Supplementary Information

Aluminium Siting in Zeolite RTH From a Combined Machine Learning – NMR Approach

Daniel Willimetz[1], Joaquin Martínez-Ortigosa[2,3], Deborah Brako-Amoafo[1,4], Lukáš Grajciar[1], J. Alejandro Vidal-Moya[2], Carlos Bornes[1], Vincent Sarou-Kanian[5], Andreas Erlebach[1], Fernando Rey[2], Teresa Blasco[2*], Christopher J. Heard[1*]

[1]Department of Physical and Macromolecular Chemistry, Charles University in Prague, Hlavova 8, 12843, Praha 2, Czech Republic

[2]Instituto de Tecnología Química, Universitat Politècnica de València - Consejo Superior de Investigaciones Científicas (UPV-CSIC), Avenida Los Naranjos s/n, 46022 Valencia, Spain

[3]Laboratorio de Nanotecnología Molecular, Departamento de Química Inorgánica, Universidad de Alicante, Ctra. San Vicente-Alicante s/n, 03690, Alicante, Spain [4]EaStChem School of Chemistry, University of St Andrews, Purdie Building, St Andrews KY16 9ST, UK

[4]CEMHTI-CNRS, CNRS UPR3079, 1D Avenue de la Recherche Scientifique, 45071 Orléans cedex 2, France

[*]Corresponding authors: heardc@natur.cuni.cz, tblasco@itq.upv.es

S1 Molecular dynamics

Molecular dynamics simulations were performed in the NVT ensemble using the Atomic Simulation Environment (ASE) version 3.23.0. [1] Reactive neural network potentials for water-loaded acidic zeolites [2] with a radial cutoff of 6 Å and created by the SchNet-Pack 1.0.0 [3] package, were employed as interaction potentials. These potentials provide DFT-level accuracy while offering a computational speed-up of approximately 1000× compared to traditional *ab initio* calculations. [2] A Nosé-Hoover thermostat was used to maintain the temperature at 350 K. The simulations were run for 1 ns with a timestep of 0.5 fs, saving every 100th structure for subsequent structural analysis or chemical shift predictions.

Energy convergence for each MD simulation was verified through visual inspection of energy versus time plots. An example of energy equilibration is shown in Figure S1.

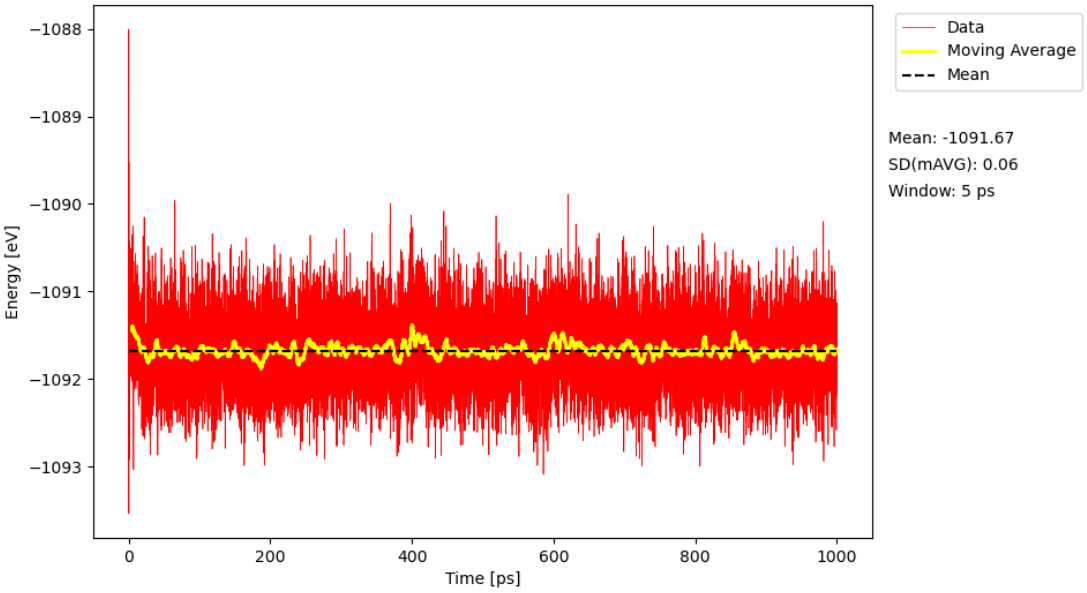

Figure S1: The energy as a function of time (red) is shown for MLIP MD simulations conducted at 350 K for the RTH zeolite with a Si/Al ratio of 31 and five water molecules per unit cell. A moving average with a window of 5 ps is applied to smooth the energy profile (yellow).

The stability of the MLIP and the ASE driver for MD simulations was tested against Born-Oppenheimer *ab initio* molecular dynamics (AIMD) simulations in our previous work, demonstrating an excellent match between the two methods. [4]

S2 DFT NMR calculation

DFT NMR calculations were performed using CASTEP with the PBEsol functional. The convergence test for the RTH and $Al(acac)_3$ is in Figure S2.

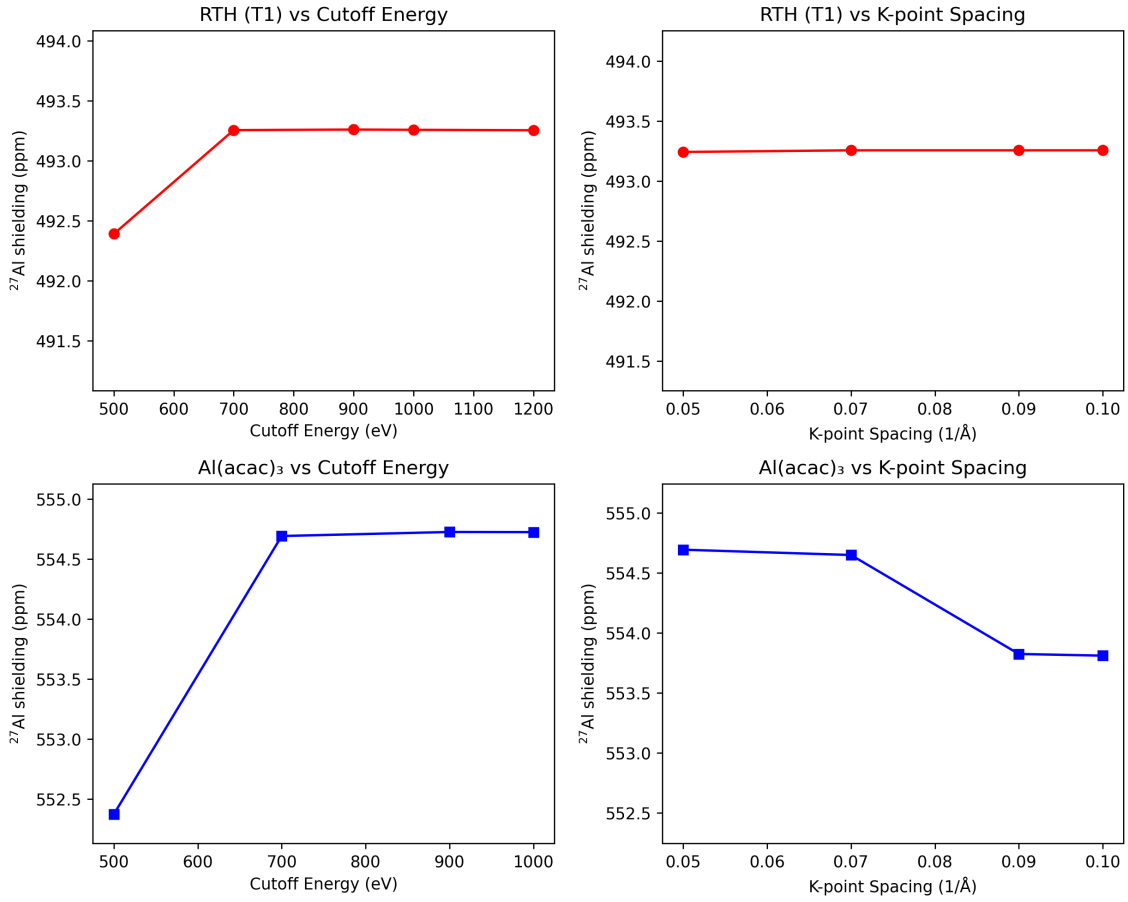

Figure S2: Benchmark of plane-wave energy cutoff (left) and k-point spacing (right) for single-point shielding calculations of RTH with the aluminum atom at the T1 site (top) and $Al(acac)_3$ (bottom).

For RTH, plane-wave energy cutoff of 700 eV and a Monkhorst–Pack k-point grid of $2 \times 1 \times 2$ was employed. For the reference compound $Al(acac)_3$, a finer k-point sampling of 0.05 and cutoff of 1000 eV was utilized to ensure full convergence of the shielding tensor calculations for precise referencing of the chemical shieldings. A 10 ps AIMD of $Al(acac)_3$ was performed, from which 100 frames were extracted and recalculated using DFT to obtain a time-averaged shielding value of 552.8 ppm. This value was used as the reference shielding for the conversion from shielding to chemical shift.

