Supporting Information

Exploring the Guest-Host Relationship for Zeolite Y: A Synergistic Structural and Theoretical Investigation

Agnieszka Seremak[1], Ruben Goeminne[2], Izar Capel Berdiell[1], Lars Lundegaard[3*], Veronique Van Speybroeck[2], Stian Svelle[1*]

[1]*Department of Chemistry, University of Oslo, Blindern, P.O. Box 1033, 0315, Oslo, Norway*

[2]*Center for Molecular Modeling, Ghent University, 9000 Ghent, Belgium*

[3]*Topsoe A/S, Haldor Topsøes Allé 1, 2800, Kgs. Lyngby, Denmark*

*Corresponding author

Email: lafl@topsoe.com

Email: stian.svelle@kjemi.uio.no

Contents

S1 Cationic sites explanation.

In this work we used the naming and localization of the four distinctive conventional cationic sites as reported by Frising *et al*. with the details provided in Table S1.

Table S1. Nature, multiplicity and localisation of the different conventional cation sites in the faujasite structure.

Nature of cationic site	Maximum number of cation/u.c.	Localization of cationic sites		
		Cavity	Definition	Coordinates
Site I	16	Double 6-membered ring (D6R)	In the double six-membered ring	*0, 0, 0* or *x, x, x*
Site I'	32	Sodalite Cage	In the sodalite cage, close to the hexagonal window to the D6R	*x, x, x*
Site II'	32		In the sodalite cage, close to the hexagonal window to the supercage	*x, x, x*
Site II	32	Supercage	In the supercage, close to the hexagonal window to the sodalite cage.	*x, x, x*

S2 Atomic coordinates used in the Rietveld refinement model.

In situ XRD (wavelength = 0.496632 Å) was refined in a surface refinement using TOPAS software[1] from 1.4 to 33 range. The parameters constant across the whole refinement matrix were:

- 5 background parameters and "one_on_x".
- 4 Thompson-Cox-Hastings pseudo-Voigt broadening parameters.
- Zero error and both the position and intensity of the amorphous bump from the capillary reactor.

The list of local parameters allowed to refine with temperature is as follows:

- Thermal parameters of O (in the framework), T site and sodium. The occupancy of the framework was fixed to 1, while individual sodium atoms and dummy water were refined independently.
- The lattice parameter of the unit cell.
- The constrained x and y atomic coordinates of O (in the framework) and the scale factor.

Atomic coordinates used in the refinement protocol are listed in Table S2 (for zeolite framework atoms), Table S3 (for sodium cations), and Table S4 (for dummy oxygen atoms representing water molecules). The constraints of atomic coordinates for oxygen atoms in the framework are listed below Table S2.

Table S2. Atomic coordinates of the atoms of the zeolite framework..

Label	Multiplicity	x fract	y fract	z fract	Occupancy	Biso
site Si1	192	x 0.12557	y 0.94652	z 0.03645	occ Si 1	beq =BSi
site O1	96	x =O1x	y =O1x	z 0.96457	occ O 1	beq =BO
site O2	96	x =O2x	y =O2x	z 0.32096	occ O 1	beq =BO
site O3	96	x =O3x	y =O3x	z 0.14056	occ O 1	beq =BO
site O4	96	x =O4x	y =1-O4x	z 0	occ O 1	beq =BO

Constraints of atomic coordinates:
O1x: max 0.19 min 0.15
O2x: max 0.20 min 0.17
O3x: max 0.27 min 0.23
O4x: max 0.12 min 0.08

Table S3. Atomic coordinates of the sodium cations.

Label	Multiplicity	x fract	y fract	z fract	Occupancy	Biso
site I	16	x 0	y 0	z 0	occ Na =occ I	beq =BNa
site I'	32	x 0.071381	y 0.071381	z 0.071381	occ Na =occ Im	beq =BNa
site II'	32	x 0.165628	y 0.165628	z 0.165628	occ Na =occ IIm	beq =BNa
site II	32	x 0.250346	y 0.250346	z 0.250346	occ Na =occ II	beq =BNa

Table S4. Atomic coordinates for dummy oxygen atoms representing water molecules.

Label	Multiplicity	x fract	y fract	z fract	Occupancy	Biso
site A	192	x 0.452953	y 0.077137	z 0.028691	occ O =occ_A	beq = 5
site B	192	x 0.339861	y 0.207472	z 0.351690	occ O =occ_B	beq = 5
site C	192	x 0.598430	y 0.557106	z 0.535525	occ O =occ_C	beq = 5

S3 Computational Details of GCMC simulations.

