Showcasing research from Professor Adam R. Urbach's laboratory, Department of Chemistry, Trinity University, San Antonio, Texas, USA.

Peptide recognition by a synthetic receptor at subnanomolar concentrations

Inspired by the recent imaging of celestial objects by the James Webb Space Telescope, the binding pair composed of cucurbit[8]uril and *N*-terminal dipeptide are depicted as a supermassive black hole and star-filled nebulae, respectively. Just like the strong pull of a black hole on matter in the vacuum of space, cucurbit[8]uril has been shown to bind *N*-terminal dipeptide sites in water with exceptionally strong affinity, and thus at low concentrations.

As featured in:

See Adam R. Urbach *et al.*,
Chem. Sci., 2024, **15**, 5133.

ROYAL SOCIETY
OF **CHEMISTRY**

rsc.li/chemical-science

EDGE ARTICLE

Check for updates

All publication charges for this article have been paid for by the Royal Society of Chemistry

Peptide recognition by a synthetic receptor at subnanomolar concentrations†

Paolo Suating, [ID] [a] Marc B. Ewe, [ID] [a] Lauren B. Kimberly,[a] Hadi D. Arman,[b] Daniel J. Wherritt [ID] [b] and Adam R. Urbach [ID] *[a]

This paper describes the discovery and characterization of a dipeptide sequence, Lys−Phe, that binds to the synthetic receptor cucurbit[8]uril (Q8) in neutral aqueous solution with subnanomolar affinity when located at the N-terminus. The thermodynamic and structural basis for the binding of Q8 to a series of four pentapeptides was characterized by isothermal titration calorimetry, NMR spectroscopy, and X-ray crystallography. Submicromolar binding affinity was observed for the peptides Phe-Lys-Gly-Gly-Tyr (FKGGY, 0.3 µM) and Tyr-Leu-Gly-Gly-Gly (YLGGG, 0.2 µM), whereas the corresponding sequence isomers Lys-Phe-Gly-Gly-Tyr (KFGGY, 0.3 nM) and Leu-Tyr-Gly-Gly-Gly (LYGGG, 1.2 nM) bound to Q8 with 1000-fold and 170-fold increases in affinity, respectively. To our knowledge, these are the highest affinities reported between a synthetic receptor and an unmodified peptide. The high-resolution crystal structures of the Q8·Tyr-Leu-Gly-Gly-Gly and Q8·Leu-Tyr-Gly-Gly-Gly complexes have enabled a detailed analysis of the structural determinants for molecular recognition. The high affinity, sequence-selectivity, minimal size of the target binding site, reversibility in the presence of a competitive guest, compatibility with aqueous media, and low toxicity of Q8 should aid in the development of applications involving low concentrations of target polypeptides.

Introduction

Biomedical science relies on methods for adding ligand binding sites to proteins. Such affinity tags are essential for protein affinity purification, detecting and quantifying proteins of interest, improving protein solubility, and adding functionality to proteins such as fluorescent labels or sites for chemical modification.[1–4] Protein affinity tags are often incorporated *via* recombinant DNA methodology and may comprise large fusion domains such as glutathione-*S*-transferase and maltose-binding protein, or small (5–20 residue) oligopeptide epitope tags such as His, Myc, and FLAG tags.[4–7] Although affinity tags assist greatly in purification, they often need to be removed afterward due to unwanted influence on the structure and/or function of the target protein. For example, His tags add several positive charges. Removal of the tag (*e.g.*, *via* endopeptidase cleavage) and subsequent purification leads to considerable loss of material. We posit that an affinity tag comprising only two residues should impose minimally on protein

structure/function and therefore reduce the likelihood that the tag would need to be removed.

In order to be useful in biological applications, oligopeptide affinity tags must be recognized with high affinity and sequence-selectivity by their cognate receptor. Creating minimal affinity tags is especially challenging due to the limited structural information and surface area inherent to a small target site. To date, the smallest oligopeptide affinity tags are 4–6 residues in length and comprise metal chelators (*e.g.*, oligo-His), polyionic tags for ion-exchange chromatography (*e.g.*, oligo-Glu and oligo-Arg), or tags recognized sequence-selectively by an antibody (*e.g.*, C-tag).[4] In general, sequence-selectivity has been challenging to achieve with synthetic ligands. The most promising ligands to date include molecularly imprinted polymers (MIPs) and synthetic receptors. MIPs are polymeric nanoparticles formed by templating a binding cavity into a crosslinked polymer using specific epitopes.[8] MIPs can be highly selective for their targets and can have sub-nanomolar detection limits for large proteins, but the affinities of MIPs for oligopeptide targets are limited to the high nanomolar range.[9] Moreover, the development of MIPs for biological applications is hindered by their heterogeneous compositions and limited molecular characterization. Synthetic receptors, by contrast, are pure organic compounds, typically macrocycles, that are highly stable and relatively inexpensive but, with the exception of cucurbit[n]urils, lack the affinities and selectivities necessary to be useful in many biological applications. The

[a]Department of Chemistry, Trinity University, 1 Trinity Place, San Antonio, TX 78212, USA. E-mail: aurbach@trinity.edu

[b]Department of Chemistry, University of Texas at San Antonio, 1 UTSA Circle, San Antonio, TX 78249, USA

† Electronic supplementary information (ESI) available: Experimental details, crystallographic data, ITC data, ^{1}H NMR data, and mass spectrometry data. CCDC 2313004 and 2314758. For ESI and crystallographic data in CIF or other electronic format see DOI: https://doi.org/10.1039/d4sc01122h

present work aims to make progress toward overcoming these challenges.[9]

Cucurbit[n]urils are a family of synthetic, barrel-shaped receptors comprising n repeating glycoluril units bridged by $2n$ methylene groups.[10,11] These rigid macrocyclic hosts have a weakly hydrated inner cavity that is accessible *via* two symmetric portals, each lined by n ureido carbonyl groups. This structure creates a ring of negative electrostatic potential at the constricted portals, which, in combination with the weakly hydrated cavity, is well-suited to guests containing both nonpolar and cationic groups. Prior work has shown that cucurbit[7]uril (Q7) and cucurbit[8]uril (Q8) can bind oligopeptides sequence-selectively and with sub-millimolar to mid-nanomolar binding affinities.[12–21] Q7 and Q8 recognize the aromatic residues Phe (F), Tyr (Y), and Trp (W) by inclusion of the aromatic side chain within the Qn cavity (Fig. 1a). Selectivity for the N-terminal position is mediated by electrostatic attraction between the N-terminal ammonium group and C=O groups on the Qn. While Q7 has a cavity volume sufficient to accommodate a single side chain, the larger cavity of Q8 can accommodate the side chains of up to two residues. Q8 can noncovalently dimerize two peptides by including a single aromatic residue from each peptide (Fig. 1b).[14] Q8 can also bind to two neighboring residues on a single peptide. In the latter case, known as the "pair-inclusion motif," the peptide backbone folds in order to accommodate the simultaneous inclusion of the pair of neighboring side chains (Fig. 1c).[19–22] At non-terminal sites, the backbone folds into a type II β-turn.[23]

