J|A|C|S
ARTICLES
Published on Web 09/02/2006

Sequence-Specific Recognition and Cooperative Dimerization of N-Terminal Aromatic Peptides in Aqueous Solution by a Synthetic Host

Lisa M. Heitmann,[†] Alexander B. Taylor,[‡,§] P. John Hart,[‡,§,||] and Adam R. Urbach*[,†]

Contribution of the Department of Chemistry, Trinity University, One Trinity Place, San Antonio, Texas 78212, Department of Biochemistry, the X-ray Crystallography Core Laboratory, and the Geriatric Research, Education, and Clinical Center, Department of Veteran's Affairs, South Texas Health Care System, The University of Texas Health Science Center at San Antonio, San Antonio, Texas 78229

Received June 19, 2006; E-mail: aurbach@trinity.edu

Abstract: This article describes the selective recognition and noncovalent dimerization of N-terminal aromatic peptides in aqueous solution by the synthetic host compound, cucurbit[8]uril (Q8). Q8 is known to bind two aromatic guests simultaneously and, in the presence of methyl viologen, to recognize N-terminal tryptophan over internal and C-terminal sequence isomers. Here, the binding of Q8 to aromatic peptides in the absence of methyl viologen was studied by isothermal titration calorimetry (ITC), ^{1}H NMR spectroscopy, and X-ray crystallography. The peptides studied were of sequence X-Gly-Gly, Gly-X-Gly, and Gly-Gly-X (X = Trp, Phe, Tyr, and His). Q8 selectively binds and dimerizes Trp-Gly-Gly (**1**) and Phe-Gly-Gly (**4**) with high affinity (ternary $K = 10^9-10^{11}$ M^{-2}); binding constants for the other 10 peptides were too small to be measured by ITC. Both peptides bound in a stepwise manner, and peptide **4** bound with positive cooperativity. Crystal structures of Q8·**1** and Q8·**4**$_2$ reveal the basis for selective recognition as simultaneous inclusion of the hydrophobic aromatic side chain into the cavity of Q8 and chelation of the proximal N-terminal ammonium group by carbonyl groups of Q8. The peptide sequence selectivity and positively cooperative dimerization reported here are, to the best of our knowledge, unprecedented for synthetic hosts in aqueous solution. Specific peptide recognition and dimerization by synthetic hosts such as Q8 should be important in the study of dimer-mediated biochemical processes and for the separation of peptides and proteins.

Introduction

Synthetic compounds that recognize predetermined peptide sequences in aqueous solution with high affinity and selectivity would be useful in basic biochemical research, in the separation of protein mixtures, and possibly in the diagnosis and treatment of human diseases.[1] Despite significant progress in the design of such compounds,[2,3] and in the development of pattern-based peptide recognition,[4] few synthetic hosts have been shown to bind peptides with binary equilibrium association constants in excess of 10^4 M^{-1} or with sequence selectivity >100-fold in purely aqueous solution; these characteristics would facilitate use at low concentrations, in complex mixtures and in biochemically relevant environments. Moreover, there is significant interest in developing chemical inducers of dimerization (CIDs) for the study of signal transduction and other biological processes.[5] CIDs are low molecular weight compounds that bind simultaneously to two target molecules. A general strategy for designing CIDs targeted to any molecule of interest would allow the study of a broad range of dimer-mediated biochemical processes. Here we show that the synthetic host compound, cucurbit[8]uril (Q8, Figure 1), can act as a selective receptor and CID for N-terminal aromatic peptides.

[†] Trinity University.
[‡] Department of Biochemistry, The University of Texas Health Science Center at San Antonio.
[§] X-ray Crystallography Core Laboratory, The University of Texas Health Science Center at San Antonio.
[||] Geriatric Research, Education, and Clinical Center, The University of Texas Health Science Center at San Antonio.

(1) Peczuh, M. W.; Hamilton, A. D. *Chem. Rev.* **2000**, *100*, 2479−2493.
(2) Bush, M. E.; Bouley, N. D.; Urbach, A. R. *J. Am. Chem. Soc.* **2005**, *127*, 14511−14517.
(3) (a) Shepherd, J.; Gale, T.; Jensen, K. B.; Kilburn, J. D. *Chem.−Eur. J.* **2006**, *12*, 713−720. (b) Tashiro, S.; Tominaga, M.; Kawano, M.; Therrien, B.; Ozeki, T.; Fujita, M. *J. Am. Chem. Soc.* **2005**, *127*, 4546−4547. (c) Schmuck, C.; Heil, M.; Scheiber, J.; Baumann, K. *Angew. Chem., Int. Ed.* **2005**, *44*, 7208−7212. (d) Buschmann, H.-J.; Mutihac, L.; Mutihac, R.-C.; Schollmeyer, E. *Thermochim. Acta* **2005**, *430*, 79−82. (e) Wright, A. T.; Anslyn, E. V. *Org. Lett.* **2004**, *6*, 1341−1344. (f) Ojida, A.; Mito-oka, Y.; Inoue, M.; Hamachi, I. *J. Am. Chem. Soc.* **2002**, *124*, 6256−6258. (g) Sirish, M.; Schneider, H.-J. *Chem. Commun.* **1999**, 907−908. (h) Wagner, H.; Still, W. C.; Chen, C.-T. *Science* **1998**, *279*, 851−853. (i) Hossain, M. A.; Schneider, H.-J. *J. Am. Chem. Soc.* **1998**, *120*, 11208−11209. (j) Breslow, R.; Yang, Z.; Ching, R.; Trojandt, G.; Odobel, F. *J. Am. Chem. Soc.* **1998**, *120*, 3536−3537.
(4) (a) Zhou, H.; Baldini, L.; Hong, J.; Wilson, A. J.; Hamilton, A. D. *J. Am. Chem. Soc.* **2006**, *128*, 2421−2425. (b) Wright, A. T.; Anslyn, E. V.; McDevitt, J. T. *J. Am. Chem. Soc.* **2005**, *127*, 17405−17411. (c) Buryak, A.; Severin, K. *J. Am. Chem. Soc.* **2005**, *127*, 3700−3701.
(5) (a) Kiessling, L. L.; Gestwicki, J. E.; Strong, L. E. *Angew. Chem., Int. Ed.* **2006**, *45*, 2348−2368. (b) Buskirk, A. R.; Liu, D. R. *Chem. Biol.* **2005**, *12*, 151−161. (c) Kley, N. *Chem. Biol.* **2004**, *11*, 599−608. (d) Berg, T. *Angew. Chem., Int. Ed.* **2003**, *42*, 2462−2481. (e) Weiss, A.; Schlessinger, J. *Cell* **1998**, *94*, 277−280. (f) Ho, S. N.; Biggar, S. R.; Spencer, D. M.; Schreiber, S. L.; Crabtree, G. R. *Nature* **1996**, *382*, 822−826. (g) Spencer, D. M.; Wandless, T. J.; Schreiber, S. L.; Crabtree, G. R. *Science* **1993**, *262*, 1019−1024.