Van der Waals interactions between guest-host and guest-guest molecules were described by Lennard-Jones (LJ) potentials, truncated at a *cutoff* distance of 12Å. For water, the all-atom TIP4P model was utilized.[2] Interactions between the zeolite framework, sodium cations, and water were modeled using parameters reported by Erdös et al.[3], while the atomic charges for the zeolite framework were adopted from Jaramillo et al.[4] All corresponding force-field parameters are listed in Table S5.

Lorentz-Berthelot mixing rules were applied to parameterize adsorbate-adsorbent interactions, and electrostatic interactions were managed using the Coulomb potential and Ewald summation. The ideal gas approximation was considered for water, due to the pressures being lower than 10 kPa. Water movements included insertion/deletion of molecules (accounting for 40% of the moves), as well as translation, rotation, and reinsertion (each representing 20% of the moves). Cation mobility was facilitated through translation and random translation movements, both with equal probability. Each GCMC simulation consisted of 50,000

equilibration cycles followed by 300,000 production cycles, with each cycle containing a number of Monte Carlo moves equal to the number of molecules present in the system. To accurately model zeolite Na-Y with adsorbed water molecules, the GCMC simulations were performed at varying temperatures (T = {50, 86, 100, 125, 140, 165, 220, 250, 300, 350, 400} °C) and pressures (p = {0.01, 0.05, 0.1, 1, 10, 100} Pa). The pressures were varied as the goal was to reproduce the TGA calculated number of adsorbed water molecules and not to check the adsorption capacity of zeolite Y

Table S5. **A:** Lennard-Jones parameters and atomic charges used in the simulation. **B:** Lennard-Jones parameters for the interactions between specific interaction sites. Subscript 'water' represents atom of the water molecule, 'zeolite' represents atom of the zeolite.

A

Atom type	ε [K]	σ [Å]	Atomic charge
O_{water}	78	3.154	0
M_{water}			-1.04
H_{water}			0.52
$Na_{zeolite}$	251.78	3.144	1.00
$Al_{zeolite}$			1.75
$Si_{zeolite}$			2.05
$O_{zeolite}$			-1.025
$O_{Al\text{-}zeolite}$			-1.20

B

Site 1	Site 2	ε [K]	σ [Å]
$Na_{zeolite}$	$O_{zeolite}$	33	3.2
$Na_{zeolite}$	$O_{Al\text{-}zeolite}$	23	3.4
$Na_{zeolite}$	O_{water}	75	2.39
$O_{zeolite}$	O_{water}	13.71	3.3765
$O_{Al\text{-}zeolite}$	O_{water}	13.71	3.3765

S4 Details of MLPs training.

To train the Machine Learning Potentials four active learning loops were performed with the details illustrated in the Figure S1. Starting from a small initial training set (950 snapshots), the goal is to gradually explore more and more of the phase space. To do that efficiently we use an iterative process of employing the MLP in a series of parallel NPT-MD simulations, recomputing the energies and forces of the last snapshot with DFT and adding those structures to the training set with which an updated MLP is trained.

The first three iterations of MLP NPT-MD were conducted at 200 K and run for 0.1 ps (200 MD steps, each 0.5 fs). Then, five more iterations of MLP NPT-MD were run at 500 K for 3000 MD steps, followed by three iterations at 700 K for 2000 MD steps. Finally, three iterations at 900 K for 2000 MD steps were performed. After each iteration, the final MD snapshot was recomputed with DFT and added to the total training set.

After the fourth active learning loop, the final MACE MLP was trained to the energies and forces with a cutoff radius of 5 Å, 32 channels, L=3 equivariant messages, 2 layers, and a body order

of 3. The models were trained for 600 epochs with weights of 100 for the energy and 1 for the forces. All models were trained to validation errors of at most 10 meV/atom on the energy and 200 meV/Å on the forces.

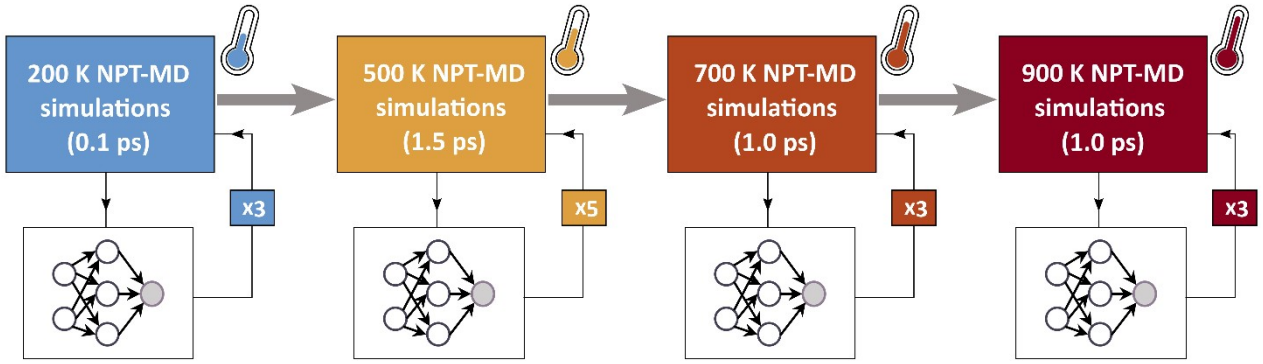

Figure S1. Schematic representation of the four active learning loops performed to train MLPs.