Numerous applications have derived from the sequence-selective recognition of peptides by Q7 and Q8, including peptide sensing,[12,19,24–27] peptide and cell capture and release,[28–30] enhancing peptide detection,[31] inhibiting and measuring protease activity,[32,33] and supramolecular polymers and hydrogels.[34–38] Recognition properties are retained when oligopeptide epitopes for Q7 and Q8 are incorporated onto larger proteins,[17] enabling applications including protein dimerization,[39–41] oligomerization,[42] and polymerization,[43,44] modulation of protein aggregation,[45,46] protein capture on surfaces,[47,48] and protein drug formulation.[49,50] These applications were aided by the biocompatibility, commercial availability, and stability (thermal, chemical, and metabolic) of Q7 and Q8, the derivatization and conjugation strategies of Q7, the predictable and well-understood sequence-selectivity of these systems, and the reversibility of binding using small molecule competitors across a wide range of binding affinities.[11,51]

Applications of cucurbit[n]uril-oligopeptide interactions can be limited, however, by binding affinity. In order to target the many peptides and proteins that exist in biological systems at or below the nanomolar concentration range, it is necessary for the equilibrium dissociation constants to be at or below the nanomolar range. Although there have been a few reports of K_d values in the nanomolar range,[15,18–21,23] to the best of our knowledge there have been no reports to date of sub-nanomolar binding of unmodified peptides by synthetic receptors. We recently investigated the binding of Q8 to non-terminal dipeptide sites.[23] Although binding was expected to be weaker at non-terminal sites due to the lack of involvement of the N-terminal ammonium group, we discovered that Q8 binds to non-terminal KF and FK dipeptide sites in the pair-inclusion motif with K_d values of 60 and 86 nM, respectively. These affinities were higher than that of non-terminal YL and LY, which were the highest affinity sites observed in previous work. We hypothesized that moving the KF and FK sites to the N-terminus could yield higher binding affinity than previously observed.
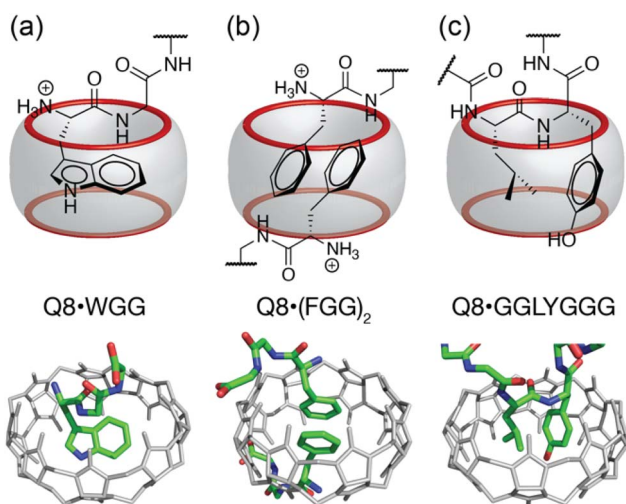
Results and discussion

Thermodynamic studies

Two pentapeptides, H-Lys-Phe-Gly-Gly-Tyr-OH (**KF**GGY) and H-Phe-Lys-Gly-Gly-Tyr-OH (**FK**GGY), were designed to contain the target dipeptide site (bold) at the N-terminus and were obtained commercially (see ESI† for experimental details). C-terminal Tyr was included to aid in quantitation by UV spectroscopy. The thermodynamics of complexation for Q8 binding with the two peptides were determined using isothermal titration calorimetry at 300 K in 10 mM sodium phosphate, pH 7.0 (Table 1 and Fig. S3–S5†). Direct titration of Q8 with KFGGY resulted in a steep binding curve (Fig. S3†), suggestive of an affinity that is too strong to be determined confidently by ITC.[52] Therefore, a competition binding experiment was performed by ITC using methyl viologen (MV, *N*,*N*′-dimethyl-4,4′-bipyridinium dichloride) as the weak competitor in 100-fold stoichiometric excess of Q8 (Fig. S4†), as described previously.[18,53] This experiment

Fig. 1 Three unique binding motifs of Q8·peptide complexes, with schematics on top and crystal structures on bottom. (a) The Q8·WGG complex shows recognition of N-terminal Trp on a single peptide, with the indole side chain bound within the Q8 cavity and chelation of the N-terminal ammonium group by three C=O oxygens; (b) the Q8·(FGG)₂ complex shows recognition of the N-terminal Phe on each of two peptides, with staggered face-to-face stacking of the benzyl side chains and binding of each N-terminal ammonium group at opposite Q8 portals. (c) the Q8·GGLYGGG complex shows recognition of the Leu–Tyr dipeptide site, with inclusion of both side chains and folding of the peptide backbone into a type II β-turn. Host molecules are shown in gray, and peptides are colored as follows: carbon in green, nitrogen in blue, and oxygen in red. Hydrogen atoms and water molecules were removed for clarity.

Table 1 Thermodynamic binding data for Q8·peptide complexes

Guest	K_d (nM)	ΔH (kcal mol^{-1})	$-T\Delta S$ (kcal mol^{-1})
KFGGY	0.33 ± 0.08[a]	−16.6 ± 0.4[a]	3.6 ± 0.3[a]
KFGGY	0.27 ± 0.05[b]	nd	nd
FKGGY	330 ± 40[c]	−13.9 ± 0.4[c]	5.0 ± 0.3[c]
LYGGG	1.2 ± 0.1[a]	−17.0 ± 0.1[a]	4.7 ± 0.1[a]
YLGGG	150 ± 20[c]	−14.9 ± 0.2[c]	5.5 ± 0.2[c]

[a] Average values (± standard deviations) for data collected in triplicate *via* ITC competition experiments using methyl viologen dichloride (MV) at a 1 : 100 ratio of Q8 : MV. K_d (MV) = 6.8 (±0.2) × 10^{-7} M; ΔH (MV) = −4.8 (±0.2) kcal mol^{-1}; $-T\Delta S$ (MV) = −3.7 ± 0.1 kcal mol^{-1}. [b] Average values (± standard deviations) for data collected in triplicate *via* NMR competition experiments with (ferrocenylmethyl)trimethylammonium. The affinity of FcNMe$_3^+$ was determined by ITC competition with MV. K_d (FcNMe$_3$) = (4.4 ± 0.7) × 10^{-10} M. [c] Average values (± standard deviations) for data collected in triplicate *via* ITC with direct titration of guest to host. nd = not determined.

yielded a K_d value of 0.33 (±0.08) nM. To our knowledge, this is the highest affinity observed between a synthetic receptor and an unmodified peptide. In order to increase confidence in this result, we also measured the K_d value for the Q8·KFGGY complex by the NMR competition method reported by Isaacs and co-workers (Fig. S16†).[51] Using (ferrocenylmethyl)trimethylammonium chloride (FcNMe$_3$) as the competitor (Fig. S1 and S2†), we determined a K_d value of 0.27 (±0.05) nM, which is very similar to the value determined by competitive ITC. Interestingly, the sequence isomer, FKGGY, bound Q8 with K_d value of 330 nM, which is three orders of magnitude weaker in affinity. When KF and FK dipeptide sites were located at nonterminal positions, their affinities for Q8 differed by less than 2-fold.[23]

