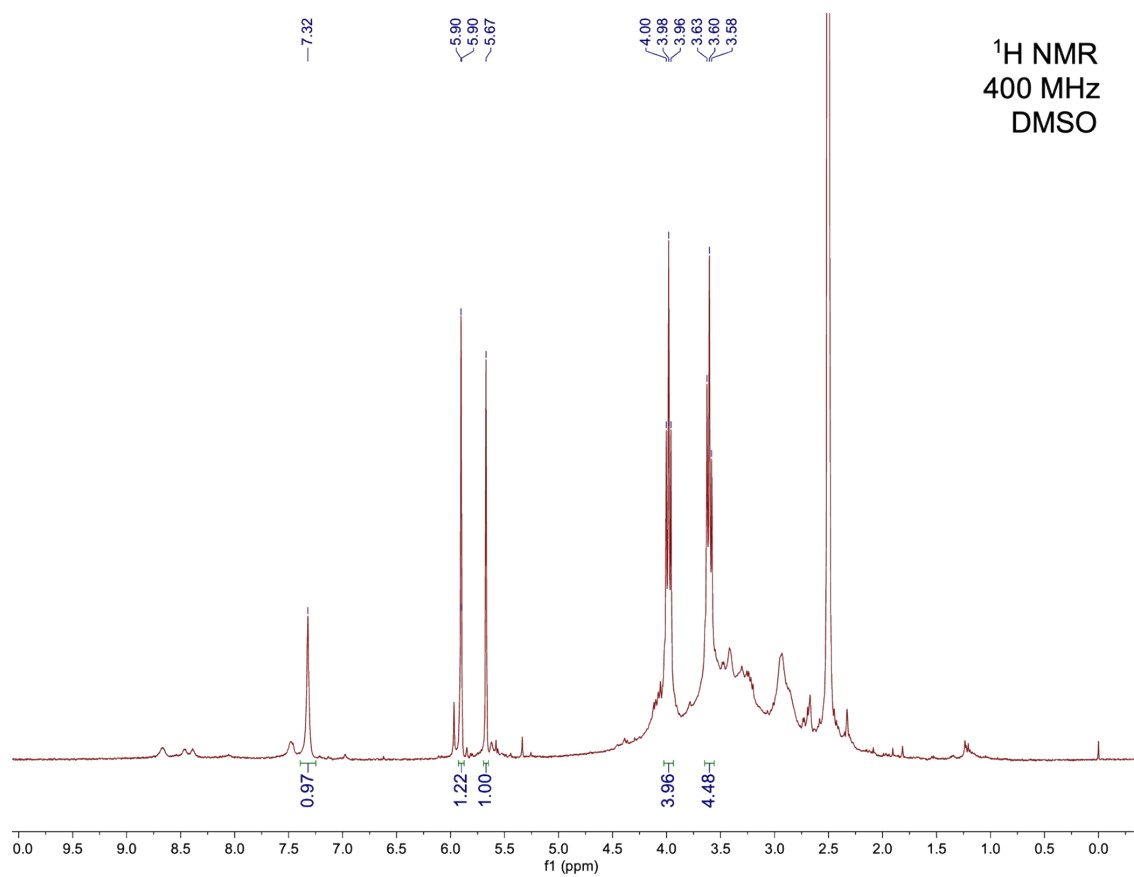
^1H NMR of the sample **CD-210-1** (the same sample, re-measured in DMSO)



^1H NMR of the sample **CD-240-1** (CA+EDA 1:1 heated at 240°C for 1h)



^1H NMR of the sample **CD-240-1** (^1H NMR of the same sample re-measured in DMSO)



4. Detailed NMR analysis of sample CD-150-1 - structure determination of 12

Signals from unreacted EDA: ^1H NMR (600 MHz, D_2O) δ 3.27 (s, 4H); ^{13}C NMR (101 MHz, D_2O) δ 44.8.

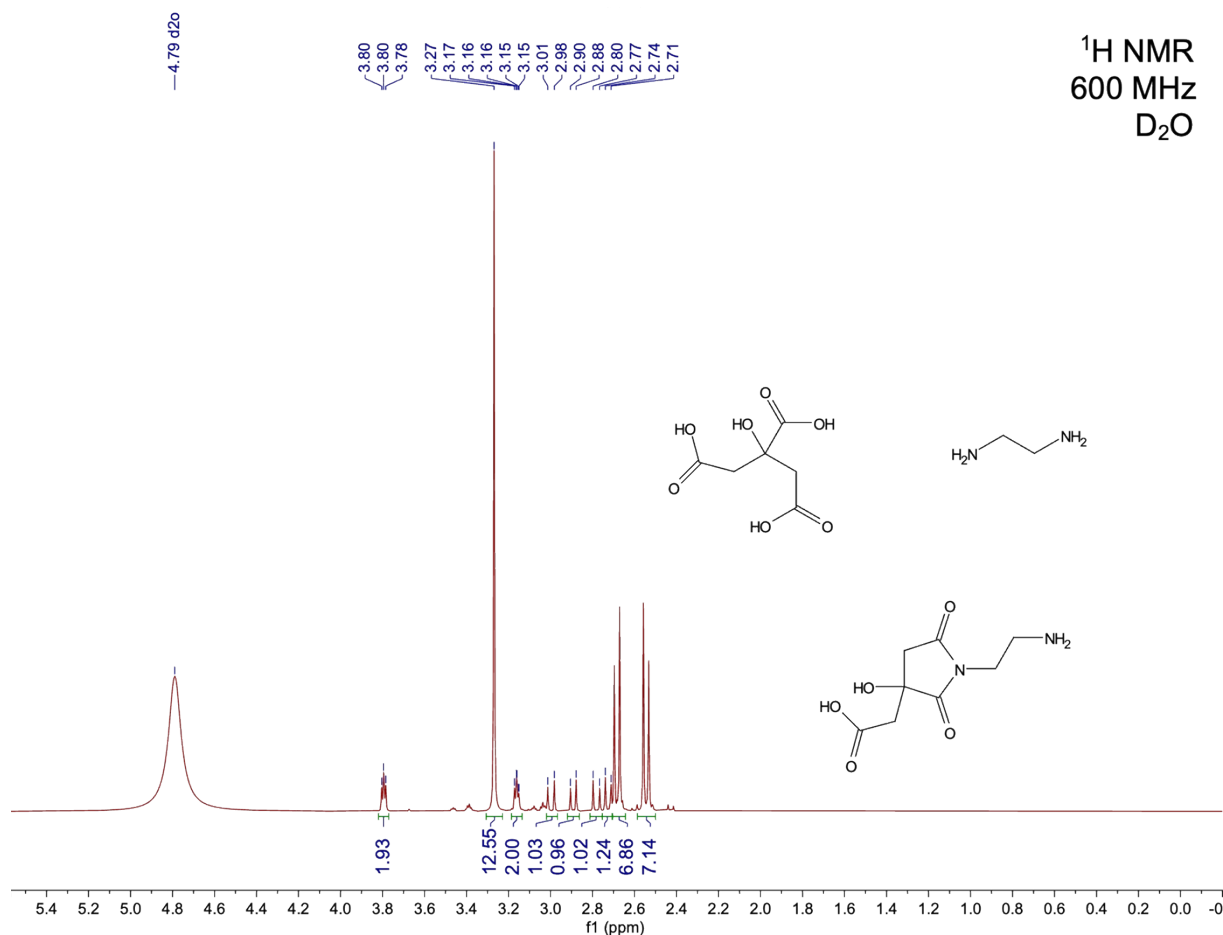
Signals from unreacted CA: ^1H NMR (600 MHz, D_2O) δ 2.68 (d, $J = 15.3$ Hz, 7H), 2.55 (d, $J = 15.3$ Hz, 7H); ^{13}C NMR (101 MHz, D_2O) δ 180.2, 176.7, 74.6, 44.8, 36.7.

Signals from compound **12**:

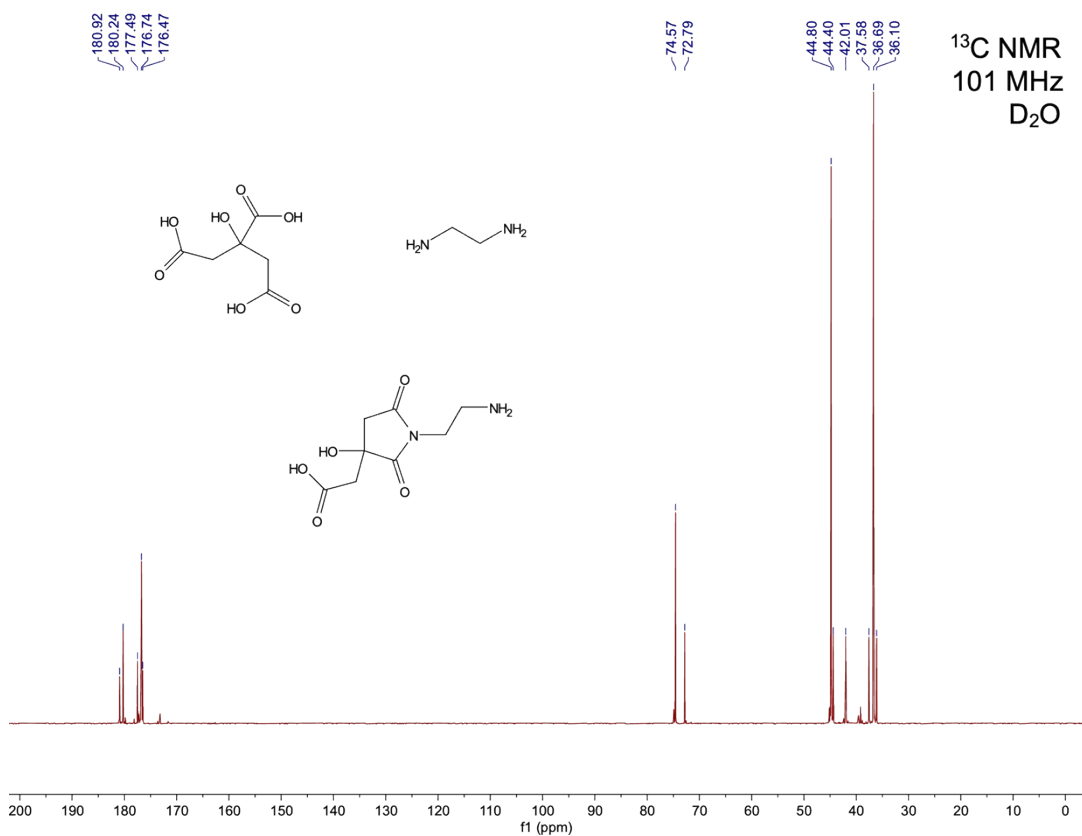
^1H NMR (600 MHz, D_2O) δ 3.79 (t, 2H), 3.16 (t, 2H), 3.00 (d, $J = 18.7$ Hz, 1H), 2.89 (d, $J = 16.1$ Hz, 1H), 2.78 (d, $J = 18.6$ Hz, 1H), 2.72 (d, $J = 16.1$ Hz, 1H).

^{13}C NMR (101 MHz, D_2O) δ 180.9, 177.5, 176.5, 72.8, 44.4, 42.0, 37.6, 36.1.

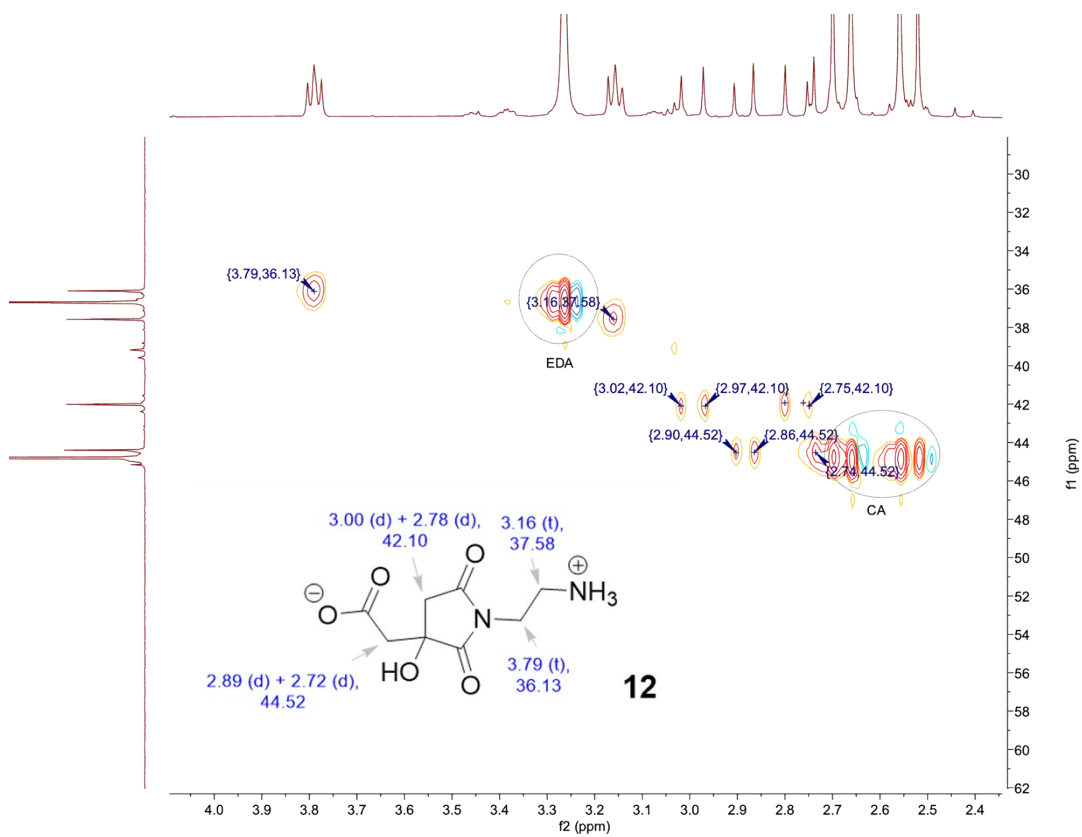
^1H NMR of sample CD-150-1



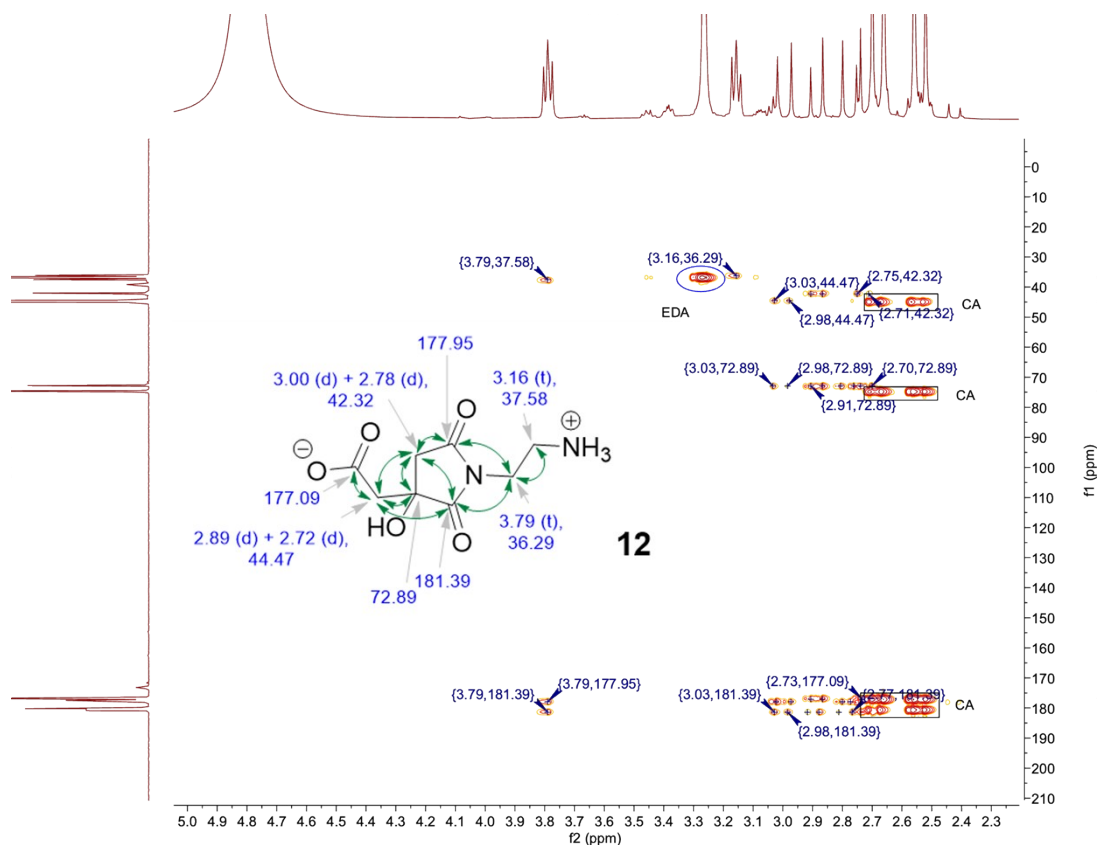
^{13}C NMR of the sample **CD-150-1**



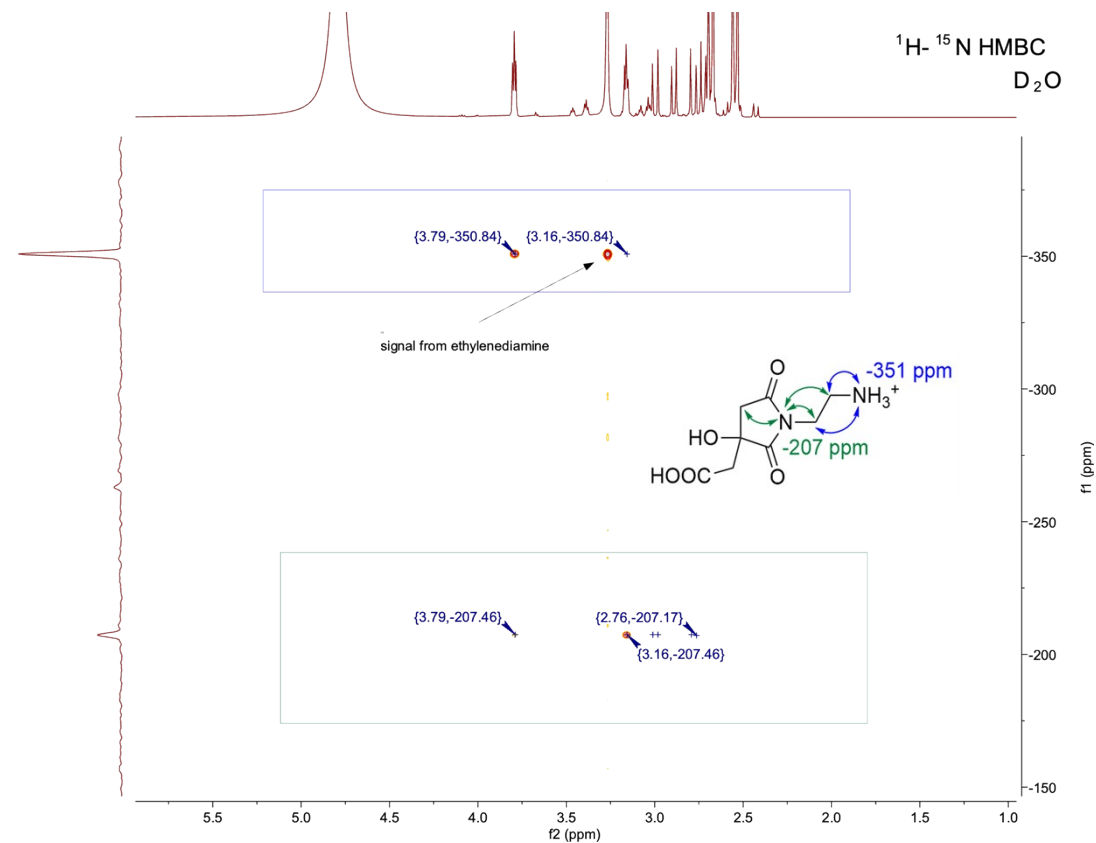
$^1\text{H} - ^{13}\text{C}$ HSQC NMR of the sample **CD-150-1**



