

Supplementary information: Multi-step nucleation pathway of C-S-H during cement hydration from atomistic simulations

Xabier M. Aretxabaleta^{1,*}, Jon López-Zorrilla¹, Iñigo Etxebarria^{1,2}, and Hegoi Manzano^{1,*}

¹Fisika saila, Euskal Herriko Unibertsitatea UPV/EHU, Sarriena Auzoa z/g, 48940 Leioa, Basque Country, Spain

²EHU Quantum Center, Euskal Herriko Unibertsitatea, UPV/EHU, Spain

*xabier.mendez@ehu.eus

ABSTRACT

C-S-H nucleation is a crucial step during cement hydration and determines to a great extent the rheology, microstructure, and properties of the cement paste. Recent evidence indicates that the C-S-H nucleation involves at least two steps, yet the underlying atomic scale mechanism, the nature of the primary particles and their stability, or how they merge/aggregate to form larger structures is unknown. In this work, we use atomistic simulation methods, specifically DFT, evolutionary algorithms (EA), and Molecular Dynamics (MD), to investigate the structure and formation of C-S-H primary particles (PPs) from the ions in solution, and then discuss a possible formation pathway for the C-S-H nucleation. Our simulations indicate that even for small sizes the most stable clusters encode C-S-H structural motifs, and we identified a $C_4S_4H_2$ cluster candidate to be the C-S-H basic building block. We suggest a formation path in which small clusters formed by silicate dimers merge into large elongated aggregates. Upon dehydration, the C-S-H basic building blocks can be formed within the aggregates, and eventually crystallize.

Ca and Si speciation

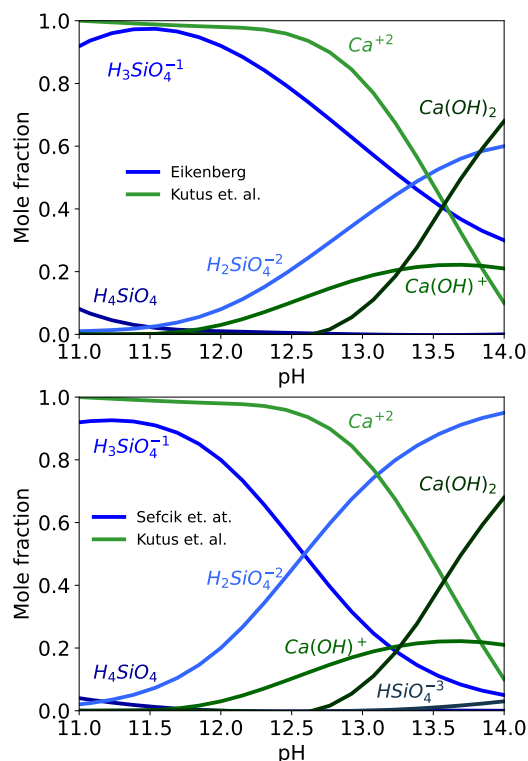


Figure S1. Ca and Si speciation in terms of pH from different sources. The data was obtained from ref¹⁻³.

Evolutionary Algorithms calculations energy evolution.

Figures S2-S10 show the energy evolution of all the individuals of every EA calculation. Pink dots represent structures created randomly, blue dots structures created by heredity operators, orange dots structures created by mutation operators and green dots structures that were kept for the next generation. The black line represents the lowest energy structure for every generation and the filled green curve represents the energy density of all the created structures.

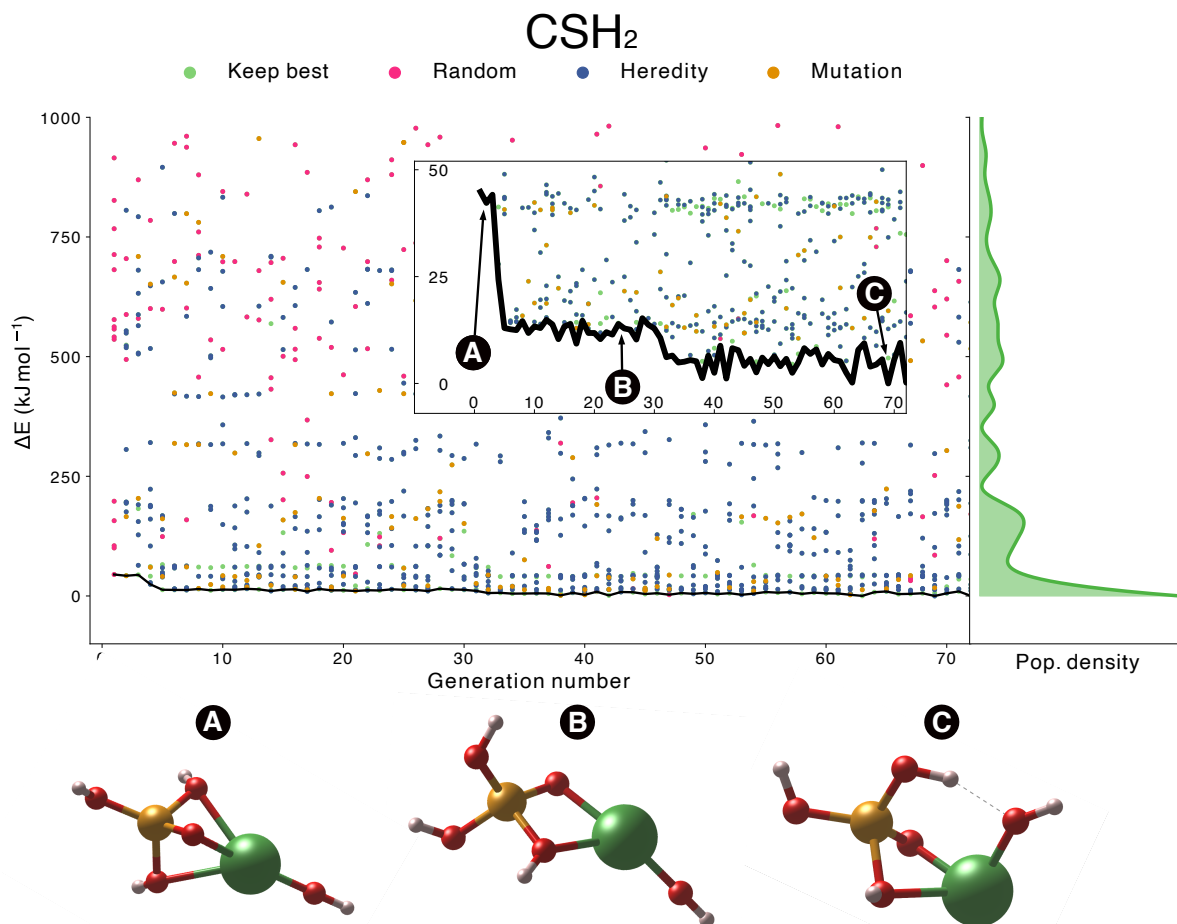


Figure S2. Energy evolution and energy population density of the Evolutionary Algorithms calculation for CSH₂ PPs. Each point represents a structure. Green points represent structures that has been kept from the previous generation. Pink points represent structures that have been created randomly. Blue points represent structures that have been created by using the heredity operator. Orange points represent structures that have been created using the mutation operator. The structures A, B, and C represent the best PP in generations 1, 25 and 70.