In our first publication on the pair-inclusion motif,[19] we reported the K_d value for Q8 in complex with the tripeptide H-Tyr-Leu-Ala-NH$_2$ as 8 nM using competitive ITC. Scherman and coworkers also studied the YL dipeptide binding site and reported K_d values for Q8 binding to H-Tyr-Leu-Ala-NH$_2$ (120 nM), H-Tyr-Leu-NH$_2$ (120 nM), and H-Tyr-Leu-Ala-Ala-NH$_2$ (140 nM) using ITC by direct titration.[22,23] Puzzled by these reports, we revisited the H-Tyr-Leu-Ala-NH$_2$ and agree that the binding constant is closer to 100 nM than 10 nM. This is important because YL has been the benchmark target dipeptide site in four studies on the pair-inclusion motif,[19,21–23] and similarly we wanted to compare the binding affinities observed for KFGGY and FKGGY to the YL benchmark. Therefore, two additional pentapeptides, H-Tyr-Leu-Gly-Gly-Gly-OH (**YLGGG**) and H-Leu-Tyr-Gly-Gly-Gly-OH (**LYGGG**) were obtained commercially and characterized by ITC (Table 1, Fig. S6–S8†). We chose the buffer condition of 10 mM sodium phosphate, pH 7.0, to allow direct comparison to prior work on Q8·peptide molecular recognition. YLGGG was found to bind Q8 with a K_d value of 150 nM, as determined by direct titration, whereas its sequence isomer, LYGGG, bound ∼100-fold more tightly (1.2 nM), as determined by competitive titration using MV as competitor. In all cases, binding is enthalpically driven and entropically unfavourable. For each peptide with an aliphatic N-terminal residue and aromatic 2nd residue (*i.e.*, KFGGY and LYGGG), binding affinity

was stronger, more exothermic, and less entropically unfavourable than for its sequence isomer (*i.e.*, FKGGY and YLGGG, respectively). The three-fold difference in binding affinity between FKGGY and YLGGG shows a lack of selectivity between these two sites. ITC and mass spectrometry data corroborate the Q8 : peptide binding stoichiometry of 1 : 1 in all cases (Fig. S20–S27†).

Structural studies by ^{1}H NMR spectroscopy

Spectra were collected at 500 MHz at 21 °C on samples dissolved in 10 mM sodium phosphate-buffered D$_2$O, pH$_{apparent}$ 7.1.[54] Upon addition of a substoichiometric quantity of Q8 to KFGGY or FKGGY, we observed the appearance of a second set of peptide signals, with each set corresponding to either the free peptide or the peptide in the 1 : 1 complex (Fig. S9 and S17†). This result indicates slow chemical exchange on the NMR time scale and is consistent with other studies on the pair-inclusion motif. In the second set of peaks, we observed a perturbation to lower chemical shift values of the signals corresponding to the Lys and Phe side chains, indicating the shielding of the side chains from the external magnetic field due to binding within the Q8 cavity.[55–57] The signals corresponding to the Tyr side chain do not perturb upon the addition of Q8, showing that Tyr does not interact significantly with the Q8 cavity or portals.

^{1}H–^{1}H correlation spectroscopy (COSY) was used to assign the signals of the Lys sidechain (Fig. S10 and S13†) in the spectra for KFGGY. In the bound state, the geminal protons were resolved for all methylene groups in the chain except δ-CH$_2$. Peak separations for geminal pairs were observed to be as small as 0.1 ppm in the case of β-CH$_2$ and as large as 0.8 ppm in the case of γ-CH$_2$. The resolution of signals for the ε-CH$_2$ geminal protons is likely aided by their proximity to the Phe aromatic side chain.

To further investigate the solution structure of the Q8·KFGGY complex, 2-D ^{1}H–^{1}H Nuclear Overhauser Effect Spectroscopy (NOESY) data were collected and analysed for the free peptide, whereas the Q8·peptide complex was analysed by 2-D ^{1}H–^{1}H rotating-frame NOESY (*i.e.*, ROESY) (Fig. 2). In the unbound state, we observed cross-peaks between the signals corresponding to the side chains of Lys, Phe, and Tyr, for example Phe-*para* with Lys-γ, and Tyr-*meta* with Lys-γ. These NOEs suggest that on the NMR time scale the solution-state free peptide exists, on average, in a somewhat stably folded state. In the presence of Q8, however, we did not observe those inter-side-chain NOEs. Instead, in the presence of Q8 we observed cross-peaks correlating all Phe *ortho*-, *meta*-, and *para*- protons with ε-Lys, δ-Lys, and β-Lys protons. This result, in combination with the upfield chemical shift perturbation of the Phe and Lys side chains, is consistent with the proximity of Lys and Phe side chains within the Q8 cavity and the position of the C-terminal Tyr distal to the binding site. It is also clear that the Q8-induced fold of the peptide is different than the fold adopted in the absence of Q8. This result is also consistent with a cation–π interaction between the Lys ε-NH$_3^+$ group and the Phe aromatic ring, a motif that is commonly observed in protein crystal structures.[58–61] The magnetic field anisotropy of the
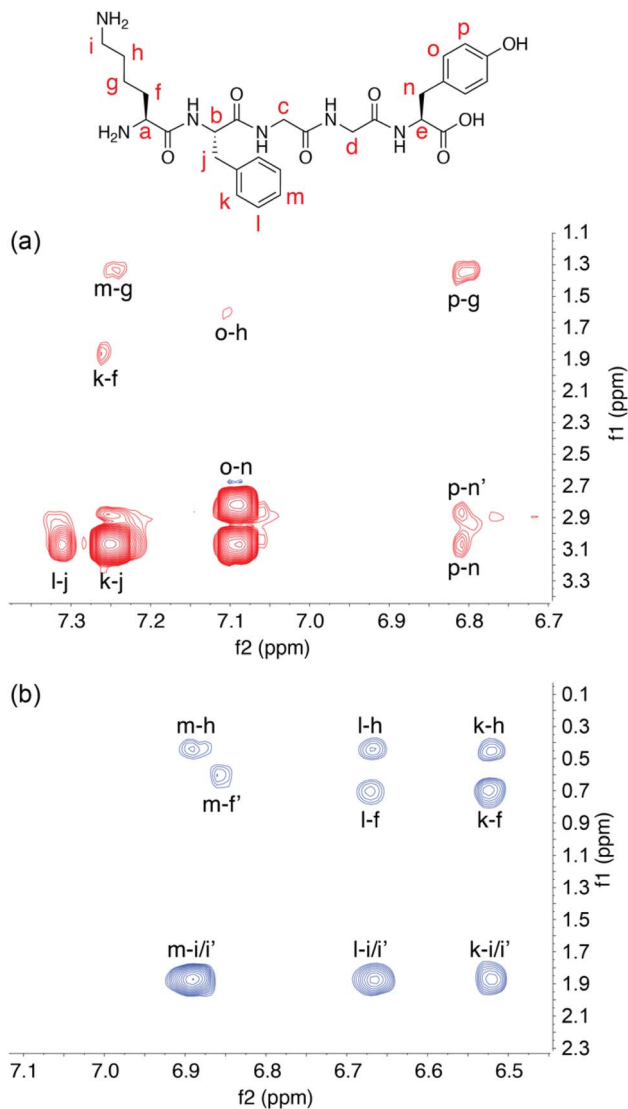
Fig. 2 Two-dimensional ^{1}H NMR spectroscopy. (a) Aromatic−aliphatic region of a 2-D NOESY spectrum of free KFGGY, showing cross peaks between the side chains of the Lys, Phe, and Tyr residues. (b) The same region of a 2-D ROESY spectrum of KFGGY in complex with Q8, showing only cross peaks between the side chains of Lys and Phe residues. Letters are used to correlate hydrogens on the peptide structural formula with corresponding signals.

aromatic ring imparts additional shielding on the Lys methylene groups beyond what is induced by binding within the host cavity.[62] This is evident not only in the differentiation of the diastereotopic protons in Lys, but also in the large perturbation of the Lys methylene signals to lower chemical shift values upon Q8 binding.

