High Affinity Recognition of a Selected Amino Acid Epitope within a Protein by Cucurbit[8]uril Complexation

Silvia Sonzini[+], Alessio Marcozzi[+], Raphael J. Gubeli, Christopher F. van der Walle, Peter Ravn, Andreas Herrmann, and Oren A. Scherman**

Abstract: Supramolecular interactions between the host cucurbit[8]uril (CB[8]) and amino acids have been widely interrogated, but recognition of specific motifs within a protein domain have never been reported. A phage display approach was herein used to select motifs with the highest binding affinity for the heteroternary complex with methyl viologen and CB[8] (MV·CB[8]) within a vast pool of cyclic peptide sequences. From the selected motifs, an epitope consisting of three amino acid was extrapolated and incorporated into a solvent-exposed loop of a protein domain; the protein exhibited micromolar binding affinity for the MV·CB[8] complex, matching that of the cyclic peptide. By achieving selective CB[8]-mediated conjugation of a small molecule to a recombinant protein scaffold we pave the way to biomedical applications of this simple ternary system.

Over the past few decades efforts to obtain libraries of compounds capable of binding therapeutically relevant targets have been made, in order to improve drug delivery efficiency. This research has been especially focused on antibody derivatives and synthetic receptor drug conjugates.[1–3] Monoclonal antibodies are characterized by tight and specific binding, low toxicity, and immunogenicity, however, it is often difficult to develop suitable conjugation chemistry to drugs of interest, without losing any potency, and still retaining a high production yield.[1,4] In the last few years, host–guest chemistry has played an important role in tether-

ing peptides and proteins together, opening new approaches towards biomedical targets.[5–11]

In particular, several examples of interactions of peptides and proteins with different members of the cucurbit[n]uril (CB[n]) family have been reported,[12–19] however, they have mainly been restricted to N-termini. CB[n]s are a family of macrocyles prepared by condensation of glycouril at high temperature under acidic conditions.[20,21] They feature both hydrophobic (within the cavity) and polar interactions (at the carbonyl rims) with guests;[21] moreover, these macrocycles exhibit strong complexation in an aqueous environment with low cytotoxicity.[22,23] Within the CB[n] family, the eight-membered ring, CB[8], has assumed a prominent place on account of its ability to bind two aromatic guests simultaneously in its large cavity.[7–9,12,13,24,25]

Herein, short amino acid (AA) motifs displayed within a peptide sequence are identified as specific binders with the preformed 1:1 methyl viologen–CB[8] (MV·CB[8]) complex (Figure 1). The most promising three AA epitope was translated into a protein domain, which subsequently showed retention of specific binding for MV·CB[8] with the same affinity. We thus report the first example of specific CB[8]-mediated conjugation of a small molecule to a short AA motif expressed within a protein domain, enabling the application of host–guest chemistry strategies in protein–small molecule conjugation.

[*] Dr. S. Sonzini,[+] Prof. O. A. Scherman
Melville Laboratory for Polymer Synthesis
Department of Chemistry, University of Cambridge
Lensfield Road, Cambridge, CB2 1EW (UK)
E-mail: oas23@cam.ac.uk

Dr. A. Marcozzi,[+] Prof. A. Herrmann
Zernike Institute for Advanced Materials
Dept. of Polymer Chemistry, University of Groningen
Nijenborgh 4, 9747 AG, Groningen (The Netherlands)
E-mail: a.herrmann@rug.nl

Dr. R. J. Gubeli, Dr. C. F. van der Walle
Formulation Sciences, MedImmune Ltd.
Granta Park, Cambridge CB21 6GH (UK)

Dr. P. Ravn
Antibody Discovery and Protein Engineering, MedImmune Ltd.
Granta Park, Cambridge CB21 6GH (UK)

Dr. S. Sonzini[+]
Present address: Formulation Sciences, MedImmune Ltd.
Granta Park, Cambridge CB21 6GH (UK)

[+] These authors contributed equally to this work.

