Journal of Molecular Liquids 328 (2021) 115479

Contents lists available at ScienceDirect

Journal of Molecular Liquids

journal homepage: www.elsevier.com/locate/molliq

Sequence-dependent nanomolar binding of tripeptides containing N-terminal phenylalanine by Cucurbit[7]uril: A theoretical study

Fenfen Ma [a], Xiaoyan Zheng [a,*], Liangxu Xie [b], Zesheng Li [a,*]

[a] Key Laboratory of Cluster Science of Ministry of Education, Beijing Key laboratory of Photoelectronic/Electro-photonic Conversion Materials, School of Chemistry and Chemical Engineering, Beijing Institute of Technology, Beijing 100081, China
[b] Institute of Bioinformatics and Medical Engineering, Jiangsu University of Technology, Changzhou 213001, China

ARTICLE INFO

Article history:
Received 12 November 2020
Received in revised form 16 January 2021
Accepted 22 January 2021
Available online 26 January 2021

Keywords:
Aqueous solution
Molecular dynamics simulations
Host-guest interactions
Cucurbit[7]uril
Peptide
Sequence selectivity

ABSTRACT

Molecular recognition towards peptides with high affinity and sequence selectivity by synthetic macrocyclic receptors is a great challenge in supramolecular chemistry. In this work, we investigate the molecular recognition of tripeptides containing N-terminal phenylalanine (Phe, F) by synthetic macrocyclic receptor Cucurbit[7]uril (CB[7]) in aqueous solution by molecular dynamics simulations. We found that among twenty amino acids (AAs), the binding of CB[7] to Phe is the strongest. Based on that a series of tripeptides containing N-terminal Phe (with the second residue Gly (G) fixed) are designed to explore the effects of adjacent AA residues on the binding to CB[7]. It is indicated that, in contrast to FGG, only when the 3rd-residue is Glu (E), Lys (K), or Arg (R), the binding affinity with CB[7] can reach the nanomolar level. For the most prominent FGE, the binding free energy with CB[7] is −12.8 kcal/mol, with association constant (K_a) of 2.1×10^9 M^{-1}. It is surprising that once reversing the sequence order of the 2nd-residue and 3rd-residue from FGX to FXG (X = E, K, R), or inserting a second Gly in the middle of FGX, such as FGGX, the binding affinity is significantly changed, with K_a decreased more than 3 orders of magnitude. These results predicted the high binding affinity and sequence selectivity of CB[7] to containing N-terminal Phe peptides, which are beneficial to predict the recognition sites of specific proteins/peptides and to design host-guest complexes with high affinity.

1. Introduction

Synthetic macrocyclic receptors that target guest molecules with high affinity and selectivity play key roles in supramolecular chemistry and biochemistry [1–5]. Lots of studies have indicated that macrocyclic receptors could recognize protein/peptide [6,7], control proteins assemblies [8], and modulate protein interactions by targeting special amino acids (AAs) in them [9], which are crucial to applications in drug delivery [3,10], drug design [7,11], self-assembly process [8,9] and biomedical applications [11,12]. Noncovalent interactions, such as ion-dipole, hydrophobic, hydrogen bonding (H-bond), and van der Waals (*v*dW) are the major driving forces in these recognition processes.

Cucurbit[7]uril (CB[7]), derived from the Cucurbiturils family with the high affinity [13], owns good water solubility [14] and extraordinary binding affinities with various guest molecules in aqueous solution [14], which makes it exhibit promising applications in biochemical [14,15], biomedical [11,12], and material science [8,16]. Chinai et al. found that human insulin could be recognized by CB[7] receptor via binding with N-terminal Phe (F) residue [6], which had been predicted accurately from the study on short peptides [17]. Wei and co-workers screened out native insulin and human growth hormone from human serum via targeting of CB[7] towards N-terminal Phe residue [11]. In addition, Webber et al. proposed that covalent modified CB[7] could recognize insulin by encapsulating the N-terminal Phe residue, and the stability of insulin could be increased without affecting the bioactivity of insulin [18]. It is also found that CB[7] could inhibit the fibrillation of human calcitonin (by binding with the aromatic groups of Phe or Tyr) [7] and insulin/amyloid (by capturing Phe residue) [19], and then improved their bioactivity. Based on the above surveys, CB[7] recognizes proteins/peptides by targeting single aromatic AA residue, and the high affinity and selectivity of CB[7] towards AAs are key to recognitions. Therefore, the binding energies of several aromatic CB[7]/AAs in water had been researched by the theoretical and experimental methods [17,20], which indicated that the binding of CB[7] and aromatic Phe is the most strongest.

Compared with single AA, the structures of short peptides can better reflect the actual configuration of protein. Therefore, the studies of molecular recognition on short peptides are more practical than that of single AA. And in the folded protein, only the binding site is small enough (2–3 AA residues), the residues involved are adjacent in the sequence. Therefore, lots of researchers pay attention to the study on the

* Corresponding authors.
E-mail address: xiaoyanzheng@bit.edu.cn (X. Zheng).

https://doi.org/10.1016/j.molliq.2021.115479
0167-7322/

molecular recognition between macrocyclic host molecule and short peptides [17,21–27]. Urbach et al. found that cucurbit[8]uril (CB[8]) could bind with tripeptide Tyr-Leu-Ala in aqueous solution with nanomolar affinity via simultaneous inclusion of two neighboring side chains [23]. Besides, CB[8] could also bind with nonaromatic peptides with submicromolar affinities, such as Met-Terminated Peptides, by simultaneously encapsulating two neighboring non-aromatic-type residues into CB[8] cavity [24]. Due to the restriction of cavity volume, CB[7] can only encapsulate single AA forming 1:1 ratio complex [14], it is impossible to regulate the binding strength by simultaneously encapsulating two adjacent AA residues into the cavity like CB[8]. Nevertheless, the binding affinities of CB[7] to peptides are also sequence-dependent [17,27], for example, the binding affinity of CB[7] to Phe-Gly-Gly (FGG) is stronger than GFG and GGF [17]. Therefore, both the residue inside CB[7] cavity and the neighboring residues play important roles in the molecular recognition of CB[7] to short peptides. However, the specific impact of adjacent residues is still unclear.

Here we firstly systematically study the binding properties between CB[7] and AAs in aqueous solution to elucidate the binding difference between CB[7] and different AAs, and clarify the binding preference for aromatic AAs, especially for Phe. Then, we design a series of tripeptides containing N-terminal Phe and study the binding characteristics with CB[7] to explore the sequence dependence of tripeptides containing N-terminal Phe binding with CB[7]. We found that the molecular recognition between CB[7] and tripeptides containing N-terminal Phe has significant sequence selectivity and the binding affinity between CB[7] and tripeptide Phe-Gly-Glu (FGE) is at the nanomolar level.

