Supporting Information

Generation-Dependent Molecular Recognition Controls Self-Assembly in Supramolecular Dendron-Virus Complexes

Giovanni Doni, [†,a,b] *Mauri A. Kostiainen,*[c] *Andrea Danani*[a] *and Giovanni M. Pavan*[†,a,]*

a: University of Applied Sciences of Southern Switzerland (SUPSI) – Mathematical and Physical Sciences Research Unit (SMF),Centro Galleria 2, Manno, 6928, Switzerland

giovanni.pavan@supsi.ch

b: University of Lugano – Institute of Computational Science, Via Giuseppe Buffi 13, 6906 Lugano

c: Radboud University Nijmegen – Institute for Molecules and Materials, Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands

†: these authors contributed equally to the work

TABLE OF CONTENTS

Computational details – the molecular systems

The 1:1 systems – study of the dendron-virus molecular recognition

The equilibrated molecular models of **G1** and **G2** spermine-base UV-degradable dendrons were taken from our previous work.[S1] They were composed of three residues: a central core (COR), the repetitive branched (REP) and the spermine (SPM) surface units. At pH~7.4, each SPM was considered to carry a +3 e charge each. The total charge of **G1** and **G2** is +9 e and +27 e respectively at neutral pH. The Cowpea Chlorotic Mottle Virus (CCMV) structure was taken from the internet at the Viper Particle Explorer web site (http://viperdb.scripps.edu/info_page.php?VDB=1cwp – structure 1CWP.pdb in the Protein Data Bank). Since we were interested in the surface of the virus, all RNA present in the capsid was removed and not considered in the calculations. The whole capsid surface is constituted by three proteins (chain A, B and C in Figure S1a). Due to the periodic nature of the virus capsid (Figure S1b), we calculated the electrostatic potential surface (Figure 1Sc) in the narrow zone of one of the 60 pores, located at the quasi-threefold axis, in the middle of the three coat proteins that constitutes the surface of CCMV. This was done by using the Adaptive Poisson-Boltzmann Solver (APBS plugin version 1.1)[S2] using the PDB2PQR web portal (http://kryptonite.nbcr.net/pdb2pqr/).[S3] Without the viral RNA inside the capsid, the total charge of the CCMV protein-cage is -660 e. Each of the 60 pores, generated by the mutual disposition of the three base-protein chains that compose the whole capsid, carries a net charge of -9 e (Figure S1a) – the 82% of the global negative charge of the CCMV capsid is focused in these region.

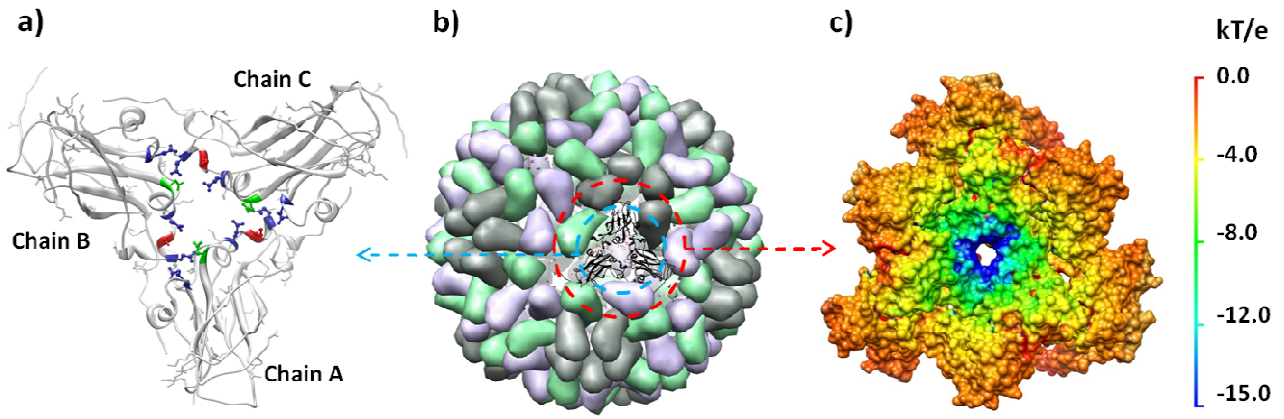


Figure S1. a) The pore zone has an overall charge of -9 e generated by 9 glutamic acids (GLU in blue), 3 aspartic acids (ASP: green) negatively charged and 3 positive lysines (LYS: in red). b) The whole surface of the CCMV capsid is formed by the three protein chains A,B and C that constitute also the pore. c) The electrostatic potential was found to be consistently negative close to the pore (blue and green regions) and practically null outside.

This evidence finds consistency also in previous studies on the Cowpea Chlorotic Mottle Virus (CCMV).[S4] For this reason, each of the pores in Figure S1a was considered as a potential binding site for the positively charged dendrons. As a binding site we chose a capsid portion, centered on the pore, large enough (Figure S1b, red circle and Figure S1c) to guarantee the correct dynamic behavior of the dendron during the binding (Figure S2a: total charge of the virus barrier: +27 e). Following a well validated procedure for the study of the molecular recognition between dendrons and organic molecules,[S5] the dendrons (**G1** and **G2**) were put in close proximity of the pore and a periodic rectangular parallelepiped box of TIP3P[S6] water molecules was created extending 12 angstroms from the last atom of the dendron (COR) and 0 angstroms from the atoms of the virus surface portion using the *leap* module within the AMBER 11 suite of programs.[S7]

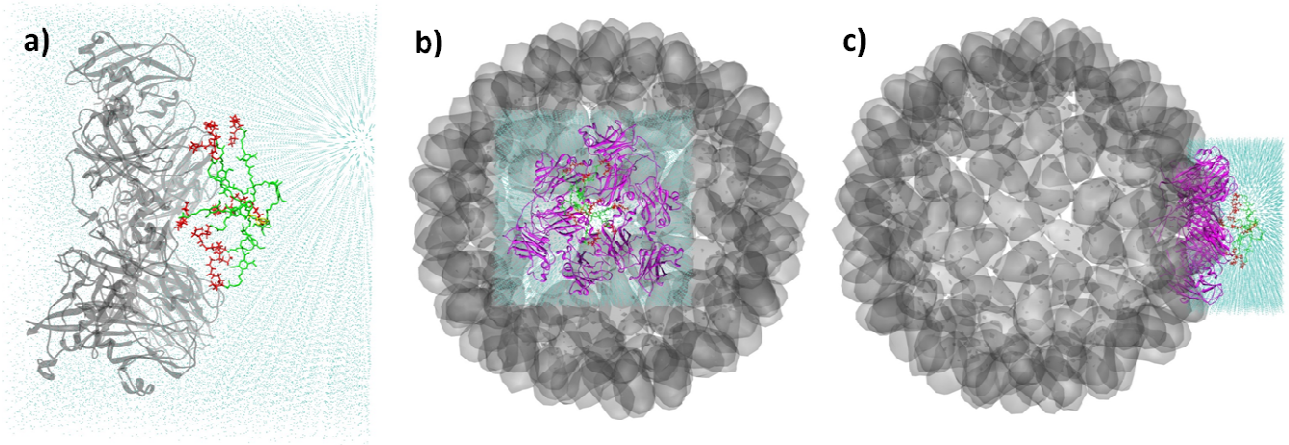


Figure S2. The **G2**+CCMV periodic system after solvation (a) – the capsid portion is represented as a dark shadowed ribbon, within **G2** COR and REP are colored in green, the SPM ligands in red. Water Oxygen atoms are colored in cyan, counterions and water Hydrogen atoms are not shown for clarity. Front (a) and side view (b) of periodic simulated system with reference to the virus capsid (black shadowed). The pore zone is represented as a purple ribbon.