Supporting information for this article can be found under:
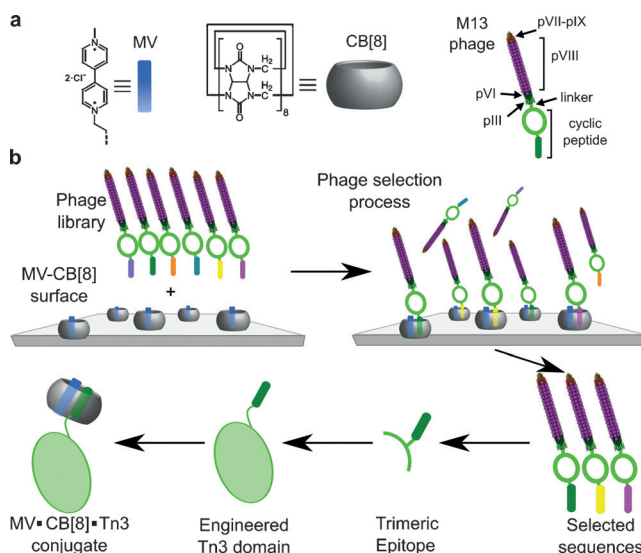http://dx.doi.org/10.1002/anie.201606763.

Figure 1. Schematic representation of the pathway from the phage display selection process to the recognition of an internal motif expressed into a Tn3 domain by MV·CB[8] complex.
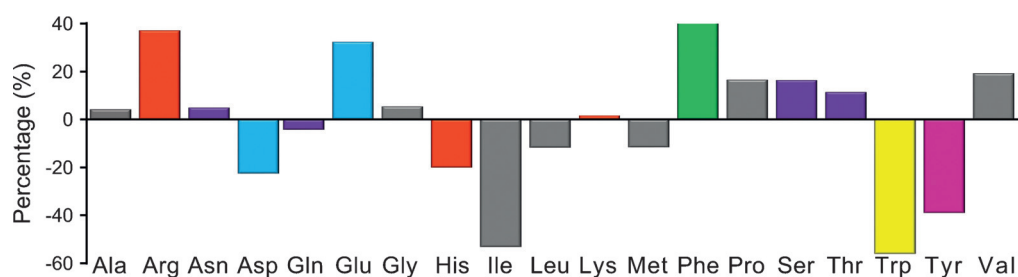
Figure 2. Selection results from the phage display data. Bar chart representing the percentage difference in abundance after the selection, relative to before the selection. The amino acids are represented with different colors depending on their inherent characteristics: orange for positively charged, light-blue for negatively charged, purple for polar and gray for lack of any specific feature; the aromatic amino acids Phe, Trp and Tyr have been highlighted with three different colours: green, yellow and pink, respectively. Cys is generally underrepresented within the phage libraries; in our analyses it was found only once before selection and never after selection, thus it was excluded from this chart.

Phage display is widely applied to isolate monoclonal antibodies as well as in the identification of epitopes, but also to characterize peptide/protein interactions towards a wide variety of molecules;[26–28] hence, it was selected as a promising technique to screen an extremely large number of sequences against binding with MV·CB[8]. In order to avoid any bias deriving from aromatic N-terminal residues, a cyclic phage library, PhD™-C7C, was analyzed. Although N-terminal residues are known to be good guests,[17] the analysis of those sequences would then restrict the incorporation of a recognition motif only at a protein N-terminus.

The PhD™-C7C library based on bacteriophage M13 is composed of cyclic peptides containing seven random amino acids between two cysteines and followed by a GGGS linker fused with the capsid protein pIII. An extra N-terminal alanine is also present leading to the following schematic sequence: NH_2-ACX$_7$CGGGS-p3-COOH. The phage library was cast on covalently bound MV surfaces, which were fully complexed with CB[8]. Three rounds of selection were carried out: one round with a neutral phosphate buffer solution (PB, pH 7.4), followed by two rounds with acidic PB (pH 5.0). The phages eluted at the end of the third round of selection were sequenced and analyzed to obtain any information on selected motifs and amino acid composition.

