

AI-Driven Discovery and Molecular Engineering Design for Enhancing Interface Stability of Black Phosphorus

Chao Peng⁺,* Bing Wang⁺, Lie Wu⁺, Haoqu Jin, Yutang Li, Wenxia Gao, Jie Zhou, Guolai Jiang, Chen Wang, Jiahong Wang,* Xingchen He, Denis Kramer, Paul K. Chu, and Xue-Feng Yu*

Abstract: Molecular engineering offers significant potential for developing advanced interfacial materials, yet the complexity of organic molecules poses challenges in discovering optimal structures. This study leveraged large language model (LLM) and machine learning (ML) to accelerate molecular discovery and guide molecular engineering for enhancing the stability of black phosphorus (BP), a promising 2D semiconductor but rapidly degrades when exposed to oxygen and moisture. By utilizing GPT-4o, molecular groups such as $-\text{SiR}_3$, $-\text{PR}_2$, $-\text{SH}$, and $=\text{NH}$ that interact effectively with BP were identified and a high-throughput workflow employing graph neural networks (GNNs) models was developed to successfully predict and screen 662 promising candidates from over 117 million molecules. These candidates were validated by density functional theory (DFT) simulations and experiments, with synthesis protocols guided by GPT-4o, achieving great interfacial stabilization of BP for up to 24 days under ambient conditions. Furthermore, a new synergistic molecular engineering strategy was proposed by incorporating functional head, linker, and tail groups of molecules to even enable the use of hydrophilic molecules to stabilize BP surface, overcoming traditional design limitations. This work highlights the AI technologies not only in optimizing BP interfacial stability but also in broader aspects of molecular engineering for various materials.

Introduction

Molecular engineering has gained significant attention due to the inherent advantages of organic molecules, such as their diverse functional groups, tunable structures, cost-effectiveness, and easy fabrication. These features render them ideal for designing advanced interfacial materials across solid–solid, solid–liquid, and solid–gas interfaces.^[1,2] However, the diversity and complexity of functional groups present a major challenge in how to systematically design optimal molecular structures to obtain the desired interfacial properties. In this respect, recent advancements in material informatics driven by artificial intelligence (AI) uncover a new paradigm to discover potential molecules and guide

their synthesis.^[3–5] For instance, machine learning (ML) and large language models (LLM) offer transformative tools to address this challenge.^[6,7] ML models have great potential in predicting molecular properties and guiding design by uncovering structure–property relationships, while LLMs, such as OpenAI's generative pre-trained transformer 4 (GPT-4), leverage vast scientific literature to provide guidance for molecular design and synthesis procedures.^[7] By taking advantage of the collective strengths of LLM and ML, the development of ML training datasets can be expedited according to insights suggested by LLM. All in all, the integration streamlines ML model training and enables high-throughput scientific discovery more efficiently.

[*] Dr. C. Peng⁺, Dr. B. Wang⁺, Dr. L. Wu⁺, H. Jin, Y. Li, W. Gao, J. Zhou, Dr. G. Jiang, C. Wang, Dr. J. Wang, Dr. X. He, Prof. X.-F. Yu
Materials Artificial Intelligence Center, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Avenue, Shenzhen 518055, P.R. China
E-mail: chao.peng@siat.ac.cn
jh.wang1@siat.ac.cn
xf.yu@siat.ac.cn

Prof. D. Kramer
Helmut-Schmidt-University, University of the Armed Forces,
Holstenhofweg 85 22043, Hamburg, Germany


Prof. P. K. Chu
Department of Physics, Department of Materials Science and Engineering, and Department of Biomedical Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong 999077, P.R. China

Dr. C. Peng⁺, W. Gao, Dr. G. Jiang, C. Wang, Dr. J. Wang, Dr. X. He, Prof. X.-F. Yu
University of Chinese Academy of Sciences, Beijing 100049, P.R. China

Dr. J. Wang, Prof. X.-F. Yu
Key Laboratory of Biomedical Imaging Science and System, Chinese Academy of Sciences, State Key Laboratory of Biomedical Imaging Science and System, 1068 Xueyuan Avenue, Shenzhen 518055, P.R. China

Dr. L. Wu⁺
State Key Laboratory of Advanced Refractories, Wuhan University of Science & Technology, 947 Heping Avenue, Wuhan 430081, P.R. China

[†] These authors contributed equally to this work.

 Additional supporting information can be found online in the Supporting Information section

Two-dimensional (2D) layered black phosphorus (BP), with its unique physicochemical and electronic properties such as a layer-dependent direct bandgap (0.3–2.0 eV), high carrier mobility, and strong anisotropy,^[8,9] has shown great promise in applications such as electronics, optoelectronics, and energy storage.^[10] However, BP suffers from intrinsic interface instability under ambient conditions^[11] due to oxidative degradation arising from the reactive lone pair electrons of phosphorus atoms, consequently limiting wider commercial practice.^[12] In recent years, various interfacial materials have been designed to enhance the stability of BP by surface passivation (e.g., covalent functionalization with alkyl iodides^[13] or noncovalent interactions like π - π stacking^[14] and van der Waals interactions.^[15–17] Despite these efforts, the development of stabilization strategies has largely relied on trial-and-error approaches, which lack systematic methodologies or guiding principles to determine the optimal molecular passivators. Similar challenges are encountered by other 2D nanomaterials such as perovskites,^[18] while interfacial stabilization remains inadequate and poorly understood.

In this work, LLMs with ML were integrated to accelerate molecular discovery and establish an intelligent framework for the molecular engineering design of BP interfacial stability. LLMs were employed to extract information from 286 papers and identify molecular groups with strong affinities to aid interfacial stabilization. Based on the knowledge, a dataset comprising linear self-assembled organic molecules was constructed from open-source PubChem chemical databases, and graph neural network (GNN) models were applied to establish the structure–stability relationship in order to choose the optimal molecular candidates. By means of this AI-assisted approach, we were able to rapidly identify a series of previously underexplored molecular passivators with excellent stabilization effects, and the methodology outperforms traditional trial-and-error techniques substantially. Furthermore, LLM provides guidance on synthesis procedures based on knowledge from 285 papers, enabling experimental validation of the top three candidates. Using BP as an example, the synergistic interfacial strategy was demonstrated to overcome the traditional design limitation. For example, it provides guidance to enhance the BP stability using molecules with hydrophilic tails by coordinated design of head, linker, and tail groups and challenges conventional perspectives with detrimental effects of hydrophilic molecules on BP interface stability. Overall, this framework offers a scalable and generalizable solution for the identification of the optimal interfacial materials beyond the specific case of BP.

Results and Discussion

Recommendation Schemes for Organic Functional Head Groups from LLM

The design and fabrication of the ideal molecular protection layer on BP by molecular engineering needs to satisfy several key criteria. First, the head groups of organic molecules should bind stably to the BP surface to form densely packed molecular layers. Secondly, the tail groups of organic

molecules should effectively prevent water molecules and oxygen from penetrating and degrading the BP surface. Thus, it is important to identify and screen head groups that exhibit strong binding affinity with the phosphorus (P) sites on the BP surface. Herein, we fine-tuned an Open AI's ChatGPT model (see details in Method) and implemented recommendation schemes for identifying the suitable head groups with binding capabilities with BP based on the knowledge obtained from 286 papers in the literature (Figure 1). The GPT-4o model recommended several types of organic head groups, including $-\text{SiR}_3$, $-\text{PR}_2$, $-\text{SH}$, and $=\text{NH}$. Among these, $-\text{SiR}_3$, $-\text{PR}_2$, and $-\text{SH}$ head groups have been rarely reported in the context of BP stability, unlike the frequently studied and experimentally validated $=\text{NH}$ group, which is known for its strong performance. In addition, GPT-4o also offered a detailed explanation of the potential passivation mechanisms of the selected head groups, thereby providing deeper insights into how these groups stabilize BP surfaces (Figure S1).

GNN Machine-Learning Model and High-Throughput Prediction Workflow

The desirable molecule candidate for BP surface modification is expected to have a strong binding affinity to the BP surface while shielding the surface from interactions with water and oxygen. Therefore, the screening process was developed precisely based on three critical properties: (1) the interaction energy between one terminal group of the molecule and the BP surface ($E_{\text{int_BP}}$), and (2) the interaction energies between the opposite terminal group and either water ($E_{\text{int_H}_2\text{O}}$) or oxygen ($E_{\text{int_O}_2}$). These interaction energies serve as key indicators of the ability to form stable and protective molecular layers on the BP surface and minimize water- and oxygen-induced degradation at the molecular level. The suitability of these molecules can be validated theoretically by DFT calculations. However, this approach is time-intensive and impractical for large datasets. Therefore, a workflow integrating ML methods was implemented to accomplish high-throughput screening for promising candidates with the desired properties as shown in Figure 2. We aim to establish a candidate database comprising molecules with excellent passivation properties to pave the way for efficient and scalable molecular design for BP surface protection.

Prior to ML model training, it is necessary to construct an initial dataset. The initial dataset was comprised by two key steps (Figure 2a): (1) construction of a molecular library with diverse heading groups and (2) establishing a representative dataset containing molecular structures and their corresponding target properties ($E_{\text{int_BP}}$, $E_{\text{int_H}_2\text{O}}$, and $E_{\text{int_O}_2}$) by performing DFT calculations. The molecular library was constructed through a systematic combination of three types of structural blocks: head blocks, linker blocks, and tail blocks (Figure 2a). The head blocks were designed and selected from heteroatom-containing functional groups recommended by GPT-4o for their capacity to interact with the BP surface, such as $\text{C}=\text{NH}$, CHO , SH , PH_2 , and $\text{Si}(\text{OH})_3$. These groups are expected to adsorb directly onto the BP surface and are considered primary contributors to

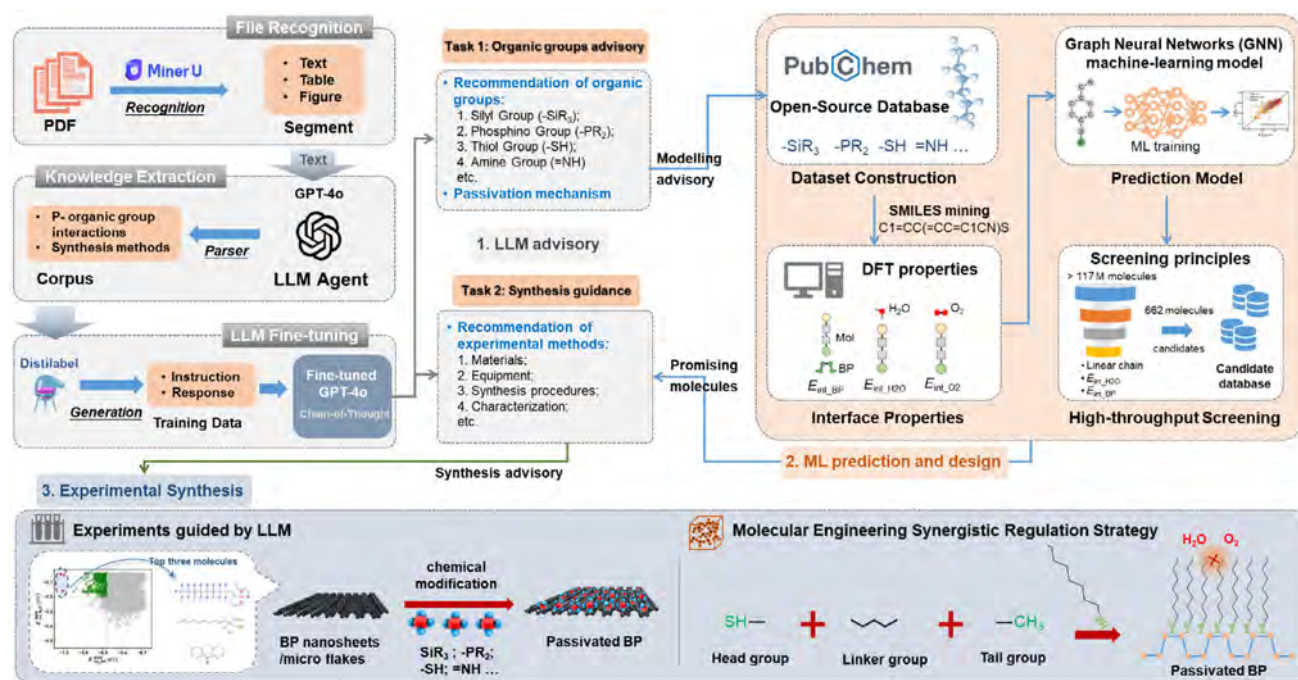


Figure 1. Schematic of intelligent molecule design for enhancing interface stability of BP via data-driven molecular engineering.

the molecule-BP interaction energy ($E_{\text{int_BP}}$). The proper linear organic molecules with those head blocks from the open-source PubChem database were screened manually to construct the molecular library. The tail blocks, located at the terminal end of the molecule, are supposed to be exposed to the ambient environment. Both polar (COOH, NH₂) and nonpolar (CH₃) functional groups were included to cover a broad range of interaction energies with water and oxygen. Linker blocks, such as CH₂, phenyl, and NH moieties, function as linker units between head and tail blocks. This is essential to yield linear-like compounds. The systematical combination of one head block, several linker blocks, and one tail block results in a library of approximately 1.7k unique molecules. This approach ensures both chemical diversity and structural relevance for BP surface modification.