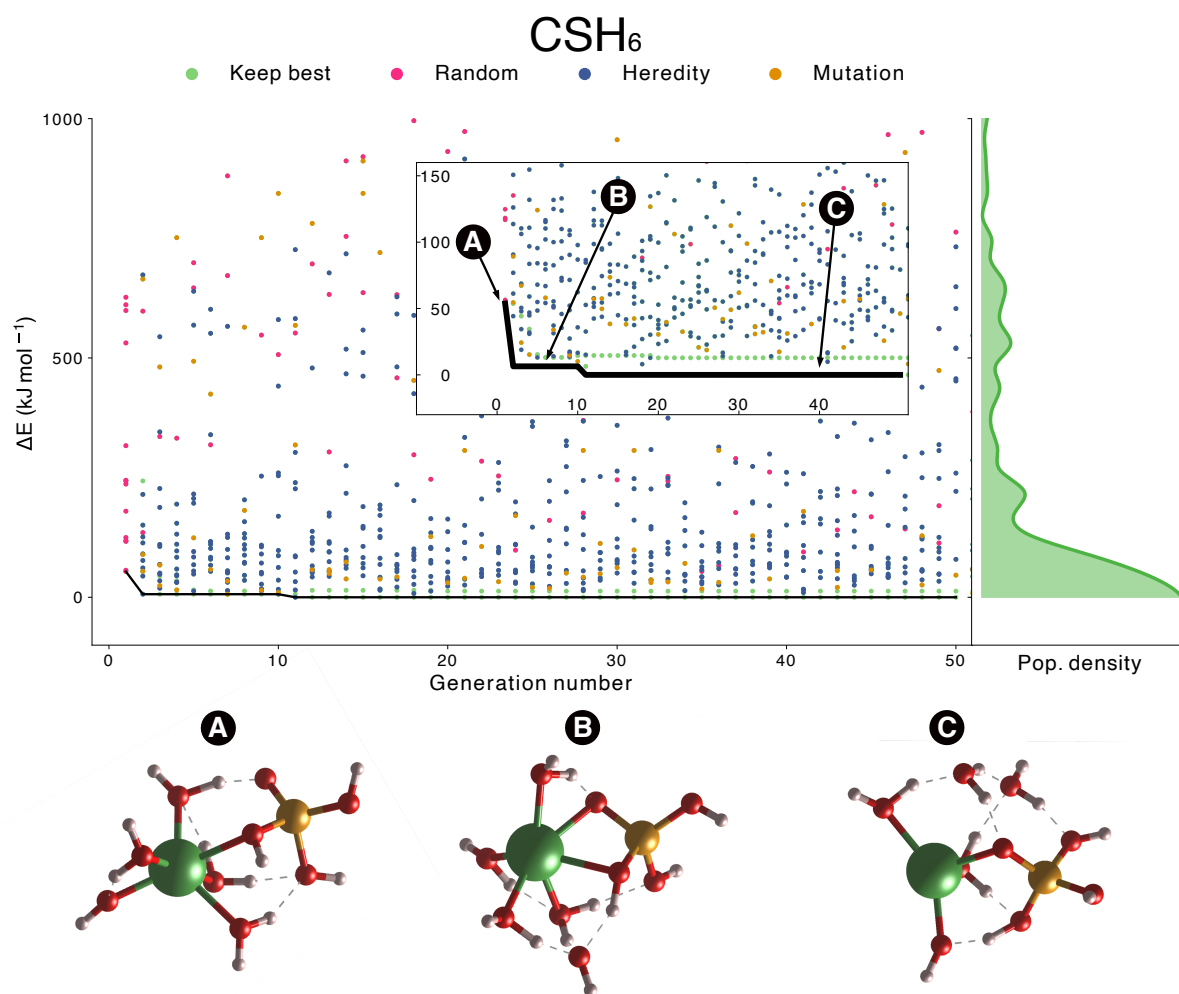


Figure S3. Energy evolution and energy population density of the Evolutionary Algorithms calculation for CSH₆ PPs. The structures A, B, and C represent the best PP in generations 1, 5 and 40.

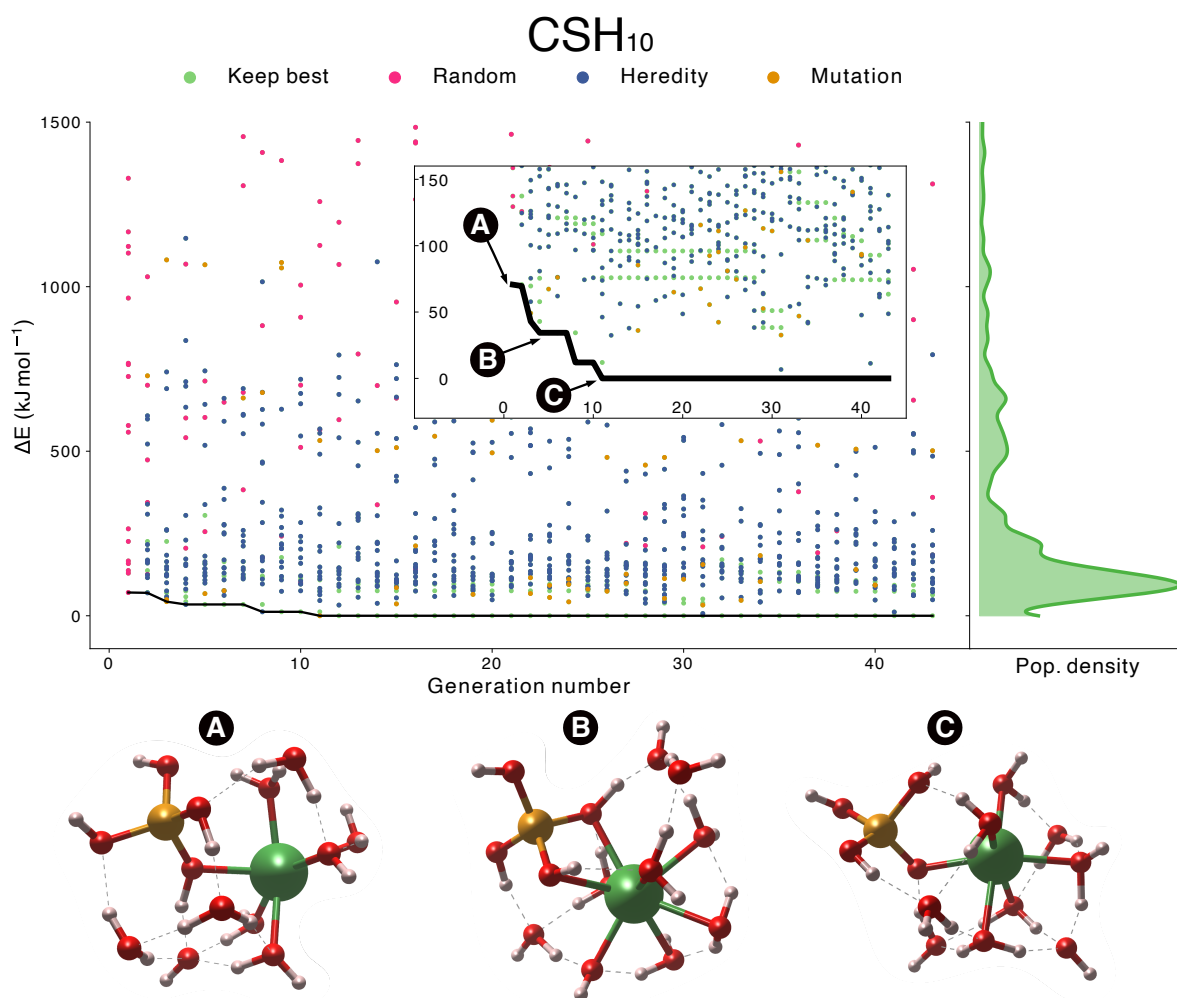


Figure S4. Energy evolution and energy population density of the Evolutionary Algorithms calculation for CSH₁₀ PPs. Each point represents a structure. Green points represent structures that has been kept from the previous generation. Pink points represent structures that have been created randomly. Blue points represent structures that have been created by using the heredity operator. Orange points represent structures that have been created using the mutation operator. The structures A, B, and C represent the best PP in generations 1, 5 and 12.

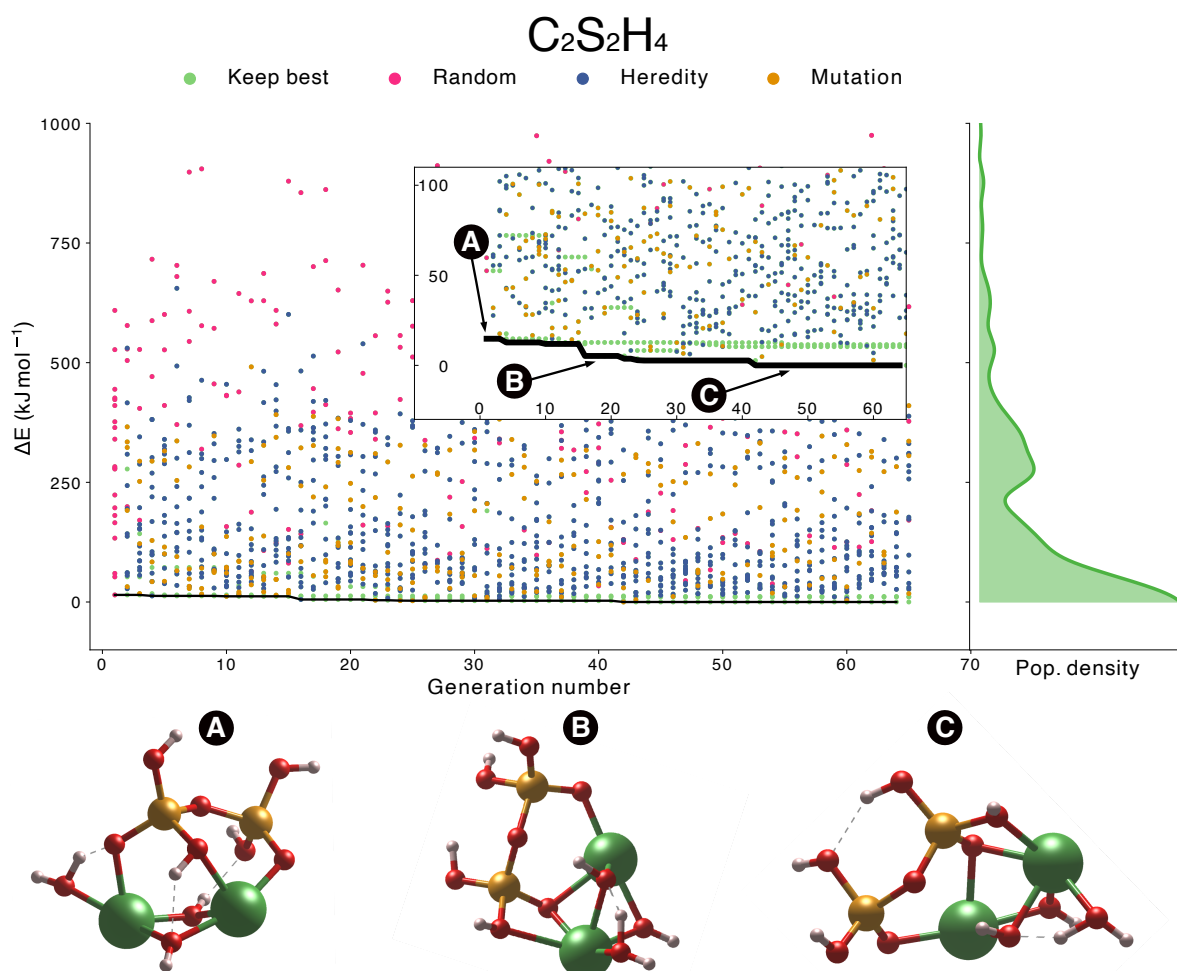


Figure S5. Energy evolution and energy population density of the Evolutionary Algorithms calculation for $C_2S_2H_4$ PPs. The structures A, B, and C represent the best PP in generations 1, 20 and 50.

