Electronic Supplementary Material (ESI) for ChemComm.


# A guanidino-γ-cyclodextrin superdimer generates a twin receptor for phosphate dimers assembled by anti-electrostatic hydrogen bonds

## Supplementary information

Emmanuel Saridakis, Eleni-Marina Kasimati, Konstantina Yannakopoulou,* Irene M. Mavridis*

*Institute of Nanoscience & Nanotechnology, National Center for Scientific Research "Demokritos" Patr. Grigoriou E' & 27 Neapoleos str., Aghia Paraskevi Attikis 15341, Greece*

**Table of contents**

# Experimental

*1. Materials and methods*

The following reagents were used as received. Native γ-cyclodextrin was a product of CycloLab S.A. (Batch No.: CYL-1815). *N*-bromosuccinimide (NBS, 99%, CAS: 128-08-5) was a product of Merck. *N*,*N*-diisopropylethylamine, (DIPEA, 98%), and diethyl ether (> 99.5%, CAS 60-29-7) were Fluka products. Dry *N*,*N*-dimethylformamide (DMF, anhydrous, 99.8%, CAS 68-12-2), triphenylphosphine (TPP, 99%, CAS: 603-35-0) and benzoylated cellulose dialysis tubing, 32 mm, cut-off MW 2000, was obtained from Sigma-Aldrich. Sodium azide (CAS: 26628-22-8) was a product of Riedel-de Haen. 2-Methyl-2,4-pentanediol (MPD) (CAS: 107-41-5) was purchased from Fluka Chemika and Tris buffer ultrapure from Sigma-Aldrich. Ammonium phosphate, monobasic ($NH_4H_2PO_4$, CAS: 131126) and ammonium sulfate (CAS 7783-20-2) were products of Panreac Quimica SLU. The starting per(6-bromo-6-deoxy)-γCD, was prepared as described previously.[1] All moisture-sensitive reactions were carried out under a nitrogen atmosphere. Thin-layer chromatography was performed with silica gel with fluorescent indicator on aluminum plates purchased from Sigma-Aldrich. All deuterated solvents were products of Deutero GmbH.

1D and 2D NMR spectra were acquired on a 500 MHz on a Bruker Avance NMR spectrometer either in deuterated $D_2O$ or deuterated HCl-borate buffer. $^{13}C$ NMR spectra for the p*K*a experiments were acquired at 62.9 MHz on a Bruker Avance III 250 MHz NMR spectrometer. All spectra were processed with Topspin 4.0.9 software.

Dynamic Light Scattering (DLS) measurements of size (d, nm) were performed using a Zetasizer Nano Series (Malvern Instruments Ltd, Worcestershire, UK) at 25.0 ± 0.1 $^oC$. The wavelength of detection was 632.8 nm at a fixed angle of 173$^o$. The experimental data were processed using the Zetasizer software version 7.11 (Malvern Instruments Ltd).

*2. Synthesis*

The compounds *octakis(6-azido-6-deoxy)-γ-cyclodextrin*[2] and *octakis(6-amino-6-deoxy)-γ-cyclodextrin*[3] were prepared following published procedures.

*Octakis(6-guanidino-6-deoxy)-γ-cyclodextrin hydrochloride (**gguan**):* The compound was prepared according to our previously published procedure[3] with modifications. Briefly, octakis(6-amino-6-deoxy)-γCD hydrochloride (147 mg, 0.11 mmol), was dispersed in dry DMF (1.5 mL) and to the mixture 1*H*-pyrazolecarboxamidine hydrochloride (58 eq., 6.38 mmol, 0.935 g) and DIPEA (40 eq., 4.4 mmol, 0.77 mL) were added. The whole was stirred at 75 $^{o}$C for 24 h under a nitrogen atmosphere and then a second addition of the same quantities of 1*H*-pyrazolecarboxamidine hydrochloride and DIPEA as before, followed. Heating at 70 $^{\circ}$C and stirring continued for a further 4 days under nitrogen. Then, diethyl ether (50 mL) was added dropwise and the suspension formed was stirred for 2 h. The solvent was decanted and the collected sticky solid was dissolved in a very small amount of water (1.5 mL). The pH was adjusted to neutral with aq. HCl (1*N*) and the solution was dialysed for 4 days. The solution was then concentrated under vacuum in order to obtain the desired product as the hydrochloride salt (110 mg, 52%). $^{1}$H NMR (500 MHz, $D_2O$, 300 K) $\delta$ 5.09 (br.s, 7H, H1), 3.97 (t, $J$ = 10 Hz, 7H, H5), 3.85 (t, $J$ = 10 Hz, 7H, H3), 3.63 – 3.55 (m, 14H, H2, H6), 3.47–3.41 (m, 14H, H4, H6) ppm; $^{13}$C NMR (125 MHz, $D_2O$, 300 K) $\delta$ 158.5 (C=), 102.2 (C1), 82.7 (C4), 73.0 (C3), 72.6 (C2), 71.4 (C5), 42.9 (C6) ppm, in agreement with the literature data.[3]

*3. Crystallisation, data collection and crystallographic structure determination*

Solid **gguan** was disolved in distilled water, at a concentration of ca. 8 mM. Crystallisation trials were set up following procedures that are regularly used for the crystallisation of macromolecular samples.[4] Hanging-drop vapour diffusion trials were set up with the commercially available crystallisation screens Crystal Screen and Crystal Screen 2 (Hampton Research, Aliso Viejo CA, USA) in Linbro-type crystallisation trays using siliconised glass coverslips (Hampton Research). For each of the 98 trial conditions of the two screens, 1 μL of **gguan** stock solution (ca. 8 mM) was mixed with 1 μL trial solution on a coverslip, which was then sealed over a reservoir containing 1 mL of the trial solution. Vapour diffusion then proceeded within the airtight droplet-reservoir system, thus slowly reconcentrating the ingredients of the drop. Only one condition (50% (v/v) MPD, 0.1 M Tris pH 8.5, 0.2 M monobasic ammonium phosphate) yielded clusters of microcrystals and was optimised to 17.5%-20% (v/v) MPD, 0.3 M ammonium phosphate and 0.1 M Tris pH 8.5 that gave crystals, diffracting to ca. 1.3 Å. In order to further improve the diffraction limits of the crystals, trials were set up with lower MPD concentrations (where spontaneous crystallisation did not occur) and the use of a heterogeneous

nucleation-inducing agent, a specially designed Bioglass [5-7] (Naomi's Nucleant, Molecular Dimensions Ltd). The lower supersaturation at which crystallisation could proceed with the use of that nucleant led to slower growth and thus larger and better diffracting crystals. After extensive trials, the crystal from which data were collected, to 1.1 Å, was grown on a grain of Bioglass in 15% (v/v) MPD, 0.3 M ammonium phosphate, 0.1 M Tris pH 8.5. Finally, crystallisation trials with ammonium sulfate instead of monobasic ammonium phosphate, using the above conditions and close variations (17, 20, 23 and 30% MPD, at 0.2, 0.3, and 0.5 M ammonium sulfate for each MPD concentration) gave no crystals.

X-ray data were collected at 100 K at the Protein Crystallography Facility, Synchrotron Radiation Source, Daresbury U.K., by the oscillation method. The XDS software[8] was used for data processing and scaling. The unit cell parameters and their esds were determined by the least squares method from the collected data. The structure was solved by direct methods with the SHELXD program,[9] using a di-glucopyranose fragment from γCD as initial model for fragment seeding. The structure solution and the refinement based on $F^2$ were carried out with the SHELXL program.[10] The coordinates of the non-hydrogen atoms of two **gguan** molecules, seven phosphate ions, one ammonium cation and the co-crystallised water molecules of partial occupancy (41) per asymmetric unit (a.u.) were determined by successive cycles of difference maps. These were refined by anisotropic thermal treatment. Geometric restraints on distances and angles were employed for all atoms of **gguan**, as well as rigid bond restraints (DELU). Similar anisotropic displacement restraints for spatially adjacent atoms (SIMU) were used for two guanidinium moieties in molecules **A** and **B**. The fully occupied phosphate ions were refined unrestrained. For the partially occupied phosphate ions (P4, P5, P6, P7) geometric restraints were employed with the same target values for the P-O distances and angles, as well as SIMU restraints. Finally, 16 water molecules were restrained (ISOR) to have approximately isotropic thermal parameters. Hydrogen atoms were placed on the atoms of **gguan** at idealised positions and were refined by the riding model (UH = 1.20 UC). The above restraints allowed the refinement of the structure by full matrix least squares to finish smoothly without oscillations. This converged to R1 = 0.1290 for Fo > 4σ(Fo), wR2 = 0.3299 and restrained Goodness-of-fit = 2.38 for all data. The final difference Fourier map showed maximum residual density of 0.20 e/$Å^3$ and minimum of -0.13 e/$Å^3$. Crystal data and analysis details are given in **Table S1**.