Subsequently, DFT calculations were performed for each molecule in the library to determine the molecule-BP interaction energy ($E_{\text{int_BP}}$), molecule-water interaction energy ($E_{\text{int_H}_2\text{O}}$), and molecule-oxygen interaction energy ($E_{\text{int_O}_2}$). These values serve as target properties of the dataset, with the corresponding distributions illustrated in Figures S2 and S3. Significant variations were observed in $E_{\text{int_BP}}$ and $E_{\text{int_H}_2\text{O}}$ among different molecules. Specifically, the $E_{\text{int_BP}}$ values range from -1.28 to -0.09 eV, and $E_{\text{int_H}_2\text{O}}$ from -0.60 to 0.00 eV. Analysis of the molecules based on the heteroatoms in the head groups reveals an overall trend in interaction strength with BP, ranked as $\text{P} > \text{S} \approx \text{Si} > \text{N} > \text{O}$. Molecules with polar tail groups, such as COOH, exhibit stronger molecule-water interactions compared to those with nonpolar tail groups like CH₃, in agreement with the established chemical knowledge. Meanwhile, almost all the $E_{\text{int_O}_2}$ values are higher than -0.20 eV, indicating that all the polar and nonpolar tail groups have weak interactions with oxygen. Therefore, the molecule-oxygen interaction is no longer considered as its impact is

negligible. The subsequent stages primarily focus on the other two key interaction properties: $E_{\text{int_BP}}$ and $E_{\text{int_H}_2\text{O}}$.

To enable high-throughput predictions of molecular interaction energies, the machine learning models were trained based on two established graph neural network (GNN) architectures for $E_{\text{int_BP}}$ and $E_{\text{int_H}_2\text{O}}$ prediction tasks (Figure 2a). With regard to $E_{\text{int_BP}}$ prediction, the Graph-based Adsorption on Metal Energy-neural Network (GAME-NET) architecture was employed for its demonstrated success in predicting molecular adsorption energies, due to the incorporation of surface topology information intrinsically.^[19] As for $E_{\text{int_H}_2\text{O}}$ prediction, the directed message passing neural network (D-MPNN) architecture implemented in Chemprop was adopted^[20,21] on account of the proven performance in predicting tasks for a variety of molecular properties. The molecular structures formulated by Simplified Molecular Input Line Entry System (SMILES) strings in the initial dataset were encoded as graphs, with atoms represented as vertices and bonds as edges. The GNN models learn molecular representations through several graph convolution layers, where atomic features are propagated and aggregated across the molecular graph, and capture both local chemical environments and global structural information. We trained both models using an 8:1:1 split for training, validation, and test sets. The trained models exhibited excellent predictive accuracy on the test set (Figure 2b,c), as manifested by root mean squared errors (RMSEs) as small as 0.115 and 0.048 eV for $E_{\text{int_BP}}$ and $E_{\text{int_H}_2\text{O}}$ predictions, respectively. The small prediction errors compared to DFT-calculated values corroborate the reliability of the trained GNN models in high-throughput virtual screening of molecular candidates.

By utilizing the validated models, high-throughput virtual screening was conducted on a vastly expanded molecule library from the PubChem database (Figure 2d). We

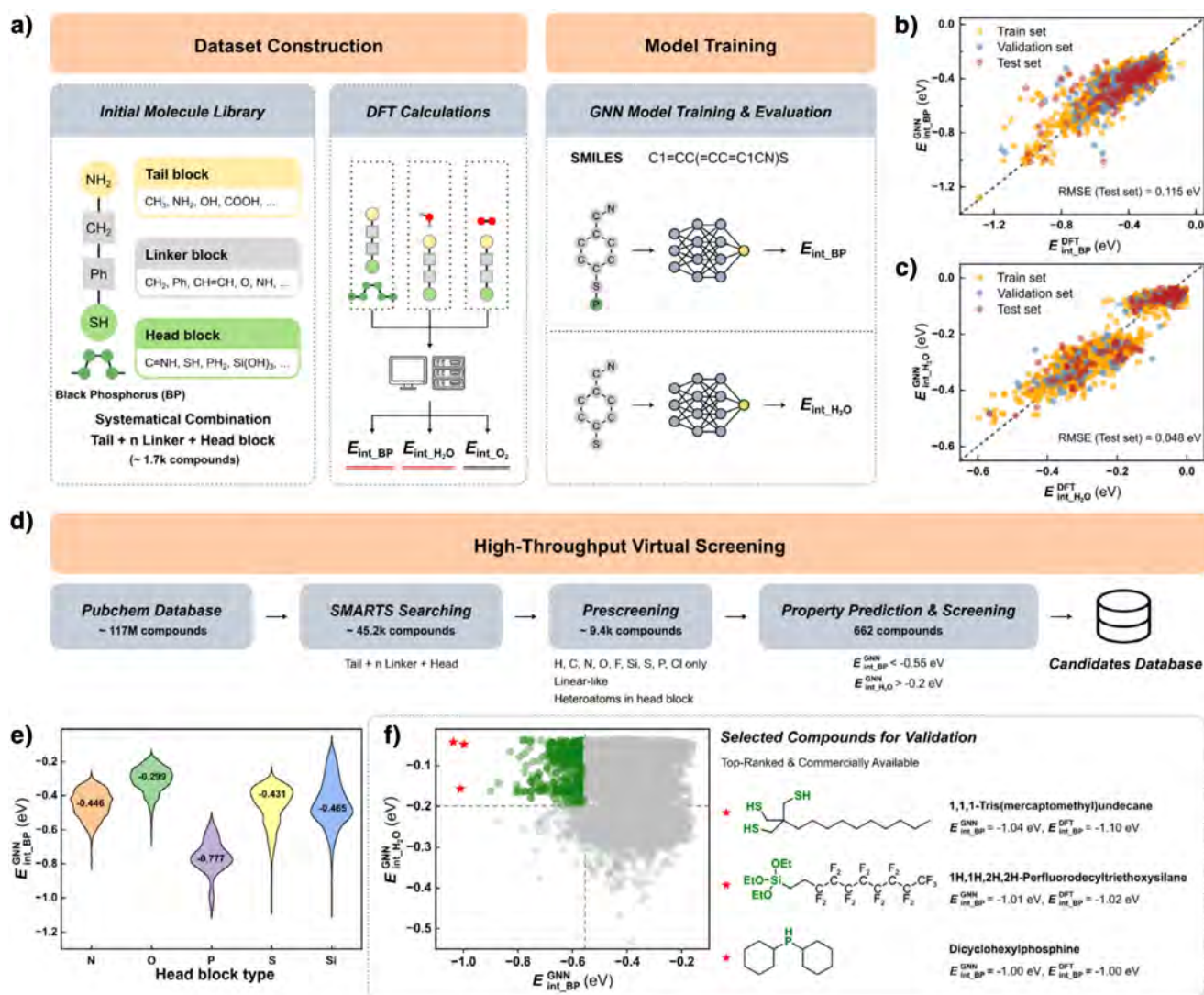


Figure 2. Overview of the machine learning and high-throughput virtual screening process. a) Workflow for machine learning, including the construction of a representative dataset and the GNN model training process. The dataset was built from a systematically generated molecular library comprising ~1.7k compounds, with target properties ($E_{\text{int_BP}}$, $E_{\text{int_H}_2\text{O}}$, and $E_{\text{int_O}_2}$) calculated by DFT simulations. b) and c) GNN predictions versus calculated DFT values for $E_{\text{int_BP}}$ and $E_{\text{int_H}_2\text{O}}$ and root mean square error (RMSE) for $E_{\text{int_BP}}$ and $E_{\text{int_H}_2\text{O}}$, respectively. d) Workflow for high-throughput virtual screening. e) Distribution of predicted $E_{\text{int_BP}}$ values grouped by molecular head groups. Median values of each group are reported. f) Scatter plot of predicted $E_{\text{int_BP}}$ and $E_{\text{int_H}_2\text{O}}$ values for individual molecules, with promising candidates indicated by green dots. Top-ranked, commercially available candidates (marked with red stars) were selected for further evaluation, and the corresponding structures, along with GNN-predicted and DFT-calculated $E_{\text{int_BP}}$ values are presented.

employed SMILES Arbitrary Target Specification (SMARTS, see details in Supporting Information) searching to perform substructure matching based on the PubChem database, which contains approximately 117 million organic molecules with diverse chemical and structural properties. As shown in Figure 2a, SMARTS strings were generated by combining the head block, the linker block, and the tail block SMARTS, similar to the previous process. Additional constraints on heavy atom counts and complexity were used to narrow down the search space. This step quickly identified approximately 45.2k molecules with specific structural patterns from the PubChem database. To further refine the candidate set, a series of pre-screening steps were performed. Molecules containing elements other than H, C, N, O, F, Si, P, S, and Cl

were excluded, as well as highly branched molecules, which are less likely to form a continuous and protective layer on the BP surface due to considerable steric hindrance. Moreover, molecules lacking heteroatoms at the terminal groups were also removed because the presence of heteroatoms like Si, P, and S is crucial for the head groups of molecules binding on the BP surface based on our hypothesis and GPT-4o recommendation. These filtering steps produce a refined set with approximately 9.4k molecules, as summarized in Figure 2d.

After the pre-screening step, the trained GNN models were employed to predict $E_{\text{int_BP}}$ and $E_{\text{int_H}_2\text{O}}$ for each molecule in the refined set. This allows us to evaluate the binding affinity of molecules with the BP surface and water and provides quantitative information about their potential

as surface modifiers. As shown in Figure 2e, molecules containing P-, S-, Si- and N-based head groups exhibit stronger molecule–BP surface interactions compared to those with O-containing head groups, consistent with our initial observation in the model training stage. Furthermore, molecules containing P-, S-, and Si-based head groups demonstrated remarkably strong binding capabilities on the BP surface, as shown by predicted $E_{\text{int_BP}}$ values lower than -0.75 eV. To select the most promising candidates, we implemented a dual-criteria filtering approach: $E_{\text{int_BP}} < -0.55$ eV to ensure effective surface binding as well as $E_{\text{int_H}_2\text{O}} > -0.2$ eV to minimize water interaction (Figure 2f). This screening yielded a final dataset of 662 distinct compounds that are expected to be promising candidates to enhance the interface stability of BP in ambient air. Those promising candidates listed with their SMILES, names and predicted energy values were compiled into a dataset as summarized in Data S1.

To validate the reliability of the screening methodology, we selected three top-ranked and commercially available compounds for further investigation (marked with stars in Figure 2f). DFT calculations show that their calculated molecule–BP interaction energies agree well with our GNN predictions, thus providing preliminary validation of the screening approach's accuracy and reliability. Furthermore, those three molecules were purchased commercially and undergo experimental validation.

Experimental Validation and Characterization

To validate the passivation effects of the top-ranked three molecules (P_1 , Si_1 , and S_1) predicted by our graph neural network (GNN) models, experiments were carried out. For a comprehensive comparison, a series of molecules from lower-ranked positions in the 9.4k database in the prescreening stage were also selected, with the focus on molecules featuring N-, S-, Si-, and P-based head groups. The synthesis procedures for coating these molecules onto BP nanosheets were formulated based on the developed GPT-4o recommendations, as illustrated in Figures S4–S6. By adopting the synthesis protocol recommended by GPT-4o, which involves liquid phase exfoliation and modification methods, we successfully prepared and modified BP nanosheets. In brief, bulk BP crystals were initially ground with *N*-methyl pyrrolidone (NMP) for 30 min to create a coarse dispersion. This was followed by 10 h of bath sonication to produce the BP nanosheets (~ 300 nm in size). Subsequently, an excess of the predicted molecules was added. The mixture was stirred for 20 h in darkness under argon to obtain the modified BP samples (denoted as P_1 -BPs, Si_1 -BPs, S_1 -BPs, etc., see Figure 3a).