The suitable number of Cl^- and Na^+ counterions was inserted with the *leap* module of AMBER 11 to reproduce the relevant ionic concentrations of 10 and 150 mM NaCl in solution (Table S1 reports the details on the atom and ion number constituting each system). In particular, the salt ions were added in the periodic systems with the standard *addIons* utility of *leap* – Na^+ and Cl^- ions were placed onto a shell around the solute using a Coulombic potential on a grid – water molecule was replaced with the ion if eventual superposition occurred. Four 1:1 systems for the study of the molecular recognition between the CCMV virus and spermine-based dendrons (**G1** and **G2**) were thus obtained (details in Table S1).

Table S1. The main features of the 1:1 molecular systems simulated in this work to study the molecular recognition between the CCMV virus and spermine-based dendrons (**G1** and **G2**).

Complex	[NaCl] [a] (mM)	Dendron charge[b] (e)	CCMV barrier charge	Box volume (Å^3)	Number of Na^+ and Cl^- atoms[c] in	Number of water molecules in	Total number of atom in

			(e)		the system	the system	the system
G1+CCMV	10	+9	-27	1074314	18	29073	108731
G2+CCMV	10	+27	-27	1265338	32	35225	127810
G1+CCMV	150	+9	-27	1069516	198	28980	108668
G2+CCMV	150	+27	-27	1263405	232	35182	127881

[a] Experimental ionic concentration in solution. [b] Each SPM residue carry a +3 e charge, thus resultant charge is +9 e for **G1** and +27 e for **G2**). [c] The total amount of Na^+ and Cl^- atoms to the system for the neutralization and to reproduce the experimental ionic concentration reported in 2nd column.

The study of the virus-virus interface

As described in the text, these four systems (**G1**+CCMV and **G2**+CCMV at 10 and 150 mM NaCl) were used to explore the molecular recognition between the dendrons and the capsid. As a second step in the simulation work we created other molecular systems in order to study the interface of the virus-virus assembly generated by **G1** and **G2** dendrons. Figure S3b reports the TEM image of the virus-assembly generated by the presence of **G1**.[S8] Figure S3c shows the base concept that is at the origin of the creation of these systems: zooming at the virus-virus (V1-V2) assembly interface.

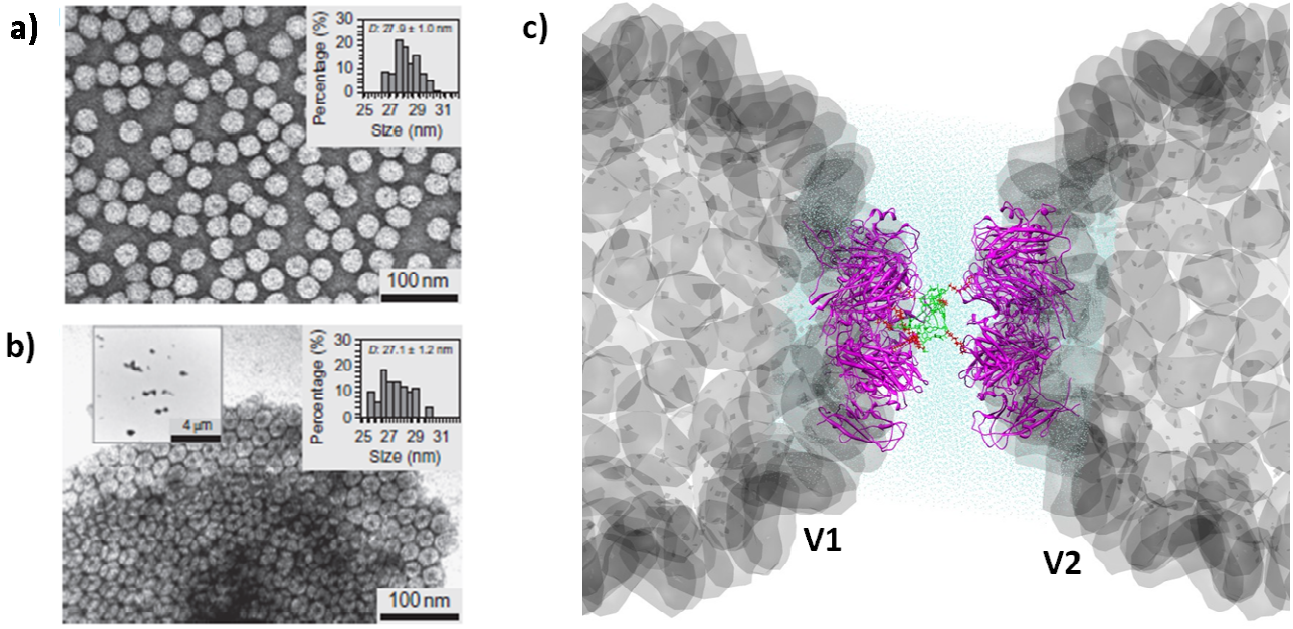


Figure S3. The CCMV compaction (a,b) driven by the presence of the dendrons in solution (TEM images: a,b).[S8] The V1-V2+**G2** simulated periodic system (c) is constructed at the V1-V2 interface – the pore zone of V1 and V2 is represented as a purple ribbon, within **G2** COR and REP are colored in green and the SPM ligands in red. Hydrogen atoms of water and ions are not shown for clarity.

The equilibrated configuration of the 1:1 **G1**+CCMV and **G2**+CCMV systems were taken as a starting point. In principle, these systems were created by removing the existing water and ions from **G1**+CCMV and **G2**+CCMV. The pore region of the CCMV capsid was then copied, reversed and used to form the interface (the opposite side of the periodic simulation box in Figure S3c). The different virus compaction induced by **G1** and **G2** is described in details in the paper – for this reason we created the two interface systems V1-V2+**G1** and V1-V2+**G2** differently. The base of the interface system that was created to simulate the assembly generated by **G1** was constituted by two virus capsid portions and by two **G1** dendrons (i.e. to create this system both the part of virus capsid portions and the dendron were copied and reversed) – the two hydrophobic COR residues of **G1**s were put in close proximity (Figure 4a in the text and Figure S7). The solute of the **G2** assembly-interface system (V1-V2+**G2**), on the other hand, was constituted by the two mirror capsid barriers confining a single **G2** molecule (the V2 barrier was put in close correspondence of the SPM ligands of **G2** oriented toward the external solution: Figure S3c, S8 and Figure 4b in the text). V1-V2+**G1** and V1-V2+**G2** ensembles were again solvated in a periodic rectangular parallelepiped box of TIP3P water molecules[S6] with no extension from the solute atoms. The suitable number of counterions necessary to guarantee the neutrality of the systems and the salt concentrations of 10 and 150 mM NaCl in solution were added with the leap module of AMBER 11 as described previously. Other four systems were thus created modeling the V1-V2 interface generated by **G1** and **G2** at both salt concentrations – Table S2 reports the details of the molecular systems.

Table S2. The main features of the molecular systems used in this work for the simulation of the virus-assembly interface (V1-V2) induced by **G1** and **G2**.

