Acta Physico-Chimica Sinica

LASPAI: AI-powered Platform for the Future Atomic Simulation

--Manuscript Draft--

Manuscript Number:	ACTPHY-D-25-00613
Full Title:	LASPAI: AI-powered Platform for the Future Atomic Simulation
Article Type:	VSI: AI chemistry - Perspective
Keywords:	atomic simulation; generative model; global machine learning potential; LASPAI platform
Corresponding Author:	Zhi-Pan Liu Fudan University CHINA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Fudan University
Corresponding Author's Secondary Institution:	
First Author:	Han-Zhi Luo
First Author Secondary Information:	
Order of Authors:	Han-Zhi Luo
	Qi-Ming Liang
	Zi-Xing Guo
	Xin-Tian Xie
	Jin-Peng Tang
	Tong Guan
	Si-Cong Ma
	Ye-Fei Li
	Zhen-Xiong Wang
	Cheng Shang
	Zhi-Pan Liu
Order of Authors Secondary Information:	
Abstract:	Atomic simulation is becoming a vital tool in modern science, bridging the gap between theory and experiments. Since its birth in 1950s, the balance between accuracy and speed has been the main theme in simulating atomic world and in recent years machine learning potential based methods emerged as a promising alternative to density functional theory calculations for exploring complex potential energy surface (PES). Here we report our implementation of LASPAI (www.laspai.com), a web-based platform for future atomic simulations, which is built using the generalized global neural network potential for fast PES evaluation as implemented in LASP software, together with a series of general diffusion generative models, stochastic surface walking (SSW) global optimization, and other common simulation tools for the PES exploration of molecules and materials. We show that LASPAI platform offers a task-orientated, user-friendly, web-based graphical user interface (GUI) to greatly simplify and speed-up atomic simulations for a wide range of scientific areas, ranging from molecule and material structure prediction to solid-gas, solid-liquid, solid-solid interface identification, and reaction pathway simulations. It aims to provide a fast chemical knowledge delivery for scientists to design new materials and reactions.

Declaration of interests

☒The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

LASPAI: AI-powered Platform for the Future Atomic Simulation

Han-Zhi Luo[a], Qi-Ming Liang[a], Zi-Xing Guo[a], Xin-Tian Xie[a], Jin-Peng Tang[a], Tong Guan[a], Ye-Fei Li[a], Si-Cong Ma[b], Zhen-Xiong Wang[a*], Cheng Shang[a*] and Zhi-Pan Liu[a,b*]

[a]State Key Laboratory of Porous Materials for Separation and Conversion, Collaborative Innovation Center of Chemistry for Energy Material, Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Key Laboratory of Computational Physical Science, Department of Chemistry, Fudan University, Shanghai 200433, China

[b]State Key Laboratory of Metal Organic Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China

[*]Email: wangzhenxiong@fudan.edu.cn; cshang@fudan.edu.cn; zpliu@fudan.edu.cn

Abstract

Atomic simulation is becoming a vital tool in modern science, bridging the gap between theory and experiments. Since its birth in 1950s, the balance between accuracy and speed has been the main theme in simulating atomic world and in recent years machine learning potential based methods emerged as a promising alternative to density functional theory calculations for exploring complex potential energy surface (PES). Here we report our implementation of LASPAI (www.laspai.com), a web-based platform for future atomic simulations, which is built using the generalized global neural network potential for fast PES evaluation as implemented in LASP software, together with a series of general diffusion generative models, stochastic surface walking (SSW) global optimization, and other common simulation tools for the PES exploration of molecules and materials. We show that LASPAI platform offers a task-orientated, user-friendly, web-based graphical user interface (GUI) to greatly simplify and speed-up atomic simulations for a wide range of scientific areas, ranging from molecule and material structure prediction to solid-gas, solid-liquid, solid-solid interface identification, and reaction pathway simulations. It aims to provide a fast chemical knowledge delivery for scientists to design new materials and reactions.

Keywords: atomic simulation, generative model, global machine learning potential, LASPAI platform

1 Introduction

Atomic simulation is an indispensable tool in modern chemistry, physics, and materials science, offering a microscopic view into the atomic interactions that govern material properties[1–4]. Historically, a significant challenge in this field has been the trade-off between the accuracy and the speed, which is generally manifested by the computational cost[1,5,6]. Recently, machine learning potentials (MLPs) have emerged as a transformative

approach to overcome this accuracy-efficiency dilemma with highly flexible surrogate models trained to directly learn the high-dimensional quantum mechanical potential energy surface (PES) from extensive databases of reference quantum mechanics (QM) calculations[7–12]. Since the first report of Behler-Parrinello Neural Network (BPNN)[9], many different MLP models were developed, to name a few representatives, the BPNN-derived models such as many-body function corrected neural network (MBNN)[13] implemented in our LASP program and deep potential[14,15], Gaussian Approximation Potentials (GAP)[8,16], Message Passing Neural Networks (MPNNs)[17,18], and various Equivariant Graph Neural Networks (EGNNs)[19] potentials such as PaiNN[20], Allegro[21] and MACE[22] emerged in recent years. The advent of MLPs enables long-time atomic simulations at high accuracy for the tasks using molecular dynamics (MD)[23–25] and global PES exploration[26], where DFT calculations are struggled to achieve enough sampling for the PES.

One of the major concerns of MLPs is the transferability and generality of the model. For example, DFT calculations can readily cover the whole periodic table but MLPs are generally limited to a few chemical elements of the training dataset. A few general-purpose MLPs emerged in recent years, including CHGNET[27] and DPA family[28–30] , which were trained on open datasets and lack the great PES data variety. By developing LASP software and global PES exploration methods, our group has accumulated a large quantity of global PES dataset in the past 20 years, which leads to the development of the High-order Pair-reduced Neural Network (HPNN) architecture[31] and the generalized global neural network (GGNN) potential that was trained on more than 6.4 million global PES dataset covering 89 elements. It allows for the first time the simulation of diverse systems without additional training while keeping the cost for simulation low. The success of GGNN encourages us to establish the LASPAI platform to deliver fast and reliable atomic simulation for everyone.

A generalized global neural network potential is not enough to achieve the general-purpose atomic simulation: one biggest challenge is how to transform the chemical knowledge/language into 3-D atomic models. This is where the recently-developed deep generative models[32–37] can come into help. Pioneering frameworks, particularly diffusion generative models (DGMs) have demonstrated success in generating conformations for small organic molecules and inorganic crystals within 20 atoms. For example, the diffusion-based models as represented by GeoDiff offer a fast and cost-effective route for generating three-dimensional (3D) molecular structures directly from two-dimensional (2D) molecular graphs[38–42] based on SchNet spatial encoder and Graph Isomorphism Network (GIN)[43] message passing[38,41]. One step further, the OA-ReactDiff model[44] utilizes a SE(3) equivariant neural network by training the Transition1x database[45], resulting in transition state (TS) generation for gas phase reactions based on the initial and final state. However, these models suffer from high time complexity due to the $O(N^2)$ computational complexity of full connected graphs, hindering the scalability of the model. Recently, by using the same High-order Pair-reduced network but further incorporating edge and time information (HPNN-ET), we managed to directly generate 3D structures of initial state (IS), TS and final state (FS) from 2D molecular graphs with high accuracy and high speed. This

leads to a Bootstrapping Diffusion Generative Model coupled with Potential Energy Surface (BDGM-PES) method[46] to explore automatedly the reaction space via a self-regressive pipeline with the integration of single-ended TS searching algorithm[47].

Given the powerful PES evaluation from GGNN potential and the 3D structure generation of HPNN-ET, we started to establish the LASPAI platform in the early of 2025. Our aim is to design a user-friendly, web-based portal to democratize access to advanced atomic simulations. LASPAI functions as an intuitive GUI for the LASP engine and expands its capabilities with task-oriented modules. By operating entirely within a standard web browser, it eliminates all setup and maintenance overhead for the end-user, requiring no compiling, installation, or configuration files. The platform's interactive GUI enables researchers not only to launch calculations with a few clicks but also to visualize, rotate, and modify molecular structures in real-time, fostering a more intuitive research process. By abstracting complex command-line syntax and system administration tasks, LASPAI empowers experimental scientists to directly apply the predictive power of computational chemistry to their work, thereby accelerating discovery and promoting a more integrated approach to scientific research. In the following sections, we overview the theoretical foundation of LASPAI platform and introduce briefly the major functionalities.

2 Theoretical foundation

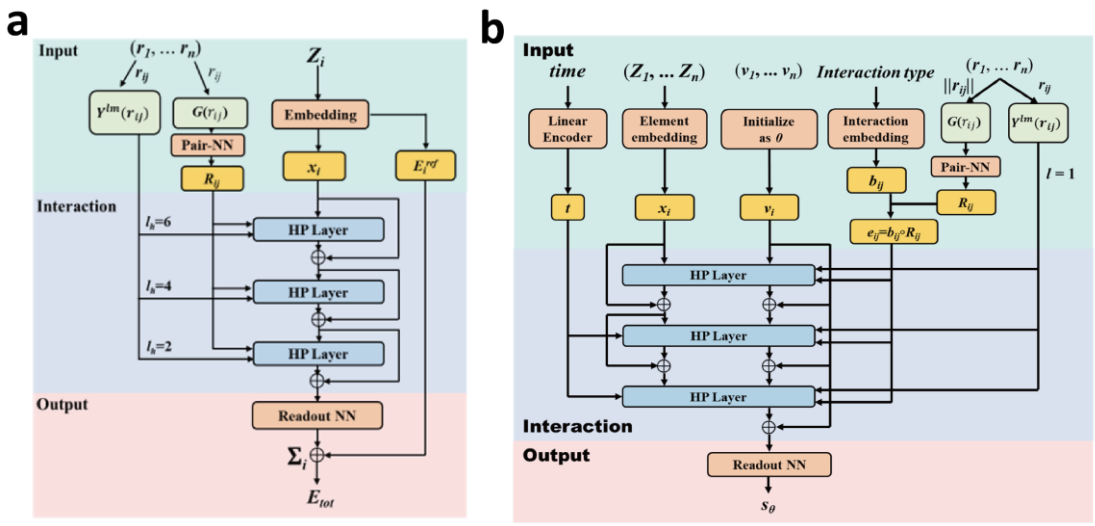


Figure 1. The HPNN architecture for (a) PES evaluation (generalized global neural network) and (b) structure generation (diffusion generative models). The figures are extracted from works by Yang et al.[31] and Guo et al. [46] respectively.

