Stoichiometry alone can steer supramolecular systems on complex free energy surfaces with high selectivity

Komáromy, Dávid; Tiemersma-Wegman, Theodora; Kemmink, Johan; Portale, Giuseppe; Adamski, Paul R.; Blokhuis, Alex; Aalbers, Friso S.; Marić, Ivana; Santiago, Guillermo Monreal; Ottelé, Jim

Published in: Chem

DOI: 10.1016/j.chempr.2021.05.020

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date: 2021

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Stoichiometry alone can steer supramolecular systems on complex free energy surfaces with high selectivity

Dynamic combinatorial chemistry allows the selective formation of three remarkably different disulfide macrocycles (self-assembling into fibers, hexagonal fiber bundles, or a [c3]-daisy chain pseudorotaxane, respectively) from two dithiol building blocks possessing complementary functionalities. The macrocycles can be obtained with high selectivity, which only depends on the stoichiometry of the building blocks. We show that self-assembly can navigate systems on complex Gibbs energy landscapes to selectively access structures beyond the individually most stable ones while remaining under thermodynamic control.
Stoichiometry alone can steer supramolecular systems on complex free energy surfaces with high selectivity

Dávid Komáromy,1,* Theodora Tiemersma-Wegman,1 Johan Kemmink,1 Giuseppe Portale,2 Paul R. Adamski,1 Alex Blokhuis,1 Friso S. Aalbers,3 Ivana Marić,1 Guillermo Monreal Santiago,1 Jim Ottelé,1 Ankush Sood,1 Vittorio Saggiomo,1,4 Bin Liu,1 Pieter van der Meulen,1 and Sijbren Otto1,5,*

SUMMARY
Self-assembly enables access to complex molecular architectures but is traditionally confined to structures that represent global minima on the Gibbs energy landscape. We now report that it is possible to access structures other than those with the lowest individual Gibbs energy, while remaining under thermodynamic control. We prepared dynamic combinatorial libraries from two building blocks possessing complementary binding motifs. Depending on the building block stoichiometry, one of three competing self-assembling macrocycles can be formed with remarkable selectivity. By mixing the same two building blocks, we could access a self-replicating octamer, self-assembling into fibers; a hexamer with a precise 4:2 building block stoichiometry, assembling into hexagonally packed fiber bundles; and a tetramer with 1:3 building block ratio affording a [c3]daisy chain pseudorotaxane. Thus, systems chemistry approaches can enhance the versatility of self-assembly by allowing to navigate complex Gibbs energy landscapes to access structures beyond those that are individually the most stable.

INTRODUCTION
Chemistry involves navigating through complex Gibbs energy landscapes. In covalent chemistry (i.e., organic synthesis), the challenge resides in guiding starting material(s) across the right saddle points (activation barriers) to selectively end up in the energy well that corresponds to the desired product. In most cases, the reactions are controlled by kinetics (i.e., heights of activation barriers). Great advances have been made in this area of covalent chemistry and, given enough time and resources, essentially any target molecule can now be accessed by synthesis.

Until recently supramolecular chemistry1–13 has followed a different paradigm. Constructing a specific supramolecular assembly has traditionally required it to be the thermodynamically most stable structure. Not barrier heights but depths of the wells in the Gibbs energy landscape typically determine which assembly is formed. This approach to self-assembly is limiting, as for a given system, only assemblies that correspond to the thermodynamic minimum can be obtained. More recently it has become appreciated that Gibbs energy landscapes pertaining to non-covalent interactions can feature energy barriers that cannot be transversed using only thermal energy, allowing specific supramolecular assemblies to be accessed by kinetic control.14–20 In principle, such pathway control greatly expands the range of structures that may be accessed by self-assembly. In practice, the still limited mechanism.
insights in self-assembly processes hamper our ability to make full use of the opportunities that are now opening up.

Dynamic combinatorial chemistry (DCC) has proven to be a powerful tool for probing complex Gibbs energy landscapes of systems featuring non-covalent interactions.\textsuperscript{21,22} In DCC, reactions between building blocks equipped with functional groups capable of reversible bond formation give rise to a dynamic combinatorial library (DCL), i.e., a mixture of interconverting oligomers. Reversibility implies that the DCL composition can be shifted toward one (or more) specific oligomer(s) by non-covalent interactions either among DCL members or between an external template and (a) DCL member(s). Most of the work on DCLs made use of external templates.\textsuperscript{23–26} Only more recently DCC has been used for exploring self-assembly processes, leading to foldamers,\textsuperscript{27} interlocked structures,\textsuperscript{28–32} surfactant assemblies,\textsuperscript{33–35} or assembly-driven self-replication.\textsuperscript{36–39} In most of these systems, a single product was obtained. Yet DCC should allow for accessing multiple different structures, even when controlled by thermodynamics. The simplest, almost trivial, approach is to vary building block stoichiometry in a DCL prepared from more than one building block. In systems in which library members with different stoichiometries compete for common building blocks, it is not necessarily the most stable product that dominates. Instead, the library member with a building block stoichiometry that matches the composition of the library may become the dominant product, as the system can produce more of this compound than of the competing species with different stoichiometries. Thus, the system may produce many species that, individually, feature weaker non-covalent interactions, as opposed to a smaller number of species that form individually stronger interactions. It is well established that such mass-balance-related effects can override the individual host-guest binding strengths in DCLs of synthetic receptors.\textsuperscript{40–43} We now show that similar effects can be utilized to navigate multi-well self-assembly Gibbs energy landscapes. We introduce a systems chemistry approach by which the stoichiometry of the components controls, with remarkable selectivity, which well in the Gibbs energy landscape becomes populated, hence, which assembly is produced. We show that, in a DCL made from two building blocks with complementary binding sites, three markedly different types of self-assembling structures are formed, i.e., fibers, fiber bundles, or molecular daisy chains (Scheme 1). The system can be instructed to afford a specific assembly solely through building block stoichiometry, even when, individually, this assembly does not correspond to the individually most stable structure.

