ROYAL SOCIETY
OF CHEMISTRY

Catalytic templated length-controlled oligomerization

Bartosz Lewandowski, [ID] * Rebecca J. B. Schäfer, Etienne Cotter, [ID] Dora Harangozo [ID] and Helma Wennemers [ID] *

Templated synthesis is an intriguing strategy for the length-controlled synthesis of oligomers. Traditionally, such reactions require stoichiometric amounts of the template with respect to the product. Recently we reported catalytic macrocyclic templates that promote oligomerization of a small molecule substrate with a remarkable degree of length control. Herein we present our efforts toward creating linear templates for catalytic length-controlled oligomer synthesis.

Introduction

Nature uses templated synthesis to produce oligomers from monomer building blocks.[1] For example, genetic information is processed by transcription of DNA into RNA and subsequent translation into peptides and proteins. Scientists have engineered the natural DNA-based oligomerization machinery to allow for the preparation of complementary DNA or RNA strands[2–4] as well as the sequence- and length-controlled synthesis of non-natural oligomers.[5–10] Templated oligomerization has also been achieved with non-DNA-based synthetic templates[11] and enabled access to macrocycles,[12,13] cages[14] and even topologically more complex products.[15,16] Templating has also allowed the preparation of linear oligomers with length control.[17–19] A common limitation of templated oligomerization is the tight binding between the template and the oligomer. The newly formed oligomer is therefore only accessible in stoichiometric amounts relative to the template, and the complex between the template and the synthetic oligomer needs to be disassembled in a subsequent release step.[20] Recently, we reported the first catalytic length-controlled oligomerization.[21] Macrocyclic templates **T1–T3** catalyzed the formation of oligomers through aldol-type reactions of bifunctional substrate **1** bearing aldehyde and malonic acid half thioester moieties (Fig. 1b). The templates consist of two rigid oligoproline moieties that are functionalized with thiourea and amine catalytic sites positioned at defined distances at both sides of the macrocyclic cavity (Fig. 1a). Only 10 mol% of the template sufficed to

Laboratory of Organic Chemistry, D-CHAB, ETH Zurich, Vladimir-Prelog-Weg 3, 8093 Zurich, Switzerland.
E-mail: Bartosz.Lewandowski@org.chem.ethz.ch; Helma.Wennemers@org.chem.ethz.ch
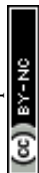
Fig. 1 (a) Catalytic macrocyclic templates **T1–T3**; (b) catalytic length-controlled oligo-merization of malonic acid half thioester (MAHT)-aldehyde bifunctional substrate **1**; and (c) correlation between the number of catalytic sites on the template with the length of the formed oligomer.

achieve full conversion of the substrate, predominantly to a single oligomer. The length of the oligomer correlated with the size of the macrocyclic catalysts (Fig. 1c). Thus, the number of monomer molecules incorporated into the olig-omer is determined by the dimensions of the macrocycle and the number of its catalytic sites.

Building on these findings, we are intrigued whether catalytic length-controlled oligomerization can also be achieved with linear templates. Such linear templates would be less complex and thus easier to access and modify, *e.g.*, by extension of the length, than macrocyclic templates. Herein we present our investigations towards the development of linear templates for catalytic length-controlled oligomer synthesis.

Results and discussion

General design of a catalytic linear oligomerization template

We envisioned that a linear template for catalytic length-controlled oligomeri-zation needs to contain a recognition site (Fig. 2, dark green) for an "initiator" building block and activation sites (Fig. 2, dark blue) that allow for the activation of the initiator and incoming "propagator" building blocks. The initiator should contain a binding site (Fig. 2, green) that binds to the recognition site of the
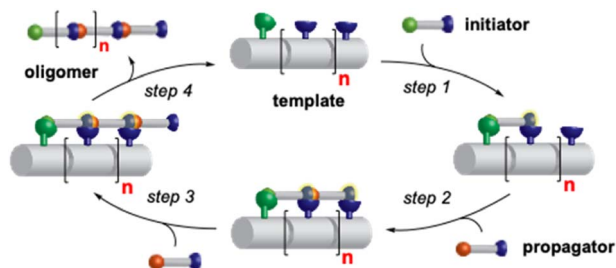
Fig. 2 Concept of catalytic length-controlled oligomerization with a linear synthetic template.

template. Additionally, the initiator needs a functional group (Fig. 2, blue) that can be activated by the activation site on the template for reaction with the reactive site of the propagator. The propagator should contain the reactive site (Fig. 2, brown) as well as the same activatable functional group as the initiator. Upon binding of the initiator to the template (step 1), its other functional group will become activated by the activating site of the template and react with the reactive site on the propagator (step 2). This reaction will bring the activatable functionality of the propagator into the vicinity of the next activation site on the template. The template-bound propagator will then become activated, and react with the next propagator molecule (step 3). These propagating reactions will continue until the last activation site at the end of the template is reached. Consequently, the length of the formed oligomer will correlate with the number of activation sites on the template.