2. Computational methods

All-atom MD simulations have been performed by GROMACS (version 5.1.5) package [28]. Initial structures of AAs and tripeptides were created using GaussView6.0 software package, and the structure of CB[7] was obtained from the X-ray crystal structure [29]. Initial conformations of the host-guest complexes were obtained by randomly placing the guests into the CB[7] cavity. The force field parameters of CB[7] and all AAs were modeled using the General Amber Force Field (GAFF) [30]. The partial charge of each atom which could reproduce the electrostatic potential calculated by Gaussian 09 [31] at the ωB97X-D/6-311++G** level, was obtained by the restrained electrostatic potential (RESP) scheme [32]. For all peptides, Amber 03 force field [33] was used. For all systems: free CB[7], free guest, CB[7]/guest and pure solvent, a cubic box of $5 \times 5 \times 5$ nm^3 were setup and explicitly solvated by 3000 TIP3P [34] water molecules. Besides, counter ion: Na^+ or Cl^- was added adaptively in charged systems to keep the system neutral. For each system, the steepest descent energy minimization was firstly performed, then 10 ns pre-equilibrium MD simulation was performed under the NVT ensemble, finally 500 ns long time production MD simulations were performed under NPT ensemble at constant temperature (300*K*) and pressure (1 bar). The last 400 ns of the production simulation of each system is used to the data analysis. The temperature was controlled by velocity rescaling thermostat [35] with 0.1 ps coupling constant, and the pressure is maintained at 1 ps interval by Parrinello-Rahman barostat [36]. The time step for the MD simulation is 2 fs [37,38]. Electrostatic interactions were evaluated using the Particle Mesh Ewald (PME) method [39], and Lennard-Jones interactions were calculated with a cutoff of 1.0 nm.

The solvent-balance method is used to simulate the interaction between host and guest molecules [40]. As shown in Fig. 1, the initial states are the free host and guest solvated in explicit 3000 water molecules, and the final states are the host-guest complexes in solvent and the pure aqueous solution with the same amount of water molecules as the initial states, which exactly balances the solvent of the initial and final states. The binding enthalpy (ΔH) between CB[7] and each guest molecule is the difference of the average potential energies for four systems, which is computed by Eq. (1): [40].

$$\Delta H = E_{\mathrm{Complex}} + E_{\mathrm{Water}} - (E_{\mathrm{Host}} + E_{\mathrm{Guest}}) \quad (1)$$

where E_{complex}, $E_{\mathrm{CB[7]}}$ and E_{guest} are the average potential energy of CB[7]/guest complex, free CB[7] and free guest in aqueous solution, and E_{water} is the average potential energy of pure water solution.

The entropy change of host CB[7] (ΔS_{Host}) and guest molecule ($\Delta S_{\mathrm{Guest}}$) in the binding process is computed separately by normal mode analysis [41,42] based on the following equation: $\Delta S_{\mathrm{Host}} = S_{\mathrm{Host/complex}} - S_{\mathrm{Host/free}}$ and $\Delta S_{\mathrm{Guest}} = S_{\mathrm{Guest/complex}} - S_{\mathrm{Guest/free}}$, where $S_{\mathrm{Host/complex}}$ and $S_{\mathrm{Guest/complex}}$ are the average entropy of CB[7] and guest in the complex state, and $S_{\mathrm{Host/free}}$ and $S_{\mathrm{Guest/free}}$ represent the average entropy of CB[7] and guest in the free state. The total entropy change (ΔS) of the complex is computed by Eq. (2):

$$\Delta S = \Delta S_{\mathrm{Host}} + \Delta S_{\mathrm{Guest}} \quad (2)$$

The binding free energy (ΔG) at $T = 300$ K is obtained by Eq. (3):

$$\Delta G = \Delta H - \mathrm{T}\Delta S \quad (3)$$

The noncovalent interactions and electrostatic potential (ESP) are analysed by Multiwfn 3.7 program [43]. Visualization and data analysis of MD trajectories were carried out by the utilities in the GROMACS package [28] and the VMD package [44].

3. Result and discussion

3.1. Host-guest interaction between CB[7] and AAs

A systematic understanding of the binding characteristics between CB[7] and AAs (the basic unit of peptides/proteins) is crucial to the in-depth investigation of the molecular recognition of CB[7] towards peptides/proteins. The binding between CB[7] and 20 AAs in a neutral aqueous solution is firstly studied by all-atom MD simulations. The corresponding chemical structure of 20 AAs and CB[7] are shown in Fig. S1. According to the relative position of AA to CB[7] cavity and the water numbers embedded inside the cavity, the binding manners between CB[7] and 20 AAs can be classified into three groups: (I) exclusion complex, (II) partial inclusion complex and (III) inclusion complex, see the cartoon models in Fig. 2a. The average water numbers inside the CB[7]

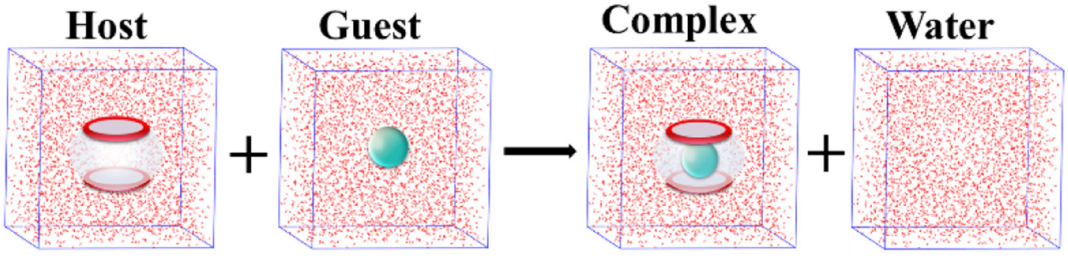

Fig. 1. Diagram of binding enthalpy computation. The barrel and ball in cyan refer to CB[7] and amino acid, respectively, the red points refer to oxygen atoms of water solvent.