V1-V2 assembly interface system	[NaCl] [a] (mM)	Dendron charge[b] (e)	CCMV barrier charge (V1+V2) [c] (e)	Box volume (Å^3)	Number of Na^+ and Cl- atoms[d] in the system	Number of water molecules in the system	Total number of atom in the system
V1-V2+**G1**	10	+9	-54	1615789	35	40949	165862
V1-V2+**G2**	10	+27	-54	1982429	57	51761	198596
V1-V2+**G1**	150	+9	-54	1718218	173	44210	175783
V1-V2+**G2**	150	+27	-54	2074540	182	54737	207662

[a] Experimental ionic concentration in solution. [b] The dendron charge is +9 e for **G1** and +27 e for **G2**). [c] The total charge of the CCMV virus portion is given by the sum of the charges of V1 and V2. [d] The total amount of Na^+ and Cl^- atoms to the system for the neutralization and to reproduce the experimental ionic concentration reported in the 2nd column.

Computational details – the simulation procedure

The procedure adopted for the molecular dynamics simulations was adapted from our previous work on the simulations of dendrimers and dendrons.[S9,S5] All calculations were conducted using the AMBER 11 software.[S7] The force field types and the partial charges for the residues that constitute **G1** and **G2** dendrons were taken from our previous work on UV-degradable dendrons[S1] – these parameters were obtained using the AM1-BCC[S10] calculation method within the *antechamber*[S11] module of AmberTools 1.4 (AMBER11).[S7] Since *antechamber* assigns parameters and force field types that are consistent with the "general AMBER force field (GAFF)" (*gaff.dat*),[S12] the calculated parameters already demonstrated to be well consistent and compatible with the force field parameters of AMBER 11.[S1] The creation of the systems was discussed and described in details in the previous section. All of the molecular systems created (both the 1:1 and the interface ones) were initially minimized and then two first steps of molecular dynamics simulations (MD) were run for 50 ps in NVT and NPT conditions respectively to reach the simulation temperature of 300 K and to

relax the density of the solvent within the periodic box. After these initial steps, the production phase lasted for 10 ns under NPT periodic boundary condition at 300 K and 1 atm, using a time step of 2 femtoseconds, the Langevin thermostat and a 10 Å cutoff. The particle mesh Ewald[S13] (PME) approach was adopted to treat the long-range electrostatic effects and the SHAKE algorithm was used on the bonds involving Hydrogen atoms.[S14] All the production molecular dynamics were carried out by using the *pmemd* and the *pmemd.cuda* modules with the *parm99* all-atom force field by Cornell et al.,[S15] working in parallel on 128 processors of the Cray XT5 at CSCS Swiss National Supercomputer Center of Manno (Switzerland) and on our NVIDIA Tesla 2050 GPU cards.

The Root Mean Square Deviation (RMSD) data were obtained from the molecular dynamics trajectories in order to verify that all of the systems converged to the equilibrium with good stability. All of the binding affinity energies (ΔG_{bind}) were obtained from molecular dynamic simulations and calculated over 200 of snapshots taken from the equilibrated phase of the dynamic trajectories (last 2 nanoseconds of the simulation) using the well-known MM-PBSA approach.[S16] The free energy of binding ΔG_{bind} is composed by an enthalpic (ΔH_{bind}) and an entropic ($-T\Delta S_{bind}$) term as given in Eq. (S1):

$$\Delta G_{bind} = \Delta H_{bind} - T\Delta S_{bind} \quad \text{(S1)}$$

$$\Delta H_{bind} = \Delta E_{gas} + \Delta G_{sol} \quad \text{(S2)}$$

ΔH_{bind} can be split into total gas-phase *in vacuo* non-bond energy (ΔE_{gas}), composed by an electrostatic and a van der Waals term (ΔE_{ele} and ΔE_{vdW}), and a solvation energetic term ($\Delta G_{solv} = \Delta G_{PB} + \Delta G_{NP}$)[S17] as described in Eq. (S2). The polar component of ΔG_{PB} was evaluated using the Poisson-Boltzmann[S18] (PB) approach with a numerical solver implemented in the *pbsa* program of AMBER 11.[S19] The non-polar contribution to the solvation energy was calculated as $\Delta G_{NP} = \gamma$ (SASA) + β, in which $\gamma = 0.00542$ kcal/Å^2, $\beta = 0.92$ kcal/mol, and SASA is the solvent-accessible

surface estimated with the MSMS program.[S20] Finally, the normal-mode[S21] approach was used to compute the entropic term.

The per-residue decomposition energies were obtained using the *mm_pbsa.pl* script of AMBER 11 in order to explore the uniformity of the binding interaction from a quantitative point of view – this approach can quantify the affinity of each residue of the dendron (CEN, REP and SPM) to CCMV capsid surface. This procedure is particularly useful in the study of the interactions between the residues that mostly interact in the system (i.e. the interactions between positive SPM ligands and negative amino acids (Asp and Glu) within the capsid). Each energetic component can be defined in terms of Eq. (S3). These values represent the difference between the energy of the molecular complex ($E_{complex}$) and the sum of the energies of the dendron and the CCMV pore zone taken separately ($E_{dendron} + E_{CCMV}$). Negative energy (ΔE) values indicate attractive forces and the thermodynamic tendency to form a complex.

$$\Delta E = E_{complex} - (E_{dendron} + E_{CCMV}) \quad \text{(S3)}$$

Gas-phase energies (E_{gas}) for each residue are composed of electrostatic and van der Waals interaction contributions (E_{ele} and E_{vdw}, respectively):

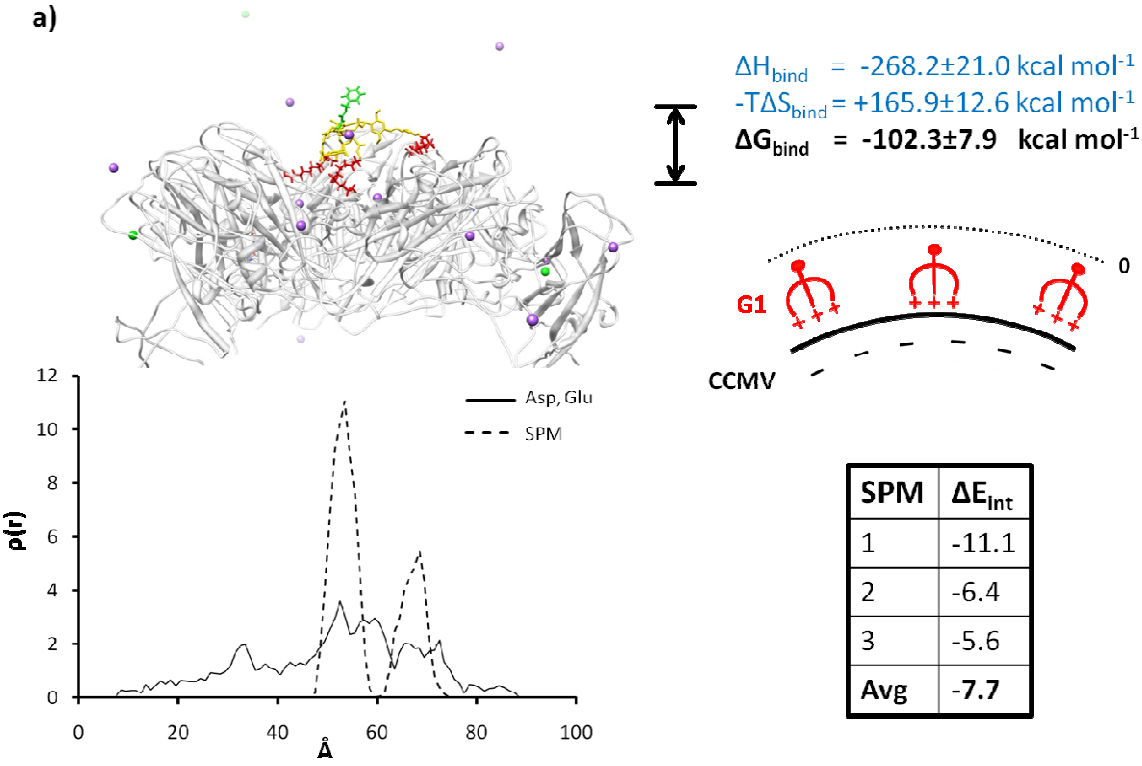
$$\Delta E_{gas} = \Delta E_{ele} + \Delta E_{vdW} \quad \text{(S4)}$$

