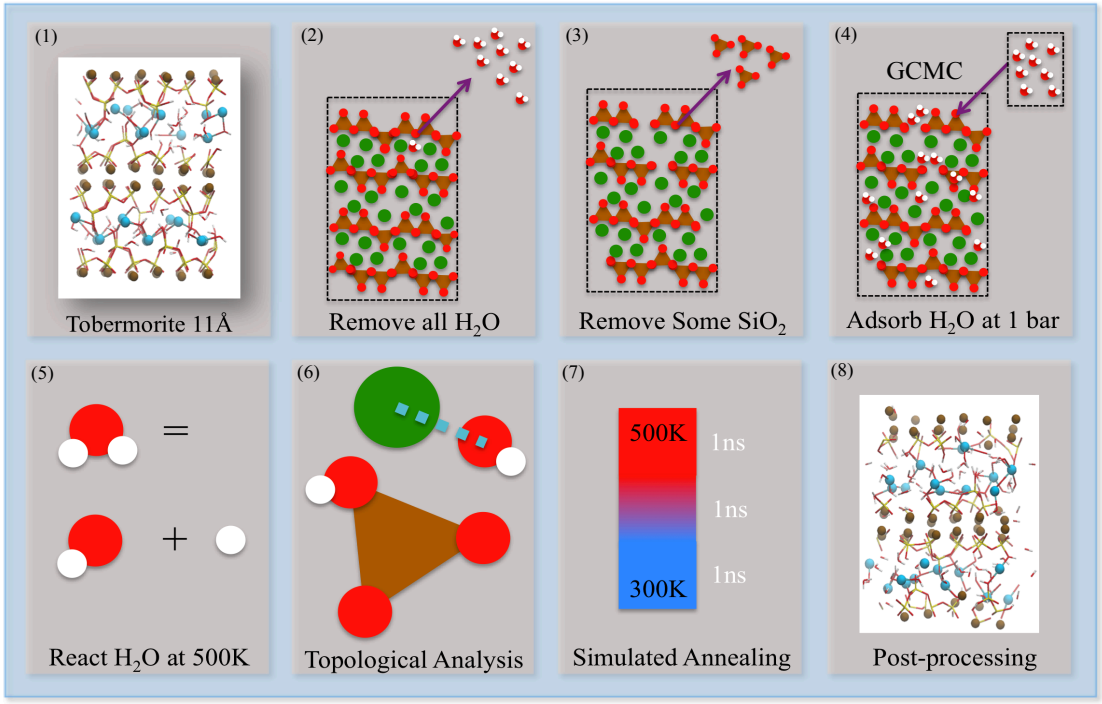
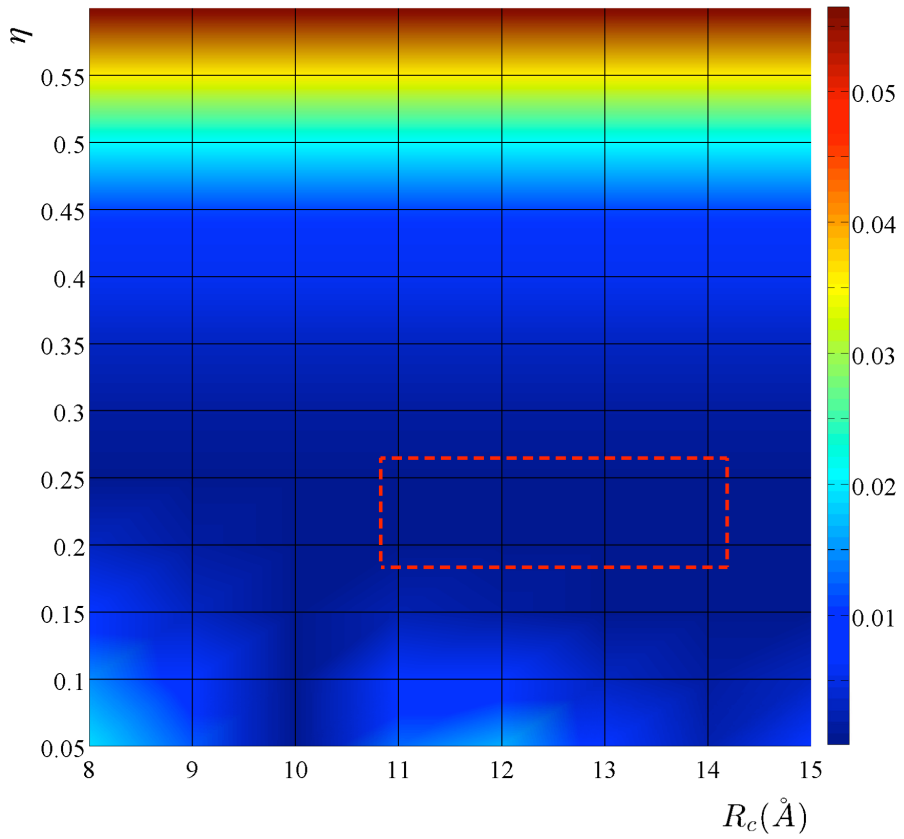
Supplementary Information

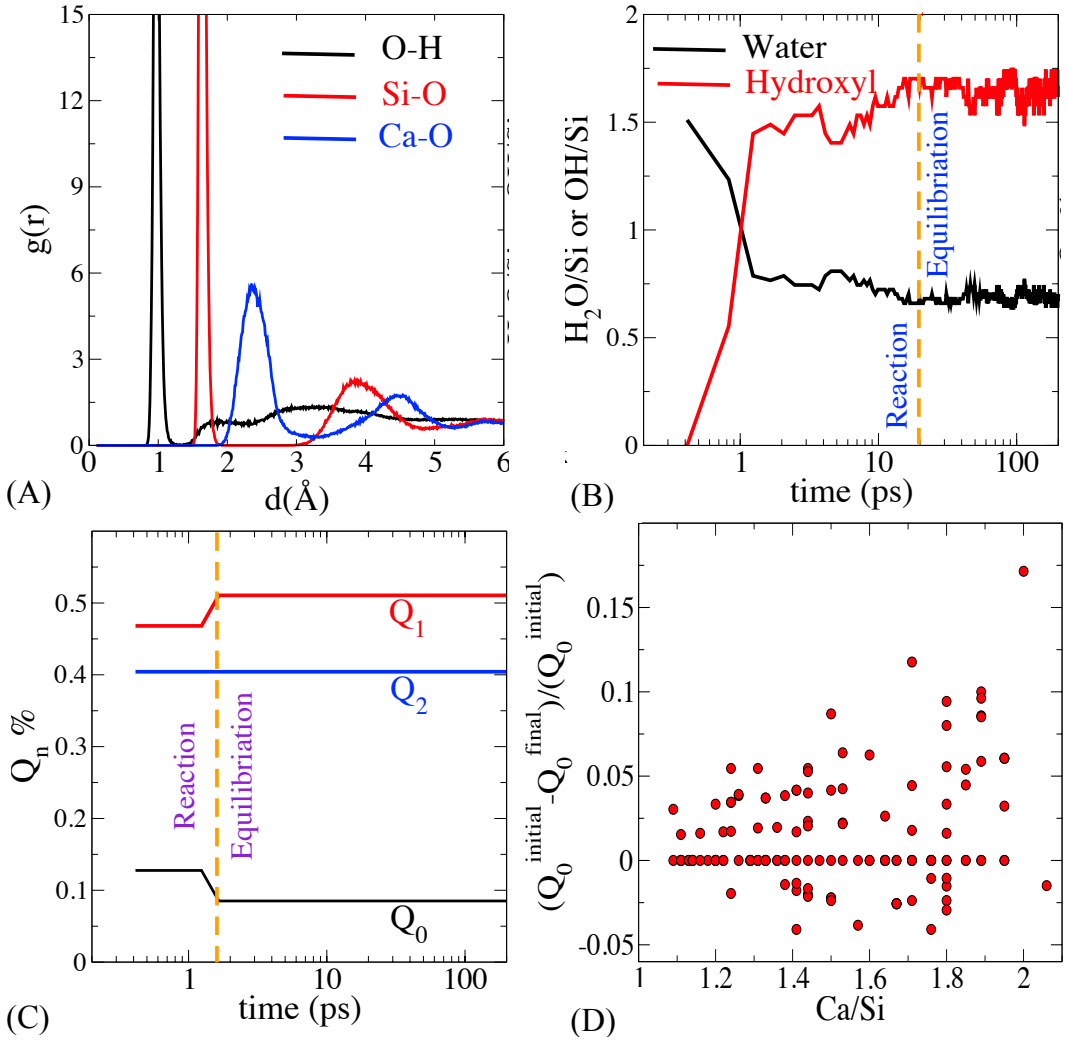
Abdolhosseini Qomi et al., "Combinatorial Molecular Optimization of Cement Hydrates".



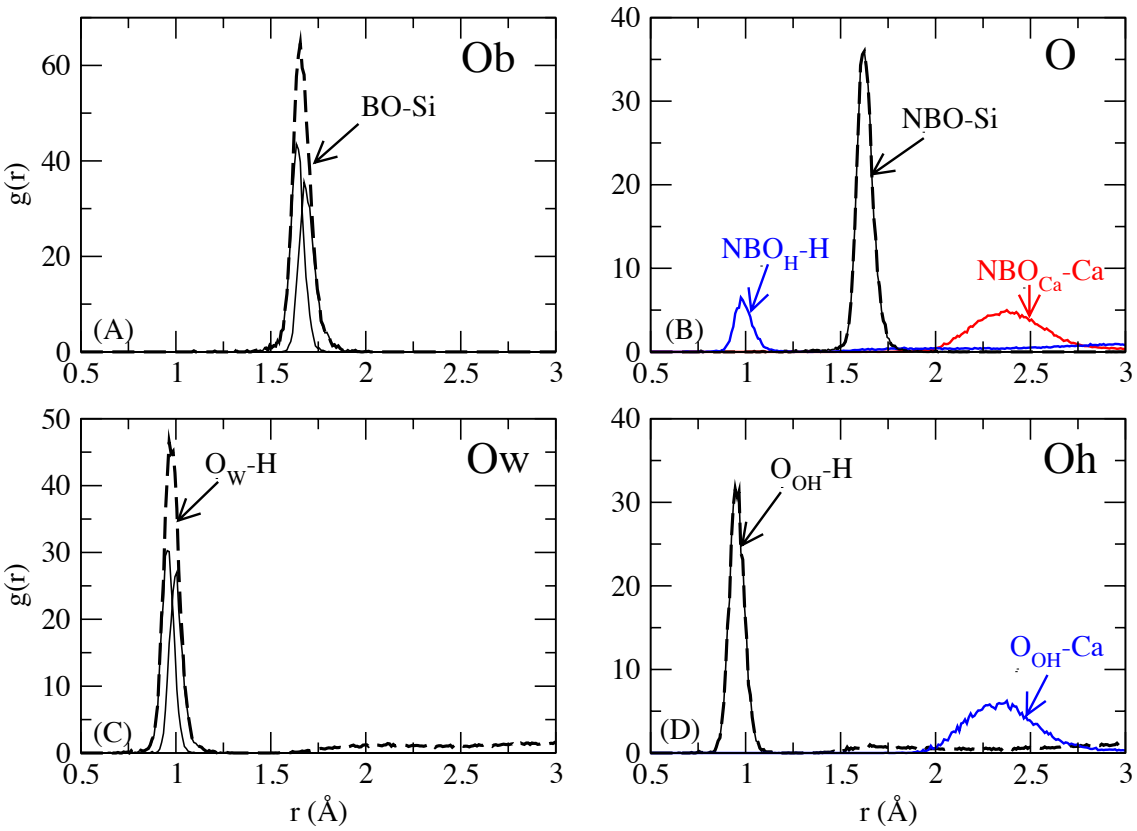
Supplementary Figure 1: The eight stages of model construction. These stages are strictly followed 150 times to produce C-S-H samples with varying Ca/Si ratio ranging from 1.1 to 2.1.



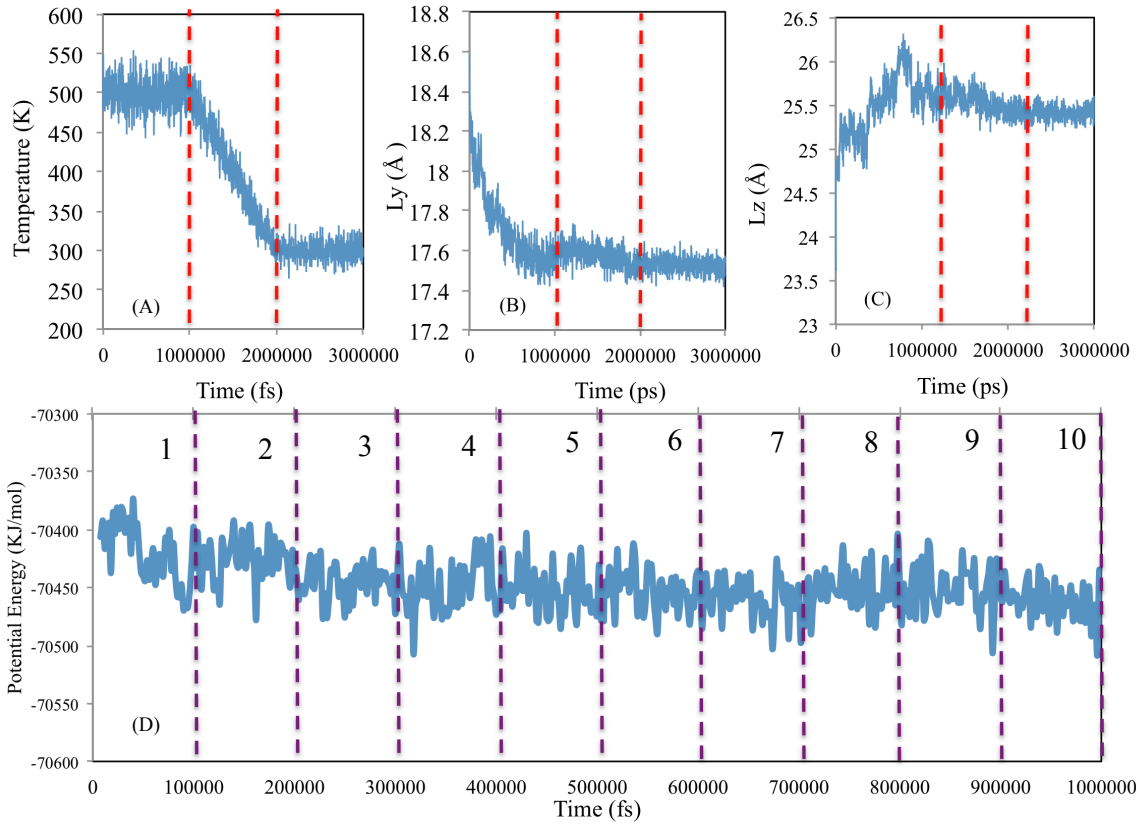
Supplementary Figure 2: Distribution of in R_c- space used for identification of optimal parameters of Wolf potential for C-S-H samples. The dashed rectangles located the region in which the normalized error is less than 10^{-4}. R_c and are set to 12Å and 0.25 in all simulations, respectively.



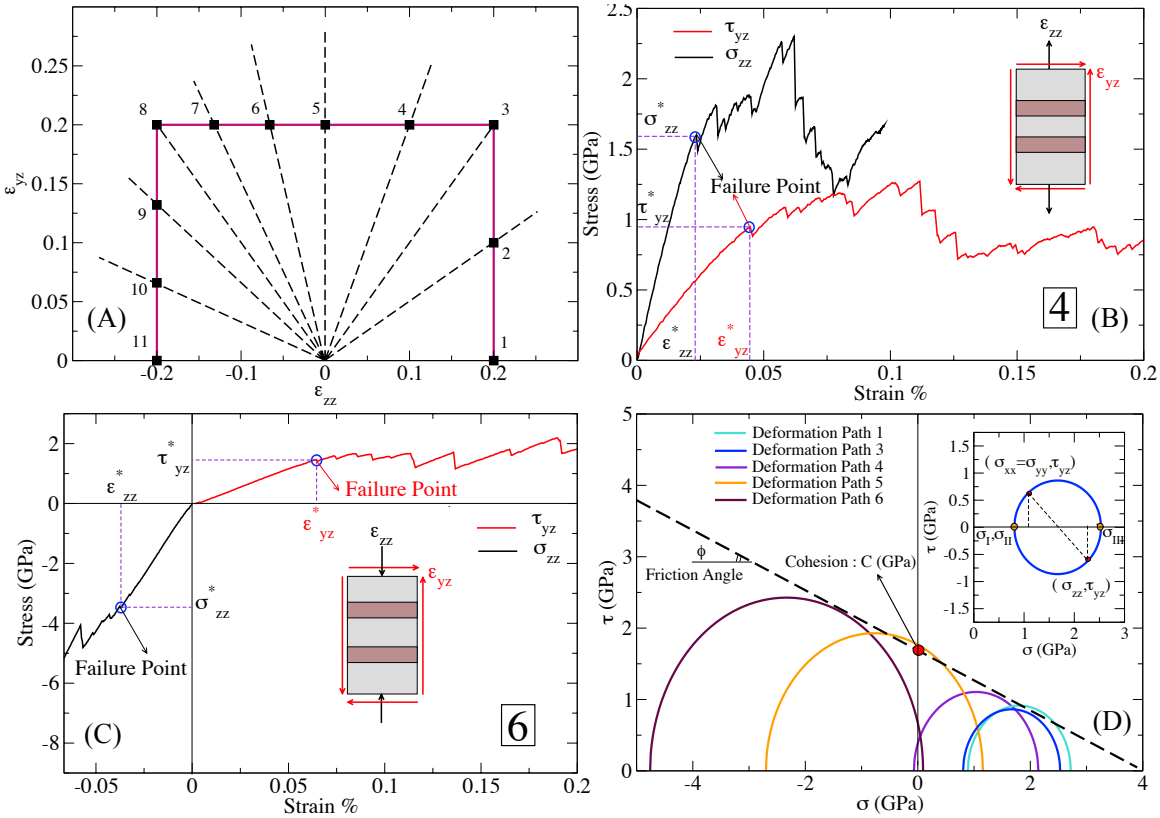
Supplementary Figure 3: Reactive force field modeling in canonical ensemble. (A) Radial distribution function for Si-O, Ca-O and O-H bonds. (B) Dissociation of water molecules in C-S-H and production of hydroxyl groups. (C) Condensation of silica groups in the silicate chain. (D) Condensation of silicate chains as a function of Ca/Si ratio.



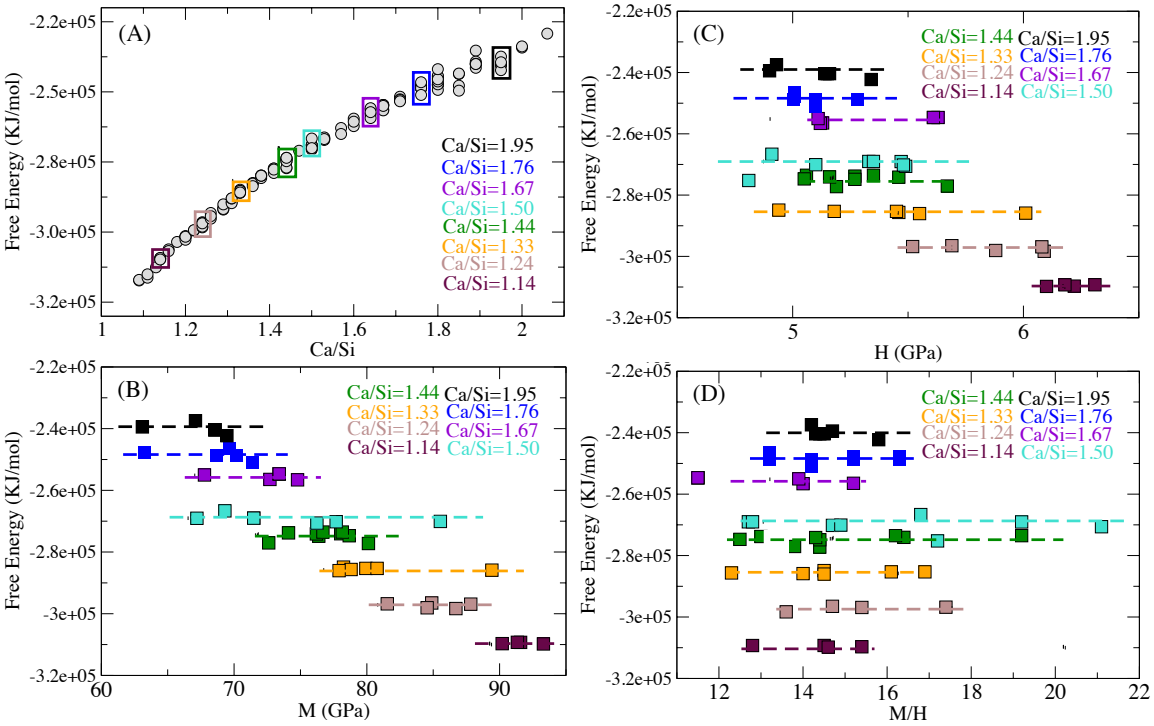
Supplementary Figure 4: Determination of the local environment of oxygen. Radial distribution functions of Si, Ca, and H cations around Ob (A), O (B), Ow (C) and Oh (D). Figures (A) and (B) show the contributions of the two nearest Si neighbors and figures (C) and (D) the contributions of the two nearest H neighbors.



