Supplemental Information

t-Stochastic Neighborhood Embedding (t-SNE)

We utilize the t-Stochastic neighborhood embedding, as implemented in Scikit-learn[1–4], to map the 13-dimensional vector representing the relative position and orientation of a particle pair to two-dimensions (2D), allowing for particle pair motifs to be clustered and identified. An example of the analysis pipeline is shown in Figure 1. As discussed in the main text, the vector is

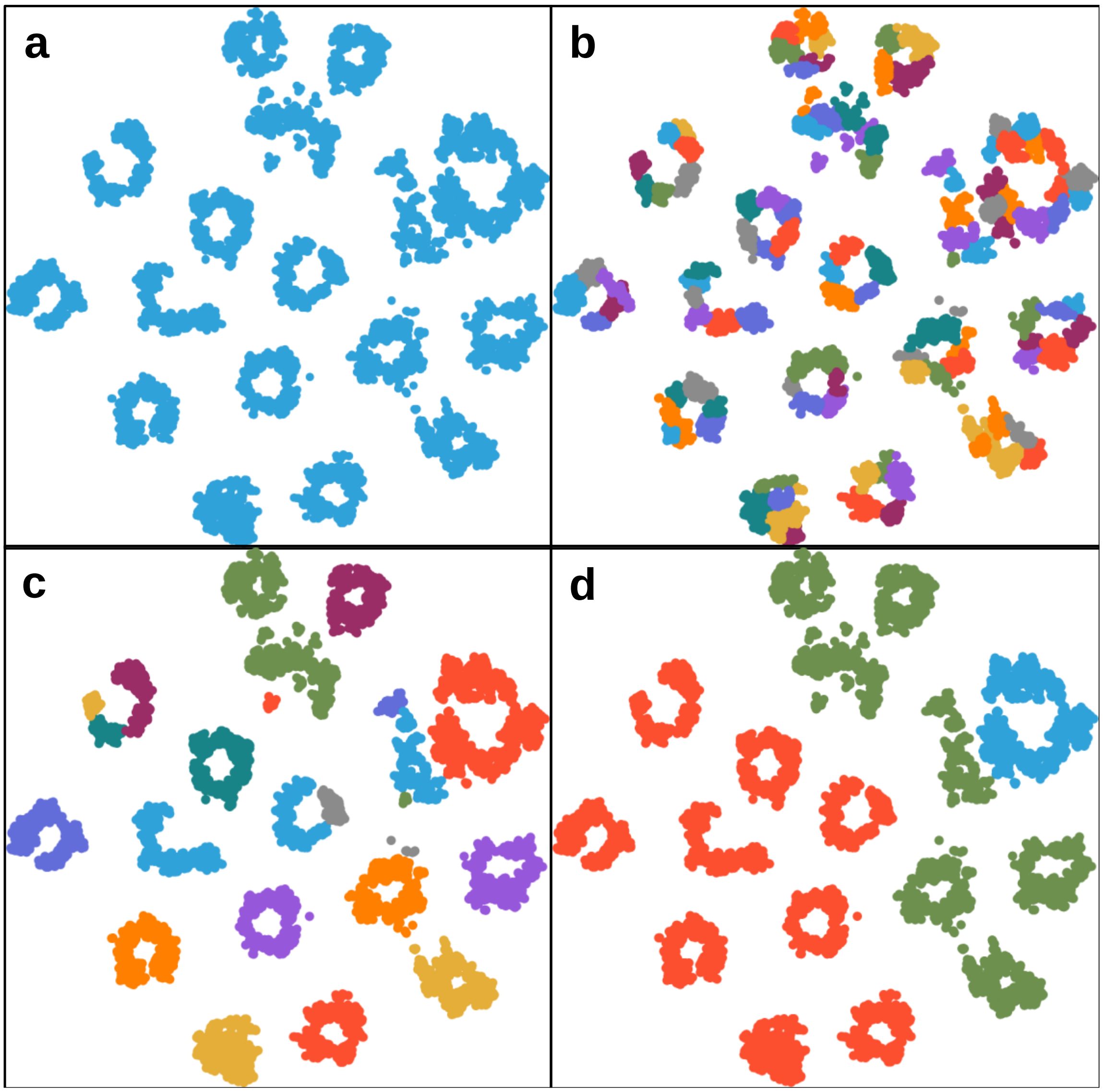

Figure 1 Example of t-SNE analysis pipeline. **a**. Output of the t-SNE dimensionality reduction technique, mapping the 13-dimensional vector describing the position and orientation of a particle pair into two-dimensional (2D) space. **b.** Use of a large number ($n_{\text{clusters}} \gg n_{\text{motifs}}$) of gaussians to generate possible groupings of similar motifs. Each color represents a different cluster as identified by the gaussian mixture model (note: the number of clusters is $n_{\text{clusters}} > 50$, so same colors do not necessarily represent the same cluster. **c.** Use of entropy-based cluster merging to reduce the number of identified clusters. Now that the number of clusters is reduced, contiguous clusters colored the same represent the same cluster. **d.** Hand-identified clusters representing each unique class of motif. Each color represents a unique motif class.