For length-controlled oligomerization, the "initiator" and the growing oligomer should only react with the "propagator" when bound and activated by the template. For the controlled activation of the reactants, the distances between the recognition and activation sites on the template need to match those of the functional groups of the initiator and the propagator. For catalytic turn-over, the template needs to bind less tightly to the oligomer than to the monomeric building blocks. These prerequisites are likely best fulfilled by rigid templates that can be functionalized with recognition and catalytic sites at geometrically well-defined positions. Ideal templates are composed of modular components that allow for facile length and functional group variations. Oligoprolines fulfil these requirements as they already adopt, at a chain-length of six residues, a polyproline II (PPII) helix, in which every third residue is located at the same face of the helix at a distance of ~9 Å (Fig. 3).[22,23] These peptides are accessible with different lengths by standard solid-phase peptide synthesis (SPPS), are
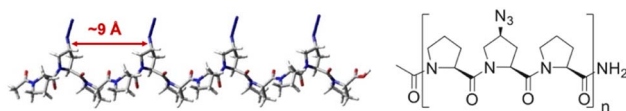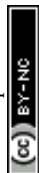

Fig. 3 Model of a PPII-helical oligoproline 12-mer (left) and general structure of oligoprolines with (2S,4S)-Azp residues in every third position (right).

soluble in aqueous as well as organic solvents, and can be derivatized if γ-azi-doproline (Azp) residues are used as building blocks.[23] Thus, oligoprolines allow for positioning recognition and activation sites on the same side of the helical scaffold.

Design of specific initiator, propagator, and template

Building on our catalytic length-controlled oligomerization with macrocyclic templates[21] and expertise with conjugate addition reactions using nitroolefins,[24–26] we started to explore addition reactions between β-ketoesters and functionalized nitroolefins with *p*-(carboxy)-nitrostyrene (**I**) as initiator and *p*-(β-ketoester)-nitrostyrene (**P**) as propagator (Fig. 4a). We started with the β-ketoester as a donor moiety *en route* to β-ketoacids since the latter can undergo non-productive decarboxylation and require careful control of the reaction condi-tions.[24,25] As linear catalytic templates, we envisioned oligoproline 6-mer T_L1 and 9-mer T_L2, each functionalized with a guanidinium moiety at C^γ of the second Pro residue for coordination with the carboxylic acid of the initiator and thiourea moieties for coordination and activation of the nitroolefin (Fig. 4b). Within propagator **P**, the distance between the nitro group and the β-ketoester is approximately 9 Å, thus, matching the distance between functional groups installed at every third residue of the oligoproline template.

Catalytic turnover requires (1) binding of the initiator to the template and (2) a lower binding affinity of the oligomer to the template compared to the initiator and propagator. The carboxylic acid moiety of the initiator **I** should ensure binding to the templates T_L1 and T_L2. Product release is more challenging but should be feasible since the conformation of the oligomer differs significantly from that of the initiator and propagator. The monomeric building blocks are predominantly flat with sp^2-hybridized carbons, whereas the oligomer contains sp^3-hybridized carbons and, therefore, tetrahedral moieties. It is noteworthy that the propagator is predominantly planar due to enolization (Fig. 4). Since the geometry of the templates is tailored to accommodate that of the initiator, the

Fig. 4 (a) Structures and calculated lowest energy structures of the initiator I, propagator P, and oligomers; (b) templates T_L1 and T_L2; and (c) plausible binding and activation of initiator I by template T_L1.
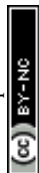
oligomer should have a lower binding affinity to the template compared to the monomers.

Synthesis of initiator, propagator, and template

Templates T_L1 and T_L2 were prepared by SPPS using established procedures.[21] Initiator **I** was prepared from 4-carboxybenzaldehyde *via* a Henry reaction followed by dehydration. Propagator **P** was also readily available from the same starting material in 5 steps.

Conformation and binding affinity studies

Circular dichroism (CD) spectra of templates T_L1 and T_L2 show the typical signature of PPII helicity; thus, the templates adopt the expected helical structure. The binding of the initiator to the template is a prerequisite for templated oligomerization. Since the binding affinity between a carboxylic acid and a guanidinium group depends on the pH and the solvent,[27] we determined the binding affinity of the complex between **I** and T_L1 in different environments. Isothermal titration calorimetry (ITC) studies in methanol revealed an association constant $K_a = 8201 \pm 600$ M^{-1}. In our previous studies with the macrocyclic template, the interaction between the template and the substrate was significantly weaker ($K_a \sim 20$ M^{-1}), which agrees with other interactions observed in organocatalytic reactions.[21,28] Thus, the observed association constant is rather high and likely too strong to allow for catalytic turnover in the templated oligomerization reaction. In a methanol/water mixture, the binding is, as expected, weaker, $K_a = 1187 \pm 36$ M^{-1}, as revealed by titration monitored with ^{1}H NMR spectroscopy (CD$_3$OD/D$_2$O = 9:1). (Note, we used NMR spectroscopic monitoring in this case since large solvent-mixing heats can override the effects arising from host–guest binding.)

Templated oligomerization

Upon mixing **I** and **P** in a 1:3 ratio and adding 0.1 equivalents of T_L1 (10 mol%) in methanol (Fig. 5a), hardly any product formation was observed after 48 hours (Fig. 5b, top). In contrast, in a mixture of methanol and water (CD$_3$OD/D$_2$O 9:1) the expected dimer O_L1 formed in approximately 50% yield (Fig. 5b, bottom). This experiment showed that template T_L1 served as a catalyst and corroborated the basic design of the catalytic oligomerization. Over the course of the reaction, O_L1 precipitated, indicating poor solubility.