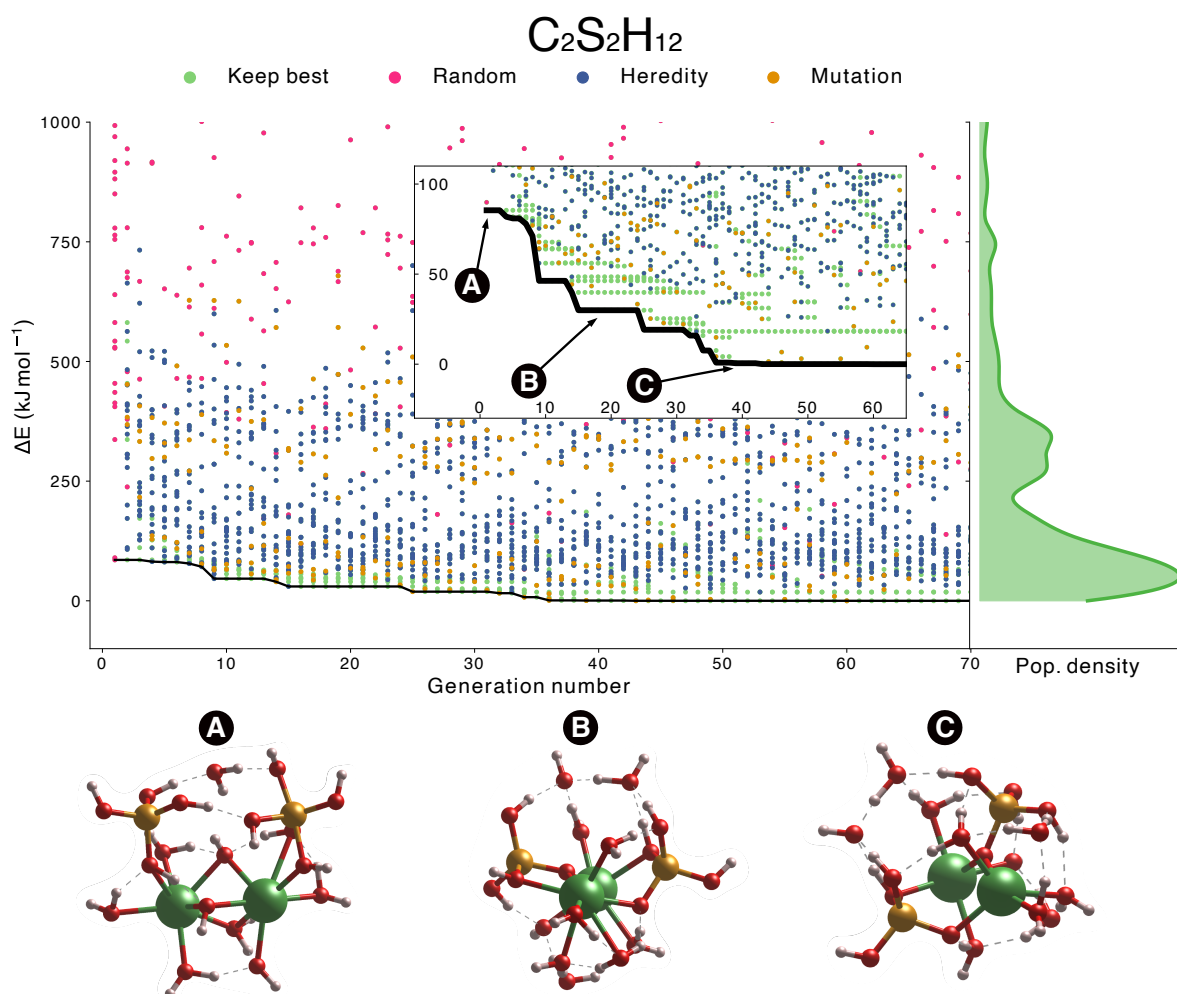


Figure S6. Energy evolution and energy population density of the Evolutionary Algorithms calculation for $C_2S_2H_{12}$ PPs. Each point represents a structure. Green points represent structures that have been kept from the previous generation. Pink points represent structures that have been created randomly. Blue points represent structures that have been created by using the heredity operator. Orange points represent structures that have been created using the mutation operator. The structures A, B, and C represent the best PP in generations 1, 20 and 40.

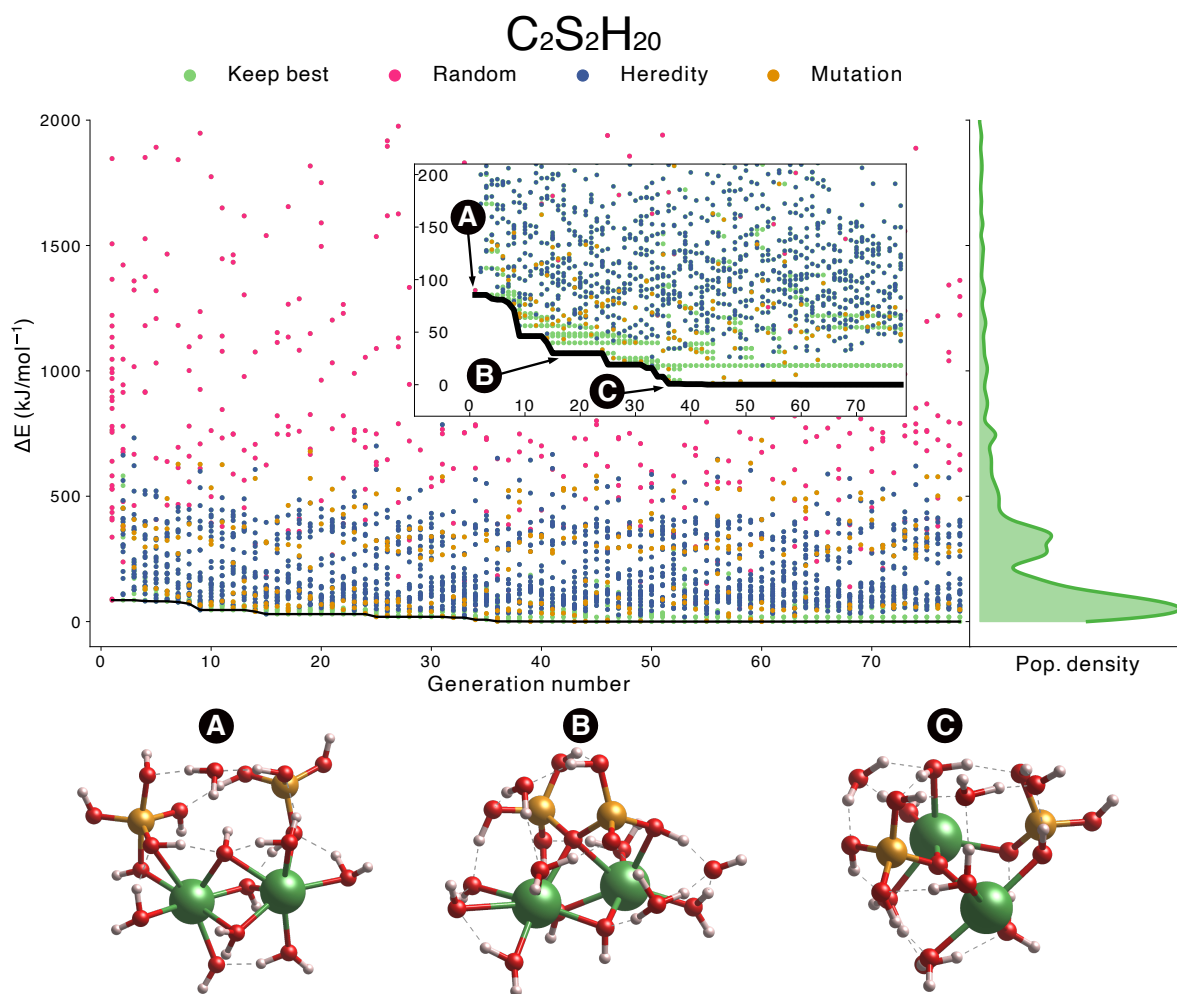


Figure S7. Energy evolution and energy population density of the Evolutionary Algorithms calculation for $C_2S_2H_{20}$ PPs. The structures A, B, and C represent the best PP in generations 1, 20 and 40.

