Supporting Information

^{27}Al NMR chemical shifts in zeolite MFI via machine learning acceleration of structure sampling and shift prediction

Daniel Willimetz,* Andreas Erlebach, Christopher J. Heard and Lukáš Grajciar*

Department of Physical and Macromolecular Chemistry, Charles University,
Hlavova 8, Prague 2, Prague 128 43, Czech Republic

*E-mail: daniel.willimetz@natur.cuni.cz, lukas.grajciar@natur.cuni.cz

S1. Models and molecular dynamics simulation

The initial structure employed in this study was derived from the siliceous MFI structure obtained from the IZA database. Unit cell parameters were systematically scaled by factors of 0.96, 0.98, 1.00, 1.02, 1.04, and 1.06, followed by a 2 ns molecular dynamics (MD) simulation. A cubic polynomial was used to fit the unit cell volume against the system's average energy. For the model with a Si/Al ratio of 95 and 7 water molecules per unit cell, the resulting fit is shown in Fig. S1. The unit cell volume of different systems corresponding to the lowest energy, as predicted by the polynomial fit, along with the associated energy values, is presented in Table S1.
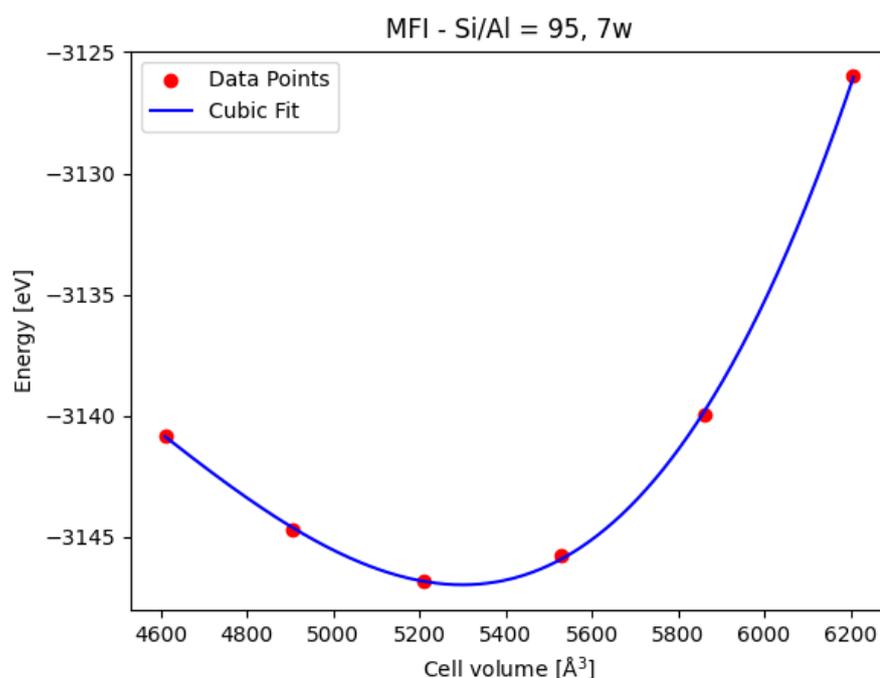
Fig. S1: Cubic polynomial fit of unit cell volume versus average energy for the MFI structure with a Si/Al ratio of 95 and 7 water molecules per unit cell. The lowest energy configuration corresponding to an optimal volume of 5300 Å³.

Table S1: Optimal unit cell volumes for different systems.

Si/Al ratio	Water loading	Energy [eV]	Volume [Å³]
95	7	–3147.0	5300.8
95	0	-3034.1	5317.5
11	17	–3320.3	5300.5
siliceous	0	–3033.1	5313.6

The unit cell volumes for the different systems containing aluminium atoms are within 3 %, indicating minimal structural change across different system configurations.

The structure obtained from the IZA database was scaled to an optimal unit cell volume of 5300 Å^3 and aluminium atoms and water molecules were added according to the model configuration. Geometry optimization was performed until the maximum force on each atom converged to less than 10^{-2} eV/Å. The SchNet Python package was employed to model the interaction potential, with a cutoff distance of 6 Å. The interaction potential (SiAlOH1) used was a neural network potential from the work of Erlebach et al. [1] To accurately represent experimental conditions, simulations were conducted at 350 K using a Nosé-Hoover thermostat. The thermostat's coupling parameter was set to 40 fs. Molecular dynamics (MD) simulations were carried out with a timestep of 0.5 fs over 2 million steps, and structural snapshots were recorded every 100 steps for analysis. Visual MD energy analysis was performed to confirm energy equilibration, as shown in Fig. S2, which depicts the time evolution of both energy and the KRR-predicted chemical shift.
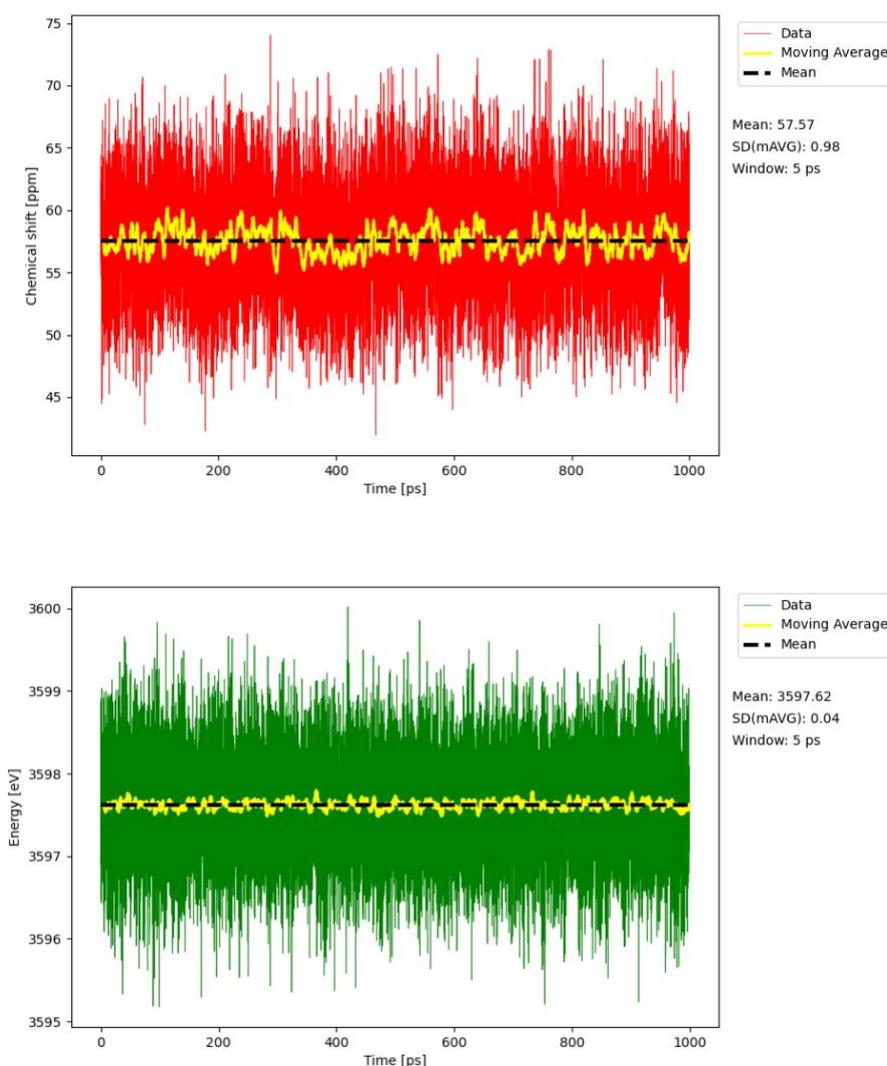
Fig. S2: Time evolution of KRR-predicted chemical shift (top) and potential energy (bottom) during the molecular dynamics simulation of the MFI structure at 350 K. The yellow line

represents the moving average with a window of 5 ps, showing fluctuations around the mean value.

The stability and reliability of the NNP (using ASE as a MD-driver) for MD simulations in water-loaded protonic aluminosilicate zeolites has been extensively demonstrated previously in the work in which these potentials were generated, [1] Nevertheless, to further test the reliability of the NNP-driven MD in Atomic simulation environment (ASE), a 5 ps NNP MD and AIMD simulations were performed from identical initial structures for the H-FAU(Si/Al=1) model with 48 H_2O in the channel. The AIMD trajectory was taken from Erlebach et. al.[1] The distribution of average Al-O-Si angles and Al-O bond lengths is visualized in Figure S3 below. The similarity between the structural parameters further illustrate the accuracy of NNP MD simulations.
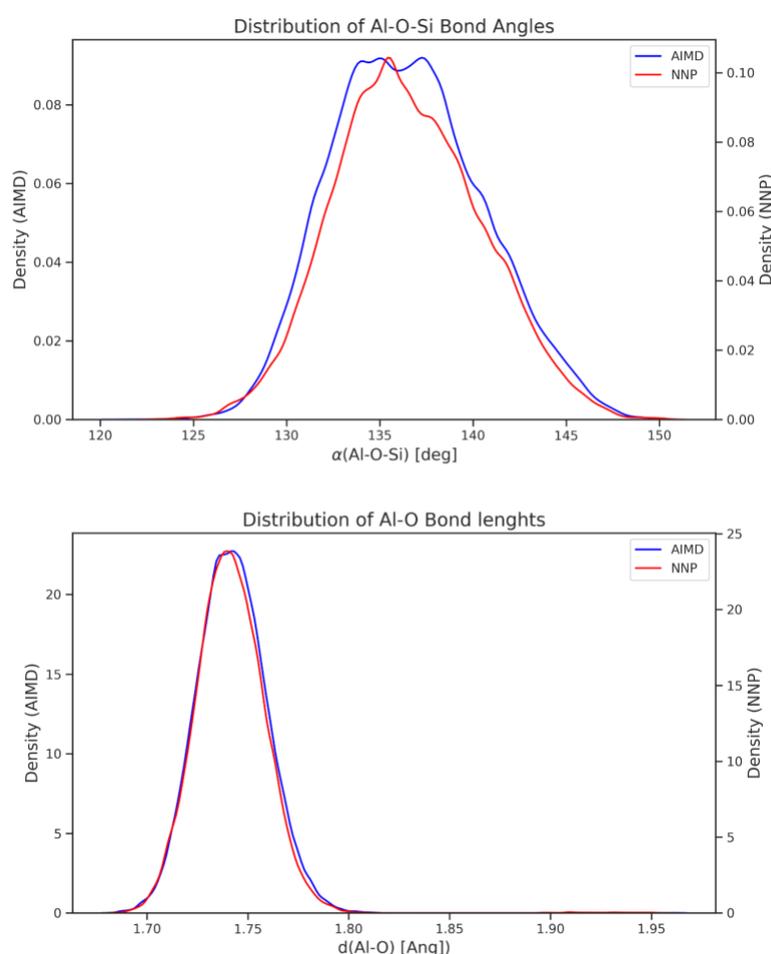
Fig. S3: The structural parameters (Al-O-Si angles and Al-O bond lengths) comparison between NNP MD (read) and AIMD (blue) of the same length starting from the same initial structure for the H-FAU(Si/Al=1) model with 48 H2O in the channel taken from Ref. [1].