LASPAI platform equips with many machine learning models, including the GGNN potential for evaluation PES and a few generative models structure generation. These models cooperate seamlessly to achieve a high efficiency workflow towards sophisticated atomic simulation tasks. Despite different purposes of these machine learning models, all of them are related to high-order pair-reduced neural network architecture, i.e. HPNN, as elaborated in Figure 1.

2.1 PES evaluation

HPNN[31], as shown in Figure 1a, incorporates high-order spherical harmonics up to l=6 without using computationally intensive Clebsch-Gordan (CG) tensor product[48]. Instead of performing a single, complex tensor product, HPNN uses a series of three interaction layers, each of which incorporates spherical harmonics of different orders in a hierarchical fashion. The first layer uses the highest order ($l_h = 6$) to capture fine-grained local geometric details. Subsequent layers use progressively lower orders ($l = 4$ and $l = 2$) to aggregate more global, non-local information. A crucial element for efficiency is the reduction of the dimensionality of atomic features before they enter the pair-based message passing operation. The feature dimension is shrunk by a linear layer and then restored after the operation is complete. This significantly reduces the computational load of the most expensive part of the calculation without compromising the model's overall parameter space, as subsequent operations are performed on the full-dimension features. The message passing is formulated as a Hadamard product (element-wise multiplication) of the reduced-dimensional atomic features, followed by an outer product with individual spherical harmonics. The resulting angular messages are then transformed into a scalar representation and concatenated. This simpler approach achieves a linear computational complexity related to spherical harmonics, making it feasible to include very high orders (up to l=6). By combining these features, the HPNN model creates a lightweight yet powerful architecture capable of learning from a vast and complex global PES dataset. The HPNN architecture was initially trained on a massive dataset of 5.84 million configurations covering 83 elements that now expands to 6.4 million configurations covering 89 elements for LASPAI platform service, achieving a balance of high accuracy and high speed with a great generality for periodic table elements that was previously unattainable. The dataset was obtained by plane wave DFT calculations with GGA-PBE functional. The model reaches the root-mean-square errors of 7.3 meV/atom for energy and 0.16 eV/Å for force. To better account for the van der Waals interaction, we also implemented a highly efficient DFT-D3 algorithm on GPU platform, which allows for fast dispersion energy and force evaluation for large systems up to 100,000,000 atoms on a single Nvidia RTX 4090 GPU.

2.2 Automated Structure Generation

Our general generative model is based on HPNN-ET, which is a SE(3)-equivariant machine-learning model for generating 3D structures from molecular graph. HPNN-ET follows HPNN architecture in collecting the message between nodes (atoms). The model[46] is shown in Figure 1b. The input of HPNN-ET takes all available molecular information including atomic properties, bond connectivity from the 2D graph, interatomic distances, and temporal data from the diffusion process into a single, unified model. This allows it to efficiently process and learn from diverse data streams. The model uses high-order spherical harmonics to achieve more sensitive and accurate discrimination against the spatial arrangement of atoms, which is crucial for generating a large variety of complex geometries with high precision. The HPNN-ET model processes input by first embedding atomic and pairwise interaction data. This information is then updated through a series of high-order pair-reduced layers that perform message passing between atoms. Similar to our

HPNN model, the generative model reduces the feature dimension during the computationally intensive pairwise operations, which significantly enhances efficiency without sacrificing the model's overall complexity or accuracy. This allows the model to generate large molecules (over 100 atoms) with exceptional precision (less than 0.05 Å error) and speed. By combining this powerful generative model with machine learning potential for exploring the reaction's energy landscape, the HPNN-ET framework can generate reaction pathways autonomously, including intermediate and transition states, for complex catalytic systems.

LASPAI now includes 5 generative models, all derived from HPNN-ET architecture, but each aims for different tasks. Here we list the functionality, the dataset size and the efficiency of these five models.

i. Molecule generation

The molecule model generates 3D structures of molecules from molecular graph. The model was trained on a dataset of 437,952 molecules, including structures screened from PubChem database and drug-like molecules obtained from GEOM-DRUG[49] together with structures for organometallic catalytic reactions[50]. The molecule generation takes generally less than 15 seconds, e.g. large organic molecules with 40 atoms taking about 8 seconds.

ii. Solid generation

The solid model generates 3D structure of inorganic crystals from chemical formulas. It was trained on a large dataset containing over 682,637 structures, including 45,231 experimentally verified stable inorganic crystal structures with up to 20 atoms from Materials Project[51,52] and the rest from Alex-MP20 dataset provided by MatterGen[53]. The model can generate large structures containing 1000 atoms within merely 20 seconds.

iii. Molecular crystal generation

The molecular crystal model generates 3D structure of molecular crystals from molecule SMILES names. The rdkit[54] is used to obtain the connectivity between atoms. The generation process mixes the molecule generation and solid generation algorithm, adding embedded connectivity information during the message passing process. The connectivity is calculated in real-time during training process to guarantee correctness for molecules at the edge of the cells. It was trained on 320,000 molecular crystals from literature[55–62]. The model can generate a large cubic molecular crystal with 720 atoms and 10% occupation ratio in 21 seconds.

iv. Molecules inside cages

A derived generative model based on the molecular crystal model is developed to fill molecules inside cages. This is achieved by modifying the atom initialization process, where the Voronoi algorithm first identifies the cage region and the initial atom coordinates are restricted in a cubic region at the center of the cage. The inference process is then performed for the newly added molecules while fixing the cage atoms. The model can generate three candidate structures of water molecules inside an $AlPO_4$ zeolite cage of 288 atoms in 20 seconds.

v. Reaction generation

The reaction generation model generates TS structures of reactions from the IS or FS structures. The initial training dataset is based on the dataset for the molecule generation. To better predict the TS, we utilize an iterative procedure to train the reaction generation model. First, additional constraints to reaction atoms, i.e. the distance between these atoms, are applied to generate the guessed TS. Constrained Broyden Dimer (CBD) algorithm[47,63] is then utilized to find the TS structure on the GGNN PES. The reaction model is retrained by adding these newly identified TS structures. The TS structure of ~ 40-atom organic reaction system can be generated within 10 seconds.

3 Architecture and workflow

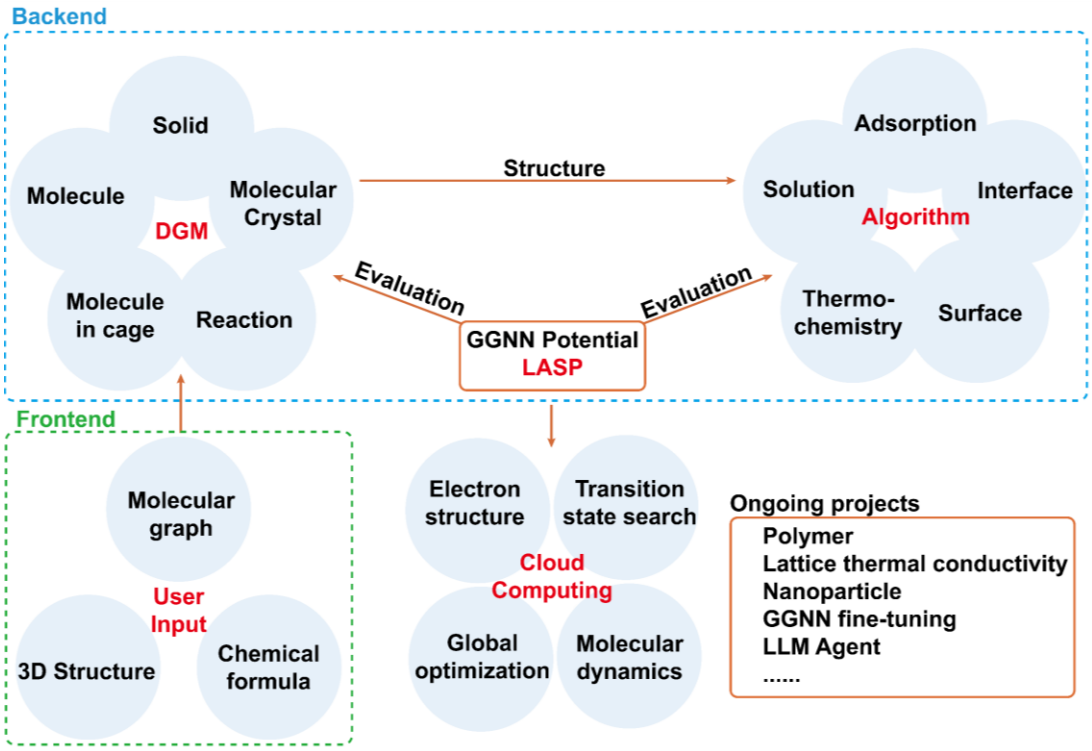


Figure 2. The architecture and user workflow of LASPAI website. The blue and green dotted box shows the architecture of the backend and frontend, respectively.

3.1 Web-based architecture

LASPAI platform is realized on a frontend on the web browser served by a backend workstation, each playing a crucial role in delivering seamless user experience for atomic simulation. The LASPAI architecture is shown in Figure 2. The frontend comprises various integrated components, including essential elements like headers and sidebars, along with the specialized visualization components. It is specifically designed for graphic interactive structure generation and simulation, taking the chemical language from human via sophisticated graphic software, such as ketcher[64] for molecular graph input, 3dmol[65] for 3D structure visualization and house-developed LASPView for 3D structure operation. Each functionality operates as a single-page application (SPA), and the platform ensures that the navigation between different pages occurs smoothly without the need to reload

the whole browser. The communication between the frontend and the backend is facilitated through asynchronous HTTP requests that achieve an efficient data exchange.