Of the seven phosphate anions ($HPO_4^{2-}$) per a.u. in the refinement, three are of full occupancy and four of partial (0.4). In **Table S1** full occupancy for all phosphates is reported (they account for 14

electrons), because it is expected that the four disordered $HPO_4^{2-}$ anions are indeed located in the reported positions. These anions alone do not balance the 16+ charge of the two **gguan**$^{8+}$ molecules per a.u., therefore some more negatively charged species, which are highly disordered and cannot be located, must exist in the structure. By solvent masking (using Olex2[11]) on the refined hydrated **gguan** structure, we find 2 identical solvent accessible voids per a.u. of 1690 electrons each, which can contain two $HPO_4^{2-}$ anions, one ammonium cation and 31 water molecules. Thus, in order to balance the charges in the structure, the molecular formula (**Table S1**) includes the 7 modelled and 2 non-modeled $HPO_4^{2-}$ anions of total charge 18 electrons that match perfectly the 18+ charge of two **gguan**$^{8+}$ molecules plus two ammonium cations, one modeled and one non-modeled. CCDC number: 2133070.

*4. NMR and DLS experiments.*

*4.1. pKa determinations by $^{13}C$ NMRspectoscopy* -Each **gguan** sample was initially dissolved in $H_2O$ acidified to pH = 3 with aq. HCl (18% v/v) and the $^{13}C$ NMR spectrum (62.9 MHz) was recorded. A sealed capillary containing deuterated DMSO-$d_6$ was used as external chemical shift reference and lock solvent. Subsequently, the solution was titrated with NaOH (1 *N* aq. solution or solid, at the high pH region). The $^{13}C$ NMR spectra were recorded at specific pH values (estimated error less than 6%) and the chemical shift differences ($\Delta\delta_{obs}$) of the C atoms were monitored (T = 295 K, NS ≈ 1800-20000). The data were collected as $\Delta\delta_{obs} = \delta_{obs} - \delta_{acidic}$ and plotted *vs* pH. These data points were fitted to a sigmoidal "Slogistic1" curve (Equation S1) in the Origin 9.0 software. The p*K*a was extracted from the fitting result by finding the midpoint (xc) of this sigmoidal curve from equation S1:

$$y = a/(1+ e^{(-k(x-xc))}) \qquad (S1)$$

Where, $a = \Delta\delta_{max}$, k = coefficient, x = pH, xc: p*K*a, $y = \Delta\delta$.

So S1 can be written: $\Delta\delta = \Delta\delta_{max}/(1+ e^{(-k(pH-pKa))})$

$$\Delta\delta/\Delta\delta_{max} = 1/(1+ e^{(-k(pH-pKa))})$$

$$\ln(\Delta\delta/\Delta\delta_{max}) = pKa - pH$$

$$pH = pKa - \ln([HA]/[A^-]) \text{ or}$$

$$pH = pKa + \ln([A^-]/[HA]), \text{ (S2- the Henderson-Hasselbalch equation)}$$

During titrations, in all $^{13}C$ NMR spectra the signals of guanidino C7 (~157 ppm), and CD core C6 and C5 were monitored (for numbering see Fig. S9). In plain $H_2O$ solution the maximum pH reached was 11.5. In $H_2O$/1 M KCl solution the maximum pH reached was ~13 and near this pH a lot of

precipitation was observed, while the $\Delta\delta$ values continued to change. Fitting of the titration data in $H_2O$ to equation S1 ($\Delta\delta^{13}C$ *vs* pH) afforded p*K*a values of 10.5 to 11 with errors ~ ±0.3. The corresponding fitting trials of the data in $H_2O$/1 M KCl gave irrational p*K*a values or values around 14 (Fig. S10).

*4.2. DLS experiments*- A Tris solution (100 mM) and a borate buffer (100 mM) solution, each at pH = 8.5, were measured. Subsequently, **gguan** was dissolved in each of the two buffers to afford a concentration of 8 mM (15.4 mg/mL) and each solution was again measured. Immediately after that, $NH_4H_2PO_4$ (200 mM) was added to both samples (the pH dropped to ~8) and an extra measurement was conducted at $t_o$ for each. For a blank comparison, a **gguan** in Tris buffer solution was prepared as above, without addition of $NH_4H_2PO_4$. All samples were monitored with time and measurements were taken at the same time intervals for each after 5 h or 10 h, 1 day, 3 days and 6 days. The measurements were conducted in triplicates. Moreover, DLS experiments with ammonium sulfate (200 mM) instead of $NH_4H_2PO_4$, using the above conditions were performed, for control.

*5. Crystallography* ***Tables S1-S5***

**Table S1.** Details of crystal and structure refinement data

| | |
|---|---|
| Molecular formula | 2(C56 H104 N24 O32) 9(HPO4) 2(H4N) 12.32(H2O) |
| Formula weight | 4373.09 |
| Temperature | 100 K |
| Radiation/Wavelength | 0.9120 Å |
| Space group | $C222_1$ |
| *A* | 26.90(2) Å |
| *B* | 36.40(2) Å |
| *C* | 53.20(2) Å |
| Volume/Z | 52091.32 / 8 |
| Density (calculated) | 1.115 $Mg/m^3$ |
| 2θ range for data collection | 8.75 °- 24.49° |
| Index ranges | -24<h<24, -33<k<33, -48<l<48 |
| Reflections collected/unique | 108839 / 19490 |
| Solution method | *Ab initio* structure expansion with fragment seeding |
| Refinement method | Full-matrix least-squares on $F^2$ |
| Data[Fo > 4σ(Fo)] /restraints/parameters | 15727 / 1274 / 2706 |
| Goodness-of-fit on $F^2$ (restrained) | 2.376 |
| R indices [$F_o$>4σ($F_o$)] | R1=0.1290 |
| R indices (all data) | R1=0.1397, wR2=0.3299 |
| Largest diff. peak and hole | 0.20 and -0.13 |

**Table S2**. Geometrical parameters of **gguan** monomers

| Glucose | D[a] (Å) | φ[b] (°) | d[c] (Å) | Tilt Angles[d] (°) | D3[e] (Å) | Torsion Angles[f] (°) $O5_n$-$C5_n$-$C6_n$-$N1_n$ |
|---|---|---|---|---|---|---|
| **gguan Molecule A** | | | | | | |
| **G1** | 4.41 (3) | 131.0 (5) | 0.01 (1) | 6 (1) | 2.98 (4) | 71 (2) |
| **G2** | 4.45 (3) | 132.4 (4) | 0.19 (1) | 20 (1) | 2.90 (3) | 56 (3) |
| **G3** | 4.66 (2) | 140.6 (4) | -0.05 (1) | 23 (1) | 2.94 (2) | 67 (2) |
| **G4** | 4.42 (2) | 134.0 (4) | -0.15 (1) | 21 (1) | 3.02 (2) | 68 (3) |
| **G5** | 4.50 (2) | 132.1 (4) | 0.04 (1) | 10 (1) | 2.78 (2) | 74 (2) |
| **G6** | 4.55(2) | 136.3 (4) | 0.11 (1) | 14 (1) | 2.79 (2) | 62 (2) |
| **G7** | 4.52 (2) | 133.4 ( 4) | 0.03 (1) | 18 (1) | 2.85 (2) | 77 (2) |
| **G8** | 4.57 (2) | 139.8 ( 4) | -0.18 (1) | 16 (1) | 2.92 (2) | 76 (3) |
| **gguan Molecule B** | | | | | | |
| **G1** | 4.34 (3) | 132.0 (4) | -0.12 (1) | 3 (1) | 2.81 (3) | 67 (4) |
| **G2** | 4.61 (2) | 135.7 (4) | 0.19 (1) | 12 (2) | 2.80 (4) | 60 (5) |
| **G3** | 4.47 (2) | 136.6 (5) | 0.16 (1) | 16 (2) | 2.75 (3) | 67 (4) |
| **G4** | 4.37 (2) | 136.0 (5) | -0.21 (1) | 17 (1) | 2.74 (3) | 73 (3) |
| **G5** | 4.43 (2) | 131.6 (4) | -0.18 (1) | 5 (1) | 2.74 (2) | 74 (3) |
| **G6** | 4.55 (2) | 134.4 (4) | 0.24 (1) | 9 (1) | 2.86 (2) | 66 (2) |
| **G7** | 4.55 (2) | 138.1 (4) | 0.15 (1) | 24 (1) | 2.82 (2) | 63 (3) |
| **G8** | 4.44 (3) | 133.9 (5) | -0.24 (1) | 15 (1) | 2.78 (2) | 66 (4) |

[a] **O4*n*…O4(*n*+1)**; [b] **O4(*n*-1)…O4*n*…O4(*n*+1)** angles; [c] Deviations of the **O4*n*** atoms from their least-squares optimum plane; [d] Tilt angles between the optimum **O4*n*** plane and the mean planes through atoms **O4(*n*-1)**, **C1*n***, **C4*n***, **O4*n***; [e] Intramolecular H-bonds between **O3*n*…O2(*n*+1)**; [f] orientation of the **C6*n*-N1*n*** bond.