10.1021/ja064323s CCC: $33.50

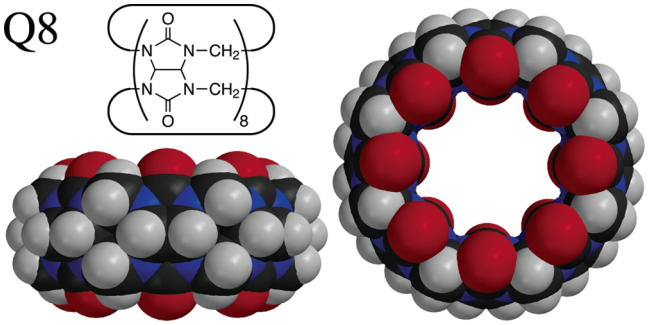

Figure 1. Structure of cucurbit[8]uril (Q8).

Q8 is the eight-membered homologue of the cucurbit[n]uril (Qn) family of glycoluril-based macrocycles.[6,7] The Qn's are water-soluble, donut-shaped compounds with a hydrophobic cavity that is accessible via two constricted, carbonyl-fringed portals. Qn's bind cationic guests in aqueous solution with equilibrium association constants in the range 10^2-10^{12} M^{-1}.[8] Binding is driven by ion−dipole interactions between the guest molecule(s) and the carbonyl groups on the Q8 portals and by the inclusion of hydrophobic groups inside the cavity.

The supramolecular chemistry and potential applications of the Qn family have been greatly advanced by the discovery of homologues containing different numbers of glycoluril units (n = 5, 7, 8, and 10)[7,9] and by the development of methods for their synthesis[10,11] and functionalization[12] and for making derivatives[13] and analogues.[14] While Q6 binds well to small alkylamines,[8] the larger cavity of Q7 accommodates larger guests,[15] and importantly, Q8 and Q10 can bind multiple guests simultaneously.[2,9,16] For example, Q8 binds to 1 (and only 1) equiv of dicationic methyl viologen (MV), and the Q8·MV complex can then bind to 1 equiv of a second aromatic guest, such as 2,6-dihydroxynaphthalene,[16a] dopamine,[7] and the aromatic amino acids tryptophan, phenylalanine, and tyrosine.[2,7]

1	***Trp***-Gly-Gly	**7**	***Tyr***-Gly-Gly
2	Gly-***Trp***-Gly	**8**	Gly-***Tyr***-Gly
3	Gly-Gly-***Trp***	**9**	Gly-Gly-***Tyr***
4	***Phe***-Gly-Gly	**10**	***His***-Gly-Gly
5	Gly-***Phe***-Gly	**11**	Gly-***His***-Gly
6	Gly-Gly-***Phe***	**12**	Gly-Gly-***His***

Figure 2. The twelve unprotected aromatic peptides studied.

We studied the effects of electrostatic charge on the binding of Q8·MV to tryptophan and found that the positively charged ammonium group was important for binding. This led to the discovery that Q8·MV can discriminate N-terminal from C-terminal and internal Trp residues on the basis of electrostatic charge.[2] From these results, we hypothesized that Q8 should bind 2 equiv of a peptide containing N-terminal Trp *in the absence of MV*. In other words, Q8 should not only recognize N-terminal aromatic peptides but also noncovalently dimerize them.

Here, we examine the potential for Q8 to recognize and dimerize aromatic peptides. The peptides studied were of sequence X-Gly-Gly, Gly-X-Gly, and Gly-Gly-X (X = Trp, Phe, Tyr, and His) (Figure 2). Using isothermal titration calorimetry, we found that Q8 binds to Trp-Gly-Gly (**1**) and Phe-Gly-Gly (**4**) in a 2:1 (peptide:Q8) stoichiometry and with *ternary* equilibrium association constants (K_{ter}) in the range 10^9-10^{11} M^{-2}. None of the other 10 compounds showed measurable binding affinities. This result, to the best of our knowledge, demonstrates unprecedented peptide sequence selectivity by a synthetic host in an aqueous solution. Binding studies and crystal structures are presented to explain the mechanism of binding and sequence selection.

Results and Discussion

Sequence Selectivity. We had previously demonstrated that the Q8·MV complex could discriminate between sequence isomers **1**, **2**, and **3**.[2] It was therefore important for the study at hand to consider not only the type but also the placement of aromatic residues.[17] The series of tripeptides shown in Figure 2 places each of the four aromatic amino acids at N-terminal, internal, and C-terminal positions of unprotected tripeptides, with Gly residues at the other two positions in each peptide.[18]

The binding of Q8 to peptides **1**−**12** was followed by isothermal titration calorimetry (ITC) (27 °C, 10 mM sodium phosphate, pH 7.0). Surprisingly, binding constants were measurable only for Trp-Gly-Gly (**1**) and Phe-Gly-Gly (**4**) (Figure 3), indicating that equilibrium association constants for the other 10 peptides were $<10^2$ M^{-1} for a single binding event (see Supporting Information). The stoichiometry of binding for peptides **1** and **4** was 2:1 (peptide:Q8). This finding, in itself,