It is established that the crucial features for predicting [27]Al chemical shifts are the Al-O bond lengths, the Al-O-Si angles and the O-Al-O angles. [2] In Figure S4 is an analysis of such features from all available MD trajectories of MFI (Si/Al = 95) with different water loading (0,

1, 2, 3 and 17 water molecules per unit cell) and different initial conditions (aluminium atom positions, BAS configurations etc.), compared to the training database. As illustrated in Figure S4, the KRR model does not extrapolate for these parameters, and thus is expected to give accurate predictions, which are verified against experiment for MFI in Fig. 8.
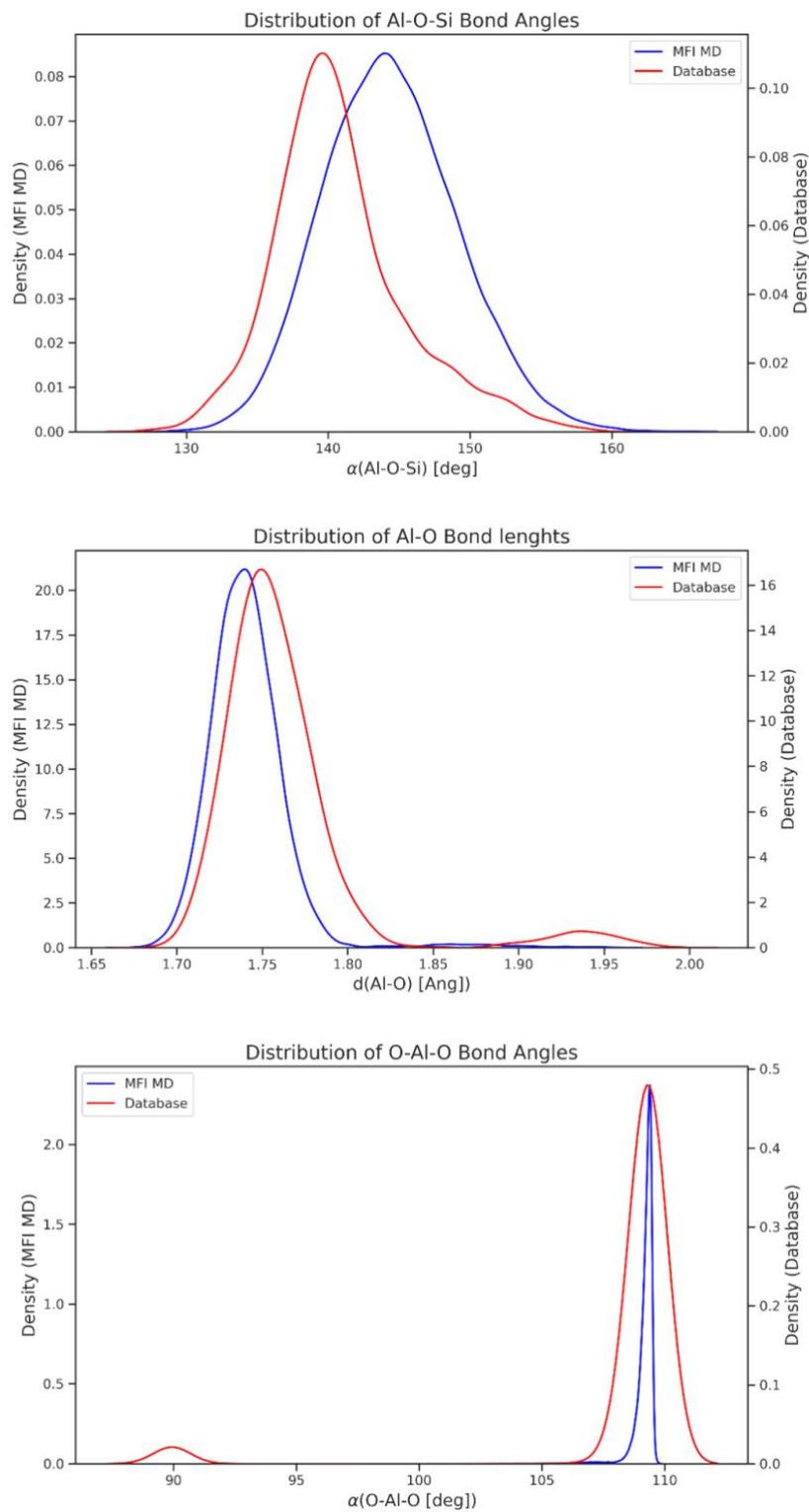
Fig. S4: Comparison of three distributions of structural parameters in all MFI (Si/Al = 95) MD trajectories (blue) and the training database (red).

S2. Database Information

Most of the MOR and CHA structures utilized in this study were sourced from previous research by Lei et al. [2]. Additional CHA models were added to better generalize the KRR model, incorporating systems with 1 to 3 Al atoms balanced by Na or H, and water content ranging from 0 to 6 molecules. The Born-Oppenheimer *ab initio* molecular dynamics (AIMD) simulations for these models ranged from 75 to 150 ps.

To further improve the model's predictive accuracy, we added 100 additional MFI (Si/Al = 95) structures. These structures were selected using the furthest point sampling (FPS) algorithm from a pool of 4 million structures generated through NNP-driven molecular dynamics simulations at 350 K with varying initial configurations and different aluminium placement. The FPS method ensured the selection of the most diverse structures in terms of geometry. The MFI dataset was further categorized based on water loading, ranging from 0 to 17 water molecules per unit cell, with an equal representation of 20 structures per water loading scenario. This balanced approach ensured adequate representation of various water content. The input for the FPS algorithm consisted of SOAP descriptors for the Al atoms, which were generated in the same manner as those used for training the KRR model (Section S3). Fig. S5 presents the Euclidean distance between the SOAP descriptors of the sampled structures, illustrating the diversity captured by the FPS sampling. The convergence of the distance suggests effective sampling of the different Al environments in MFI was achieved, thus enhancing the model's ability to generalize across water loadings and structural variations.
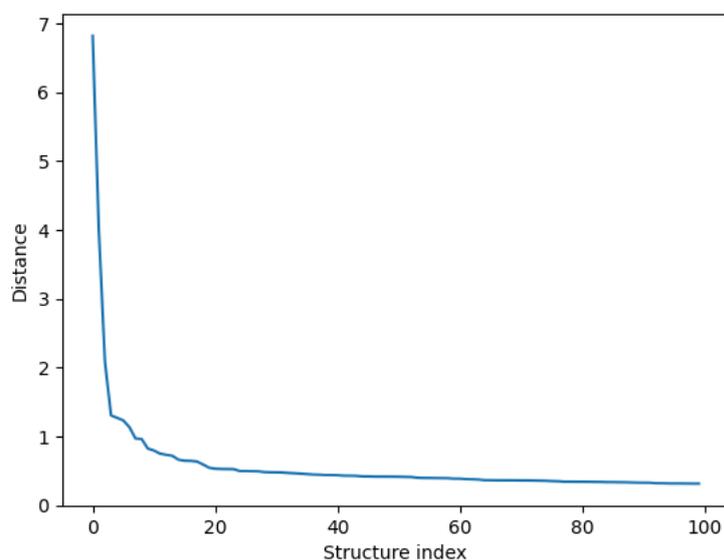
Fig. S5: Euclidean distance between SOAP descriptors for the first 100 structures in the FPS-sampled MFI database.

The inclusion of MOR zeolite in the structural database is important due to its variety of ring sizes. While the CHA framework only features 4-, 6-, and 8-membered rings, the MOR framework includes 4-, 5-, 8-, and 12-membered rings. Zeolites like MFI contain T sites within 5-membered rings, making the inclusion of MOR necessary for broader applicability. Additionally, 100 MFI structures were incorporated to account for 10-membered rings, enhancing the database's accuracy for this framework. Furthermore, the database includes MFI structures that contain intrazeolitic hydrogen bond, which is not well represented in the

MOR and CHA database, where this structural motif is not as common. To be more specific, the intrazeolitic hydrogen bond is 25/100 MFI structures, 25/1769 CHA structures and 0/260 MOR structures. This intrazeolitic hydrogen bond is defined as a proton bonded to a framework oxygen atom that is within 2 Å of another framework oxygen atom.

S3. KRR Training

The Kernel Ridge Regression (KRR) model was trained using SOAP descriptors of Al atoms, generated by the DScribe package.[3] The model included species with atomic numbers 1 (hydrogen), 8 (oxygen), 11 (sodium), 13 (aluminium), and 14 (silicon), with a cutoff radius (rcut) of 5.0 Å to define the atomic neighbourhood. Radial basis functions were expanded up to a maximum degree (nmax) of 6, and spherical harmonics expansion was set to a maximum angular momentum (lmax) of 6. Gaussian smearing with a width (σ) of 0.5 Å was applied, and a power spectrum exponent (ζ) of 2 was used to adjust the sensitivity of the SOAP kernel to structural variations.

