Organic & Biomolecular Chemistry

Site-Selective Modification of Proteins Using Cucurbit[7]uril as Supramolecular Protection for N-terminal Aromatic Amino Acids

Journal:	*Organic & Biomolecular Chemistry*
Manuscript ID	OB-COM-05-2020-001004
Article Type:	Communication
Date Submitted by the Author:	15-May-2020
Complete List of Authors:	Smith, Anton; Stanford University, Materials Science & Engineering Maikawa, Caitlin; Stanford University, Department of Bioengineering Roth, Gillie; Stanford University, Department of Bioengineering Appel, Eric; Stanford University, Materials Science & Engineering

ARTICLE

Site-Selective Modification of Proteins Using Cucurbit[7]uril as Supramolecular Protection for *N*-terminal Aromatic Amino Acids

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Anton A. A. Smith,[a,b] Caitlin L. Maikawa,[c] Gillie A. Roth[c] and Eric A. Appel [a,c]

Cucurbit[7,8]urils are known to form inclusion complexes with aromatic amino acids, hosting the hydrohobic side chains within the cavity and adjacent cations within the portal of the macrocyclic host. Here we show that cucurbit[7]uril binding with *N*-terminal phenylalanine significantly reduces the nucleophilicity of the amine, likely due to an increase in stability of the ammonium ion, rendering it unreactive at neutral pH. Using insulin as a model protein, we show that this supramolecular protection strategy can drive selectivity of *N*-terminal amine conjugation away from the preferred B chain *N*-terminal phenylalanine towards the A chain *N*-terminal glycine. Cucurbit[7]uril can therefore be used as a supramolecular protecting group for site-selective protein modification.

Introduction

Polymer conjugation of proteins and peptides has resulted in several PEGylated therapeutic proteins widely used clinically. PEGylation is typically used to enhance protein/peptide stability and/or modify drug pharmacokinetics in a favourable manner. Generally, the conjugation sites on proteins/peptides consist of nucleophilic amines and thiols that can be modified with PEG polymers using activated esters or EDC coupling strategies. [1, 2] Amines on proteins/peptides can be located either on the *N*-terminus of a protein/peptide chain or on the side-chains of lysine residues. Yet, many proteins/peptides contain multiple amines, posing a challenge for site specific conjugations. Current approaches to specific modification of proteins/peptides depend on modulation of the pH of the reaction medium as a measure of control by altering the protonation state (therefore rendering non-nucleophilic) of the various amine residues. The pK_a values for the corresponding acid of many *N*-terminal amines are typically 8.5 or below, while the pK_a values for lysine side chains are in the vicinity of 10.5. Accordingly, couplings favouring *N*-terminal amines can be carried out at a pH where the lysine amines are mostly protonated, rendering them less nucleophilic. [3] Conversely, raising the pH of the reaction medium to the pKa of the corresponding acid of lysine ε-amines increases their relative nucleophilicity.[4]

Insulin presents an interesting challenge for site specific conjugation as it has three amines: (i) an *N*-terminal glycine on the A chain, (ii) an *N*-terminal phenylalanine on the B chain, and (iii) a lysine in the B29 position on the B chain. Site specific conjugation has been performed on the B29 lysine by raising the pH.[5] Consequently, lower pH values should render both *N*-termini amenable to conjugation; however, the B1 phenylalanine of insulin is reported to exhibit preferential reactivity over the A1 glycine for conjugation.[6] This natural selectivity (hypothesized to arise from sterics) has been exploited to install functionality at this position; however, it

