npj Computational Materials

www.nature.com/npjcompumats

ARTICLE OPEN

Check for updates

# Supramolecular gating of guest release from cucurbit[7]uril using de novo design

Hugues Lambert[1,2,3], Alvaro Castillo Bonillo[2,3,4], Qiang Zhu[4], Yong-Wei Zhang[1]✉ and Tung-Chun Lee[2,3]✉

Herein we computationally explore the modulation of the release kinetics of an encapsulated guest molecule from the cucurbit[7]uril (**CB7**) cavity by ligands binding to the host portal. We uncovered a correlation between the ligand-binding affinity with **CB7** and the guest residence time, allowing us to rapidly predict the release kinetics through straightforward energy minimization calculations. These high-throughput predictions in turn enable a Monte-Carlo Tree Search (MCTS) to *de novo* design a series of cap-shaped ligand molecules with large binding affinities and boosting guest residence times by up to 7 orders of magnitude. Notably, halogenated aromatic compounds emerge as top-ranking ligands. Detailed modeling suggests the presence of halogen-bonding between the ligands and the CB7 portal. Meanwhile, the binding of top-ranked ligands is supported by $^{1}$H NMR and 2D DOSY-NMR. Our findings open up possibilities in gating of molecular transport through a nanoscale cavity with potential applications in nanopore technology and controlled drug release.

*npj Computational Materials* (2022)8:21 ; https://doi.org/10.1038/s41524-022-00702-0

## INTRODUCTION

Gating the release and the transport of molecules underpins a range of essential biological functions, especially in advanced organisms. For instance, exocytosis of neurotransmitters into a synapse is mediated by the fusion of secretory vesicles and porosomes[1,2], whereas the membrane potential of a cell is controlled by the gating of ion fluxes through its ion channels[3]. In enzymes, the access of solvent molecules and substrates to the active site can be restricted through conformational gating[4] and salt concentration[5]. The gating of substrates and products can even serve the purpose of controlling and synchronizing reaction chains, as is the case in ammonia-transferring enzymes[6]. While these biological processes are often mediated by complex chemical networks, the gating of molecular release is less common in simple synthetic host-guest systems.

Cucurbit[*n*]urils (**CB***n*, $n = 5$–8,10) are synthetic macrocycles composed of repeating glycoluril units. These organic macrocyclic compounds possess carbonyl-rich portals and a hydrophobic cavity. They are able to form extremely stable complexes with guests, with binding affinities surpassing that of the avidin-biotin complex[7,8]. Despite their deceptively simple symmetric structure, cucurbiturils have proved to be extremely versatile molecules. Cucurbiturils have been shown to promote the stabilization of degradation-prone species[9] and short-lived electronic states[10]. In catalysis, **CB6** was found to dramatically enhance the 1,3 dipolar cycloaddition between azides and acetylenes, or as it is known today, in situ click chemistry[11], while a selective retro-Diels-Alder reaction was demonstrated in the gas phase[12]. Recently, **CB6** has also been suggested to assist the isomerization of xylene[13]. A complete review of the role of **CB***n* in catalysis can be found elsewhere[14]. A variety of guests have been reported to form host-guest complexes with **CB***n* with a range of binding affinities[15]. Building on these successes, **CB***n* have been included in several supramolecular architectures[16] such as hydrogels[17], micelles[18,19], hybrid nanoparticles[20–22], and nanomachines[23,24].

In 2004, Kim leveraged **CB5** and **CB6** as part of egg yolk L-*α*-phosphatidylcholine (EYPC) membranes to create artificial selective ion channels able to respond to pH variations[25]. Additional improvements on the earliest passive ion channels based on cyclodextrins[26] involved the design of voltage-, light-, mechano- and ligand-gating mechanisms[27]. Ligand-based gating, in particular, is especially interesting given that its prevalence in natural systems[28] has yet to be matched in artificial constructs. In a bio-inspired example of ligand-gated ion channel, a synthetic chiral receptor embedded in a lipid membrane can allow ion transport to occur across a channel when the addition of a ligand induces a conformational change. The addition of *β*-cyclodextrin would scavenge the ligand and return the channel to a non-conductive state[29].

Previous works have highlighted that the cavity of **CB6** can be gated through modulation of pH[30] and salt concentrations[31] but the gated release of a neutral molecular species from **CB7** has not been reported to the best of our knowledge due to its larger portal. Here we design a *de novo* approach to highlight strong ligands for the portal of **CB7** that are able to hinder the entry and exit of the guest from the cavity of the macrocycle. Machine Learning is used to explore the chemical space and complex potential energy surfaces[32–34]. In particular, ligands with a strong binding affinity with **CB7**'s portal (up to $-41.85\,\mathrm{kcal\,mol^{-1}}$) are able to boost a guest residence time by over 7 orders of magnitude, to 350 s, thanks to *π*-orbital-driven halogen bonds. We believe that these ligands will stand as a valuable addition to the toolkit of designers of **CB7**-based channels and hope our de novo design approach will inspire other designs of supramolecular gating systems.

## RESULTS AND DISCUSSION

### Overview of design procedure

The general workflow used to design strong ligands for **CB7** is illustrated in Fig. 1. To keep computations tractable, a binary host-

[1]Institute of High Performance Computing, A*STAR, 1 Fusionopolis Way, Connexis Tower, 138632 Singapore, Singapore. [2]Department of Chemistry, Christopher Ingold Building, University College London (UCL), 20 Gordon Street, London WC1H 0AJ, UK. [3]Institute for Materials Discovery, University College London (UCL), London, UK. [4]Institute of Materials Research and Engineering, A*STAR, 2 Fusionopolis Way, Innovis Tower, 138634 Singapore, Singapore. ✉email: zhangyw@ihpc.a-star.edu.sg; tungchun.lee@ucl.ac.uk

guest complex composed of adamantanone (ADA) encapsulated inside of **CB7** is considered. From preliminary studies of known ligands for the portal of **CB7**, a correlation between the binding affinity of the ligands with the portal and their ability to prevent the escape of the ADA guest is observed. Further investigations revealed that this correlation is a Linear Free-Energy Relationship (LFER) that can be used as a heuristic to quickly estimate the 'gating power', or the ability to prevent a guest from escaping, of a given ligand candidate. Instead of systematically screening a library of molecular compounds, a generative algorithm (a Monte Carlo Tree Search with a recurrent neural network used in the roll-out phase to speed up the exploration, based on ChemTS[33]) trained on chemical data is set up to suggest ligand candidates with up to 12 non-hydrogen atoms (see the reduction of the model loss during its training in Supplementary Fig. 3). The candidates are quickly evaluated via their computed binding affinity $\Delta G^{(1)}_{bind}$, given as:

$$\Delta G^{(1)}_{bind} = G[\text{ADA @}\mathbf{CB7} + 1\text{ ligand}] - G[\text{ADA@CB7}] - G[1\text{ ligand}], \quad (1)$$

with the **CB7** portal thanks to the cheap and accurate GFN2-xTB tight-binding method[35]. This binding affinity is used to improve the algorithm's subsequent guesses as part of a fitness score $J(S)$ written as,

$$J(S) = (\text{portal's occluded fraction})^2 * \Delta G^{(1)}_{bind} \quad (2)$$

Eventually, strong ligand candidates resulting from this iterative process and their commercially available derivatives have their