S5 Identification of cationic sites from MLP NPT-MD simulations

To identify and assign each sodium cation to one of the four conventional cationic sites in faujasite zeolite, as described in the S1, a distance-based method was applied. Using the coordinates for each cationic site position obtained from Rietveld refinement (S2), a specific cutoff radius was set for each site: site I – 1.8 Å, site I' – 2.0 Å, site II' – 2.0 Å, and site II – 2.4 Å. Each cation was examined to determine if it fell within the sphere defined by the cutoff radius for a particular site. If it was located within this sphere, it was assigned to that site. If a cation was not found to belong to any of the conventional sites, it remained unclassified. Figure S2 is a graphical representation of the cutoff regions corresponding to each site type.

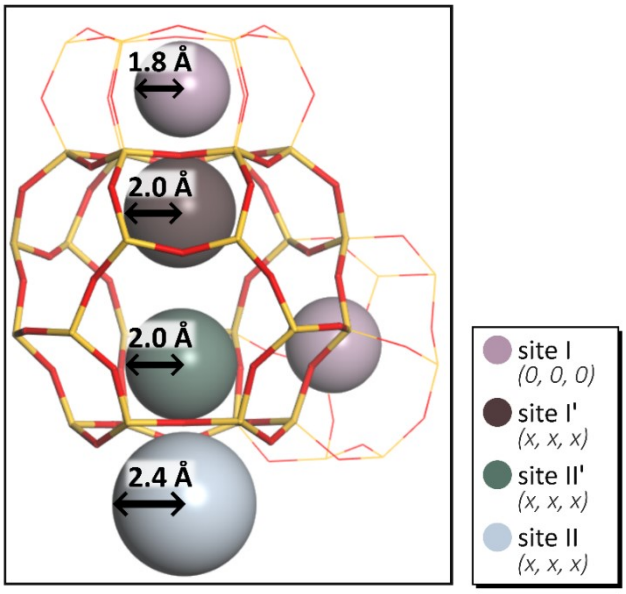

Figure S2. Graphical representation of identification of sodium sites from MLP NPT-MD simulations.

S6 TGA Results.

Figure S2presents the TGA plot for the CBV-100 sample. It can be observed that the dehydrated zeolite Y mass corresponds to the 78.25% of the hydrated sample.

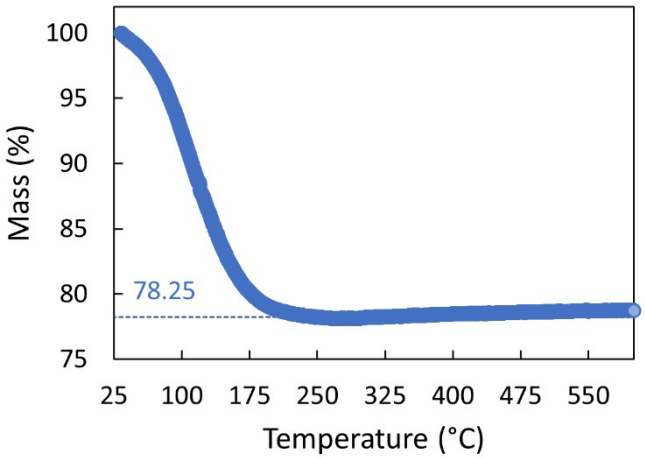

Figure S3. Weight percent loss during thermogravimetric analysis of the zeolite Na-Y sample (CBV-100).

To assess the sensitivity of zeolite Na-Y to external conditions such as humidity, the same sample was subjected to thermogravimetric analysis under different conditions compared to the original measurement. Figure S4 illustrates the TGA plot recorded during these alternate conditions. It is evident from the plot that the sample adsorbed a greater number of water molecules than observed in the initial TGA measurement (Figure S3).