The relative percentage increase (or decrease) of each AA residue following the selection was analyzed (Figures 2 and S6 in the Supporting Information). We expected an over-representation of all aromatic residues, especially Trp, which has been reported to be the strongest binder to MV·CB[8];[17] nevertheless, the bar charts shown in both Figures 2 and S6, clearly indicate a significant increase in Phe, but a decrease in both Trp and Tyr. Residue representation analysis was further applied to understand whether specific epitopes were selected. Figure S6 shows that Ser and Thr were the most represented residues, followed by polar and positively charged residues such as Asn, Lys, Arg, and the rigid Pro. Nevertheless, the negative residue Glu was as over-represented in the selected sequences as the positively charged Arg, while other highly represented residues, such as Lys, only retained their abundance from the original pool (see Figure 2). Although phage selection was performed in slightly

acidic conditions (pH 5.0), the selection of Glu (side chain pK_a 4.1) was unexpected as interactions between the carboxylate side chain and the carbonyl portals of CB[8] are typically unfavorable. However, this result emphasizes the capacity for phage display to select epitopes that may not be rationally predicted. Therefore, the overall residue representation analysis highlighted the selection of polar (Arg and Glu), bulky and constrain bearing (Pro and Val) with only one aromatic (Phe) residue.

Additional in-depth analysis of peptide selection revealed the presence of repeated sequences and three AA motifs, reported in Table 1. Interestingly, all of the repeated sequences contained a different aromatic residue, respectively, Tyr (Seq 1), Phe (Seq 2), and Trp (Seq 3). Hence, the residue representation and peptide selection analyses together suggested that MV·CB[8]·Phe was overall the most stable ternary complex, as later confirmed by further affinity studies. Nevertheless, the selected sequences contained other aromatic residues, highlighting the importance of the neighboring environment towards the formation of the MV·CB-[8]·peptide complex and not just of the moiety binding within the cavity itself.

Table 1: Most repeated sequences and 3-*mer* motifs obtained from analyses of the phage display results.

	Sequence	3-*mer* motif	Aromatic AA
Seq 1	NTGSPYE	SPY(E)[a]	Tyr
Seq 2	QNPNQKF	QKF	Phe
Seq 3	LKLGEKW	EKW	Trp

[a] SPY and PYE motifs were repeated the same number of times and they are both part of Seq 1.

In order to verify the results obtained by phage display, the three selected sequences were synthesized and their binding with the MV·CB[8] binary complex analyzed by fluorescence spectroscopy (Figures 2b and S7,8) and isothermal titration calorimetry (ITC) (Figures 2c,d and S7). The peptides were synthesized as: NH_2-ACX$_7$C-CONH$_2$; the C-terminal amide, N-terminal Ala and the two Cys residues (in positions 2 and 10) were introduced in order to obtain cyclic 10-*mer* sequences (Seq-*c*) resembling the conformation of the cyclic peptides displayed on the phage coat pIII protein.

