Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

ELECTRONIC SUPPLEMENTARY INFORMATION

Energetics and Exchange of Xenon and Water in a Prototypic Cryptophane-A Biosensor Structure

Perttu Hilla and Juha Vaara

perttu.hilla@oulu.fi; juha.vaara@iki.fi

NMR Research Unit
P.O. Box 3000, FI-90014 University of Oulu, Finland

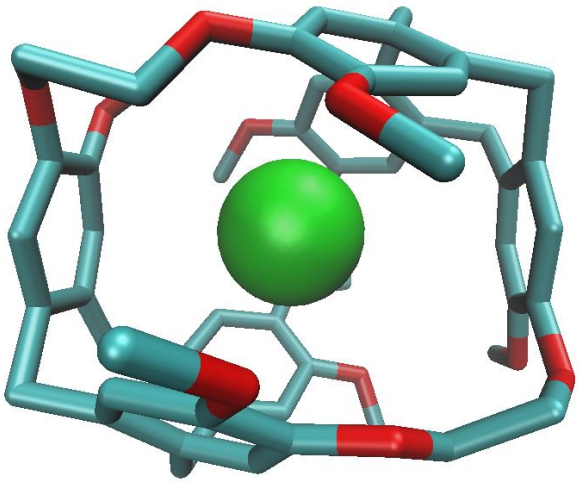

Figure S1: Structure of the Xe@CrA biosensor. The Xe atom dissociates through one of the three portals of the structure: one in the front of the image and two in the back. Hydrogen atoms are not shown for clarity.

Contents

List of Figures

List of Tables

1 System preparation

The initial coordinates for the CrA host were obtained from open-source protein data bank crystallographic data (structure 3CYU) [1, 2] by removing the excess atoms of a Xe biosensor bound to a larger protein. The Packmol [3] code was used to construct a spherical droplet model for each site, as the xTB software does not currently fully support periodic boundary conditions [4]. To approximate the volume of the sites, a polarizability volume calculation was carried out on the Gaussian 16 software [5] for the Xe atom, H_2O molecule and the CrA host (Table S1). CrA and Xe were found to have the volume of *ca.* 62 and 3 water molecules, respectively. The droplet volume was, therefore, increased by the amount corresponding to the volume of these numbers of water molecules (at 300 K), as compared to site (i), to accomodate CrA and Xe. By doing this, the number of water molecules was kept constant, at 500, in each simulation. This is helpful in estimating of ΔA_{Bind} through the thermodynamic cycle, as "background" interaction energy between the water molecules then cancels out. Also, a realistic initial configuration is required because GFN-FF generates the force field based on the initial structure [6]. Prior to the simulations, the site clusters were geometry-optimized using xTB at the GFN2 level of theory. Default parameters were chosen.

Table S1: Quantum-chemically[a] calculated polarizability volumes $\alpha' = \alpha_{\mathrm{ISO}}/4\pi\epsilon_0$ of the species involved in the simulations.

Molecule	α'/a_0^3	$\alpha'/\alpha'(H_2O)$
H_2O	9.69	1
Xe	27.84	≈ 3
CrA	598.01	≈ 62

[a] Geometry optimization followed by polarizability calculation using density-functional theory, PBE0 functional [7], pcseg1 basis set [8] for C, H, O in CrA, aug-pcseg-2 basis set [8] for H_2O, and quasirelativistic ECP28MDF effective core potential [9] with def2-QZVPD valence basis set [9, 10] for Xe.

2 MD simulations

A total of 12 simulations (4 sites, each at 3 levels of theory) were performed at constant temperature and volume, using a Berendsen thermostat [11] at 300 K as the heat bath. The Leapfrog algorithm [12] with 1 fs step was used, and a snapshot was stored every 0.1 ps. Thermal equilibrium was obtained before each production period, as judged from a plateau in the potential energy and disappearance of systematic changes in characteristic radial distribution functions (RDFs). The lengths of production trajectories were 300 ps for GFN2, 1 ns for GFN0, and 8.5 ns for GFN-FF. The SHAKE algorithm [13] for the covalent bonds to hydrogen atoms was used in the GFN0 and GFN-FF simulations.

