ARTICLES
https://doi.org/10.1038/s41557-020-00565-2

nature chemistry

Emergence of low-symmetry foldamers from single monomers

Charalampos G. Pappas[1], Pradeep K. Mandal[2], Bin Liu[1], Brice Kauffmann[3], Xiaoming Miao[1], Dávid Komáromy[1], Waldemar Hoffmann[4,5], Christian Manz[4,5], Rayoon Chang[4,5], Kai Liu[1], Kevin Pagel[4,5], Ivan Huc[2]✉ and Sijbren Otto[1]✉

Self-assembly is a powerful method to obtain large discrete functional molecular architectures. When using a single building block, self-assembly generally yields symmetrical objects in which all the subunits relate similarly to their neighbours. Here we report the discovery of a family of self-constructing cyclic macromolecules with stable folded conformations of low symmetry, which include some with a prime number (13, 17 and 23) of units, despite being formed from a single component. The formation of these objects amounts to the production of polymers with a perfectly uniform length. Design rules for the spontaneous emergence of such macromolecules include endowing monomers with a strong potential for non-covalent interactions that remain frustrated in competing entropically favoured yet conformationally restrained smaller cycles. The process can also be templated by a guest molecule that itself has an asymmetrical structure, which paves the way to molecular imprinting techniques at the level of single polymer chains.

The self-assembly of a defined number of identical molecular units into discrete objects is to some extent understood and amenable to design. No matter how large, these objects generally lack complexity in that their subcomponents engage in similar interactions with their neighbours, which results in the high symmetry of the assembly, as various types of rings and cages illustrate[1–4]. Homomeric assemblies with no symmetry may occur in the asymmetric unit of crystals lattices[5,6], but these assemblies are not discrete and may not exist in solution. Homomeric bundles of n α-helices expected to possess C_n or D_n symmetry may also collapse and lose all or part of their symmetry but preserve most helix–helix interactions[7]. Here we report the discovery of the self-templated formation of exceptionally large dynamic macrocycles that have well-defined folded conformations with a low symmetry and substantial structural complexity. Each macrocycle consists of identical subunits, but their roles and interactions with neighbour units in the final structure differ. In the chemical space explored so far of poly(disulfides) of dipeptidyl aryl-dithiols, the occurrence of such structures is both frequent and diverse. It is driven by folding into well-defined conformations akin to those of biopolymers with the remarkable difference that the sequences described here are homomeric and largely devoid of secondary structure elements. Thus, a 16-mer (8.3 kDa) with eight distinct environments and a 23-mer (10.9 kDa) with 12 distinct environments spontaneously form from their respective building block at the exclusion of shorter or longer sequences. On subtle chemical variations, a 9-mer, a 17-mer and a 20-mer also emerged. Structure elucidation shows that, in contrast with biopolymers, the folded conformations of these macrocycles show a limited hierarchy. Metal ions or small molecules template the formation of other ring sizes (12-mer and 13-mer). The resulting macrocycles were then capable of binding their templates. Altogether, these results shed new light on molecular self-assembly and on the complexity of the objects that may be obtained by this method. They also introduce an unanticipated approach to the synthesis of macromolecules with a perfectly uniform length.

Results and discussion

Building block design. Dynamic bonds, whether covalent or non-covalent, are essential to access discrete self-assembled structures, as they allow for error correction during the assembly processes, which gives rise to the thermodynamically most stable object[8–12]. In dynamic combinatorial chemistry, these characteristics are combined with the ability to screen structural space for the most stable assemblies from among many related structures[13–15]. Dynamic combinatorial approaches to foldamers[16–19] involve creating mixtures of molecules that exchange building blocks by reversible covalent chemistry. Non-covalent interactions within folded structures cause these to be more stable than their non-folded counterparts, which shifts the product distribution of the dynamic combinatorial library (DCL) in the direction of, ideally, the most stably folded structures. Despite early proof-of-principle results[20], this approach has so far not lived up to the promise of delivering foldamers with unexpected and unpredictable structures. Encouraged by our recent discovery of a new foldamer based on a chimeric nucleobase–amino acid conjugate[21] we intensified our exploration of structural space focused on building blocks with a 1,3-dimercaptobenzene core that carries a side chain in the 5-position (Fig. 1a). We reasoned that for a specific foldamer to form selectively, it needs to be stabilized by selective intramolecular non-covalent interactions more effectively than the other members of the DCL with which it competes for building blocks. This consideration suggests that specific directional non-covalent interactions, such as hydrogen bonds, are needed. Previous work already showed that when interactions are repulsive (such as when using the carboxylic acid-functionalized building block **1a**, which is negatively charged at neutral pH) only

[1]Centre for Systems Chemistry, Stratingh Institute, Groningen, the Netherlands. [2]Department of Pharmacy and Center for Integrated Protein Science, Ludwig-Maximilians Universität, Munich, Germany. [3]Université de Bordeaux, CNRS, INSERM, UMS3033, Institut Européen de Chimie et Biologie, Pessac, France. [4]Institute of Chemistry and Biochemistry, Freie Universität Berlin, Berlin, Germany. [5]Fritz Haber Institute of the Max Planck Society, Berlin, Germany. ✉e-mail: ivan.huc@cup.lmu.de; s.otto@rug.nl