**Table S3**. Geometry and conformation of the guanidinium groups.

| Glucose | D[a] $N3_n…O5_n$ (Å) | $D_{shortest}$[b] (Å) | Dihedral_1 $G_n$... $G_{(n+1)}$[c] (°) | Dihedral_2 Tilt_2[d] (°) | [e]Torsion Angles (°) $N2_n$-$C7_n$-$N1_n$-$C6_n$ | [f]Torsion Angles (°) $N3_n$-$C7_n$-$N1_n$-$C6_n$ |
|---|---|---|---|---|---|---|
| **gguan Molecule A** | | | | | | |
| **G1** | 2.96 (3) | - | 59 (2) | 62 (1) | -163 (2) | 20 (3) |
| **G2** | 2.94 (3) | 3.5 (1) | 44 (2) | 53 (1) | -167 (4) | 10 (6) |
| **G3** | 3.35 (2) | 3.6 (1) | 29 (2) | 66 (2) | -162 (4) | 17 (7) |
| **G4** | 3.07 (2) | 3.6 (1) | 23 (2) | 69 (1) | -158 (2)[g] | 22 (5) |
| **G5** | 2.90 (2) | - | 60 (1) | 58 (1) | -169 (2) | 22 (3) |
| **G6** | 4.59 (3) | 3.4 (1) | 25 (1) | 90 (1) | -163 (2) | 5 (4) |
| **G7** | 4.10 (3) | 3.4 (1) | 13 (1) | 76 (1) | -174 (2) | 1 (4) |
| **G8** | 3.44 (2) | [g]3.5 (1) | 24 (1)[g] | 68 (1) | -161 (2) | 21 (3) |
| **gguan Molecule B** | | | | | | |
| **G1** | 2.87 (3) | - | 66 (2) | 60 (1) | -171 (3) | 12 (9) |
| **G2** | 4.48 (3) | 3.4 (1) | 18 (1) | 74 (1) | -173 (5) | 3 (8) |
| **G3** | 2.95 (3) | 3.6 (1) | 27 (2) | 57 (2) | -174 (3) | 6 (7) |
| **G4** | 2.94 (2) | 3.8 (1) | 23 (2) | 63 (2) | -165 (3) | 20 (5) |
| **G5** | 2.92(2) | - | 61(1) | 64 (1) | -168 (2) | 4 (4) |
| **G6** | 4.48 (2) | 3.5 (1) | 25 (2) | 81 (1) | -172 (3) | 6 (4) |
| **G7** | 3.14 (3) | 3.6 (1) | 13 (3) | 58 (1) | -177 (3) | 5 (5) |
| **G8** | 3.13 (2) | 3.6 (1)[g] | 22 (2)[g] | 58 (1) | -176 (3) | 22 (8) |

[a]Distance between **N3n**…**O5n**; [b]Shortest distance between planar guanidine group of **Gn** and **G(n+1)**; [c]Dihedral angles between the guanidine plane **Gn** and **G(n+1**); [d]Dihedral angles between the guanidine plane **Gn** and the least-squares optimum plane of the **O4n** atoms; [e]Torsion angles **N2n-C7n-N1n-C6n** exhibiting the orientation of the **N2n** towards the cavity; [f]Torsion Angles **N3n-C7n-N1n-C6n** pointing outwards. [g]The value refers to **G8** and **G1** distance (or dihedral angle).

**Table S4.** Intermolecular H-bond distances between the hydroxyl O atoms of dimer $\mathbf{AA_{sym}}$.

| $\mathbf{O_A \ldots O_{Asym}}$ | **Distance (Å)** | $\mathbf{C_A\text{-}O_A \ldots O_{Asym}}$ ($^o$) | $\mathbf{O_A \ldots O_{Asym}\text{-}C_{Asym}}$ ($^o$) |
|---|---|---|---|
| **OA21 - OA32$_s$** | 2.66 (2) | 124 (2) | 126 (2) |
| **OA32 - OA21$_s$** | 2.62 (2) | 126 (2) | 124 (2) |
| **OA33 - OA38$_s$** | 2.90 (2) | 110 (1) | 122 (1) |
| **OA34 - OA37$_s$** | 2.86 (2) | 113 (1) | 119 (1) |
| **OA35 - OA36$_s$** | 2.78 (2) | 113 (1) | 118 (1) |
| **OA36 - OA35$_s$** | 2.78 (2) | 118 (1) | 113 (1) |
| **OA37 - OA34$_s$** | 2.86 (2) | 119 (1) | 113 (1) |
| **OA38 - OA33$_s$** | 2.90 (2) | 122 (1) | 110 (1) |

*Subscript **s** denotes the equivalent atom of molecule **A** generated by the 2-fold screw axis, (-X, Y, -Z+1/2).

**Table S5.** Intermolecular H-bond distances of the phosphate groups.

| P-O…N(O)C(P) | Distance (Å) | P-O… N(O)(°) | O… N(O)C(P) (°) |
|---|---|---|---|
| P1 - P1 non-conventional phosphate dimer | | | |
| **P1-O4 … $O4_s$ –$P1_s$** | 2.54 (5) | 124 (1) | 124 (1) |
| **P1-O1 …NA31–CA71** | 2.85 (4) | 121 (1) | 118 (2) |
| **P1-O3 …NA21– CA71** | 2.83 (3) | 122 (1) | 122 (1) |
| **P1-O1 … NA2** | 2.78 (4) | 112 (2) | - |
| **$P1_s$ –$O2_s$ … NA2** | 2.55 (3) | 145 (2) | - |
| **P1 –O2 … OW7** | 2.72 (3) | 123 (1) | - |
| **P1 –O3 … OW15** | 2.68 (4) | 147 (1) | - |
| P2 – P3 archetypal phosphate dimer | | | |
| **P2-O5 … O12 –P3** | 2.95(7) | 111 (2) | 113 (3) |
| **P2-O6 … O11 –P3** | 2.45(4) | 123 (2) | 127 (2) |
| **P2-O5 … NA23–CA73** | 2.54 (2) | 129 (3) | 135 (2) |
| **P2-O6 … NA28–CA78** | 2.80 (2) | 135 (2) | 121 (2) |
| **P2-O7 … NA12–CA72** | 2.93 (4) | 114 (1) | 112 (2) |
| **P2 –O7 … OW3** | 2.67 (4) | 110 (1) | - |
| **P2 –O7 … OW12** | 2.76 (4) | 107 (2) | - |
| **P2-O8 … NA21–CA71** | 2.95 (4) | 113 (1) | 112 (1) |
| **P2-O8 … NA22–CA72** | 2.72 (5) | 111 (1) | 131 (2) |
| **P3-O9 … NA26–CA76** | 3.02 (5) | 96 (2) | 132 (2) |
| **P3-O9 … NA25–CA75** | 2.88 (5) | 117 (2) | 113(2) |
| **P3-O10 … NA16–CA76** | 3.10 (4) | 115 (2) | 109(2) |
| **P3-O10 – NA15–CA75** | 3.37 (5) | 93 (2) | 130(2) |
| **P3–O10 … OW8** | 2.57 (5) | 112 (2) | - |
| **P3-O11 … NA27–CA77** | 2.86 (4) | 135 (2) | 131 (2) |
| **P3-O12 … NA14–CA74** | 2.98 (6) | 129 (3) | 108 (3) |
| **P3-O12 … NA24–CA74** | 3.12 (5) | 136 (2) | 103 (3) |
| P4 –P7 “a dissociating dimer” | | | |
| **P4-O13 – O26 –P7** | 3.58 (9) | 142 (4) | 100 (3) |
| **P4-O13 – O27 –P7** | 4.03(9) | 151 (4) | 90 (3) |
| **P4-O14 – NB22–CB72** | 3.89 (8) | 116 (4) | 137 (3) |