A grid search was performed to find the optimal regularization parameter (λ), ranging from 10^{-7} to 10^{-2}. The model with $\lambda = 5.5 \cdot 10^{-6}$ achieved the lowest error. Performance metrics for this optimal model are provided in Table S2.

Table S2: Performance metrics of the best KRR model.

	Train	Test
R^2	0.9996	0.9966
MAE [ppm]	0.204	0.510
RSME [ppm]	0.070	0.554

The model was trained using an 80/20 train-test split to evaluate its predictive accuracy. Multiple data splits were tested to ensure robustness, with test errors across different splits varying within a narrow margin of 0.1 ppm. This consistency indicates that the model's performance is relatively stable across different train-test splits. Detailed results, including the random states used and their corresponding train and test errors, are presented in Table S3.

Table S3: Training and testing errors measured by the MAE metric for various random states used in splitting the training dataset into an 80/20 train-test ratio.

Random state	Train MAE [ppm]	Test MAE [ppm]
124513453	0.28	0.49
2454353	0.23	0.55
54636554	0.23	0.57
93457	0.27	0.55
644474322	0.29	0.54
7645645	0.31	0.55
12334253	0.26	0.54
88460530	0.30	0.54
240452203	0.26	0.56
36364042	0.26	0.53

S4. Benchmarking on MFI Structures

To evaluate the performance of the KRR model and other predictive models for zeolites with MFI topology, we created a benchmark set of MFI structures. The benchmarking process involved predicting chemical shifts using the KRR method and then identifying molecular dynamics (MD) simulations with the highest and lowest predicted chemical shifts to represent the full range of chemical shifts. Initial Brønsted acid sites (BAS) configurations corresponding to these extreme chemical shifts are listed in Table S4. Subsequently, 10 structures from the MD simulations were randomly selected, and their chemical shifts were recalculated using the DFT method.

These selected structures and their corresponding chemical shifts serve as a reference to assess the accuracy and reliability of the KRR model in predicting chemical shifts across various conditions.

Table S4: Selected initial BAS configurations that exhibited the lowest (top) and highest (bottom) chemical shifts for each tested water loading scenario (i.e., number of water molecules per unit cell).

Water loading	BAS
0	T4O4
	T6O10
1	T7O18
	T9O15
2	T7O17
	T9O15
3	T7O17
	T9O15
17	T11O16
	T9O15

S5. Comparison with LASSO Model

Lei et al. [2] applied a LASSO model to a dataset of 1400 structures from MOR and CHA, obtaining a mean absolute error (MAE) of 1.3 ppm. Using the same dataset, the KRR model demonstrated a lower test error of 0.4 ppm. For MFI structures, the LASSO model yielded MAE values below 2 ppm, whereas the KRR model, evaluated on the same MFI dataset, exhibited an error of 1.0 ppm. When trained on a larger dataset comprising over 4,000 distinct Al environments (including MFI), the KRR model achieved an MAE of 0.7 ppm for MFI structures.

Fig. S6 shows the correlation between predicted chemical shifts and those calculated by DFT for the MFI test set across different water loadings. The comparison includes LASSO models (5p-LASSO and 2p-LASSO), a KRR model trained on the same database as the LASSO models (KRR_dtb), and a KRR model trained on a comprehensive database including CHA, MOR, and MFI structures (KRR_MFI).
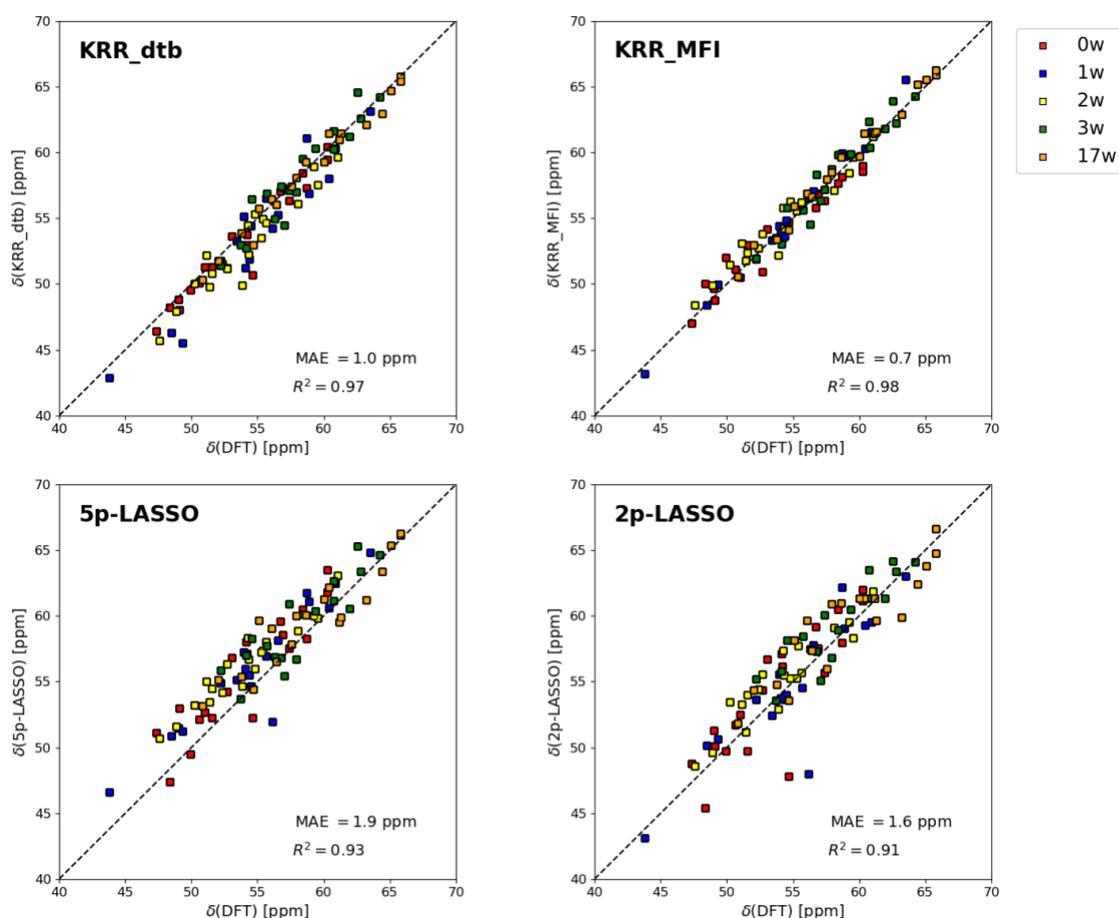
Fig. S6: Correlation between predicted chemical shifts and DFT-calculated values for MFI test set at various water loadings, comparing LASSO models (5p-LASSO and 2p-LASSO), KRR trained on the same database (KRR_dtb), and KRR trained on a combined database of CHA, MOR, and MFI structures (KRR_MFI).

The KRR model trained on the same database as the LASSO models outperforms LASSO in both MAE and R^2 scores, indicating superior predictive performance for chemical shifts in MFI structures.

S6. Managing Multiple BAS Configurations

In the zeolite framework, each T-site is coordinated by four oxygen atoms, leading to four distinct configurations of Brønsted acid sites (BAS) around each aluminium atom. The distribution of acidic protons among these configurations follows the Boltzmann probability distribution, where more stable configurations are favoured due to their higher Boltzmann factors. To derive a single representative value for the chemical shift at each T-site, we average the chemical shift values across various initial BAS configurations using normalized Boltzmann factors. This Boltzmann averaging helps to mitigate the influence of the initial configuration and provides a more stable representation of the chemical shift.

When water molecules are present, BAS configurations can fluctuate during the simulation as protons move between framework oxygens through interactions with water molecules. Consequently, the energy levels of different BAS configurations become similar, as shown in Table S5. For the system with 2 water molecules per unit cell, the energy differences between configurations are within 1 kJ/mol, leading to comparable normalized Boltzmann factors.

Table S5: Boltzmann averaging example for T1 across four BAS configurations. The presence of water molecules results in similar energy values for different BAS configurations due to extensive sampling, irrespective of the initial state.

T-site	BAS	Energy [eV]		Normalized Boltzmann factors	
		0w	2w	0w	2w
T1	O1	-3031.502	-3064.222	0.26	0.29
	O2	-3031.518	-3064.213	0.45	0.22
	O3	-3031.463	-3064.216	0.07	0.24
	O4	-3031.496	-3064.216	0.22	0.25