To assess the passivation effects on BP surfaces, dispersions of P_1 -BPs, S_1 -BPs, Si_1 -BPs, N_1 -BPs, N_2 -BPs, and bare BPs with a concentration of 38 ppm were exposed to air. As depicted in the white-light photos in Figure 3b, the BPs dispersions containing the top-ranked molecules (P_1 -BPs, Si_1 -BPs, and S_1 -BPs) retained their brown color after 20 days, whereas the bare BPs became transparent within just 12 days. This observation was also verified by the absorption spectra shown in Figure S7, which indicated a similar trend. Notably,

the top-ranked new molecules P_1 -BPs, Si_1 -BPs, and S_1 -BPs exhibited markedly improved stability over N_1 -BPs and N_2 -BPs. This stability enhancement underscores the reliability of our ML prediction models. Furthermore, the superior performance of molecules containing newly identified P-, S-, or Si-based head groups, compared to traditionally reported N-based head groups, highlights the generalization capability of our ML model in discovering new effective molecules. Figure 3c illustrates molecules with N-, P-, S-, and Si-based head groups, where it was observed that predicted P-, S-, and Si-based molecules significantly improved the BP stability compared to N-based alternatives (see structures in Figure S8). P_1 -BPs, Si_1 -BPs, and S_1 -BPs maintained stability for over 20 days, demonstrating that the selected organic molecules provide enhanced ambient stability and effectiveness. As highlighted in Table 1, previous studies have made notable advances through both covalent (e.g., alkyl iodide functionalization) and noncovalent strategies (e.g., π – π stacking, van der Waals interactions). However, due to the vastness of chemical structural space and the multidimensional nature of passivation performance, it is exceedingly challenging for human chemists to rationally design and screen such a broad array of candidate molecules in a high-throughput manner. Our AI-driven strategy, combining LLM and ML, enables rapid molecular engineering of the BP interface beyond what can currently be achieved through manual or intuition-based methods. Importantly, molecules such as P_1 , Si_1 , and S_1 , identified through our automated pipeline, exhibited excellent ambient stability exceeding 20 days, outperforming most previously reported organic passivators. These results highlight the power of our workflow not only in accelerating discovery but in uncovering novel molecular structures.

This surface modification strategy was extended to BPs microflakes (BPs MFs). The BP MFs were mechanically exfoliated from the bulk crystal using scotch tape and transferred to a Si/SiO₂ substrate for passivation. The organic molecule passivation method was performed as described earlier. As shown in Figure 3d, the lateral dimension of the flakes is about 10–20 μm . After exposure for 12 days, the atomic force microscopy (AFM) images of bare BP flake, N_1 -BPs MFs, and N_2 -BPs MFs show dense droplets and become larger at 20 days (Figure 3c). In contrast, the surfaces of P_1 -BPs MFs, S_1 -BPs MFs, and Si_1 -BPs MFs on the substrate are unchanged. These results suggest that the organic molecule passivation method can be employed to protect different BPs structures from oxidation and degradation. To verify the formed morphology of molecular layers on the BPs surfaces, high-resolution transmission electron microscopy (HR-TEM) was carried out. As depicted in Figure 3e, the HR-TEM images of Si_1 -BPs illustrate uniform ~ 3.7 nm organic layers adhered to the BPs surface. HR-TEM also shows lattice fringes of 2.3 Å corresponding to the (014) plane of BPs, confirming that surface modification does not compromise the structural integrity of BPs. Furthermore, the ¹H nuclear magnetic resonance (¹H NMR) spectra for N_2 -BPs, P_1 -BPs, Si_1 -BPs, and S_1 -BPs show signals corresponding to the modified molecules, confirming their effective anchoring to the BP surface (Figure S9). In comparison with N_1 -BPs, the Fourier transform infrared spectroscopy (FTIR) of

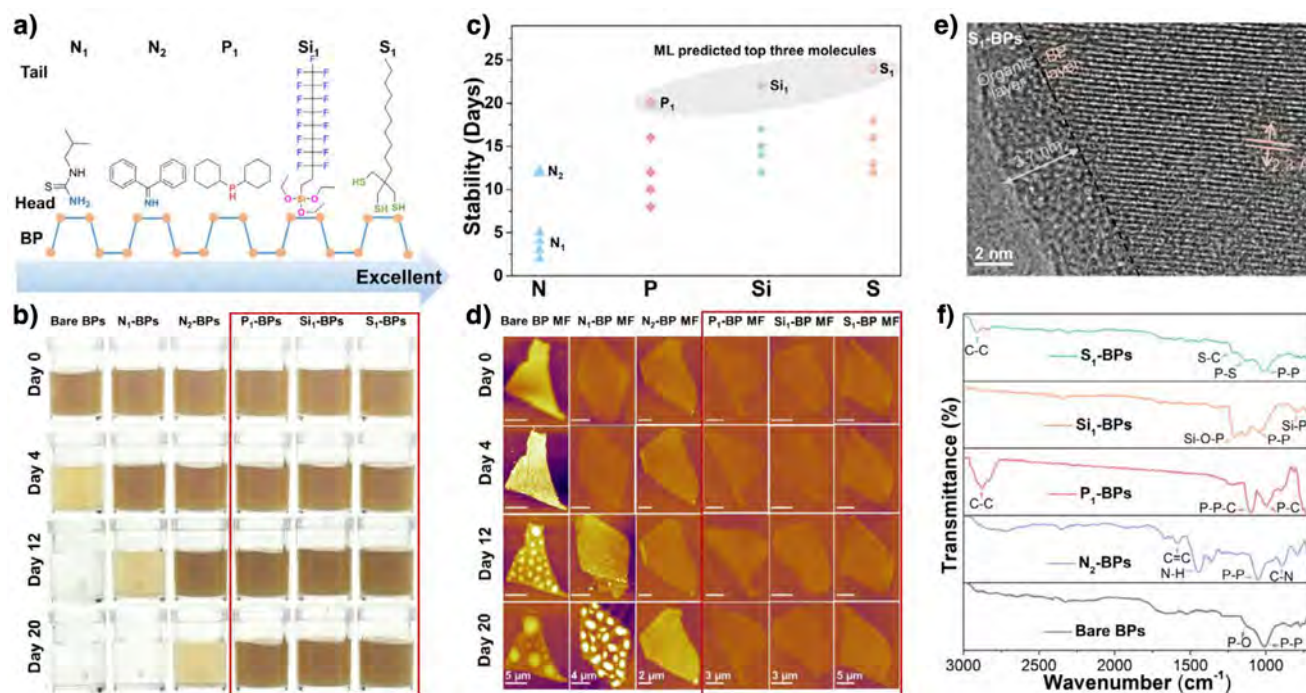


Figure 3. Experimental validation of the passivation effect of predicted molecules on the BP surface. a) Schematic illustration of the molecular passivation strategy for stabilizing the BP surfaces. b) White-light photos of bare BPs and BPs passivated with organic molecules after storing in water for different time durations. c) Stable days for BPs passivated with molecules containing different head groups. d) AFM photos of bare BP MF and BP MFs passivated with organic molecules exposed to air for 0, 4, 12, and 20 days. e) HR-TEM image of S_1 -BPs. f) FT-IR spectra of bare BPs and molecularly passivated BPs.

Table 1: Comparison of the ambient stability of molecular engineering with previously reported organic molecule passivation techniques.

Passivation methods	Organic molecules	Ambient stability	Refs.
Noncovalent passivation	Organic solvent	7–10 days	[22,23]
Noncovalent passivation	Phospholipids	>16 days	[24]
Noncovalent passivation	Perylene diimides	>8 days	[25]
Noncovalent passivation	Surfactants	21–30 days	[26]
Noncovalent passivation	Tripeptide	>4 days	[27]
Covalent passivation	Aryl diazonium salts	3–21 days	[10,28]
Covalent passivation	Alkali metal hydrides	18 days	[29]
Coordination passivation	Titanium sulfonate	>7 days	[30]
Coordination passivation	Lanthanide sulfonate	>8 days	[31]
Molecular engineering	–P, –S, –Si groups	>20 days	This work

P_1 -BPs, Si_1 -BPs, and S_1 -BPs reveal peaks at 1106, 1208, and 1126 cm^{-1} , corresponding to P–P–C, Si–O–P, and P–S bonds respectively, suggesting strong interactions between molecules with P, S, and Si groups and BPs (Figure 3f). These experimental findings confirm that the P-, S-, and Si-based molecules predicted by our ML models can significantly enhance the BP stability.

Synergistic Design Strategies for Interfacial Molecular Engineering

We are now at a stage to understand how molecular engineering enhances the interface stability of BP. As shown in Figure 4a, with an increasing number of S-based head

groups in the packed molecules, the binding energy between the molecule and BP(010) surface increases from -0.86 to 1.10 eV. A similar trend was observed from the Si-based head group, as increasing the number of $-Si$ groups leads to stronger binding. This is attributed to the strong interactions between the $-S$ and $-Si$ head groups and BP surface. The protective effects of these groups on the BP surface are quantified by monitoring the stability durations in the experiments. Figure 4b illustrates that as the number of $-S$ groups increase from 1 to 3, the stability duration rises from 13 to 24 days. Similarly, for the $-Si$ head group, increasing the number of $-Si$ groups improves the stability durations from 12 to 18 days. This suggests that enhancing the binding between head groups and the BP surface effectively improves the passivation effect. Moreover, maintaining the same number of