Next, we performed the oligomerization reaction under the same conditions using template T_L2 in catalytic amounts. Conversion of the initiator and the propagator took place (~50%) along with the formation of a precipitate. LC-MS analysis of the precipitate revealed that it consisted exclusively of dimer O_L1. Trimer O_L2 was not detected in the precipitate or the reaction mixture. This finding suggests that the initially formed dimer precipitates before a second template-controlled reaction with another molecule of the propagator **P** to form trimer O_L2 can take place. Thus, the results are not conclusive as to whether the current template design allows for length-controlled oligomerization.†

---

† Aside from precipitation, the oligomerization could also stop due to a non-favorable binding geometry between the dimer and the template.
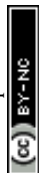
Fig. 5 (a) Catalytic templated oligomerization with T_L1; (b) ^1H NMR spectra (selected region) of the reaction in CD_3OD (top) and CD_3OD/D_2O 9 : 1 (bottom) recorded after 24 h – signals corresponding to the product are labelled with a green dot; and (c) LC-MS analysis of the precipitate formed in the reaction in CD_3OD/D_2O 9 : 1 (inset: ESI$^-$ spectrum of the peak at 3.67 min).

Conclusions

We proposed here linear templates for catalytic length-controlled syntheses of oligomers. The templates are based on rigid oligoproline scaffolds and contain one binding site and one or more activation sites located at precise distances along the scaffold. The templates are designed to bind bifunctional substrates, an initiator and propagators, and control the length of the arising oligomer by the length of the template and the number of activation sites thereon. The work provided insights into how molecular recognition and catalysis can be harnessed to achieve length-controlled assembly of oligomers from monomeric building blocks.

A catalytically active template for dimer formation has been established, but poor solubility of the dimer has prevented elongation to longer oligomers. Future studies will therefore focus on fine-tuning the structure of the template and the substrates, the solvent, and other reaction parameters, such as the concentration and the temperature. These will include the use of weaker coordinating groups on the template for binding to the initiator (e.g., amino instead of guanidinium groups), and variations of the substrates (e.g., use of MAHTs instead of β-ketoesters, and use of aldehyde instead of nitro groups). Analysis of the thermodynamic parameters will provide insight into the roles of enthalpy and entropy in the oligomerization reaction.

Materials and instruments

Solvents and reagents were of the highest commercially available grade and used without further purification. They were purchased from Sigma Aldrich, Fisher

Scientific, Fluka, Bachem, BioMatrix, Biotage, IRIS Biotech, Protein Technologies and Acros Organics. Solvents used for HPLC were HPLC-grade quality. Water used for HPLC purifications and Staudinger reactions was filtered in a Milli-Q system (Millipore).

Solid-phase peptide synthesis (SPPS)

SPPS was performed using the Fmoc-strategy and Rink amide-PS as resin. Amino acids and coupling reagents were purchased from Bachem, IRIS Biotech and Protein Technologies. For automated peptide synthesis, a SYRO Robot peptide synthesizer (Biotage, Sweden) controlled by the SyroXP peptide software version 2.0.126 (MultiSynTech GmbH, Germany) was used.

Preparative reverse phase high-pressure liquid chromatography (RP-HPLC)

Purifications of the oligoprolines were carried out on a Semiprep UltiMate 3000 chromatography system (Dionex, USA), using a Reprosil gold C18 (150×16 mm) column with a flow of 6 mL min^{-1}. The system was controlled using the Chromeleon software version 6.80. Two different solvents were used. Solvent A was HPLC-grade ACN and solvent B was $1000:10:1$ $H_2O:ACN:TFA$.

Preparative medium-pressure liquid chromatography (MPLC)

Purifications of the building blocks and bifunctional substrates were carried out on a CombiFlash EZ Prep flash chromatography system (Teledyne ISCO). Two different solvent systems were used: for the building blocks for peptide functionalization, solvent A was HPLC-grade CH_2Cl_2 without a stabilizer and solvent B was HPLC-grade methanol. For the bifunctional substrates, solvent A was HPLC-grade hexane and solvent B was HPLC-grade EtOAc.

Analytical reverse phase high-pressure liquid chromatography (RP-HPLC)

Reverse phase HPLC (RP-HPLC) analysis of the oligoprolines was performed on a Dionex UHPLC, UltiMate 3000. A Reprosil gold 120 C18 (150×4 mm, 5 μm) analytical column with a flow of 1.0 mL min^{-1} was used. The analyses were performed using a two-solvent system. Solvent A was HPLC-grade acetonitrile and solvent B was a mixture of 1% acetonitrile and 0.1% TFA in Milli-Q pure water.

Lyophilization

An Alpha 2-4 LD plus lyophilizer (Christ, Germany) was used.

Thin-layer chromatography (TLC)

TLC was conducted on aluminum sheets coated with silica gel 60 F254 (Merck) using UV fluorescence (254 and 366 nm). Analytical grade solvents were used.

Liquid chromatography-mass spectrometry (LC-MS)

Analytical reverse phase HPLC was performed on an Agilent 1260 Infinity II Prime LC System. An Agilent Eclipse Plus C18 (150×4.6 mm, 5 μm) column with a flow of 0.6 mL min^{-1} was used. Two different solvents were used. Solvent A was

a mixture of 1% formic acid in Milli-Q pure water, and solvent B was a mixture of 1% formic acid in acetonitrile. The gradient used for the HPLC analyses of the templates was 95% A to 5% A over 11 minutes. The mass analysis was performed on a maXis ESI-QTOF spectrometer (Bruker, USA). The obtained mass values are listed as the ratio of atom mass per charge (m/z).