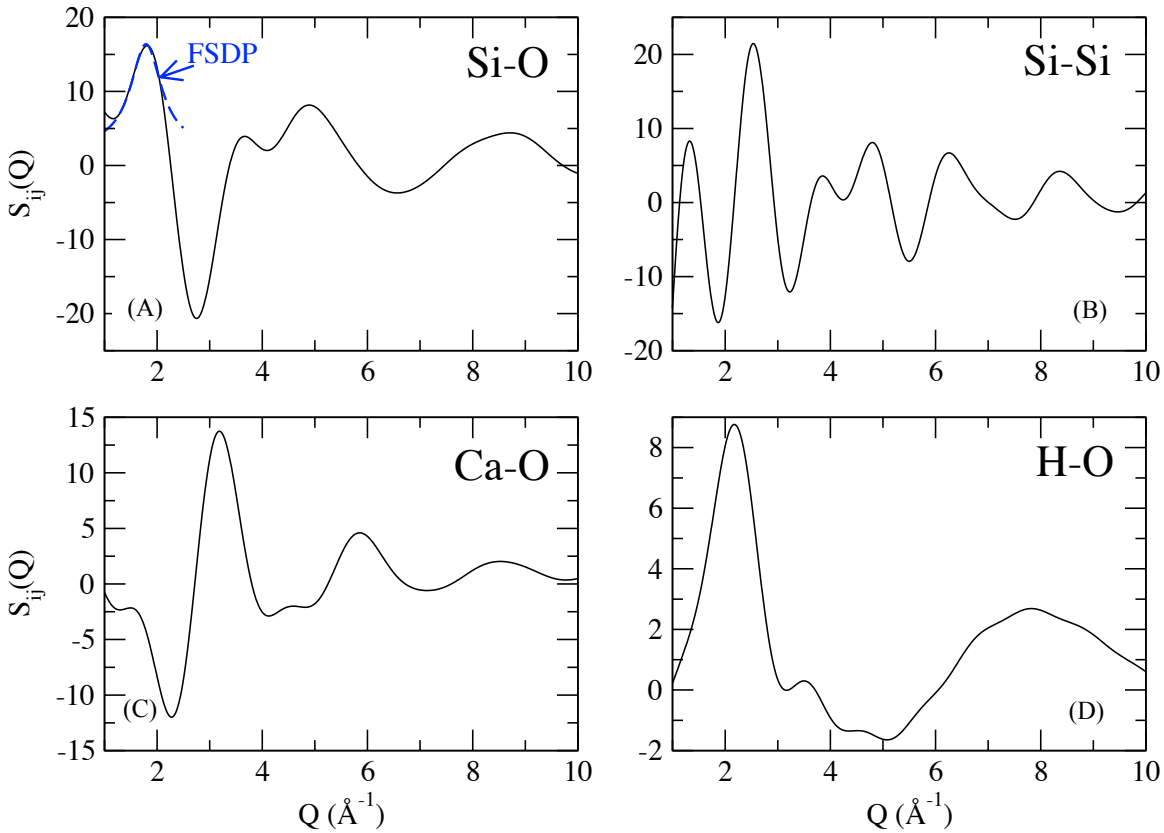
Supplementary Figure 5: Simulated annealing and sampling process from the energy landscape. (A) The temperature profile in annealing process. (B) and (C) the simulation box size during the quenching period. (D) Sampling from the energy landscape for estimation of mechanical properties, where M was calculated for 150 conformations within each of the 10 portions of this equilibration trajectory, to produce 10 points for each polymorph structure. M in Figs. 3 and 4 show standard deviation among those 10 points for each polymorph.



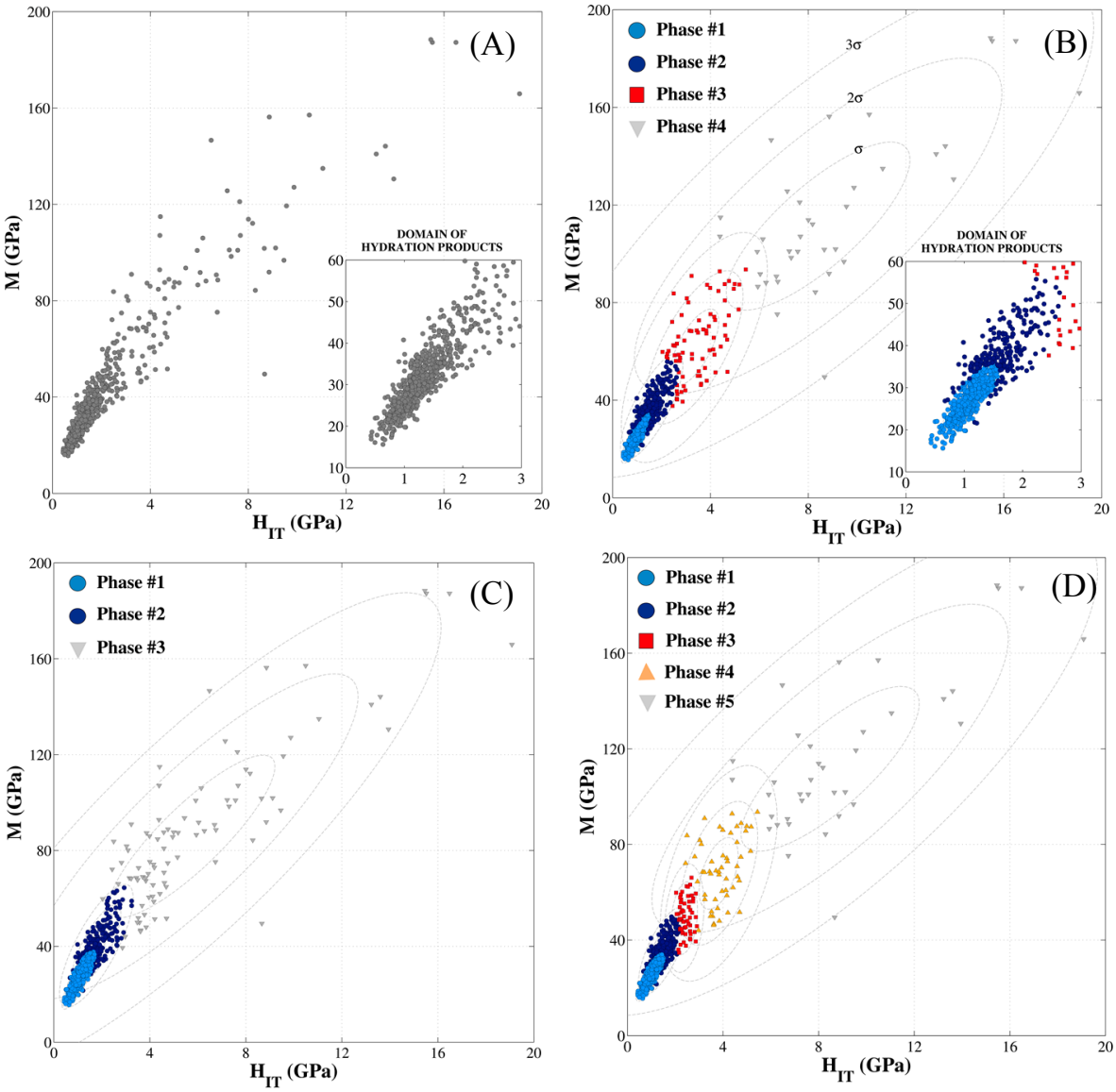
Supplementary Figure 6: calculation of hardness in atomistic simulation**.** (A) 11 independent deformation paths to deform C-S-H sample. Biaxial strain-control deformation analysis of C-S-H in (B) deformation path #4 and (C) deformation path #6. (D) Calculation of cohesion and friction angle from the linear elastic instability point using Mohr-Coulomb criterion. The circled point indicates stress at which deviation from a linear elastic stress-strain response is identified; We call this point instability or failure point.



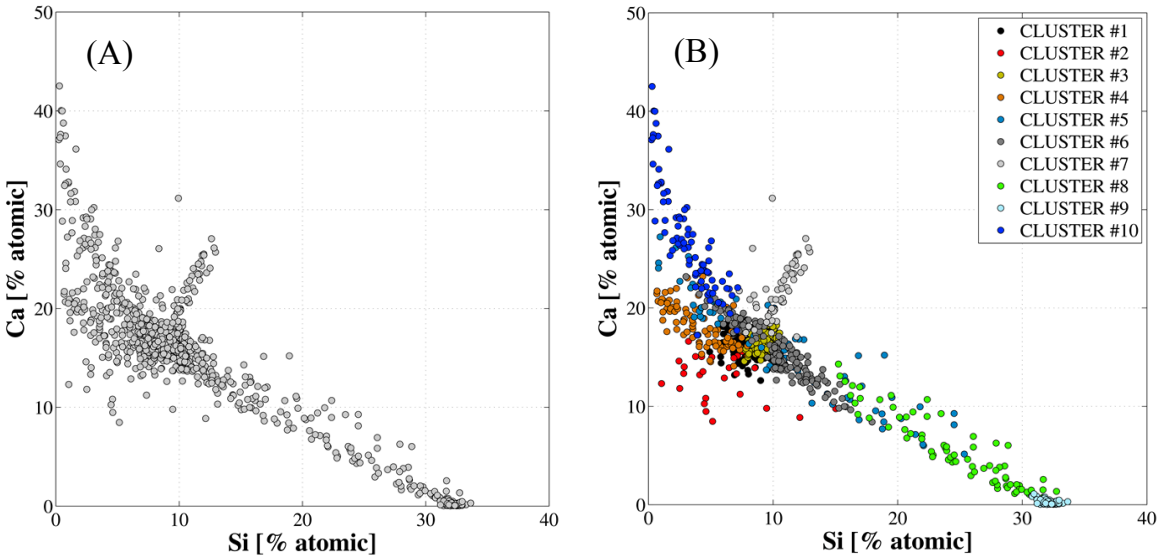
Supplementary Figure 7: Polymorphism in C-S-H. (A) Free energy as function of Ca/Si ratio. Free Energy as a function of B) and M C) H and D) M/H.



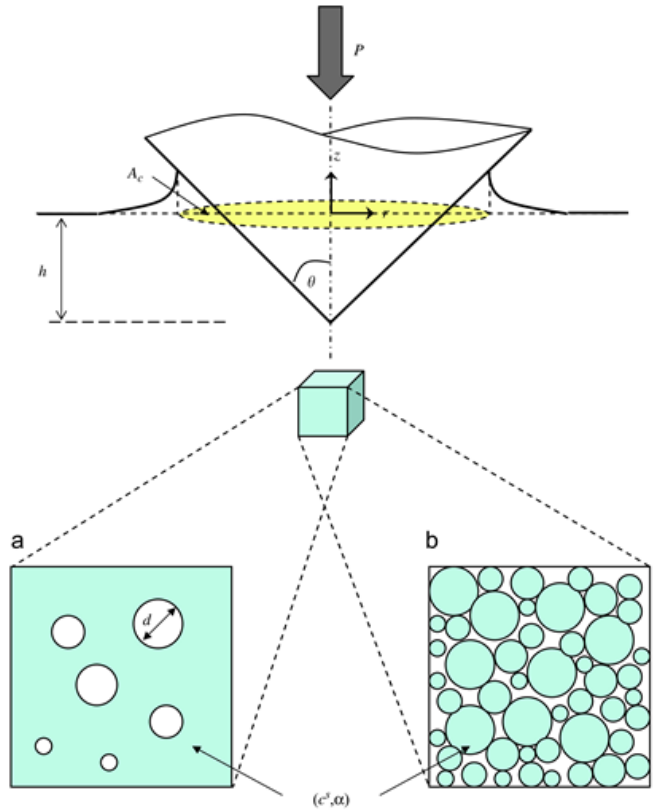
Supplementary Figure 8: Partial Structure factors in C-S-H. Computed (A) Si-O, (B) Si-Si, (C) Ca-O and (D) H-O structure factors at Ca/Si=1.7. The blue broken line shows a fit of the first sharp diffraction peak with a Lorentzian function.



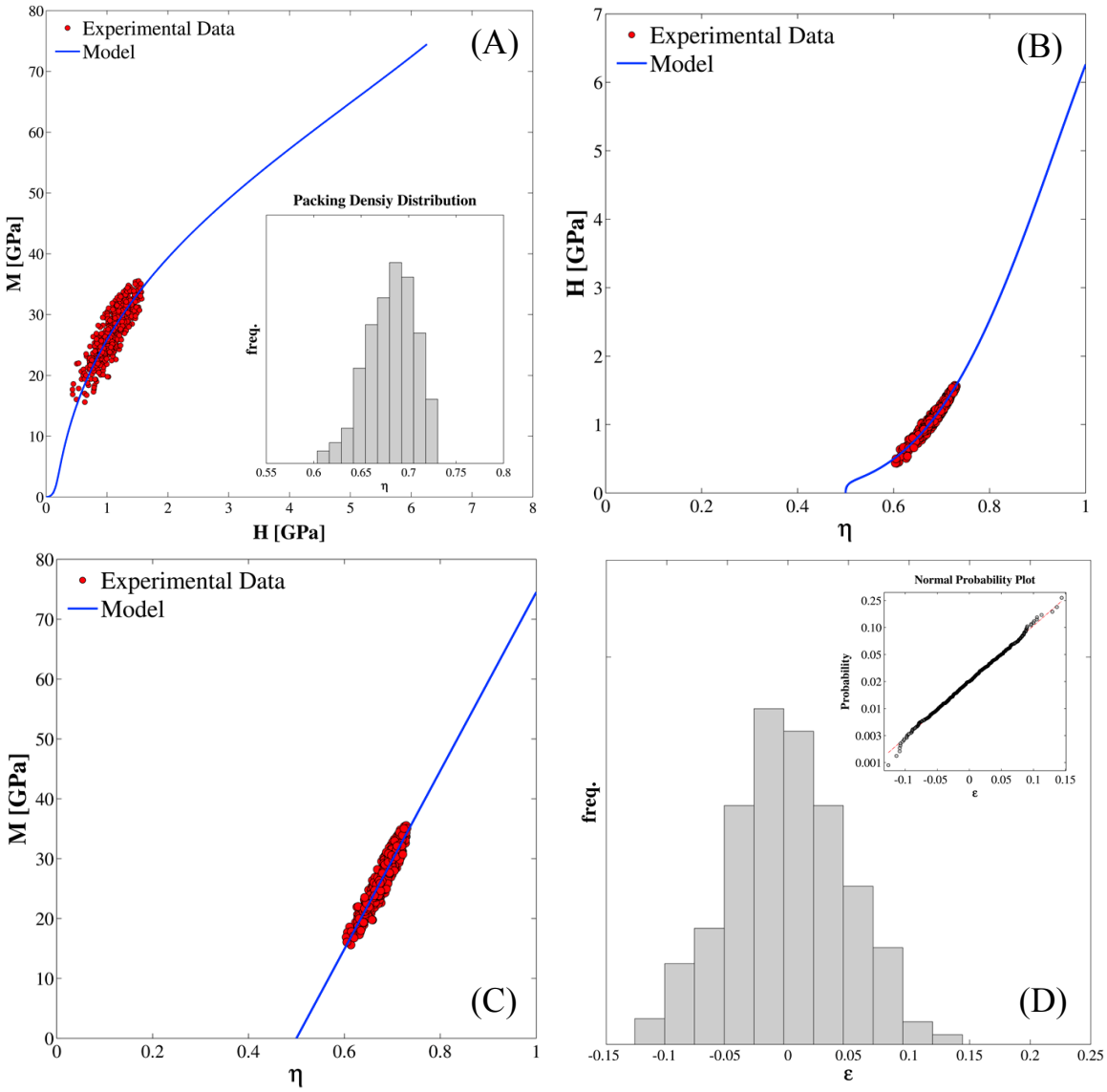
Supplementary Figure 9: Maximum Likelihood based deconvolution of the grid indentation data on OPC. (A) global plot of the experimental records, with a focus on the expected domain of hydration products, (B) clustering of the experimental data into four component Gaussian Mixture Model corresponding to maximized penalized likelihood BIC statistics, (C-D) clustering of the experimental data with respect to the model with ±1 number of optimum mechanically active phases. H-indentation hardness, M-indentation modulus, adapted from the Dr. Matthieu Vandamme PhD thesis (MIT 2008).



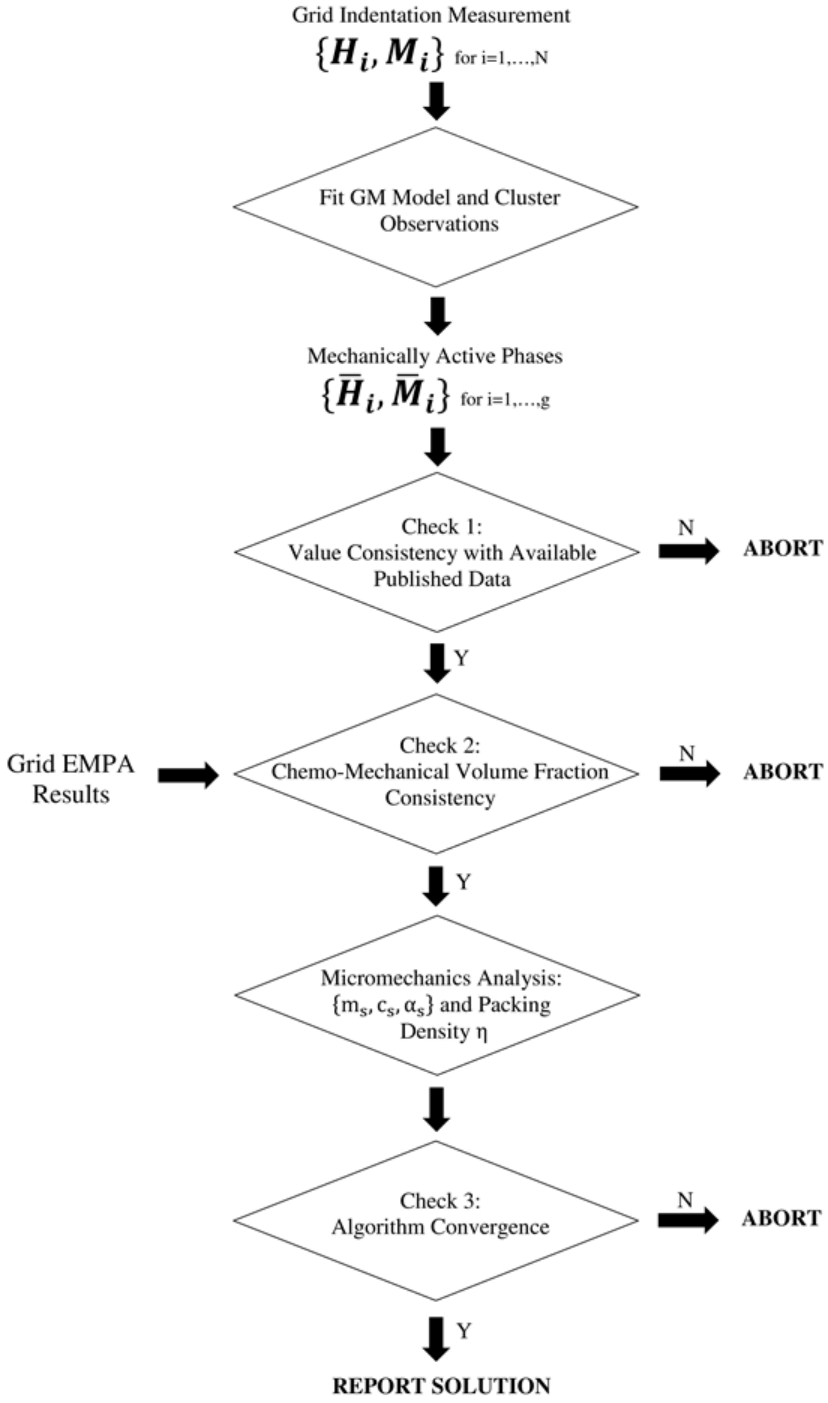
Supplementary Figure 10: Statistical analysis of the data obtained in the grid electron microprobe measurement on the system cement Class G with silica addition. a) a global plot of the experimental record in the Si – Ca domain, b) Maximum – Likelihood estimation and clustering of the experimental data into Gaussian Mixture Model corresponding to maximized penalized likelihood BIC statistics.