Quenching of the fluorescence emission upon addition of MV·CB[8] is shown in Figure 3b. All of the residues showed a degree of fluorescence quenching and reached a plateau at a 1:1 molar ratio (MV·CB[8] to Seq-*c*), suggesting the
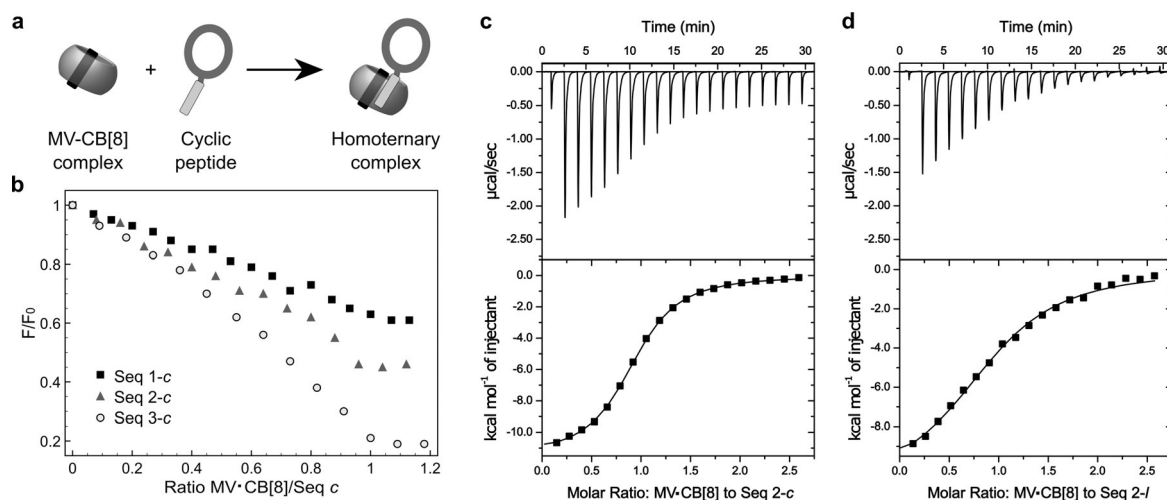
Figure 3. Binding evaluation of Seq-*c* upon MV·CB[8] addition. a) Schematic representation of heteroternary binding between MV·CB[8] and one of the cyclic peptide sequences. b) Fluorescence quenching vs. concentration ratio of MV·CB[8] and Seq 1–3 in their cyclic version. c) ITC between MV·CB[8]· Seq 2-*c* and d) MV·CB[8]·Seq 2-*l*. The analyses were performed titrating MV·CB[8] complex (1 mM) into a peptide solution (20–30 and 80 μM, for the fluorescence and ITC data, respectively) in 10 mM PB pH 7.4.

formation of the heteroternary complex. Consistent with the fluorescence data, ITC analyses (Figures 3c and S9) showed binding at a 1:1 molar ratio for all of the sequences, although only Seq 2-*c* exhibited a higher binding constant (Table 2) compared to previously recorded N-terminal Phe binding to MV·CB[8].[17] Corroborating with previous reports,[17] formation of the heteroternary complex is mainly enthalpically driven, in only one sequence (Seq 1-*c*) is an additional favorable entropic contribution evident. The positive entropic contribution is likely related to the release of additional water molecules, both from the solvating sphere of the peptide and possibly the CB[8] cavity.[29]

Table 2: Binding constants for heteroternary complexes of MV·CB[8] with Seq 1–3.

	$K_a^{[a]}$ [10^4 M^{-1}]	K_d [μM]	ΔH [kcal mol^{-1}]	$T\Delta S^{[b]}$ [kcal mol^{-1}]
Seq 1-*c*	4.3 ± 0.8	23 ± 4.3	−1.3 ± 0.1	5.0 ± 0.2
Seq 2-*c*	22 ± 1	4.5 ± 0.2	−11 ± 0.1	−3.9 ± 0.1
Seq 3-*c*	4.4 ± 0.7	23 ± 3.6	−12 ± 0.7	−5.5 ± 0.7
Seq 1-*l*[c]	–	–	–	–
Seq 2-*l*	5.9 ± 0.6	17 ± 1.7	−11 ± 0.4	−4.6 ± 0.4
Seq 3-*l*[c]	–	–	–	–

[a] ITC experiments were run in triplicate at 25 °C in 10 mM PB pH 7.4; [b] Entropic contributions to ΔG were calculated from the K_a and ΔH values, with error propagated from those of K_a and ΔH; [c] The values obtained from the titration were too low to be fit.

