Live-Cell-Templated Dynamic Combinatorial Chemistry

Daniel Carbajo,[a] Yolanda Pérez,[b] Jordi Bujons,[a] and Ignacio Alfonso*[a]

Dedicated to the memory of Prof. Kilian Muñiz

[a] Dr. Daniel Carbajo, Dr. Jordi Bujons, Dr. Ignacio Alfonso
Department of Biological Chemistry
Institute of Advanced Chemistry of Catalonia (IQAC-CSIC)
Jordi Girona 18-26, 08034, Barcelona, Spain.
E-mail: ignacio.alfonso@iqac.csic.es
[b] Dr. Yolanda Pérez
NMR Facility (IQAC-CSIC)
Jordi Girona 18-26, 08034, Barcelona, Spain.

Supporting information for this article is at the end of the document

Abstract: Dynamic covalent chemistry combines in a single step the screening and synthesis of ligands for biomolecular recognition. To that, a chemical entity is used as template within a dynamic combinatorial library of interconverting species, so that the stronger binders are amplified due to the efficient interaction with the target. Here we employed whole A549 living cells as template in a dynamic mixture of imines, for which amplification reflects the efficient and selective interaction with the corresponding extracellular matrix. The amplified polyamine showed strong interaction with the A549 extracellular matrix by on-cell NMR experiments, while combination of NMR, SPR and molecular dynamics simulations in model systems revealed insights on the molecular recognition event. Noticeably, our work pioneers the use of whole living cells in dynamic combinatorial chemistry, which paves the way towards the discovery of new bioactive molecules in a more bio-relevant environment.

Dynamic combinatorial chemistry (DCC) proposes the use of dynamic libraries (DCL) for the generation of species able to exchange through reversible covalent bonds.[1] These molecular systems are responsive to external stimuli by modifying the DCL composition,[2] with the stabilized members being amplified at the expense of the other components in the mixture.[3] Within the chemical biology field, the DCC approach has led to the discovery of new protein ligands,[4] nucleic acids binders[5] or even replicators.[6] For biological applications, it should be desirable that the conditions used for the DCC screening resemble those in the place of action.[7] Inspired by Sander's comparison of DCC with the immune system,[8] we envisioned to target the extracellular matrix (ECM).[9] The external surface of the cells is formed by a complex network of glycoproteins and anionic polysaccharides that is fundamental for processes such as cell communication,[10] regeneration,[11] metastasis[12] or host-pathogen infection.[13] The ECM is the first barrier for a molecule (i.e. a drug) to enter inside the cell; thus, navigating the ECM is fundamental in biomedicine and in chemical biology.[14] However, the chemical and structural complexity of the ECM have hindered its detailed molecular characterization, and frustrated the rational design of synthetic ligands.[15] Paradoxically, the intrinsic complexity of the ECM offers an ideal playground for the realization of the self-organizing features of DCC (Figure 1A), which has demonstrated its power for the discovery of strong binders to challenging biomolecules.[1-8] Considering our recent results in the identification of a strong heparin binder[16] and the chemical similarity between heparin and the glycosaminoglycans (GAGs) of the ECM,[14] we designed a library (Figure 1B) combining spermine as a cationic polyamine scaffold[17] with a set of aromatic aldehydes, which would mediate binding through CH-aryl interactions with the saccharide units.[18] Thus, the dynamic mixture of imines (**2XY**) obtained by the reaction between spermine (**1**) and an equimolecular mixture of four aromatic aldehydes (**A**, **B**, **K**, **L**) was incubated with living cells and reduced in situ with $NaBH_3CN$ to the corresponding polyamines (**3XY**). As initial model, we used the A549 human lung adenocarcinoma cell line since the ECM of these cells is rich in anionic GAGs.[19] The supernatant was analyzed by UPLC-MS allowing the identification and quantification of each member of the library (Figure 1C). The normalized area of the UPLC-MS peaks for the reactions performed in the presence (A_T) and in the absence (A_0) of cells was compared by the calculation of the corresponding amplification factors ($AF = A_T/A_0$).

The A549-templated library displayed interesting AF values (black bars in Figure 1C) with a molecule (**3AL**) specially favored by the presence of the cells. Interestingly, the experiment performed with another cell line (HeLa) produced a different amplification pattern (gray bars in Figure 1C) as expected for the different GAG-composition of its ECM.[20] Control experiments were designed to get clues about the origin of these observations. The amplification of **3AL** was drastically reduced when the A549-templated reaction was performed in the presence of protamine (blue bars in Figure 1C), an arginine-rich protein with well-known GAG-binding properties,[21] supporting the role of GAGs in the amplification process. Additionally, we also tested A549 cells grown in the presence of sodium chlorate (green bars in Figure 1C), which inhibits the GAG sulfation in the cell.[22] Here again, the AF of **3AL** drastically decreased. Both experiments underlined the importance of electrostatic interactions with the GAGs present in the ECM. Moreover, the behavior of different members of the library suggested that additional non-covalent contacts (hydrogen bonds, hydrophobic or CH-π interactions) might be also playing an important role. Thus, the selection of the corresponding aromatic moieties in **3AL** might reflect an optimal balance between H-bonding sites and aromatic surfaces for carbohydrate CH-π interactions. In all the cases, the cell viability after the reaction was checked using the MTT assay rendering >60% of cell survival (Figure S1).

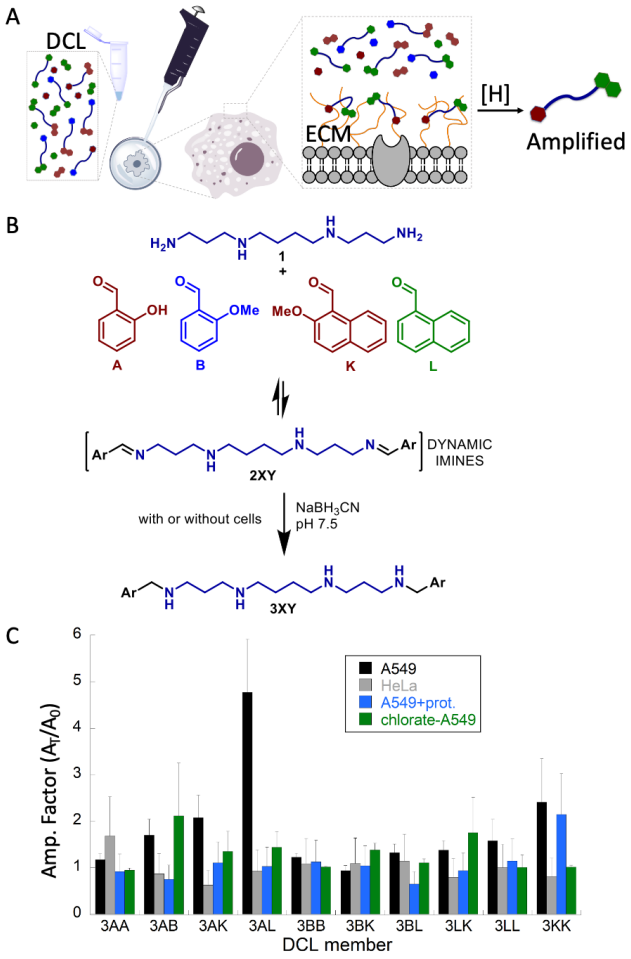


Figure 1. (A) Cartoon for the cell-templated DCC. (B) Scheme of the DCL. (C) Plot of the AF (UPLC peak area in the templated/non-templated reactions) using living cells as templates: A549 (black), HeLa (gray), A549 in the presence of protamine (blue) and A549 grown with sodium chlorate (green). The values in (C) show the average of at least three replicates.

In order to confirm that the observed amplification is due to a stabilizing interaction with the ECM, we synthesized **3AL** at preparative scale and studied its binding ability with a set of on-cell NMR[23] experiments (Figure 2). The NMR tube was filled with A549 cells in order to obtain a homogenous sample (Figure S2). The corresponding solution ^{1}H NMR spectrum showed the signals for the most fluxional parts of the cell (lipids, GAGs, collagen and some excreted metabolites), in agreement with those reported in the literature[24] and confirmed by 1D/2D NMR experiments (Figures S3-S4). We used two different NMR experiments for studying the **3AL**-ECM interaction: relaxation-filtered[25] and WaterLOGSY[26] sequences. The first experiment takes advantage of the fast transverse relaxation of large molecules: the implementation of a T_2-filter attenuates the NMR signals of species showing slow tumbling in solution. When small molecules interact with large biopolymers (or cells), their signals will decrease with respect to the unbound molecules in the bulk. On the other hand, the WaterLOGSY is a NOE-based experiment that starts with the excitation of the water protons, which transfer their magnetization to the hydrated molecules by NOE and spin diffusion mechanisms (Figure 2A,B). As in NOESY, the signals of small and large species have opposed phase. Accordingly, the binding of a small molecule to a large entity (like a cell) produces a change in the sign of its signals in the WaterLOGSY experiment. Thus, the T_2-filtered ^{1}H NMR spectrum of **3AL** in the presence of A549 cells showed much less intense signals than the corresponding spectrum of the free ligand in solution (compare traces (ii) vs. (iii) in Figure 2C). Additionally, the WaterLOGSY spectra of **3AL** in the absence and in the presence of A549 cells displayed opposed signs (traces (v) vs. (vi) in Figure 2C). Both results support a strong binding of **3AL** to the external surface of A549 cells.

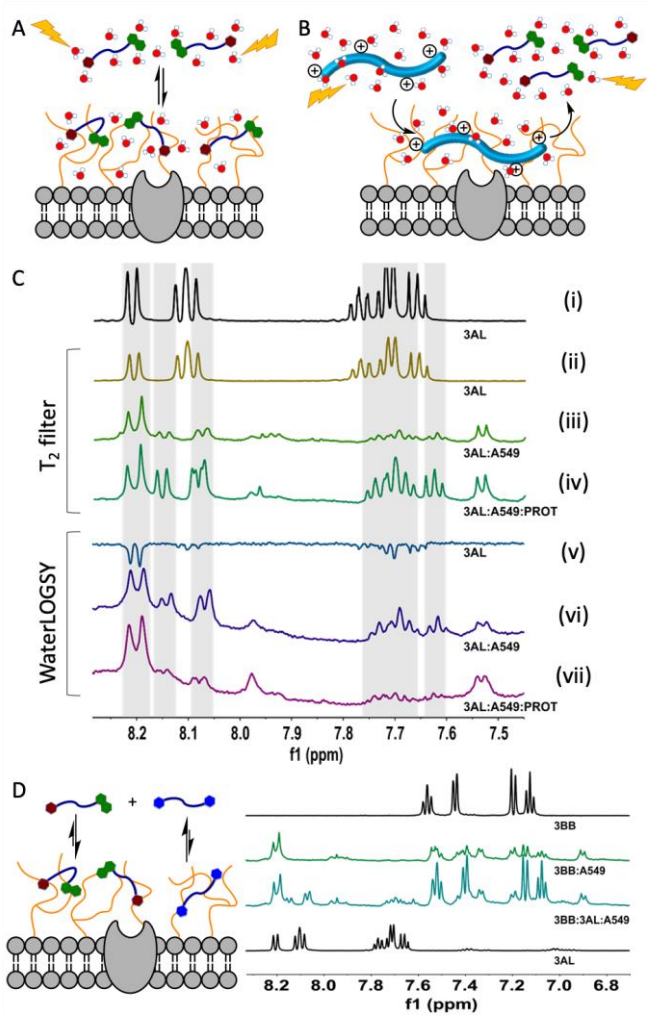


Figure 2. Binding of ligands to A549 whole cells: (A) Cartoon representation for the interaction of **3AL** with the ECM of A549 live cells. (B) Cartoon representation of the protamine-mediated displacement of the ECM-bound **3AL**. (C) NMR binding experiments (T_2-filtered (200 ms) and WaterLOGSY spectra) supporting the processes depicted in (A,B). From up-down in (C): ^{1}H NMR reference spectra of **3AL** (i); T_2-filtered spectra of **3AL** (ii), **3AL** : A549 (iii), and **3AL** : A549 : protamine (PROT) (iv) samples; WaterLOGSY of **3AL** (v), **3AL** : A549 (vi), and **3AL** : A549 : PROT (vii) samples. (D) Competition experiments for the **3AL**/**3BB** binding to A549 by T_2-filtered (200 ms) NMR (from up-down): ^{1}H NMR spectrum of **3BB**, T_2-filtered of **3BB** : A549, **3BB** : **3AL** : A549 samples, and ^{1}H NMR of **3AL**.

We also used protamine to perform competition experiments (Figure 2B). Thus the addition of protamine to the **3AL** : A549 sample induced the recovery of the intensity of the **3AL** signals in the T_2-filtered spectrum (trace (iv) in Figure 2C) and a decrease

of the corresponding WaterLOGSY effect (trace (vii) in Figure 2C). This is a consequence of the **3AL** release back to the bulk due to the efficient displacement from its complex with the GAGs of the ECM. On the other hand, the different AF observed for the relatively similar small molecules of the DCL is an intriguing observation, which could be related to a selective binding to the ECM. To confirm this assumption, we also synthesized a member of the library that does not vary its relative concentration in any of the DCC experiments (**3BB**, Figure 1C). The T_2-filtered experiment with a sample containing **3BB** and A549 cells showed that this molecule also binds the ECM (second trace in Figure 2D), although this interaction is weaker than the **3AL**-ECM complex, as confirmed by the corresponding **3BB**/**3AL** competition experiment (third trace in Figure 2D).

The ligand **3AL** showed low toxicity to A549 cells (>90% cell survival up to 250 µM **3AL**, Figure S7), but effectively reduced the staining of the cells with methylene blue (MB, Figure S8), a known dye for the GAGs of the ECM. This observation confirms the strong binding of **3AL** to the ECM, thus precluding the MB-ECM interaction.

To unravel the process at the molecular level, we envisioned chondroitin sulfate (CDS) as a convenient GAG model considering its abundance (65% of total GAG) in the ECM of the A549 cells.[27] First, selected NMR interaction experiments were performed with **3AL** and CDS from shark cartilage (Figure 3A). Also in this case, the addition of CDS to a **3AL** sample produced a clear attenuation of the T_2-filtered NMR signals and a change in the sign of the WaterLOGSY spectrum, supporting the efficient **3AL**-CDS interaction in solution. Remarkably, the corresponding NMR experiments with **3BB** under the same experimental conditions showed an imperceptible attenuation of the T_2-filtered signals and zero WaterLOGSY effects (Figure 3B). These results suggested a stronger interaction of **3AL** to CDS, as compared to **3BB**. Moreover, a quantitative estimation of the binding was also carried out by SPR on CDS-functionalized chips, revealing a more intense response with **3AL** than with **3BB** (Figure 3C), in line with the NMR results in solution. The global fitting of the SPR data rendered a dissociation constant (K_d) of 2.1 µM for **3AL**, while 33.6 µM for **3BB** (Figure S9,S10). All these results correlate with the observed AF in the DCC assays and the NMR binding experiments, both performed with the whole A549 cells.