(6) Lagona, J.; Mukhopadhyay, P.; Chakrabarti, S.; Isaacs, L. *Angew. Chem., Int. Ed.* **2005**, *44*, 4844−4870.
(7) Lee, J. W.; Samal, S.; Selvapalam, N.; Kim, H.; Kim, K. *Acc. Chem. Res.* **2003**, *36*, 621−630.
(8) (a) Liu, S.; Ruspic, C.; Mukhopadhyay, P.; Chakrabarti, S.; Zavalij, P. Y.; Isaacs, L. *J. Am. Chem. Soc.* **2005**, *127*, 15959−15967. (b) Jeon, W. S.; Moon, K.; Park, S. H.; Chun, H.; Ko, Y. H.; Lee, J. Y.; Lee, E. S.; Samal, S.; Selvapalam, N.; Rekharsky, M. V.; Sindelar, V.; Sobransingh, D.; Inoue, Y.; Kaifer, A. E.; Kim, K. *J. Am. Chem. Soc.* **2005**, *127*, 12984−12989. (c) Buschmann, H.-J.; Cleve, E.; Schollmeyer, E. *Inorg. Chim. Acta* **1992**, *193*, 93−97. (d) Mock, W. L.; Shih, N.-Y. *J. Am. Chem. Soc.* **1988**, *110*, 4706−4710. (e) Mock, W. L.; Shih, N.-Y. *J. Org. Chem.* **1986**, *51*, 4440−4446. (f) Mock, W. L.; Shih, N.-Y. *J. Org. Chem.* **1983**, *48*, 3618−3619.
(9) Kim, J.; Jung, I.-S.; Kim, S.-Y.; Lee, E.; Kang, J.-K.; Sakamoto, S.; Yamaguchi, K.; Kim, K. *J. Am. Chem. Soc.* **2000**, *122*, 540−541.
(10) Day, A. I.; Arnold, A. P.; Blanch, R. J.; Snushall, B. *J. Org. Chem.* **2001**, *66*, 8094−8100.
(11) Liu, S.; Zavalij, P. Y.; Isaacs, L. *J. Am. Chem. Soc.* **2005**, *127*, 16798−16799.
(12) (a) Jon, S. Y.; Selvapalam, N.; Oh, D. H.; Kang, J.-K.; Kim, S.-Y.; Jeon, Y. J.; Lee, J. W.; Kim, K. *J. Am. Chem. Soc.* **2003**, *125*, 10186−10187. (b) Lee, H.-K.; Park, K. M.; Jeon, Y. J.; Kim, D.; Oh, D. H.; Kim, H. S.; Park, C. K.; Kim, K. *J. Am. Chem. Soc.* **2005**, *127*, 5006−5007.
(13) (a) Isobe, H.; Sato, S.; Nakamura, E. *Org. Lett.* **2002**, *4*, 1287−1289. (b) Day, A.; Arnold, A. P.; Blanch, R. J. *Molecules* **2003**, *8*, 74−84. (c) Zhao, Y.; Xue, S.; Zhu, Q.; Tao, Z.; Zhang, J.; Wei, Z.; Long, L.; Hu, M.; Xiao, H.; Day, A. *Chin. Sci. Bull.* **2004**, *49*, 1111−1116.
(14) (a) Isaacs, L.; Park, S.-K.; Liu, S.; Ko, Y. H.; Selvapalam, N.; Kim, Y.; Kim, H.; Zavalij, P. Y.; Kim, G.-H.; Lee, H.-S.; Kim, K. *J. Am. Chem. Soc.* **2005**, *127*, 18000−18001. (b) Lagona, J.; Fettinger, J. C.; Isaacs, L. *J. Org. Chem.* **2005**, *70*, 10381−10392. (c) Lagona, J.; Wagner, B. D.; Isaacs, L. *J. Org. Chem.* **2006**, *71*, 1181−1190.
(15) (a) Blanch, R. J.; Sleeman, A. J.; White, J. T.; Arnold, A. P.; Day, A. I. *Nano Lett.* **2002**, *2*, 147−149. (b) Kim, H.-J.; Jeon, W. S.; Ko, Y. H.; Kim, K. *PNAS* **2002**, *99*, 5007−5011. (c) Ong, W.; Gomez−Kaifer, M.; Kaifer, A. E. *Org. Lett.* **2002**, *4*, 1791−1794. (d) Sindelar, V.; Moon, K.; Kaifer, A. E. *Org. Lett.* **2004**, *6*, 2665−2668.
(16) (a) Kim, H.-J.; Heo, J.; Jeon, W. S.; Lee, E.; Kim, J.; Sakamoto, S.; Yamaguchi, K.; Kim, K. *Angew. Chem., Int. Ed.* **2001**, *40*, 1526−1529. (b) Jon, S. Y.; Ko, Y. H.; Park, S. H.; Kim, H.-J.; Kim, K. *Chem. Commun.* **2001**, 1938−1939. (c) Jeon, W. S.; Kim, H.; Lee, C.; Kim, K. *Chem. Commun.* **2002**, 1828−1829. (d) Ziganshina, A.; Ko, Y. H.; Jeon, W. S.; Kim, K. *Chem. Commun.* **2004**, 806−807. (e) Sindelar, V.; Cejas, M. A.; Raymo, F. M.; Chen, W.; Parker, S.; Kaifer, A. E. *Chem.−Eur. J.* **2005**, *11*, 7054−7059.
(17) We focused on the aromatic amino acids for the likelihood that their binding properties would be similar in nature to prior work in this area and, therefore, more straightforward to characterize.
(18) We studied Gly-Gly-Gly as a control for the possible binding of Gly residues to Q8 and, as expected, found no binding to Q8 by ITC (see Supporting Information).

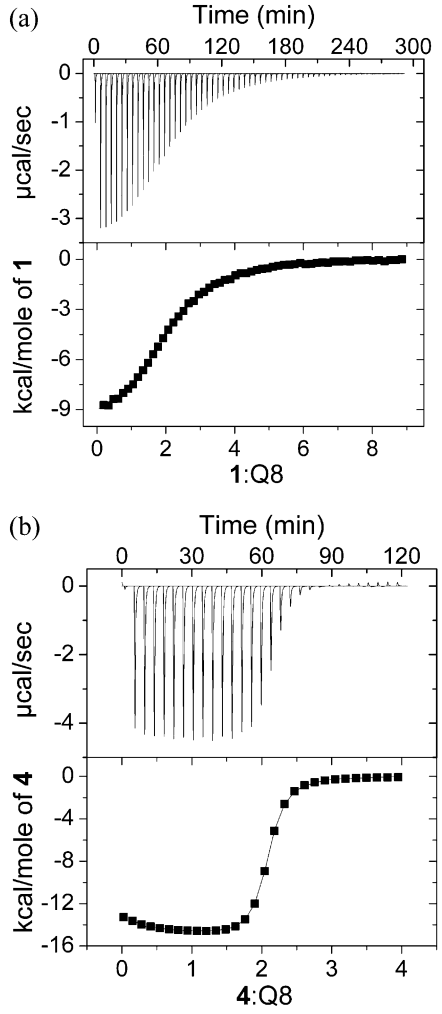

Figure 3. ITC data for (a) peptide **1** and (b) peptide **4** at 27 °C in 10 mM sodium phosphate (pH 7.0). Peptide **1** was titrated at 1 mM into a 0.05 mM solution of Q8. Peptide **4** was titrated at 2 mM into a 0.1 mM solution of Q8. At top is the raw data for power versus time. At bottom are integrated enthalpy values versus the molar ratio of peptide:Q8. These data were fit to the "sequential binding model" using Origin 7.0 software.

shows that Q8 is highly selective not only for Trp- and Phe-containing peptides but also for the placement of these residues at the N-terminal position.

These results are consistent with prior work on the recognition of N-terminal Trp by the Q8·MV complex,[2] where recognition was dependent on the N-terminal positive charge. In that study Q8·MV bound to Gly-Trp-Gly (**2**) and Gly-Gly-Trp (**3**) with measurable, albeit low, affinities. In the current study, however, we are unable to measure the binding of Q8 to compounds **2** or **3**. This result demonstrates that Q8 is substantially more sequence selective in the absence of MV.