The *in vacuo* gas-phase interaction energy for each residue (ΔE_{gas}) is then corrected for solvation to give the total energy ΔE_{int} (Figure S4). The Generalized Born method available in the *mm_pbsa.pl* script of AMBER 11 was used to correct the gas-phase energies for solvation. Energy decomposition for non-polar contributions to desolvation is performed using the LCPO method.[S22]

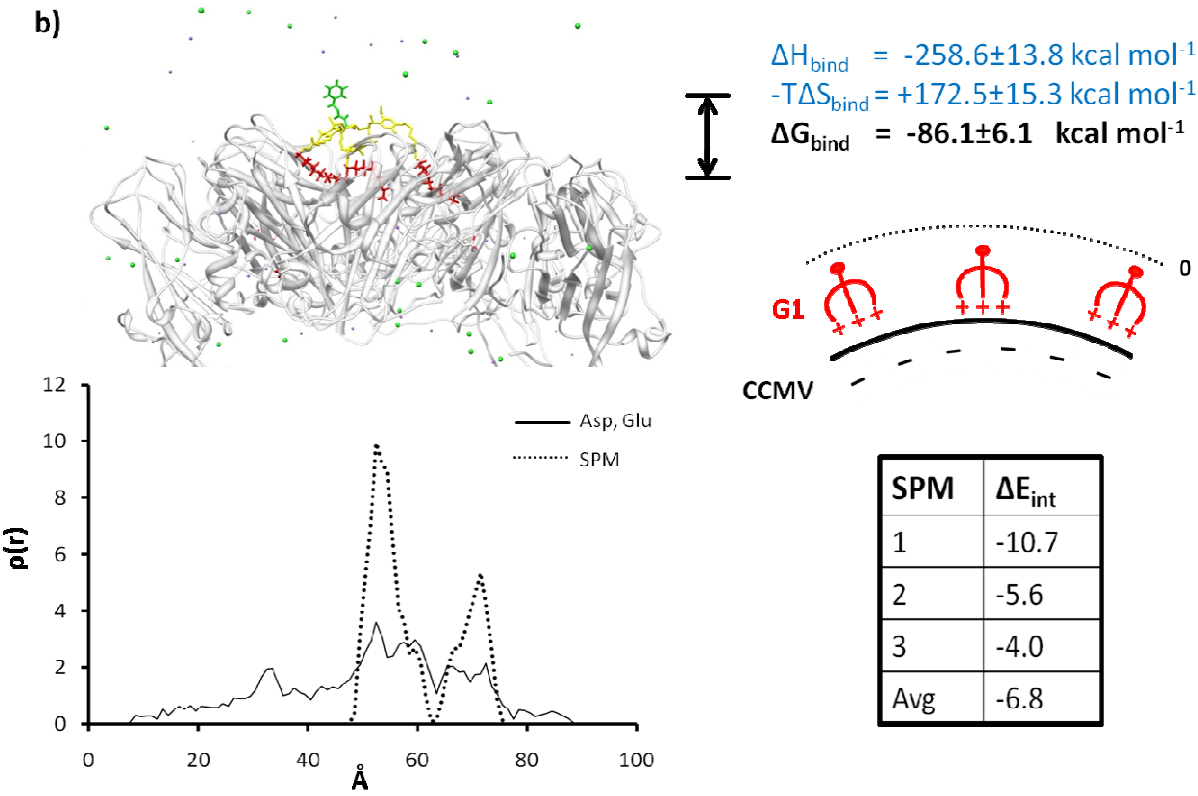
The equilibrated trajectories were finally processed with the *ptraj* module in order to obtain the radial distribution functions (RDF), which are very useful for the understanding of the structural and dynamic behavior of **G1** and **G2** during the binding event.

Extended data on the molecular recognition of the (1:1) dendrons-CCMV systems

The molecular recognition between **G1** and **G2** and the CCMV virus was studied from the simulation of the 1:1 systems. The molecular dynamics trajectories of these systems were processed in order to obtain free energy of binding, structural, dynamic and decomposition-energies data. Figure S4 and Figure S5 reports dynamic snapshots, RDF plots and energetic of the binding event for **G1** and **G2** at low (10 mM NaCl) and high (150 mM NaCl) salt concentration in solution.



SPM	ΔE_{Int}
1	-11.1
2	-6.4
3	-5.6
Avg	**-7.7**



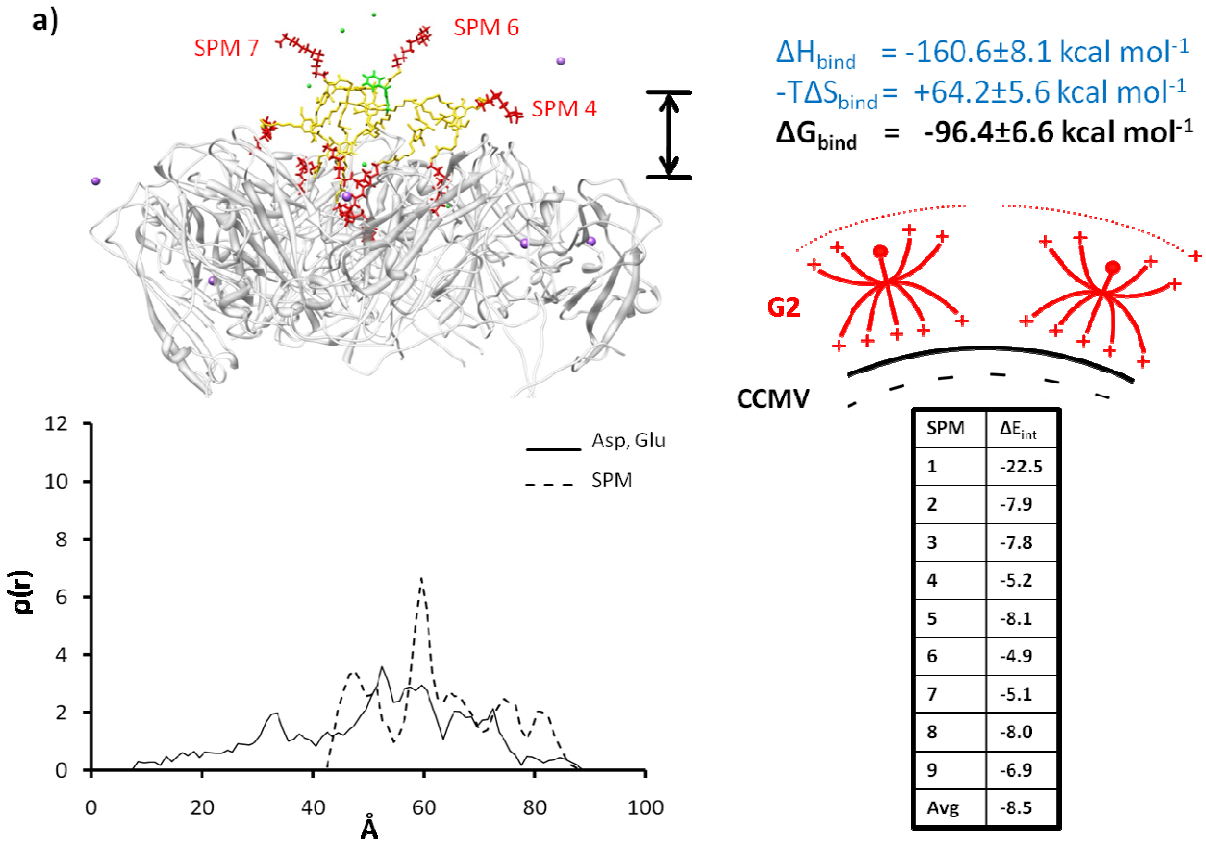
SPM	ΔE_{int}
1	-10.7
2	-5.6
3	-4.0
Avg	-6.8