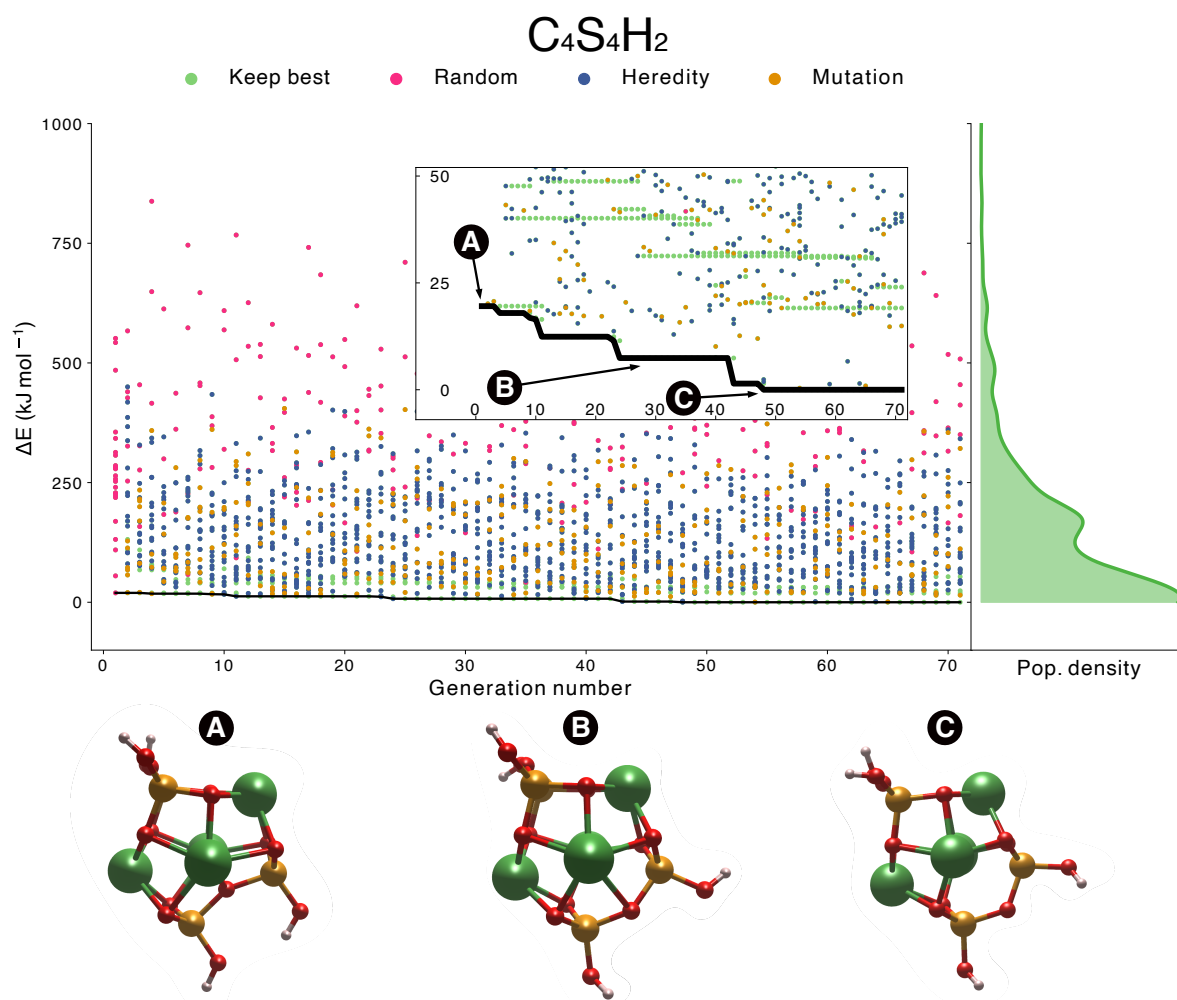


Figure S8. Energy evolution and energy population density of the Evolutionary Algorithms calculation for $C_4S_4H_2$ PPs. Each point represents a structure. Green points represent structures that have been kept from the previous generation. Pink points represent structures that have been created randomly. Blue points represent structures that have been created by using the heredity operator. Orange points represent structures that have been created using the mutation operator. The structures A, B, and C represent the best PP in generations 1, 30 and 50.

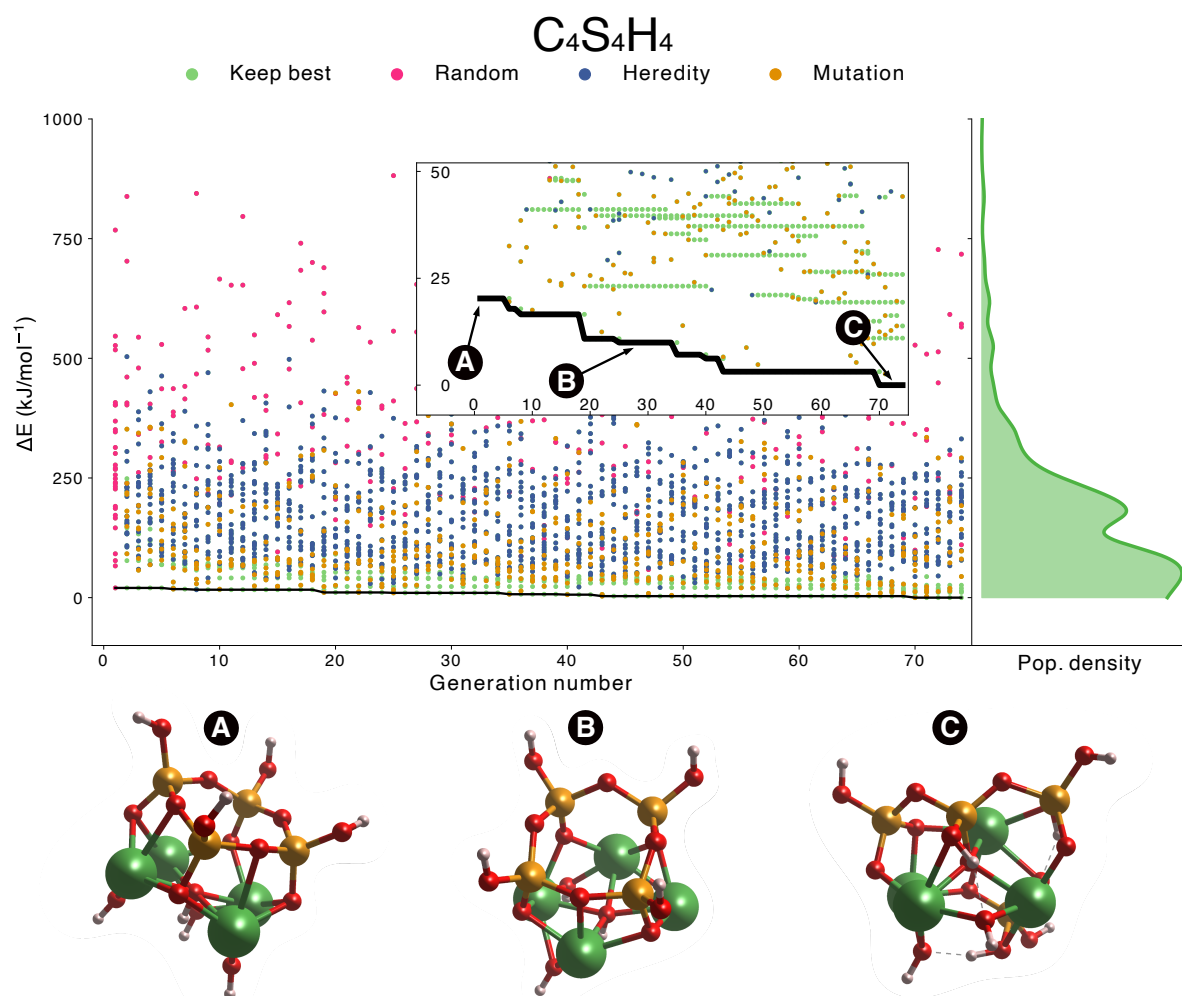


Figure S9. Energy evolution and energy population density of the Evolutionary Algorithms calculation for C₄S₄H₄ PPs. Each point represents a structure. Green points represent structures that has been kept from the previous generation. Pink points represent structures that have been created randomly. Blue points represent structures that have been created by using the heredity operator. Orange points represent structures that have been created using the mutation operator. The structures A, B, and C represent the best PP in generations 1,30 and 72.