Structural studies by X-ray crystallography

Further structural investigation was carried out by X-ray crystallography. The poor solubility of the complexes of Q8 with LYGGG and YLGGG impeded analysis by ^{1}H NMR spectroscopy but facilitated the growth of single co-crystals for analysis by X-ray crystallography. Plate-like co-crystals of the Q8·LYGGG and

Q8·YLGGG complexes were grown from approximately equimolar mixtures of host and guest at 3–4 mM concentration in 10 mM sodium phosphate-buffered D_2O, $pH_{apparent}$ 7.1. Crystal structures were obtained at resolutions of 0.82 Å for Q8·LYGGG and 0.79 Å for Q8·YLGGG (Tables S1 and S2,† CCDC 2314758 and 2313004). The Q8·YLGGG complex crystallized in the $P2_1$ space group, with one Q8·YLGGG complex and eight associated water molecules in the asymmetric unit and two asymmetric units per unit cell. The Q8·LYGGG complex crystallized in the C_2 space group, with one Q8·LYGGG complex and 44 associated water molecules in the asymmetric unit and one asymmetric unit per unit cell. All of the observed water molecules for the Q8·LYGGG complex are located outside of the Q8 cavity.

In both structures, Q8 is deformed from the ideal D_{8h} symmetry observed in its un-complexed state[63] to a distorted ellipse, with greater deformation and a small degree of twisting observed in the Q8·YLGGG complex (Fig. 3). In both structures the side chains of Tyr and Leu residues are included within the Q8 cavity, and there is a CH–π interaction between the Leu γ-CH and the centre of the electron-rich Tyr phenol ring,[64,65] as expected in this motif.[19,23] The Leu γ-CH is more closely aligned with the centre of the Tyr phenol ring in the Q8·LYGGG structure.

The peptide bond between the Tyr and Leu residues is positioned such that the Tyr–Leu peptide bond of YLGGG is coplanar with the ring of carbonyl oxygens of the host portal (Fig. 4), which is similar to that observed in the crystal structure of Q8 in complex with GGLYGGG.[23] In the Q8·LYGGG structure, however, the plane of the Leu–Tyr peptide bond is positioned at a 30° dihedral angle with respect to the ring of the Q8 carbonyl oxygens, with the amide NH group canted inward toward the Q8 cavity. In both structures, the peptide folds in the same direction, in which the α–CH bond of the N-terminal residue is anti-coplanar with the vicinal C=O bond. This fold is similar to our semiempirical structure of the Q8·YLA complex[19] and similar to a model subsequently predicted by Scherman and co-workers for a Q8·YL complex,[22] but contrary to their prediction for a Q8·LY complex.[22]

The Tyr and Leu side chains in the Q8·YLGGG complex are bound more deeply within the Q8 cavity than they are in the Q8·LYGGG complex (Fig. 4). The positions of the Leu and Tyr α-carbons in the Q8·YLGGG structure are approximately 0.5 Å inside of the plane of Q8 carbonyl oxygens, whereas the Tyr α carbons in the Q8·LYGGG are aligned with the Q8 carbonyl oxygens. The side chains of the Tyr and Leu residues in the Q8·YLGGG structure are therefore pushed more deeply into the cavity than in the Q8·LYGGG structure. The geometric constraint forces the Tyr side chain of YLGGG at an angle (compared to the Q8·LYGGG structure) and positions the Tyr phenolic OH group outside of the plane of Q8 carbonyl oxygens and in the centre of the portal, such that a water molecule is needed to bridge them (Fig. 5). In the Q8·LYGGG structure, however, the Tyr phenolic OH group forms a direct hydrogen with a proximal C=O group on Q8.

In both structures there are several hydrogen bonds between the peptide backbone and the proximal carbonyl groups on Q8 (Fig. 6).‡[66,67] In the Q8·LYGGG structure, the Tyr NH forms
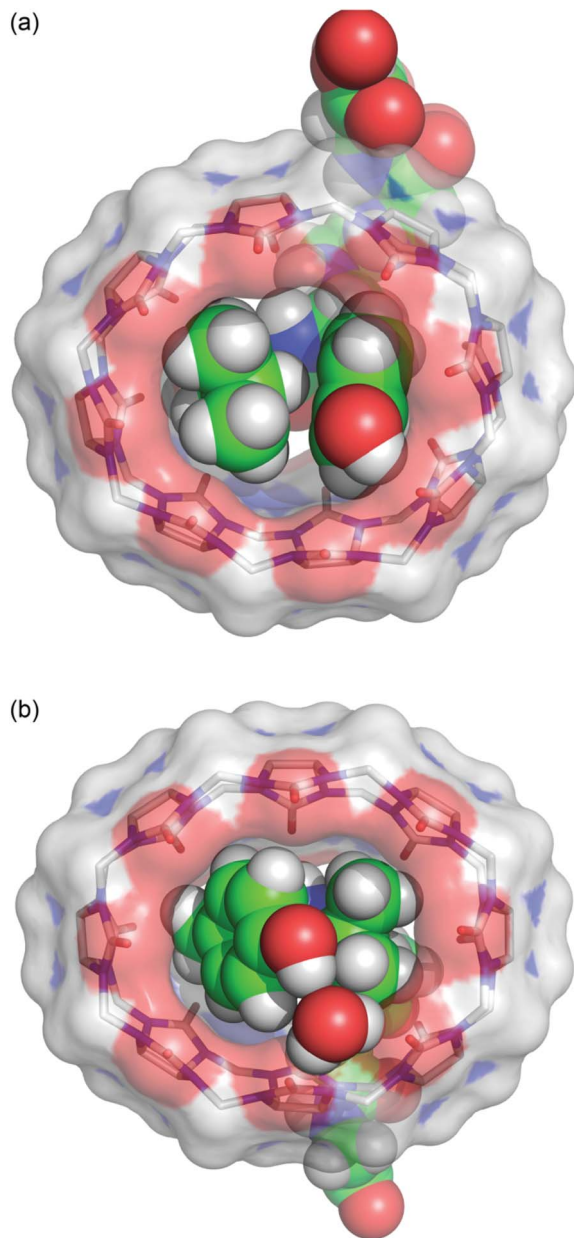
(a)

(b)

Fig. 3 Bottom view of the crystal structures of (a) Q8·LYGGG and (b) Q8·YLGGG. The inclusion of Leu and Tyr side chains is apparent in both structures, as is the electrostatic interaction between the Tyr phenolic OH group and the Q8 portal, either *via* direct hydrogen bond (Q8·LYGGG) or *via* a bridging water molecule (Q8·YLGGG). The shape of the host is considerably puckered in the Q8·YLGGG complex but less so in the Q8·LYGGG complex. Host molecules are rendered as sticks with transparent solvent-accessible surface, and peptides and the bridging water are rendered as space-filling models. The molecules are coloured as follows: Q8 carbons in grey, peptide carbons in green, nitrogens in blue, oxygens in red, and hydrogens in white.

