Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Oligopeptide-CB[8] complexation with switchable binding pathways†

Guanglu Wu,‡[a] David E. Clarke,‡[a] Ce Wu,[a] and Oren A. Scherman*[a]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Host-guest complexes exhibiting a 1:1 binding stoichiometry need not consist of a single host and guest. A series of oligopeptides, which were previously reported to have abnormally high binding enthalpies were investigated to deduce whether they exist as a 2:2 quaternary or a 1:1 binary complex with cucurbit[8]uril (CB[8]). Through a systematic study of the sequence-specific binding pathways of peptide-CB[8] association, a phenylalanine-leucine dipeptide was found to be capable of switching from a 1:1 stoichiometric complex to a 2:1 complex. By studying the differences in size-based diffusion properties of these two binding modes, the presence of a 1:1 pairwise inclusion complex was verified for the regime where CB[8] is in excess. Findings in this study can be utilised to 'customise' the precise CB[8]-oligopeptide self-assembly pathway, acting as a useful toolbox in the design of supramolecular systems.

1 Introduction

Cucurbit[n]urils (CB[n]) are a family of synthetic macrocyclic host molecules capable of encapsulating a wide range of organic guests in aqueous media with exceptional affinity.[1–4] They have also been shown to bind with biomolecules such as amino acids, peptides, and proteins; demonstrating a preference for aromatic residues such as tryptophan (W), phenylalanine (F), and tyrosine (Y).[5,6] Among the cucurbituril homologues, CB[8] is versatile on account of its ability to form both hetero- and homoternary complexes with these amino acids, accommodating two different guests (e.g. methyl viologen and tryptophan[5]) or two identical guests, respectively (e.g. phenylalanine[6]). The discovery of CB[8]-mediated ternary complexes has led to their application in molecular[4] or chiral[7] recognition, supramolecular catalysis,[8–10] and engineered nanostructures.[11,12]

The host-guest binding of CB[n] in aqueous media is mainly driven by the release of energetically-frustrated cavity-bound water (high-energy water),[13,14] whereby free energy is contributed mostly from enthalpy and is directly related to the number of water molecules that are displaced. For instance, the formation of a binary complex in aqueous solution typically will release less heat than the formation of a ternary complex, as only one guest displaces a smaller number of cavity-bound water molecules.[15] This means that the enthalpy change from binding can provide insight into the exact composition of the CB[8]-mediated host-guest complex.

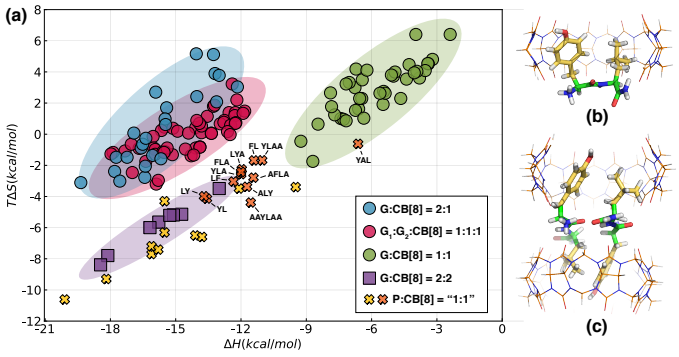

Fig. 1 a) Thermodynamic data of CB[8]-mediated complexations including 2:1 homoternary (blue circles), 1:1:1 heteroternary (red circles), 1:1 binary (green circles), 2:2 quaternary (purple square),[15] and "1:1" peptide-CB[8] complexes (crosses), where the yellow crosses represent data previously reported by Urbach *et al.*[16,17] and orange crosses represent data collected in this work. The two possible binding motifs to be distinguished and verified are b) one CB[8] encapsulating two neighboring amino acid residues from one peptide within its cavity and c) two CB[8] each of which encapsulates two residues from two antiparallel-aligned peptides. N.B. The dipeptide backbones (with green carbons) are not derived from computational simulation results but are directly taken from existed FF dipeptide crystals.

Recently, unique CB[8]-mediated complexes of 1:1 binding stoichiometry have also been observed.[15–19] These complexes have exhibited abnormally large binding enthalpies when compared to typical enthalpy changes for 1:1 binary complexation (Fig. 1a).

[a] *Melville Laboratory for Polymer Synthesis, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: oas23@cam.ac.uk*

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

‡ These authors contributed equally to this work.

Through further exploration of this phenomena, we found that diaryl-viologens attributed with electron-donating substituents as well as azobenzene derivatives preferred to form 2:2 quaternary complexes with CB[8].[15,18] In these complexes, two partially-stacked guests are 'clamped' together by two CB[8] macrocycles resulting in a multi-component complex whose exchange dynamics are substantially retarded compared to typical binary complexes (henceforth referred to as 'static' throughout). Other examples of 1:1 binding stoichometry exhibiting abnormally high binding enthalpy are the association of CB[8] with sequence-specific tripeptides such as tyrosine-leucine-alanine (YLA)[16] and methionine-leucine-alanine (MLA)[17] both reported by Urbach *et al.*[16,17] Although a binding motif was proposed where one CB[8] accommodated two neighbouring amino acid residues within the cavity (Fig. 1b),[16] their computational simulations were solely based on a 1:1 binary model. The simulations did not take into account other possible n:n stoichometric binding modes such as a 2:2 quaternary complex, where each CB[8] encapsulates two residues from two antiparallel-aligned peptides (Fig. 1c).