S7. Structural characteristics in models with different water loadings

Fig. S7 and Table S6 show the average Al-O-Si angle and average Al-O bond length for models with various water loadings. Notably, the average Al-O bond length decreases consistently as the water loading increases. The most significant reduction occurs between 1 and 2 water molecules per unit cell, with a decrease of 0.005 Å. This decrease in bond length is likely responsible for the observed increase in chemical shift from 1 to 2 water molecules. The variations in the average T-O-T angles appear somewhat random between 0 and 2 water molecules, but there is a general trend toward an increase in the angle with higher water loadings.
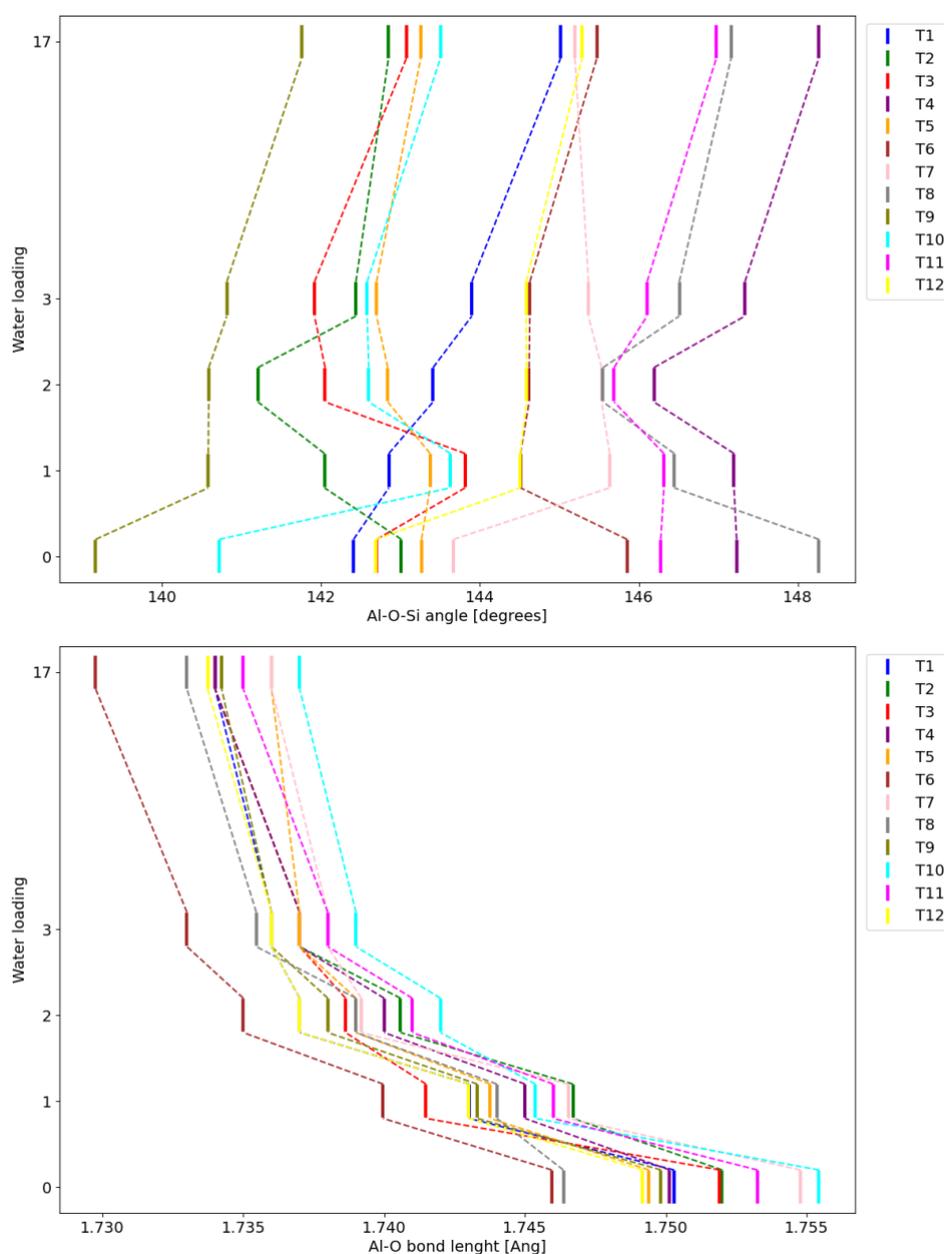
Fig. S7: Structural parameters (Al-O-Si angle - top and Al-O bond - bottom) at different water loadings.

Table S6: Average Al-O bond lengths and Al-O-Si angles in models with different water loading

Water loading	Average d(Al-O) [Å]	Average α(T-O-T) [°]
0	1.751	143.77
1	1.744	144.24
2	1.739	143.74
3	1.737	144.07
17	1.734	144.82

Table S7 shows that as water loading increases, the solvation percentage and the average distance between the oxygen in H_3O^+ ion and the Al atom both rise. As water loading increases, solvation reaches 82.6% with 2 water molecules and is nearly complete (99.5% and 99.7%) with higher loadings. Correspondingly, the average distance grows from 3.2 Å with 1 water molecule to 12.0 Å with 17 water molecules. The proton is considered to be solvated if the oxygen atom of any water molecule is closer to the proton than any oxygen atom from the zeolite framework. This is consistent with our prior work, [4] and allows for a well-defined boundary to be defined between common scenarios, such as water H-bonded to a proton on the framework, and hydronium ions close to a deprotonated framework site.

Table S7: Average solvation percentage and average distance between the oxygen in H_3O^+ and the Al atom.

Water loading	Solvation [%]	Average d(H_3O^+-Al) [Å]
0	0.0	-
1	4.0	3.2
2	82.6	3.4
3	99.5	3.7
17	99.7	12.0

To better analyze the similarities between the behaviour of different T-sites as a function of the water loading, a principal component analysis (PCA) was carried out on vectors composed of relative changes in chemical shifts with water loading for each site, and the resulting principal components for each T-site were clustered with the K-means algorithm. This resulted in separation of the T-sites into three distinct groups (see Fig. S8). The first cluster contains T7, T9, T10, and T12 and it corresponds to T-sites located along the sides of the 8 ring, which form the sinusoidal channel. The second cluster includes T2, T6, and T8, which are situated on the same side of the straight channel. The third cluster contains T1, T4, T5, T11, and T3 which do not share any structural parameters in the framework. However, T1, T4, T5, and T11 are situated in the same 6 ring near the intersection. The first principal component explains 77 % of the variance and is most strongly influenced by the change in chemical shift as the number of water molecules increases from 0 to 1. The second principal component, which explains 16 % of the variance, is predominantly affected by the change from 1 to 2 water

molecules. This indicates that the behavior of chemical shift in water is largely governed by the changes in chemical shift at low water concentrations.
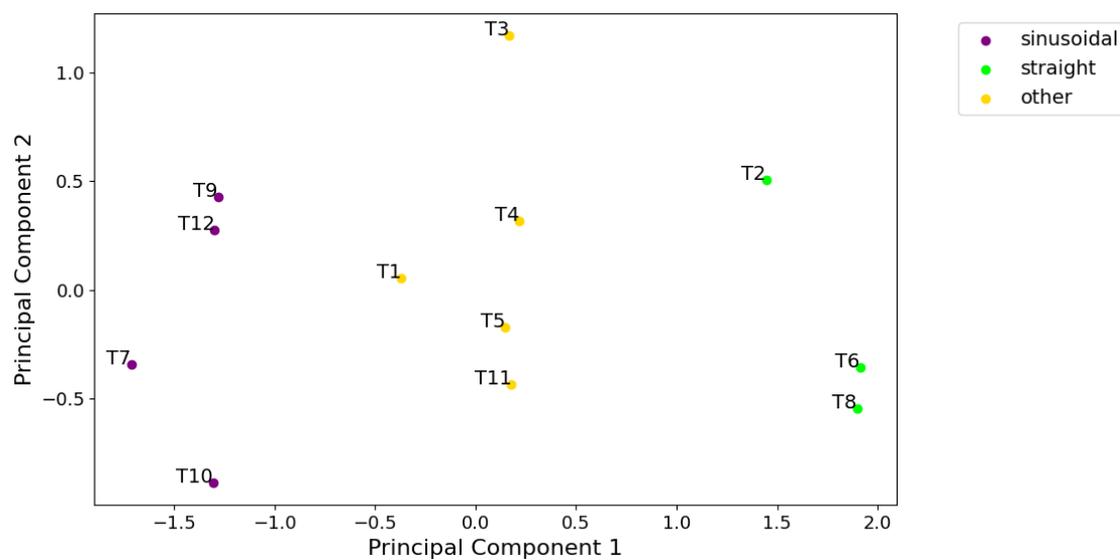
Fig. S8: A principal component analysis (PCA) carried out on vectors composed of relative changes in chemical shifts with water loading for each T-site (processing data shown in Fig. 4), followed by K-means-based clustering.

S8. Dynamic sampling and high chemical shift of T9

To assess the impact of molecular dynamics on the T-O-T angles, we compared three models: the initial (siliceous) MFI structure from the IZA database, the same structure after local optimization using a neural network potential (NNP), and the average structure from molecular dynamics (MD) simulation. Table S8 shows the average T-O-T angles for each model. Notably, the T9 site has the smallest average T-O-T angle only in the MD model, indicating that dynamic sampling significantly influences this angle. This reduction in angle is associated with the high chemical shift observed for the T9 site in the orthorhombic MFI cell. The average T-O-T angles are in Table S8 and visualized in Fig. S9.

Table S8: Average T-O-T angles [in degrees] for different structural models: MFI structure from the IZA database, locally optimized (LO) structure, and molecular dynamics (MD) simulations of the same structure.

T-site	MFI	LO structure	MD
T1	149.3	147.1	147.8
T2	141.7	142.3	148.4
T3	147.6	152.1	149.7
T4	154.6	153.8	152.1
T5	149.1	150.3	148.0
T6	151.2	155.4	149.3
T7	146.8	146.6	147.7
T8	144.7	147.3	150.7
T9	147.3	148.1	146.2
T10	150.0	151.0	146.5
T11	148.9	152.8	149.9
T12	145.8	152.7	149.1
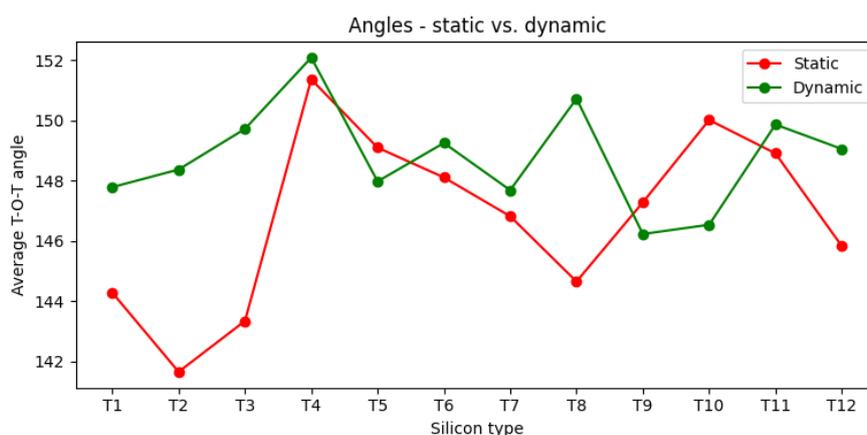
Fig. S9: Average angles in the siliceous MFI unit cell. The red line represents the average angles from the locally optimized MFI structure, while the green line shows the average T-O-Si angles from a 250 ps molecular dynamics simulation.