RESULTS

Selective formation of three different molecules in a two-building-block DCL

In order to achieve self-assembly in DCLs, building blocks should be designed to be complementary to each other. This principle has proven potential in protein assembly. For example, inserting complementary amino acids at more or less well-defined positions of known proteins can result in their assembly into non-natural types of aggregates.\textsuperscript{44,45} In order to explore building block complementarity in DCC, we made use of already well-characterized systems arising from building blocks 1 or 4 (Scheme 1). The peptide-derived building block 1 features two lysine residues, while the tetramer macrocycle derived from carboxylate functionalized building block 4 has a high affinity for protonated alkylamines, such as spermine (Sp), which are bound through a combination of hydrogen bonds, electrostatic interaction, and hydrophobic interactions.\textsuperscript{46}
Thus, the ingredients are in place for binding between (oligomers of) 4 and the lysine groups of 1.

Oxidation of a stirred aqueous solution of dithiol building block 1 (Scheme 1) gives rise to a small DCL of macrocyclic disulfide oligomers of different ring sizes, among which octamer $1_8$ is stabilized by self-assembly into non-discrete supramolecular polymer nanofibers and is, thus, formed preferentially.

In order to probe whether interactions between 1 and 4 can indeed give rise to unprecedented self-assembling species, we prepared a series of DCLs containing the same total amount of building blocks ($[1] + [4] = 2.0$ mM) with increasing molar ratios...
of 4, ranging from 0 to 100 mol %. The libraries were stirred at 1,200 rpm at room temperature and their composition was monitored with UPLC-MS (supplemental information, Sections 2.2.1 and 2.3; Figures S5–S27) until no further change was detected (typically 7–8 days). We anticipated that, in the absence of additional second-ary interactions, a large number of small oligomers would form and the building block composition for a given oligomer size would follow a binomial distribution, reflecting the building block stoichiometry. Instead, we observed the selective and preferential formation of three disulfide macrocycles with different ring sizes and stoichiometries, in a reproducible manner (Figure 1A, traces (1)–(5); Figures S59–S61). In libraries containing 0–8 mol % of 4, the octamer 1₈ dominates, as expected. Between 8–20 mol % of 4, we could no longer detect 1₈. Instead, a broad distribution of products was formed, differing in ring size and composition. At 20 mol % of 4, a mixed hexamer with stoichiometry 1₄4₂ appeared, becoming the dominant species (90 mol %) at 35 mol % of 4. Between 35–75 mol % of 4, the mixed hexamer persists, however, also tetramer 1₄₃ appears as the sole other product. Oligomers 1₄₃ and 1₄₃ coexist up to 75 mol % of 4. At 80 mol % of 4, the mixed tetramer is the dominant species, whereas upon further increasing the relative amount of 4 to 95 mol %, the mixed tetramer is still present, eventually giving way to 4-only oligomers when the mol % of 4 approached 100 % (see Figures S59 and S61 for detailed chromatograms).

Importantly, each macrocycle forms in exceptionally high yields (corresponding to 90% of the total UPLC peak area) (note that at the applied detection wavelength [254 nm], the extinction coefficients of the two building blocks differ by ca. 30%; Figure S51). Thus, peak areas do not accurately reflect concentrations. However, as our estimations show, the calculated relative molar percentage values obtained from the relative UPLC peak areas [by taking into consideration the difference in extinction coefficients] differ at most 10% from the actual relative UPLC peak areas; supplemental information, Section 1.13) if the initial building block ratio approximates the macrocycle stoichiometry, i.e., 0, 35, and 80 mol % of 4 for 1₈, 1₄₄₂, and 1₄₃₃, respectively (Figures 1B and S60). The high yields indicate that, at the specified stoichiometries, the corresponding macrocycle is formed preferentially. Note that between 35–80 mol % of 4, only 1₄₄₂ and 1₄₃₃ are formed, representing a novel type of
stoichiometry-specific social self-sorting,\textsuperscript{48} where the outcome of the self-sorting process does not only depend on the structure of the assemblies but can also be controlled by changing the stoichiometry of the building blocks, across a wide stoichiometric range.

**Structure and assembly of octamer 1\textsubscript{8}**

The systems chemistry of a 1-only system has been extensively studied before.\textsuperscript{47,49} Upon oxidation of a stirred aqueous solution of 1, initially, mainly cyclic trimer and tetramer disulfides are formed, alongside a small amount of octamer 1\textsubscript{8}. The latter self-assembles into fibrous stacks, which further self-assemble into laterally assembled helical fibers (Figure S73 shows a transmission electron microscopy [TEM] micrograph).\textsuperscript{50} Fiber growth is autocatalytic, but the activation barrier for octamer formation and assembly nucleation is relatively high. As the stacks grow beyond a certain length, they become susceptible to mechanical breakage (conferring by stirring), whereby each breakage event redoubles the number of catalytic fiber ends. Finally, due to reversible thiol-disulfide exchange, the DCL material is channeled almost exclusively into octamer stacks via a fiber growth-breakage mechanism, thereby realizing exponential self-replication.\textsuperscript{38}