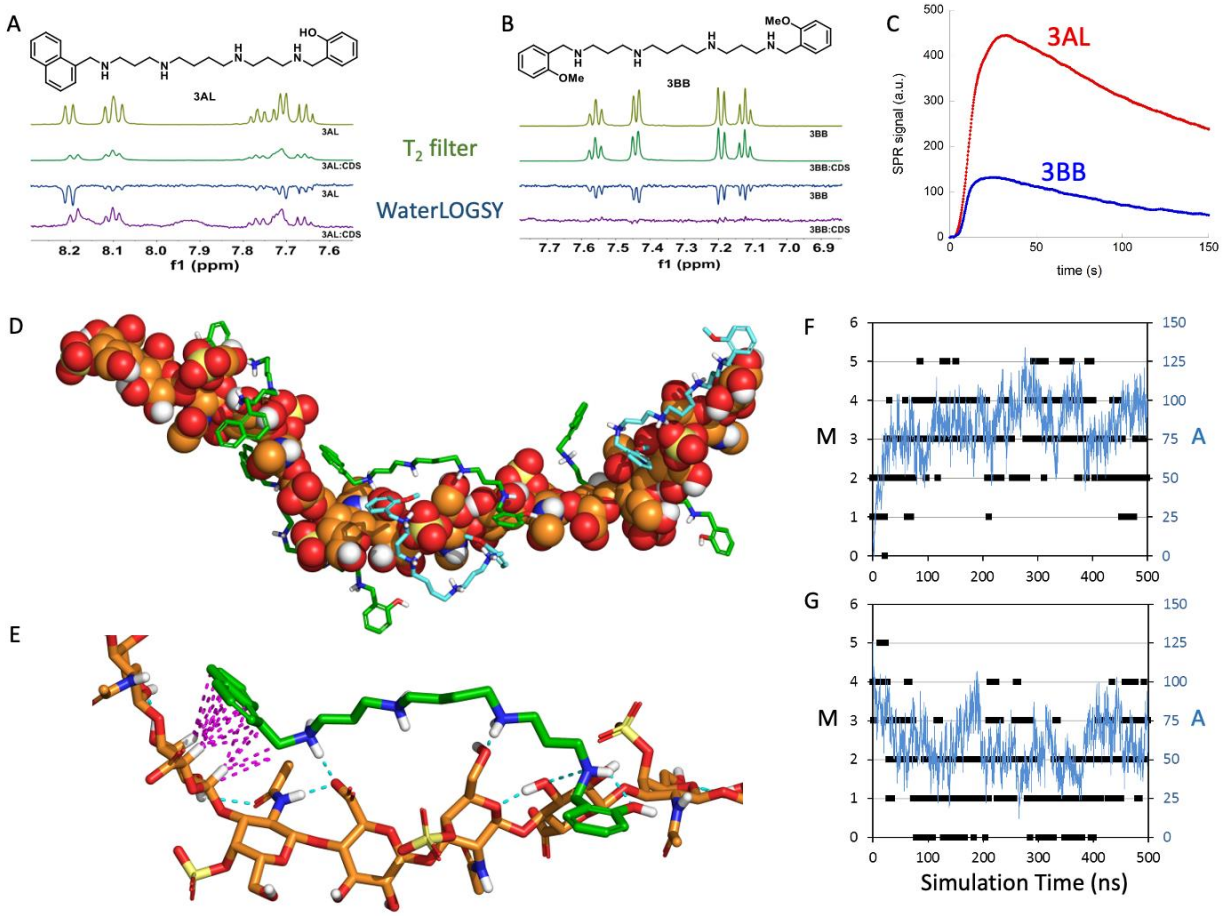


Figure 3. Binding of CDS with **3AL** or **3BB**. (A) NMR binding experiments with **3AL** and CDS (from up-down): T_2-filtered of **3AL**, and **3AL** : CDS samples; WaterLOGSY of **3AL**, and **3AL** : CDS samples. (B) Corresponding NMR binding experiments with **3BB**. (C) SPR sensograms of **3AL** (red) and **3BB** (blue) binding to CDS-functionalized chips. (D-G) Summary of results obtained from 500 ns MD simulations of a system containing a CDS-A hexadecamer, with 5 molecules of **3AL** and 5 molecules of **3BB** in TIP3P water. (D) Representative snapshot of the CDS-A/**3AL**+**3BB** simulation (CDS-A: spheres, orange C-atoms; **3AL**: sticks, green C-atoms; **3BB**: sticks, cyan C-atoms), ligands at >5 Å, water molecules and ions are omitted. (E) Detail showing the interactions between **3AL** and CDS-A (H-bonds: cyan dashed lines; CH-π interactions: magenta dashed lines). (F,G) Graphics representing the number of ligand molecules (M, black squares, left axis) and the number of ligand atoms (A, blue lines, right axis) within a distance of 3 Å from the CDS-A molecule vs. simulation time. Average values: (F) $\overline{M}$(**3AL**) = 3.19±0.99, $\overline{A}$(**3AL**) = 83±18; (G) $\overline{M}$(**3BB**) = 1.84±1.10, $\overline{A}$(**3BB**) = 60±16.

The molecular recognition process was also visualized with the help of molecular dynamics simulations in explicit water (Figure 3D-G). We built the corresponding hexadecamer of CDS-A, since this is the main GAG in the ECM of A549 cells (33% of total GAGs)[27] and the one used for the SPR assays. The individual binding of **3AL** and **3BB** to CDS-A was confirmed by 500 ns simulations of a CDS-A hexadecamer (fully ionized) and 5 molecules of either **3AL** or **3BB** in their corresponding tetraprotonated forms, adding chloride anions to neutrality. In both cases, the results showed the efficient interaction of the ligands with CDS-A, with no evident differences between the two polyamines (Figures S11,S12,S14A,B). Then, we performed the corresponding molecular dynamics simulations including both compounds (5 molecules of **3AL** and 5 molecules of **3BB**). In this competition set up, a clear preference for **3AL** was observed, since the most populated cluster of the simulation showed four molecules of **3AL** but two molecules of **3BB** bound to CDS-A (Figure 3D, Figures S13,S16 and Movie). The **3AL** : CDS-A molecular recognition is sustained by H-bonding with the ammonium and phenol groups, as well as carbohydrate CH-π contacts with the naphthyl ring (Figure 3E). As initially hypothesized, the synergy of the two types of interactions is optimized for the amplified ligand. The selectivity towards **3AL** is clearly reflected in the plots of the contacts (molecules and atoms) between the ligands and the CDS-A during the simulation time (Figure 3F,G, Figures S14-S16). A more detailed analysis shows that **3AL** establishes a higher number of direct H-bonding, ionic, and water mediated interactions than **3BB** during most of the simulation (Figure S17). Furthermore, an estimation of the number of CH-π contacts also points to the same direction (Figure S18). Thus, the experimental and theoretical results on the interaction of the ligands with this GAG can satisfactorily explain the observed selectivity in the DCC experiments.

Our results demonstrate, for the first time, that dynamic covalent chemistry can be successfully applied with living templates thus expanding its application in chemical biology, and opening new venues for a fast and meaningful approach to discover new ligands for challenging biomolecular targets in realistic media. Recalling the original inspiration from immune system, the efficient molecular recognition of the cell surface opens the possibility to use DCC to differentiate, for instance, healthy cells from different organs against cancer cells or pathogens. Moreover, the modular nature of the dynamic combinatorial library allows expanding its molecular diversity to virtually cover the wide chemical and compositional variety of the extracellular matrix in different cell types. Therefore, this approach has the potential to contribute to many different aspects at the intersection between chemistry and biology.

Acknowledgements

We thank Prof. A. Messeguer for helpful discussion. The ICTS “NANOBIOSIS” (CAbS, IQAC-CSIC, CIBER-BBN) is acknowledged for the assistance and support with the use of OpenSPR. This work was supported by the Spanish Ministry of Science and Innovation / Spanish Research Agency (MCI/AEI/FEDER, RTI2018-096182-B-I00, CSIC13-4E-2076) and AGAUR (2017 SGR 208).

Keywords: dynamic combinatorial chemistry • extracellular matrix • glycosaminoglycans • molecular recognition • on-cell NMR spectroscopy

[1] a) P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J.-L. Wietor, J. K. M. Sanders, S. Otto, *Chem. Rev.* **2006**, *106*, 3652-3711; b) J.-M. Lehn, A. V. Eliseev, *Science* **2001**, *291*, 2331-2332.

[2] J.-M. Lehn, *Chem. Soc. Rev.* **2007**, *36*, 151-160.

[3] J.-M. Lehn, in *Constitutional Dynamic Chemistry* (Ed.: M. Barboiu), Springer Berlin Heidelberg, Berlin, Heidelberg, **2012**, pp. 1-32.

[4] a) M. Jaegle, E. L. Wong, C. Tauber, E. Nawrotzky, C. Arkona, J. Rademann, *Angew. Chem. Int. Ed.* **2017**, *56*, 7358-7378; *Angew.Chem.* **2017**, *129*, 7464-7485; b) M. Mondal, A. K. H. Hirsch, *Chem. Soc. Rev.* **2015**, *44*, 2455-2488.

[5] A. M. Whitney, S. Ladame, S. Balasubramanian, *Angew. Chem. Int. Ed.* **2004**, *43*, 1143-1146; *Angew. Chem.* **2004**, *116*, 1163-1166.

[6] S. Otto, *Acc. Chem. Res.* **2012**, *45*, 2200-2210.

[7] A. Herrmann, *Chem. Soc. Rev.* **2014**, *43*, 1899-1933.

[8] F. B. L. Cougnon, J. K. M. Sanders, *Acc. Chem. Res.* **2012**, *45*, 2211-2221.

[9] C. Frantz, K. M. Stewart, V. M. Weaver, *J. Cell Sci.* **2010**, *123*, 4195-4200.

[10] L. Xu, L. Tang, L. Zhang, in *Progress in Molecular Biology and Translational Science, Vol. 162* (Ed.: L. Zhang), Academic Press, **2019**, pp. 59-92.

[11] a) T. Pap, J. Bertrand, *Nat. Rev. Rheumatol.* **2013**, *9*, 43-55; b) M. Schnabelrauch, D. Scharnweber, J. Schiller, *Curr. Med. Chem.* **2013**, *20*, 2501-2523.

[12] A. Pudelko, G. Wisowski, K. Olczyk, E. M. Kozma, *FEBS J.* **2019**, *286*, 1815-1837.

[13] A. H. Bartlett, P. W. Park, *Expert Rev. Mol. Med.* **2010**, *12*, e5.

[14] N. K. Karamanos, Z. Piperigkou, A. D. Theocharis, H. Watanabe, M. Franchi, S. Baud, S. Brézillon, M. Götte, A. Passi, D. Vigetti, S. Ricard-Blum, R. D. Sanderson, T. Neill, R. V. Iozzo, *Chem. Rev.* **2018**, *118*, 9152-9232.

[15] A. D. Theocharis, S. S. Skandalis, C. Gialeli, N. K. Karamanos, *Adv. Drug Deliv. Rev.* **2016**, *97*, 4-27.

[16] M. Corredor, D. Carbajo, C. Domingo, Y. Pérez, J. Bujons, A. Messeguer, I. Alfonso, *Angew. Chem. Int. Ed.* **2018**, *57*, 11973-11977; *Angew.Chem.* **2018**, *130*, 12149-12153.

[17] A. De Robertis, C. De Stefano, A. Gianguzza, S. Sammartano, *Talanta* **1999**, *48*, 119-126.

[18] J. L. Asensio, A. Ardá, F. J. Cañada, J. Jiménez-Barbero, *Acc. Chem. Res.* **2013**, *46*, 946-954.

[19] M. P. Rangel, V. K. de Sa, T. Prieto, J. R. M. Martins, E. R. Olivieri, D. Carraro, T. Takagaki, V. L. Capelozzi, *Glycoconj. J.* **2018**, *35*, 233-242.

[20] N. Fujitani, J.-i. Furukawa, K. Araki, T. Fujioka, Y. Takegawa, J. Piao, T. Nishioka, T. Tamura, T. Nikaido, M. Ito, Y. Nakamura, Y. Shinohara, *Proc. Natl. Acad. Sci.* **2013**, 201214233.

[21] X. Guo, I. S. Han, V. C. Yang, M. E. Meyerhoff, *Anal. Biochem.* **1996**, *235*, 153-160.

[22] D. E. Humphries, J. E. Silbert, *Biochem. Biophys. Res. Commun.* **1988**, *154*, 365-371.

[23] D. I. Freedberg, P. Selenko, in *Ann. Rev. Biophys.* **2014**, *43*, 171-192.

[24] W. Ling, R. R. Regatte, M. E. Schweitzer, A. Jerschow, *NMR Biomed.* **2008**, *21*, 289-295.

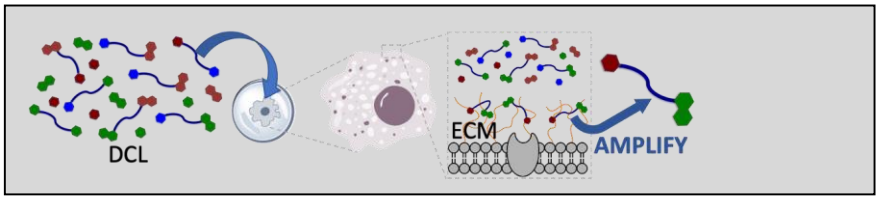
[25] M. Liu, J. K. Nicholson, J. C. Lindon, *Anal. Chem.* **1996**, *68*, 3370-3376.

[26] C. Dalvit, G. Fogliatto, A. Stewart, M. Veronesi, B. Stockman, *J. Biomol. NMR* **2001**, *21*, 349-359.

[27] W. Wang, N. Han, R. Li, W. Han, X. Zhang, F. Li, *Anal. Chem.* **2015**, *87*, 9302-9307.

Entry for the Table of Contents

Insert graphic for Table of Contents here.



Whole living cells were used as template in a dynamic combinatorial chemistry system. The amplified member from the dynamic library showed strong and selective binding to the main glycosaminoglycan of the extracellular matrix of the corresponding cell. This work demonstrates the power of dynamic covalent screening in chemical biology, even in highly challenging biorelevant media.