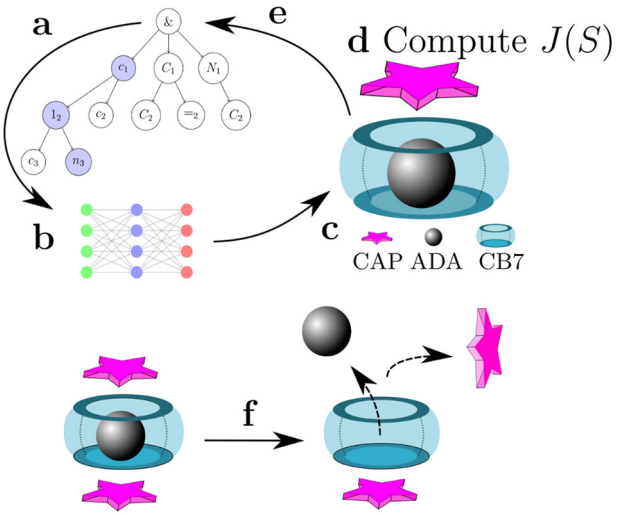


**Fig. 1 Overview of the design procedure for** CB7 **strong ligands.** Overview of the design procedure for **CB7** strong ligands. In (**a**), the Monte Carlo Tree Search generative algorithm suggests fragments of ligand candidates in the SMILES format. This fragment of SMILES is expanded in (**b**) thanks to a recurrent neural network until a complete SMILES candidate is obtained. If the SMILES candidate is deemed valid by the RDKit in (**c**), a 3D structure of the ligand candidate (CAP) is generated. In (**d**), the ligand candidate is docked on **CB7** with adamantanone (ADA) as an inclusion guest, then the ternary complex is optimized and its binding free energy is computed. In step (**d**), the fraction of the portal occluded by the ligand is also estimated to obtain a fitness score $J(S)$ (see Supplementary Fig. 6 and Supplementary Methods 2.3 for details). Using the fitness score of the ligand candidate, the tree nodes are updated in (**e**) and a new iteration of the tree exploration can begin. After several iterations of the **a** → **e** procedure, promising candidates are further evaluated in (**f**) using infrequent metadynamics, the ligand residence time $t_{1/2}$ is estimated, with stronger ligands expected to show a longer residence time.

average residence time $t_{1/2}$ computed through biased molecular dynamics and are finally evaluated experimentally. Additional details on the computation of the fitness function is provided in Supplementary Methods 2.3 and illustrated in Supplementary Fig. 6. The structure of the [ADA@**CB7**] binary complex is illustrated in Supplementary Fig. 5 while the ability of ADA to enter **CB7**'s cavity is shown in Supplementary Fig. 10 through $^1$H NMR data. The ADA guest is chosen as it almost completely fills the **CB7**'s cavity and therefore favors ligand interactions with the **CB7**'s portals instead of the inner cavity.

## Linear free-energy relationship

Linear free-energy relationships (LFER) connect the equilibrium constants and reaction rates of related reactions. As such, a LFER can be seen as a correlation between thermodynamics and kinetics which, while empirical in nature, has found useful applications such as in the Hammet equation[36] or the Brønsted catalysis equation[37]. LFER have been described in the context of supramolecular host-guest binding before[38], through the relation between the number of salt bridges to a host-guest binding affinity but a LFER between a host-guest binding kinetics and the binding affinity of a separate constituent of the system (the ligand in the present case) has, to the best of our knowledge, not been reported.

Here we highlight a LFER between the binding affinity of a ligand with **CB7**'s portal and the decomplexation kinetics of an [ADA@**CB7**] complex on which two identical ligands are bonded on **CB7**'s portals. The [ADA@**CB7**+2 ligands] quaternary complex is seen as a simple model to probe gating processes involving **CB7**. In this case, a modulation of the ADA release time ($t_{1/2}$) from an ultrashort channel, consisting of the **CB7** cavity only, is used as a metric to quantify the 'gating power' of a ligand on the portals of **CB7**. Intuitively, a correlation between $\Delta G^{(1)}_{bind}$ (from equation (1)) and log $t_{1/2}$ is expected as a large guest such as ADA cannot escape **CB7**'s cavity if both of its portals are occluded by ligands. Such a relationship can help power a search of the chemical space to spot strong ligands as $\Delta G^{(1)}_{bind}$ is much cheaper to evaluate computationally than $t_{1/2}$. The correlation between $t_{1/2}$ (representing the kinetics) and $\Delta G^{(full)}_{bind}$ (representing the thermodynamics), which is written as,

$$\Delta G^{(full)}_{bind} = G[\text{ADA @}\mathbf{CB7} + 2\text{ ligands}] - G[\text{ADA}] - G[\mathbf{CB7}] - 2 * G[1\text{ ligand}], \quad (3)$$

is evaluated on a selection of organic cations. Indeed, experimental reports suggest that organic cations such as guanidinium[39] and ammonium[40] are able to interact with **CB7**'s portal, likely through electrostatic interaction with the electron-rich carbonyl groups of **CB7**'s portals. Therefore, 10 organic cations were selected from the ChEBI[41] database for benchmarking purposes. Table 1 shows the log $t_{1/2}$, $\Delta G^{(full)}_{bind}$ and $\Delta G^{(1)}_{bind}$ values of the manually selected ligands. The binding affinity of ADA inside the capped cavity, $\Delta G^{(ADA)}_{bind}$, is also given as

$$\Delta G^{(ADA)}_{bind} = G[\text{ADA @}\mathbf{CB7} + 2\text{ ligands}] - G[\mathbf{CB7} + 2\text{ ligands}] - G[\text{ADA}]. \quad (4)$$

It is expected that $t_{1/2}$ is a reliable estimator of a ligand's relative ability to modify a guest's complexation kinetics due to the extensive sampling that takes place during its computation. Given that the evaluation procedure of $t_{1/2}$ does not take into account the discrete nature of water, which is known to be important in the descriptions of systems involving **CB***n*[42], $t_{1/2}$ values are considered to provide relative information on the ligands rather than to be taken as quantitative estimates. In Fig. 2, it can be seen that $t_{1/2}$ correlates relatively well with $\Delta G^{(full)}_{bind}$ which is an

**Table 1.** Overview of log $t_{1/2}$ and $\Delta G^{(x)}_{bind}$ values for the manually selected ligands arranged in increasing order of log $t_{1/2}$ (all $\Delta G^{(x)}_{bind}$ values are given in kcal mol$^{-1}$).

| Ligand | | CAS n° | log $t_{1/2}$ (s) | $\Delta G^{(ADA)}_{bind}$ | $\Delta G^{(full)}_{bind}$ | $\Delta G^{(1)}_{bind}$ |
|---|---|---|---|---|---|---|
| **1** | Guanidinium | 25215-10-5 | −5.20 | −7.77 | −11.32 | −1.43 |
| **2** | Pyridinium | 628-13-7 | −4.85 | −11.35 | −17.65 | −2.83 |
| **3** | No ligand | N/A | −4.79 | −15.58 | −15.58 | N/A |
| **4** | Ammonium | 12125-02-09 | −4.63 | −13.79 | −11.94 | −0.80 |
| **5** | Tetramethylammonium | 75-57-0 | −4.52 | −6.50 | −12.65 | −0.88 |
| **6** | [Melamine+H]$^+$ | 108-78-1 | −4.46 | −13.68 | −16.27 | −2.56 |
| **7** | Pyrylium | 289-67-8 | −4.32 | −9.81 | −16.53 | −3.36 |
| **8** | Tetrazolium | 298-96-4 | −4.03 | −6.67 | −25.96 | −5.78 |
| **9** | Triphenylsulfonium | 18393-55-0 | −3.99 | −12.61 | −12.95 | −0.72 |
| **10** | Trytilium | 341-02-06 | −3.92 | −10.76 | −19.20 | −2.97 |
| **11** | Benzenediazonium | 100-34-5 | −3.64 | −11.35 | −23.12 | −5.58 |