| | | | |
|---|---|---|---|
| **P4-O15 – NB12–CB72** | 2.95 (7) | 122 (4) | 146 (5) |
| **P4-O16 – NB18–CB78** | 2.08 (8) | 126 (4) | 126 (4) |
| **P4-O16 – NB28–CB78** | 3.14 (8) | 119 (3) | 77 (3) |
| **P4-O16 – OW40** | **3.11 (8)** | **112 (3)** | |
| **P7-O25 – O22–P6** | 3.07 (7) | 124 (3) | 122(3) |
| **P7-O25 – NB26–CB76** | 2.96 (5) | 113 (2) | 130 (2) |
| **P7-O25 – NB15–CB75** | 2.97 (6) | 115 (3) | 97 (2) |
| **P7-O25 – NB25–CB75** | 2.93 (7) | 121 (2) | 98 (3) |
| **P7-O26 – NB27–CB77** | 2.87 (7) | 148 (3) | 129 (3) |
| **P7-O27 – NB14–CB74** | 2.34 (7) | 141 (3) | 111 (4) |
| **P7-O27 – NB24–CB74** | 2.88 (8) | 104 (3) | 84 (3) |
| **P7-O26 – OW35** | **2.55 (7)** | **106 (3)** | |
| P5 phosphate | | | |
| **P5-O17 – OB21–CB21** | 2.59 (6) | 125 (3) | 92 (3) |
| **P5-O17 – OA27–CA27** | 2.62 (6) | 133 (3) | 92 (3) |
| **P5-O18 – NB32–CB72** | 3.00 (6) | 139 (3) | 146 (3) |
| **P5-O18 – OW11** | 2.69 (6) | 124 (2) | - |
| **P5-O18 – OW10** | 2.23(7) | 142 (3) | - |
| **P5 –O19– OW18** | 3.10 (6) | 116 (2) | - |
| **P5 –O20 – OW19** | 2.41 (6) | 138 (2) | - |
| **P5-O20 – OA25–CA25** | 2.82 (2) | 118(1) | 101 (3) |
| P6 phosphate | | | |
| **P6-O21 – OB23–CB23** | 2.97 (6) | 107(3) | 116 (2) |
| **P6-O22 – O25** | 3.07 (7) | 122 (3) | 124 (3) |
| **P6-O22 – OW9** | 2.80 (2) | 92 (3) | - |
| **P6-O23 – OW9** | 2.68 (1) | 97 (3) | - |
| **P6-O24 – OW36** | 2.42 (6) | 143(3) | - |
| **P6-O24 – NA35–CA75** | 3.06 (6) | 110 (3) | 105 (5) |

Subscript **s** denotes the equivalent atom generated by the 2-fold screw axis, (-X, Y, -Z+1/2).

*5. Crystallography Figures S1-S6*

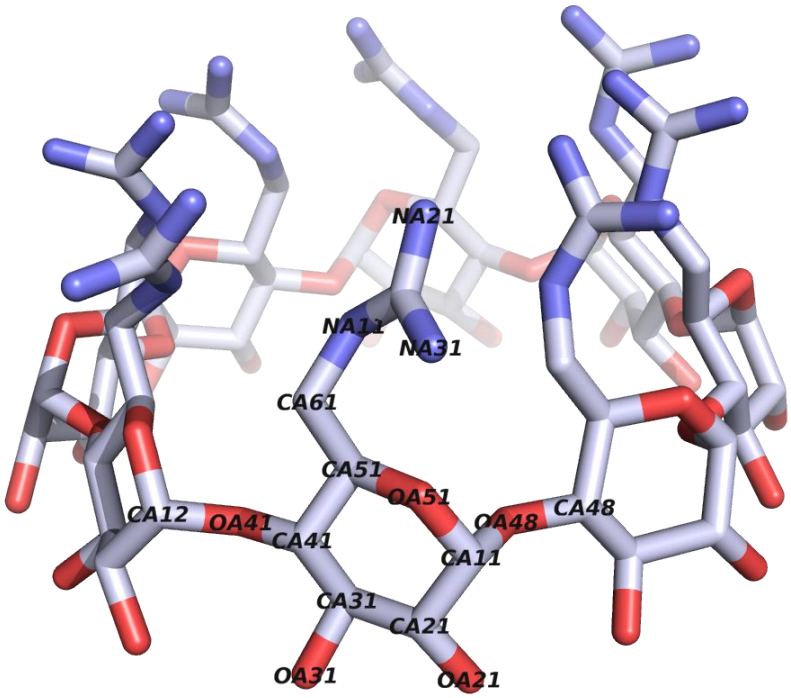


**Fig. S1**. The numbering scheme of **gguan** ( shown residue G1): CA(B)mn, OA(B)mn, NA(B)mn, m being the atom number of the glucopyranose moiety of monomer **A** or **B** and n(1-8) the residue number (Gn). Generated by Mercury.[12]

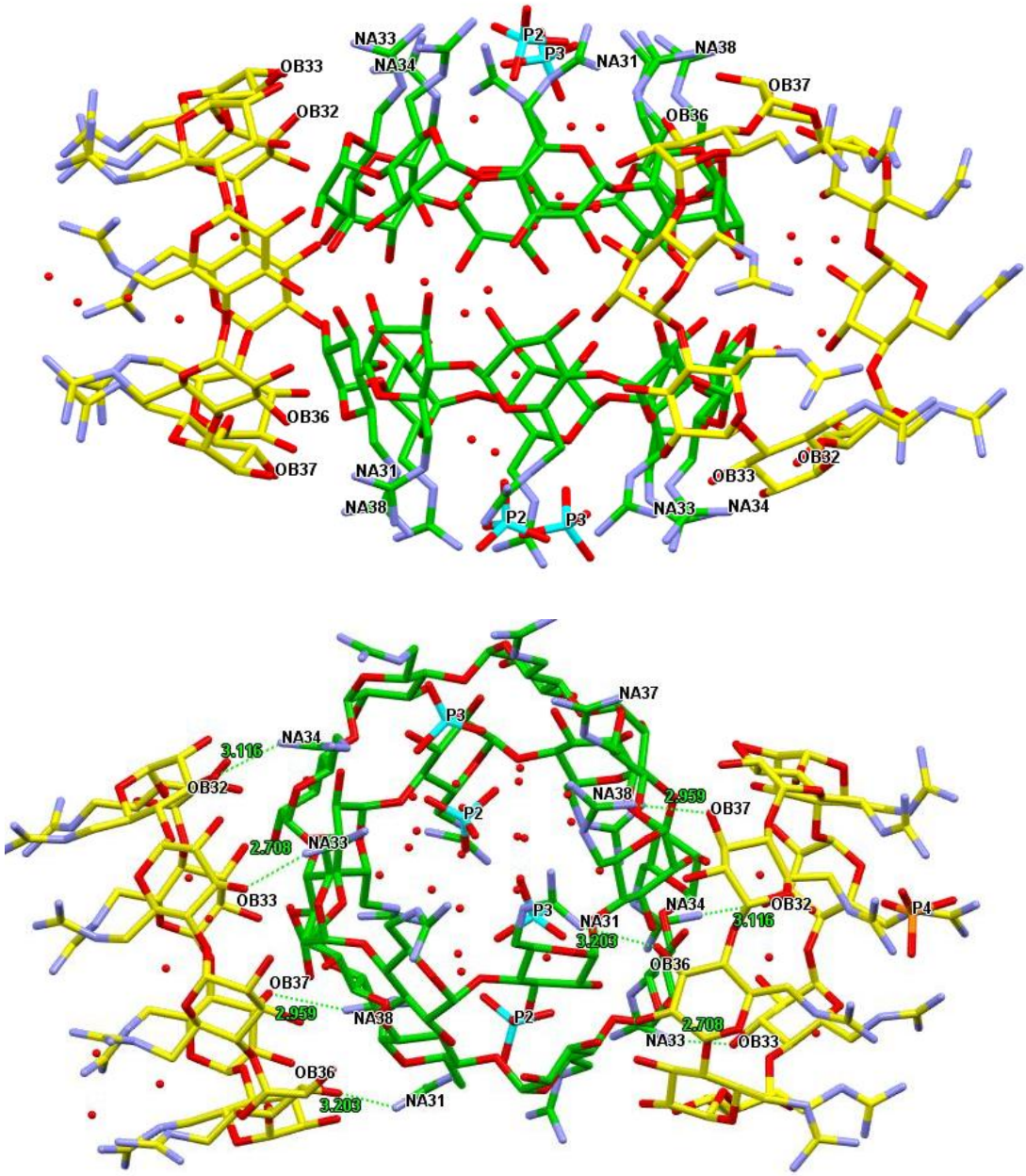


**Fig. S2.** (Top) Symmetric view of the superdimer. The eight H-bonds between hydroxy groups of **B** and $\mathbf{B}_{sym}$ and guanidinium nitrogen atoms of **A** and $\mathbf{A}_{sym}$ are shown: OB32…NA34 and OB33…NA33 (upper left) and OB36…NA31 and OB37…NA38 (upper right), as well as their symmetry equivalent in the lower part. (Bottom) In this view it is apparent that the 'embracing' of $\mathbf{AA}_{sym}$ at its intradimer region by **B** and $\mathbf{B}_{sym}$ is not symmetric. The reason is that the hydroxyl pairs OB32, OB33 and OB36 and OB37 involved in these H-bonds are related by the pseudo 2-fold axis, whereas the guanidinium nitrogen pairs NA34, NA33 and NA38, NA31 are not. The embracing would be symmetric if the involved guanidinium pairs were (upper left): NA34, NA33 and the NA38, NA37 that are related by the pseudo 2-fold axis to the former. This probably does not happen, because the OB36…OB37 distance (5.26 Å) matches better the NA38…NA31 distance (4.94 Å) for the formation of the H-bonds, instead of the NA38…NA37 nitrogen distance of 4.12 Å that is rather small. Generated by Mercury.[12]