Institute and/or researcher Twitter usernames:

Institute: @IQAC_CSIC,

Ignacio Alfonso: @nacho_alfonso,

Yolanda Pérez: @yolandapereznmr

Table of Contents:

Materials and Methods

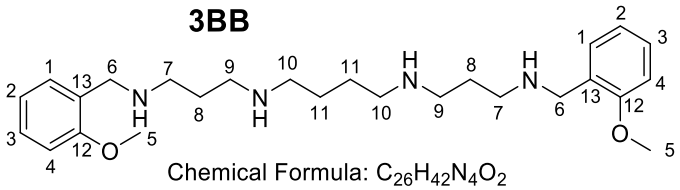
Materials

Reagents and solvents were purchased from commercial suppliers (Aldrich, Fluka or Merck) and were used without further purification.

Synthesis of **3AL*:*** polyamine **3AL** was synthesized as previously described[1].

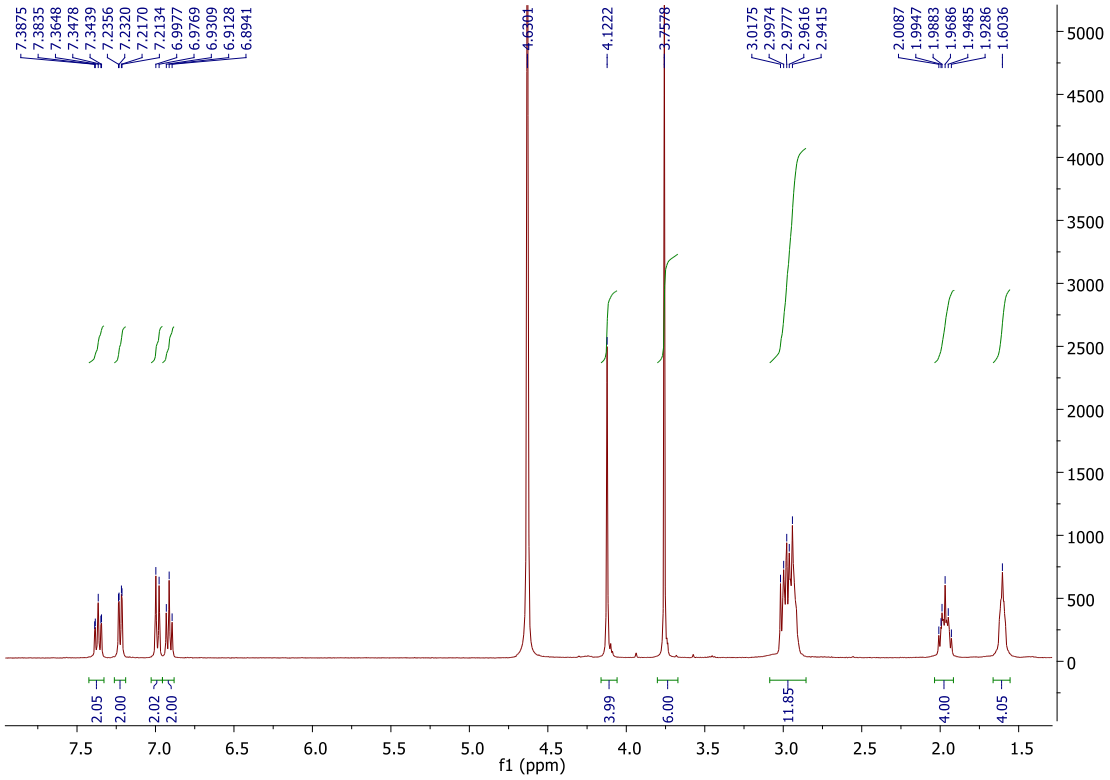
Synthesis of **3BB**: Spermine (70 mg, 0.35 mmol) was dissolved in 25 mL of anhydrous THF at 0°C. Then, *o*-anisaldehyde (17 mg, 0.7 mmol) was added dissolved in 7 mL of anhydrous THF. The solution was stirred overnight. Then, $NaBH_3CN$ (87 mg, 1.4 mmol) was added and the reaction was stirred 24h. After addition of H_2O (2 mL) and 1M HCl (2 mL), the reaction was stirred for 1h. THF was evaporated in vacuum. Reaction mixture was purified by reverse phase chromatography with a gradient of ACN (1% TFA) and water (1% TFA) to yield 24 mg (16%) of pure product (>99% by HPLC).

3BB

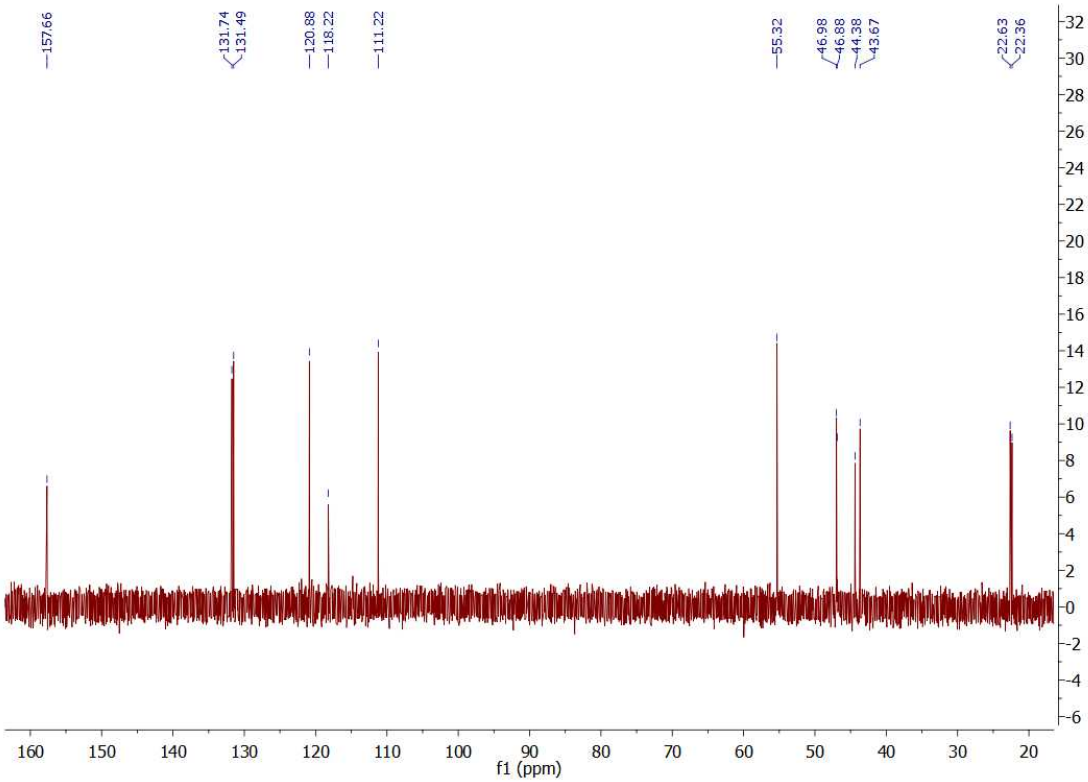


Chemical Formula: $C_{26}H_{42}N_4O_2$
Exact Mass: 442.3308
Molecular Weight: 442.6480

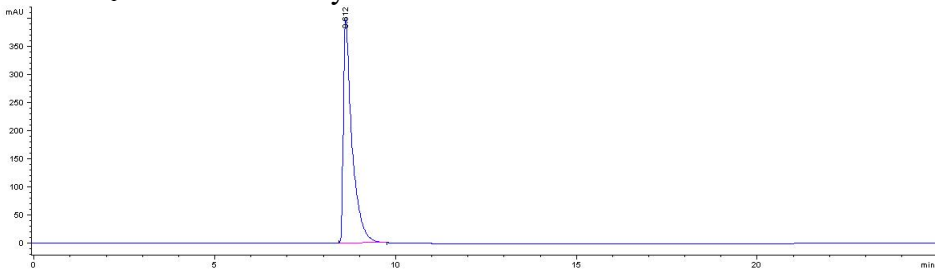
^{1}H RMN (H_2O/D_2O): δ 7.37 (ddd; 2H; J_1=7.5Hz, J_2=8.3Hz, J_3=1.4Hz; H3), 7.22 (dd; 2H, J_4=7.5Hz, J_3= 1.4Hz; H1), 6.99 (d; 2H; J_2=8.3Hz; H4), 6.91 (dd; 2H, J_1=7.5Hz; J_4=7.5Hz; H2), 4.10 (s; 4H, H6), 3.74 (s; 6H, H5), 2.98 (m, 12H, H7, H9, H10), 1.97 (m, 4H, H8), 1.6 (m, 4H, H11).



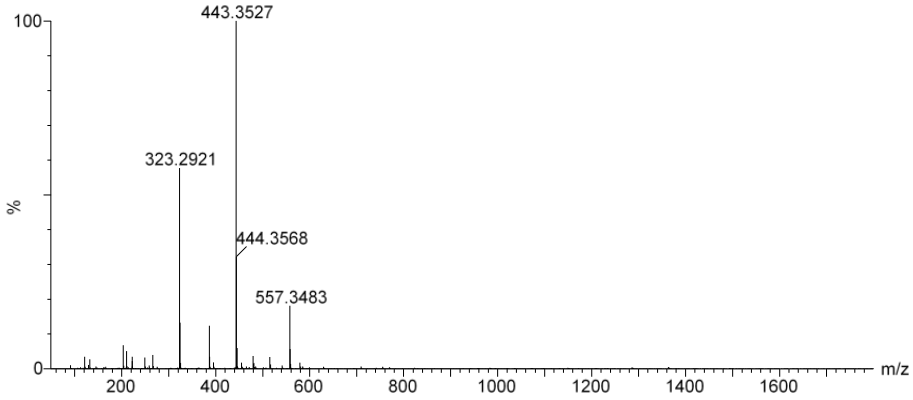
^{13}C RMN (H_2O/D_2O): δ 157.7 (C12), 131.7 (C1), 131.5 (C3), 120.9 (C13), 118.2 (C2), 111.2 (C4), 55.3 (C5), 47.0 (C10), 46.9 (C6), 44.4 (C9), 43.7 (C7), 22.6 (C11), 22.4 (C8).



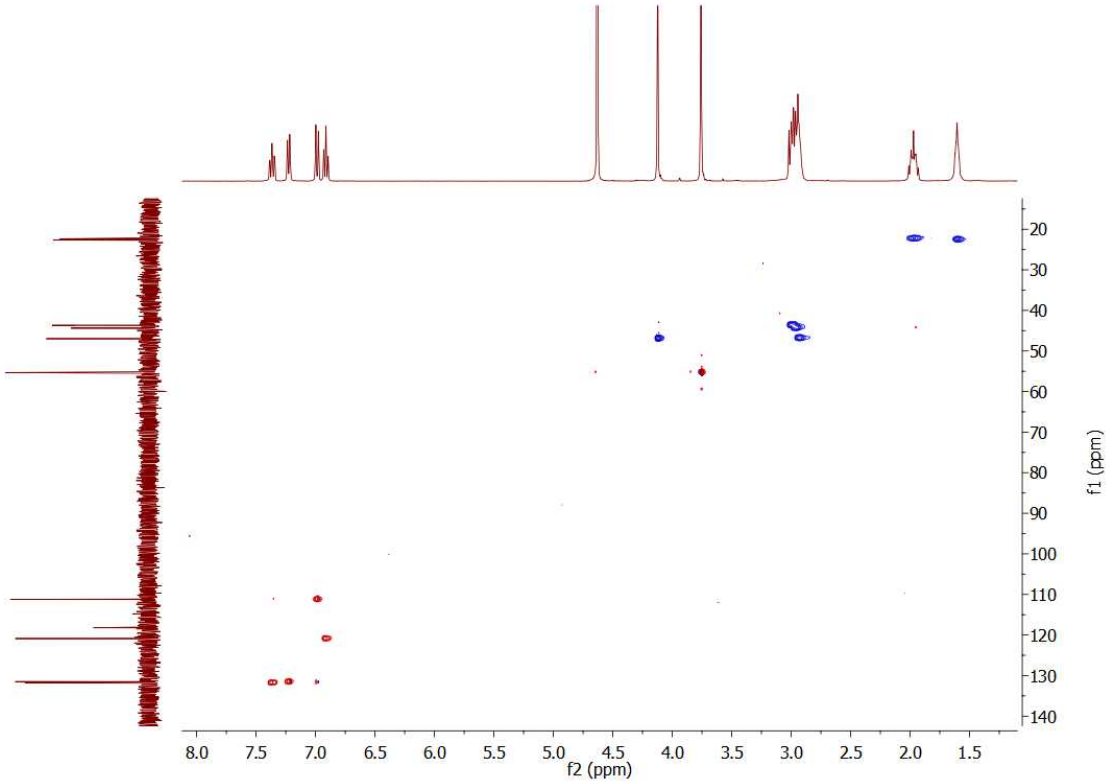
HPLC: t_r => 8.6min Purity >99% at 220 nm.



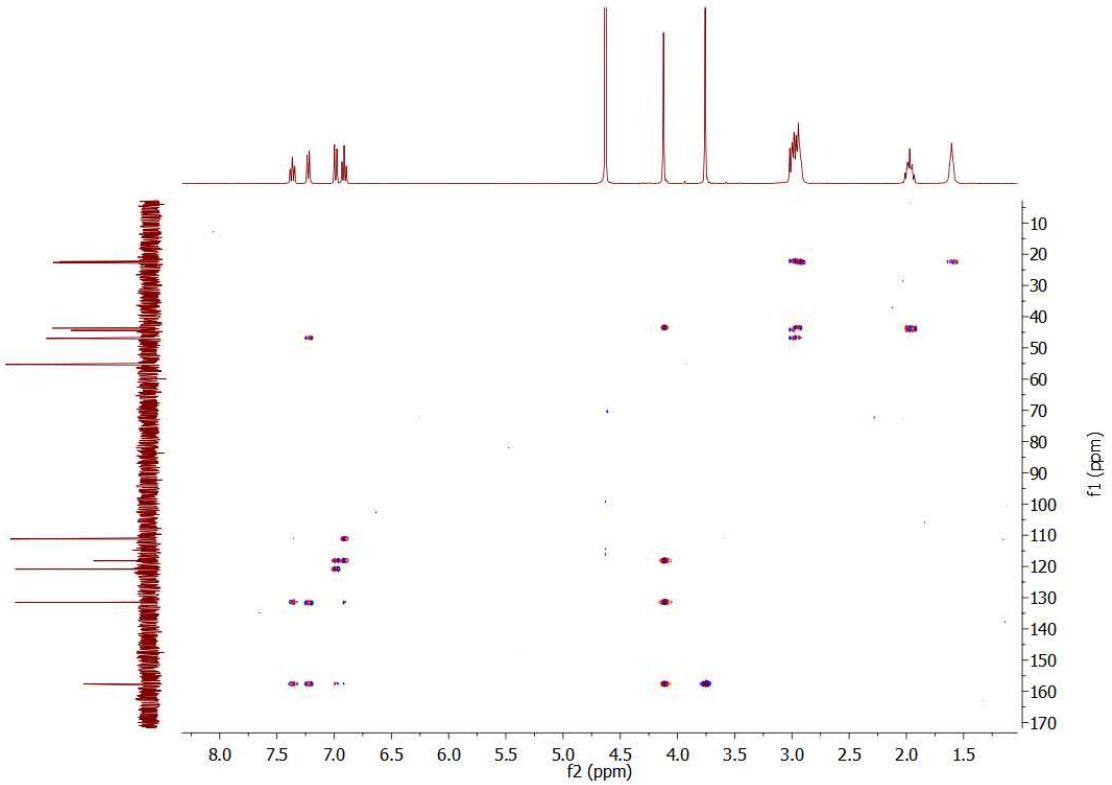
MS: Calculated MS for $(M+H)^+$ $C_{26}H_{43}N_4O_2$ 443.3381; found 443.3527



HSQC:



HMBC:



Methods:

Cell growth: A549 and HeLa cells were maintained in DMEM (4500 mg/mL glucose) culture medium (Sigma) containing 10% fetal calf serum (FCS), 2 mM glutamine, 50 u/mL penicillin, and 0.05 g/mL streptomycin, at 37°C under 5% CO_2 atmosphere. For experiments in cell medium doped with 50 mM $NaClO_3$, cells were grown in chlorate supplemented medium during 72h in advance.