Nuclear magnetic resonance (NMR) spectroscopy

1D and 2D NMR spectra were recorded on 400 and 500 MHz Ultrashield spectrometers (Bruker, USA). ^{1}H-NMR chemical shifts (δ_H) are quoted in parts per million (ppm) downfield from TMS and coupling constants (J) are quoted in Hertz (Hz). Abbreviations for NMR data are s (singlet), d (doublet), t (triplet), q (quartet), and m (multiplet).

MALDI-TOF (matrix-assisted laser desorption/ionization–time-of-flight) mass spectrometry

MALDI-TOF analyses were carried out using a Bruker SolariX 94 spectrometer (Bruker, USA). α-Cyano-4-hydroxycinnamic acid (10 mg mL^{-1} in ACN/H$_2$O 1 : 1 and 1 μL TFA) was used as the matrix. The obtained mass values are listed as the ratio of atom mass per charge (m/z).

High-resolution mass spectrometry (HR-MS)

HR-MS analyses were performed by the Molecular and Biomolecular Analysis Service of ETH Zurich (MOBIAS). High-resolution electrospray ionization (HR-ESI) spectra were measured on a Bruker maXis spectrometer.

Circular dichroism spectroscopy (CD)

CD spectroscopic analyses were carried out with a Chirascan™ Plus (Applied Photophysics Ltd, Leatherhead, UK). Quartz cuvettes with a path length of 1.0 mm (Hellma 110-QS) were used. The samples were prepared at 200 μM concentration in MeOH as a solvent.

### Isothermal titration calorimetry (ITC)

ITC measurements were carried out on a MicroCal PEAQ-ITC (Malvern Panalytical Ltd, Malvern, UK). The raw calorimetry data were analyzed using the analysis software provided by Malvern Panalytical Ltd.

Synthesis

Linear templates T$_L$1 and T$_L$2

The peptides were prepared following previously reported procedures[21] using Fmoc-Pro-OH, Fmoc-Azp-OH (Azp)[29] and Fmoc-di-(Boc)-Gup-OH (Gup)[30,31] as individual amino-acid building blocks.
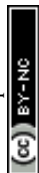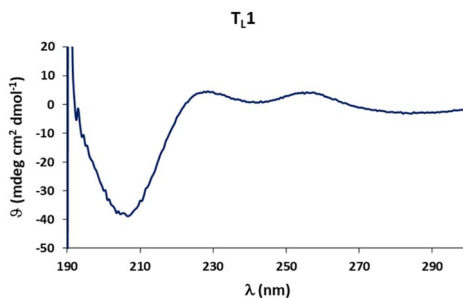
6-mer template (T$_L$1)
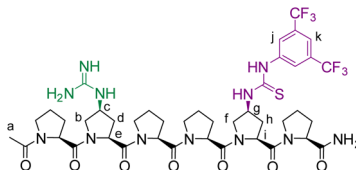
^{1}H NMR (400 MHz, methanol-d_4). δ 8.23 (s, 2H, H$_j$), 7.67 (s, 1H, H$_k$), 5.36–5.10 (m, 1H, H$_g$), 4.90–4.82 (u – partial overlap with solvent peak, 2H, H$_\alpha$ + H$_i$), 4.76 (dd, J = 8.6, 4.4 Hz, 1H, H$_\alpha$), 4.66 (td, J = 8.7, 4.0 Hz, 2H, H$_c$ + H$_e$), 4.45 (dd, J = 8.4, 5.4 Hz, 1H, H$_\alpha$), 4.32 (q, J = 5.5 Hz, 1H, H$_\alpha$), 4.21 (ddd, J = 15.2, 10.7, 6.2 Hz, 2H, H$_b$ + H$_f$), 3.96–3.51 (m, 10H, 8 × H$_\gamma$ + H$_{b'}$ + H$_{f'}$), 2.80–2.63 (m, 2H, H$_d$ + H$_h$), 2.41–2.22 (m, 4H, 4 × H$_\beta$), 2.16–1.88 (m, 6H, 4 × H$_\beta$ + H$_{d'}$ + H$_{h'}$), 2.10 (s, 3H, H$_a$).
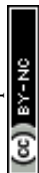
HPLC. t_R = 8.8 min; Reprosil gold 120 C$_{18}$ (150 × 4 mm, 5 μm); gradient 80% to 40% B over 30 min at 50 °C (PDA @ 214 nm), flow 1.0 mL min^{-1}.

HR-MALDI-MS: m/z: 985.3927 [M + H]$^+$, calculated for C$_{42}$H$_{55}$F$_6$N$_{12}$O$_7$S$^+$: 985.3936.

CD.