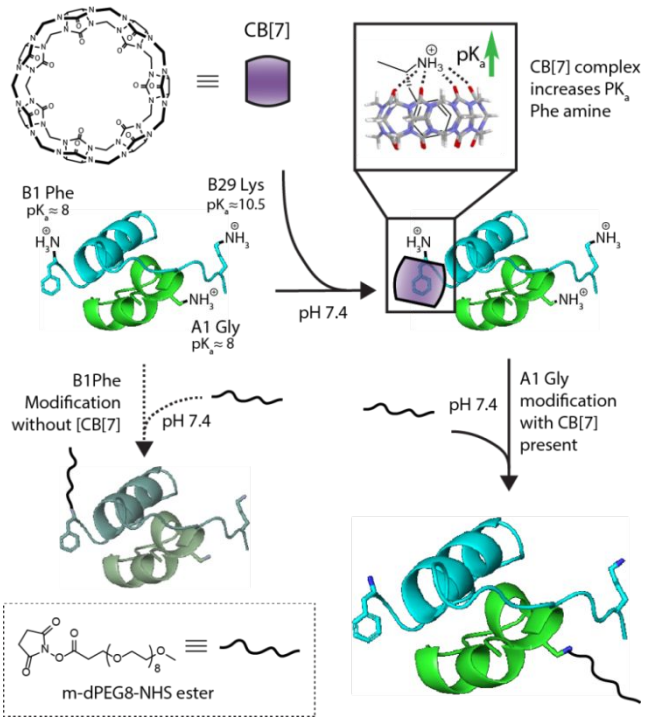


Figure 1. CB[7] binds to the N-terminal phenylalanine of the insulin B chain. This complexation event increases the pKa of the ammonium ion, resulting in more extensive protonation at pH 7.4 and reducing the nucleophilicity of the corresponding amine. This "supramolecular protection" of the B1 Phe enables selective modification of the N-terminal glycine of the insulin A chain.

[a.] *Department of Materials Science & Engineering, Stanford University, Stanford CA 94305, USA*
[b.] *Department of Chemistry, Aarhus University, 8000 Denmark.*
[c.] *Department of Bioengineering, Stanford University, Stanford CA 94305, USA.*
Supplementary information is available…
DOI: 10.1039/x0xx00000x

has also been used to install Fmoc as a protecting group on the B1 *N*-terminus to enable selective functionalization of the A1 *N*-terminus, though with an overall yield of only 10% after Fmoc removal.[6] We hypothesized that inclusion complexes between molecular host molecules and *N*-terminal aromatic amino acids could be exploited as a strategy for supramolecular protection of these amines complimenting traditional protection group chemistries.

Cucurbit[7]uril (CB[7]) demonstrates extreme binding affinities with numerous guests[7-27] and the formation of guest-host complexes with amine-functional guests has been shown to increase the pK_a of protic ammonium salts upon binding in water. [28-32] These behaviours extend to aromatic moieties, especially when the aromatic residue is adjacent to a positive charge such as in *N*-terminal aromatic amino acids. [33-37] These properties have been exploited for numerous applications, such as sequestering drugs from serum,[22] enabling selective protein modification through click chemistry,[38] selective sequestering of organic cations through pH modulation,[39] sequestering proteins with *N*-terminal aromatic amino acids,[37,35,40] enhancing the solubility of hydrophobic drugs,[41,42] and enabling supramolecular PEGylation of proteins. [43, 44] Consequently, we hypothesized that CB[7] complexation could be used to block reactivity of *N*-terminal aromatic amino acids, thus driving selective modification of other amines on the protein. Here we use insulin as a model protein and report the use of CB[7] as an effective non-covalent protection group for the B1 *N*-terminal phenylalanine (Figure 1), thus enabling selective modification of the A1 *N*-terminal glycine in a one-pot reaction step. This work demonstrates the use of CB[7] as a supramolecular protection group for *N*-terminal aromatic amino acids as another tool for site selective modification of proteins and peptides.

Results and discussion

In this work we sought to modify insulin with a short poly(ethylene glycol) (PEG) chain as PEGylation is often used in development of new chemical entities as drug candidates. Insulin was allowed to react with m-dPEG$_8$-NHS ester, with or without CB[7] present, at pH 7.4 in PBS buffer. With 2.2 equivalents of the NHS ester in the absence of CB[7], LC-MS analysis of the resulting modified protein shows a heterogeneous mixture of mono-, di-, and tri-m-dPEG functionalized insulin, as well as unmodified insulin. When 5 equivalents of CB[7] are present in the reaction medium, a reduction of multi-functionalized species, and an increase in mono-functionalized insulin, is observed, suggesting one predominately reactive site (Figure S1).