General DCC reaction: Exponentially growing cells were detached from the culture flasks using a trypsin-0.25% EDTA solution. The cell suspension was seeded on 25 mm plates (Corning) at 21.4 x 10^3 cells/cm^2. The cells were incubated 24h. Then, the cell media was replaced with PBS (x3) and the imine mixture was added as described just below.

A stock solution of imines was prepared by dissolving the proper amount of spermine and the corresponding aldehydes in water rendering a final concentration of 7.5 mM in all reagents. The solution was shaken overnight. Then, two reaction mixtures were prepared in different falcon centrifuge tubes by mixing 50 µL of the stock solution with 450 µL of PBS, they were added to the 25 mm plates (Corning) either with cells or without them. Immediately, $NaBH_3CN$ from a stock solution at 50 mM was added, rendering a final concentration of 1 mM of the reducing reagent. Plates were incubated during 1h. After that time, medium was extracted to Eppendorfs and reaction was stopped by the addition of 50 µL of 1M HCl. Reaction mixture was then centrifuged (5 min at 6 Krpm). Supernatant was kept, two-fold diluted with water and analyzed by UPLC-MS. The assignment of the peaks observed in the reactions was done on the basis of the m/z values and confirmed by injection of samples obtained from deconvoluted sublibraries. The amplification factors were calculated by dividing the normalized areas of the corresponding UPLC peaks in the presence of cells (A_T) by the areas obtained in the absence of cells (A_o).

DCC in the presence of protamine: The DCC procedure was performed as previously described, but 500 µM of protamine (from a stock solution at 25 mM) was added before the addition of imines.

DCC with cells grown in the presence of $NaClO_3$: The DCC procedure was performed as previously described, but all cell media was supplemented with 50 mM $NaClO_3$ 72h in advance of seeding the cells into the 25 mm plates.

MTT Assays: The viability of A549 cells was tested using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay. In it, exponentially growing cells were detached from the culture flasks using a trypsin-0.25% EDTA solution, and the cell suspension was seeded onto 96-well (Nunclon) at a concentration of 7k cells/well. Toxicity of aldehydes, spermine, **3AL** and $NaBH_3CN$ was tested as follows: 24h after seeding, culture medium was discarded and replaced by compound solution in PBS that were diluted with cell culture medium to their final concentration ([aldehydes, Spermine] = 0.75 mM, [$NaBH_3CN$] = 1 mM, [**3AL**] = 50 to 500 µM). After 1h of incubation at 37°C under 5% CO_2 atmosphere, solvent was discarded and replaced with MTT (0.5 mg/mL). After 2 hours of incubation with MTT, the medium was discarded by aspiration and DMSO was added to dissolve formazan, a dark blue colored crystal observed in the wells. Absorbance was measured at 570 nm in a spectrophotometric Biotek Sinergy 2 Microplate Readed (Agilent), 30 minutes after the addition of DMSO. Cell viability is expressed as an absorbance percent ratio of cells treated with compound to untreated cells, which were used as a control. Results are the average from three independent experiments. Toxicity of imines and reaction conditions was tested as follows: 24h after seeding, culture medium was discarded and replaced with a solution of either the reaction cocktail or just the imine mixture ([imines]=0.75 mM each) in PBS. After 1h of

incubation, solvent is discarded and cells are treated with MTT (0.5mg/mL) as previously described.

A549 cells preparation for NMR experiments: Cells were normally grown until they reach 80-85x10^6 cells, enough to completely fill the Shigemi NMR tube (Fig. S2). Then, culture flasks were washed with PBS and cells were detached using Versene (1:5000, Thermofisher), the cells were centrifuged and resuspended in 0.5xPBS (x3). The final pellet was resuspended in 500-600 µl 0.5XPBS and passed to a Shigemi tube for solution NMR. The cells were centrifuged to the bottom of the NMR tube using a hand-driven centrifuge, the excess of PBS was eliminated and 10% of D2O (in volume) was added for deuterium locking purpose. Once the round of NMR experiments had finished, the stability of the cells was tested by Trypan Blue.

A549 cells solution state NMR experiments: Experiments were carried out on a Bruker Avance^{TM} III HD spectrometer operating at 500 MHz (^{1}H resonance frequency) and equipped with a 5mm TCI (^{1}H/^{13}C/^{15}N) cryoprobe. NMR spectra were acquired and processed using Bruker TopSpin 3.5 and MNova software, respectively. All NMR spectra were acquired at 37ºC and the cells were inside the NMR tube/magnet not more than 5 hours. The setup included locking (90% H_2O and 10% D_2O), tuning and matching, shimming, and calibration of the (^{1}H) pulse for every cell suspension sample. The pulse programs used for spectra acquisition were zggpw5 (1D ^{1}H with water suppression using watergate W5 pulse sequence), cmpgpr1D (1D ^{1}H T2-filtered, 200-ms filter, and hsqcetgpprsisp2.2 (2D ^{1}H-^{13}C HSQC with presaturation during relaxation delay) from Bruker pulse sequence library.

Trypan Blue assay: Cell viability of cell suspension in PBS was determined by Trypan Blue assay. 10 µL of the cell suspensions was pipetted to an Eppendorf and 10 µL of the Trypan Blue solution (0.4%, Sigma) were added. The Eppendorf was gently shaken 1 minute. Then, non-viable (stained) and viable (non-stained) cells were counted. Stability is expressed as percent ratio of viable cells to total number of cells.

Methylene Blue staining: A549 cells were normally grown in Nunc^{TM} Glass Botton Dishes until they reached roughly 70% confluence. Then, they were treated with 0.05% methylene blue in PBS w/o presence of 100 µM **3AL** during 5min. After thorough cleaning with PBS, pictures of live cells in PBS were taken using a Nikon Eclipse TS100 microscope equipped with a Nikon DS-2Mv Digital Sight camera, using Nikon Imaging Software (NIS) Elements F v4.3 within the next 30 minutes after staining. Pictures were taken with 1600x1600 resolution, using a 1.4xgain filter with 25 ms exposure and the Default contrast provided by the imaging software. These conditions were kept constants to allow the comparison between different experiments. Although difference in staining could easily be spotted, comparison between cell color was determined by following an adapted method from reference[3]. Image processing with ImageJ software was used to extract the RGB intensity of the cells within each condition (n>80). As brighter cells would just have higher intensity values, comparison of the different conditions has been performed by looking at the proportion of each RGB color in each case. Stained cells are more bluish/greenish than non-stained ones while presence of **3AL** reduces this color apparition. Same process was done with background to avoid artifacts produced by non-specific staining of the whole plate.

Surface Plasmon Resonance (SPR): Affinity experiments between **3AL**/**3BB** and chondroitin sulfate were performed on an Open SPR^{TM} (Nicoya). All measurements were performed at 25°C using a working buffer of 25 mM Tris at pH 7.5. Biotin-loaded sensor chips (NICOYA) were further functionalized with streptavidine (50 µg/mL) and later with biotin-chondroitin sulfate (50 µg/mL). Binding experiments to chondroitin were performed by injecting **3AL**/**3BB** at desired concentrations and at a rate of 40 µL/min. Between binding

assays, the surface was regenerated by exposure to an injection of 10 mM HCl. Fitting has been performed by Trace Drawer software using a 'one-to-one two-state algorithm', which considers a possible interaction into two different independent reacting places. Results obtained from three independent experiments at three different concentrations of the ligands were fit globally to render the corresponding on/off rate constants and the dissociation constants (K_d). The somehow large errors in the determined K_d can be a consequence of the rough approximations assumed with the applied binding mode and also due to the polydispersity of the chondroitin sulfate sample. Anyway, since the difference in the corresponding dissociation constants largely exceeds the estimated errors, we are confident that they can be used for comparison purposes. SPR experiments have been performed with the support of the ICTS "NANBIOSIS", more specifically by the Unit U2, Custom Antibody Service (CAbS, CIBER-BBN, IQAC-CSIC).

Biotin-Chondroitin Sulfate: Chondroitin sulfate was purchased from SIGMA (Chondroitin sulfate A sodium salt from bovine trachea, purchased as a mixture of polyanion chains, with most chains in the order of 10kDa), sulfosuccinimidyl 6-(biotinamido)hexanoate was obtained from Merck. GAG biotin-functionalization was performed as previously reported[2].

Molecular Dynamics Simulation Methods: All molecular simulations were carried out with the package Schrödinger Suite 2019[4], through its graphical interface Maestro[5]. The program Macromodel[6], with its default force field OPLS3[7] and GB/SA water solvation conditions[8], was used for energy minimization. Molecular dynamics simulations were performed with the program Desmond[9,10], using the OPLS3 force field. A chondroitin-A hexadecamer model (dp16 CDS-A) was built from the reported crystal structure of a chondroitin sulphate A hexamer (PDB 1C4S)[11]. Ligands **3AL** and **3BB** were built within Maestro. The dp16 CDS-A molecule was modelled with all its sulfate and carboxylate groups in their ionized state (total charge -16), while the **3AL** and **3BB** ligands were modelled with their four amino groups protonated. All structures were energy minimized before building the simulation systems. Since each CDS-A oligomer can potentially bind several molecules of the ligands, systems composed of a dp16 CDS-A molecule plus five unbound and randomly placed **3AL** or **3BB** molecules were set up in a 100x100x100 Å cubic box, with added Cl^- ions to reach neutrality and the whole system solvated with TIP3P water (~32600 TIP3P molecules), using the System Builder of the Maestro-Desmond interface[12]. Similarly, a system to analyze the competence between **3AL** and **3BB** for binding to CDS-A was built starting from a snapshot of the CDS-A/**3BB** simulation where the five **3BB** molecules were bound to CDS-A. After removing the waters and ions from this system, 5 additional unbound molecules of **3AL** were randomly placed around it and then it was resolvated as before.

Simulations were performed as previously described[1]. Thus, the full systems (~99000 atoms) were initially subjected to steepest descent minimization, first with the solute restrained and then without restraints until a gradient threshold of 0.1 kcal/mol/Å was reached. Then they were heated stepwise to 300 K with short MD runs under periodic boundary conditions (PBC, NPT, Berendsen thermostat-barostat) (12 ps at 0.1, 10, 100 and 300 K), and equilibrated for 2 ns at the last temperature and 1.0 bar in the NPT ensemble. Production MD simulations (500 ns, 2 fs time step) were performed under the same conditions (PBC, NPT ensemble, 300 K and 1.0 bar) using the Nose-Hoover thermostat method[13-14] with a relaxation time of 1.0 ps and the Martyna-Tobias-Klein barostat method[15] with isotropic coupling and a relaxation time of 2 ps. Integration was carried out with the RESPA integrator[16] using time steps of 2.0, 2.0, and 6.0 fs for the bonded van der Waals and short range and long range electrostatic interactions, respectively. A cut-off of 9 Å was applied to van der Waals and short-range electrostatic interactions, while long-range electrostatic interactions were computed using the smooth

particle mesh Ewald method with an Ewald tolerance of 10^{-9} [17,18]. Bond lengths to hydrogen atoms were constrained using the Shake algorithm[19]. Coordinates were saved every 200 ps, hence 2500 snapshots (frames) were obtained for each simulation. The Simulation Event Analysis application included in the Desmond-Maestro interface and different ad-hoc scripts were used to analyze the simulations results. The 2500 snapshots were clustered based on the atomic RMSD of the heavy atoms using the hierarchical clustering with average linkage method implemented in the Schrodinger Suite and fixing the number of clusters to 10.

Supplementary Figures

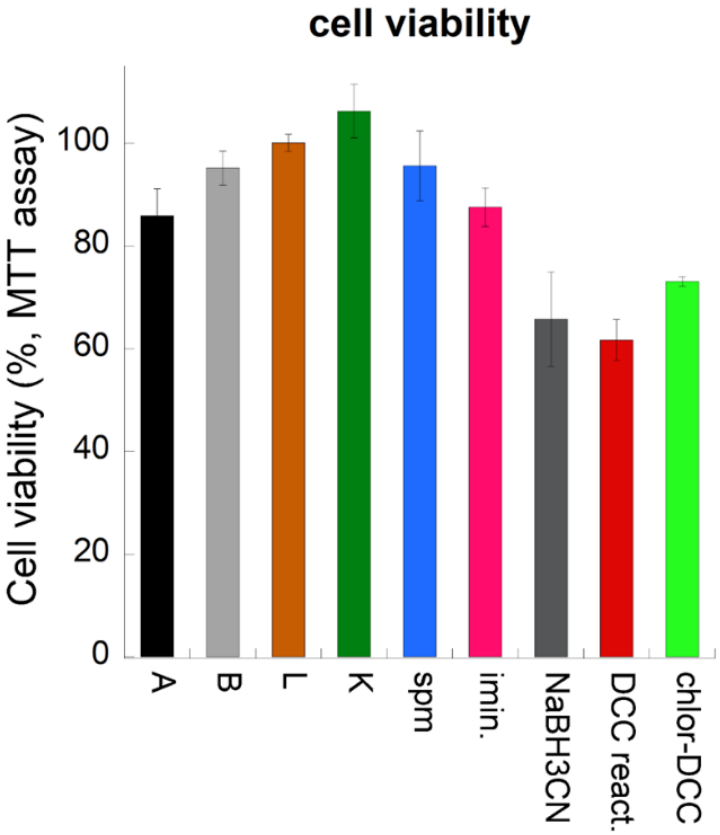


Figure S1. Cell viability (%, MTT assay) under different reaction conditions and molecules: aldehydes **A** (black), **B** (gray), **L** (brown) and **K** (green), spermine (blue), imines reaction (magenta), $NaBH_3CN$ (dark gray), full DCC reaction conditions (red) and full DCC reaction conditions on cells grown in the presence of sodium chlorate (light green).