9-mer template (T$_L$2)
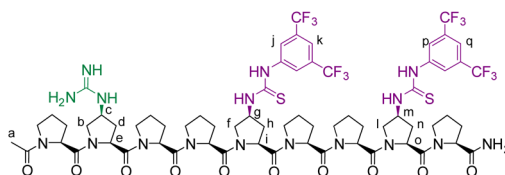
^{1}H NMR (400 MHz, methanol-d_4). δ 8.18 (s, 4H, H$_j$ + H$_p$), 7.64 (s, 2H, H$_k$ + H$_q$), 5.13 (bs, 2H, H$_g$ + H$_m$), 4.90–4.82 (u – partial overlap with solvent peak, 3H, H$_\alpha$ + H$_i$ + H$_o$), 4.77–4.54 (m, 4H, 4 × H$_\alpha$), 4.43 (dd, J = 8.4, 5.3 Hz, 1H, H$_\alpha$), 4.39–4.26 (m, 1H, H$_e$), 4.22 (dd, J = 10.5, 6.6 Hz, 2H, H$_f$ + H$_l$), 3.92–3.75 (m, 4H, 3 × H$_\delta$ + H$_c$), 3.75–3.48 (m, 13H, 9 × H$_\delta$, H$_b$ + H$_f$ + H$_{l'}$), 2.78–2.57 (m, 3H, H$_d$ + H$_h$ + H$_n$), 2.43–2.20 (m, 6H, 6 × H$_\beta$), 2.18–1.83 (m, 9H, 6 × H$_b$ + H$_{d'}$ + H$_{h'}$ + H$_{n'}$), 2.10 (s, 3H, H$_a$).
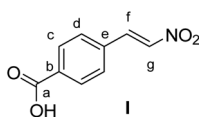
HPLC. t_R = 19.4 min; Reprosil gold 120 C_{18} (150 × 4 mm, 5 μm); gradient 85% to 30% B over 30 min at 50 °C (PDA @ 214 nm), flow 1.0 mL min^{-1}.

HR-MALDI-MS. *m/z*: 1562.5505 [M + H]$^+$, calculated for $C_{66}H_{80}F_{12}N_{17}O_{10}S_2{}^+$: 1562.5518.

CD.

T_L2

Initiator I

4-Carboxybenzaldehyde (0.3 g, 2.0 mmol, 1 equiv.) and NH$_4$OAc (0.385 g, 5 mmol, 2.5 equiv.) were dissolved in AcOH (8 mL). CH$_3$NO$_2$ (4 mL) was then added. The reaction mixture was stirred under reflux for 1 h. The reaction mixture was poured into a beaker containing water (15 mL). The aqueous mixture was extracted with EtOAc (3 × 15 mL). The combined organic extracts were washed with brine, dried over MgSO$_4$, filtered and the solvent was removed under reduced pressure. The crude product was purified by recrystallization from MeOH. 4-(E)-nitrostyrene-benzoic acid **I3** was obtained as yellow powder (0.34 g, 1.76 mmol, 88%).

^{1}H NMR (400 MHz, DMSO-d_6). δ 8.31 (d, *J* = 13.7 Hz, 1H, H$_f$), 8.19 (d, *J* = 13.7 Hz, 1H, H$_g$), 8.04–7.87 (m, 4H, H$_c$ + H$_d$).

^{13}C NMR (101 MHz, DMSO-d_6). δ 167.1 (C$_a$), 140.2 (C$_f$), 138.3 (C$_g$), 134.9 (C$_d$), 133.8 (C$_b$), 130.3, 130.3 (C$_c$ + C$_d$).

HR-ESI-MS. *m/z*: 216.0281 [M + Na$^+$], calculated for $C_9H_7NNaO_4$: 216.0273.
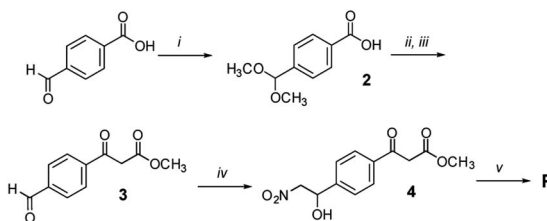
Propagator P (Fig. 6)

Fig. 6 Synthesis of propagator **P**: (i) CH$_3$OH, NH$_4$Cl, reflux, o/n, 75%; (ii) Meldrum's acid, EDC × HCl, DMAP, CH$_2$Cl$_2$, r. t., 48 h, 70%; (iii) 1 M HCl, THF, 40 °C, 2 h, 95%; (iv) CH$_3$NO$_2$, NaOH$_{aq.}$, CH$_3$OH, r. t., 65%; (v) Ac$_2$O, pyridine, CH$_2$Cl$_2$, 0 °C, 2 h, 75%.

4-(Dimethoxymethyl)benzoic acid (2).

4-Formyl benzoic acid (500 mg, 3.3 mmol, 1 equiv.) and NH_4Cl (1 g, 18.7 mmol, 5.7 equiv.) were suspended in dry MeOH (10 mL) in a 20 mL microwave vial. The vial was sealed and the mixture was heated to reflux for 36 h, upon which the solution became clear. The solvent was then removed under reduced pressure and the solid residue was crystallized from hexane to yield **2** as glassy crystals (463 mg, 2.4 mmol, 71% yield).

^{1}H NMR (400 MHz, DMSO-d_6). δ 7.98–7.94 (m, 2H, H_e), 7.54–7.47 (m, 2H, H_d), 5.45 (s, 1H, H_b), 3.26 (s, 6H, H_a).

^{13}C NMR (101 MHz, DMSO-d_6). δ 167.5 (C_g), 143.3 (C_f), 131.2 (C_c), 129.7 (C_e), 127.2 (C_d), 102.5 (C_b), 53.1 (C_a).