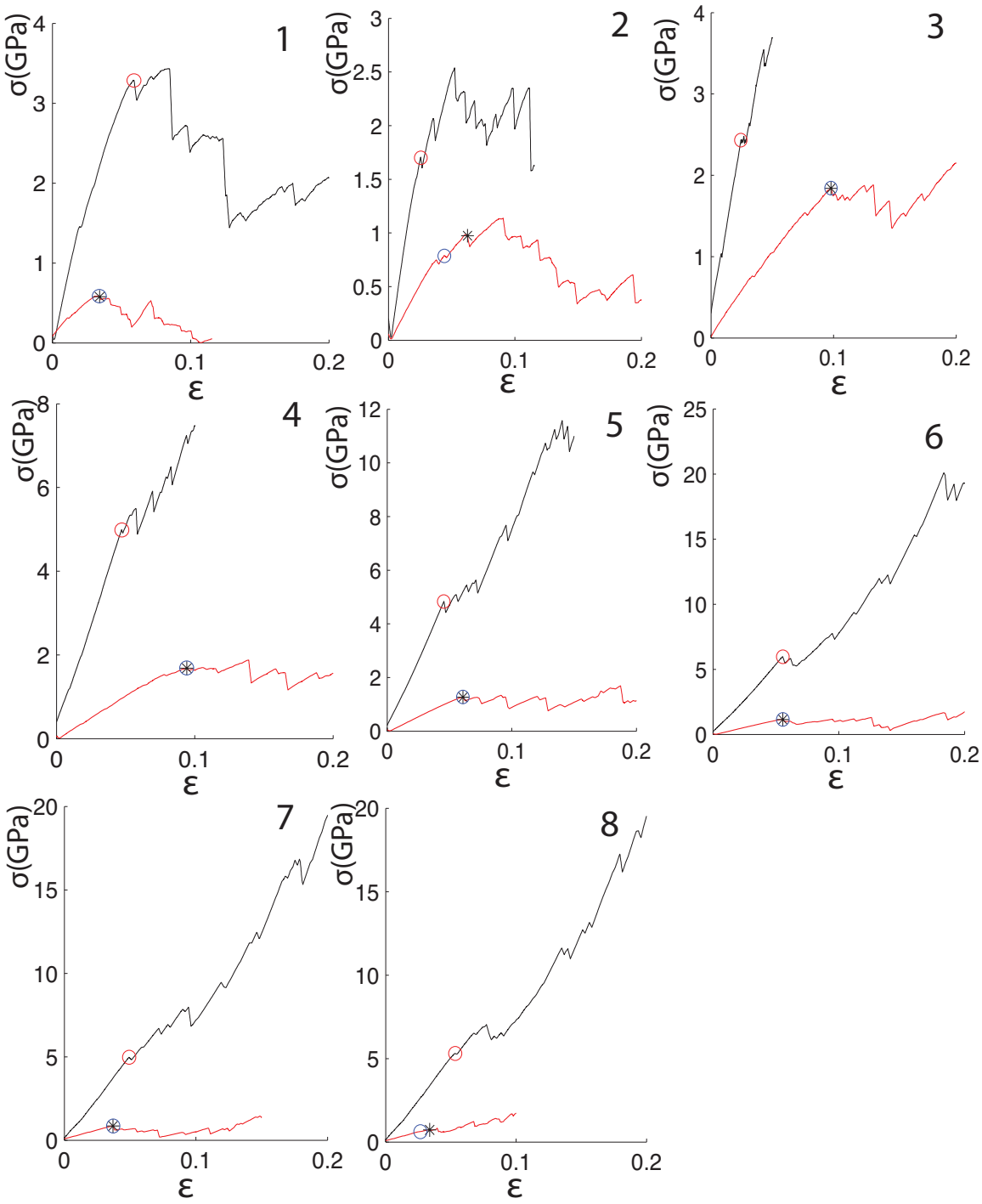
Supplementary Figure 11: Conical indentation in a cohesive-frictional porous material half-space. P is the indentation force, A_c is the projected contact area (projected on the initial sample surface). The porous composite (RVE) is composed of a solid phase (cohesion c_s and friction angle α) and pore space. Two solid–pore morphologies are presented: (a) matrix–pore inclusion morphology (η_0=0), (b) polycrystal morphology (η_0=1/2) assumed in the present modeling, after ref. 50.



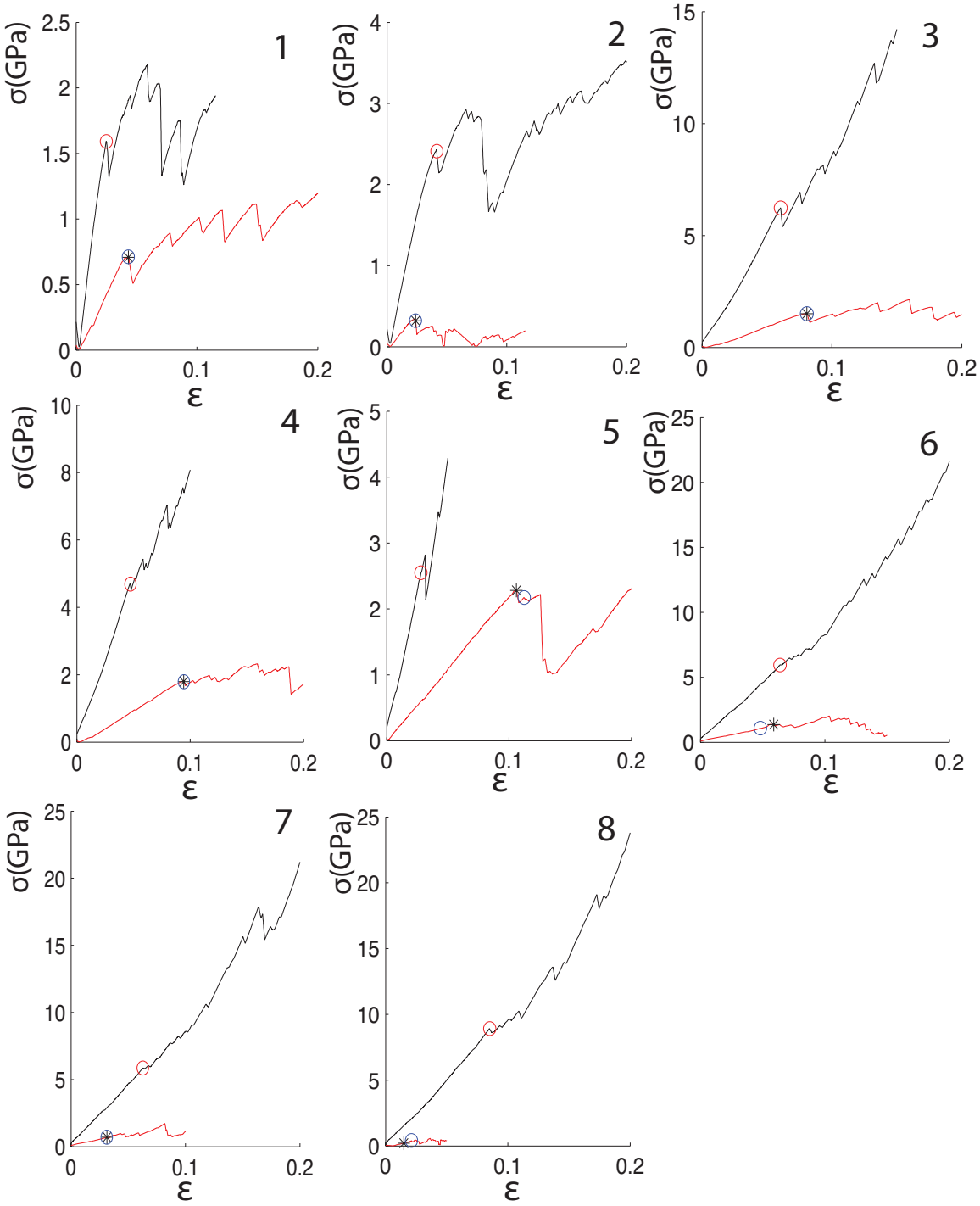
Supplementary Figure 12: Micromechanics based inference of the skeleton solid properties of the C-S-H dominated matrix for Class G neat cement paste with cured at room temperature under 3000psi pressure; (a) scatter diagram of indentation results with overlaid optimum micromechanics fit and corresponding packing density distribution with mean and standard deviation , the estimated solid hardness of C-S-H particles is , (b) scaling of the indentation hardness H and indentation modulus M (c) with packing density η , (d) distribution of model residuals ε.



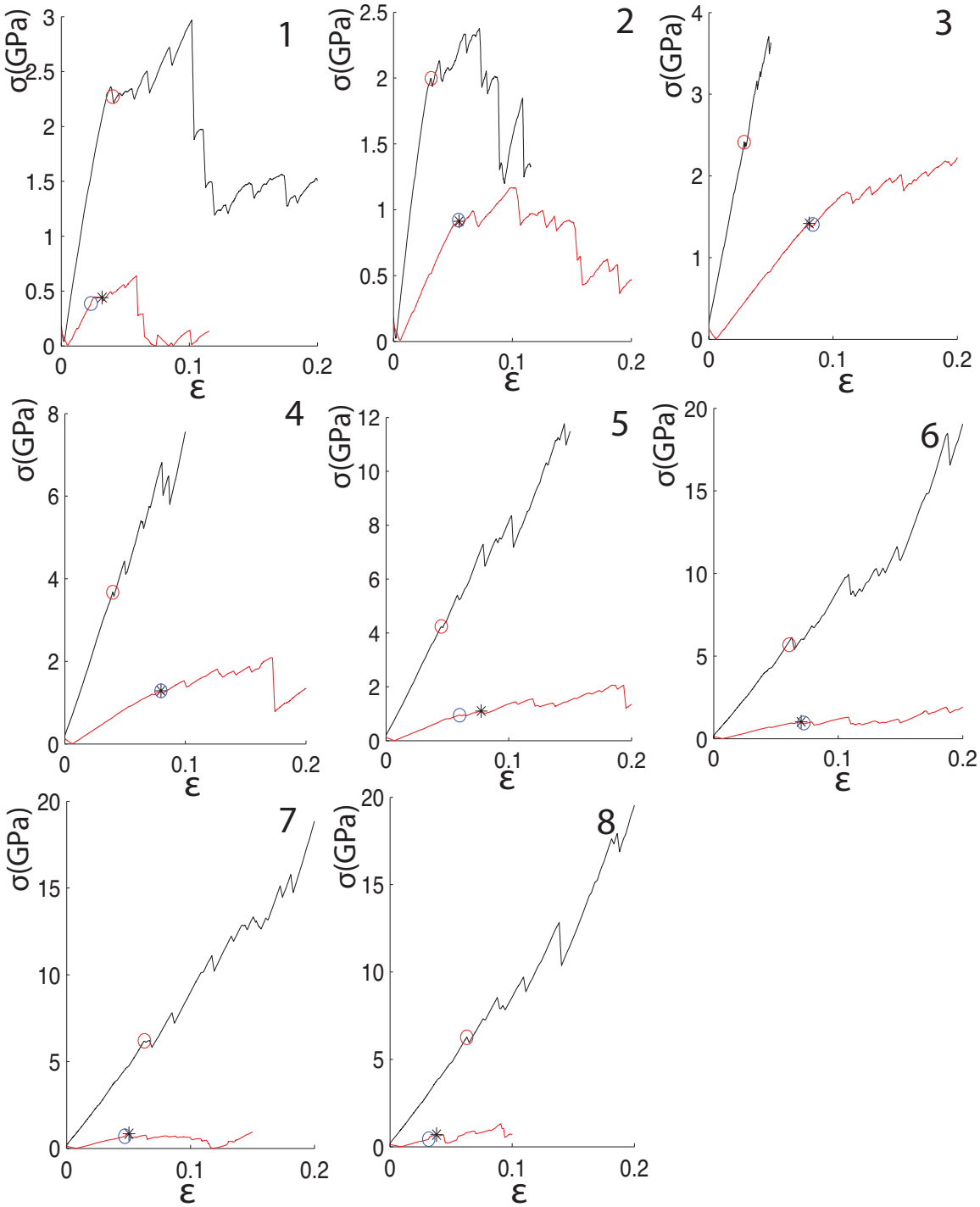
Supplementary Figure 13: Schematic of the step-by-step process of data analysis and indirect chemo-mechanic coupling. H_i and M_i are indentation hardness and modulus, respectively, obtained at the microscale via indentation, from which the M and H of the monolithic C-S-H solid particle are computed as described in Supp. Methods.



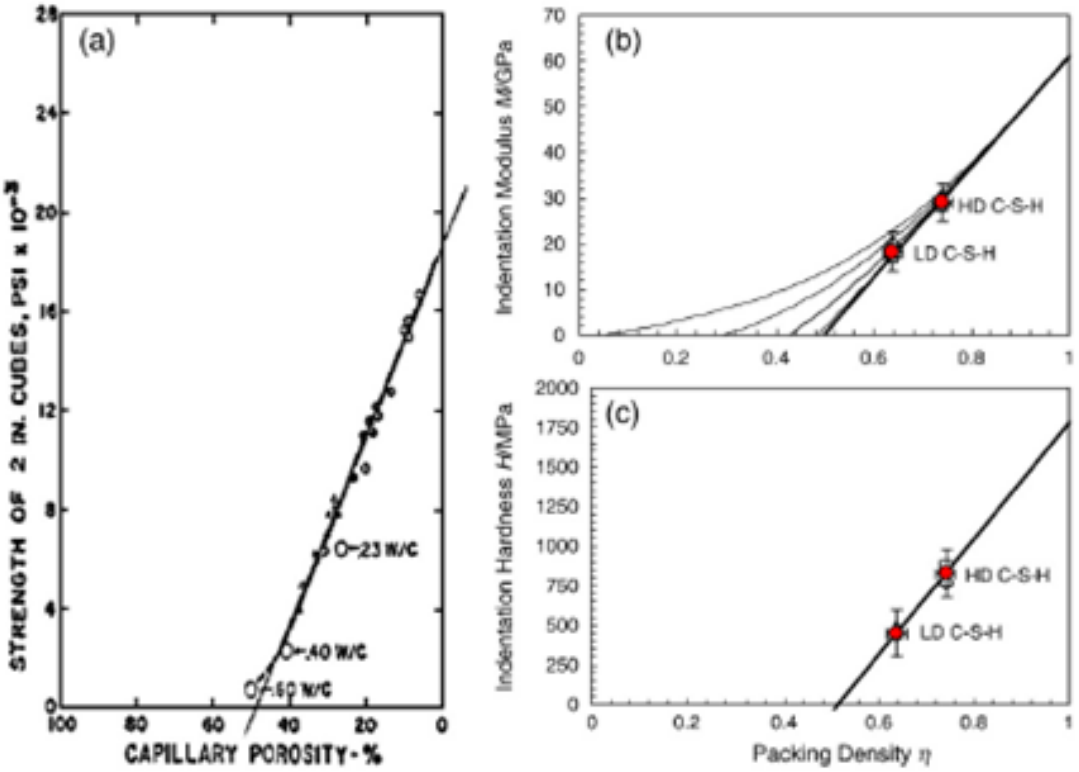
Supplementary Figure 14: Simulated stress-strain curves along the 8 deformation paths of a sample with C/S=1.5, one of the three polymorphs in Fig. 4 that exhibited the highest M/H; here, M/H = 21. The black and red circles identify the instability point (elastic limit strains) defined as the deviation from linear elastic stress-strain response for normal (black) and shear (red) deformation. The asterisks confirm that the ratio of the elastic limit strains in the normal and shear deformation corresponds to that particular deformation path.



Supplementary Figure 15: Simulated stress-strain curves along the 8 deformation paths of a sample with C/S=1.5, one of the three polymorphs in Fig. 4 that exhibited the highest M/H; here, M/H = 20. The black and red circles identify the instability point (elastic limit strains) defined as the deviation from linear elastic stress-strain response for normal (black) and shear (red) deformation. The asterisks confirm that the ratio of the elastic limit strains in the normal and shear deformation corresponds to that particular deformation path.



Supplementary Figure 16: Simulated stress-strain curves along the 8 deformation paths of a sample with C/S=1.48, one of the three polymorphs in Fig. 4 that exhibited the highest M/H; here, M/H = 19.5. The black and red circles identify the instability point (elastic limit strains) defined as the deviation from linear elastic stress-strain response for normal (black) and shear (red) deformation. The asterisks confirm that the ratio of the elastic limit strains in the normal and shear deformation corresponds to that particular deformation path.



Supplementary Figure 17: Scaling of mechanical properties with porosity (which is 1 - packing density, η). (a) macroscale compressive strength of cement paste; (b) indentation modulus, M, and indentation hardness, H, of C-S-H regions within cement paste. LD and HD are low packing density and high packing density, respectively, of the C-S-H solid nanoparticles within those porous C-S-H regions. (Figure taken from Ref. 91.)

Supplementary Table 1: Partial Charges in CSH-FF potential.

spec	Si	Ca	Cw	Ow	Hw
Charge	1.722357	1.435466	1.705529	-0.82	0.41
Spec	H	O	Oh	Ob	
Charge	0.425	[-1.15,-1.26]	[-0.93,-0.97]	[-1.01,-1.13]	

Supplementary Table 2 : Lennard-Jones potential parameters used in CSH-FF potential.

$spec_i$	$spec_j$	ϵ (Kcal/mol)	σ (Å)	$spec_i$	$spec_j$	ϵ (Kcal/mol)	σ (Å)
O	Ca	8.694E-4	4.365	Oh	Oh	6.180E-2	3.448
Ca	Ob	7.057E-4	5.452	O	Ob	4.543E-2	3.626
Ca	Oh	8.694E-4	4.365	O	Oh	4.543E-2	3.626
O	Cw	1.460E-3	4.365	O	Ob	6.180E-2	3.448
Cw	Ob	1.040E-3	4.446	Ca	Ow	8.760E-4	4.365
Cw	Oh	1.040E-3	4.446	Cw	Ow	6.042E-4	4.472
O	Si	5.604E-4	3.270	O	Ow	5.258E-3	4.241
Si	Ob	5.950E-4	3.261	Ow	Ob	8.717E-1	2.895
Si	Oh	5.950E-4	3.261	Ow	Oh	8.717E-1	2.895
O	O	1.243E+0	2.735	Ow	Ow	1.540E-1	3.163
Ob	Ob	6.180E-02	3.448				

Supplementary Table 3: Radial interaction Constants in CSH-FF potential.

$spec_i$	$spec_j$	k_r (Kcal/mol/Å^2)	r_0 (Å)
Hw	Ow	554.13	1
H	O	554.135	1

Supplementary Table 4: Angular interaction within water molecules in CSH-FF potential.

$spec_i$	$spec_j$	$spec_k$	k_θ (Kcal/mol/ θ^2)	θ_0
Hw	Ow	Hw	45.770	109.47

Supplementary Table 5: List of measured elements with associated standard materials and diffracting crystals used in the WDS measurements.