To further understand the role of rigidity on ternary complex formation (introduced into the system by the cystine bridge) the affinity of the linear version (Seq-*l*) of the selected sequences was also evaluated. Fluorescence titrations carried out on linear sequences also showed quenching, but generally to a lower extent compared to the cyclic sequences (as marked by dashed lines in Figure S9), suggesting lower affinity as previously reported for Trp containing peptides

binding to MV·CB[8].[30] The ITC data, reported in Table 2 and in Figures 3d and S9, showed a general decrease in affinity for all of the sequences. While Seq 2-*l* showed an increase of more than 3.5 fold in dissociation constant, Seq 1-*l* did not present any binding and Seq 3-*l* displayed a low enthalpic interaction that did not allow for fitting of the data points (Figure S9b). The main difference in thermodynamic contributions between Seq 2-*c* and *l* is in the entropic term (more unfavorable for the linear sequence), suggesting an introduction of rigidity in the linear peptide upon binding. On account of the overall data, it is possible to infer that the steric constraints introduced by the cystine bridge are fundamental towards the enhancement in affinity.

Encouraged by the promising results obtained, the best performing sequence, Seq 2, was engineered into a larger biomolecule and its ability to bind the MV·CB[8] complex evaluated. The protein of choice for these further studies was Tn3, which is derived from the third fibronectin type-III domain of tenascin C.[31,32] Tn3 are immunoglobulin-like domains and adopt a β-fold with non-contiguous loops at both poles, which allow for a large binding surface (Figure 4 a,b). Even if binding can occur at both poles of the Tn3 domain,[32] three out of the six loops exposed at the surface, namely BC, DE and FG (Figure 4b), are structurally similar to an antibody complementarity determining region (CDR) and, thus, are the best candidates for modification and introduction of specific binding motifs.[4,32]

From the epitope analysis, a short three AA motif (QKF) was extrapolated, which is part of Seq 2 (Table 1), but even more frequently repeated within the peptides sequenced. In order to avoid the modification of a large number of residues, this epitope was inserted instead of the full Seq 2 in the BC, DE and FG loops of the Tn3 domain. The binding of the modified Tn3 domains with precomplexed MV·CB[8] was evaluated by ITC in similar conditions used for the 10-*mer* sequences. The ITC data for each domain, shown in Figures 4c and S10, were fit to obtain thermodynamic data as
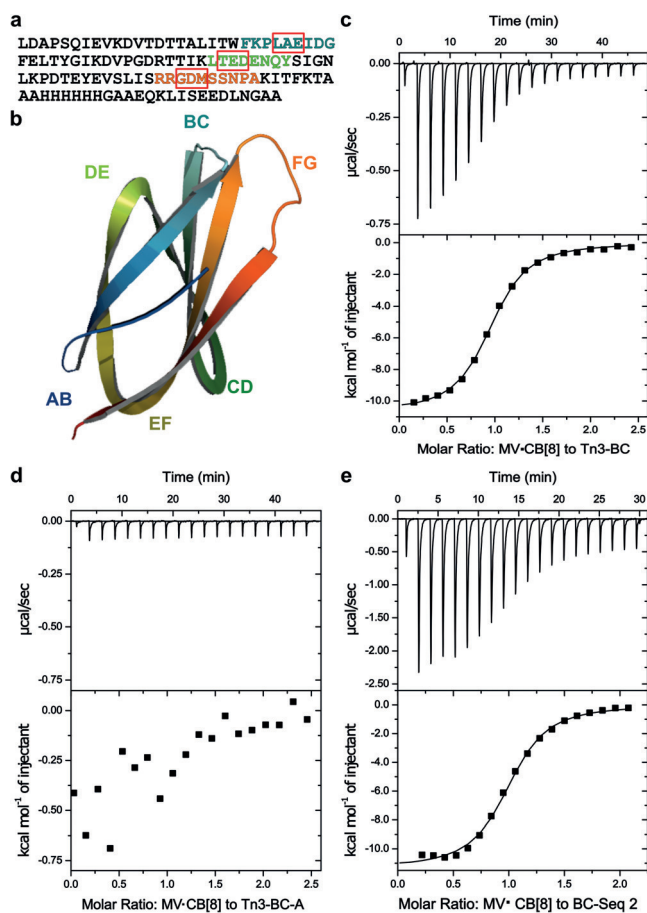
Figure 4. a) Sequence and b) crystallographic structure of starting Tn3 domain (1FNA PDB); the loops are highlighted in different colors and red boxes are added around the residues that were changed in the modified domains. ITC of MV·CB[8] complex (0.5 or 1 mM) into c) Tn3 domain with BC loop modified (Tn3-BC, 50 μM), d) Tn3-BC domain with Ala instead of Phe (Tn3-BC-A, 50 μM) and e) cycled peptide from the Tn3-BC loop (BC-Seq 2, 80 μM). All the solutions were prepared in 10 mM PB pH 7.4.

