Supplementary Information (SI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2026

## 1. Simulation System Modeling

We constructed a total of six boxes. Boxes 1 (Figure S1a), 3, and 5 simulated experimental conditions with THF solvent dilution, while boxes 2 (Figure S1b), 4, and 6 simulated conditions without solvent addition. Specific details of the boxes are provided in Table S1.

For simulations with solution systems, the system needs to be pre-equilibrated before production. OPLS/AA force field parameters[1] for all molecules were generated by the LigParGen web server[2]. Classical molecular dynamics at 300 K, 1 bar NpT ensemble conditions were performed using the Gromacs program[3]. Equilibrated system configurations are required for ab initio molecular dynamics (AIMD).

**Figure S1**

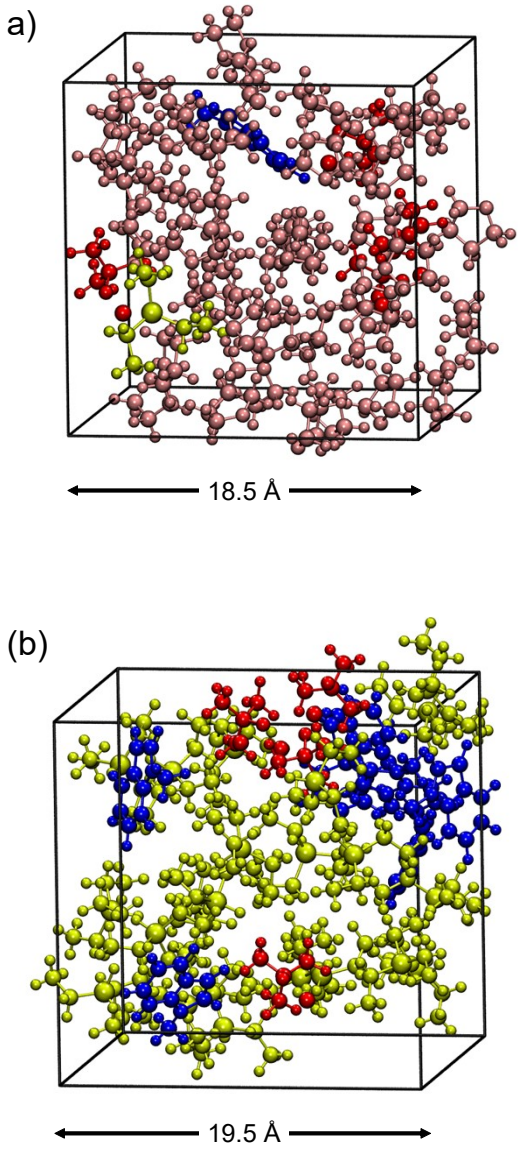


**Figure S1.** (a) and (b) Boxes of THF environment and neat environment after classical molecular dynamics equilibrium. Atoms are colored using molecular species, KO$^t$Bu in red, $Et_3SiH$ in yellow, N-methylindole in blue, and THF in pink.

**Table S1**

| No. | $N_{KO^tBu}$ | $N_{Et_3SiH}$ | $N_{N-methylindole}$ | $N_{THF}$ | Box Size (Å) |
|---|---|---|---|---|---|
| 1 | 4 | 1 | 1 | 40 | 18.5 |
| 2 | 4 | 21 | 7 | 0 | 19.5 |
| 3 | 4 | 1 | 1 | 400 | 25.6 |
| 4 | 4 | 70 | 30 | 0 | 29 |
| 5 | 4 | 1 | 1 | 120 | 37.6 |
| 6 | 4 | 140 | 60 | 0 | 36.4 |

**Table S1.** Details of the simulation box

## 2. Neural Networks Potential

The machine learning molecular dynamics (MLMD) in this work were performed using a machine learning interatomic potential constructed within the Deep Potential Molecular Dynamics (DeePMD)[4] framework, as implemented in the DeePMD-kit package. DeePMD employs deep neural networks (DNNs) to represent high-dimensional potential energy surfaces with near first-principles accuracy while having considerable computational efficiency.

In the DeePMD formalism, the total potential energy of a system containing $N$ atoms are decomposed into a sum of atomic energy contributions,

$$E = \sum_{i=1}^{N} E_i ,$$

where each atomic energy $E_i$ depends only on the local chemical environment of atom $i$ within a finite cutoff radius. This locality assumption enables linear scaling with system size and is essential for large-scale molecular simulations.