To assess the impact of MD sampling and the accuracy of the kernel ridge regression

(KRR) model, 50 snapshots were extracted from a 1 ns MD trajectory and evaluated using DFT. These same configurations were also evaluated using the KRR model to directly compare with the DFT results. Furthermore, KRR results are compared with the 2p-LASSO method, introduced by Lei et al.[5], which predicts chemical shifts as a function of both T–O–T angles and Al–O bond lengths, and with the correlation based solely on the T–O–T angle proposed by Lippmaa.[6] The outcomes are summarized in Table S1. The mean absolute error (MAE) between the DFT and KRR predictions over the 50 frames is 0.33 ppm for ^{27}Al and 0.31 ppm for ^{29}Si, demonstrating the high accuracy of the KRR models. The MAE between the MD-averaged KRR prediction and the average over the 50 KRR-evaluated snapshots is 0.34 ppm for ^{27}Al and 0.14 ppm for ^{29}Si, indicating that the selected subset of 50 frames effectively captures the structural variability of the full trajectory. However, the standard error of the mean estimated from the autocorrelation function is 0.7 ppm for ^{27}Al and 0.4 ppm for ^{29}Si, whereas the statistical error of the 1 ns MD average is only around 0.05 ppm for both types atoms, underscoring the robustness of the machine learning approach. Moreover, the range of chemical shifts over the 50 frames exceeds 10 ppm for each T site, highlighting the strong sensitivity of the chemical shift to dynamical effects. For a single locally optimized structure, the ^{27}Al chemical shifts are 61.2, 58.0, 59.8, and 56.8 ppm for T1, T2, T3, and T4, respectively. This clearly demonstrates the necessity of dynamical sampling, in agreement with our previous works.[4, 5]

Table S1: Comparison of predicted ^{27}Al chemical shifts (in ppm) using different approaches, KRR, 2p-LASSO,[5] and Lippmaa[6] refers to the prediction based on MD-averaged structures with their respective methods, DFT (50) is the average from 50 frames with direct DFT calculation, and KRR (50) refers to kernel ridge regression applied to the same 50 frames and errors obtained using the autocorrelation function.[7]

T-site	KRR	2p-LASSO	Lippmaa	DFT (50)	KRR (50)
T1	59.9 ± 0.04	58.9 ± 0.03	58.9 ± 0.03	59.2 ± 0.6	59.8 ± 0.6
T2	58.6 ± 0.04	59.1 ± 0.04	59.1 ± 0.04	59.2 ± 0.7	58.7 ± 0.6
T3	57.4 ± 0.05	58.1 ± 0.05	58.9 ± 0.05	56.7 ± 0.7	56.6 ± 0.7
T4	61.8 ± 0.03	62.4 ± 0.04	61.8 ± 0.04	62.4 ± 0.6	62.2 ± 0.6

S3 KRR training

Both ^{27}Al and ^{27}Si chemical shift prediction models using Kernel Ridge Regression (KRR) implemented in the scikit-learn 1.1.3 Python package [8] were trained on SOAP descriptors generated with DScribe 1.2.2 [9]. The following SOAP hyperparameters were applied:

```
species = 1 8 11 13 14
r_cut = 5.
n_max = 6
l_max = 6
sigma = 0.5
average = "off"
```

SOAP descriptor averaging was disabled to ensure that the predictions were based on individual atomic environments rather than the entire structure. This approach allows for the development of more generalized predictive models.

The KRR models were trained using isotropic chemical shieldings calculated with the CASTEP software. [10] These calculations employed the PBEsol exchange-correlation functional [11] and the GIPAW method, [12] with a plane-wave energy cutoff set to 700 eV. The same approach was employed in our previous works. [4, 5]

The predicted shielding values must be converted to chemical shift values for comparison with experimental data. This transformation is performed using simple linear regression based on reference experimental chemical shift values.

The prediction of the NNP-KRR model was compared with a set of DFT chemical shift calculations on two hundred structures uniformly sampled from the MD trajectories, using the CASTEP software,[10] for the purposes of verification (Figure S3).

S3.1 ^{27}Al KRR

The ^{27}Al KRR model was adopted from our previous work on the zeolite MFI. [4] This model was trained using a database comprising of approximately 2000 zeolite structures obtained from AIMD simulations conducted at various temperatures. The database in-

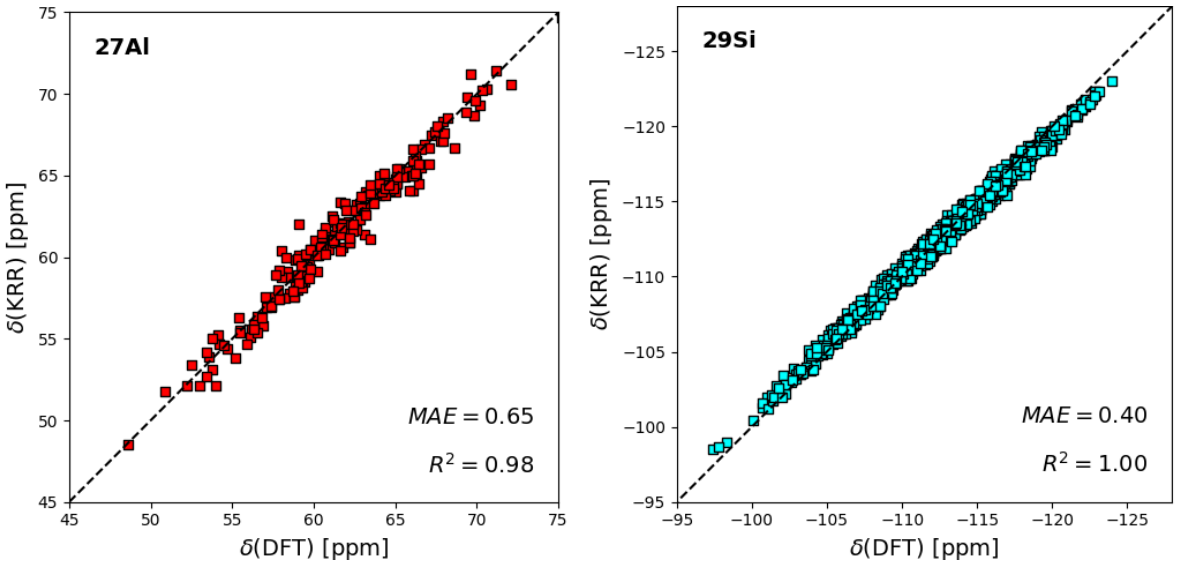

Figure S3: Correlation of the ^{27}Al and ^{29}Si chemical shift predicted by the KRR and the DFT method.

cludes zeolites with diverse topologies, varying Si/Al ratios, and different chemical compositions. In total, the training dataset contains over 4000 unique aluminum environments. For further details, we refer the reader to our prior publication. [4]

Before applying the ^{27}Al KRR model, it is essential to compare the structural parameters of the RTH zeolite with those in the training database to ensure that the predictions do not involve significant extrapolation. Such extrapolation has been observed to yield unrealistic chemical shifts, as seen in the case of zeolites TON and MTT. [4] The selected structural parameters for comparison include the Al-O bond length, the average Al-O-Si angle, and the O-Al-O angles. These parameters have been identified as the most critical for predicting ^{27}Al chemical shifts. [5] As shown in Figure S4, all structural parameters fall within the range provided by the database, with none being extrapolated.

To calculate the chemical shift from the chemical shielding, Al(acac)$_3$ was used as a reference. This compound serves as a solid-state reference in experimental measurements and has a chemical shift of 0.03 ppm. [13] The chemical shielding of Al(acac)$_3$ was determined to be 552.8 ppm based on 20 ps AIMD simulations. [5] The slope between the chemical shift and shielding was assumed to be -1. For more details, refer to our previous work. [4, 5] Thus, the chemical shielding σ is transformed to chemical shift δ

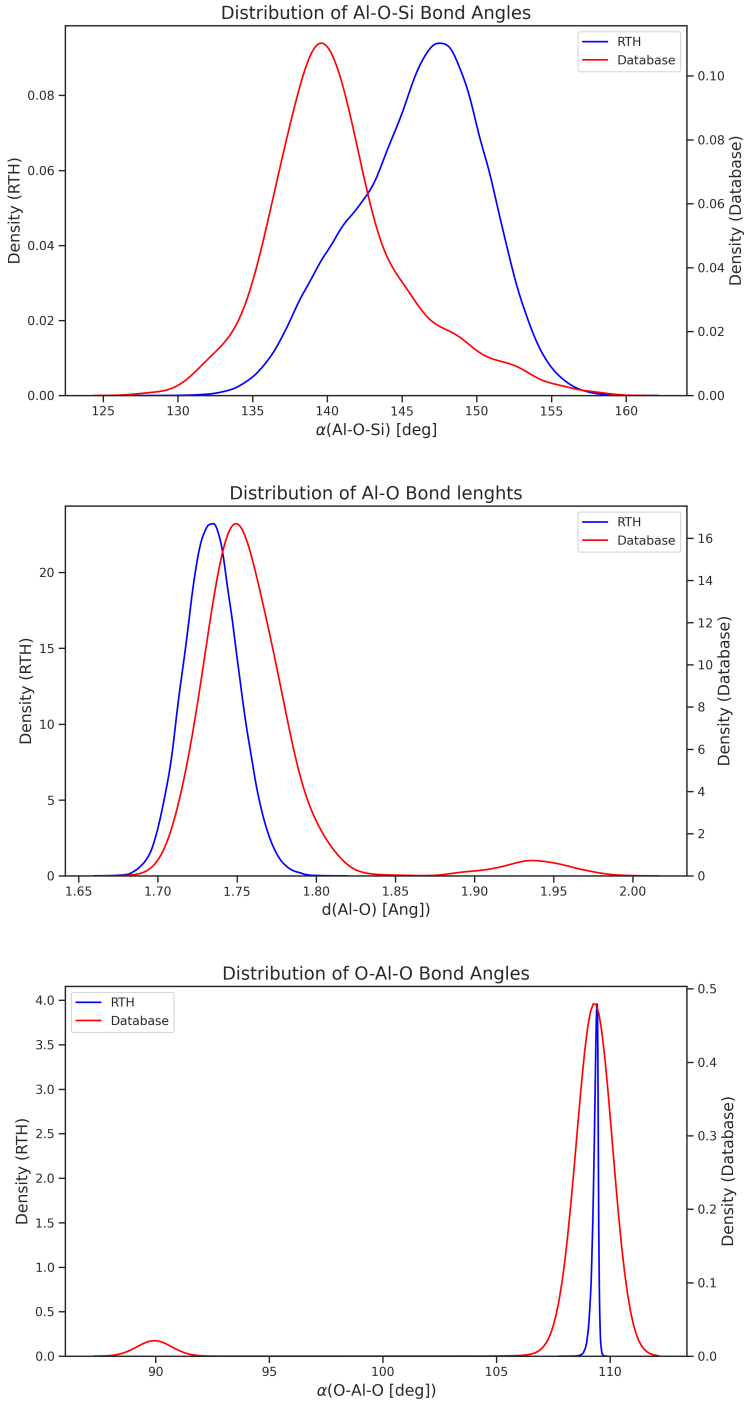

Figure S4: Structural comparison between the KRR training database and RTH is presented for the Al-O-Si angle (top), Al-O bond lengths (middle), and O-Al-O bond lengths (bottom). The structural parameters for RTH were derived from MLIP MD simulations at 350 K, with a Si/Al ratio of 31 and five water molecules per unit cell.

using following equation:

$$\delta(^{27}\mathrm{Al}) = -1 \cdot \sigma(^{27}\mathrm{Al}) + 552.8\,\mathrm{ppm} \tag{S1}$$

S3.2 ^{29}Si KRR

The ^{29}Si KRR model was trained using subsampled AIMD simulations of various zeolites (MOR, CHA, ABW, FER, and LTA) at different Si/Al ratios and water loadings at a temperature of 300 K.

