ACS OMEGA

Open Access

This article is licensed under CC-BY-NC-ND 4.0

http://pubs.acs.org/journal/acsodf

Article

Machine Learning-Assisted Computational Screening of Adhesive Molecules Derived from Dihydroxyphenyl Alanine

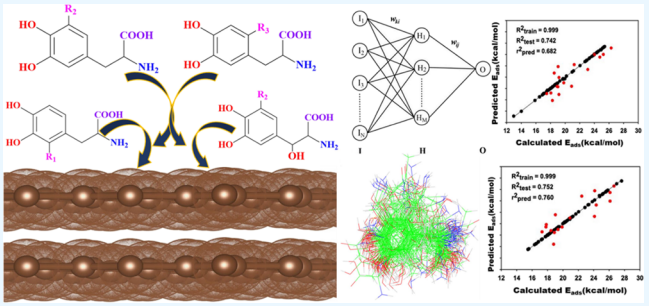
Srimai Vuppala,[§] Ramesh Kumar Chitumalla,[§] Seyong Choi, Taeho Kim, Hwangseo Park,* and Joonkyung Jang*

Cite This: *ACS Omega* 2024, 9, 994−1000

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Marine mussels adhere to virtually any surface via 3,4-dihydroxyphenyl-L-alanines (L-DOPA), an amino acid largely contained in their foot proteins. The biofriendly, water-repellent, and strong adhesion of L-DOPA are unparalleled by any synthetic adhesive. Inspired by this, we computationally designed diverse derivatives of DOPA and studied their potential as adhesives or coating materials. We used first-principles calculations to investigate the adsorption of the DOPA derivatives on graphite. The presence of an electron-withdrawing group, such as nitrogen dioxide, strengthens the adsorption by increasing the $\pi-\pi$ interaction between DOPA and graphite. To quantify the distribution of electron charge and to gain insights into the charge distribution at interfaces, we performed Bader charge analysis and examined charge density difference plots. We developed a quantitative structure−property relationship (QSPR) model using an artificial neural network (ANN) to predict the adsorption energy. Using the three-dimensional and quantum mechanical electrostatic potential of a molecule as a descriptor, the present quantum NN model shows promising performance as a predictive QSPR model.



1. INTRODUCTION

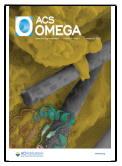
Marine mussels have genetically developed an extraordinary aptitude for adhering to wet (underwater) surfaces via their *byssuses*. So far, researchers have identified several proteins in the byssal plaques of mussels: among these, *mussel foot protein* (MFP)-3, -5, and -6 are known to play important roles in adhesion.[1,2] Interestingly, all of the MFPs undergo the posttranslational modification of tyrosine to *3,4-dihydroxyphenyl*-L-alanine (L-DOPA) which constitutes up to 30% of MFP-5.

The current consensus is that L-DOPA is responsible for the remarkable adhesiveness of MFP[1−4]: L-DOPAs play vital roles both in the initial *anchoring* of an MFP on a surface[5,6] and in the subsequent *cross-linking* formed between MFPs.[7,8] These essential roles of L-DOPAs[9−14] have triggered extensive efforts to synthesize adhesives analogous to DOPA. In particular, *polydopamines* (PDAs) have been widely used as adhesives or coating materials on various (organic and inorganic) substrates.[6,15−18]

However, diverse chemical derivatives of DOPA, other than DA, are possible by attaching various functional groups. Such substituent effects have not been studied in the context of DOPA-inspired adhesives. Herein, by using *first-principles* calculations, we simulate chemically diverse derivatives of DOPA adsorbed on a graphite surface. Graphite is chosen because it is a representative nonpolar surface utilized in wide applications: for example, a graphite coated with PDA shows an enhanced rate capability as an anode material in a lithium-ion battery.[19−23] Graphite is a versatile surface that can be used to study a wide range of adhesive interactions including van der Waals forces, electrostatic forces, and covalent bonds. This makes it a good model surface for studying the adhesive properties.

We construct a data set containing the molecular structures and properties of 170 derivatives of DOPA (Figures 1 and S1). The present data set also contains the geometries and energies of these derivatives adsorbed on a graphite surface. We extract molecular insights into how the adsorption energy is related to the structure and physicochemical properties of an adsorbate molecule. In addition, an *artificial neural network* (ANN) model is used to draw a *quantitative structure−property relationship* (QSPR) model which predicts the adsorption energy for a given molecule. By using the *three-dimensional* (3D) and *quantum mechanical* (QM) molecular descriptors, the present ANN model illustrates a promising performance in predicting the adsorption energy of a DOPA derivative.

Received: September 19, 2023
Revised: November 28, 2023
Accepted: November 30, 2023
Published: December 19, 2023



© 2023 The Authors. Published by American Chemical Society

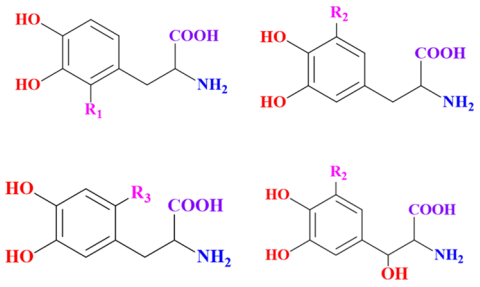


Figure 1. Computational design of a molecule derived from DOPA. Chemically distinct substituents are attached to one of the three (R_1, R_2, and R_3) positions of the aromatic ring of DOPA. Also designed are the molecules doubly substituted with R_2 and with −OH at the alkyl side chain (bottom right).

2. COMPUTATIONAL METHODS

2.1. Adsorption of a DOPA Derivative on Graphite. We designed 170 adsorbate molecules by attaching 35 chemically distinct substituents at various positions of DOPA. We attach a substituent at one of the three different locations in the aromatic ring of DOPA (R_1, R_2, and R_3 in Figure 1). We also designed adsorbates with double substituents with R_2 and −OH at the side alkyl chain of DOPA (Figure 1). These substituents are −Cl, −F, −CN, $-NO_2$, −CHO, $-COCH_3$, $-CH_3$, $-CF_3$, $-CH(CH_3)_3$, −OH, $-CH_2OH$, $-CH_2Cl$, −SH, $-CH_2F$, $-CH_2CN$, $-NH_2$, −Br, −OCN, $-OCH_3$, $-CH_2NH_2$, $-SCH_3$, $-CH_2CH_3$, $-OCH_2F$, $-COOCH_3$, $-CH{=}CH_2$, $-C_2H_2$, $-CH{=}CH{-}CH_3$, $-CH{\equiv}C{-}CH_3$, $-CH_2OCH_3$, −COCl, $-CONH_2$, −COOH, $-CH_2C_6H_5$, $-CH_2COC_6H_5$, and $-CH_2OC_6H_5$. The substituents differ in their electronic, steric, and functional characteristics. For example, the data set includes strong electron-withdrawing groups, such as $-NO_2$ and −CN, and the moderate-to-weak electron-withdrawing groups, such as −CHO and −COOR. Also included in the data set are the moderate-to-weak electron-donating groups such as $-CH_3$ and $-CH_2CH_3$ and strong electron-donating groups such as $-NH_2$ and −OH.

By systematically tuning the chemical properties and placement of a substituent, we investigated the influence of these variations on the adsorption of a DOPA derivative. This study not only provides insights into the interplay between molecular structure and adsorption characteristics but also provides basic guidelines for the rational design of an adhesive molecule for wide applications. We optimize the molecular geometries of adsorbates by using the *density functional theory* (DFT) at the level of B3LYP/6−31G(d)[24] implemented in the Gaussian16 suite.[25]

Upon completion of geometry optimization, an adsorbate molecule is deposited on a graphite surface. The geometry and energy of the adsorption are simulated by using the DFT at the level of the Perdew-Burke-Ernzerhof (PBE) functional with the generalized gradient approximation (GGA) for the exchange correlation. The projector-augmented-wave (PAW) potentials are used to simulate the electronic−ionic core interactions for atoms, such as hydrogen, carbon, nitrogen, oxygen, sulfur, and bromine. The spin-polarized DFT simulations are carried out by employing the Vienna ab initio simulation package (VASP).[26−28] As the dispersion plays a significant role in evaluating the binding energy of the DOPA derivatives on the graphite surface, we considered the dispersion correction using Grimme's DFT+D3 approach.[29]

A graphite surface is modeled as an AB-stacked[30−35] bilayer of graphene. The graphite surface is represented by a 6 × 6 supercell containing 144 carbon atoms under the periodic boundary conditions ($a = b = 14.76$ Å and $c = 45$ Å). We fix atoms in the bottom layer of graphite in simulation. For the Brillouin-zone sampling, the Monkhorst-Pack scheme with 1 × 1 × 1 k-points is employed, along with a plane-wave cutoff energy of 450 eV. We calculate the *adsorption energy*, E_{ads}, defined as

$$E_{ads} = (E_{surf} + E_{mol}) - E_{mol@surf} \quad (1)$$

where E_{surf}, E_{mol}, and $E_{mol@surf}$ are the energies of the surface, adsorbate, and adsorbate@surface, respectively.

2.2. QM-3D Descriptor. A QSPR model to predict the adsorption energy E_{ads} of a given molecule requires a molecular descriptor as input. In principle, such a descriptor should accurately capture both the structure and physicochemical properties of a molecule. We propose that the 3D distribution of the *electrostatic potential* (ESP) of a molecule should be a reasonable descriptor. Moreover, the 3D ESP can be accurately calculated by using first-principles methods such as the DFT or *Hartree−Fock* (HF) theory. Previously, this QM-3D descriptor has proven to be successful for QSPR modeling of various physicochemical properties of drug candidate molecules.[36,37]

In order to implement the present QM-3D ESP descriptor, the adsorbate molecules in the present data set should be structurally aligned first. Such a structural alignment is nontrivial, however, especially in the case where molecules have a wide range of *molecular weights* (MWs).[38] It is necessary to select a representative molecule that serves as a *template* to align with the rest of molecules. Considering the wide range of MWs of the present adsorbates, we categorize the adsorbates into two subgroups with different ranges of MW: each subset includes half of the molecules (85) with an MW range of 110.1−238.2 or 238.2−331.3 amu (Table 1). As usual in the prior machine-learning (ML) studies,[39,40] 80 and 20% of molecules in each subset are used for the train and test set of the present QSPR model, respectively.

Table 1. Two Molecular Subsets Used To Derive and Validate the Predictive QSPR Model for the Adsorption Energy of a DOPA Derivative

molecular subset	MW range (a.m.u.)	adsorption energy (kcal/mol)	no. of molecules in the train set	no. of molecules in the test set
subset 1	110.1−238.2	12.9−26.3	68	17
subset 2	238.2−331.3	15.4−27.6	68	17

We align molecular structures by putting them in a common rectangular box represented by a cubic grid with a spacing of 0.106 Å. The box is big enough to contain the van der Waals volumes of all of the adsorbate molecules. We add an extra margin of 2.7 Å along each axis of the common box in order to secure sufficient space for the subsequent translational and rotational movements required in the structural alignment (see later).

The initial structure in aligning a molecule is obtained by geometry optimization at the HF/6−31G (d,p) level of theory. The structural alignment of a molecule with respect to the template proceeds by translating and rotating the molecule so that the overlap with the template is maximized. In each subset, the structure of the heaviest molecule is taken to be the

https://doi.org/10.1021/acsomega.3c07208
ACS Omega 2024, 9, 994−1000

template. For a given molecule, 2000 *rotamers* are generated by incremental sampling in the SO (3) rotation group employing the Hopf fibration method.[41] The *charge density distribution* $\rho(x, y, z)$ is calculated for the initial structure of a molecule only, and the ρs of the subsequent rotamers are interpolated from that of the initial structure.

For the jth molecule, the charge density of a rotamer, $\rho_j(x, y,$ and $z)$, is calculated with the HF/6−31G (d,p) level of theory. Each rotamer is translated to give the optimal alignment with respect to stationary template molecule i. Incremental translations are iterated to reach the maximum of the cross-correlation, E_{ij}, defined as

$$E_{ij} = \iiint_V \phi_i(x, y, z)\rho_j(x, y, z)\mathrm{d}V \tag{2}$$

where $\phi_i(x, y, z)$ is the ESP due to the template molecule. E_{ij} is calculated by using the fast Fourier transform algorithm.[42] Among 2000 rotamers, the rotamer giving the highest value of E_{ij} is taken to be optimally aligned and is further used to calculate ESP as the descriptor. This protocol of alignment proves to be useful in deriving an accurate 3D-QSPR model for various molecular properties.[36,37] All of the molecules are kept neutral in calculating E_{ij}, in order to retain their original electronic and structural features.