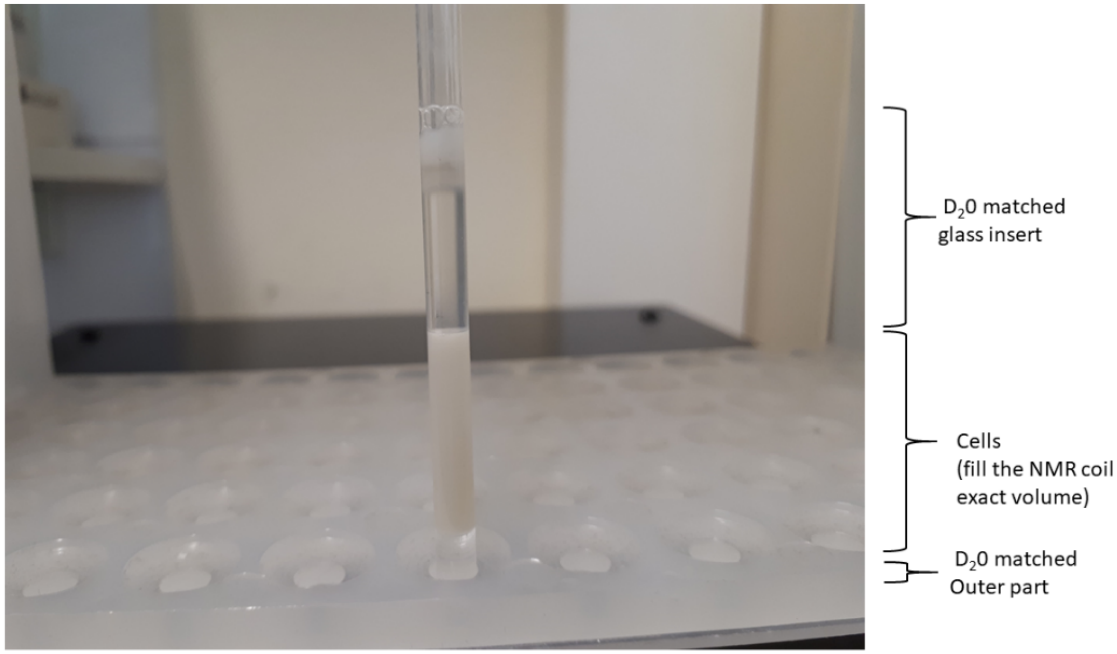


Figure S2. Picture of a sample of A549 cells suspension inside a Shigemi NMR tube.

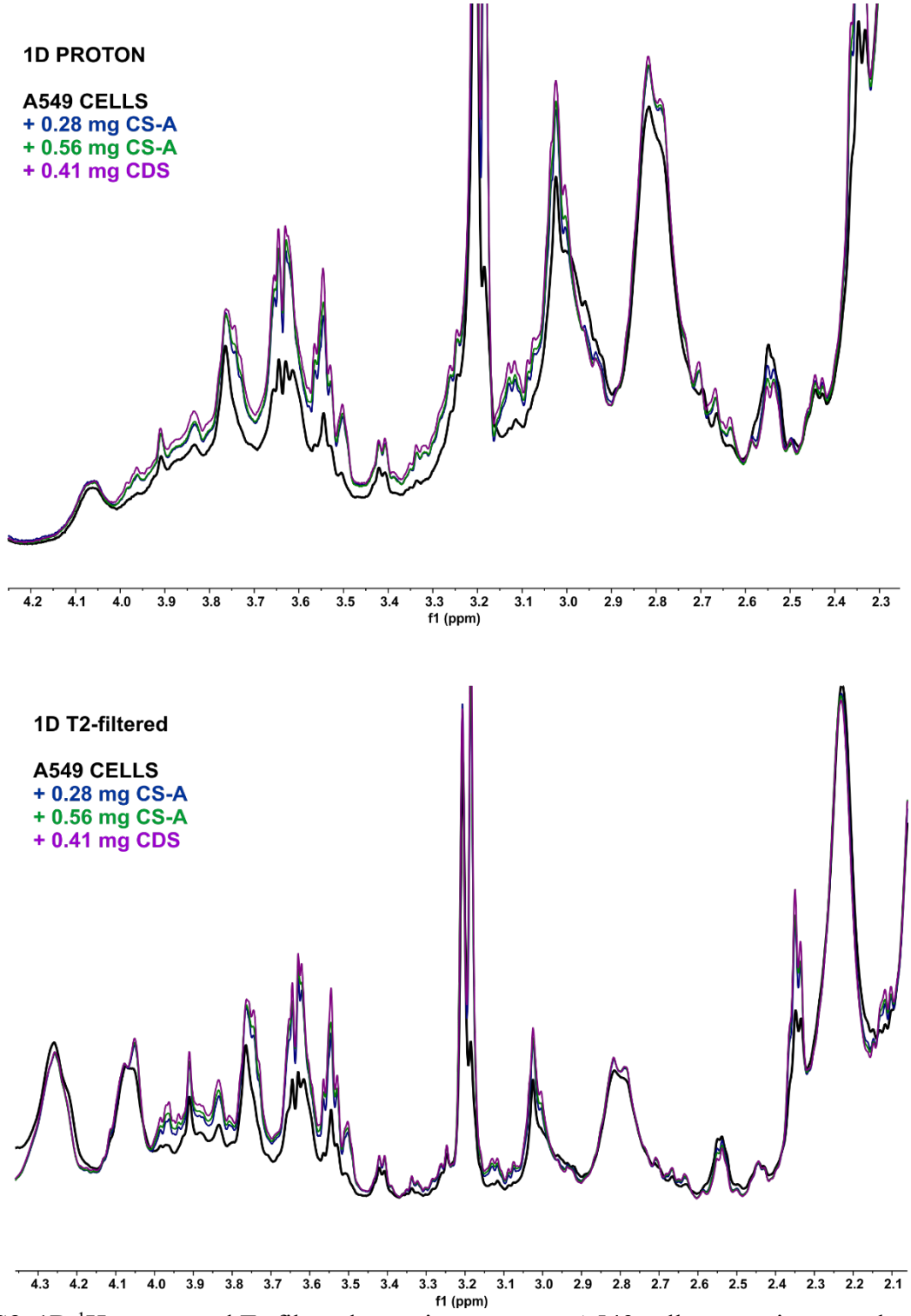


Figure S3. 1D ^{1}H proton and T_2-filtered experiments on an A549 cell suspension sample with glycosaminoglycan spiking: the amount of CDS-A or CDS (from stock concentrated solutions prepared in 0.5XPBS) added into the NMR tube A549 cells suspension is indicated.

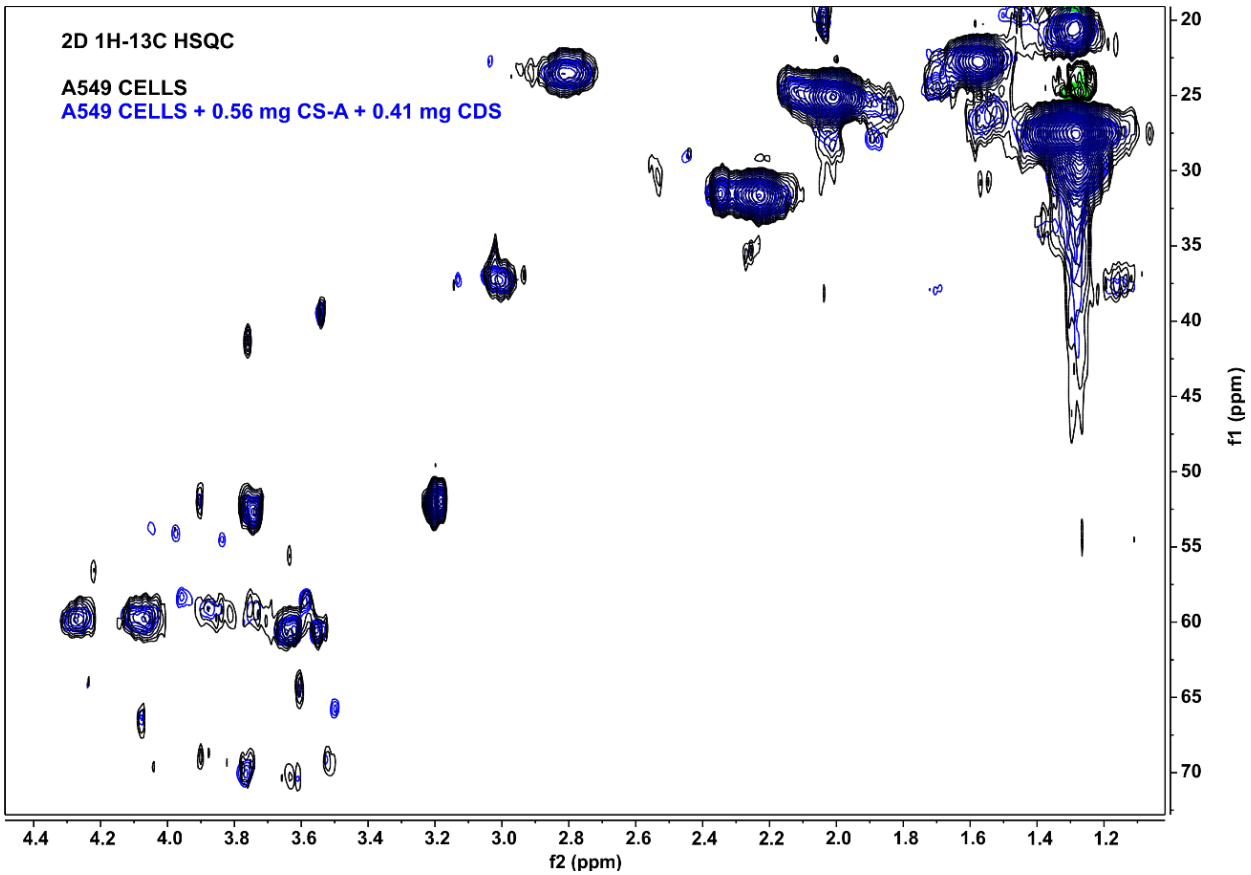


Figure S4. 2D ^{1}H-^{13}C HSQC experiment with glycosaminoglycan spiking: the amount of CDS-A or CDS added (from stock concentrated solutions prepared in 0.5XPBS) to the NMR tube containing the A549 cells suspension is indicated.

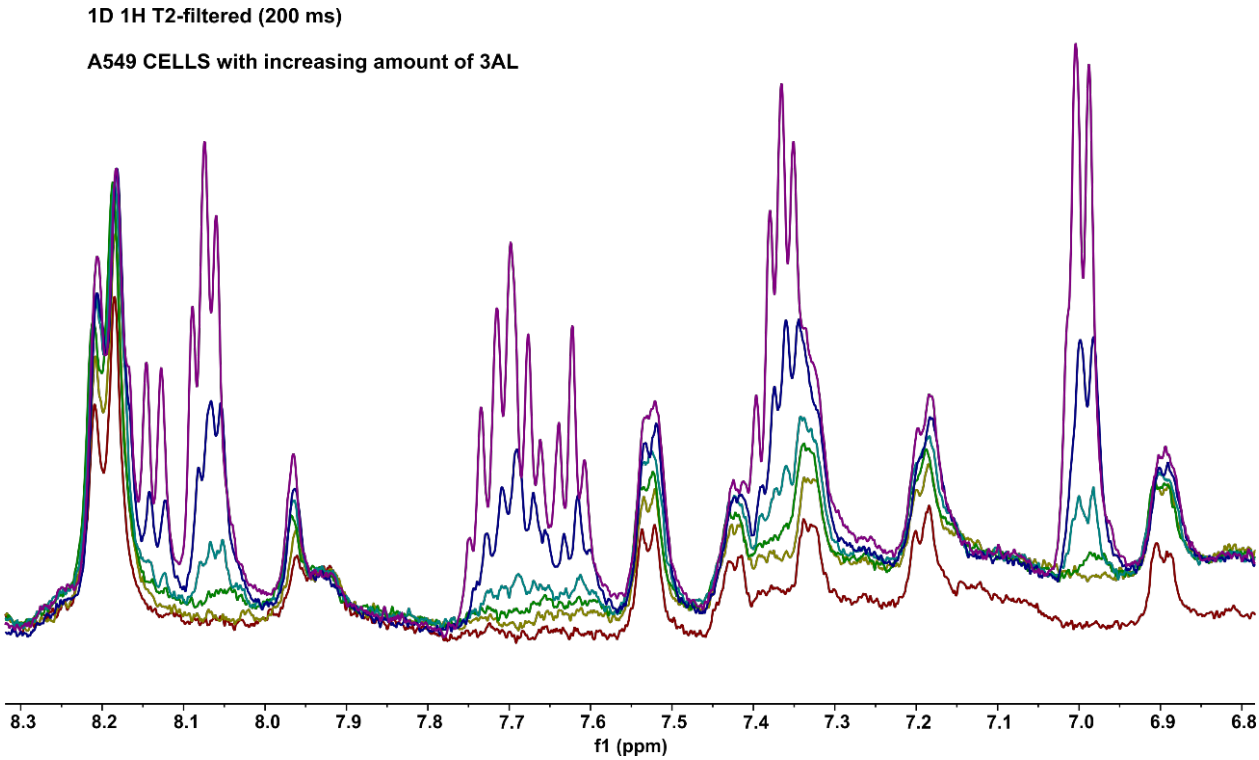


Figure S5. 1D ^{1}H T_2-filtered spectra of **3AL** titration in A549 cells suspension (0.0, 0.8, 2.3, 3.7, 5.1 and 6.4 mM; from a 63 mM stock in DMSO-d6).