The endoprotease Glu-C was used to determine sites of modification on the insulin, whereby insulin digested with Glu-C yielded 4 distinct fragments expected from cleaving on the C-terminal of glutamate (Figure S2, Table 1). Modifications of the three amine sites could be observed in fragments I, II, and III across the different experimental conditions. While all modified fragments were observed, the presence of CB[7] was found to reduce the ionic count for the modified B1 containing

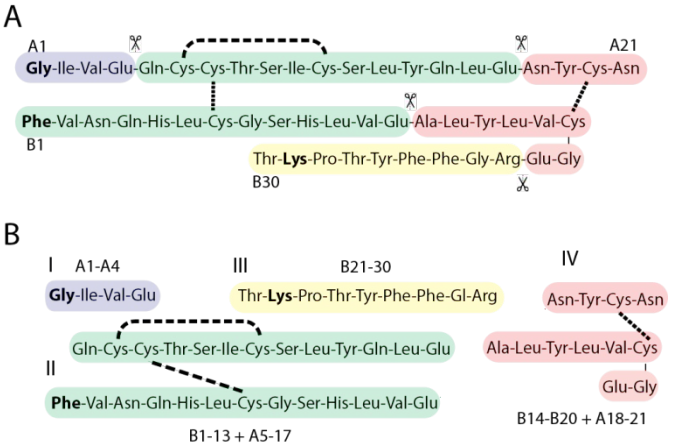


Figure 2. (A) Insulin with Glu-C endoprotease cleaving sites marked. (B) Fragments with possible modifications (bolded amino acids) if treated with m-dPEG$_8$-NHS ester prior to digestion.

fragment. The extent of modification of each fragment was evaluated using fragment IV as an internal standard, whereby ion count intensities could be normalized by the intensity of fragment IV as it exists at a fixed ratio to the other fragments (Figure 3A, Figure S6, and Figure S7). The sum of the intensity ratio of modified and unmodified fragments I and III was approximately equal to the intensity ratio of the corresponding fragments in non-modified insulin (Figure S3), providing support that this analysis is sufficiently quantitative to make comparisons between reaction conditions, despite differences in ionization.

Modification of insulin in the presence of CB[7] revealed a decrease of native fragment I counts, and an increase in the prevalence of the corresponding modified fragment. Yet, both native and modified fragment II decreased with CB[7] present, (Figure S3). We attribute these observations to the complexation of CB[7] to fragment II, which likely affects ionization of native fragment II. Native fragment III decreased with CB[7] present, and the count for modified fragment III increased correspondingly. This trend is the same observed for fragment I. Together these observations suggest a redirection of reactivity of the amine groups on insulin in the presence of CB[7] as the complexation of CB[7] to the B1 Phe on fragment II significantly hinders the B1 terminal amine from reacting as a nucleophile. The experiment was repeated with 50 equivalents of m-dPEG$_8$-NHS to test if the supramolecular protection effect of CB[7] binding to the B1 Phe could be overcome. It was observed that all three amines react with the NHS ester in the absence of CB[7], with each modified fragment being detected (Table 1). When 2.7 equivalents of CB[7] was added, the

Table 1. M/Z of insulin fragments from Glu-c digestion. m/z from fragment II is calculated from its +2H$^+$ ion. Fragments I, III and IV m/z are from their +1H$^+$ ion.

	Native		Modified	
	calc.	found	calc.	Found
I	417.2344	417.2336	811.4547	811.4532
II	2972.3348	2972.3301	3365.5444	3365.5455
III	1116.5837	1116.5820	1511.8113	1511.8084
IV	1377.5814	1377.5803		