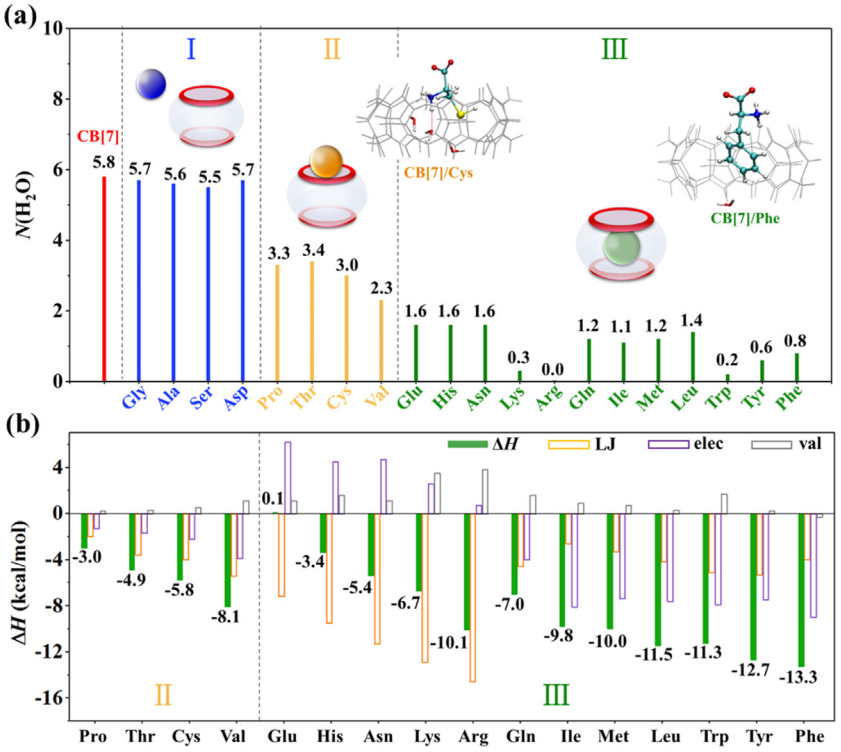

Fig. 2. (a) The cartoon models and the average number of water molecules inside CB[7] cavity (marked in black) of three different binding manners between CB[7] and AAs: (I) exclusion complex, (II) partial inclusion complex and (III) inclusion complex. The balls in blue, orange and teal refer to AAs in different binding manners, and CB[7] is presented by the barrel. As the examples of manners II and III, the conformation of CB[7]/Cys and CB[7]/Phe and their inside waters are shown in the inset, with C in cyan, H in white, N in blue, O in red, S in yellow, respectively. H-bonds are indicated by the red dotted line. All the color scheme is the same in this work. (b) The binding enthalpies Δ*H* for complexes in manners II and III, the decomposed terms: LJ (Lennard-Jones interactions), Elec (electrostatic interactions) and Val (contribution from intramolecular structural changes). The values of Δ*H* are labeled in black.

cavity of CB[7]/AA complexes are shown in Fig. 2a. It is reported that the recognition process of CB[7] towards AA is enthalpy-driven, that is, binding enthalpy plays a major role in the process, the entropy contribution is neglectable because of the rigid nature of CB[7] and small size of AA [20,45]. The host-guest binding enthalpies Δ*H* and the corresponding energy decomposition terms are shown in Fig. 2b and Table S1. Here the Δ*H* for each system is decomposed into three contributions: one is the change of Lennard-Jones force field term (LJ), which primarily models *v*dW forces; one is the contribution of Coulombic electrostatic interaction (elec); one is the change of valence terms (val), comprising of bond-stretch, angle-bend, and dihedral terms [40]. The more negative Δ*H* is, the more strong binding strength of the complex is.

3.1.1. Manner I

For the exclusion complex, the side chain of AA is too small (Gly, Ala, Ser) or negatively charged (Asp), which are hard to encapsulate into the hydrophobic cavity, therefore the guest AA is far away from CB[7] in aqueous solution. Similar to free CB[7], the hydrophobic cavity of the exclusion complex is filled by water molecules with a rectangular H-bonds network due to spatial confinement (Fig. S2), consistent with previous research about waters inside CB[n] [46]. The number of embedded water molecules for exclusion complexes is about 5.5– 5.7, close to that of free CB[7] (5.8), see Fig. 2a. Since AAs are not encapsulated into CB[7] cavity for complexes in manner I, the binding enthalpies are tiny and even positive (−0.7– 1.8 kcal/mol), see Table S1.

3.1.2. Manner II

For partial inclusion complex, the side chain of AA is hydrophobic (Pro, Cys and Val) or bigger (Thr) than AA in manner I [47,48]. As shown in Fig. 2a, the number of embedded water molecules is about 3, which are less than those of complexes in the manner I. In each complex, AA is located at one portal of CB[7], leaving a partial hydrophobic cavity and the other portal for water molecules (see the representative conformation for CB[7]/Cys in Fig. 2a). It is clear that the corresponding binding strength increases as the decrease of the number of embedded water molecules, such as CB[7]/Thr, CB[7]/Cys and CB[7]/Val (Fig. 2), which is in line with the result that the release of water molecules within CB[7] cavity could provide an essential driving force for host-guest complexation [46]. In addition, both LJ and elec play important roles in the binding strength, because of the hydrophobic interaction between CB[7] cavity and the side chain of AA, and the ion-dipole interaction between backbone $-NH_3^+$ and one CB[7] portal. And the contribution of LJ is always larger than the corresponding elec.

3.1.3. Manner III

For inclusion complex, the side chain of AA is relatively large and completely embedded inside the cavity, leading to the release of almost all water molecules (see Fig. 2a). The representative conformation (CB[7]/Phe) is shown in Fig. 2a, and the others are listed in Figs. S3–4. According to the different characteristics (hydrophilia or hydrophobicity) of the side chains of AAs, the complexes in manner III were further divided into two subgroups: hydrophilic AAs and hydrophobic AAs, respectively.

For each inclusion complex of hydrophilic AAs (Glu, His, Asn, Lys, Arg and Gln), the ring-like structure is formed inside the cavity by the intramolecular electrostatic interaction of AA, such as the intramolecular H-bonds between backbone $-COO^-$ and side chain $-NH_3^+$ in CB[7]/Lys complex (see Fig. S3). The ring-like AA is completely encapsulated into CB[7] cavity, which leads to a large LJ contribution, however, the unfavorable electrostatic repulsion between $-COO^-$ and the electronegative inner wall of CB[7] cavity partially cancels the large LJ. Thus, the Δ*H* of the inclusion complexes for hydrophilic AAs are not so large (Fig. 2b). Meanwhile, the ring-like structure of AA inside the CB[7] cavity is much

different from that of free AA, thus the contribution of val also plays some role. For Gln, there is weak electrostatic repulsion between side chain -$CONH_2$ group and CB[7] cavity, while the electrostatic attractions between backbone -NH_3^+, side chain –NH_2 of Gln and two portals of CB[7] are relatively strong, therefore, elec is favorable for CB[7]/Gln.

For AAs with hydrophobic side chains, such as aliphatic AAs (Ile, Met and Leu) and aromatic AAs (Trp, Tyr, and Phe), the hydrophobic side chains of each AA is embedded into CB[7] cavity by hydrophobic interaction, and electropositive backbone -NH_3^+ combines with one portal of CB[7] by ion-dipole interaction (see Fig. S4). The binding manners of aromatic CB[7]/AAs complexes are consistent with previous results [17,20]. It is obvious that both the LJ and elec are in favor of the host-guest binding (Fig. 2b), which renders the large binding strength between CB[7] and these AAs. Especially for aromatic AAs, they show the highest binding strength with CB[7] due to the hydrophobic aromatic ring of the sidechain suiting in the CB[7] cavity, supported by the favorable binding enthalpy in the order: Phe (−13.3 kcal/mol) < Tyr (−12.7 kcal/mol) < Trp (−11.3 kcal/mol). Therefore, for aromatic AAs, the order of binding strength to CB[7] is Phe > Tyr > Trp, which are qualitatively consistent with the experimental results that the binding affinity of aromatic AAs to CB[7] is in the order: Phe (1.8×10^5 M^{-1}) > Tyr (1.6×10^4 M^{-1}) > Trp (1.2×10^3 M^{-1}) [20].