For the solvent water molecules, a spherical confinement potential with a radius of 30.10 a_0, corresponding to site (iv) (see the main text), was utilized. For both Xe and the CrA host structure, a tighter confinement potential with a radius of 15.05 a_0 was additionally pplied. The rest of the MD-specific parameters were taken as default.

The obtained averages of the total energies and the lengths of the MD production periods of the four simulated sites at the three different levels of theory are shown in Table S2.

Table S2: Averages of the total energies $\langle E_{\text{tot}} \rangle$ and the corresponding errors δ_E (taken as the standard errors of mean) for the GFN-FF, GFN0 and GFN2 levels of theory. The lengths of the MD production runs are also shown. The four data columns denote the four different simulated sites, as in the main text.

		Site			
		(i)	(ii)	(iii)	(iv)
GFN-FF (8.5 ns)	$\langle E_{\text{tot}} \rangle$ (Eh)	-167.4472	-167.4417	-189.4632	-189.4648
	δ_E (Eh)	0.0004	0.0004	0.0008	0.0005
GFN0 (1 ps)	$\langle E_{\text{tot}} \rangle$ (Eh)	-2188.934	-2192.8756	-2371.508	-2375.4664
	δ_E (Eh)	0.001	0.0011	0.003	0.0012
GFN2 (300 ps)	$\langle E_{\text{tot}} \rangle$ (Eh)	-2542.215	-2546.074	-2733.508	-2737.400
	δ_E (Eh)	0.003	0.004	0.005	0.004

3 Error margins

The statistical standard errors of mean (SEMs) of the site energies $\langle E \rangle$ were estimated by the data-halving method [14] at the GFN0 and GFN2 levels. The absolute value of the error of ΔA_{Bind} was calculated by adding up the SEMs of the individual sites. At the GFN-FF level, due to a program artifact in the total energy outputs, the error margin was estimated from the GFN0 error value by assuming a common statistically uncorrelated block length of values for both levels of theory, and scaling using the number of data points of the GFN-FF simulations.

4 MTD simulations

MTD is a method of enhanced sampling [15] that uses a history-dependent bias potential to discourage the simulated system from entering the already-sampled regions of the phase space, parameterized in terms of a collective geometrical variable. In the MTD implementation within the xTB code [16], narrow Gaussian potentials are added to the system as the simulation proceeds. Root-mean-square deviations (RMSD) are used as the collective variables. The RMSD between atoms is measured between a set of reference

structures, which are the previous configurations of the system on its trajectory. The number of reference structures and the set of atoms included in the MTD calculation are determined by the user.

Suitable parameters that define the strength of the bias potential are not evident, as practically no literature is available. After a few runs of careful bias strength tuning, the realm of values that started to produce the rare event was reached. The MTD-specific parameters were mildly (and quite arbitrarily) varied between the simulations. No systematic difference in the bias potential parameters between the simulations in which the rare event occurred, and those in which it did not, was observed.

The RMSD measure was computed between the positions of the Xe atom and either the two O and C atoms of one of the three -O-C=C-O- ethylenedioxy linkers connecting the CTV bowls of the cryptophane host, or the six C atoms of one of the three benzene rings of a single CTV bowl. The parameters used to push the Xe guest out from the CrA host are shown in Table S3.

Table S3: MTD-specific parameters (see Ref. [16]) of the six MTD simulations that produced the Xe dissociation event. In the top row: the temperature T, value of the push-pull parameter k, width of the bias Gaussian α, number of reference structures n, and atoms chosen for the RMSD calculation. The type of dissociation process by which the Xe atom was found to exit the cryptophane-A host, is shown in the last column (see the main text).

MTD simulation number	T (K)	k (mE_h)	α ($1/a_0$)	n	Atoms[a]	Dissociation process
1	300	5	2	10	Linker	*i*
2	300	10	5	10	Linker	*iii*
3	300	10	5	10	Linker	*ii*
4	300	10	10	10	Linker	*ii*
5	300	5	10	100	Benzene	*i*
6	600	5	10	100	Benzene	*ii*

[a] "Benzene" refers to the six carbon atoms of one of the benzene rings of a CTV bowl, and "Linker" to the four atoms of one -O-C=C-O- ethylenedioxy chain connecting the two CTV bowls (Figure S1).