(a)

(b)

Fig. 4 Side view of the crystal structures of (a) Q8·LYGGG and (b) Q8·YLGGG. In both structures, the N-terminal ammonium group is proximal to the Q8 portal, and the Leu and Tyr side chains are closely stacked and buried within the Q8 cavity. The peptide is buried more deeply in the Q8·YLGGG structure than in the Q8·LYGGG structure. The Tyr side chain of YLGGG is projected at an angle, positioning the phenolic OH group just outside the centre of the Q8 portal, too far to form a direct hydrogen bond. Host molecules are rendered as sticks with hydrogens omitted for clarity, and peptides are rendered as space-filling models with crystallographic disorder removed. The molecules are coloured as follows: Q8 carbons in grey, peptide carbons in green, nitrogens in blue, oxygens in red, and hydrogens in white.

a hydrogen bond with a proximal C=O oxygen on Q8, which also forms a bifurcated hydrogen bond with the Gly3 NH. The Gly4 NH forms a hydrogen bond with the neighbouring C=O group. In the Q8·YLGGG structure, however, the Tyr NH, Gly3 NH, and Gly4 NH each form a hydrogen bond with a unique, proximal C=O oxygen. The protonated N-terminal ammonium groups in

both complexes form hydrogen bonds with proximal C=O groups on Q8. In the Q8·LYGGG structure, the Leu nitrogen is ≤3.0 Å from two carbonyl oxygens, whereas in the Q8·YLGGG structure, the Tyr nitrogen is ≤3.0 Å from three carbonyl oxygens. The C-terminal carboxylate groups in both complexes do not
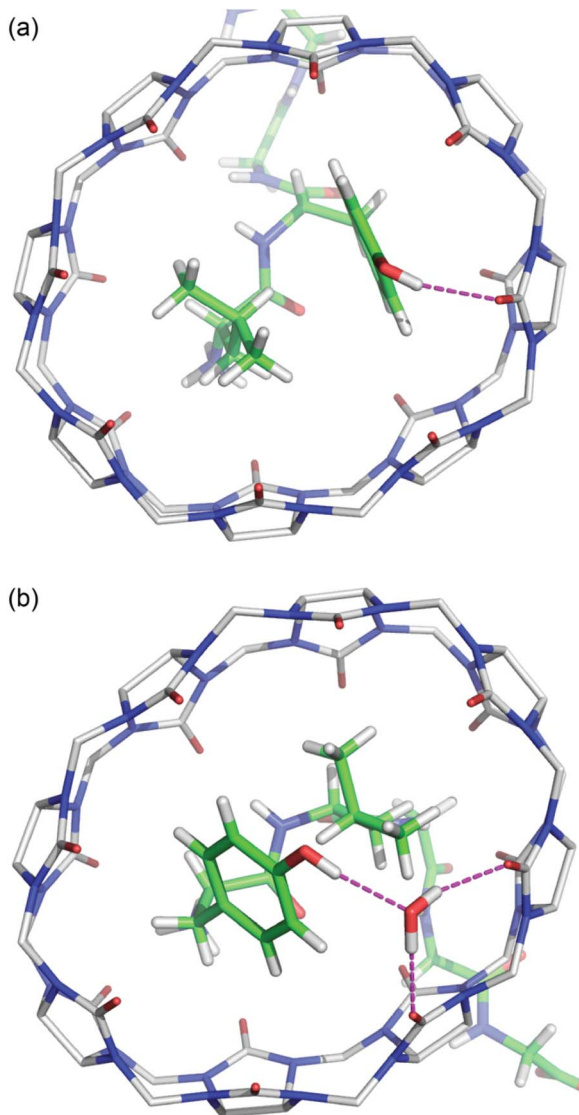
Fig. 5 Bottom view of the crystal structures of (a) Q8·LYGGG and (b) Q8·YLGGG. Highlighted here are electrostatic interactions between the Tyr phenolic OH group and the Q8 portal. Host molecules have hydrogens omitted for clarity. The molecules are coloured as follows: Q8 carbons in grey, peptide carbons in green, nitrogens in blue, oxygens in red, and hydrogens in white.

Fig. 6 Crystal structures of (a) Q8·LYGGG and (b) Q8·YLGGG showing the hydrogen bonds as dashed lines between the peptide backbone and the Q8 portal. Q8 hydrogens have been omitted for clarity. The molecules are coloured as follows: Q8 carbons in grey, peptide carbons in green, nitrogens in blue, oxygens in red, and hydrogens in white.

directly interact with the equator of neighbouring complexes, as observed in the Q8·GGLYGGG structure,[23] but rather they occupy the spaces in between the complexes in the crystal and form hydrogen bonds with the interstitial water molecules. This observation is similar to the occupancy of sulfate, oxonium, and disordered water in the inter-macrocycle spaces in the crystal structure of unbound Q8.[63]

Although the crystal structures provide information about the binding behaviour of Q8 in the solid state, we believe they suggest a molecular basis for the extraordinary selectivity of Q8 for LYGGG *vs.* YLGGG observed in solution. Simultaneous inclusion of the Tyr and Leu side chains within the Q8 cavity presents significant steric demands, and Q8 is highly rigid with significant torsional strain. In the Q8·LYGGG complex, the host

shows minimal puckering, whereas in the Q8·YLGGG complex, the significant deformation of Q8 indicates poor shape complementarity and an energetic cost for binding compared to the Q8·LYGGG complex. In both structures, the N-terminal ammonium group is positioned proximal to the portal, driven by highly stabilizing ion–dipole interactions. This position forces the side chain of the N-terminal residue to insert fully within the Q8 cavity. If the first residue is Tyr, then the large side chain is forced into a conformation in which the ring protrudes from the opposite portal, costing a direct hydrogen bond with C=O groups at that portal and possibly a significant amount of complex stability compared to the Q8·LYGGG complex. If the first residue is Leu, then the aliphatic side chain is able to be fully buried within the Q8 cavity. Therefore, we propose that the 1.2 nM affinity observed in the Q8·LYGGG complex is conferred by the optimal shape of the aliphatic residue at the first position

and aromatic residue at the second position that enables inclusion within the Q8 cavity without significant distortion of the macrocycle and while forming a direct hydrogen bond between the Tyr OH group and a Q8 C=O group and maintaining numerous electrostatic interactions between the peptide backbone and the Q8 portal.

Although we were unable to obtain crystal structures for Q8 in complex with KFGGY or FKGGY, we believe that the subnanomolar affinity of Q8 for KFGGY and the 1000-fold selectivity *versus* FKGGY might be explained by factors similar to those observed in the Q8·LYGGG and Q8·YLGGG crystal structures. In both sets of peptides, the higher affinity complex has an aliphatic residue at the first position and an aromatic residue at the second position (*i.e.*, LY *vs.* YL and KF *vs.* FK). Pair inclusion of KF would be expected to induce significantly less puckering of Q8 than inclusion of FK, while allowing the positioning of both the N-terminal ammonium and Lys ε-ammonium proximal to Q8 carbonyl oxygens. Further structural studies would be needed to confirm this hypothesis.