created from the vector and quaternion representing the position and orientation of a pair of particles: $(\vec{v}_i, \vec{q}_i), (\vec{v}_j, \vec{q}_j)$. These are combined to create the 13D vector: $(\hat{r}_{ij}, \hat{u}_{ij}, \hat{r}_{ji}, \hat{u}_{ji}, |\hat{r}|)$. The t-SNE dimensionality reduction (Figure 1**a**) results in clusters in which the particle pairs are similarly arranged. These clusters can be automatically identified *via* Gaussian mixture models

(Figure 1**b**) and combined (Figure 1**c**)[3,5,6] to deliver a small number of classes of motifs. These classes of motifs are then classified by hand and associated with particular configurations of particle pairs (Figure 1**d**).

t-SNE analysis of cube derivatives

Below we present the t-SNE dimensionality reduction each of the cube derivatives (TP: Figure 2, S3: Figure 3, HFP and RFP: Figure 4), with each particle pair colored by the motif class to which it belongs. The number of pairs belonging to each motif is then used to understand the ability for a particular derivative to self-assemble the target cubic crystal (see main text Figure 6).

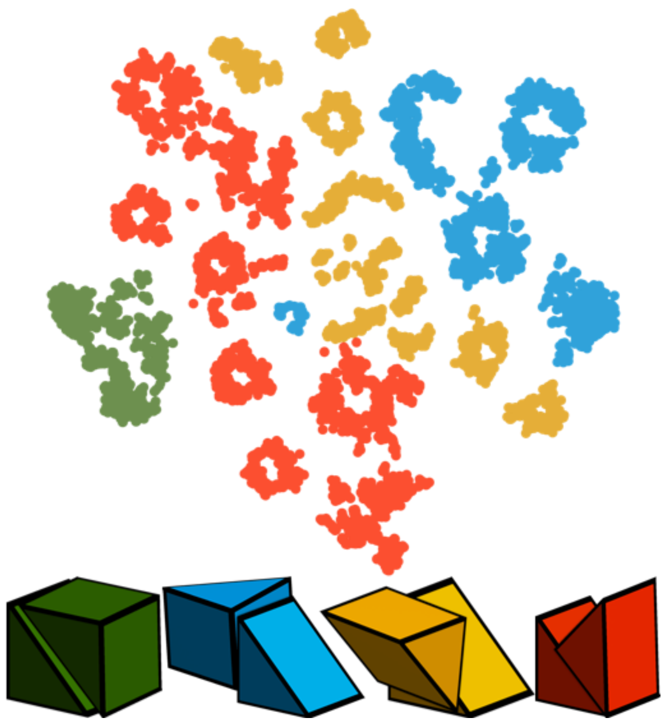

Figure 2 t-SNE analysis of the fluid phase for the TP. Color-coded representative motifs are included below: cut-face pair (green), square-face pair (blue), contributing motif (yellow), and competing motif (red).

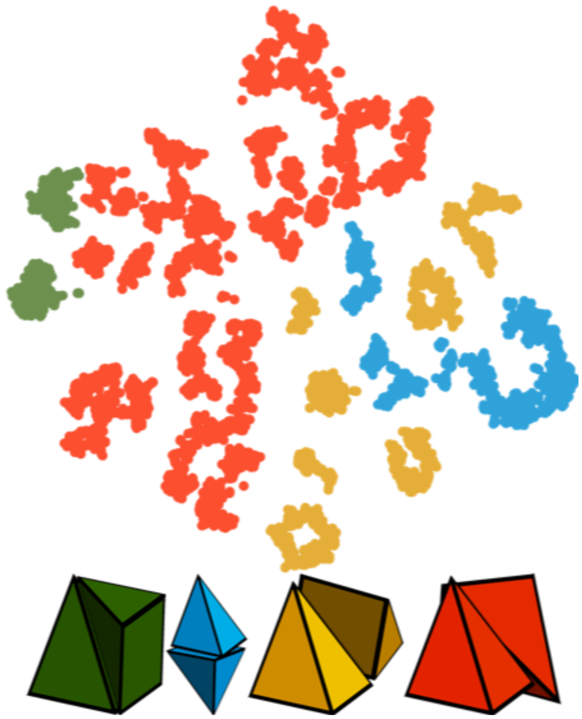

Figure 3 t-SNE analysis of the fluid phase for the S3 shape. Color-coded representative motifs are included below: cut-face pair (green), square-face pair (blue), contributing motif (yellow), and competing motif (red).