Element	Standard	Crystal
Si	Synthetic Diopside-Jadeite	TAP
Ca		PET
Al		TAP
Mg		TAP
Na		TAP
Fe	Hematite	LiF
O		LDE1
K	Synthetic Orthoclase	PET
Ti	Rutile	LiF
S	NiS	PET

Supplementary Table 6: Experimental data for nanoindentation and EMPA/WDS analysis. Data indicated as mean +/- standard deviation, where standard deviation is calculated by bootstrap method.

Ca/Si	Indentation Modulus (M in GPa)	Hardness (H in GPa)
*0.75±0.37	90.3±9.3	11.6±4.6
*0.78±0.28	115.8±24.7	13.8±4.9
0.88±0.06	93.3±6.6	12.4±1.8
0.98±0.04	98.1±4.9	8.2±0.7
1.69±0.11	69.0±6.8	5.3±0.9
**1.79±0.10	77.8±6.6	6.0±0.6
1.9±0.15	69.1±13.3	6.0±2.1
2.13±0.23	74.5±2.6	6.3±0.6
*2.1±0.18	100.9±8.3	8.5±1.0
**2.10±0.25	64.1±4.5	3.7±0.3

* - This sample type is typically rich in unreacted clinker, leading to unresolved mechanical properties of C-S-H with our cluster approach. Those samples have pronounced composite effect of CSH-clinker phases and therefore they are not illustrated in figure 3 and 4. As it can be seen in possible ranges of H and M values, they are close to those measured for clinker phases C_2S and C_3S[1,2]. ** - values obtained from direct measurement of chemistry and indentation properties at the grid point location on C-S-H dominated phase

Supplementary Discussion:

Comparison of Combinatorial Models with Richardson's[3] T/CH model

In this subsection, we include the comparison of our combinatorial model with the so-called Tobermorite-'solid solution' calcium-hydroxide (T/CH) model. This model is initially proposed by Richardson and Groove[3,4] and is intended to provide a "crystal-chemical" picture of C-S-H at different C/S starting from 14Å Tobermorite. Following the notation of Richardson[5], the chemical composition of T/CH model can be written as:

$$Ca_x H_{(6n-2x)} Si_{(3n-1)} O_{(9n-2)}.\ zCa(OH)_2\ .\ mH_2O \qquad (1)$$

Where x is the number of Ca^{2+} ions that to charge balance the silica chains, n is the integer number representing the Drierketten structure of the silicate chain (1 for dimer, 2 for pentamer and so on), z and m indicate the number of calcium-hydroxide and intact water molecules in the interlayer spacing of C-S-H.

In our combinatorial atomistic simulation approach, we start from 11Å Tobermorite and construct our samples carefully following an eight-staged process including proper consideration of chemical reactions via ReaxFF potential[6]. As we showed in Fig. 2 in the main text, our simulations can reproduce drying, elastic and inelastic neutron scattering and ^{29}Si NMR experiments. The chemical composition of C-S-H, as seen through our combinatorial approach, can be written as:

$$Ca_y\ Si\ O_{2.75}\ (OH^{Ca})_{0.85(y-1)+a}\ (OH^{Si})_{1.15(y-1)+0.5+b}\ .\ [0.8(y-1)+0.3+c]H_2O \qquad (2)$$

Where y is the Ca/Si ratio and a, b, c represent the variability in the nanostructure of C-S-H at a given C/S ($a=b=c=0$ corresponds to average pattern for a given C/S and $|a|$, $|b|$, $|c|<0.2$ to only account for polymorphism). The variability in the structure of C-S-H is due to the change in the structure of calcium-silicate backbone at a given Ca/Si ratio. In the supplementary materials under the section of free energy calculation, we have shown

that these structures are energetically competitive and hence have polymorphic nature. Note that OH^{Ca} and OH^{Si} show hydroxyl groups in calcium hydroxide and silanol groups, respectively.

Having properly identified the stoichiometry of both T/CH and combinatorial model, we dedicate the forthcoming paragraphs to discuss seven points of agreement and disparity between the two models.

1) In T/CH model, *n* is an integer number, which gives a mean chain length of (3n-1). Therefore, the T/CH model fails to describe the variability of chain length at the nano-scale for a given C/S. In fact, in a realistic cement paste a combination of monomers, dimers, pentamers and so on exist which rises to a non-integer mean chain length (MCL). Unlike T/CH, our molecular modeling along with NMR experiments of Chen et al.[7] asserts that MCL can have a value between an upper and lower bound. Our own NMR experiments on a subset of samples synthesized and characterized in this work (see Methods) also finds MCL to be within this range. This is mathematically expressed as:

$$\frac{2y-1.75}{y-1} < \overline{MCL(y)} < \frac{2.5y-2}{y-1} \quad (3)$$

where $\overline{MCL(y)}$ denotes the average mean chain length at C/S=*y*. According to eq. (3), the $\overline{MCL(1)}$ is infinite indicating Tobermoritic chains. At C/S>1, $\overline{MCL}$ is between the two limits, which depend on the structural configuration of the calcium silicate backbone. We recognize that the exact prediction of MCL is controlled by the kinetics of cement hydration process (and not through equilibrium conditions as shown by our detailed free energy calculations), which depends on curing conditions, stoichiometry of reactants, water to cement ratio along with other factors.

2) At 1<C/S<1.5, T/CH model proposes that by gradually removing bridging SiO_2 groups, the calcium to silicon ratio can be increased from 1 to 1.5. At C/S>1.5, T/CH model assumes that the molecular structure of C-S-H is combination of Tobermorite

devoid of bridging silica group sandwiched with calcium-hydroxide sheets in the form of nano-scale Portlandite layers (CH). At this point, the model fails to answer concerns regarding the mixing of CH with defective Tobermorite. Where these layers are located? How are they bonded to the calcium-silicate layers? What is the size distribution of these Portlandite nano-layers? What is the molecular structure of C-S-H at Ca/Si=1.75?

Recognizing these problems, we adopt a different approach, which resolves this issue. At Ca/Si>1.5, we start removing the pairing sites in the defected Tobermorite structure such that the number of monomers is kept minimal to agree with NMR experiments. Moreover, we run reactive molecular dynamics simulations to prompt reactions in the silicate regions to polymerize existing monomers if needed (see the inset of figure 2.c in the main text). Having done this, we show that the MCL of combinatorial structure are within the bounds observed in ^{29}Si NMR experiment by Chen et al[7]. Therefore, our combinatorial model does not need the additional CH layers to describe the high C/S.

3) T/CH model (eq. (1)) does not provide any relation between the C/S and *m* (the structural water content). In fact, T/CH model fails to provide quantitative or even qualitative description for the correlation between the defect content in C-S-H and the water adsorbed in the interlayer spacing and voids created by removing silica groups. By contrast, our combinatorial model provides a clear image and identifies a linear correlation between the water content and Ca/Si ratio. This is achieved by performing Grand Canonical Monte Carlo simulation to adsorb water in C-S-H upon creation of defects in the structure of Tobermorite. As discussed in Fig. 2.a of the main text, the total equivalent water (comprising both structural water and hydroxyl groups) matches perfectly with drying[8] and elastic neutron scattering[9] experiments. We emphasize that both neutron scattering and drying experiments explore the total hydrogen content rather than hydrogen in structural water molecules per se. This indicates that our combinatorial approach and the proposed chemical composition fully display the interplay between chemistry of calcium-silicate backbone and the water content, which is missing in T/CH model.

4) T/CH model (eq. (1)) does not provide any relation between the Ca/Si ratio and *z* (the number of calcium-hydroxide bonds). In fact, Thomas et al.[10] showed that there is a linear correlation between Ca-OH content and C/S. Our reactive molecular dynamics simulation *without any tuning parameters* indicated the very same linear behavior (see Fig. 2b). This means that the environment of Ca is fully captured in our model without the need to introduce artificial CH in the composition. Given that our composition captures both total equivalent H_2O and Ca-OH content, this entails that Si-OH (silanol) content should also be properly captured in our model simply because of conservation of mass. However, there are no experimental data so far for correlation of C/S and silanol content.

5) Unlike T/CH model which is controlled by 4 parameters (*x,n.z,m*), the combinatorial model is majorly controlled by 1 parameter (C/S=*y*). In fact, by assuming a=b=c=0, one reaches good agreement (in average) with experimental observations. We emphasize that the effect of (*a,b,c*) is marginal with respect to y, as these parameter are describing local textural variability at a given C/S.

6) T/CH model provides a "crystal-chemical" interpretation of C-S-H structure. As extensively discussed by Richardson[5], the defective Tobermorite remains fairly crystalline and also the additional CH is assumed to be crystalline at high Ca/Si ratios. Therefore, the molecular structure proposed by Richardson and his co-workers has crystalline nature. This is against the observations by Pellenq and his co-workers.[11,12] that C-S-H at high C/S presents pair-correlation or structure factor responses, which resemble that of defective amorphous and 3D glassy materials.

7) Finally, the very fundamental difference between the T/CH model and the combinatorial models is the predictive power of these two models. While T/CH gives a description of molecular structure of C-S-H, it is basically incapable of providing qualitative or quantitative insights regarding the physical properties of C-S-H. Unlike T/CH, as shown in Figs. 3,4 and 5 in the main text, combinatorial models are able to predict physical properties of C-S-H at the nanoscale such as indentation elastic modulus

and hardness. This is a clear advantage for the combinatorial model as placing it as a foundation for computational material design.

The above 7 facts fully differentiates the combinatorial approach with respect to the model proposed by Richardson and co-workers[3–5,13,14].

Supplementary Methods:

This supplementary material is intended to explain the details required to reproduce the results of the atomistic simulations and experimental data provided in the main text. We have dedicated the first section to different aspects of atomistic simulation and model construction. We have fully described different stages required for construction of realistic C-S-H models. Furthermore, we have extensively discussed how these numerical samples can be utilized to calculate elastic properties, strength, free energy, structure factor and total X-ray pair correlation function. At the end, we have discussed the details of experimental techniques employed to validate certain predictions of the simulations. It includes specifics of Transmission Electron Microscopy (TEM) imaging and chemomechanical analysis of C-S-H using statistically coupled nanoindentation and Wavelength Dispersive Spectroscopy (WDS) experiments.

Atomistic Simulation and Model Construction

The computational strategy used here, aiming to vary Ca/Si systematically and explore resulting properties, is a combination of several classical atomistic simulation techniques. These methods are deployed at different stages of the model preparation. The complex hydrate materials have different types of atoms with varying chemical environments, which is discussed in details in following subsections. These simulations, which have classical nature, are performed using ReaxFF [6] and CSH-FF [15] potentials. As schematically presented in Supplementary Fig. 1, there are eight explicit stages to construct a proper molecular structure of calcium-silicate-hydrates (C-S-H). To eliminate the finite size effect in computational modeling, a super-cell of the molecular structure of

Tobermorite 11Å due to Hamid [16] is prepared in the first step. 11Å Tobermorite has no hydroxyl groups and the length of its silica chains is infinite. In the second step, the entire water molecules are removed from the interlayer spacing. In the third stage, 150 numerical models are constructed by randomly cutting the silica chains (removing charge-neutral SiO_2 groups). This increases the Ca/Si ratio (C/S) of the models. This provides a wide range of C/S to investigate the effect of stoichiometry on physical properties of C-S-H. At this stage, first the interlayer calcium atoms and subsequently all the atomic species and cell dimensions are relaxed using energy minimization at 0 K. In the next stage, water molecules are adsorbed back to the interlayer spacing and nano-voids created by removing SiO_2 groups using the Grand Canonical Monte Carlo simulation. At this point, the interlayer water is reacted at 500 K with the defective calcium-silicate backbone of C-S-H using ReaxFF potential. In this stage part of the interlayer water dissociates in to hydroxyl groups and proton. Minor condensations of silica chains are also observed in some samples. In the sixth stage, a comprehensive topological analysis is performed to identify the local environment of each atom. This facilitates the transfer of model from reactive environments provided via ReaxFF potential to non-reactive environment in CSH-FF force field. In stage seven, a 3ns long simulated annealing is performed on each individual sample to reduce the temperature from 500 K to 300K at ambient pressure. Finally, the samples are relaxed and proper sampling of energy space is performed to measure elastic constants, hardness measurements, free energy, medium-range order and total X-ray pair correlation function.

Descriptions of Atomic Species:

Although minor impurities might be existing in C-S-H [17,18], the main binding phase in hardened cement paste is generally considered to be consisted of Si, Ca, O and H atoms. These atoms adopt different characteristics depending on their local chemical environments. To appropriately address such variations, classical potentials utilize different labels for a given element to describe these chemical environments. However, ReaxFF potential shows the ability to calculate the interatomic interactions without the need to designate different labels for various chemical environments. For instance, due to

charge compensation, the O-H bonding would be different inside a water molecule and a hydroxyl group. Therefore, for classical simulations using CSH-FF potential, one requires the a priori knowledge about the local atomic environments in the molecular structure of material. More specifically, CSH-FF uses nine labels to describe the four types of elements in C-S-H. Our convention for these nine labels is as follows: Ca (intralayer calcium), Cw (intralayer calcium), Si (silicon), Oh (oxygen in hydroxyl group), H (hydrogen in hydroxyl groups), Ob (bridging oxygen in silica chain), O (non-bridging oxygen in silica chain or calcium layers not in a hydroxyl group), Ow (oxygen in water molecules), Hw (hydrogen in water molecule). The partial charges of all species are provided in Supplementary Table 1. The partial charges of all species are constant except oxygen atoms. In fact, the partial charges of oxygen species in calcium-silicate layers are set to vary in order to ensure charge-neutrality within numerical samples. Since ReaxFF calculations are performed using electronegativity equalization method (EEM) [19] and the species of atoms are subjected to change as a result of chemical reactions, the unit cells might not be charge-balanced after reactions. A comprehensive sensitivity analysis on the elastic properties is carefully performed with different charge-balancing schemes. Due to the small charge-imbalance in the samples, charge-neutralization schemes do not strongly affect the elastic properties. In this work, the excess charge is equally distributed among O, Ob and Oh atoms.

Interatomic Potentials

In this work, we use two interatomic potentials: ReaxFF and CSH-FF. ReaxFF is a reactive force field originally developed on the same notion as bond order potentials [20–22] and it is further enriched with polarizable charge model [19]. ReaxFF is a powerful tool for modeling formation and breakage of bonds during chemical reactions and modification of charge distribution during molecular dynamics (MD) simulation. In this force field, the total potential energy is written as a sum of various energetic contributions:

$$E_{tot}^{REAXFF} = E_{Col} + E_{VdW} + E_{radial} + E_{angular} + E_{lp} + E_{OverC} + E_{UnnderC} \quad (4)$$

where E_{Col}, E_{VdW}, E_{radial}, $E_{angular}$ and E_{lp} are Coulomb, Van der Waals, radial bond stretching, angular bond bending and lone pair contributions to the potential energy, respectively. E_{OverC} and $E_{UnnderC}$ are defined as over- and under-coordination energy contributions to impose energetic penalties for high and low coordination numbers. Each of these terms depends on the local environment of atoms described via a bond order expression. In addition, a screened taper function is applied on the non-bonded Coulomb and Van der Waals interactions. In addition to previously defined Si-OH interactions [23], ReaxFF was further developed to include Ca-OH interactions. This was particularly done to incorporate calcium within ReaxFF framework. This facilitates the description of chemical reactions within C-S-H ultraconfined medium. For more detailed discussion on the ReaxFF potential, interested readers are referred to the work of Resso et al. [24].

The CSH-FF interatomic/intermolecular potential is a core-only model for hydrated calcium-silicates, in which atoms are described as interacting points through Columbic and short-range potential functions. The total interatomic potential for CSH-FF potential reads:

$$E_{tot}^{CSHFF} = E_{Col} + E_{lj} + E_{radial} + E_{angular} \quad (5)$$

where E_{Col}, E_{lj}, E_{radial}, and $E_{angular}$ are the electrostatic, short-range Van der Waals in the form of Lennard-Jones potential, bond stretching and bond bending contributions to the potential energy, respectively. CSH-FF potential is a customized force field for hydrated calcium-silicates, and it is not transferable to other oxides with different chemical environments. Therefore, CSH-FF can be seen as an improved version of the generic ClayFF potential [25] (originally developed for structural properties of clay minerals) for hydrated calcium-silicates such as cement hydrates. To describe the molecular structure of water and its interaction with other water molecules, both CSH-FF and Clay-FF potential models are built around the flexible SPC model for water. In particular, CSH-FF was designed to reproduce both structural and elastic data of Tobermorite minerals obtained from ab initio calculations [15]. This was achieved by adjusting a set of 29 potential parameters as described in Ref [15]. The original form of

CSH-FF was employed in the stage 1 through 4 of the model preparation where the hydroxyl groups were absent in the model.

In order to extend the predictability of the CSH-FF model, the latter was slightly modified to include a couple of new features. Primarily to reduce the computational costs incurred by the calculation of long-range forces, the columbic interactions are calculated using Wolf method [26]. In Ewald summation scheme, the inverse distance in electrostatic interaction is rewritten as its Laplace transformation and then split into two moderately fast converging summations between real and reciprocal spaces. The distribution of the summation between reciprocal and real space is controlled by γ. The final Ewald expression reads:

$$E_{Col}^{Ewald} = \frac{1}{2}\frac{4}{\pi V}\sum_G \frac{exp\left(\frac{-G^2}{4\gamma}\right)}{G^2}\sum_i\sum_j q_i q_j \exp\left(-iGr_{ij}\right) + \frac{1}{2}\sum_i\sum_j \frac{q_i q_j \mathrm{erfc}(\gamma^{1/2} r_{ij})}{r_{ij}} \quad (6)$$

where q is charge, r is the interatomic distance, V is the volume of the crystal's unit cell and G is a vector of the reciprocal lattice. The *erfc* stands for complimentary error function in mathematics. The notion of Wolf summation is based on the idea of charge neutralization of a spherical subset of atoms in condensed ionic systems. Originally, Wolf proposed a two-body real space summation, which significantly reduces the computational cost of Ewald summation in real and reciprocal spaces. The Wolf summation reads:

$$E_{Col}^{Wolf} = \frac{1}{2}\sum_i \sum_{j\,(r_{ij}<R_c)} \left(\frac{q_i q_j \mathrm{erfc}(\gamma^{1/2} r_{ij})}{r_{ij}} - \lim_{r_{ij}\to R_c}\left(\frac{q_i q_j \mathrm{erfc}(\gamma^{1/2} r_{ij})}{r_{ij}}\right)\right) - \left(\frac{\mathrm{erfc}\left(\gamma^{\frac{1}{2}} r_{ij}\right)}{2R_c} + \frac{\eta}{\pi^{1/2}}\right)\sum_i q_i^2 \quad (7)$$

where N is the number of atoms in the unit cell, q is the charge, r is the interatomic distance, R_c is the cut-off distance for which the neutralization technique is imposed and η is the damping factor. While Ewald summation scales with O($N^{3/2}$), Wolf summation scales with O(N). This is computationally advantageous, especially for extensive MD studies involving thousands of simulations. However, the success of the Wolf summation

is tied to the proper identification of the Wolf's free parameters, R_c and η. To obtain the proper values for C-S-H's molecular structure, the absolute difference between the Ewald and Wolf summations is calculated for different values of R_c and η. The normalized error is given as:

$$\lambda = \frac{\left|E_{Col}^{Ewald} - E_{Col}^{Wolf}\right|}{E_{Col}^{Ewald}} \quad (8)$$

As it is shown in Supplementary Figure 2, λ is quite negligible (in the order of 10^{-4}) in the window of $11 < R_c < 14$Å and $0.2 < \eta < 0.25$. To apply a uniform cut-off radius in the simulation box for both long- and short-range interactions, R_c is set to 12Å. The damping factor, η, is set to 0.25 in all simulations. The second modification in the CSH-FF potential pertains to incorporation of hydroxyl groups. The force field parameters for hydroxyl groups are included in the extended form of the CSH-FF potential.