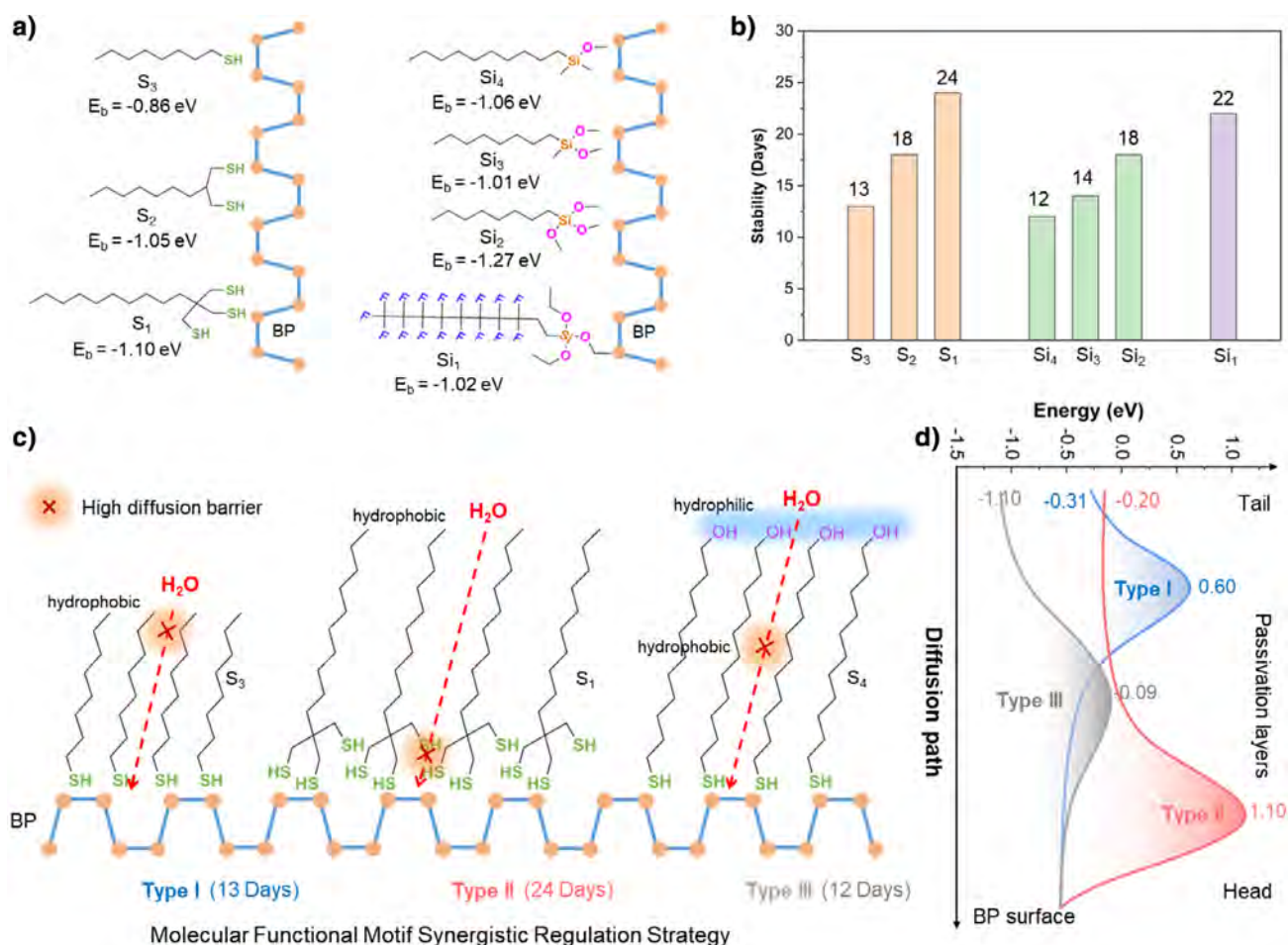


Figure 4. Molecular engineering strategies to enhance the interface stability of BP. a) Adsorption of various packing molecules with $-\text{S}$ and $-\text{Si}$ head groups on the surface of BP. b) Stability durations (in days) of the molecules shown in (a) demonstrating their effectiveness in passivating the BP surface. c) Boosting the ambient stability of BP via a synergistic regulation strategy, achieved by combining functions from the head, linker, and tail groups of the molecules. d) Diffusion barriers for a water molecule diffusing from the tail groups of the molecular passivation layers with different $-\text{S}$ head groups down to the BP(010) surface as indicated by the red dashed arrow in (c).

$-\text{Si}$ head groups but replacing the tail and linker groups with hydrophobic groups such as $-\text{F}$ enhances the stability to 22 days. This finding reveals that reinforcing the binding of head groups on the BP surface, in combination with integrating hydrophobic groups at the tail and linker positions, can synergistically improve the passivation duration for BP. This also demonstrates that a synergistic approach involving multiple molecular group interactions is quite efficient via molecular engineering. This synergistic molecular engineering strategy is further supported by the model presented in Figure 4c,d. Specifically, we examined three types of $-\text{S}$ based molecules, differing in the number of $-\text{S}$ head groups and tail groups including hydrophobic $-\text{CH}_3$ and hydrophilic $-\text{OH}$ (referred to as S₁, S₃, and S₄). A water molecule was used as an indicator to assess the passivation effect of these molecules on the BP surface. Comparing SH-based and (SH)₃-based molecules, aside from the expected influence of varying $-\text{S}$ group numbers on the surface binding energies (where the (SH)₃-based molecule provides better protection), the diffusion barrier (0.91 eV) for H₂O approaching the BP surface is initially raised by the dense-packed molecular layers

and hydrophobic $-\text{CH}_3$ and $-\text{CH}_2$ groups in the tail and linker segments in the S₃ molecule. However, for the S₁ molecule, the high diffusion barrier for H₂O passage is not attributed to the tail and linker groups. Instead, owing to steric hindrance from large head groups, the intermolecular channel is widened to allow easier H₂O passage, as reflected by a lower potential energy surface along the diffusion path of tail and linker blocks in Figure 4d (Type III). The high energy barrier (1.30 eV) arises at the location of the $-(\text{SH})_3$ head group due to the dense packing of $-\text{SH}$ on the BP surface and significant repulsion between $-\text{SH}$ and H₂O.

Notably, substituting the hydrophobic $-\text{CH}_3$ with a hydrophilic $-\text{OH}$ group in S₄ is typically detrimental to BP passivation, as the hydrophilic nature of $-\text{OH}$ facilitates water access to the surface.^[32,33] However, S₄ shows stable passivation effects on BP for about 12 days, similar to the M1 molecule. This is due to the synergistic function of the organic groups in S₄. While the hydrophilic $-\text{OH}$ group attracts H₂O, as confirmed by a lower adsorption energy of -1.10 eV compared to -0.31 and -0.20 eV for H₂O accessing the hydrophobic $-\text{CH}_3$ of S₁ and S₃ in Figure 4d,

the presence of considerable hydrophobic $-\text{CH}_2$ groups in the linker block results in a higher diffusion barrier (1.01 eV), which impedes H_2O to reach the BP surface. Therefore, S_4 still offers a comparable passivation effect on the S_3 molecule by combining tail and linker blocks. According to the above analysis, it is evident that the molecular functional motif design can synergistically regulate the properties of molecular layers and shed light on broader molecular design strategies applicable to other systems.

Implications of AI-Driven Discovery and Molecular Engineering Design for BP Interface Stability

It is worth noting that recent advances in LLM- and ML-driven chemical intelligence have demonstrated transformative capabilities for molecule discovery, synthesis planning, and automated experimentation. For instance, Gensch et al. introduced *kraken*, a ligand discovery platform incorporating DFT-derived descriptors and ML predictions to screen over 300 000 organophosphorus(III) ligands for homogeneous catalysis.^[34] Gomes et al. developed Coscientist, a GPT-4-driven architecture capable of designing, optimizing, and executing complex chemical experiments via integration of tools such as document search, code execution, and automated lab facilities, enabling closed-loop reaction optimization and cross-coupling catalysis.^[35] Likewise, ChemCrow, developed by Bran et al., showed that an LLM agent augmented by 18 expert-designed tools can autonomously plan and execute organic syntheses, with tasks ranging from organocatalyst discovery to chromophore design.^[36] These works highlight a paradigm shift toward AI-enabled closed-loop scientific discovery, with LLM evolving from pure text generators to powerful agents of scientific reasoning and physical experimentation.

Building upon these capabilities, this work specifically targets the interfacial engineering of 2D materials, with a focus on addressing the critical challenge of BP instability under ambient conditions. Notably, we incorporate LLM not as standalone agents for autonomous chemistry but as an expert assistant and an integrated knowledge engine that supports expert-designed workflows throughout the research pipeline. GPT-4o was used to systematically analyze hundreds of scientific literature to recommend synthesis protocols and mine underexplored headgroups (e.g., $-\text{SiR}_3$, $-\text{PR}_2$, $-\text{SH}$) that had been overlooked in prior BP passivation literature, which directly guided the data construction and downstream machine learning model training. Using GNN prediction models trained on DFT data, we screened over 117 million PubChem molecules and identified 662 candidates, three of which achieved over 20-day ambient stabilization of BP. More importantly, a synergistic molecular engineering strategy was introduced where molecular head, linker, and tail groups are co-engineered to control both surface adsorption and moisture exclusion, enabling the counterintuitive use of hydrophilic molecules for BP stabilization. Thus, this work not only introduces a new class of molecular stabilizers but also bridges LLM-guided hypothesis generation, ML-driven screening, and experimental validation, offer-

ing a scalable template for interfacial engineering beyond BP.

Finally, there are several potential directions to extend this framework. One is to integrate the molecular design loop into a closed-loop AI-material discovery system, wherein candidate predictions are iteratively updated based on experimental feedback. Another is to expand the framework beyond BP and apply the molecular engineering paradigm to other air-sensitive 2D materials such as perovskites and MXenes. In addition, coupling LLM reasoning with autonomous synthesis platforms could enable real-time exploration of interfacial materials under dynamic environments, thus advancing toward truly intelligent interface design and stabilization.

Conclusion

A knowledge- and data-driven framework was developed that integrates LLM-based literature analysis, DFT calculations, GNN prediction, and experimental validation to enable high-throughput screening of optimal passivating molecules. The proposed synergistic approach facilitates molecular-level engineering for enhanced interfacial stability of BP. Specifically, the GPT-4o model was utilized to analyze 286 papers in the literature and develop a recommendation scheme for identifying molecular groups with strong interactions with the BP surface. Key candidates, such as $-\text{SiR}_3$, $-\text{PR}_2$, $-\text{SH}$, and $=\text{NH}$, were recommended for their high affinity to BP to form the foundation for the subsequent ML process. A GNN model and a high-throughput molecular screening workflow were then established based on datasets constructed from the GPT-4o suggestions. Out of 117 million molecules in the PubChem database, 662 candidates were efficiently screened, and the top three molecules were selected for further validation. Their outstanding properties were firstly confirmed by DFT calculations. Experimental synthesis guided by LLMs based on 285 papers in the literature was further performed to demonstrate the exceptional passivation effects and stability of BP for over 20 days, which are substantially longer than those of unmodified BP and most previously reported systems. The molecular-passivated BP structures demonstrated clear advantages in preventing oxidation and degradation, as verified by various characterization techniques. Furthermore, the molecular functional motifs designed with head, linker, and tail groups showcased synergistic stabilization effects and enhanced interfacial properties beyond conventional designs. For instance, new P-, S-, and Si-based molecular groups, identified by the AI-driven workflows, offered drastically improved stability than traditional N-based groups. Such modifications suppressed surface degradation and facilitated the formation of uniform molecular layers with good protective capabilities.

Unlike traditional trial-and-error experiments, which are limited in terms of scalability and efficiency, our AI-driven approach offers significant advantages for molecular engineering and design. By combining LLM and ML, we were able to analyze literature quickly, systematically screen a vast list of candidates, and validate a portfolio of effective

molecular stabilizers. This streamlined workflow eliminates the laborious manual effort in traditional screening and design of interfacial materials, consequently enabling more expeditious discovery of higher-performing candidates in a fraction of the time. In addition, our proposed interfacial molecular engineering strategy introduces a novel synergistic approach that overcomes traditional limitations in the molecular mechanisms responsible for stability enhancement, resulting in the expansion of design principles for interfacial materials. For instance, hydrophilic molecules are typically considered detrimental to the surface stability of BP. However, by performing precise and synergistic design of the head, linker, and tail groups, significant improvement is accomplished in the stability of BP surfaces, even for molecules with hydrophilic tails. More importantly, these principles are not only applicable to BP but can also be generalized to other 2D materials requiring advanced interface stabilization, such as perovskite and MXene. In general, this study demonstrates the power of combining AI-driven data analytics with molecular engineering to address complex challenges. The framework yields scalable and efficient solutions and a robust, theoretical foundation for the systematic design of new interfacial materials.

Supporting Information

The authors have cited additional references within the Supporting Information.^[37–48]

Acknowledgements

We are grateful for the financial support from the National Key R&D Program of China (2023YFA0915600), the Shenzhen Medical Research Fund (No. B2302028), the National Natural Science Foundation of China (52203303 and 32471459), the Shenzhen Science and Technology Program (SGDX20211123151002003, GJHZ20220913142812025, RCJC20200714114435061, ZDSYS20220527171406014, KJZD20230923114703007, and JCYJ20220531095813031), the International Partnership Program of the Chinese Academy of Sciences (321GJHZ2023189FN), the SIAT International Joint Lab (E5G108), the City University of Hong Kong Donation Research Grants (Nos. DON-RMG 9229021 and 9229021), the Guangdong Basic and Applied Basic Research Foundation (2025A1515011408, 2023A1515110255, and 2025B1515020088), and the China Postdoctoral Science Foundation (2023M743671).

Conflict of Interests

The authors declare no conflict of interest.

Code Availability

The source code, including the details of the large language model and machine learning models used in this work,

is accessible on our GitHub repository: <https://github.com/ChaoPeng716/BP-Interface-Stability.git>.

Data Availability Statement

The data supporting the findings of this study are available within the published article and its Supporting Information.