S9. Temperature Effects

To examine how temperature affects chemical shifts, we modelled all Brønsted acid site (BAS) configurations for the T5 site and applied Boltzmann averaging to obtain a representative chemical shift for each temperature (see Section S6). Table S9 presents the chemical shifts for various temperatures and water loadings. Fig. S10 highlights the structural changes that correspond to these temperature-dependent shifts. The observed trends in chemical shift align with the structural variations seen across different temperatures.

Table S9: Chemical shift values in ppm for every water loading and temperature.

Temperature [K]	0w	1w	2w	3w	17w
250	56.3	56.3	57.3	57.4	57.3
300	56.2	56.3	57.6	57.3	56.7
350	56.2	56.3	57.4	57.2	56.5
400	56.9	56.2	57.1	56.8	56.2
450	57.1	56.3	56.8	56.5	56.0
500	57.1	56.3	56.5	56.4	55.8
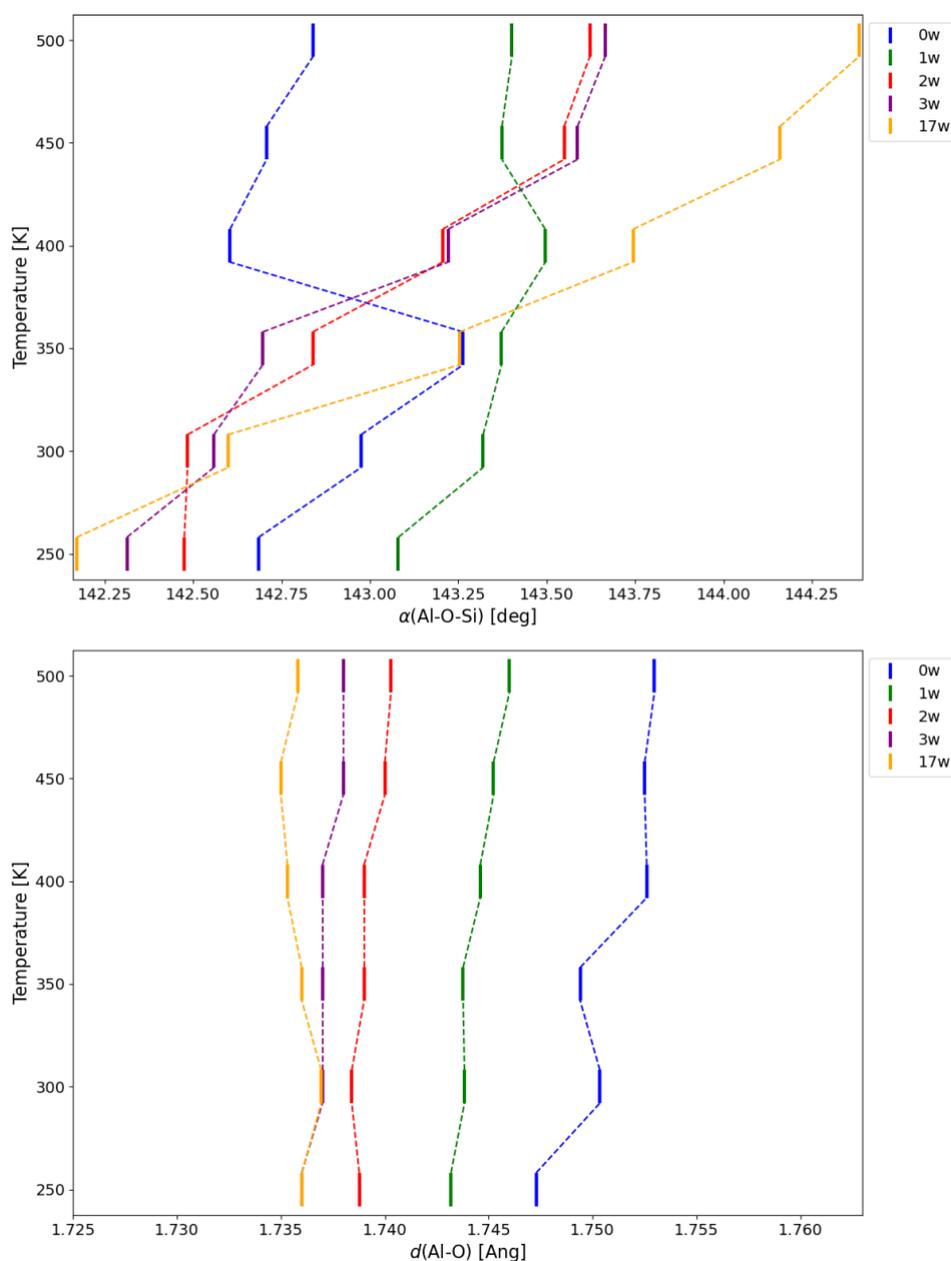
Fig. S10: Structural parameters at different temperatures - average Al-O-Si angle (left) and average Al-O bond length.

The abrupt changes observed in the T-O-T angle and Al-O bond length are due to shifts in BAS configuration. Specifically, when the BAS proton is situated on O14, it forms a strong hydrogen bond (intrazeolitic hydrogen bond - IZB) with the framework. Below 350 K, the hydrogen bond fraction remains high, ranging from 0.8 to 0.9. However, this fraction decreases significantly to around 0.6 to 0.5 at temperatures above 400 K, which correlates with the rise in chemical shift. Fig. S11 illustrates this distinctive structural motif and its temperature-dependent behaviour.
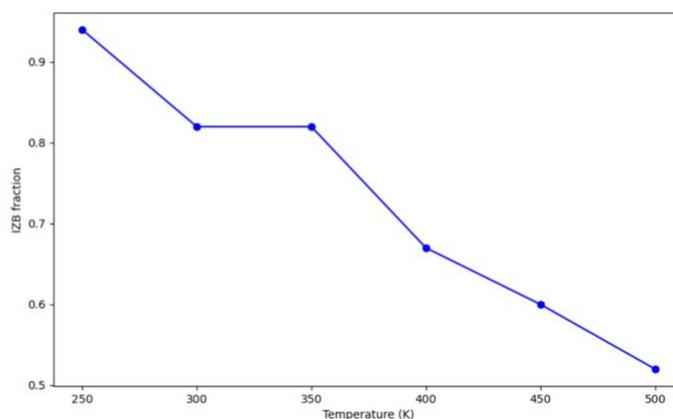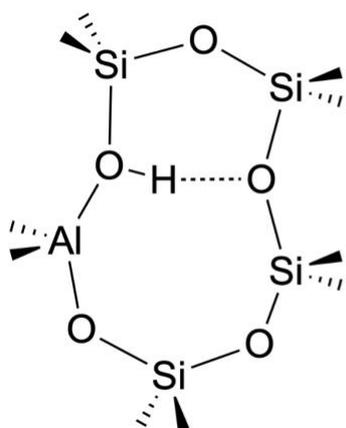
Fig. S11: Intrazeolitic hydrogen bond (left) and the fraction of the MD simulation during which the IZB is present (right).

Table S10 shows the average distance between oxygen atoms near the aluminium center and the closest hydrogen atom. As the water content increases, hydrogen atoms move further from the aluminium center. This increased distance allows the framework to relax, leading to larger T-O-T angles. This relaxation effect is evident in the structural trends observed with higher water loadings, where the greater separation between hydrogen atoms and the aluminium center contributes to the observed increase in T-O-T angles.

Table S10: Average distance [in Å] between oxygen atoms near the Al-center and the closest hydrogen atom.

Water loading	250 K	300 K	350 K	400 K	450 K	500 K
0w	2.61	2.60	2.60	2.60	2.60	2.60
1w	2.56	2.57	2.57	2.57	2.58	2.59
2w	2.68	2.57	2.59	2.69	2.76	2.76
3w	2.57	2.56	2.59	2.66	2.70	2.73
17w	2.81	3.02	3.18	3.40	4.02	5.12

Table S11 presents the solvation percentages across various water loadings and temperatures. As the temperature rises, both the solvation percentage and the average distance between the oxygen in H_3O^+ and the aluminium atoms increase.

Table S11: Solvation and proton dynamics analysis for models with different water loadings and different temperatures.

Water loading	Temperature [K]	Solvation [%]	Distance H_3O^+-Al [Å]
1	250	2.9	3.17
	300	3.1	3.18
	350	3.3	3.18
	400	3.7	3.18
	450	3.6	3.18
	500	3.8	3.20
2	250	98.2	3.36
	300	97.4	3.37
	350	93.2	3.36
	400	87.0	3.37
	450	77.6	3.36
	500	70.7	3.38
3	250	100.0	3.76
	300	100.0	3.71
	350	99.9	3.70
	400	99.1	3.69
	450	97.8	3.66
	500	96.6	3.70
17	250	100.0	9.09
	300	99.2	9.76
	350	100.0	11.57
	400	100.0	12.05
	450	99.9	19.78
	500	100.0	26.56

S10. Aluminium pairs

A total of 20 aluminium pairs were modelled, ensuring that at least one of the aluminium atoms (BAS 1) was T12O8. The dynamically averaged structural parameters for these models are summarized in Table S12. In the hydrated model with isolated Al atom in the T12 position, the chemical shift is 55.3 ppm, the average T-O-T angle is 145.39 degrees, and the average Al-O bond length is 1.733 Å.

Table S12: Details regarding the paired models with BAS configurations, including the chemical shift of each Al atom and other structural parameters, are provided. The structural parameters are provided for fully hydrated models (17 water molecules per unit cell).