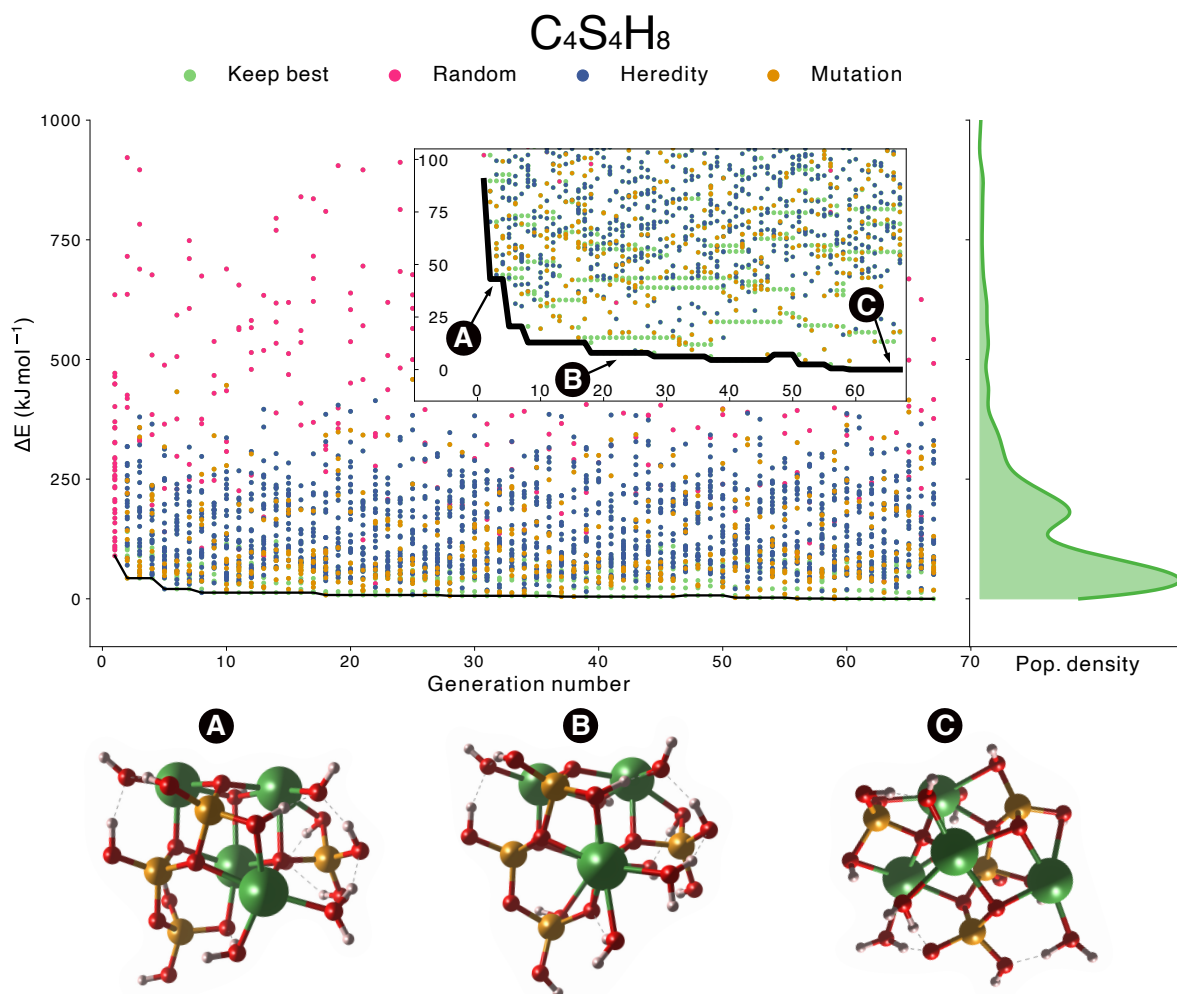


Figure S10. Energy evolution and energy population density of the Evolutionary Algorithms calculation for C₄S₄H₈ PPs. Each point represents a structure. Green points represent structures that has been kept from the previous generation. Pink points represent structures that have been created randomly. Blue points represent structures that have been created by using the heredity operator. Orange points represent structures that have been created using the mutation operator. The structures A, B, and C represent the best PP in generations 2, 25 and 65.

Sketch-maps

Figures S11-S16 show the sketch-map of following stoichiometries: CSH₂, CSH₆, C₂S₂H₄, C₂S₂H₂₀, C₄S₄H₂ and C₄S₄H₄. These figures also shown different properties of the clusters by means of different colormaps. For a deeper analysis of the results, .json files are also included as supplementary material, ready to be used with the Chemiscope online tool <http://www.chemiscope.org>. The files are named as followed:

1. CSH₂ clusters → CSH2_chemiscope.json.gz
2. CSH₆ clusters → CSH6_chemiscope.json.gz
3. CSH₁₀ clusters → CSH10_chemiscope.json.gz
4. C₂S₂H₄ clusters → C2S2H4_chemiscope.json.gz
5. C₂S₂H₁₂ clusters → C2S2H12_chemiscope.json.gz
6. C₂S₂H₂₀ clusters → C2S2H20_chemiscope.json.gz
7. C₄S₄H₂ clusters → C4S4H2_chemiscope.json.gz

8. $C_4S_4H_4$ clusters $\rightarrow$ C4S4H4_chemiscope.json.gz

9. $C_4S_4H_8$ clusters $\rightarrow$ C4S4H8_chemiscope.json.gz

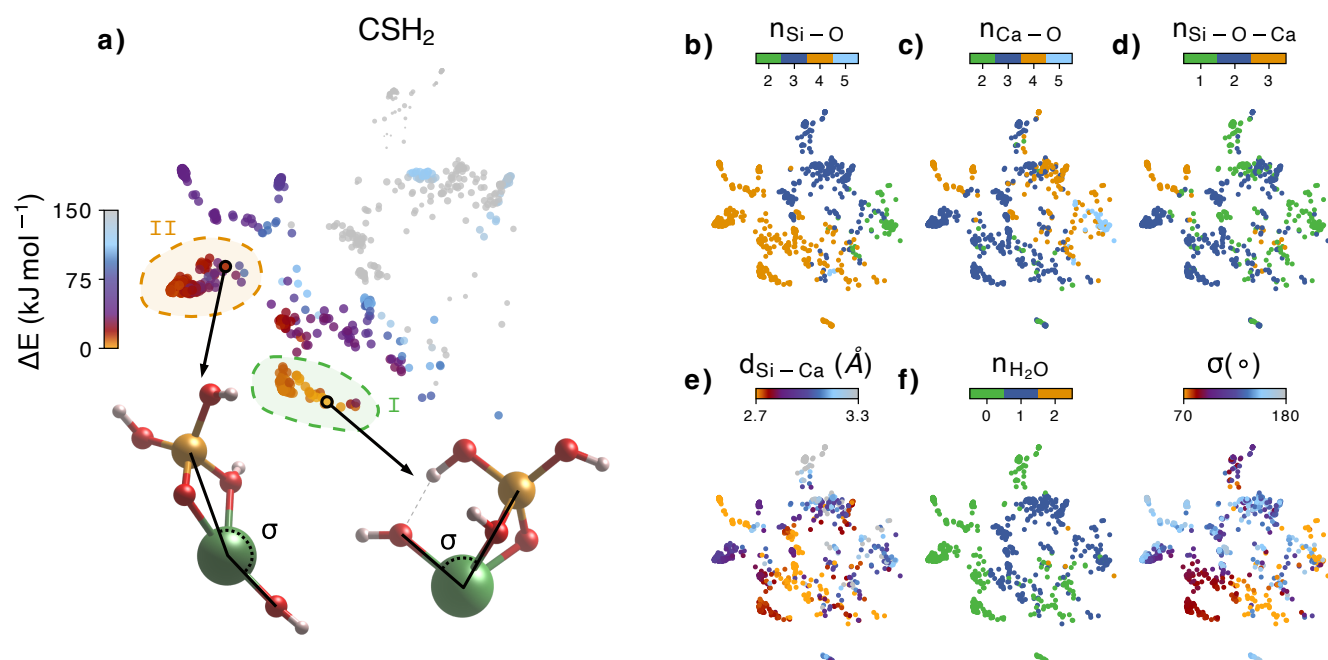


Figure S11. Sketch-map representation of the CSH_2 clusters. The colormap in (a) represents the energy difference ΔE . (b-g) Collection of structural properties of the CSH_2 clusters mapped into the same Sketch-map. n_{Si-O} refers to the number of Si-O bonds, n_{Ca-O} refers to the number of Ca-O bonds, $n_{Si-O-Ca}$ refers to the number of Si-O-Ca bonds, d_{Si-Ca} refers to the Si-Ca distance, n_{H_2O} refers to the number of water molecules and σ refers to the number of OH groups. Regions I and II each contain clusters with similar structural properties.

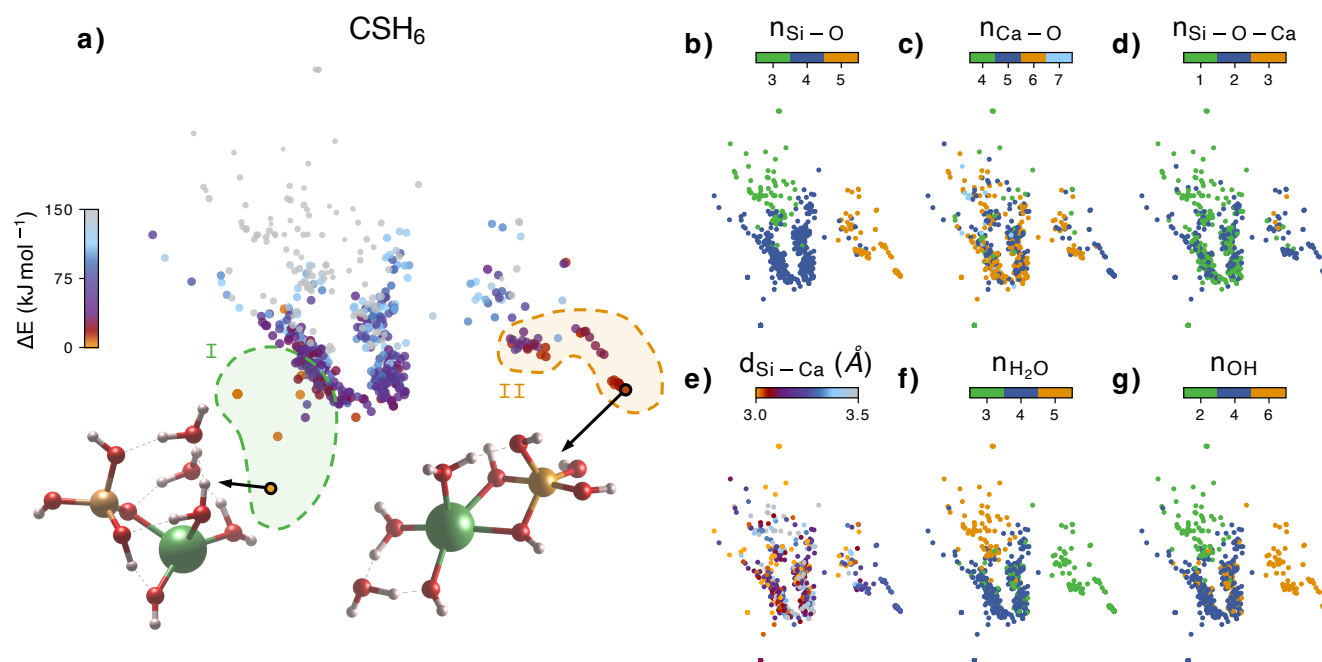


Figure S12. Sketch-map representation of the CSH_6 clusters. The colormap in (a) represents the energy difference ΔE . (b-g) Collection of structural properties of the CSH_6 clusters mapped into the same Sketch-map. $n_{\text{Si-O}}$ refers to the number of Si-O bonds, $n_{\text{Ca-O}}$ refers to the number of Ca-O bonds, $n_{\text{Si-O-Si}}$ refers to the number of Si-O-Si bonds, $n_{\text{Ca-O-Ca}}$ refers to the Ca-O-Ca bonds, $n_{\text{H}_2\text{O}}$ refers to the number of water molecules and n_{OH} refers to the number of OH groups. Regions I and II each contain one clusters with similar structural properties.