Binding Mode. Here we consider the mechanism of binding of Q8 with compounds **1** and **4.** The ITC plots in Figure 3 are consistent with a complex stoichiometry of 2:1 (peptide:Q8). Both 1:1 and 2:1 complexes were observed for peptides **1** and **4** by electrospray mass spectrometry (see Supporting Information). The crystal structure of Q8·$\mathbf{4}_2$, discussed in the next section, shows a 2:1 complex in which the two phenyl groups interact extensively inside the binding cavity of Q8. Although we do not have a crystal structure of Q8·$\mathbf{1}_2$, it is reasonable to predict from the structure of Q8·$\mathbf{4}_2$, from modeling, and from prior work on Q8 complexes[2,9,16] that two indole groups should fit inside the cavity of Q8, stacked face-to-face. To test for dimerization of peptides in the absence of Q8, we titrated compounds **1** and **4** into 10 mM phosphate buffer (pH 7.0) by ITC over the range 1−1000 μM and by ^{1}H NMR over the range 0.1−10 mM and found no measurable binding; this result indicates that the peptides do not dimerize when free in solution under the conditions of our binding experiments (data not shown). These observations and arguments, as a whole, support a model in which (i) both peptides **1** and **4** can bind as 1:1 and 2:1 (peptide:Q8) complexes and (ii) the presence of the first bound peptide influences the binding of the second. It is therefore most appropriate to treat the model of binding as stepwise:

$$\text{Q8} + \text{peptide} \rightleftharpoons \text{Q8}\cdot\text{peptide} \quad (1)$$

$$\text{Q8}\cdot\text{peptide} + \text{peptide} \rightleftharpoons \text{Q8}\cdot\text{peptide}_2 \quad (2)$$

In a stepwise binding model, it is important to consider the nature of cooperativity; that is, whether the second stepwise equilibrium association constant (K_{s2}, eq 2) is greater than (positively cooperative), less than (negatively cooperative), or similar to (noncooperative) the first stepwise equilibrium association constant (K_{s1}, eq 1). We obtained ^{1}H NMR spectra of compounds **1** and **4** in 1:1, 2:1, and 3:1 (peptide:Q8) stoichiometric mixtures (D_2O, 25 °C). Compound **1** showed significant exchange broadening, which makes the spectra inconclusive with regard to identifying the species present, their amounts, and their possible interactions (see Supporting Information).

For compound **4** (Figure 4), a common set of peaks is observed at the same chemical shift values in the 1:1, 2:1, and 3:1 stoichiometric mixtures. Beyond a stoichiometric ratio of 2:1, a second set of peaks, which corresponds to the free guest, is observed. Below a stoichiometric ratio of 2:1, a second set of peaks, which does not correspond to free guest, is observed. These observations, most notably that the chemical shifts of the common set of peaks remain constant, strongly support the assignment of these peaks as the 2:1 complex. The new set of peaks in the 1:1 mixture is therefore likely the 1:1 complex. Based on these assignments, the 2:1 complex is the only observable species in the 2:1 stoichiometric mixture and is a significant species in the 1:1 mixture. At the concentrations used in these NMR experiments, the ternary equilibrium constant obtained from ITC experiments (1.5×10^{11} M^{-2}, discussed below) was used to calculate the possible distributions of 1:1 and 2:1 complexes in the 1:1 and 2:1 stoichiometric mixtures over a range of stepwise association constants. This simple analysis reveals that a significant quantity (>5%) of 2:1 complex can only be present (as observed) in the 1:1 stoichiometric mixture if binding is positively cooperative.

The ITC data for compounds **1** and **4** were fit to the "sequential binding" model using Origin 7.0 software, and it was observed that this model provides an excellent fit to the negatively cooperative or noncooperative binding isotherm of peptide **1** but a less satisfactory fit to the positively cooperative binding isotherm[19] of peptide **4** (as expected[20]). For the formation of Q8·$\mathbf{1}_2$, we were able to determine stepwise bimolecular association constants, K_{s1} ($(1.3 \pm 0.2) \times 10^5$ M^{-1}) and K_{s2} ($(2.8 \pm 0.3) \times 10^4$ M^{-1}), as well as the enthalpy and entropy of binding for each step (Table 1). It is important to

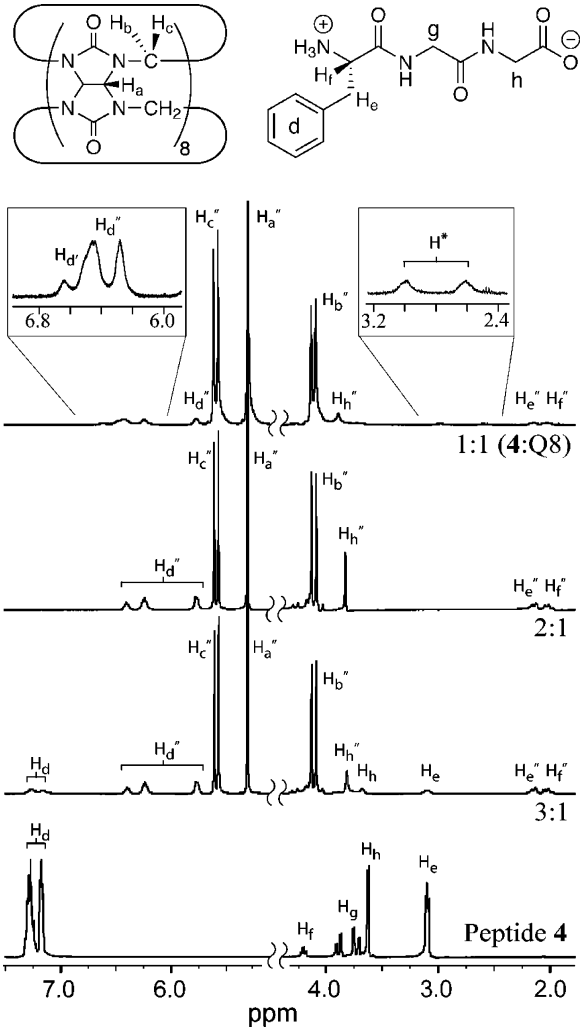

Figure 4. ^{1}H NMR titration of peptide **4** with Q8 in deuterium oxide at 25 °C. In all mixtures, the concentration of Q8 is 1 mM. Peptide:Q8 ratios are listed at right. The spectrum of **4** is at the bottom for reference. In the proton labeling scheme, a single prime indicates the 1:1 complex, a double prime indicates the 2:1 complex, and no prime indicates free peptide. H_e'' and H_g'' were assigned based on the crystal structure of Q8·$\mathbf{4}_2$. H* is assigned generally to the 1:1 complex.

note that, for a 2:1 complex, a stepwise binding model would predict $K_{s1} = 4K_{s2}$ if the intrinsic binding constants at the two sites are identical.[21] That we observe this relationship between K_{s1} and K_{s2} precludes any distinction between negatively cooperative and noncooperative binding on the basis of this data alone, and therefore the binding data in Table 1 should be considered with this caveat in mind. Regardless of the model chosen, however, binding is enthalpically driven and entropically unfavorable, as observed for peptides binding to the Q8·MV complex.[2]

The determination of stepwise thermodynamic constants for the formation of Q8·$\mathbf{4}_2$ is more problematic because there is

(19) The isotherms observed for Q8·$\mathbf{4}_2$ are described as "positively cooperative" due to their shape. The binding experiments show that saturation occurs at a 2:1 peptide:Q8 ratio. The isotherms show a single prominent sigmoidal phase at approximately this ratio. This characteristic indicates that the second stepwise equilibrium association constant is either greater than or equal to the first. In addition, all experiments show a characteristic, shallow "smile" early in the titration (at left), indicating that the first binding transition is less steep (and thus the first K value is smaller) than the second.