Conclusions

This study presents the unexpected observation of subnanomolar binding affinity between a synthetic receptor and an unmodified peptide in neutral aqueous buffer. This result is remarkable in light of the exceptionally small binding site, only two amino acid residues. Unlike other studies of the pair-inclusion motif, the results presented here demonstrate extraordinary selectivity of Q8 for the dipeptide binding sites Lys–Phe and Leu–Tyr *versus* their sequence isomers Phe–Lys and Tyr–Leu, respectively. Comparative analysis of the isomeric crystal structures of Q8·LYGGG and Q8·YLGGG reveals remarkable electrostatic and shape complementarity between host and guest that confers such stability and sequence selectivity. Given the many reports demonstrating successful incorporation of the tripeptide FGG affinity tag at the N-terminus of proteins for micromolar recognition by Q8, we anticipate that incorporation of KF or LY as N-terminal protein affinity tags should enable predictive and selective recognition by Q8 at subnanomolar concentrations. This capability would be an important step toward the development of applications such as protein sensing or inhibition under physiologically relevant conditions while having minimal impact on the structure and properties of the tagged protein. The future development of derivatization strategies for Q8 should greatly enhance the applicability of this approach to protein recognition.

Data availability

All data including experimental and analytical details are in the ESI.†

Author contributions

P. S., M. B. E., L. B. K., H. D. A., D. J. W., and A. R. U. conducted experiments and analytical studies. P. S., M. B. E., and A. R. U. wrote the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

‡ We follow the definition of hydrogen bonding from the 2011 recommendations of the IUPAC.[66,67] Briefly, in a D–H···A system, the bonding is an attractive interaction stemming from either electrostatics, dispersion, or charge transfer; the donor (D) and acceptor (A) are more electronegative than H; the DHA angle approaches 180° with a minimum of 110°; the H···A distance ranges from ~1.0 to 2.6 Å, and where there is either spectroscopic or crystallographic evidence for bond formation.

1 C. L. Young, Z. T. Britton and A. S. Robinson, Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications, *Biotechnol. J.*, 2012, **7**, 620–634, DOI: **10.1002/biot.201100155**.

2 *Protein Purification: Principles, High Resolution Methods, and Applications*, ed. J.-C. Janson, Wiley, Hoboken, USA, 3rd edn, 2011.

3 O. Boutureira and G. J. L. Bernardes, Advances in Chemical Protein Modification, *Chem. Rev.*, 2015, **115**, 2174–2195, DOI: **10.1021/cr500399p**.

4 V. Mishra, Affinity Tags for Protein Purification, *Curr. Protein Pept. Sci.*, 2020, **21**, 821–830, DOI: **10.2174/1389203721666200606220109**.

5 P. Hengen, Purification of His-Tag fusion proteins from Escherichia coli, *Trends Biochem. Sci.*, 1995, **20**, 285–286, DOI: **10.1016/s0968-0004(00)89045-3**.

6 C. Zhao, L. M. Hellman, X. Zhan, W. S. Bowman, S. W. Whiteheart and M. G. Fried, Hexahistidine-tag-specific optical probes for analyses of proteins and their interactions, *Anal. Biochem.*, 2010, **399**, 237–245, DOI: **10.1016/j.ab.2009.12.028**.

7 J. Arnau, C. Lauritzen, G. E. Petersen and J. Pedersen, Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins, *Protein Expression Purif.*, 2006, **48**, 1–13, DOI: **10.1016/j.pep.2005.12.002**.

8 J. J. BelBruno, Molecularly Imprinted Polymers, *Chem. Rev.*, 2019, **119**, 94–119, DOI: **10.1021/acs.chemrev.8b00171**.

9 J. Murray, K. Kim, T. Ogoshi, W. Yao and B. C. Gibb, The aqueous supramolecular chemistry of cucurbit[n]urils,

pillar[n]arenes and deep-cavity cavitands, *Chem. Soc. Rev.*, 2017, **46**, 2479–2496, DOI: **10.1039/c7cs00095b**.

10 J. Lagona, P. Mukhopadhyay, S. Chakrabarti and L. Isaacs, The Cucurbit[n]uril Family, *Angew. Chem., Int. Ed.*, 2005, **44**, 4844–4870, DOI: **10.1002/anie.200460675**.

11 S. J. Barrow, S. Kasera, M. J. Rowland, J. del Barrio and O. A. Scherman, Cucurbituril-Based Molecular Recognition, *Chem. Rev.*, 2015, **115**, 12320–12406, DOI: **10.1021/acs.chemrev.5b00341**.

12 M. E. Bush, N. D. Bouley and A. R. Urbach, Charge-Mediated Recognition of *N*-Terminal Tryptophan in Aqueous Solution by a Synthetic Host, *J. Am. Chem. Soc.*, 2005, **127**, 14511–14517, DOI: **10.1021/ja0548440**.

13 M. V. Rekharsky, H. Yamamura, C. Inoue, M. Kawai, I. Osaka, R. Arakawa, K. Shiba, A. Sato, Y. H. Ko, N. Selvapalam, *et al.*, Chiral recognition in cucurbituril cavities, *J. Am. Chem. Soc.*, 2006, **128**, 14871–14880, DOI: **10.1021/ja063323p**.

14 L. M. Heitmann, A. B. Taylor, P. J. Hart and A. R. Urbach, Sequence-Specific Recognition and Cooperative Dimerization of *N*-Terminal Aromatic Peptides in Aqueous Solution by a Synthetic Host, *J. Am. Chem. Soc.*, 2006, **128**, 12574–12581, DOI: **10.1021/ja064323s**.

15 M. V. Rekharsky, H. Yamamura, Y. H. Ko, N. Selvapalam, K. Kim and Y. Inoue, Sequence recognition and self-sorting of a dipeptide by cucurbit[6]uril and cucurbit[7] uril, *Chem. Commun.*, 2008, 2236–2238, DOI: **10.1039/b719902c**.

16 A. R. Urbach and V. Ramalingam, Molecular Recognition of Amino Acids, Peptides, and Proteins by Cucurbit[*n*]uril Receptors, *Isr. J. Chem.*, 2011, **51**, 664–678, DOI: **10.1002/ijch.201100035**.

17 J. M. Chinai, A. B. Taylor, L. M. Ryno, N. D. Hargreaves, C. A. Morris, P. J. Hart and A. R. Urbach, Molecular Recognition of Insulin by a Synthetic Receptor, *J. Am. Chem. Soc.*, 2011, **133**, 8810–8813, DOI: **10.1021/ja201581x**.

18 L. A. Logsdon, C. L. Schardon, V. Ramalingam, S. K. Kwee and A. R. Urbach, Nanomolar binding of peptides containing noncanonical amino acids by a synthetic receptor, *J. Am. Chem. Soc.*, 2011, **133**, 17087–17092, DOI: **10.1021/ja207825y**.

19 L. C. Smith, D. G. Leach, B. E. Blaylock, O. A. Ali and A. R. Urbach, Sequence-Specific, Nanomolar Peptide Binding via Cucurbit[8]uril-Induced Folding and Inclusion of Neighboring Side Chains, *J. Am. Chem. Soc.*, 2015, **137**, 3663–3669, DOI: **10.1021/jacs.5b00718**.