BAS 1	BAS2	Al-Al	Shift 1	Shift 2	T-O-T 1	T-O-T 2	Al-O 1	Al-O 2
T12O8	T6O15	4.77	55.6	56.0	146.01	146.22	1.735	1.73
T12O8	T4O9	4.85	56.3	52.8	144.64	147.73	1.736	1.736
T12O8	T2O5	5.49	57.7	57.3	142.69	143.07	1.736	1.732
T12O8	T5O13	5.03	56.2	57.4	144.07	141.61	1.736	1.739
T12O8	T7O17	5.75	55.4	55.0	145.28	144.31	1.734	1.736
T12O8	T10O24	4.37	55.3	56.0	145.52	142.24	1.736	1.739
T12O8	T3O10	5.81	55.2	56.5	145.26	143.26	1.734	1.734
T12O8	T11O14	5.06	56.0	53.6	144.61	145.92	1.735	1.737
T12O8	T8O17	5.03	56.2	53.3	144.69	146.32	1.736	1.743
T12O8	T2O5	5.2	55.8	57.6	144.41	142.61	1.739	1.733
T12O8	T7O17	4.56	56.4	55.6	144.05	144.15	1.736	1.738
T12O8	T9O15	4.65	55.1	58.9	145.27	140.68	1.736	1.736
T12O8	T1O1	7.44	55.1	56.1	144.99	144.41	1.734	1.735
T12O8	T8O6	7.75	55.8	53.6	144.51	147.71	1.735	1.732
T12O8	T7O17	6.53	54.8	54.4	145.81	144.56	1.734	1.738
T12O8	T1O1	7.06	56.4	56.7	143.90	143.58	1.735	1.736
T12O8	T8O6	7.65	55.2	55.9	145.50	143.94	1.734	1.734
T12O8	T9O15	7.2	55.3	57.8	145.27	141.60	1.734	1.735
T12O8	T4O9	6.45	55.8	52.5	144.59	147.78	1.735	1.736
T12O8	T3O10	6.71	56.0	56.7	144.61	143.07	1.735	1.735

To validate the accuracy of the KRR model for description of aluminium pairs, the lowest-energy structures from the NNP-driven MD trajectories of the hydrated (17 water molecules) MFI models were locally optimized using NNPs and then re-optimized with DFT with PBE functional. For these structures, the chemical shieldings predicted by DFT and KRR are

compared (Figure S12). The overall MAE of the KRR prediction is 0.8 ppm, which is consistent with the MAE of KRR prediction for MFI models with a single aluminium per unit cell. This finding verifies the predictive consistency of the KRR method with NNP data against the DFT reference.
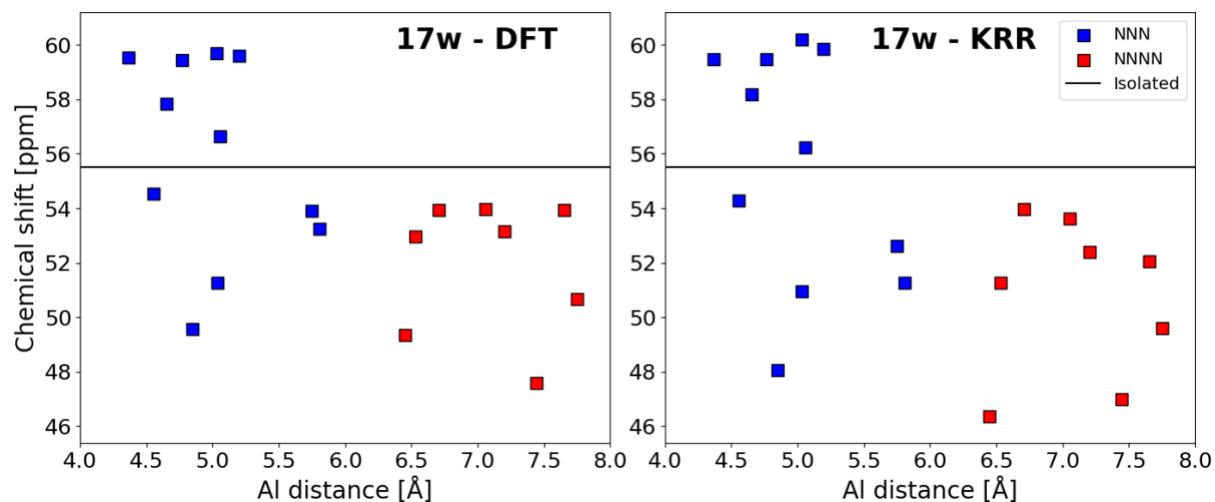
Fig. S12: The comparison between chemical shift predicted by the DFT (left) and KRR (right) method for aluminium paris with the hydrated MFI model (17 waters per unit cell).

S11. Higher aluminium content

A total of 45 models featuring multiple aluminium atoms were generated. The positions of the aluminium atoms were randomly assigned, and structures containing Al-O-Al motifs were excluded following Löwenstein's rule. This resulted in three distinct sets of models: 15 samples with 4 aluminium atoms (Si/Al = 22), and 15 samples with 5 aluminium atoms (Si/Al=17), 15 samples with 6 aluminium atoms (Si/Al ratio of 14). Table S13 presents the chemical shift values of these models for every T-site, specifically for the fully hydrated case with 17 water molecules per unit cell.

Table S13: T-sites in each model of higher Al content and their corresponding chemical shifts. The table is organized into three sections based on Si/Al ratios, with different models separated by thicker borders.

Si/Al = 23					
T-site	**Shift [ppm]**	**T-site**	**Shift [ppm]**	**T-site**	**Shift [ppm]**
T10	62.0	T12	61.6	T7	60.2
T10	61.9	T7	60.8	T8	61.9
T5	61.5	T10	61.9	T12	61.6
T11	62.3	T5	63.8	T1	61.3
T8	62.0	T9	64.3	T10	62.1
T5	63.4	T8	62.8	T6	63.4
T12	62.2	T4	61.8	T5	63.8
T12	62.7	T12	61.5	T7	60.4
T5	64.1	T7	60.6	T9	64.7
T9	64.4	T10	61.3	T6	61.7
T5	62.8	T12	61.4	T12	61.4
T12	62.8	T5	66.0	T5	63.7
T8	62.6	T7	61.7	T12	62.9
T9	65.3	T1	62.0	T1	62.6
T12	60.2	T9	64.1	T5	63.4
T10	63.0	T7	61.1	T3	62.0
T6	63.9	T11	60.5	T7	61.0
T12	61.4	T4	61.3	T6	61.1
T4	61.2	T9	63.6	T1	64.0
T10	61.8	T4	60.7	T12	60.9

Si/Al = 18					
T-site	**Shift [ppm]**	**T-site**	**Shift [ppm]**	**T-site**	**Shift [ppm]**
T5	62.3	T7	60.3	T9	64.4
T4	59.9	T2	62.2	T9	64.0
T7	61.6	T6	62.7	T7	64.1
T12	63.1	T3	62.1	T11	61.9
T11	60.4	T11	60.9	T5	61.8
T9	64.2	T10	63.6	T7	61.0
T12	61.9	T6	62.1	T3	61.0
T5	61.9	T6	63.0	T6	61.4
T10	62.4	T7	60.0	T9	64.6
T6	62.9	T9	65.0	T10	61.9
T7	62.2	T10	61.2	T12	61.8
T12	63.3	T9	65.2	T5	63.2
T9	63.5	T11	62.9	T7	62.4
T7	61.4	T9	66.1	T1	61.8
T6	62.9	T12	62.6	T7	62.3
T12	63.0	T6	61.8	T10	62.4
T1	61.6	T7	61.5	T12	61.5
T7	59.5	T9	65.8	T9	63.5
T10	61.5	T6	62.0	T4	60.5
T6	62.8	T1	62.9	T5	62.7
T10	62.4	T7	61.0	T3	61.0
T10	63.6	T9	64.5	T5	64.2
T5	62.8	T9	64.4	T7	60.3
T9	64.6	T9	63.3	T6	62.1
T9	63.9	T7	60.8	T12	62.9

Si/Al = 15					
T-site	Shift [ppm]	T-site	Shift [ppm]	T-site	Shift [ppm]
T9	66.5	T6	61.3	T12	61.6
T9	65.0	T5	62.0	T9	63.1
T10	59.5	T6	61.7	T7	61.9
T6	61.7	T9	65.6	T12	61.6
T12	63.6	T1	62.5	T7	61.7
T1	63.4	T5	62.3	T7	63.2
T9	63.8	T12	62.9	T7	60.9
T7	61.0	T1	61.8	T2	61.6
T1	64.4	T9	65.5	T11	60.1
T12	58.9	T3	63.1	T7	61.6
T7	59.6	T7	60.8	T11	62.7
T9	64.5	T10	60.9	T10	63.4
T9	64.5	T10	60.9	T6	62.4
T7	64.5	T8	62.3	T1	62.9
T10	62.3	T10	62.4	T11	61.1
T9	63.7	T5	61.3	T12	62.2
T2	62.0	T10	62.9	T4	59.4
T9	63.5	T4	60.7	T9	65.3
T7	59.6	T9	64.8	T7	62.1
T10	61.9	T9	66.4	T9	64.5
T11	62.9	T11	60.3	T12	60.8
T9	65.4	T6	61.3	T10	64.0
T12	64.8	T12	63.0	T12	61.6
T9	64.9	T1	62.9	T3	63.7
T3	58.8	T3	64.5	T2	63.3
T6	61.9	T9	64.2	T7	61.0
T7	57.0	T10	62.1	T1	62.8