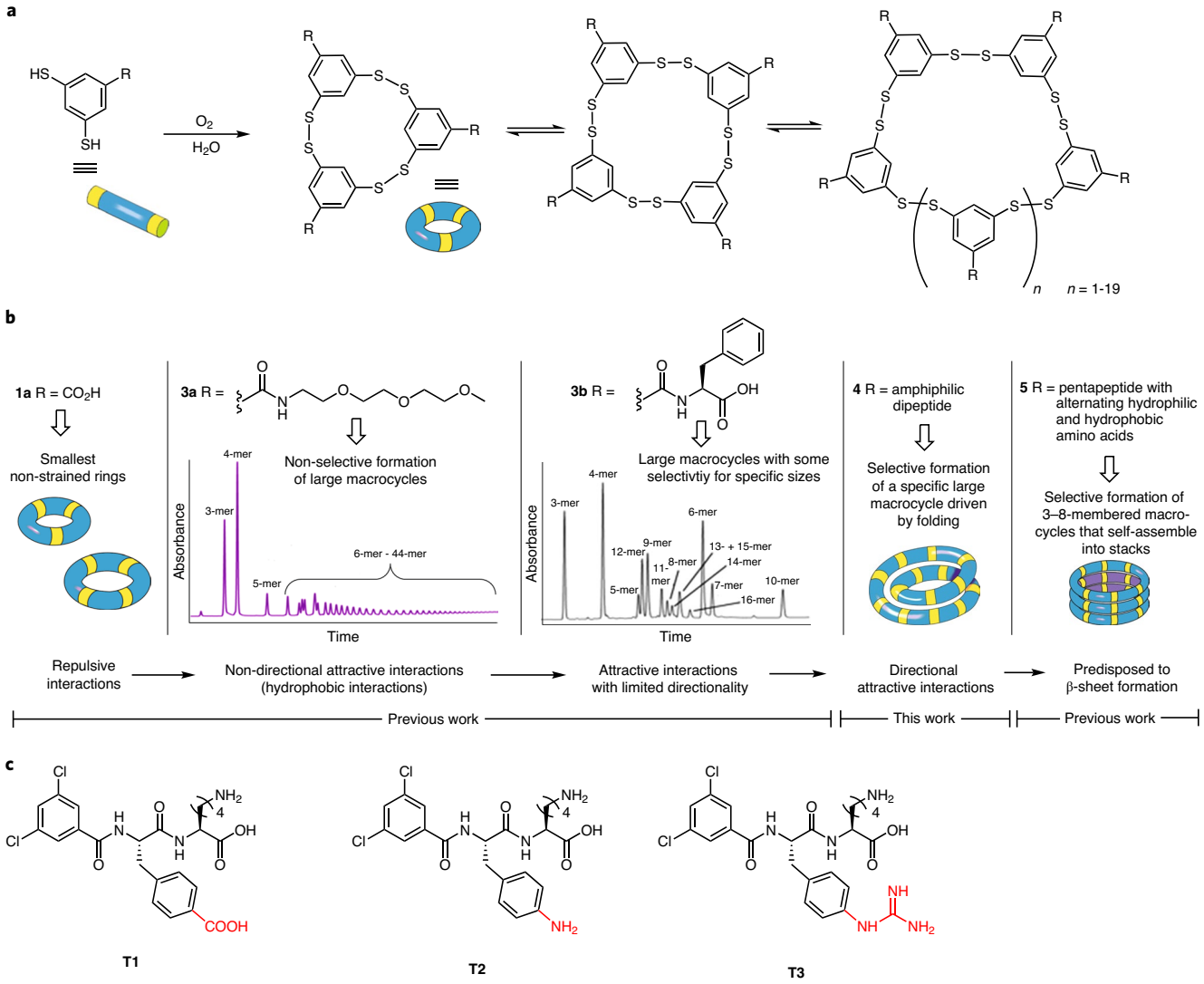

Fig. 1 | How foldamer formation depends on building block design. a, Dithiol building blocks oxidize to form DCLs of differently sized disulfide macrocycles. **b**, Depending on the substituent R, DCLs can either give rise to a few small non-folded macrocycles, a wide range of different ring sizes, the selective formation of a specific foldamer or self-assembled stacks of rings. **c**, Peptides used to template foldamer formation.

small non-strained rings form[22] (trimers and tetramers (Fig. 1b)). When building blocks may engage in favourable yet non-directional contacts, for example, due to hydrophobic interactions, many different library members benefit to a similar extent, which leads to a broad distribution of oligomer sizes, as the oligomerization of **3a** illustrates (Fig. 1b)[23]. Enabling more specific interactions by appending an amino acid to the building block leads to the first signs of a bias in favour of certain oligomers at the expense of others, as shown by the behaviour of DCLs made from **3b**[24]. When interactions between building blocks become highly specific on introducing a peptide group that is predisposed to β-sheet formation (by alternating hydrophobic and hydrophilic amino acids, as in **5**) we observed the selective formation of a specific library member as a result of the assembly of β-sheets. However, in this case, the interactions take place intermolecularly, which leads to the stacking of rings rather than foldamer formation[25] (Fig. 1b). These observations suggest that there may be a window for foldamer formation in which interactions are not as specific as for **5**, but more specific than for **3b**. This induced us to focus on dipeptides. Within this class we found that amphiphilic peptides bearing substituents that carry hydrogen-bonding functionalities in addition to main-chain peptide bonds have a pronounced tendency to selectively form a specific foldamer (see below).

Low-symmetry foldamer identification. DCLs were generated by dissolving building blocks at a 2.0 mM concentration in buffered aqueous solution (pH = 8.0). The presence of oxygen from the air is sufficient to oxidize the thiols into disulfides, which exchange with each other as long as residual thiolate is present in the medium[25]. As a benchmark, a small DCL was made from the parent derivative **1a** with a carboxylate group in the 5-position. Analysis of the library composition using ultra performance liquid chromatography/mass spectrometry (UPLC/MS) showed that it was dominated by small three- and four-membered rings (Fig. 2a). This product distribution reflects the entropic tendency of relatively dilute DCLs to be dominated by the smallest non-strained rings. A survey of a series of dipeptide derivatives of **1a** revealed three types of behaviour (Table 1 and Fig. 2a): only small rings, a range of large rings or mostly one large ring. DCLs made from building block families **1** (which feature hydrophilic peptides) and **2** (which feature hydrophobic amino acid side chains) gave rise to small rings, similar to the behaviour observed for the parent molecule **1a**. Much larger rings

form from building block families **3** and **4**. For building block family **4**, we observed the remarkably selective formation of large rings that ranged from 9-mers to 23-mers, of which several have a prime number of units (13, 17 and 23) rarely encountered in self-assembly and incompatible with symmetry elements other than unlikely high-order axes (C_{13}, C_{17} or C_{23}, respectively). Specifically, on modifying building block **2a**, which formed only a small macrocycle, with a carboxylic acid group (to give **4a**) a 9-mer macrocycle $(\mathbf{4a})_9$ accounted for 75% of the library material. The analogous building block **4b**, which carried a guanidinium group, selectively yielded the 16-mer $(\mathbf{4b})_{16}$. The corresponding chromatogram showed two distinct signals with the same mass. The ^{1}H NMR spectrum of the early eluting peak shows sharp signals, which suggests a conformationally defined species (Supplementary Fig. 90). UPLC analysis of the original sample after lyophilization followed by dissolution in dimethylformamide gave only one main peak that eluted late (Supplementary Fig. 60). ^{1}H NMR analysis of this material revealed broad signals (Supplementary Fig. 92), which indicates a disordered conformation. Thus, the second peak most probably arises from partial unfolding of the material during analysis. Introducing an aminomethyl substituent on the aromatic ring (**4c**) also gave rise to a 16-mer, which was now accompanied by 10% of a 23-mer. Substitution by an amino group (**4d**) led to a 95% conversion into 23-mer $(\mathbf{4d})_{23}$. Methylation of the lysine residue in **4a** and **4d** (to give **4e** and **4f**, respectively) had limited impact on the library composition. In contrast, monomers **4h** and **4g**, the constitutional isomers of **4b** and **4d**, respectively, in which the amino acid sequence was reversed, yielded macrocycles $(\mathbf{4g})_{20}$ and $(\mathbf{4h})_{17}$. UPLC/MS characterization of the samples above is provided in the Supplementary Information (Supplementary Figs. 1–86). We monitored the kinetics of the emergence of the foldamers. For $(\mathbf{4a})_9$ and $(\mathbf{4b})_{16}$, a lag phase was observed during which trimers and tetramers built up prior to the emergence of the foldamers (Supplementary Fig. 87). In contrast, $(\mathbf{4d})_{23}$ was formed without a lag phase, which suggests that the exact pathway by which foldamers form varies between different foldamers. To assess whether the obtained foldamers correspond to the thermodynamic products, we set up libraries in two different ways: either starting from dithiol building blocks or quickly pre-oxidizing 70% of these with sodium perborate, followed by a slower further oxidation mediated by atmospheric oxygen. Both pathways led to identical product distributions (Supplementary Fig. 88), consistent with the products obtained corresponding to the lowest energy state of the system.