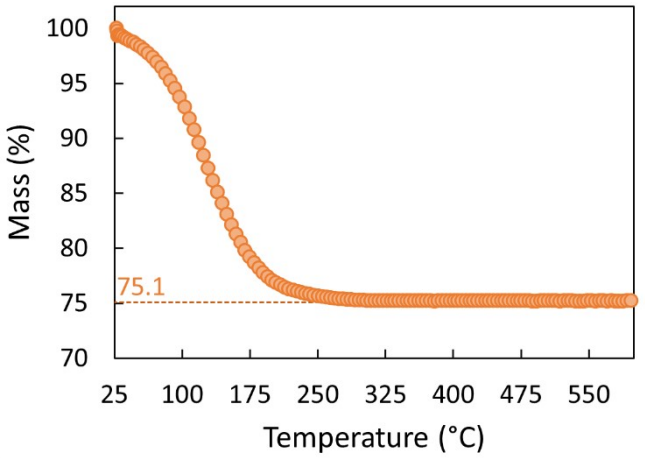

Figure S4. Weight percent loss during TGA of the zeolite Na-Y sample (CBV-100) performed during different external conditions.

S7 Mean Square Displacement and self-diffusion coefficient

Mean Square Displacement (MSD) were calculated using Einstein formula:

$$MSD(r_d) = \left\langle \frac{1}{N}\sum_{i=1}^{N} |r_d - r_d(t_0)|^2 \right\rangle_{t_0}$$

where N is the number of equivalent particles the MSD is calculated over, r are their coordinates and d is the desired dimensionality of the MSD.

MDAnalysis Python library [5,6] was used to perform the calculations of MSD by the Einstein relation and then to calculate the self-diffusivity. Self-diffusivity can be calculated from the MSD by performing linear least square regression to fit the model with respect to the lag-time.

$$D = \frac{1}{2d}\lim_{t\to\infty} \frac{d}{dt} MSD(r_d)$$

The line was fitted in the lag time range 50 – 300 ps. The MSD curves for water molecule and sodium are presented in Figure S4 and the self-diffusion coefficient with error bars given by the corresponding standard error are plotted in Figure S5.

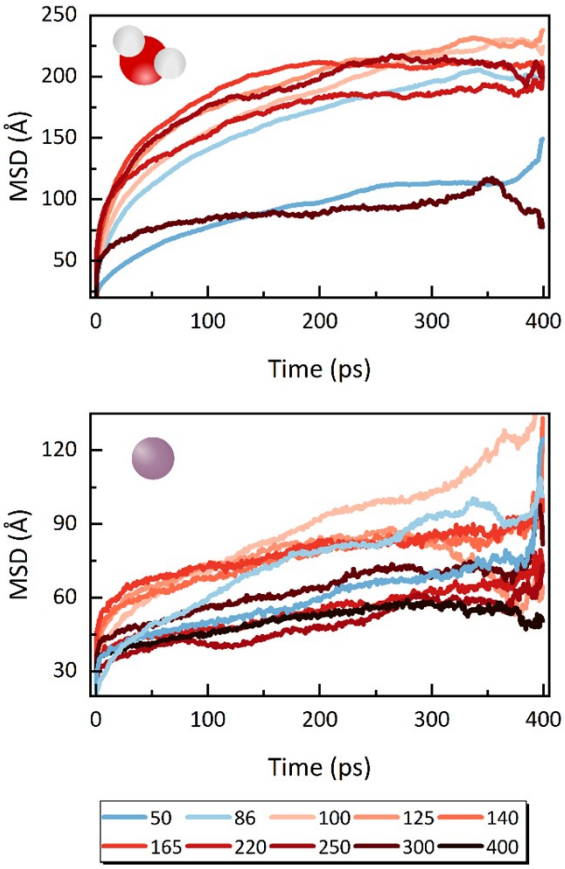

Figure S5. Mean Square Displacement (MSD) as obtained from trajectiories of the MLP-MD simulations. **Top**: MSD for water molecules. **Bottom**: MSD for sodium cations. Data from simulations at various temperatures (represented by different colors).