Figure S4. Side view of the binding between **G1** and the CCMV pore zone at 10 (a) and 150 mM (b) and scheme of the mechanism of binding. Within the dendron, CEN is colored in green, REP in yellow and SPM in red. Both the energetic of binding (ΔG_{bind}, ΔH_{bind} and $-T\Delta S_{bind}$) and the per-residue decomposition interaction energies between SPMs and virus (ΔE_{int}) are expressed in kcal mol^{-1}. The RDF plots are represented as a dotted (SPM) and a continuous line (Glu, Asp) and calculated with respect (origin of the graphs) to the left side of the simulation periodic box.

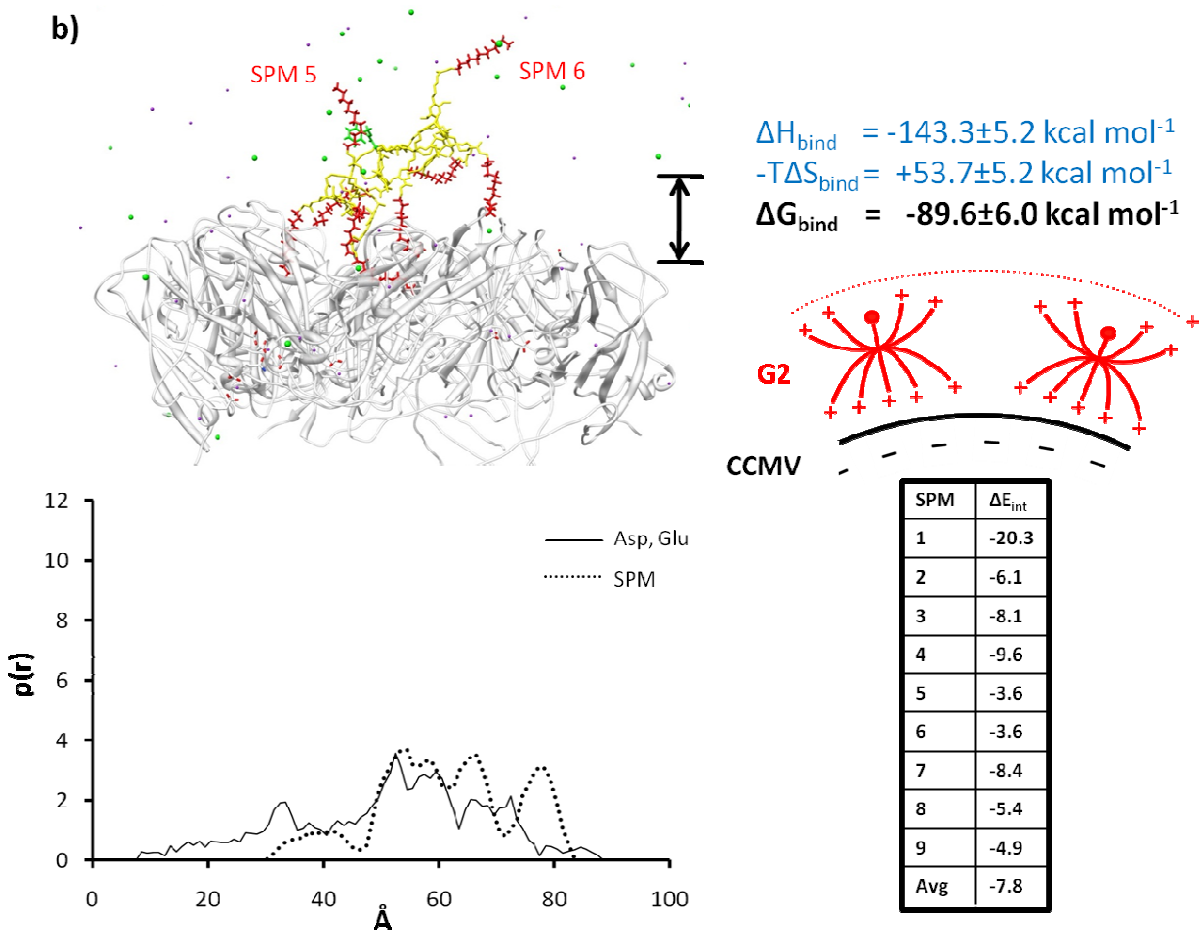
In order to allow a direct comparison with **G2**, the binding energetic values (ΔG_{bind}, ΔH_{bind} and $-T\Delta S_{bind}$) were multiplied per three (charge of **G1**: +9 e, charge of **G2**: +27 e) according to our previous studies on similar molecules.[S23] We calculated the radial distribution functions (RDF) profiles of the atoms that are most actively participating to the binding in order to have a dynamic interpretation of the binding event – the SPM ligands of dendrons and the negatively charged residue within the protein-cage around the pore of the virus (Asp and Glu). RDF represents the density and the distribution of SPM, Glu and Asp atoms in space with respect to the respect to the left side of the simulation periodic box (origin of the RDF graphs). This gives indications about the

presence of SPM, Glu and Asp atoms in a certain zone of the system (spatial density). However, since these curves are calculated at each step of the simulation, and they are reported in the plots as averaged over the equilibrated phases of the dynamic trajectories, they give information also on the dynamics of the system – they provide indication on the time period in which a certain atom is present in a certain area in the space (dynamic density).

In RDF plots, high and narrow peaks in a small area of these graphs mean not only high density of atoms in a certain zone, but also high localization and low mobility of these atoms. On the other hand, broad and low intensity peaks indicate low density and high vibrations. Importantly, while **G1** penetrates deeply into the asperities of the virus surface (high focused peaks), **G2**, due to its higher rigidity, diverse shape and larger dimensions binds the pore zone as an external body, orienting few (2-3) of its SPM ligands toward the external environment, free to fluctuate (lower SPM peaks).