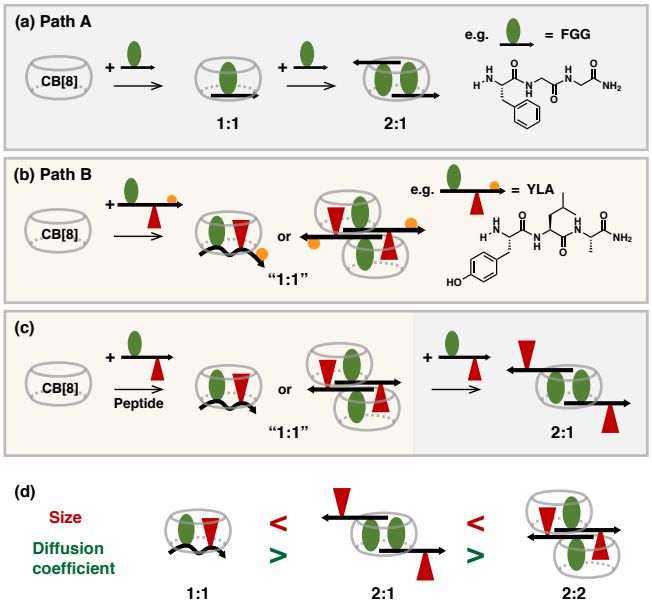

Fig. 2 The association of peptides with CB[8] often takes place along two typical pathways: a) Path A, involves a sequential formation of the binary and homo-ternary complex with increasing peptide to CB[8] ratio, a typical case is Phe-Gly-Gly (FGG);[6,20] b) Path B, involves the formation of a "1:1" stoichometric complex, in which CB[8] encapsulates two residues in a ternary manner and remains unaltered even in the presence of excess peptide, a typical case is Tyr-Leu-Ala (YLA).[16] Herein, we pursue a peptide that c) associates with CB[8] along Path B when excess CB[8] is present and turns to Path A when peptide is in excess. d) Through comparing the size or diffusion coefficent of the complex in these two different configurations, we can confirm the binding model of the "1:1" stoichometric complex.

The practicality of distinguishing between binding motifs of the same stoichometry, especially for systems with small differences such as oligopeptide-CB[8] complexes, remains extremely challenging. Both 1:1 and 2:2 binding motifs depicted in Fig. 1 can readily explain the large enthalpy changes observed (Fig. 1, Table S1), as the CB[8] in either motif includes two pendant groups within its cavity, resembling ternary complexation. The ^{1}H-NMR and ROESY studies of the YLA@CB[8] complex revealed that both the side-chains of the Y and L resided inside the CB[8] cavity and were in close proximity to one another.[16] However, it cannot be distinguished whether the Y and L is derived from the same tripeptide (Fig. 1b) or from two antiparallel-aligned YLA peptides (Fig. 1c).

The most abundant signals detected by mass spectrometry (in the gas phase) do not necessarily correlate to the most abundant species in solution, especially for such complexes, which is driven by desolvation.[15,21] Direct injection of a 2:2 diarylviologen-CB[8] solution yields an intense signal, which corresponds to the complex consisting of one guest with one CB[8], suggesting the dissociation of 2:2 complexes in the gas phase.[15] Analysing signals of lower intensity in the mass spectrum of YLA@CB[8] we observed various aggregation species that contained multiple CB[8] and multiple peptide molecules (Fig. S14), which did not allow us to distinguish between the two binding modes. Moreover, YAL@CB[8] exhibits almost the same mass spectrum as YLA@CB[8] in gas phase (Fig. S13-16), although their bindings in aqueous solution are completely different. In light of these observations, it is not clear whether the 1:1 stoichometric binding mode of a peptide-CB[8] complex exists as either a 1:1 binary complex or a 2:2 quaternary complex.

One possible approach is to detect size-based diffusion properties of different binding modes using *in situ* techniques like diffusion ordered spectroscopy (DOSY). However, the results derived from DOSY NMR are often susceptible to molecular properties and strongly correlate to the experimental and instrumental conditions. For instance, one cannot readily distinguish if the small difference in diffusion coefficients observed between the FGG_2@CB[8] and YLA@CB[8] systems is caused by the structural differences of the oligopeptides or by their different binding modes with CB[8], or merely due to apparatus error. Therefore, devising a system that can display different binding modes with a minimum perturbation to the system parameters is highly desirable.