**Structure and assembly of the mixed hexamer 1\textsubscript{4}4\textsubscript{2}**

Subsequently, we attempted to uncover the factors bringing about the highly specific stoichiometry and ring size observed for hexamer 1\textsubscript{4}4\textsubscript{2}. At pH 8.2, each unit of 1 has a single positive charge (two protonated lysine side chains and one terminal carboxylate), whereas each unit of 4 carries two negative charges (two carboxylates). Consequently, in oligomers composed of 1 and 4 in a 2:1 molar ratio, charge neutrality is potentially satisfied. Moreover, a DCL with the same stoichiometry and concentration ([1] + [4] = 2.0 mM, [1]:[4] = 2:1) at considerably lower pH (50 mM acetate buffer, pH = 3.2) yielded only a minor amount of 1\textsubscript{4}4\textsubscript{2}, alongside other oligomers (Figure 1A, trace [6]). Notably, at this pH, both lysines and the terminal carboxylates of 1 (p\textsubscript{K\textsubscript{a}}1 = 2.2), as well as the carboxylates of 4, are partially or fully protonated (for the structurally related thiosalicylic acid the p\textsubscript{K\textsubscript{a}}= 3.5,\textsuperscript{51} whereas for terephthalic acid p\textsubscript{K\textsubscript{a}1} = 3.5 and p\textsubscript{K\textsubscript{a}2} = 4.8\textsuperscript{52}), thus possible oligomers are positively charged at this pH. We conclude that 1\textsubscript{4}4\textsubscript{2} is formed preferentially under conditions where its net charge is zero and its self-assembly is not hampered by charge repulsion.

Next, we investigated the building block sequence of 1\textsubscript{4}4\textsubscript{2}. Three regioisomers can potentially form: ortho, meta, and para, similar to disubstituted benzene derivatives (Figure 2A, inset). UPLC analysis of 1\textsubscript{4}4\textsubscript{2} only revealed a single peak under all chromatographic conditions tried, which suggests that a single regioisomer dominates (although we cannot rigorously exclude persistent co-elution of different regioisomers). Microscopy (TEM and atomic force microscopy [AFM]) indicated the presence of large nanoscale aggregates in DCLs containing the mixed hexamer (Figures 2B and 2C; \textit{vide infra}), precluding analysis by NMR spectrometry due to slow relaxation dynamics. ESI-Orbitrap tandem mass spectrometry of the selected parent ion [1\textsubscript{4}4\textsubscript{2}]\textsuperscript{3+} (m/z = 1,083 Da) at low collision voltage (15V) gave rise to several fragments (Figure 2A, bottom). Most importantly, tetramer fragment [1\textsubscript{4}]\textsuperscript{3+} (m/z = 931 Da) was observed, besides [1\textsubscript{4}4\textsubscript{1}]\textsuperscript{3+} (m/z = 1,007 Da) and [1\textsubscript{4}4\textsubscript{1}]\textsuperscript{2+} (m/z = 812 Da). Whereas the two latter can form from any of the three regioisomers, [1\textsubscript{4}]\textsuperscript{3+} can only originate from the ortho isomer, featuring four contiguous 1-units (see Figure S62 for details). At higher collision voltages, we observed water loss as well as the formation of b\textsubscript{3} fragments (i.e., loss of the C-terminal lysine) of one peptide unit within the macrocycle (Figure S63), both of which
are well-described phenomena in gas-phase fragmentation of serine-containing peptide cations. Notably, the presence of non-specific fragments does not allow us to exclude the initial presence of the other positional isomers. However, the ion intensity of the diagnostic fragment $1_4^{3+}$ is the highest compared with that of $1_2^{2+}$ and $1_3^{4+}$ fragments (Figure S64), suggesting that the ortho isomer is formed as the main product.

Figure 2. Elucidating the molecular and supramolecular organization of the mixed hexamer $1_4^{4+}$
(A) ESI-Orbitrap mass spectrum of $1_4^{4+}$ after mass selection of the $[1_4^{4+}]^{3+}$ (m/z = 1083 Da) ion at collision voltages of 0 V (top) and 15 V (bottom). Inset: three possible positional isomers of $1_4^{4+}$. The appearance of $[1_4^{4+}]^{3+}$ indicates the presence of mainly the ortho isomer. Asterisks indicate consecutive water loss from the indicated parent ions.
(B) Cryo-TEM shows the presence of a dense network of fibers (see also Figure S76 for better magnification). Scale bar: 200 nm.
(C) AFM shows densely associated fibers forming linear, bundled (red arrow), and circular (blue arrow) assemblies, which are 5–15 μm in length. Scale bar: 3 μm.
(D) CD spectroscopy shows characteristic signatures at 195 and 205 nm, which are also characteristic of peptide β-sheets.
(E) ThT fluorescence assay indicates the presence of peptide β-sheets.