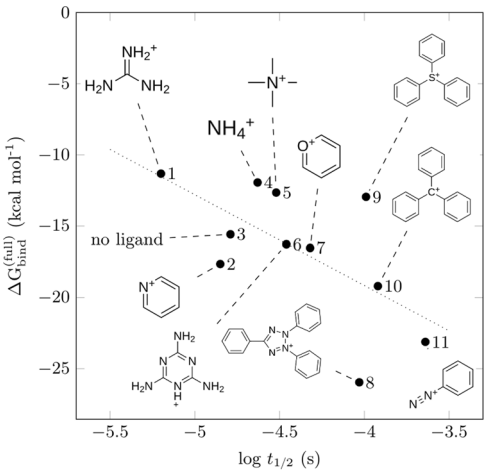


**Fig. 2 Relationship between log $t_{1/2}$ and $\Delta G^{(full)}_{bind}$ for selected organic cations.** Relationship between log $t_{1/2}$ and $\Delta G^{(full)}_{bind}$ for a set of manually selected organic cations which are expected to strongly bind to **CB7**'s portal. The dashed trend line has an equation of $f(x) = -6.36x - 44.6$ while the data points' labels are given in Table 1. The $R^2$ value is 0.408.

encouraging result. Notably, a direct evaluation of $\Delta G^{(full)}_{bind}$ might take 20-30 minutes on a single CPU using the tight-binding method GFN2-xTB[43] while the evaluation of $t_{1/2}$ using the same method and infrequent metadynamics may take hours or days on 50 CPU. In addition, $\Delta G^{(full)}_{bind}$ correlates well with $\Delta G^{(1)}_{bind}$ as shown in Supplementary Fig. 1, which allows to estimate $t_{1/2}$ through $\Delta G^{(1)}_{bind}$ and further reduces the computation time per ligand evaluated.

Interestingly, there does not seem to be a correlation between $\Delta G^{(ADA)}_{bind}$ and $t_{1/2}$ as shown in Supplementary Figure 2. All ligands tend to lower the binding affinity of ADA ($\Delta G^{(ADA)}_{bind}$) inside the cavity even though most are seen to increase ADA's residence time in the cavity as shown in Fig. 2. They are therefore expected to extend $t_{1/2}$ purely through a kinetic effect by providing an additional energy barrier to the escape event rather than by stabilizing ADA in the cavity, which is consistent with experimental observations of a **CB5** derivative capped by metal cations[44]. This kinetic barrier effect can be understood as an extension of the portal's 'constrictive binding' often described in **CB7**[45].

The control case of adamantanone with no ligands (labeled as 'no ligand') appears with an estimated average residence time of 16 µs. This is larger than that estimated for xylene in **CB7** using a similar procedure[13], which is consistent with the fact that adamantanone is a bulkier guest for **CB7** (68% packing coefficient in **CB7** vs 57% for p-xylene). It is worth noting that guanidinium and pyridinium capping of the [ADA@**CB7**] complex yields a lower $t_{1/2}$ than in the uncapped case. This is unsurprising given that while guanidinium was successfully used in the gas phase to trap benzene-sized guests in **CB7**, slightly bigger guests like xylene were not observed due to size exclusion. As ADA is large, it is likely that it clashes with the tight-fitting guanidinium ligands which would therefore undermine the complex kinetic stability. Confident in the ability of $\Delta G^{(1)}_{bind}$ to spot strong ligands, the Monte Carlo Tree Search is initiated using a fitness function that takes $\Delta G^{(1)}_{bind}$ into account.

### Hunting for strong ligands using Monte Carlo Tree Search

The Monte Carlo Tree Search procedure is run 20 times in parallel, each using 4 cores for 3 days and evaluated a total of 5568 molecules. Additional details on the MCTS implementation are provided in Supplementary Methods 2.2 and 2.3. We chose a binding free energy threshold of −20 kcal mol$^{-1}$ to isolate the top ligand candidates, in light of the very stable (in the fM range) biotin/avidin complex[46] itself possessing a binding affinity of −20.4 kcal mol$^{-1}$ and the ~2 kcal mol$^{-1}$ mean absolute error in binding energy expected from the GFN2-xTB[43] method. Such threshold should ensure only very strong ligands qualify and are further considered. Of the 122 molecules passing the selection criterion, only 48 were chemically stable and possessed at most one positive charge. Computations of binding affinities however can be subject to substantial inaccuracies and binding affinities of ligands larger than −20 kcal mol$^{-1}$ are probably overestimates. Nevertheless, with the GFN2-xTB model used, the strong neutral guest amantadine (SMILES: C1C2CC3CC1CC(C2)(C3)N), has a predicted binding affinity of −11.14 kcal mol$^{-1}$ ($K_d$ = 7.5 nM) for **CB7**'s inner cavity, qualitatively comparable with the experimental value of −15.9 kcal mol$^{-1}$[15], which supports the view that GFN2-xTB is able to spot strong interactions. As can be seen in Supplementary Fig. 4, several candidates such as the numerous phosphonium derivatives are either unstable in water or unlikely to be synthetically available. Additional discussions on the manual sorting of the top-ranked ligands suggested by the MCTS are provided in Supplementary Discussion 3.1. Common to almost all of the top-ranked candidates, is a presence of halogens and in particular iodine and bromine. Benzene rings extensively

**Table 2.** Overview of log $t_{1/2}$ and $\Delta G_{bind}^{(x)}$ values for a selection of the best ligands suggested by the MCTS (All $\Delta G_{bind}^{(x)}$ values are given in kcal mol$^{-1}$ and the compounds are ranked by increasing $t_{1/2}$).

| Ligand | | CAS n° | log $t_{1/2}$ (s) | $\Delta G_{bind}^{(ADA)}$ | $\Delta G_{bind}^{(full)}$ | $\Delta G_{bind}^{(1)}$ |
|---|---|---|---|---|---|---|
| **3** | No ligand | N/A | −4.79 | −15.58 | −15.58 | N/A |
| **12** | Iodobenzene | 591-50-4 | −4.18 | −11.59 | −19.35 | −2.11 |
| **13** | 1h-imidazole-2-diazonium | 53485-21-5 | −3.89 | −12.25 | −26.29 | −8.02 |
| **14** | 1,3 diiodobenzene | 626-00-6 | −3.64 | −9.38 | −31.10 | −8.13 |
| **15** | Fluoropyridinium | 140623-89-8 | −3.63 | −10.81 | −29.49 | −8.75 |
| **16** | 1,2,3-benzodithiazol-1-ium | 273-81-4 | −3.60 | −14.00 | −29.66 | −9.59 |
| **17** | 1,4 diiodobenzene | 624-38-4 | −3.21 | −10.46 | −27.31 | −6.44 |
| **18** | 1,2 diiodobenzene | 615-42-9 | −3.21 | −10.21 | −33.21 | −9.28 |
| **19** | 1,2,4 trioodobenzene | 615-68-9 | −3.04 | −8.73 | −39.28 | −13.58 |
| **20** | 1,3,5 triiodobenzene | 626-44-8 | −3.03 | −9.77 | −37.00 | −11.34 |
| **21** | 1,2,3 triiodobenzene | 608-29-7 | −2.71 | −8.33 | −48.20 | −17.29 |
| **22** | 1,2,4,5 tetraiodobenzene | 636-31-7 | −2.01 | −10.48 | −52.69 | −19.87 |
| **23** | 1,2,3,5 tetraiodobenzene | 634-92-4 | −1.78 | −11.75 | −53.91 | −19.97 |
| **24** | 1,2,3,4 tetraiodobenzene | 634-68-4 | −1.34 | −8.31 | −61.07 | −24.78 |
| **25** | Pentabromofluorobenzene | 827-05-4 | −0.83 | −9.86 | −64.69 | −26.83 |
| **26** | Hexabromobenzene | 87-82-1 | −0.40 | −10.10 | −69.30 | −28.23 |
| **27** | Iopamidol | 60166-93-0 | −0.40 | 0.36 | −55.11 | −23.60 |
| **28** | Iodoxamic acid | 31127-82-9 | −0.36 | 20.08 | −28.25 | −1.97 |
| **29** | Acetrizoate | 129-63-5 | 0.32 | −6.95 | −65.98 | −29.41 |
| **30** | Pentaiodobenzene | 608-96-8 | 0.64 | −10.67 | −73.14 | −30.79 |
| **31** | Diatrizoate | 737-31-5 | 1.52 | −22.37 | −89.58 | −32.81 |
| **32** | Hexaiodobenzene | 608-74-2 | 2.55 | −11.88 | −93.94 | −41.85 |

substituted with bromine and iodine make up almost a quarter of the 50 best candidates and half of the 10 best candidates.