To ensure translational, rotational, and permutational invariance of the potential energy surface, the atomic coordinates are transformed into symmetry-preserving descriptors that characterize the local environment of each atom. For atom $i$, descriptors are constructed from relative position vectors $R_{ij} = R_j - R_i$ of neighboring atoms and expressed in a locally defined coordinate frame. Depending on the neighbor distance, either full (radial and angular) information or radial-only information is retained. A representative descriptor component takes the form

$$D_{ij} = \left\{ \frac{1}{R_{ij}}, \frac{x_{ij}}{R_{ij}}, \frac{y_{ij}}{R_{ij}}, \frac{z_{ij}}{R_{ij}} \right\},$$

Where $R_{ij} = |R_{ij}|$, and $(x_{ij}, y_{ij}, z_{ij})$ are the neighbor coordinates in the local reference frame. Neighbors are sorted by chemical species and distance, which naturally enforces permutational symmetry. In this work, the $se\_e2\_a$ descriptor we use contains full (radial and angular) information.

The atomic energy is obtained by mapping the descriptor vector $D_i$ to a scalar value through a feedforward deep neural network specific to the atomic species,

$E_i = N_{s(i)}(D_i)$.

The neural network consists of multiple fully connected hidden layers with nonlinear activation functions, followed by a linear output layer. Separate networks are used for different chemical species, allowing flexible and chemically transferable representations.

Atomic forces are computed analytically as the negative gradient of the total energy with respect to atomic positions, ensuring strict consistency between energies and forces. The virial tensor is obtained analogously from analytical derivatives, enabling constant-pressure simulations and stress analysis.

The neural network parameters are optimized by minimizing a composite loss function that incorporates errors in energies, forces, and virials,

$$L = \frac{p_e}{N}\Delta E^2 + \frac{p_f}{3N}\sum_i |\Delta F_i|^2 + \frac{p_\xi}{9N}\|\Delta \Xi\|^2,$$

where $p_e$, $p_f$, and $p_\xi$ are the energy, force and the virial tensor prefactors, respectively. The optimization is carried out using the Adam stochastic gradient descent algorithm with an exponentially decaying learning rate. The prefactor of the energy term in the loss function was changed from 0.02 to 2, the force term from 1,000 to 1, and the virial term from 0.01 to 1, respectively.

## 3. Model Accuracy

The accuracy of the trained neural network potential was validated by comparing atomic forces predicted by the model against DFT calculated forces. A validation dataset was constructed by randomly extracting configurations from simulated trajectories. The configurations in the validation set were subjected to DFT calculations using the same parameters as the computed data. The dataset is divided into three sets based on system behavior.: All validation sets, including biased and non-reactive unbiased simulation trajectories (Figure S2a); Hydride generation (Figure S2b); C2 and C3 deprotonation of N-methylindole (Figure S2c); Silylation of C2 and C3 in N-methylindole (Figure S3d). The validation dataset covers all intermediates and transition states involved in the reactions studied. In total, approximately 3,000 configurations were included. The mean absolute error (MAE) and root mean square error (RMSE) of atomic forces were 0.0383 eV/Å and 0.0570 eV/Å, respectively. Results are shown in Figure S2.