Next, we investigated the supramolecular organization of $1_4^{4+}$. Cryo-TEM and AFM analysis of a 2-day-old sample (containing exclusively $1_4^{4+}$) indicated the presence of a network of nanoscale fibers (Figures 2B and 2C). The length of these fibers extended to the micrometer range (5–15 μm) and they were several hundred nanometers wide, as indicated by TEM (Figures S74F–S74J). AFM and negative staining TEM (Figure S74) showed that the fibers further assembled into thick bundles of diverse morphologies (linear, curled, and circular). In contrast, fibers assembled from peptide-only replicators (i.e., $1_8^{4+}$, see Figure S73) are significantly shorter (0.5–5 μm) and only 2–5 nm wide. CD spectroscopy (Figure 2D) and thioflavin T (ThT) fluorescence assays (Figure 2E) both indicated a (partial) β-sheet structure, similar to those of peptide fibers. A comparison of the spectroscopic signatures of octamer $1_8^{4+}$ and mixed hexamer $1_4^{4+}$ reveals a noticeable difference: whereas the intensity of the CD spectra is comparable, that of the ThT assay is markedly lower for $1_4^{4+}$ (Figure S65). This observation is in agreement with the lateral association of the mixed hexamer fibers: whereas the intercalation of ThT molecules between the macrocycles is more hindered for the dense fiber bundles of $1_4^{4+}$ than for the more accessible fibers of $1_8^{4+}$, no such difference holds for the interaction with circularly polarized light. ThT fluorescence experiments conducted on thermally annealed samples of $1_4^{4+}$ and $1_8^{4+}$ (Figures S66–S72) failed to improve access of the structures to ThT. In both cases, only minor changes in ThT fluorescence were observed upon incubating the samples at elevated temperatures.
In order to obtain insight into the supramolecular structure of macrocycles $1_8$ and $1_{442}$ in the native (solution) state, we monitored their self-assembly process with small-angle X-ray scattering (SAXS). Two DCLs were prepared from dithiol monomers under identical conditions (50 mM phosphate buffer, pH = 8.2, $[1]+[4] = 2.0$ mM, 1,200 rpm stirring rate); the first composed of both building blocks ($[1]:[4] = 2:1$), whereas the second one was composed only of $1$. The temporal evolution of the two systems showed remarkable differences: the SAXS profile of the DCL leading to the formation of $1_{442}$ (Figure 3A) showed a slope in the low q-range region of the log-log plot of about $q^{1.4}$, close to the characteristic $q^{-1}$ dependence, indicative of the formation of long fibers. SAXS intensity modeling provides an estimate of the fiber diameter of ca. 2.0 nm (close to the theoretically calculated value of 1.8 nm, Figure S78) after 1 day (Figure 3C). In line with this observation, the UPLC peak area of hexamer (corresponding to the molecular composition) and the characteristic CD intensity (corresponding to the supramolecular chirality) increased simultaneously (Figure 4A) in this period, indicating that, as soon as hexamers are formed, they self-assemble into fibers. Moreover, after 7 days, three characteristic signals (Figure 3A, indicated with asterisks) were observed. The relative intensities of these signals increased over time, indicating progressive higher-order association of the fibers. The peaks were centered at $d = 5.8$, 3.4, and 2.9 nm, corresponding to a
1 : $\sqrt{3}$ : 2 ratio in the q-space, which indicates a hexagonal arrangement of the fibers (Figure 3D). In contrast, the second DCL resulting in the formation of 1$_8$ (Figure 3B) does not feature any fibrous aggregate after 1 day (Figure 3C). However, after 4 days, a broad peak centered at d = 4.8 nm is observed, most possibly corresponding to the pairwise association of fibers. The relative intensity of this signal increases over time until 11 days, concomitantly with the formation of 1$_8$, as indicated by UPLC analysis. These results show that the self-assembly of 1$_4$A$_2$ occurs in two steps: fibers are formed first (concomitantly with the evolution of the macrocycle at the molecular level), which later self-assemble into a hexagonal bundle. In contrast, 1$_8$ self-assembles in one step (parallelly to the evolution of the macrocycle), giving rise to pairwise fibers.

Notably, monitoring the structural evolution of 1$_4$A$_2$ with negative staining TEM did not reveal a clear trend, as fiber bundles of various morphologies were observed throughout an observation time of 2 h up to 8 days (Figure S74), although the bundles became more regular after 3 days (Figure S74I). This discrepancy can be a consequence of drying effects during TEM sample preparation, which are often disregarded but could be consequential, especially for samples which are highly prone to secondary association. 56

Given the remarkably high self-assembly propensity of the mixed hexamer, we finally investigated the kinetics of its formation and compared it to that of the one-component peptide replicators (e.g., 1$_8$). The latter are only formed under mechanical agitation due to the aforementioned breakage-growth mechanism and their evolution features an initial lag phase, followed by an exponential growth phase, both taking several days. In sharp contrast, the time evolution of the mixed hexamer 1$_4$A$_2$ from its building blocks upon agitation with the same stir rate used for the one-component peptide replicators (1,200 rpm) featured a much shorter lag phase of ca. 1 h (Figure 4A). Even more surprisingly, a similar kinetic profile was observed for an unagitated sample, showing that the formation of 1$_4$A$_2$ is not mechano-sensitive (Figures 4B and 4C). Furthermore, also monomer oxidation is two orders of magnitude faster than for the peptide-only replicators (3–6 h versus 4–5 days, Figure 4B). These characteristics only change slightly by decreasing the stir rate to 400 rpm (Figure S79B) or the total building block concentration to 50 μM (Figure 4C, black line). These results indicate that the formation of 1$_4$A$_2$ from its building blocks and its self-assembly proceeds with a low kinetic barrier and most likely follows a different mechanism than that of 1$_8$ replicators.
Figure 5. Structure elucidation and self-assembly of [c3]-[1,4]$_3$

(A) Schematic structure of acyclic ([a]n) and cyclic ([c]n) daisy chains from (1,4)$_3$.