The backend, hosted on a remote server, is responsible for processing requests initiated by the user. Upon receiving a request, it validates the provided parameters and subsequently creates a pending new task entry. This streamlined approach to backend operations is designed to provide instant feedback to users, preventing any perceived delays or the website from becoming unresponsive. Concurrently, independent worker processes are continuously listening to new pending tasks. Once a worker identifies a pending task, it invokes pre-compiled, high-performance LASP program and projects including phononpy[66], rdkit[54], ase[67], pymatgen[68] and packmol[69] to execute the required computational task. Throughout this process, the workers periodically update the task's status to the backend, which in turn relays this information to the frontend, ultimately presenting the status to the user. After a task is successfully completed, the worker process marks it as complete, and both the backend and frontend, through application programming interface (API) endpoint polling, update this status information to the user, providing a comprehensive overview of their simulation progress.

LASPAI platform has four main modules, namely the GGNN potential module, the generation module, the reaction module and the cloud computing module, each module contains a few independent web pages. The GGNN potential module provides four functionalities arranged into four pages, including single point energy calculations, (variable-cell) structure optimization, PES exploration trajectories, solid phonon and vibrational frequency calculations. These can be utilized to fast validate the performance of GGNN potential, from energy to the second derivative of energy, on user-provided structures.

The generation module equipped with various structure generation algorithms and general generative models can generate a wide range of structures, which are arranged into 11 pages, including molecules, molecular crystal, solid, polymers, nanoparticles, crystals and surfaces, interfaces, structure filling (e.g. solutions and molecules inside cages) and preequilibrium. Preequilibrium simulations are utilized to run short-time molecular dynamics and global optimization for these as-generated structures.

The reaction module aims to identify the TS of molecular reactions, surface reactions and solid-solid phase transition from different initial structures, such as molecular graph, guessed TS structure, IS and FS structures. The cloud computing module provides graphic interfaces to run long simulation jobs for global optimization, MD simulation, TS location and electron structure calculation. All web pages utilize modern web-based architecture and the task-oriented page layout to maximumly simplify the procedure of atomic simulation.

3.2 Workflow of Simulation

LASPAI platform aims to offer a simple and continuous workflow of atomic simulations for both experienced computational chemists and nonprofessionals. In most cases, users simply need to specify the chemical formula or draw the molecule graph to initiate simulations. Throughout the simulation process, the website provides intuitive, real-time re-

sults, allowing for easy viewing and analysis of structures and idea proofing. Here we illustrate the website's workflow using a common scenario, i.e. to build a composite structure and perform molecular dynamics to validate the constructed model.

Multiple-component composite systems, such as mixed solutions, solid-solid-interfaces, are prevalent in material world. However, constructing a desired and physically meaningful composite structure is often challenging due to their multi-component and non-standard nature. LASPAI provides a four-step workflow to solve this challenge, namely, structure construction, local optimization, preequilibrium and cloud computing, where the first three steps can be finished conveniently on the LASPAI platform and the final production step is achieved by submitting the task to a remote supercomputer.

The structure filling page is the main tool to address the challenge of composite generation. Currently, this page can process solid-liquid interfaces, liquid-liquid mixtures, solutions, and molecules within caged materials. By specifying the components, the desired composite structures can be generated by mouse clicking. These structures can then be finely modified and locally optimized using LASP, providing an ideal starting point for subsequent atomic simulations. By sending the constructed composite structure to preequilibrium page, one can then perform SSW global optimization and MD simulations directly. A GUI is provided to directly view the structure together with the input boxes to set the initial settings of simulation. After the simulation task is finished, the user can view the result and the trajectory of the simulation. Taking the preequilibrated structure, users may further perform cloud computing for long-time simulations, which could finally produce the concerned physiochemical properties.

4 Examples of LASPAI

LASPAI can streamline a variety of sophisticated simulation tasks by combining the elementary functions with different pages. The prerequisites for computational experience and simulation knowledge are maximally reduced, including the code/script writing, program switching, manual file transfers, and invoking external visualization tools as required in traditional atomic simulation. Leveraging the power of LASP software, generative models and LASPAI workflow, the platform offers a simple, intuitive, fast, and precise experience towards the future of atomic simulation.

4.1 Solid structure prediction

One important feature of LASPAI is its AI-powered structure generation. It allows users to generate solid structures with the least information input, making it particularly effective for predicting solid structures with complex PES. Taking TiO_2 as an example, we illustrate how LASPAI can predict the crystal structure of TiO_2. TiO_2, as an important material, is widely used in many fields. It primarily exhibits two stable crystal structures: rutile and anatase[70], with space groups $P4_2/mnm$ and $I4_1/amd$, respectively. The transition between these two crystal phases occurs above 800 K, suggesting a relatively high barrier of phase transition [71]. By utilizing the solid generation page on LASPAI, we can obtain their crystal structures and assess their relative stability.

a

Unit Energy eV/f.u.	Chemical Formula	Energy eV	Space Group	Volume Å³	Unit Volume Å³/f.u.
-27.0064	O4-Ti2	-54.0128	I4_1/am...	69.6363	34.8181
-26.9532	O4-Ti2	-53.9064	P4_2/m...	63.8473	31.9236
-26.8459	O10-Ti5	-134.2296	C2 (5)	174.7062	34.9412
-26.7925	O8-Ti4	-107.1700	C2/m (12)	141.2947	35.3237
-26.7815	O10-Ti5	-133.9077	P1 (1)	180.2376	36.0475
-26.7795	O6-Ti3	-80.3385	P-31m (...	93.5795	31.1932
-26.7790	O8-Ti4	-107.1159	C2/m (12)	133.8940	33.4735
-26.7482	O4-Ti2	-53.4964	P2_1/m ...	70.2028	35.1014
-26.6993	O6-Ti3	-80.0980	Cm (8)	131.5900	43.8633

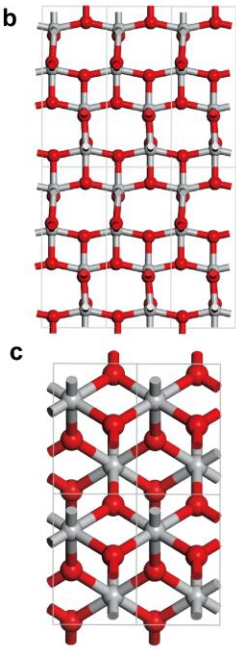


Figure 3. The generation of different polymorphs of TiO_2 using the solid generation page. (a) The list of generated TiO_2 structures sorted by unit energy. (b) and (c) are the first and second entries in (a), corresponding to the generated and optimized structures of anatase and rutile, respectively. The red and grey spheres represent O and Ti atoms, respectively.

To start, we simply specify the chemical formula "TiO2" on the page and specify the number of atoms per cell as random. Other specifications of the generation like the lattice and occupation ratio parameters are optional. With three iterations of generation (clicking three times generation button), we obtain a list of generated structures, the top-ranked ones among which are shown in Figure 3a. The first two structures in the list share identical space groups with rutile and anatase, and by viewing their atomic structure, as shown in Figure 3b and c, we confirm that the generated structures correspond to these two stable crystal forms. It should be mentioned that traditional PES algorithms would require significantly more time to explore the structures of TiO_2[23]. With the aid of our generative model and the GGNN PES evaluation, it is now possible to obtain stable solid structures conveniently and visualize them intuitively, thereby significantly enhancing the material design. Each generation takes about 20 seconds and produces around 10

structures. It takes approximately 1 minute in total to identify the rutile and anatase phases from scratch.

4.2 Adsorption structure prediction

Adsorption is a fundamental process for molecule interaction with materials. LASPAI platform facilitates rapid exploration of molecular adsorption on surfaces. To give an example, we show how LASPAI can rapidly predict the pyridine adsorption on a γ-Al_2O_3 surface. The pyridine adsorption is a well-established probe reaction in surface chemistry for identifying surface acidic sites through characteristic peaks in infrared spectroscopy. The workflow of this adsorption task is shown in Figure 4a. It involves four pages from the generation module, namely the molecule generation page, the solid generation page, the surface cutting page and the adsorption generation page. The initial bulk structure may come from several sources, e.g. our solid generative model, the database of crystal structures and custom input structures.

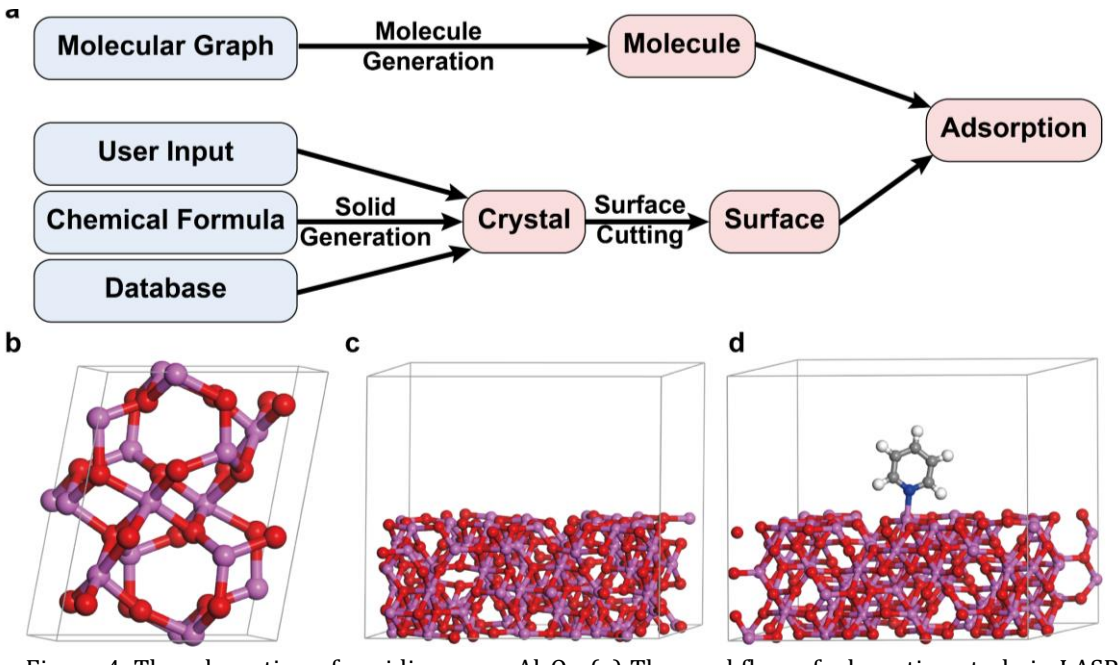


Figure 4. The adsorption of pyridine on γ-Al_2O_3. (a) The workflow of adsorption study in LASPAI. (b) The minimum γ-Al_2O_3 (γ-AD) structure found by Yang et al.[72] (c) The (512) facet of γ-AD structure, corresponding to (110) facet of γ-Al_2O_3. (d) The generated adsorption structure of pyridine on Al_2O_3. The white, grey, blue, red and pink spheres represent H, C, N, O and Al atoms, respectively.