The above results investigate systematically the binding manners and binding strength between CB[7] and all natural AAs, providing a complete library of binding strength for all CB[7]/AA complexes. The strongest binding of CB[7] to Phe gives an explicit support to previous many practical applications that CB[7] prefers to bind peptides or proteins with Phe at N-terminus [7,11,18].

3.2. Potential of mean force (PMF) calculations for CB[7]/Phe

To further characterize the interactions between CB[7] and Phe, and show the most stable binding configuration of CB[7]/Phe complex, the PMF for Phe to pass through the cavity of CB[7] from one side to the other was calculated by umbrella sampling method [49,50]. The COM distance between CB[7] and the phenyl ring of Phe side chain, projected along the central z-axis of CB[7] cavity, was used to monitor the translocation process (see Fig. 3a), and it is constrained to certain reference values in the successive umbrella sampling simulations.

As shown in Fig. 3b, as the hydrophobic phenyl ring of Phe entering into CB[7] cavity, the inside water molecules are gradually released, and the PMF gradually decreases until the phenyl ring completely embedded (structure S_1) with the strongest binding free energy −14.4 kcal/mol, which is close to the calculated Δ*H* of CB[7]/Phe (−13.3 kcal/mol), confirming the neglectable contribution of entropy. Then the free energy increases as the phenyl ring of Phe side chain is gradually pulled down to the other portal of CB[7], and when the phenyl ring is in the outside the cavity and meanwhile the hydrophilic backbone of Phe accesses the cavity, the PMF reaches to maximum (structure S_2, 7.1 kcal/mol). There is the other free energy minimum when the phenyl ring of Phe completely exposes to bulk water, and simultaneously the polar backbone of Phe forms intermolecular H-bonds with the portal of CB[7], at this time, the cavity is filled by water molecules again (structure S_3, −1.9 kcal/mol). It is obvious that a slight configuration change of CB[7]/Phe could produce a great impact on the binding energy, therefore, designing Phe-containing peptides could be an effective way for manipulating the binding strength between CB[7] and peptides.

3.3. Binding characteristics of CB[7] towards peptides containing N-terminal Phe

In practical application, AAs mostly exist as the building block of proteins/peptides, therefore although only one AA residue can be encapsulated into CB[7] cavity [14], the adjacent residues are also able to influence the binding stability between CB[7] and AAs via the interaction with electronegative CB[7] portals and electropositive outside wall of CB[7] cavity as shown in Fig. 3a, while the effect of different adjacent AA residues is indistinct. The experimental study on the molecular recognition between CB[7] and Phe-containing peptides revealed that the binding is more favored when Phe locates at the N-terminus of the peptide [17, 27]. Thereby, we take Phe as the N-terminus to explore the influence of different adjacent AA residues on the binding strength towards CB[7].

Firstly, the FGG is selected as a reference, because Gly (G) is the smallest AA with only one hydrogen atom in the sidechain. For CB[7]/FGG complex, the Δ*H* (−12.4 kcal/mol) is similar to the experimental value (−13.4 ± 0.4 kcal/mol) [17]. Due to the large flexibility of tripeptides, we also take the change of entropy into account in the calculation. The results indicate that -TΔ*S* is unfavorable for the binding process of CB[7] and FGG, which is in agreement with the previous experimental result [17]. The value of -TΔ*S* (6.7 kcal/mol) is a little more than the experimental value (4.4 kcal/mol), leading to the larger Δ*G* (−5.7 kcal/mol) than −9.0 ± 0.3 kcal/mol [17].

Observing IGM result of CB[7]/FGG complex in Fig. 4b, the intermolecular interactions between CB[7] and FGG consist of hydrophobic interaction between CB[7] cavity and phenyl ring of Phe side chain, ion-dipole interaction between CB[7] portal and N-terminal -NH_3^+ of FGG, *v*dW interactions between CB[7] and the 2nd-residue Gly_1, 3rd-residue Gly_2. Different from Gly_1 above CB[7] portal, Gly_2 completely exposes outside of CB[7] cavity, as shown in Fig. 4b. Substituting the 3rd-residue Gly_2 into other large side chains AAs and introducing new binding sites between CB[7] and tripeptide (such as the outside wall of CB[7]) may be an effective way to enhance the binding affinity. Therefore, we design a serial of tripeptides (FGX) to explore

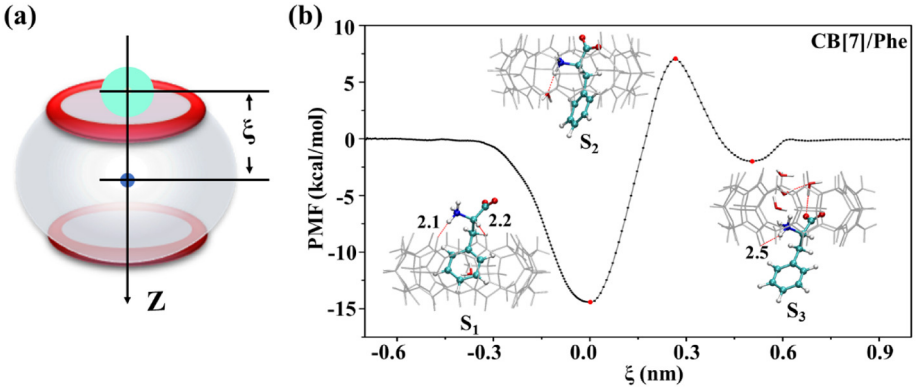

Fig. 3. (a) Illustrative representation of the reaction coordinate (ξ) used in umbrella sampling simulations. ξ presents the COM distance between CB[7] and the phenyl ring of the Phe side chain. (b) The PMF for Phe to translocate through the cavity of CB[7]. The typical structures along the reaction coordinate are displayed.

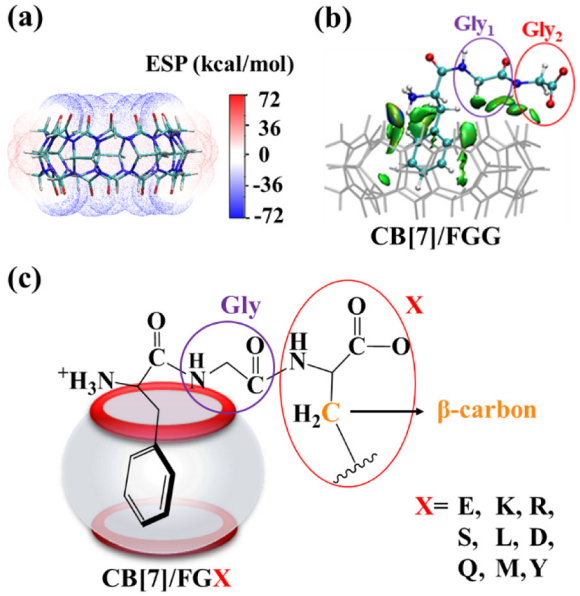

Fig. 4. (a) The electrostatic potential (ESP) surface of CB[7] (isovalue = 0.001 a.u.). (b) Independent Gradient Method (IGM) isosurface analysis of CB[7]/FGG. Gly_1 and Gly_2 represent the second and third residue of the tripeptide. (c) Schematic diagram of the complex between CB[7] and tripeptide FGX. The β-carbon of the third residue is marked by orange.

the influence of the different 3rd-residue on binding between CB[7] and tripeptides containing N-terminal Phe in this section.