HR-ESI-MS. m/z: 219.0639 [M + Na$^+$], calculated for $C_{10}H_{12}NaO_4$: 219.0633.

Methyl 3-(4-formylphenyl)-3-oxopropanoate (3)
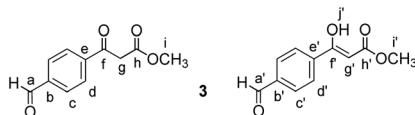
Compound **2** (450 mg, 2.3 mmol, 1 equiv.) was dissolved in dry CH_2Cl_2 (3 mL). Meldrum's acid (397 mg, 3.45 mmol, 1.5 equiv.) followed by DMAP (220 mg, 2.75 mmol, 1.2 equiv.) and EDC × HCl (660 mg, 3.45 mmol, 1.5 equiv.) were added. After 22 hours, the reaction was quenched by addition of water (5 mL) and diluted with CH_2Cl_2 (7 mL). The layers were separated. The aqueous layer was extracted with CH_2Cl_2 (6 × 5 mL). The combined organic extracts were washed with 1 N HCl aqueous solution (2 × 5 mL) and brine (2 × 5 mL), dried over $MgSO_4$, filtered and the solvent was evaporated under reduced pressure. The crude product (515 mg, 1.6 mmol, 70% yield) was immediately used in the next step without purification. It was dissolved in dry MeOH (10 mL) and then refluxed under an Ar atmosphere for 2 h. Afterwards MeOH was removed under reduced pressure and the crude residue was dissolved in a THF : H_2O mixture (4 : 1, 5 mL). 2 N aqueous HCl solution was added (1 mL) and the mixture was stirred on a rotary evaporator bath for 45 minutes at 40 °C. Then it was cooled down to room temperature and CH_2Cl_2 (10 mL) and H_2O (10 mL) were added. The layers were separated and the aqueous one was extracted with CH_2Cl_2 (3 × 5 mL). The combined organic extracts were washed with brine (5 mL), dried over $MgSO_4$, filtered and the solvent was removed under reduced pressure. The crude product was purified by MPLC (hexane : EtOAc, gradient 0% to 20% EtOAc in 15 min). Compound **3** was obtained as a colourless oil (313 mg, 1.52 mmol, 95% yield).
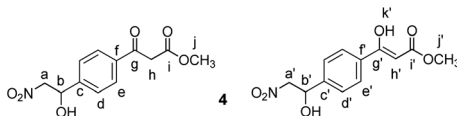
^{1}H NMR (300 MHz, CDCl$_3$, a 2 : 1 mixture of the keto and enol forms). δ 12.47 (s, 0.33H, $H_{j'}$), 10.12 (s, 0.67H, H_a), 10.06 (s, 0.33H, $H_{a'}$), 8.10 (d, J = 8.2 Hz, 1.3H,

H$_c$), 8.04–7.97 (m, 1.3H, H$_d$), 7.94 (d, J = 0.9 Hz, 1.3H, H$_{c'}$ + H$_{d'}$), 5.77 (s, 0.33H, H$_{g'}$), 4.05 (s, 1.3H, H$_g$), 3.83 (s, 1H, H$_{i'}$), 3.77 (s, 2H, H$_i$).

HR-ESI-MS. m/z: 229.0484 [M + Na$^+$], calculated for C$_{11}$H$_{10}$NaO$_4$: 229.0477.

Methyl 3-(4-(1-hydroxy-2-nitroethyl)phenyl)-3-oxopropanoate (4)


Nitromethane (27 µL, 0.4 mmol, 1.5 equiv.) was dissolved in MeOH (0.5 mL). 10 M aqueous NaOH (45 µL, 0.34 mmol, 1.3 equiv.) was added to the solution. The mixture was cooled to 0 °C using an ice bath. Then a solution of compound **3** (70 mg, 0.26 mmol, 1 equiv.) in MeOH (0.5 mL) was added dropwise. The reaction was stirred at 0 °C for 30 minutes and then allowed to warm to room temperature. Subsequently it was stirred at room temperature for 2 h. A few drops of AcOH, followed by H$_2$O (5 mL) and CH$_2$Cl$_2$ (5 mL) were added and the layers separated. The aqueous layer was extracted with CH$_2$Cl$_2$ (4 × 5 mL). The combined organic extracts were washed with brine (5 mL), dried over MgSO$_4$, filtered and the solvents evaporated under reduced pressure. The crude mixture was separated by MPLC (hexane : ethyl acetate, gradient 0% to 40% over 15 minutes) to yield **4** as a pale yellow solid (53 mg, 0.17 mmol, 65%).

^{1}H NMR (300 MHz, CDCl$_3$, an 8 : 1 mixture of the keto and enol forms). δ 12.46 (s, 0.11H, H$_{k'}$), 8.02–7.84 (m, 1.78H, H$_f$), 7.79 (d, J = 8.2 Hz, 0.22H, H$_f$), 7.59–7.48 (m, 1.78H, H$_d$), 7.46 (d, J = 8.3 Hz, 0.22H, H$_d$), 5.67 (s, 0.11H, H$_{h'}$), 5.58–5.49 (m, 1H, H$_a$ + H$_{a'}$), 4.74–4.35 (m, 2H, H$_b$ + H$_{b'}$), 3.98 (s, 1.8H, H$_h$), 3.80 (s, 0.33H, H$_{j'}$), 3.74 (s, 2.67H, H$_j$).