Herein, we revisit the sequence-specific binding behaviour of oligopeptides with CB[8], to deduce whether they exist as a 2:2 or a 1:1 binary complex with CB[8] (Fig. 2). We hypothesised that by investigating the differences in size-based diffusion properties, it would allow us to distinguish between the different binding modes. To facilitate this size comparison, we designed and studied oligopeptide sequences in an attempt to program a versatile complex that can exhibit both a 1:1 stoichometry (either 1:1 or 2:2) when excess CB[8] is present and a 2:1 complex when the peptide is in excess (Fig. 2c). Subsequently, by comparing the sizes of the formed complexes in these two regimes, we will be able to distinguish a 1:1 pairwise inclusion from a 2:2 quaternary complex, as the size of 1:1, 2:1, and 2:2 species have ascending molecular sizes (Fig. 2d).

2 Results and Discussion

To generate the desired versatile system, which can switch between 1:1 and 2:1 binding modes, our peptide design would need

to suppress the 1:1 path B and maximise the 2:1 path A in conditions with excess peptide. Therefore, the effects of altering the sequence on binding behaviour was studied, including the sequential order of the effective residues, as well as chain extension at both the C and N termini.

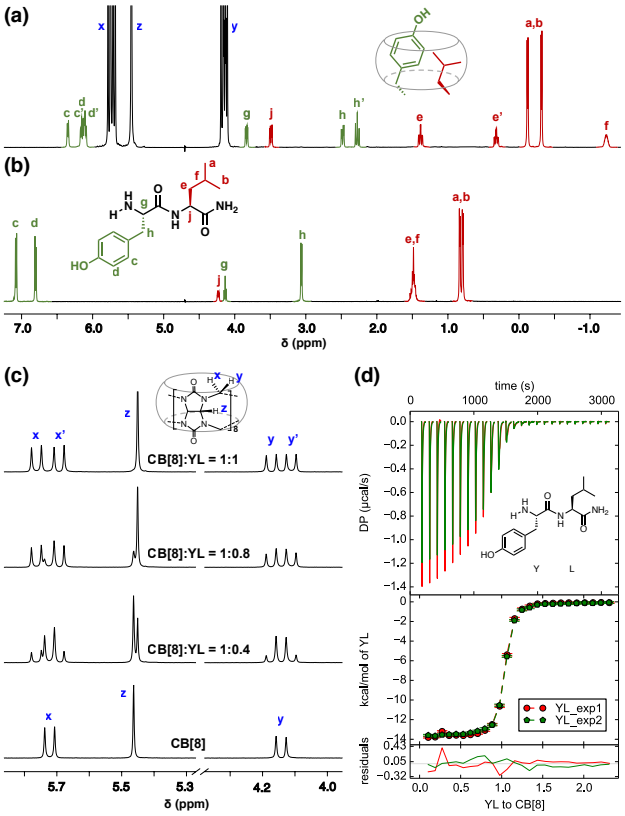

Fig. 3 Association of YL with CB[8] following Path B, demonstrated by ^{1}H-NMR spectra of a) the YL@CB[8] complex with 1:1 stoichiometry, b) YL in D_2O solution at 298 K, where the concentration of CB[8] is 0.1 mM, and c) the signal splitting of CB[8] protons during the titration of YL, as well as d) the isothermal titration calorimetry of YL into CB[8] in 10 mM sodium phosphate buffer (pH 7.0) at 298 K. Two individual titrations were performed for a subsequent global fitting. N.B. the intensity of H^j in b) is smaller than it supposes to be due to solvent suppression in NMR.

2.1 Association of YL and LY oligopeptide derivatives with CB[8]

Several YL and LY oligopeptide derivatives including YL, YLA, YLAA, AAYLAA, LY, LYA, ALY were synthesised and their complexation with CB[8] studied using a combination of NMR titrations and isothermal titration calorimetry (ITC) techniques. NMR titration studies of YL and LY oligopeptide derivatives against CB[8] all exhibited similar proton shifts and splitting patterns corresponding to binding typical for Path B, where both Y and L sidechains resided inside the CB[8] cavity in close proximity to one another. As exemplified by YL@CB[8] in Fig. 3a, the proton signals from the side-chains of Y ($H^{c,d,h}$) and L ($H^{a,b,e,f}$) were significantly shifted upfield compared with the signals of unbound YL (Fig. 3b), demonstrating the encapsulation of both side-chains inside the CB[8] cavity. The inclusion of a proton inside a CB[8] cavity contributes to a shift of approximately 1 ppm[22], thus the upfield shift of H^f by more than 2.5 ppm (Fig. 3a,b) must also be a product of the aromatic ring current from the tyrosine sidechain and confirms the close proximity of the phenol ring of Y and the isobutyl group of L inside the CB[8] cavity (Fig. 3a). This results in an asymmetric chemical environment within the cavity where previously equivalent protons are no longer chemically equivalent, resulting in two set of signals for each proton in $H^{c,d,h,e}$ (Fig. 3b). Meanwhile, the observation of inequivalent protons as two sets of signals also suggests the formation of a 'static' YL@CB[8] complex with a slow flipping rate of the enclosed phenol and isobutyl groups on the NMR timescale. This asymmetric environment coupled with a slow flipping rate can also be demonstrated by the signal splitting of CB[8] protons ($H^{x,y,z}$), as shown in Fig. 3c.