Halogenated benzenes likely rank highly as they appear to be a natural fit for the **CB7**'s carbonyl-rich portal as carbonyls are known to be able to form halogen bonds[47]. Iodine-carbonyl bonds, in particular, have been shown to be central to design strongly binding inhibitors for certain enzymes[48,49]. Lower down the ranking, in the −10 to −20 kcal mol$^{-1}$, chlorine and charged nitrogen atoms become more abundant. Diazonium groups also show up with high frequency.

A sample of promising, commercially available, ligands is selected and considered for further analysis using metadynamics. In particular, three organic cations labeled as numbers **13**, **15**, and **16** and three halogenated benzene compounds corresponding to numbers **25**, **26**, and **32**, as recorded in Table 2, are further investigated using metadynamics.

## Investigation of the best leads

The residence time $t_{1/2}$ of the promising ligands suggested by the MCTS (numbers **13**, **15**, **16**, **25**, **26**, and **32** in Table 2) is computed and reveals that the 'gating power' of halogenated ligands dramatically exceeds that of the benchmarking organic cations, sometimes by as much as 7 orders of magnitude.

As **CB7** is water-soluble and water is a non-toxic and popular solvent, the insolubility of the promising halogenated benzenes is not optimal. The PubChem database was therefore queried for other halogenated benzenes that might be water-soluble. The search highlighted a class of X-ray radiocontrast agents using 1,3,5-triiodobenzene (**20**) as their absorbing core which has been optimized for water-solubility and biocompatibility decades ago already[50]. Four water-soluble iodinated contrast agents (**27**, **28**, **29**, and **31**) were further considered and displayed large $t_{1/2}$ values. This large predicted $t_{1/2}$ for water-soluble derivatives of 1,3,5-triiodobenzene is encouraging and might hopefully lead to drug delivery applications in vivo thanks to their modest toxicity[51].

Drawing from the observation that halogen-, and in particular iodine-, substituted benzenes are able to bring about strong ligand-portal interactions, several halogenated benzene rings are further characterized using infrequent metadynamics with the objective to obtain a spectrum of ligand-binding affinities. The results for iodobenzene (**12**), diiodobenzene (**14**, **17**, **18**), triiodobenzene (**19**, **20**, **21**), tetraiodobenzene (**22**, **23**, **24**) and pentaiodobenzene (**30**) are listed in Table 2.

Increasing the number of halogen substituents on a benzene ring tends to lower its aromaticity, decrease its HOMO-LUMO gap and increase its polarizability[52]. Atomic polarizabilities are responsible for dispersion interactions which are known to be important in halogen bonds[53]. As shown in Table 2, periodinated compounds are successful in modulating the predicted host-guest binding kinetics over almost 7 orders of magnitude.

Interestingly, ligands with an explicit negative charge due to a carboxylate group at neutral pH such as diatrizoate (**31**) and acetrizoate (**29**) still display a large enhancement in $t_{1/2}$. Following a line of thought similar to our assumption that cationic ligands should bind **CB7**'s portals strongly, negatively charged ligands were expected to show a poorer binding to **CB7**'s portals. In the context of the `alpb` solvation model as implemented by `xtb 6.3.3`, this does not seem to be the case. Iopamidol (**27**) stands out by strongly destabilizing the [ADA@**CB7**] complex (with $\Delta G_{bind}^{(ADA)} > 0$) despite still extending ADA's residence time in the cavity. The 1,3,5-triiodobenzene (**20**) framework itself is outperformed by almost 4 orders of magnitude, as measured by $t_{1/2}$, by the four ligands based on it (**27**, **28**, **29** and **31**). This discrepancy originates in the side chains of the contrast agents as they are rich in groups capable of forming hydrogen bonds with the portals such as alcohols and secondary amines that can be seen forming during the MD trajectory.

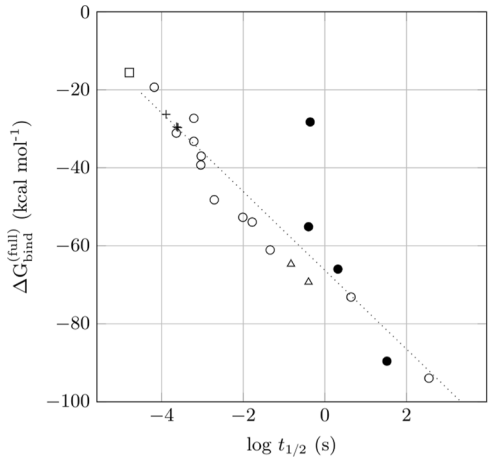


**Fig. 3 Relationship between log $t_{1/2}$ and $\Delta G_{bind}^{(full)}$ for ligands suggested by the MCTS.** Relationship between log $t_{1/2}$ and $\Delta G_{bind}^{(full)}$ for a range of ligands suggested by the MCTS and manually selected derivatives in the case of [ADA@**CB7**]. Organic cations (**13**, **15**, **16**) are labeled '+' while the 'no ligand' reference is labeled □. Periodinated benzenes are labeled ○ while water-soluble commercial derivatives (**27**, **28**, **29**, **31**) are labeled •. Other halogenated benzenes (**25**, **26**) are finally labeled as Δ. There is a strong linear relationship between the free energy $\Delta G_{bind}^{(full)}$ and $t_{1/2}$ that holds over 8 orders of magnitude of $t_{1/2}$. Water-soluble derivatives yield large values of $t_{1/2}$ which is encouraging for water-based applications of the capped **CB7**. The periodinated compounds with one to six substituted iodine atoms show a log $t_{1/2}$ that spans a large −4.18 to 2.55 range which hints at its ability to effectively modulate host-guest decomplexation kinetics. The dashed trend line has an equation of $f(x) = -10.1x - 66.3$. The $R^2$ value is 0.832.

One outlier is iodoxamic acid (**28**) which, despite showing low binding affinities between the ligand and portal, displays a remarkably high guest residence time $t_{1/2}$. This phenomenon is attributed to the structure of iodoxamic acid which consists in two ligands which are, unlike all other ligands, linked together by a molecular chain. This effect is expected to boost the binding affinity as iodoxamic acid behaves as a multivalent ligand where one ligand binding the portal improves the binding affinity of the second ligand[54].

All promising ligands were further displayed in Fig. 3 and shown to follow the hypothesized LFER between $\Delta G_{bind}^{(full)}$ and $t_{1/2}$ over a range of $t_{1/2}$ spanning 7 orders of magnitude. The well-defined LFER in the case of capped **CB7** with a large guest (ADA) begs the question whether such a relationship would hold for other host-guest complexes. We argue that for a ligand to create a kinetic barrier to the decomplexation of a guest, it only needs to bind somewhere on the macrocycle where it hinders the escape path of the guest over the energetically accessible conformational space of the capped host-guest complex. However, the LFER is only expected to become apparent if the different ligands considered bind through a similar mechanism. For example, a smaller guest exiting a hexaiodobenzene (**32**) capped **CB7** would be required to break several halogen bonds to exit the cavity while in the case of a 1,2,3 triiodobenzene (**21**) ligand, the ligand might be able to just shift without requiring a breakup of as many halogen bonds. For this hypothetical small guest, the LFER between $\Delta G_{bind}^{(full)}$ and $t_{1/2}$ might not be apparent. In the case of ADA, the LFER is well-defined as the whole ligand needs to be removed from the portal to allow the bulky guest to exit the cavity and therefore $\Delta G_{bind}^{(full)}$ correlates well with $t_{1/2}$. Interestingly, cyclohexanone, a guest smaller than ADA with a weaker affinity for **CB7** (−11 kcal mol$^{-1}$ at the GFN2-xTB level), still displays the LFER as illustrated in Supplementary Fig. 7.