HR-ESI-MS. m/z: 290.0634 [M + Na$^+$], calculated for C$_{12}$H$_{13}$NNaO$_6$: 290.0641.

Methyl (E)-3-(4-(2-nitrovinyl)phenyl)-3-oxopropanoate (P)
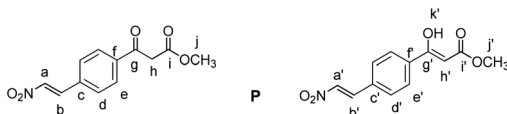
Compound **4** (40 mg, 0.15 mmol, 1 equiv.) was dissolved in CH$_2$Cl$_2$ (1 mL). The solution was cooled down to 0 °C. Subsequently, dry pyridine (12 µL, 0.15 mmol, 1 equiv.) and acetic anhydride (16.5 µL, 0.15 mmol, 1 equiv.) were added. The mixture was allowed to warm to room temperature and was then stirred for 4 hours. Next, CH$_2$Cl$_2$ (2 mL) and 0.1 M aqueous HCl (1 mL) were added and the layers were separated. The aqueous layer was extracted with CH$_2$Cl$_2$ (3 × 2 mL). The combined organic layers were dried over MgSO$_4$, filtered and the solvent was evaporated under reduced pressure. The crude mixture was separated by MPLC (hexane : ethyl acetate, gradient 10% to 20% over 20 minutes) to yield **P** as a yellow solid (28 mg, 0.113 mmol, 75%).

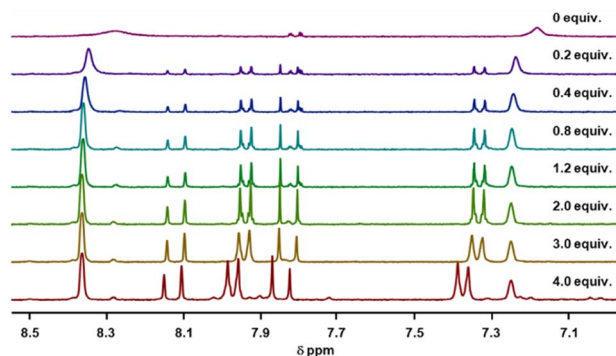**Fig. 7** ITC analysis of the interaction between I and T_L1 in CH_3OH: (a) raw data; (b) fitted curve.

Fig. 8 Stacked ^1H NMR spectra of T_L1 ($CD_3OD : D_2O = 9 : 1$) upon titration of increasing portions of I.

^1H NMR (400 MHz, $CDCl_3$, a 3 : 2 mixture of the keto and enol forms). δ 12.50 (s, 0.4H, H_k), 8.07–8.01 (m, 2.2H, $H_a + H_{a'} + H_d$), 7.88 (d, $J = 8.5$ Hz, 0.8H, $H_{d'}$), 7.72–7.65 (m, 1.8H, $H_b + H_e$), 7.64–7.58 (m, 1.2H, $H_{b'} + H_{e'}$), 5.76 (s, 0.4H, $H_{h'}$), 4.05 (s, 1.2H, H_h), 3.85 (s, 1.2H, $H_{j'}$), 3.79 (s, 1.8H, H_j).

^{13}C NMR (101 MHz, $CDCl_3$). δ 191.4 (C_g), 173.2 ($C_{i'}$), 169.5 ($C_{g'}$), 167.5 (C_i), 139.1, 138.1 ($2 \times C$, $C_b + C_{b'}$), 138.1, 137.8, 137.1, 136.7 ($4 \times C$, $C_b + C_{b'} + C_e + C_{e'}$), 134.9 ($C_f$), 132.6 ($C_{f'}$), 129.4, 129.4 ($2 \times C$, $C_a + C_{a'}$), 129.2 (C_d), 127.0 ($C_{d'}$), 88.6 ($C_{h'}$), 52.7 (C_j), 51.7 ($C_{j'}$), 45.8 (C_h).

HR-ESI-MS. m/z: 272.0542 $[M + Na^+]$, calculated for $C_{12}H_{11}NNaO_5$: 272.0535.

Binding affinity measurements

ITC experiments. The experiments were performed in methanol at room temperature. The concentrations of the template T_L1 and the initiator I were 1 mM and 30 mM, respectively. The solution of the guest was added in 1.5 μL portions to the solution of the host. The obtained data was fitted to a 'One set of sites' model in the Malvern ITC Analysis software (Fig. 7).

Titrations monitored by ^1H NMR spectroscopy. The titration experiments were carried out in $CD_3OD : D_2O = 9 : 1$, at room temperature. The concentrations of
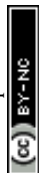
Fig. 9 Titration curve correlating the changes in the chemical shifts of aromatic protons in T_L1 with the changes in the concentration of **I** during the titration (left); K_d values obtained for both fits and the average K_a value (right).

the host (**I**) and the guest (**T_L1**) solutions were 6 mM and 300 mM, respectively (Fig. 8). The data were fitted to a single-site non-competitive binding equation using Prism ver.9.2.0 (GraphPad) (Fig. 9).