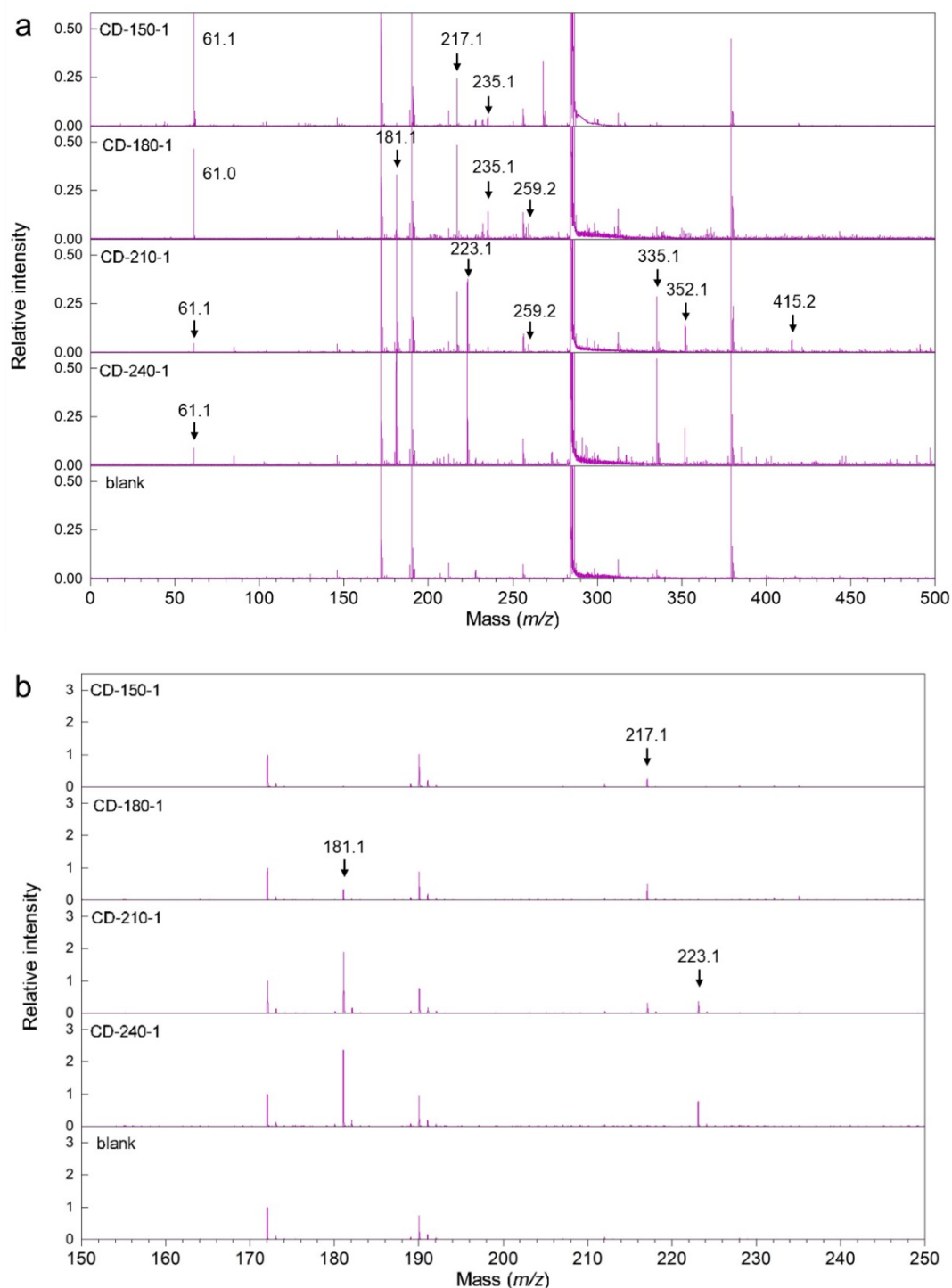
$^1\text{H} - ^{13}\text{C}$ HMBC NMR of the sample **CD-150-1**



$^1\text{H} - ^{15}\text{N}$ HMBC NMR of sample **CD-150-1**



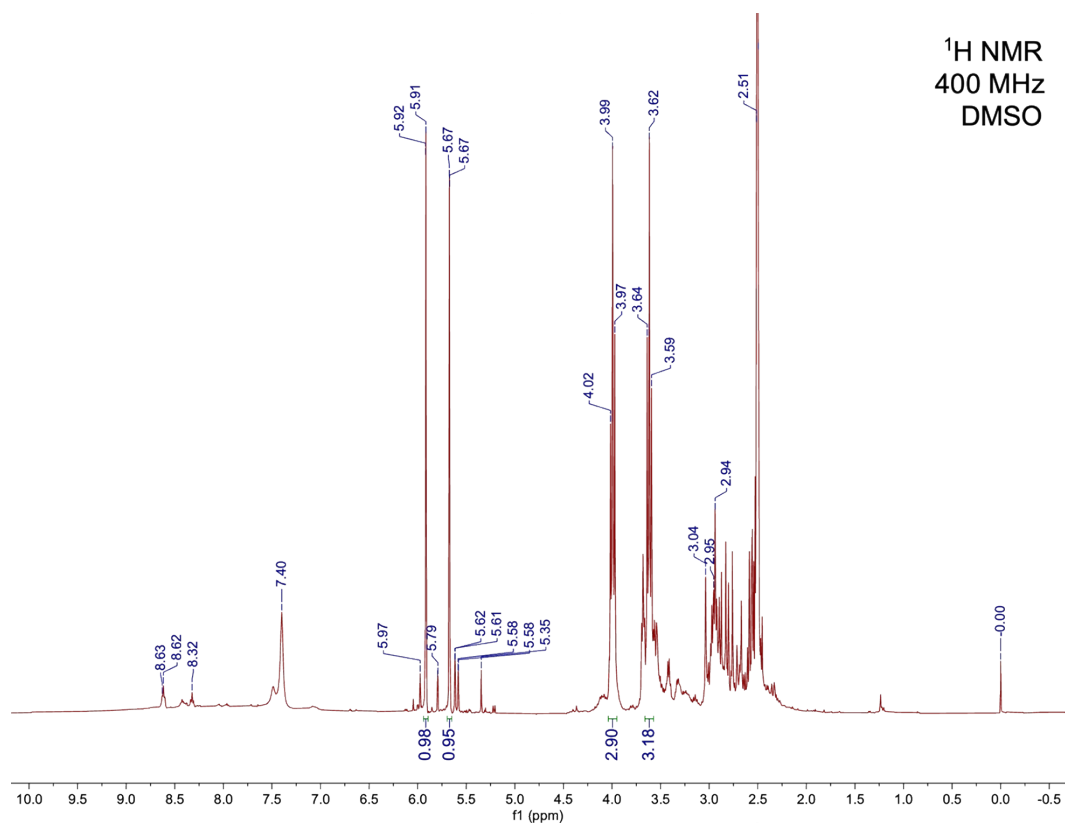
5. MALDI-TOF analysis of samples CD-150-1, CD-180-1, CD-210-1 and CD-240-1



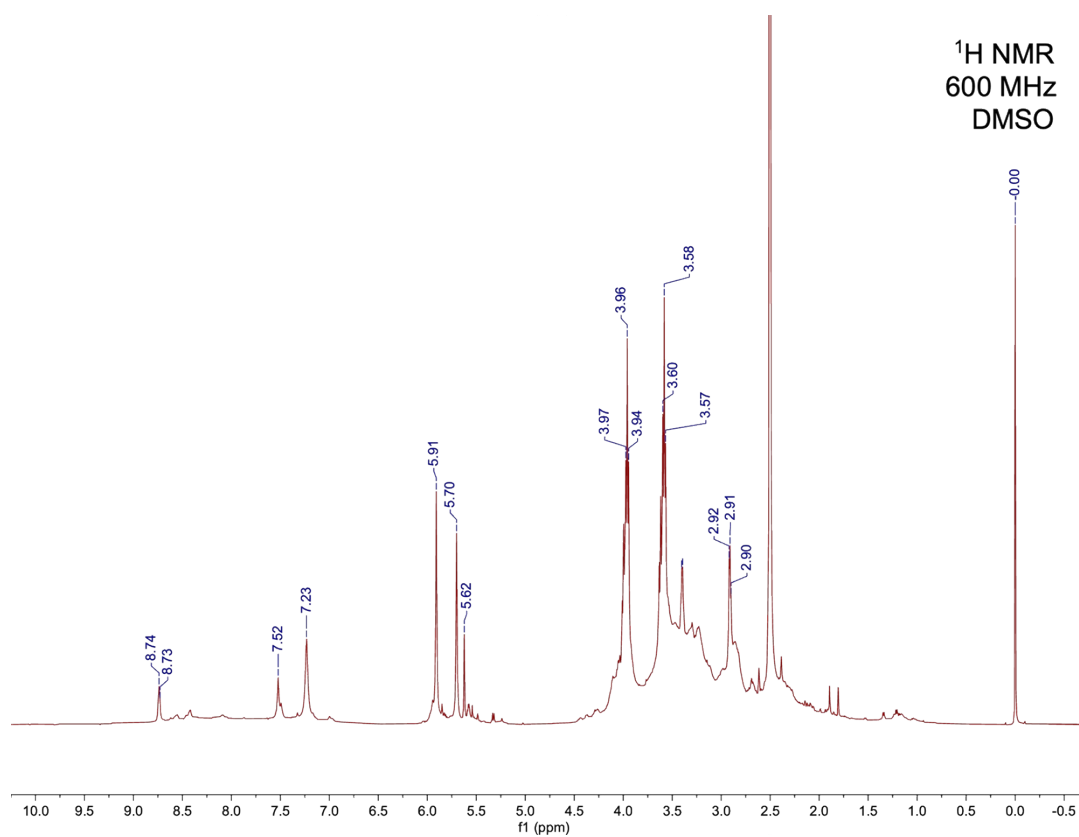
MALDI-TOF analysis of samples heated to 150–240 °C for 1 h. a) full view (m/z interval: 0–500), b) close-up (m/z interval: 150–250) showing the presence of intermediate **12** (m/z 217.1), which gradually converts to IPCA (**10**, m/z 181.1) and its amide **11** (m/z 223.1). The measured MS spectra were normalized so that the matrix signal at $m/z = 172.0$ was set to 1.

6. ^1H NMR of samples prepared by heating of CA/EDA mixtures (10 h/150–240 °C)

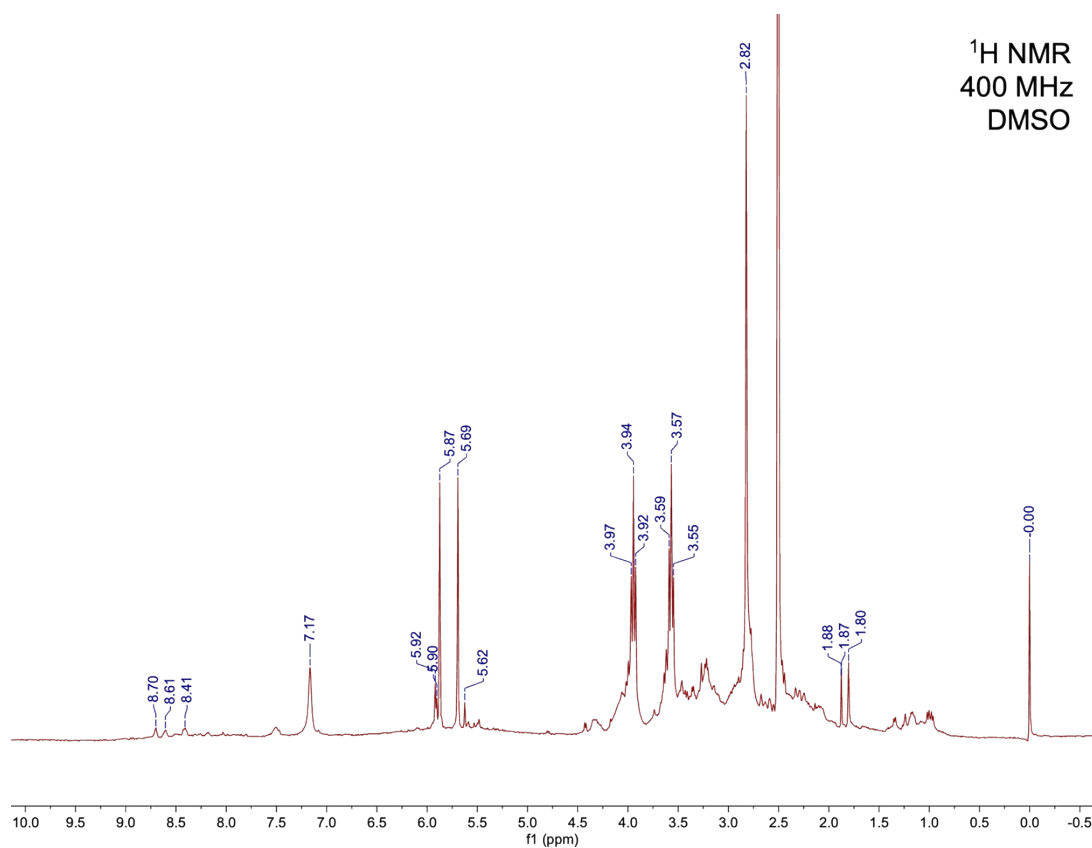
^1H NMR of the sample **CD-150-10** (CA+EDA 1:1 heated at 150 °C for 10h)



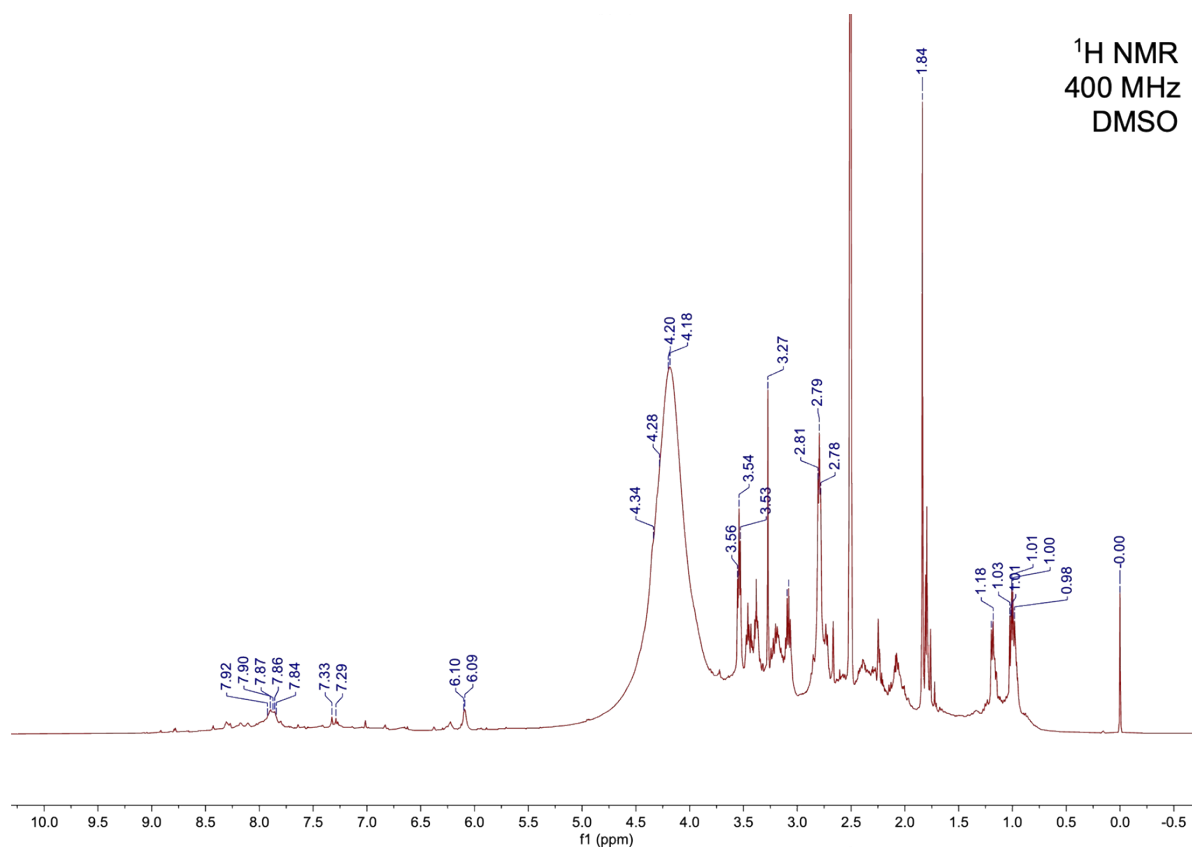
^1H NMR of the sample **CD-180-10** (CA+EDA 1:1 heated at 180°C for 10h)



^1H NMR of the sample **CD-210-10** (CA+EDA 1:1 heated at 210°C for 10h)



^1H NMR of the sample **CD-240-10** (CA+EDA 1:1 heated at 240°C for 10h)



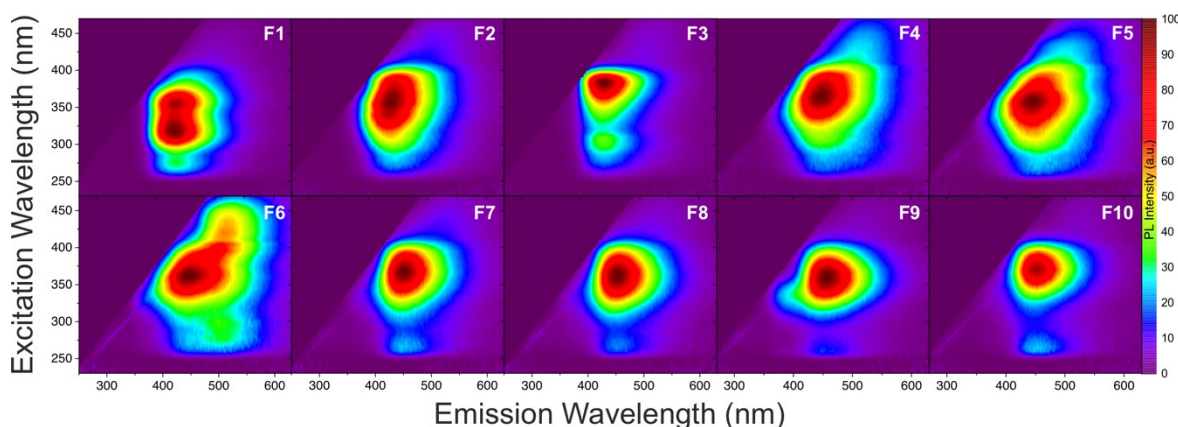
7. Detailed analysis of the sample CD-180-10

The sample CD-180-10 (a part of mother liquor obtained after crystallization of IPCA evaporated to dryness, 350 mg) was dissolved in water (4 mL) and filtrated via a syringe filter to remove any insoluble particles. This solution was transferred on the top of Sephadex G-10 size-exclusion chromatography column and eluted using deionized water. The first collected fraction had intense brownish color. Together, 9 fractions were collected (each with volume approx. 20 mL). After separation, the column exhibited weak blue fluorescence and was therefore purged with deionized water (approx. 2 L). Concentrating this solution (denoted as fraction 10) resulted in isolation of compound **11**.

The isolated fractions were analyzed by UV-vis spectroscopy (Figure 4d). Highly concentrated fractions 1-3 were diluted with deionized water (20x) to obtain reasonable absorbance, all other fractions were analyzed as eluted from the chromatography column.

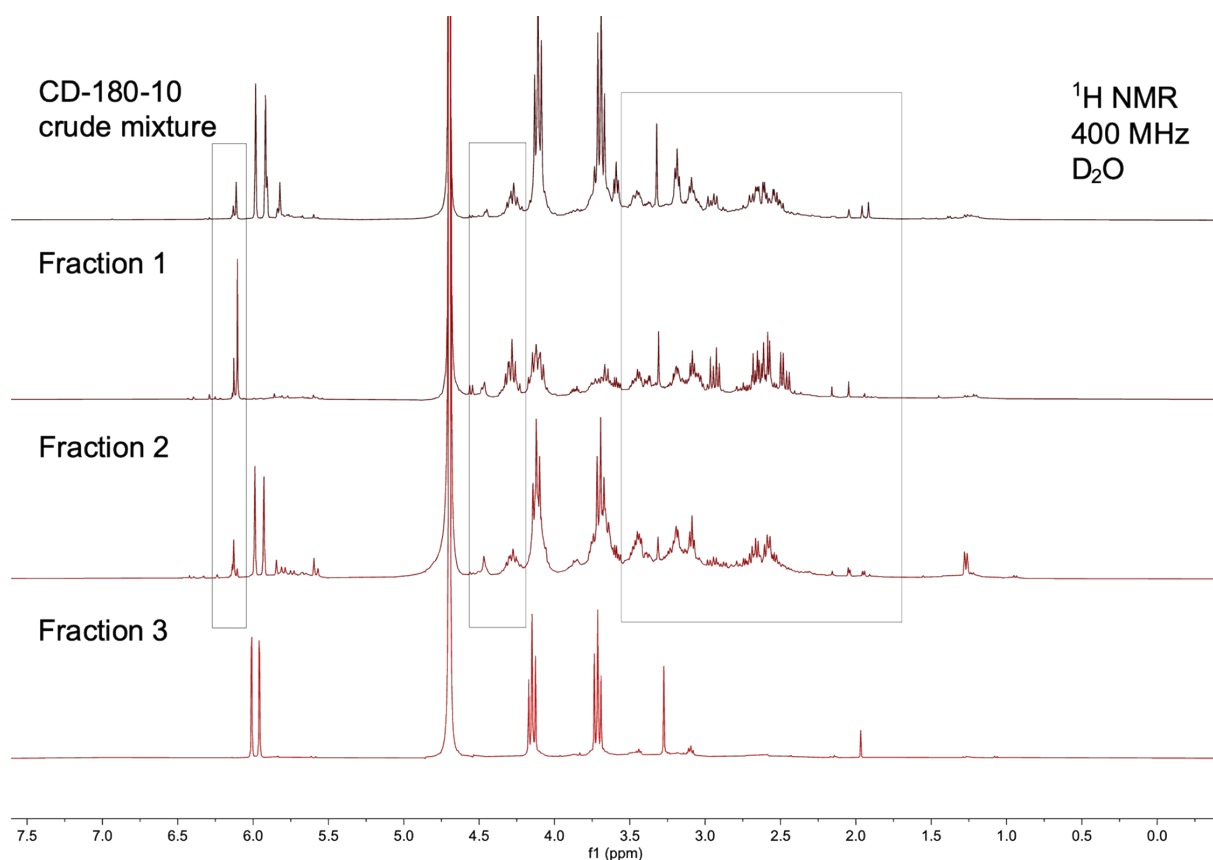
Photoluminescence spectra of the fractions 1-10 were recorded in diluted solutions (so that the maximum absorbance in the 300-400 nm region did not exceed 0.1).

PL Excitation–emission color maps of individual extracted fractions are presented below.