ITC studies (Fig. 3d) also demonstrated a 1:1 stoichiometric binding of YL with CB[8], having an affinity of 8.2×10^6 M^{-1} and an enthalpy change of -13.6±0.1 kcal/mol (Table 1), which corresponds to the association of CB[8] in a ternary manner with a typical enthalpy change between -12 and -18 kcal/mol(Fig. 1).

Table 1 Thermodynamic data for the associated CB[8] with peptide derivatives of YL, LY, and FL. All the data is obtained from ITC at 298 K in 10 mM sodium phosphate buffer (pH 7.0), whose individual titration plots can be found in the ESI.

Peptides	K_a (M^{-1} or M^{-2})	ΔG ($kcal/mol$)	ΔH ($kcal/mol$)	$T\Delta S$ ($kcal/mol$)
YL	8.2×10^6	-9.4±0.1	-13.6±0.1	-4.2±0.2
YLA	8.1×10^6	-9.4±0.1	-12.0±0.1	-2.6±0.2
YLAA	7.1×10^6	-9.3±0.1	-11.0±0.1	-1.7±0.2
AAYLAA	1.8×10^5	-7.2±0.1	-11.6±0.1	-4.4±0.2
LY	1.3×10^7	-9.7±0.2	-13.7±0.2	-4.0±0.4
LYA	1.3×10^7	-9.7±0.2	-12.0±0.2	-2.3±0.3
ALY	1.3×10^6	-8.4±0.1	-11.7±0.1	-3.4±0.1
FLA	1.0×10^7	-9.6±0.1	-12.0±0.1	-2.4±0.1
AFLA	2.1×10^6	-8.6±0.1	-11.4±0.1	-2.8±0.1
LF	6.6×10^6	-9.3±0.2	-12.3±0.2	-3.0±0.4
FL (Path B)	1.3×10^7	-9.7±0.8	-11.4±0.2	-1.7±1.0
FL (Path A)	1.9×10^{11}	-15.4±2.4	-26.7±1.4	-11.3±3.8

2.2 Effects of the chain extension from C- or N-terminus and sequential order of residues on Path B

As YL@CB[8] exhibits exactly the same binding behaviour as Urbach *et al.* previously reported for their tripeptide systems such as YLA@CB[8],[16] we can confirm here that the shortest oligopeptide chains capable of forming 1:1 binding along Path B are dipeptide sequences such as YL and LY. The thermodynamic data of various oligopeptide derivatives of YL and LY in Table 1 suggest that their affinities with CB[8] are not significantly altered by extending the peptide chain at the C-terminus. YL, YLA, and YLAA exhibited a uniform binding constant (K_a) and ΔG regardless of the additional Ala extended from C-terminus of leucine. The extra Ala units were found to affect association through a reduction in binding enthalpy (ΔH), however, this was compenstated by a change in entropy likely due to solvent reorgnization.[23]

A reduction in binding affinity can be achieved by shifting the effective binding pair away from the N-terminus. For instance,

AAYLAA contains an additional two Ala units extended from the N-terminus of tyrosine exhibiting a binding affinity with CB[8] of $1.8 \times 10^5\,M^{-1}$, which is weaker than that of YL, YLA, and YLAA by more than one order of magnitude (Table 1). The same was observed for LY, LYA and ALY, where the binding affinity of LY@CB[8] is the same as that for LYA@CB[8], but is one order of magnitude stronger than that of ALY@CB[8] (Table 1). Therefore, the 1:1 binding along Path B may be suppressed by chain extension from the N-terminus but is not significantly influenced by chain extension at the C-terminus.

Comparing the changes in the free energies of association for YL@CB[8] and YLA@CB[8] (-9.4±0.1 kcal/mol) with LY@CB[8] and LYA@CB[8] (-9.7±0.2 kcal/mol) (Table 1), the similar values dectected suggest that the sequential order of the effective binding pair does not significantly influence binding affinity.