20 Z. Hirani, H. F. Taylor, E. F. Babcock, A. T. Bockus, C. D. Varnado, C. W. Bielawski and A. R. Urbach, Molecular Recognition of Methionine-Terminated Peptides by Cucurbit[8]uril, *J. Am. Chem. Soc.*, 2018, **140**, 12263–12269, DOI: **10.1021/jacs.8b07865**.

21 G. Wu, D. E. Clarke, C. Wu and O. A. Scherman, Oligopeptide-CB[8] complexation with switchable binding pathways, *Org. Biomol. Chem.*, 2019, **17**, 3514–3520, DOI: **10.1039/c9ob00592g**.

22 D. E. Clarke, G. Wu, C. Wu and O. A. Scherman, Host–Guest Induced Peptide Folding with Sequence-Specific Structural

Chirality, *J. Am. Chem. Soc.*, 2021, **143**, 6323–6327, DOI: **10.1021/jacs.1c00342**.

23 P. Suating, L. B. Kimberly, M. B. Ewe, S. L. Chang, J. M. Fontenot, P. R. Sultane, C. W. Bielawski, D. A. Decato, O. B. Berryman, A. B. Taylor and A. R. Urbach, Cucurbit[8]uril Binds Nonterminal Dipeptide Sites with High Affinity and Induces a Type II β-Turn, *J. Am. Chem. Soc.*, 2024, DOI: **10.1021/jacs.3c14045**.

24 Y. Ling, W. Wang and A. E. Kaifer, A new cucurbit[8]uril-based fluorescent receptor for indole derivatives, *Chem. Commun.*, 2007, 610–612, DOI: **10.1039/b611559d**.

25 F. Biedermann and W. M. Nau, Noncovalent chirality sensing ensembles for the detection and reaction monitoring of amino acids, peptides, proteins, and aromatic drugs, *Angew. Chem., Int. Ed.*, 2014, **53**, 5694–5699, DOI: **10.1002/anie.201400718**.

26 F. Biedermann, G. Ghale, A. Hennig and W. M. Nau, Fluorescent artificial receptor-based membrane assay (FARMA) for spatiotemporally resolved monitoring of biomembrane permeability, *Commun. Biol.*, 2020, **3**, 383, DOI: **10.1038/s42003-020-1108-9**.

27 H. Li, H. Xie, Y. Cao, X. Ding, Y. Yin and G. Li, A general way to assay protein by coupling peptide with signal reporter via supermolecule formation, *Anal. Chem.*, 2013, **85**, 1047–1052, DOI: **10.1021/ac302906c**.

28 F. Tian, M. Cziferszky, D. Jiao, K. Wahlstrom, J. Geng and O. A. Scherman, Peptide separation through a CB[8]-mediated supramolecular trap-and-release process, *Langmuir*, 2011, **27**, 1387–1390, DOI: **10.1021/la104346k**.

29 S. Sonzini, A. Marcozzi, R. J. Gubeli, C. F. Van Der Walle, P. Ravn, A. Herrmann and O. A. Scherman, High Affinity Recognition of a Selected Amino Acid Epitope within a Protein by Cucurbit[8]uril Complexation, *Angew. Chem., Int. Ed.*, 2016, **55**, 14000–14004, DOI: **10.1002/anie.201606763**.

30 Q. An, J. Brinkmann, J. Huskens, S. Krabbenborg, J. de Boer and P. Jonkheijm, A supramolecular system for the electrochemically controlled release of cells, *Angew. Chem., Int. Ed.*, 2012, **51**, 12233–12237, DOI: **10.1002/anie.201205651**.

31 J. W. Lee, M. H. Shin, W. Mobley, A. R. Urbach and H. I. Kim, Supramolecular Enhancement of Protein Analysis via the Recognition of Phenylalanine with Cucurbit[7]uril, *J. Am. Chem. Soc.*, 2015, **137**, 15322–15329, DOI: **10.1021/jacs.5b10648**.

32 G. Ghale, V. Ramalingam, A. R. Urbach and W. M. Nau, Determining protease substrate selectivity and inhibition by label-free supramolecular tandem enzyme assays, *J. Am. Chem. Soc.*, 2011, **133**, 7528–7535, DOI: **10.1021/ja2013467**.

33 L. A. Logsdon and A. R. Urbach, Sequence-specific inhibition of a nonspecific protease, *J. Am. Chem. Soc.*, 2013, **135**, 11414–11416, DOI: **10.1021/ja406032x**.

34 X. Tan, L. Yang, Y. Liu, Z. Huang, H. Yang, Z. Wang and X. Zhang, Water-soluble supramolecular polymers fabricated through specific interactions between cucurbit[8]uril and a tripeptide of Phe-Gly-Gly, *Polym. Chem.*, 2013, **4**, 5378–5381, DOI: **10.1039/c3py00888f**.

35 Z. Huang, Y. Fang, Q. Luo, S. Liu, G. An, C. Hou, C. Lang, J. Xu, Z. Dong and J. Liu, Construction of supramolecular polymer by enzyme-triggered covalent condensation of CB [8]-FGG-based supramonomer, *Chem. Commun.*, 2016, **52**, 2083–2086, DOI: **10.1039/c5cc09103a**.

36 M. J. Rowland, E. A. Appel, R. J. Coulston and O. A. Scherman, Dynamically crosslinked materials via recognition of amino acids by cucurbit[8]uril, *J. Mater. Chem. B*, 2013, **1**, 2904–2910, DOI: **10.1039/c3tb20180e**.

37 Q. Song, Y. Gao, J. F. Xu, B. Qin, M. J. Serpe and X. Zhang, Supramolecular Microgels Fabricated from Supramonomers, *ACS Macro Lett.*, 2016, **5**, 1084–1088, DOI: **10.1021/acsmacrolett.6b00592**.

38 S. Li, N. Jiang, W. Zhao, Y. F. Ding, Y. Zheng, L. H. Wang, J. Zheng and R. Wang, An eco-friendly in situ activatable antibiotic via cucurbit[8]uril-mediated supramolecular crosslinking of branched polyethylenimine, *Chem. Commun.*, 2017, **53**, 5870–5873, DOI: **10.1039/c7cc02466e**.

39 H. D. Nguyen, D. T. Dang, J. L. van Dongen and L. Brunsveld, Protein Dimerization Induced by Supramolecular Interactions with Cucurbit[8]uril, *Angew. Chem., Int. Ed. Engl.*, 2010, **49**, 895–898, DOI: **10.1002/anie.200904413**.

40 D. T. Dang, H. D. Nguyen, M. Merkx and L. Brunsveld, Supramolecular control of enzyme activity through cucurbit[8]uril-mediated dimerization, *Angew. Chem., Int. Ed. Engl.*, 2013, **52**, 2915–2919, DOI: **10.1002/anie.201208239**.

41 R. P. Bosmans, J. M. Briels, L. G. Milroy, T. F. de Greef, M. Merkx and L. Brunsveld, Supramolecular Control over Split-Luciferase Complementation, *Angew. Chem., Int. Ed. Engl.*, 2016, **55**, 8899–8903, DOI: **10.1002/anie.201602807**.