For γ-Al_2O_3, we adopt the γ-AD model suggested by Yang et al[72] (Figure 4b) and its (512) facets is shown in Figure 4c, corresponding to the conventional (110) facet of γ-Al_2O_3 using O-sublattice. To start, we upload the γ-AD structure on the surface page and cut the (512) slab; the pyridine molecule is also generated and optimized on the molecule page. Both the slab structure and the molecule are then transferred to the adsorption page. After the adsorption generation, a list of adsorption structures is obtained and the most stable is shown in Figure 4d, where the adsorption energy is reported to be 2.11 eV. The pyridine molecule bonds with the Lewis acid site (Al) via its N-end in a vertical configuration with the N-Al bond length of 1.97 Å. Subsequent vibrational frequency analysis, also available on the adsorption page, reveals a characteristic ring stretching mode at 1491

cm^{-1}, corresponding to the v19a frequency[73], red-shifted by 10 cm^{-1} compared to the gas phase pyridine[74]. This computed frequency is well consistent with experimental results reported by Phung et al.[75] of 1490 cm^{-1}.

The molecule generation takes 5 seconds from molecular graph to 3D structure. Surface cutting and optimization takes less than 10 seconds, and the adsorption structure generation takes around 25 seconds for one iteration. The frequency analysis takes less than 30 seconds. The whole procedure takes approximately 70 seconds.

4.3 Liquid phase separation

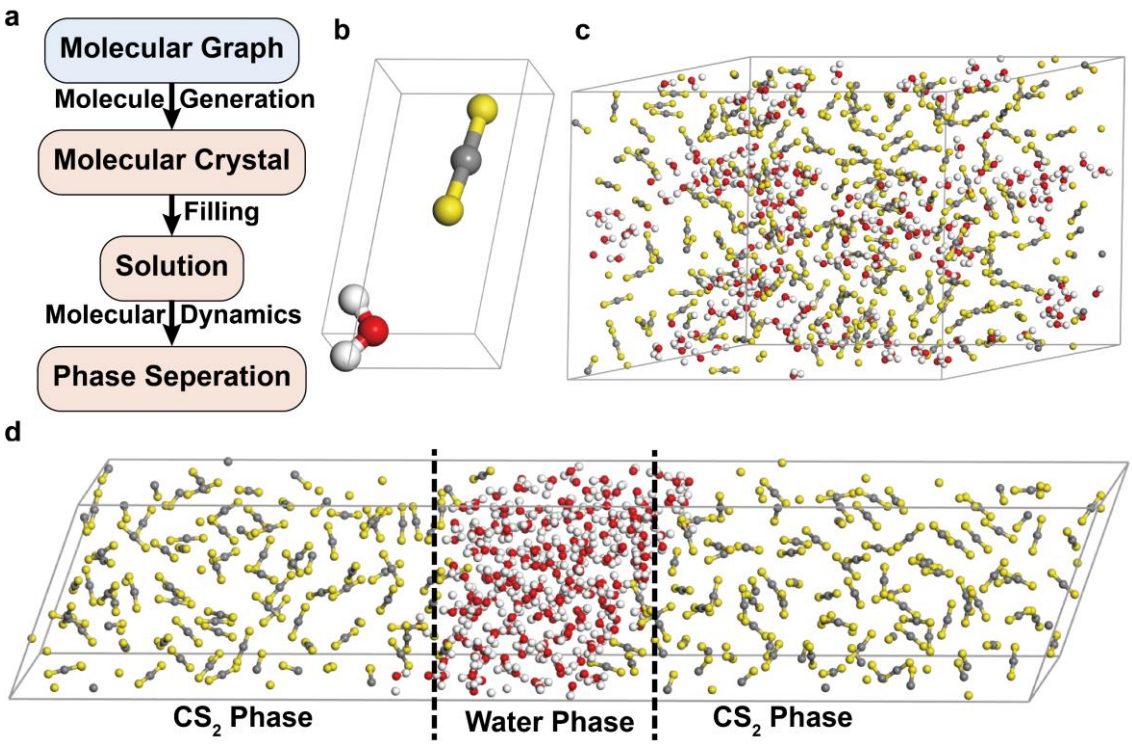


Figure 5. Atomic simulation of oil/water phase separation. (a) The workflow for performing phase separation simulation on LASPAI. (b) The generated 1:1 molecular crystal of water and CS_2. (c) The 1:1 mixed starting solution structure. (d) The structure after MD simulation with the two phases separated. The white, grey, red and yellow spheres represent H, C, O and S atoms, respectively.

Liquid phase separation is a common experimental operation in chemistry laboratory. The separation of two phases from a mixed solution can be simulated through long-time MD simulations. As summarized in Figure 5a, LASPAI platform offers a convenient workflow to finish the simulation task by simply specifying the chemical formulas of the solvent and solute as input. This process involves four LASPAI pages included in the generation module and cloud computing module, namely the molecule crystal generation, the filling, the preequilibrium and the molecular dynamics page in cloud computing.

Taking the phase separation between water and carbon disulfide (CS_2) as an example, we can first generate the mixture as molecular crystal by specifying the molecular graphs of the two molecules and the best molecular crystal obtained is shown in Figure 5b, which already tells the weak interaction between two molecules from the long distance separation of the two molecules. This molecular crystal structure is then transferred to the filling

page to populate a large cell with the two molecules. The resulting structure, as shown in Figure 5c, is a large solution system comprising 1086 atoms, with 181 water molecules and 181 CS_2 molecules. A preequilibrium using MD simulation of ~ 2.5 picoseconds is then performed on the solution mixture. After preequilibrium, the volume of the system shrinks from 33,072 Å^3 to 30,058 Å^3, resulting in a more stable mixture of water and CS_2 with density of 0.94 g/cm^3. This structure is taken as the starting structure for long-time MD simulation by submitting the task to the cloud computing platform and the task finishes until a total simulation time of 320 picoseconds is reached. Both the preequilibrium and MD simulation process is performed under the constant-pressure-temperature (NPT) ensemble at 300 K, with a time step of 0.5 femtoseconds. The MD simulation process can be monitored on the website, and the structures during the simulation can be viewed directly. As shown in Figure 5d, the separation of the water and CS_2 phases is clearly revealed by MD, in agreement with the known immiscibility of water and CS_2.

The generation of the molecular crystal takes about 40 seconds, and the filling procedure takes approximately 20 seconds to build the solution cell. The preequilibrium process takes 175 seconds, and the final long-time MD simulation takes 15 hours and 18 minutes.

5 Conclusion and Outlook

This work overviews the architecture and workflow of our recently-released LASPAI platform (www.laspai.com) born for the next-generation atomic simulation, which integrates a user-friendly web frontend and a powerful backend server equipped with the latest generalized global MLP (GGNN) and the general generative models. We introduced the theoretical foundation of LASPAI, the interactive web GUI with its backend server, the workflow and finally presented three examples finished by LASPAI, from crystal structure prediction to molecular adsorption and to liquid-phase separation. We show that with the help of generalized MLP and the powerful generative model, LASPAI expands greatly the frontier of atomic simulation, not only being more convenient and autonomous, but also being more intelligent towards material and molecule design from first principles. While there are still many ongoing projects of LASPAI, e.g. polymer modelling, lattice thermal conductivity computation, GGNN fine-tuning and the combination of force-field potentials with GGNN, the next major move of LASPAI will be the integration of large language model to drive the LASPAI tools, facilitating higher degrees of automation and intelligence. We believe that by abstracting complex command-line syntax and deep knowledge on computation methodology, LASPAI empowers all areas of science students and researchers to directly exploit the predictive power of computational chemistry, thereby accelerating discovery and promoting AI-integrated modern scientific research.

Acknowledgement

This work was supported by the National Key Research and Development Program of China (2024YFA1509600), the National Science Foundation of China (12188101, 22033003, 92472113), the Fundamental Research Funds for the Central Universities (20720250005, 20720220011), Science & Technology Commission of Shanghai Municipality (2024ZDSYS02), the robotic AI-Scientist platform of Chinese Academy of Science, and the Tencent Foundation for XPLORER PRIZE. The computations in this research were performed using the CFFF platform of Fudan University.

References

(1) Cazorla, C.; Boronat, J. Simulation and Understanding of Atomic and Molecular Quantum Crystals. *Rev. Mod. Phys.* **2017**, *89* (3), 035003. https://doi.org/10.1103/RevModPhys.89.035003.

(2) Foucaud, Y.; Badawi, M.; Filippov, L.; Filippova, I.; Lebègue, S. A Review of Atomistic Simulation Methods for Surface Physical-Chemistry Phenomena Applied to Froth Flotation. *Minerals Engineering* **2019**, *143*, 106020. https://doi.org/10.1016/j.mineng.2019.106020.

(3) Liu, H.; Zhao, Z.; Zhou, Q.; Chen, R.; Yang, K.; Wang, Z.; Tang, L.; Bauchy, M. Challenges and opportunities in atomistic simulations of glasses: a review. *Comptes Rendus. Géoscience* **2022**, *354* (S1), 35–77. https://doi.org/10.5802/crgeos.116.