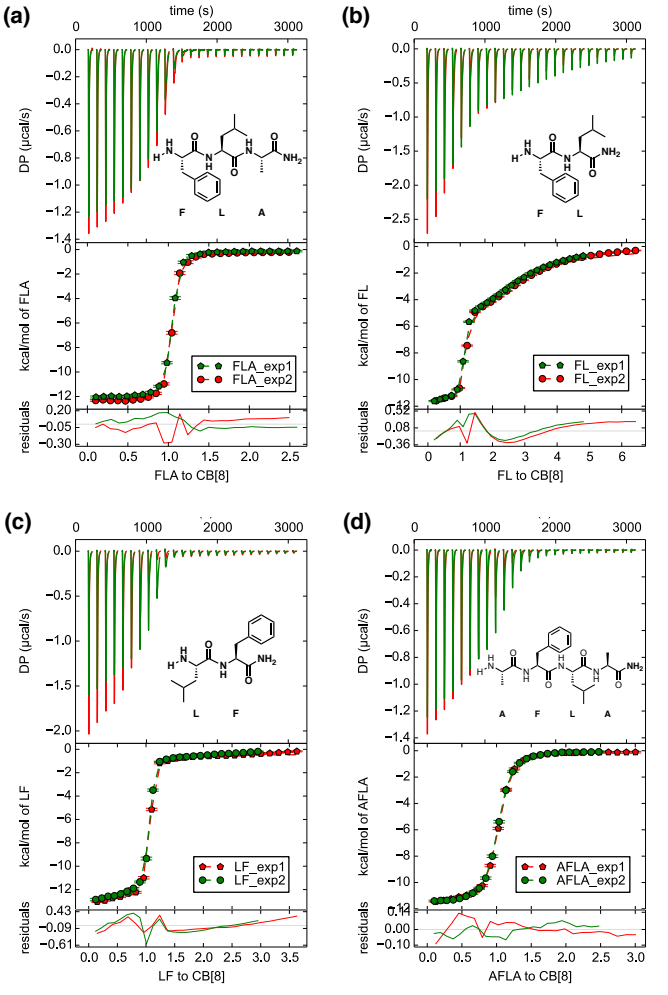

Fig. 4 Isothermal titration calorimetry of a) FLA, b) FL, c) LF, and d) AFLA into CB[8] in 10 mM sodium phosphate buffer (pH 7.0) at 298 K. For each species, at least two individual titrations were performed for the subsequent global fitting. In order to turn the binding from Path B to Path A when excess peptide is present, a dipeptide is required rather than a tri- or tetra- peptide, with phenylalanine at the N-terminus and leucine at the C-terminus, *i.e.* FL.

2.3 Effects of C- or N-terminus chain extension on Path A

Effective 2:1 homoternary binding following Path A requires the presence of F (phenylalanine) and is not observed with other aromatic residues such as Y and W (tryptophan).[6] In a 2:1 binding mode, only one residue per oligopeptide is engaged with CB[8] so there is no obvious influence from the sequential order of the peptide chain. However, the binding of a 2:1 homoternary complex with CB[8] is highly susceptible to the position of F in the peptide chain. With F positioned at the N-terminus, the homoternary complex of FGG with CB[8] exhibited a strong affinity; whilst no effective binding was observed for the GFG and GGF sequences, which have chain extension from N-terminus of F.[6]

Although the additional two glycine residues extended from the C-terminus of F were found to facilitate the formation of a homoternary complex with CB[8],[6] Cavatorta *et al.* demonstrated that further chain extensions from C-terminus reduce the binding affinity.[20] Thus, 2:1 binding along Path A can be suppressed by chain extension from either the C- or N-terminus. In other words, Path A may be maximised by choosing the shortest sequence that contains F at the N-terminus.

2.4 To pursue a sequence with desired binding pathway

In our effort to identify an oligopeptide that can exhibit both Path A and Path B at different ratios of peptide to CB[8], we decided to investigate a peptide similar to YLA (so that it can form a 1:1 complex) but with a F residue in the N-terminus in order to potentially display a 2:1 complex as well. Thus we settled on the tripeptide FLA, which only exhibits 1:1 stoichometric complexation,[16] whereby no 2:1 complex is detected even as the molar ratio of FLA to CB[8] approaches a value of 2.5 (Fig. 4a). As FLA does not lead to a 2:1 complexation product, it suggests that the FLA tripeptide prefers to assemble along Path B solely. However, based on the relationship between oligopeptide sequence and binding behaviour described above, one may be able to customise the overall assembly pathway by manipulating the sequence further.

We first attempt to invert the sequence order of F and L, investigating thermodynamic data of the dipeptide LF. The ITC of LF with CB[8] is shown in Fig. 4c, which exhibits a binding profile corresponding to an absolute 1:1 stoichiometric complex. Second, we explore what happens when FLA is extended at the N-terminus, leading to AFLA, which remains a 1:1 complex as shown in Fig. 4d. Learning from YLA and YLAA, one can also infer that chain extension of FLA from the C-terminus (FLAA) would not lead to a significant change of binding. Finally, shortening the peptide backbone of FLA from the C-terminus leads to the dipeptide FL, which may result in a peptide with the desired properties.

2.5 Coexistence of two binding pathways of FL with CB[8]

The ITC of the FL dipeptide (Fig. 4b) exhibits an obvious two-stage binding profile that differs from the typical 1:1 binding profile exhibited by FLA (Fig. 4a). A turning point is evident in the ITC at the FL:CB[8] ratio around 1:1, which likely corresponds to a 1:1 binding behaviour represented in Path B. A second stage

in the binding curve appears above the 1:1 ratio resembling a weaker binding process, which may correspond to competitive binding.