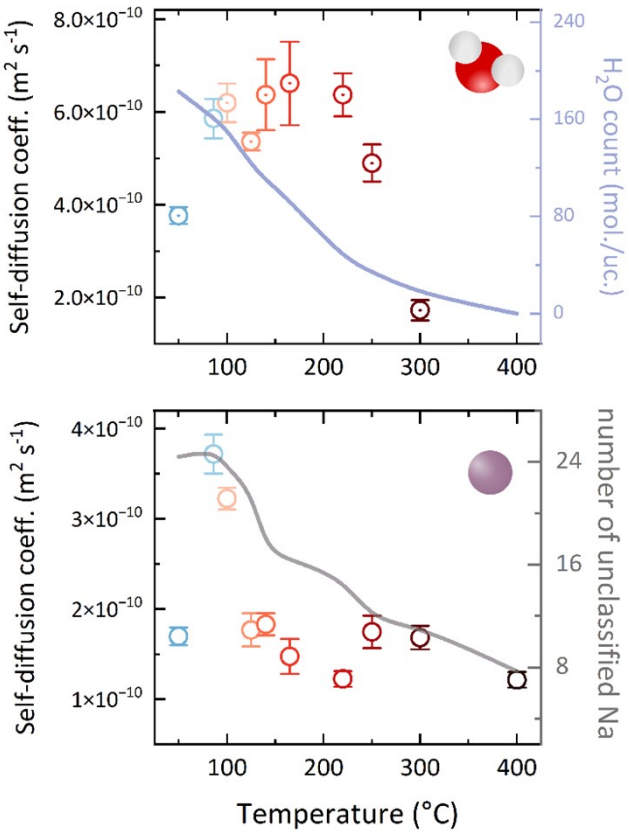

Figure S6. Self-diffusion coefficients from MSD data with standard error bars.**Top**: Water molecule diffusion coefficients as a function of temperature, with a secondary Y-axis indicating changes in water molecule count. **Bottom**: Sodium cation diffusion coefficients against temperature, with a secondary Y-axis indicating changes in number of unclassified Na cations.

Diffusion coefficients for water molecules in zeolite Y range from 2 - 7 x 10^{-10} m^2 s^{-1}, depending on the loading and temperature. The highest values are observed in models with an intermediate water count and within the temperature range of 140–250°C. Surprisingly, at higher temperatures, the diffusion coefficient does not increase, which is counterintuitive given the expectation of increased kinetic activity. Demontis *et al.*[7] studied the diffusion of water in zeolite NaY using quasi-elastic neutron scattering and molecular dynamics simulations (of at least 5 ns) at various water loadings and different temperatures. Their results showed that at 100°C and a water loading of 150 molecules per unit cell (mol./uc.), the calculated diffusion coefficient was 8.9 x 10^{-10} m^2 s^{-1}, which is similar to our results. However, at 300°C and a water loading of 20 mol./uc., the diffusion coefficient for water was 24 x 10^{-10} m^2 s^{-1}, which is one order of magnitude higher than our results. The discrepancies at higher temperatures are most likely due to the inadequately short MLP-MD simulations, which did not allow for an accurate description of water mobility.

In comparison, the diffusion coefficient for sodium was lower than that for water molecules. The highest sodium mobility was observed at 100°C and 125°C, corresponding to a reduced

number of adsorbed water molecules compared to the initial state. This increase in mobility can be attributed to the less confined space, as sodium, initially located in the supercage, migrates toward the zeolite framework. Despite increasing temperatures, the diffusion coefficients for sodium do not increase. As explained in the main text, this is primarily due to sodium migration to sites II and I, which are stable sites, resulting in sodium diffusing mostly within areas close to these sites. Another reason for the lack of increased sodium mobility could be the simulation length, which might have been too short to accurately describe the movement.

S8 Rietveld refinement - total number of sodium and weighted profile residuals (Rwp)

The total number of refined sodium cations was calculated and is reported in Figure S7. Additionally, the weighted profile residuals (Rwp) were calculated at each refinement step. Overall, the fit of the refinement model is good, with Rwp values falling within accepted ranges (up to 10%).

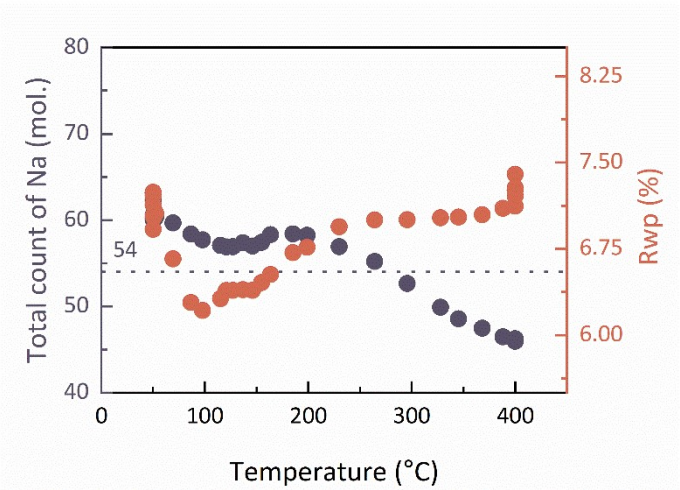

Figure S7. Total number of identified sodium cations and the weighted profile residual (Rwp) as a function of temperature. Data obtained from Rietveld refinement.

S9 Relative model for the XRD data – H_2O distribution

Figure S8 visualization provides a comparative model for the X-ray diffraction data, with the plot showing changes in the population of each water site. All water sites used in the model were located within the supercage.