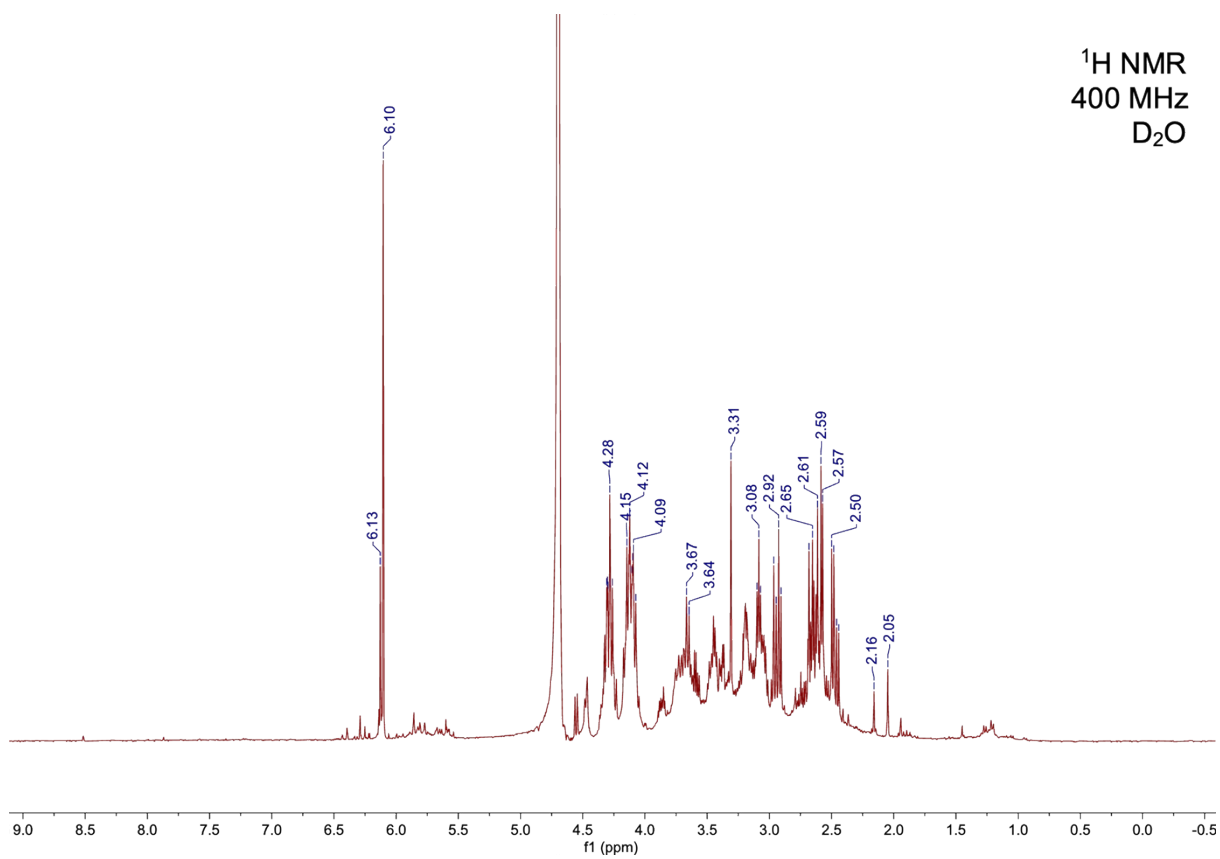
Then, the fractions were concentrated to approx. 2 mL and analyzed by TLC chromatography. The samples were spotted to TLC plate (SiO₂) as solutions in water, TLC plate was dried with a heating gun and subsequently developed using ethyl-acetate/methanol (1:1, v/v) as the eluent. After evaporation to dryness on the rotary evaporator, the most abundant fractions (1-3) together with fraction 10 were analyzed by ¹H NMR spectroscopy.

^1H NMR spectra of the most abundant fractions (1-3) were in good agreement with ^1H NMR of the full (non-separated) crude mixture, pointing that ^1H NMR spectroscopy is a useful method for rapid identification of the most dominant intermediates during the formation of CDs without the need for time-consuming purification.

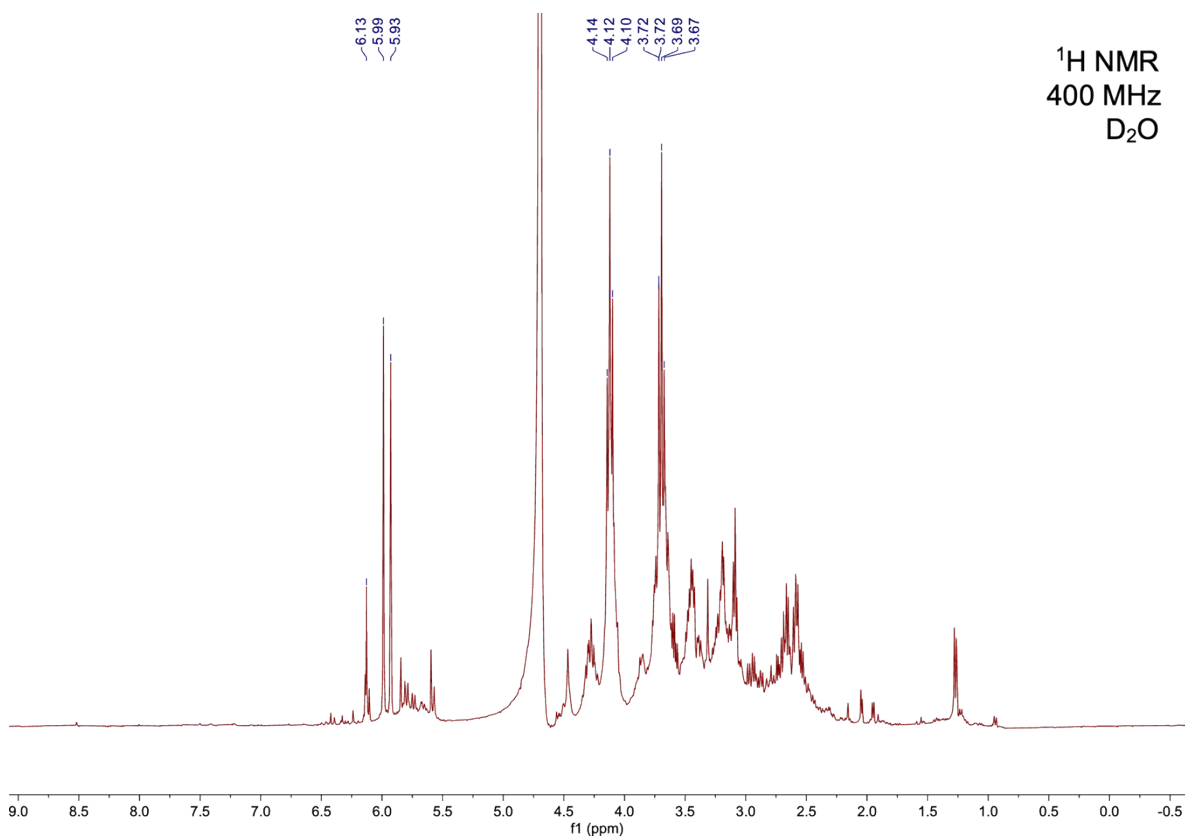


Detailed results from NMR measurements are summarized below.

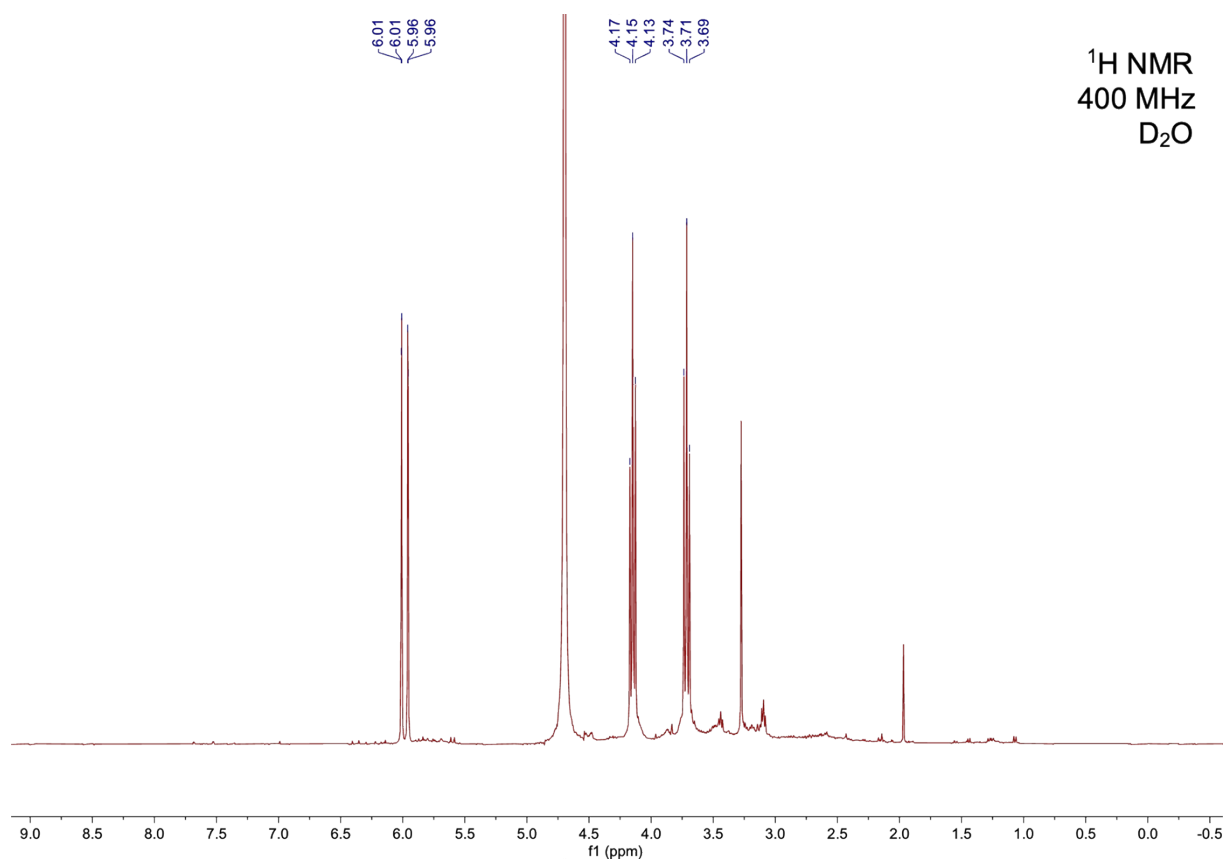
^1H NMR spectroscopy of fraction 1



^1H NMR spectroscopy of fraction 2



^1H NMR spectroscopy of fraction 3



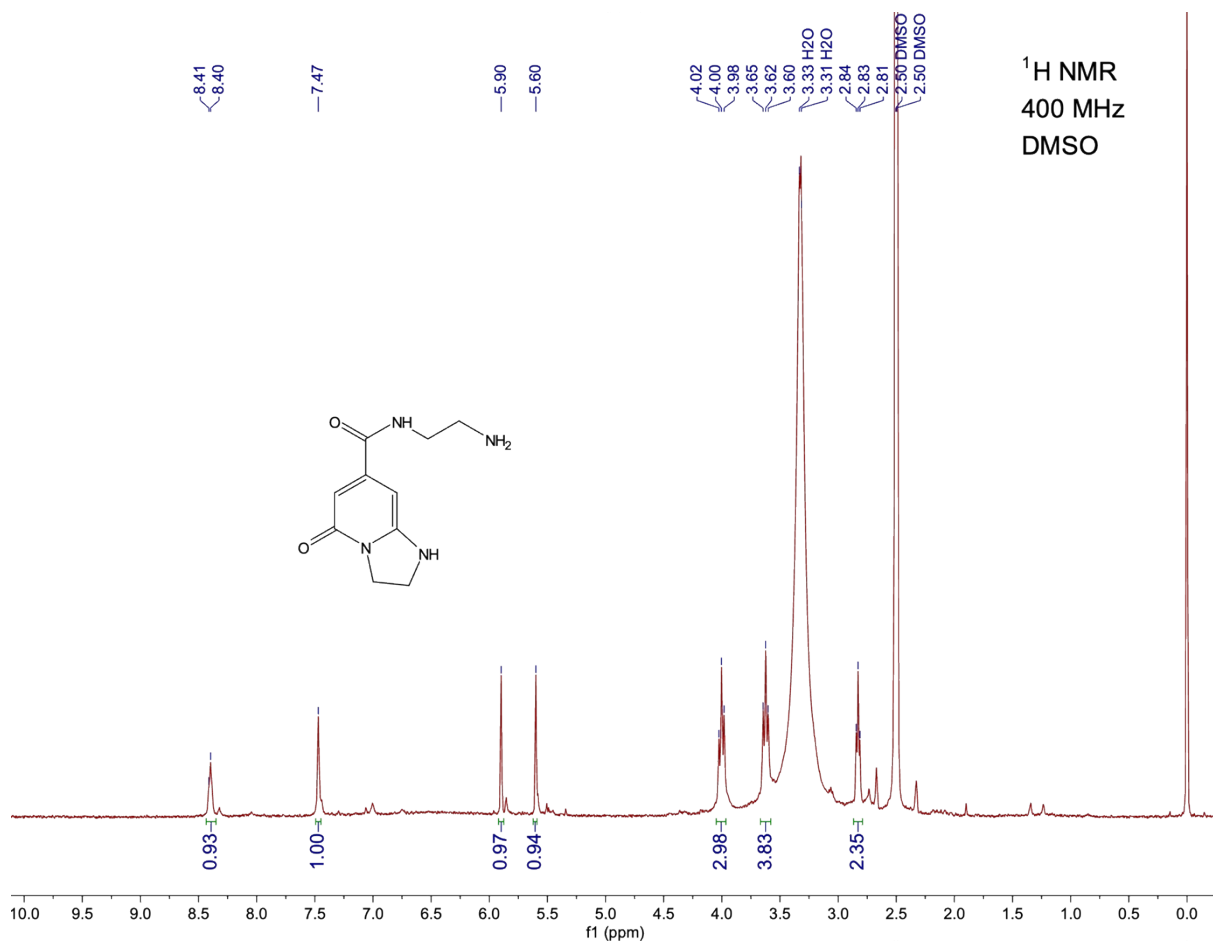
NMR analysis of fraction 10 – isolation of a new fluorophore (11)

Compound **11** was obtained only in trace amounts, therefore acquisition of well-resolved ^{13}C NMR spectra was not possible. Some of the ^{13}C NMR resonances were observed in HSQC and HMBC spectra:

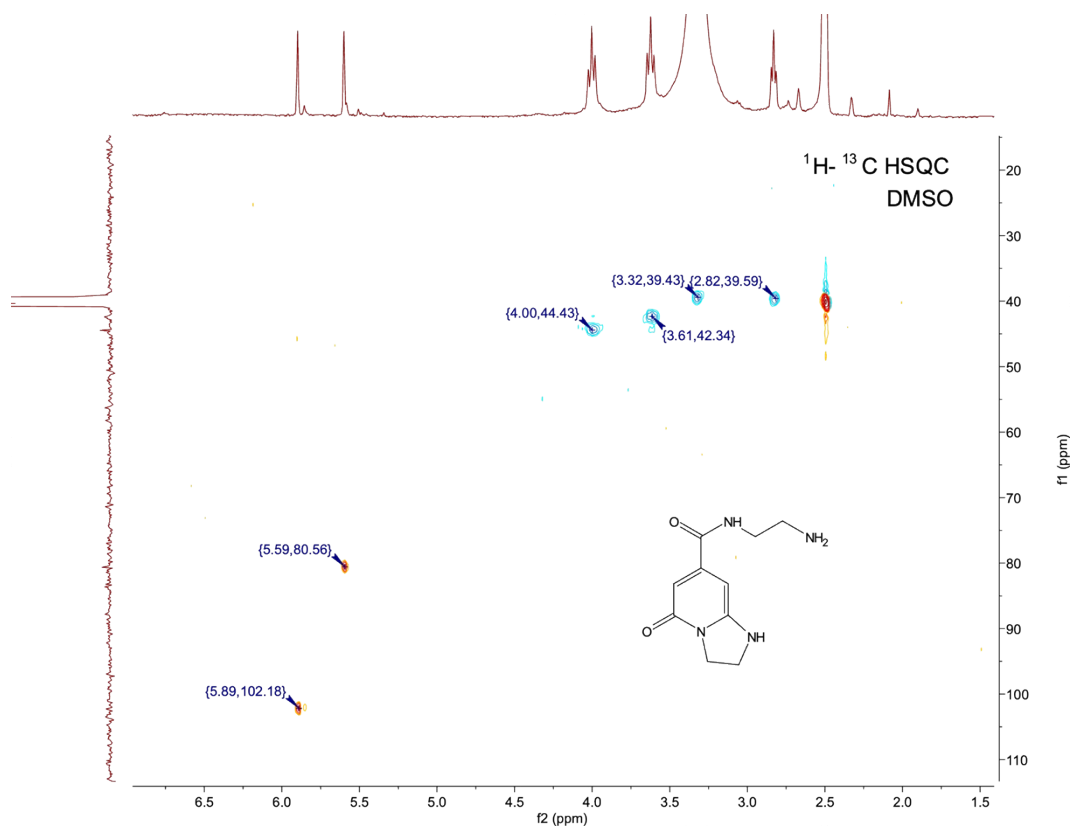
^1H NMR (400 MHz, DMSO) δ 8.40 (s, 1H), 7.47 (s, 1H), 5.90 (s, 1H), 5.60 (s, 1H), 4.00 (t, $J = 8.8$ Hz, 2H), 3.62 (t, $J = 8.9$ Hz, 2H), 3.32 (t, 2H) 2.83 (t, $J = 6.4$ Hz, 2H).

^{13}C NMR (400 MHz, DMSO) δ 166.6, 154.5, 102.2, 80.6, 44.4, 42.3, 39.6, 39.4.

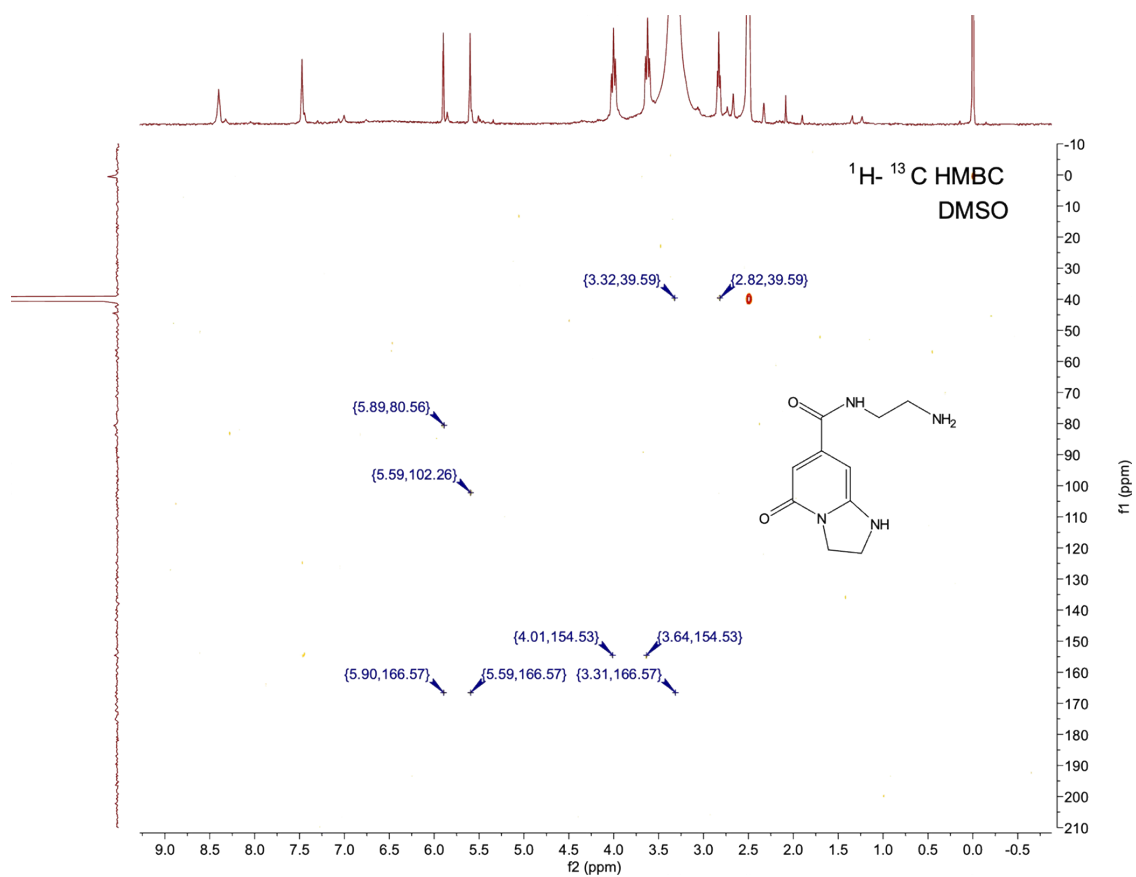
^1H NMR of fraction 10 (compound **11)**



$^1\text{H} - ^{13}\text{C}$ HSQC NMR of fraction 10 (compound **11**)



$^1\text{H} - ^{13}\text{C}$ HMBC NMR of fraction 10 (compound **11**)



8. The transformation of **12** to IPCA

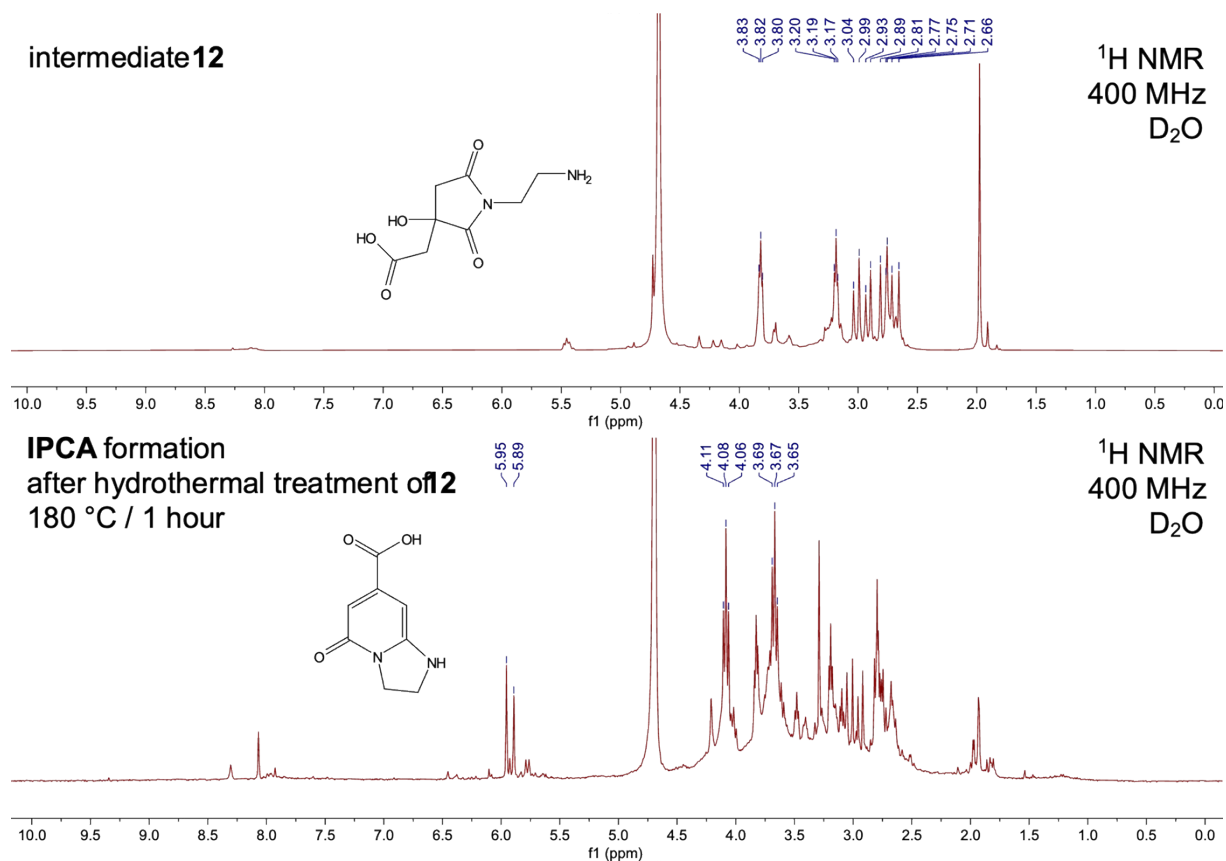
Intermediate **12** (150–200 mg) prepared by the route described above was dissolved in water (10 mL) and the solution was heated in the hydrothermal autoclave reactor as indicated below. After cooling to RT, the reaction mixture was evaporated to dryness at the rotary evaporator and analyzed by ^1H NMR.