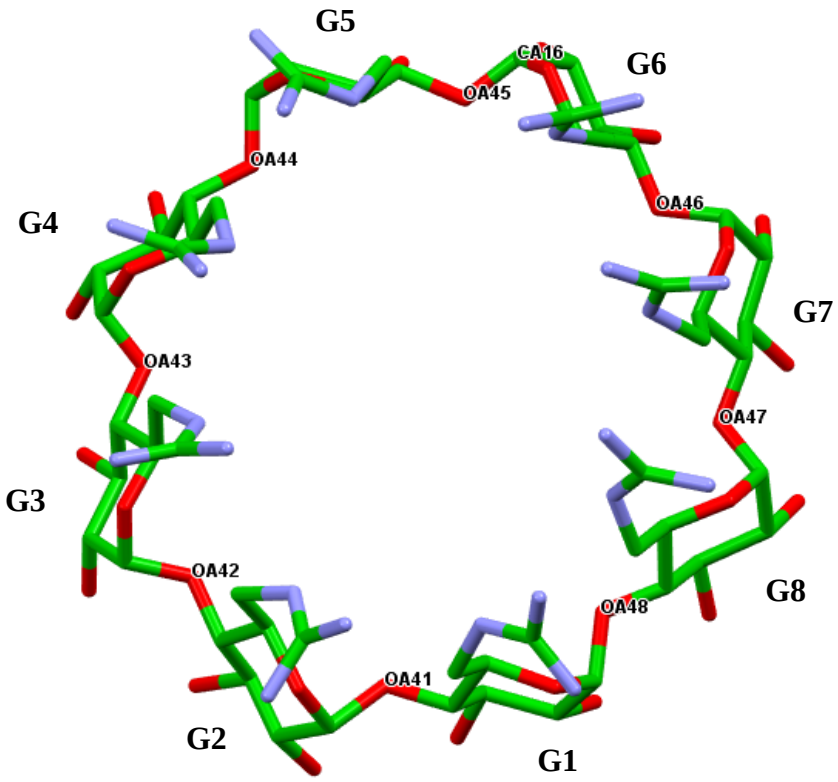


**Fig. S3.** The γCD torus exhibits a pseudo 2-fold axis that relates residues **$G_2$, $G_3$, $G_4$, $G_5$** to **$G_6$, $G_7$, $G_8$, $G_1$** , respectively. This is more prominent in the guanidinium cations' conformations. Related distances in the two sets are very close. Generated by Mercury.[12]

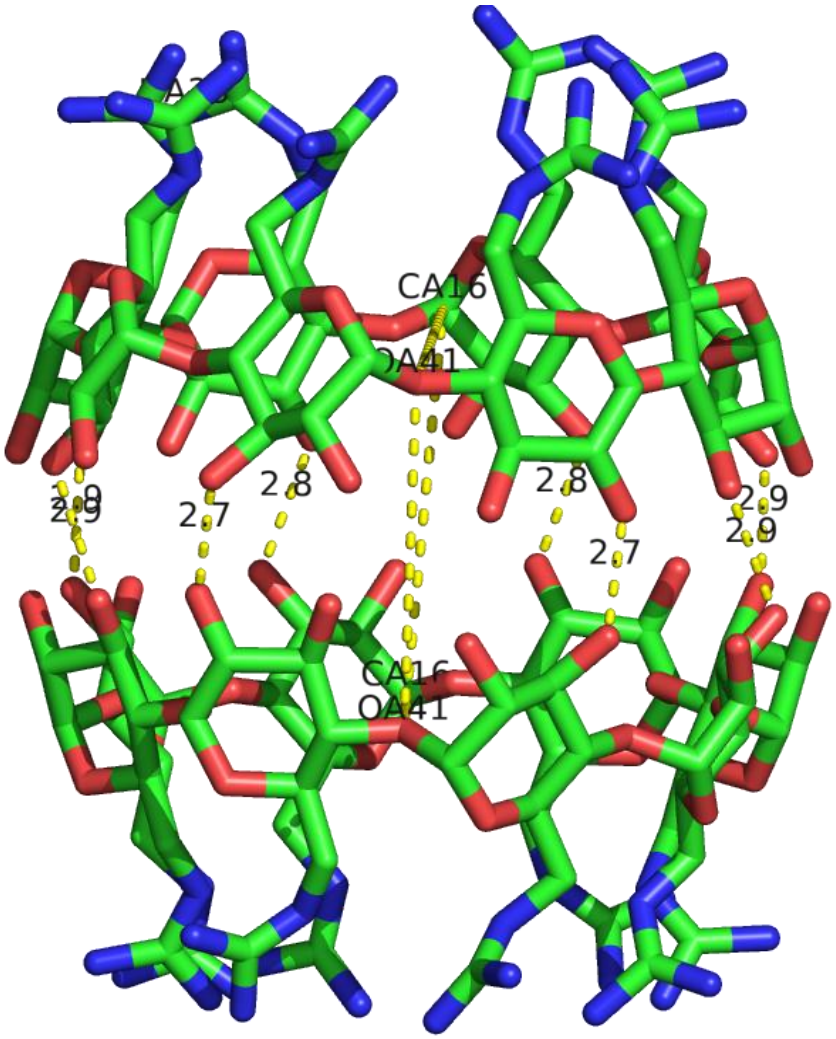


**Fig. S4.** Two sets of equal intermolecular H-bonds in the $\mathbf{AA}_{\mathbf{sym}}$ dimer. Atoms OA41 and CA61 and their symmetry equivalent $OA41_{sym}$ and $CA61_{sym}$ (located at opposite sides of **gguan**) define a plane that contains the 2-fold **b**-axis relating the **A**, $\mathbf{A}_{\mathbf{sym}}$ monomers. Generated by PyMOL.[13]

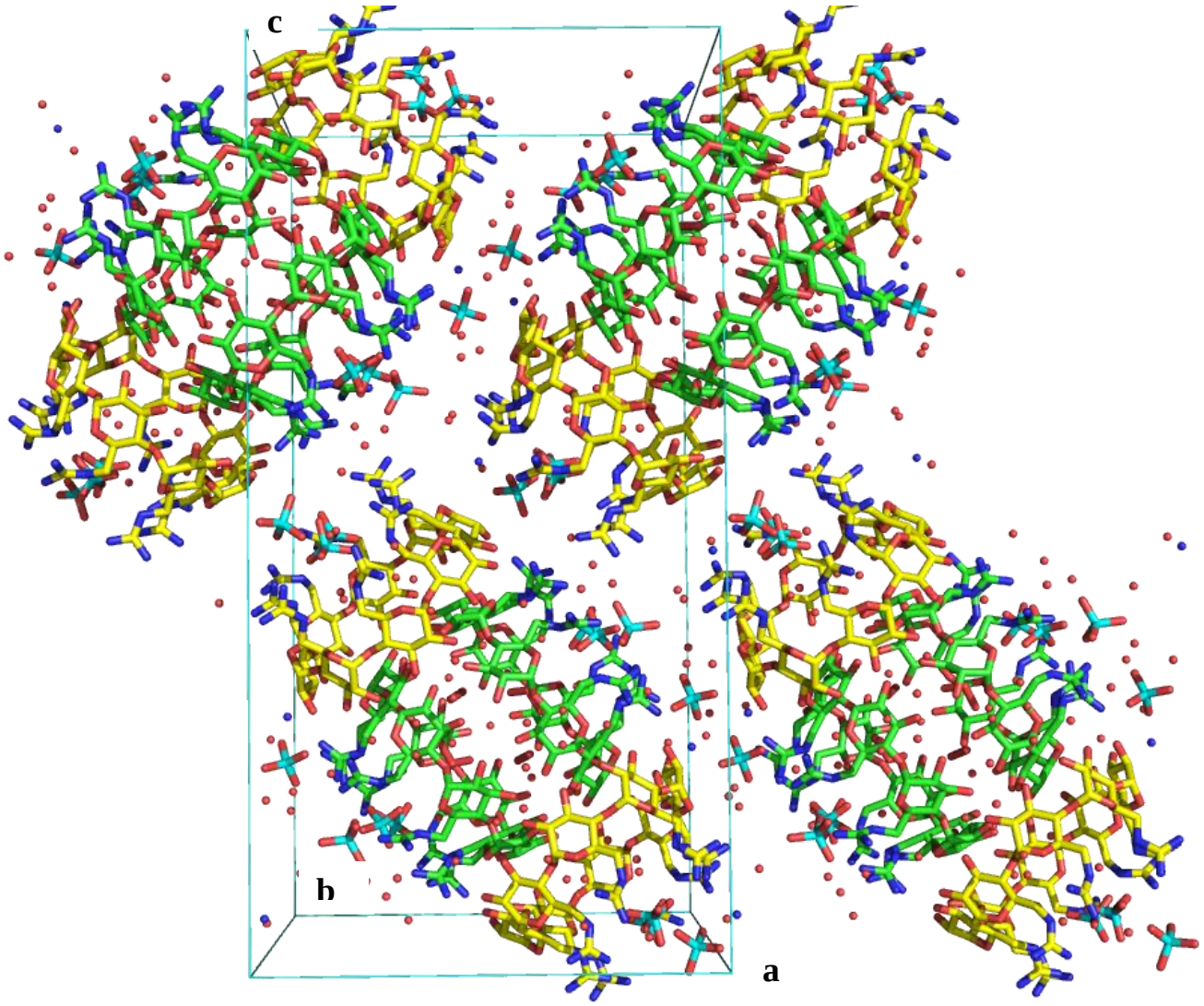


**Fig. S5a**. In front, the **ac** plane. The superdimers are packed in layers along the **c**-axis, as well as along the **b**-axis; shown is one layer of the **gguan** superdimers, perpendicular to the **b**-axis. Within each layer the superdimers are connected by H-bonds *via* the phosphates P1, P5 and P6 as well as numerous water molecules. Generated by PyMOL.[13]