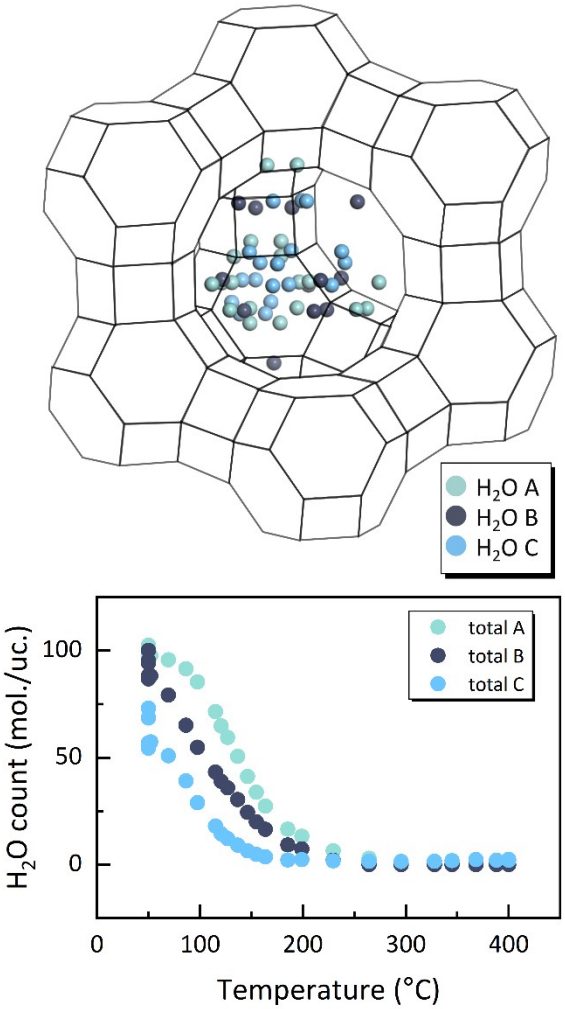

Figure S8. **Top**: Line representation of the Y zeolite framework showing one supercage and several sodalite cages together with the refined water (A, B, and C) sites according to the models used in the refinement (Supporting Information 2). **Bottom**: The number of adsorbed water molecules per unit cell obtained from Rietveld refinement as a function of temperature.

S10 Modeling the dehydration of siliceous zeolite Y

To examine changes in the unit cell parameter upon dehydration of siliceous zeolite Y, molecular dynamics simulations with trained MLPs were performed using OpenMM software[9] in the NPT ensemble. The initial structures contained the same number of water molecules as the corresponding Na-Y model at given temperature. The simulations were performed at T = 50, 125, 165, 250 and 400 °C with the number of water molecules equal to 183, 121, 94, 34 and 0, respectively.

Temperature and pressure were kept constant, with a timestep of 1.0 fs. Each simulation ran for 250 ps (250 000 MD steps). Temperature was controlled using a Langevin thermostat with a time constant of 100 fs, and pressure was controlled using a Monte Carlo barostat with a time constant of 1 ps. Data on energy, temperature, and unit cell volume were recorded every 100 steps, while trajectories were captured every 1 000 steps.

The unit cell parameter was calculated as an average of the data recorded during the last 100 ps of the MD simulations. It was assumed that the framework symmetry remained unchanged, and that zeolite Y maintained a cubic structure throughout each MD simulation.

S11 Validation of the Machine Learning Potentials

To validate the Machine Learning Potentials, the final 10 snapshots from the MLP-MD simulations at each studied temperature were extracted, in total 120 snapshots. Energies and forces for each snapshot were calculated using Density Functional Theory (DFT) with the PBE+D3 functionals [10,11], as implemented in version 8.2 of the CP2K simulation package. Differences between the *ab initio* values and those obtained from the MLPs were calculated as errors. The errors were determined using the following equations:

$$\frac{E_{DFT} - E_{MLP}}{n_{atoms}} \#(1)$$

$$F_{DFT} - F_{MLP} \#(2)$$

Equation (1) calculates the energy error per atom, while Equation (2) calculates the average force error. The results of these validations are shown in Figure S9.

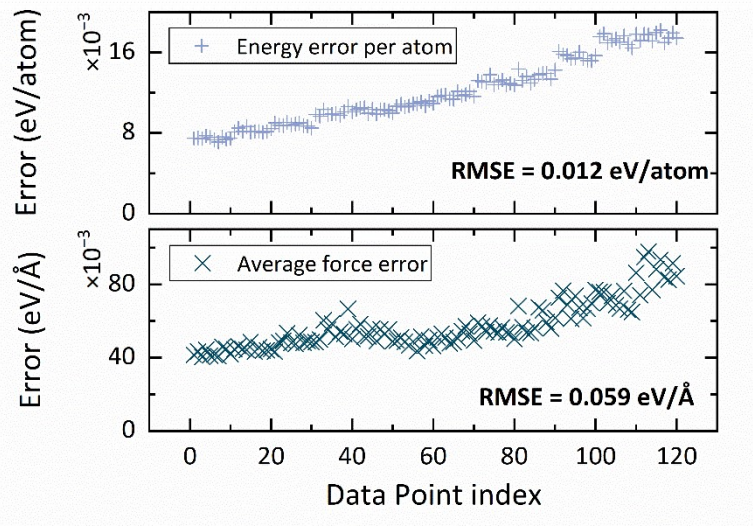

Figure S9. Energy and forces calculated errors on each compared system.