The same training workflow was applied as in training the ^{27}Al KRR model, which is described in detail in our previous work. [4] The regularization parameter λ that produced the best-performing models was $4 \cdot 10^{-6}$. The test MAE is 0.4 ppm, similar to the ^{27}Al KRR model (0.5 ppm). The R^2 score exceeds 0.99, demonstrating an excellent correlation with the reference DFT data.

To convert the predicted ^{29}Si shielding $\sigma(^{29}\mathrm{Si})$ from the KRR model into the ^{29}Si chemical shift $\delta(^{29}\mathrm{Si})$, a calibration equation must be determined. For this purpose, 1 ns molecular dynamics simulations of five siliceous structures (FER, TON, CHA, cristobalite, and α-quartz) were performed using MLIP. The chemical shielding was then calculated with the trained KRR model.

A comparison between the KRR predicted shielding and experimental chemical shifts [14–17] is presented in Table S2. The data was fitted using linear regression, yielding the calibration equation:

$$\delta(^{29}\mathrm{Si}) = -0.9173 \cdot \sigma(^{29}\mathrm{Si}) + 287.47\,\mathrm{ppm}, \tag{S2}$$

which serves as a reference for converting chemical shielding into chemical shift.

S4 Computational models and zeolite samples

The siliceous RTH (Si-RTH) models were obtained directly from the IZA database. [18] The Al-RTH models were generated by replacing one silicon atom with aluminum. A proton was added to an oxygen atom neighboring the aluminum atom, acting as a charge-compensating cation and creating a Brønsted acid site (BAS). Furthermore, five water molecules were initialized to allow immediate access to both the aluminum and the hydrogen atom. The unit cell dimensions of Al-RTH were determined by X-ray diffraction

Table S2: ^{29}Si shieldings and chemical shifts for siliceous structures. Experimental ^{29}Si chemical shifts for FER [14], TON [15], CHA [16], cristobalite [17], and α-quartz [17] were taken from the corresponding literature.

Structure	Si type	σ(KRR) [ppm]	δ(exp) [ppm]
FER	Si1	439.4	-116.5
	Si2	435.1	-112.3
	Si3	437.0	-111.9
	Si4	437.0	-117.2
	Si5	440.9	-116.2
TON	Si1	436.4	-113.1
	Si2	434.7	-112.8
	Si3	434.5	-110.9
	Si4	436.7	-114.3
CHA	Si1	435.6	-111.4
Cristobalite	Si1	433.8	-108.5
Quartz	Si1	428.9	-107.1

(XRD) of the experimental sample with Si/Al ratio of 28 (Table S7).

Each aluminum atom can adopt four possible BAS configurations. Two BAS configurations per T-site were tested, chosen based on the orientation of the hydrogen atoms pointing towards the RTH pore channels. The chemical shift differences between the various starting BAS configurations were found to be less than 0.2 ppm for ^{27}Al and 0.1 ppm for ^{29}Si in Al-RTH, indicating that the initial configurations have minimal impact on the chemical shifts using the machine learning approach.

The models were created to reproduce the experimental conditions. The chemical composition of the zeolite samples and their XRD patterns are shown in Table S3 and Figure S5.

Table S3: Table of chemical composition of RTH samples.

Sample	Si/Al	Al/u.c.	OSDA/Al
RTH_{16}	16	1.1	27
RTH_{28}	28	2.0	39
RTH-SiO_2	∞	-	-

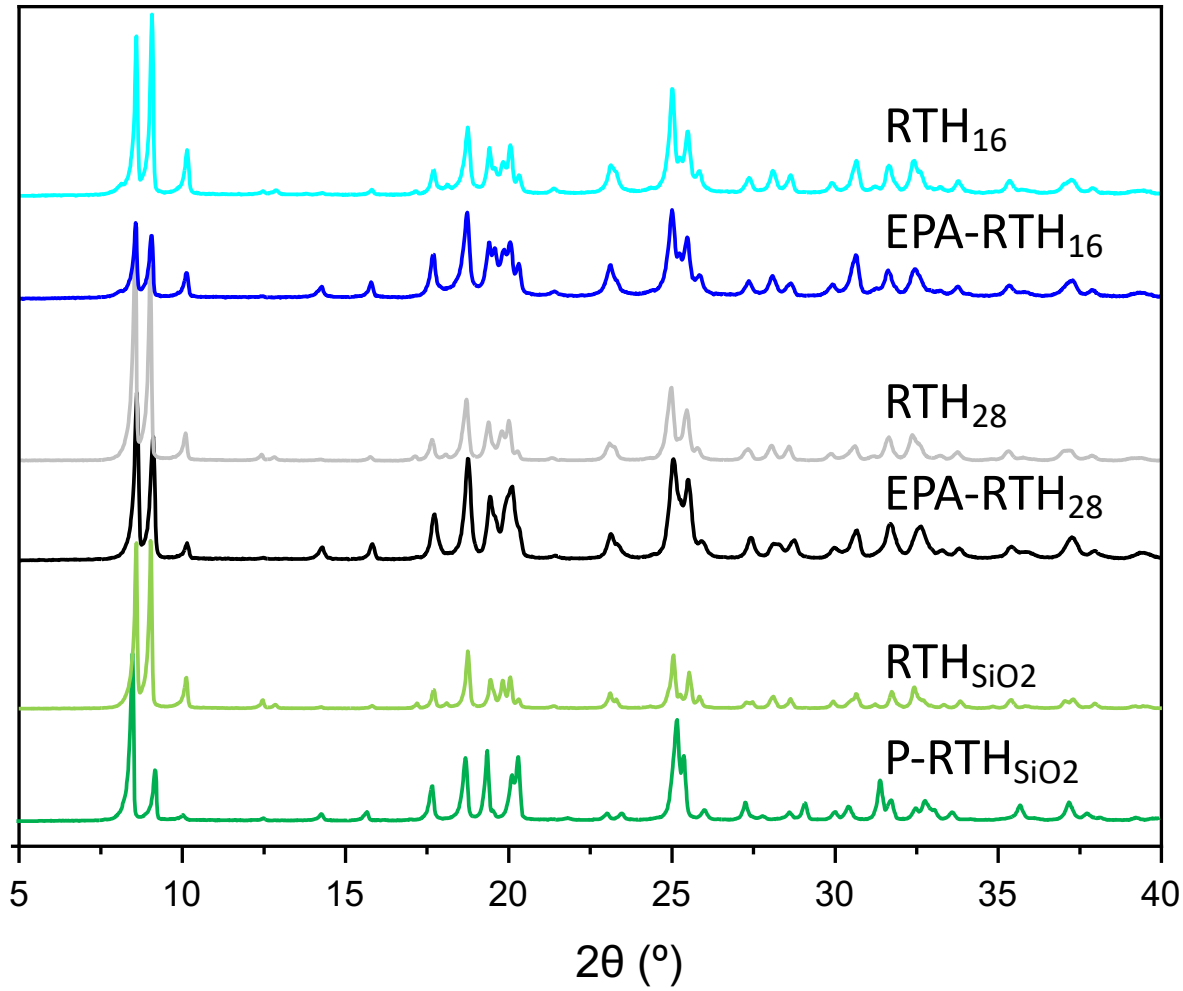

Figure S5: XRD patterns of RTH samples.

S5 ^{27}Al MQMAS NMR

The ^{27}Al MQMAS NMR spectra of the RTH_{16} and RTH_{28} samples are in Figure S6. The ^{27}Al sheared MQMAS NMR spectra of the RTH_{16} and RTH_{28} samples (Figure 2) exhibits two resonances in the isotropic chemical shift dimension. Both signals lie close to the F2=F1 diagonal, indicating that second-order quadrupolar broadening is minimal at 750 MHz for all T-sites. The spread along the isotropic axis reflects a distribution of chemical environments. Notably, the lower chemical shift peak exhibits a broader distribution, consistent with its assignment to three distinct T-sites (T1–T3), whereas the correlation peak centered at 64 ppm corresponds to Al at the unique T4 site.