(B) Proton assignment of [c3]-[1,4]$_3$, highlighting the K3 lysine side chain threaded through the macrocycle.
Furthermore, 1-g can form autocatalytically, whereas 1,4-d shows only weak autocatalytic activity. In a non-agitated, partially oxidized DCL of 1 octamer formation is very slow, whereas seeding of this sample with a small amount of pre-formed 1-g results in fast octamer formation. In contrast, in a slightly oxidized, non-agitated mixture of 1 and 4 ([1] + [4] = 2.0 mM, [1]:[4] = 2:1), fast autonomous conversion (70 mol % hexamer in 50 min) into 1,4-d was observed (Figure S82). This result showed that the non-catalyzed reaction is very fast leaving little room for autocatalysis. Additionally, TEM showed (Figure S75) that in sheared hexamer seeds, mostly the degree of lateral association decreases, while fiber length is essentially retained. Thus, it proved difficult to produce a seed with a large number of fiber ends which are the sites of potential autocatalysis.

A somewhat increased fluorescence intensity of ThT was observed (Figure S77) for the sheared seeds, suggesting better accessibility of the hexamer molecules for dye intercalation. Decreasing the oxidation level, i.e., using a monomer mixture as the starting material, results in mostly the degree of lateral association decreases, while fiber length is essentially retained. Thus, it proved difficult to produce a seed with a large number of fiber ends which are the sites of potential autocatalysis.

Structure and assembly of the mixed tetramer 1,4-d

In contrast to the mixed hexamer, no nanoscale assemblies were detected in samples containing mixed tetramer 1,4-d. However, its efficient (80% by UPLC) and stoichiometry-specific formation suggest that non-covalent interactions stabilize this particular oligomer. As no nanoscale objects were detected by TEM (Figure S84) and kinetic analysis did not provide any indication for autocatalysis (Figure S83), it seemed more likely that a discrete assembly was formed. Thus, 1,4-d was isolated and characterized using NMR, ESI-MS, and SDS-PAGE. Nearly all protons of 1 and 1,4-d (Figure 5B) were assigned following an extensive series of TOCSY, NOESY, and HSQC experiments (supplemental information, Section 2.5.2; Figures S85–S97). The most striking feature in the NMR spectrum of 1,4-d (Figure 5D) compared with that of 1 (Figure 5C) was the large upfield shift of the lysine side chain protons K3β-K3γ (∆δ = −2.9 ppm for K3γ-K3α) and the smaller but also significant upfield shifts for K3α and K3ω. Furthermore, NOE cross-peaks were detected between the protons of the aromatic macrocycle and the aforementioned side chain protons (Figure 5E), indicating that these protons are in close spatial proximity. These observations indicate that the lysine side chain of K3 is threaded through a cavity of the macrocycle. The threaded structure is further supported by the observed splitting of the K3β-K3γ protons, indicating that the two protons belonging to the same CH2 group experience a different chemical environment, as their rotation is restricted in the confined space of the cavity (Figure 5D). Moreover, in contrast to 1, the lysine ammonium protons (K3NH3+, K5NH3+) are detectable in the

**Figure 5. Continued**  
(C) 1H-NMR spectrum of peptide building block 1 (500 MHz, H2O/D2O 9:1, phosphate buffer, pH = 7.2, 278 K). *, dioxane (internal standard); **, TCEP and TCEP-oxide.  
(D) 1H-NMR spectrum of [c3]1,4-d3 (600 MHz, H2O/D2O 9:1, phosphate buffer, pH = 7.2, 278 K), highlighting large upfield shifts of the protons of the K3 side chain and considerable downfield shifts in the L2 and S4 amide protons.  
(E) 1H-1H NOESY spectrum of [c3]1,4-d3, highlighting NOE cross-peaks between the K3 side chain and the aromatic protons of the macrocycle.  
(F) DOSY spectrum of [c3]1,4-d3, confirming the presence of one species with a diffusion constant D = 2.1 × 10−6 cm2 s−1.  
(G) SDS-PAGE analysis of [c3]1,4-d3 in ammonium acetate buffer (50 mM, pH = 7.0). BPB, bromophenol blue (m = 669 Da).  
(H) Negative mode ESI-Orbitrap spectrum of 1,4-d, in ammonium acetate buffer (50 mM, pH = 7.0), showing the presence of daisy chain [c3]1,4-d3 as well as dimer fragment (1,4-d2) and monomer 1,4-d.
The presence of only two lysine \( K\) protons (a threaded and a non-threaded one) in \( 1,4_3 \) suggests a mixture of cyclic \([c_n]daisy chain(s)\) (Figure 5A)\(^{57}\) or lasso \([1]pseudorotaxane\) \( 4,\text{Sp} \), which is stabilized by hydrogen bonding between ammonium protons and the macrocyclic carboxylates, as well as by the threading of nonpolar alkyl chains through the hydrophobic aromatic cavity.\(^{56}\) Given the structural similarity between \( \text{Sp} \) and the lysine side chains, the proposed threaded structure seems plausible.

In order to determine the exact value of \( n \) (i.e., the number of macrocycles in one daisy chain unit), native SDS-PAGE experiments (both in phosphate and ammonium acetate buffer) were performed (Figure 5G). These indicated the presence of one species, with a molecular mass slightly bigger than 3.5 kDa, which most possibly corresponds to the \([c_3]\)- (4,142 Da) rather than the \([c_2]\)- (2,764 Da) or the \([c_4]\)- (5,528 Da) species. Finally, native Orbitrap-ESI-MS in ammonium acetate buffer (Figure 5H) showed the presence of the \([c_3]\) species \(([1,4_3)_3]^{3+} \), 1,035 Da; \([1,4_3]_3]^{3+} \), 690 Da) alongside with dimer \(([1,4_3)_2]^{3+} \), 920 Da) and monomer \(([1,4_3]^{-} \), 690 Da) fragments, but no higher oligomers (i.e., \([1,4_3)_4]^{5+} \), 1,840 Da). The extensive fragmentation may be due to the absence of stabilization through hydrophobic effects in the gas phase (i.e., in the absence of aqueous solvent). We conclude that \( 1,4_3 \) selectively assembles into a \([c_3]\) daisy chain as shown in Scheme 1.