SPM	ΔE_{int}
1	-22.5
2	-7.9
3	-7.8
4	-5.2
5	-8.1
6	-4.9
7	-5.1
8	-8.0
9	-6.9
Avg	-8.5



SPM	ΔE_{int}
1	-20.3
2	-6.1
3	-8.1
4	-9.6
5	-3.6
6	-3.6
7	-8.4
8	-5.4
9	-4.9
Avg	-7.8

Figure S5. Side view of the binding between **G2** and the CCMV pore zone at 10 (a) and 150 mM (b) NaCl and scheme of the mechanism of binding. Within the Dendron, CEN is colored in green, REP in yellow and SPM in red. Both the energetic of binding (ΔG_{bind}, ΔH_{bind} and $-T\Delta S_{bind}$) and the per-residue decomposition interaction energies between SPMs and virus (ΔE_{int}) are expressed in kcal mol^{-1}. The SPM residues that do not participate actively in the binding with the CCMV surface are indicated in red in the pictures and are identified by the lower ΔE_{int} in the table. The RDF plots are represented as a dotted (SPM) and a continuous line (Glu, Asp) and calculated with respect (origin of the graphs) to the left side of the simulation periodic box.

These data highlight the different manners how **G1** and **G2** bind to the capsid surface despite a similar binding efficiency. This is true independently on the salt concentration – in this framework, the amount of ions in solution only marginally (ΔG_{bind} can be considered on the same level from the

statistical point of view). This generates two different kinds of self assemblies – extended data on the interface systems are reported in the next section.

Details and comments on experimental evidences

Relevant experimental data for the self-assembly systems are taken from our previous experimental work on the topic[S8] and reported in Table S3.

Table S3. Relevant experimental evidences on the virus-assembly induced by **G1** and **G2**.

CCMV assembly[a]	[NaCl][b] (mM)	Dendron charge[c] (e)	Dendron MW[d] (kDa)	% of CCMV surface charge that is located in the pores[e]	N° of CCMV negatively charged pores[f]	N° of dendrons per-virus that induces a 100% CCMV assembly[g]	ζ-potential values of the dendron-virus complexes[h] (mV)
V1-V2+**G1**	10	+9	2.196	~82%	60	~324	-1.9
V1-V2+**G2**	10	+27	6.604	~82%	60	~14	+23.7
V1-V2+**G1**	150	+9	2.196	~82%	60	>5000	-1.9
V1-V2+**G2**	150	+27	6.604	~82%	60	~90	+23.7

[a] Data related to the virus assembly generated by **G1** and **G2** are taken from our previous experimental work.[S8] [b] Experimental ionic concentration in solution. [c] The dendron charge is +9 e for **G1** and +27 e for **G2**). [d] Molecular weight of **G1** and **G2** expressed in kDa (molecular weight of a single CCMV is 4738.16 kDa). [e] The total surface charge of CCMV virus is -660 e. Each of the pores has an overall charge of -9 e (total 60 pore charge: -540 e) [f] Number of negatively charged pores present on the surface of a single CCMV virus. [g] Data calculated from the graphs in Figure 1c,d in the text – number of dendrons per-single-virus that generates a complete association of CCMV capsids in solution. [h] ζ-potential of the CCMV-dendrons super-assemblies are taken from our previous experimental work.[S8] ζ-potential value for the CCMV free in solution is -18.5 mV.

We already highlighted how most of the surface charge of the virus is located in the pores (Figure 1a), that is why we considered each of the pore as a potential binding site for the dendrons. Interestingly, if we convert the data (expressed in weight/volume) in the graphs of Figure 1c,d in the

text in terms of molar concentration, we find the number of **G1** and **G2** molecules per CCMV virus that creates a 100% compaction of viruses in solution – a sort of stoichiometry of the assembly process. Data are really interesting – **G1** is able to compact viruses only as soon as its number exceeds the number of pores. For **G2**, on the other hand, the necessary number of dendrons is more than one order of magnitude lower (Table S3). This suggests a different self-assembly of viruses induced by **G1** and **G2**. **G1** tends to saturate the charged pores of CCMV – the result is an almost neutral and hydrophobic sphere. The virus assembly in this case is characterized by a hydrophobic association of hydrophobic capsids in solution. **G2** on the other hand acts as a "molecular glue", compacting CCMV viruses at lower stoichiometry than **G1** in low (1:14) but also high (1:90) salt concentrated solution – a virus association driven by electrostatic intermolecular interactions. This is evidenced also by the ζ-potential values V1-V2+**G1** (-1.9 mV – almost neutral surface of the complexes) and V1-V2+**G2** (+23.7 mV – positively charged surface) – the value for the free CCMV virus in solution is -18.5 mV. Figure S6 represents the different virus assembly generated by **G1** and **G2**.