The ^{27}Al MQMAS NMR spectrum in Figure S7 was fitted assuming a Czjzek distribution[19] to extract the isotropic chemical shifts and compare them with theoretical

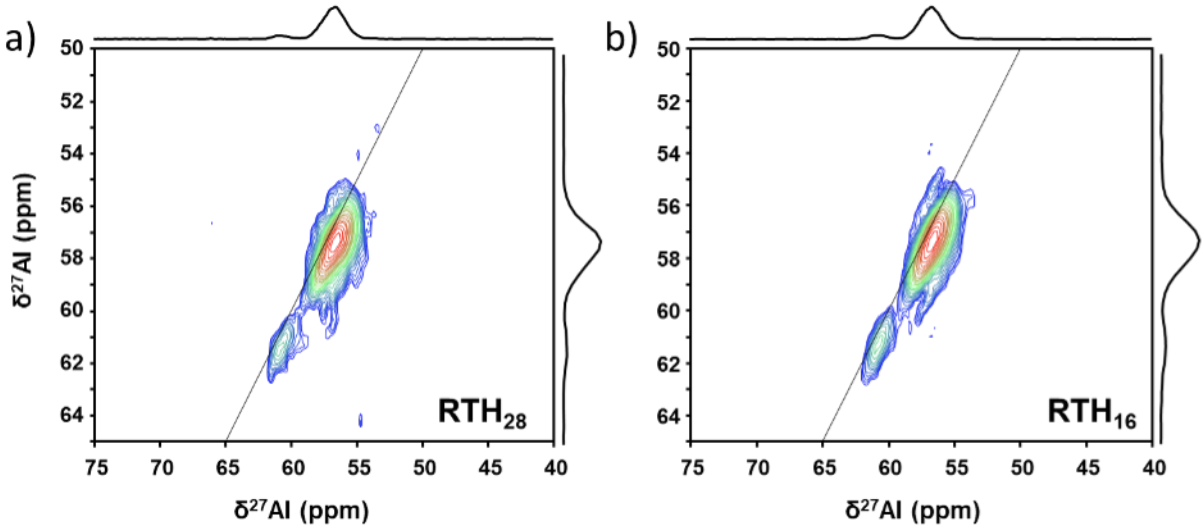

Figure S6: ^{27}Al MQ-MAS NMR measured at 750 MHz of the two RTH samples with Si/Al ratios a) 28 and b) 16.

calculations, all collected in Table 2.

S6 2D recoupling NMR experiments

Figure S8 presents the ^{29}Si-^{29}Si DQ-SQ MAS NMR spectra at different coupling times, demonstrating that 4 ms is the best recoupling time to observe only the nearest-neighbour (NN) silicon, since increasing the recoupling time shows up new signals corresponding to the next-nearest-neighbour (NNN) silicon.

Figure S9 shows an unmodified overlay of Figure 5. The ^{27}Al chemical shifts are higher than the center of the peak due to the quadrupolar nature of the nucleus. To improve the clarity and understanding of the plot, the ^{27}Al chemical shifts were adjusted by a constant amount of 1.3 ppm and 0.5 ppm for ^{29}Si to align with the experimental peaks. The offset is likely caused by imperfections in the shielding-to-shift conversion, inaccuracies in the DFT calculations used to train the KRR model or non optimal unit cell. However, since the chemical shift is a relative metric, applying a constant shift does not affect the model's accuracy or its ability to predict experimental measurements.

To investigate the effect around (57,-113) ppm peak, the transfer of magnetization (i.e., the intensity of the cross peak) was calculated for multiple silicon atoms near the aluminum atom. For each T-site, at least five silicon atoms were considered: four silicon

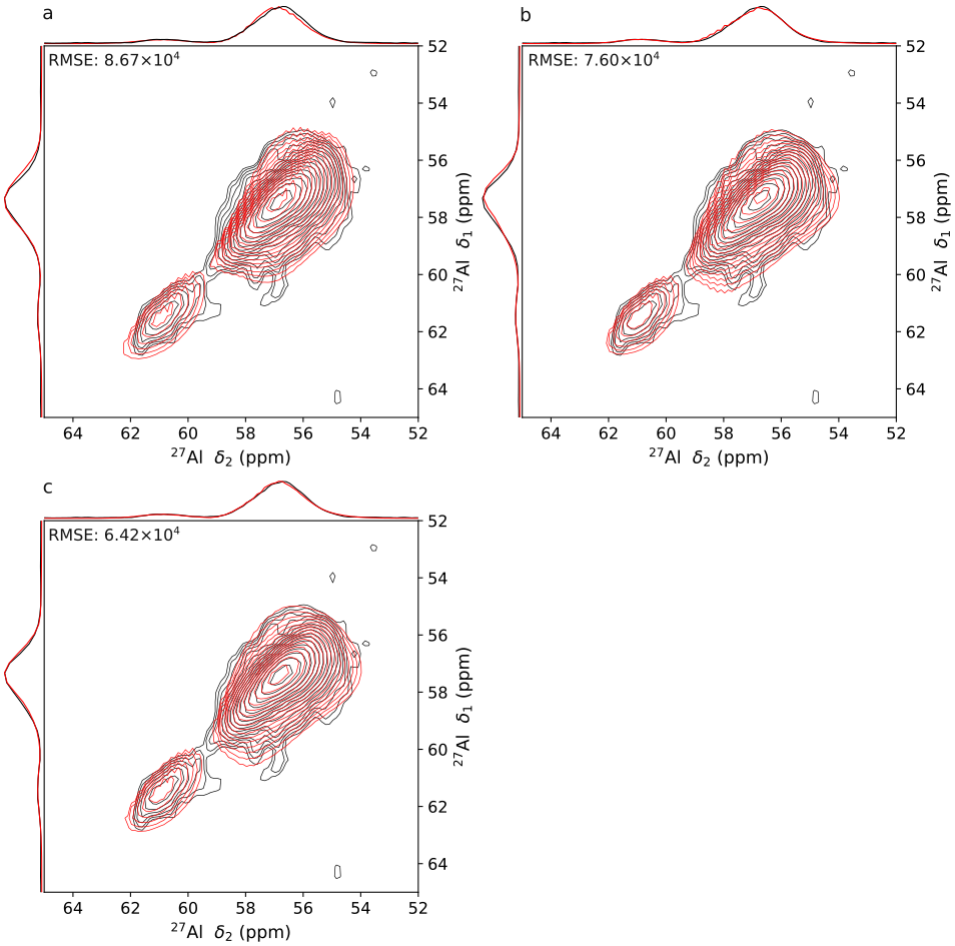

Figure S7: ^{27}Al MQMAS of RTH_{28} sample at 750 MHz and fitted a) 1+1, b) 2+1 and c) 3+1 Czjzek distribution curves. RMSEs were calculated from pointwise residuals within the fit window.

atoms were directly adjacent to the aluminum atom, and one silicon atom across the 4 ring in RTH (1 silicon atom in between), approximately 4.5 Å away from the aluminum atom. For T3, a silicon atom across a 6 ring, less than 5 Å from the aluminum, was also considered.

Numerical simulations of ^{29}Si-^{27}Al magnetization transfer efficiency of D-HMQC experimental using $SR4_1^2$ recoupling element were performed using Simpson 4.2.1.[20, 21] All simulations were conducted at a spinning rate of 10 kHz and a magnetic field of 750 MHz, and pulse length matching those used experimentally, and employing 2145 ZCW orientations for powder averaging.[22–24] Input files comprised isolated dipolar-coupled ^{29}Si-^{27}Al pairs and were constructed with Soprano and Simpyson,[25, 26] considering the spin system ^{27}Al's C_Q and η_Q parameters obtained from CASTEP calculations. Quadrupolar effects were included up to second order. To avoid carrier-frequency sensitivity losses

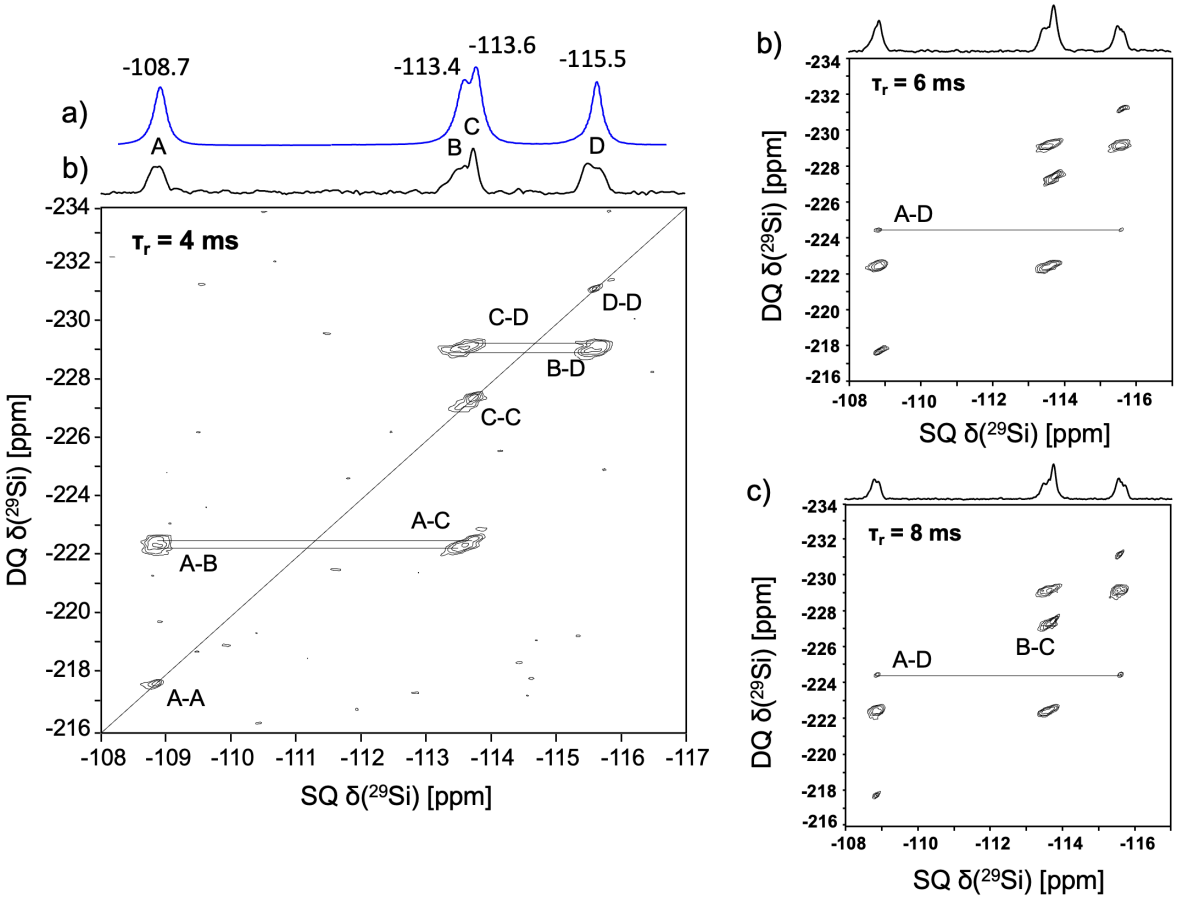

Figure S8: 2D ^{29}Si-^{29}Si DQ-SQ MAS NMR spectra at a) 4 ms, b) 6 ms, and b) 8 ms of recoupling time for the sample RTH-SiO_2.

characteristic of symmetry-based schemes, we omitted chemical shift information in the simulations.