ARTICLE

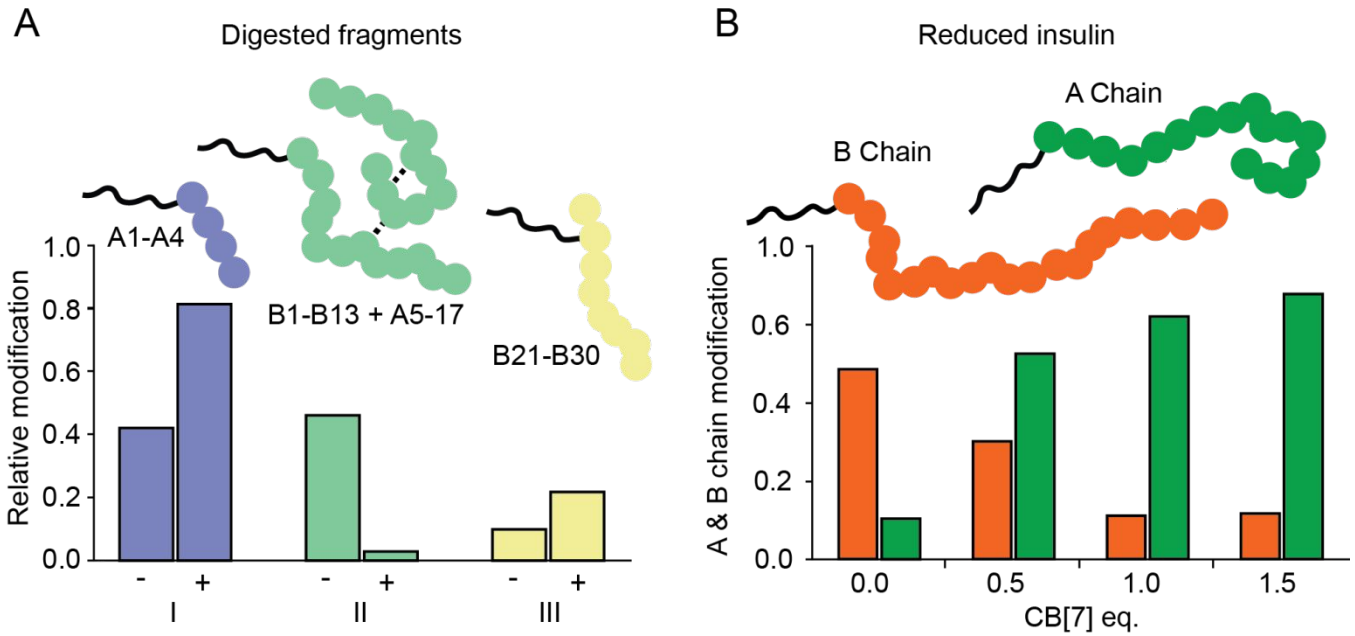


Figure 3:(A) Relative modification of each fragment of insulin with m-dPEG8-NHS (2.2 eq) calculated from the intensity ratio of modified fragments I, II, and III to fragment IV, relative to their respective native fragments ratio to fragment IV, with (+) and without (-) CB[7] present in the reaction. The addition of CB[7] results in the near elimination of modification of fragment II as well as increased modification of fragment I, suggesting blockage of the nucleophilicity of the B1 amine by CB[7] complexation. (B) Approximate chain modification with m-dPEG8-NHS 1.5 eq, with various equivalents of CB[7] added. A close to complete selectivity to the A chain is observed with 1.5 eq of CB[7].

modification of fragment II was greatly reduced (Figure S4). In contrast, fragment I was completely modified with 50 equivalents of m-dPEG NHS ester, whether CB[7] was present or not. These results corroborate those discussed above, whereby CB[7] complexation to the B1 Phe was found to effectively block modification of the amine on the N-terminal B1 Phe.

To validate that CB[7] complexation is capable of blocking the reactivity of the B1 amine, we incubated insulin with 1.5 equivalents of m-dPEG$_8$-NHS ester and various concentrations of CB[7], followed by reduction of the insulin disulphide bonds with DTT. While exclusive functionalization of the B1 Phe was observed in the absence of CB[7], the addition of 1.5 equivalents of CB[7] to insulin was sufficient to severely reduce the modification of the B1 Phe and shift reactivity to the A1 Gly (Figure 3B Figure S5). This experiment revealed almost complete selective to A1 functionalization at 1.5 eq CB[7] and 1.5 eq m-dPEG$_8$-NHS, likely stemming from reduced excess of m-dPEG$_8$-NHS that would react at B27 or B1 when A1 become less abundant as the reaction proceeds. These results corroborate literature reports suggesting that B1 preferentially reacts before A1, and show that this trend can be reversed through the use of CB[7] as a supramolecular protection group on the B1 Phe. Moreover, due to reversible nature of the protection group, the mixed ACN:water solvent for the LC-MS was sufficient to separate the CB[7] from modified and intact insulin, Figure S8.