The short-range interactions in CSH-FF and ClayFF force fields are described via Lennard-Jones potential:

$$E_{lj} = 4\epsilon \sum_i \sum_{j>i} \left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \quad (9)$$

where σ and ϵ have the dimensions of distance and energy, respectively. These parameters are provided for short-range interactions between pairs of elements in the model. The potential parameters are provided in Supplementary Table 1. Similar to ClayFF potential, the hydrogen species do not have short-range interactions. The radial and angular excursions in water molecules and hydroxyl groups are described via simple harmonic potentials:

$$E_{radial} = \sum_{i=1}^{N_{OH}} \frac{1}{2} k_r \left(r_{ij} - r_0\right)^2 \quad (10)$$

$$E_{angular} = \sum_{i=1}^{N_{H_2O}} \frac{1}{2} k_\theta \left(\theta_{ij} - \theta_0\right)^2 \quad (11)$$

where k_r and k_θ are radial and angular stiffness, respectively. Parameters r_0 and θ_0 denote the equilibrium hydroxyl bond length and Hw-Ow-Hw angle in water molecules, respectively. N_{OH} and N_{H2O} are the total number of O-H bonds in hydroxyl groups and water molecules and total number of water molecules, respectively. The radial and angular constants are provided in the Supplementary Tables 3 and 4.

Generating different C/S

In this work, the C-S-H models describe cement hydrates as defective calcium-silicates with short- to medium-range order. These models are suggestive of Tobermorite layer-to-layer long-range correlations. In contrast to Tobermorite polymorphs, cement hydrates have a different stoichiometry with significantly larger calcium (and water) content: C/S in Tobermorite is unity or smaller while it is around 1.7 for ordinary Portland cement hydrates[9]. This large C/S can be achieved by engineering the silica chains lengths (these chains are infinite in the case of Tobermorite minerals). In other words, the C-S-H model of cement hydrates contains many defects and also a larger water content compared to that of Tobermorite polymorphs. Silica chain defects in the C-S-H models were implemented by randomly removing SiO_2 groups from an anhydrous 11Å Tobermorite crystal so the targeted C/S was obtained. This is precisely the strategy that was deployed in the present work aiming at exploring the effect of varying this C/S on texture and mechanical properties. It is fairly well known that in real cement paste the C/S spans majorly in the range of 1 to 2.1.

In this section, we implemented the above strategy to impose a given chemistry in terms of C/S. As for all the C-S-H models, we started from a periodic simulation box containing $2\times3\times1$ unit cells of Hamid's structure of 11Å Tobermorite [16]. At this stage, we constructed an anhydrous version of this super-cell as we were only interested in creating the solid backbone of the model itself; the addition of water was carried out at the fourth stage using Grand Canonical Monte-Carlo simulation utilized to describe adsorption/desorption processes [27]. The simulation boxes are large enough to accommodate silica dimers, pentamers and even octamers as these types of silica chain are known to be dominantly present in real cement paste. One can map out each initial

infinite silica chain onto a line that contains "1" for pairing tetrahedral and "2" for the bridging ones. In Tobermorite, bridging and pairing silica tetrahedrals form the so-called Drierketten pattern that in this simple notation writes "112". Thus within our periodic simulation box, silica structure of Tobermorite writes "112112112". Juxtaposing all the eight chains (2 per side of each layer, two layers), we can assemble the following matrix for the abovementioned 11Å Tobermorite super-cell:

$$\begin{pmatrix} 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 \\ 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 \\ 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 \\ 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 \\ 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 \end{pmatrix}$$

Note that the above super-cell contains 72 calcium and silicon atoms (C/S=1). To construct a model with a give C/S, we first developed a simple computer program that uses a random number generator to replace some elements of the above matrix by "0" with the constraint of matching a given C/S (within ± 0.1) and producing as much dimers ("0110") pentamers ("0112110") and octamers ("0112112110") as possible. For instance, the "cutting" matrix for C/S=1.30 can be written as:

$$\begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 & 0 & 1 & 1 & 2 & 1 \\ 1 & 0 & 1 & 1 & 2 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

The second step is to realize the combinatorial aspect of the approach, as there are many ways to "cut" for a given stoichiometry. Here, we considered 10 to 15 different samples

for each C/S value. Second, each "cutting matrix" is projected onto our simulation box. Each "0" corresponds to a silicon atom that has to be removed along with two of its neighboring oxygen atoms. Since SiO_2 is a charge-neutral group, the unit-cells electro-neutrality is maintained. We have made sure that silicon in ending sites (Q_1 in solid state NMR nomenclature) had the necessary four oxygen neighbors. The preparation of monomers in C/S>1.5 is only plausible if the pairing sites are removed from the silica chains. In fact, upon removal of all the bridging silica sites from the initial crystalline structure, the Ca/Si ratio is found to be 1.5. Therefore, some monomers, Q_0, inevitably exist in the initial solid skeleton of C-S-H samples. However, the percentage of monomers in the samples is reduced via reactive modeling through which condensation of silicate chains is energetically plausible.

Grand Canonical Monte-Carlo technique for water adsorption:

In this work, we first produced an anhydrous version of C-S-H models by removing all the interlayer water. We subsequently calculated the maximum amount of water that can be accommodated in its pore voids via the Grand Canonical Monte-Carlo (GCMC) simulation technique. GCMC is a well-suited technique to study adsorption/desorption processes in nanoporous materials. GCMC determines the properties of a system at a constant volume V (the pore with the adsorbed phase) in equilibrium with an infinite fictitious reservoir of particles imposing its chemical potential μ and its temperature T [27]. For different values of μ, the absolute adsorption isotherm can be determined as an ensemble average of the adsorbed atom numbers in the system versus the pressure of the gas reservoir P (the latter can be obtained from the chemical potential according to the equation of state for the bulk gas). The adsorption and desorption processes can be respectively simulated by increasing or decreasing the chemical potential of the reservoir; the final configuration of a simulation is the initial state for the next point. Periodic boundary conditions were used in all directions of space as for the energy minimization procedure. An equal number attempt for translation, rotation, creation or destruction of molecules has been chosen. The isotherm has been calculated for 300 K. Acknowledging the very restricted available space in between Tobermorite layers, one should not expect capillary condensation to occur by contrast to larger pore systems such as Vycor [28,29] or

in the larger inter C-S-H grain pore space [30]. In our case, the adsorption/desorption process is expected to be close to that observed for microporous zeolite as far as water adsorption is concerned[29]. We did not calculate the entire water adsorption/desorption isotherm but perform a single GCMC simulation with the water chemical potential fixed to a value that corresponds to the bulk liquid phase with a density of 1 g/cm^3 at room temperature (μ = 0 eV for flexible SPC water model). CSH-FF potential was used in all GCMC calculations via GULP [31,32].

Reactive and Non-Reactive MD in the NPT and NVT ensembles

All reactive simulations are performed using LAMMPS [33]. We employ the velocity Verlet algorithm for our time integration schemes. In order to appropriately describe chemical reactions, the time step is set to 0.1 fs. To further trigger chemical reactions, ReaxFF simulations are carried at 500 K to facilitate the reaction of water molecules with defective calcium-silicate layers. According to Arrhenius law, this assists the water molecules to overcome the reaction energy barriers. This reduces the computational expense of reactive modeling by reducing the required simulation time to reach equilibrium. Initially, all samples are relaxed in canonical ensemble (NVT) for at 200 ps using Nose-Hoover thermostat. Subsequently, the samples are further relaxed in isothermal-isobaric ensemble (NPT) at the 500 K and 1atm using Parrinello-Rahman method[34]. The overall performance of a C-S-H sample with Ca/Si = 1.5 is presented in Supplementary Figure 2.

Supplementary Figure 2.A provides the radial distribution function, RDF or *g(r)*, of Si-O, Ca-O and O-H bonds. Since the peaks in *g(r)* are intact, the slightly elevated temperature has not affected the solid backbone of C-S-H. It should be also noted that while Si-O and O-H RDFs have a very defined first sharp peak, the Ca-O RDF exhibits a wider peak. This means that the radial excursions of Si-O and H-O bonds are very limited. However, Ca-O bond has a much larger radial excursion. This means that Ca-O bond is not as stiff as Si-O and H-O bonds. As it can be seen in Supplementary Figure 2.B, in the very first few picoseconds of the simulation, a large portion of water molecules dissociates to form OH^- and H^+ ionic pairs. While the hydroxyl groups coordinate with Ca^{2+} ions, protons

react with non-bridging oxygen atoms in silicate groups. Non-bridging oxygen atoms are only attached to one silicon atom in the silicate chain. No hydrogen atom was found coordinated to the bridging oxygen shared between adjacent silicate groups. In fact, over-coordination energy penalty makes such a reaction nearly energetically implausible. This hydrophobic behavior was originally seen in siloxane oxygen atoms[35] and further confirmed for C-S-H[36]. The dissociation mechanism affects the large portion of interlayer water molecules in the C-S-H and the dissociation percentage depends on the C/S which is directly proportional to the number of non-bridging oxygen atoms in silicate chains. Reactive modeling also affects the solid skeleton of C-S-H.

Supplementary Figure 3.C shows the effect of reactivity on condensation of silicate groups. Unlike the recent work of Manzano et al. [37], it is observed that silicate groups condensate in a number of C-S-H models. In the sample presented in Supplementary Figure 3, a pair of monomers has condensed to form a dimer. This is in agreement with energetic arguments provided by DFT calculations of isolated silicate groups by Manzano et al.[38] showing dimers have high chemical stability index. The observed condensation mechanism is triggered by two silicate tetrahedra pairing through which a dimer is produced and an oxygen atom is released. The free oxygen pairs with a proton and the resultant hydroxyl group coordinate with a Cw^{2+} cation to complete the reaction path. Supplementary Figure 3.D presents the difference in percentage of silica monomers before and after reactive modeling for all the C-S-H models. It is observed that some samples are intact, in some condensation occurs and dissociations of longer instable chains increases monomer content slightly in a few samples. As shown in Supplementary Figure 3.C, the condensation occurs in the very first couple of picoseconds of simulation. This signifies that due to random "cutting" of chains, some monomers have extremely unrealistic configuration. This unrealistic state is relaxed via polymerization of monomers producing energetically more stable dimers. The energy barriers for the condensation of silica chains are large, in the order of a few *eVs*, and the relevant simulation time required to capture such a rare condensation events exponentially increases according to the well-known transition state theory. The methods and approaches to overcome the barriers in energy landscape and subsequent time marching

strategies such as Activation-Relaxation Technique (ART)[39] or Autonomous Basin Climbing (ABC) [40] are beyond the scope of this paper.

Topological Analysis

ReaxFF simulations are computationally expensive to undertake. In order to overcome the prohibiting computational cost of reactive modeling, required for extensive mechanical testing on each individual sample, all samples are transferred to non-reactive environment after completion of reactions. As the direct consequence of reactive modeling, the local environment of atoms undergoes drastic changes. This necessitates a precise characterization of local environment of each single atom. Such a precise analysis can be performed using the fundamental ideas of topological constraint analysis [41]. In C-S-H's molecular structure, Si, Ca and H cations are always connected to O atoms, the latter showing different local environment (Supplementary Figure 4). In order to differentiate between these species, we performed a careful topological analysis of the coordination of each O atom. Bridging oxygen atoms (Ob) are thus defined as oxygen atoms showing exactly two Si neighbors in their O-Si first coordination shell (FCS), which limit has been fixed as the first minimum of the $g_{O\text{-}Si}$ RDF, typically 2.5Å. On the contrary, non-bridging oxygen atoms (O) show only one Si in their O-Si FCS. Those species arise from the depolymerisation of the silicate network by Ca and H cations and can belong to hydroxyl groups O-H (Oh) if they show an H neighbor in their O-H FCS (minimum of the $g_{O\text{-}H}$ RDF, typically 1.3Å) or are connected to Ca atoms otherwise. Remaining O atoms that are not part of the silicate network have then been respectively differentiated into oxygen in water molecules (Ow) and hydroxyl group (Oh) when they show two and one hydrogen atom in their O-H FCS, respectively.

Simulated Annealing

Following the topological analysis, all samples are transferred to the non-reactive environment of CSH-FF potential. Since the conditions at the reactive simulations were performed at elevated temperature, simulated annealing technique (SAT) was used to lower the temperature to that of ambient conditions (Supplementary Figure 5). During SAT, each sample is relaxed for 1 ns at 500 K using CSH-FF potential in isobaric-

isothermal ensemble with time steps of 1 fs. All samples are subsequently quenched over 1 ns to 300K using 100 annealing steps. Finally, each sample is further relaxed in NPT ensemble for another 1 ns at 300K. The temperature profile during the quenching stage is provided in Supplementary Figure 5.A. The simulation box sizes were tracked to ensure proper convergence of SAT (Supplementary Figure 5.B and 5.C). We note that the most of box length adjustments occur during the model relaxation at 500 K and not during the quenching stage. These slight adjustments are mainly due to the discrepancies between the ReaxFF and CSH-FF potentials. However, as shown in Supplementary Figure 5.A-C, these differences are negligible.

Calculation of Elastic Properties

After the simulated annealing stage, all models are further relaxed for 1 ns at 300 K. Next, 10 independent frames each 100 ps apart are consecutively extracted from MD trajectories. This procedure is strictly followed to ensure proper unbiased statistics for each sample. Afterwards, all samples are relaxed via energy minimization at constant volume (CONV)[31,32] with rational functional optimization (RFO)[31,32] module for residuals less than 10^{-3}. This ensures that the Hessian matrix is positive definite, i.e. all the eigenvalues of the Hessian matrix are positive. All the relaxed instances of all C-S-H models are further relaxed via energy minimization at constant pressure (CONP)[31,32] with RFO option. Through this multi-step scheme, it is made sure that all the available degrees of freedom, both the atomic positions and the unit cell, are properly relaxed. Elastic properties were calculated for 1500 configurations along this MD relaxation trajectory, at 0 K. Each of the 10 frames contained 150 of these configurations, providing a mean value for each frame. These 10 points were averaged for the M shown in Figs. 3 and 4 of the manuscript. The stiffness tensor, C_{ijkl} in tensor notation and C_{ij} in Voigt notation, is subsequently calculated from the second derivative of the energy density.

$$C_{ijkl} = \frac{1}{V}\left(\frac{\partial^2 E}{\partial \epsilon_{ij} \partial \epsilon_{kl}}\right) \tag{12}$$

where E and V are internal energy and volume of the unit cell, respectively. ϵ_{ij} denotes the strain tensor. It should be noted that positive-definiteness of Hessian matrix entails

positive-definiteness of the stiffness tensor. A wide range of elastic constants can be calculated from stiffness and compliance ($S=C^{-1}$) tensors. Two of the commonly used isotropic measures of the stiffness are bulk (K) and shear (G) moduli. The Voight-Reuss-Hill bounds of bulk and shear modulus are given by:

$$K_{Voight} = \frac{1}{9}\left(C_{11} + C_{22} + C_{33} + 2(C_{12} + C_{13} + C_{23})\right) \quad (13)$$

$$K_{Reuss} = \left(S_{11} + S_{22} + S_{33} + 2(S_{12} + S_{13} + S_{23})\right) \quad (14)$$

$$G_{Voight} = \frac{1}{15}(C_{11} + C_{22} + C_{33} + 3(C_{44} + C_{55} + C_{66}) - C_{12} - C_{13} - C_{23}) \quad (15)$$

$$G_{Reuss} = \frac{15}{4(S_{11}+S_{22}+S_{33}-S_{12}-S_{13}-S_{23})+3(S_{44}+S_{55}+S_{66})} \quad (16)$$

The Hill bounds are simply the average of Voight and Reuss bounds, which are called Voight-Reuss-Hill bounds[31]. The indentation modulus (M) can be directly calculated from bulk and shear modulus:

$$M = 4G\frac{3K+G}{3K+4G} \quad (17)$$

In fact these bounds are isotropic descriptions of a generally anisotropic medium. A wide range of materials in nature including clay [42] and Tobermorite minerals [17] exhibit anisotropic behavior due to their molecular structure.

There are numerous approaches to quantify anisotropy at nanoscale. The stiffness of materials under the uniaxial tension is called Young Modulus. In anisotropic materials, the Young's Modulus in different directions is different in uniaxial deformation path ($Y_i = d\sigma_{ii}/d\epsilon_{ii}$). Another method to assess the level of elastic anisotropy in C-S-H is through the measurement of orthotropic indentation modulus in the plane and perpendicular to the calcium-silicate layers. Following the derivation of Delafargue and Ulm [43]:

$$M_1 = \sqrt{\frac{C_{11}C_{33}-C_{13}^2}{C_{11}}\left(\frac{1}{C_{44}} + \frac{2}{\sqrt{C_{11}C_{33}}+C_{13}}\right)^{-1}} \quad (18)$$

$$M_3 = \sqrt{\sqrt{\frac{C_{11}}{C_{33}}}\frac{C_{11}^2 - C_{22}^2}{C_{11}} M_1} \quad (19)$$

One can think of $|M_1\text{-}M_3|/M_1$ as a measure of anisotropy in the structure of materials. There are many other ways to assess the level of anisotropy in materials such as Thompson index based on anisotropy in acoustic velocities of materials [44]. In this work, C^{iso} is mathematically defined by enforcing isotropic conditions on a general stiffness tensor:

$$C_{11}^{iso} = C_{22}^{iso} = C_{33}^{iso} = \frac{C_{11}+C_{22}+C_{33}}{3} \quad (20)$$

$$C_{44}^{iso} = C_{55}^{iso} = C_{66}^{iso} = \frac{C_{44}+C_{55}+C_{66}}{3} \quad (21)$$

$$C_{12}^{iso} = C_{12}^{iso} = C_{23}^{iso} = C_{11}^{iso} - 2C_{44}^{iso} \quad (22)$$

The rest of the entries in the C^{iso} are set to zero. A new index for anisotropy is defined using Euclidean, d_E, and Riemannian, d_R, Norms of a general stiffness tensor from its isotropic part, C^{iso}, defined as:

$$d_E(C^{iso}, C) = \left\|C^{iso}, C\right\|_E = \sqrt{tr\left((C^{iso} - C)^T (C^{iso} - C)\right)} \quad (23)$$

$$d_R(C^{iso}, C) = \left\|ln\left(C^{\frac{1}{2}}(C^{iso})^{-1}C^{\frac{1}{2}}\right)\right\|_E = \sqrt{\sum_{i=1}^{n} \ln^2 \lambda_i} \quad (24)$$

where λ denotes set of eigenvalues and *tr()* stands for the trace of the matrix.

Hardness Prediction via Biaxial Deformation:

In this work, a new method is designed for assessment of hardness at nano-scale. As in the calculation of M, hardness H was calculated in a non-reactive potential and at 0 K. The notion behind this approach is originated from laboratory tri-axial testing of materials. Very similar to tri-axial testing, the atomistic simulation cell can be isotropically deformed in different deformation paths. Due to the application of periodic boundary conditions, the simulation cell is representative of the bulk of the material. There are six independent deformation paths to deform a material, which consists of three normal, and three shear paths[45]. In continuum mechanics, the deformation gradient tensor, **F**, describes the state of deformation in the deformed configuration compared to the undeformed body as:

$$dx = F_{ij}dX \tag{25}$$

where *dX* and *dx* denote the infinitesimal vectors in undeformed and deformed configurations, respectively. The Green-Lagrange strain tensor is a rotation-independent measure of local relative deformation defined as

$$\epsilon_{ij} = \frac{1}{2}\left(F_{ki}F_{kj} - \delta_{ij}\right) \tag{26}$$

where δ_{ij} is the Kronecker delta. Following the assumption of Cauchy-Born, it is assumed that the unit cell of the material at the atomic scale deforms with the macroscopic deformation gradient[46–49]. In other words, the deformation at the molecular scale can be stated as:

$$a_i = F_{ij}A_j \tag{27}$$

where *A* and *a* denote the lattice vectors of the original undeformed lattice and the deformed structure, respectively. This entails an approximation due to the finite size of the lattice. In fact, Cauchy-Born rule implies that the macroscopic deformation gradient minimizes the free energy, Ψ, of the lattice at atomic-scale. In solid mechanics, the Cauchy stress is defined as:

$$\sigma_{ij}^{continuum} = \frac{d\Psi}{d\epsilon_{ij}} \tag{28}$$

The above formulation defines the stress tensor in a hyperplastic sense for energy-conserving solid undergoing either linearly or non-linearly elastic deformation. As defined in thermodynamics, the free energy, $\Psi = E - TS$, is the sum of molecular interaction and entropic contributions where T and S are temperature and entropy of the system. At considerably low temperatures (below Debye temperature), both configurational and vibrational entropy can be assumed to be negligible. Therefore, the stress can be measured as the derivative of internal energy with respect to strain. In

molecular mechanics, the internal energy can be stated as a function describing interaction between particles. This gives rise to a semi-analytical formulation for stress tensor at the atomic scale known as Virial stress:

$$\sigma_{ij}^{Virial} = \langle \frac{1}{\Omega} \sum_k \sum_l \left(r_{kl}^i f_{kl}^j \right) \rangle \tag{29}$$

where Ω is the unit cell volume, r_{kl}^i and f_{kl}^j is the distance and pair-wise force between the pair of k^{th} and l^{th} atom in the i^{th} direction. The <> sign represents the ensemble average which is extensible to the time average under the ergodic assumption. Considering kinematic equations, eqs. (26) and (27), along with Virial stress , eq. (29), the stress-strain relationship at the scale of unit cell can be described via atomistic simulation. The two-body form of Wolf summation makes it possible to calculate stress for ionic systems. The concise theory presented above builds the foundation for measuring hardness at the atomic-scale.