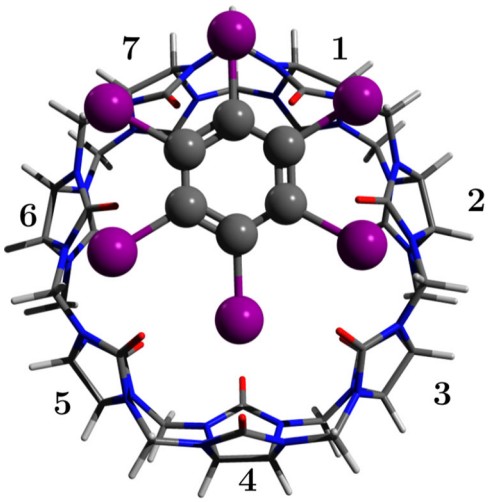


**Fig. 4 Hexaiodobenzene (32) binding to the** CB7**'s portal as optimised using GFN2-xTB.** Hexaiodobenzene (**32**) binding to the **CB7**'s portal as optimised using GFN2-xTB. Glycoluril units are numbered from **1** to **7**. The staggered configuration of the ligand appears to maximise electrostatic interactions of the ligand with the glycoluril unit **4**.

## Halogen bonds

As the strong ligands involve halogens in close proximity to the portal's carbonyls, the presence of halogen bonds as the source of the ligand-portal interaction is checked. Despite the fact that halogen bonds are usually understood as highly directional and linear, there is evidence that halogen bonds can occur over a large range of angles from linear to lower than 90°, including in **CBs**[55].

In the case of halogen bonds involving amide carbonyl groups, $\pi$-orbitals predominantly act as donors in contrast to the more frequent halogen bond involving the oxygen's lone electron pair. These $\pi$-orbitals-driven bonds can display N-C=O···I angles much lower than in the common case where only n-orbitals are involved[53]. For example, iodobenzene (**12**) in water forms an angle C-I···O of 179.3° and an angle N-C=O···I of 123.2° with a single glycoluril unit as optimized at the wB97XD/CPCM/Def2-SVPP level. To explore the possible occurrence of such halogen bonds in the case of the hexaiodobenzene ligand, the N-C=O···I and C-I···O angles of a [ADA@**CB7** + 2 hexaiodobenzene] complex were recorded at every timestep when the O···I distance appeared below 3.5 Å (see Supplementary Figs. 8 and 9) over a molecular dynamics trajectory. Interestingly, a high number of bonds (up to 5 bonds per ligand per snapshot on average) is able to form. The range of both angles closely matches that of O···I bonds found along with carbonyl-bearing amino acids in biological molecules[56]. Indeed the C-I···O distribution shows a significant peak in the 80° to 120° while it peaks at around 75° in the histogram shown in Supplementary Fig. 9. Similarly the N-C=O···I distribution peaks at 140° in the histogram in Supplementary Figure 8 while it peaks at 115° in the reference data. Due to the fact that ligands lying flat on top of **CB7**'s portal would naturally exhibit a 120° angle with the pre-organized portal's carbonyl groups, a ligand rich in halogens and able to be involved in halogen bonds involving $\pi$-orbitals would be expected to bind strongly to the portal while occluding it.

In addition, a Symmetry-adapted perturbation theory (SAPT) energy decomposition on a model system consisting of a single hexaiodobenzene (**32**) ligand docked on **CB7** and optimized at the GFN2-xTB level is shown in Fig. 4. Each glycoluril unit is

**Table 3.** SAPT energy decomposition of the interaction between hexaiodobenzene (**32**) and each of the seven glycoluril units in **CB7** at the sSAPT0/Def2-TZVP level in kcal mol$^{-1}$ (the sum of all energy contributions: electrostatic $E_{elst}$, exchange $E_{exch}$, induction $E_{ind}$ and dispersion $E_{disp}$ is labeled $E_{sSAPT0}$).

| Unit (#) | $E_{elst}$ | $E_{exch}$ | $E_{ind}$ | $E_{disp}$ | $E_{sSAPT0}$ |
|---|---|---|---|---|---|
| **1** | −7.48 | 15.77 | −3.17 | −8.70 | −3.58 |
| **2** | −5.92 | 13.59 | −2.81 | −7.47 | −2.60 |
| **3** | −3.45 | 13.00 | −2.53 | −5.76 | 1.26 |
| **4** | −10.19 | 14.62 | −4.32 | −5.51 | −5.40 |
| **5** | −3.44 | 12.98 | −2.52 | −5.77 | 1.25 |
| **6** | −5.92 | 13.61 | −2.81 | −7.49 | −2.61 |
| **7** | −7.48 | 15.80 | −3.17 | −8.70 | −3.56 |

considered separately by deleting the other units and replacing the four connecting carbon atoms by hydrogen atoms. These newly added four hydrogen atoms are further optimized with all other atomic positions kept frozen. The SAPT energy decomposition is performed using Psi4[57] with sSAPT0 and the Def2-TVZP basis set with RI and JK approximations. Results from Table 3 indicate that all glycoluril units are able to sustain strong dispersion interactions with the polarizable iodine atoms of hexaiodobenzene (**32**). Strikingly, however, there is a large variability in $E_{elst}$ and $E_{ind}$. Electrostatic, induction (under the form of charge transfer), and dispersion interactions are all known to be driving forces for the formation of halogen bonds[53]. It is most interesting that the interaction of glycoluril unit 4 in hexaiodobenzene (**32**) with **CB7**'s portal yields the most negative $E_{sSAPT0}$ of all glycoluril units thanks to strong $E_{elst}$ and $E_{ind}$ interactions which are compatible with the energetics of a halogen bond. For glycoluril unit 4, the angle C-I···O is 163° while the angle N-C=O···I is 139.4°. The staggered configuration of the hexaiodobenzene (**32**) ligand on **CB7**'s portal, therefore, maximizes electrostatic interactions with at least an iodine-oxygen couple while still allowing for dispersion interactions to take place between the other iodine and oxygen atoms. In light of the results of Table 3, the interaction of periodinated benzenes with the carbonyl portal appears driven by both $E_{elst}$ and $E_{disp}$ with the probable formations of halogen bonds.

## Experimental validation

Experiments were carried out with two objectives. The first was to verify the ability of the ligands to modify the host-guest exchange rate by using $^1$H NMR, and the second was to provide evidence that the best ligands predicted by the infrequent metadynamics investigation are able to form quaternary complexes with **CB7** with an encapsulated guest by using DOSY-NMR.

An observation that the ligands are able to induce a fast to slow shift in host-guest exchange kinetics on the $^1$H NMR timescale would be a strong clue that they are able to influence the ingression and egression kinetics of a guest, presumably by binding to **CB7**'s portal. The ADA guest used in the in silico ligand exploration is a known strong guest of **CB7** exhibiting slow exchange kinetics (as shown in Supplementary Fig. 21) and is therefore inadequate to highlight a fast to slow shift in host-guest exchange kinetics.