reported in Table 3. The three AA motif was inserted into the three solvent exposed loops of the Tn3 domain but only substitution into loop BC (Tn3-BC) generated a scaffold with micromolar affinity for the MV·CB[8] complex. The other domains did not present a measureable binding event. In order to confirm that the heteroternary complex MV·CB-[8]·Tn3-BC was related to the inserted Phe within the QKF

Table 3: ITC Data for MV·CB[8] complexed with Tn3 modified domains.

	$K_a^{[a]}$ [10^4 M^{-1}]	K_d [μM]	ΔH [kcal mol^{-1}]	$T\Delta S^{[b]}$ [kcal mol^{-1}]
Tn3-BC	61 ± 3	1.6 ± 0.1	−11 ± 0.1	−2.8 ± 0.1
Tn3-DE	8.9 ± 3.3	11 ± 4.1	−2.1 ± 0.3	4.7 ± 0.4
Tn3-FG	6.4 ± 2.6	16 ± 6.5	−1.4 ± 0.3	5.2 ± 0.4
BC-Seq 2	32 ± 3	3.1 ± 0.3	−11 ± 0.1	−3.8 ± 0.1
BC-Seq 2-P	21 ± 1	4.8 ± 0.2	−13 ± 0.1	−5.6 ± 0.1
Tn3-BC-P	35 ± 3	2.9 ± 0.2	−12 ± 0.3	−4.1 ± 0.3

[a] ITC experiments at 25 °C in 10 mM PB pH 7.4; [b] Entropic contributions to ΔG were calculated from the K_a and ΔH values, with error propagated from those of K_a and ΔH.

motif, the affinity of the same Tn3 domain but with Ala substituted for Phe (Tn3-BC-A) was evaluated and shown to have no binding (Figure 4d). It is important to note that the first AA in the BC loop is a Phe residue (Figure 4a), which is solvent exposed, but does not present any affinity for the MV·CB[8] complex. To further ensure selectivity of the system, we evaluated the binding of MV·CB[8] to the Gipg013 Fragment antigen-binding (Fab), which is characterized by the presence of a different Phe-based epitope (PTF) in its complementarity determining region (CDR, Figure S12a).[33] Although the crystal structure (Figure S12a) suggests that the Phe residue should be accessible, no binding event with MV·CB[8] was observed by ITC (Figure S12b), thus confirming the specificity of the system. These experiments further support the fundamental requirement to maintain a favorable amino acid side chain orientation within the selected epitope in order to achieve high affinity heteroternary complex formation.

To investigate the nature of this selective binding, a cyclic 10-*mer* peptide consisting of the 7-*mer* sequence expressed in the BC loop (BC-Seq 2) was synthesized. BC-Seq 2 was tested for binding towards MV·CB[8] by ITC and showed an affinity comparable to Tn3-BC, as reported in Table 3 and Figure 4e. Residue representation analysis combined with the evaluation of analogies and differences between the three loop sequences and Seq 2, highlighted the presence of a Pro in the sequence. To clarify the importance of this residue towards heteroternary complex formation, an Ala substitution in BC-Seq 2 (BC-Seq 2-P) and in Tn3-BC (Tn3-BC-P) was carried out. As shown in Figure S11a and b, the peptide and the protein were tested for binding towards MV·CB[8] and they both showed a slight decrease in affinity, suggesting the Pro might be an important, but still not fundamental, residue towards complex formation.