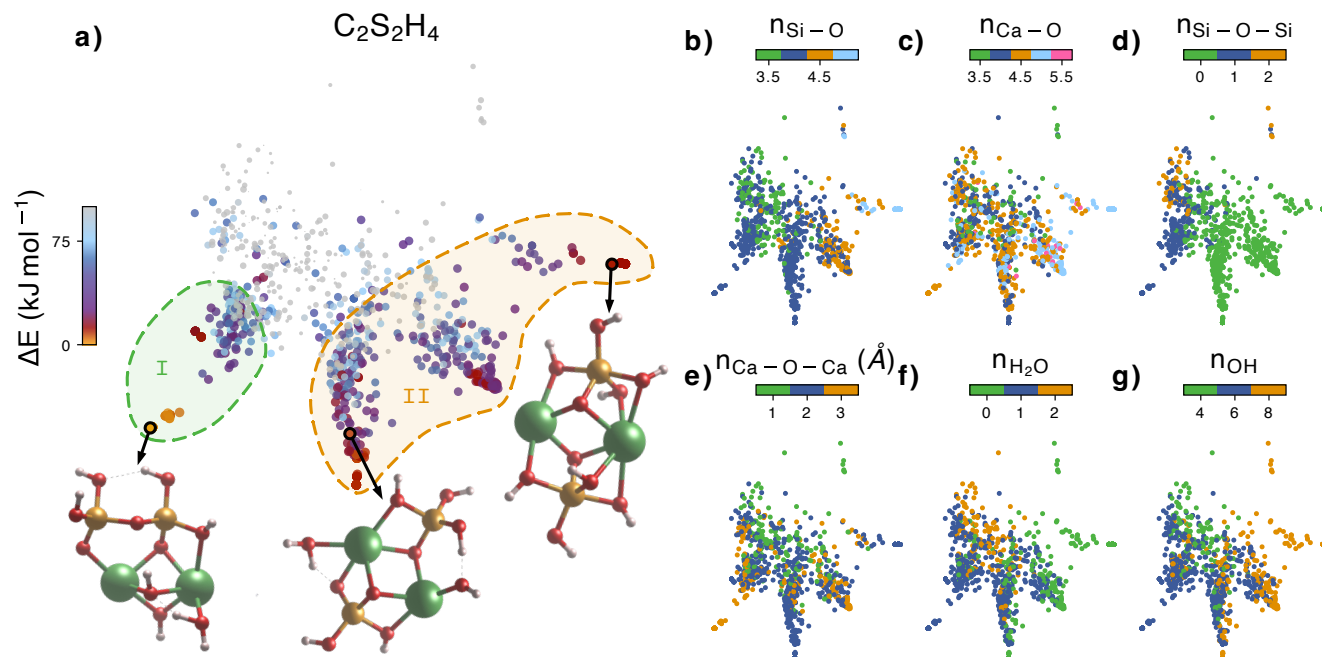


Figure S13. Sketch-map representation of the $\text{C}_2\text{S}_2\text{H}_4$ clusters. The colormap in (a) represents the energy difference ΔE . (b-g) Collection of structural properties of the $\text{C}_2\text{S}_2\text{H}_4$ clusters mapped into the same Sketch-map. $n_{\text{Si-O}}$ refers to the number of Si-O bonds, $n_{\text{Ca-O}}$ refers to the number of Ca-O bonds, $n_{\text{Si-O-Si}}$ refers to the number of Si-O-Si bonds, $n_{\text{Ca-O-Ca}}$ refers to the Ca-O-Ca bonds, $n_{\text{H}_2\text{O}}$ refers to the number of water molecules and n_{OH} refers to the number of OH groups. Regions I and II each contain clusters with similar structural properties.

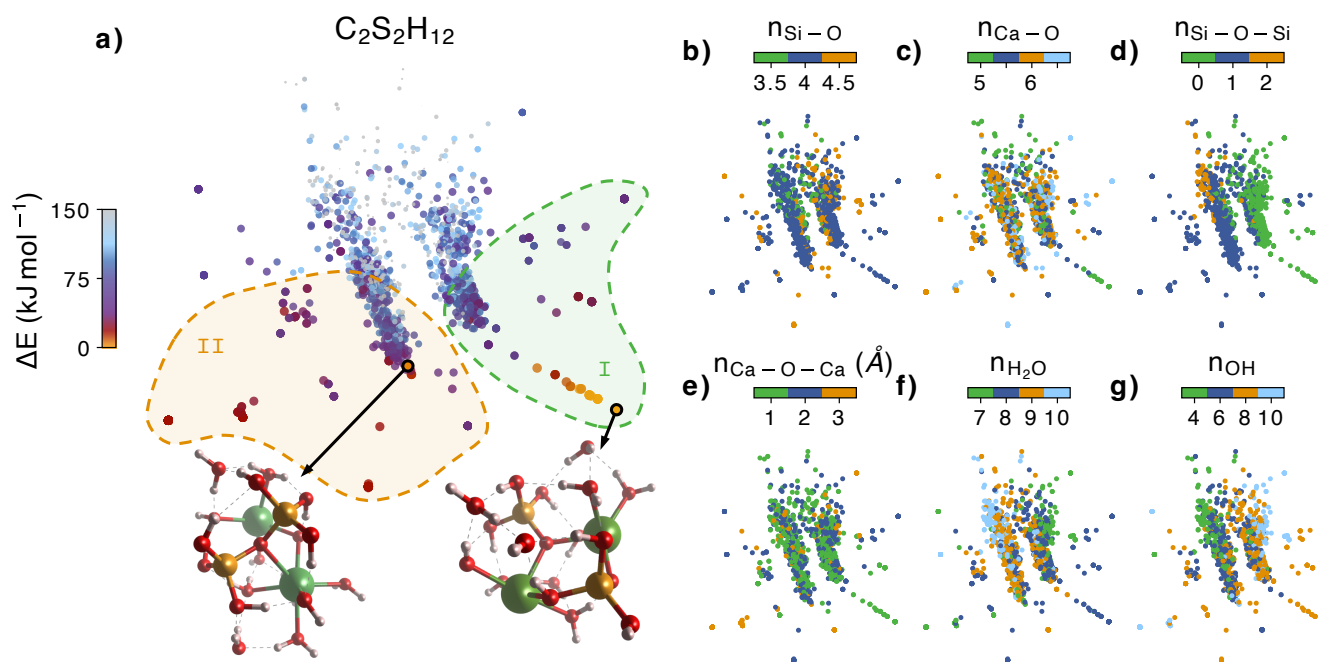


Figure S14. Sketch-map representation of the $C_2S_2H_{12}$ clusters. The colormap in (a) represents the energy difference ΔE . (b-g) Collection of structural properties of the $C_2S_2H_{12}$ clusters mapped into the same Sketch-map. n_{Si-O} refers to the number of Si-O bonds, n_{Ca-O} refers to the number of Ca-O bonds, $n_{Si-O-Si}$ refers to the number of Si-O-Si bonds, $n_{Ca-O-Ca}$ refers to the Ca-O-Ca bonds, n_{H_2O} refers to the number of water molecules and n_{OH} refers to the number of OH groups. Regions I and II each contain clusters with similar structural properties.

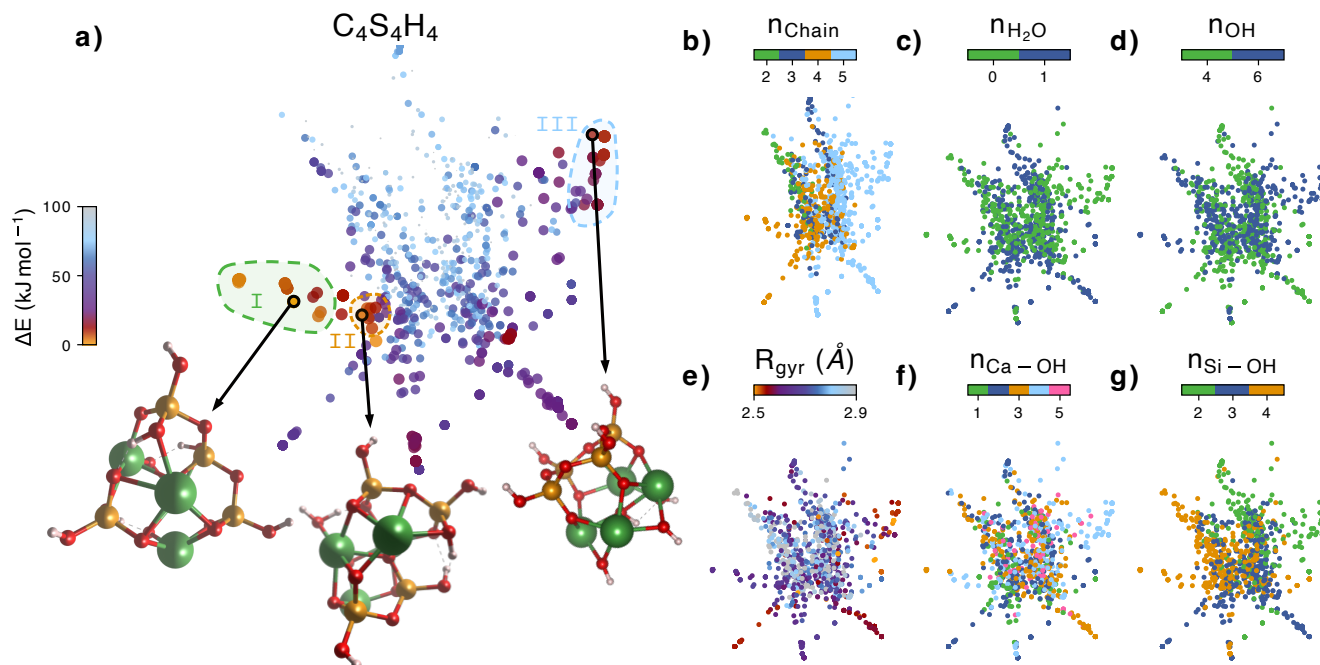


Figure S15. Sketch-map representation of the $C_4S_4H_4$ clusters. The colormap in (a) represents the energy difference ΔE . (b-g) Collection of structural properties of the $C_4S_4H_4$ clusters mapped into the same Sketch-map. n_{chain} refers to the different silicate chain arrangements defined in Fig. 5 of the main manuscript, n_{H_2O} refers to the number of water molecules, n_{OH} refers to the number of OH groups, R_{gyr} refers to the gyration radius of the clusters, n_{Ca-OH} refers to the number of Ca-OH bonds and n_{Si-OH} refers to the number of Si-OH bonds. Regions I, II and III each contain clusters with similar structural properties.