S12 Elemental analysis - EDX

The elemental composition, Si/Al ratio and Na/Al ratio were determined by energy-disperse X-ray spectroscopy (EDX) at a 20 kV accelerating voltage on 200 × 200 µm areas of pelletized material. Quantification was performed using the Bruker Quantax system consisting of a X-flash 6|10 detector and Esprit software. The results are presented in Table S6.

Table S6. EDX analysis of sample CBV 100.

Sample ID	SiO_2/Al_2O_3 mole ratio	Nominal Cation Form	Si/Al nominal	Si/Al EDX	Na/Al EDX	Formula	Calculated Formula (EDX)
CBV 100	5.1	Sodium	2.55	2.238	0.998	$Na_{54}(Si_{138}Al_{54})O_{384}$	$Na_{59}(Si_{133}Al_{59})O_{384}$

S13 ^{27}Al solid-state-NMR

A Bruker AVANCE Neo spectrometer operating at a magnetic field of 18.8 T (1H resonance frequency 800 MHz) was applied for the solid-state NMR experiment using a 3.2 mm double channel Magic Angel Spinning (MAS) probe at room temperature. The MAS rate was 20 kHz. We applied a pulse acquire type experiment with a short hard pulse (pulse width less than π/12). The spectrum was referenced to 1M Al(aq) set to be 0 ppm. To obtain the spectrum the raw data (FID) was apodised with a line broadening factor of 50 Hz and baseline corrected. A full spectrum is presented in Figure S10.a, with a large peak as a central transition and surrounding small ones are spinning sidebands. The large peak is associated with Al in T-sites and has a chemical shift of 62.0 ppm. The zoom-in of the spectrum is presented in Figure S10.b, with a large magnification in a rectangle with blue background. A very small peak at 11.3 ppm

can be attributed to extra-framework Al or octahedrally coordinated Al. Estimated peak intensities shows that this small peak should be less than 0.5% of the total peak areas.