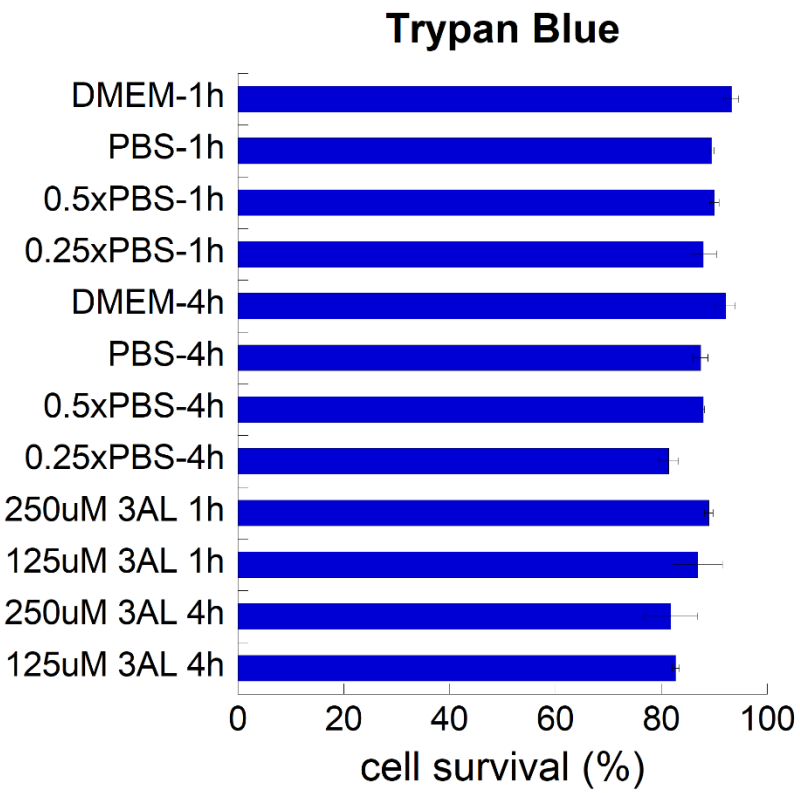
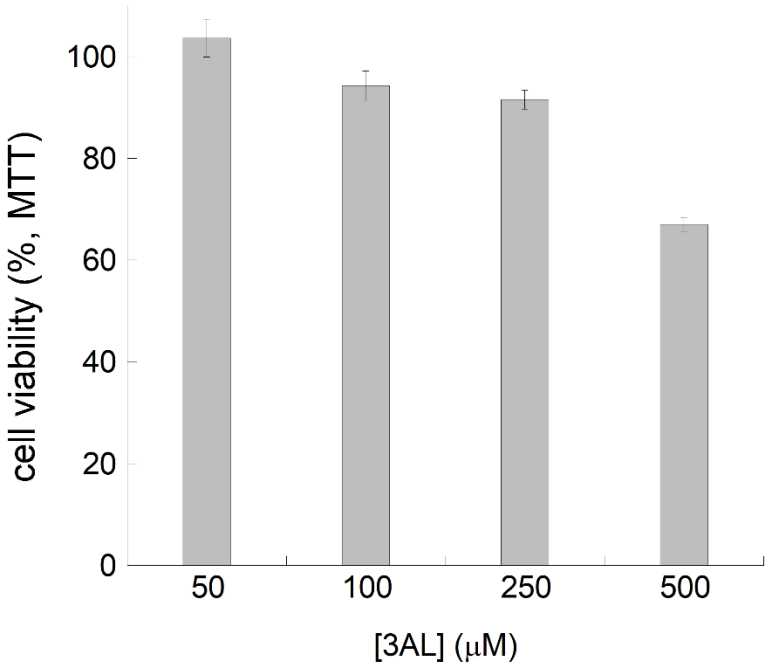


Figure S6. Cell survival (%, Trypan Blue assay) at different experimental conditions similar to those used in the NMR experiments, as determined with Trypan Blue assay (the experiments performed with **3AL** were carried out in 0.5xPBS).



Figuree S7. Cell viability (%, MTT assay) in the presence of different concentrations of **3AL**.

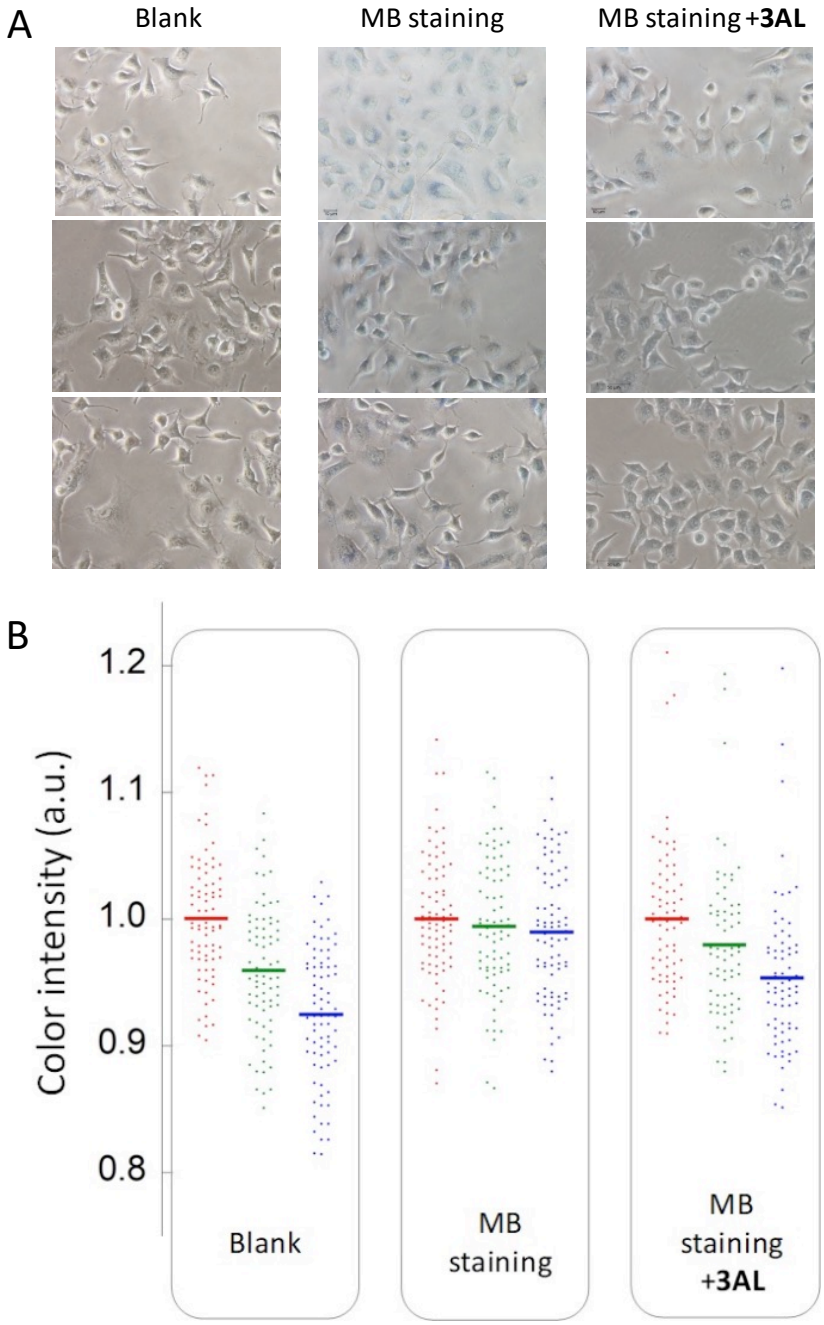


Figure S8. MB staining experiments: (A) Selected pictures of the A549 cells upon different conditions: blank (left column), stained with methylene blue dye (middle), and stained with methylene blue in the presence of 100 μM **3AL** (right). (B) Scatter plot of the color intensities of the cells upon the different staining conditions. The normalized red, green and blue intensities are shown in their respective colors with the solid lines as the mean values.

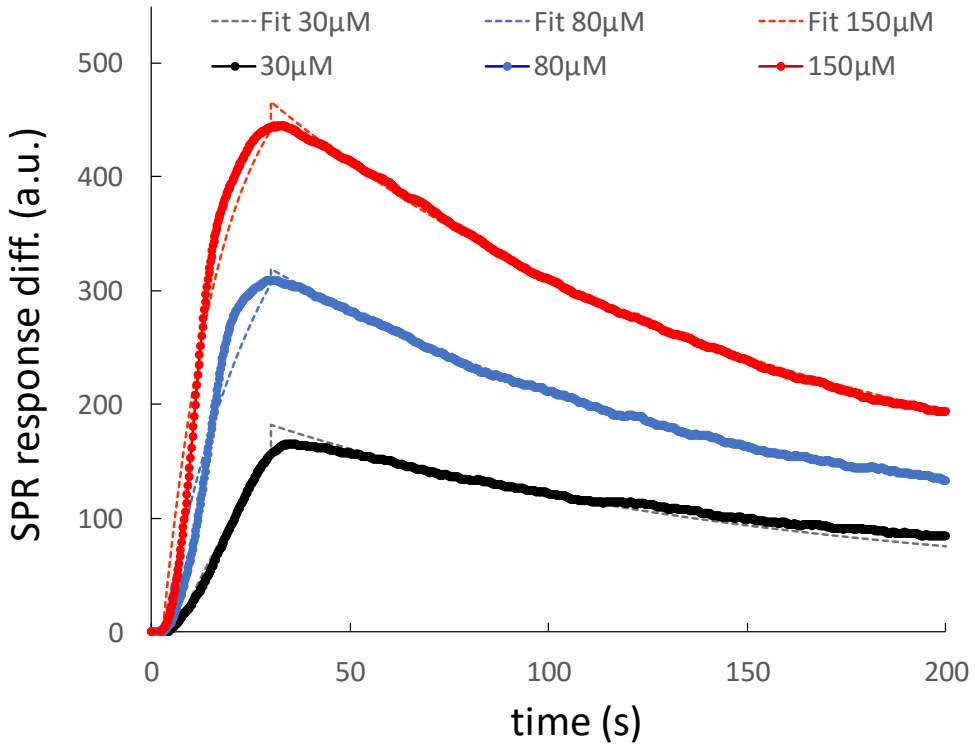


Figure S9. SPR sensograms for the binding of **3AL** to chondroitin sulfate functionalized chips. The concentrations of **3AL** in the flowing solutions were: 30 µM (black), 80 µM (blue) and 150 µM (red). Solid symbols represent experimental values while the dashed line is the best fitting to the 'one-to-one two-state' model. Despite only one experiment per concentration is represented here for clarity, three independent replicates for each concentration were performed with very similar results, and all the curves were globally fitted to converge to a K_d = 2.1±1.0 µM.

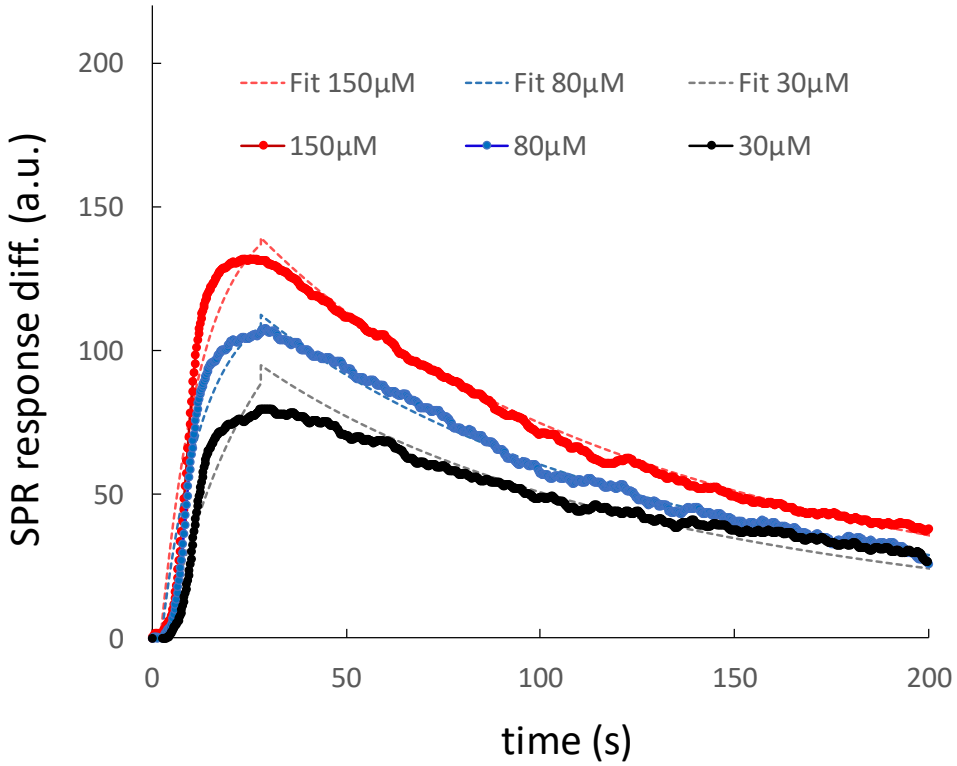


Figure S10. SPR sensograms for the binding of **3BB** to chondroitin sulfate functionalized chips. The concentrations of **3BB** in the flowing solutions were: 30 µM (black), 80 µM (blue) and 150 µM (red). Solid symbols represent experimental values while the dashed line is the best fitting to the 'one-to-one two-state' model. Despite only one experiment per concentration is represented here for clarity, three independent replicates for each concentration were performed with very similar results, and all the curves were globally fitted to converge to a *K*d = 33.6±11.2 µM.

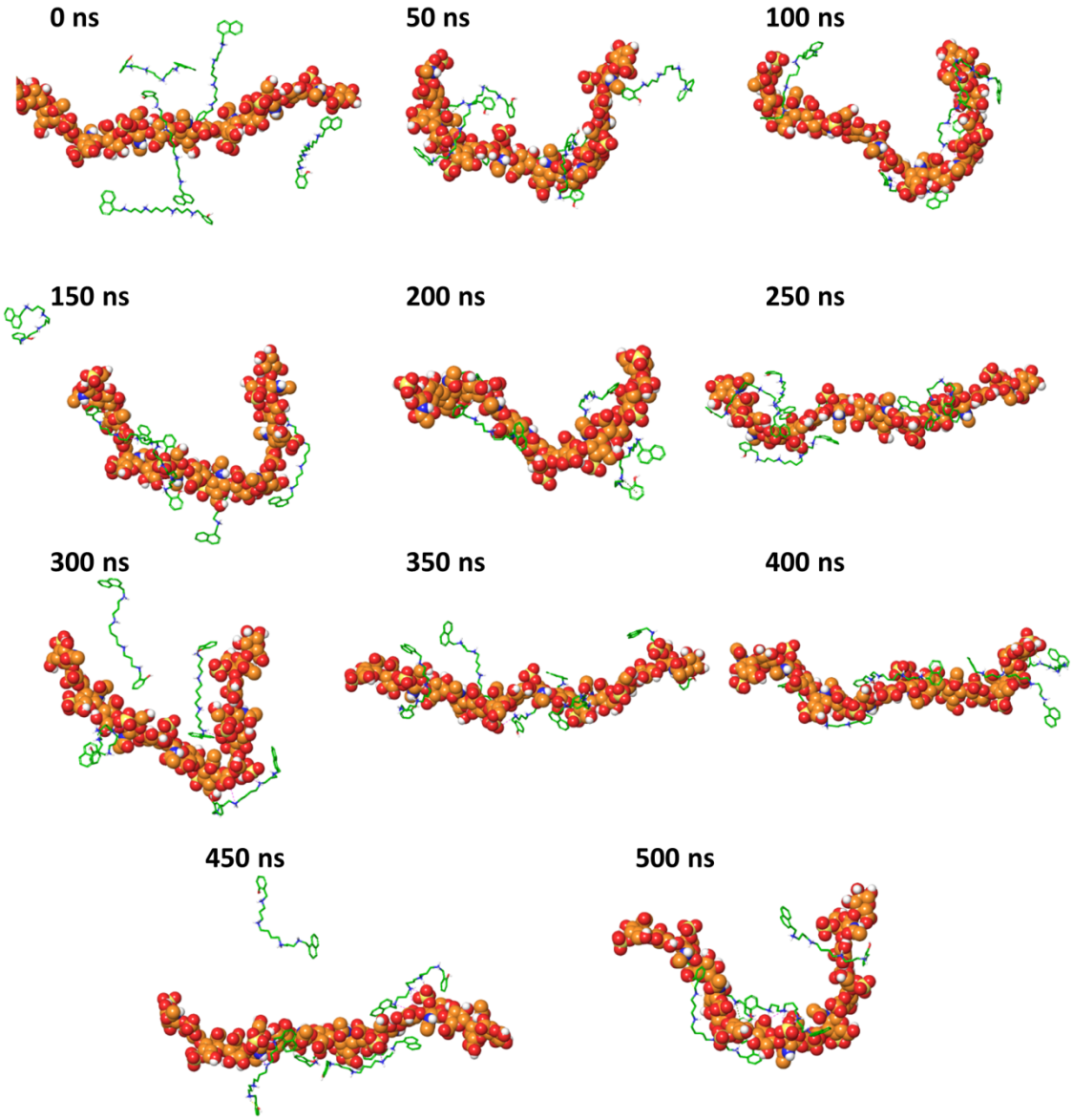


Figure S11. Snapshots from the 500 ns MD simulation of a system containing one dp16 CDS-A model (spheres) and five **3AL** (green sticks) molecules. Water molecules, ions and **3AL** molecules that are too far from CDS-A are omitted for clarity.