42 D. T. Dang, J. Schill and L. Brunsveld, Cucurbit[8]uril-mediated protein homotetramerization, *Chem. Sci.*, 2012, **3**, 2679–2684, DOI: **10.1039/c2sc20625k**.

43 C. Hou, J. Li, L. Zhao, W. Zhang, Q. Luo, Z. Dong, J. Xu and J. Liu, Construction of protein nanowires through cucurbit[8]uril-based highly specific host-guest interactions: an approach to the assembly of functional proteins, *Angew. Chem., Int. Ed. Engl.*, 2013, **52**, 5590–5593, DOI: **10.1002/anie.201300692**.

44 X. Li, Y. Bai, Z. Huang, C. Si, Z. Dong, Q. Luo and J. Liu, A highly controllable protein self-assembly system with morphological versatility induced by reengineered host-guest interactions, *Nanoscale*, 2017, **9**, 7991–7997, DOI: **10.1039/c7nr01612c**.

45 H. H. Lee, T. S. Choi, S. J. Lee, J. W. Lee, J. Park, Y. H. Ko, W. J. Kim, K. Kim and H. I. Kim, Supramolecular inhibition of amyloid fibrillation by cucurbit[7]uril, *Angew. Chem., Int. Ed. Engl.*, 2014, **53**, 7461–7465, DOI: **10.1002/anie.201402496**.

46 N. E. de Almeida, T. D. Do, M. Tro, N. E. LaPointe, S. C. Feinstein, J. E. Shea and M. T. Bowers, Opposing Effects of Cucurbit[7]uril and 1,2,3,4,6-Penta-*O*-galloyl-β-D-glucopyranose on Amyloid β25-35 Assembly, *ACS Chem. Neurosci.*, 2016, **7**, 218–226, DOI: **10.1021/acschemneuro.5b00280**.

47 S. Sankaran, M. de Ruiter, J. J. Cornelissen and P. Jonkheijm, Supramolecular Surface Immobilization of Knottin Derivatives for Dynamic Display of High Affinity Binders, *Bioconjugate Chem.*, 2015, **26**, 1972–1980, DOI: **10.1021/acs.bioconjchem.5b00419**.

48 R. P. Bosmans, W. E. Hendriksen, M. Verheijden, R. Eelkema, P. Jonkheijm, J. H. van Esch and L. Brunsveld, Supramolecular Protein Immobilization on Lipid Bilayers, *Chem.–Eur. J.*, 2015, **21**, 18466–18473, DOI: **10.1002/chem.201502461**.

49 M. J. Webber, E. A. Appel, B. Vinciguerra, A. B. Cortinas, L. S. Thapa, S. Jhunjhunwala, L. Isaacs, R. Langer and D. G. Anderson, Supramolecular PEGylation of biopharmaceuticals, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 14189–14194, DOI: **10.1073/pnas.1616639113**.

50 C. L. Maikawa, A. A. A. Smith, L. Zou, G. A. Roth, E. C. Gale, L. M. Stapleton, S. W. Baker, J. L. Mann, A. C. Yu, S. Correa, *et al.*, A co-formulation of supramolecularly stabilized insulin and pramlintide enhances mealtime glucagon suppression in diabetic pigs, *Nat. Biomed. Eng.*, 2020, **4**, 507–517, DOI: **10.1038/s41551-020-0555-4**.

51 S. Liu, C. Ruspic, P. Mukhopadhyay, S. Chakrabarti, P. Y. Zavalij and L. Isaacs, The cucurbit[n]uril family: prime components for self-sorting systems, *J. Am. Chem. Soc.*, 2005, **127**, 15959–15967, DOI: **10.1021/ja055013x**.

52 C. L. D. Gibb and B. C. Gibb, in *Supramolecular Chemistry: From Molecules to Nanomaterials*, ed. P. A. Gale and J. W. Steed, John Wiley & Sons, Ltd, 2012, vol. 2, DOI: **10.1002/9780470661345.smc005**.

53 M. V. Rekharsky, T. Mori, C. Yang, Y. H. Ko, N. Selvapalam, H. Kim, D. Sobransingh, A. E. Kaifer, S. Liu, L. Isaacs, *et al.*, A synthetic host-guest system achieves avidin-biotin affinity by overcoming enthalpy-entropy compensation, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 20737–20742, DOI: **10.1073/pnas.0706407105**.

54 A. Krężel and W. Bal, A formula for correlating p*K*a values determined in D_2O and H_2O, *J. Inorg. Biochem.*, 2004, **98**, 161–166, DOI: **10.1016/j.jinorgbio.2003.10.001**.

55 H. Günther, *NMR Spectroscopy: Basic Principles, Concepts and Applications in Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 3rd edn, 2013.

56 O. Zerbe and S. Jurt, *Applied NMR Spectroscopy for Chemists and Life Scientists*, Wiley-VCH, Weinheim, 2014.

57 K. Moon and A. E. Kaifer, Modes of binding interaction between viologen guests and the cucurbit[7]uril host, *Org. Lett.*, 2004, **6**, 185–188, DOI: **10.1021/ol035967x.**

58 D. A. Dougherty, Cation-π interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp, *Science*, 1996, **271**, 163–168, DOI: **10.1126/science.271.5246.163.**

59 J. C. Ma and D. A. Dougherty, The Cation–π Interaction, *Chem. Rev.*, 1997, **97**, 1303–1324, DOI: **10.1021/cr9603744**.

60 J. P. Gallivan and D. A. Dougherty, Cation-π interactions in structural biology, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9459–9464, DOI: **10.1073/pnas.96.17.9459**.

61 X. Xiu, N. L. Puskar, J. A. Shanata, H. A. Lester and D. A. Dougherty, Nicotine binding to brain receptors

requires a strong cation-pi interaction, *Nature (London, U. K.)*, 2009, **458**, 534–537, DOI: **10.1038/nature07768**.

62 R. A. Pascal, R. B. Grossman and D. Van Engen, Synthesis of *in*-[34,10][7]metacyclophane: projection of an aliphatic hydrogen toward the center of an aromatic ring, *J. Am. Chem. Soc.*, 1987, **109**, 6878–6880, DOI: **10.1021/ja00256a067**.

63 J. W. Lee, S. Samal, N. Selvapalam, H. J. Kim and K. Kim, Cucurbituril homologues and derivatives: new opportunities in supramolecular chemistry, *Acc. Chem. Res.*, 2003, **36**, 621–630, DOI: **10.1021/ar020254k**.

64 S. Tsuzuki and A. Fujii, Nature and physical origin of CH/π interaction: significant difference from conventional hydrogen bonds, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2584–2594, DOI: **10.1039/B718656H**.

65 S. Tsuzuki, K. Honda, A. Fujii, T. Uchimaru and M. Mikami, CH/π interactions in methane clusters with polycyclic aromatic hydrocarbons, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2860, DOI: **10.1039/b718550b**.

66 E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, *et al.*, Definition of the hydrogen bond (IUPAC Recommendations 2011), *Pure Appl. Chem.*, 2011, **83**, 1637–1641, DOI: **10.1351/pac-rec-10-01-02**.

67 A. L. Thompson and N. G. White, Hydrogen atoms in supramolecular chemistry: a structural perspective. Where are they, and why does it matter?, *Chem. Soc. Rev.*, 2023, **52**, 6254–6269, DOI: **10.1039/d3cs00516j**.