**Figure S2**

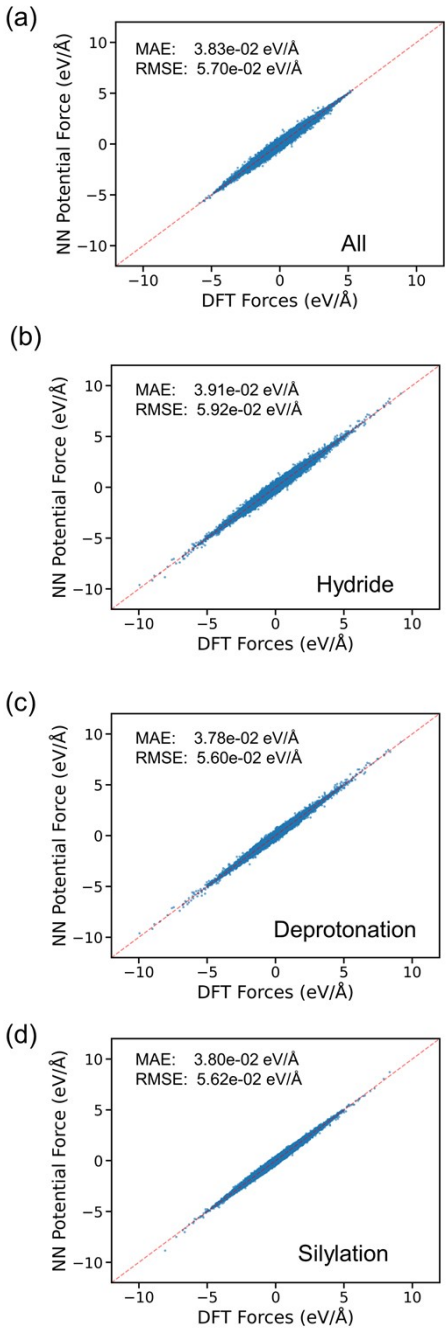


**Figure S2.** Force prediction errors of the neural network potential on the validation dataset. (a) All validation sets, including b, c and non-reactive unbiased simulation trajectories (~3000 configurations). (b) Hydride generation from the trajectory of metadynamics in Figure 1d, S2 (~1,000 configurations). (c) C2 and C3 deprotonation of N-methylindole from metadynamics trajectories of Figure 2b-c and S2 (~1,000 configurations). (d) C2 and C3 silylation of N-methylindole from metadynamics trajectories of Figure S10 (~1,000 configurations).

## 4. Collective Variable（CV）

The formula for calculating the coordination number and PATH[5] used is shown in Table S1. In the simulation of deprotonation reactions, we use Path Collective Variables[6] to define the progress of the current state in the deprotonation reaction. For example, in the C3 deprotonation reaction initial state (IS), $^{C_{HH_2}} \rightarrow 0$, $^{C_{C_3H_3}} \rightarrow 1$ in $s.Path$; final state (FS), $^{C_{HH_2}} \rightarrow 1$ , $^{C_{C_3H_3}} \rightarrow 0$ in $s.Path$.

**Table S2.**

| CV [a] | Definition | Parameters |
|---|---|---|
| $C_{OO}$ | $C_{OO} = \sum_{i \in A} \sum_{j \in B} \left( \frac{r_{ij} - r_0}{r_{max} - r_0} - 1 \right)^2 \left(1 + \frac{2(r_{ij} - r_0)}{r_{max} - r_0}\right)$ | $r_{ij}$: distance between atoms O and O; $r_0 = 4.0$ Å, $r_{max} = 4.2$ |
| $C_{SiO}$ | $C_{SiO} = \sum_{i \in A} \sum_{j \in B} \left( \frac{r_{ij} - r_0}{r_{max} - r_0} - 1 \right)^2 \left(1 + \frac{2(r_{ij} - r_0)}{r_{max} - r_0}\right)$ | $r_{ij}$: distance between atoms Si and O; $r_0 = 1.8$ Å, $r_{max} = 2{,}0$ |
| $C_{HH_2}$ | $C_{HH_2} = \sum_{i \in A} \sum_{j \in B} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^n}{1 - \left(\frac{r_{ij}}{r_0}\right)^m}$ | $r_{ij}$: distance between atoms H and $H_2$; $r_0 = 1.2$ Å, $m = 6$, $n = 8$ |
| $C_{HH_3}$ | $C_{HH_3} = \sum_{i \in A} \sum_{j \in B} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^n}{1 - \left(\frac{r_{ij}}{r_0}\right)^m}$ | $r_{ij}$: distance between atoms H and $H_3$; $r_0 = 1.2$ Å, $m = 6$, $n = 8$ |
| $C_{KH}$ | $C_{KH} = \sum_{i \in A} \sum_{j \in B} \left( \frac{r_{ij} - r_0}{r_{max} - r_0} - 1 \right)^2 \left(1 + \frac{2(r_{ij} - r_0)}{r_{max} - r_0}\right)$ | $r_{ij}$: distance between atoms K and H; $r_0 = 2.8$ Å, $r_{max} = 3.2$ |
| $C_{C_2H_2}$ | $C_{C_2H_2} = \sum_{i \in A} \sum_{j \in B} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^n}{1 - \left(\frac{r_{ij}}{r_0}\right)^m}$ | $r_{ij}$: distance between atoms $C_2$ and $H_2$; $r_0 = 1.2$ Å, $m = 6$, $n = 8$ |
| $C_{C_3H_3}$ | $C_{C_3H_3} = \sum_{i \in A} \sum_{j \in B} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^n}{1 - \left(\frac{r_{ij}}{r_0}\right)^m}$ | $r_{ij}$: distance between atoms $C_3$ and $H_3$; $r_0 = 1.2$ Å, $m = 6$, $n = 8$ |