Relative stabilities of the different states of the \((1+4+\text{Sp})\)-system

In view of the substantially different self-assembly modes of the three preferred macrocycles in the system, we were particularly interested in assessing relative stabilities of the various states of the system and in probing whether any kinetic barriers exist to their interconversion. We, thus, prepared samples of \( 1_8, 1_4, 1_4, 1_4_3 \) as well as of spermine complex \( 4,\text{Sp} \) and let them react with \( 1, 4 \) (oxidized into mixtures of small macrocycles) or \( \text{Sp} \) (2.0 mM total building block concentration \( pH = 8.2, 50 \text{ mM phosphate buffer, stir rate 1,200 rpm; reactions were conducted at 20 mol % thiol content in order to keep the systems dynamic) so that conversion to another assembly could potentially occur (see overview in Schemes 2 and 3). In each of these experiments, we added components together at concentrations such that roughly equimolar amounts of two different competing assemblies could potentially be formed. We calculated the product distributions that, given the chosen concentrations, would arise if \( K \) were to equal 1 (i.e., the Gibbs energy change associated with the interconversion of the two states would be zero) using the law of mass action (supplemental information, Section 2.6). A comparison between this expected (in case of \( K = 1 \)) and the observed product distributions then informs on whether the interconversion involves a negative or positive \( \Delta G \) value. The results of this series of experiments are summarized in Figure 6, showing the UPLC peak areas that would correspond to the expected \( (K = 1) \) product concentrations in red, the experimentally observed product distribution upon equilibration in blue, and the starting state of the system in white (see Figures S98–S99 for the UPLC chromatograms). As for most reactions, the expected and observed concentrations differ by several orders of magnitude, the concentrations are shown on a logarithmic scale. The corresponding graphs plotted using a linear scale are shown in Figure S100.
The results of Figure 6 are summarized in Scheme 2, indicating the positions of the various equilibria, and in Scheme 3, showing which interconversions were found to be feasible and showing the relative stabilities of some of the states of the system. Any given assembly can be converted to at least one other assembly, showing that none of them resides in a kinetically trapped state. Of the reactions shown in Figure 6, only one (Figure 6D) did not show a significant change in the amount of the starting assembly $1_4$4 (even though the same assembly does interconvert in reaction C). In order to further confirm that the behavior of the reaction shown in Figure 6D was not the result of a kinetic trap, we performed the reaction in the opposite direction (Section 2.6.2; Figures S101 and S102). Both forward and reverse experiments converged on the same product distribution, suggesting that this distribution corresponds to the equilibrium state of this system.

The fact that all assemblies can be converted to other ones also indicates that none of the assemblies represents a free energy well that is so deep that the system cannot be lifted out of it. The latter notion is nicely illustrated by the experiments shown in Figures 6C and 6F. In the former, $1_44_2$ is partially converted into $1_44_1$ by adding 4, while in the latter $1_44_3$ is partially converted into $1_44_2$ by the addition of 1 (note that these two reactions do not form a thermodynamic cycle, but that each transformation consumes material in the form of 1 or 4; see Scheme 3B for the thermodynamics). Thus, whichever of the two assemblies might be individually more stable, it can be (at least partially) converted to the other assembly upon providing the required building block. This observation shows that stoichiometry can be used to overcome inherent differences in thermodynamic stability between $1_44_2$ and $1_44_3$.

The fact that the interconversions between different assemblies shown in Scheme 2 tend to be associated with the consumption or liberation of building blocks 1 or 4 (see Scheme 3B), impairs a direct comparison of the thermodynamic stability of the individual assemblies $1_8$, $1_44_2$, and $1_44_3$. The only direct comparison between the stabilities of the three assemblies is through equilibrium (J) in Scheme 2. This equation can be obtained as a linear combination of equations (E) and (D) (see supplemental information, Section 2.6.3), both of which are shifted toward the product

**Scheme 2. Positions of the equilibria involving the interconversion between different types of assemblies made from 1, 4, and Sp**

Reactions (A)–(I) correspond to (A)–(I) in Figure 6. Reaction equation (J) can be obtained as a linear combination of equations (E) and (D). Building blocks 1 and 4 are shown in quotation marks to indicate that they represent oxidized forms of 1 or 4, respectively (predominantly trimeric and tetrameric cyclic disulfides) present at equilibrium.
side (Scheme 2). Thus, the resulting equation is also shifted to the right, hinting at a relatively high stability of the $1_4 4_2$ complex.

Previously, simulations on thermodynamically governed DCLs have shown that the formation of a large number of smaller weakly binding macrocycles may be preferred over the formation of a lower number of larger more strongly binding ones.40,43,60 Most of the reactions shown in Figure 6 show similar behavior, with the equilibrium favoring the side that produces the largest number of self-assembling species (for example in reaction [A] forming two molecules of $1_4 4_2$ is favored over one molecule of $1_8$). An exception to this behavior is the reaction shown in Figure 6E, preferring 7 assembling species (2 molecules of $1_4 4_2$ and 5 molecules of $4_4 Sp$) over 8 molecules of $1_1 4_3$. (supplemental information, Section 2.8.)