(20) Tochtrop, G. P.; Richter, K.; Tang, C.; Toner, J. J.; Covey, D. F.; Cistola, D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1847−1852.

(21) Connors, K. A. *Binding Constants. The Measure of Molecular Complex Stability*; John Wiley & Sons, Inc.: New York, 1987.

Table 1. Stepwise Thermodynamic Binding Data for Peptide **1** in Complex with Q8

	K_s[a] (M^{-1})	ΔG_s[b] (kcal/mol)	ΔH_s[c] (kcal/mol)	$-T\Delta S_s$[d] (kcal/mol)
step 1	$(1.3 \pm 0.2) \times 10^5$	−7.0 (±0.2)	−10.5 (±0.3)	3.4 (±0.4)
step 2	$(2.8 \pm 0.3) \times 10^4$	−6.1 (±0.1)	−12.3 (±0.2)	6.2 (±0.3)

[a] Mean values measured from at least three ITC experiments at 27 °C in 10 mM sodium phosphate, pH 7.0. Standard deviations are given in parentheses. [b] Gibbs free energy values calculated from K_s values. Standard deviations for ΔG_s values were calculated as the relative error observed in K_s, due to their relationship by a natural logarithm. [c] Enthalpy values measured by ITC. [d] Entropic contributions to ΔG_s calculated from K_s and ΔH_s values.

Table 2. Thermodynamic Binding Data for Ternary Complexes of **1** and **4** with Q8

complex	K_{ter}[a] (M^{-2})	ΔG_{ter}[b] (kcal/mol)	ΔH_{ter}[c] (kcal/mol)	$-T\Delta S_{ter}$[d] (kcal/mol)
Q8·$\mathbf{1}_2$	$(3.6 \pm 0.9) \times 10^9$	−13.1 (±0.2)	−22.8 (±0.5)	9.7 (±0.6)
Q8·$\mathbf{4}_2$	$(1.5 \pm 0.2) \times 10^{11}$	−15.4 (±0.1)	−29.6 (±0.2)	14.2 (±0.3)

[a] Mean values measured from at least three ITC experiments at 27 °C in 10 mM sodium phosphate, pH 7.0. Standard deviations are given in parentheses. [b] Gibbs free energy values calculated from K_{ter} values. Standard deviations for ΔG_{ter} values were calculated as the relative error observed in K_{ter}, due to their relationship by a natural logarithm. [c] Enthalpy values measured by ITC. [d] Entropic contributions to ΔG_{ter} calculated from K_{ter} and ΔH_{ter} values.

insufficient data to define the first binding transition, and thus a wider range of values for the two binding constants can produce similar quality fits. This problem was described in detail by Tochtrop, Cistola, and co-workers,[20] who performed a Bayesian analysis of positively cooperative ITC data sets and found that a range of stepwise constants satisfied the stepwise binding model. An interesting and perhaps overlooked result of their analysis is that while the values of the pairs of stepwise binding constants in each fit varied significantly, the product of each pair of constants, i.e., the overall ternary equilibrium association constant, K_{ter} (eq 3), hardly varied.

$$\text{Q8} + 2\ \text{peptide} \rightleftharpoons \text{Q8}\cdot\text{peptide}_2 \qquad (3)$$

Therefore, we report the *ternary* equilibrium association constant, K_{ter} ($(1.5 \pm 0.2) \times 10^{11}\ M^{-2}$) for the formation of Q8·$\mathbf{4}_2$ (Table 2) as derived from the product of the (less accurate) stepwise binding constants obtained from Origin 7.0. To further probe the validity of this approach, we also derived the value of K_{ter} from the ITC data by treating the cumulative heat at each step of the titration as a fraction of the total enthalpy of formation of the ternary complex and assuming that this fractional enthalpy is equivalent to the mole fraction of ternary complex formed (see Supporting Information); this assumption is acceptable in a positively cooperative system under conditions in which the experimental concentrations are in excess of K_{s1}^{-1} and K_{s2}^{-1}. The resulting value of K_{ter} ($(1.5 \pm 0.2) \times 10^{11}\ M^{-2}$) was equivalent to that obtained using the stepwise model with Origin 7.0 software.

We did not expect Q8 to be selective for Phe- over Trp-based peptides. Trp has a larger hydrophobic surface area than Phe, and prior work from our group has shown that tryptophan binds with higher affinity than phenylalanine to the Q8·MV complex.[2] Here we observe that K_{ter} for the formation of Q8·$\mathbf{4}_2$ ($1.5 \times 10^{11}\ M^{-2}$) is greater than that for Q8·$\mathbf{1}_2$ ($3.6 \times 10^9\ M^{-2}$) by ∼40-fold. Both ternary complexes form in a process that is

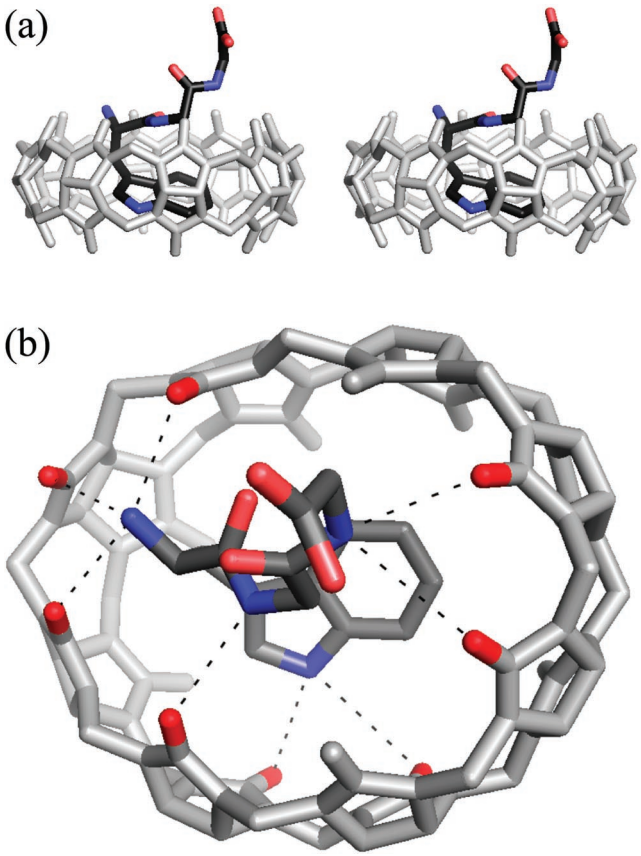

Figure 5. Crystal structure of Q8·**1**; hydrogens and solvating water were removed for clarity. (a) Cross-eyed stereoview; (b) the dashed lines show a total of six out of eight available carbonyl groups (red) on one portal and two on the other portal, engaged in binding the peptide via ion−dipole and dipole−dipole interactions. The N-terminal ammonium group is chelated by three carbonyls.

enthalpically driven and entropically unfavorable. The formation of Q8·$\mathbf{4}_2$ is more enthalpically favorable than Q8·$\mathbf{1}_2$ by ∼7 kcal/mol yet more entropically unfavorable by ∼4.5 kcal/mol. Based on the nature of cooperativity in this system, this enthalpy/entropy compensation is likely due to a greater degree of electrostatic attraction between the two ligands in the binding cavity of Q8 (more exothermic), at the entropic cost of reduced mobility.[22] Further interpretation of these results is facilitated by high-resolution structural analysis, as described below.