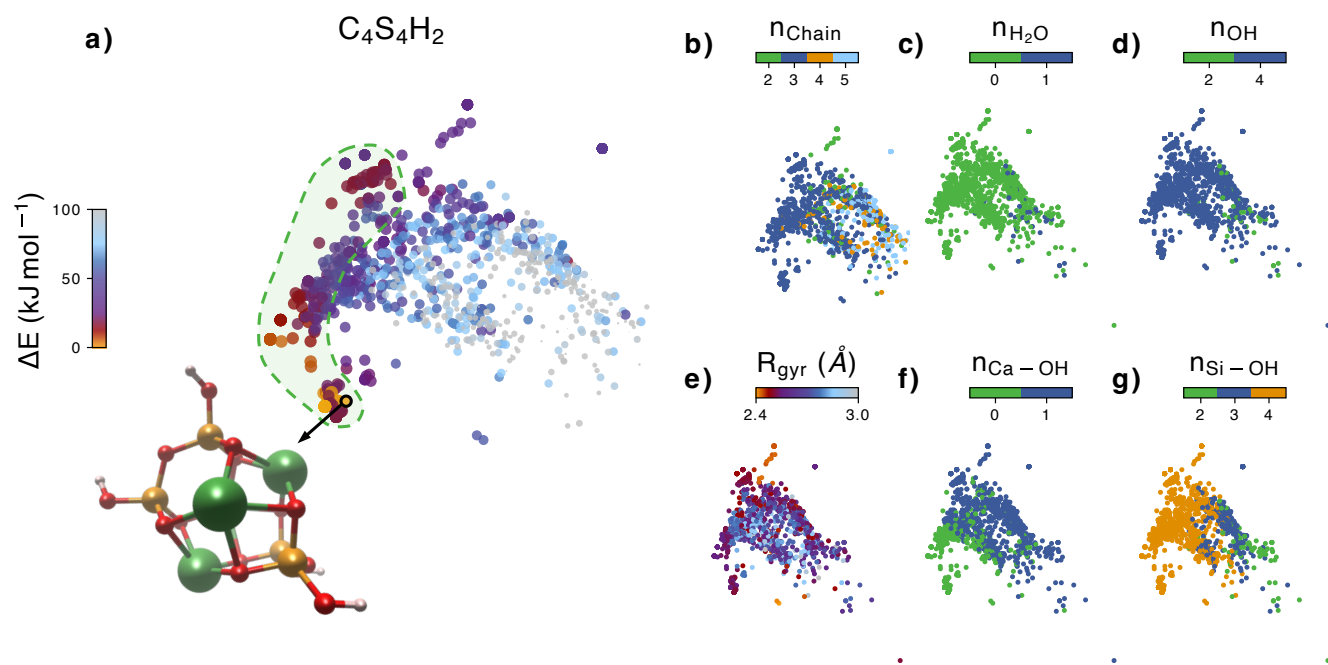


Figure S16. Sketch-map representation of the $C_4S_4H_2$ clusters. The colormap in (a) represents the energy difference ΔE . (b-g) Collection of structural properties of the $C_4S_4H_2$ clusters mapped into the same Sketch-map. n_{chain} refers to the different silicate chain arrangements defined in Fig. 5 of the main manuscript, n_{H_2O} refers to the number of water molecules, n_{OH} refers to the number of OH groups, R_{gyr} refers to the gyration radius of the clusters, n_{Ca-OH} refers to the number of Ca-OH bonds and n_{Si-OH} refers to the number of Si-OH bonds.

Free energies of aggregation

Product	Reagents	ΔG (KJ/mol)
CSH ₁₀ -I	CaO ₂ H ₃ ⁺ + SiO ₅ H ₅ ⁻ + 6H ₂ O	-91.07
	CaO ₆ H ₁₂ ⁺² + SiO ₉ H ₁₂ ⁻² - 2H ₂ O	-156,75
C ₂ S ₂ H ₁₂ -I	2CaO ₂ H ₃ ⁺ + 2SiO ₅ H ₅ ⁻ + 4H ₂ O	-124.13
	2CaO ₆ H ₁₂ ⁺² + 2SiO ₉ H ₁₂ ⁻² - 12H ₂ O	-189.81
	2CSH ₂₀ -I - 8H ₂ O	-66.11
C ₂ S ₂ H ₁₂ -II	2CaO ₂ H ₃ ⁺ + 2SiO ₅ H ₅ ⁻ + 4H ₂ O	-111.97
	2CaO ₆ H ₁₂ ⁺² + 2SiO ₉ H ₁₂ ⁻² - 12H ₂ O	-177.65
	2CSH ₁₀ -I - 8H ₂ O	-41.80
C ₂ S ₂ H ₂₀ -I	2CaO ₂ H ₃ ⁺ + 2SiO ₅ H ₅ ⁻ + 12H ₂ O	-127.55
	2CaO ₆ H ₁₂ ⁺² + 2SiO ₉ H ₁₂ ⁻² - 4H ₂ O	-193.23
	2 CSH ₁₀	-36.48
C ₂ S ₂ H ₂₀ -II	2CaO ₂ H ₃ ⁺ + 2SiO ₅ H ₅ ⁻ + 12H ₂ O	-134.23
	2CaO ₆ H ₁₂ ⁺² + 2SiO ₉ H ₁₂ ⁻² - 4H ₂ O	-199.90
	2 CSH ₁₀	-43.15
C ₄ S ₄ H ₂	4CaO ₂ H ₃ ⁺ + 4SiO ₅ H ₅ ⁻ - 14H ₂ O	-140,36
	4CaO ₆ H ₁₂ ⁺² + 4SiO ₉ H ₁₂ ⁻² - 46H ₂ O	-206.04
	4CSH ₁₀ -I - 38H ₂ O	-49.29
	2CSH ₁₀ -I + C ₂ S ₂ H ₂₀ -I - 38 H ₂ O	-31.05
	2C ₂ S ₂ H ₂₀ -I - 38H ₂ O	-12.81
C ₄ S ₄ H ₄	4CaO ₂ H ₃ ⁺ + 4SiO ₅ H ₅ ⁻ - 12H ₂ O	-132,01
	4CaO ₆ H ₁₂ ⁺² + 4SiO ₉ H ₁₂ ⁻² - 44H ₂ O	-197.69
	4CSH ₁₀ -I - 36H ₂ O	-40.94
	2CSH ₁₀ -I + C ₂ S ₂ H ₂₀ -I - 36 H ₂ O	-22.70
	2C ₂ S ₂ H ₂₀ -I - 36H ₂ O	-4.46
C ₄ S ₄ H ₈	4CaO ₂ H ₃ ⁺ + 4SiO ₅ H ₅ ⁻ - 8H ₂ O	-132,74
	4CaO ₆ H ₁₂ ⁺² + 4SiO ₉ H ₁₂ ⁻² - 40H ₂ O	-198.42
	4CSH ₁₀ -I - 16H ₂ O	-8.61
	2CSH ₁₀ -I + C ₂ S ₂ H ₂₀ -I - 32 H ₂ O	-23.42
	2C ₂ S ₂ H ₂₀ -I - 32H ₂ O	-5.19

Table S1. Formation free energies.

Additional data of cluster aggregation

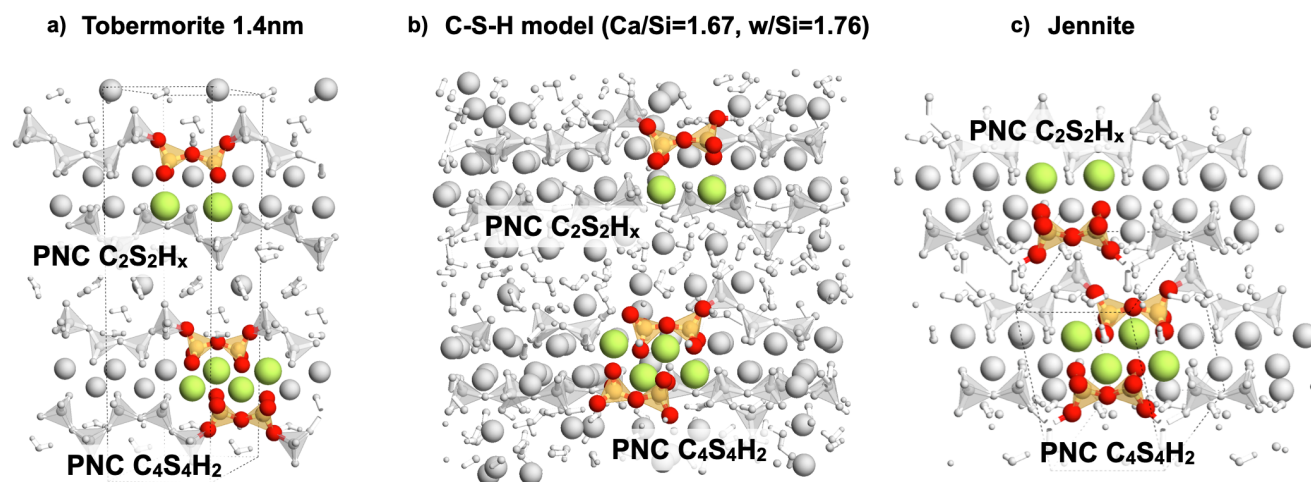


Figure S17. Representation of the atomic structure of (a) tobermorite 14, (b) C-S-H model and (c) Jennite together with the $C_2S_2H_x$ and $C_4S_4H_2$ PPs that act as basic building blocks.