Table 1 | Product distribution of DCLs prepared from different building blocks

Entry	R	Ring size
1a	COOH	3, 4
1b	CO-Asp-OH	3, 4
1c	CO-Asp-Lys-OH	3, 4
1d	CO-4-Pro-Lys-OH	3, 4
2a	CO-Phe-Lys-OH	4
2b	CO-Phe(4-CF_3)-Lys-OH	4
2c	CO-Phe(4-OCH_3)-Lys-OH	4
2d	CO-Phe(4-F)-Lys-OH	4
2e	CO-Phe(4-NO_2)-Lys-OH	4
3a	CO-NH-$(CH_2)_2$-$(OCH_2CH_2)_2$-OCH_3	Family of oligomers
3b	CO-Phe-OH	Family of oligomers
3c	CO-Phe(4-OH)-Lys-OH	Family of oligomers
3d	CO-Lys-Phe(4-OH)-OH	Family of oligomers
3e	CO-Ser-Lys-OH	Family of oligomers
4a	CO-Phe(4-COOH)-Lys-OH	9, 12, 13
4b	CO-Phe(4-guanidinium)-Lys-OH	16
4c	CO-Phe(4-CH_2NH_2)-Lys-OH	16, 23
4d	CO-Phe(4-NH_2)-Lys-OH	23
4e	CO-Phe(4-COOH)-Lys(N^6-$(CH_3)_2$)-OH	9
4f	CO-Phe(4-NH_2)-Lys(N^6-(CH_3)-OH	23
4g	CO-Lys-Phe(4-NH_2)-OH	20
4h	CO-Lys-Phe(4-guanidinium)-OH	17
5	CO-Gly-Leu-Lys-X-Lys-OH (X = Phe, Tyr, Ser, Leu)	3, 4, 5, 6, 7, 8

Structure elucidation. The selectivity with which the large rings form for building block family **4** suggests that the rings benefit from specific non-covalent interactions that occur most efficiently with the particular observed ring size; that is, these rings appear to be foldamers. This hypothesis was confirmed through circular dichroism (CD) and NMR spectroscopy, ion mobility–mass spectrometry (IM-MS) and X-ray crystallography. Initial evidence that suggested the large macrocycles adopt a specific folded conformation came from IM-MS. The change in rotationally averaged collision cross-sections (CCS) as a function of the number of subunits in a macrocycle provides information as to the extent of folding. If the macrocycles can be approximated as two-dimensional objects, the CCS is expected to increase linearly with the number of subunits, whereas the formation of compact three-dimensional shapes gives rise to a smaller CCS than that predicted by the linear growth trend[26]. Regardless of the building block used, the smaller macrocycles of all libraries follow such a linear growth trend up to hexamers. In contrast, $(\mathbf{4a})_9$ and, in particular, $(\mathbf{4b})_{16}$ and $(\mathbf{4d})_{23}$ adopt considerably more compact structures than predicted by the linear growth trend (Fig. 2c–e and Supplementary Tables 1–3).

Further evidence that suggests well-defined assemblies came from CD and NMR spectroscopy. Although solutions of the monomers **4a**, **4b** and **4d** show only very weak CD signals, intense and distinct CD bands were observed for $(\mathbf{4a})_9$, $(\mathbf{4b})_{16}$ and $(\mathbf{4d})_{23}$, which indicates that these structures have substantially different folds (Fig. 2f). Solution-phase ^{1}H NMR spectra of the three large macrocycles (D_2O, 298 K) also demonstrated well-defined conformations and limited flexibility, as evident from the presence of remarkably sharp peaks (Supplementary Figs. 89–91). Temperature-dependent NMR (Supplementary Figs. 93–95) and CD experiments (Fig. 2g) show that $(\mathbf{4b})_{16}$ and $(\mathbf{4d})_{23}$ start to unfold only on heating above 70 °C, whereas $(\mathbf{4a})_9$ unfolds at lower temperatures. Refolding was observed for $(\mathbf{4b})_{16}$ and $(\mathbf{4a})_9$ after a heat–cool cycle (Supplementary Figs. 96–99), but for $(\mathbf{4d})_{23}$, the original ellipticity was not fully recovered, which suggests that refolding is not complete. Subsequent UPLC measurements revealed that the composition remained unchanged for all the samples. The ultraviolet spectra of the foldamers remained unchanged on heating (Supplementary Fig. 100).

Decisive evidence for foldamer formation came from analysis by X-ray crystallography of macrocycles $(\mathbf{4b})_{16}$ and $(\mathbf{4d})_{23}$ (Fig. 3, Supplementary Videos 1 and 2 and Supplementary Figs. 101–110). Each structure revealed a well-defined and novel fold. In both cases, the 1,3-dimercaptobenzene macrocycle collapsed to form a dense hydrophobic core, as observed in globular proteins. However, unlike proteins, no hierarchy between secondary and tertiary folding could be distinguished, as distinct secondary elements (such as helices and sheets) were lacking. The dipeptides have peripheral positions