Notably, these conclusions also hold if, instead of treating the macrocycles as individual equilibrating molecules, the assemblies are treated a separate phases (supplemental information, Section 2.8). Where the multicomponent equilibria between macrocycles within a single solution-phase qualitatively explain that one or another species forms selectively, depending on the stoichiometry, multiphase behavior enhances selectivity since phase rules restrict the number of species that are allowed to coexist. Treatment of the system in terms of phases also confirmed that it is possible to access structures other than those with the lowest individual Gibbs energy through appropriate choice of stoichiometry (supplemental information, Section 2.8.5).

**Substrate scope**

Finally, we investigated how structural modification of the building blocks affects the described system’s behavior. For this purpose, we prepared DCLs in 2:1 and 1:3 molar ratios, either substituting 1 for peptide building blocks 2, 3, or 5 (containing an alanine, threonine, and phenylalanine moiety in place of serine, respectively) or 4 for regioisomer 6 (Scheme 1 and Figure 7). The selective formation of mixed hexamers ($2_4 4_2$ and $3_4 4_2$) as well as of mixed tetramers ($2_4 4_3$ and $3_4 4_3$) was observed for building blocks 2 (Figure 7A) and 3 (Figure 7B), respectively, showing that small structural modifications near the

**Scheme 3. Systems chemistry of 1, 4, and Sp**

(A) Available pathways for interconverting assemblies $1_8$, $1_4 4_2$, $1_4 4_3$, and $4_4 Sp$ through the reactions probed in Figure 6.

(B) Qualitative comparison of the relative Gibbs energies of the various states of the system made from building blocks 1 and 4. Note that, in order to be able to include reactions (C) and (F) in the same energy diagram, we added “1” to both sides of reaction (C) in the energy diagram. This does not alter the $\Delta G$ of reaction (C) as the formally added “1” is neither consumed nor produced in reaction (C) but it is needed in reaction (F).
Figure 6. Comparison of the relative stabilities of self-assembling macrocycles $1_8$, $1_44_2$, $1_44_3$, and $4_4Sp$ by comparing the logarithm of their expected (in case of $K=1$; red bars) and observed (blue bars) conversion in various interconversion processes starting from the macrocycles indicated by white bars.

(A) The reaction between $1_8$ and $4$ (1:1.3 molar ratio) affords $1_44_2$ in higher yield (26% of total UPLC peak area) than expected for the $K=1$ case where substrate and product would be equally stable (3.3 $\times$ 10$^{-4}$% of total UPLC peak area).

(B) The reaction between $1_8$ and $4$ (1:24 molar ratio; 33 mol% thiolate) produces $1_44_3$ in larger quantities (57% of total UPLC peak area) than expected for the $K=1$ case (6.0 $\times$ 10$^{-4}$% of total UPLC peak area).

(C) The reaction between $1_44_2$ and $4$ (1:2 molar ratio) affords $1_44_3$ in considerably higher yield (8% of total UPLC peak area) than expected for the $K=1$ case (1.9 $\times$ 10$^{-4}$% of total UPLC peak area).

(D) The reaction between $1_44_2$ and $Sp$ (1:16 molar ratio) fails to produce detectable quantities of $1_8$, whereas for the $K=1$ case $1_8$ should account for 8.2% of UPLC peak area.

(E) The reaction between $1_44_3$ and $Sp$ (2:5 molar ratio) affords (besides $4_4Sp$) $1_44_2$ in considerably higher yield (47% of total UPLC peak area) than expected for the $K=1$ case (6.7 $\times$ 10$^{-2}$% of total UPLC peak area).

(F) The reaction between $1_44_3$ and $1$ (2:1 molar ratio) affords $1_44_2$ in considerably higher yield (46% of total UPLC peak area) than expected for the $K=1$ case (3.9 $\times$ 10$^{-2}$% of total UPLC peak area).

(G) The reaction between $1_44_3$ and $Sp$ (1:11 molar ratio) fails to produce $1_8$ and $4_4Sp$, affording $1_44_2$ and $4_4Sp$ instead, consistent with the outcome of experiment (D).
guest binding site (lysine side chain) do not affect the self-assembly propensity (Figures S28–S37 and S104–S106). In contrast, 5 and 4 form hexamer 5_4_2 with only poor selectivity, but are capable of forming 5_4_3 with high selectivity at the right stoichiometries (Figures 7C and S38–S43). This discrepancy suggests that the formation of the mixed hexamer is more susceptible to structural changes (phenylalanine instead of serine) than that of the mixed tetramer. We speculate that this difference reflects the fact that, in the fiber bundles, the building blocks are brought into close proximity, which is spatially more demanding than the much more open arrangement of building blocks in the daisy chains (whose formation is supposedly the driving force also in the case for the other peptide analogs). Upon substituting 4 for 6, no selective formation of any specific macrocycle was observed upon mixing with 1 (Figures 7D and S44–S58), although building blocks 4 and 6 only differ in the position of one carboxylate group. Instead, a complex mixture of different ring sizes and compositions was obtained. Apparently, the position of the carboxylate binding site has a pronounced effect on the self-assembly propensity. Notably, the regioisomers of several mixed macrocycles (1_2_6_n, n = 3–6) could be resolved (Figure 7D, red asterisks) applying the same UPLC gradient as used for the 1+4 system. As the polarity of 4 and 6 is expected to be similar, this finding supports the notion that, in the case of 1_4_2, indeed only one regioisomer was formed.