Cyclohexanone is a small molecule able to form inclusion complexes with **CB7** with a fast exchange kinetics, with respect to the measurement time scale of a 500 MHz instrument, as do other cyclohexane derivatives[58]. At the bottom of Supplementary Fig. 11, it can be seen that the $^1$H NMR peaks of free cyclohexanone in solution at 1.6, 1.75, and 2.25 ppm shift to 0.85, 0.95, and 1.4 ppm upon addition of **CB7**, exhibiting a fast host-guest exchange kinetics on the 500 MHz $^1$H NMR timescale. When ligands such as hexabromobenzene, hexaiodobenzene, iopamidol, diatrizoate, and acetrizoate are added to the [cyclohexanone@**CB7**] solution, the host-guest exchange kinetics switches to slow, in 500 MHz $^1$H NMR time scale, as the ligands hinder the ingression and egression of the cyclohexanone guest (see Supplementary Fig. 11). Under slow host-guest exchange kinetics, the $^1$H NMR peaks of both the free and encapsulated cyclohexanone appear on the same spectrum for each ligand probed. The fast to slow exchange kinetics switch is a qualitative indication that the ligands are able to influence the host-guest complexation process. Such modulation of the exchange kinetics has been reported before by the means of pH[30] and salt concentration variations[31] but not using neutral molecular ligands.

DOSY-NMR experiments further reveal the existence of quaternary complexes involving **CB7**, two ligands and a cyclohexanone inclusion guest. In the case of iopamidol (**27**) in Supplementary Figure 20, the ligands' protons clearly appear near the same diffusion coefficient as **CB7** and cyclohexanone at a diffusion coefficient of $1.36 \times 10^{-10}$ m$^2$ s$^{-1}$. This represents almost a 50% reduction in diffusion coefficient going from the [cyclohexanone@**CB7**] binary complex (shown in Supplementary Fig. 17) to the quaternary complex involving two additional iopamidol ligands. The isolated cyclohexanone diffusion coefficient is measured to be $6.36 \times 10^{-10}$ m$^2$ s$^{-1}$ in Supplementary Fig. 13 while the diffusion coefficients of isolated **CB7** and iopamidol are only $2.71 \times 10^{-10}$ m$^2$ s$^{-1}$ as shown in Supplementary Figs. 12 and 16 respectively. The presence of hexaiodobenzene (**32**) and hexabromobenzene (**26**) also causes the diffusion coefficient of the [cyclohexanone@**CB7**] to drop by about 15% as shown in Supplementary Fig. 18 and Supplementary Figure 19 respectively. As neither of these perhalogenated compounds possesses protons, they do not yield $^1$H NMR peaks and do not appear on the DOSY-NMR plots (Supplementary Figs. 14 and 15) and therefore only provide indirect evidence of the formation of a quaternary complex.

## Discussion

A range of ligands able to modulate the guest release kinetics from the cavity of **CB7** thanks to interactions with the portals of the macrocycle is reported. The strong ligands were designed using an efficient exploration of the chemical space through the Monte Carlo Tree Search generative model which is able to improve its suggestions in a closed-loop, iterative process. The search is enabled by the discovery of a Linear Free-Energy Relationship between the computationally cheap binding affinity of a ligand and its ability to modulate the residence time of a guest inside of the cavity of **CB7**. This residence time, quantifying the host-guest decomplexation kinetics, is itself used as a proxy for the 'gating power' of a ligand. Among the ligands investigated, organic cations were found to mildly extend or shorten a guest's residence time while periodinated ligands are able to boost it by up to 7 orders of magnitude. The high binding affinity of the perhalogenated ligands with **CB7**'s carbonyl-rich portals is found to originate from strong halogen bonds with a non-linear geometry involving the $\pi$-orbital of the carbonyl groups. The best ligands are shown to experimentally form quaternary complexes with **CB7** and a guest, hinting at their 'gating power'. Their ability to influence the host-guest exchange dynamics was confirmed using $^1$H NMR and DOSY-NMR. It is expected that these ligands will provide flexible tools for the study of ion fluxes in synthetic, **CB7**-based, ionic channels. In addition, the [**CB7**+ 2 ligands] systems studied here might be of interest in the context of drug delivery and nanocontainers design. These encouraging results obtained in the context of a moderate-size supramolecular system shed light on the exciting potential of generative design models to assist the construction of other supramolecular gating systems.

## METHODS

### Estimation of a guest's decomplexation kinetics

The ability of a ligand to act as a capping agent for **CB7** is investigated using metadynamics[59] as implemented by `xtb 6.3.3`[35]. The ligands are docked at both portals and optimized with a convergence criteria of $10^{-6}$ Ha using the `alpb` implicit solvation model for water. The RMSD of **CB7** and adamantanone is chosen as the default collective variable while the ligands are not included. The system is then propagated in time for up to 1 ns in 30 independent trajectories. The mass of the hydrogen atoms is set to 4 u and the time step is set to 4 fs. Only covalent bonds involving hydrogen atoms are maintained using the SHAKE algorithm. The temperature is set to 298 K and controlled using `xtb`'s default thermostat. The $k_{\mathrm{push}}$ parameter is set to 0.05 Ha while the Gaussian width $\alpha$ is set to 0.6 $\mathrm{Bohr}^{-1}$. An estimate of the unbiased time to dissociation eventually built from the distribution of biased escape times using infrequent metadynamics[60]. Additional details are given in Supplementary Methods 2.1.

### Monte Carlo Tree Search

This model is adapted from ChemTS[33]. ChemTS is retrained to generate small molecules by retraining its roll-out neural network on molecules with at most 12 non-hydrogen atoms and a maximum SMILES length of 31 characters. To serve as training dataset, the first 250.000 SMILES from the PubChem database with at most 12 heavy atoms from the set of elements C, N, O, S, P, F, Cl, Br, and I are chosen. Entries with isotopic information were discarded before inclusion as well as entries containing SMILES characters not considered in the original implementation of ChemTS ('\n', '&', 'C', '(', ')', 'c', '1', '2', 'o', '=', 'O', 'N', '3', 'F', '[C@@H]', 'n', '-', '#', 'S', 'Cl', '[O-]', '[C@H]', '[NH+]', '[C@]', 's', 'Br', '/', '[nH]', '[NH3+]', '4', '[NH2+]', '[C@@]', '[N+]', '[nH+]', '\\', '[S@]', '5', '[N-]', '[n+]', '[S@@]', '[S-]', '6', '7', 'I', '[n-]', 'P', '[OH+]', '[NH-]', '[P@@H]', '[P@@]', '[PH2]', '[P@]', '[P+]', '[S+]', '[o+]', '[CH2-]', '[CH-]', '[SH+]', '[O+]', '[s+]', '[PH+]', '[PH]', '8', '[S@@+]'). The SMILES were further canonicalized using Open Babel[61]. The rest of the code was adapted to work with a SMILES length of 31 characters.

The number of Gated Recurrent Units was set to 512 while the learning rate was set to 0.001, the network was trained for 10 epochs with a batch size of 512 and the training loss can be seen in Supplementary Fig. 3. Character prediction accuracy reaches 83.8% for the training set and 82.8% for the validation set. Total training time for the 10 epochs was 4 h on 10 cores from a Intel(R) Xeon(R) Gold 6140 node (2.3 GHz).

The rewards function for a molecule $S$ with a score $J(S)$ as computed by the fitness function of the original implementation of ChemTS which is written as:

$$r(S) = \begin{cases} \frac{-aJ(S)}{1+a|J(S)|} & \text{valid SMILES} \\ -1 & \text{otherwise} \end{cases} \quad (5)$$

where $a = 0.8$. Given that the fitness function proposed here tends to return values in the $[-50;0]$ range which is larger than the default ChemTS one, the parameter $a$ is set to 0.05. Note that since high docking scores, which are taken into account in the fitness function, are negative, a strongly binding molecule will lead to a positive reward. Additional details on the SMILES preprocessing and calculation of a ligand's fitness score $J(S)$ can be found in Supplementary Methods 2.2 and Supplementary Methods 2.3.