(4) Steinhauser, M. O.; Hiermaier, S. A Review of Computational Methods in Materials Science: Examples from Shock-Wave and Polymer Physics. *International Journal of Molecular Sciences* **2009**, *10* (12), 5135–5216. https://doi.org/10.3390/ijms10125135.

(5) Kulichenko, M.; Nebgen, B.; Lubbers, N.; Smith, J. S.; Barros, K.; Allen, A. E. A.; Habib, A.; Shinkle, E.; Fedik, N.; Li, Y. W.; Messerly, R. A.; Tretiak, S. Data Generation for Machine Learning Interatomic Potentials and Beyond. *Chem. Rev.* **2024**, *124* (24), 13681–13714. https://doi.org/10.1021/acs.chemrev.4c00572.

(6) Li, Y.; Zhang, X.; Shen, L. A Critical Review of Machine Learning Interatomic Potentials and Hamiltonian. *jmi* **2025**, *5* (4). https://doi.org/10.20517/jmi.2025.17.

(7) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108* (5), 058301. https://doi.org/10.1103/PhysRevLett.108.058301.

(8) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104* (13), 136403. https://doi.org/10.1103/PhysRevLett.104.136403.

(9) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98* (14), 146401. https://doi.org/10.1103/PhysRevLett.98.146401.

(10) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B* **2013**, *87* (18), 184115. https://doi.org/10.1103/PhysRevB.87.184115.

(11) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121* (16), 10142–10186. https://doi.org/10.1021/acs.chemrev.0c01111.

(12) Zhang, Y.; Xia, J.; Jiang, B. Physically Motivated Recursively Embedded Atom Neural Networks: Incorporating Local Completeness and Nonlocality. *Phys. Rev. Lett.* **2021**, *127* (15), 156002. https://doi.org/10.1103/PhysRevLett.127.156002.

(13) Yang, Z.-X.; Xie, X.-T.; Kang, P.-L.; Wang, Z.-X.; Shang, C.; Liu, Z.-P. Many-Body Function Corrected Neural Network with Atomic Attention (MBNN-Att) for Molecular Property Prediction. *J. Chem. Theory Comput.* **2024**, *20* (15), 6717–6727. https://doi.org/10.1021/acs.jctc.4c00660.

(14) Zhang, L.; Lin, D.-Y.; Wang, H.; Car, R.; E, W. Active Learning of Uniformly Accurate Inter-Atomic Potentials for Materials Simulation. *Phys. Rev. Materials* **2019**, *3* (2), 023804. https://doi.org/10.1103/PhysRevMaterials.3.023804.

(15) Wang, H.; Guo, X.; Zhang, L.; Wang, H.; Xue, J. Deep Learning Inter-Atomic Potential Model for Accurate Irradiation Damage Simulations. *Applied Physics Letters* **2019**, *114* (24), 244101. https://doi.org/10.1063/1.5098061.

(16) Klawohn, S.; Darby, J. P.; Kermode, J. R.; Csányi, G.; Caro, M. A.; Bartók, A. P. Gaussian Approximation Potentials: Theory, Software Implementation and Application Examples. *J. Chem. Phys.* **2023**, *159* (17), 174108. https://doi.org/10.1063/5.0160898.

(17) Liu, Y.; Wang, L.; Liu, M.; Zhang, X.; Oztekin, B.; Ji, S. Spherical Message Passing for 3D Graph Networks. arXiv November 24, 2022. https://doi.org/10.48550/arXiv.2102.05013.

(18) Gasteiger, J.; Yeshwanth, C.; Günnemann, S. Directional Message Passing on Molecular Graphs via Synthetic Coordinates. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc., 2021; Vol. 34, pp 15421–15433.

(19) Satorras, V. G.; Hoogeboom, E.; Welling, M. E(n) Equivariant Graph Neural Networks. In *Proceedings of the 38th International Conference on Machine Learning*; PMLR, 2021; pp 9323–9332.

(20) Schütt, K. T.; Unke, O. T.; Gastegger, M. Equivariant Message Passing for the Prediction of Tensorial Properties and Molecular Spectra. arXiv June 7, 2021. https://doi.org/10.48550/arXiv.2102.03150.

(21) Musaelian, A.; Batzner, S.; Johansson, A.; Sun, L.; Owen, C. J.; Kornbluth, M.; Kozinsky, B. Learning Local Equivariant Representations for Large-Scale Atomistic Dynamics. *Nat Commun* **2023**, *14* (1), 579. https://doi.org/10.1038/s41467-023-36329-y.

(22) Batatia, I.; Kovacs, D. P.; Simm, G.; Ortner, C.; Csanyi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. *Advances in Neural Information Processing Systems* **2022**, *35*, 11423–11436.

(23) Huang, S.-D.; Shang, C.; Kang, P.-L.; Liu, Z.-P. Atomic Structure of Boron Resolved Using Machine Learning and Global Sampling. *Chem. Sci.* **2018**, *9* (46), 8644–8655. https://doi.org/10.1039/C8SC03427C.

(24) Kang, P.-L.; Shang, C.; Liu, Z.-P. Glucose to 5-Hydroxymethylfurfural: Origin of Site-Selectivity Resolved by Machine Learning Based Reaction Sampling. *J. Am. Chem. Soc.* **2019**, *141* (51), 20525–20536. https://doi.org/10.1021/jacs.9b11535.

(25) Liu, Q.-Y.; Shang, C.; Liu, Z.-P. In Situ Active Site for CO Activation in Fe-Catalyzed Fischer–Tropsch Synthesis from Machine Learning. *J. Am. Chem. Soc.* **2021**, *143* (29), 11109–11120. https://doi.org/10.1021/jacs.1c04624.

(26) Huang, S.-D.; Shang, C.; Zhang, X.-J.; Liu, Z.-P. Material Discovery by Combining Stochastic Surface Walking Global Optimization with a Neural Network. *Chem. Sci.* **2017**, *8* (9), 6327–6337. https://doi.org/10.1039/C7SC01459G.

(27) Deng, B.; Zhong, P.; Jun, K.; Riebesell, J.; Han, K.; Bartel, C. J.; Ceder, G. CHGNet as a Pretrained Universal Neural Network Potential for Charge-Informed Atomistic Modelling. *Nat Mach Intell* **2023**, *5* (9), 1031–1041. https://doi.org/10.1038/s42256-023-00716-3.

(28) Zhang, D.; Bi, H.; Dai, F.-Z.; Jiang, W.; Liu, X.; Zhang, L.; Wang, H. Pretraining of Attention-Based Deep Learning Potential Model for Molecular Simulation. *npj Comput Mater* **2024**, *10* (1), 94. https://doi.org/10.1038/s41524-024-01278-7.

(29) Zhang, D.; Liu, X.; Zhang, X.; Zhang, C.; Cai, C.; Bi, H.; Du, Y.; Qin, X.; Peng, A.; Huang, J.; Li, B.; Shan, Y.; Zeng, J.; Zhang, Y.; Liu, S.; Li, Y.; Chang, J.; Wang, X.; Zhou, S.; Liu, J.; Luo, X.; Wang, Z.; Jiang, W.; Wu, J.; Yang, Y.; Yang, J.; Yang, M.; Gong, F.-Q.; Zhang, L.; Shi, M.; Dai, F.-Z.; York, D. M.; Liu, S.; Zhu, T.; Zhong, Z.; Lv, J.; Cheng, J.; Jia, W.; Chen, M.; Ke, G.; E, W.; Zhang, L.; Wang, H. DPA-2: A Large Atomic Model as a Multi-Task Learner. *npj Comput Mater* **2024**, *10* (1), 293. https://doi.org/10.1038/s41524-024-01493-2.

(30) Zhang, D.; Peng, A.; Cai, C.; Li, W.; Zhou, Y.; Zeng, J.; Guo, M.; Zhang, C.; Li, B.; Jiang, H.; Zhu, T.; Jia, W.; Zhang, L.; Wang, H. A Graph Neural Network for the Era of Large Atomistic Models. arXiv June 9, 2025. https://doi.org/10.48550/arXiv.2506.01686.

(31) Zheng-Xin, Y.; Xin-Tian, X.; Zhen-Xiong, W.; Dong-Xiao, C.; Zi-Xing, G.; Jia-Jie, D.; Qian-Yu, L.; Cheng, S.; Zhi-Pan, L. High-Order Pair-Reduced Neural Network Architecture for Global Potential Energy Surface Exploration Across the Periodic Table. *Sci. China. Chemistry*. https://doi.org/10.1007/s11426-025-3054-y.

(32) Makoś, M. Z.; Verma, N.; Larson, E. C.; Freindorf, M.; Kraka, E. Generative Adversarial Networks for Transition State Geometry Prediction. *J. Chem. Phys.* **2021**, *155* (2), 024116. https://doi.org/10.1063/5.0055094.

(33) Ganea, O.-E.; Pattanaik, L.; Coley, C. W.; Barzilay, R.; Jensen, K. F.; Green, W. H.; Jaakkola, T. S. GeoMol: Torsional Geometric Generation of Molecular 3D Conformer Ensembles. arXiv June 8, 2021. https://doi.org/10.48550/arXiv.2106.07802.

(34) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.

(35) Westermayr, J.; Gilkes, J.; Barrett, R.; Maurer, R. J. High-Throughput Property-Driven Generative Design of Functional Organic Molecules. *Nat Comput Sci* **2023**, *3* (2), 139–148. https://doi.org/10.1038/s43588-022-00391-1.

(36) Lim, J.; Ryu, S.; Kim, J. W.; Kim, W. Y. Molecular Generative Model Based on Conditional Variational Autoencoder for de Novo Molecular Design. *J Cheminform* **2018**, *10* (1), 31. https://doi.org/10.1186/s13321-018-0286-7.

(37) Choi, S. Prediction of Transition State Structures of Gas-Phase Chemical Reactions via Machine Learning. *Nat Commun* **2023**, *14* (1), 1168. https://doi.org/10.1038/s41467-023-36823-3.

