Supporting Information for:
Beyond icosahedral symmetry in packings of proteins in spherical shells

Majid Mosayebiab, Deborah Shoemarkbcd, Jordan M. Fletcherd, Richard B. Sessionsbcd, Noah Lindena, Derek N. Woolfsonbcd, and Tanniemola B. Liverpoolab

aSchool of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK; bBrisSynBio, Tyndall Avenue, Bristol BS8 1TQ, UK; cSchool of Biochemistry, University of Bristol, Bristol BS8 1TD, UK; dSchool of Chemistry, University of Bristol, Cantock’s Close, Bristol BS8 1TS, UK

S I. CC-level model of SAGEs

We performed molecular dynamics simulation, where the mass, energy, and length units are chosen to be $m_0 = 876.7$ Da, $\epsilon_0 = 4.14 \times 10^{-21}$ J, and $l_0 = 1$ nm, implying a time unit of $\tau_0 = (m_0 l_0^2/\epsilon_0)^{1/2} = 1.87 \times 10^{-11}$ s. In the CC-level model of SAGEs, each CC peptide has a mass of $m_0$ and is modeled as a rigid body composed of, three spherical particles on top of each other (each with a mass, $m = 0.17 m_0$). The three spheres are visualized as a cylinder in Fig. 1-B, and three/six distinct attractive patches ($m = 0.082 m_0$ for trimer-forming CC patches and $m = 0.17 m_0$ for dimer-forming CC patches) on its surface (53). The six trimer patches are arranged in two lines with an angle of $60^\circ$ with respect to the origin of the CC, hence favoring formation of tree-fold symmetric trimers. Whereas, the single line of three patches in dimer-forming CCs favors formation of two-fold symmetric dimers. The relative positions of all particles within a rigid body are kept fixed during the simulation using an integration technique based on the Richardson iterations which is implemented in the LAMMPS molecular dynamics package (55). We took the integration time step of $dt = 10^{-3} \tau_0$, and used the Langevin thermostat with $\tau_{\text{damp}} = 10 \tau_0$ for thermostating the translational and rotational degrees of freedom of our rigid bodies.

The attraction between patches and the excluded volume interactions are modeled with a pair-wise LJ interaction of the form

$$V_{\text{LJ}}(r) = 4\epsilon_{\text{LJ}} \left[ (\sigma_{\text{LJ}}/r)^{12} - (\sigma_{\text{LJ}}/r)^{6} \right],$$

where $r$ is the separation, $\sigma_{\text{LJ}}$ is the LJ diameter, and $\epsilon_{\text{LJ}}$ is the interaction strength of the two interacting sites. The interactions are truncated and shifted to zero at a cut-off distance $r_c$. The parameters of the potentials are listed in Table S I. The strength of attraction between complementary and non-complementary dimer-forming patches were tuned to reproduce the experimental melting temperatures (56, 57). Because we were only interested in simulating the assembly of pre-formed hubs, we have chosen a stronger attraction between the patches of the trimer-forming CCs to prevent dissociation of trimeric CC bundles.

A harmonic bond of length $l_0$ with a stiffness of $400 \epsilon_0/l_0^2$ used to permanently link a trimeric CC to a dimeric one. This bond models the disulfide bond in the SAGE design. Moreover, the bend angles $\theta_i$ and the twist angle $\phi$ (see Fig. S1) are controlled with the following potentials

$$V_{\text{bend}}(\theta) = K_\theta (\theta - \theta_0)^2, \quad \theta_0 = \frac{1.056}{2} \pi,$$

Fig. S1. CC-level CG model of SAGEs. A. side view of the elementary unit of the SAGE assembly, which is two CCs linked together with a permanent harmonic bond (thick black line). Each CC was modeled as a rigid body made from three LJ particles (represented by circles) with attractive patchy particles on its surface. B. The top-view of the complementary attractive LJ patches (shown with dots). The permanent bond also penalizes the bend angles $\theta_1$ and twist angle $\phi$ of the linked CCs from their respective equilibrium values.
Table SI. LJ interaction parameters of the CC-level CG model of SAGEs.