| | | |
|---|---|---|
| $s.Path$ | $s.Path = \frac{\sum_{i=1}^{N} i exp(-\lambda R[X - X_i])}{\sum_{i=1}^{N} exp(-\lambda R[X - X_i])}$ | $X$: $C_{HH_2}$(or $C_{HH_3}$) and $C_{C_2H_2}$ (or $C_{C_3H_3}$); $\lambda = 4$ |
| $z.Path$ | $z.Path = -\frac{1}{\lambda}\ln[\sum_{i=1}^{N} i exp(-\lambda R[X - X_i])]$ | $X$: $C_{HH_2}$(or $C_{HH_3}$) and $C_{C_2H_2}$ (or $C_{C_3H_3}$); $\lambda = 4$ |

[a] Specific objects referred to here by individual atoms: $H_2$,$C_2$,$H_3$,$C_3$ represent the H and C atoms at the C2 and C3 positions of the N-methylindole; H represents hydride.

**Table S2.** CVs used in the simulation and their definitions.

## 5. Convergence of the Metadynamics Simulations

**Figure S3**

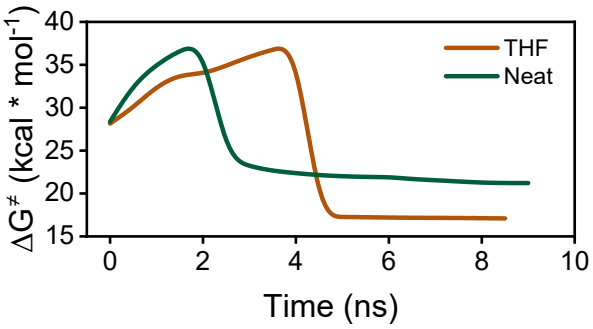


**Figure S3.** Convergence trend of free energy barriers over time in the main text Figure 2d.

**Figure S4**

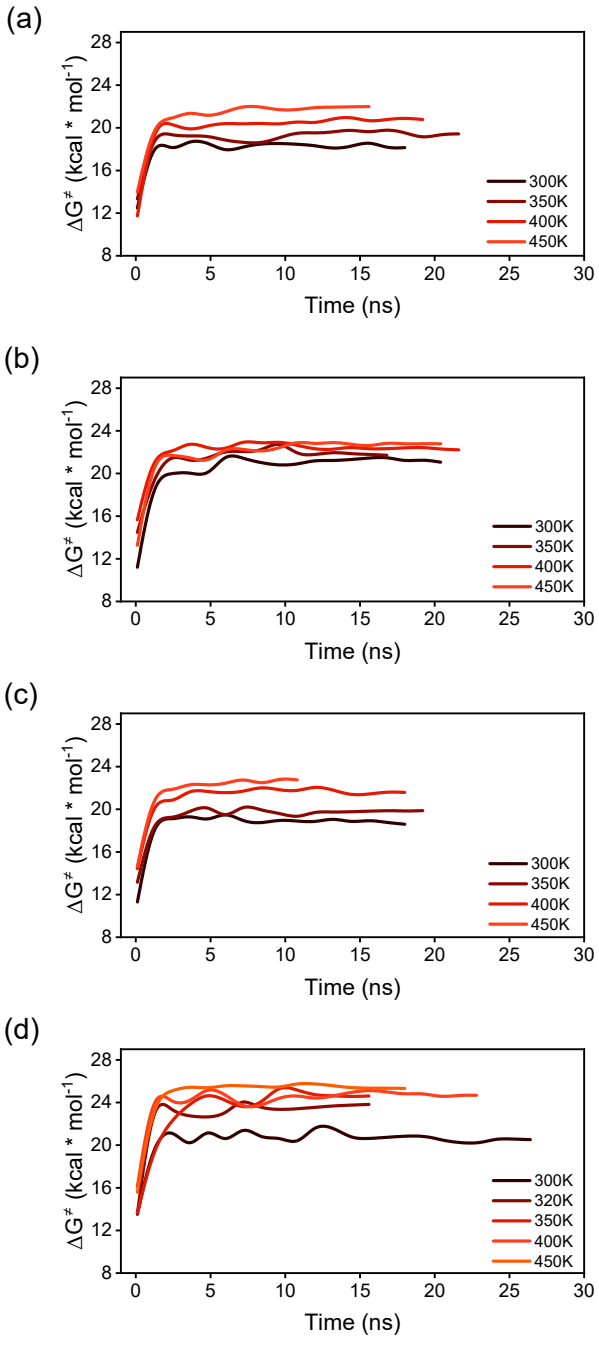


**Figure S4.** (a-d) Convergence trend of free energy barriers over time in the main text Figure 3c-d.

## 6. Simulation Results

**Figure S5**

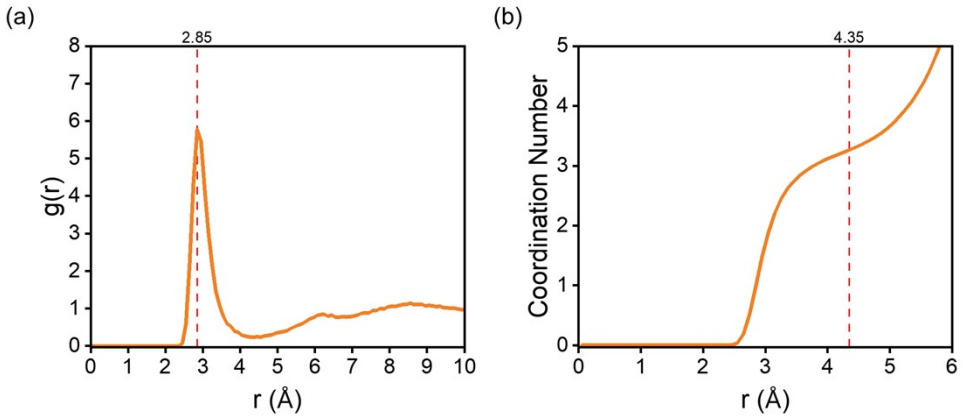


**Figure S5.** (a) and (b) At 300K in a THF environment, the RDF of potassium ions from the KO$^t$Bu catalyst with O atoms in THF and its integral in the first shell region (average coordination number).

**Figure S6**

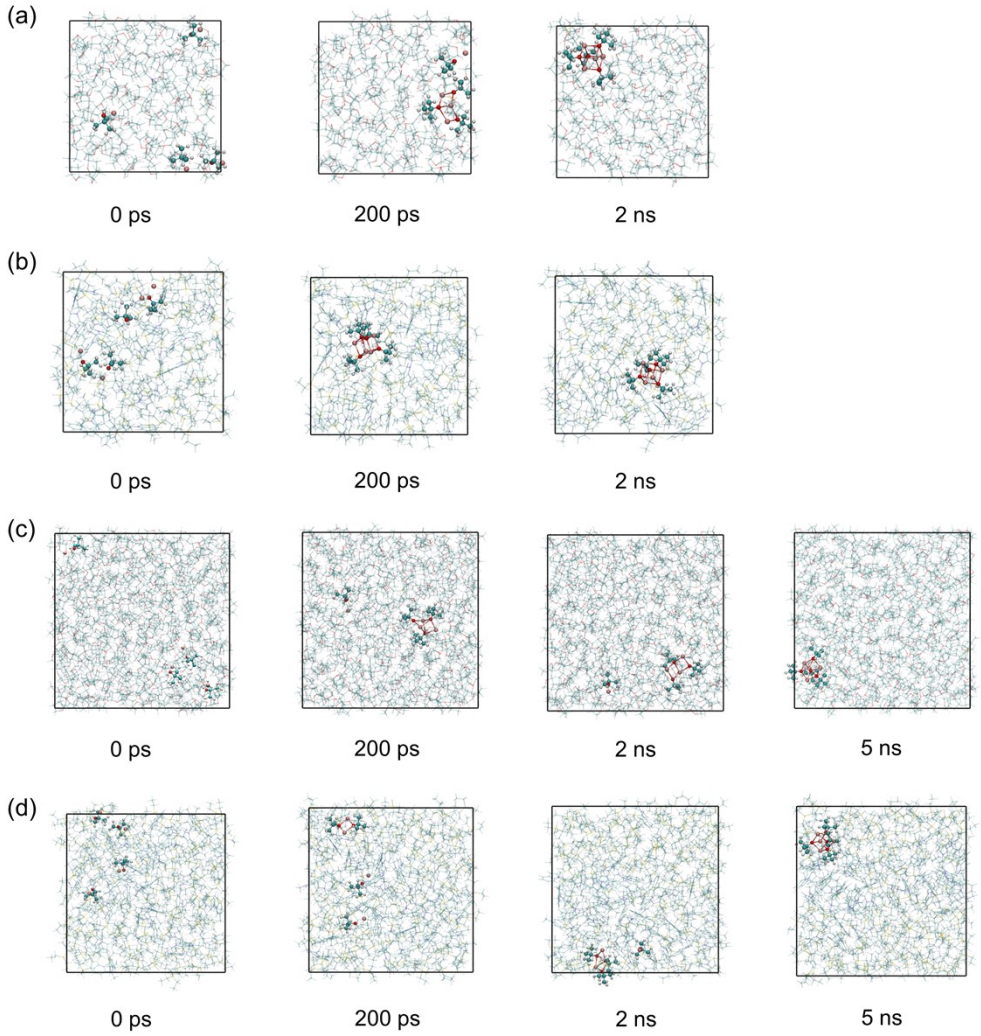


**Figure S6.** Expanded unbiased simulation snapshots at 0 ps, 200 ps, 2 ns, and 5 ns, highlighting KO$^t$Bu molecules. a, b, c, d correspond to boxes 3, 4, 5, and 6 in Table S1.

**Figure S7**

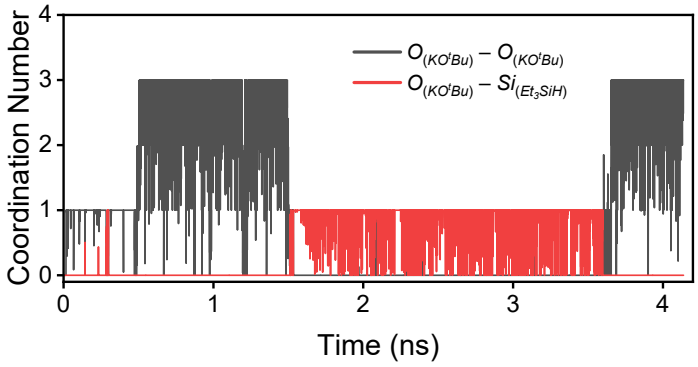


**Figure S7.** The mutual coordination number of O atoms in KO$^t$Bu molecules, and the coordination number of oxygen atoms in KO$^t$Bu with Si atoms in $Et_3SiH$ molecules over time.

**Figure S8**

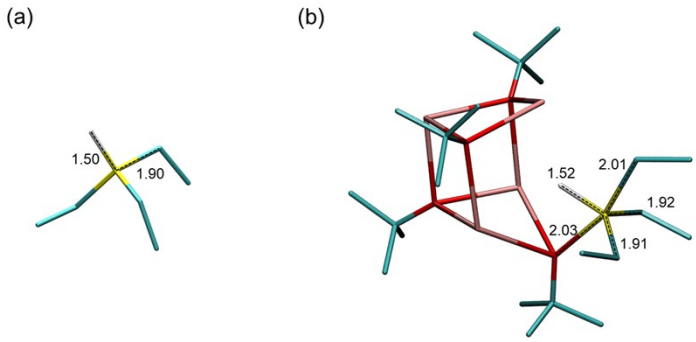


**Figure S8.** Structures of $Et_3SiH$ and pentacoordinate silicon captured in simulations, with H atoms on C atoms hidden. Geometry optimization performed using the Gaussian 16 program[7] at the B3LYP/6-31G(d) level of theory.