Schematic pathway of processes (i) and (iii), along with a graph that shows the distance of the Xe atom and water molecules from the center of the CrA cage against MTD simulation time, are presented in Figures S2 and S3.

It is noteworthy that in the MTD simulation at 600 K, the CrA host entered the collapsed conformation [17] incapable of hosting further guest molecules. Two water molecules were displaced from the host as a result.

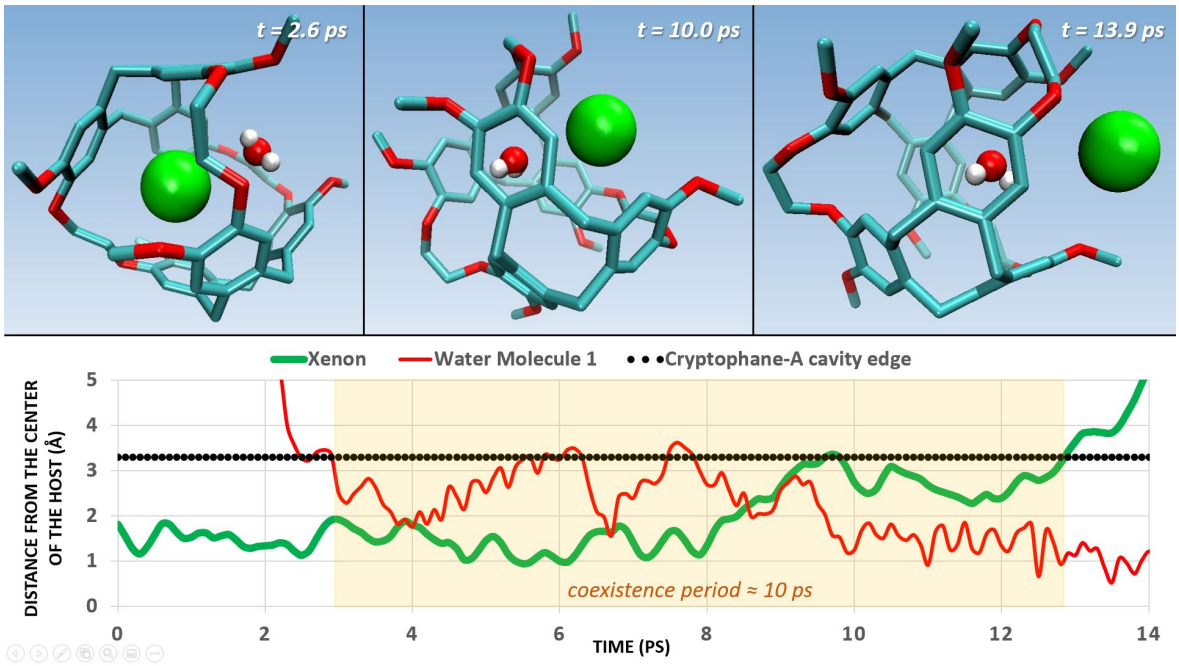

Figure S2: As in Figure 2 of the main text, but for a Xe dissociation event of type (i). Coexistence period of *ca.* 10 ps is highlighted. From the MTD simulation number 1 (Table S3).