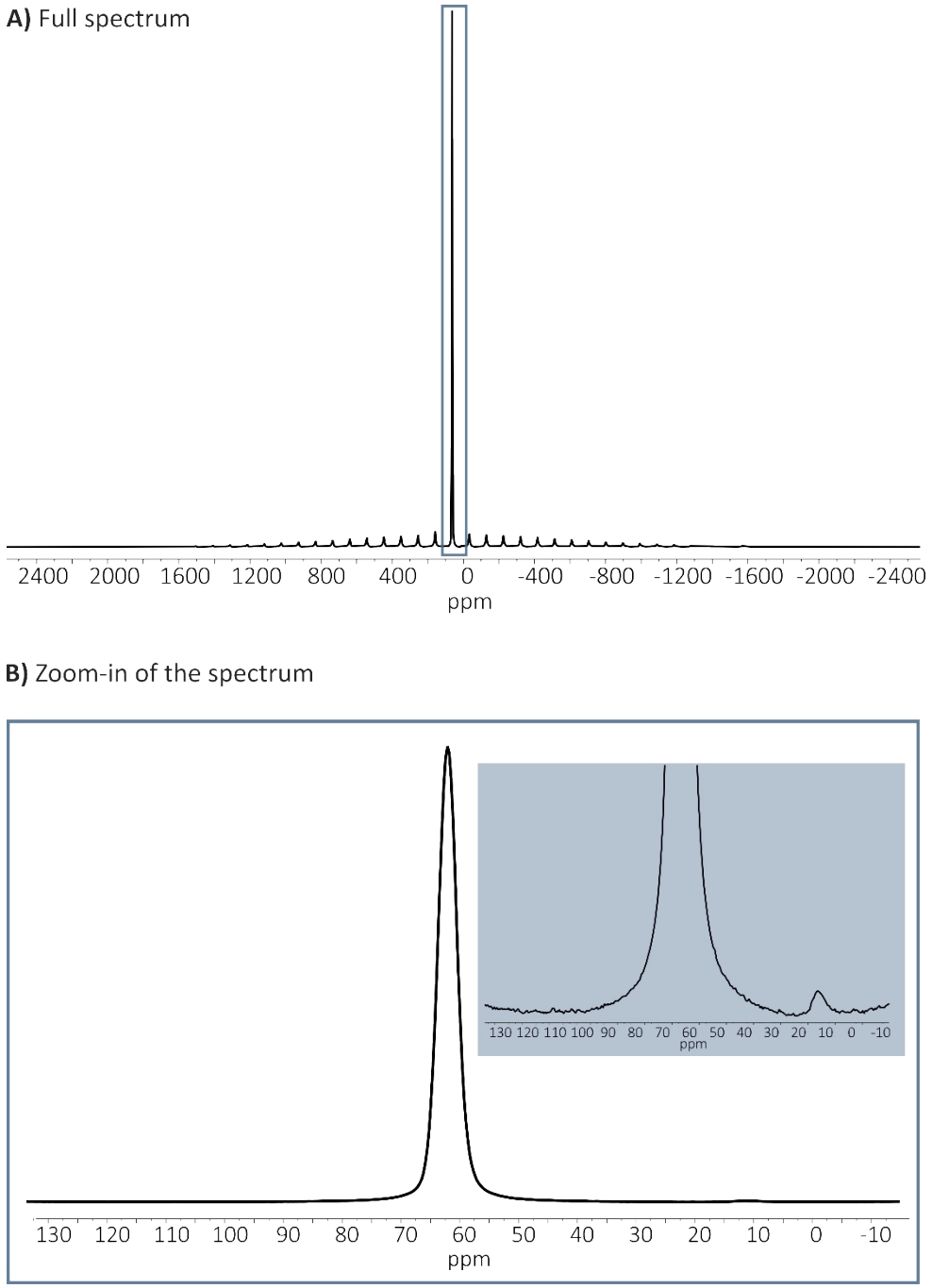

Figure S10.a) Full ^{27}Al-NMR spectrum of as-received Na-Y, b) Zoom-in to the central peak at 62.0 ppm, and small peak at 11.3 ppm.

References

(1) Coelho, A. A. TOPAS and TOPAS-Academic: An Optimization Program Integrating Computer Algebra and Crystallographic Objects Written in C++. *J Appl Cryst* 2018, *51* (1), 210–218. https://doi.org/10.1107/S1600576718000183.

(2) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* 1983, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(3) Erdős, M.; Geerdink, D. F.; Martin-Calvo, A.; Pidko, E. A.; van den Broeke, L. J. P.; Calero, S.; Vlugt, T. J. H.; Moultos, O. A. In Silico Screening of Zeolites for High-Pressure Hydrogen Drying. *ACS Appl. Mater. Interfaces* 2021, *13* (7), 8383–8394. https://doi.org/10.1021/acsami.0c20892.

(4) Jaramillo, E.; Auerbach, S. M. New Force Field for Na Cations in Faujasite-Type Zeolites. *J. Phys. Chem. B* 1999, *103* (44), 9589–9594. https://doi.org/10.1021/jp991387z.

(5) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *Journal of Computational Chemistry* 2011, *32* (10), 2319–2327. https://doi.org/10.1002/jcc.21787.

(6) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *scipy* 2016. https://doi.org/10.25080/Majora-629e541a-00e.

(7) Demontis, P.; Jobic, H.; Gonzalez, M. A.; Suffritti, G. B. Diffusion of Water in Zeolites NaX and NaY Studied by Quasi-Elastic Neutron Scattering and Computer Simulation. *J. Phys. Chem. C* 2009, *113* (28), 12373–12379. https://doi.org/10.1021/jp901587a.

(8) Furukawa, S.; Goda, K.; Zhang, Y.; Nitta, T. Molecular Simulation Study on Adsorption and Diffusion Behavior of Ethanol/Water Molecules in NaA Zeolite Crystal. *J. Chem. Eng. Japan / JCEJ* 2004, *37* (1), 67–74. https://doi.org/10.1252/jcej.37.67.

(9) Eastman, P.; Galvelis, R.; Peláez, R. P.; Abreu, C. R. A.; Farr, S. E.; Gallicchio, E.; Gorenko, A.; Henry, M. M.; Hu, F.; Huang, J.; Krämer, A.; Michel, J.; Mitchell, J. A.; Pande, V. S.; Rodrigues, J. P.; Rodriguez-Guerra, J.; Simmonett, A. C.; Singh, S.; Swails, J.; Turner, P.; Wang, Y.; Zhang, I.; Chodera, J. D.; De Fabritiis, G.; Markland, T. E. OpenMM 8: Molecular Dynamics Simulation with Machine Learning Potentials. *J. Phys. Chem. B* 2024, *128* (1), 109–116. https://doi.org/10.1021/acs.jpcb.3c06662.

(10) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* 1996, *77* (18), 3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865.

(11) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *The Journal of Chemical Physics* 2010, *132* (15), 154104. https://doi.org/10.1063/1.3382344.