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 J. M. Berg and J. L. Tymoczko, *Biochemistry*, W. H. Freeman, New York, 2006.
2 U. Gubler and B. J. Hoffman, *Gene*, 1983, **25**, 263–269.
3 J. Tian, H. Gong, N. Sheng, X. Zhou, E. Gulari, X. Gao and G. Church, *Nature*, 2004, **432**, 1050–1054.
4 N. K. Vaish, A. W. Fraley, J. W. Szostak and L. W. McLaughlin, *Nucleic Acids Res.*, 2000, **28**, 3316–3322.
5 Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder and D. R. Liu, *Science*, 2004, **305**, 1601–1605.
6 S. Liao and N. C. Seeman, *Science*, 2004, **306**, 2072–2074.
7 K. Josephson, M. C. T. Hartman and J. W. Szostak, *J. Am. Chem. Soc.*, 2005, **127**, 11727–11735.
8 Y. Ura, J. M. Beierle, L. J. Leman, L. E. Orgel and M. R. Ghadiri, *Science*, 2009, **325**, 73–77.
9 J. Niu, R. Hili and D. R. Liu, *Nat. Chem.*, 2013, **5**, 282–292.
10 W. Meng, R. A. Muscat, M. L. McKee, P. J. Milnes, A. H. El-Sagheer, J. Bath, B. G. Davis, T. Brown, R. K. O'Reilly and A. J. Turberfield, *Nat. Chem.*, 2016, **8**, 542–548.

11 F. Diederich and P. J. Stang, *Templated Organic Synthesis*, Wiley-VCH, Weinheim, 2000.

12 M. C. O'Sullivan, J. K. Sprafke, D. V. Kondratuk, C. Rinfay, T. D. W. Claridge, A. Saywell, M. O Blunt, J. N. O'Shea, P. H. Beton, M. Malfois and H. Anderson, *Nature*, 2011, **469**, 72–75.

13 J. M. A. Carnall, C. A. Waudby, A. M. Belenguer, M. C. A. Stuart, J. J.-P. Peyralans and S. Otto, *Science*, 2010, **327**, 1502–1506.

14 J. Cremers, R. Haver, M. Rickhaus, J. Q. Gong, L. Favereau, M. D. Peeks, T. D. W. Claridge, L. M. Herz and H. L Anderson, *J. Am. Chem. Soc.*, 2018, **140**, 5352–5355.

15 Y. Segawa, M. Kuwayama, Y. Hijikata, M. Fushimi, T. Nishihara, J. Pirillo, J. Shirasaki, N. Kubota and K. Itami, *Science*, 2019, **365**, 272–276.

16 Z. Ashbridge, E. Kreidt, L. Pirvu, F. Schaufelberger, J. H. Stenlid, F. Abild-Petersen and D. A. Leigh, *Science*, 2022, **375**, 1035–1041.

17 K. S. Feldman and Y. B. J. Lee, *J. Am. Chem. Soc.*, 1987, **109**, 5850–5851.

18 B. Lewandowski, G. De Bo, J. W. Ward, M. Papmeyer, S. Kuschel, M. J. Aldegunde, P. M. E. Gramlich, D. Heckmann, S. M. Goldup, D. M. D'Souza, A. E. Fernandes and D. A. Leigh, *Science*, 2013, **339**, 189–193.

19 D. Núñez-Villanueva, M. Ciaccia, G. Iadevaia, E. Sanna and C. A. Hunter, *Chem. Sci.*, 2019, **10**, 5258–5266.

20 X. Li and D. R. Liu, *Angew. Chem., Int. Ed.*, 2004, **43**, 4848–4870.

21 B. Lewandowski, D. Schmid, R. Borrmann, D. Zetschok, M. Schnurr and H. Wennemers, *Nat. Synth.*, 2023, **2**, 331–337.

22 P. Wilhelm, B. Lewandowski, N. Trapp and H. Wennemers, *J. Am. Chem. Soc.*, 2014, **136**, 15829–15832.

23 S. Dobitz, M. R. Aronoff and H. Wennemers, *Acc. Chem. Res.*, 2017, **50**, 2420–2428.

24 J. Saadi and H. Wennemers, *Nat. Chem.*, 2016, **8**, 276–280.

25 J. Lubkoll and H. Wennemers, *Angew. Chem., Int. Ed.*, 2007, **46**, 6841–6844.

26 T. Schnitzer, A. Budinská and H. Wennemers, *Nat. Catal.*, 2020, **3**, 143–147.

27 B. Linton and A. D. Hamilton, *Tetrahedron*, 1999, **55**, 6027–6038.

28 C. E. Grünenfelder, J. K. Kisunzu and H. Wennemers, *Angew. Chem., Int. Ed.*, 2016, **55**, 8571–8574.

29 H. Wennemers, J. Kisunzu, B. Lewandowski and U. Lewandowska, (2S,4S)-(9H-Fluoren-9-ylmethoxycarbonyl)-4-azidoproline, *e-EROS Encycl. Reagents Org. Synth.*, 2016, DOI: **10.1002/047084289X.rn01954**.

30 Y. A. Nagel, P. S. Raschle and H. Wennemers, *Angew. Chem., Int. Ed.*, 2017, **56**, 122–126.

31 M. Li, R. Puschmann, A. Herdlitschka, D. Fiedler and H. Wennemers, *Angew. Chem., Int. Ed.*, 2020, **59**, 15586–15589.