Altogether, this work shows that CB[7] can be used as a supramolecular protection group to effectively lower the nucleophilicity of *N*-terminal aromatic amino acids at pH 7.4. The observations reported here with insulin, whereby the preferred modification site on insulin can be switched from the B1 Phe to the A1 Gly, can likely be attributed to a combination of an increase in pKa of the corresponding acid of the N-terminal B1 Phe and added steric bulk due to complexation with CB[7]. As such, CB[7] can be used as a non-covalent protection group in aqueous conditions, with applications in site-specific modification of proteins.

Conflicts of interest

There are no conflicts to declare.

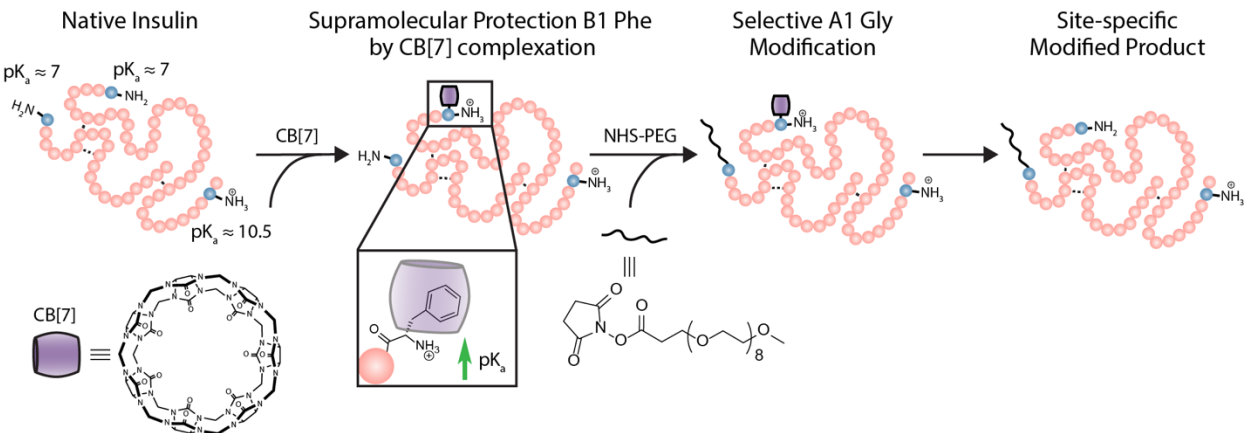
Acknowledgements

This work was performed with the assistance of the Vincent Coates Foundation Mass Spectrometry Laboratory at Stanford University. This work was supported in part by NIDDK R01 (NIH grant #R01DK119254) and NIH P30 CA124435 utilizing the Stanford Cancer Institute Proteomics/Mass Spectrometry

Shared Resource. A.A.A.S. was funded by grant NNF18OC0030896 from the Novo Nordisk Foundation and the Stanford Bio-X Program. C.L.M. was supported by the NSERC Postgraduate Scholarship and the Stanford BioX Bowes Graduate Student Fellowship. G.A.R. is grateful for the NSF Graduate Research Fellowship (DGE-114747)