T1	61.6	T8	62.3	T11	65.0
T12	61.4	T7	62.1	T12	61.5
T11	61.2	T12	63.0	T7	61.2

S12. Modelling experimental spectra

To compare our chemical shift KRR predictions with experimental spectra, we need to determine the parameters C_Q and η. For consistency with Holzinger et al., [5] we utilized their reported experimental C_Q values, selecting the C_Q corresponding to the experimental resonance closest to each predicted chemical shift. We further assumed the η to be zero to isolate the pure chemical shift component. The resulting simulated spectrum using SIMPSON [6] for a 950 MHz NMR spectrometer is shown in Fig. S13.
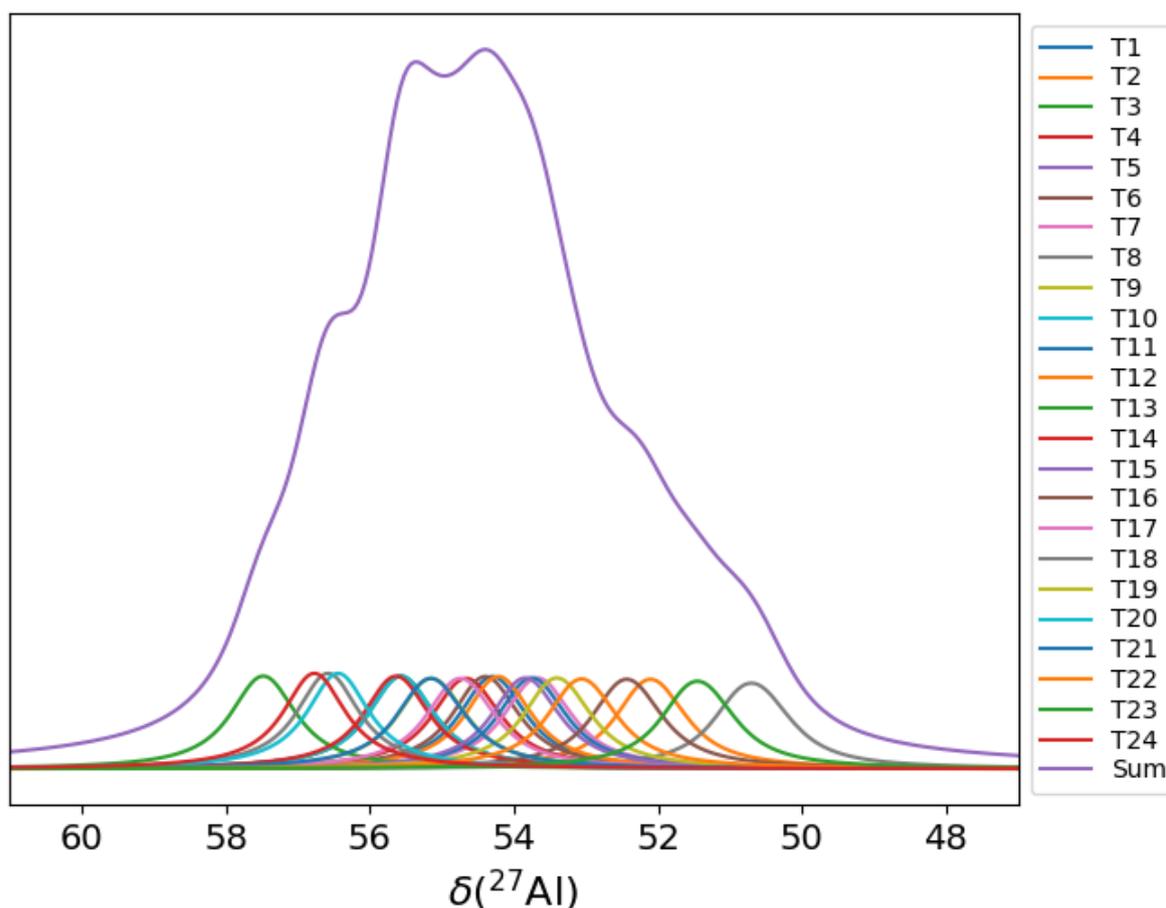
Fig. S13: NMR spectrum predicted by Kernel Ridge Regression (KRR) for a 950 MHz spectrometer, using C_Q values from Holzinger et al. and assuming $\eta = 0$.

We have also computed theoretical NMR spectra using a single representative structure, which was the locally optimized most stable structure obtained from the MD simulations. The theoretical DFT-calculated values (C_Q and η) were used to predict the signal shape, and the chemical shift was adjusted to the value obtained by MD and KRR approach. The resulting spectrum is shown in Figure S14a. This spectrum displayed considerable broadening, attributed to the use of a single structure. To refine this, we selected the narrowest peak (T7 - C_Q(T7) = 3 MHz) and replicated the chemical shifts with the same peak shape. The resulting spectrum is presented in Figure S14b.
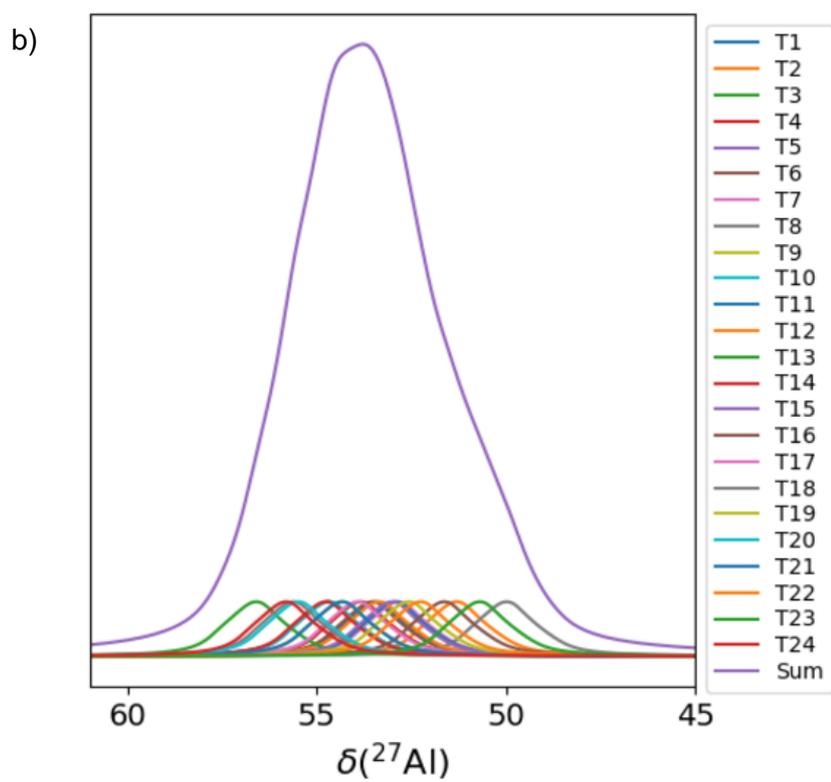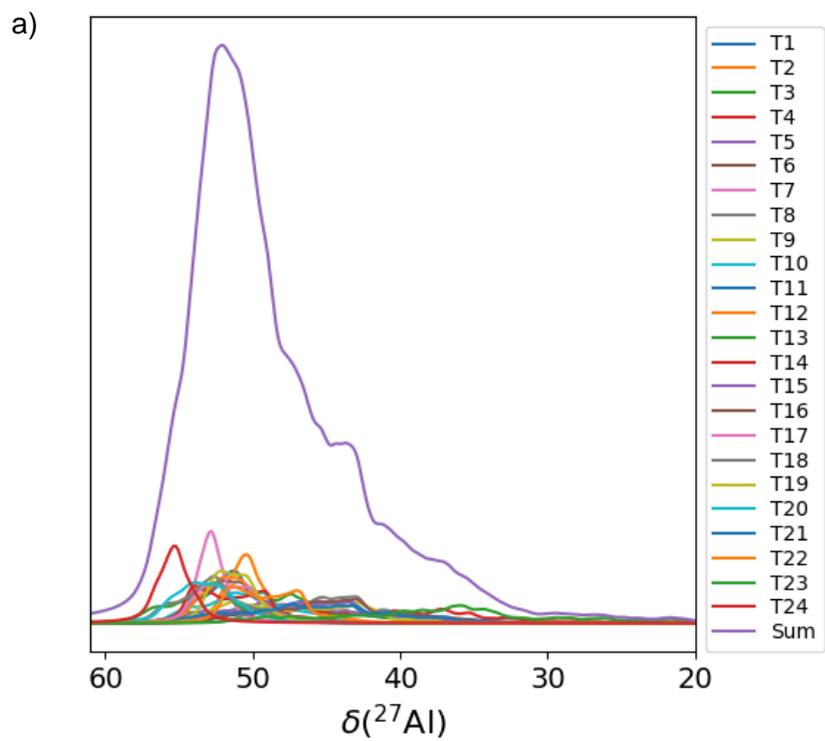
Fig. S14: Theoretical spectra using a) a single representative structure and b) selecting the narrowest peak for every T-site ($C_Q(T7) = 3$ MHz).

All of the modeled experimental NMR spectra exhibit significant signal overlap, complicating the assignment process when relying solely on chemical shifts.

The experimentally observed ^{27}Al NMR parameters from Holzinger et. al. [5] are in Table S14.

Table S14: Experimentally observed ^{27}Al NMR parameters taken from Ref. [5]

Resonance	Shift [ppm]	CQ [MHz]
1	52.0	2.06
2	52.8	1.95
3	53.4	1.83
4	54.1	1.81
5	54.8	1.74
6	55.4	1.63
7	56.1	1.77
8	56.7	1.6
9	57.4	1.29
10	58.4	1.59

S13. Other zeolites - MTT, TON, MOR, CHA

To extend the discussion of generality, and test the general application of the KRR model, we have included four additional zeolites which represent a stern, potentially out-of-domain challenge to the model: TON, MTT, CHA and MOR. For smaller unit cell dimensions, a supercell was created to minimize aluminium pairing. The model specifications are listed in Table S14. Molecular dynamics simulations were carried out following the protocol detailed in S1.

Table S15: Si/Al ratios, number of water molecules per aluminium atom for each zeolite framework, and supercell parameters

Zeolite	Si/Al	water/Al	Supercell
TON	47	4	1 × 1 × 2
MTT	71	5	3 × 1 × 1
MOR	95	15	1 × 1 × 2
CHA	35	10	1 × 1 × 1

Both TON and MTT have high TOT angles, which are not represented in the training database, as illustrated in Figure S15. This lack of coverage leads to the KRR model extrapolation, which in turn results in unrealistically low chemical shifts (see Table S15). Thus, we cannot conclude that the model in its current form can predict the shifts in such zeolites. To make a more general KRR model, one would have to include these structures to the training database, which is a straightforward extension for any user of the model presented in this work.