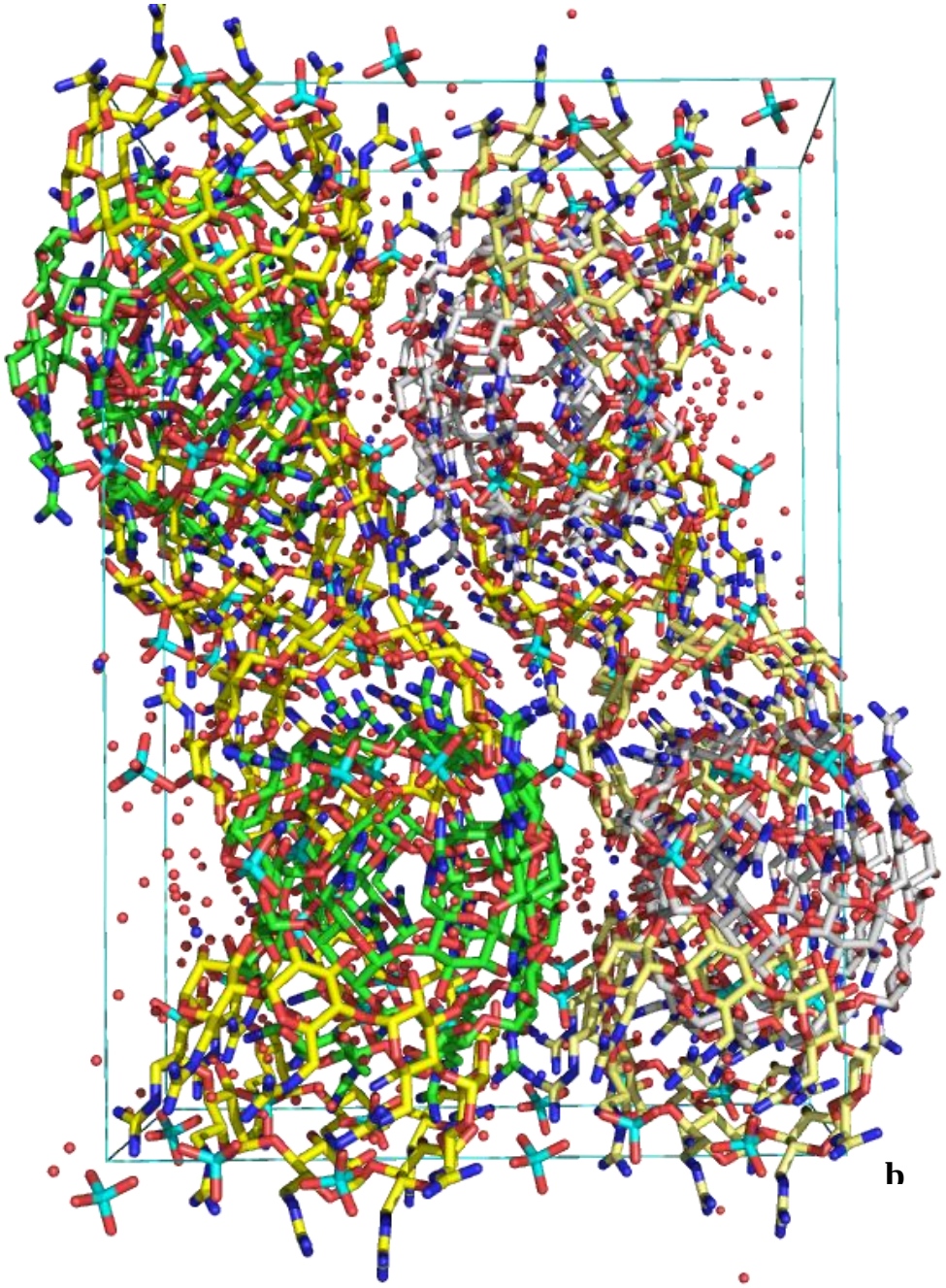


**Fig. S5b.** In front, the **bc** plane. Two layers of the **gguan** superdimers, perpendicular to the **b**-axis. Between layers, the superdimers are connected by H-bonds *via* the phosphates P1, P5 and P6 as well as numerous water molecules. Generated by PyMOL.[13]

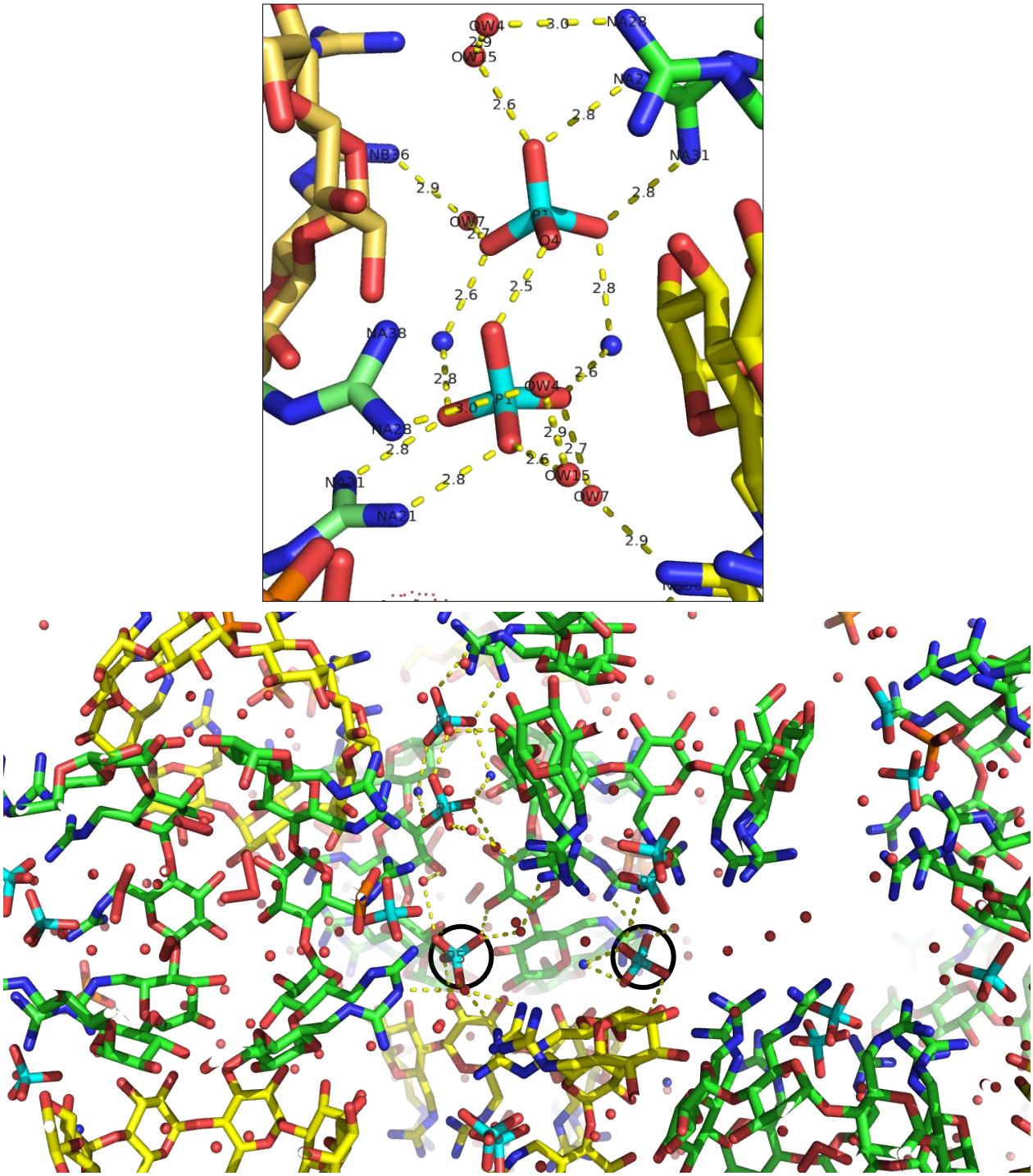


**Fig. S6.** (Top) Distances shown for a phosphate P1-P1 dimer formed by a direct anion-anion H-bond and two ammonium ion pairs. The dimer is stabilised by electrostatic H-bonds donated directly by guanidinium cations, as well as by H-bonding to water molecules. (Bottom) Intermolecular interactions of phosphate anions between superdimers in a layer perpendicular to the **b**-axis. Phosphates P5 and P6 (circled) are stabilized by similar electrostatic H-bonds as in the P1 phosphate. Generated by PyMOL.[13]