The simulated recoupling efficiencies for all T sites as a function of mixing time are shown in Figure S10. As expected for a dipolar-based magnetization transfer, the effective recoupling scales with $\frac{1}{r^3}$, and thus NN ($r \approx 2.5-3.5$ Å) builds up rapidly at short mixing times, whereas NNN transfer is only appreciable at longer mixing times for ^{29}Si-^{27}Al pairs across 4 ring, ≈ 4.5 Å, while those across 6 rings remain too weak to be detected under our experimental conditions. We verified the robustness of these trends by a grid search over experimentally relevant parameter ranges for hydrated RTH, by considering Al-Si distances between 2.0 and 6.0 Å, ^{27}Al's C_Q from 1.0 to 4.0 MHz and η_Q from 0 to 1. We confirmed that across all this space $SR4_1^2$-based D-HMQC provided consistent higher transfer efficiencies for NN over NNN, with only the short 4-ring NNN contributing at longer mixing times, Figure S11 and S12.

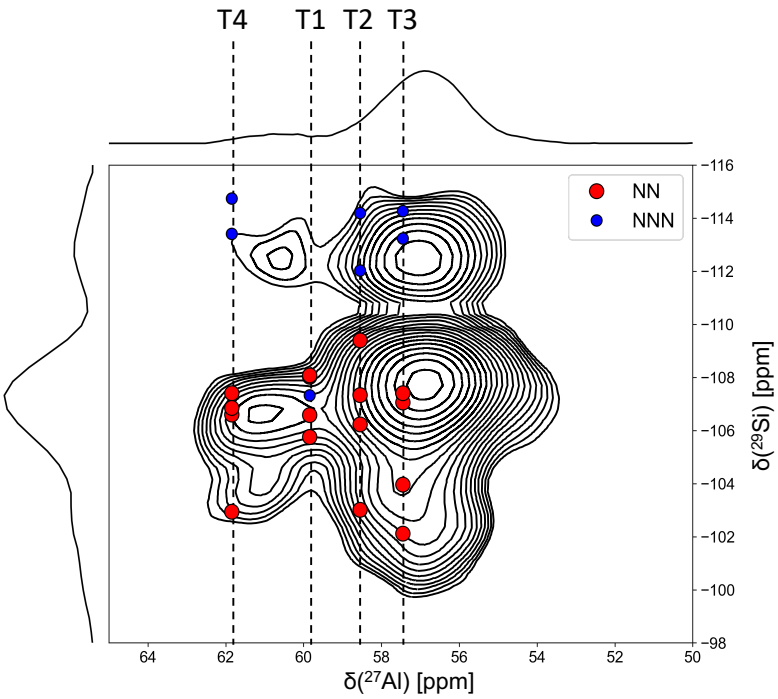

Figure S9: Overlay of unmodified predicted chemical shifts and experimental 2D ^{29}Si-^{27}Al (D)-HMQC MAS-NMR spectrum of the RTH$_{28}$ zeolite at 750 MHz.

To achieve a semi-quantitative interpretation of the 2D ^{29}Si-^{27}Al D-HMQC spectra, correlation-peak intensities must be corrected for distance-dependent recoupling efficiency. For simulating this effect, we extracted, from long ML-MD trajectories, the minimum, average, and maximum Al–Si distances for NN and NNN of each T site, and constructed ^{29}Si-^{27}Al pairs spanning 20 distances across each range and considering a C_Q of 2.0 MHz resembling those determined experimentally. This simulation suggests that at 1.8 ms the 4-ring NNN signals are around 35% of the NN intensities.

S7 Water loading and Si/Al ratio

S7.1 Water loading

Aluminosilicate zeolites are hygroscopic, with water molecules typically present in their pores. The water content in individual zeolites can vary, making it necessary to investigate different levels of water loading to ensure that its effect on both ^{27}Al and ^{29}Si chemical shifts is negligible.

To demonstrate the robustness of our chemical shift predictions and the effect of two

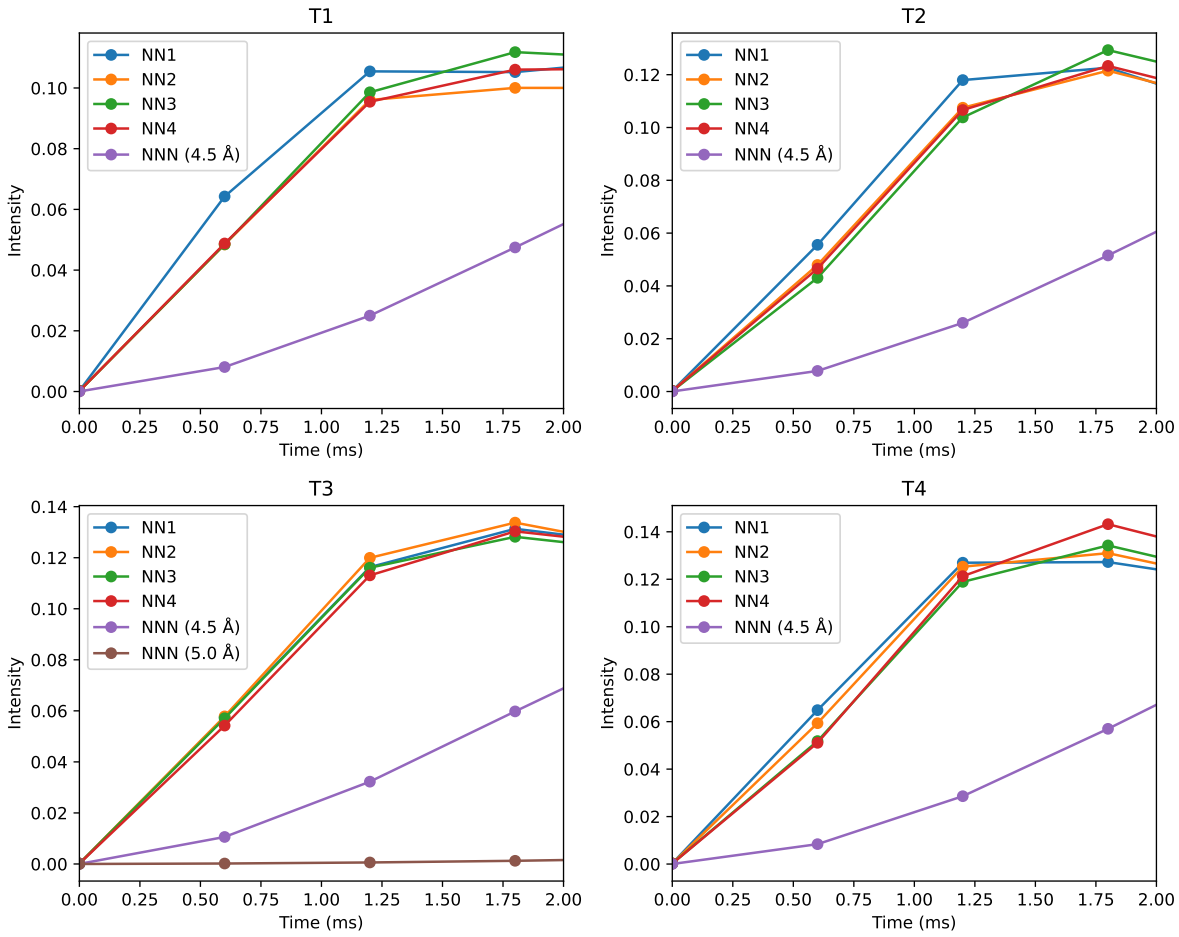

Figure S10: The calculated transfer of magnetization for all T-sites (RTH_{31} with 5 water molecules per unit cell) for multiple silicon atoms with varying mixing times is shown. Blue, orange, green, and red represent silicon atoms adjacent to aluminum. Purple represents silicon atoms located across a 4 ring within 4.5 Å, while brown represents silicon atoms across a 6 ring within 5 Å.

experimentally obtainable parameters on the chemical shifts, multiple modified models were tested. First, different water loadings were examined around the experimentally observed 4 wt%, which corresponds to 4.5 water molecules per unit cell in the RTH zeolite. A range of water loadings (3, 4, 5, 6, and 7 water molecules per unit cell) was tested, corresponding to 2.7, 3.6, 4.4, 5.3, and 6.2 wt%. The ^{27}Al and ^{29}Si chemical shifts based on the water loading are presented in Table S4 and Table S5, respectively.