Supramolecular interactions between CB[8] and amino acid residues have been widely interrogated in the literature,[12–18] but they have typically been limited to N-terminal aromatic moieties for conjugation. Here, phage display was successfully applied for the first time to identify motifs with high affinity for the MV·CB[8] complex without appending an N-terminal aromatic residue. The data obtained underline the importance of having at least one aromatic residue in the guest sequence, in accordance with previous findings. Sequencing analyses, however, point to Phe being a higher affinity second guest than Trp for the MV·CB[8] complex when the aromatic residue is not located at the N-terminus. The results also suggested that structural constraints play a role in binding as the cyclic sequences performed better than their corresponding linear analogs.

Phage display screening resulted in selection of a three AA epitope (QKF), which was engineered into a protein domain and maintained the same high binding affinity for the MV·CB[8] complex as observed for the free peptide. The specificity of binding towards only one of the modified Tn3 domains implies that the local 3-dimensional structure of the epitope on the protein surface is fundamental towards the recognition properties of the CB[8] complex. Consequently, only proteins exhibiting a preferential conformation of the residues will display high affinity, thus conferring selectivity to

the system. For the first time, this simple supramolecular system was exploited to target protein domains without the restriction and difficulty arising from chemical modifications or aromatic residue insertion at the N-terminus; the appropriate "guest epitope" could be readily expressed as a non-terminal tag only three AA in length. While the effects derived by such a modification on the folding and function of an enzyme must be further investigated, we envision that this approach can be used to easily tag structural domains like the fragment crystallizable (Fc) domains of antibodies thus creating a novel generation of antibody-small molecule conjugates. The CB[8]-mediated conjugation of the small molecule MV to a protein through the recognition of a specifically designed short amino acid motif casts new light on the possible role of host–guest chemistry in protein–small molecule conjugates and paves the way towards the application of CB[8] supramolecular chemistry in biotechnology.

[1] H.-P. Peng, K. H. Lee, J.-W. Jian, A.-S. Yang, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E2656–E2665.

[2] M. W. Peczuh, A. D. Hamilton, *Chem. Rev.* **2000**, *100*, 2479–2494.

[3] H. Yin, A. D. Hamilton, *Angew. Chem. Int. Ed.* **2005**, *44*, 4130–4163; *Angew. Chem.* **2005**, *117*, 4200–4235.

[4] D. Lipovšek, *Protein Eng. Des. Sel.* **2011**, *24*, 3–9.

[5] S. B. Nimse, T. Kim, *Chem. Soc. Rev.* **2013**, *42*, 366–386.

[6] S. V. Kurkov, T. Loftsson, *Int. J. Pharm.* **2013**, *453*, 167–180.

[7] F. Biedermann, U. Rauwald, J. M. Zayed, O. A. Scherman, *Chem. Sci.* **2011**, *2*, 279–286.

[8] D. T. Dang, J. Schill, L. Brunsveld, *Chem. Sci.* **2012**, *3*, 2679–2684.

[9] H. Nguyen, D. Dang, J. van Dongen, L. Brunsveld, *Angew. Chem. Int. Ed.* **2010**, *49*, 895–898; *Angew. Chem.* **2010**, *122*, 907–910.

[10] D. T. Dang, H. D. Nguyen, M. Merkx, L. Brunsveld, *Angew. Chem. Int. Ed.* **2013**, *52*, 2915–2919; *Angew. Chem.* **2013**, *125*, 2987–2991.

[11] K. N. Houk, A. G. Leach, S. P. Kim, X. Zhang, *Angew. Chem. Int. Ed.* **2003**, *42*, 4872–4897; *Angew. Chem.* **2003**, *115*, 5020–5046.

[12] M. E. Bush, N. D. Bouley, A. R. Urbach, *J. Am. Chem. Soc.* **2005**, *127*, 14511–14517.