*7. DLS and NMR Figures S7-S10*

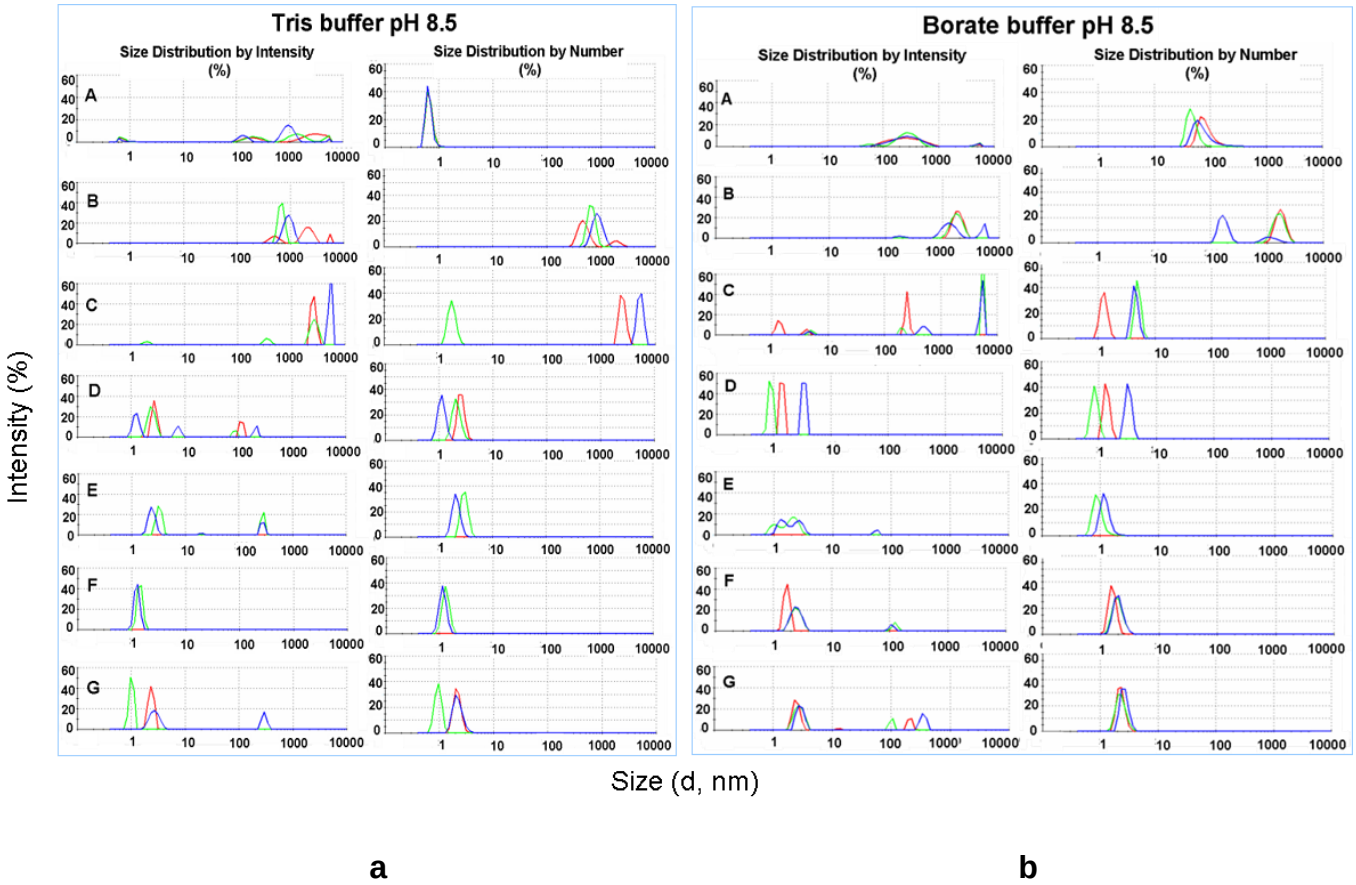


**Fig. S7**. DLS measurements plotted as both % intensity and % number *vs* size distribution. a) Tris buffer (100 mM pH 8.5) measurements and b) borate buffer in $H_2O$ (100 mM pH 8.5) measurements: A) buffer alone; B) + **gguan** (8 mM); C) + $NH_4H_2PO_4$ (200 mM); D) after 5 h; E) after 1 d; F) after 3 d; G) after 6 d.

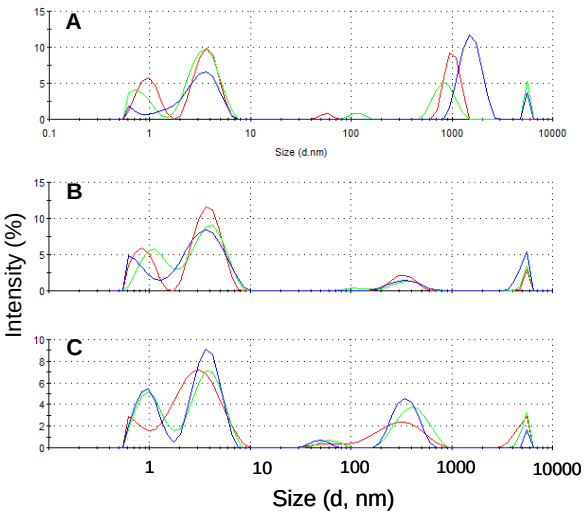


**Fig. S8**. DLS measurements plotted as % intensity *vs* size distribution of **gguan** (8 mM) in Tris buffer (100 mM pH 8.5 with $(NH_4)_2SO_4$ (200 mM) A) after 1 d; B) after 3 d; C) after 6 d.

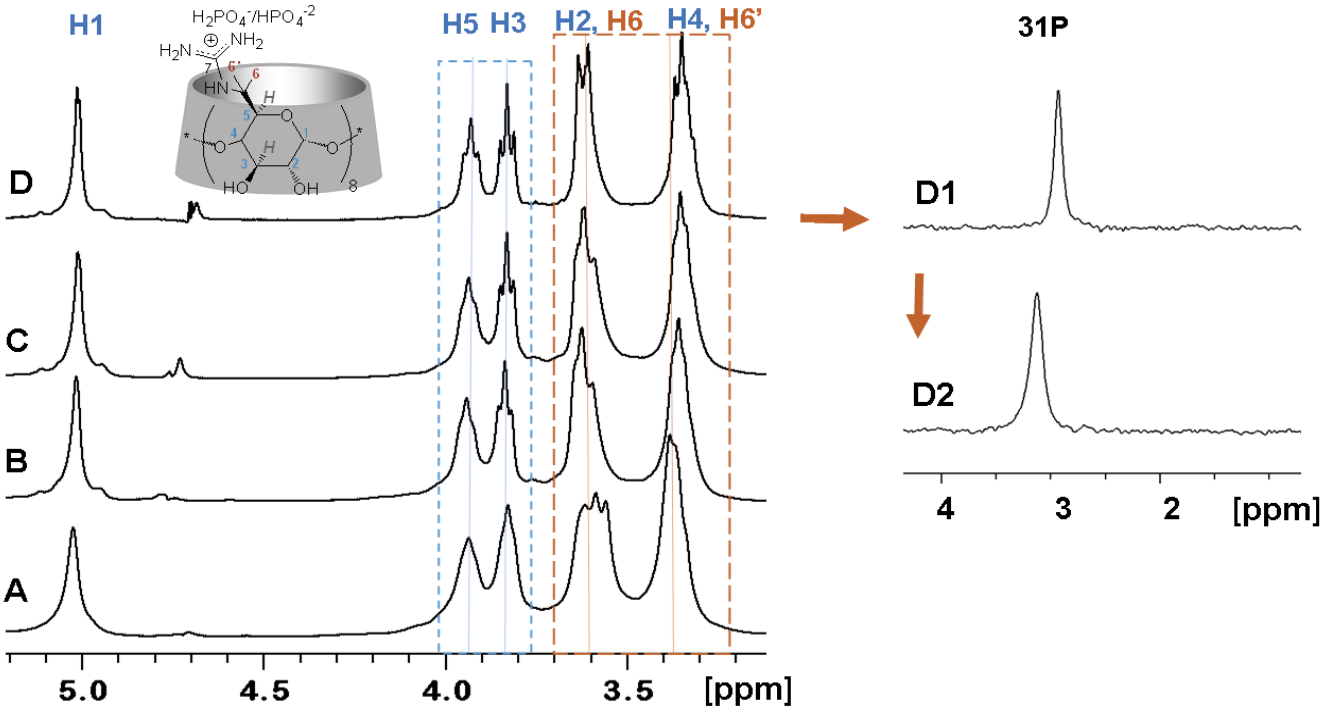


**Fig. S9.** $^{1}$H NMR spectra of **gguan** (8 mM) in $D_2O$, 500 MHz A) in 100 mM borate buffer in $D_2O$, pH 8.5; B) + NH4H2PO4 (200 mM), with H6, H6' affected; **C**) 3 days later, no change in the spectrum D) the **gguan** sample from DLS experiments in Tris, after overnight dialysis: Tris signals are absent and the spectrum is practically the same as in C; D1) the corresponding $^{31}$P spectrum of the dialysed sample denotes the presence of phosphates; D2) after a 2$^{nd}$ overnight dialysis the phosphates are still present.