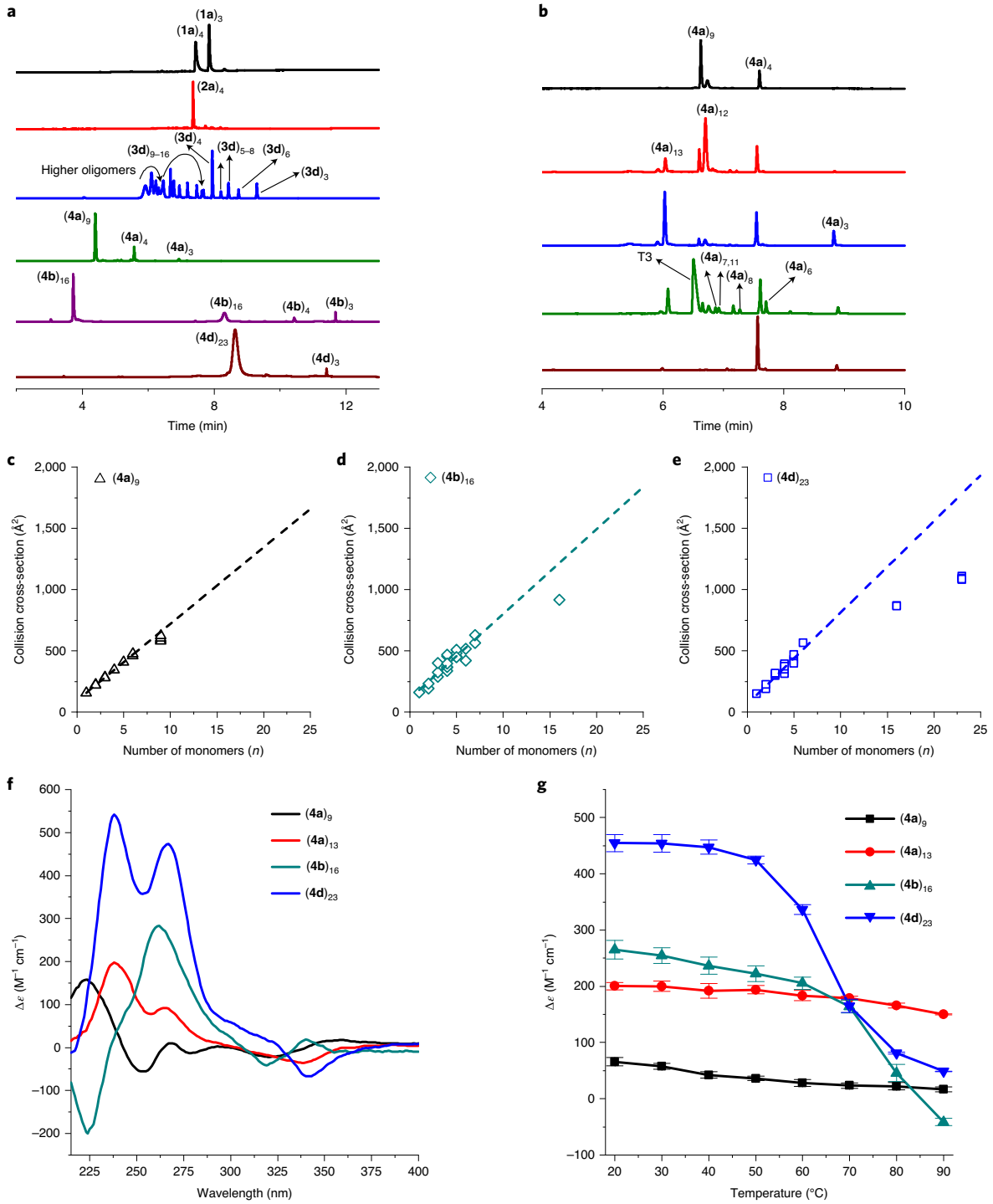

Fig. 2 | Spontaneous formation of large folded macrocycles. a, UPLC chromatograms (absorption at 254 nm) showing compositions after 16 days of stirring of DCLs made from 2.0 mM of **1a**, **2a**, **3d**, **4a**, **4b** and **4d** in 12.5 mM borate buffer pH = 8.0. **b**, UPLC chromatograms of DCLs made from 2.0 mM of **4a** in the presence of (from top to bottom) 1.0 M NaCl, 1.0 M $MgCl_2$, 1.0 M $MnCl_2$, 5.0 mM template **T3** and 1.0 M guanidinium chloride. **c**–**e**, CCS for library members indicate the formation of compact conformations for $(\mathbf{4a})_9$ (**c**), $(\mathbf{4b})_{16}$ (**d**) and $(\mathbf{4d})_{23}$ (**e**). The statistical error of the measured CCSs is smaller than the symbol size. **f**, CD spectra of the libraries prepared from **4a**, **4b** and **4d** after 16 days. **g**, Changes in molar ellipticity at specified wavelengths in the CD spectra of the libraries prepared from **4a** (at 253 nm for $(\mathbf{4a})_9$ and 238 nm for $(\mathbf{4a})_{13}$), **4b** (at 262 nm) and **4d** (at 266 nm) on heating, which indicates different unfolding temperatures. For $(\mathbf{4a})_9$, the absolute value of the CD signal at 253 nm was taken for a better comparison. Data points are the average of three independent experiments with single standard deviations shown.

and form multiple hydrogen bonds. The phenylalanine side chains partially shield the hydrophobic cores and 4-substituents are often involved in intramolecular interaction networks. The charged functions are exposed to the solvent and frequently disordered. The overall structure of $(\mathbf{4b})_{16}$ has only one crystallographic C_2 axis, which thus defines eight distinct environments for **4b**. The symmetry of the core is higher and reflected in the ^{1}H NMR spectrum (Supplementary Fig. 90), which suggests that the arrays of dipeptide

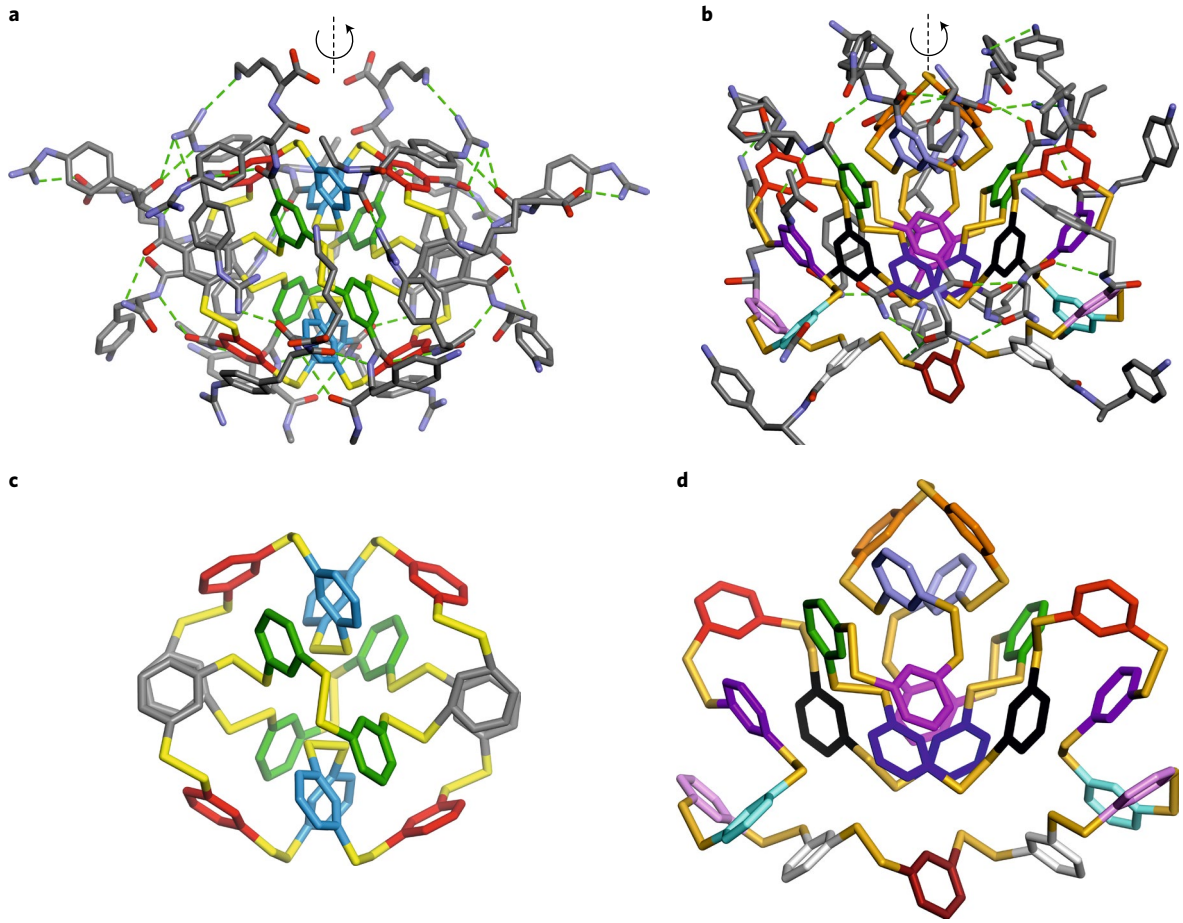

Fig. 3 | Crystal structures of low-symmetry-foldamers. a–d, Crystal structures of $(\mathbf{4b})_{16}$ (**a**,**c**) and $(\mathbf{4d})_{23}$ (**b**,**d**). Hydrogen bonds that involve peptidic side chains are shown as green dashed lines in **a** and **b**. The views in **c** and **d** show the 1,3-dimercaptobenzene hydrophobic backbones alone. Benzene rings are colour coded according to their environment. The core of $(\mathbf{4b})_{16}$ (**c**) possesses three apparent C_2 symmetry axes but two of them do not apply to the dipeptide appendages, which eventually generates eight different environments. The pseudo-C_2 axis in $(\mathbf{4d})_{23}$ generates 12 different environments.

side chains exchange rapidly on the NMR timescale. Similarly, a pseudo-C_2 axis in the structure of $(\mathbf{4d})_{23}$ defines 12 distinct environments for **4d**. This low symmetry contrasts with the C_5-symmetrical, propeller-shaped, chimeric nucleobase–amino acid conjugate found previously[21], and with all the self-assembled homomeric structures described to date, in which the same building block only rarely occupies more than one distinct environment[27].