1st experiment:

12 (as prepared by reaction of anhydromethylene acid anhydride and EDA, without any purification), heated to 180 °C / 1 hour

Top: ^1H NMR before hydrothermal treatment,

Bottom: ^1H NMR of the reaction mixture after evaporation

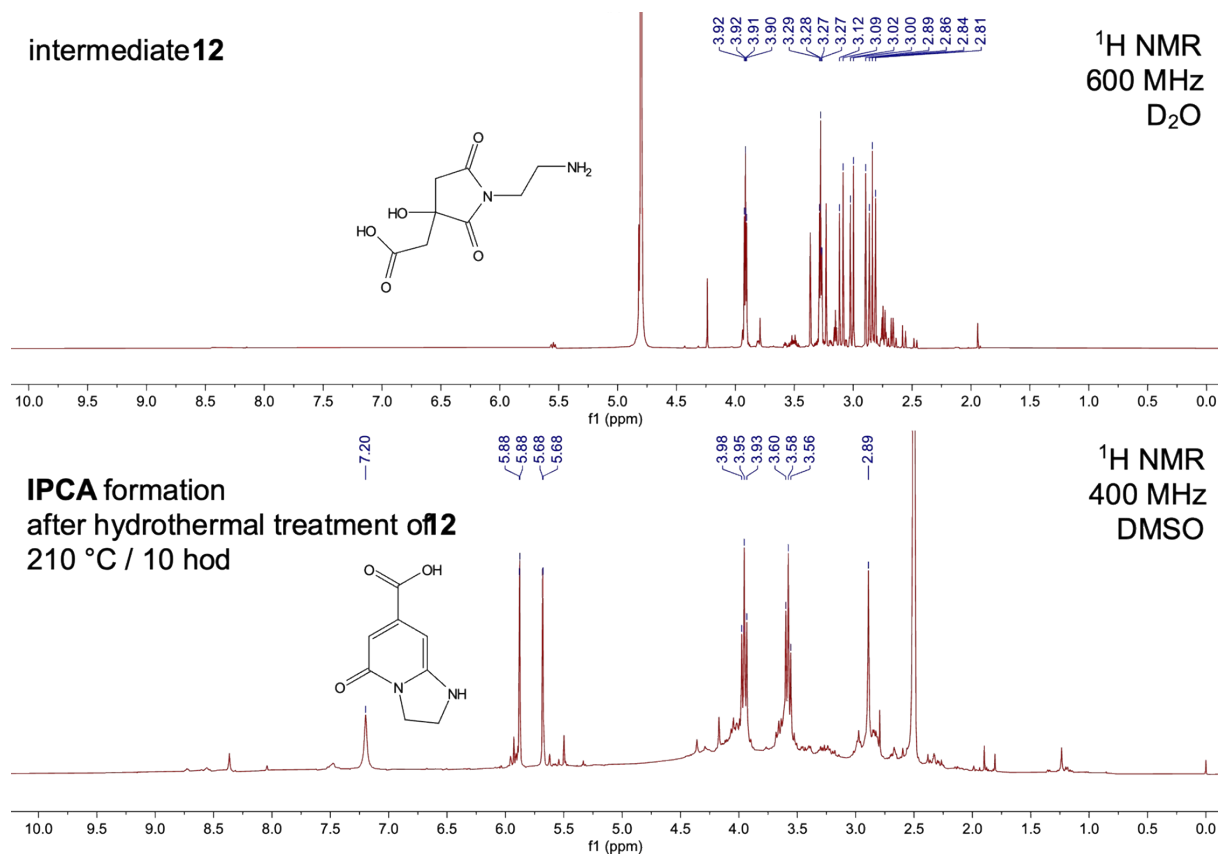


2nd experiment:

12 (purified by sephadex chromatography) / heated to 210 °C / 10 h

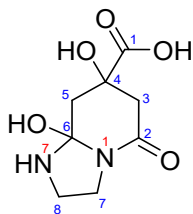
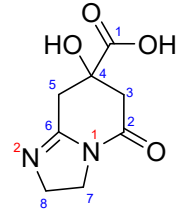
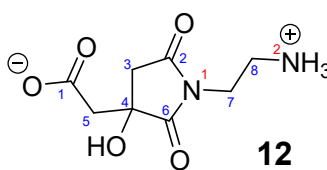
Top: ¹H NMR before hydrothermal treatment,

Bottom: ¹H NMR of the reaction mixture after evaporation



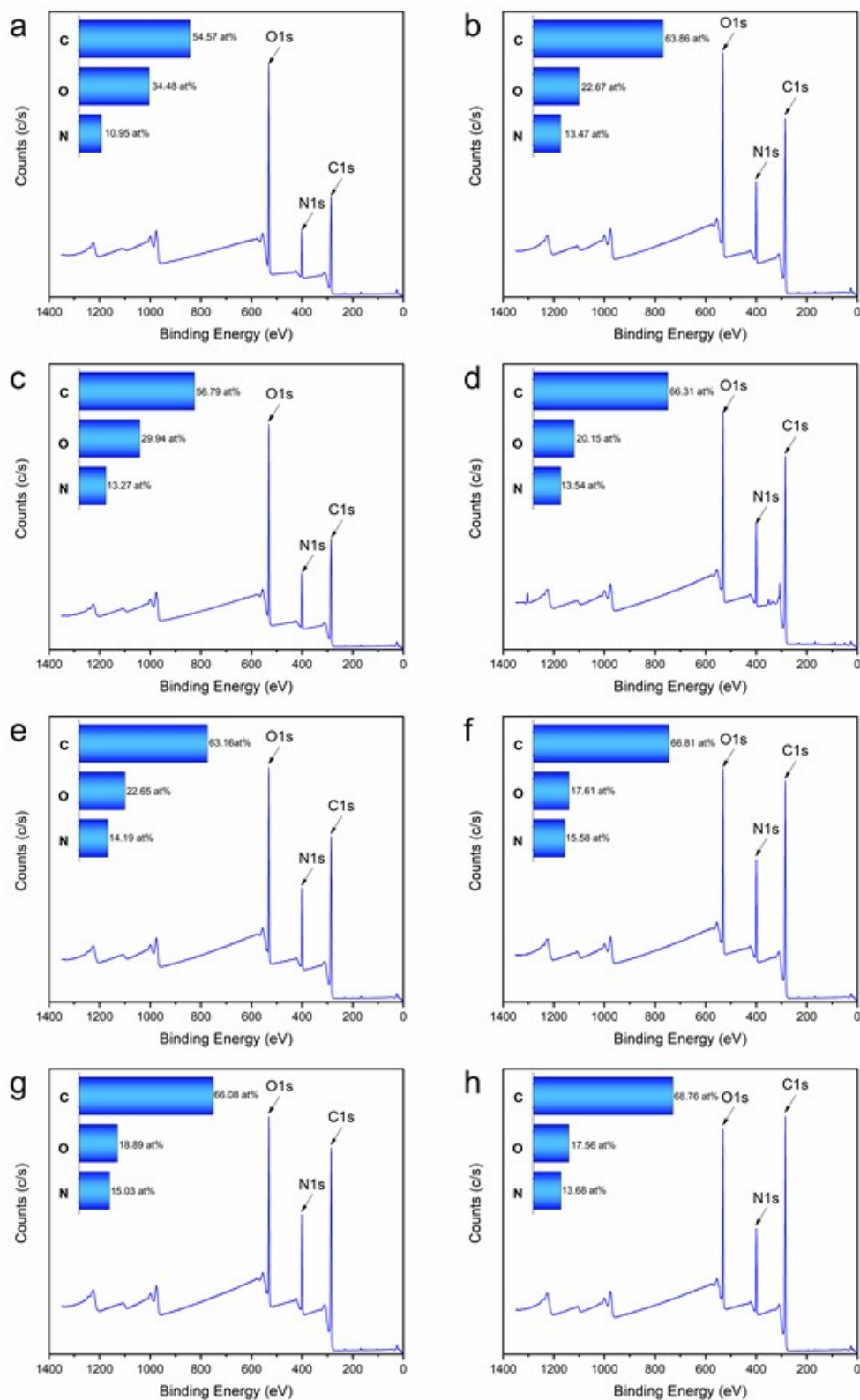
9. Calculated and experimental ^{13}C and ^{15}N NMR shifts for intermediates 7, 8 and 12

Experimental values were determined by ^1H - ^{13}C HMBC and ^1H - ^{15}N HSQC experiments and are reported using tetramethylsilane (for ^{13}C NMR) or nitromethane (for ^{15}N NMR) as a standard. The NMR chemical shifts were calculated employing the Gauge-Including Atomic Orbitals (GIAO) method^[3–5] on the structures optimized using the density functional theory (DFT) with the PBE0 functional,^[6] as implemented in the Gaussian16^[7] program, using the 6-311+G(2d,p) basis set. The geometry optimization and the shielding constant calculations were performed at the same theory level, modelling the solvent effect by the Polarizable Continuum Model with Integral Equation Formalism (IEFPCM).^[8–11] To compare the calculated shielding constant to the experimental chemical shift we used the following references: (a) the ^{13}C NMR value that we experimentally measured for the carbonyl group of citrazinic acid in DMSO (166.8 ppm) for which we calculated a shielding constant of 17.2 ppm; (b) the ^{15}N value reported for pyridine in ref.^[12] (–63.2 ppm after conversion to the nitromethane ^{15}N scale) for which we calculated a shielding constant of 78.8 ppm. The reference molecules were chosen to ensure that the reference atoms are chemically similar to those under investigation, to allow for the cancellation of systematic errors. The reference molecules were simulated using the same theory level as that used for the investigated molecules.

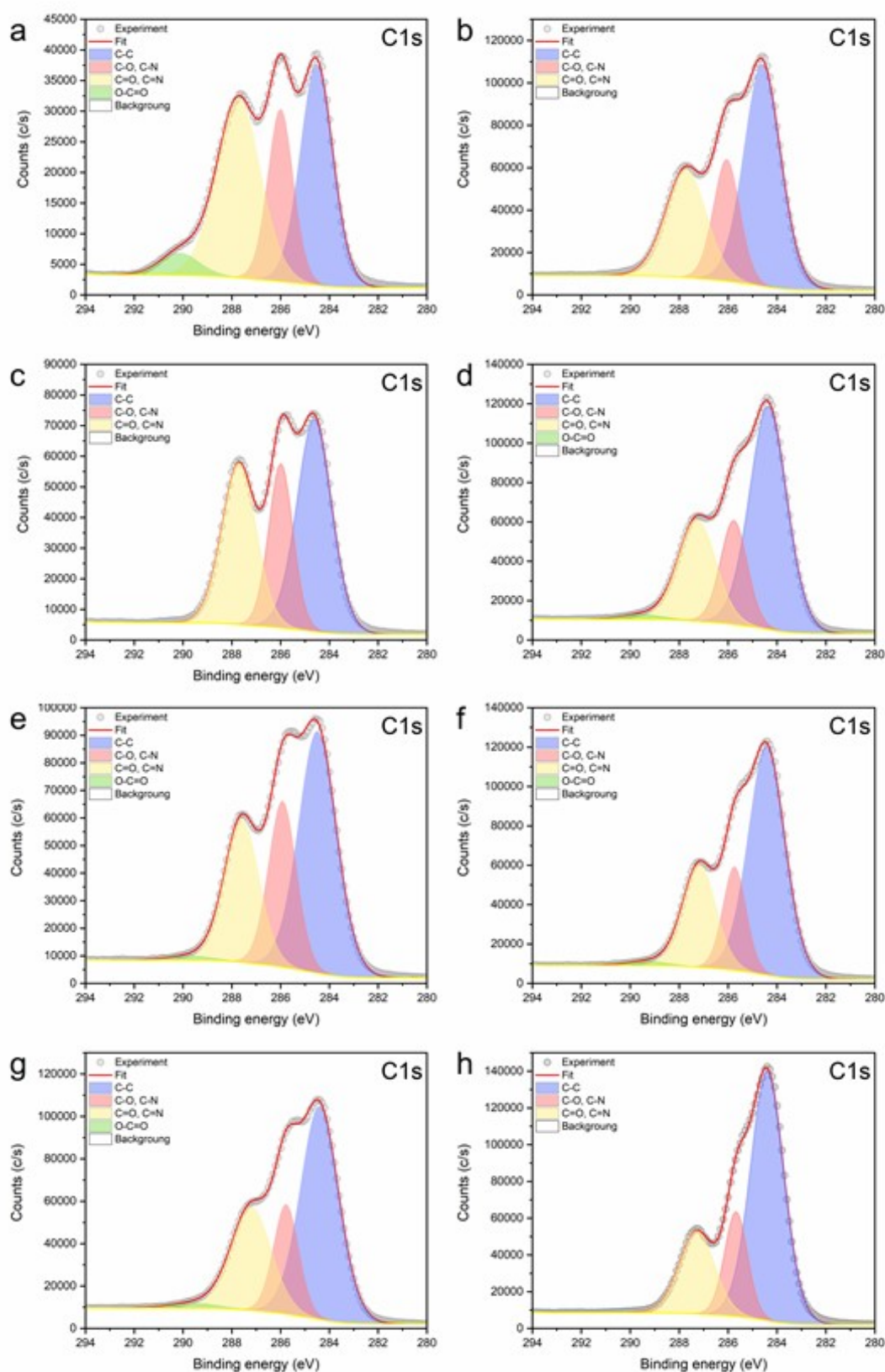
Molecule	Chemical shift (ppm)									
 7		C¹	C²	C³	C⁴	C⁵	C⁶	C⁷	C⁸	
	δ_{calcd}	178.2	165.6	39.1	70.4	39.7	95.3	40.5	39.0	
		N¹	N²							
	δ_{calcd}	−233	−319							
 8		C¹	C²	C³	C⁴	C⁵	C⁶	C⁷	C⁸	
	δ_{calcd}	178.8	167.7	41.6	72.5	34.4	159.4	42.2	53.9	
		N¹	N²							
	δ_{calcd}	−210	−130							
 12		C¹	C²	C³	C⁴	C⁵	C⁶	C⁷	C⁸	
	δ_{calcd}	176.5	177.6	42.9	73.2	38.5	183.2	31.8	36.5	
	δ_{exp}	177.1	177.9	42.3	72.9	44.5	181.4	36.3	37.6	
		N¹	N²							
	δ_{calcd}	−210	−361							
	δ_{exp}	−207	−351							

10. XPS measurements

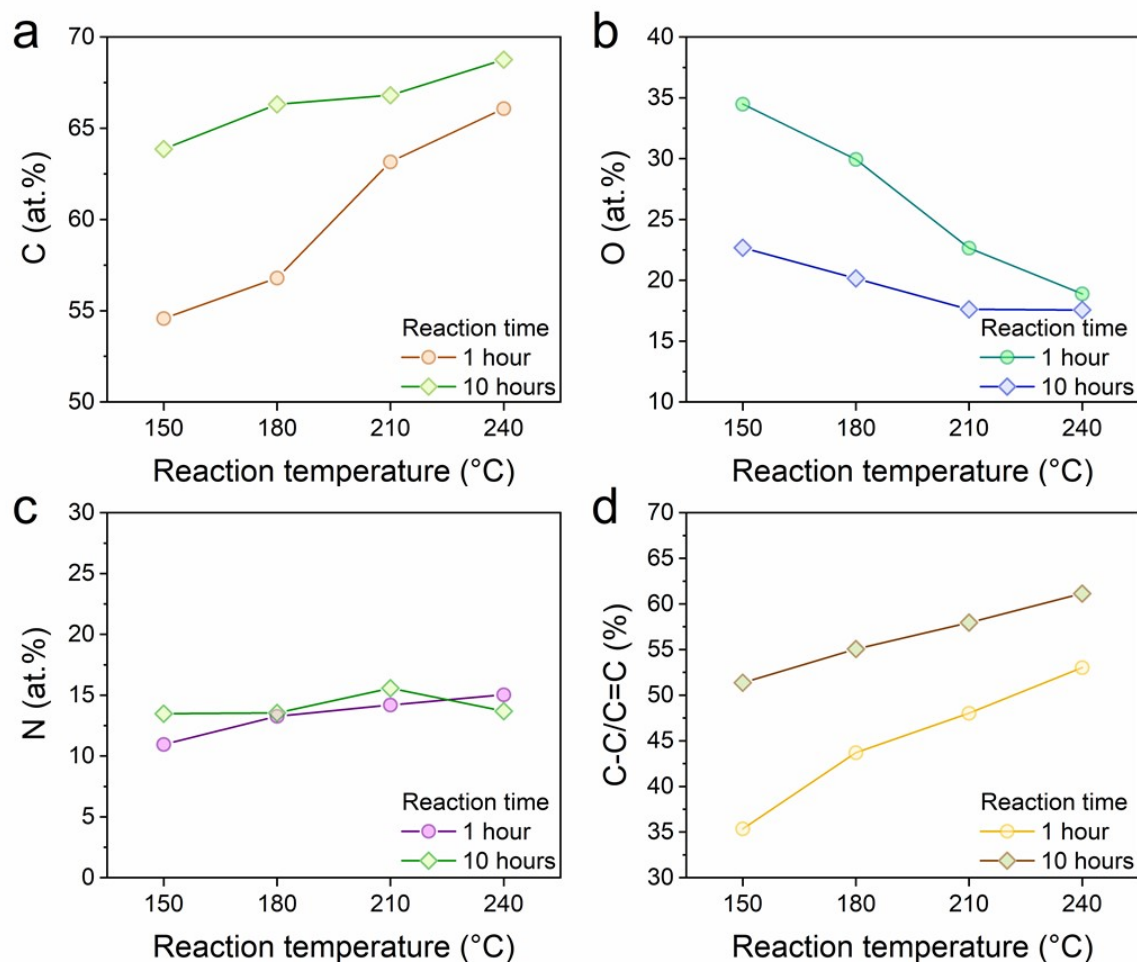
X-ray photoelectron spectroscopy (XPS) survey spectra where the insets show the elemental composition of samples (CD-150-1 (a), CD-150-10 (b), CD-180-1 (c), CD-180-10 (d), CD-210-1 (e), CD-210-10 (f), CD-240-1(g) and CD-240-10 (h)).



High-resolution X-ray photoelectron spectroscopy (HR-XPS) spectra of C1s region (CD-150-1 (a), CD-150-10 (b), CD-180-1 (c), CD-180-10 (d), CD-210-1 (e), CD-210-10 (f), CD-240-1 (g) and CD-240-10 (h)).



Dependence of the total content of individual elements (carbon (a), oxygen (b) and nitrogen (c)) derived from X-ray photoelectron spectroscopy (XPS) survey spectra and percentage representation of C-C/C=C bonds derived from deconvoluted HR-XPS spectra of the C1s region (d) on the reaction temperature and the reaction time.



11. Computational details

Our model system of reactants consisted of one molecule of citric acid (CA), one molecule of ethylenediamine (EDA) and eight explicit water molecules (**Figure S1**).