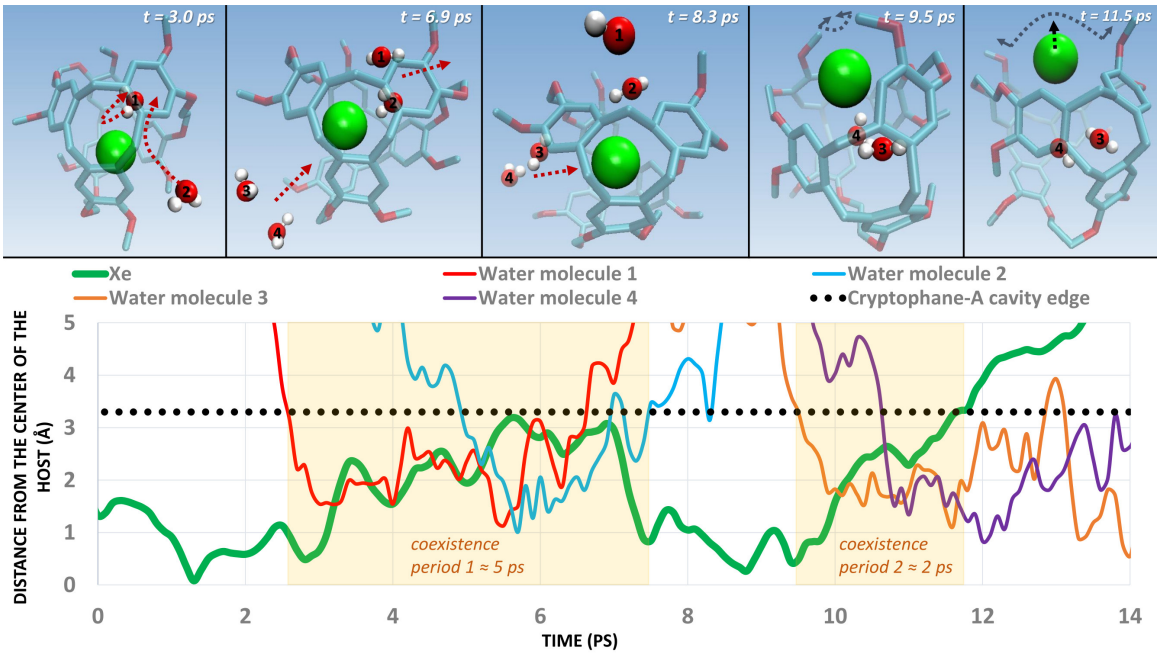

Figure S3: Xe dissociation event of type (iii). Coexistence period 1, of *ca.* 5 ps, and coexistence period 2, of *ca.* 2 ps, are highlighted and the gating mechanism is pointed out. From the MTD simulation number 2 (Table S3).

5 Water dynamics

The distribution of the number of water molecules N_W inside the CrA cavity was calculated by computing the RMS distances between the center of the host and the closest

oxygen atoms of the 500 solvent water molecules as a function of time, and considering water guests within 3.3 Å (*vide infra*) from the center to be encapsulated. An in-house Python script was used for the numerical calculations. The results are shown in Figure S4.

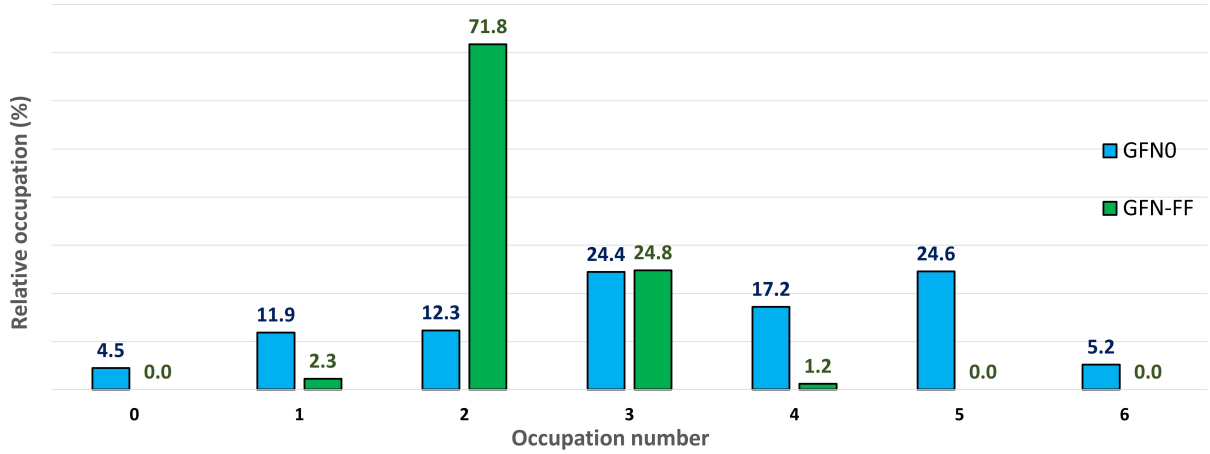

Figure S4: Simulated distribution (in %) of occupation numbers of water molecules in the CrA cage.