Keywords: Black phosphorus • Large language model • Machine learning • Molecular engineering

- [1] G. Ashkenasy, D. Cahen, R. Cohen, A. Shanzer, A. Vilan, *Acc. Chem. Res.* **2002**, *35*, 121–128.
- [2] C. Sanchez, C. Boissiere, S. Cassaignon, C. Chaneac, O. Durupthy, M. Faustini, D. Grosso, C. Laberty-Robert, L. Nicole, D. Portehault, F. Ribot, L. Rozes, C. Sassoey, *Chem. Mater.* **2014**, *26*, 221–238.
- [3] T. Mueller, A. G. Kusne, R. Ramprasad, In *Reviews in Computational Chemistry*, Wiley **2016**, pp. 186–273, <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119148739.ch4>.
- [4] A. Ajagekar, F. You, *NPJ Comput. Mater.* **2023**, *9*, 143.
- [5] H. Li, H. Zheng, T. Yue, Z. Xie, S. Yu, J. Zhou, T. Kapri, Y. Wang, Z. Cao, H. Zhao, A. Kemelbay, J. He, G. Zhang, P. F. Pieters, E. A. Dailing, J. R. Cappiello, M. Salmeron, X. Gu, T. Xu, P. Wu, Y. Li, K. B. Sharpless, Y. Liu, *Nat. Energy* **2025**, *10*, 90–100.
- [6] Z. Yao, Y. Lum, A. Johnston, L. M. Mejia-Mendoza, X. Zhou, Y. Wen, A. Aspuru-Guzik, E. H. Sargent, Z. W. Seh, *Nat. Rev. Mater.* **2023**, *8*, 202–215.
- [7] Q. Zhang, Y. Hu, J. Yan, H. Zhang, X. Xie, J. Zhu, H. Li, X. Niu, L. Li, Y. Sun, W. Hu, *Adv. Mater.* **2024**, *36*, 2405163.
- [8] X. Han, H. Gong, H. Li, J. Sun, *Chem. Rev.* **2024**, *124*, 6903–6951.
- [9] Y. Liu, P. Gao, T. Zhang, X. Zhu, M. Zhang, M. Chen, P. Du, G. W. Wang, H. Ji, J. Yang, S. Yang, *Angew. Chem. Int. Ed.* **2019**, *58*, 1479–1483.
- [10] Y. Zhao, L. Tong, Z. Li, N. Yang, H. Fu, L. Wu, H. Cui, W. Zhou, J. Wang, H. Wang, P. K. Chu, X.-F. Yu, *Chem. Mater.* **2017**, *29*, 7131–7139.
- [11] Q. Zhou, Q. Chen, Y. Tong, J. Wang, *Angew. Chem. Int. Ed.* **2016**, *55*, 11437–11441.
- [12] Z. Guo, S. Chen, Z. Wang, Z. Yang, F. Liu, Y. Xu, J. Wang, Y. Yi, H. Zhang, L. Liao, P. K. Chu, X.-F. Yu, *Adv. Mater.* **2017**, *29*, 1703811.
- [13] Z. Sofer, J. Luxa, D. Bousa, D. Sedmidubsky, P. Lazar, T. Hartman, H. Hardtdegen, M. Pumera, *Angew. Chem. Int. Ed.* **2017**, *56*, 9891–9896.
- [14] R. Gusmao, Z. Sofer, M. Pumera, *ACS Nano* **2018**, *12*, 5666–5673.
- [15] Y. Xiong, Y. Wang, R. Zhu, H. Xu, C. Wu, J. Chen, Y. Ma, Y. Liu, Y. Chen, K. Watanabe, T. Taniguchi, M. Shi, X. Chen, Y. Lu, P. Zhan, Y. Hao, F. Xu, *Sci. Adv.* **2022**, *8*, eabo0375.
- [16] H. Shi, S. Fu, Y. Liu, C. Neumann, M. Wang, H. Dong, P. Kot, M. Bonn, H. I. Wang, A. Turchanin, O. G. Schmidt, A. Shaygan Nia, S. Yang, X. Feng, *Adv. Mater.* **2021**, *33*, e2105694.
- [17] Y. Liu, M. Chen, S. Yang, *InfoMat* **2021**, *3*, 231–251.
- [18] S. Teale, M. Degani, B. Chen, E. H. Sargent, G. Grancini, *Nat. Energy* **2024**, *9*, 779–792.
- [19] S. Pablo-García, S. Morandi, R. A. Vargas-Hernández, K. Jorner, Ž. Ivković, N. López, A. Aspuru-Guzik, *Nat. Comput. Sci.* **2023**, *3*, 433–442.
- [20] K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, A. Palmer, V.

- Settels, T. Jaakkola, K. Jensen, R. Barzilay, *J. Chem. Inf. Model.* **2019**, *59*, 3370–3388.
- [21] E. Heid, K. P. Greenman, Y. Chung, S.-C. Li, D. E. Graff, F. H. Vermeire, H. Wu, W. H. Green, C. J. McGill, *J. Chem. Inf. Model.* **2024**, *64*, 9–17.
- [22] J. Kang, J. D. Wood, S. A. Wells, J.-H. Lee, X. Liu, K.-S. Chen, M. C. Hersam, *ACS Nano* **2015**, *9*, 3596–3604.
- [23] Y. Zhao, Q. Zhou, Q. Li, X. Yao, J. Wang, *Adv. Mater.* **2017**, *29*, 1603990.
- [24] Q. Zhang, Y. Liu, J. Lai, S. Qi, C. An, Y. Lu, X. Duan, W. Pang, D. Zhang, D. Sun, J.-H. Chen, J. Liu, *2D Mater.* **2018**, *5*, 025012.
- [25] V. Lloret, E. Nuin, M. Kohring, S. Wild, M. Löffler, C. Neiss, M. Krieger, F. Hauke, A. Görling, H. B. Weber, G. Abellán, A. Hirsch, *Adv. Mater. Interfaces* **2020**, *7*, 2001290.
- [26] R. Jain, Y. Singh, S.-Y. Cho, S. P. Sasikala, S. H. Koo, R. Narayan, H.-T. Jung, Y. Jung, S. O. Kim, *Chem. Mater.* **2019**, *31*, 2786–2794.
- [27] H. Wang, K. Hu, Z. Li, C. Wang, M. Yu, Z. Li, Z. Li, *Small* **2018**, *14*, 1801701.
- [28] C. R. Ryder, J. D. Wood, S. A. Wells, Y. Yang, D. Jariwala, T. J. Marks, G. C. Schatz, M. C. Hersam, *Nat. Chem.* **2016**, *8*, 597–602.
- [29] S. J. R. Tan, I. Abdelwahab, L. Chu, S. M. Poh, Y. Liu, J. Lu, W. Chen, K. P. Loh, *Adv. Mater.* **2018**, *30*, 1704619.
- [30] Y. Zhao, H. Wang, H. Huang, Q. Xiao, Y. Xu, Z. Guo, H. Xie, J. Shao, Z. Sun, W. Han, X.-F. Yu, P. Li, P. K. Chu, *Angew. Chem. Int. Ed.* **2016**, *55*, 5003–5007.
- [31] L. Wu, J. Wang, J. Lu, D. Liu, N. Yang, H. Huang, P. K. Chu, X.-F. Yu, *Small* **2018**, *14*, 1801405.
- [32] Y. Huang, J. Qiao, K. He, S. Bliznakov, E. Sutter, X. Chen, D. Luo, F. Meng, D. Su, J. Decker, W. Ji, R. S. Ruoff, P. Sutter, *Chem. Mater.* **2016**, *28*, 8330–8339.
- [33] Y. H. Pai, C. H. Chen, *Nanoscale* **2021**, *13*, 763–775.
- [34] T. Gensch, G. dos Passos Gomes, P. Friederich, E. Peters, T. Gaudin, R. Pollice, K. Jorner, A. Nigam, M. Lindner, D'Addario, M. S. Sigman, A. Aspuru-Guzik, *J. Am. Chem. Soc.* **2022**, *144*, 1205–1217.
- [35] D. A. Boiko, R. MacKnight, B. Kline, G. Gomes, *Nature* **2023**, *624*, 570–578.
- [36] A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White, P. Schwaller, *Nat. Mach. Intell.* **2024**, *6*, 525–535.
- [37] B. Wang, C. Xu, X. Zhao, L. Ouyang, F. Wu, Z. Zhao, R. Xu, K. Liu, Y. Qu, F. Shang, *arXiv* **2024**, 2409–18839, <https://doi.org/10.48550/arXiv.2409.18839>.
- [38] X. Li, P. Yu, C. Zhou, T. Schick, O. Levy, L. Zettlemoyer, J. Weston, M. Lewis, *arXiv* **2023**, 2308–06259, <https://doi.org/10.48550/arXiv.2308.06259>.
- [39] Á. B. Del Canto, G. M. Blázquez, A. P. Lajarín, D. V. Suero, *GitHub Repository* **2024**, <https://github.com/argilla-io/distilabel>.
- [40] W. Liu, W. Zeng, K. He, Y. Jiang, J. He, *arXiv* **2023**, 2312–15685, <https://doi.org/10.48550/arXiv.2312.15685>.
- [41] S. Kim, J. Suk, S. Longpre, B. Y. Lin, J. Shin, S. Welleck, G. Neubig, M. Lee, K. Lee, M. Seo, *arXiv* **2024**, 2405–01535, <https://doi.org/10.48550/arXiv.2405.01535>.
- [42] G. Kresse, J. Furthmüller, *Comput. Mater. Sci.* **1996**, *6*, 15–50.
- [43] G. Kresse, J. Furthmüller, *Phys. Rev. B* **1996**, *54*, 11169–11186.
- [44] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- [45] G. Kresse, D. Joubert, *Phys. Rev. B* **1999**, *59*, 1758–1775.
- [46] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
- [47] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- [48] G. Henkelman, B. P. Uberuaga, H. Jónsson, *J. Chem. Phys.* **2000**, *113*, 9901–9904.

Manuscript received: April 16, 2025

Revised manuscript received: July 05, 2025

Accepted manuscript online: July 06, 2025

Version of record online: August 08, 2025

Supplementary Information

AI-Driven Discovery and Molecular Engineering Design for Enhancing Interface Stability of Black Phosphorus

Chao Peng,^{*[a,d]} Bing Wang,^[a] Lie Wu,^[a] Haoqu Jin,^[a] Yutang Li,^[a] Wenxia Gao,^[a,d] Jie Zhou,^[a] Guolai Jiang,^[a,d] Chen Wang,^[a,d] Jiahong Wang,^{*[a,d,e]} Xingchen He,^[a,d] Denis Kramer,^[b] Paul K. Chu,^[c] Xue-Feng Yu,^{*[a,d,e]}

[a] Dr. C. Peng, Dr. B. Wang, Dr. L. Wu, H. Q. Jin, Y. T. Li, W. X. Gao, J. Zhou, Dr. G. L. Jiang, C. Wang, Dr. J. H. Wang, Dr. X. C. He, Prof. X.-F. Yu
Materials Artificial Intelligence Center
Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
1068 Xueyuan Avenue, Shenzhen 518055, P. R. China
E-mail: xf.yu@siat.ac.cn; chao.peng@siat.ac.cn; jh.wang1@siat.ac.cn

[b] Prof. D. Kramer
Helmut-Schmidt-University, University of the Armed Forces
Holstenhofweg 85, Hamburg 22043, Germany

[c] Prof. P. K. Chu
Department of Physics, Department of Materials Science and Engineering, and Department of Biomedical Engineering,
City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong 999077, China

[d] Dr. C. Peng, W. X. Gao, Dr. G. L. Jiang, C. Wang, Dr. J. H. Wang, Dr. X. C. He, Prof. X.-F. Yu
University of Chinese Academy of Sciences, Beijing 100049, P. R. China

[e] Dr. J. H. Wang, Prof. X.-F. Yu
Key Laboratory of Biomedical Imaging Science and System, Chinese Academy of Sciences, State Key Laboratory of Biomedical Imaging Science and System
1068 Xueyuan Avenue, Shenzhen 518055, P. R. China

[f] Dr. L. Wu
State Key Laboratory of Advanced Refractories, Wuhan University of Science & Technology
947 Heping Avenue, Wuhan 430081, China

I. Supporting Text

1. Methods

1.1. Large-Language-Model approach

Data collection and processing: This study established a corpus database focused on enhancing the stability of black phosphorus by modifying its surface with functional groups. We primarily focused on the literature about phosphorus published in the last 20 years. In the final dataset, we selected 286 papers based on functional group recommendations and 285 papers according to the synthesis procedures. In order to extract high-quality textual information from the PDF documents, these documents were first processed using the MinerU tool,¹ performing paragraph segmentation and extraction of contents such as text, tables, and figures.