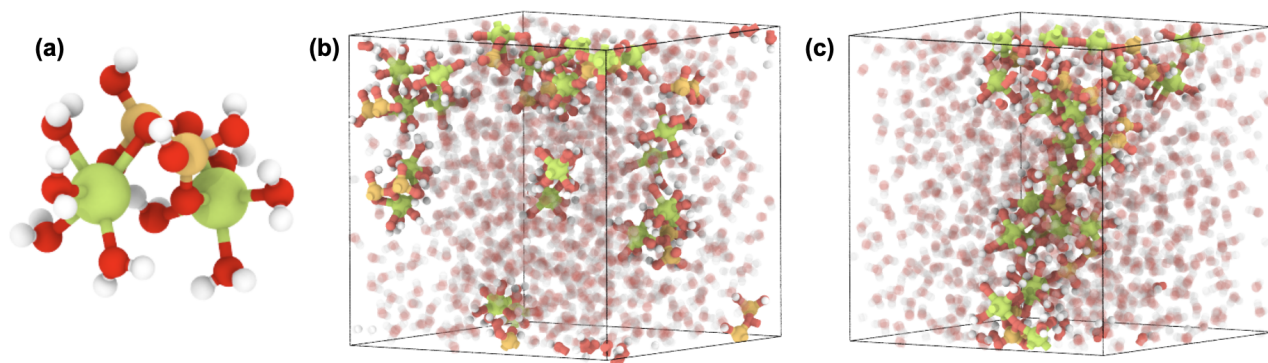


Figure S18. (a) Image of the clusters used in the MD simulation, (b) the initial random configuration of one of the simulations, (c) and the final configuration after the aggregation (15ns).

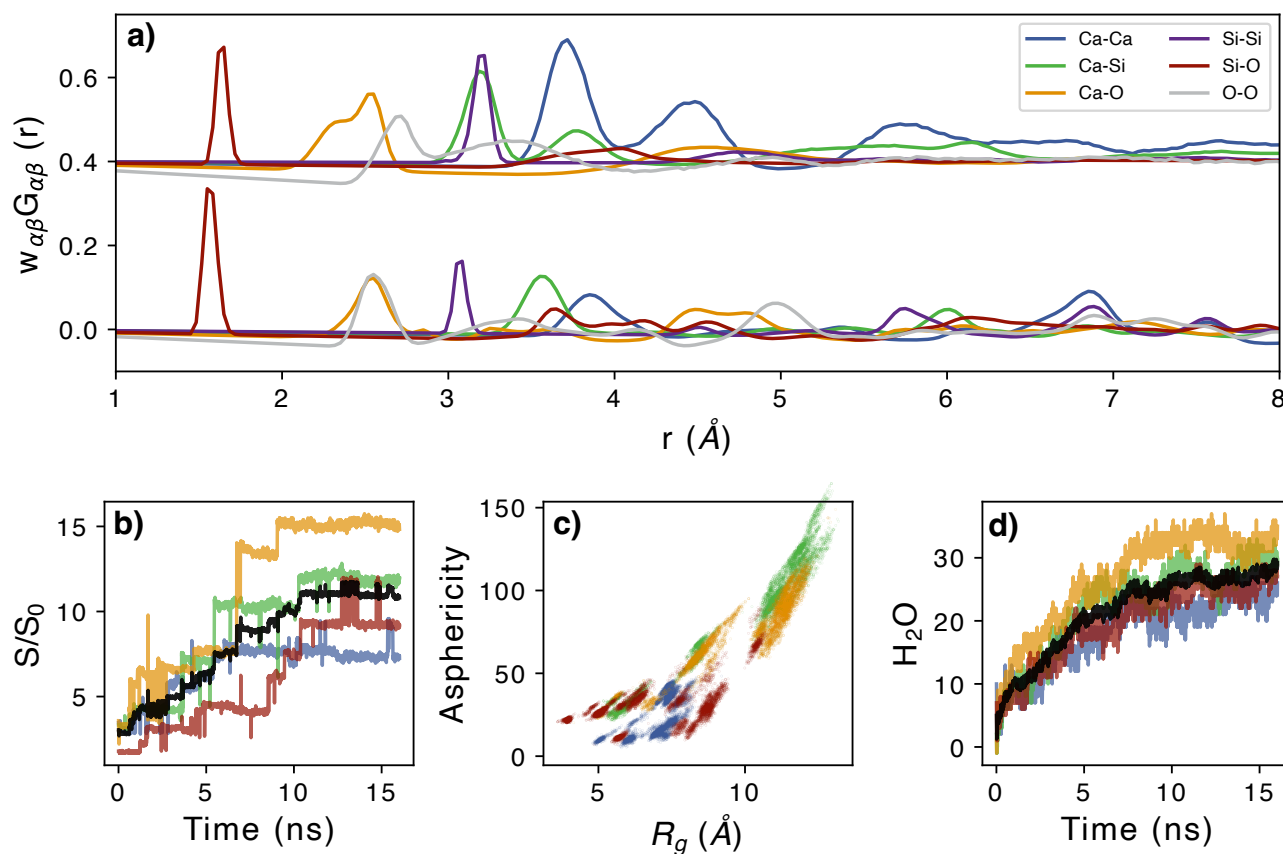


Figure S19. (a) Weighted partial pair distribution functions $w_{\alpha\beta}G_{\alpha\beta}$ for the aggregate (top) and tobermorite 14 (bottom). α and β stand for each of the atomic species, and $w_{\alpha\beta}$ is the weight of the partial distribution $G_{\alpha\beta}$ in the total one. This analysis has been helpful to understand the contribution of each atomic pair to the total $G(r)$ and identify the peaks. (b) Evolution of the aggregate size S over time for each MD simulation (average in black). The size is defined in terms of number of atoms that for the aggregate normalised by the number of atoms in a primary particle S_0 . (c) Asphericity of the aggregate as a function of the gyration radius R_g for each MD simulation. (d) Water loss $\langle H_2O \rangle$ from the cluster over time for each MD simulation.

Input Files for the simulation codes

Sample inputs are given in the following files:

- `USPEX_input.txt` : USPEX input to perform EA. The population size is set to 25 structures and each generation the 60% of structures are used to generate the following generation. From those, 60% are created by the heredity operator, 20% by random generation, 10% by permutation and the remaining 10% by atoms mutation. A 15 minimum distance is set for
- `siesta_input.fdf` : SIESTA input for the relaxation of the clusters with DFT. GGA-PBE was used as the exchange-correlation functional and double- ζ polarized (DZP) basis set was used for the pseudo atomic orbitals. The Grimme dispersion correction was used to include the van der Waals forces. The convergence criterion for the self-consistent cycle (SCF) was set to 10^{-4} Å. The conjugate gradient (CG) method was used to minimize the positions of the atoms until the maximum force was less than 0.04 eV/Å.
- `gulp_input.in` : Input to minimize the structures using the GULP code with the ReaxFF force-field. The ReaxFF parameters are available for download from https://github.com/hegoimanzano/ReaxFF_cement⁴ in formatted files ready to use with GULP or LAMMPS. The energy was minimized using the NR/BFGS minimiser using a maximum of 50000 steps.
- `gaussian_input.com` : Gaussian input to compute the free energy of the selected clusters using the Polarizable Continuum

Model (PCM). GGA-PBE was used as the exchange-correlation functional using 6-311+G(d) basis set which includes the Grimme dispersion. The dielectric constant value for the PCM was set to $\epsilon = 56$.

- `lammps_input.in` : LAMMPS input file for the molecular dynamics simulations with ReaxFF. The charge equilibration method was used, with a cut-off of 10 Å and a convergence criterion of 10^{-5} . Three simulation stages are done: an energy minimization with conjugate gradient as, a 200 ps for density equilibration in the NPT ensemble, and 6 ns in the NVT ensemble for production. The time step was set to 0.2 fs using the Verlet integrator. The Nose-Hoover thermostat and barostat damping constants were set to 20 and 200 fs respectively.

Supplementary references

1. Eikenberg, J. On the problem of silica solubility at high ph. Tech. Rep., Paul Scherrer Inst.(PSI) (1990).
2. Kutus, B. *et al.* A comprehensive study on the dominant formation of the dissolved $\text{Ca}(\text{OH})_2(\text{aq})$ in strongly alkaline solutions saturated by $\text{Ca}(\text{II})$. *RSC advances* **6**, 45231–45240 (2016).
3. Šefčík, J. & McCormick, A. V. Thermochemistry of aqueous silicate solution precursors to ceramics. *AIChE J.* **43**, 2773–2784 (1997).
4. Manzano, H. `hegoimanzano/reaxff_cement`: v01, DOI: [10.5281/zenodo.8379302](https://doi.org/10.5281/zenodo.8379302) (2023).