Due to transversely isotropic nature of C-S-H, especially at low Ca/Si ratios, the direction perpendicular to the calcium silicate layers, z direction, shows lower strength. This is directly attributed to the lack of covalent bonding, Q_3 and Q_4 sites, bridging between the layers. This suggests that the failure analysis should be conducted in ϵ_{zz} and the two other relevant shear strains ϵ_{xz} and ϵ_{yz}. The response of the C-S-H against the two shear paths is almost identical as they both describe shear localization at the interlayer spacing. Therefore, the six dimensional strain spaces reduced effectively to a two dimensional space. This considerably reduces the number of computations associated with calculation of hardness. Subsequently as displayed in Supplementary Figure 6.A, eleven independent deformation paths are considered in $\epsilon_{zz} - \epsilon_{yz}$ space. Each line is a linear combination of ϵ_{zz} and ϵ_{yz}. The normal strain, ϵ_{zz}, covers the range of -0.2 to 0.2, which includes both tensile and compressive modes. In fact, mutual application of shear deformation along with tensile or compressive deformation resonates the confinement effect at the atomic-scale. This mechanism is at a level below the scale of friction and wear, describing how solid bodies are rubbed against each other.

For each of the 150 C-S-H structures considered, six different configurations were selected from the MD relaxation trajectory. Each of the six configurations was spaced at least 100 ps apart along this 1 ns trajectory. Each configuration was deformed in strain-control following the above 11 deformation paths to obtain a calculated hardness *H*. This resulted in 150 mean values of hardness, one for each C-S-H structure that represented the mean hardness computed among the six configurations for that structure. To eliminate the rate effect, all simulations are performed using energy minimization approach. An extensive convergence analysis is performed to identify the size of the strain increments. It is found that strain steps as large as 10^{-3} are somehow large enforcing the stress path to deviate from the equilibrium deformation path. This leads to improper identification of the failure point. Therefore, the strain increment in all simulations is set to 10^{-4}.

The stress-strain response is plotted accordingly for each individual deformation test. Supplementary Figure 6.A-B presents the stress-strain diagram in deformation path 4 and 6, respectively. Supplementary Figures 6.A-B are instances of C-S-H under biaxial tensile-shear and compressive-shear mixed-mode failure analysis, respectively. Both tensile and shear modes exhibit linear elastic response before the failure point. As indicated by circles locating the failure point, upon the irreversible deformation at the atomic scale, the stress relaxes in both deformation modes simultaneously. This is an indicative of plastic deformation, which can be verified by unloading the sample before and after the yielding point and identifying the extent of residual strain by unloading the sample to its original state. As illustrated in Supplementary Figure 6.D, the failure analysis identifies the strain, $(\epsilon_{zz}^*, \epsilon_{yz}^*)$, and stress $(\sigma_{zz}^*, \sigma_{yz}^*)$ at which the material fails upon loading in a particular deformation path. Identification of failure is certainly subjected to user-bias if performed manually. To eliminate such an effect from the simulation results, a simple conditional statement is implemented such that if the slope of stress-strain curve falls below a certain threshold, that point is taken as the failure limit. Since the deformation is confined in the other axial directions and because of Poisson effect, a planar stress evolves in the material while undergoing constrained biaxial deformation. The planar stress is direction-independent, $\sigma_{xx}^* = \sigma_{yy}^* = \sigma_{pl}^*$, due to isotropy

in calcium-silicate sheets. Therefore, the stress tensor at failure point, $\boldsymbol{\sigma}^*$, can be viewed as:

$$\boldsymbol{\sigma}^* = \begin{bmatrix} \sigma_{pl}^* & 0 & 0 \\ 0 & \sigma_{pl}^* & \sigma_{yz}^* \\ 0 & \sigma_{yz}^* & \sigma_{zz}^* \end{bmatrix} \tag{30}$$

The failure stress tensor can be stated in terms of principal failure stress components $(\sigma_I^* = \sigma_{II}^* = \sigma_{III}^*)$ by simply finding the eigenvalues of failure stress tensor. Therefore, the principal stresses can be written as:

$$\sigma_I^* = \sigma_{pl}^* \tag{31}$$

$$\sigma_{II}^* = \frac{1}{2}\left(-\sqrt{\left(\sigma_{pl}^* - \sigma_{zz}^*\right)^2 + 4{\sigma_{pl}^*}^2} + \sigma_{pl}^* - \sigma_{zz}^*\right) \tag{32}$$

$$\sigma_{III}^* = \frac{1}{2}\left(\sqrt{\left(\sigma_{pl}^* - \sigma_{zz}^*\right)^2 + 4{\sigma_{pl}^*}^2} + \sigma_{pl}^* - \sigma_{zz}^*\right) \tag{33}$$

The transformation of normal and shear stresses to the principal stress components is graphically represented in the inset of Supplementary Figure 6.D. Using the Mohr representation, the principal components of the stress can be identified as the intersection of Mohr circles with σ-axis. The maximum and minimum of principal stresses for k^{th} deformation path are Σ_I^k and Σ_{II}^k, respectively. These two principal stresses and enclosing Mohr circles define the limit stress state at which atomic structures of C-S-H undergoes an irreversible deformation. As shown in Supplementary Figure 6.D, a set of Mohr failure circles on independent deformation paths can define the failure mechanism and the elastic limit of the material. Supplementary Figure 6.D denotes that C-S-H at the atomic scale follows a Mohr-Coulomb-like failure criterion. Application of Mohr-Coulomb in mechanics of granular materials is prevalent due to proper description of inter-granular friction. However, the increase in strength of C-S-H at compressive domain is attributed to the confinement effect. There is no dissipative mechanism such as friction or wear at the molecular scale and the increase in the strength is mainly because of elastic confinement effect making the failure harder for materials at high pressures. Particularly, in the case of C-S-H, the failure envelope is linear in the stress domains related to nano-

indentation of cementitious materials. Therefore, C-S-H at the nano-scale can be described as a cohesive-frictional material identified via cohesion (C) constant and friction angle (ϕ). Cohesion is the strength of the C-S-H at zero confinement pressure. Following the work of Ganneau et al. [50], the hardness (H) of a cohesive-frictional material can be stated as:

$$\frac{H}{C} = \frac{\delta(\phi,\theta)}{tan\phi} = \frac{1}{tan\phi}\sum_{m=1}^{N}(a_k(\theta)tan\phi)^k \quad (34)$$

where θ is the indenter apex angle and a_k coefficients are fitting parameter for given indenter geometry. For the special case of C-S-H materials, the friction angle is on the order of 6-9 degrees in which tan (ϕ) is quite negligible and as a result, the hardness can be approximated with 5.8C. In current work, friction angle, cohesion and hardness were independently measured at 6 independent points for each sample. 6 different configurations selected from the MD relaxation trajectory for each polymorph; this is fewer than the 1500 calculations of M for each polymorph due to the higher computational cost associated with this hardness calculation. The reported hardness and the error bars are the mean and standard deviations for each sample. See supplementary figures 14-16 for strain-stress curves of the three polymorphs exhibiting the highest M/H in Fig. 4 of the manuscript.

Phonons and Free Energy Calculations

Due to vibrational entropy at finite temperature in solids, the atoms constantly oscillate around their equilibrium position. At low temperatures, these oscillations can be decomposed to a linear superposition of vibrational modes known as '*phonons*'. In the case of infinite 3D solids, there are infinite numbers of phonons. The phonons are characterized by their values at a point in the reciprocal space, usually called k-point, within the first Brillouin zone. Hence, there are 3N phonons per k-point. The lowest three phonons are known as acoustic branches, which tend to zero at the center of Brillouin zone, known as Γ-point. The rest of the phonons belong to the optical branches, which have a finite frequency at the Γ-point. Identification of phonons in a 3D periodic solid involves numerical calculation of the eigenvalues of dynamical matrix at each k-point:

$$D_{i\alpha j\beta} = \frac{1}{\sqrt{m_i m_j}} \sum_R \left(\frac{\partial^2 U}{\partial\alpha\ \partial\beta}\right) exp\left(ik\left(r_{ij} + R\right)\right) \quad (35)$$

where m is the mass of atom, i and j indices denote two arbitrary atoms within the cut-off radius, α and β show the components in the real space, k is the reciprocal vector and R represents the sum over lattice vectors within the cut-off radius. Having phonons in hand, statistical mechanics can be employed as a powerful tool to bridge microscopic energy levels to macroscopic thermodynamical observables such as entropy, Helmholtz and Gibbs free energies and many others. The ubiquitous assumption in molecular physics is that all forms of the energy are independent and as a result the total energy can be decomposed to translational, rotational, vibrational and the contributions to the electronic state of the system. In the special case of solids at low temperatures (considerably below the melting point), the rotational and translational contributions can be neglected. In atomistic simulations, the electronic contribution is directly calculated via interatomic force field. Therefore, the challenge simplifies to the measurement of vibrational contributions. This is achieved via calculation of partition function, Z_{vib}, in the framework of lattice harmonic approximation. The vibrational energy level for a harmonic oscillator of the m^{th} mode can be written as

$$U_{vib}^{m}(n,k) = \left(n + \frac{1}{2}\right)\hbar\omega(m,k) \quad (36)$$

where n is the number of phonons occupying that particular energy level. The number of phonons occupying an energy state is given by Bose-Einstein statistics:

$$n = \frac{1}{\exp\left(-\frac{\hbar\omega}{k_B T}\right) - 1} \quad (37)$$

where T is the temperature and k_B is Boltzman constant. The internal vibrational energy, U_{vib} can be written as the superposition of all the energy levels by considering that the energy at each state is the number of phonons occupying that state times the energy of each phonon:

$$U_{vib} = \sum_m \sum_k w_k \left(\frac{1}{2} \hbar\omega + \frac{\hbar\omega}{\exp\left(-\frac{\hbar\omega}{k_B T}\right)-1} \right) \quad (38)$$

where w_k represents the weight of a particular point in the Brillouin zone such that the sum of the weights equal one. To calculate the internal energy, the summation should be over all oscillation modes, *m*, and reciprocal lattice vectors, *k*. It should be noted that the fist term in the internal energy is just due to zero point vibrational energy. The partition function in canonical ensemble can be written as:

$$Z_{vib} = \sum_m \sum_k w_k \frac{\exp\left(-\frac{\hbar\omega}{k_B T}\right)}{\exp\left(-\frac{\hbar\omega}{k_B T}\right)-1} \quad (39)$$

It should be emphasized that the frequencies are explicit function of *k* and *m*, which are omitted from the above equation for brevity. The Helmholtz free energy, Ψ_{vib}, and Gibbs free energy, $\mathcal{L}_{vib}$, are readily calculated using:

$$\Psi_{vib} = \sum_m \sum_k w_k \left(\frac{1}{2} \hbar\omega + ln\left(1 - \exp\left(-\frac{\hbar\omega}{k_B T}\right)\right) \right) \quad (40)$$

$$\mathcal{L}_{vib} = \Psi_{vib} + PV \quad (41)$$

where *P* and *V* are pressure and volume of the unit cell, respectively. In practice, the above double summation is reduced to integration over the phonon density of states (DOS). Generally, the integration involves a numerical scheme over a discrete set of points. Due to the lack of symmetry in C-S-H's molecular structures, the most general set of points in the first Brillouin zone known as Monkhorst-Pack scheme is used in this work. A set of 1099 evenly spaced mesh of k-points, given by shrinking factors along each axis in reciprocal space, is adopted in all the simulations. Subsequently, the DOS is distributed over 60 equal bins and all the thermodynamic observables are calculated following the above-mentioned approach. The free energy as a function of C/S is presented in Supplementary Figure 8.A. This shows that the free energy content decreases with C/S. Supplementary Figure 8.B-Ds present the correlation between free

energy and mechanical response of the C-S-H (M, H and M/H). They all show that at a given free energy level, the mechanical properties can vary which is a clear sign of polymorphism in C-S-H molecular structures.

Medium Range Order Analysis

As presented before, radial distribution functions (RDF) give insights into the short-range order of the structure of the material, this length scale being mostly driven by coordination numbers and bond lengths. As C-S-H polymorphs are characterized by the same short range order but show different densities, we analyzed their medium range order to understand the origin of this density variability. To do so, we largely relied on tools introduced in glass science. The medium range order can be captured by the partial structure factors, which can be computed from the RDFs $g_{ij}(r)$[12,51,52]:

$$S_{ij} = 1 + \rho_0 \int_0^R 4\pi r^2 \left(g_{ij}(r) - 1\right) \frac{sin(Qr)}{Qr} F_L dr \quad (42)$$

where $Q=2\pi/r$ is the scattering length, ρ_0 is the average atom number density and R is the maximum value of the integration in real space (here $R=6$ Å). The $F_L(r)=sin(\pi r/R)/(\pi r/R)$ term is a Lortch-type window function used to reduce the effect of the finite cutoff of r in the integration [53]. As discussed elsewhere [54], the use of this function reduces the ripples at low Q but can induce a broadening of the structure factor peaks. Supplementary Figure 8 shows Si-O, Si-Si, Ca-O and H-O partial structure factors at Ca/Si=1.7, which are by all means similar to typical structure factors observed in silicate glasses [54].

Among the different features of structure factors, the so-called first sharp diffraction peak (*FSDP*) has received a lot of attention in glass science. Starting from the principal peak, which position $Q_P=2\pi/r_P$ is associated to the nearest-neighbor distance r_P, the *FSDP* is defined as the next peak located at lower $Q < Q_P$. It is thus related to the structural correlations in the medium range order at $r > r_p$. This feature has been observed both in

covalent[55,56] and ionic[57] amorphous system. In ionic systems, this medium-range order has been associated to the forced separation between cations because of their mutual Coulomb repulsion, thus producing a pre-peak in the cation-cation structure factor[58]. Pre-peaks can also arise from size effects of the atoms of the network[59]. However, the network formation itself can have a major role since the *FSDP* is also observed in the monoatomic tetravalent systems a-Si and a-Ge [60,61]. The *FSDP* origin is now usually explained by using a void-based model[62,63] in which ordering of interstitial voids occurs in the structure. It is worth noticing that FSDPs are not simply the first of the many peaks of any diffraction pattern but display many anomalous behaviors as a function of temperature, pressure and composition[62,64].

The *FSDPs* we obtained from simulations were further studied by fitting them with Lorentzian functions (example of fitted function can be seen on Supplementary Figure 8). This choice is supported by the fact that the experimental results in neutron scattering factor of silica can be better fitted with a Lorentzian function than with a Gaussian one[65]. It should be noted that the fit has been done on the low Q part of the *FSDP* to avoid the contribution of the following peaks. This allows us to track precisely the *FSDP* peak intensity, position and full-width at half maximum (*FWHM*) for each sample.

Coming back to the real space structural correlations, the *FSDP* peak position Q_{FSDP} is usually related to a characteristic repetition distance $d=2\pi /Q_{FSDP}$ and the *FWHM* to a correlation length $L=2\pi/FWHM$, sometimes also called coherence length, due to atomic density fluctuations [66]. The effect of irradiation [65,67], water content [66,68] and alkali content [54] on the *FSDP* have been studied, leading to the idea that a depolymerization of the network (a decrease of the atomic order) is associated to a decrease of the intensity of the FSDP and a decrease in the characteristic distance [69]. A global understanding of the correlation length L is still lacking and seems to be highly system dependent.

Calculation of X-ray total pair correlation from atomistic simulation

We calculated the total X-ray pair correlation function $N_x(r)$ from atomistic simulation. $N_x(r)$ is related to the total X-ray pair distribution function $G_x(r)$ via:

$$N_x(r) = 4\pi r^2 \rho_0 G_x(\mathrm{r}) \tag{43}$$

where ρ_0 is the average number density in the medium. $G_x(r)$ is related to the partial pair correlation functions $g_{ij}(r)$ where i and j indicate different species (Ca, Si, O and H) in the simulation:

$$G_x(r) = \sum_{ij} \frac{c_i c_j f_i(Q) f_j(Q)}{\langle f(Q) \rangle} g_{ij}(r), \quad \langle f(Q) \rangle = [\sum_i c_i f_i(Q)]^2 \tag{44}$$

where c_i is the concentration of i^{th} specie, $f_i(Q)$ is the atomic form factor describing the Q-dependent scattering length for the i^{th} specie[70]. The atomic form factor is calculated using:

$$f(Q) = \sum_{i=1}^{4} a_i exp\left(-b_i \left(\frac{q}{4\pi}\right)^2\right) + c \tag{45}$$

where a_i, b_i and c are element-dependent form factor constants. Due to the broadening effect induced by termination of X-ray structure factor at high Q values in experiments, we follow the suggestion of Wright[53,65,71] to calculate the broadened molecular dynamics response $G_x^b(r)$ as follow:

$$G_x^b(r) = \int_{-\infty}^{\infty} G_x(x) * H(r - x)\, dx \tag{46}$$

where H is a Gaussian with standard deviation set to 0.055. We use $G_x^b(r)$ to calculate $N_x(r)$ from eq. (43). Subsequently, we calculated the coordination number of Ca atoms in the simulation using:

$$N_{Ca} = \int_0^{r_{min}} 4\pi r^2 \rho_O g_{CaO}(r) dr \tag{47}$$

where ρ_O is the average number density of oxygen atoms in the simulation box.

Materials and Sample Synthesis

Several cement based systems with different Ca/Si ratio were synthesized and characterized; eight samples are reported in Figs. 3-4. Two cement pastes with C-S-H of Ca/Si ≈ 1.75 were based on the ordinary Portland cement Class I, with water-to-cement ratio [0.4, 0.5, 0.6] and cured at room temperature. Two cement pastes with C-S-H with lower calcium content (Ca/Si ≈ 1) were synthesized from a mixture of Portland cement Class G and crystalline silica, under hydrothermal conditions at the temperatures 130°C, 175°C and 200°C and a pressure of 3000 psi. For these samples, silica flour was added in the amount of 35% by weight of cement, at the water-to-cement ratio 0.45.Two cement pastes with C-S-H of lower calcium content (Ca/Si ~0.75) were obtained via synthesis of Ductal®, a proprietary (Lafarge) ultra-high performance concrete. Two cement pastes with higher calcium content (Ca/Si ~2) were obtained by varying the clinker type, initial pH of the water-based electrolyte, curing temperature, and potential additives. Specifically, Ca/Si was 2.10 for cement paste synthesized from white Portland cement (Lehigh-Heidelberg Cement Group, Brooklyn, NY) at water/cement ratio of 0.4 and initial pH of 7, and a curing temperature of 60^{o}C; and Ca/Si was 1.89 for pure tricalcium silicate clinker at water/cement of 0.4 and initial pH of 4, high-carbon fly ash addition (Brayton Point; Headwater Resources, Taylorsville, GA), and a curing temperature of 5^{o}C.