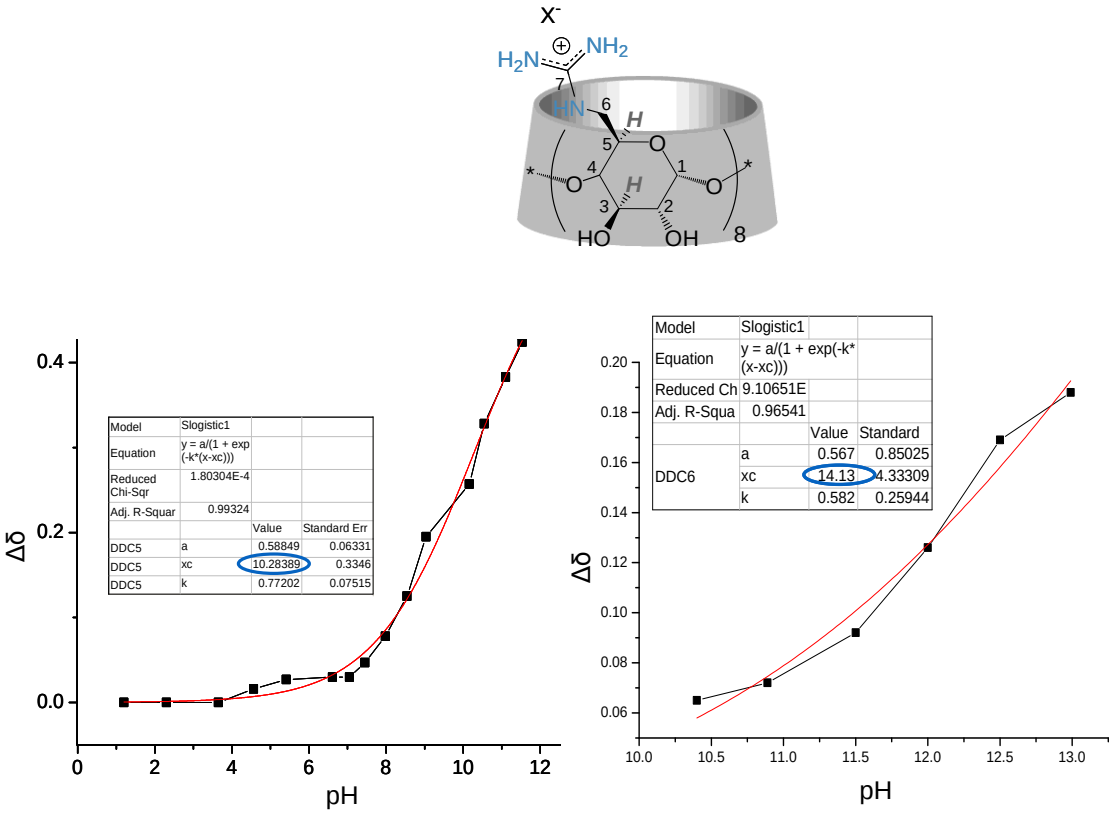


**Fig. S10.** Representative p*K*a calculation trials by monitoring the $^{13}$C NMR (62.9 MHz) signals of **gguan** C5, C6 or C7 during pH titrations. Left: in $H_2O$; right: in $H_2O$/1M KCl.

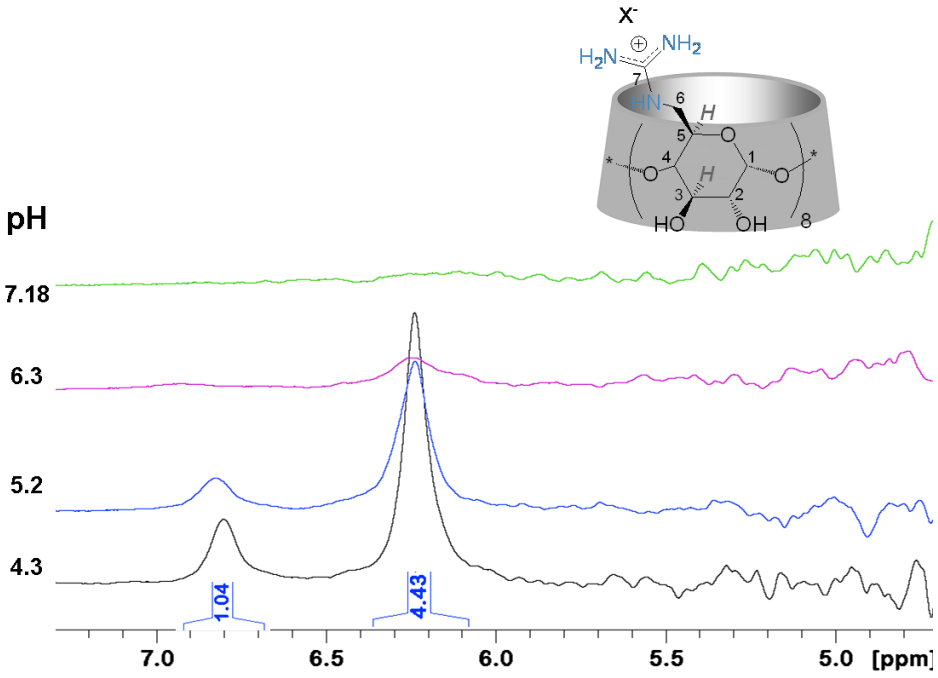


**Fig. S11.** Partial $^{1}$H NMR spectra showing the protonated guanidino group of **gguan** (20 mg/0.6 mL in $H_2O$ with external DMSO-$d_6$ as reference, 250 MHz) at different pH values, obtained with presaturation of the residual solvent peak. The protonated guanidino signals are clearly observed (ratio ~1:4, as measured at pH 4.3). These signals diminish with increase of pH, but they can still be detected up to pH ~7.

*8. References*


1. D. Vizitiu, C. S. Walkinshaw, B. I. Gorin, G. R. J. Thatcher, Synthesis of monofacially functionalized cyclodextrins bearing amino pendent groups. *J. Org. Chem.* 1997, **62**, 8760 –8766.

2. P. R. Ashton, R. Koniger, J. F. Stoddart, D. Alker, V. D. Harding, Amino acid derivatives of beta-cyclodextrin. *J. Org. Chem.* 1996**, 61**(3), 903-908.

3. N. Mourtzis, K. Eliadou, C. Aggelidou, V. Sophianopoulou, I. M.Mavridis, K. Yannakopoulou, Per(6-guanidino-6-deoxy)cyclodextrins: synthesis, characterisation and binding behaviour toward selected small molecules and DNA. *Org. Biomol. Chem.* 2007**, 5**(1), 125-131.

4. N. E. Chayen, E. Saridakis, Protein crystallisation: from purified protein to diffraction-quality crystal. *Nat. Methods* 2008**, 5**(2), 147-153.

5. N. E. Chayen, E. Saridakis, R. P. Sear, Experiment and theory for heterogeneous nucleation of protein crystals in a porous medium. *Proc. Natl. Acad. Sci. USA* 2006**, 103**(3), 597-601.

6. S. Khurshid, E. Saridakis, L. Govada, N. E. Chayen, Porous nucleating agents for protein crystallisation. *Nat. Protoc.* 2014**, 9**(7), 1621-1633.

7. E. Saridakis and N. E. Chayen, Towards a 'universal' nucleant for protein crystallization *Trends Biotechnol.*, 2009, **27**, 99-106.

8. W. Kabsch, XDS. *Acta Cryst.* 2010**, D66**(2), 125-132.

9. G. M Sheldrick, Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Cryst.* 2010**, D66**, 479-485.

10. G. M. Sheldrick, T. R. Schneider, SHELXL: High-resolution refinement. *Methods Enzymol.* 1997, **277**, 319-343.

11. O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. K. Howard and H. Puschmann. "OLEX2: a complete structure solution, refinement and analysis program". *J. Appl. Cryst.* 2009, **42**, 339- 341

12. C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek and P. A. Wood, Mercury CSD 2.0 - new features for the visualization and investigation of crystal structures, *J. Appl. Cryst*. 2008, **41**, 466-470.

13. W. L. DeLano, *The PyMOL Molecular Graphics System, DeLano Scientific LLC, http://www.pymol.org*, 2002.