The previous study on the binding of CB[8] (with larger cavity than inhering CB[7]) towards tripeptides indicated that the branches at β-carbon of AAs are unfavorable for the binding due to its steric hindrance [24]. To avoid the large steric effect, the nine AAs (X = E, K, R, S, L, D, Q, M, Y) without branch at the β-carbon were chosen to substitute the 3rd-residue of FGG. The structural formulas of tripeptides and the configurations of CB[7]/FGX complex are shown in Fig. S5 and Fig. 4c, respectively.

3.3.1. Binding properties of CB[7] towards designed tripeptide FGX

Due to the flexibility of tripeptides, entropy contribution was taken into account in all systems of CB[7] and peptides. As shown in Fig. 5a and Table 1, both enthalpy (ΔH) and entropy($-T\Delta S$) play important roles in host-guest binding, because of their comparable value, especially for FGY. For each CB[7]/FGX complex, ΔH is favorable, and $-T\Delta S$ is unfavorable, which is in line with the previous result of CB[7] binding peptides [17, 27]. In contrast to CB[7]/FGG complex, the three tripeptides show higher binding strength to CB[7], such as CB[7]/FGE, CB[7]/FGK, and CB[7]/FGR, with free energy differences −7.2, −6.2 and − 5.9 kcal/mol, respectively. The more negative ΔG of CB[7]/FGE, CB[7]/FGK and CB[7]/FGR than FGG are mainly contributed by ΔH, because of their similar entropy contributions. At the same time, the other tripeptide reflects the worse binding ability to CB[7] than FGG. The most negative ΔH of CB[7]/FGE could be attributed to the introduction of both the intermolecular H-bonds (between CB[7] portal and the backbone of 2nd-residue) and vdW interactions (between the outside wall of CB[7] and the side chain of 3rd-residue), as shown in IGM result (blue and green region) in Fig. 5f. For CB[7]/FGK, due to the long and flexible Lys side chain, both the backbone of 2nd-residue and the $-NH_3^+$ of 3rd-residue (Lys) side chain hang above CB[7] portal and form intermolecular H-bond and ion-dipole interactions with CB[7] portal (blue region in Fig. 5g). Due to the long side chain of residue R in CB[7]/FGR, the favorable host-guest binding interactions are intermolecular H-bond contributed by the 2nd-residue and the vdW interactions between the guanidine group and the outside wall of CB[7].

The average number of H-bond increases to 1.5, 1.7 and 1.6 for CB[7]/FGE, CB[7]/FGK and CB[7]/FGR (Table S3), respectively, compared with 1.3 of CB[7]/FGG, which also supports the increase of the larger binding strength. While the binding ability of the other six CB[7]/FGX (K, S, L, Q, M, Y) complexes are all worse than CB[7]/FGG, that is, only when the 3rd-residue of tripeptide is Glu (E), Lys (K) or Arg (R), the binding strength of CB[7] and FGX are increased.

The association constant (K_a) could be calculated by $\Delta G = -RT\ln K_a$ [51,52], where R and T are ideal gas constant and room temperature, respectively. As shown in Table 1, the K_a of CB[7]/FGE ($2.1 \times 10^9\ M^{-1}$), CB[7]/FGK ($4.7 \times 10^8\ M^{-1}$) and CB[7]/FGR ($2.8 \times 10^8\ M^{-1}$) is 4– 5 orders of magnitude larger than that of CB[7]/FGG (see Table 1). The corresponding ΔG (−12.8, −11.9 and − 11.6 kcal/mol) of the three CB[7]/FGX (E, K, R) are larger than that of CB[8]/YLA (−11.2 kcal/mol), which was reported with nanomolar binding affinity [23], therefore, the binding affinities of our designing three complexes (CB[7]/FGE, CB[7]/FGK and CB[7]/FGR) should also be at the nanomolar level.

For each CB[7]/FGX complex, entropy plays a key role in the binding process. The main contribution is from the entropy change of guest molecule ($-T\Delta S_{Guest}$), and the corresponding contribution of the host molecule ($-T\Delta S_{Host}$) could be neglectable because of its rigid feature (Fig. 5c). These could be further demonstrated by the conformational change of host and guest molecules before and after complex formation. Take CB[7]/FGY as an example, the distance projections of two representative carbon-atom pairs of CB[7] are almost the same (Fig. 5d), thus the values of $-T\Delta S_{Host}$ are usually small (Fig. 5c), while the distance projections of the other two representative carbon-atom pairs (backbone and side chain of FGY) for free and complex states are significantly different (Fig. 5e), the more localized distributions of the distance projections for FGY after complexation indicates the conformation of FGY is strongly restricted after binding with CB[7], therefore the value of $-T\Delta S_{Guest}$ is large.

3.3.2. Effect of sequence order on binding affinity

The influence of sequence order on the binding affinity of CB[7] and tripeptides was further investigated by reversing the sequence order of the 2nd-residue and 3rd-residue for the above selected three tripeptides, from FGX (X = E, K, R) to FXG (X = E, K, R), see Fig. 6a. It is obvious that the binding free energies of CB[7]/FXG (X = E, K, R) are more positive than the corresponding CB[7]/FGX (X = E, K, R), increasing by 5.4, 7.2 and 3.6 kcal/mol, respectively (Fig. 6b and Table S4), thus K_a decreases by 3– 5 orders of magnitude (see Table. S6), that is, the binding affinity of CB[7]/FXG (X = E, K, R) complexes are significantly worse than those of CB[7]/FGX (X = E, K, R), which indicates the binding of CB[7] towards tripeptides containing N-terminal Phe is significantly sequence-dependent.

3.3.3. Effect of the chain length of peptides on binding affinity

The effect of chain length of the peptide on the binding affinity between CB[7] and the peptide was studied by inserting the second Gly residue into between the original 2nd-residue and 3rd-residue, from FGX (X = E, K, R) to FGGX (X = E, K, R), see Fig. 6a. It is obvious that the binding free energies of FGGX (X = E, K, R) significantly increase by 9.9, 11.5 and 14.1 kcal/mol, respectively, and the corresponding K_a decreases about 7– 10 orders of magnitude (Fig. 6b and Table S4). Besides, the ΔG for CB[7]/FGGR is even positive, due to the influence of the fourth residue Arg (R), that is, the binding of CB[7] towards FGGR is difficult. Therefore, inserting a second Gly in the middle of FGX is unfavorable to binding with CB[7]. In respect to CB[7]/FGG (−12.4 kcal/mol), ΔH of CB[7]/FGGK and CB[7]/FGGR are more positive (−8.8 and − 8.3 kcal/mol), because the interaction between the fourth residue Lys/Arg and CB[7] is weak (see the IGM results in Fig. S7), moreover the conformations of phenyl ring inside CB[7] cavity are also different from that of CB[7]/Phe, which indicates that the hydrophobic interactions decrease after extending the chain length of peptides. Therefore, extending the chain length of the peptide by inserting the second Gly residue