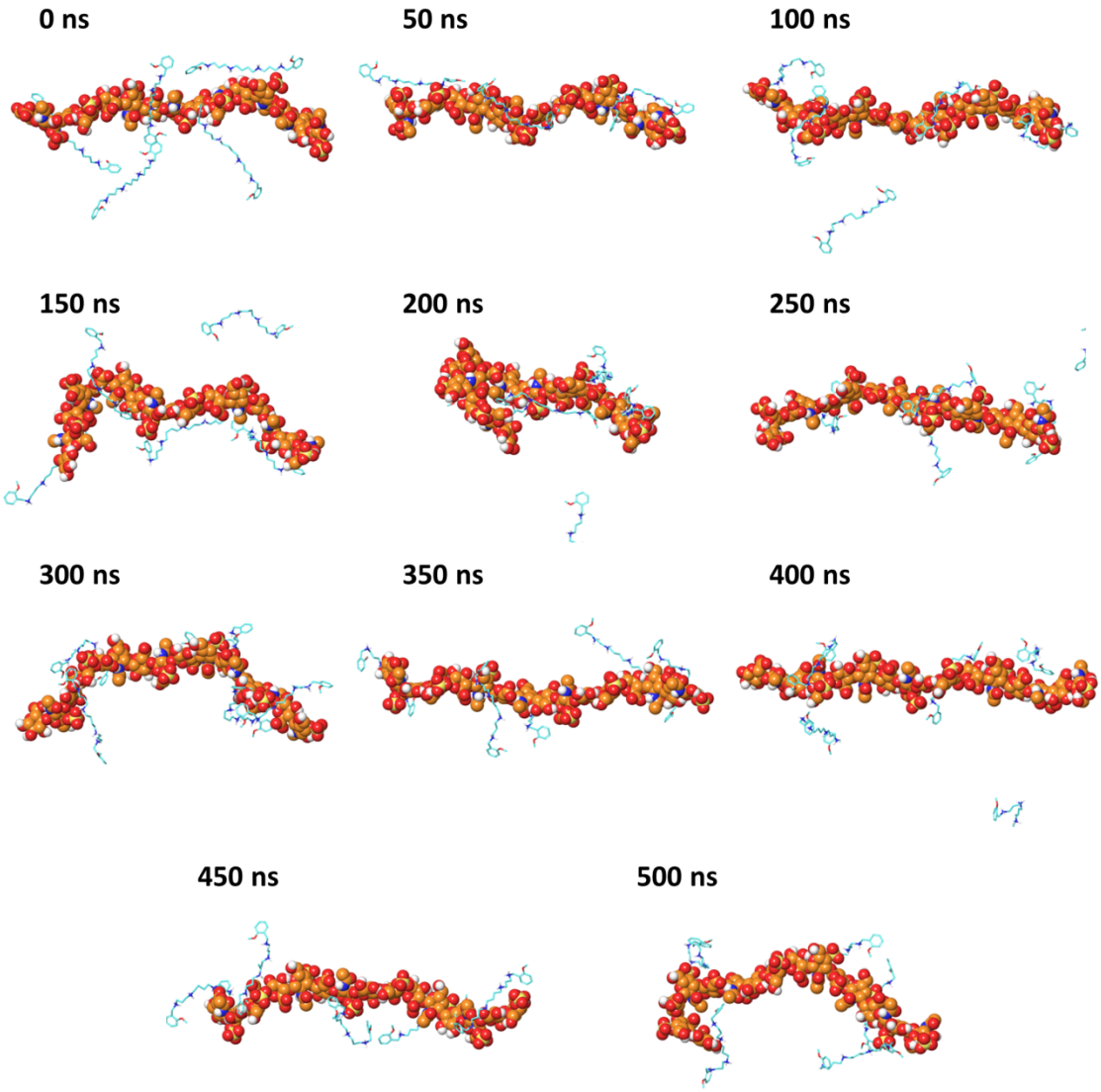


Figure S12. Snapshots from the 500 ns MD simulation of a system containing one dp16 CDS-A model (spheres) and five **3BB** (cyan sticks) molecules. Water molecules, ions and **3BB** molecules that are too far from CDS-A are omitted for clarity.

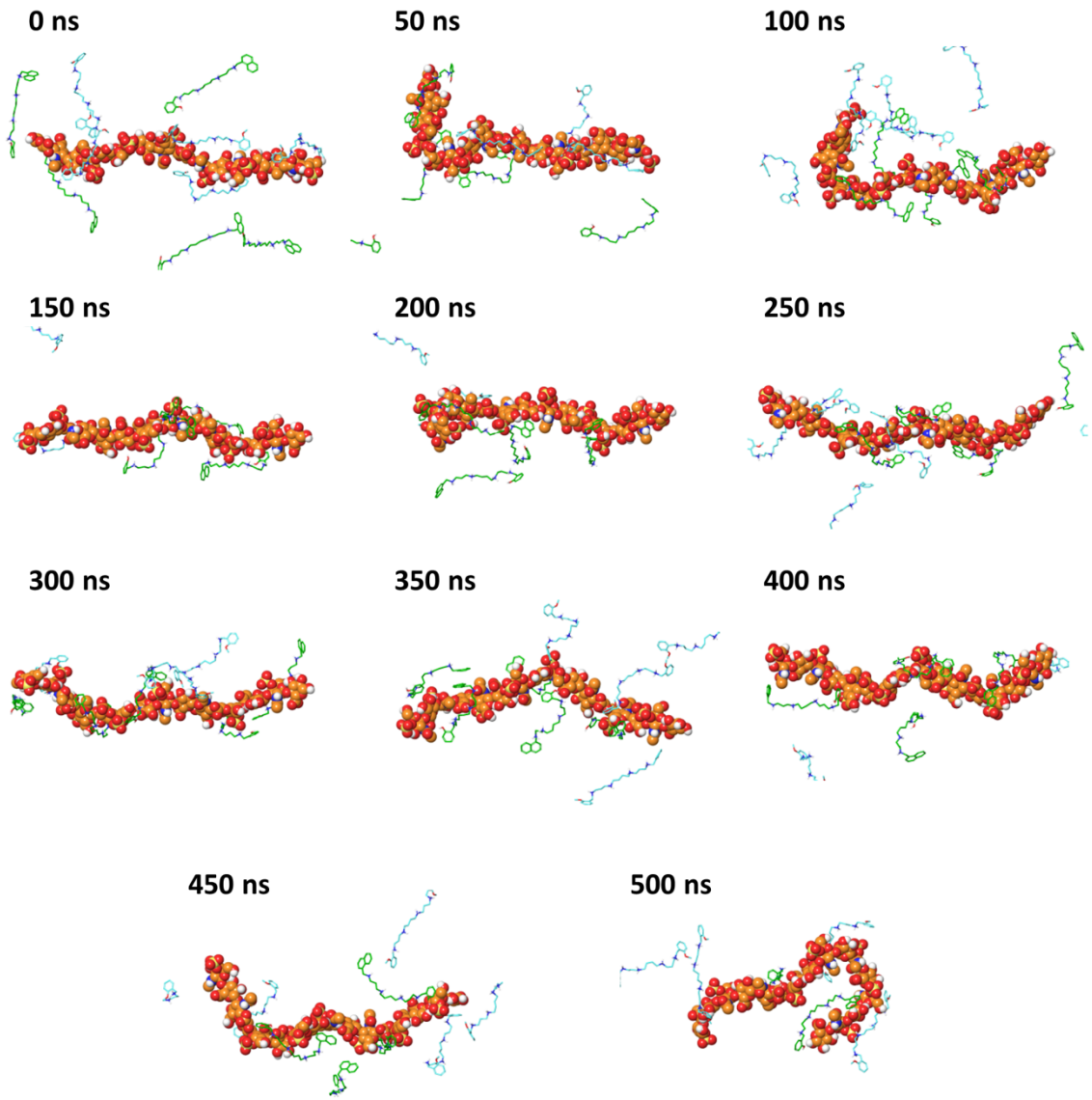


Figure S13. Snapshots from the 500 ns MD simulation of a system containing one dp16 CDS-A model (spheres), five **3AL** (green sticks) and five **3BB** (cyan sticks) molecules. Water molecules, ions and **3AL** or **3BB** molecules that are too far from CDS-A are omitted for clarity.

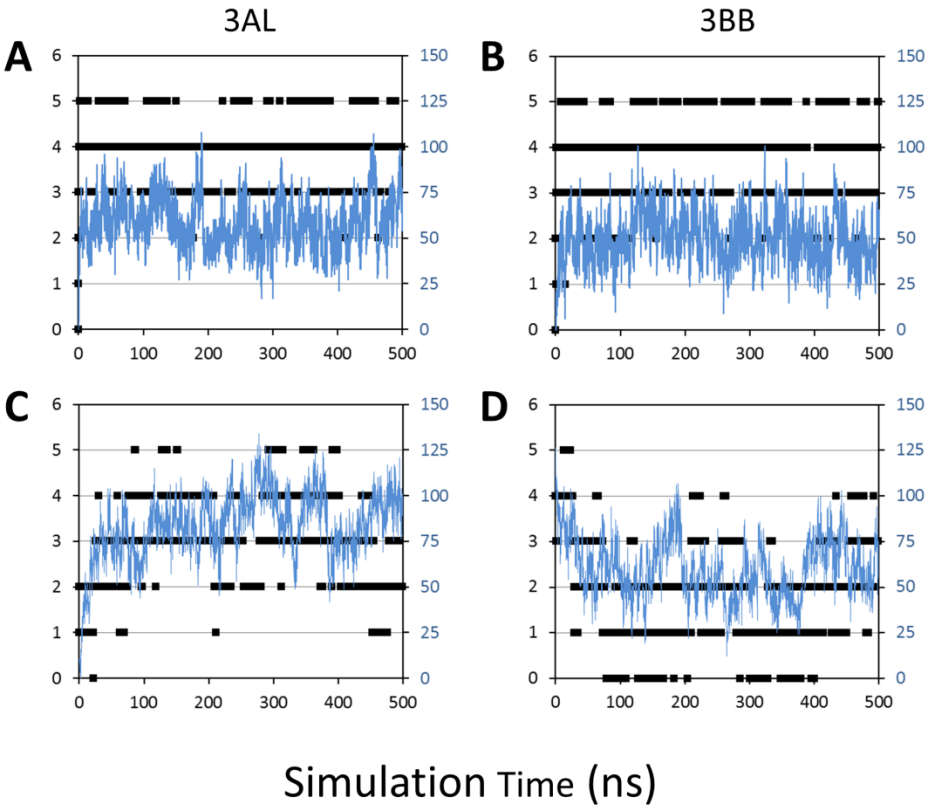


Figure S14. Summary of results obtained from 500 ns MD simulations of three different systems containing a CDS-A hexadecamer with all its sulfate and carboxylate groups ionized (total charge -16), and the following ligands: (A) 5 molecules of **3AL** with their four amino groups protonated (total charge +20); (B) 5 molecules of **3BB** with their four amino groups protonated (total charge +20); or (C-D) 5 molecules of **3AL** and 5 molecules of **3BB**, all with their four amino groups protonated (total charge +40). The three systems included enough Cl^- ions to achieve neutrality and they were immersed in a 100 x 100 x 100 Å cubic box of explicit TIP3P water. The graphs represent the number of ligand molecules (black squares, left axis) and the number of ligand atoms (blue lines, right axis) within a distance of 3 Å from the CDS-A molecule *vs.* simulation time. Average values of ligand molecules ($\overline{M}$) and ligand atoms ($\overline{A}$): (A) $\overline{M}$(**3AL**) = 4.12±0.64 , $\overline{A}$(**3AL**) = 58±14 ; (B) $\overline{M}$(**3BB**) = 4.12±0.74 , $\overline{A}$(**3BB**) = 50±14 ; (C) $\overline{M}$(**3AL**)= 3.19±0.99 , $\overline{A}$(**3AL**) = 83±18 ; (D) $\overline{M}$(**3BB**)= 1.84±1.10 , $\overline{A}$(**3BB**) = 60±16 .