Notes and references

1. M. J. Roberts, M. D. Bentley and J. M. Harris, *Advanced Drug Delivery Reviews*, 2002, **54**, 459-476.
2. S. Jevševar, M. Kunstelj and V. G. Porekar, *Biotechnology Journal*, 2010, **5**, 113-128.
3. C. B. Rosen and M. B. Francis, *Nature Chemical Biology*, 2017, **13**, 697-705.
4. R. Wetzel, R. Halualani, J. T. Stults and C. Quan, *Bioconjug Chem*, 1990, **1**, 114-122.
5. D. H. Chou, M. J. Webber, B. C. Tang, A. B. Lin, L. S. Thapa, D. Deng, J. V. Truong, A. B. Cortinas, R. Langer and D. G. Anderson, *Proc Natl Acad Sci U S A*, 2015, **112**, 2401-2406.
6. D. Chen, M. M. Disotuar, X. Xiong, Y. Wang and D. H. Chou, *Chem Sci*, 2017, **8**, 2717-2722.
7. S. Choi, R. D. Mukhopadhyay, Y. Kim, I. C. Hwang, W. Hwang, S. K. Ghosh, K. Baek and K. Kim, *Angew Chem Int Ed Engl*, 2019, **58**, 16850-16853.
8. S. Choi, S. H. Park, A. Y. Ziganshina, Y. H. Ko, J. W. Lee and K. Kim, *Chem Commun (Camb)*, 2003, DOI: 10.1039/b306832c, 2176-2177.
9. Y. J. Jeon, S. Y. Kim, Y. H. Ko, S. Sakamoto, K. Yamaguchi and K. Kim, *Org Biomol Chem*, 2005, **3**, 2122-2125.
10. H. J. Kim, J. Heo, W. S. Jeon, E. Lee, J. Kim, S. Sakamoto, K. Yamaguchi and K. Kim, *Angew Chem Int Ed Engl*, 2001, **40**, 1526-1529.
11. H. J. Kim, W. S. Jeon, Y. H. Ko and K. Kim, *Proc Natl Acad Sci U S A*, 2002, **99**, 5007-5011.
12. Y. Kim, H. Kim, Y. H. Ko, N. Selvapalam, M. V. Rekharsky, Y. Inoue and K. Kim, *Chemistry*, 2009, **15**, 6143-6151.
13. S. Lim, H. Kim, N. Selvapalam, K. J. Kim, S. J. Cho, G. Seo and K. Kim, *Angew Chem Int Ed Engl*, 2008, **47**, 3352-3355.
14. K. M. Park, K. Baek, Y. H. Ko, A. Shrinidhi, J. Murray, W. H. Jang, K. H. Kim, J. S. Lee, J. Yoo, S. Kim and K. Kim, *Angew Chem Int Ed Engl*, 2018, **57**, 3132-3136.
15. K. M. Park, M. Y. Hur, S. K. Ghosh, D. R. Boraste, S. Kim and K. Kim, *Chem Commun (Camb)*, 2019, **55**, 10654-10664.
16. F. Biedermann, V. D. Uzunova, O. A. Scherman, W. M. Nau and A. De Simone, *J Am Chem Soc*, 2012, **134**, 15318-15323.
17. H. Lambert, N. Mohan and T. C. Lee, *Phys Chem Chem Phys*, 2019, **21**, 14521-14529.
18. D. Ma, P. Y. Zavalij and L. Isaacs, *J Org Chem*, 2010, **75**, 4786-4795.
19. S. Moghaddam, C. Yang, M. Rekharsky, Y. H. Ko, K. Kim, Y. Inoue and M. K. Gilson, *J Am Chem Soc*, 2011, **133**, 3570-3581.
20. L. Cao, G. Hettiarachchi, V. Briken and L. Isaacs, *Angew Chem Int Ed Engl*, 2013, **52**, 12033-12037.
21. D. Diaz-Gil, F. Haerter, S. Falcinelli, S. Ganapati, G. K. Hettiarachchi, J. C. Simons, B. Zhang, S. D. Grabitz, I. Moreno Duarte, J. F. Cotten, K. Eikermann-Haerter, H. Deng, N. L. Chamberlin, L. Isaacs, V. Briken and M. Eikermann, *Anesthesiology*, 2016, **125**, 333-345.
22. S. Ganapati, S. D. Grabitz, S. Murkli, F. Scheffenbichler, M. I. Rudolph, P. Y. Zavalij, M. Eikermann and L. Isaacs, *Chembiochem*, 2017, **18**, 1583-1588.