For ^{27}Al chemical shifts, water loadings greater than 3 water molecules per unit cell result in nearly constant values, varying by only 0.3 ppm. A slight decrease is observed with higher water content, a trend also noted for MFI in our previous work, which is attributed to proton and water dynamics. [4] At higher water loadings, the proton becomes

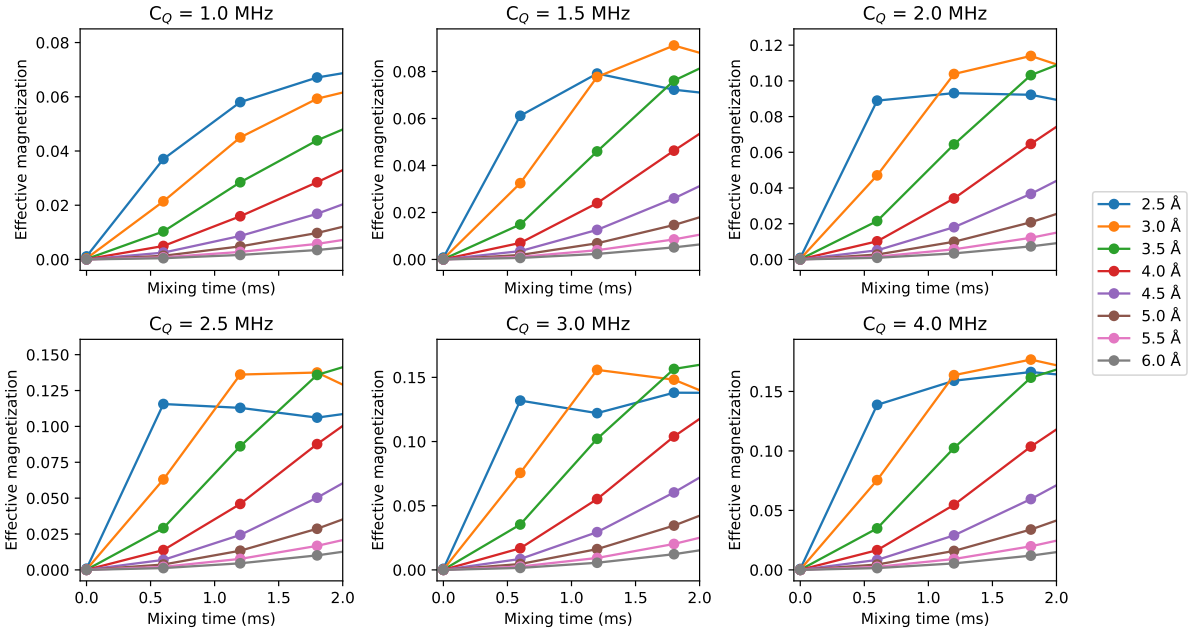

Figure S11: Effect of ^{27}Al C_Q on magnetization transfer efficiency at distinct mixing times and Si-Al distances ranging from 2.5 to 6.0 Å.

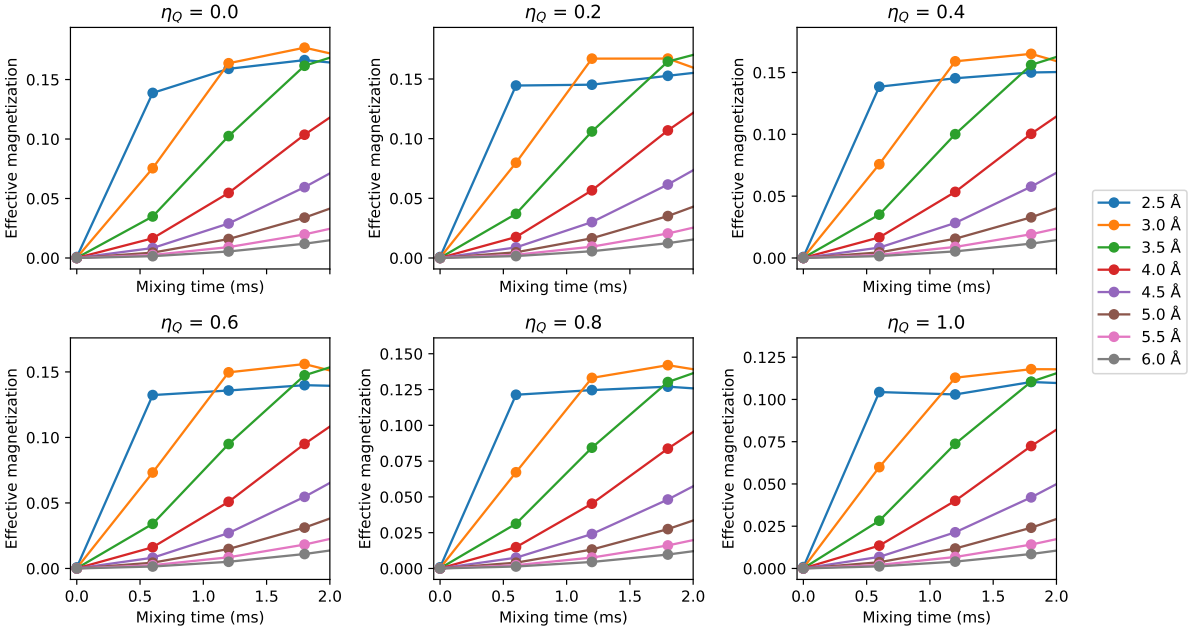

Figure S12: Effect of ^{27}Al η_Q on magnetization transfer efficiency at distinct mixing times and Si-Al distances ranging from 2.5 to 6.0 Å.

fully solvated, leading to a more symmetrical aluminum environment and sharper experimental peaks. However, the water molecules continue to interact with the framework, highlighting the importance of explicitly modeling the water molecules within the zeolite channels. The ^{29}Si chemical shift appears to be unaffected by water content.

Table S4: ^{27}Al chemical shifts in ppm at RTH_{31} with different water loading.

Water (wt %)	Al-T1	Al-T2	Al-T3	Al-T4
3 (2.7)	60.2	58.4	57.3	62.3
4 (3.6)	60.0	58.6	57.6	61.9
5 (4.4)	59.9	58.6	57.4	61.8
6 (5.3)	59.8	58.3	57.5	62.0
7 (6.2)	59.8	58.4	57.4	61.8

Table S5: ^{29}Si chemical shifts averaged over T-sites in ppm at RTH_{31} with different water loading.

Water (wt %)	Si-T1	Si-T2	Si-T3	Si-T4
3 (2.7)	-114.7	-112.8	-112.5	-107.0
4 (3.6)	-114.7	-112.8	-112.5	-107.0
5 (4.4)	-114.7	-112.8	-112.5	-107.0
6 (5.3)	-114.7	-112.8	-112.5	-107.0
7 (6.2)	-114.8	-112.7	-112.4	-107.0

S7.2 Si/Al ratio

Interpreting NMR spectra should primarily be performed on samples with a high Si/Al atomic ratio, as aluminum pairing can significantly affect the ^{27}Al chemical shift in a non-systematic manner,[4, 27] due to the flexible nature of the zeolite framework near aluminum atoms. Two Si/Al ratios were tested: 63 (RTH_{63}), which was modeled by creating a supercell along the z direction, and 15 (RTH_{15}), modeled by adding a second aluminum atom to the unit cell. In total, four aluminum pairing configurations with varying Al-Al distances were tested per T-site.

The chemical shift of the RTH with a Si/Al ratio of 63 is shown in Table S6. The ^{27}Al chemical shift does not significantly change compared to the models with a Si/Al ratio of 31.

For RTH_{15}, the change from RTH_{31} was more significant, with an average deviation of 0.5 ppm (Table S6). However, some trends remained consistent even at lower Si/Al ratios, such as T4 exhibiting the highest chemical shift. This consistency is observable in the 1D ^{27}Al MAS NMR spectra (Figure 4a). However, changes in chemical shift of up to 4 ppm have been observed in zeolite MFI [4], indicating that these aluminum pairs should be modeled explicitly rather than assuming that the trends remain constant.

The same analysis is not applicable to ^{29}Si chemical shifts, as these shifts vary significantly depending on the distance to aluminum atoms. However, as shown in Figure 4b, the ^{29}Si peaks are broader due to the higher aluminum content.

Table S6: ^{27}Al chemical shifts at different Si/Al ratios (31, 15, and 63). For models with Si/Al ratio of 15, four different pair configurations were tested with different distance between aluminum atoms.

T-site	δ(**^{27}Al-31**) **[ppm]**	δ(**^{27}Al-15**) **[ppm]**	d(**Al-15**) **[Å]**	δ(**^{27}Al-63**) **[ppm]**
T1	60.1	60.1	7.1	59.5
		59.7	7.3	
		59.8	9.7	
		60.0	9.1	
T2	58.6	59.9	5.0	58.7
		59.1	4.9	
		59.8	7.3	
		59.3	4.2	
T3	57.4	57.4	9.7	57.4
		57.3	9.7	
		57.2	9.5	
		57.2	9.7	
T4	61.8	61.4	6.8	61.3
		63.8	8.4	
		62.1	4.9	
		61.5	6.6	

S8 Aluminum distribution

To estimate the volume of different peaks in the (D)-HMQC spectra (Figure 5), integration of the 1D projection was performed, as shown in Figure S13. This integration provides an approximation of peak intensities, visualized in Figure S14. A more rigorous approach would involve deconvoluting the peaks into a predefined number of components. However, given that the (D)-HMQC experiment is not explicitly quantitative and deconvoluting the silicon projection would require a high number of peaks, a simpler integration method was chosen. Despite its limitations, this approach should provide a reasonable estimate of the tentative aluminum distribution. Future work will focus on a more detailed analysis, incorporating full shielding and EFG tensor predictions to improve the accuracy and quantitative reliability of these results.