Data cleaning and instruction dataset construction: All the data processing operations were conducted in a Python 3.10 environment using the OpenAI package version 1.10.0. We distinguished different types of information by classifying text paragraphs. The tables, figures, and images were extracted by MinerU; GPT-4's multimodal capabilities were further utilized to summarize these images, reconstruct tables in LaTeX format, and classify text paragraphs. We predefined three categories: functional group recommendations, synthesis procedures, and other information. To generate high-quality instruction-response pairs, the InstructionBacktranslation method² from the Distilabel toolkit³ was employed to generate instructions and responses that aligned with research domain characteristics based on the aforementioned corpus. To further improve the quality of the generated data, the QualityScore⁴ and PrometheusEval⁵ methods provided by Distilabel were adopted to score and filter these instruction-response pairs to ensure that only high-quality data were included in the fine-tuning dataset.

Model fine-tuning and evaluation: The selected high-quality instruction-response pairs were used to tune the GPT-4 model. During fine-tuning, all the parameters adopted the default

configuration recommended by OpenAI. Through this process, the fine-tuned model was better adapted to the needs of the research field for BP interfacial stability, and its expertise and knowledge accuracy were significantly improved in this domain.

1.2. Machine-learning method

Two established graph neural network (GNN) architectures were adopted. The Graph-based Adsorption on Metal Energy Neural Network (GAME-Net) was implemented to predict the interaction energy between the molecule and BP surface ($E_{\text{int_BP}}$),⁶ while the Directed Message-Passing Neural Network (D-MPNN), as implemented in Chemprop (version 2.0.4),^{7,8} was utilized to predict the interaction energy between the terminal group and water molecule ($E_{\text{int_H}_2\text{O}}$). The molecular structures were encoded as graphs with atoms as vertices and bonds as edges. In the $E_{\text{int_BP}}$ prediction, the phosphorus atom serves as the adsorption site on the black phosphorus surface in the graph representation to capture the molecule-BP interactions⁶. Both models process the graph representations through multiple convolutional layers, followed by either a global pooling function (for $E_{\text{int_BP}}$ model) or an attentive aggregation layer with a feed-forward neural network (for $E_{\text{int_H}_2\text{O}}$ model) to generate the final predictions. The models were trained on our systematically generated dataset using an 8:1:1 split for training, validation, and test sets. The model performance was evaluated using the root mean square error (RMSE) between GNN-predicted and DFT-calculated energies. Hyperparameter optimization was conducted on the validation set to improve the prediction accuracy. The optimized hyperparameters were listed in Table S1 and Table S2.

1.3. DFT calculations

Density functional theory (DFT) computations were carried out using the Vienna Ab Initio Simulation Package (VASP).^{9,10} For exchange-correlation interactions, the Generalized Gradient

Approximation (GGA) was employed,¹¹ specifically utilizing the Perdew–Burke–Ernzerhof (PBE) functional. The core electrons were represented by the projector-augmented wave (PAW) method, and the valence electrons were modeled using plane-wave basis sets.¹² To account for dispersion forces, DFT-D3 correction with Becke–Johnson (BJ) damping was applied.^{13,14} A plane-wave energy cutoff of 450 eV was used, and the force convergence threshold was set to 0.05 eV/Å. The energy barrier of H₂O molecule diffusing through molecular layers was calculated by employing the climbing image nudged elastic band method (CI-NEB) method.¹⁵ The adsorption energy (E_{ads}) was determined by the following equation:

$$E_{\text{ads}} = E(\text{Surf|adsorbate}) - E(\text{adsorbate}) - E(\text{Surf}), \quad (1)$$

where $E(\text{Surf|adsorbate})$ refers to the total energy of the system with the adsorbate on BP(010) surface, $E(\text{adsorbate})$ is the energy of the isolated adsorbate, and $E(\text{Surf})$ is the energy of the BP(010) slab.

1.4. Experimental synthesis and method

Materials: The BP crystals were sourced from Mophos (www.Mophos.cn) and stored in a dark argon glove box. 2-Methylpropylthiourea (N₁) and N-methylglycinamide (N₄) were acquired from Nanjing Chemlin Chemical Co., Ltd. Diphenylmethanimine (N₂), octan-1-amine (N₃), and 2-amino-N-[2-(methylamino)-2-oxoethyl]acetamide (N₅) were purchased from Aladdin Reagents. Dicyclohexylphosphane (P₁), octylphosphane (P₂), 2-[(2-methylpropan-2-yl)oxy]ethylphosphane (P₃), pentylphosphane (P₄), and (4-methylphenyl)phosphane (P₅) were obtained from Wuhan Shiruikai Technology Co., Ltd. 1H,1H,2H,2H-Perfluorodecyltriethoxysilane (Si₁), trimethoxy(octyl)silane (Si₂), dimethoxy(methyl)octylsilane (Si₃), n-octyl dimethyl methoxysilane (Si₄), and decyl-methoxy-dimethylsilane (Si₅) were sourced from J&K Scientific. 1,1,1-Tris(mercaptomethyl)undecane (S₁), 2-heptylpropane-1,3-dithiol (S₂), 1-octanethiol (S₃), 1-

undecanethiol (S₄), and 11-mercapto-1-undecanol (S₅) were procured from Energy Chemical. NMP (99.5%, anhydrous), DMF (99%), DMSO (99.5%), and ethanol (EtOH) were also purchased from Aladdin Reagents. All chemicals used in this study were of analytical reagent grade and were used without further purification.

Synthesis of BP nanosheets (BPs): The BP nanosheets (BPs) were prepared by a solvent exfoliation method. Specifically, bulk BP (10 mg) was dispersed in 20 mL of NMP and sonicated in a 40 kHz ice-cooled ultrasonic bath for 6 hours. The solution was centrifuged at 7,000 rpm for 15 minutes on a Thermo ST-16 centrifuge to remove unexfoliated bulk BP. The supernatant containing the exfoliated BPs was collected and stored for further use.

Mechanical exfoliation of BP microflakes (BP MFs): The thin BP MFs were prepared by a scotch tape-based mechanical exfoliation method. The microflakes were peeled from the bulk crystal and transferred onto silicon (Si)/silicon dioxide (SiO₂) substrates, which had been sonicated in ethanol (EtOH) and dried with a nitrogen (N₂) stream.

General methods for surface passivation of BPs: The BPs were dispersed in 1 mL of organic solvent (10 μmol), and excessive organic molecules were added to the solution. The mixture was stirred under argon and in darkness for 20 hours. Subsequently, the mixture was centrifuged at 12,000 rpm for 20 minutes to collect the precipitate of passivated BP. To obtain the BP microflakes, the samples were immersed in organic solutions for 10 hours. After modification, the samples were rinsed with ethanol (EtOH) and stored for further use.

Materials Characterization: The transmission electron microscopy (TEM) and high-resolution TEM (HR-TEM) images were obtained on the JEM-F200 (JEOL, Japan). Atomic force microscopy (AFM) was conducted on the Cypher S (Asylum Research, USA) and Fourier transform infrared spectroscopy (FTIR) was performed on a NICOLET iS50 spectrometer

(Thermo Scientific, USA). The UV-Vis absorption spectra were recorded on the A390 spectrophotometer (AOE, China), and the ^1H NMR spectra were acquired on a Bruker Advance DRX-300 spectrometer (Karlsruhe, Germany) at 25 °C. The optical images of the electrospun samples were acquired on the BX53M metallurgical microscope (Olympus, Japan).

1.5. Details of data processing

The data processing pipeline consisted of four key steps, each addressing specific challenges in transforming raw literature into high-quality fine-tuning data. First, the PDF literature was converted to Markdown format to facilitate text processing. Second, the language models were employed to segment texts by primary headings such as Abstract, Experimental Methods. It was crucial to allow to focus on key sections like “Experimental Methods” that contain rich information about organic molecules, functional groups and synthesis details, while avoiding incomplete information extraction directly from long-text processing. Third, “functional group information” and “synthesis scheme information” were specifically extracted from segmented text blocks. These two data types were of central importance for generating recommendations of organic functional group as they directly relate to chemical transformations and molecular structures. This step also served as an essential denoising and refinement procedure. Finally, high-quality question-answer (QA) pairs were constructed by generating candidate questions from extracted information, selecting optimal questions, and formulating precise answers. This multi-step QA construction process functioned as a deep cleaning and validation layer, ensuring the relevance, accuracy, and quality of the training data.

1.6. Detailed description of the SMARTS searching process

With the validated models in place, we initiated the first step of high-throughput virtual screening by constructing a candidate compound library through substructure-based searches of

the PubChem database, which contains approximately 117 million organic compounds with broad chemical and structural diversity. To identify molecules containing the desired core scaffold and functional groups while allowing for structural variability, we used SMILES Arbitrary Target Specification (SMARTS), a molecular pattern language widely used for defining substructures in cheminformatics, to perform substructure matching. SMARTS strings were constructed following the same block-based strategy previously applied in the creation of the initial molecular library. By systematically combining SMARTS patterns for each structural block (also including one head block, several linker blocks and one tail block), we generated composite query strings for targeted searching within PubChem. In addition, filtering criteria were applied based on heavy atom count (≤ 20) and molecular complexity (≤ 150) to further narrow down the search space. All retrieved compounds were subsequently deduplicated, and any molecules already present in the training set were excluded. This step quickly identified approximately 45.2k molecules with specific structural patterns from the PubChem database.

II. Supporting Figures



Could you recommend which functional groups might interact with the surface of black phosphorus to improve its stability? Please provide a list of structures and a reasonable interaction mechanism, and summarize the recommended functional groups in a table format.



Certainly! To enhance the stability of black phosphorus (BP) in water and oxygen environments via surface modification with organic small molecules, we need to create a molecular protective layer that effectively interacts with and shields the surface of BP. This can be achieved by utilizing functional groups that can form strong chemical bonds or physical interactions with BP, reducing its reactivity with water and oxygen.

Recommended Functional Groups and Their Interaction Mechanisms

1. -PR₂ (Phosphino Group)

Mechanism: Phosphino groups can form covalent bonds with the phosphorus atoms in BP. These bonds create a robust protective layer that enhances stability by preventing oxidation and hydrolysis.

2. -SiR₃ (Silyl Group)

Mechanism: Silyl groups can form Si-P bonds, which result in a hydrophobic layer. This reduces exposure to moisture and oxygen, thereby improving BP stability.

3. -SH (Thiol Group)

Mechanism: Thiol groups can react with BP through a Lewis acid-base interaction, forming S-P bonds. Additionally, thiols can undergo oxidation to form disulfides, further passivating the surface.

4. -NH₂ (Amine Group)

Mechanism: Amine groups can form hydrogen bonds with surface oxidation products of BP (e.g., P-OH functional groups). They can also act as Lewis bases to stabilize BP.

5. C=NH (Imine Group)

Mechanism: Imines, due to their conjugated structure, can form stronger interactions with BP through hydrogen bonding or covalent interactions with surface phosphorus oxides.

.....

Summary of Recommended Functional Groups

Below is a table summarizing the above functional groups along with their mechanisms:

Functional Group	Mechanism
-PR ₂ (Phosphino)	Covalent bonding, forming a protective layer
-SiR ₃ (Silyl)	Si-P bonding, hydrophobic layer formation
-SH (Thiol)	S-P bonding, oxidation to form passivating layer
-NH ₂ (Amine)	Hydrogen bonding, Lewis base stabilization
C=NH (Imine)	Hydrogen bonding, conjugated system for stability

These functional groups contribute to the stability of black phosphorus by forming protective barriers, either through chemical bonding or physical interactions, thereby reducing its exposure to oxidative and hydrolytic degradation in water and oxygen.

This systematic approach not only helps in choosing appropriate functional groups for BP modification but also ensures an effective enhancement of its surface stability.

Figure S1. ChatGPT-4o-BP: Dialogue example image: Recommended dialogue sample for functional group modification on the surface of black phosphorus.