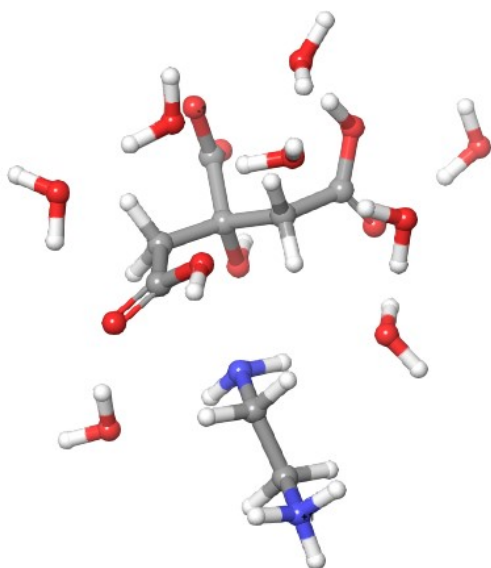


Figure S1. The initial geometry of the model CA + EDA system.

The addition of explicit water molecules was needed to properly model the proton transfer reaction steps, as the water molecules often act as a bridge. Additionally, the explicit water molecules stabilize the charged groups of the species involved in the reaction by forming hydrogen bonds which are not accurately accounted for via implicit solvation models. The most abundant forms of CA and EDA corresponding to reaction conditions (pH 5–6) are CA(2–), CA(3–) and di-protonated cations EDA(2+), respectively (**Table S1**). As the activation barrier for a direct condensation of the charged ammonium group with carboxylate is too high (31.2 kcal mol^{−1}), it can be expected that the reaction proceeds via less charged forms (CA(2–) and EDA(+)) with a much lower barrier (20.0 kcal mol^{−1}), systematically shifting the acid-base equilibrium towards these forms due to their transformation to condensates. However, the next cyclization reaction also proceeds via nucleophilic attack of carboxylate to amino group, thus the protonated form of carboxylic group is also required. In order to simplify the calculation and avoid the change of the net charge between the reaction steps we have used as initial protonation states [C₃H₄OH (COOH)₂ (COO)][−] for CA (with the central carboxylate group deprotonated]) and [C₂H₄(NH₂)(NH₄⁺)] for EDA.

Table S1. Equilibrium concentrations of CA and EDA forms in the aqueous solution at pH = 6.1 with total concentrations $c_A = 0.08 \text{ mol}\cdot\text{dm}^{-3}$ and $c_B = 0.12 \text{ mol}\cdot\text{dm}^{-3}$, respectively (CA: $pK_{a1} = 3.13$, $pK_{a2} = 4.76$, $pK_{a3} = 6.39$; di-protonated EDA: $pK_{a1} = 9.92$ and $pK_{a2} = 6.86$).^[13]

Form	Equilibrium concentrations ($\text{mol}\cdot\text{dm}^{-3}$)	Relative distribution (%)
CA	2.51E-06	0%
CA ⁻	2.35E-03	3%
CA ²⁻	5.13E-02	64%
CA ³⁻	2.63E-02	33%
EDA	2.69E-06	0%
EDA ⁺	1.78E-02	15%
EDA ²⁺	1.02E-01	85%

QM calculations

The initial geometries of reactants were obtained by building the molecules from scratch using Maestro.^[14] The initial intermolecular geometry between CA and EDA has been constructed in order to facilitate the first step of the reaction (nucleophilic attack of $-\text{NH}_2$ of EDA to $-\text{COOH}$ group of CA). The explicit water molecules were added to the system (Figure S1) using a system builder in Maestro program. First, the system was solvated by explicit water molecules in a large solvation box around the system. Then, all water molecules except eight molecules from the first solvation shell in the vicinity of the potential proton transfer sites were removed. The impact of more distant water molecules was accounted for by the COSMO implicit solvation model.^[15]

Density functional theory (DFT) calculations were performed with Jaguar program^[16] in Schrödinger. We employed the M06-2X functional^[17] and the 6-31+G(d,p) basis set. Empirical dispersion corrections were included via the DFT-D3 approach^[18] as implemented in Jaguar. M06-2X is suitable functional for calculation of weakly interacting systems as well as the systems with ionic interactions.^[19]

The domain-based local pair natural orbital (DLPNO)^[20] coupled cluster with singles and doubles and perturbative triples (CCSD(T))^[21,22] single-point calculations were performed with the def2-TZVPP basis set^[23] for all the structures previously optimized at the M06-2X/6-31+G(d,p) level using the software ORCA,^[24,25] ver. 5.0.4, applying the TightPNO settings as defined in Liakos et al.^[26] The solvent effects were considered employing the implicit Solvation Model based on Density (SMD).^[27] The Gibbs energies were then calculated as the sum of the DLPNO-CCSD(T)/def2-TZVPP electronic energy and the thermal and zero-point vibrational contributions obtained from M06-2X/6-31+G(d,p) calculations. This composite approach is hereafter referred to as DLPNO-CCSD(T)/def2-TZVPP//M06-2X/6-31+G(d,p).

Geometry optimization and the search of the transition states

The geometries of the reactants, intermediates and products were optimized by Schlegel quasi-newton method^[28] as implemented in Jaguar. The reactions were tracked by relaxed geometry scans along the most probable reaction paths. These typically involved scans along the C \cdots N distance between –NH and –COOH groups for condensation and cyclization reactions, along the H \cdots N(O) distance for proton transfer reactions, and along the C \cdots O distance for the removal of a water molecule. The proton transfer happens mostly between –NH and –OH groups and bridging water molecules. The geometries of the initial guess of the TS were taken from the relaxed geometry scan (at the point with the highest energy). In some cases, 2-D relaxed geometry scans were performed (e.g., a 2-D scan along C \cdots N distance between –NH and –COOH groups and the H \cdots N(O) distance for proton transfer) in order to obtain the geometry of the initial guess of the TS.

The TS geometries were optimized using quadratic synchronous transit (QST) method^[29] requiring an initial guess of TS and geometries of the reactant and the product. Schlegel quasi-newton method^[28] with the initial Hessian calculated by QM was employed for the geometry optimization of TS. The vibrational frequency analyses were carried out for the optimized structures to confirm the local minimum character of reactants, intermediates and products as well as to verify the occurrence of a single negative hessian eigenvalue required for the true TS.

Heuristic molecular dynamics

The MD simulations were performed using the combination of GROMACS (version 5.1.2)^[30] and LAMMPS (version 2Aug2023)^[31] software packages. GROMACS was used mainly for equilibration purposes and free MD simulations and subsequent analyses. To investigate the reaction between CA and EDA we used the REACTER algorithm^[32,33] as implemented in LAMMPS software. To put it simply, this algorithm makes possible to model reactions using classical MD by updating the topology of individual molecules and restarting the simulation at regular intervals based on predefined reaction templates. A reaction will occur if the initiator atoms approached the selected reaction cutoff distance and at the same time the reacting molecules match the specific and well-defined pre-reaction templates while meeting other user-specified reaction constraints. The protocol will update the connectivity according to the post-reaction template and the simulation continues. A comprehensive description is described in the original works of Gissinger et al.^[32,33]

In the present study we have included an extensive set of possible reaction involving the reaction of CA and EDA in different protonation states. Both molecules were modelled using the Amber99 force field.^[34] Partial charges of individual protonation states were derived using the RESP method^[35] at the HF/ 6-31G* level of theory on several conformers and then averaged for each protonation state. We tried to cover the whole scale of different protonation states; therefore, we simulated 2 independent systems. System **A** consisted of CA in its single (–) double (2–) and fully (3–) deprotonated form and EDA molecules in all possible protonation states (neutral, cationic and dicationic form, labeled as (0), (+) and (2+), respectively) and system **B**, where we artificially shifted the dissociation curve to the lower pH, so it did not contain the (3–) form of CA. Since we were unable to model properly the proton exchange reactions between reactants and the solvent related to the acid-base equilibria, we considered also the EDA(0) species in our systems to enable a wider spectrum of polymerization reactions. Additionally, to find out the effect of the thermodynamically most preferred product (IPCA) on the reaction mixture, we added IPCA molecules in their deprotonated form to the systems. Thus, these molecules as well as the fully dissociated CA and EDA in its dicationic form therefore did not directly participate in the reaction in the first phase. The detailed composition is summarized in the following **Table S3**. Each system was randomly distributed into a $10 \times 10 \times 10$ nm box (maintaining the similar concentration as in the original work of Song et al.^[36]) and solvated using the SPC/E water model.^[37] Systems were thermalized to 200 °C, near the actual reaction temperature, and pre-equilibrated in the NpT ensemble for 10 ns before the cross-linking MD using the REACTER protocol. Alternatively, system **B** was also prepared for 150 °C and 250 °C, respectively, to study the effect of the temperature on the product formation. Reaction distance cutoff for the step-growth reaction was set at 0.35 nm. Potential reactive sites were searched for every 10 steps (except for the water release procedure that was allowed for every step) and the probability of a polymeric reaction occurring in the first phase and dehydrogenation was set to 1, in the later ensuring the rapid release of the water molecule from the intermediate. Probabilities for individual competitive cyclic reactions were set to 2×10^{-4} and 1×10^{-6} for the formation of a 5- and 6-membered ring, respectively (corresponding to the $\sim 7\text{--}8$ kcal mol⁻¹ energy difference between condensation and cyclization reaction and $\sim 4\text{--}5$ kcal mol⁻¹ between 5- and 6-membered ring cyclizations). Unity probability was also chosen in the case of the charged-end reactions in the second phase. Reacting atoms were dynamically stabilized for 60 fs after each reaction. The particle-particle particle-mesh (PPPM) method^[38] was employed to calculate long-range Coulombic interactions. The non-bonded van der Waals interactions were modelled using a standard 12/6 Lennard-Jones

potential. A cutoff distance of 1 nm was utilized for electrostatic and non-bonded interactions. Production runs were performed in the NVT ensemble using the Nose–Hoover thermostat with a time step of 0.5 fs.

In total we used a set of 50 possible mutual reactions which each of them consists of two partial reactions: the first reaction facilitates the elongation of the polymeric chain, while the second reaction releases the water molecule to complete the reaction. All allowed reactions are shown in **Scheme S2**. Based on our QM calculations the first stage is expected to occur much faster than the subsequent reactions between charged species. Therefore, to simplify the modelling of this complex reaction, we divided it into two primary stages: first, the reaction of neutral groups (including self-cyclization processes), and second, the more energetically demanding cross-linking of the charged polymer ends. Both phases were simulated for 2 ns. Thus, this division ensures more manageable modelling framework and more accurate representation of each reaction step within the simulated time scale.

For analysis of minimal distances between neutral groups of individual CA and EDA we used the gmx mindist tool. The analyses were made from the first 50 ns of a free simulation. MD snapshots were generated using PyMOL software.^[39]

Classical MD simulations

The classical MD simulations were performed in GROMACS 2018 and 2024. The classical MD simulations were used for investigation of the behavior of system composed of a) multiple smaller reaction fragments with IPCA and b) reaction polymer (700 g mol^{-1}) in charged and uncharged state with and without attached IPCA. For the simulations of mixture of fragments, we used the same force field parameters as in the reaction simulations and transferred the parameters from LAMMPS format. For the simulations of polymers with and without IPCA, we optimized the structure and calculated RESP charges by B3LYP/cc-pVDZ level of theory. GAFF bonded and nonbonded parameters were used for preparation of the molecular topologies. Simulations were performed with a 2 fs time step for 100 ns at 200 °C, maintained by a Nosé–Hoover thermostat, with isotropic pressure coupling using a Parrinello–Rahman barostat. The simulation of a mixture of multiple fragments was performed on the structure extracted from reaction simulations. The simulations of charged and uncharged polymer with and without IPCAs were prepared by randomly inserting four polymers into the simulations box, its solvation and neutralization.

12. Computational analysis of the reaction mechanism

A comparison of the standard reaction Gibbs energies ($\Delta_r G^\circ(T = 180\text{ }^\circ\text{C})$) and the activation barriers ($\Delta G^\ddagger(T = 180\text{ }^\circ\text{C})$) of individual reaction steps between CA and EDA in water leading to IPCA computed at the M06-2X-D3/6-31+G(d,p) level and the composite DLPNO-CCSD(T)/def2-TZVPP//M06-2X/6-31+G approach are shown in **Figure S2**. All the computed data are collected in **Table S2**.

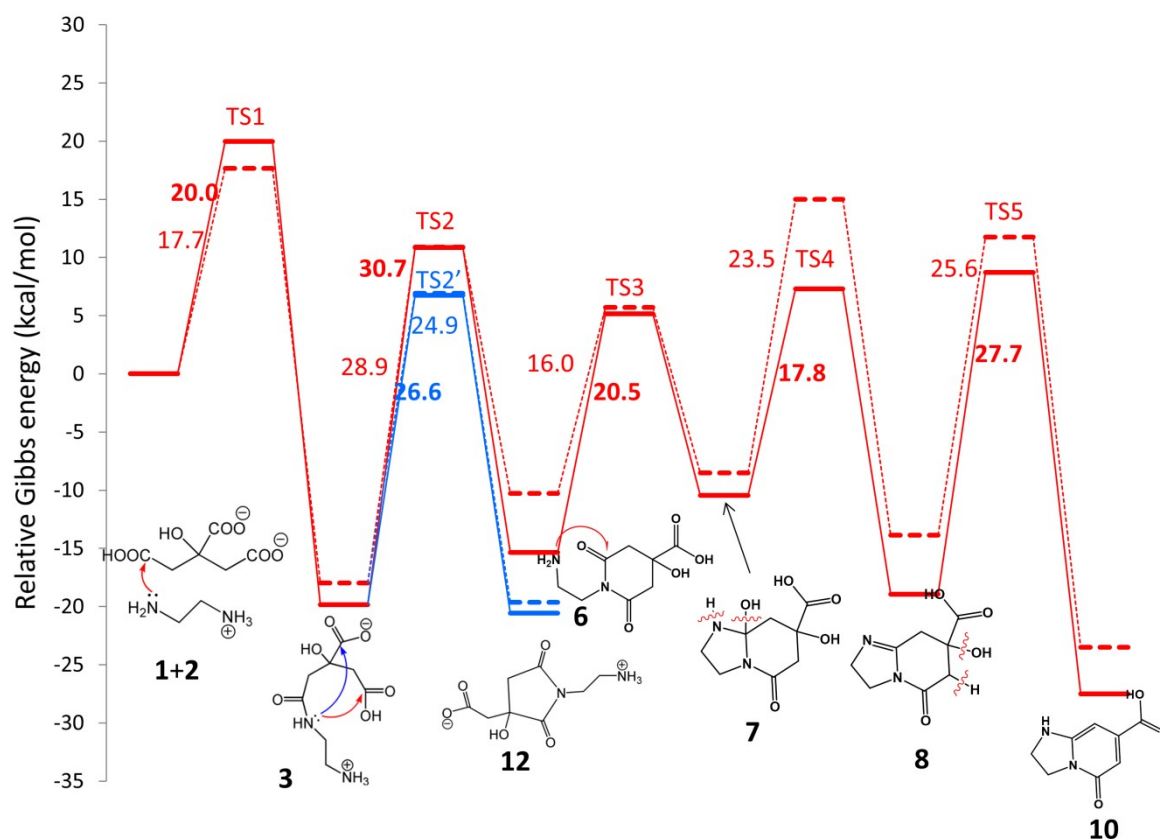


Figure S2. Standard Gibbs energy diagram ($T = 180\text{ }^\circ\text{C}$) of the early stages of the reaction between CA and EDA in water. The full and dashed lines represent DLPNO-CCSD(T)/def2-TZVPP//M06-2X/6-31+G(d,p) and M06-2X/6-31+G(d,p) energies, respectively. Gibbs activation barriers are given in kcal mol^{-1} .

The formation of **3** starts with a nucleophilic attack of the amine group onto protonated carboxylate via transition state **TS1a** with a small barrier of 4.8 kcal mol^{-1} (at the composite CC/DFT level of theory, **Figure S3**) leading to an intermediate **3a** which transforms to **3** via transition state (**TS1**) by chain proton transfer reaction facilitated by surrounding water molecules ending with detachment of a water molecule (**Figure S4**). The Gibbs energy barrier for this reaction (calculated as the energy difference between **TS1** and **3a**) is $17.6\text{ kcal mol}^{-1}$ (**Figure S3**). **TS1** is the rate determining transition state of $1 + 2 \rightarrow 3$ reaction, thus the energy

barrier of the reaction (calculated as energy difference between **TS1** and **1+2** in CA(−) and EDA(+) form) is 20.0 kcal mol^{−1}.

However, the most abundant forms of CA and EDA corresponding to reaction conditions (pH 5–6) are ionic forms (**Table S1**). The reaction of ionic groups of **1** and **2** to form **3a** is preceded by proton transfer from −NH₃⁺ of EDA to −COO[−] of CA facilitated by water molecules to form a carboxylic group at CA and a neutral amino group at EDA. This step requires additional 11.2 kcal mol^{−1} to the reaction barrier of **1+2** to **3** reaction, implying the total barrier 31.2 kcal mol^{−1}.

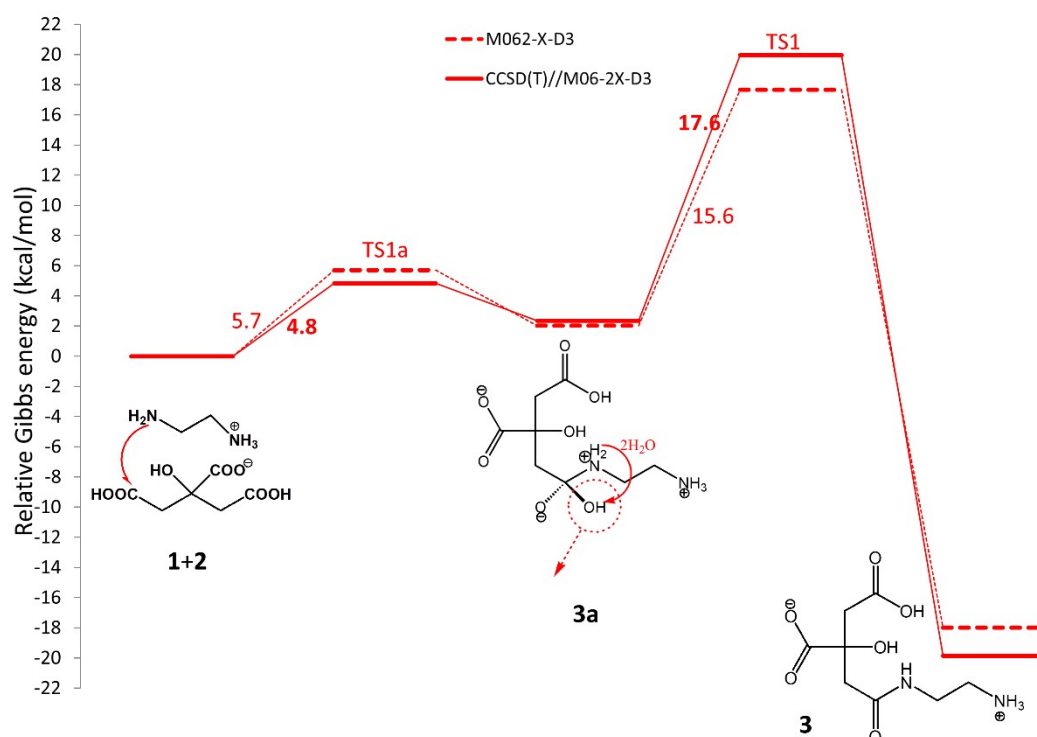


Figure S3. The calculated relative Gibbs energy profile (in kcal mol^{−1}) at 180 °C for the reaction of ethylenediamine (**1**) with citric acid (**2**) to form **3**. The full and dashed lines represent DLPNO-CCSD(T)/def2-TZVPP/M06-2X/6-31+G(d,p) and M06-2X/6-31+G(d,p) energies, respectively.

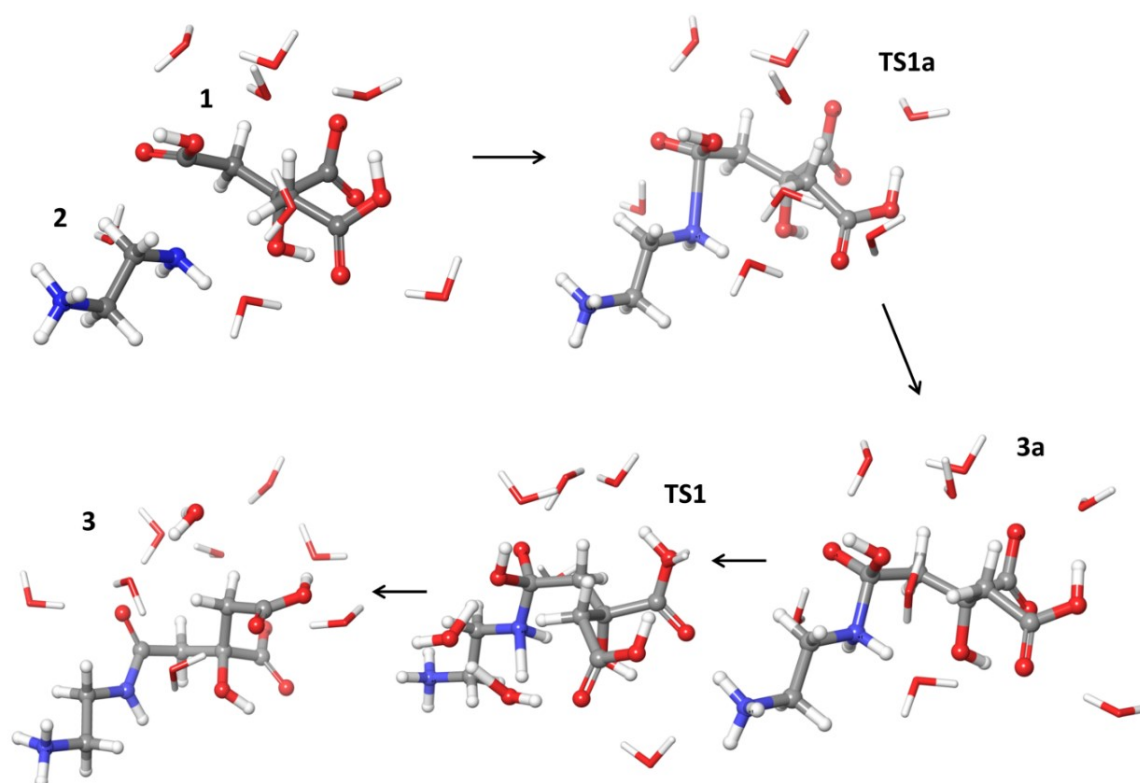


Figure S4. The geometry of intermediates and transition states for the reaction of ethylenediamine (**2**) with citric acid (**1**) to form **3**. Transition state **TS1** of proton transfer mediated by 2 water molecules is depicted in ball-and-stick representation. The new water molecule after formation of **3** is depicted in ball-and-stick representation.