[13] L. M. Heitmann, A. B. Taylor, P. J. Hart, A. R. Urbach, *J. Am. Chem. Soc.* **2006**, *128*, 12574–12581.

[14] L. A. Logsdon, C. L. Schardon, V. Ramalingam, S. K. Kwee, A. R. Urbach, *J. Am. Chem. Soc.* **2011**, *133*, 17087–17092.

[15] G. Ghale, V. Ramalingam, A. R. Urbach, W. M. Nau, *J. Am. Chem. Soc.* **2011**, *133*, 7528–7535.

[16] L. C. Smith, D. G. Leach, B. E. Blaylock, O. A. Ali, A. R. Urbach, *J. Am. Chem. Soc.* **2015**, *137*, 3663–3669.

[17] P. Rajgariah, A. Urbach, *J. Inclusion Phenom. Macrocyclic Chem.* **2008**, *62*, 251–254.

[18] J. W. Lee, H. H. L. Lee, Y. H. Ko, K. Kim, H. I. Kim, *J. Phys. Chem. B* **2015**, *119*, 4628–4636.

[19] R. J. Gubeli, S. Sonzini, A. Podmore, P. Ravn, O. A. Scherman, C. F. van der Walle, *Chem. Commun.* **2016**, *52*, 4235–4238.

[20] J. W. Lee, S. Samal, N. Selvapalam, H.-J. Kim, K. Kim, *Acc. Chem. Res.* **2003**, *36*, 621–630.

[21] S. J. Barrow, S. Kasera, M. J. Rowland, J. del Barrio, O. A. Scherman, *Chem. Rev.* **2015**, *115*, 12320–12406.

[22] V. D. Uzunova, C. Cullinane, K. Brix, W. M. Nau, A. I. Day, *Org. Biomol. Chem.* **2010**, *8*, 2037–2042.

[23] G. Hettiarachchi, D. Nguyen, J. Wu, D. Lucas, D. Ma, L. Isaacs, V. Briken, *PLoS ONE* **2010**, *5*, e10514.

[24] S. Sonzini, S. T. J. Ryan, O. A. Scherman, *Chem. Commun.* **2013**, *49*, 8779–8781.

[25] M. Ramaekers, S. P. W. Wijnands, J. L. J. van Dongen, L. Brunsveld, P. Y. W. Dankers, *Chem. Commun.* **2015**, *51*, 3147–3150.

[26] J. Scott, G. Smith, *Science* **1990**, *249*, 386–390.

[27] "Cancer Genomics and Proteomics": R. Brissette, N. I. Goldstein, in *Methods in Molecularbiology* (Ed.: P. Fisher), Humana Press, Totowa, NJ, **2007**, pp. 203–213.

[28] "Bacteriophages": K. Hertveldt, T. Beliën, G. Volckaert, in *Methods in Molecular Biology* (Eds.: M. R. Clokie, A. M. Kropinski), Humana Press, Totowa, NJ, **2009**, pp. 321–339.

[29] F. Biedermann, V. D. Uzunova, O. A. Scherman, W. M. Nau, A. D. Simone, *J. Am. Chem. Soc.* **2012**, *134*, 15318–15323.

[30] O. A. Ali, E. M. Olson, A. R. Urbach, *Supramol. Chem.* **2013**, *25*, 863–868.

[31] D. Leahy, W. Hendrickson, I. Aukhil, H. Erickson, *Science* **1992**, *258*, 987–991.

[32] L. Bloom, V. Calabro, *Drug Discovery Today* **2009**, *14*, 949–955.

[33] P. Ravn, C. Madhurantakam, S. Kunze, E. Matthews, C. Priest, S. O'Brien, A. Collinson, M. Papworth, M. Fritsch-Fredin, L. Jermutus, L. Benthem, M. Gruetter, R. H. Jackson, *J. Biol. Chem.* **2013**, *288*, 19760–19772.