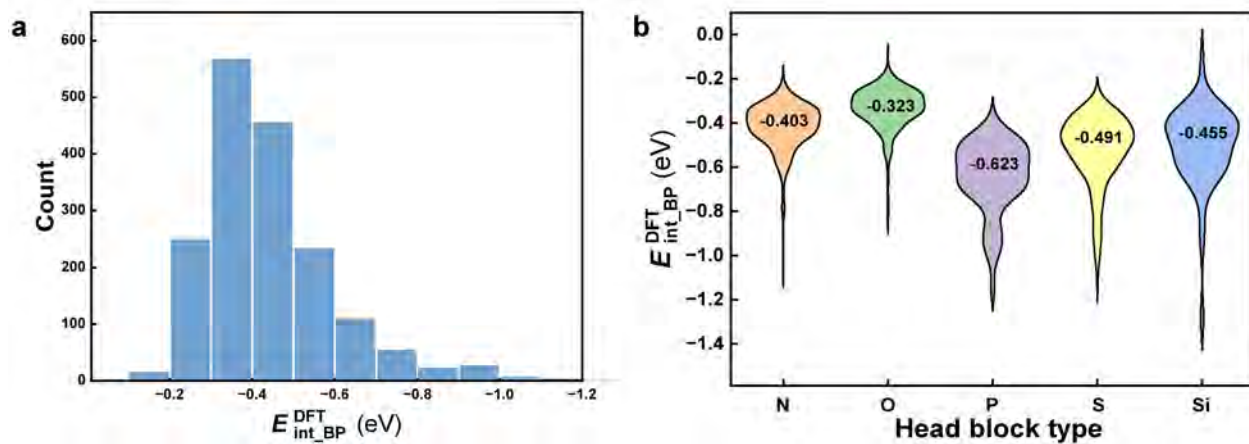


Figure S2. Distribution of DFT-calculated $E_{\text{int_BP}}$ values of the dataset. **a**, Overall distribution of interaction energies. **b**, $E_{\text{int_BP}}$ values categorized by head block type, with median values reported for each group.

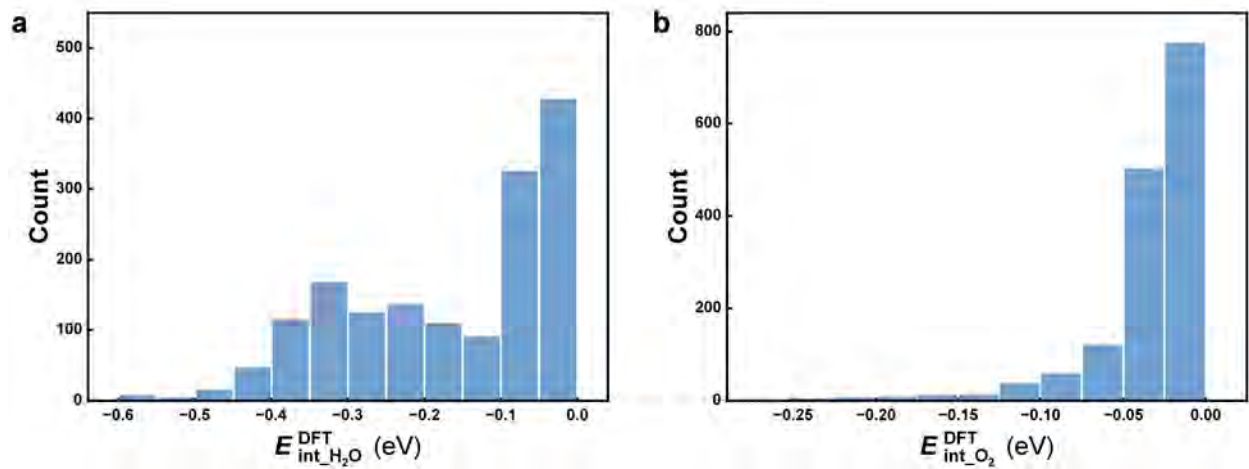


Figure S3. **a**, Distribution of DFT-calculated $E_{\text{int_H}_2\text{O}}$ values of the dataset. **b**, Distribution of DFT-calculated $E_{\text{int_O}_2}$ values of the dataset.



We plan to modify the surface of black phosphorus with organic molecules to form a molecular protective layer, thereby improving the surface stability of the black phosphorus material. Please recommend an experimental preparation plan for liquid phase modification of black phosphorus nanosheets based on the molecule CCCCCCCCCCCC(CS)(CS)CS (1,1,1-Tris(mercaptomethyl)undecane) I provided.



To address your question about improving the surface stability of black phosphorus (BP) using 1,1,1-Tris(mercaptomethyl)undecane (CCCCCCCCCCCC(CS)(CS)CS), we'll approach this through a liquid phase modification process. This involves creating a stable covalent bond or interaction between the BP surface and your organic molecule.

Here is a detailed experimental preparation plan:

Key Points to Consider:

Stability: The main goal is to create a protective molecular layer that limits BP's degradation when exposed to air.

Functional Groups: The thiol (-SH) groups in the molecule enable strong interaction with the BP surface, potentially forming P-S bonds.

Reaction Medium: A suitable solvent must be chosen to disperse both BP and 1,1,1-Tris(mercaptomethyl)undecane effectively.

Temperature and Time: Precise conditions are necessary to ensure effective modification without degrading BP.

Experimental Preparation Plan:

Materials:

- **Black Phosphorus:** Exfoliated nanosheets from bulk BP.
- **1,1,1-Tris(mercaptomethyl)undecane (CCCCCCCCCCCC(CS)(CS)CS):** The modifying agent.
- **Solvent:** Ethanol or anhydrous toluene. Ethanol is more commonly used for thiol-based reactions, but toluene may help in achieving better dispersion.
- **Argon or Nitrogen gas:** To maintain an inert atmosphere.
- **Vacuum oven or desiccator:** For drying.

Procedure:

- **Exfoliation of Black Phosphorus:**
 - Start with bulk BP crystals.
 - Disperse ~10 mg of BP in 20 mL of an anhydrous and oxygen-free solvent (ethanol or toluene).
 - Use ultrasonication (bath or probe sonicator) for 6 hour under an ice bath to avoid heating.
 - Work under an inert atmosphere by bubbling argon or nitrogen to prevent oxidation.
- **Preparation of 1,1,1-Tris(mercaptomethyl)undecane Solution:**
 - Dissolve the organic molecule in the same solvent (ethanol or toluene) at a concentration of about 10 mM.
 - Ensure that the solvent is anhydrous to avoid premature reaction of thiol groups.
- **Modification Reaction:**
 - Mix the exfoliated BP dispersion with the 1,1,1-Tris(mercaptomethyl)undecane solution in a 1:5 molar ratio of phosphorus atoms to the thiol compound.
 - Stir the mixture at room temperature for 12-24 hours under an inert atmosphere to allow sufficient time for the organic molecules to attach to the BP surface.
- **Purification:**
 - After the reaction, centrifuge the mixture at 5000 rpm for 10 minutes to separate the modified BP from the solvent.
 - Wash the precipitate with anhydrous ethanol/toluene at least three times to remove unreacted molecules.
 - Dry the modified BP under vacuum at room temperature to avoid thermal degradation.
- **Characterization:**
 - Use Raman spectroscopy to confirm the presence of BP and check for alterations in vibrational modes indicative of chemical modification.
 - X-ray photoelectron spectroscopy (XPS) to verify P-S bonding on the BP surface.
 - Atomic force microscopy (AFM) or scanning electron microscopy (SEM) to observe morphological changes.
 - UV-Vis spectroscopy to analyze any optical property changes.

Notes:

- **Safety:** Ensure proper handling of black phosphorus and thiol compounds in an inert atmosphere to prevent oxidation and degradation.
- **Optimization:** The reaction time, temperature, and ratios may need optimization depending on the specific quality of BP and molecule purity.

By following this method, you should form a stable, protective layer on the black phosphorus surface to enhance its air stability effectively.

Figure S4. ChatGPT-4o-BP: Dialogue example image: Recommended scheme of experimental synthesis procedures for passivating BP surfaces with CCCCCCCCCCCC(CS)(CS)CS.



We plan to modify the surface of black phosphorus with organic molecules to form a molecular protective layer, thereby improving the surface stability of the black phosphorus material. Please recommend an experimental preparation plan for liquid phase modification of black phosphorus nanosheets based on the molecule C1CCC(CC1)PC2CCCCC2 (dicyclohexylphosphane) I provided.



1. Materials Preparation

Black Phosphorus Bulk Crystals: Source high-purity BP crystals.

Dicyclohexylphosphane: Use this as the modifying molecule.

Organic Solvent: Choose anhydrous solvents like N-methyl-2-pyrrolidone (NMP) or dimethyl sulfoxide (DMSO) to prevent premature oxidation of BP.

Optional Chemicals: Isopropanol (IPA) and deionized (DI) water for washing.

2. Synthesis and Exfoliation of BP Nanosheets

Exfoliation Process

- **Materials:** Use bulk BP, anhydrous NMP or acetone, and nitrogen or argon gas for an inert atmosphere.
- **Procedure:**
 - **Dissolve Bulk BP:** Add 10–20 mg of BP crystals to 10 mL of anhydrous NMP in a well-sealed vial, ensuring the operation occurs in an inert atmosphere, such as a nitrogen-filled glovebox.
 - **Sonication:** Sonicate the mixture for 6 hours using a probe sonicator at 10–20 kHz to exfoliate the BP into nanosheets. Keep the vial cooled using an ice bath to prevent thermal degradation.
 - **Centrifugation:** Centrifuge the sonicated dispersion at 1500 rpm for 30 minutes to remove unexfoliated BP. Collect the supernatant, which contains the exfoliated nanosheets.

Key Points

- Ensure all steps occur under strictly inert conditions to prevent oxidation.
- Use nitrogen purging or work inside a glovebox.

3. Surface Modification with Dicyclohexylphosphane

Modification Process

- **Materials:**
 - Exfoliated BP nanosheet dispersion (in NMP).
 - Dicyclohexylphosphane.
 - Additional anhydrous NMP if dilution is necessary.
- **Procedure:**
 - **Preparation of Modifier Solution:** Dissolve a suitable amount of dicyclohexylphosphane (e.g., 10 mg) in 5 mL of anhydrous NMP. Stir under inert conditions.
 - **Add Modifier to BP Solution:** Slowly add the phosphane solution to the BP nanosheet dispersion under continuous stirring. Maintain an inert atmosphere (e.g., nitrogen bubbling) throughout the process.
 - **Reaction Conditions:** Allow the mixture to stir at room temperature for 12–24 hours. This duration ensures sufficient interaction between the dicyclohexylphosphane molecule and BP, forming P-BP bonds.

Key Points

- Ensure the molar ratio of dicyclohexylphosphane to BP surface phosphorus atoms is adequate to form a complete protective layer.
- Maintain oxygen-free conditions throughout to prevent BP oxidation.

4. Post-Treatment and Characterization

Isolation and Washing

- **Procedure:**
 - **Precipitation:** Add a non-solvent like ethanol to precipitate the modified BP nanosheets.
 - **Centrifugation:** Centrifuge at 3000 rpm for 10 minutes to collect the modified nanosheets.
 - **Washing:** Wash the nanosheets with anhydrous IPA to remove unreacted molecules. Repeat washing three times.

Drying

Dry the product under vacuum at room temperature to avoid decomposition or oxidation.

Characterization

- **Techniques:**
 - **Raman Spectroscopy:** Confirm structural integrity of BP.
 - **XPS (X-ray Photoelectron Spectroscopy):** Verify P-BP bond formation.
 - **AFM (Atomic Force Microscopy):** Confirm nanosheet thickness.
 - **TEM (Transmission Electron Microscopy):** Check morphology and dispersion of nanosheets.

Conclusion

By following this carefully controlled preparation plan, a stable, covalently bonded protective layer of dicyclohexylphosphane can be formed on BP, significantly enhancing its stability against water and oxygen degradation. This approach takes advantage of the strong phosphine chemistry to provide effective surface passivation.

Figure S5. ChatGPT-4o-BP: Dialogue example image: Recommended scheme of experimental synthesis procedures for passivating BP surfaces with C1CCC(CC1)PC2CCCCC2.