Alternatively, the $1 + 2 \rightarrow 3$ reaction can start from CA anhydride instead of CA. The initial formation of both symmetrical and asymmetrical CA anhydrides was considered. The CA anhydride then reacts with EDA to form the product **3** via an intermediate analogous to **3a** (**Figure S5**). The structure of **TS1'** for the reaction starting from asymmetric CA anhydride is shown in **Figure S6**. The Gibbs energy for the formation of asymmetric CA anhydride from CA is $7.2 \text{ kcal mol}^{-1}$. The energy of the **TS1'** relative to the energy of CA+EDA structure was $19.6 \text{ kcal mol}^{-1}$ as oppose to $20.0 \text{ kcal mol}^{-1}$ for the direct reaction of $1 + 2 \rightarrow 3$ (**Figure S5**). Therefore, this alternative reaction path for $1 + 2 \rightarrow 3$ reaction via CA anhydride is possible and energetically comparable to the direct reaction. However, the main reaction profile is depicted with direct reaction.

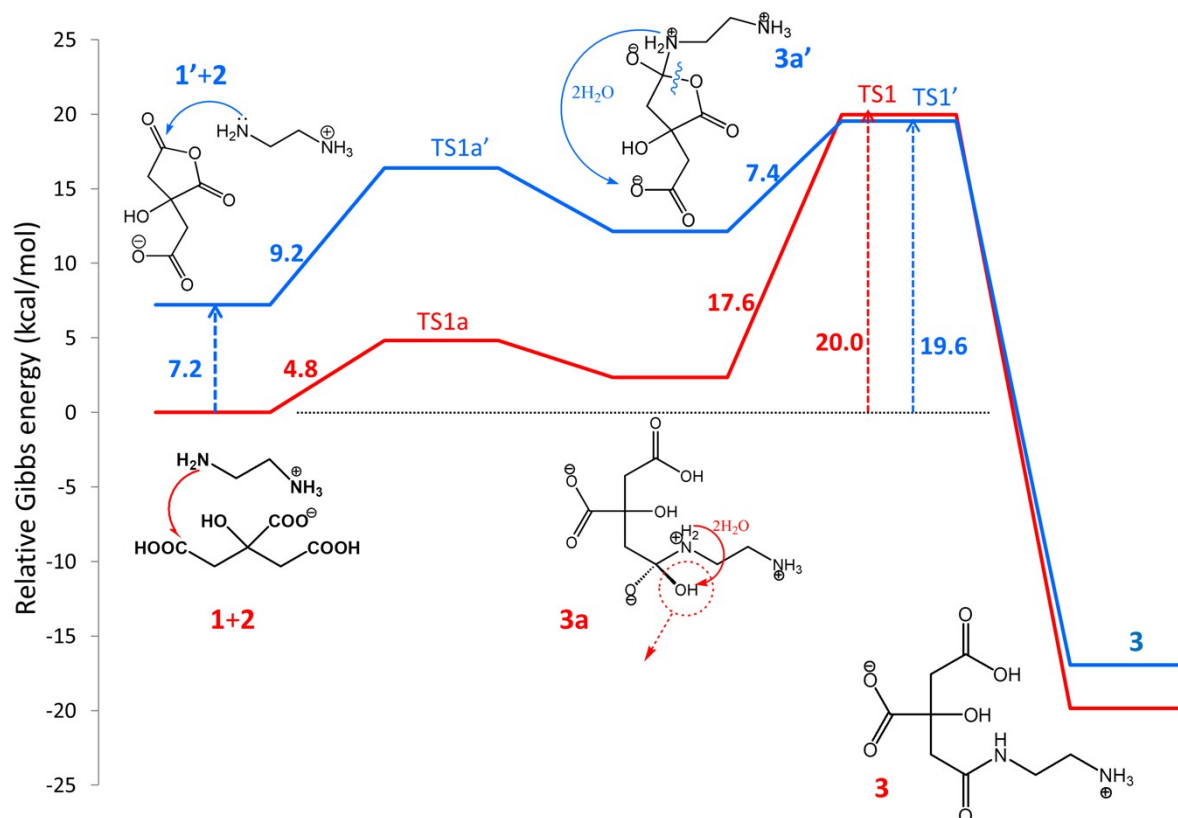


Figure S5. The Gibbs energy profile (in kcal mol⁻¹) at 180 °C for the reaction of ethylenediamine (1) with citric acid (2) to form 3, comparison between direct (red) and CA anhydride (1') path (blue) obtained using the DLPNO-CCSD(T)/def2-TZVPP//M06-2X/6-31+G(d,p) approach.

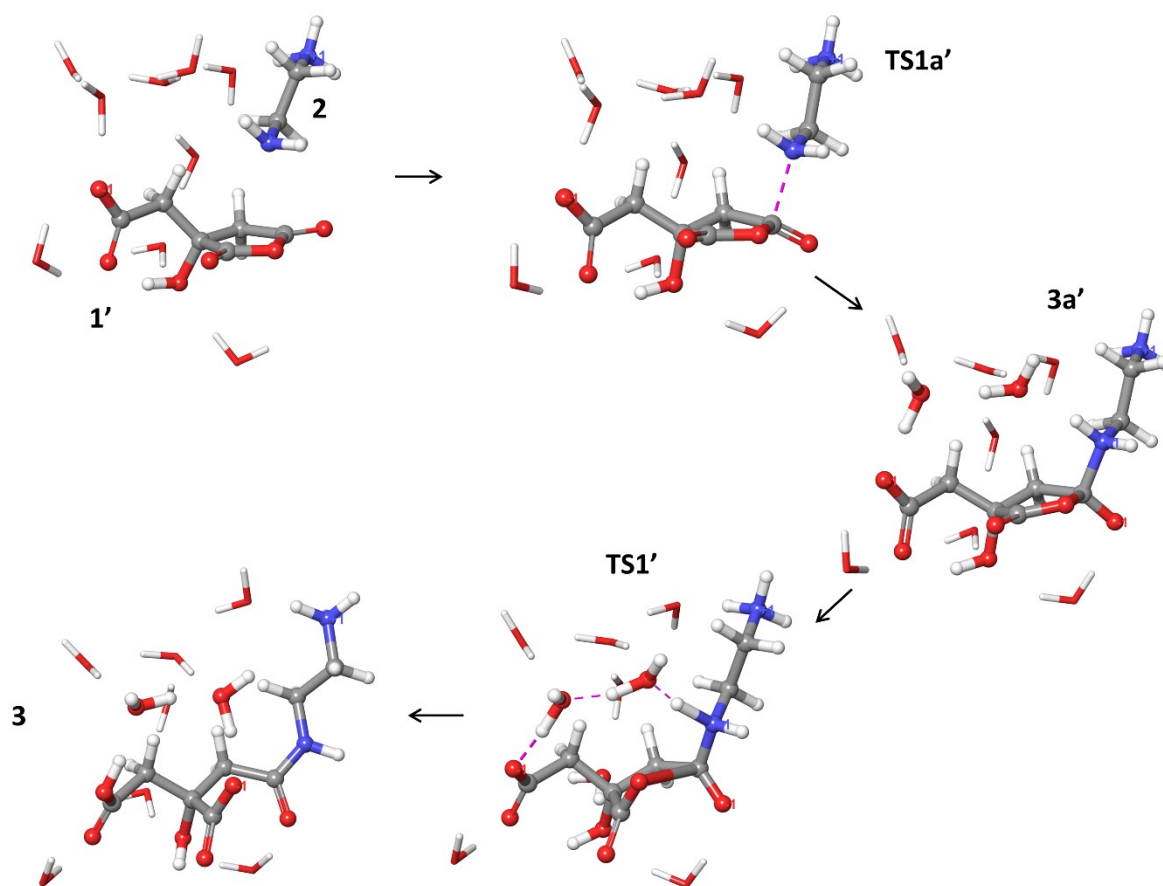


Figure S6. The geometry of intermediates and transition states for the alternative reaction of ethylenediamine (**2**) with asymmetric anhydride of citric acid (**1'**) to form **3**. Transition state **TS1'** of proton transfer mediated by 2 water molecules depicted in ball-and-stick representation.

The next step is the first cyclization reaction. As already mentioned, the cyclization can result in the formation of either 6-membered ring (**6**) or 5-membered ring (**12**) intermediates, depending on which carboxylic group of **3** the nucleophilic attack takes place. The formation of the 5-member ring (**12**) starts by proton transfer from the one -COOH to the central -COO^- group (**Figure S7**) resulting in a different protonation state **3'**. The Gibbs energy of **3'** in the conformation allowing the nucleophilic attack of -NH- to the protonated carboxylic group is by $2.4 \text{ kcal mol}^{-1}$ higher than **3** (**Figure S8**) in the original conformation and protonation state. The reaction proceeds via transition state **TS2'** with nucleophilic attack of -NH- group to protonated carboxylic group and, at the same time, the proton transfer from -NH- to -OH of the same -COOH via 2 water molecules acting as a bridge. The formed -OH_2 would immediately leave as the new water molecule, resulting in the formation of **12** (see **Figure S7**).

The energy barrier (calculated as Gibbs energy difference between **TS2'** and **3** in the original conformation and protonation state) for this reaction is 26.6 kcal mol⁻¹ (**Figure S8**).

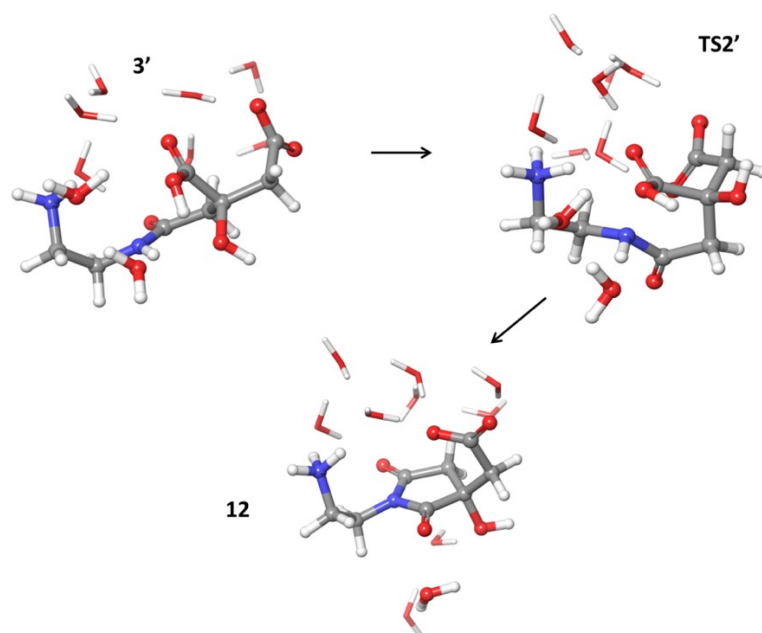


Figure S7. The geometry of intermediates and transition states for the cyclization reaction of **3** to **12** (5-member cycle). **3'** is near the attack conformation of **3** with different carboxylic group protonated (see **3** in **Figure S4, S6**). The water molecule mediating the proton transfer in transition state **TS2'** is depicted in ball-and-stick representation. The new water molecule after formation of **12** is depicted in ball-and-stick representation

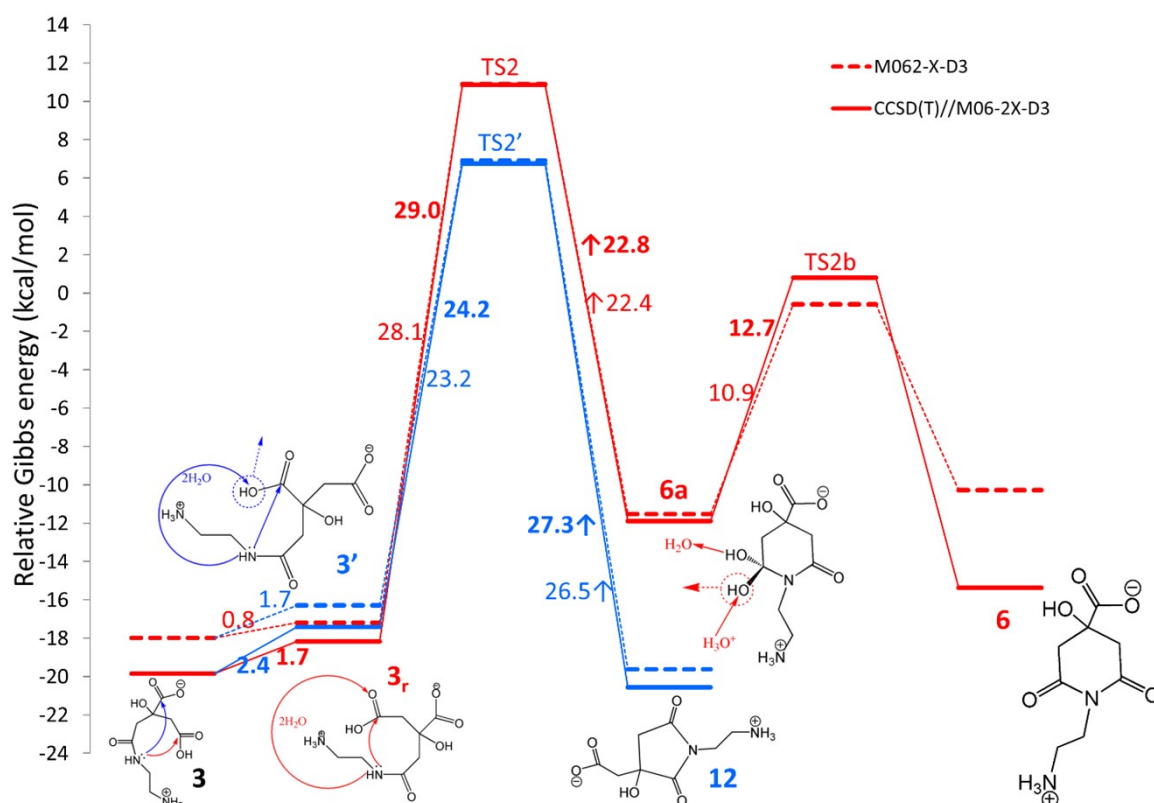


Figure S8. The calculated Gibbs energy profile (in kcal mol⁻¹) at 180 °C for the cyclization reactions **3** → **12** (blue) and **3** → **6** (red). The full and dashed lines represent DLPNO-

CCSD(T)/def2-TZVPP//M06-2X/6-31+G(d,p) and M06-2X/6-31+G(d,p) energies, respectively. The zero energy level corresponds to the system of **1+2** in their CA(−) and EDA(+) forms.

An alternative reaction is the formation of the 6-member ring (**6**), which starts by nucleophilic attack of −NH− group to the protonated carboxylic group (located furthest from −NH− group). The reaction proceeds via transition state **TS2** (**Figure S9**) with a nucleophilic attack of −NH− group to protonated carboxylic group accompanied by the proton transfer from −NH− to =O of the carboxylic group (involved in nucleophilic attack) mediated via two water molecules resulting in formation of the intermediate **6a** (**Figure S9**). The energy barrier (calculated as Gibbs energy difference between **TS2** and **3** in the original conformation) for this reaction is 30.7 kcal mol^{−1} (**Figure S8**). The reaction then proceeds via **TS2b** (**Figure S9**) by the proton transfer from a nearby protonated water molecule (H₃O⁺) to the −OH of the former carboxylic group (one attacked by −NH− during the previous step) creating −OH₂ that would immediately leave as the new water molecule accompanied by the proton transfer from the other −OH of the former carboxylic group to the nearby water (forming H₃O⁺) resulting in the formation of **6** (**Figure S9**). The energy barrier (calculated as energy difference between **TS2b** and **3a**) for this reaction is 12.7 kcal mol^{−1} (**Figure S8**), thus the rate determining transition state would be the previous **TS2**.

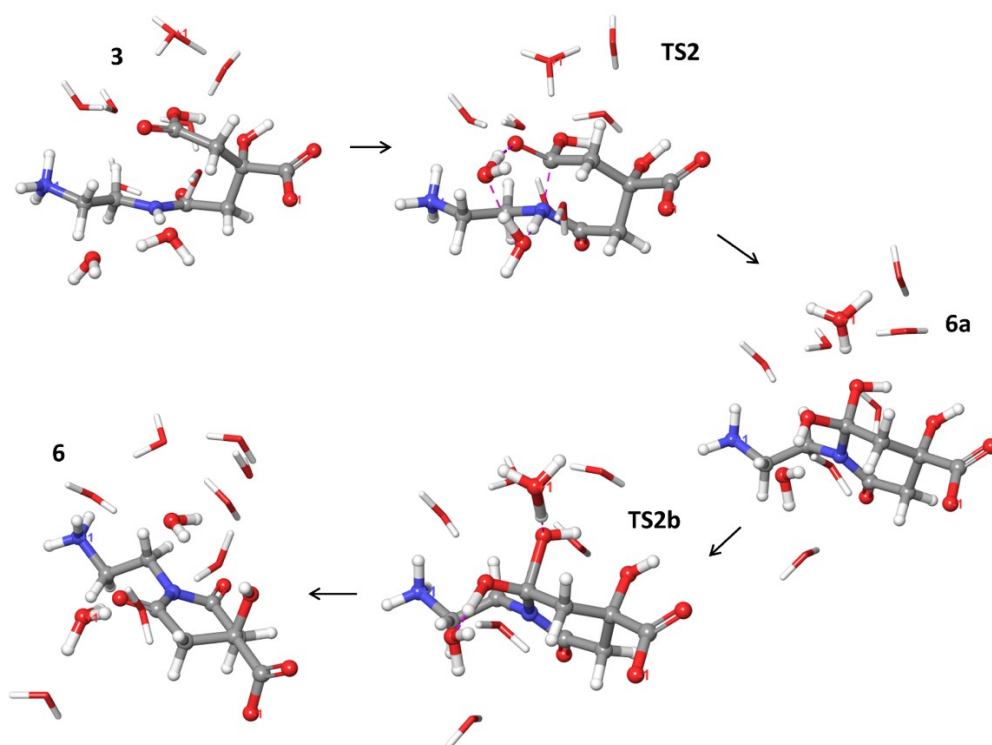
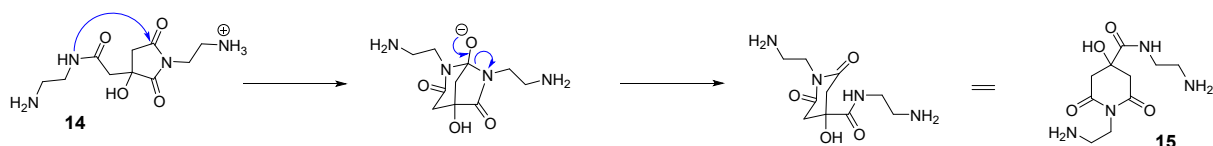


Figure S9. The geometry of intermediates and transition states for the cyclization reaction of **3** to **6** (6-membered cycle). **3** is near the attack conformation. The water molecules mediating the proton transfer in transition state **TS2** are depicted in ball-and-stick representation (**3**, **TS2**).

The protonated water and water accepting proton are depicted in ball-and-stick representation (**6a**, **TS2b**). The new water molecule after formation of **6** is depicted in ball-and-stick representation.

Compound **12** is both kinetically and thermodynamically favoured compared to **6**, which is in line with the experimental evidence of **12** at the lower temperature (150–180 °C) regimes. Although the 5-membered ring structure can be involved in subsequent condensation reactions leading, e.g., to surmised structure **13** and thus can be present in the polymeric structures formed using longer reaction times, its transformation to IPCA (**10**) would require complex rearrangements (**Scheme S1**). Therefore, we hypothesize that at higher temperatures the 5-member ring of **12** reversibly opens back to **3**, and the reaction further proceeds via the competitive 6-membered ring cyclization towards the thermodynamically most stable product **10**, which is IPCA. The reverse reaction from **12** to **3** requires proceed via 27.3 kcal mol⁻¹ Gibbs energy barrier (**Figure S8**).



Scheme S1. The transformation of intermediate **14** to **15** requiring formation of a bicyclic intermediate.