### Experimental details

Cyclohexanone was purchased from Sigma-Aldrich with 99% purity. Hexaiodobenzene and hexabromobenzene were purchased from Sigma-Aldrich with 98% purity. Iopamidol was purchased from Sigma-Aldrich with a 99.9% purity. Sodium acetrizoate hydrate and sodium diatrizoate hydrate were purchased from Sigma-Aldrich. All reagents were used as received without further purification steps. **CB7** was synthesized by our group using the classical method developed by Day[62] and Kim[63]. Both $^1$H NMR and DOSY-NMR studies were carried out on a 500 MHz JNM-ECZR Research NMR Spectrometer by JEOL. For fast to slow host-guest exchange kinetics studies using $^1$H NMR, a 6:1:6 cyclohexanone:**CB7**:ligand ratio was used while a 1:1:2 cyclohexanone:**CB7**:ligand ratio was used for all DOSY-NMR studies. For DOSY-NMR, the temperature was set to a constant of 25° C, and sample rotation was disconnected to aid reproducible scanning of the sample at a different pulsed-field gradient strength ranging from 0.3 $\mathrm{mT\,m^{-1}}$ to 0.5 $\mu\mathrm{T\,m^{-1}}$.

## DATA AVAILABILITY

The SMILES used to train the recurrent neural network are provided in the `data` folder in the public Github repository https://github.com/sozenoid/find-ligand-for-CB7.git while the weights of the trained model are located in the `RNN-model` folder of the same repository. Additional data can be obtained from the corresponding author upon reasonable request.

## CODE AVAILABILITY

The scripts used in this manuscript to run the Monte-Carlo Tree Search procedure as well as to evaluate the fitness function are available in the `cap_design_for_CB7_XTB` folder on Github at https://github.com/sozenoid/find-ligand-for-CB7.git in a public repository.

Received: 25 July 2021; Accepted: 22 December 2021;
Published online: 27 January 2022

## REFERENCES

1. Cho, J. W., Lee, J.-S. & Jena, B. P. Conformation states of the neuronal porosome complex. *Cell Biol. Int.* **34**, 1129–1132 (2010).
2. Kovari, L. C. et al. X-ray solution structure of the native neuronal porosome-synaptic vesicle complex: implication in neurotransmitter release. *Micron* **56**, 37–43 (2014).
3. Alexander, S., Mathie, A. & Peters, J. Ion channels. *Brit. J. Pharmacol.* **164**, S137–S174 (2011).
4. Karplus, M., McCammon, J. A. & Peticolas, W. L. The internal dynamics of globular protein. *Crit. Rev. Biochem.* **9**, 293–349 (1981).
5. Tóth, K., Sedlák, E., Sprinzl, M. & Žoldák, G. Flexibility and enzyme activity of NADH oxidase from thermus thermophilus in the presence of monovalent cations of hofmeister series. *BBA-Proteins Proteom.* **1784**, 789–795 (2008).
6. Gora, A., Brezovsky, J. & Damborsky, J. Gates of enzymes. *Chem. Rev.* **113**, 5871–5923 (2013).
7. Rekharsky, M. V. et al. A synthetic host-guest system achieves avidin-biotin affinity by overcoming enthalpy–entropy compensation. *Proc. Natl Acad. Sci. USA* **104**, 20737–20742 (2007).
8. Lambert, H., Mohan, N. & Lee, T.-C. Ultrahigh binding affinity of a hydrocarbon guest inside cucurbit[7]uril enhanced by strong host–guest charge matching. *Phys. Chem. Chem. Phys.* **21**, 14521–14529 (2019).
9. Mohanty, J. & Nau, W. M. Ultrastable rhodamine with cucurbituril. *Angew. Chem. Int. Edit.* **117**, 3816–3820 (2005).
10. Montes-Navajas, P. & Garcia, H. Cucurbituril complexation enhances intersystem crossing and triplet lifetime of 2, 4, 6-triphenylpyrylium ion. *J. Phys. Chem. C* **114**, 2034–2038 (2010).
11. Mock, W., Irra, T., Wepsiec, J. & Manimaran, T. Cycloaddition induced by cucurbituril. a case of pauling principle catalysis. *J. Org. Chem.* **48**, 3619–3620 (1983).
12. Lee, T.-C. et al. Chemistry inside molecular containers in the gas phase. *Nat. Chem.* **5**, 376–382 (2013).
13. Lambert, H., Zhang, Y.-W. & Lee, T.-C. Supramolecular catalysis of m-xylene isomerization by cucurbiturils: transition state stabilization, vibrational coupling, and dynamic binding equilibrium. *J. Phys. Chem. C* **124**, 11469–11479 (2020).
14. Assaf, K. I. & Nau, W. M. Cucurbiturils: from synthesis to high-affinity binding and catalysis. *Chem. Soc. Rev.* **44**, 394–418 (2015).
15. Barrow, S. J., Kasera, S., Rowland, M. J., del Barrio, J. & Scherman, O. A. Cucurbituril-based molecular recognition. *Chem. Rev.* **115**, 12320–12406 (2015).
16. Gürbüz, S., Idris, M. & Tuncel, D. Cucurbituril-based supramolecular engineered nanostructured materials. *Org. Biomol. Chem.* **13**, 330–347 (2015).
17. Appel, E. A. et al. Ultrahigh-water-content supramolecular hydrogels exhibiting multistimuli responsiveness. *J. Am. Chem. Soc.* **134**, 11767–11773 (2012).
18. Loh, X. J. et al. Triply triggered doxorubicin release from supramolecular nanocontainers. *Biomacromolecules* **13**, 84–91 (2012).
19. Jiao, D. et al. Supramolecular peptide amphiphile vesicles through host–guest complexation. *Angew. Chem. Int. Edit.* **51**, 9633–9637 (2012).
20. Lee, T.-C. & Scherman, O. A. A facile synthesis of dynamic supramolecular aggregates of cucurbit[n]uril (n= 5–8) capped with gold nanoparticles in aqueous media. *Chem.-Eur. J.* **18**, 1628–1633 (2012).
21. Loget, G. et al. Direct visualization of symmetry breaking during janus nanoparticle formation. *Small* **8**, 2698–2703 (2012).
22. Peveler, W. J. et al. Cucurbituril-mediated quantum dot aggregates formed by aqueous self-assembly for sensing applications. *Chem. Commun.* **55**, 5495–5498 (2019).
23. Sinha, M. K., Reany, O., Yefet, M., Botoshansky, M. & Keinan, E. Bistable cucurbituril rotaxanes without stoppers. *Chem.-Eur. J.* **18**, 5589–5605 (2012).
24. Ellis, E., Moorthy, S., Chio, W.-I. K. & Lee, T.-C. Artificial molecular and nanostructures for advanced nanomachinery. *Chem. Commun.* **54**, 4075–4090 (2018).