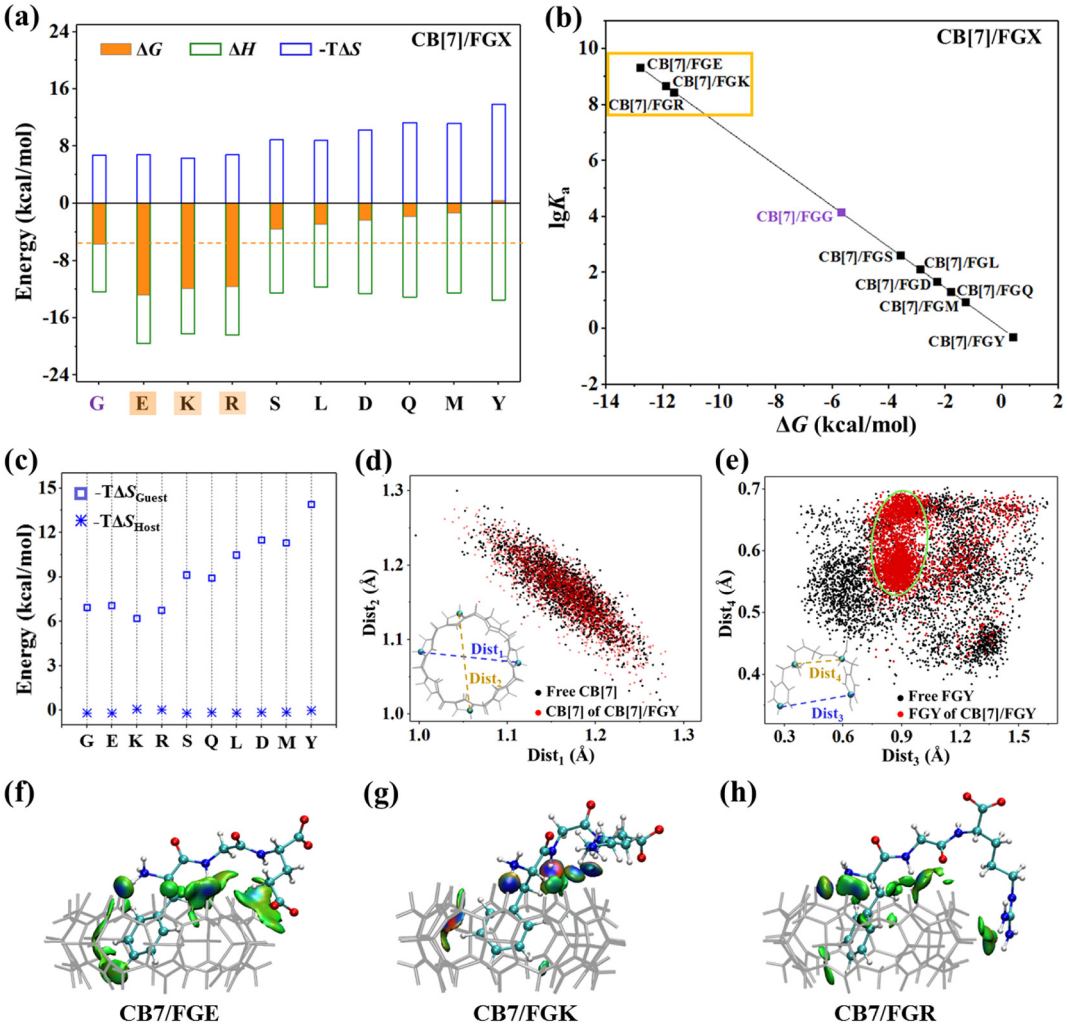

Fig. 5. (a) Thermodynamic properties of inclusion complexes of CB[7] and FGX, including binding free energy (ΔG), enthalpy (ΔH) and entropy ($-T\Delta S$). The orange dotted line is the ΔG of CB[7]/FGG complex. (b) The plot of the logarithm of association constant K_a (lgK_a) concerning ΔG. Strongly bound complexes are highlighted by the orange squares. (c) The entropy changes of the host ($-T\Delta S_{Host}$) and guest ($-T\Delta S_{Guest}$). (d) Distance projections of two pairs of carbon-atom in CB[7] ($Dist_1$ and $Dist_2$ marked in the inset) in free (black) and complex (red) state. (e) Distance projections of two pairs of carbon-atom with the farthest distance on the backbone and side chain of tripeptide FGY, respectively, in both free (black) and complex (red) state ($Dist_3$ and $Dist_4$ marked in the inset). (f-h) IGM isosurface analysis of CB[7]/FGE, CB[7]/FGK and CB[7]/FGR.

Table 1
Thermodynamic properties (binding free energy ΔG, enthalpy ΔH and entropy $-T\Delta S$, in kcal/mol) and the association constant K_a in M^{-1} of CB[7]/FGX complexes. $\Delta\Delta G = \Delta G_{(CB[7]/FGX)} - \Delta G_{(CB[7]/FGG)}$. The value of ΔG, K_a *and* $\Delta\Delta G$ *for the complexes that the binding affinities reach the nanomolar level* are bold.

	ΔG	$\Delta\Delta G$	K_a	ΔH	$-T\Delta S$
CB[7]/FGG	−5.7	0	1.4×10^4	−12.4	6.7
CB[7]/FGE	**−12.8**	**−7.1**	**2.1×10^9**	−19.6	6.8
CB[7]/FGK	**−11.9**	**−6.2**	**4.7×10^8**	−18.2	6.3
CB[7]/FGR	**−11.6**	**−5.9**	**2.8×10^8**	−18.4	6.8
CB[7]/FGS	−3.6	2.1	4.2×10^2	−12.5	8.9
CB[7]/FGL	−2.9	2.8	1.3×10^2	−11.7	8.8
CB[7]/FGD	−2.3	3.4	4.7×10	−12.6	10.3
CB[7]/FGQ	−1.8	3.9	2.1×10	−13.1	11.3
CB[7]/FGM	−1.3	4.4	8.9	−12.5	11.2
CB[7]/FGY	0.4	6.1	0.5	−13.5	13.9

largely reduces the binding affinity between CB[7] and peptides containing N-terminal Phe.

4. Conclusions

In this work, we systematically investigated the binding characteristics between CB[7] and all twenty AAs in aqueous solution and explored the binding strength of CB[7] towards a series of tripeptides by the MD method. It is indicated that there are three binding manners between CB[7] and single AA: (I) exclusion complex (Gly, Ala, Ser and Asp), (II) partial inclusion complex (Pro, Thr, Cys and Val), and (III) inclusion complex (all other AAs with large side chains). Among them, the binding affinity of CB[7] towards Phe is the strongest, therefore CB[7] prefers to recognize Phe residue in the real environment, which is consistent with many experimental results [21]. Next, a series of tripeptides containing N-terminal Phe are further designed to explore the effects of adjacent AA residues on the binding with CB[7]. The result indicates that, in contrast to FGG, only when the 3rd-residue is Glu, Lys or Arg, that is FGX (X = E, K or R), the binding strength can increase obviously. For the most prominent CB[7/]FGE, the binding free energy is −12.8 kcal/mol, with the corresponding association constant (K_a) of 2.1×10^9 M^{-1}, reflecting the nanomolar binding affinity. It is surprising that their binding affinities are significantly reduced once reversing the sequence order of the 2nd-residue and 3rd-residue (from FGX to FXG), or inserting the second Gly in FGX middle (from FGE to FGGE), with K_a decreased more than 3 orders of magnitude. These results indicate that the recognition of CB[7] towards the peptides containing N-terminal Phe is significant sequence selectivity, which could provide useful clues to CB[7] recognizing specific proteins/peptides and designing reasonable host-guest complexes with high affinities.