Surface Preparation Protocol

The nanoindentation technique used requires a flat polish of the sample surface to minimize the role of surface geometry on measured mechanical properties. This was achieved using a semi-automatic polishing procedure developed and successfully applied to each specimen. In the first pass, cylindrical sub-samples, ≈1.0 cm in diameter and 0.5 cm in height, were cut from each bulk sample, with the use of the low speed diamond saw IsoMet, Buehler (Lake Bluff, USA). The cylindrical samples where then mounted separately with cyanoacrylate glue on the standard AFM base (Ted Pella, USA), to be

later automatically leveled and ground in a specially designed jig and the post system. Details of this system may be found elsewhere [72,73]. The sample, inside the jig, was ground on a hard perforated cloth TexMetP (Buehler, USA) with a 9 μm oil-based diamond suspension MetaDi (Buehler, USA). In case of high-temperature cured specimens, the jig was loaded with an additional mass of 150 g. Grinding pass was performed for approximately 20 min. This preparation time does not include the 3 [min] ultrasonic cleaning with n-Decane, (TCI, Tokyo, Japan) which is performed after every 5 [min] of grinding. In the second step, the ground and cleaned samples were manually polished on the alumina abrasive disks FibrMet, Buehler (Lake Bluff, USA), with the respective reduced abrasive sizes 9, 3 and 1 μm. At each pad of different abrasive size the samples were polished for around 5 min. In this step the used pad was exchanged for a fresh one every 2.5 min. These changes of the pad, as well as abrasive size, were accompanied by 1 min cleaning of the sample in the ultrasonic bath with n-Decane. So obtained final surface was characterized by rare occurrence of the scratches on its surface, visible under 20x and 50x objective of optical microscope, and high mirror-like reflectivity of the surface.

Nuclear Magnetic Resonance (NMR) to estimate silicate chain length

We used ^{1}H-^{29}Si cross polarization (CP) magic angle spinning (MAS) NMR spectroscopy to estimate the silicate chain length in C-S-H. In cementitious materials, the ^{29}Si chemical shift depends predominantly on the atomic environment of the silicon- oxygen tetrahedra, denoted by Q_n, where n represents the number of bridging oxygen atoms [14]. Note that dimers consist of two Q_1 groups. With the proportion V_1 of different Q_n groups in a cement paste, one can determine the average silicate chain length l_{Si} as $2\ (1 + V_2/V_1)$. Thus, if the proportion of silicate dimers is high, or $V_1 >> V_2$, chain length is approximately 2.

CP MAS NMR is a double resonance experiment in which energy is transferred from ^{1}H nucleus to ^{29}Si nucleus, and thus requires them to be in close proximity. This condition is not fulfilled in anhydrous silicates, and hence only silicates from C-S-H should contribute to the NMR spectrum. CP MAS NMR experiments were conducted on two samples at Bruker Biochem Corp., Billerica MA by Dr. J. Struppe. These cement

paste samples were based on either tricalcium silicate-rich clinker or dicalcium silicate-rich clinker, both at water/cement ratio of 0.40 and synthesized by CIMPOR colleagues Drs. J. Pereire and S. Lebreiro. We then deconvoluted the observed spectrum as a sum of two Lorentzian peaks. The proportion of V_n of Q_n groups was then calculated as the ratio of area enclosed under the Lorentzian peak corresponding to Q_n and the total area enclosed by the sum of all Lorentzian peaks. For the two points shown in Fig. 2c of the main text, dimer content was 86% (for l_{Si} = 2.33) and 50% (for l_{Si} = 3.98).

Instrumented Nanoindentation Measurement Protocol

Indentation tests in this study were conducted using either the CSM Nanoindentation platform, equipped with the vibration isolation table as well as an environmental chamber to control testing conditions, temperature (18°C) and relative humidity (RH≈20%), respectively; or the MicroMaterials NanoTest600 platform, also with vibration isolation and environmental control of temperature and humidity. All indentations were conducted using the Berkovich diamond indenter, a three-sided pyramid with the half angle of 65.35°. Certified fused silica sample, with an elastic modulus of 73.1±0.7 GPa was used to calibrate the tip area function in the load range 0.5 – 100 mN. The indentation protocol consisted of three phases: a loading segment to a maximum load of 2 mN at a rate of 8 mN/sec; a holding period of 5 sec at maximum load; and an unloading segment at 8 mN/sec. Measurements on each sample were acquired in a square grid array, with at least 400 points within each grid and the grid spacing large enough for the successive measurements to be considered as statistically independent based on prior works[74,75]. Here, this spacing was typically 20 μm center-center spacing between adjacent indentations.

Indentation elastic modulus M and indentation hardness H were determined from the measured load-depth responses according to the method proposed by Oliver and Pharr [76,77], with each response corrected for point of contact and frame compliance. In other contexts, indentation is used to infer Young's elastic modulus E of the sample, and this calculation requires assumption or measurement of the unitless Poisson's ratio ν. We define indentation modulus M as $E/1-\nu^2$; in this way, we do not assume or measure ν and

simply report a stiffness that is related to the more commonly reported E by this undetermined elastic property.

Statistical Analysis of the Indentation Data

Statistical analysis of the indentation data was carried out in the framework of Finite Mixture Models (see Eq. 48) with the component probability densities of the mixture assumed to be of Gaussian type (Eq. 48-50) [78,79].

$$f(\boldsymbol{y_j};\boldsymbol{\Psi}) = \sum_{i=1}^{g} \pi_i f_i(\boldsymbol{y}_j;\boldsymbol{\theta}_i), \qquad 0 \le \pi_i \le 1\ (i = 0, \dots, g), \qquad \sum_{i=1}^{g} \pi_i = 1 \tag{48}$$

$$f_i(\boldsymbol{y}_j;\boldsymbol{\theta}_i) = \phi(\boldsymbol{y}_j;\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i) \tag{49}$$

$$\phi(\boldsymbol{y}_j;\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i) = (2\pi_i)^{-\frac{p}{2}} |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} exp\,\{-\tfrac{1}{2}(\boldsymbol{y}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_j - \boldsymbol{\mu}_i)\} \tag{50}$$

Therefore, by design of the indentation experiment, at each location on the indentation grid the pair of properties under study is measured $\boldsymbol{y}_j = (H_j, M_j)$. where $j = 1, \dots, N$ with N the total number of measurement points. A possible outcome of single realization of random variable $\boldsymbol{Y}$ is $\boldsymbol{y}_j$ belonging to one of g mechanically active phases within material microstructure e.g., residual clinker grains, calcium-silica-hydrate gel (C-S-H). Therefore each of such phases has a uniquely defined pair of indentation hardness and modulus represented by the mean vector of properties $\mu_i = (H_i, M_i)$, with $i = 1, \dots, g$. Additionally, in the framework of this statistical model, variances on both phase properties as well as correlations among them are captured in the covariance matrices Σ_i. The last parameter of the model, π_i, represents the surface fraction of a given phase. All unknown parameters $\boldsymbol{\Psi} = (\pi_1, \dots, \pi_g, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$ of the assumed GMM model are estimated with the maximum likelihood (ML) fitting approach via the expectation maximization (EM) algorithm [80]. The estimation of model parameters as well as the clustering of observations based on the Bayes rule of allocation is carried out with the use

of the freely available program, EMMIX (http://www.maths.uq.edu.au/~gjm/emmix/emmix.html).

Estimation of the parameters of the assumed Gaussian Mixture Model requires *a priori* selection of the number of mixture components g. In this study, the selection of model order, which corresponds to the number of mechanically active phases is assessed with Bayesian Information Criterion (BIC), Eq. (48-50) [79,81]. Each grid indentation data set is analyzed for the range of components $g = 1, \dots, 8$, and the one which maximizes penalized log likelihood statistics is selected as optimum fit; k_g is the number of model parameters and $L_g(\widehat{\boldsymbol{\Psi}})$ corresponds to the likelihood of a model with g components.

$$\log L_g(\widehat{\boldsymbol{\Psi}}) - \frac{1}{2} k_g log N \tag{51}$$

A typical example of the applied statistical data analysis in the GMM framework and the inference of the optimum model are presented on the Supplementary Figure 9. Presented indentation results were obtained on the ordinary Portland cement paste with $w/c = 0.45$.

Electron Microprobe Measurement Protocol

Ca/Si for the C-S-H within each sample was determined via electron microprobe analysis (EMPA), via wavelength dispersive spectroscopy (WDS). Back-scattered electron (BSE) images, elemental X-ray maps, and chemical compositions of phases were obtained by a JEOL JXA-8200 Superprobe electron probe micro-analyzer (EPMA). The polished samples were analyzed by Wavelength Dispersive Spectrometry (WDS) using an accelerating voltage of 15 keV, a beam current of 10 nA, a 1 μm beam diameter, and 10 s counting times per element with background correction at every point. A square grid of points was collected for each sample. As in the case of instrumented indentation, there were approximately points per sample that were each spaced by approximately 20 μm. The raw data were corrected for matrix effects with the CITZAF package [82] using the ϕ(ρz)-method (stopping power, back-scatter corrections and ϕ(0) from Love and Scott,

mean ionization potential from Berger and Seltzer, Heinrich's tabulation of mass absorption coefficients, and Armstrong's modified equations for absorption and fluorescence corrections [83]. The set of measured elements, their associated standard materials, and the diffraction crystals are presented in Supplementary Table 5.

The obtained experimental data were statistically analyzed in the framework of Gaussian Mixture Models (§2.3, Eqns. 48-50), as in the case of the nanoindentation results. However, contrary to the indentation data analysis results, chemical phases were inferred assuming a multivariate model with the principal variables being four elements dominant in concentration: {Ca;Si;Al;Fe}. Such selection of variables provided sufficient "statistical contrast" for the chemical phases to be accurately inferred and clustered on a probabilistic basis. An example of electron microprobe data analysis in the assumed framework, obtained on the system cement Class G and crystalline silica in the amount 35% by weight of cement, is presented in Supplementary Figure 10.

Assumed Micromechanics Model of Cement-Based Matrix, and Fitting Procedure

The intrinsic indentation elastic modulus *M* and hardness *H* of monolithic calcium-silicate-hydrate were determined from an indentation experiment using a micromechanics model based on the assumption of a granular solid (Supplementary Figure 11[84]). Below, we refer to that monolithic version as the "solid" to distinguish it from the porous phases that are indented, and to compare it to the monolithic versions of C-S-H that were simulated. The porous phases that were investigated experimentally can be envisioned to include this solid in the form of particles that are arranged at a particular packing density. The relationship between the phase hardness, H, measured in an indentation test on a porous solid at the scale of a RVE, and the intrinsic properties of the isotropic solid, cohesion c_s and friction coefficient (friction angle) α, is given by:

$$\frac{H}{h_s(c_s,\alpha)} = \Pi_h(\alpha,\eta,\eta_0) \tag{52}$$

The dependence of H on the local packing density η, percolation threshold η_0, and friction coefficient α, is captured in the dimensionless function Π_h. In the limiting case, $h_s = h_s(c_s, \alpha) = \lim_{\eta \to 1} H$ is the asymptotic hardness of a cohesive-frictional solid phase that obeys the Drucker-Prager criterion [84,85]. The solid hardness relates to the solid's cohesion c_s, and friction coefficient, α, by the model function:

$$h_s(c_s, \alpha) = c_s \times a(1 + b\alpha + (c\alpha)^3 + (d\alpha)^{10}) \tag{53}$$

In eq. 42 a, b, c, and d are the fitted parameters: a=4.7644, b=2.5934, c=2.1860, d=1.6777 respectively. The dimensionless function Π_h depends on the local solid packing density η and the friction coefficient α of the solid, for more details see ref. [84,86].

The indentation modulus M extracted from the indentation test, relates to the elastic modulus of the solid m_s, packing density η of the solid within RVE, percolation threshold η_0, and Poisson's ratio υ_s. This relation is written in the dimensionless form as:

$$\frac{M}{m_s} = \Pi_M(\upsilon_s, \eta, \eta_0) \tag{54}$$

$$\Pi_M(\upsilon_s = 0.2, \eta, \eta_0 = 0.5) = 2\eta - 1 \tag{55}$$

For a perfectly disordered porous material system (Supplementary Figure 12.B), the solid percolation threshold, representing the solid volume fraction required to provide a continuous force path through the system, approaches $\eta_0 = 0.5$. Assuming Poisson's ratio $\upsilon_s = 0.2$, this relationship reduces to linear scaling of the indentation modulus with the packing density Eqn. (55). However, note that we do not assume a value of Poisson's ratio when calculating M (which is simply $E/1\text{-}\nu^2$).

Fitting procedure

The micromechanics model chosen links the measured indentation hardness H and modulus M to the set of three intrinsic properties $\{m_s, c_s, \alpha_s\}$ and packing density distribution η of the solid phase. Therefore, provided the existence of the unique solid C-

S-H phase, the scaling relationships Eqn. (48) and (49) may be used to probe the microstructure sensed in the grid indentation experiment.

The solid properties and the packing density were determined by minimizing the quadratic error (Eqn. (50)) between the experimentally measured values $X = (M, H)$ and the model predictions using the scaling relationships. Because of different orders of magnitude between measured hardness and indentation modulus, the error minimization is carried out in the normalized space with the normalization factors $\max(H, M)$. Such normalization prevents bias in the fitted model toward one of the variables, since errors on both have been assigned equal importance.

$$\min_{\substack{\kappa^S_{X=(m_S,c_S)} \\ \chi^S_{X=(\upsilon_S=0.2,\alpha_S)} \\ \eta_i=1,N}} \sum_{i=1}^{N} \sum_{X=(M,H)} \left(\frac{X_i - \kappa^S_X \Pi_X(\chi^S_X, \eta_i)}{\max(X)} \right)^2 \tag{56}$$

The orthogonal least square fitting algorithm has been coded and executed in the Matlab® environment using the trust-region-reflective algorithm for solving the nonlinear minimization [87]. In order to minimize the risk of obtaining the solution corresponding to local minimum, the minimization problem have been initiated from $n = 50$ random initial vectors of the form $\{m_s^0, c_s^0, \alpha_s^0\}$ drawn from the parameter space $m_s^0 = [0; 140], c_s^0 = [0; 2], \alpha_s^0 = [0; \frac{1}{\sqrt{3}}]$, and the solution with minimum error was selected. The termination tolerance on the objective function value as well as model parameters was set to 10^{-7}. An illustrative example of the results of error minimization procedure is presented in Supplementary Figure 12.

Procedure to identify valid experimental data

In this study, Supplementary Figure 13 was used to identify experimental datasets (chemical and mechanical) that were considered valid comparisons with computational predictions. This procedure recognizes that the mechanical properties of monolithic C-S-H are not measured directly via experiment, but rather inferred by the Gaussian Mixture and micromechanics modeling described above and thus subject to some constraints. This procedure also recognizes that the mechanical properties and chemical properties were not obtained on the exact same grid points on the sample. Rather, a grid at one location of

the sample provided *M* and *H*, and a grid at another location on the sample provided Ca/Si.

This procedure for each sample included acquisition of *M* and *H* for points within an indentation grid and for Ca/Si within an EMPA/WDS grid. Indentation data were clustered to assign points to C-S-H of finite packing density (i.e., not the extrapolation to monolithic C-S-H, but rather what was actually measured at each grid point), and also to other "phases" such as unreacted clinker. The first check of these data was comparison of the *M* and *H* for this porous C-S-H to a database that some of us (K. Krakowiak and F.-J. Ulm) have assembled from prior published reports that also measured *M* and *H* for C-S-H (and other mechanically active phases) at this scale via indentation. If the measured mechanical data for a sample grid agreed poorly with previous reports, the sample was not further analyzed. We recognize this as a potential source of bias.

If the mechanical data agreed with previous reports, the second check was considered. This check was to compare the volume fraction of C-S-H inferred from the mechanical experiments (i.e., related to how many points within the indentation grid were attributed to C-S-H vs. other phases) to the volume fraction estimated from chemical experiments (i.e., the volume fraction of the phase identified as C-S-H via EMPA/WDS). These two experimental techniques have different objectives, chemistry assessment (e.g. Ca/Si ratio) and mechanical properties (H, M). However, the estimated volume fractions of C-S-H obtained by both approaches is reasonably expected to be comparable, because of:

- reliance on probing an area large enough to be considered representative of the bulk (typically a 20 x 20 point grid array with center-center spacing of 20 μm)
- material volume sensed by both techniques are comparable (depths of ≈ 1.5 μm)
- shared technique for statistical modeling and analysis of the experimental data.

To ensure confidence, a threshold of 10% relative difference in the estimated volume fractions of the hydration products by both techniques was enforced. A pair of measurements fulfilling this criterion was considered for the micromechanics analysis that determined *M* and *H* of the monolithic or solid C-S-H. Finally, the third check was whether the error minimization algorithm converged in the estimation of solid properties

via this micromechanics algorithm. The experimental results reported in this work fulfilled these three criteria and are provided in Supplementary Table 6. However, only a subset of data in Supp. Table 6 appear in Figs. 3 and 4, which we discuss below. We note that many of the cement paste samples that we have generated over several years, and then analyzed via instrumented indentation grids and EMPA/WDS grids, did not fulfill at least one of these three criteria.

In fact, even among the eight samples included here and listed in SI Table 6, which were all analyzed according to the protocol outlined in SI Fig. 13, there can be uncertainty as to whether the chemical and mechanical data attributed to C-S-H (as opposed to other phases) is accurate. For example, the sample with Ca/Si = 0.78 passed the first check that microscale *M* and *H* sampled by the indentation grid were within previously reported ranges, and passed the second check that the volume fraction of C-S-H identified via our clustering algorithms was within 10% when estimated from WDS and nanoindentation, and the third check that the fitting solution converged. However, *M* of the monolithic or solid C-S-H extracted from these data exceeded 100 GPa (which was unexpected based on the bounds we imposed on our allowable ranges to microscale porous C-S-H and monolithic solid C-S-H). Also, the Ca/Si ratio of 0.78 had a standard deviation of 0.28, which was higher than that in some other samples. Though *a priori* such results are possible and the sample passed our checks, we note that this sample type is typically rich in unreacted clinker, which may make it challenging to resolve the mechanical properties of C-S-H with this approach in such samples. For similar subjective reasons based on our collective experience in analyzing such samples, results for other cement paste samples may appear in Supporting Figure 17, but not in Figs 3a, c and Fig. 4 of the manuscript. Additionally, one experimental point at C/S = 0.88 appears only in Fig. 3a and not in Figs. 3c or 4 because the mean value of H is higher than the vertical axis limit of Fig. 3c and the mean value of M/H is lower than the vertical axis limit of Fig. 4.

Calculation of uncertainty in M, H, and Ca/Si

In Figures 3 and 4 and Supplementary Table 6, uncertainty is indicated in *M*, *H*, *M/H* or Ca/Si. For M and H we report uncertainty as standard deviations on each parameter

estimated in nonlinear regression analysis. The uncertainty on Ca/Si ratio is reported as sample standard deviation of the experimental measurements attributed to C-S-H product, or dominated by this phase. Additionally, statistical bootstrapping[88] was employed for selected samples to gain insight into the probability distribution of investigated parameters, and estimates of standard errors. Briefly, a measured set of n values of either M_i or H_i (indentation modulus or hardness at a point of the indentation grid on the cement paste) was sampled with replacement in order to create m independent data sets with an equal number of values n. For each bootstrapped sample, the estimate of M or H for the monolithic C-S-H solid particle was obtained via the fitting procedure of the micromechanics model described previously. The number of bootstrapping iterations, m, was typically > 50 in order to assure convergence of the standard deviations. In Figure 4, the uncertainty in the quotient M/H was obtained according to propagation of errors theorem, assuming M and H are statistically independent and normally distributed. The standard deviations are calculated from standard errors by dividing it to the square root of the sample size.

TEM Imaging Conditions

The cement samples are prepared to obtain 30 microns thick blades (stuck on optical glass blades). A micro-pincher (O'Medenbach) is used to drill 3mm diameters cement paste discs. Cu slots for TEM are stuck with a fine layer of araldite glue on these discs. The unit 'disc+slot' is then unstuck from the optical glass. After drying, the unit is subjected to ionic abrasion using a PIPS slimming ionic beam (Gatan) until the complete drilling of the cement in the central zone of the slot. The edges of the home are then ready for TEM Observations. Carbon metallization is not necessary to obtain the electronic conduction of the cement paste samples. Observation are performed on a high resolution field emission gun TEM JEOL 2010F (structural resolution of 0.18 nm, probe size of 0.5 nm, KEVEW EDS system, GIF loss of energy system, vacuum conditions 10^{-6} torr).