Diffusion-ordered NMR experiments for $(\mathbf{4b})_{16}$ and $(\mathbf{4d})_{23}$ revealed a diffusion coefficient of $1.2\times10^{-10}\,m^2\,s^{-1}$ for both compounds. The hydrodynamic radius of the molecules (20.5 Å) is consistent with the size observed in the crystal structures (Supplementary Figs. 111 and 112). Additional evidence that folding takes place not only in the solid state but also in solution is provided by temperature-dependent NMR experiments (Supplementary Figs. 93–95), in which, at room temperature, amide protons in the building block were found to exchange rapidly, whereas those in the folded macrocycles exchanged much more slowly. However, on heat-induced (partial) unfolding, they readily exchanged.

Guiding principles for a dynamic combinatorial approach to foldamers. The formation of the large folded macrocycles appears to be the result of a balance between two antagonistic parameters: (1) the high potential of each monomer to form strong interactions via hydrophobic effects, hydrogen bonds, chalcogen bonds and salt bridges observed in the solid-state structures and (2) the geometrical constraints imposed by the semirigidity of the 1,3-dimercaptobenzene backbone and macrocyclization. Small macrocycles may not provide enough degrees of freedom to fulfil a large number of interactions intramolecularly. It is remarkable that, for many derivatives, a unique structure exists that fulfils interactions so much better than others that it constitutes a dominant product. This is not always the case, though, and **3a**–**3e**, which possess fewer hydrogen bonding units than **4a**–**4h**, produce a range of large rings. Remarkably, macrocycle growth and folding occur in preference over aggregation, which may have constituted alternative pathways to fulfil interactions[25], albeit not necessarily more favourable on entropic grounds.

Our survey of structural space allows us to formulate guidance to access new foldamers by dynamic combinatorial chemistry. First, the main chain should have some degree of rigidity. Too flexible a main chain gives more chances to fulfil a large amount of interactions in small entropically favoured structures. Furthermore, semirigidity favours distinct main-chain conformations for different ring sizes. Second, each monomer should be endowed with both multiple and diverse sites for non-covalent interactions. Multiple interactions are required to provide the necessary energy to stabilize otherwise disfavoured large rings. Diversity is required to prevent the fulfilment of all interactions in a simple mode, be it intramolecular or intermolecular, which would lead to the commonly observed high symmetry of self-assembled structures. In other words, a certain degree of frustration is necessary to obtain asymmetrical objects[5]. From the different monomers in the solid-state structures of $(\mathbf{4b})_{16}$

and $(\mathbf{4d})_{23}$, it is remarkable that their potential for interactions is not always fulfilled and not always in the same manner: some frustration remains, but it has been minimized. These structures also suggest that, in water, amphiphilic building blocks are beneficial as they allow for the formation of a hydrophobic core surrounded by a hydrophilic shell. The hydrophobic/hydrophilic balance should ideally be such that packing away all the hydrophobic surface area is only possible in a large oligomer. The balance between the multiplicity and diversity of interactions is delicate. For instance, elongating the peptide appendages, as in **5**, does not increase diversity, but leads to the emergence of a dominant β-sheet pattern. Third, and obviously, some of the interactions must be directional to make the different library members have different structures and therefore different energies. Directional interactions may preferably be in part conflicting to avoid the prevalence of symmetrical objects. Admittedly, these guidelines do not allow us to predict which ring size and which structure may emerge, but they indicate in which part of the chemical space one should focus.

Template-induced foldamer formation. The dynamic combinatorial approach to folded macromolecules has as yet unrealized potential for the development of functional foldamers. Specifically, the addition of a template molecule that can bind to a constitutionally dynamic foldamer should lead to the amplification of this species at the expense of other library members. After isolation, the amplified foldamer should be able to selectively recognize and bind to its template. Our attempts to implement this strategy focused on the DCL prepared from **4a**, which gave foldamer $(\mathbf{4a})_9$. We reasoned that the relatively low stability of this foldamer (Supplementary Fig. 93) would make the template-induced formation of other foldamers easier. Furthermore, the presence of carboxylate groups in **4a** makes this building block potentially suitable for binding to metal ions. Indeed, the addition of NaCl, $MgCl_2$ or $MnCl_2$ to a solution of **4a** (top three traces in Fig. 2b) resulted in the formation of $(\mathbf{4a})_{12}$ and $(\mathbf{4a})_{13}$. In the presence of 1.0 M guanidinium chloride (a widely used denaturant of proteins), these effects were suppressed (bottom trace in Fig. 2b). These template effects are due to the metal cations, as different anions did not affect the composition of the library, except for I^-, which favoured $(\mathbf{4a})_4$ (Supplementary Fig. 113). Macrocycle $(\mathbf{4a})_{13}$ was isolated in 90% purity using flash column chromatography (Supplementary Fig. 114) and its binding to $MnCl_2$ was investigated using isothermal titration calorimetry (ITC), which revealed a binding constant of $4\times10^4\,M^{-1}$ (Supplementary Fig. 115). Interestingly, $(\mathbf{4a})_9$ and $(\mathbf{4a})_{13}$ exhibit substantially different folds (Fig. 2f), even though they are made from the same building block.

We also investigated whether peptides that resemble the dithiol monomers would template the formation of foldamers (Fig. 1c). Thus, dipeptides **T1**–**T3** were mixed with libraries made from building blocks **4a**, **4b** or **4d**. We found that **T3** strongly influenced the macrocyclization of **4a**, and eventually favoured $(\mathbf{4a})_{13}$ (35% conversion at a 1:2.5 **4a**:**T3** ratio Fig. 2b). ITC experiments (Supplementary Fig. 116) revealed that $(\mathbf{4a})_{13}$ binds to **T3** with a K_a of $8.1\times10^3\,M^{-1}$. **T1** and **T2** showed considerably lower affinities (Supplementary Figs. 117 and 118). Additionally, smaller rings, such as $(\mathbf{4a})_4$, showed a limited affinity to **T3** (Supplementary Fig. 119). No template effects were observed in libraries made from **4b** and **4d** (Supplementary Figs. 120–126), possibly due to the higher thermodynamic stability of $(\mathbf{4b})_{16}$ and $(\mathbf{4d})_{23}$ compared with that of $(\mathbf{4a})_9$.