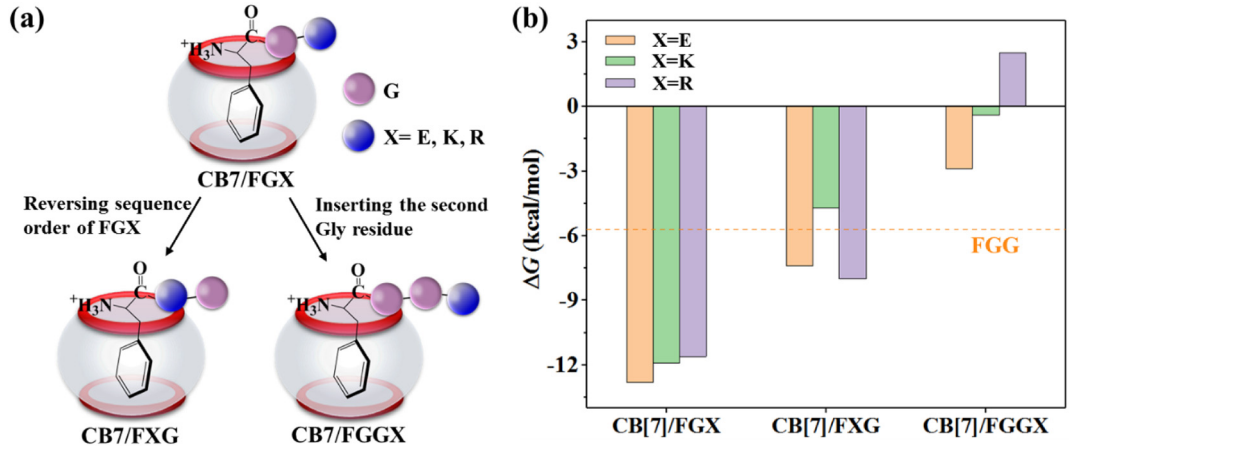

Fig. 6. (a) Schematic representation of the design principle of controlled simulations: one is reversing sequence order from FGX to FXG; the other is inserting the second Gly residue, from FGX to FGGX. (b) The binding free energies (ΔG) for FGX, FXG and FGGX systems, here X = E, K, R.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the financial support of the National Natural Science Foundation of China (Grants 21803007, 21473010, 91544223), Beijing Institute of Technology (BIT) Research Fund Program for Young Scholar.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.molliq.2021.115479.

References

[1] W. Liu, S.K. Samanta, B.D. Smith, L. Isaacs, Synthetic mimics of biotin/(strept)avidin, Chem. Soc. Rev. 46 (2017) 2391–2403.
[2] S. van Dun, C. Ottmann, L.-G. Milroy, L. Brunsveld, Supramolecular chemistry targeting proteins, J. Am. Chem. Soc. 139 (2017) 13960–13968.
[3] J. Zhou, G. Yu, F. Huang, Supramolecular chemotherapy based on host-guest molecular recognition: a novel strategy in the battle against cancer with a bright future, Chem. Soc. Rev. 46 (2017) 7021–7053.
[4] Z. Liu, S.K.M. Nalluri, J.F. Stoddart, Surveying macrocyclic chemistry: from flexible crown ethers to rigid cyclophanes, Chem. Soc. Rev. 46 (2017) 2459–2478.
[5] G. Yu, K. Jie, F. Huang, Supramolecular amphiphiles based on host–guest molecular recognition motifs, Chem. Rev. 115 (2015) 7240–7303.
[6] J.M. Chinai, A.B. Taylor, L.M. Ryno, N.D. Hargreaves, C.A. Morris, P.J. Hart, A.R. Urbach, Molecular recognition of insulin by a synthetic receptor, J. Am. Chem. Soc. 133 (2011) 8810–8813.
[7] H. Shang, A. Zhou, J. Jiang, Y. Liu, J. Xie, S. Li, Y. Chen, X. Zhu, H. Tan, J. Li, Inhibition of the fibrillation of highly amyloidogenic human calcitonin by cucurbit[7]uril with improved bioactivity, Acta Biomater. 78 (2018) 178–188.
[8] F. Guagnini, P.M. Antonik, M.L. Rennie, P. O'Byrne, A.R. Khan, R. Pinalli, E. Dalcanale, P.B. Crowley, Cucurbit[7]uril-dimethyllysine recognition in a model protein, Angew. Chem. Int. Ed. 57 (2018) 7126–7130.
[9] R.P. Bosmans, J.M. Briels, L.G. Milroy, T.F. de Greef, M. Merkx, L. Brunsveld, Supramolecular control over split-luciferase complementation, Angew. Chem. Int. Ed. 55 (2016) 8899–8903.
[10] M.J. Webber, R. Langer, Drug delivery by supramolecular design, Chem. Soc. Rev. 46 (2017) 6600–6620.
[11] W. Li, A.T. Bockus, B. Vinciguerra, L. Isaacs, A.R. Urbach, Predictive recognition of native proteins by cucurbit[7]uril in a complex mixture, Chem. Commun. 52 (2016) 8537–8540.
[12] A.A.A. Smith, C.L. Maikawa, G.A. Roth, E.A. Appel, Site-selective modification of proteins using cucurbit[7]uril as supramolecular protection for N-terminal aromatic amino acids, Org. Biomol. Chem. 18 (2020) 4371–4375.
[13] K.I. Assaf, W.M. Nau, Cucurbiturils: from synthesis to high-affinity binding and catalysis, Chem. Soc. Rev. 44 (2015) 394–418.
[14] S.J. Barrow, S. Kasera, M.J. Rowland, J. del Barrio, O.A. Scherman, Cucurbituril-based molecular recognition, Chem. Rev. 115 (2015) 12320–12406.
[15] J.K.K. Dinesh Shetty, Kyeng Min Parkad, Kimoon Kim, Can we beat the biotin–avidin pair?: cucurbit[7]uril-based ultrahigh affinity host–guest complexes and their applications, Chem. Soc. Rev. 44 (2015) 8747–8761.
[16] A.G. Mullins, N.K. Pinkin, J.A. Hardin, M.L. Waters, Achieving high affinity and selectivity for asymmetric dimethylarginine by putting a lid on a box, Angew. Chem. Int. Ed. 58 (2019) 5282–5285.
[17] L.A. Logsdon, C.L. Schardon, V. Ramalingam, S.K. Kwee, A.R. Urbach, Nanomolar binding of peptides containing noncanonical amino acids by a synthetic receptor, J. Am. Chem. Soc. 133 (2011) 17087–17092.
[18] M.J. Webber, E.A. Appel, B. Vinciguerra, A.B. Cortinas, L.S. Thapa, S. Jhunjhunwala, L. Isaacs, R. Langer, D.G. Anderson, Supramolecular PEGylation of biopharmaceuticals, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) 14189–14194.
[19] H.H. Lee, T.S. Choi, S.J.C. Lee, J.W. Lee, J. Park, Y.H. Ko, W.J. Kim, K. Kim, H.I. Kim, Supramolecular inhibition of amyloid fibrillation by cucurbit[7]uril, Angew. Chem. Int. Ed. 53 (2014) 7461–7465.
[20] J.W. Lee, H.H.L. Lee, Y.H. Ko, K. Kim, H.I. Kim, Deciphering the specific high-affinity binding of cucurbit[7]uril to amino acids in water, J. Phys. Chem. B 119 (2015) 4628–4636.
[21] A.R. Urbach, V. Ramalingam, Molecular recognition of amino acids, peptides, and proteins by cucurbit[n]uril receptors, Isr. J. Chem. 51 (2011) 664–678.
[22] A. Bandyopadhyay, S.K. Pati, Role of donor-acceptor macrocycles in sequence specific peptide recognition and their optoelectronic properties: a detailed computational insight, Phys. Chem. Chem. Phys. 18 (2016) 20682–20690.
[23] L.C. Smith, D.G. Leach, B.E. Blaylock, O.A. Ali, A.R. Urbach, Sequence-specific, nanomolar peptide binding via cucurbit[8]uril-induced folding and inclusion of neighboring side chains, J. Am. Chem. Soc. 137 (2015) 3663–3669.
[24] Z. Hirani, H.F. Taylor, E.F. Babcock, A.T. Bockus, C.D. Varnado, C.W. Bielawski, A.R. Urbach, Molecular recognition of methionine-terminated peptides by cucurbit[8]uril, J. Am. Chem. Soc. 140 (2018) 12263–12269.
[25] S. Fa, Y. Zhao, General method for peptide recognition in water through bioinspired complementarity, Chem. Mater. 31 (2019) 4889–4896.
[26] R.J. Gubeli, S. Sonzini, A. Podmore, P. Ravn, O.A. Scherman, C.F. van der Walle, Selective, non-covalent conjugation of synthetic peptides with recombinant proteins mediated by host–guest chemistry, Chem. Commun. 52 (2016) 4235–4238.
[27] M.V. Rekharsky, H. Yamamura, Y.H. Ko, N. Selvapalam, K. Kim, Y. Inoue, Sequence recognition and self-sorting of a dipeptide by cucurbit[6]uril and cucurbit[7]uril, Chem. Commun. (2008) 2236–2238.
[28] M.J. Abraham, D.v.d. Spoel, E. Lindahl, B. Hess, The GROMACS Development Team, GROMACS User Manual Version 5.1, 2015.
[29] J. Kim, I.-S. Jung, S.-Y. Kim, E. Lee, J.-K. Kang, S. Sakamoto, K. Yamaguchi, K. Kim, New cucurbituril homologues: syntheses, isolation, characterization, and X-ray crystal structures of cucurbit[n]uril (n = 5, 7, and 8), J. Am. Chem. Soc. 122 (2000) 540–541.
[30] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157–1174.
[31] G.W.T.M.J. Frisch, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V.B.B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F.I.J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K.T.R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H.N.T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E.B.K.N. Kudin, V.N. Star-overov, T. Keith, R. Kobayashi, J. Normand, K.R.A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M.M.M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R.G.R.E. Strat-Mann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W.O.R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P.