<table>
<thead>
<tr>
<th>Interacting sites</th>
<th>$\sigma_{LJ}/l_0$</th>
<th>$\epsilon_{LJ}/\epsilon_0$</th>
<th>$r_c/l_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC particles</td>
<td>0.9</td>
<td>60</td>
<td>1.0</td>
</tr>
<tr>
<td>Complementary dimer-forming patches</td>
<td>0.36</td>
<td>16</td>
<td>0.6</td>
</tr>
<tr>
<td>Non-complementary dimer-forming patches</td>
<td>0.36</td>
<td>8</td>
<td>0.6</td>
</tr>
<tr>
<td>Complementary trimer-forming patches</td>
<td>0.36</td>
<td>18</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Fig. S2. Angular specificity of interacting hubs in CG and atomistic simulations. Distribution of bending $\Theta_i$ and twist $\Phi$ angles between hub pairs in the CC-level CG model (left) compared with the results from atomistic simulations (right). $\Theta = 0^\circ$ corresponds to the situation where the hub axis is perpendicular to the connecting vector between hub pairs. Lines are fits to the Gaussian distributions $P(\alpha) = \left(2\pi \sigma_\alpha^2\right)^{-1/2} \exp \left[-\frac{(\alpha - \mu_\alpha)^2}{2\sigma_\alpha^2}\right]$. In the CC-level CG model, widths of the distributions $\{\sigma_\alpha\}$ are approximately 6 times narrower than in atomistic simulations.

and

$$V_{\text{twist}}(\phi) = K_{\phi} \left(1 - \cos(2\phi)\right), \quad [S3]$$

where $K_\theta = 400\epsilon_0/rad^2$ and $K_\phi = 120\epsilon_0$. To prevent the CCs from freely rotating along their central axis, deviation of the angle $\eta$ from $\eta_0 = \pi$ (see Fig. S1) was penalized with a harmonic potential of the form Eq. S2 with $K_\eta = 500\epsilon_0/rad^2$. We note that the fluctuations of the angles about their equilibrium values are coupled in this model. The bending and twisting stiffnesses of interactions are chosen such that the overall angular specificity of the hub pairs in the CC-level model, which arises from the finite range of attractive LJ interactions between the patches and also from the stiffness of the permanent bonds, was much less than the angular specificity of the atomistic simulations (See Fig. S2), hence facilitating the formation of the error-free honeycomb network in numerical simulations.

S II. Details of the atomistic simulations

Atomistic simulations were performed under the Amber99SB-ildn forcefield using GROMACS 4.6.7 under conditions of constant pressure (1 bar) and temperature (300 K) with the PME method used for long range electrostatics. The simulation box was 4 nm larger than the SAGE model in each dimension, and filled with TIP3P water and 0.1 M sodium chloride. Simulations were run out to 1 $\mu$s using the UK HPC Facility Archer. Full details will be published elsewhere.
Fig. S3. Ideal spherical packing composed of hexagons and pentagons with $R = 1.603\sigma$ at $kT = 0.15\sigma$. The ratio of activities is $z_P/z_H = \exp(\Delta\mu/kT) = 12/20$. 

$\langle \hat{Q}_6 \rangle = 0.761851$

$\langle \hat{W}_6 \rangle = 0.916486$

$2.5$

$2.3$

$E/\epsilon$

$\langle E \rangle = -2.61678$

$\langle N \rangle = 32.0007$

$N$

$N_H$

$N_P$

$\langle \{N_i\} \rangle = 17.9503, 14.0503$

$0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6$

$\hat{Q}_6$

$\hat{W}_6$
Fig. S4. Perturbed spherical packing composed of hexagons, pentagons and squares with $R = 1.603\sigma$ at $kT = 0.15\epsilon$ with $z_P/z_H = 12/20$ and $z_S/z_P = 0.1$. The packing in Fig. S6-D with a $D_{5h}$ symmetry is responsible for the drop of $E$ when $\hat{Q}_6 \rightarrow 0$. 

\[
\langle \hat{Q}_6 \rangle = 0.171576
\]

\[
\langle \hat{W}_6 \rangle = 0.033391
\]

\[
\langle E \rangle = -2.58978
\]

\[
\langle N \rangle = 35.9776
\]
The perfect icosahedral arrangement with a slightly larger energy is also shown in Fig. S5. A view of these 3D configurations.

\[ \beta = 0 \]

Fig. S5. 3D interactive illustrations of the snapshots in Fig. 3. i–w are typical ideal packings at \( kT = 0.15 \) (i–iii) and \( kT = 0.3 \) (iv), and v is a typical perturbed packing at \( kT = 0.15 \) and \( z_S/z_H = 0.01 \). Copper, blue and cyan particles represent hexagons, pentagons and squares respectively. Adobe Acrobat can be used to interactively change the viewpoints of these 3D configurations.