Conclusion

Using a systems chemistry approach, we discovered the spontaneous or template-induced formation of foldamers of unprecedented complexity. Unlike in the large constructs obtained previously[1–4,27], building blocks that were identical at the start of the experiment end up in as many as 12 distinct environments on their incorporation into a foldamer. The 1,3-dimercaptobenzene core proved to be versatile, as it adopted varied conformations depending on the side chain it carried. This is similar to how peptide side chains direct protein folding with the major difference that complex folds are shown here to emerge from homomeric sequences. Rules are starting to emerge for the design of the building blocks that give rise to foldamers in DCLs. Folding is a cooperative phenomenon that requires large molecules. Thus, the inherent tendency to form small-ring assemblies needs to be overcome, which can be achieved using hydrophobically frustrated building blocks that are unable to shield their hydrophobic parts from water when part of a small oligomer. The preferential formation of one particular foldamer requires it to be substantially more stable than other potential foldamers of similar sizes. Such differences in stability probably require the presence of a sufficiently large number of non-covalent interactions with relatively strong angle and distance dependence, such as, for example, hydrogen bonds. Indeed, the main difference between the behaviour of building blocks in family **3** (to form a range of large rings) and family **4** (to form one specific foldamer) is the higher propensity to form hydrogen bonds in the latter. Our results have relevance in diverse areas of chemistry. The spontaneous and uniquely selective oligomerization of simple monomer units into oligomers that are of uniform length and are conformationally well-defined, as demonstrated here, has been a long-standing objective in polymer science. Dendrimers meet some of these objectives[28]. Yet, unlike dendrimers, which require multistep synthesis, the present foldamers can be obtained in a one-pot procedure and are conformationally better defined. We also showed that templates can direct the formation of new functional foldamers. The resulting structures carry imprints of the templates and are therefore capable of binding these. This behaviour resembles that of molecularly imprinted polymers[29], with the important distinction that the selected oligomers are uniform and conformationally well-defined, whereas molecularly imprinted polymers suffer from heterogeneity in structure, shape and size.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41557-020-00565-2.

Received: 5 December 2019; Accepted: 21 September 2020;
Published online: 20 November 2020

References

1. Fujita, D. et al. Self-assembly of tetravalent Goldberg polyhedra from 144 small components. *Nature* **540**, 563–566 (2016).
2. Bale, J. B. et al. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science* **353**, 389–394 (2016).
3. Pasquale, S., Sattin, S., Escudero-Adan, E. C., Martinez-Belmonte, M. & de Mendoza, J. Giant regular polyhedra from calixarene carboxylates and uranyl. *Nat. Commun.* **3**, 785 (2012).
4. Sasaki, E. et al. Structure and assembly of scalable porous protein cages. *Nat. Commun.* **8**, 14663 (2017).
5. Anderson, K. M., Goeta, A. E. & Steed, J. W. Supramolecular synthon frustration leads to crystal structures with $Z'>1$. *Cryst. Growth Des.* **8**, 2517–2524 (2008).
6. Banerjee, R., Bhatt, P. M., Kirchner, M. T. & Desiraju, G. R. Structural studies of the system Na(saccharinate)·nH$_2$O: a model for crystallization. *Angew. Chem. Int. Ed.* **44**, 2515–2520 (2005).
7. Rhys, G. G. et al. Maintaining and breaking symmetry in homomeric coiled-coil assemblies. *Nat. Commun.* **9**, 4132 (2018).
8. Whitesides, G. M. & Ismagilov, R. F. Complexity in chemistry. *Science* **284**, 89–92 (1999).
9. Lehn, J.-M. From supramolecular chemistry towards constitutional dynamic chemistry and adaptive chemistry. *Chem. Soc. Rev.* **36**, 151–160 (2007).
10. Lehn, J.-M. Constitutional dynamic chemistry: bridge from supramolecular chemistry to adaptive chemistry. *Top. Curr. Chem.* **322**, 1–32 (2012).

11. Vantomme, G. & Meijer, E. W. The construction of supramolecular systems. *Science* **363**, 1396–1397 (2019).
12. Nitschke, J. R. Systems chemistry: molecular networks come of age. *Nature* **462**, 736–738 (2009).
13. Li, J., Nowak, P. & Otto, S. Dynamic combinatorial libraries: from exploring molecular recognition to systems chemistry. *J. Am. Chem. Soc.* **135**, 9222–9239 (2013).
14. Corbett, P. T. et al. Dynamic combinatorial chemistry. *Chem. Rev.* **106**, 3652–3711 (2006).
15. Cougnon, F. B. L. & Sanders, J. K. M. Evolution of dynamic combinatorial chemistry. *Acc. Chem. Res.* **45**, 2211–2221 (2012).
16. Tsiamantas, C. et al. Selective dynamic assembly of disulfide macrocyclic helical foldamers with remote communication of handedness. *Angew. Chem. Int. Ed.* **55**, 6848–6852 (2016).
17. Steinkruger, J. D., Woolfson, D. N. & Gellman, S. H. Side-chain pairing preferences in the parallel coiled-coil dimer motif: insight on ion pairing between core and flanking sites. *J. Am. Chem. Soc.* **132**, 7586–7588 (2010).
18. Hadley, E. B., Testa, O. D., Woolfson, D. N. & Gellman, S. H. Preferred side-chain constellations at antiparallel coiled-coil interfaces. *Proc. Natl Acad. Sci. USA* **105**, 530–535 (2008).
19. Krishnan-Ghosh, Y. & Balasubramanian, S. Dynamic covalent chemistry on self-templating peptides: formation of a disulfide-linked beta-hairpin mimic. *Angew. Chem. Int. Ed.* **42**, 2171–2173 (2003).
20. Oh, K., Jeong, K. S. & Moore, J. S. Folding-driven synthesis of oligomers. *Nature* **414**, 889–893 (2001).
21. Liu, B. et al. Complex molecules that fold like proteins can emerge spontaneously. *J. Am. Chem. Soc.* **141**, 1685–1689 (2019).
22. Carnall, J. M. A. et al. Mechanosensitive self-replication driven by self-organization. *Science* **327**, 1502–1506 (2010).
23. Komáromy, D. et al. Self-assembly can direct dynamic covalent bond formation toward diversity or specificity. *J. Am. Chem. Soc.* **139**, 6234–6241 (2017).
24. Bartolec, B., Altay, M. & Otto, S. Template-promoted self-replication in dynamic combinatorial libraries made from a simple building block. *Chem. Commun.* **54**, 13096–13098 (2018).
25. Sadownik, J. W., Mattia, E., Nowak, P. & Otto, S. Diversification of self-replicating molecules. *Nat. Chem.* **8**, 264–269 (2016).
26. Seo, J. et al. An infrared spectroscopy approach to follow β-sheet formation in peptide amyloid assemblies. *Nat. Chem.* **9**, 39–44 (2016).
27. Käseborn, M., Holstein, J. J., Clever, G. H. & Lützen, A. A rotaxane-like cage-in-ring structural motif for a metallosupramolecular Pd_6L_{12} aggregate. *Angew. Chem. Int. Ed.* **57**, 12171–12175 (2018).
28. Huang, B., Prantil, M. A., Gustafson, T. L. & Parquette, J. R. The effect of global compaction on the local secondary structure of folded dendrimers. *J. Am. Chem. Soc.* **125**, 14518–14530 (2003).
29. BelBruno, J. J. Molecularly imprinted polymers. *Chem. Rev.* **119**, 94–119 (2019).

Methods

General procedures. All the chemicals, unless otherwise stated, were purchased from Sigma-Aldrich and used as received. Amino acid resins were purchased from Novabiochem and Fmoc-modified phenylalanine amino acid residues from Chem Impex International and Iris Biotech. Acetonitrile (ULC-MS grade), water (ULC-MS grade) and trifluoroacetic acid (HPLC grade) were purchased from Biosolve BV. Flash column chromatography was performed on a Reveleris X2 Flash Chromatography System (Grace Davison Discovery Sciences) on normal or reversed phase silica cartridges. NMR spectra were recorded on Bruker 500 and 600 MHz spectrometers.