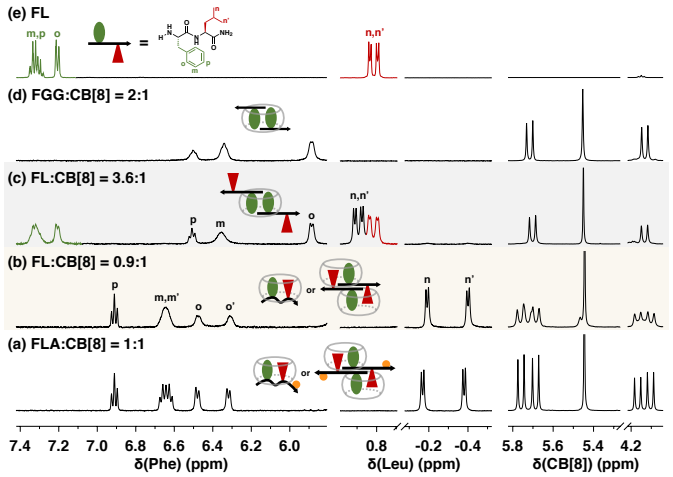

Fig. 5 ^{1}H-NMR spectra of a) FLA:CB[8]=1:1, b) FL:CB[8]=0.9:1, c) FL:CB[8]=3.6:1, d) FGG:CB[8]=2:1, e) FL in D_2O solution at 298 K, where concentrations of CB[8] are all 0.1 mM and demonstrates the evolution of binding from FLA-type pathway to FGG-type pathway. A complete NMR titration spectra can be found in Fig. S36

Analysis of NMR titration spectra at different FL:CB[8] ratios (Fig.S36) resulted in a few important observations. First, in the presence of excess CB[8], *e.g.* FL:CB[8]=0.9:1 (Fig. 5b), the NMR spectrum replicates that of FLA@CB[8] (Fig. 5a) with a characteristic 1:1 stoichometric binding mode. Second, both the CB[8] and F residue protons exhibit signal splitting. Third, the L residue protons exhibit significant upfield shifts. All of these observations indicate the formation of a 1:1 YLA@CB[8]-type complex that corresponds to Path B.

Upon further increasing the FL:CB[8] ratio, a completely different NMR spectrum is observed (Fig. 5 & Fig.S36), where the peaks corresponding to the protons of the F and L side-chains broaden and merge together. At a ratio of FL:CB[8]=3.6:1 (Fig. 5c), the signals from the isobutyl group of L (*e.g* $H^{n,n'}$) shift downfield compared to that of both the 1:1 complex (Fig. 5b) and uncomplexed FL peptide (Fig. 5e). This indicates that the side-chain of L no longer resides in the CB[8] cavity and is now situated outside in close proximity to the portal.

Instead, a homoternary 2:1 complex is formed where two phenyl groups are encapsulated in one CB[8] cavity. This complex displays a greater upfield shift of the aromatic proton signals (Fig. 5c) compared to the spectrum at a ratio of FL:CB[8]=0.9:1 (Fig. 5b), attributed to the aromatic ring-current effect. The characteristic splitting of the CB[8] peaks also disappear at this higher ratio, resembling that of a symmetric and equivalent environment around the two portals. Comparing these observations to that of a typical FGG_2@CB[8] NMR spectrum (Fig. 5d), the signals corresponding to the binding mode in the second stage are almost the same except for the presence of signals from excess, uncomplexed FL peptides.

Therefore, using a combination of ITC and NMR titration techniques, a coexistence of two binding pathways can be confirmed for the association of the FL dipeptide with CB[8]. In addition, two distinct complexation products (1:1 and 2:1 complexation) can be obtained by carefully controlling the FL:CB[8] ratio.

2.6 Distinguishing between 1:1 and 2:2 binding modes by DOSY

The FL peptide exhibits the coexistence of two binding pathways and is capable of switching from a 1:1 stoichiometric complex (either 1:1 or 2:2) to a 2:1 complex as a function of the peptide to CB[8] ratio. We therefore used this system to deduce whether the characteristic 1:1 stoichometric binding mode is in fact a 1:1 binary or a 2:2 quaternary complex. Through isolating specific peptide:CB[8] ratios corresponding to the different binding modes and analysing them using DOSY NMR, size-based diffusion coefficients (D) could be mapped.

First, a preliminary correlation between D and the binding modes was obtained from the study of three well-understood CB[8]-mediated complexes. The complexes selected include the 1:1 binary complex formed by N,N'-dimethyl-4,4-diazapyrenium iodide (dzpy) guest, the FGG_2@CB[8] complex (2:1) and the known 2:2 complex formed with the 1,1'-bis(4-(dimethylamino)phenyl)-[4,4'-bipyridine]-1,1'-diium ($VNMe_2$) guest.

Table 2 Diffusion coefficients (D) of different CB[8] complexes and their respective guests. All the data is obtained from DOSY measurements in D_2O, where the individual DOSY spectra can be found in the ESI.