(38) Xu, M.; Yu, L.; Song, Y.; Shi, C.; Ermon, S.; Tang, J. GeoDiff: A Geometric Diffusion Model for Molecular Conformation Generation. arXiv March 6, 2022. https://doi.org/10.48550/arXiv.2203.02923.

(39) Jing, B.; Corso, G.; Chang, J.; Barzilay, R.; Jaakkola, T. Torsional Diffusion for Molecular Conformer Generation. arXiv March 1, 2023. https://doi.org/10.48550/arXiv.2206.01729.

(40) Morehead, A.; Cheng, J. Geometry-Complete Diffusion for 3D Molecule Generation and Optimization. *Commun Chem* **2024**, *7* (1), 150. https://doi.org/10.1038/s42004-024-01233-z.

(41) Kim, S.; Woo, J.; Kim, W. Y. Diffusion-Based Generative AI for Exploring Transition States from 2D Molecular Graphs. *Nat Commun* **2024**, *15* (1), 341. https://doi.org/10.1038/s41467-023-44629-6.

(42) Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; Poole, B. Score-Based Generative Modeling through Stochastic Differential Equations. arXiv February 10, 2021. https://doi.org/10.48550/arXiv.2011.13456.

(43) Xu, K.; Hu, W.; Leskovec, J.; Jegelka, S. How Powerful Are Graph Neural Networks? arXiv February 22, 2019. https://doi.org/10.48550/arXiv.1810.00826.

(44) Duan, C.; Du, Y.; Jia, H.; Kulik, H. J. Accurate Transition State Generation with an Object-Aware Equivariant Elementary Reaction Diffusion Model. *Nat Comput Sci* **2023**, *3* (12), 1045–1055. https://doi.org/10.1038/s43588-023-00563-7.

(45) Schreiner, M.; Bhowmik, A.; Vegge, T.; Busk, J.; Winther, O. Transition1x - a Dataset for Building Generalizable Reactive Machine Learning Potentials. *Sci Data* **2022**, *9* (1), 779. https://doi.org/10.1038/s41597-022-01870-w.

(46) Zi-Xing, G.; Jin-Peng, T.; Zhen-Xiong, W.; Qi-Ming, Q.-M.; Si-Cong, M.; Cheng, S.; Lin, C.; Zhi-Pan, L. Reaction Design from Artificial Intelligence Guided Potential Energy Surface Exploration: Suzuki Coupling from Theory to Experiment.

(47) Shang, C.; Liu, Z.-P. Constrained Broyden Minimization Combined with the Dimer Method for Locating Transition State of Complex Reactions. *J. Chem. Theory Comput.* **2010**, *6* (4), 1136–1144. https://doi.org/10.1021/ct9005147.

(48) Thomas, N.; Smidt, T.; Kearnes, S.; Yang, L.; Li, L.; Kohlhoff, K.; Riley, P. Tensor Field Networks: Rotation- and Translation-Equivariant Neural Networks for 3D Point Clouds. arXiv May 18, 2018. https://doi.org/10.48550/arXiv.1802.08219.

(49) Axelrod, S.; Gómez-Bombarelli, R. GEOM, Energy-Annotated Molecular Conformations for Property Prediction and Molecular Generation. *Sci Data* **2022**, *9* (1), 185. https://doi.org/10.1038/s41597-022-01288-4.

(50) Ma, S.; Shang, C.; Wang, C.-M.; Liu, Z.-P. Thermodynamic Rules for Zeolite Formation from Machine Learning Based Global Optimization. *Chem. Sci.* **2020**, *11* (37), 10113–10118. https://doi.org/10.1039/D0SC03918G.

(51) Horton, M. K.; Huck, P.; Yang, R. X.; Munro, J. M.; Dwaraknath, S.; Ganose, A. M.; Kingsbury, R. S.; Wen, M.; Shen, J. X.; Mathis, T. S.; Kaplan, A. D.; Berket, K.; Riebesell, J.; George, J.; Rosen, A. S.; Spotte-Smith, E. W. C.; McDermott, M. J.; Cohen, O. A.; Dunn, A.; Kuner, M. C.; Rignanese, G.-M.; Petretto, G.; Waroquiers, D.; Griffin, S. M.; Neaton, J. B.; Chrzan, D. C.; Asta, M.; Hautier, G.; Cholia, S.; Ceder, G.; Ong, S. P.; Jain, A.; Persson, K. A. Accelerated Data-Driven Materials Science with the Materials Project. *Nat. Mater.* **2025**, *24* (10), 1522–1532. https://doi.org/10.1038/s41563-025-02272-0.

(52) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1* (1), 011002. https://doi.org/10.1063/1.4812323.

(53) Zeni, C.; Pinsler, R.; Zügner, D.; Fowler, A.; Horton, M.; Fu, X.; Wang, Z.; Shysheya, A.; Crabbé, J.; Ueda, S.; Sordillo, R.; Sun, L.; Smith, J.; Nguyen, B.; Schulz, H.; Lewis, S.; Huang, C.-W.; Lu, Z.; Zhou, Y.; Yang, H.; Hao, H.; Li, J.; Yang, C.; Li, W.; Tomioka, R.; Xie, T. A Generative Model for Inorganic Materials Design. *Nature* **2025**, *639* (8055), 624–632. https://doi.org/10.1038/s41586-025-08628-5.

(54) *RDKit*. https://www.rdkit.org/ (accessed 2025-10-31).

(55) Vaitkus, A.; Merkys, A.; Sander, T.; Quirós, M.; Thiessen, P. A.; Bolton, E. E.; Gražulis, S. A Workflow for Deriving Chemical Entities from Crystallographic Data and Its Application to the Crystallography Open Database. *Journal of Cheminformatics* **2023**, *15* (1), 123. https://doi.org/10.1186/s13321-023-00780-2.

(56) Merkys, A.; Vaitkus, A.; Grybauskas, A.; Konovalovas, A.; Quirós, M.; Gražulis, S. Graph Isomorphism-Based Algorithm for Cross-Checking Chemical and Crystallographic Descriptions. *Journal of Cheminformatics* **2023**, *15* (1), 25. https://doi.org/10.1186/s13321-023-00692-1.

(57) Vaitkus, A.; Merkys, A.; Gražulis, S. Validation of the Crystallography Open Database Using the Crystallographic Information Framework. *J Appl Cryst* **2021**, *54* (2), 661–672. https://doi.org/10.1107/S1600576720016532.

(58) Quirós, M.; Gražulis, S.; Girdzijauskaitė, S.; Merkys, A.; Vaitkus, A. Using SMILES Strings for the Description of Chemical Connectivity in the Crystallography Open Database. *Journal of Cheminformatics* **2018**, *10* (1), 23. https://doi.org/10.1186/s13321-018-0279-6.

(59) Gražulis, S.; Merkys, A.; Vaitkus, A.; Okulič-Kazarinas, M. Computing Stoichiometric Molecular Composition from Crystal Structures. *J Appl Cryst* **2015**, *48* (1), 85–91. https://doi.org/10.1107/S1600576714025904.

(60) Gražulis, S.; Daškevič, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quirós, M.; Serebryanaya, N. R.; Moeck, P.; Downs, R. T.; Le Bail, A. Crystallography Open Database (COD): An Open-Access Collection of Crystal Structures and Platform for World-Wide Collaboration. *Nucleic Acids Res* **2012**, *40* (D1), D420–D427. https://doi.org/10.1093/nar/gkr900.

(61) Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A. Crystallography Open Database – an Open-Access Collection of Crystal Structures. *J Appl Cryst* **2009**, *42* (4), 726–729. https://doi.org/10.1107/S0021889809016690.

(62) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Cryst B* **2016**, *72* (2), 171–179. https://doi.org/10.1107/S2052520616003954.

(63) Shang, C.; Liu, Z.-P. Constrained Broyden Dimer Method with Bias Potential for Exploring Potential Energy Surface of Multistep Reaction Process. *J. Chem. Theory Comput.* **2012**, *8* (7), 2215–2222. https://doi.org/10.1021/ct300250h.

(64) Karulin, B.; Kozhevnikov, M. Ketcher: Web-Based Chemical Structure Editor. *Journal of Cheminformatics* **2011**, *3* (1), P3. https://doi.org/10.1186/1758-2946-3-S1-P3.

(65) Rego, N.; Koes, D. 3Dmol.Js: Molecular Visualization with WebGL. *Bioinformatics* **2015**, *31* (8), 1322–1324. https://doi.org/10.1093/bioinformatics/btu829.

(66) Togo, A. First-Principles Phonon Calculations with Phonopy and Phono3py. *J. Phys. Soc. Jpn.* **2023**, *92* (1), 012001. https://doi.org/10.7566/JPSJ.92.012001.

(67) Hjorth Larsen, A.; Jørgen Mortensen, 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.; Bjerre Jensen, P.; Kermode, J.; Kitchin, J. R.; Leonhard Kolsbjerg, E.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Bergmann Maronsson, J.; 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. *J. Phys.: Condens. Matter* **2017**, *29* (27), 273002. https://doi.org/10.1088/1361-648X/aa680e.

(68) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (Pymatgen): A Robust, Open-Source Python Library for Materials Analysis. *Computational Materials Science* **2013**, *68*, 314–319. https://doi.org/10.1016/j.commatsci.2012.10.028.

(69) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J Comput Chem* **2009**, *30* (13), 2157–2164. https://doi.org/10.1002/jcc.21224.

(70) Cromer, D. T.; Herrington, K. The Structures of Anatase and Rutile. *J. Am. Chem. Soc.* **1955**, *77* (18), 4708–4709. https://doi.org/10.1021/ja01623a004.

(71) Zhu, S.-C.; Xie, S.-H.; Liu, Z.-P. Nature of Rutile Nuclei in Anatase-to-Rutile Phase Transition. *J. Am. Chem. Soc.* **2015**, *137* (35), 11532–11539. https://doi.org/10.1021/jacs.5b07734.

(72) Yang, X.; Shang, C.; Liu, Z.-P. Resolving the Atomic Structure of γ-Alumina: A Non-Spinel Phase with a Distorted Anion Lattice and Three Adjacent Long Channels. *J. Mater. Chem. A* **2025**, *13* (23), 17429–17433. https://doi.org/10.1039/D5TA01715G.

(73) Morterra, C. Infrared Study of the Adsorption of Pyridine on $alpha;-Al2O3. *Journal of Catalysis* **1978**, *54* (3), 348–364. https://doi.org/10.1016/0021-9517(78)90083-0.

(74) Kline, C. H., Jr.; Turkevich, J. The Vibrational Spectrum of Pyridine and the Thermodynamic Properties of Pyridine Vapors. *J. Chem. Phys.* **1944**, *12* (7), 300–309. https://doi.org/10.1063/1.1723943.

(75) Phung, T. K.; Herrera, C.; Larrubia, M. Á.; García-Diéguez, M.; Finocchio, E.; Alemany, L. J.; Busca, G. Surface and Catalytic Properties of Some γ-Al2O3 Powders. *Applied Catalysis A: General* **2014**, *483*, 41–51. https://doi.org/10.1016/j.apcata.2014.06.020.

Figure 1

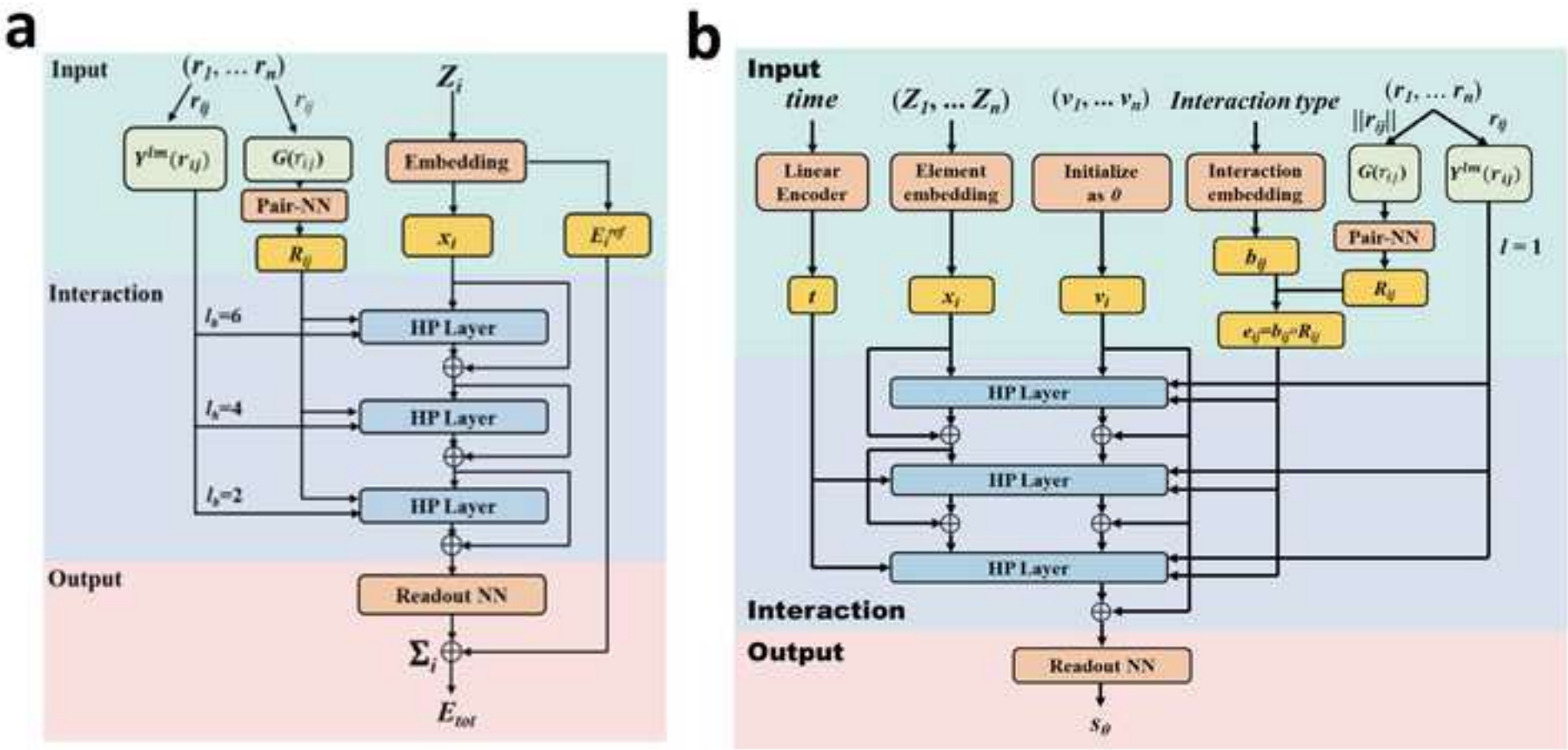


Figure 2

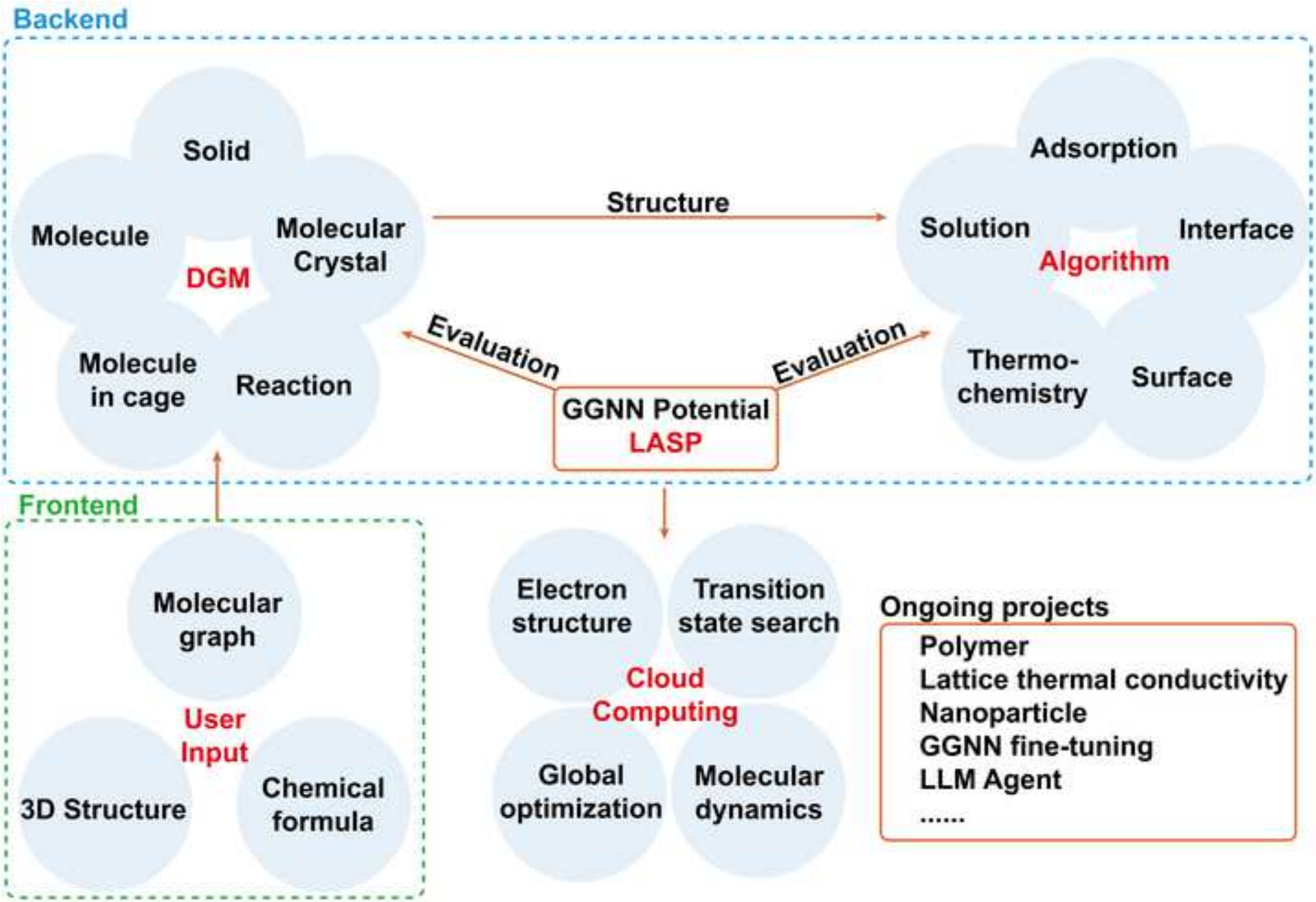
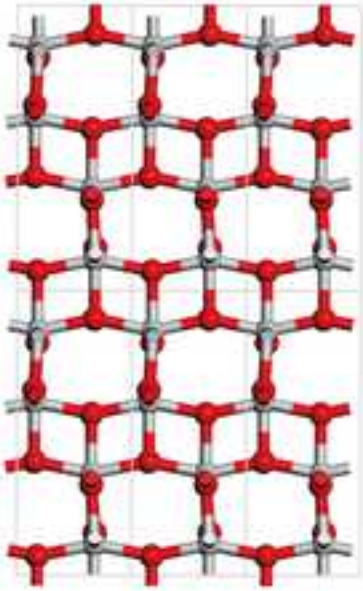


Figure 3

a

Unit Energy eV/f.u.	Chemical Formula	Energy eV	Space Group	Volume Å³	Unit Volume Å³/f.u.
-27.0064	O4-Ti2	-54.0128	I4_1/am...	69.6363	34.8181
-26.9532	O4-Ti2	-53.9064	P4_2/m...	63.8473	31.9236
-26.8459	O10-Ti5	-134.2296	C2 (5)	174.7062	34.9412
-26.7925	O8-Ti4	-107.1700	C2/m (12)	141.2947	35.3237
-26.7815	O10-Ti5	-133.9077	P1 (1)	180.2376	36.0475
-26.7795	O6-Ti3	-80.3385	P-31m (...	93.5795	31.1932
-26.7790	O8-Ti4	-107.1159	C2/m (12)	133.8940	33.4735
-26.7482	O4-Ti2	-53.4964	P2_1/m ...	70.2028	35.1014
-26.6993	O6-Ti3	-80.0980	Cm (8)	131.5900	43.8633

b



c

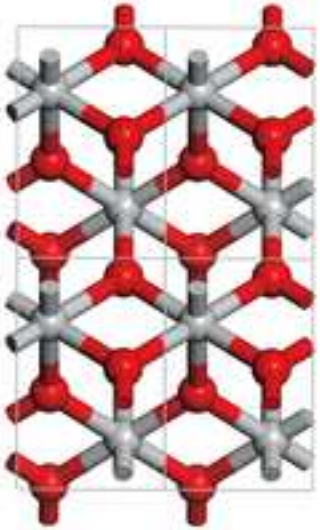


Figure 4

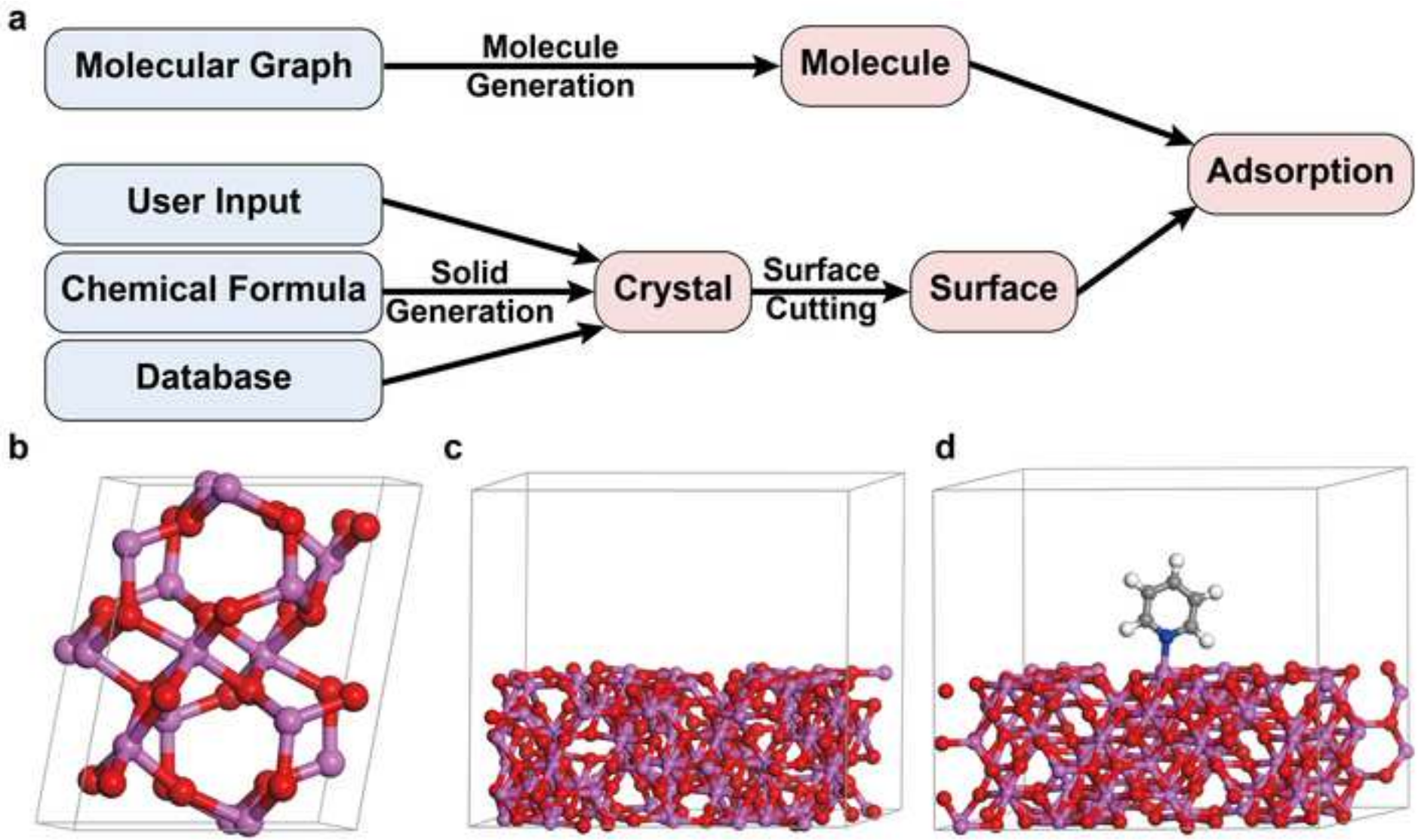


Figure 5

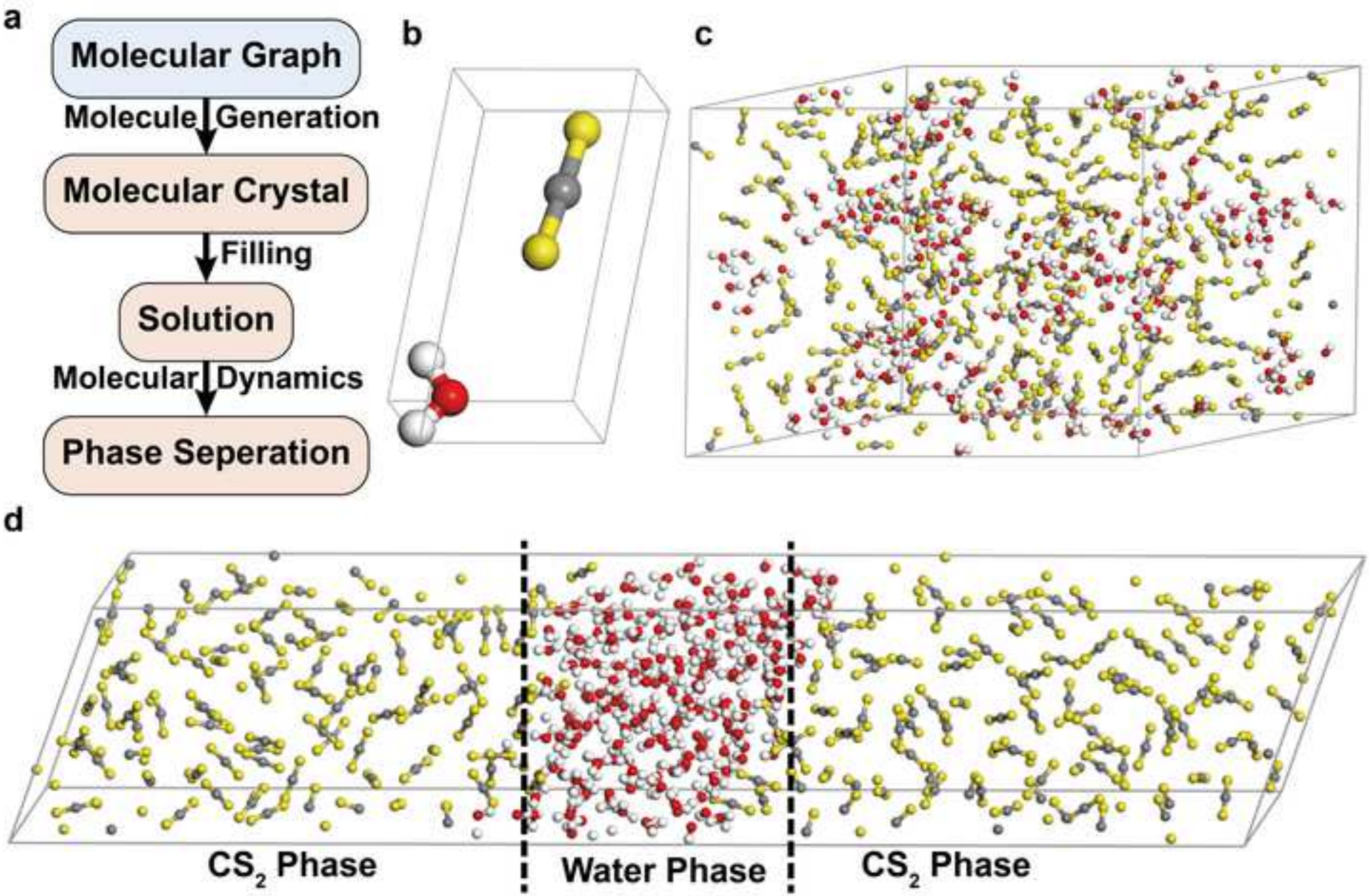


Figure 1. The HPNN architecture for (a) PES evaluation (generalized global neural network) and (b) structure generation (diffusion generative models). The figures are extracted from works by Yang et al.[31] and Guo et al. [46] respectively.

Figure 2. The architecture and user workflow of LASPAI website. The blue and green dotted box shows the architecture of the backend and frontend, respectively.

Figure 3. The generation of different polymorphs of TiO_2 using the solid generation page. (a) The list of generated TiO_2 structures sorted by unit energy. (b) and (c) are the first and second entries in (a), corresponding to the generated and optimized structures of anatase and rutile, respectively. The red and grey spheres represent O and Ti atoms, respectively.

Figure 4. The adsorption of pyridine on γ-Al_2O_3. (a) The workflow of adsorption study in LASPAI. (b) The minimum γ-Al_2O_3 (γ-AD) structure found by Yang et al.[72] (c) The (512) facet of γ-AD structure, corresponding to (110) facet of γ-Al_2O_3. (d) The generated adsorption structure of pyridine on Al_2O_3. The white, grey, blue, red and pink spheres represent H, C, N, O and Al atoms, respectively.

Figure 5. Atomic simulation of oil/water phase separation. (a) The workflow for performing phase separation simulation on LASPAI. (b) The generated 1:1 molecular crystal of water and CS_2. (c) The 1:1 mixed starting solution structure. (d) The structure after MD simulation with the two phases separated. The white, grey, red and yellow spheres represent H, C, O and S atoms, respectively.