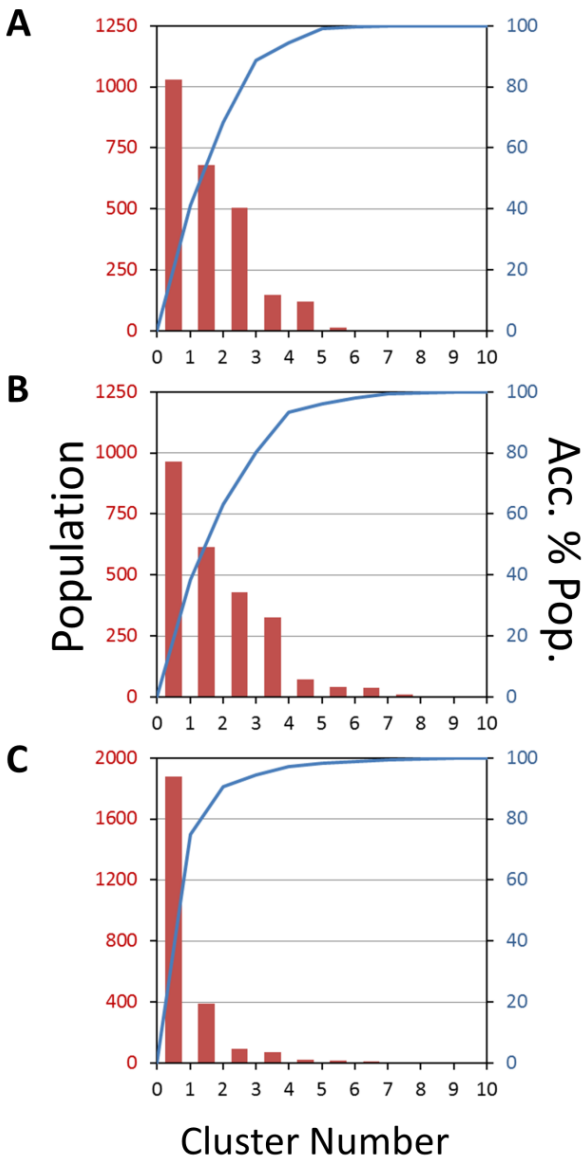


Figure S15. Results from clustering the 2500 snapshots of each MD simulation, by applying a hierarchical clustering method with 10 clusters. Simulations: (A) CDS-A/**3AL**, (B) CDS-A/**3BB**, and (C) CDS-A/**3AL**+**3BB**. The graphs show the population distribution among the clusters (red bars, left axis) and the accumulated % of population (blue line, right axis).

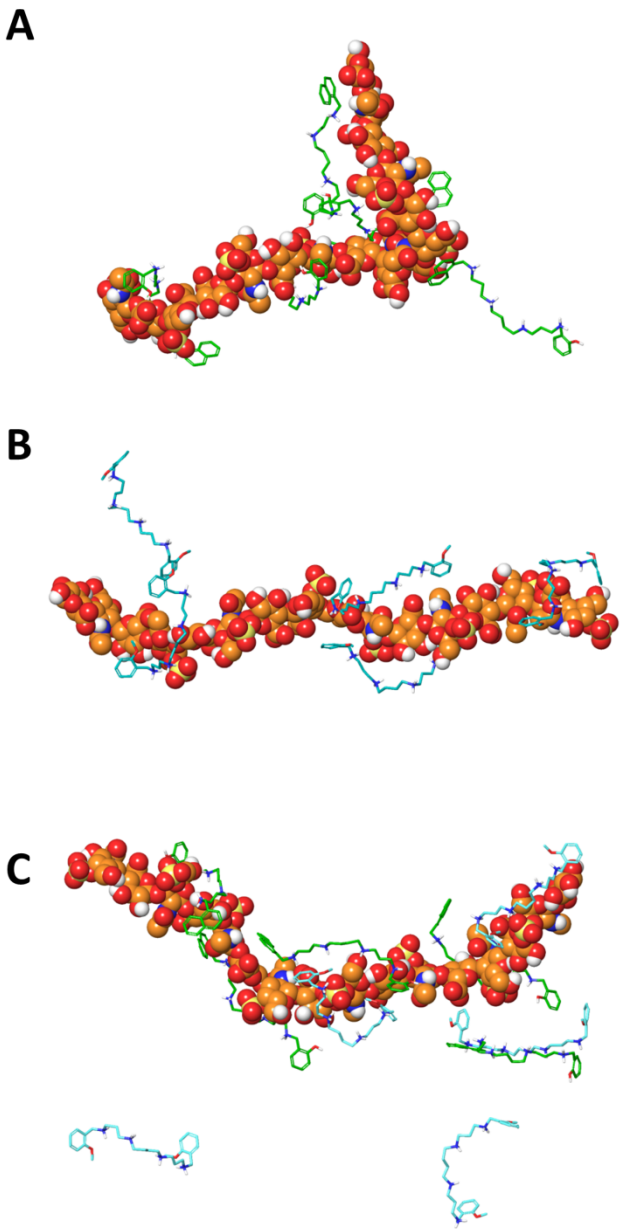


Figure S16. Representative snapshots of the most populated clusters for each simulation. Simulations: (A) CDS-A/**3AL**, (B) CDS-A/**3BB**, and (C) CDS-A/**3AL**+**3BB**. The CDS-A model molecule is shown as spheres, and the five **3AL** and **3BB** molecules are shown as green and cyan sticks, respectively. Water molecules and ions are omitted for clarity.

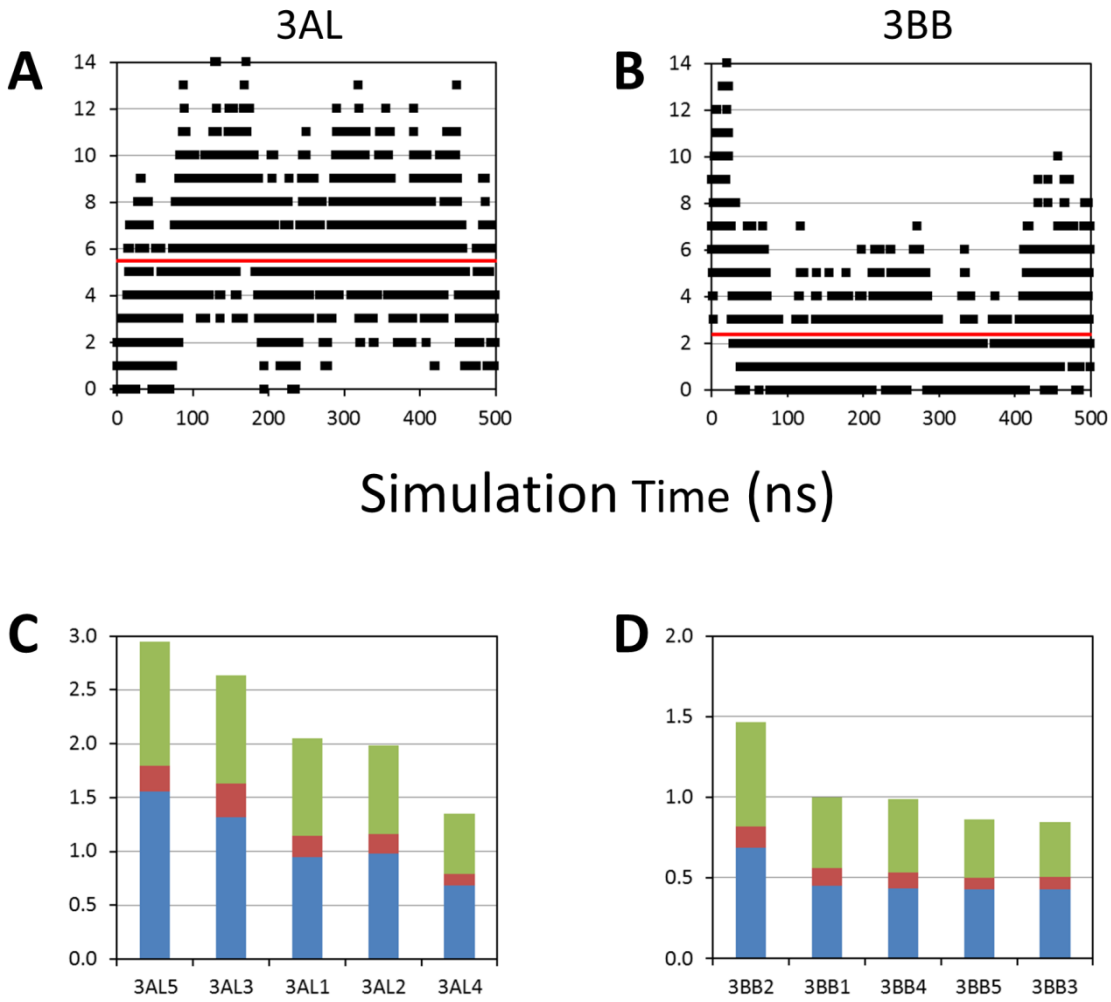


Figure S17. Total number of direct H-bonds stablished between the CDS-A molecule and the (A) **3AL** or (B) **3BB** ligands during the competition simulation of CDS-A/**3AL**/**3BB**. The red lines represent the average over the whole simulation (**3AL**: 5.5 ± 2.7; **3BB**: 2.4 ± 2.2). Polar contacts between CDS-A and each **3AL** (C) or **3BB** (D) molecule, averaged over the 500 ns simulation. Contacts are classified by type: H-bonds (blue bars), ionic (red bars) or water bridges (green bars).

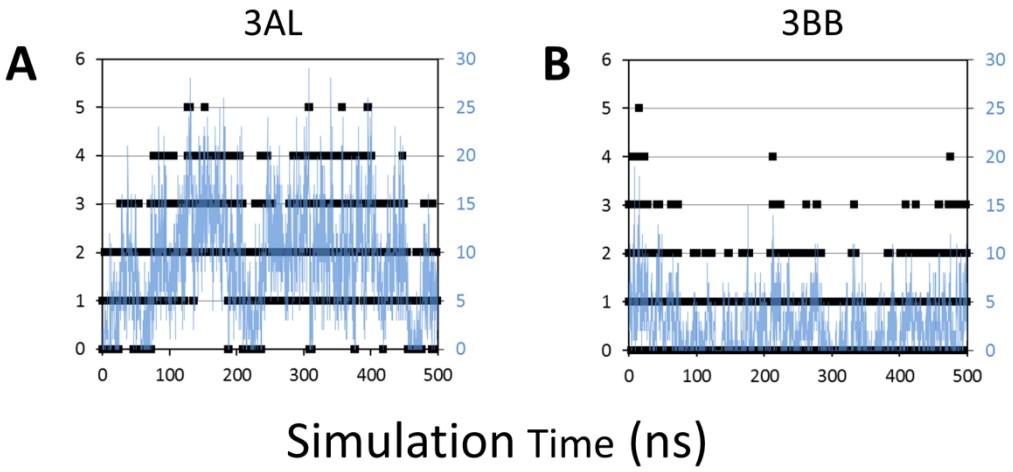


Figure S18. Number of ligand molecules (black squares, left axes) and ligand atoms (blue lines, right axes) involved in CH-π contacts between CDS-A and (A) **3AL** or (B) **3BB** during the competition simulation of CDS-A/**3AL**/**3BB**. The CH-π contacts were determined as those involving a sp3 C-atom of CDS-A and a sp2 C-atom of the ligands separated by a distance < 4 Å. In CDS-A, the sp3 C-atoms include the CH-groups of the glucuronic and galactosamine rings, the methyl of the 2-acetamido substituent and the 6-CH_2 of the galactosamine moieties. In **3AL** and **3BB** the sp2 C-atoms correspond to those of the aromatic substituents at the extremes of both ligand molecules. Average values of ligand molecules ($\overline{M}$) and ligand atoms ($\overline{A}$): (**A**) $\overline{M}$ (**3AL**) = 2.2 ± 1.1, $\overline{A}$ (**3AL**) = 9.0 ± 5.6; (**B**) $\overline{M}$ (**3BB**) = 0.9 ± 0.8, $\overline{A}$ (**3BB**) = 2.4 ± 2.7.

Supporting Movie. Movie showing selected snapshots from the 500 ns molecular dynamics simulation for the CDS-A hexadecamer (orange spheres) + 5 molecules of **3AL** (green sticks) + 5 molecules of **3BB** (cyan sticks). Only the ligands at a distance shorter than 8 Å are shown.

References

1) M. Corredor, D. Carbajo, C. Domingo, Y. Pérez, J. Bujons, A. Messeguer, I. Alfonso, Dynamic Covalent Identification of an Efficient Heparin Ligand. *Angewandte Chemie International Edition* **57**, 11973-11977 (2018). DOI: 10.1002/anie.201806770.
2) R. I. W. Osmond, W. C. Kett, S. E. Skett, D. R. Coombe, Protein–heparin interactions measured by BIAcore 2000 are affected by the method of heparin immobilization. *Anal. Biochem.* **310**, 199-207 (2002). DOI: 10.1016/S0003-2697(02)00396-2
3) A. K. Yetisen, R. Moreddu, S. Seifi, N. Jiang, K. Vega, X. Dong, J. Dong, H. Butt, M. Jakobi, M. Elsner, A. W. Koch, Dermal Tattoo Biosensors for Colorimetric Metabolite Detection. *Angewandte Chemie International Edition* **58**, 10506-10513 (2019). DOI: 10.1002/anie.201904416.
4) Schrödinger *Release 2019-1*, Schrödinger, LLC: New York, NY, 2019.
5) Schrödinger *Release 2019-1: Maestro*, Schrödinger, LLC: New York, NY, 2019.
6) Schrödinger *Release 2019-1: Macromodel*, Schrödinger, LLC: New York, NY, 2019.
7) E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, R. A. Friesner, OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of chemical theory and computation* **12**, 281-296 (2016). DOI: 10.1021/acs.jctc.5b00864.
8) W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson, Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**, 6127-6129 (1990). DOI: 10.1021/ja00172a038.
9) Schrödinger *Release 2019-1: Desmond Molecular Dynamics System*, D. E. Shaw Research: New York, NY, 2019.
10) K. J. Bowers, E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossváry, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan, D. E. Shaw, in *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*. (Tampa, Florida, 2006).
11) W. T. Winter, S. Arnott, D. H. Isaac, E. D. Atkins, Chondroitin 4-sulfate: the structure of a sulfated glycosaminoglycan. *J Mol Biol* **125**, 1-19 (1978). DOI: 10.1016/0022-2836(78)90251-6.
12) Schrödinger *Release 2019-1: Maestro-Desmond Interoperability Tools*, D. E. Shaw Research: New York, NY, 2019.
13) D. J. Evans, B. L. Holian, The Nose–Hoover thermostat. *The Journal of Chemical Physics* **83**, 4069-4074 (1985). DOI: 10.1063/1.449071.
14) G. J. Martyna, M. L. Klein, M. Tuckerman, Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics* **97**, 2635-2643 (1992). DOI: 10.1063/1.463940.
15) G. J. Martyna, D. J. Tobias, M. L. Klein, Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* **101**, 4177-4189 (1994). DOI: 10.1063/1.467468.
16) M. Tuckerman, B. J. Berne, G. J. Martyna, Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics* **97**, 1990-2001 (1992). DOI: 10.1063/1.463137.
17) T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089-10092 (1993). DOI: 10.1063/1.464397.

18) U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577-8593 (1995). DOI: 10.1063/1.470117.
19) V. Kräutler, W. F. van Gunsteren, P. H. Hünenberger, A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *Journal of Computational Chemistry* **22**, 501-508 (2001). DOI: 10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V.