Guest	Guest:CB[8]	D (e^{-10} m^2s^{-1})
-	0:1	3.11 (±0.01)
dzpy	1:1	3.04 (±0.01)
YL	1:1	2.95 (±0.01)
YLA	1:1	2.88 (±0.01)
FLA	1:1	2.87 (±0.01)
FL (Path B, D1)	1:1	2.86 (±0.01)
FL (Path A, D2)	2:1	2.68 (±0.01)
FGG	2:1	2.66 (±0.01)
$VNMe_2$	2:2	2.01 (±0.01)
YLAGGALY	1:2	2.08 (±0.02)
YLA-ahx6-ALY	1:2	2.07 (±0.01)

It is evident from the DOSY results (Table 2) that the D values correlate to the number of CB[8] macrocycles present in the complex. The 2:2 complex of $VNMe_2$ with CB[8] exhibits a diffusion coefficient (2.01 $e^{-10}m^2s^{-1}$) significantly slower than the D of free CB[8] (3.11 $e^{-10}m^2s^{-1}$), dzpy@CB[8] complex (3.04 $e^{-10}m^2s^{-1}$), and the FGG_2@CB[8] complex (2.66 $e^{-10}m^2s^{-1}$). In other words, if the 1:1 stoichiometric complex of FL with CB[8] is a 2:2 quaternary complex, it should exhibit a much smaller D value than the confirmed 2:1 ternary complex on account of the number of CB[8] hosts present in the complex.

The D values are also affected by the size of moieties situated outside the CB[8] cavity. For instance, the additional G residues in FGG_2@CB[8] slightly slows down the diffusion of the complex; whilst the dzpy@CB[8] complex with little situated around the CB[8] portal exhibits a D value very similar to free CB[8]. Similarly, one can deduce that the 1:1 binary complex of FL and CB[8]

should exhibit a larger D value than the 2:1 ternary complex. The simultaneous inclusion of F and L in the binary complex will result in a smaller amount of the dipeptide outside of the CB[8] cavity compared to the 2:1 FGG-type complex where the L residues are entirely situated outside the CB[8] cavity.

Therefore, the potential binary, ternary, and quaternary binding modes of FL with CB[8] should have ascending molecular sizes and descending diffusion coefficients (Fig. 2d). By comparing D of the 1:1 stoichiometric complex (FL:CB[8]=0.9:1) to the 2:1 homoternary complex (FL:CB[8]=3.6:1), one should be able to distinguish between 1:1 and 2:2 binding modes and provide insight on the unknown nature of the 1:1 stoichiometric complex. As shown in Table 2, the 1:1 stoichometric species (FL:CB[8]=0.9:1) and 2:1 homoternary species (FL:CB[8]=3.6:1) exhibit D values of 2.86 $e^{-10}m^2s^{-1}$ (D1) and 2.68 $e^{-10}m^2s^{-1}$ (D2), respectively. D2 has a similar D value to that of FGG_2@CB[8], consistent with the formation of a 2:1 homoternary complex (FL:CB[8]=3.6:1), where the L residues reside outside the CB[8] cavity. The fact that the '1:1' complex exhibits a larger D value than that of 2:1 ternary complex indicates that the 1:1 stoichiometric complex is of a smaller size than that of the 2:1 species, which verifies that the binding mode is in fact a 1:1 pairwise inclusion complex. Moreover, D1 is very close in value to that of both 1:1 complexes of YL@CB[8], YLA@CB[8], and FLA@CB[8] (Table 2), providing further verification that D1 is not a case specific towards FL, but is a characteristic D for 1:1 complexes where two side-chains of the peptide are enclosed in the same CB[8] cavity.

2.7 Peptide sequences with two effective binding pairs

To further corroborate this finding, two longer peptide sequences were designed and synthesised to contain two CB[8] binding sites at their termini. The first employs glycine-glycine as a spacer between YLA and ALY segments (YLAGGALY) and the second uses a short aliphatic chain as a spacer (YLA-ahx6-ALY) (Fig. S38-S45). If the 1:1 stoichiometric complex corresponds to a 2:2 binding mode, the resultant complexes of YLAGGALY and YLA-ahx6-ALY with CB[8] would contain four CB[8] molecules in the complex. These larger complexes should exhibit a significant reduction in their D value and differ from that of complexes containing one or two CB[8] macrocycles. However, the association of CB[8] with either YLAGGALY or YLA-ahx6-ALY have D values around $2.00\,e^{-10}m^2s^{-1}$, which is very similar to that of the 2:2 binding mode of $VNMe_2$ with CB[8] and represents a complex containing two CB[8] moieties. This observation further verifies the formation of a pairwise inclusion complex for the 1:1 stoichiometric complexation.