23. G. Hettiarachchi, S. K. Samanta, S. Falcinelli, B. Zhang, D. Moncelet, L. Isaacs and V. Briken, *Mol Pharm*, 2016, **13**, 809-818.
24. L. Isaacs, *Chem Commun (Camb)*, 2009, DOI: 10.1039/b814897j, 619-629.
25. L. Isaacs, *Acc Chem Res*, 2014, **47**, 2052-2062.
26. D. Sigwalt, M. Sekutor, L. Cao, P. Y. Zavalij, J. Hostas, H. Ajani, P. Hobza, K. Mlinaric-Majerski, R. Glaser and L. Isaacs, *J Am Chem Soc*, 2017, **139**, 3249-3258.
27. B. Vinciguerra, P. Y. Zavalij and L. Isaacs, *Org Lett*, 2015, **17**, 5068-5071.
28. J. Mohanty, A. C. Bhasikuttan, W. M. Nau and H. Pal, *J Phys Chem B*, 2006, **110**, 5132-5138.
29. M. Shaikh, J. Mohanty, A. C. Bhasikuttan, V. D. Uzunova, W. M. Nau and H. Pal, *Chem Commun (Camb)*, 2008, DOI: 10.1039/b804381g, 3681-3683.
30. M. Shaikh, J. Mohanty, P. K. Singh, W. M. Nau and H. Pal, *Photochem Photobiol Sci*, 2008, **7**, 408-414.
31. S. Zhang, L. Grimm, Z. Miskolczy, L. Biczok, F. Biedermann and W. M. Nau, *Chem Commun (Camb)*, 2019, **55**, 14131-14134.
32. S. S. Thomas and C. Bohne, *Faraday Discuss*, 2015, **185**, 381-398.
33. F. Biedermann, U. Rauwald, M. Cziferszky, K. A. Williams, L. D. Gann, B. Y. Guo, A. R. Urbach, C. W. Bielawski and O. A. Scherman, *Chemistry*, 2010, **16**, 13716-13722.
34. Z. Hirani, H. F. Taylor, E. F. Babcock, A. T. Bockus, C. D. Varnado, Jr., C. W. Bielawski and A. R. Urbach, *J Am Chem Soc*, 2018, **140**, 12263-12269.
35. J. W. Lee, M. H. Shin, W. Mobley, A. R. Urbach and H. I. Kim, *J Am Chem Soc*, 2015, **137**, 15322-15329.
36. W. Li, A. T. Bockus, B. Vinciguerra, L. Isaacs and A. R. Urbach, *Chem Commun (Camb)*, 2016, **52**, 8537-8540.
37. L. C. Smith, D. G. Leach, B. E. Blaylock, O. A. Ali and A. R. Urbach, *J Am Chem Soc*, 2015, **137**, 3663-3669.
38. J. A. Finbloom, K. Han, C. C. Slack, A. L. Furst and M. B. Francis, *J Am Chem Soc*, 2017, **139**, 9691-9697.
39. L. Mikulu, R. Michalicova, V. Iglesias, M. A. Yawer, A. E. Kaifer, P. Lubal and V. Sindelar, *Chemistry*, 2017, **23**, 2350-2355.
40. S. Sonzini, A. Marcozzi, R. J. Gubeli, C. F. van der Walle, P. Ravn, A. Herrmann and O. A. Scherman, *Angew Chem Int Ed Engl*, 2016, **55**, 14000-14004.
41. E. A. Appel, M. J. Rowland, X. J. Loh, R. M. Heywood, C. Watts and O. A. Scherman, *Chem Commun (Camb)*, 2012, **48**, 9843-9845.
42. D. Ma, G. Hettiarachchi, D. Nguyen, B. Zhang, J. B. Wittenberg, P. Y. Zavalij, V. Briken and L. Isaacs, *Nat Chem*, 2012, **4**, 503-510.
43. M. J. Webber, E. A. Appel, B. Vinciguerra, A. B. Cortinas, L. S. Thapa, S. Jhunjhunwala, L. Isaacs, R. Langer and D. G. Anderson, *Proceedings of the National Academy of Sciences*, 2016, **113**, 14189.
44. C. L. Maikawa, A. A. A. Smith, L. Zou, C. M. Meis, J. L. Mann, M. J. Webber and E. A. Appel, *Advanced Therapeutics*, 2020, **3**, 1900094.



Supramolecular protection of N-terminal aromatic amino acids through complexation with Cucurbit[7]uril can enable site-selective protein modification of unfavored motifs.