Table S16: Chemical shift values for MTT and TON

MTT T site	Shift [ppm]
T1	45.4
T2	49.0
T3	47.7
T4	46.4
T5	50.3
T6	43.7
T7	48.2

TON T site	Shift [ppm]
T1	50.1
T2	47.5
T3	50.8

T4	46.0
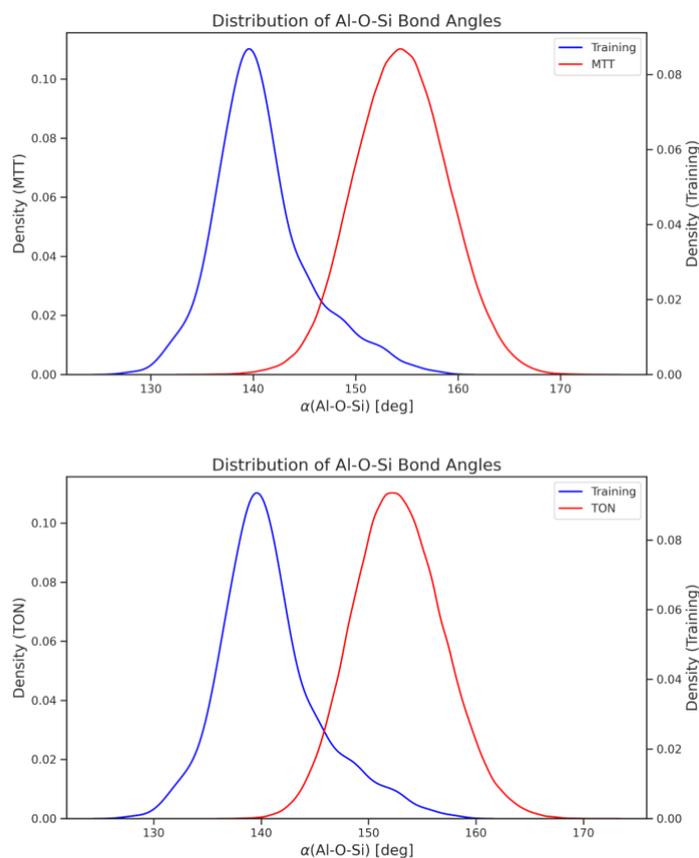
Figure S15: The distribution of average Al-O-Si angles in MTT (top) and TON (bottom) zeolites from NNP-driven MD at 350 K.

For MOR and CHA zeolites, the training database covers the range of bond angles adequately (Figure S16), and thus the KRR model does not have to extrapolate. It therefore predicts reasonable chemical shift values for both zeolites. The Al-O-Si angles distribution is in Figure S16. The chemical shifts for CHA and MOR are in Table S14. The average MOR chemical shift is 54.6 ppm, and the chemical shift of one T site for CHA is 57.6 ppm. Both of these results are 1-2 ppm lower than experimental data, [7,8] which is consistent with the results for ZSM-5.

Table S17: Chemical shift values for MOR and CHA

MOR T site	Shift [ppm]
T1	56.1
T2	54.0
T3	54.5
T4	57.8

CHA T site	Shift [ppm]
T1	57.6
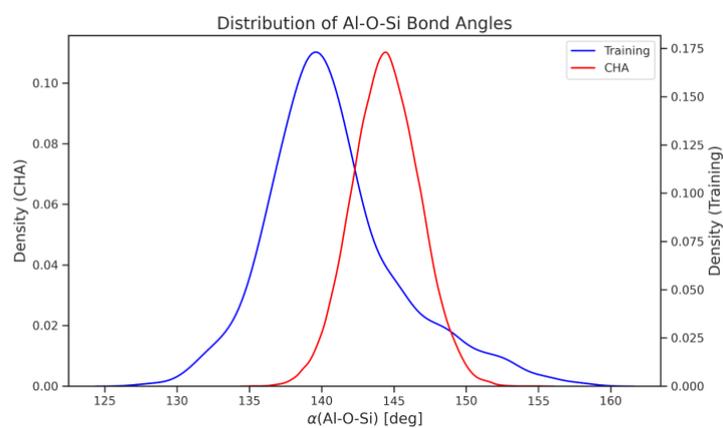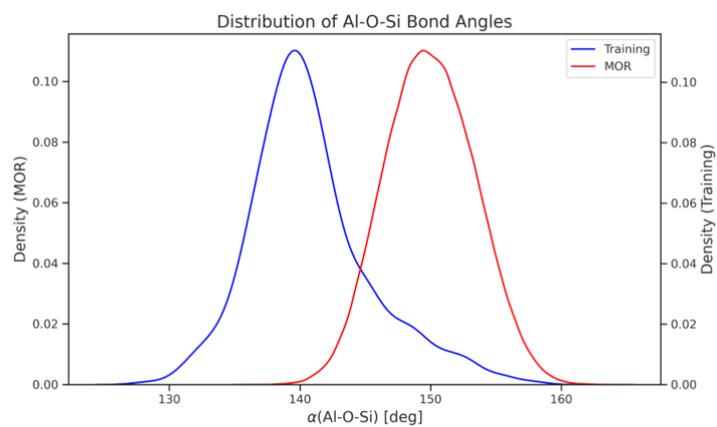
Figure S16: The distribution of average Al-O-Si angles in MOR (top) and CHA (bottom) zeolites from NNP-driven MD at 350 K.

S14. Training [29]Si KRR model

The KRR model was also trained on [29]Si chemical shifts utilizing the existing reference DFT data for H-MFI (Si/Al=95) (see Section S2). Each MFI unit cell contains 95 Si atoms and one Al atom, providing nearly 10 000 data points ([29]Si shieldings) for the KRR training. The training and testing error of the [29]SI KRR model were 0.2 ppm and 0.3 ppm, respectively. This KRR model was then applied to structures sampled from molecular dynamics simulation of the purely siliceous MFI, with the shielding-to-shift conversion assumed as a simple linear correlation taking the lowest experimental chemical shift (T9 - see Table S18) as a reference to which we pinned the lowest KRR predicted value. The experimental [29]Si data was sourced from Fyfe et. al. [9] The resulting chemical shifts are in Table S18 below. The MAE is 0.7 ppm, with a correlation coefficient of 0.87. The minor discrepancy is not surprising as the dataset is not diverse, and is not focused on siliceous systems with the training set made up from structures with aluminium and water molecules. The additional refinement of the database to improve the KRR accuracy is necessary to accurately predict [29]Si chemical shifts.

Table S18: The comparison between [29]Si KRR model and experimental values from Ref. [9].

T site	[29]Si exp. [ppm]	[29]Si KRR [ppm]	Error [ppm]
T1	-113.2	-113.3	-0.1
T2	-114.3	-115.9	-1.6
T3	-114.3	-115.7	-1.4
T4	-117.3	-116.6	0.7
T5	-113.6	-114.1	-0.5
T6	-114.8	-114.0	0.8
T7	-114.1	-113.2	0.9
T8	-116.3	-117.1	-0.8
T9	-112.2	-112.2	0.0
T10	-112.7	-112.4	0.3
T11	-115.2	-115.7	-0.5
T12	-114.4	-114.0	0.4

References

[1] Erlebach, A., Šípka, M., Saha, I. *et al.* A reactive neural network framework for water-loaded acidic zeolites. *Nat Commun* 15, 4215 (2024). https://doi.org/10.1038/s41467-024-48609-2

[2] Lei, C., Erlebach, A., Brivio, F., Grajciar, L., Tošner, Z., Heard, C. J., & Nachtigall, P. (2023). Correction: The need for operando modelling of 2727Al NMR in zeolites: the effect of temperature, topology and water. *Chemical Science*, *14*(47), 13987-13987. https://doi.org/10.1039/D3SC90195E

[3] Himanen, L., Jäger, M. O. J., Morooka, E. V., Federici Canova, F., Ranawat, Y. S., Gao, D. Z., Rinke, P., & Foster, A. S. (2020). DScribe: Library of descriptors for machine learning in materials science. *Computer Physics Communications*, *247*, 106949. https://doi.org/10.1016/j.cpc.2019.106949

[4] Heard, C. J., Grajciar, L., & Nachtigall, P. (2019). The effect of water on the validity of Löwenstein's rule. *Chem. Sci., 10* (22), 5705-5711. https://doi.org/10.1039/C9SC00725C

[5] Holzinger, J., Beato, P., Lundegaard, L. F., & Skibsted, J. (2018). Distribution of aluminium over the Tetrahedral Sites in ZSM-5 Zeolites and Their Evolution after Steam Treatment. *J. Phys. Chem. C*, 122(27), 15595-15613. https://doi.org/10.1021/acs.jpcc.8b05277

[6] Bak, M., Rasmussen, J. T., & Nielsen, N. C. (2000). SIMPSON: A general simulation program for solid-state NMR spectroscopy. *Journal of Magnetic Resonance, 147*(2), 296–330. https://doi.org/10.1006/jmre.2000.2179

[7] Bohström, Z., Arstad, B., & Lillerud, K. P. (2014). Preparation of high silica chabazite with controllable particle size. *Microporous and Mesoporous Materials*, 195, 294–302. https://doi.org/10.1016/j.micromeso.2014.03.030

[8] Petranovskii, V., Marzke, R. F., Diaz, G., Gomez, A., Bogdanchikova, N., Fuentes, S., Katada, N., Pestryakov, A., & Gurin, V. (2002). Characterization of H and Cu mordenites with varying SiO_2/Al_2O_3 ratios, by optical spectroscopy, MAS NMR of ^{29}Si, ^{27}Al, and ^{1}H, temperature programmed desorption and catalytic activity for nitrogen oxide reduction. *Studies in Surface Science and Catalysis*, 142, 815–822. https://doi.org/10.1016/S0167-2991(02)80106-9

[9] Fyfe, C. A., Grondey, H., Feng, Y., & Kokotailo, G. T. (1990). Natural-abundance two-dimensional silicon-29 MAS NMR investigation of the three-dimensional bonding connectivities in the zeolite catalyst ZSM-5. *Journal of the American Chemical Society*, 112(24), 8812–8820. https://doi.org/10.1021/ja00180a024