X-ray Crystallography. To further probe the nature of binding and selectivity in this system, we obtained crystal structures of Q8·**1** and Q8·$\mathbf{4}_2$ at 0.92 Å and 0.95 Å resolution, respectively (see Supporting Information). We were particularly interested in understanding the nature of the exquisite selectivity observed in this system, not only for Phe- and Trp-containing peptides but also for their location at the N-terminal position. It was also hoped that the structures would shed light on the nature of cooperativity (positive and negative) observed for the complexes of **4** and **1**, respectively.

Structure of Q8·1. The crystal structure of the 1:1 Q8·**1** complex is shown in Figure 5. The indole side chain is bound entirely within the Q8 cavity, with the long axis of the indole parallel to the equatorial cross section of Q8 and the short axis of the indole tilted at ∼40° from this cross section. The Q8 molecule is distorted from an "ideal" D_{8h} symmetry to an elliptical shape with a major axis that is ∼16% longer than the minor axis. This distortion sterically accommodates the sideways inclusion of the indole group. The backbone of the Trp residue and the central Gly residue folds onto the indole group, almost straight across the Q8 cavity. This folding is likely induced to maximize intramolecular van der Waals interactions in the peptide and to fill a portion of the Q8 cavity, while excluding solvent.

There are several key intermolecular interactions that shed light on the binding data (Figure 5b). Foremost among these are the ion−dipole interactions between the N-terminal nitrogen of **1** and the three proximal carbonyl oxygens of Q8 (N−O distances in the range 2.8−3.1 Å). The chelating of three carbonyl groups to one ammonium group likely provides significant driving force for binding. We believe that these ion−dipole interactions are of primary importance in the sequence selectivity of Q8 for N-terminal peptide **1** over internal and C-terminal peptides **2** and **3**. Prior work[2] has shown that the sequence selectivity of the Q8·MV complex for **1** over **2** and **3** was due to the presence of a positive charge proximal to the indole group. This structure supports that hypothesis. Moreover, the hydrogens projecting from the N-terminal nitrogen are well positioned to form hydrogen bonds to the carbonyl oxygens, which would further stabilize the complex.

Several additional interactions were observed (Figure 5b): (1) bifurcated dipole−dipole interactions between the pyrollic NH of the indole and two proximal carbonyl oxygens of Q8 (N−O distances 3.5−3.7 Å); (2) a hydrogen bond[23] between the amide NH group of the central Gly residue and the proximal carbonyl oxygen of Q8 (N−O distance 3.1 Å, N−H−O angle 163°); and (3) bifurcated dipole−dipole interactions between the amide NH of the C-terminal Gly residue and the proximal carbonyl oxygens of Q8 (N−O distances 3.2−3.5 Å). These amide NH−Q8 carbonyl interactions occupy six out of eight available carbonyls on one portal of Q8; this observation exemplifies why Q8 is such an excellent host for binding peptides: the peptide tail can adopt a conformation that places each NH group in register with an accepting carbonyl of Q8. These interactions also help to explain the previously observed increase in affinity of Q8·MV for Trp-Gly-Gly as compared to monomeric tryptophan derivatives.[2]

An interesting feature of the structure of Q8·**1** is the relatively empty space in the cavity of Q8 distal to the peptide chain, due to the presence of only one bound aromatic group. In the crystal, this space is filled by the inclusion of a portion of Q8 from an adjacent Q8·**1** complex, thus excluding solvent (Figure 6a). In this intercomplex inclusion interaction, the indole ring of one complex is stacked on top of a Q8 urea group of another complex (average distance 3.5 Å; dihedral angle between indole plane and urea plane ∼26°). Although this is not a traditional pi-pi stacking interaction, the local dipole moments of the indole group and the proximal urea group are aligned for favorable electrostatic attraction. This interaction is sufficiently strong to generate helical arrays in the crystal that propagate with a four-fold screw helical axis of symmetry. Each helical array binds to its neighbor via dipole−dipole interactions between the Q8

(22) Williams, D. H.; Stephens, E.; O'Brien, D.; Zhou, M. *Angew. Chem., Int. Ed.* **2004**, *43*, 6596−6616.

(23) The distinguishing factor between this hydrogen bond and the other close (≤3.1 Å) dipole−dipole interactions described is the N−H−O angle of 163°, which is significantly above a cutoff of 140°.

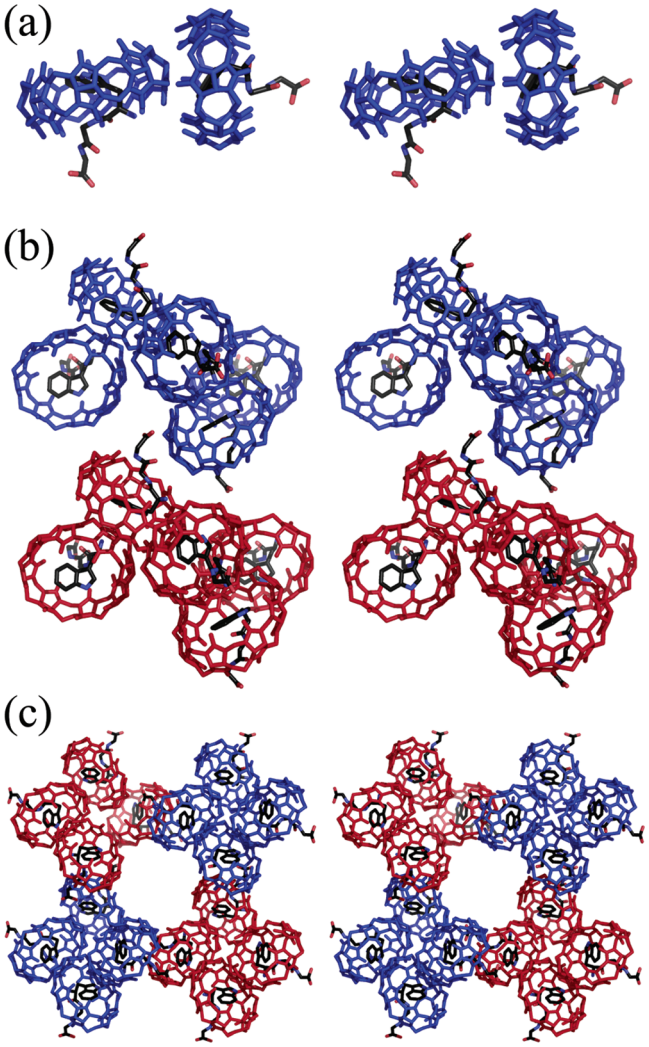

Figure 6. Cross-eyed stereoviews of the crystal packing interactions of Q8·**1**. (a) Interface between two Q8·**1** complexes. (b) Side view of two helical arrays showing the interface between the four-fold helices. (c) Axial view showing the solvent channel formed by assembly of the helical arrays.

and central Gly carbonyl groups of one Q8·**1** complex and the positive electrostatic potential on the equator of Q8 on an adjacent helical array (Figure 6b). The assembly of helical arrays creates large solvent channels (~1.5 nm diameter) in the crystal (Figure 6c).