The next step in IPCA synthesis is the second cyclization reaction from **6** to **7**. The reaction should presumably proceed via a nucleophilic attack of amino group to carboxyl group; thus, the reactive amino group has to be again in the neutral deprotonated form. However, as **6** (from the previous step) is in the zwitterion form, the reaction itself starts by proton transfer from –NH₃⁺ to –COO⁻ via two water molecules acting as a bridge (**Figure S10**). The Gibbs energy of the resulting neutral form of **6** (**6b**) in the near-the-attack conformation is by 7.8 kcal mol⁻¹ higher than zwitterion **6** in the original conformation (see **Figure S11**). The cyclization reaction itself proceeds via transition state **TS3a** with nucleophilic attack of –NH₂ to –(C=O)– forming the intermediate **7a** (see **Figure S10**). The energy barrier (calculated as energy difference between **TS3a** and **6b**) for this reaction is 6.2 kcal mol⁻¹ (see **Figure S11**). The reaction then continues via transition state **TS3** with proton transfer from –NH₂⁺ to –O⁻ mediated via one water molecule resulting in the formation of **7** (see **Figure S10**). The energy barrier (calculated as energy difference between **TS3** and **7a**) for this reaction is 12.5 kcal mol⁻¹ (see **Figure S11**). **TS3** is the rate determining transition state of **6** to **7** cyclization reaction, thus the energy barrier of **6** to **7** cyclization reaction (calculated as energy difference between **TS3** and **6** zwitterion state and original conformation) is 20.5 kcal mol⁻¹.

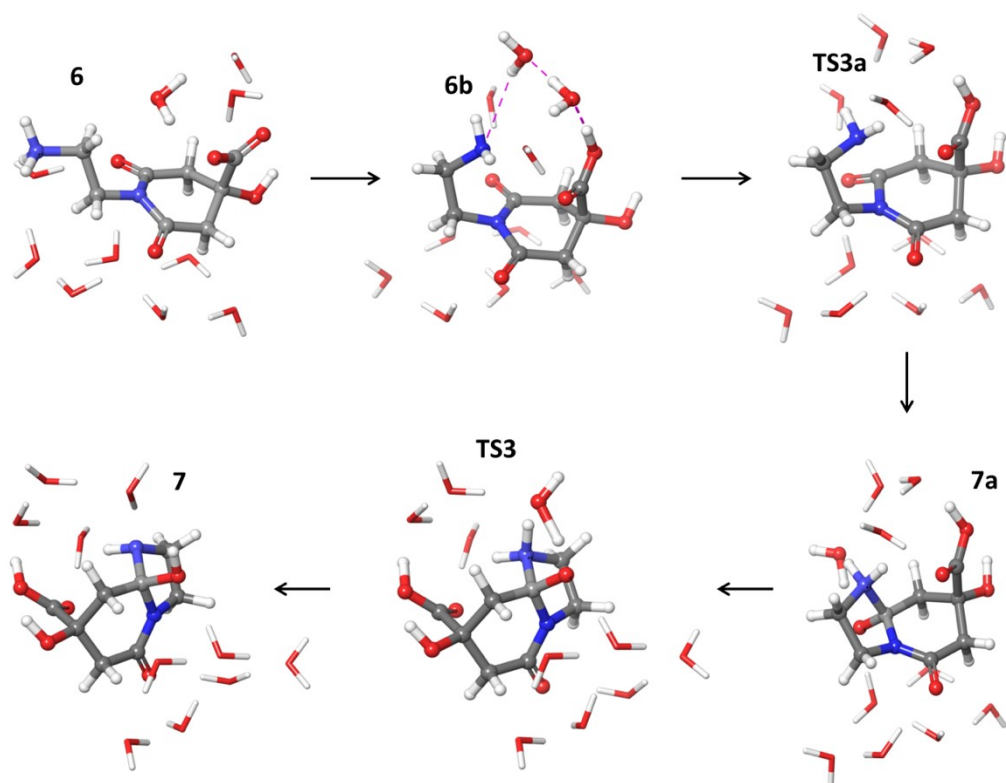


Figure S10. The geometry of intermediates and transition states for the second cyclization reaction of **6** to **7**. The 2 water molecules mediating the -NH_3^+ to COO^- proton transfer forming **6b** (neutral protonation form of **6**) are depicted in ball-and-stick representation (**6b**). The water molecule mediating proton transfer in transition state **TS3** is depicted in ball-and-stick representation. The new water molecule after formation of **7** is depicted in ball-and-stick representation.

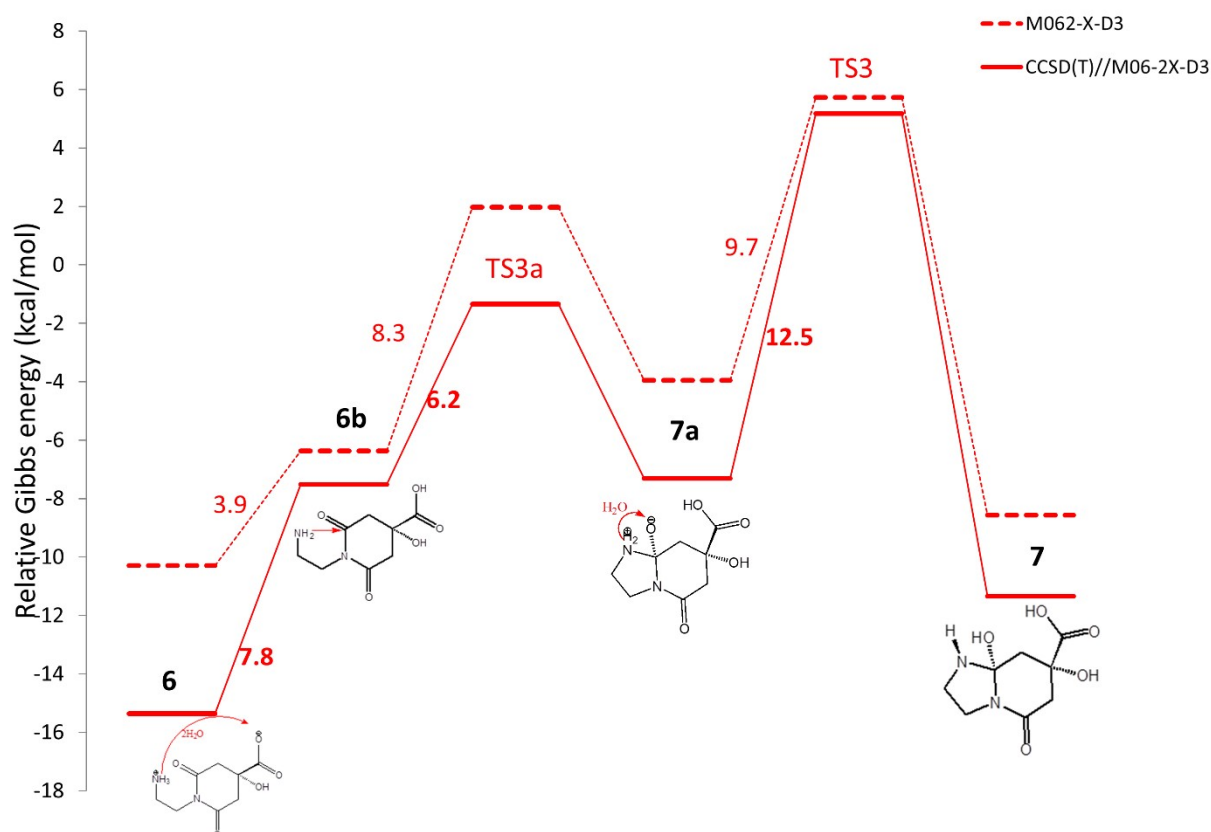


Figure S11. The calculated Gibbs energy profile (in kcal mol⁻¹) at 180 °C for the second cyclization reaction **6** → **7**. The full and dashed lines represent DLPNO-CCSD(T)/def2-TZVPP//M06-2X/6-31+G(d,p) and M06-2X/6-31+G(d,p) energies, respectively. The zero energy level corresponds to the system of **1+2** in their CA(–) and EDA(+) forms.

The next step in IPCA synthesis is the dehydration of the **7** leading to the **8**. Different conformation of **7** (see upper row of **Figure S12**), where hydrogen of –NH– group is rotated to the same side as –OH group at the neighbouring atom, was used as the starting structure. This conformation is only by 0.1 kcal mol⁻¹ higher than the original conformation of **7**. The reaction proceeds via transition state **TS4** with proton transfer from –NH– group to –OH group at the neighbouring atom mediated by one water molecule, forming –OH₂ that would immediately leave the new water molecule (**Figure S12**). The energy barrier (calculated as energy difference between **TS4** and **7**) for this reaction is 17.8 kcal mol⁻¹ (see **Figure S13**).

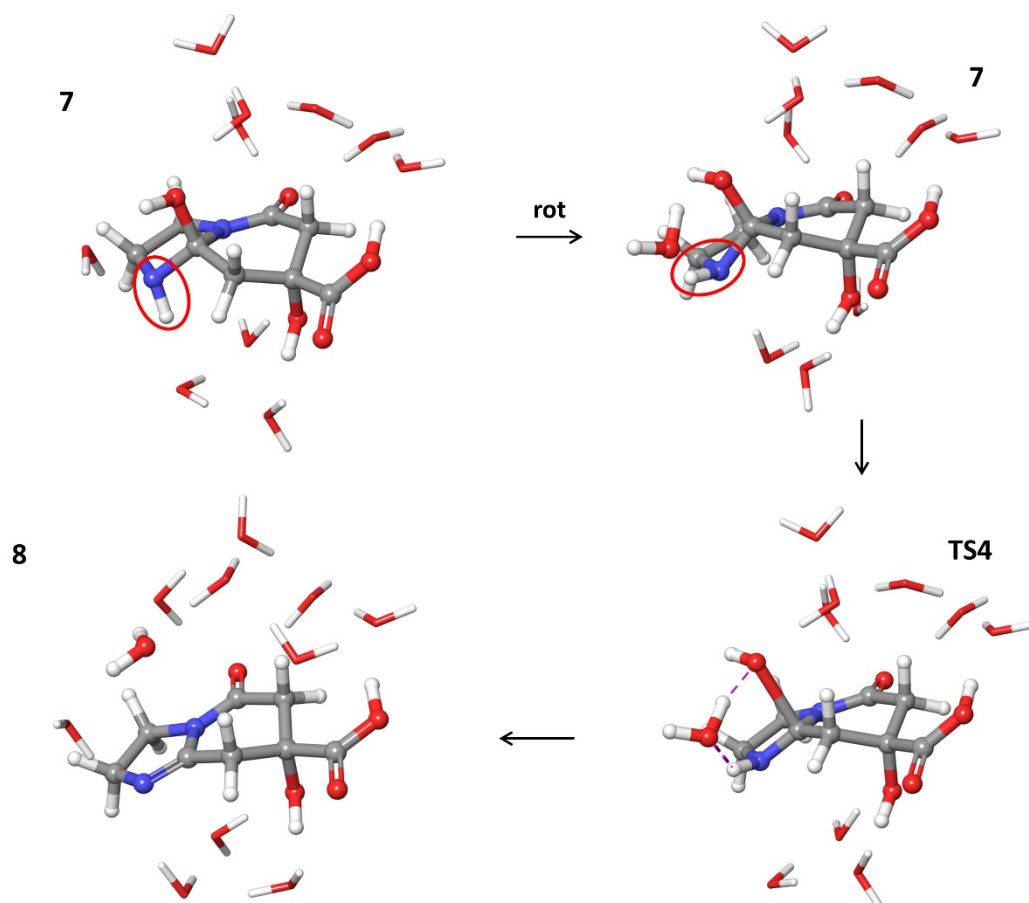


Figure S12. The geometry of intermediates and transition states for the **7** to **8** reactions. The upper row depicts conformation change (rotation of -NH- group) required by the reaction. The water molecule mediating proton transfer in **TS4** is depicted in ball-and-stick representation. The new water molecule after formation of **8** is depicted in ball-and-stick representation.

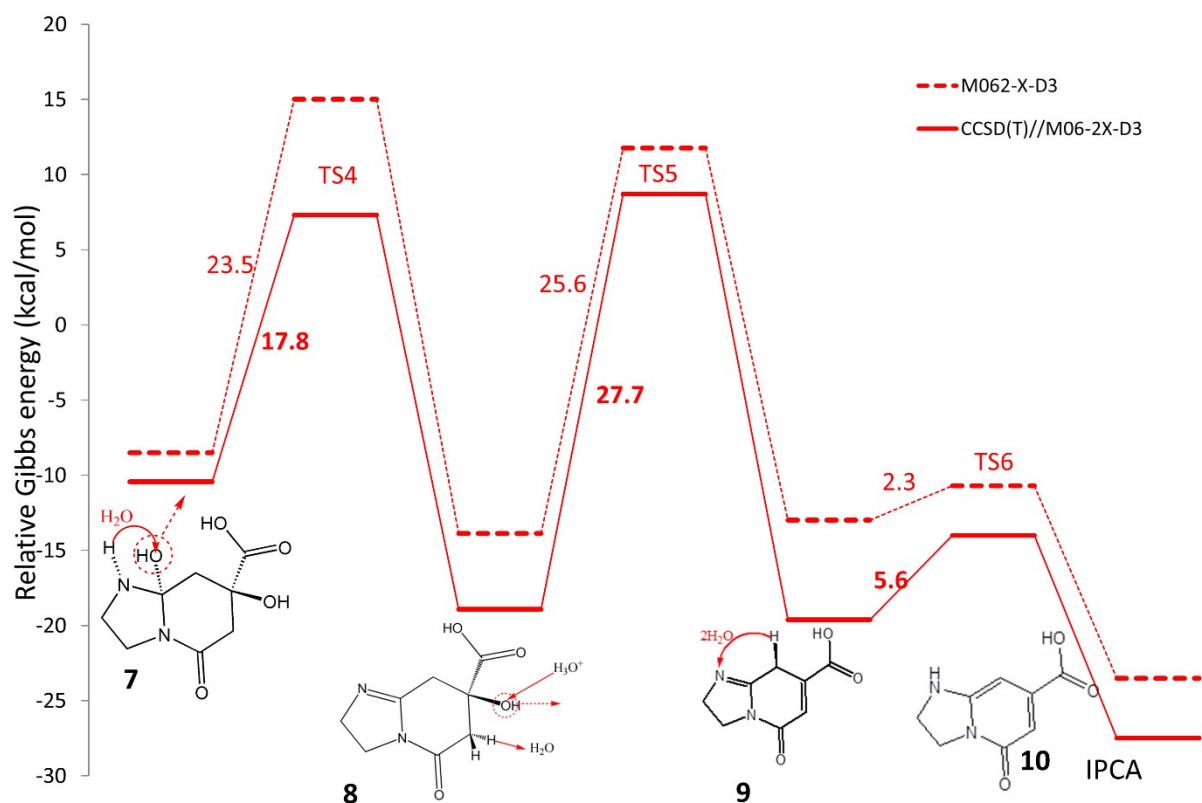


Figure S13. The calculated Gibbs energy profile (in kcal mol⁻¹) at 180 °C for the last steps of IPCA synthesis (**7** to **10** reactions). The full and dashed lines represent DLPNO-CCSD(T)/def2-TZVPP//M06-2X/6-31+G(d,p) and M06-2X/6-31+G(d,p) energies, respectively. The zero energy level corresponds to the system of **1+2** in their CA(–) and EDA(+) forms.

The reaction then continues by dehydration of **8** to form one of the IPCA tautomers (**9**). This reaction would proceed via **TS5** with the proton transfer from –CH₂– (at position 6 of imidazopyridine ring) to the nearby water molecules (forming H₃O⁺) accompanied by the proton transfer from the protonated water molecule (H₃O⁺, initially located nearby the –OH bonded to carbon at position 7 of imidazopyridine ring) to the –OH (bonded to carbon at position 7 of imidazopyridine ring) resulting in the formation of –OH₂ that would immediately leave as the new water molecule (see upper row of **Figure S14**). The energy barrier (calculated as energy difference between **TS5** and **8**) for this reaction is 27.7 kcal mol⁻¹ (see **Figure S13**).

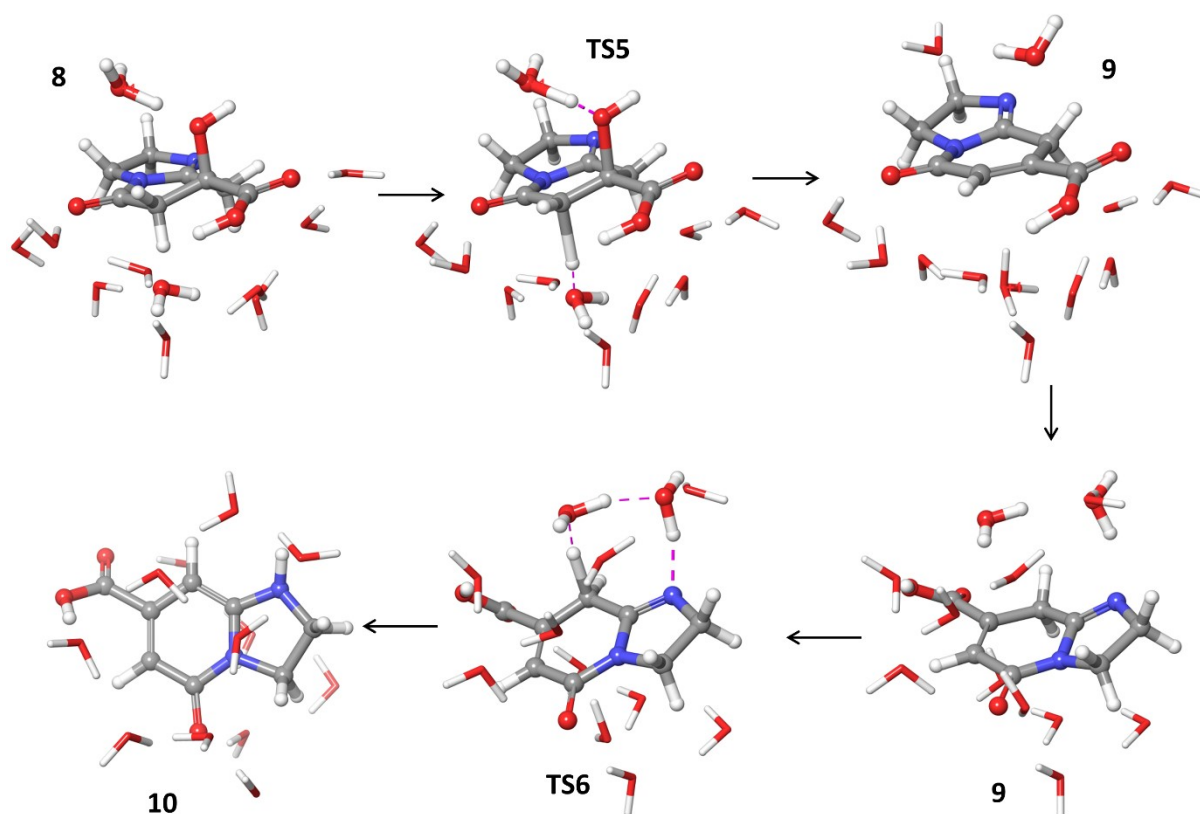


Figure S14. The geometry of intermediates and transition states for the **8** to **10** reactions. The protonated and neutral water molecules mediating proton transfer in transition state **TS5** and two water molecules mediating proton transfer in transition state **TS6** are depicted in ball-and-stick representation. The new water molecule after formation of **9** is depicted in ball-and-stick representation.

An alternative reaction path from **7** to **9** has been also investigated. In this case, the reaction would start by the elimination of a water molecule from the six membered ring of structure **7** to form structure **8'** (**Figure S15**) following a similar mechanism as in reaction **8** \rightarrow **9**. Then, a water molecule would be eliminated from the five membered ring of structure **8'** following a similar mechanism as in reaction **7** \rightarrow **8**. However, already the first reaction had much higher barrier of 32 kcal mol⁻¹ compared to the original **8** \rightarrow **9** (and **7** \rightarrow **8**) reactions (see **Figure S16**). Therefore, this alternative path was considered as unfavourable.

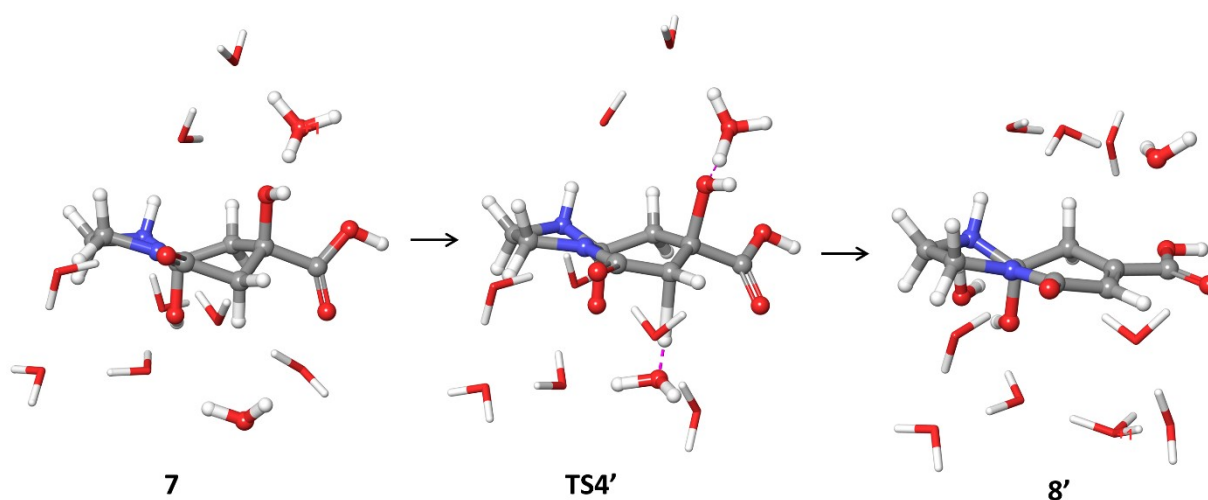


Figure S15. The geometry of intermediates and transition state for the alternative **7** to **8'** reaction. The protonated and neutral water molecules mediating proton transfer in transition state **TS4'** are depicted in ball-and-stick representation. The new water molecule after formation of **9** is depicted in ball-and-stick representation.