\[ \hat{Q}_6 = 0.737, \hat{W}_6 = 1.000 \]
\[ N_H = 240, N_P = 72 \]

\[ \hat{Q}_6 = 1.000, \hat{W}_6 = 1.000 \]
\[ N_H = 300, N_P = 12 \]

\[ \hat{Q}_6 = 0.98, \hat{W}_6 = 0.514 \]
\[ N_H = 16, N_P = 16 \]

\[ \hat{Q}_6 = 0.005, \hat{W}_6 = -0.533 \]
\[ N_H = 15, N_P = 10, N_P = 12 \]

Fig. S6. Clustering of non-hexagonal particles leads to lower than icosahedral energies for large ideal spherical packings in A, where \( \hat{E}_{\text{min}} = -2.9076\epsilon \) and \( R = 5.0678\sigma \). The perfect icosahedral arrangement with a slightly larger energy is also shown in B, where \( \hat{E}_{\text{min}} = -2.8497\epsilon \) and \( R = 5.2408\sigma \). For the smaller ideal systems that we have studied here, the icosahedral packings was the global energy minimum, and the clustering of the pentagonal particles could only lead to local minima. An example with \( R = 1.603\sigma \) that has \( D_{1h} \) point group symmetry is shown in C, where \( \hat{E}_{\text{min}}^{D_{1h}} = -2.6965\epsilon \), while the corresponding icosahedral energy is \( \hat{E}_{\text{min}}^{\text{icos}} = -2.7929\epsilon \). For the perturbed systems, however, the clustering could lead to energies lower than \( \hat{E}_{\text{min}}^{\text{icos}} \). For instance, see the packing in D with \( D_{2h} \) point group symmetry which has an energy of \( \hat{E}_{\text{min}}^{D_{2h}} = -2.8086\epsilon \). The optimum packings shown in C and D are responsible for the drop of the average energy at small \( \hat{Q}_6 \) values in Fig. 3. However, they are not typically sampled at the temperature shown in Fig. 3 (i.e. at \( kT = 0.15\epsilon \)), consistent with findings in Ref. (44). Adobe Acrobat can be used to interactively change the viewpoints of these 3D configurations.

\[ \hat{Q}_6 \approx 1.0, \hat{W}_6 \approx 1.0 \]
\[ N_H = 20, N_P = 12 \]

\[ \hat{Q}_6 \approx 0.87, \hat{W}_6 = 0.98 \]
\[ N_H = 19, N_P = 13 \]

\[ \hat{Q}_6 \approx 0.77, \hat{W}_6 = 0.95 \]
\[ N_H = 18, N_P = 14 \]

\[ \hat{Q}_6 \approx 0.08, \hat{W}_6 = -0.1 \]
\[ N_H = 3, N_P = 35 \]

\[ \hat{Q}_6 \approx 0.17, \hat{W}_6 = 0.07 \]
\[ N_H = 14, N_P = 17, N_P = 3 \]

Fig. S7. Free energy \( F \), energy \( E \) and the number of non-hexagonal particles as a function of normalized BOOs \( \hat{W}_6 \) and \( \hat{Q}_6 \) at \( kT = 0.2\epsilon \) for an ideal system composed of hexagons and pentagons (solid lines) and perturbed systems composed of hexagons, pentagons and squares with \( z_S/z_H = 1 \) (dashed lines). Blue and green lines show packings with \( R = 1.603\sigma \) (\( N = 32 \) when \( T \to 0 \)) and \( R = 2.469\sigma \) (\( N = 72 \) when \( T \to 0 \), respectively.

Mosayebi et al.
Fig. S8. Top panel plots display average normalized BOOs and the energy $E$ as a function of the temperature for the ideal packing (solid lines) and for the perturbed packing (dashed lines) with $z_{P}/z_{H} = 1$. Average number of species as a function of temperature are also plotted in the bottom panel plots. Blue and green lines show packings with $R = 1.603\sigma$ ($N = 32$ when $T \rightarrow 0$, and $z_{P}/z_{H} = 12/20$) and $R = 2.469\sigma$ ($N = 72$ when $T \rightarrow 0$, and $z_{P}/z_{H} = 12/60$), respectively. Error bars represent two standard deviations away from the mean value obtained from 20 to 40 independent simulations.

Fig. S9. Top panel plots display average normalized bond orientational order parameters and the energy as a function of $z_{P}/z_{H}$ for an ideal packing at $T = 0.15\sigma$ and $R = 1.603\sigma$. Average number of species as a function of temperature are also plotted in the bottom panel plots. Note that the definition of BOOs we used in this study suffers in regions where there are only a few non-hexagonal particles (i.e. where $z_{P} \ll z_{H}$). Error bars represent two standard deviations away from the mean value obtained from 20 independent simulations.