Peptide synthesis. The synthesis of the monomers was performed using conventional solid-phase peptide synthesis using the Fmoc/^{t}Bu protecting group strategy on Wang resin. Amino acids were introduced protected as Fmoc-Lys(Boc)-OH, Fmoc-Tyr(Boc)-OH, Fmoc-Phe(4-NH-Boc)-OH or Fmoc-Phe(4-COOtBu)-OH. Fmoc deprotection steps were carried out with 20% piperidine in dimethylformamide (v/v) for 15 min. Coupling of Fmoc amino acids was performed in dimethylformamide using *N*-diisopropylcarbodiimide and ethyl cyano(hydroxyimino)acetate (oxyma). Deprotection from the resin and removal of the protecting groups on the side chains of the amino acids were performed using a cocktail of 95% trifluoroacetic acid, 2.5% 1,2-ethanedithiol, 1.25% water and 1.25% triisopropylsilane for 3 h. Crude peptides were purified using flash column chromatography and obtained at a purity level >97%. The ≤3% impurities were dominated by disulfide oligomers.

Library preparation. Building blocks were dissolved in borate buffer (12.5 mM in $Na_2B_4O_7$, pH = 8.0). Where necessary, the pH of the solution was adjusted by the addition of 1.0 M NaOH. All the libraries were set up in an HPLC vial (12 × 32 mm) with a Teflon-coated screw cap. All the HPLC vials were equipped with a cylindrical stirrer bar (2 × 5 mm, Teflon coated, purchased from VWR) and stirred at 1,200 r.p.m. using an IKA RCT basic hot plate stirrer. All the experiments were performed at ambient conditions.

Buffer preparation. Borate buffer (12.5 mM in $Na_2B_4O_7$, pH = 8.0) was prepared by dissolving boric acid anhydride (87.0 mg, B_2O_3) in 50 ml of doubly distilled water. The pH was adjusted to 8.0 using concentrated NaOH.

UPLC analysis. UPLC analyses were performed on a Waters Acquity H and I-class system equipped with a photodiode array detector at a detection wavelength of 254 nm. Samples were injected on an Aeris WIDEPORE 3.6 µm BEH-C18 (150 × 2.1 mm) column, using ULC-MS grade water (eluent A) and ULC-MS grade acetonitrile (eluent B), which contained 0.1 v/v trifluoroacetic acid as the modifier. A flow rate of 0.3 ml min^{-1} and a column temperature of 35 °C were applied. Sample preparation was performed by diluting 5 µl of the library with 100 µl of double distilled water.

UPLC-MS analysis. UPLC-MS analyses were performed using a Waters Acquity UPLC H-class system coupled to a Waters Xevo-G2 TOF. The mass spectrometer was operated in a positive electrospray ionization mode with the ionization parameters: capillary voltage, 3 kV; sampling cone voltage, 20 V; extraction cone voltage, 4 V; source gas temperature, 120 °C; desolvation gas temperature, 450 °C; cone gas flow (nitrogen), 1 l h^{-1} and desolvation gas flow (nitrogen), 800 l h^{-1}.

CD spectroscopy. Spectra were recorded on a Jasco J-810 spectrometer with a Peltier temperature controller. Heat–cool cycles were applied from 20 to 90 °C in steps of 1° at a rate of 0.1° min^{-1} and maintained for 10 min at every temperature before measuring. Spectra were obtained as averages of three measurements from 210 to 400 nm with a scanning speed of 150 nm min^{-1} and a bandwidth of 1 nm. A quartz cuvette with a 1 cm path length was used for the measurements.

ITC. The binding studies were conducted using VP-ITC (Microcal LLC, GE Healthcare). Solutions of $MnCl_2$ and templates **T1**–**T3** (in 12.5 mM borate buffer pH = 8.0) were titrated into solutions of hosts $(\mathbf{4a})_{13}$ and $(\mathbf{4a})_4$ (in 12.5 mM borate buffer, pH = 8.0). Binding constants and enthalpies of binding were obtained by curve fitting of the titration data using the one-site binding model available in the Origin 2.9 software, as shown in the Supplementary Information.

Diffusion-ordered NMR spectroscopy. Diffusion-ordered NMR spectra were acquired at 600 MHz with a Bruker AVANCE NEO 600 spectrometer using the Bruker pulse program stebpgp1s at 298 K. The diffusion time was adjusted to 150 ms. The pulse gradient was increased from 2 to 95% of the maximum gradient strength using a linear ramp. For the F2 dimension, 65,000 data points (12 ppm) were collected with 16 data points for the F1 dimension.

IM-MS. Linear drift tube IM-MS measurements were performed on a modified Synapt G2-S HDMS (Waters Corporation) and on a home-built instrument (iMob), both described in detail elsewhere[30,31]. For the nano-electrospray, typically 5 µl of the sample were loaded and electrosprayed by applying a 0.4–0.8 kV capillary voltage. Further parameters on the modified Synapt instrument were: 100 V sampling-cone voltage, 1 V source-offset voltage, 30 °C source temperature, 150 °C desolvation gas temperature, 50 l h^{-1} cone gas flow, 500 l h^{-1} desolvation gas flow, 0 V trap CE (Collision Energy), 0 V transfer CE and 3 ml min^{-1} trap gas flow. Ion mobility parameters were: 2.2 torr helium IM-MS gas, 27–29 °C IM-MS temperature, 2.0 V trap d.c. entrance voltage, 5.0 V trap d.c. bias voltage, −10.0 V trap d.c. voltage, 5.0 V trap d.c. exit voltage, −25.0 V IM-MS d.c. entrance voltage, 60–200 V helium cell d.c. voltage, −40.0 V helium exit voltage, 60 V IM-MS bias voltage, 0 V IM-MS d.c. exit voltage, 5.0 V transfer d.c. entrance voltage, 15.0 V transfer d.c. exit voltage, 150 m s^{-1} trap wave velocity, 2.0 V trap wave height voltage, 200 m s^{-1} transfer wave velocity and 5.0 V transfer wave height voltage. Further parameters on the iMob instrument were: 10–20 V cm^{-1} electric field inside the ion mobility tubes, ~4 mbar helium IM-MS gas and ~23 °C IM-MS temperature. IM-MS spectra were recorded in the positive ion mode (except for the **4a** library which was measured in negative ion polarity) and measured drift times were converted into rotationally averaged CCS using the Mason–Schamp equation[32].

Crystallization of L-$(\mathbf{4b})_{16}$ and L/D-$(\mathbf{4d})_{23}$. Aqueous solutions of L-$(\mathbf{4b})_{16}$ and of L/D-$(\mathbf{4d})_{23}$ were prepared using pure water to a final concentration of 25 mg ml^{-1}. Racemic $(\mathbf{4d})_{23}$ was prepared by mixing the enantiopure solutions of L-$(\mathbf{4d})_{23}$ and D-$(\mathbf{4d})_{23}$. Crystallization trials of each of these solutions were carried out with commercial sparse matrix screens (JBScreen Basic, Jena Bioscience) using the standard sitting drop vapour diffusion method at 293 K. X-ray quality crystals of L-$(\mathbf{4b})_{16}$ (Supplementary Fig. 101) were observed after 5 days by the addition of 1 µl of the L-$(\mathbf{4b})_{16}$ solution and 1 µl of 30% w/v PEG 4000, 100 mM Tris buffer (pH = 8.5) and 200 mM magnesium chloride from the reservoir solution. X-ray quality crystals of L/D-$(\mathbf{4d})_{23}$ (Supplementary Fig. 101) were optimized using the hanging drop method by mixing 0.8 µl of the L/D-$(\mathbf{4d})_{23}$ solution and 2.2 µl of 30% w/v PEG 4000, 100 mM Tris buffer (pH = 8.0) and 200 mM lithium sulfate from the reservoir solution. Crystals for both L-$(\mathbf{4b})_{16}$ and L/D-$(\mathbf{4d})_{23}$ were fished using micro loops and plunged into liquid nitrogen directly such that the respective mother liquor served as cryoprotectant.