Structure of Q8·4$_2$. The crystal structure of Q8·**4**$_2$ is shown in Figure 7. The structure, which is approximately C_2-symmetric, shows the two peptides bound through opposite portals of the Q8, with their phenyl groups engaged in a slipped pi−pi stacking interaction in the center of the binding cavity and making extensive van der Waals contacts. The phenyl rings are separated at an average distance of 3.5 Å with a lateral offset of ~1 Å, and the C1−C4 axes are offset by an angle of ~140°. This ligand−ligand interaction is presumably the cause for positive cooperativity in this complex. Each phenyl ring is also in close contact with two urea groups of Q8 (average distance ~4 Å). The Q8 ring is distorted to a small extent from "circular", although to a lesser extent than the case for Q8·**1**. The Gly-Gly tails extend out of the cavity and coil around each portal of Q8. No ordered water is observed inside the cavity of Q8.

There are several important host−guest interactions (Figure 7b). As in the Q8·**1** complex, there are ion−dipole interactions

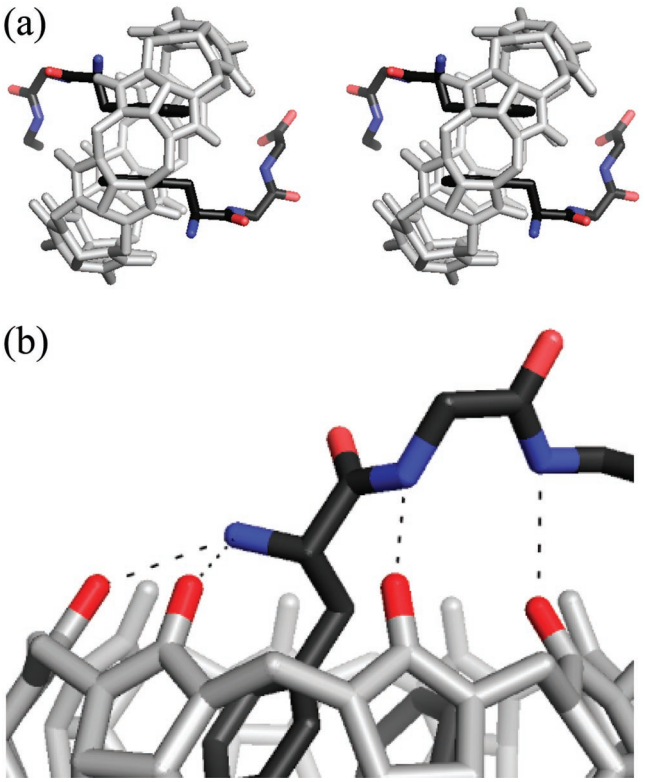

Figure 7. Crystal structure of Q8·**4**$_2$; hydrogens and solvating water were removed for clarity. (a) Cross-eyed stereoview; (b) the dashed lines indicate key electrostatic interactions.

between the N-terminal ammonium group of each peptide and the proximal carbonyl oxygens of Q8, but in this case, only two carbonyls chelate each ammonium (N−O distances 2.9−3.0 Å). In contrast to the Q8·**1** structure, the N-terminal nitrogens in the structure of Q8·**4**$_2$ are not oriented in a manner that facilitates hydrogen bonding.

In this structure there are also dipole−dipole interactions between each amide NH group and the proximal carbonyl oxygen of Q8 (distances 2.7−2.9 Å). As in the structure of Q8·**1**, these interactions demonstrate good alignment in register between amide NH groups and Q8 carbonyls.

The lack of available space in the Q8 cavity of Q8·**4**$_2$ does not allow for the type of intercomplex and, thus, crystal packing interactions observed for Q8·**1**. Consequently, the packing of complexes in the Q8·**4**$_2$ crystal structure is fundamentally different. The positive electrostatic potential on the equatorial rim of each Q8 interacts with the Q8 carbonyl groups on an adjacent complex, forming a serrated packing pattern (see Supporting Information).

Recognition and Cooperativity. The crystal structures show the expected result that binding is driven by the inclusion of hydrophobic aromatic side chains into the cavity of Q8 and by ion−dipole interactions between the guest and the host, similar to many cucurbit[*n*]uril−guest complexes.[6] What is interesting here is that all 12 peptides studied have an aromatic side chain and an N-terminal ammonium group, but Q8 binds only to **1** and **4**. We explain this result in terms of the hydrophobicity of the side chain and its proximity to the N-terminal ammonium group. Trp and Phe are more hydrophobic than the other aromatic residues. We observed previously that Q8·MV prefers to bind the amino acids with a trend in binding affinity of Trp

> Phe > Tyr > His, and because Tyr is a better electron donor than Phe (and should therefore bind with greater stability in a charge-transfer complex with methyl viologen), we reasoned that the observed trend in binding affinity is likely based on hydrophobicity.[2] In the current study, Q8 similarly prefers Trp and Phe to the other two aromatic residues. Sequence-specific recognition is most likely the result of close proximity between the aromatic group and the N-terminal ammonium group, where inclusion of an N-terminal aromatic side chain forces the proximal ammonium group (which is otherwise well solvated in water) into close proximity to Q8, thus promoting chelation by the carbonyl groups to stabilize the complex. The fact that binding is only observable for peptides **1** and **4** indicates that both interactions, the inclusion of the hydrophobic side chain and the ion−dipole contacts, contribute substantially to complex stability.

The crystal structure of Q8·$\mathbf{4}_2$ reveals close interaction between the two phenyl groups of **4**. This interaction likely provides additional enthalpic stability in the binding of the second equivalent of **4** and is presumably the basis for the positive cooperativity observed in this system. This explanation is also consistent with the thermodynamic data in Table 2, where the binding of **4** is more enthalpically favorable than that of **1**.

Conclusions

Q8 selectively recognizes and dimerizes the aromatic tripeptides Trp-Gly-Gly (**1**) and Phe-Gly-Gly (**4**) with high affinity in an aqueous solution. The former binds with negative cooperativity or noncooperativity, and the latter, with positive cooperativity. Crystal structures of Q8·**1** and Q8·$\mathbf{4}_2$ reveal the structural basis for selective recognition as the inclusion of the hydrophobic aromatic side chain and chelation of the proximal N-terminal ammonium group by carbonyl oxygens on Q8. The structures also show ordered interactions of the adjacent peptide residues, providing an aid for the discovery of better-binding peptides. This work presents an example of a synthetic host that can bind with high affinity to peptides in an aqueous solution. Moreover, the extent of sequence selective recognition and positively cooperative dimerization of peptides reported here are, to the best of our knowledge, unprecedented for synthetic hosts in an aqueous solution. Such hosts should be useful in studying dimer-mediated biochemical processes, in the separation of protein mixtures, or as generic agents for the noncovalent dimerization of Phe-modified compounds.