DISCUSSION

We showed that incorporating complementary molecular recognition units into separate building blocks 1 and 4 brings about the spontaneous and stoichiometry-selective formation of three oligomers. Each oligomer possesses a different ring size and self-assembles into markedly different nanoscale assemblies, while competing for common building blocks. The system can be directed to form specific assemblies solely through building block stoichiometry, overriding any differences in thermodynamic stabilities of these assemblies. Specifically, 1 self-assembles into fibers via hydrophobic self-stacking. Mixed hexamer 1_4_2 also self-assembles into fibers, which subsequently arrange into hexagonally packed bundles. The remarkably high efficiency of this assembly process precludes significant autocatalysis, which is essential to the formation of 1_8. Finally, if 4 is in excess, daisy chain pseudorotaxane [c_3](1_4_3)3 is obtained. Most importantly, both mixed macrocycles are stabilized by secondary interactions between the complementary binding motifs (i.e., hydrophobic effects as well as ion pairing between the lysine ammonium side chains from one building block and the carboxylates from the other).

The remarkable stoichiometry- and sequence-specific formation of 1_4_2 can be attributed to several factors. First, the formation of a neutral species renders a 2:1 stoichiometry favorable, as macrocycles with this building block ratio can potentially benefit maximally from attractive electrostatic interactions and allow for charge repulsion to be minimized. These characteristics make assembly by 1_4_2 more facile than the assembly of 1_8, which is accompanied by charge repulsion. Second, we know from the structure of 4_2_Sp that, if two or more carboxylates are in close proximity, they can engage in hydrogen bonds and ion-pairing interactions with protons of a primary ammonium ion from 1.46 We speculate that this arrangement renders the formation of hexamer 1_4_2 to be thermodynamically more favorable than that of the competing trimer 1_4_1 (not observed in significant quantities but also satisfying the 2:1 stoichiometry and entropically less demanding, as it has a smaller ring size), as
the latter has no vicinal 4-moieties. Arrangements featuring 4-dyads in 1,42 in close proximity to each other along the fiber axis create a microenvironment with four carboxylates, which may bind lysine ammonium groups protruding from other fibers. The intermolecular interactions responsible for the lateral association of the fibers are likely to be due to a combination of hydrophobic effects and electrostatic attraction (see Scheme 1 and Figure S81 and the discussion following this figure). In contrast to other nucleation-elongation processes, where recognition sites required for self-assembly are only available at the fiber ends, 1,42 can self-assemble along the fiber surface as well, further lowering the barrier for self-assembly. Most of

Figure 7. Effect of building block structure on systems behavior
(A) UPLC traces of DCLs prepared from 2 and 4 ([2] + [4] = 2.0 mM). Stoichiometry is indicated on the chromatograms.
(B) UPLC traces of DCLs prepared from 3 and 4 ([3] + [4] = 2.0 mM). Stoichiometry indicated on the chromatograms.
(C) UPLC traces of DCLs prepared from 5 and 4 ([5] + [4] = 2.0 mM). Stoichiometry indicated on the chromatograms.
(D) UPLC traces of DCLs prepared from 1 and 6 ([1] + [6] = 2.0 mM). Stoichiometry indicated on the chromatograms.
the macrocycles are buried in the tightly packed bundles of long fibers, thus only a relatively small number of catalytic fiber ends are available, explaining the limited autocatalytic activity. Furthermore, the fact that only a part of the surface of the fibers is exposed to the solvent may limit the ability of the fibers to guide the building blocks to the fiber ends, as observed recently for another fiber system. Similar reduction of autocatalytic activity upon lateral association has been observed previously.

Using the appropriate building block ratio, we can efficiently access a \([c_3]\) daisy chain assembly of \(1,4\). Molecular daisy chains are candidates for biomimicking motors, such as molecular muscles. However, their development lags behind those of other molecular motors, in part due to their demanding synthesis. Moreover, reports about daisy chains in water are scarce. Our results show that a \([c_3]\) daisy chain can form spontaneously in high yield in water through a process of self-synthesis, where non-covalent interactions direct covalent bond formation in a subcomponent self-assembly process. We attribute the high selectivity of formation of \(1,4\) to the fact that this is the smallest macrocycle that provides a binding pocket for an alkyl ammonium group (requiring at least four building blocks) while offering a matching number of alkyl ammonium groups at the same time.

Most importantly, our results demonstrate that stoichiometry can be a powerful tool to navigate Gibbs energy landscapes of supramolecular systems, enabling the remarkably selective formation of specific assemblies even where these compete with other assemblies that may, when compared one-to-one, be of higher thermodynamic stability. That stoichiometry can override thermodynamic stability is a systems effect: the system as a whole can achieve more non-covalent interactions by producing a larger number of individually less stable assemblies as opposed to a smaller number of individually more stable assemblies. Making use of such systems effects expands the scope of self-assembly beyond only populating the individually most stable species but without having to resort to kinetically controlled steps.

**EXPERIMENTAL PROCEDURES**

**Resource availability**

**Lead contact**

Requests for further information or resources should be directed to and will be fulfilled by the lead contact, Sijbren Otto (s.otto@rug.nl).

**Materials availability**

This study did not generate new unique reagents.

**Data and code availability**

This study did not generate datasets or codes.

Full experimental procedures can be found in the supplemental information.

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at [https://doi.org/10.1016/j.chempr.2021.05.020](https://doi.org/10.1016/j.chempr.2021.05.020).

**ACKNOWLEDGMENTS**

We are grateful to ERC (AdG ToDL 741774), the EU (Marie-Sklodowska-Curie grant 847675), the NWO, and the Dutch Ministry of Education, Culture, and Science (Gravitation program 024.001.035) for financial support. Bartosz Matysiak, Dr. Gaël
Schaeffer, and Dr. Charalampos G. Pappas are gratefully acknowledged for fruitful discussions.

AUTHOR CONTRIBUTIONS


DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES