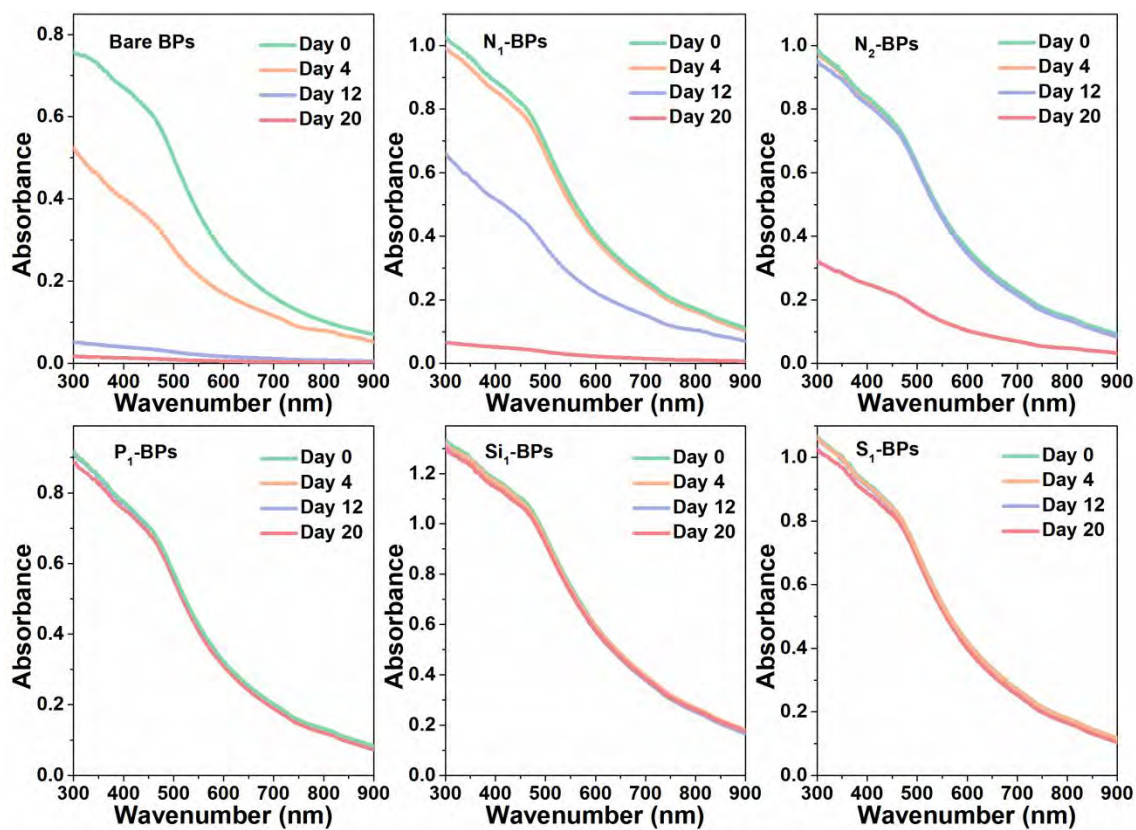


Figure S7. Absorption spectra of the bare BPs and passivated BPs by different organic molecules after storing in water for different periods of time.

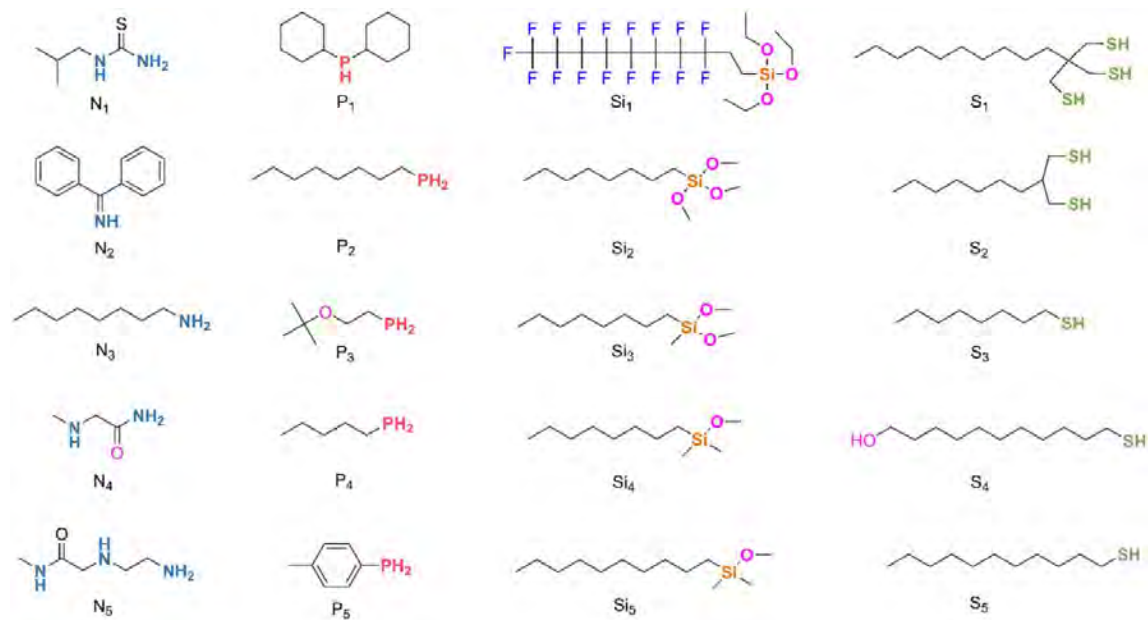


Figure S8. Structures of experimentally validated molecules featuring N-, P-, Si-, and S-based head groups.

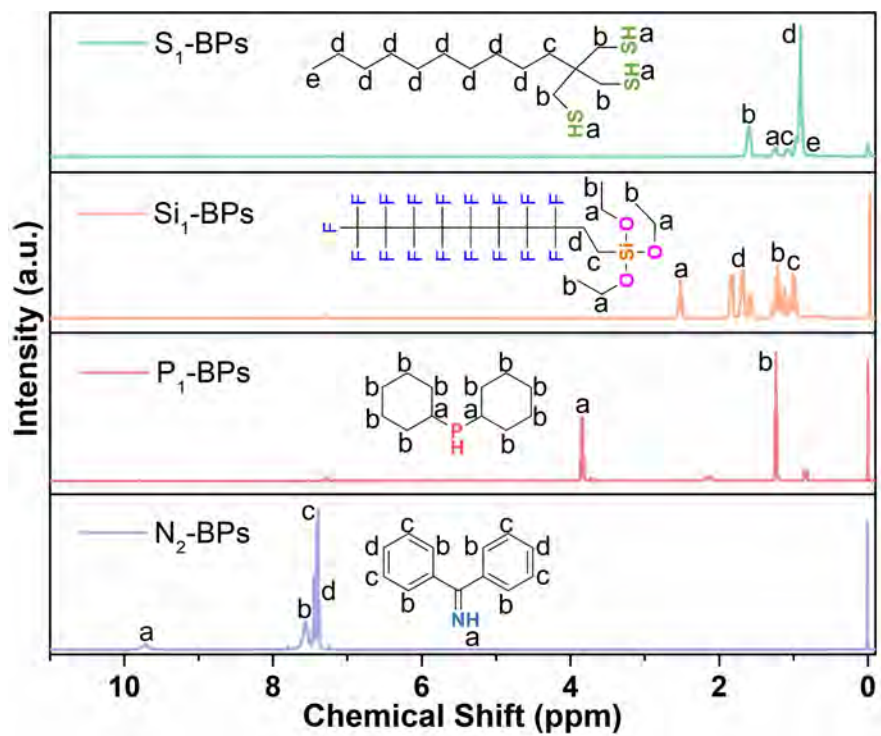


Figure S9. ^1H NMR spectra of $\text{S}_1\text{-BPs}$, $\text{Si}_1\text{-BPs}$, $\text{P}_1\text{-BPs}$, and $\text{N}_2\text{-BPs}$.

III. Supporting Tables

Table S1. Key hyperparameters applied in the $E_{\text{int_H}_2\text{O}}$ model*.

Hyperparameter	Value	Description
batch_size	16	The batch size of data
max_epoch	500	The max epochs applied in final model training process
patience_early_stopping	50	Number of epochs to wait for improvement before early stopping in final model training process
init_lr	5e-4	The initial learning rate of optimizer
final_lr	5e-7	The final learning rate of optimizer
max_lr	1e-3	The max learning rate of optimizer
d_h	300	The hidden dimension of message passing layer
depth	4	The number of message passing iterations
dropout	0.2	The dropout probability of message passing layer
Aggregation layer	Attentive Aggregation	The type of aggregation layer
batch_norm	False	No Batch normalization layer
hidden_dim	300	The hidden dimension of feed forward network layer
n_layers	3	The number of feed forward network layer
dropout	0.2	The dropout probability of feed forward network layer
activation	leakyReLU	The activation function of feed forward network layer

*All unspecified parameters retain Chemprop v2.0.4 default values.

Table S2. Key hyperparameters applied in the $E_{\text{int_BP}}$ model*.

Hyperparameter	Value	Description
batch_size	16	The batch size of data
epochs	100	The epochs applied in model training process
lr0	1e-3	The initial learning rate of optimizer
patience	5	Patience of the lr-scheduler
factor	0.7	Decreasing factor of the lr-scheduler
minlr	1e-8	Minimum lr of the lr-scheduler
dim	160	Dimension of the layers
n_linear	0	Number of fully connected layers
n_conv	3	Number of convolutional layers

*All unspecified parameters retain the optimum values reported in GAME-Net ⁶.

References

- [1] B. Wang, C. Xu, X. Zhao, L. Ouyang, F. Wu, Z. Zhao, R. Xu, K. Liu, Y. Qu, F. Shang, “Mineru: An Open-Source Solution for Precise Document Content Extraction” *arXiv* **2024**, 2409.18839.
- [2] X. Li, P. Yu, C. Zhou, T. Schick, O. Levy, L. Zettlemoyer, J. Weston, M. Lewis, “Self-Alignment with Instruction Backtranslation” *arXiv* **2023**, 2308.06259.
- [3] Á. B. Del Canto, G. M. Blázquez, A. P. Lajarín, D. V. Suero, “Distilabel: An AI Feedback (AIF) Framework for Building Datasets with and for LLMs” *GitHub repository* **2024**, <https://github.com/argilla-io/distilabel>.
- [4] W. Liu, W. Zeng, K. He, Y. Jiang, J. He, “What Makes Good Data for Alignment? A Comprehensive Study of Automatic Data Selection in Instruction Tuning” *arXiv* **2023**, 2312.15685.
- [5] S. Kim, J. Suk, S. Longpre, B. Y. Lin, J. Shin, S. Welleck, G. Neubig, M. Lee, K. Lee, M. Seo, “Prometheus 2: An Open Source Language Model Specialized in Evaluating Other Language Models” *arXiv* **2024**, 2405.01535.

- [6] S. Pablo-García, S. Morandi, R. A. Vargas-Hernández, K. Jorner, Ž. Ivković, N. López, A. Aspuru-Guzik, “Fast Evaluation of The Adsorption Energy of Organic Molecules on Metals via Graph Neural Networks” *Nat. Comput. Sci.* **2023**, *3*, 433-442.
- [7] K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, A. Palmer, V. Settels, T. Jaakkola, K. Jensen, R. Barzilay, “Analyzing Learned Molecular Representations for Property Prediction” *J. Chem. Inf. Model.* **2019**, *59*, 3370-3388.
- [8] E. Heid, K. P. Greenman, Y. Chung, S.-C. Li, D. E. Graff, F. H. Vermeire, H. Wu, W. H. Green, C. J. McGill, “Chemprop: A Machine Learning Package for Chemical Property Prediction” *J. Chem. Inf. Model.* **2024**, *64*, 9-17.
- [9] G. Kresse, J. Furthmüller, “Efficiency of Ab-initio Total Energy Calculations for Metals and Semiconductors Using A Plane-Wave Basis Set” *Comput. Mater. Sci.* **1996**, *6*, 15-50.
- [10] G. Kresse, J. Furthmüller, “Efficient Iterative Schemes for Ab-initio Total Energy Calculations Using A Plane-Wave Basis Set” *Phys. Rev. B* **1996**, *54*, 11169-11186.
- [11] J. P. Perdew, K. Burke, M. Ernzerhof, “Generalized Gradient Approximation Made Simple” *Phys. Rev. Lett.* **1996**, *77*, 3865-3868.
- [12] G. Kresse, D. Joubert, “From Ultrasoft Pseudopotentials to The Projector Augmented-Wave Method” *Phys. Rev. B* **1999**, *59*, 1758-1775.
- [13] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, “A Consistent And Accurate Ab-initio Parametrization of Density Functional Dispersion Correction (DFT-D) for The 94 Elements H-Pu” *J. Chem. Phys.* **2010**, *132*, 154104.
- [14] S. Grimme, S. Ehrlich, L. Goerigk, “Effect of The Damping Function in Dispersion Corrected Density Functional Theory” *J. Comput. Chem.* **2011**, *32*, 1456-1465.
- [15] G. Henkelman, B. P. Uberuaga, H. Jónsson, “A Climbing Image Nudged Elastic Band Method for Finding Saddle Points and Minimum Energy Paths” *J. Chem. Phys.* **2000**, *113*, 9901-9904.