The simulations show that the polymerization of KO$^t$Bu always takes precedence over the hydride generation process. That is, in all simulations, the hydride is generated under the catalytic influence of KO$^t$Bu clusters. To highlight the difference between a single KO$^t$Bu catalyst and KO$^t$Bu clusters, we simulated the hydride production process in a box containing one KO$^t$Bu and a box containing four KO$^t$Bu, respectively, with the simulation and enhancement sampling setups consistent with Figures 2a-b. The results show that free the hydride has almost no lifetime in a single catalyst

environment and that individual potassium ions have no temperature hydride capability (Figure 4a). In contrast, in catalyst clusters, the hydride relies on the electrostatic interaction of multiple potassium ions and can be stabilized (Figure 4b).

**Figure S9**

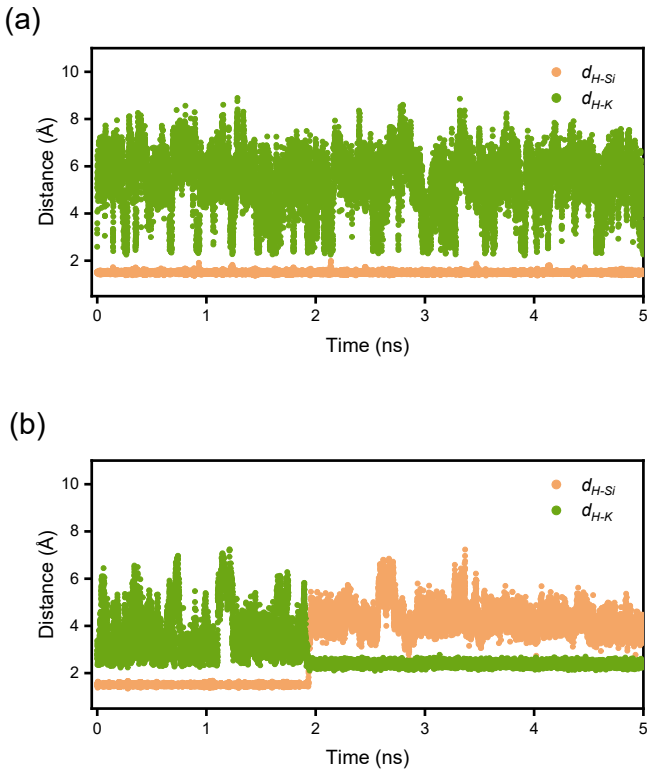


**Figure S9.** Distance of the Si-H bond in $Et_3SiH$, and the distance of silicon hydrogen from the nearest $K^+$ ion with time. (a) single catalyst system; (b) four catalyst systems

**Figure S10**

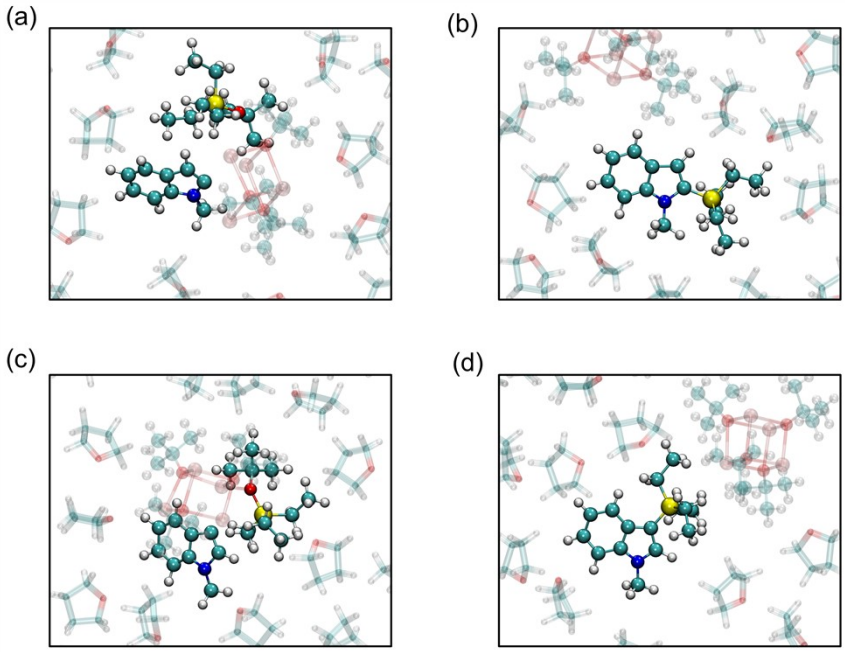


**Figure S10.** Snapshots of N-methylindole silylation extracted from trajectories. a, b: IS (initial state) and FS (final state) snapshots of 2-indole anion silylation. c, d: IS and FS snapshots of 3-indole anion silylation.