3 Conclusions

Through a systematic study on the effect of sequence towards the binding mode of oligopeptide-CB[8] complexation, we found the dipeptide FL has the capability to initially form a 1:1 stoichiometric complex, characteristic of YLA@CB[8], which upon addition of excess peptide, forms a 2:1 homoternary complex similar to that of FGG_2@CB[8]. This unique two-stage binding process allowed us to compare size-based diffusion properties of these two binding modes *using the same dipeptide with only one change in the system: the dipeptide-CB[8] ratio.* A singular change ensures that the small differences in diffusion coefficients observed by DOSY NMR are only correlated to the difference between individual binding modes. From the D values obtained we verified that the 1:1 stoichometric binding mode characteristic of YLA@CB[8] is in fact a 1:1 binary complex with am intramolecular pair inclusion rather than a 2:2 quaternary complex. These findings emphasise the importance of peptide/biomolecule design when considering a desired binding mode in applications that utilise host-guest binding motifs. In simple terms, using these principles we are able to customise the self-assembly pathway of peptides with CB[8] through simple changes in oligopeptide design. This will act as a useful toolbox in the design of peptide@CB[8] supramolecular complexes for future biomedical applications.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank the Leverhulme Trust (project: 'Natural material innovation for sustainable living'), the EPSRC ('NOtCH', EP/L027151/1), the ERC ('ASPiRe', 240629) and Marie Curie (FP7 'SASSYPOL' ITN, 607602) for funding. We also thank Dr. Aniello Palma for suggestions and Magdalena Olesińska for offering dzpy as a reference compound.

Notes and references

1 A. Day, A. P. Arnold, R. J. Blanch and B. Snushall, *J. Org. Chem.*, 2001, **66**, 8094–8100.
2 J. Lagona, P. Mukhopadhyay, S. Chakrabarti and L. Isaacs, *Angew. Chem., Int. Ed.*, 2005, **44**, 4844–4870.
3 J. W. Lee, S. Samal, N. Selvapalam, H.-J. Kim and K. Kim, *Acc. Chem. Res.*, 2003, **36**, 621–630.
4 S. J. Barrow, S. Kasera, M. J. Rowland, J. del Barrio and O. A. Scherman, *Chem. Rev.*, 2015, **115**, 12320–12406.
5 M. E. Bush, N. D. Bouley and A. R. Urbach, *J. Am. Chem. Soc.*, 2005, **127**, 14511–14517.
6 L. M. Heitmann, A. B. Taylor, P. J. Hart and A. R. Urbach, *J. Am. Chem. Soc.*, 2006, **128**, 12574–12581.
7 F. Biedermann and W. M. Nau, *Angew. Chem., Int. Ed.*, 2014, **53**, 5694–5699.
8 K. I. Assaf and W. M. Nau, *Chem. Soc. Rev.*, 2015, **44**, 394–418.
9 A. Palma, M. Artelsmair, G. Wu, X. Lu, S. J. Barrow, N. Uddin, E. Rosta, E. Masson and O. A. Scherman, *Angewandte Chemie*, 2017, **129**, 15894–15898.
10 Y. Kang, X. Tang, H. Yu, Z. Cai, Z. Huang, D. Wang, J.-F. Xu and X. Zhang, *Chemical science*, 2017, **8**, 8357–8361.
11 E. Masson, X. Ling, R. Joseph, L. Kyeremeh-Mensah and X. Lu, *RSC Adv.*, 2012, **2**, 1213–1247.
12 S. Gürbüz, M. Idris and D. Tuncel, *Org. Biomol. Chem.*, 2015, **13**, 330–347.
13 F. Biedermann, V. D. Uzunova, O. A. Scherman, W. M. Nau

and A. De Simone, *J. Am. Chem. Soc.*, 2012, **134**, 15318–15323.
14 F. Biedermann, M. Vendruscolo, O. A. Scherman, A. De Simone and W. M. Nau, *J. Am. Chem. Soc.*, 2013, **135**, 14879–14888.
15 G. Wu, M. Olesińska, Y. Wu, D. Matak-Vinkovic and O. A. Scherman, *J. Am. Chem. Soc.*, 2017, **139**, 3202–3208.
16 L. C. Smith, D. G. Leach, B. E. Blaylock, O. A. Ali and A. R. Urbach, *J. Am. Chem. Soc.*, 2015, **137**, 3663–3669.
17 Z. Hirani, H. F. Taylor, E. F. Babcock, A. T. Bockus, C. D. Varnado, C. W. Bielawski and A. R. Urbach, *J. Am. Chem. Soc.*, 2018, **140**, 12263–2269.
18 J. del Barrio, S. T. Ryan, P. G. Jambrina, E. Rosta and O. A. Scherman, *J. Am. Chem. Soc.*, 2016, **138**, 5745–5748.
19 F. Biedermann and O. A. Scherman, *J. Phys. Chem. B*, 2012, **116**, 2842–2849.
20 E. Cavatorta, P. Jonkheijm and J. Huskens, *Chem.–Eur. J.*, 2017, **23**, 4046–4050.
21 J. W. Lee, H. H. L. Lee, Y. H. Ko, K. Kim and H. I. Kim, *J. Phys. Chem. B*, 2015, **119**, 4628–4636.
22 W. L. Mock and N. Y. Shih, *J. Org. Chem.*, 1986, **51**, 4440–4446.
23 E. Grunwald and C. Steel, *J. Am. Chem. Soc.*, 1995, **117**, 5687–5692.