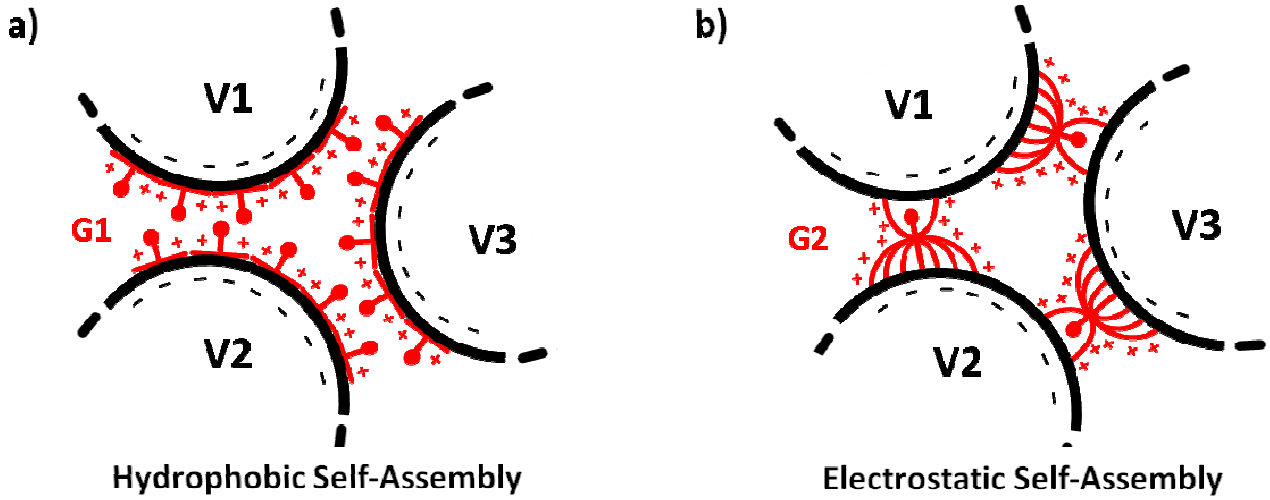


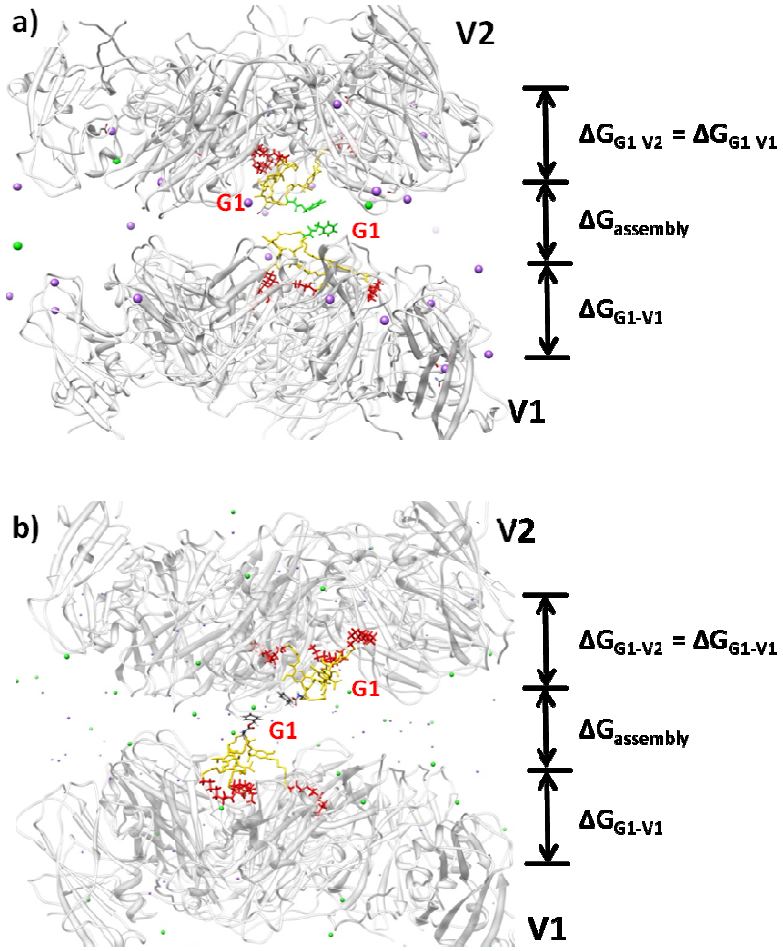
Figure S6. Schematic representation of the different virus self-assembly induced by **G1** (a) and **G2** (b).

Following this concept, we created the systems to simulate the virus-virus assembly interface induced by **G1** and **G2** accordingly.

Extended data on the molecular interactions that drive the self-assembly: the virus-virus interface

Due to the different manner how **G1** and **G2** bind the surface of CCMV virus (Figure S4 and Figure S5), we guessed that **G1** and **G2** generate two different virus assemblies (Figure S6) – **G1** induces a hydrophobic association while the presence of **G2** in solution induces an electrostatic virus association. This hypothesis finds consistency in the experimental evidences and also in the simulations on the molecular dendron-virus recognition. In particular, **G1** penetrates deep in the surface of CCMV, covering the pore with its structure and orienting the hydrophobic CEN toward the external solution. The capsids (V1 and V2) will assemble due to the tendency of hydrophobic entities to associate in solution. The spermines that do not actively participate to the binding between **G2** and a first virus (V1) are, on the other hand, able to attract a second one (V2) resulting in an electrostatic association. For this reason, to prove our hypothesis on the different self-assembly generated by a diverse molecular recognition, we created the virus-virus interface systems (V1-V2+**G1** and V1-V2+**G2**) as composed by two virus surfaces (V1 and V2) and two **G1** and a single **G2**.

Accordingly, we calculated also the energetic values that drive the V1-V2 assembly ($\Delta G_{assembly}$) differently, according to the MM-PBSA approach described in previously.[S16] Figures S7 and S8 show the side view of the V1-V2 assembly interface generated by **G1** and **G2** at 10 and 150 mM – affinity energies are also reported and expressed in kcal mol^{-1}.



ΔH_{G1-V1}	-274.2±26.4
$-T\Delta S_{G1-V1}$	+174.2±14.4
ΔG_{G1-V1}	**-100.1±8.9**
$\Delta G_{G1-V1} = \Delta G_{G1-V2}$	
$\Delta H_{assembly}$	-0.1±0.01
$-T\Delta S_{assembly}$	-3.7±0.3
$\Delta G_{assembly}$	**-3.8±0.3**

ΔH_{G1-V1}	-261.0±21.8
$-T\Delta S_{G1-V1}$	+179.3±15.9
ΔG_{G1-V1}	**-81.7±7.0**
$\Delta G_{G1-V1} = \Delta G_{G1-V2}$	
$\Delta H_{assembly}$	-0.2±0.01
$-T\Delta S_{assembly}$	-1.3±0.2
$\Delta G_{assembly}$	**-1.5±0.2**

Figure S7. The V1-V2+**G1** systems at 10 (a) and 150 mM (b) NaCl. Within the dendrons, CEN is colored in green at 10 mM and by atom at 150 mM NaCl, REP in yellow and SPM in red. The energetic of binding (ΔG, ΔH and -TΔS) are expressed in kcal mol^{-1}. The Cl^- and Na^+ atoms are colored in green and purple respectively – at 150 mM NaCl counterions are represented as smaller spheres for clarity. Water molecules are not shown for clarity.

As described in the text, the ΔG_{G1-V1} are at the same level of the binding energies reported in Figure S4 (also ΔG_{G1-V1} values are multiplied by three in order to allow comparison with ΔG_{G2-V1} and the 1:1 binding values), meaning that the presence of V2 and of the second **G1** does not have any effect in the direct V1-**G1** binding. The same is true for the second **G1**, since the binding energy ΔG_{G1-V2} is equal to ΔG_{G1-V1}. Under this light, the energetic of the V1-V2+**G1** assembly ($\Delta G_{assembly}$) was calculated as the free energy of binding between the complex formed by the first **G1** and V1 and the one composed by the second **G1** with V2. Importantly, the $\Delta G_{assembly}$ for V1-

V2+**G1** results to be entropically driven with a very low favorable enthaplic contribution independently on the amount of salt in solution – typical values of a hydrophobic association between hydrophobic molecules in solution. Figure S8 shows the same images and data for **G2**.