Scaling of macroscopic strength with C-S-H strength

Since the beginning of its industrialization in the early 20th century, the strength of concrete (i.e. cement paste plus sand and aggregate) has been recognized to be determined by the strength of the binder phase[89]: the cement paste. In its turn, groundbreaking work by Powers and collaborators at the Portland Cement Association in the 1950s and 1960s[90] established functional relations between cement paste's macroscale, uniaxial compressive strength and porosity (Supplementary Figure 17.a), while assuming constant compressive strength of hydration products (chiefly C-S-H regions) within the porous cement paste. This assumption of porosity-scaled relationships of mechanical properties has also been substantiated at the microscale for stiffness and hardness of the C-S-H regions within cement pastes, via indentation experiments over the past decade. These experiments are interpreted by describing the C-S-H regions as nanogranular. For example, Supplementary Figures 17.b-c show that indentation modulus and hardness of the C-S-H measured in microscale regions of low and high packing density among the C-S-H nanoparticles (denoted LD and HD, respectively) can be fit by micromechanics models that assume constant stiffness and hardness of the nanoparticles themselves. It is that solid nanoparticle-scale stiffness M and hardness H that we report in Figs. 3-4 of the manuscript.

It is well established that the compressive strength of concrete depends directly on the compressive strength of the hardened cement paste[91], which in turn contains this nanogranular C-S-H and pores. To first order, increasing the compressive strength of C-S-H nanoparticles could be expected to then increase the strength of the cement paste, and in turn of the concrete. We note that this expectation includes several important caveats regarding the microstructure and mechanics. These include comparison of samples with equivalent pore size distribution and cement content, as well as an assumption of equivalent cohesive forces among the C-S-H nanoparticles of high vs. low particle strength. Note that we directly report here on the mechanical properties of elastic modulus and hardness of the C-S-H nanoparticle, and not on nanoparticle compressive strength. Hardness is approximately proportional to yield strength or fracture strength in tension and compression for many monolithic materials; correlations between hardness and strength are weak for some mineral-based materials[92] or lacking in some particle-

containing composites[93]. However, as noted in the conclusion of the main text, the historical strength-porosity relation and the potential to modulate particle-scale mechanical properties together motivate future work to move the present findings toward strength, fracture toughness, and positive environmental impact on concrete design.

Supplementary References:

1. Mondal, P. Nanomechanical properties of cementitious materials. (2008).
2. Velez, K., Maximilien, S., Damidot, D., Fantozzi, G. & Sorrentino, F. Determination by nanoindentation of elastic modulus and hardness of pure constituents of Portland cement clinker. *Cem. Concr. Res.* **31,** 555–561 (2001).
3. Richardson, I. G. & Groves, G. W. the incorporation of minor and trace elements into calcium silicate hydrate (C-S-H) gel in hardened cement pastes. *Cem. Concr. Res.* **23,** 131–138 (1993).
4. Richardson, I. & Groves, G. microstructure and microanalysis of hardened cement pastes involving ground granulated blast-furnace slag. *J. Mater. Sci.* **27,** 6204–6212 (1992).
5. Richardson, I. G. tobermorite/jennite- and tobermorite/calcium hydroxide-based models for the structure of C-S-H: applicability to hardened pastes of tricalcium silicate, β-dicalcium silicate, Portland cement, and blends of Portland cement with blast-furnace slag, metakaolin, or silica fume. *Cem. Concr. Res.* **34,** 1733–1777 (2004).
6. Manzano, H., Pellenq, R. J. M., Ulm, F.-J., Buehler, M. J. & van Duin, A. C. T. hydration of calcium oxide surface predicted by reactive force field molecular dynamics. *Langmuir* **28,** 4187–4197 (2012).
7. Chen, J. J., Thomas, J. J., Taylor, H. F. W. & Jennings, H. M. solubility and structure of calcium silicate hydrate. *Cem. Concr. Res.* **34,** 1499–1519 (2004).
8. Cong, X. & Kirkpatrick, R. J. 29Si MAS NMR study of the structure of calcium silicate hydrate. *Adv. Cem. Based Mater.* **3,** 144–156 (1996).
9. Allen, A. J., Thomas, J. J. & Jennings, H. M. composition and density of nanoscale calcium–silicate–hydrate in cement. *Nat. Mater.* **6,** 311–316 (2007).
10. Thomas, J. J., Chen, J. J., Jennings, H. M. & Neumann, D. A. Ca-OH bonding in the C-S-H gel phase of tricalcium silicate and white portland cement pastes measured by inelastic neutron scattering. *Chem. Mater.* **15,** 3813–3817 (2003).
11. Pellenq, R. J.-M. *et al.* a realistic molecular model of cement hydrates. *Proc. Natl. Acad. Sci.* **106,** 16102–16107 (2009).
12. Abdolhosseini Qomi, M. J., Bauchy, M., Pellenq, R. J.-M. & Ulm, F.-J. Applying Tools from Glass Science to Study Calcium-Silicate- Hydrates. in *Mech. Phys. Creep Shrinkage Durab. Concr. Tribute Zdenk P Bazant* 78–85 (American Society of Civil Engineers, 2013). doi:10.1061/9780784413111.008
13. Richardson, I. G. the nature of the hydration products in hardened cement pastes. *Cem. Concr. Compos.* **22,** 97–113 (2000).

14. Richardson, I. G. the calcium silicate hydrates. *Cem. Concr. Res.* **38,** 137–158 (2008).
15. Shahsavari, R., Pellenq, R. J.-M. & Ulm, F.-J. empirical force fields for complex hydrated calcio-silicate layered materials. *Phys. Chem. Chem. Phys. PCCP* **13,** 1002–1011 (2011).
16. Hamid, S. the crystal-structure of the 11A natural tobermorite Ca2.25[Si3O7.5(OH)1.5].1H2O. *Z. Krist.* **154,** 189–198 (1981).
17. Abdolhosseini Qomi, M. J., Ulm, F.-J. & Pellenq, R. J.-M. evidence on the dual nature of aluminum in the calcium-silicate-hydrates based on atomistic simulations. *J. Am. Ceram. Soc.* **95,** evidence on the dual nature of aluminum in the calcium–silicate–hydrates based on atomistic simulations (2012).
18. Manzano, H. *et al.* impact of chemical impurities on the crystalline cement clinker phases determined by atomistic simulations. *Cryst. Growth Des.* **11,** 2964–2972 (2011).
19. Mortier, W. J., Ghosh, S. K. & Shankar, S. Electronegativity-equalization method for the calculation of atomic charges in molecules. *J. Am. Chem. Soc.* **108,** 4315–4320 (1986).
20. Tersoff, J. Empirical Interatomic Potential for Carbon, with Applications to Amorphous Carbon. *Phys. Rev. Lett.* **61,** 2879–2882 (1988).
21. Brenner, D. W. Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films. *Phys. Rev. B* **42,** 9458–9471 (1990).
22. Abell, G. C. Empirical chemical pseudopotential theory of molecular and metallic bonding. *Phys. Rev. B* **31,** 6184–6196 (1985).
23. Fogarty, J. C., Aktulga, H. M., Grama, A. Y., van Duin, A. C. T. & Pandit, S. A. A reactive molecular dynamics simulation of the silica-water interface. *J. Chem. Phys.* **132,** 174704–174704–10 (2010).
24. Russo Jr., M. F. & van Duin, A. C. T. Atomistic-scale simulations of chemical reactions: Bridging from quantum chemistry to engineering. *Nucl. Instrum. Methods Phys. Res. Sect. B Beam Interact. Mater. At.* **269,** 1549–1554 (2011).
25. Cygan, R. T., Liang, J.-J. & Kalinichev, A. G. Molecular Models of Hydroxide, Oxyhydroxide, and Clay Phases and the Development of a General Force Field. *J. Phys. Chem. B* **108,** 1255–1266 (2004).
26. Wolf, D., Keblinski, P., Phillpot, S. R. & Eggebrecht, J. Exact method for the simulation of Coulombic systems by spherically truncated, pairwise r−1 summation. *J. Chem. Phys.* **110,** 8254–8282 (1999).
27. Nicholson, D. & Parsonage, N. G. *Computer simulation and the statistical mechanics of adsorption*. (Academic Press, 1982).
28. Puibasset, J. & Pellenq, R. J. M. Water adsorption in disordered mesoporous silica (Vycor) at 300 K and 650 K: A Grand Canonical Monte Carlo simulation study of hysteresis. *J. Chem. Phys.* **122,** (2005).
29. Puibasset, J. & Pellenq, R. J.-M. Grand canonical Monte Carlo simulation study of water adsorption in silicalite at 300 K. *J. Phys. Chem. B* **112,** 6390–6397 (2008).
30. Bonnaud, P. A., Ji, Q., Coasne, B., Pellenq, R. J.-M. & Van Vliet, K. J. Thermodynamics of Water Confined in Porous Calcium-Silicate-Hydrates. *Langmuir* **28,** 11422–11432 (2012).

31. Gale, J. D. & Rohl, A. L. The General Utility Lattice Program (GULP). *Mol. Simul.* **29,** 291–341 (2003).
32. Gale, J. D. GULP: A computer program for the symmetry-adapted simulation of solids. *J. Chem. Soc.-Faraday Trans.* **93,** 629–637 (1997).
33. Plimpton, S. FAST PARALLEL ALGORITHMS FOR SHORT-RANGE MOLECULAR DYNAMICS. *J. Comput. Phys.* **117,** 1–19 (1995).
34. Parrinello, M. & Rahman, A. polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52,** 7182–7190 (1981).
35. Coudert, F.-X., Vuilleumier, R. & Boutin, A. Dipole Moment, Hydrogen Bonding and IR Spectrum of Confined Water. *ChemPhysChem* **7,** 2464–2467 (2006).
36. Manzano, H., Dolado, J. S. & Ayuela, A. elastic properties of the main species present in Portland cement pastes. *Acta Mater.* **57,** 1666–1674 (2009).
37. Manzano, H. *et al.* confined water dissociation in microporous defective silicates: mechanism, dipole distribution, and impact on substrate properties. *J. Am. Chem. Soc.* **134,** 2208–2215 (2012).
38. Manzano, H., Dolado, J. S. & Ayuela, A. aluminum Incorporation to Dreierketten Silicate Chains. *J. Phys. Chem. B* **113,** 2832–2839 (2009).
39. Barkema, G. T. & Mousseau, N. Event-Based Relaxation of Continuous Disordered Systems. *Phys. Rev. Lett.* **77,** 4358–4361 (1996).
40. Lau, T. T., Kushima, A. & Yip, S. Atomistic Simulation of Creep in a Nanocrystal. *Phys. Rev. Lett.* **104,** 175501 (2010).
41. Bauchy, M. & Micoulaut, M. transport anomalies and adaptative pressure-dependent topological constraints in tetrahedral liquids: evidence for a reversibility window analogue. *Phys. Rev. Lett.* **110,** 095501 (2013).
42. Ebrahimi, D., Pellenq, R. J.-M. & Whittle, A. J. nanoscale elastic properties of montmorillonite upon water adsorption. *Langmuir* **28,** 16855–16863 (2012).
43. Delafargue, A. & Ulm, F.-J. explicit approximations of the indentation modulus of elastically orthotropic solids for conical indenters. *Int. J. Solids Struct.* **41,** 7351–7360 (2004).
44. Ortega, J. A., Ulm, F.-J. & Abousleiman, Y. The nanogranular acoustic signature of shale. *Geophysics* **74,** D65–D84 (2009).
45. Malvern, L. E. INTRODUCTION TO THE MECHANICS OF A CONTINUOUS MEDIUM. (1969). at <http://trid.trb.org/view.aspx?id=199874>
46. Aghaei, A., Abdolhosseini Qomi, M. J., Kazemi, M. T. & Khoei, A. R. Stability and size-dependency of Cauchy–Born hypothesis in three-dimensional applications. *Int. J. Solids Struct.* **46,** 1925–1936 (2009).
47. Khoei, A. R., Qomi, M. J. A., Kazemi, M. T. & Aghaei, A. An investigation on the validity of Cauchy–Born hypothesis using Sutton-Chen many-body potential. *Comput. Mater. Sci.* **44,** 999–1006 (2009).
48. Khoei, A. R., Ghahremani, P., Abdolhosseini Qomi, M. J. & Banihashemi, P. Stability and size-dependency of temperature-related Cauchy–Born hypothesis. *Comput. Mater. Sci.* **50,** 1731–1743 (2011).
49. Qomi, M. J. A., Bauchy, M., Ulm, F.-J. & Pellenq, R. J.-M. anomalous composition-dependent dynamics of nanoconfined water in the interlayer of disordered calcium-silicates. *J. Chem. Phys.* **140,** 054515 (2014).

50. Ganneau, F. P., Constantinides, G. & Ulm, F.-J. dual-indentation technique for the assessment of strength properties of cohesive-frictional materials. *Int. J. Solids Struct.* **43,** 1727–1745 (2006).
51. Bauchy, M., Qomi, M. J. A., Ulm, F.-J. & Pellenq, R. J.-M. Order and disorder in calcium–silicate–hydrate. *J. Chem. Phys.* **140,** 214503 (2014).
52. Bauchy, M., Abdolhosseini Qomi, M. J., Bichara, C., Ulm, F.-J. & Pellenq, R. J.-M. Nanoscale Structure of Cement: Viewpoint of Rigidity Theory. *J. Phys. Chem. C* (2014). doi:10.1021/jp502550z
53. Wright, A. C. Neutron and X-ray amorphography. *J. Non-Cryst. Solids* **106,** 1–16 (1988).
54. J. Du, L. R. C. Compositional dependence of the first sharp diffraction peaks in alkali silicate glasses: A molecular dynamics study. *J. Non-Cryst. Solids* 3255–3269 (2006). doi:10.1016/j.jnoncrysol.2006.05.025
55. Elliott, S. R. origin of the first sharp diffraction peak in the structure factor of covalent glasses. *Phys. Rev. Lett.* **67,** 711–714 (1991).
56. Sokolov, A. P., Kisliuk, A., Soltwisch, M. & Quitmann, D. Medium-range order in glasses: Comparison of Raman and diffraction measurements. *Phys. Rev. Lett.* **69,** 1540–1543 (1992).
57. Wilson, M. & Madden, P. A. '"Prepeaks"' and '"first sharp diffraction peaks"' in computer simulations of strong and fragile ionic liquids. *Phys. Rev. Lett.* **72,** 3033–3036 (1994).
58. Tatlipinar, H., Akdeniz, Z., Pastore, G. & Tosi, M. P. Atomic size effects on local coordination and medium-range order in molten trivalent metal chlorides. *J. Phys. Condens. Matter* **4,** 8933 (1992).
59. Iyetomi, H. & Vashishta, P. Atomic-size effects on medium-range order in glasses. *Phys. Rev. B* **47,** 3063–3069 (1993).
60. Dixmier, J. Hole generation of prepeaks in diffraction patterns of glasses. *J. Phys. I* **2,** 1011–1027 (1992).
61. Uhlherr, A. & Elliott, S. R. extended-range propagated order in amorphous solids. *J. Phys. Condens. Matter* **6,** L99 (1994).
62. Elliott, S. R. extended-range order, interstitial voids and the first sharp diffraction peak of network glasses. *J. Non-Cryst. Solids* **182,** 40–48 (1995).
63. Zaug, J. M., Soper, A. K. & Clark, S. M. Pressure-dependent structures of amorphous red phosphorus and the origin of the first sharp diffraction peaks. *Nat. Mater.* **7,** 890–899 (2008).
64. Micoulaut, M. & Bauchy, M. anomalies of the first sharp diffraction peak in network glasses: Evidence for correlations with dynamic and rigidity properties. *Phys. Status Solidi B* n/a–n/a (2013). doi:10.1002/pssb.201248512
65. Wright, A. C. *et al.* The structure of some simple amorphous network solids revisited. *J. Non-Cryst. Solids* **129,** 213–232 (1991).
66. Zotov, N., Yanev, Y., Epelbaum, M. & Konstantinov, L. Effect of water on the structure of rhyolite glasses - X-ray diffraction and Raman spectroscopy studies. *J. Non-Cryst. Solids* **142,** 234–246 (1992).
67. Susman, S., Volin, K. J., Liebermann, R. C., Gwanmesia, G. D. & Wang, Y. Structural changes in irreversibly densified fused silica : implications for the chemical

resistance of high level nuclear waste glasses. *Phys. Chem. Glas.* **31,** 144–150 (1990).
68. N. Zotov, H. K. The effect of water on the structure of silicate glasses — A neutron diffraction study. *J. Non-Cryst. Solids* 153–163 (1996). doi:10.1016/0022-3093(96)00149-4
69. Bauchy, M. Structural, vibrational, and thermal properties of densified silicates: Insights from molecular dynamics. *J. Chem. Phys.* **137,** 044510–044510–12 (2012).
70. Soyer-Uzun, S., Chae, S. R., Benmore, C. J., Wenk, H.-R. & Monteiro, P. J. M. compositional evolution of calcium silicate hydrate (C–S–H) structures by total X-ray scattering. *J. Am. Ceram. Soc.* **95,** 793–798 (2012).
71. Wright, A. C. The comparison of molecular dynamics simulations with diffraction experiments. *J. Non-Cryst. Solids* **159,** 264–268 (1993).
72. Miller, M., Bobko, C., Vandamme, M. & Ulm, F.-J. Surface roughness criteria for cement paste nanoindentation. *Cem. Concr. Res.* **38,** 467–476 (2008).
73. Abuhaikal, M. Nano-ChemoMechanical assessment of Rice Husk Ash cement by wavelength dispersive spectroscopy and nanoindentation. (2011).
74. Constantinides, G. Invariant mechanical properties of calcium-silicate-hydrates (C-S-H) in cement-based materials: Instrumented nanoindentation and microporomechanical modeling. (2006).
75. G. Constantinides, K. S. R. C. grid indentation analysis of composite microstructure and mechanics: Principals and validation. *Mater. Sci. Eng. -Struct. Mater. Prop. Microstruct. Process.* **430,** 189–202
76. Oliver, W. c. & Pharr, G. m. Measurement of hardness and elastic modulus by instrumented indentation: Advances in understanding and refinements to methodology. *J. Mater. Res.* **19,** 3–20 (2004).
77. Oliver, W. c. & Pharr, G. m. An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. *J. Mater. Res.* **7,** 1564–1583 (1992).
78. Titterington, D. M., Smith, A. F. M. & Makov, U. E. *Statistical Analysis of Finite Mixture Distributions*. (John Wiley & Sons, 1986).
79. McLachlan, G. & Peel, D. *Finite Mixture Models*. (Wiley-Interscience, 2000).
80. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* **39,** 1–38 (1977).
81. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **6,** 461–464 (1978).
82. Armstrong, J. T. CITZAF - A package of correction programs for the quantitative electron microbeam X-ray analysis of thick polished materials, thinfilms, and particles. *Microbeam Anal.* **4,** 177–200 (1995).
83. Heinrich, K. F. J. & Newbury, D. *Electron Probe Quantitation*. (Springer, 1991).
84. Ulm, F.-J., Vandamme, M., Bobko, C. & Ortega, J. A. Statistical indentation techniques for hydrated nanocomposites: Concrete, bone, and shale. *J. Am. Ceram. Soc.* **90,** 2677–2692 (2007).
85. Cariou, S., Ulm, F.-J. & Dormieux, L. Hardness–packing density scaling relations for cohesive-frictional porous materials. *J. Mech. Phys. Solids* **56,** 924–952 (2008).

86. Gathier, B. Multiscale strength homogenization – application to shale nanoindentation. (2008).
87. Conn, A. R., Gould, N. I. M. & Toint, P. L. *Trust-Region Methods*. (Society for Industrial and Applied Mathematics, 1987).
88. Maloney, J. M., Lehnhardt, E., Long, A. F. & Van Vliet, K. J. Mechanical Fluidity of Fully Suspended Biological Cells. *Biophys. J.* **105,** 1767–1777 (2013).
89. Feret, R. *Étude expérimentale du ciment armé*. (Gauthier-Villars, 1906).
90. Verbeck, G. J. & Helmuth, R. H. Structures and physical properties of cement paste. in 1–44 (1969).
91. Popovics, S. *Strength and Related Properties of Concrete: A Quantitative Approach*. (John Wiley & Sons, 1998).
92. Koncagül, E. C. & Santi, P. M. Predicting the unconfined compressive strength of the Breathitt shale using slake durability, Shore hardness and rock structural properties. *Int. J. Rock Mech. Min. Sci.* **36,** 139–153 (1999).
93. Shen, Y.-L. & Chawla, N. On the correlation between hardness and tensile strength in particle reinforced metal matrix composites. *Mater. Sci. Eng. A* **297,** 44–47 (2001).