We define a survival probability function,

$$P(t) = \sum_{i=1}^{N_{\text{tot}}} \frac{1}{N_{\text{ss}} - m + 1} \sum_{t_0} p_i(t_0, t_0 + t, \delta t), \tag{S4}$$

as was done in Refs. [18–20]. Here, $N_{\text{tot}} = 500$ is the total number of water molecules, N_{ss} is the number of saved configurations (snapshots) and m is the index of the current configuration. The first sum is carried over all water molecules, and the second sum over the times t_0 corresponding to the snapshots.

For numerical evaluation of Eq. (S4), one has to define whether a water molecule i is inside or outside the host cavity. This is implemented through the binary function p_i that takes a value 1 if the i:th water molecule is inside the cavity of the host at both times t_0 and t, and has not left the cavity for a time longer than δt (to allow very brief absences). Otherwise, the value taken by p is 0. For simplicity, the cavity was assumed to have a spherical shape with a constant radius, called the cut-off distance. The values of the cut-off and the parameter δt have to be determined, and we acknowledge that there exists no self-evident choice. The RDFs of the oxygen atoms of water molecules with respect to the cryptophane center were calculated to estimate the cut-off and, hence, the behavior of p_i. The cut-off distance was determined to be 3.3 Å. For comparison, in Ref. [21], the chosen cut-off for various water-soluble CrA derivatives was 4.0 Å. To select δt, the RMS results were used to inspect the distances of the closest water molecules. The maximum time that a water molecule would spend outside the cavity of radius 3.3 Å, and still return to it, was *ca.* 1.0 ps. Additionally, different values for δt were tested, and with values larger than 1.0 ps, the results for $P(t)$ did not vary greatly.

We were interested in the proper exchange of water molecules, rather than faster "peeking" events at the perimeter of the cavity. In practice, we took the natural logarithm