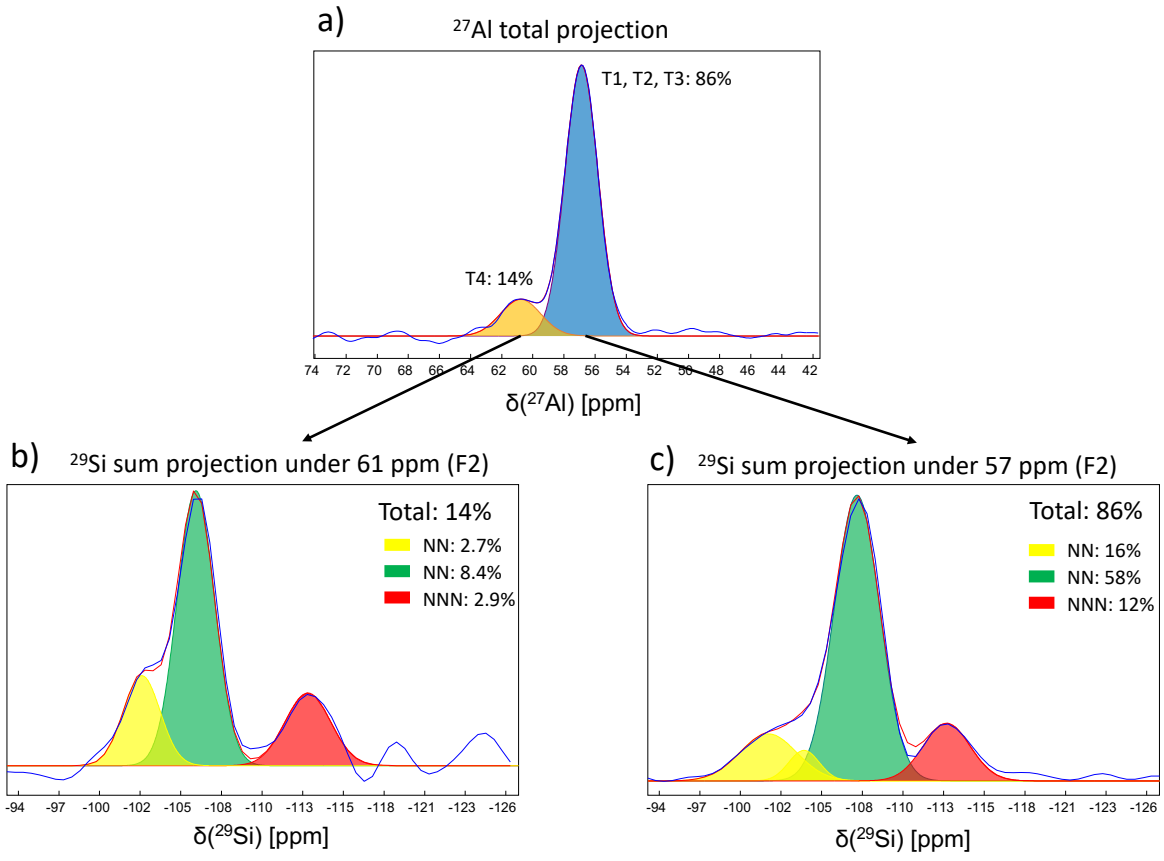

Figure S13: Integration of the projections of: a) ^{27}Al (total projection), b) ^{29}Si projections at 61 ppm in F2 of the ^{29}Si-^{27}Al (D)-HMQC NMR spectrum and c) ^{29}Si projections at 57 ppm in F2 of the ^{29}Si-^{27}Al (D)-HMQC NMR spectrum

Figure S14 shows the peak volumes and their composition in the (D)-HMQC spectra. The theoretical dipolar coupling intensities from Figure S10 determine the relative

intensities of the observed peaks. For example, the cross-peak corresponding to the next-nearest-neighbor silicon to aluminum in the T1 position at a recoupling time of 1.8 ms should have around 40 % of the intensity of correlation peaks arising from nearest neighbor should be approximately. Using this information in combination with the peak compositions from Figure S14, a system of four equations was constructed to determine the distribution across the four T-sites.

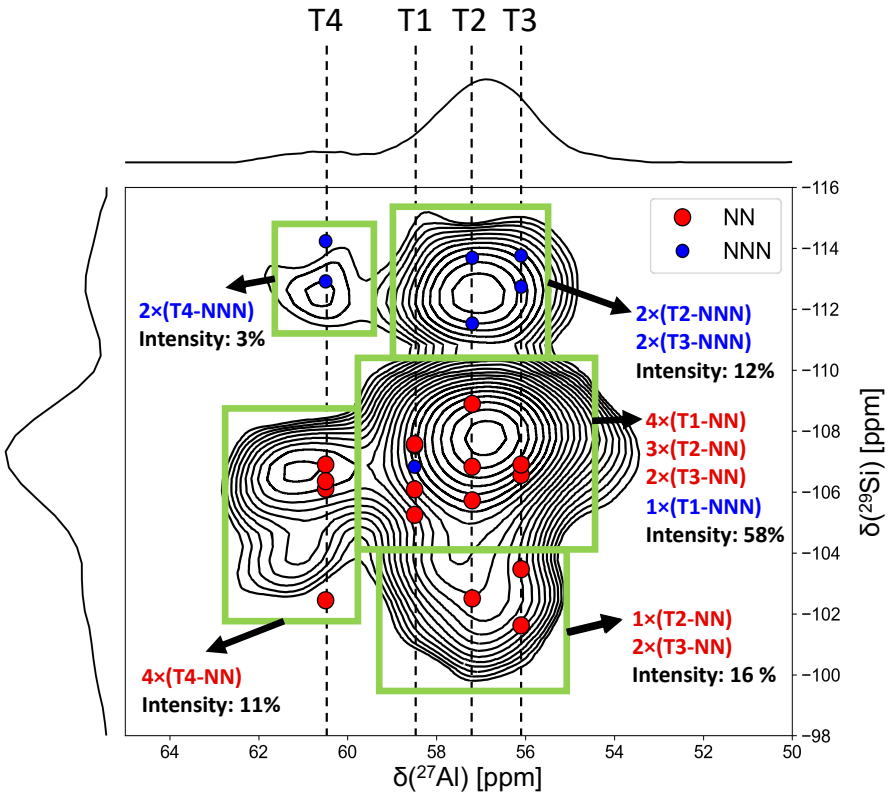

Figure S14: (D)-HMQC spectra divided into 5 peaks with their relative intensities and their composition predicted by the KRR.

$$
\begin{aligned}
0\,t_1+ \quad & 0\,t_2+ \quad & 0\,t_3+ \quad & (4T_4^{\mathrm{NN}}+2T_4^{\mathrm{NNN}})\,t_4 = 14\\
0\,t_1+ \quad & (2T_2^{\mathrm{NNN}})\,t_2+ \quad & (2T_3^{\mathrm{NNN}})\,t_3+ \quad & 0\,t_4 = 12\\
(4T_1^{\mathrm{NN}}+1T_1^{\mathrm{NNN}})\,t_1+ \quad & (3T_2^{\mathrm{NN}})\,t_2+ \quad & (2T_3^{\mathrm{NN}})\,t_3+ \quad & 0\,t_4 = 58\\
0\,t_1+ \quad & (1T_2^{\mathrm{NN}})\,t_2+ \quad & (2T_3^{\mathrm{NN}})\,t_3+ \quad & 0\,t_4 = 16
\end{aligned}
\tag{S3}
$$

The relative occupancies obtained by solving these equations are 15 %, 64 %, 8 %, and 14 % for T1, T2, T3, and T4, respectively. The results clearly indicate that T2 is

the most populated aluminum site, while T3 and T4 are the least populated.

S9 Unit cell effect

NMR is highly sensitive to the local atomic environment, making the choice of unit cell parameters crucial, as they directly influence chemical shifts through changes in T–O–T angles and bond lengths. In this study, we considered three types of unit cells: (a) the reference unit cell from the IZA database [18], (b) a relaxed unit cell obtained via DFT using the PBE functional with a plane-wave energy cutoff of 700 eV, where atomic forces were converged to 10^{-3} eV/Å, and (c) an experimental unit cell derived from XRD measurements. The unit cell parameters are in Table S7.

Table S7: Unit cell parameters for different structural models. IZA refers to the unit cell from the IZA database. Al–T indicates water-loaded structures relaxed with aluminum substituted at a specific T-site. Experimental corresponds to the unit cell of the zeolite sample used in this study.

Unit cell	a [Å]	b [Å]	c [Å]	α [°]	β [°]	γ [°]	Volume [Å^3]
IZA	9.76	20.53	10.00	90.0	96.9	90.0	1988.9
Al-T1	9.74	20.56	9.97	92.2	97.1	91.5	1979.9
Al-T2	9.75	20.72	9.97	90.0	97.2	89.7	1998.7
Al-T3	9.71	20.70	9.99	90.0	97.4	90.6	1991.8
Al-T4	9.73	20.59	9.88	89.3	96.7	90.7	1966.8
Experimental	9.73	20.71	9.88	90.0	96.6	90.0	1979.2

The ^{27}Al NMR chemical shifts predicted using the machine learning approach for different structural models are summarized in Table S8. Across all considered unit cells, the T-sites can be grouped into two categories: T1, T2, and T3 exhibit similar chemical shift values, whereas T4 consistently shows significantly higher shifts. Notably, the experimental and relaxed structures yield the same relative ordering of T-site shifts. In contrast, the IZA-based model presents a reversed ordering between T1 and T2. This discrepancy likely arises from the IZA unit cell being optimized for purely siliceous RTH zeolites, making it less suitable for aluminum containing frameworks. The differences in absolute chemical shift values between the relaxed and experimental unit cells highlight the sensitivity of NMR parameters to subtle variations in atomic structure. Such variations may contribute to systematic offsets between predicted and experimentally measured shifts. As chemical shift is inherently a relative quantity, this offset does not affect the reliability

of the proposed predictive method.

Table S8: Predicted ^{27}Al NMR chemical shifts (in ppm) for different T-sites. Results are compared across three structural models: IZA unit cell, relaxed structures, and the experimental unit cell.

T-site	IZA [ppm]	Relaxed [ppm]	Experimental [ppm]
T1	59.2	58.6	59.9
T2	60.4	57.4	58.6
T3	57.9	56.3	57.4
T4	63.7	63.2	61.8

The overlays of predicted chemical shifts from the three models with the experimental 2D ^{29}Si-^{27}Al (D)-HMQC MAS-NMR spectra are presented in Figure S15. Although the absolute chemical shift values differ, the relative composition of the individual peaks remains qualitatively consistent. The aluminum distribution for different unit cells is summarized in Table S9. While the absolute values vary, T2 is consistently the most populated T-site, with occupancies exceeding 40 % across all unit cells, making this prediction robust regardless of the unit cell choice. Similarly, T1 shows a consistent occupancy around 30 %. In contrast, T3 and T4 are the least populated sites, with occupancies under 25 %, which would be expected of a random aluminum distribution. While these aluminum distributions are approximate, the machine learning approach provides reliable estimates of the aluminum distribution. The accuracy of this method could be further enhanced by incorporating tensorial predictions, which would allow for the prediction of peak shapes and quadrupolar broadening.