25. Jeon, Y. J. et al. Artificial ion channel formed by cucurbit[n]uril derivatives with a carbonyl group fringed portal reminiscent of the selectivity filter of K+ channels. *J. Am. Chem. Soc.* **126**, 15944–15945 (2004).
26. Tabushi, I., Kuroda, Y. & Yokota, K. A, B, D, F-tetrasubstituted β-cyclodextrin as artificial channel compound. *Tetrahedron Lett.* **23**, 4601–4604 (1982).
27. Chen, J.-Y. & Hou, J.-L. Controllable synthetic ion channels. *Org. Chem. Front.* **5**, 1728–1736 (2018).
28. Sheng, M. & Pak, D. T. Ligand-gated ion channel interactions with cytoskeletal and signaling proteins. *Annu. Rev. Physiol.* **62**, 755–778 (2000).
29. Muraoka, T. et al. Reversible ion transportation switch by a ligand-gated synthetic supramolecular ion channel. *J. Am. Chem. Soc.* **136**, 15584–15595 (2014).
30. Marquez, C. & Nau, W. M. Two mechanisms of slow host–guest complexation between cucurbit[6]uril and cyclohexylmethylamine: pH-responsive supramolecular kinetics. *Angew. Chem. Int. Edit.* **40**, 3155–3160 (2001).
31. Márquez, C., Hudgins, R. R. & Nau, W. M. Mechanism of host- guest complexation by cucurbituril. *J. Am. Chem. Soc.* **126**, 5806–5816 (2004).
32. Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **361**, 360–365 (2018).
33. Yang, X., Zhang, J., Yoshizoe, K., Terayama, K. & Tsuda, K. Chemts: an efficient python library for de novo molecular generation. *Sci. Technol. Adv. Mat.* **18**, 972–976 (2017).
34. Bernstein, N., Csányi, G. & Deringer, V. L. De novo exploration and self-guided learning of potential-energy surfaces. *Npj Comput. Mater.* **5**, 1–9 (2019).
35. Grimme, S. Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations. *J. Chem. Theory. Comput.* **15**, 2847–2862 (2019).
36. Hammett, L. P. The effect of structure upon the reactions of organic compounds. benzene derivatives. *J. Am. Chem. Soc.* **59**, 96–103 (1937).
37. Bronsted, J. Acid and basic catalysis. *Chem. Rev.* **5**, 231–338 (1928).
38. Schneider, H.-J. Linear free energy relationships and pairwise interactions in supramolecular chemistry. *Chem. Soc. Rev.* **23**, 227–234 (1994).
39. Yang, F. & Dearden, D. V. Guanidinium-capped cucurbit[7]uril molecular cages in the gas phase. *Supramol. Chem.* **23**, 53–58 (2011).
40. Hwang, I. et al. Cucurbit[7]uril: a simple macrocyclic, ph-triggered hydrogelator exhibiting guest-induced stimuli-responsive behavior. *Angew. Chem. Int. Edit.* **119**, 214–217 (2007).
41. Degtyarenko, K. et al. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res.* **36**, D344–D350 (2007).
42. Nau, W. M., Florea, M. & Assaf, K. I. Deep inside cucurbiturils: physical properties and volumes of their inner cavity determine the hydrophobic driving force for host–guest complexation. *Isr. J. Chem.* **51**, 559–577 (2011).
43. Bannwarth, C., Ehlert, S. & Grimme, S. GFN2-xTB–An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory. Comput.* **15**, 1652–1671 (2019).
44. Hickenlooper, S. M., Harper, C. C., Pope, B. L., Mortensen, D. N. & Dearden, D. V. Barriers for extrusion of a guest from the interior binding cavity of a host: Gas phase experimental and computational results for ion-capped decamethylcucurbit [5]uril complexes. *J. Phys. Chem. A* **122**, 9224–9232 (2018).
45. Miskolczy, Z. & Biczók, L. Kinetics and thermodynamics of berberine inclusion in cucurbit[7]uril. *J. Phys. Chem. B* **118**, 2499–2505 (2014).
46. General, I. J., Dragomirova, R. & Meirovitch, H. Absolute free energy of binding of avidin/biotin, revisited. *J. Phys. Chem. B* **116**, 6628–6636 (2012).
47. Bissantz, C., Kuhn, B. & Stahl, M. A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **53**, 5061–5084 (2010).
48. Parks, D. J. et al. 1, 4-benzodiazepine-2, 5-diones as small molecule antagonists of the HDM2–p53 interaction: discovery and sar. *Bioorg. Med. Chem. Lett.* **15**, 765–770 (2005).
49. Maillard, M. C. et al. Design, synthesis, and crystal structure of hydroxyethyl secondary amine-based peptidomimetic inhibitors of human β-secretase. *J. Med. Chem.* **50**, 776–781 (2007).
50. Pitre, D. & Felder, E. Development, chemistry, and physical properties of iopamidol and its analogues. *Invest. Radiol.* **15**, S301–9 (1980).
51. Pasternak, J. J. & Williamson, E. E. Clinical pharmacology, uses, and adverse reactions of iodinated contrast agents: a primer for the non-radiologist. *Mayo. Clin. Proc.* **87**, 390–402 (2012).
52. Jezowski, S. R., Baer, R., Monaco, S., Mora-Perez, C. A. & Schatschneider, B. Unlocking the electronic genome of halogenobenzenes. *Phys. Chem. Chem. Phys.* **19**, 4093–4103 (2017).
53. Cavallo, G. et al. The halogen bond. *Chem. Rev.* **116**, 2478–2601 (2016).
54. Zhou, H.-X. & Gilson, M. K. Theory of free energy and entropy in noncovalent binding. *Chem. Rev.* **109**, 4092–4107 (2009).
55. El-Sheshtawy, H. S., Bassil, B. S., Assaf, K. I., Kortz, U. & Nau, W. M. Halogen bonding inside a molecular container. *J. Am. Chem. Soc.* **134**, 19935–19941 (2012).
56. Sirimulla, S., Bailey, J. B., Vegesna, R. & Narayan, M. Halogen interactions in protein–ligand complexes: implications of halogen bonding for rational drug design. *J. Chem. Inf. Model.* **53**, 2781–2791 (2013).
57. Turney, J. M. et al. Psi4: an open-source ab initio electronic structure program. *Wires. Comput. Mol. Sci.* **2**, 556–565 (2012).
58. Yu, J.-S., Wu, F.-G., Tao, L.-F., Luo, J.-J. & Yu, Z.-W. Mechanism of the fast exchange between bound and free guests in cucurbit[7]uril–guest systems. *Phys. Chem. Chem. Phys.* **13**, 3638–3641 (2011).
59. Barducci, A., Bonomi, M. & Parrinello, M. Metadynamics. *Wires. Comput. Mol. Sci.* **1**, 826–843 (2011).
60. Salvalaglio, M., Tiwary, P. & Parrinello, M. Assessing the reliability of the dynamics reconstructed from metadynamics. *J. Chem. Theory. Comput.* **10**, 1420–1425 (2014).
61. O'Boyle, N. M. et al. Open babel: an open chemical toolbox. *J. Cheminformatics* **3**, 33 (2011).
62. Day, A., Arnold, A. P., Blanch, R. J. & Snushall, B. Controlling factors in the synthesis of cucurbituril and its homologues. *J. Org. Chem.* **66**, 8094–8100 (2001).
63. Kim, J. et al. 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**, 540–541 (2000).

## ACKNOWLEDGEMENTS

H.L. and T.-C.L. are grateful to the studentship funded by the A*STAR-UCL Research Attachment Programme through the EPSRC Centre for Doctoral Training in Molecular Modelling and Materials Science (Grant EP/L015862/1). T.-C.L. is grateful to the Research Project Grant (Grant RPG-2016-393) funded by the Leverhulme Trust. We acknowledge the use of the UCL Myriad High Performance Computing Facility (Myriad@UCL), and associated support services, in the completion of this work. This work is partially supported financially by the Agency for Science, Technology and Research (A*STAR) under grant AMDM A1898b0043, and A*STAR SERC CRF Award.

## AUTHOR CONTRIBUTIONS

H.L. conceived the original idea, carried out the MCTS, infrequent metadynamics procedures and computational exploration of the nature of the halogen bonds. A.C. B performed the $^{1}$H NMR and DOSY-NMR experiments. A.C.B., H.L., and T.-C.L. analyzed the NMR data. Q.Z. and T.-C.L. supervised the NMR experiments. Y.-W.Z. and T.-C.L. supervised the project, discussed the data and provided direction. H.L. wrote the manuscript with input from Y.-W.Z. and T.-C.L. All authors read and approved the manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41524-022-00702-0.

**Correspondence** and requests for materials should be addressed to Yong-Wei Zhang or Tung-Chun Lee.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.