**Figure S11**

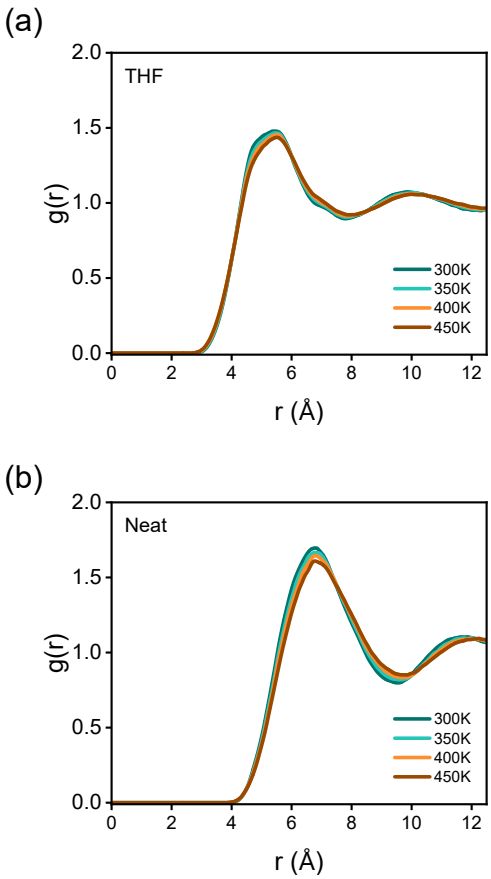


**Figure S11.** Radial distribution function (RDF) of the solvent counted in the trajectory of the deprotonation reaction simulation. (a) RDF of THF molecules in THF environment; (b) RDF of $Et_3SiH$ molecules in neat environment.

168 REFERENCES


169 1. W. L. Jorgensen and J. Tirado-Rives, Potential energy functions for atomic-level
170 simulations of water and organic and biomolecular systems, *Proceedings of the*
171 *National Academy of Sciences*, 2005, **102**, 6665-6670.
172 2. L. S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives and W. L. Jorgensen, LigParGen
173 web server: an automatic OPLS-AA parameter generator for organic ligands, *Nucleic*
174 *Acids Research*, 2017, **45**, W331-W336.
175 3. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl,
176 GROMACS: High performance molecular simulations through multi-level
177 parallelism from laptops to supercomputers, *SoftwareX*, 2015, **1-2**, 19-25.
178 4. J. Zeng, D. Zhang, D. Lu, P. Mo, Z. Li, Y. Chen, M. Rynik, L. a. Huang, Z. Li, S. Shi,
179 Y. Wang, H. Ye, P. Tuo, J. Yang, Y. Ding, Y. Li, D. Tisi, Q. Zeng, H. Bao, Y. Xia, J.
180 Huang, K. Muraoka, Y. Wang, J. Chang, F. Yuan, S. L. Bore, C. Cai, Y. Lin, B.
181 Wang, J. Xu, J.-X. Zhu, C. Luo, Y. Zhang, R. E. A. Goodall, W. Liang, A. K. Singh,
182 S. Yao, J. Zhang, R. Wentzcovitch, J. Han, J. Liu, W. Jia, D. M. York, W. E, R. Car,
183 L. Zhang and H. Wang, DeePMD-kit v2: A software package for deep potential
184 models, *The Journal of Chemical Physics*, 2023, **159**, 054801.
185 5. D. Branduardi, F. L. Gervasio and M. Parrinello, From A to B in free energy space,
186 *The Journal of Chemical Physics*, 2007, **126**, 054103.
187 6. L. Hovan, F. Comitani and F. L. Gervasio, Defining an Optimal Metric for the Path
188 Collective Variables, *Journal of Chemical Theory and Computation*, 2019, **15**, 25-32.
189 7. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R.
190 Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M.
191 Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P.
192 Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F.
193 Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.
194 G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota,
195 R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T.
196 Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark,
197 J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi,
198 J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi,
199 M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L.
200 Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian 16 Rev.
201 C.01.*Journal*, 2016.


202