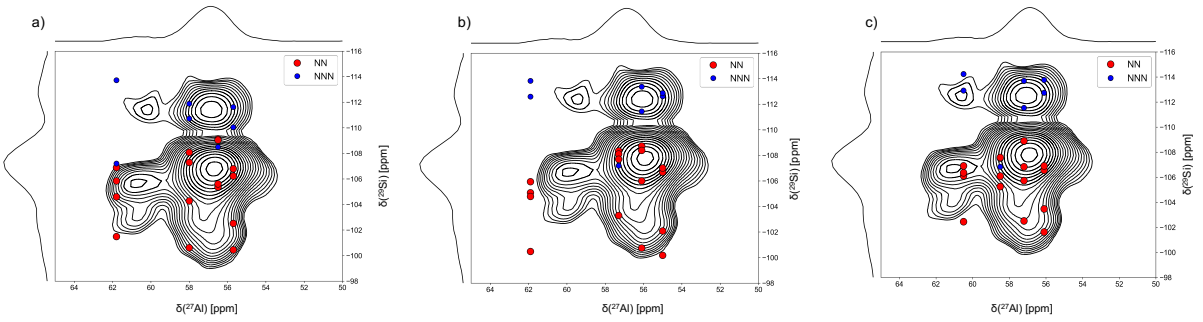

Figure S15: Overlay of predicted chemical shifts and experimental 2D ^{29}Si-^{27}Al (D)-HMQC MAS-NMR spectra for a) IZA, b) relaxed, c) experimental unit cells. The predicted chemical shift values were uniformly offset by a constant to align with the experimental values.

Table S9: Aluminum distribution across different T-sites for various structural models. Values represent the percentage of aluminum at each T-site for the IZA, relaxed, and the experimental unit cells.

T-site	IZA [%]	Relaxed [%]	Experimental [%]
T1	32	15	15
T2	51	65	64
T3	1	0	8
T4	16	20	14

S10 Exchange-correlation functional effect

In order to determine the role of the exchange-correlation functional on the calculated chemical shifts, we generated an optimised model of RTH31 for which a single Al is located in the T1 site, and calculated the chemical shifts for Al atoms in each of the four inequivalent positions, with a selection of GGA and meta-GGA functionals. It is observed that while the absolute positions of the peaks vary significantly among functionals, the spacing between peaks, and therefore the peak patterns are very similar across the set, which correlation coefficients above 0.99. Therefore, after applying a scalar offset that is specific to a given functional, each of the tested functionals give the same predictions.

Table S10: Chemical shifts calculated for the optimised unit cell of RTH31 with Al in the T1 site, as a function of DFT exchange-correlation functional.

ISIF3 (T1)	PBE	PBEsol	rSCAN	r2SCAN
	750eV	750eV	750eV	750eV
T1	57.1	57.7	52.4	52.0
T2	57.4	58.0	52.7	52.3
T3	54.3	55.0	50.0	49.5
T4	61.7	62.2	56.4	55.9

References

(1) Larsen, A. H.; Mortensen, J. J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.; Jensen, P. B.; Kermode, J.; Kitchin, J. R.; Kolsbjerg, E. L.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Maronsson, J. B.; Maxson, T.; Olsen, T.; Pastewka, L.; Peterson, A.; Rostgaard, C.; Schiøtz, J.; Schütt, O.; Strange, M.; Thygesen, K. S.; Vegge, T.; Vilhelmsen, L.; Walter, M.; Zeng, Z.; Jacobsen, K. W. The atomic simulation environment - a Python library for working with atoms. *Journal of Physics: Condensed Matter* **2017**, *29*, 273002.

(2) Erlebach, A.; Šípka, M.; Saha, I.; Nachtigall, P.; Heard, C. J.; Grajciar, L. A reactive neural network framework for water-loaded acidic zeolites. *Nature Communications* **2024**, *15*, 4215.

(3) Schütt, K. T.; Kessel, P.; Gastegger, M.; Nicoli, K. A.; Tkatchenko, A.; Müller, K.-R. SchNetPack: A Deep Learning Toolbox For Atomistic Systems. *Journal of Chemical Theory and Computation* **2018**, *15*, 448–455.

(4) Willimetz, D.; Erlebach, A.; Heard, C. J.; Grajciar, L. ^{27}Al NMR chemical shifts in zeolite MFI *via* machine learning acceleration of structure sampling and shift prediction. *Digital Discovery* **2025**, *4*, 275–288.

(5) Lei, C.; Erlebach, A.; Brivio, F.; Grajciar, L.; Tošner, Z.; Heard, C. J.; Nachtigall, P. The need for operando modelling of 27Al NMR in zeolites: the effect of temperature, topology and water. *Chemical Science* **2023**, *14*, 9101–9113.

(6) Lippmaa, E.; Samoson, A.; Magi, M. High-resolution aluminum-27 NMR of aluminosilicates. *Journal of the American Chemical Society* **1986**, *108*, 1730–1735.

(7) Saha, I.; Grajciar, L.; Willimetz, D. CRISP: Enhancing ASE Workflows with Advanced Molecular Simulation Post-Processing. *ChemRxiv* **2025**, DOI: `10.26434/chemrxiv-2025-dht8g`.

(8) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

(9) Himanen, L.; Jäger, M. O. J.; Morooka, E. V.; Federici Canova, F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. DScribe: Library of descriptors for machine learning in materials science. *Computer Physics Communications* **2020**, *247*, 106949.

(10) Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. I. J.; Refson, K.; Payne, M. C. First principles methods using CASTEP. *Zeitschrift für Kristallographie - Crystalline Materials* **2005**, *220*, 567–570.

(11) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L. A.; Zhou, X.; Burke, K. Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces. *Physical Review Letters* **2008**, *100*, 136406.

(12) Charpentier, T. The PAW/GIPAW approach for computing NMR parameters: A new dimension added to NMR study of solids. *Solid State Nuclear Magnetic Resonance* **2011**, *40*, 1–20.

(13) Schurko, R. W.; Wasylishen, R. E.; Foerster, H. Characterization of Anisotropic Aluminum Magnetic Shielding Tensors. Distorted Octahedral Complexes and Linear Molecules. *The Journal of Physical Chemistry A* **1998**, *102*, 9750–9760.

(14) Lewis, J. E.; Freyhardt, C. C.; Davis, M. E. Location of pyridine guest molecules in an electroneutral {3}[SiO4/2] host framework: Single-crystal structures of the as-synthesized and calcined forms of high-silica ferrierite. *The Journal of Physical Chemistry* **1996**, *100*, 5039–5049.

(15) Williams, J. J.; Lethbridge, Z. A.; Clarkson, G. J.; Ashbrook, S. E.; Evans, K. E.; Walton, R. I. The bulk material dissolution method with small amines for the synthesis of large crystals of the siliceous zeolites ZSM-22 and ZSM-48. *Microporous and mesoporous materials* **2009**, *119*, 259–266.

(16) Díaz-Cabañas, M.-J.; Barrett, P. A. Synthesis and structure of pure SiO 2 chabazite: the SiO 2 polymorph with the lowest framework density. *Chemical Communications* **1998**, 1881–1882.

(17) Schmahl, W. W.; Swainson, I.; Dove, M.; Graeme-Barber, A. Landau free energy and order parameter behaviour of the α/β phase transition in cristobalite. *Zeitschrift für Kristallographie* **1992**, *201*, 125–145.

(18) Baerlocher, C.; McCusker, L. B. Database of zeolite structures. http://www.iza–structure.org/databases/.

(19) D'Espinose de Lacaillerie, J.-B.; Fretigny, C.; Massiot, D. MAS NMR spectra of quadrupolar nuclei in disordered solids: The Czjzek model. *Journal of Magnetic Resonance* **2008**, *192*, 244–251.

(20) Bak, M.; Rasmussen, J. T.; Nielsen, N. C. SIMPSON: A General Simulation Program for Solid-State NMR Spectroscopy. *Journal of Magnetic Resonance* **2000**, *147*, 296–330.

(21) Juhl, D. W.; Tošner, Z.; Vosegaard, T. Versatile NMR Simulations Using SIMPSON. **2020**, *100*, 1–59.

(22) Zaremba, S. K. Good Lattice Points, Discrepancy, and Numerical Integration. *Annali di Matematica Pura ed Applicata* **1966**, *73*, 293–317.

(23) Conroy, H. Molecular Schrödinger Equation. VIII. A New Method for the Evaluation of Multidimensional Integrals. *The Journal of Chemical Physics* **1967**, *47*, 5307–5318.

(24) Cheng, V. B.; Suzukawa, H. H.; Wolfsberg, M. Investigations of a Nonrandom Numerical Method for Multidimensional Integration. *The Journal of Chemical Physics* **1973**, *59*, 3992–3999.

(25) Sturniolo, S.; Shenton, J. K.; Bornes, C. Soprano, version 0.10.1, CCP for NMR Crystallography, 2025.

(26) Bornes, C.; Soares, M.; Pereira, D. Simpyson, version 0.2, NUTS - NMR Unofficial ToolSet, 2026.

(27) Dědeček, J.; Sklenak, S.; Li, C.; Wichterlová, B.; Gábová, V.; Brus, J.; Sierka, M.; Sauer, J. Effect of Al-Si-Al and Al-Si-Si-Al Pairs in the ZSM-5 Zeolite Framework on the^{27}Al NMR Spectra. A Combined High-Resolution^{27}Al NMR and DFT/MM Study. *The Journal of Physical Chemistry C* **2009**, *113*, 1447–1458.