Experimental Details

Materials. The following commercial reagents of analytical or higher purity grade were used without further purification: deuterium oxide (Cambridge Isotope Laboratories); H−Trp-Gly-Gly−OH (WGG, **1**), H−Gly-Trp-Gly−OH (GWG, **2**), H−Gly-Gly-Trp−OH (GGW, **3**), H−Tyr-Gly-Gly−OH (YGG, **7**), and H−Gly-Gly-Tyr (GGY, **9**) (Bachem); H−Gly-Tyr-Gly-OH (GYG, **8**), H−Phe-Gly-Gly−OH (FGG, **4**), H−Gly-Phe-Gly−OH (GFG, **5**), H−Gly-Gly-Phe−OH (GGF, **6**), H−His-Gly-Gly−OH (HGG, **10**), H−Gly-His-Gly−OH (GHG, **11**), and H−Gly-Gly-His−OH (GGH, **12**) (Chem-Impex International); sodium phosphate (mono- and dibasic) (Sigma); and cucurbit[8]uril (Q8) was synthesized by the group of Dr. Anthony Day (University of New South Wales, Australia) and purchased from Unisearch. Water was obtained from a Barnstead Nanopure Infinity water system (18 MΩ cm).

A stock solution of 1.0 M sodium phosphate buffer was adjusted to pH 7.0 and sterile filtered. The pH was checked periodically. With the exception of the NMR experiments, which were run in deuterium oxide, all binding experiments were carried out in 10 mM phosphate buffer, which was made as needed by diluting the 1.0 M stock. Fresh analyte solutions were prepared every couple of days and were thoroughly dissolved by heating at 60 °C. All analytes were massed to ±0.02 mg with an accuracy of at least three significant digits. The purities of Q8, FGG, and WGG were determined by ^{1}H NMR using anhydrous *tert*-butyl alcohol as reference.

Isothermal Titration Calorimetry (ITC). Titration experiments were carried out in 10 mM sodium phosphate buffer (pH 7.0) at 27 °C on a VP-ITC calorimeter from Microcal, Inc. (http://www.microcalorimetry.com). In a typical experiment, Q8 was in the sample cell at a concentration of 0.05−0.1 mM, and peptide was in the injection syringe at a concentration of 0.5−2 mM. The titration schedule consisted of 29 or 57 consecutive injections of 2−10 μL with at least a 200 s interval between injections. Heats of dilution, measured by titrating beyond saturation, were subtracted from each data set. All solutions were degassed prior to titration. The data were analyzed using Origin 7.0 software.

^{1}H NMR Spectroscopy. One-dimensional spectra were collected in deuterium oxide at 25 °C on a Varian Inova 400 MHz spectrophotometer using a presaturation pulse to suppress the signal from residual protiated solvent.

Crystallization, Data Collection, Structure Determination, and Refinement. Colorless crystals of Q8·**1** were grown by heating 5 mL of a solution of 1.6 mM **1** and 1.6 mM Q8 in neat water to 50 °C (to dissolve) and cooling overnight to room temperature in a Dewar flask. Clear crystals of Q8·$\mathbf{4}_2$ were grown by heating 5 mL of a solution of 3.0 mM **4** and 1.5 mM Q8 in neat water to 50 °C (to dissolve) and cooling overnight to room temperature in a Dewar flask. Crystal growth was expedited with the use of seed crystals. Prior to data collection, crystals were immersed in a 30% (v/v) ethylene glycol solution and flash cooled in liquid nitrogen for data acquisition at 100 K. The Q8·$\mathbf{4}_2$ complex crystallized in space group $P2_12_12_1$ with unit cell constants a = 13.21 Å, b = 22.66 Å, c = 38.42 Å, and the Q8·**1** complex crystallized in space group $P4_1$ with unit cell constants $a = b$ = 22.65 Å, c = 21.63 Å. Synchrotron data were collected at the Advanced Light Source beamlines 5.0.2 and 4.2.2 for Q8·$\mathbf{4}_2$ and Q8·**1**, respectively, using ADSC and Westbrook CCD detectors. Diffraction data were integrated and scaled using HKL-2000.[24] The *ab initio* determination of both crystal structures was performed using Sir2004.[25] Manual adjustments including addition of solvent molecules were made to the models superimposed on electron density maps using COOT,[26] and computational least-squares refinement was completed using SHELXL.[27] Non-hydrogen atoms were refined anisotropically, and riding hydrogens were added in the final stages of refinement.

Acknowledgment. This article is dedicated to Professor Michael J. Waring on the occasion of his retirement. We thank Prof. Don Crothers for helpful discussions. This work was supported in part by grants from the Donors of the American Chemical Society Petroleum Research Fund (42220-GB4, ARU), Research Corporation (CC6517, ARU), and the Robert A. Welch Foundation (AQ-1399, PJH). We are also grateful to the Robert A. Welch Foundation (departmental grant) for a summer stipend to L.M.H. and to the W. M. Keck Foundation for generous support to Trinity University that provided the titration calorimeter and mass spectrometer used in this study.

(24) Otwinowski, Z.; Minor, W. In *International Tables for Crystallography*; Rossman, M., Arnold, W., Eds.; Kluwer: Dordrecht, The Netherlands, 2001; Vol. F.
(25) Burla, M. C.; Caliandro, R.; Camalli, M.; Carrozzini, B.; Cascarano, G. L.; De Caro, L.; Giacovazzo, C.; Polidori, G.; Spagna, R. *J. Appl. Crystallogr.* **2005**, *38*, 381−388.
(26) Emsley, P.; Cowtan, K. *Acta Crystallogr.* **2004**, *D60*, 2126−2132.
(27) Sheldrick, G. M.; Schenider, T. R. *Methods Enzymol.* **1997**, *277*, 319−343.

We thank the staff at Advanced Light Source beamlines 4.2.2 and 5.0.2 and, in particular, Jay Nix for assistance in collecting the data for Q8·**1**. Support for the X-ray Crystallography Core Laboratory at the University of Texas Health Science Center by the Institutional Executive Research Council and the San Antonio Cancer Institute is gratefully acknowledged.

Supporting Information Available: Representative ITC binding data for compounds **2**, **3**, **5**−**12**, and Gly-Gly-Gly. Mass spectrometry data for compounds **1** and **4** in complex with Q8. ^{1}H NMR titration for compound **1**. Method of K_{ter} determination. X-ray data collection and refinement statistics. Thermal ellipsoid (ORTEP) diagrams, electron-density map diagrams, and coordinate files for the crystal structures of Q8·**1** and Q8·$\mathbf{4}_2$. Diagram of the crystal packing interactions of Q8·$\mathbf{4}_2$. This material is available free of charge via the Internet at http://pubs.acs.org.

JA064323S