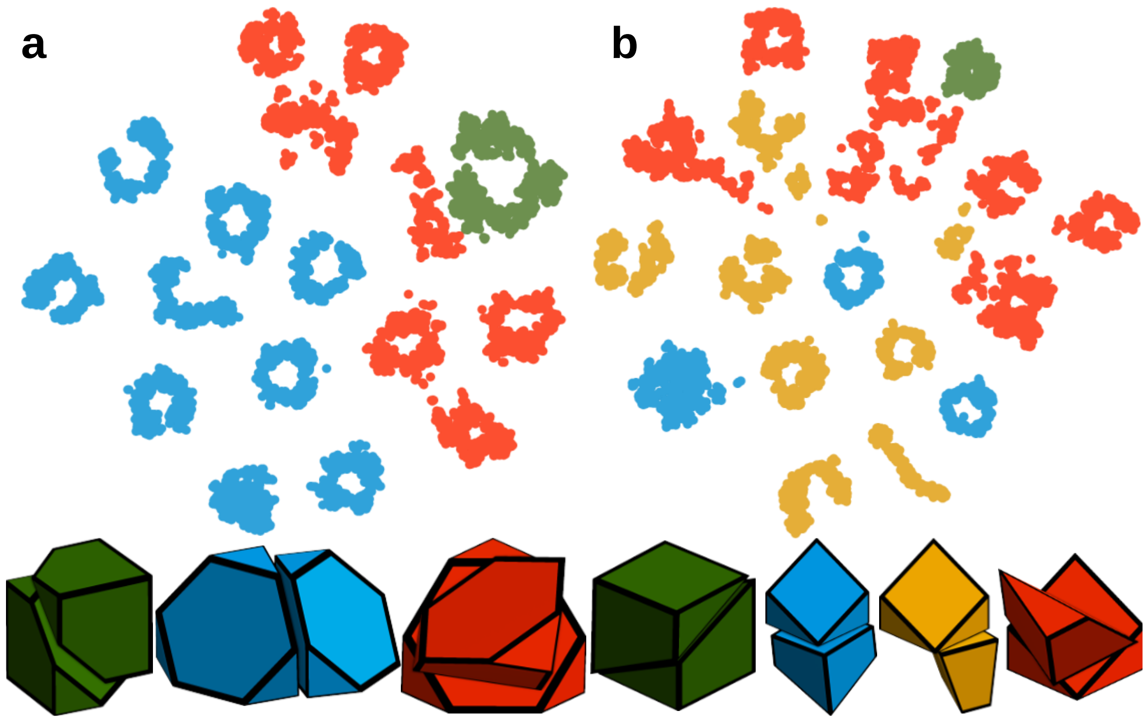

Figure 4 Comparison of the t-SNE analysis of the (color-coded) fluid phase motifs for **a** HFP and **b** RFP. Note that the HFP only forms motifs that directly contribute to the final crystal structure because its three faces originating from the square faces of the cube are congruent. This is not the case for the RFP. The square face is not congruent with the other faces that originate from the square faces (**b**, yellow motif). Also note that the two large faces originating from the square faces are not congruent; being chiral, one left-handed face must pair with a right-handed face to form a square-face pair (blue motif), while two left-handed or two right-handed faces will form a contributing motif (yellow motif).

Supplementary Material References

1 L. Van Der Maaten and G. Hinton, *J. Mach. Learn. Res. 1*, 2017, **620**, 267–284.

2 L. Van Der Maaten, *J. Mach. Learn. Res.*, 2014, **15**, 3221–3245.

3 M. Spellings and S. C. Glotzer, *AIChE J.*, 2018, **64**, 2198–2206.

4 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.

5 A. P. Dempster, N. M. Laird and D. B. Rubin, *J. R. Stat. Soc. Ser. B Methodol.*, 1977, **39**, 1–38.

6 J. P. Baudry, A. E. Raferty, Celeux Gilles, Lo Kenneth and R. Gottardo, *J. Comput. Graph. Stat.*, 2010, **19**, 332–353.