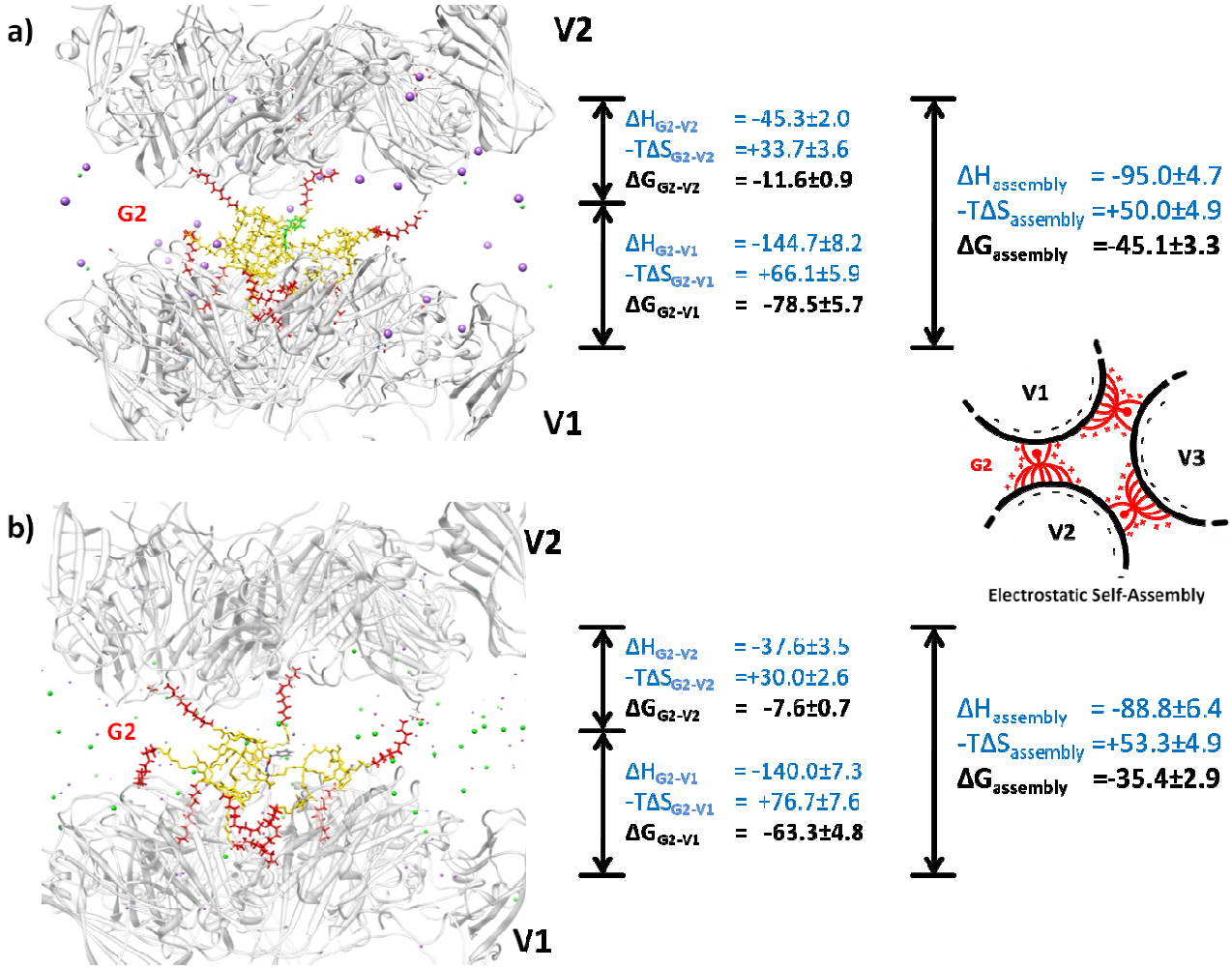


Figure S8. The V1-V2+**G2** systems at 10 (a) and 150 mM (b) NaCl. Within the dendrons, CEN is colored in green at 10 mM and by atom at 150 mM NaCl, REP in yellow and SPM in red. The energetic of binding (ΔG, ΔH and -TΔS) are expressed in kcal mol^{-1}. The Cl^- and Na^+ atoms are colored in green and purple respectively – at 150 mM NaCl counterions are represented as smaller spheres for clarity. Water molecules are not shown for clarity.

In this system ΔG_{G2-V1} is evidently not equal to ΔG_{G2-V2} due to the different number of SPM that attract V1 and V2. In particular, differently from the V1-V2+**G1** case (Figure S7), ΔG_{G2-V1} is different from the numbers reported by the 1:1 systems (Figure S5) due to the stretching of the

dendron indyced by the presence of V2 – the entropic cost is the same, but the enthaplic attraction is rather reduced. Due to the complexity of the V1-V2 interface created by **G2** and represented by the scheme in Figure S8, the $\Delta G_{assembly}$ for V1-V2+**G2** was calculated as the average between $\Delta G_{G2\text{-}V1}$ and $\Delta G_{G2\text{-}V2}$ – from the statistical point of view, in fact, this was considered as the most reliable interpretation of the cohesive energy induced by **G2** between V1 and V2. The $\Delta G_{assembly}$ generated by **G1** and **G2** is more than one order of magnitude different, and this is also more evident at 150 mM NaCl underlining the different interactions that lie behind the self-assembly. This is fully consistent with the trend in virus assembly abilities reported in Figure 1 in the manuscript, reinforcing the concept that **G1** and **G2** generates different CCMV virus assemblies due to the different manner they bind to the capsid surface.

References

S1 Pavan, G. M.; Kostiainen, M. A.; Danani, A. *J. Phys. Chem. B.*, **2010**, 114, 5686–5693.

S2 Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. USA,* **2001**, *98*, 10037-10041.

S3 a) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Res.*, **2007**, *35*, W522- W525. b) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. *Nucleic Acids Res.*, **2004**, *32*, W665-W667.

S4 Zhang, D.; Konecny, R.; Baker, N. A.; McCammon, J. A. Biopolymers, **2010**, *75*, 325-337.

S5 a) Pavan, G. M.; Danani, A.; Pricl, S.; Smith, D. K. *J. Am. Chem. Soc.*, **2009**, *131*, 9686-9694. b) Jones, S. P.; Pavan, G. M.; Danani, A.; Pricl, S.; Smith, D. K. *Chem. Eur. J.*, **2010**, *16, 4519-4532*. c) Kostiainen, M. A.; Kotimaa, J.; Laukkanen, M.-L.; Pavan, G. M. *Chem. Eur. J.,* **2010**, *16*, 6912-6918.

S6 Jorgensen W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–35.

S7 Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Robertson, B..; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D.R.; Mathews, D.H.; Seetin, M.G.; Sangui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A., AMBER 11. In University of California, San Francisco, **2010**.

S8 Kostiainen, M. A.; Kasyutich, O.; Cornelissen, J. J.; Nolte, R. J. *Nature Chem.,* **2010**, *2*, 394-399.

S9 a) Pavan, G. M.; Albertazzi, L.; Danani, A. *J. Phys. Chem. B.*, **2010**, 114, 2667-2675. b) Pavan, G. M.; Mintzer, M. A.; Simanek, E. E.; Merkel, O. M.; Kissel, T.; Danani, A. *Biomacromolecules,* **2010**, 11, 721–730. c) Pavan, G. M.; Posocco, P.; Tagliabue, A.; Maly, M.; Malek, A.; Danani, A.; Ragg, E. M.; Catapano, C.; Pricl, S. *Chem. Eur. J.*, **2010**, *16*, 7781-7795.

S10 a) Jakalian, A.; Bush, B.L.; Jack, D.B.; Bayly, C.I. *J. Comput. Chem.* **2000**, *21***,** 132-146. b) Jakalian, A.; Jack, D.B.; Bayly, C.I. *J. Comput. Chem.* **2002**, *25***,** 1623-1641.

S11 Wang, J.; Wang, W.; Kollman, P.A.; Case, D.A. *J. Mol. Graphics Model.* **2006**, *25***,** 247-260.

S12 Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25***,** 1157-1174.

S13 Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.,* **1998**, *98*, 10089-10092.

S14 a) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys*. **1977**, *23*, 327. b) Krautler, V.; van Gunsteren, W. F.; Hanenberger, P. H. *J. Comput. Chem*. **2001**, *5*, 501.

S15 Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.

S16 Srinivasan, J.; Cheatham, T. E.Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401-09.

S17 Jayaram, B.; Sprous, D.; Beveridge, D. L.; *J. Phys. Chem.* **1998**, *102*, 9571-9576.

S18 Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem*. **1994**, *98*, 1978-1988.

S19 Luo, R.; David, L.; Gilson. M. K. *J. Comput. Chem*. **2002**, *23*, 1244-1253.

S20 Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers*. **1996**, *38*, 305-20.

S21 Andricioaei, I.; Karplus, M. J. *Chem. Phys.* **2001**, *115*, 6289-92.

S22 Weiser, J.; Shenkin, P. S.; Clark Still. W. J. *Comput. Chem*. **1999**, *20*, 217-230.

S23 Pavan, G. M.; Danani, A. *Phys. Chem. Chem. Phys.*, **2010**, *12*, 13914 – 13917.