Salvador, J.J.D.S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J.C.D.J. Fox, Inc., Gaussian G09, Revision D.01, Gaussian, Wallingford, CT, 2013.
[32] C.I. Bayly, P. Cieplak, W. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, J. Phys. Chem. 97 (1993) 10269–10280.
[33] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, J. Comput. Chem. 24 (2003) 1999–2012.
[34] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[35] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, J. Chem. Phys. 126 (2007), 014101.
[36] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, J. Appl. Phys. 52 (1981) 7182–7190.
[37] W. Khuntawee, M. Karttunen, J. Wong-Ekkabut, A molecular dynamics study of conformations of beta-cyclodextrin and its eight derivatives in four different solvents, Phys. Chem. Chem. Phys. 19 (2017) 24219–24229.
[38] M. Zhang, Y. Huang, D. Hao, Y. Ji, D. Ouyang, Solvation structure and molecular interactions of ibuprofen with ethanol and water: a theoretical study, Fluid Phase Equilib. 510 (2020) 112454.
[39] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8593.
[40] A.T. Fenley, N.M. Henriksen, H.S. Muddana, M.K. Gilson, Bridging calorimetry and simulation through precise calculations of cucurbituril–guest binding enthalpies, J. Chem. Theory Comput. 10 (2014) 4069–4078.
[41] M. Levitt, C. Sander, P.S. Stern, The normal modes of a protein: native bovine pancreatic trypsin inhibitor, Int. J. Quantum Chem. 24 (1983) 181–199.
[42] B. Brooks, M. Karplus, Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor, Proc. Natl. Acad. Sci. U. S. A. 80 (1983) 6571–6575.
[43] T. Lu, F. Chen, Multiwfn: a multifunctional wavefunction analyzer, J. Comput. Chem. 33 (2012) 580–592.
[44] A.D.W. Humphrey, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[45] F.F. Ma, X.Y. Zheng, J. Xie, Z.S. Li, Binding properties of cucurbit[7]uril to neutral and protonated amino acids: a theoretical study, Int. J. Quantum Chem. 121 (2020) e26491.
[46] F. Biedermann, V.D. Uzunova, O.A. Scherman, W.M. Nau, A. De Simone, Release of high-energy water as an essential driving force for the high-affinity binding of cucurbit[n]urils, J. Am. Chem. Soc. 134 (2012) 15318–15323.
[47] C.P. Moon, K.G. Fleming, Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 10174–10177.
[48] A.A. Aboderin, An empirical hydrophobicity scale for a-amino-acids and some of its applications, Int. J. BioChem. 2 (1971) 537–544.
[49] J.G. Kirkwood, Statistical mechanics of fluid mixtures, J. Chem. Phys. 3 (1935) 300–313.
[50] B. Roux, The calculation of the potential of mean force using computer simulations, Comput. Phys. Commun. 91 (1995) 275–282.
[51] T.F.G.G. Cova, B.F. Milne, S.C.C. Nunes, A.A.C.C. Pais, Drastic stabilization of junction nodes in supramolecular structures based on host–guest complexes, Macromolecules 51 (2018) 2732–2741.
[52] N.S. Venkataramanan, A. Suvitha, Y. Kawazoe, Unravelling the nature of binding of cubane and substituted cubanes within cucurbiturils: a DFT and NCI study, J. Mol. Liq. 260 (2018) 18–29.