Data collection and structure determination of L-$(\mathbf{4b})_{16}$. The X-ray diffraction data were collected at the ID23-1 beamline at the European Synchrotron Radiation Facility with a Dectris Pilatus 6 M detector[33]. Diffraction data were measured at $T = 100$ K and $\lambda = 0.972$ Å. The crystal was exposed for 0.1 s with a 0.15° oscillation per frame. Diffraction data were processed using the program XDS[34]. The crystal belonged to the space group $P6_2$ with unit cell parameters $a = b = 36.980(5)$ Å and $c = 44.071(9)$ Å, $V = 52{,}194(15)$ Å^3 and half a molecule per asymmetric unit ($Z = 3$). The structure was solved with the program SHELXT[35] and refined by the full-matrix least-squares method on F^2 with SHELXL-2014[36] within Olex2[37]. After each refinement step, visual inspections of the model and the electron-density maps were carried out using Olex2 and Coot[38]. Some side-chain atoms of phenyl guanidium and lysine were observed to be severely disordered and were either omitted or refined with partial occupancy and isotropic-displacement parameters. AFIX, DFIX and FLAT instructions were used to improve the geometry of the molecules. Restraints on anisotropic displacement parameters were implemented with DELU, SIMU, RIGU and ISOR instructions. After several attempts to model the disordered side chains, the SQUEEZE procedure was used to flatten the electron density map[39]. Very disordered side chains and solvent molecules were removed. The calculated total potential solvent accessible void volume and electron counts per cell were 26,111 Å^3 and 9,476, respectively. Hydrogen atoms were placed at idealized positions, except for those at disordered or missing side chains.

Data collection and structure determination of L/D-$(\mathbf{4d})_{23}$. The X-ray diffraction data were collected at the EMBL P13 beamline, Hamburg, with a Dectris Pilatus 6M detector[40]. Diffraction data were measured at $T = 100$ K and $\lambda = 0.980$ Å. The crystal was exposed for 0.04 s and a 0.1° oscillation per frame. Diffraction data were processed using the program CrysAlis^{Pro}[41]. The crystal belonged to the space group $P1$ with unit cell parameters of $a = 35.114(2)$ Å, $b = 49.232(3)$ Å, $c = 61.528(5)$ Å, $\alpha = 77.641(1)°$, $\beta = 89.369(1)°$, $\gamma = 71.229(1)°$, $V = 98{,}185.4(14)$ Å^3 and four molecules per unit cell ($Z = 4$). The structure was solved with the program SHELXT[35] and refined by a full-matrix least-squares method on F^2 with SHELXL-2014[36] within Olex2[37]. The initial structure revealed most of the main-chain atoms and few of the side chains of two macrocycles out of four. After several iterations of least-squares refinement, the main-chain trace improved for the third and fourth macrocycles. Anisotropic refinement was carried out only for the sulfur atoms. AFIX, DFIX and FLAT instructions were used to improve the geometry of molecules. After several attempts to model the disordered side chains, the *SQUEEZE* procedure was used to flatten the electron density map[39]. Very disordered side chains were removed. The calculated total potential solvent accessible void volume and electron count per cell were 72,418.1 Å^3 and 2,951, respectively. Hydrogen atoms were not added due to the poor quality of the data.

Crystallographic data, refinement statistics and comments on checkcif alerts obtained from IUCr's checkcif algorithm are reported in Supplementary Table 4.

Data availability

The authors declare that all the data supporting the findings of this study are available within the article, in the source data files and in the Supplementary Information. Mass and NMR spectra are stored locally in native format and are available upon request. Crystallographic data for the structures reported in this article have been deposited at the Cambridge Crystallographic Data Centre, under deposition numbers CCDC 1942977 for $(\mathbf{4b})_{16}$ and 1999456 for $(\mathbf{4d})_{23}$, respectively. Copies of the data can be obtained free of charge via https://www.ccdc.cam.ac.uk/structures/. Source data are provided with this paper.

References

30. Allen, S. J., Giles, K., Gilbert, T. & Bush, M. F. Ion mobility mass spectrometry of peptide, protein, and protein complex ions using a radio-frequency confining drift cell. *Analyst* **141**, 884–891 (2016).
31. Warnke, S., von Helden, G. & Pagel, K. Analyzing the higher order structure of proteins with conformer-selective ultraviolet photodissociation. *Proteomics* **15**, 2804–2812 (2015).
32. Revercomb, H. E. & Mason, E. A. Theory of plasma chromatography/gaseous electrophoresis. Review. *Anal. Chem.* **47**, 970–983 (1975).
33. Nurizzo, D. et al. The ID23-1 structural biology beamline at the ESRF. *J. Synchrotron Rad* **13**, 227–238 (2006).
34. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
35. Sheldrick, G. M. SHELXT-integrated space-group and crystal-structure determination. *Acta Crystallogr. A* **71**, 3–8 (2015).
36. Sheldrick, G. M. Crystal structure refinement with SHELXL. *Acta Crystallogr. C* **71**, 3–8 (2015).
37. Dolomanov, O. V., Bourhis, L. J., Gildea, R. J., Howard, J. A. K. & Puschmann, H. OLEX2: a complete structure solution, refinement and analysis program. *J. Appl. Cryst.* **42**, 339–341 (2009).
38. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
39. Spek, A. L. Structure validation in chemical crystallography. *Acta Crystallogr. D* **65**, 148–155 (2009).
40. Cianci, M. et al. P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focussing. *J. Synchrotron Rad.* **24**, 323–332 (2017).
41. CrysAlisPRO (Agilent Technologies Ltd, 2014).

Acknowledgements

We thank P. van der Meulen and J. Kemmink for their assistance with the NMR experiments and analysing the data. We thank I. Melnikov (ID23-1, ERSF) and G. Pompidor (PETRA III, DESY) for assistance during data collection at the synchrotron beamlines. This research was supported by the ERC (AdG 741774), the EU (MCIF 745805–DSR), NWO (VICI grant), Zernike Dieptestrategie and the Dutch Ministry of Education, Culture and Science (Gravitation program 024.001.035).

Author contributions

S.O. and I.H. supervised the overall project. C.G.P. conceived and designed the study, synthesized the building blocks, analysed the DCL compositions by UPLC and isolated and characterized the foldamers by CD and NMR spectroscopy. B.L. performed the UPLC-MS experiments and analysed the data. P.K.M. and B.K. carried out the crystallographic studies. X.M. and K.L. performed the ITC experiments. D.K., W.H., C.M., R.C. and K.P. performed the IM-MS experiments and analysed the data. C.G.P., I.H. and S.O. co-wrote the paper. All the authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41557-020-00565-2.

Correspondence and requests for materials should be addressed to I.H. or S.O.

Reprints and permissions information is available at www.nature.com/reprints.