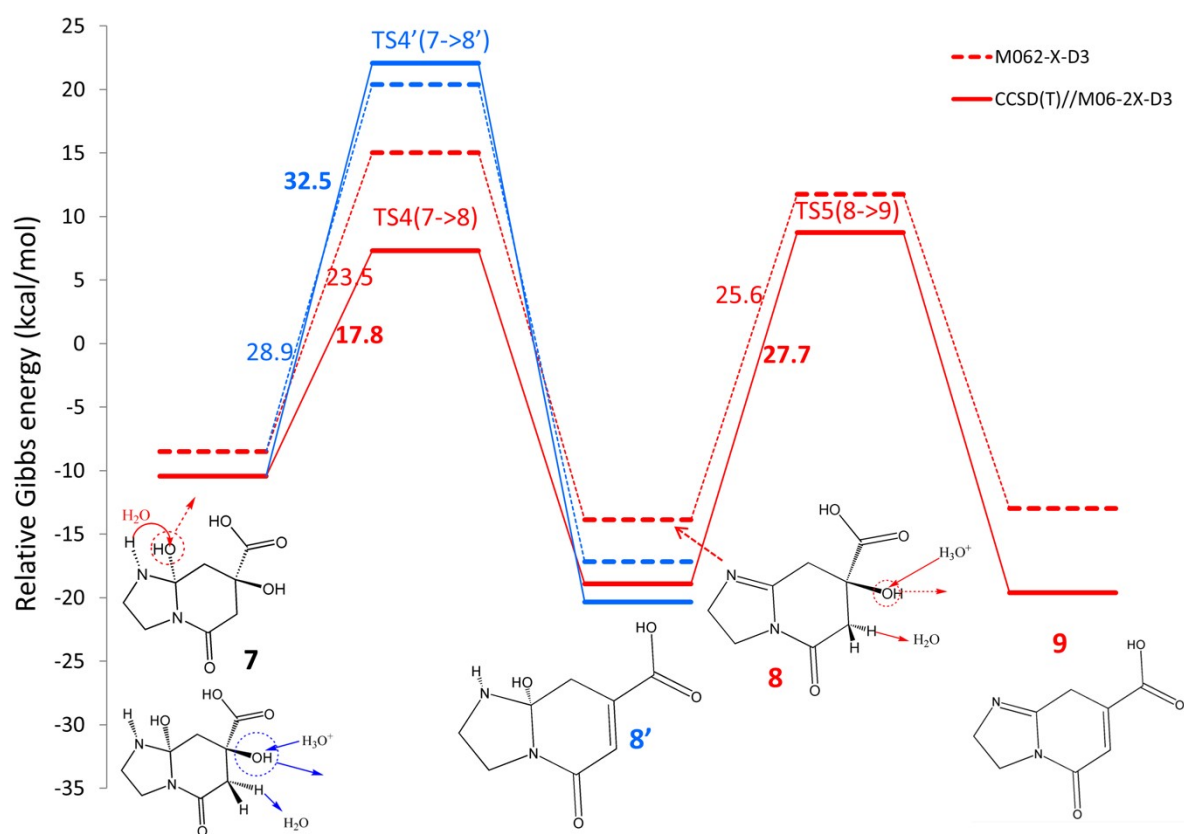


Figure S16. The Gibbs energy profile (in kcal mol⁻¹) at 180 °C for the reaction of **7** → **9**, comparison between the original path (red, same as in **Figure S2**) and alternative path (blue) starting by the elimination of the water from the six membered ring of structure **7**. The full and dashed lines represent DLPNO-CCSD(T)/def2-TZVPP//M06-2X/6-31+G(d,p) and M06-2X/6-31+G(d,p) energies, respectively. The zero energy level corresponds to the system of **1**+**2** in their CA(−) and EDA(+) forms.

The final step of IPCA synthesis is the tautomerization of **9** to form **10**. The process involves a proton transfer from $-\text{CH}_2-$ (at position 8 of imidazopyridine ring) to $-\text{N}=\text{}$ mediated by one water molecule. The tautomerization reaction proceeds via transition state **TS6** (**Figure S14**). The energy barrier for this step is $5.6 \text{ kcal mol}^{-1}$ (**Figure S13**).

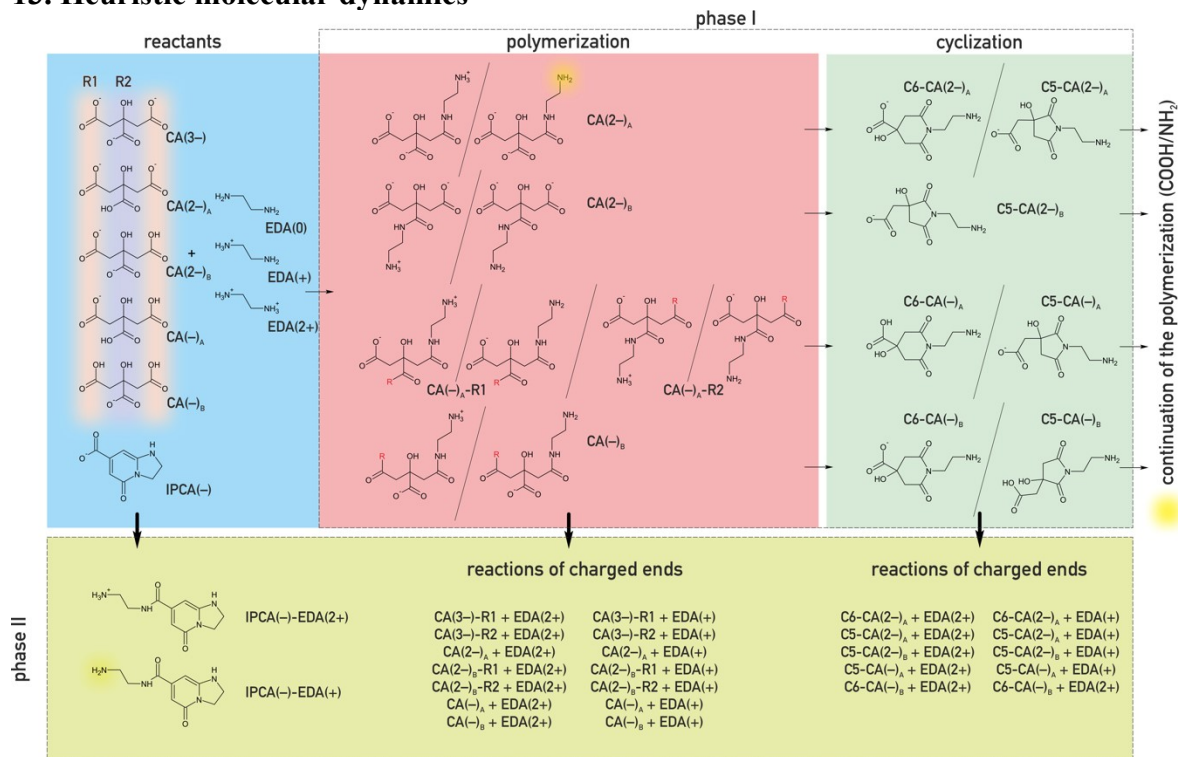
Table S2. Relative standard Gibbs energies at 180°C

System	$\Delta_r G^\circ$ (kcal mol^{-1}) CCSD(T) ^a	$\Delta_r G^\circ$ (kcal mol^{-1}) M06-2X-D3 ^b	System	$\Delta_r G^\circ$ (kcal mol^{-1}) CCSD(T) ^a	$\Delta_r G^\circ$ (kcal mol^{-1}) M06-2X-D3 ^b
1+2	0.0	0.0	6b	-7.5	-6.4
TS1a	4.8	5.7	TS3a	-1.3	2.0
3a	2.4	2.0	7a	-7.3	-4.0
TS1	20.0	17.7	TS3	5.2	5.7
3	-19.9	-18.0	7	-11.3	-8.6
3'	-17.4	-16.3	7 (rot)	-10.4	-8.5
TS2'	6.7	6.9	TS4	7.3	15.0
12	-20.6	-19.6	8	-18.9	-13.9
3_r	-18.2	-17.2	TS5	8.7	11.8
TS2	10.9	10.9	9	-19.6	-13.0
6a	-11.9	-11.5	TS6	-14.0	-10.7
TS2b	0.8	-0.6	10	-27.5	-23.5
6	-15.4	-10.3			

a Gibbs energies at DLPNO-CCSD(T)/def2-TZVPP//M06-2X-D3/6-31+G(d,p) level

b Gibbs energies at M06-2X-D3/6-31+G(d,p) level

13. Heuristic molecular dynamics



Scheme S2. Set of possible reactions that could occur within the simulated systems and their division into individual phases.

Table S3. Composition of individual systems used in our study.

Molecule	CA			EDA			IPCA	water
Protonation state	(3-)	(2-)	(-)	0	(+)	(2+)	(-)	SPC/E
System A	20	20	20	40	40	40	—	31721
System A + IPCA	20	20	20	50	50	50	30	31338
System B	—	30	30	30	30	30	—	31844
System B + IPCA	—	30	30	40	40	40	30	31459

Table S4. Number of reactions and products in the phase I depending on the composition of the system.

	System A	System A + IPCA	System A + IPCA (150 °C)	System A+ IPCA (250 °C)	System B	System B + IPCA
<i># of reactions</i>	70	72	73	77	92	105
<i># of products</i>	34	33	34	33	34	43
<i>% of reactions comprising CA(2-)</i>	42	39	40	39	38	40
<i>% of cyclic reactions</i>	21	18	20	24	17	20

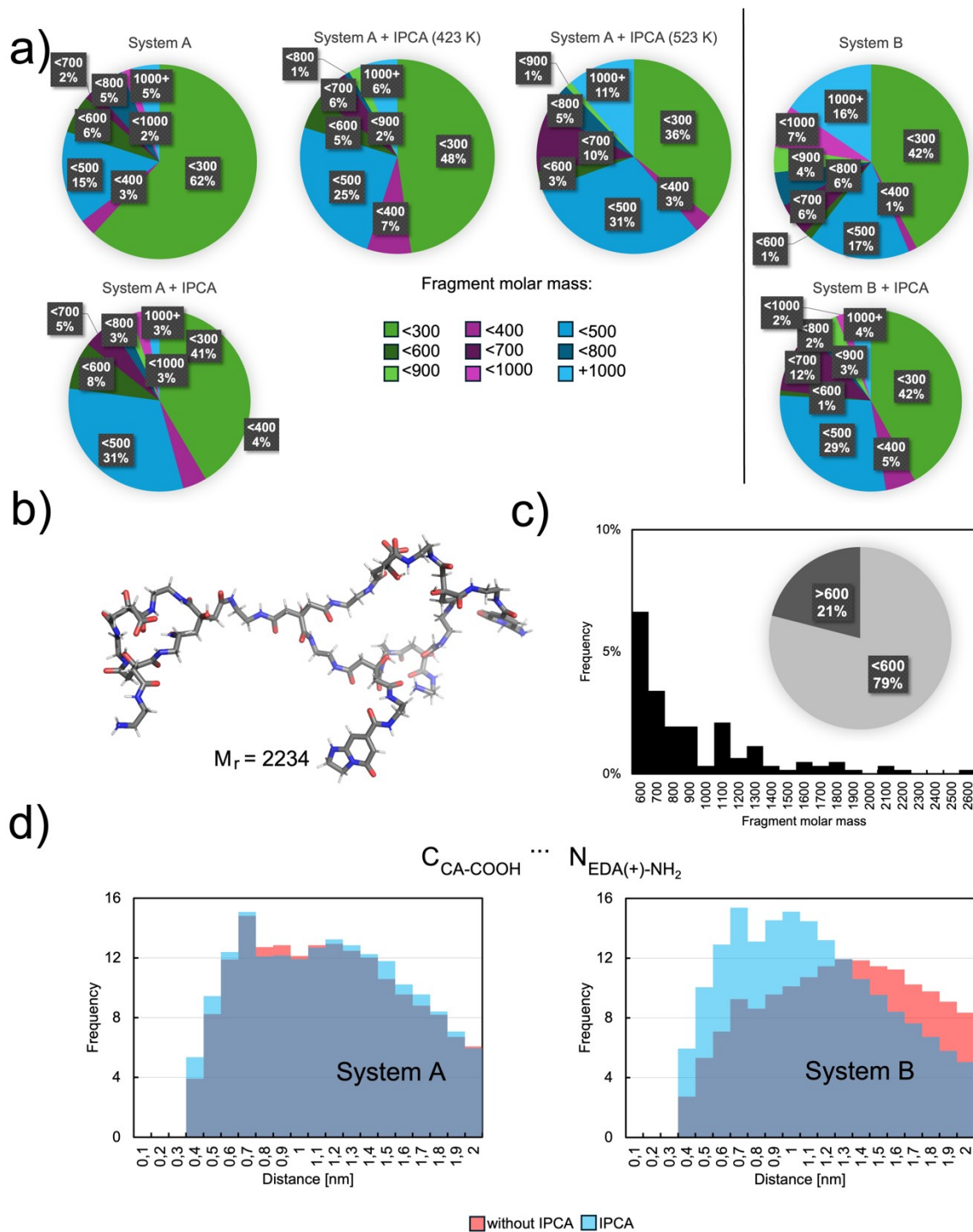


Figure S17. a) Pie charts showing the distribution by mass of the product of the System **A** and **B** from the phase II; b) Molecular structure of one of the largest formed polymers showing its extended conformation. Color code: oxygen, red; nitrogen, blue; carbon, gray; hydrogen, white; c) Histogram showing the distribution of products (with molar mass exceeding 600 g mol⁻¹) according their molar mass after the phase II; d) Minimal distance histograms between the neutral groups of any CA and EDA(+) molecules of the System **A** and **B** of the phase I. The histogram is normalized by the number of CA molecules and the simulation length.

14. Classical MD simulations

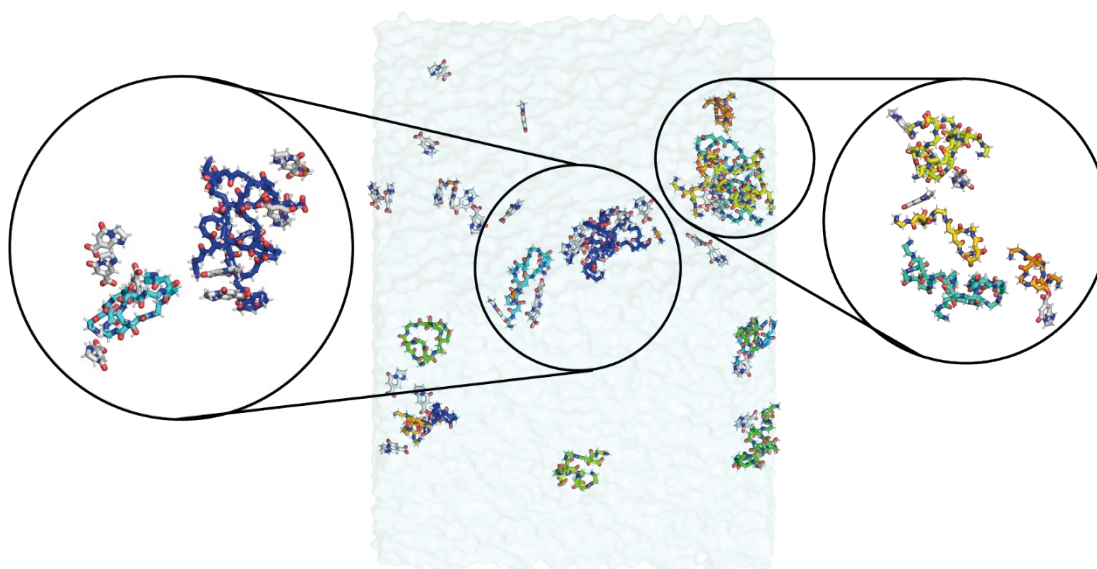


Figure S18. Snapshot from reactive simulation containing polymers and IPCA and smaller molecular fragments after 100 ns of following non-reactive MD simulation. IPCA molecules mostly interact with polymers or stack on top of each other. IPCA molecules are shown in grey, polymer carbons in colors different for each molecule. Hydrogens are shown in white, oxygens red and nitrogens blue. Solvent is shown in transparent blue surface. The insets zoom on the clusters from different point of view for better perspective, omitting water for clarity.

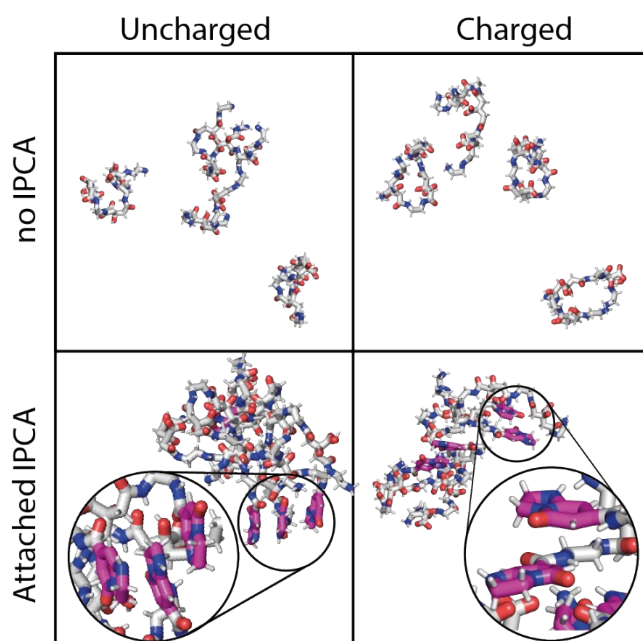


Figure S19. Clustering behavior of uncharged (left) and charged (right) polymer (700 g mol^{-1}). The polymers with attached IPCA moieties (lower panel) are more profound to cluster than without IPCA moieties (upper panel). Chosen stacked IPCAs are highlighted and zoomed. Solvent is omitted for clarity. Polymer carbons are shown in grey, IPCA carbons in magenta, hydrogens in white, oxygens red and nitrogens blue.

References

- [1] Z. Guo, Q. Zhang, H. Zou, B. Guo, J. Ni, *Anal. Chem.* **2002**, *74*, 1637.
- [2] R. Weaver, I. H. Gilbert, *Tetrahedron* **1997**, *53*, 5537.
- [3] R. Ditchfield, *Mol. Phys.* **1974**, *27*, 789.
- [4] K. Wolinski, J. F. Hinton, P. Pulay, *J. Am. Chem. Soc.* **1990**, *112*, 8251.
- [5] J. R. Cheeseman, G. W. Trucks, T. A. Keith, M. J. Frisch, *J. Chem. Phys.* **1996**, *104*, 5497.
- [6] C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158.
- [7] Gaussian 16, Revision A.03, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian, Inc., Wallingford CT, **2016**.
- [8] M. Cossi, G. Scalmani, N. Rega, V. Barone, *J. Chem. Phys.* **2002**, *117*, 43.
- [9] E. Cancès, B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *107*, 3032.
- [10] M. Cossi, V. Barone, B. Mennucci, J. Tomasi, *Chem. Phys. Lett.* **1998**, *286*, 253.
- [11] B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *106*, 5151.
- [12] K. A. Thorn, L. G. Cox, *Org. Geochem.* **2009**, *40*, 484.
- [13] W. M. Haynes, Ed. , *CRC Handbook of Chemistry and Physics*, CRC Press, **2015**.
- [14] Schrödinger Release 2022-2: Maestro, Schrödinger, LLC, New York, NY, **2022**.
- [15] A. Klamt, *WIREs Comput. Mol. Sci.* **2018**, *8*, DOI 10.1002/wcms.1338.
- [16] A. D. Bochevarov, E. Harder, T. F. Hughes, J. R. Greenwood, D. A. Braden, D. M. Philipp, D. Rinaldo, M. D. Halls, J. Zhang, R. A. Friesner, *Int. J. Quantum Chem.* **2013**, *113*, 2110.
- [17] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215.
- [18] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, DOI

10.1063/1.3382344.

- [19] M. Walker, A. J. A. Harvey, A. Sen, C. E. H. Dessent, *J. Phys. Chem. A* **2013**, *117*, 12590.
- [20] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, F. Neese, *J. Chem. Phys.* **2016**, *144*, 024109.
- [21] G. D. Purvis, R. J. Bartlett, *J. Chem. Phys.* **1982**, *76*, 1910.
- [22] K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, *Chem. Phys. Lett.* **1989**, *157*, 479.
- [23] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
- [24] F. Neese, *WIREs Comput. Mol. Sci.* **2012**, *2*, 73.
- [25] F. Neese, *WIREs Comput. Mol. Sci.* **2022**, *12*, DOI 10.1002/wcms.1606.
- [26] D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. L. Martin, F. Neese, *J. Chem. Theory Comput.* **2015**, *11*, 1525.
- [27] A. V. Marenich, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. B* **2009**, *113*, 6378.
- [28] Ö. Farkas, H. B. Schlegel, *J. Chem. Phys.* **1999**, *111*, 10806.
- [29] C. Peng, H. Bernhard Schlegel, *Isr. J. Chem.* **1993**, *33*, 449.
- [30] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701.
- [31] S. Plimpton, *J. Comput. Phys.* **1995**, *117*, 1.
- [32] J. R. Gissinger, B. D. Jensen, K. E. Wise, *Polymer (Guildf)*. **2017**, *128*, 211.
- [33] J. R. Gissinger, B. D. Jensen, K. E. Wise, *Macromolecules* **2020**, *53*, 9953.
- [34] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- [35] C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269.
- [36] Y. Song, S. Zhu, S. Zhang, Y. Fu, L. Wang, X. Zhao, B. Yang, *J. Mater. Chem. C* **2015**, *3*, 5976.
- [37] H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, *J. Phys. Chem.* **1987**, *91*, 6269.
- [38] R. W. Hockney, J. W. Eastwood, *Computer Simulation Using Particles*, Taylor & Francis, **1988**.
- [39] The PyMOL Molecular Graphics System, Version 3.0, Schrödinger, LLC.