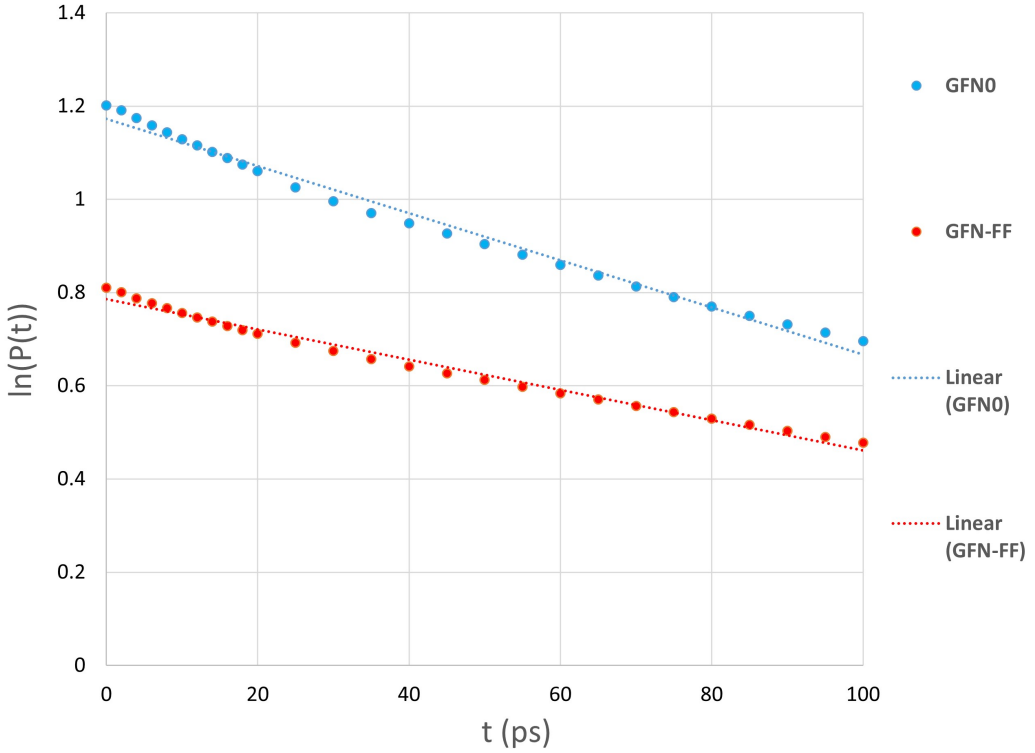

Figure S5: Values of $\ln[P(t)]$, and a linear fit. The GFN0 level of theory is shown in blue, and the GFN-FF level in red.

of $P(t)$ and did a linear fit to the logarithmic values. The inverse of the resulting slope reveals the mean residence time τ. An in-house Python script and the NumPy library [22] were used for the numerical calculation of $P(t)$ and the fits, respectively. The fit and the numerical values are given in Figure S5 and Table S4, respectively.

Table S4: Fitting parameters of the survival probability function $P(t)$. Average number $P(0)$ of water molecules inside the simulated CrA cage, the slope $-1/\tau$ of the linear regression fit to the values $\ln[P(t)]$, and the mean residence time τ, are given.

Level of theory	$P(0)$	$-1/\tau$	τ (ps)
GFN0	3.32	-0.00506 ± 0.00011	198 ± 5
GFN-FF	2.25	-0.00325 ± 0.00008	308 ± 8

References

(1) J. Aaron, J. Chambers, K. Jude, L. Costanzo, I. Dmochowski and D. Christianson, *J. Am. Chem. Soc.*, 2008, **130**, 6942–6943.

(2) *PDB Protein Data Bank*, `https://www.rcsb.org/structure/3CYU`.

(3) *Packmol*, `http://leandro.iqm.unicamp.br/m3g/packmol/home.shtml`.

(4) *GitHub: xTB*, `https://github.com/grimme-lab/xtb`.

(5) M. J. Frisch et al., *Gaussian 16 Revision C.01*, Gaussian Inc. Wallingford CT, 2016.

(6) S. Spicher and S. Grimme, *Angew. Chem. Int. Ed.*, 2020, **59**, 15665–15673.

(7) C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.

(8) F. Jensen, *J. Chem. Theory Comput.*, 2014, **10**, 1074–1085.

(9) K. Peterson, D. Figgen, E. Goll, H. Stoll and M. Dolg, *J. Chem. Phys.*, 2003, **119**, 11113–11123.

(10) D. Rappoport, *J. Chem. Phys.*, 2010, **133**, 134105.

(11) H. Berendsen, J. Postma, W. van Gunsteren, A. DiNola and J. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

(12) W. van Gunsteren and H. Berendsen, *Mol. Simul.*, 1988, **1**, 173–185.

(13) J.-P. Ryckaert, G. Ciccotti and H. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

(14) H. Flyvbjerg and H. Petersen, *J. Chem. Phys.*, 1989, **91**, 461–466.

(15) R. Bernardi, M. Melo and K. Schulten, *Biochim. Biophys. Acta*, 2015, 872–877.

(16) S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 2847–2862.

(17) G. El-Ayle and K. Travis, *In: Comprehensive Supramolecular Chemistry II*, 2017, 199–249.

(18) A. García and L. Stiller, *J. Comput. Chem.*, 1993, **14**, 1396–1406.

(19) C. Rocchi, A. Bizzarri and S. Cannistraro, *Chem. Phys.*, 1997, **214**, 261–176.

(20) A. Luise, M. Falconi and A. Desideri, *Proteins: Struct. Funct. Genet.*, 2000, **39**, 56–67.

(21) L. Gao, W. Liu, O.-S. Lee, I. Dmochowski and J. Saven, *Chem. Sci.*, 2015, **6**, 7238–7248.

(22) C. R. Harris et al., *Nature*, 2020, **585**, 357–362, DOI: `10.1038/s41586-020-2649-2`.