Figure 2 illustrates the final results for the structural alignments of the molecules in each subset. In both subsets,

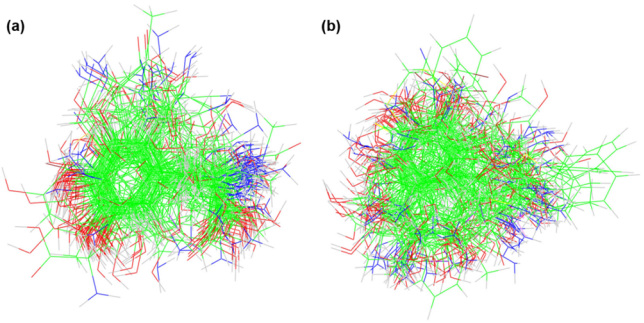


Figure 2. Structurally aligned molecules in subsets 1 (a) and 2 (b). Carbon, hydrogen, nitrogen, and oxygen atoms are colored green, gray, blue, and red, respectively.

the core structures of molecules are overlaid in a similar region, while the side chain moieties point in different directions. The present 3D structural alignment has the advantage of coping with the structurally diverse molecules because the E_{ij} scores are calculated via QM calculations. By contrast, the conventional 3D-QSPR packages employ a simple atom-to-atom matching protocol. We opt for the present QM method because an accurate structural alignment is the most important prerequisite for the predictive capability of a 3D-QSPR model.[43]

Upon completion of the structural alignment of a molecule, the 3D distribution of ESP is derived from its determinantal wave function consisting of n molecular orbitals calculated at the HF/6−31G(d,p) level of theory. Using the molecular wave function, charge density ρ is calculated at the 3D grid points in the common box. The ESP ϕ values at the grid points are calculated from ρ by solving Poisson's equation:

$$\vec{\nabla}^2 \phi(x, y, z) = \rho(x, y, z) \tag{3}$$

The present descriptor is preliminarily constructed as a K-dimensional vector consisting of the ϕ values at the grid points. Because of the huge number of grid points (K = 1,191,016), we reduce the dimensionality for the QSPR modeling. This is carried out by using *principal component analysis* (PCA) widely used to reduce high-dimensional numerical data by extracting the principal components only.[44,45] This projected ESP vector, made of 68 components, is finally used as an input to the 3D-QSPR model (see later). In short, the present descriptor represents the 3D ESP arising from a molecular charge density that is calculated by the first-principles calculations.

2.3. ANN Model for Prediction of the Adsorption Energy. We developed a QSPR model that can predict the adsorption energy of a given molecule by using the 3D ESP as an input descriptor. We employ an ANN algorithm operated in a feed-forward fashion with the backpropagation of an error network.[46] The ANN comprises an input, a hidden, and an output layer, as shown in Figure 3. The ESP vector of a

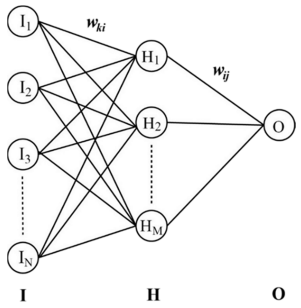


Figure 3. Schematics for the present ANN model for prediction of the adsorption energy of a DOPA derivative. Columns I, H, and O indicate the input (of 68 neurons), hidden (of 40 neurons), and output (of 1 neuron) layers, respectively. In each layer, neurons are fully connected with those in the neighboring layer through the weighting matrices w_{ki} or w_{ij}.

molecule forms 68 neurons in the input layer. All these input neurons ($\hat{I}_k$'s) are fed into the sigmoidal function after multiplying the weighting factors (w_{ki}'s) to produce 40 intermediate neurons ($\hat{H}_i$'s) in the hidden layer, which are in turn processed in the same way to generate a single output neuron ($\hat{O}$):

$$\hat{H}_i = \mathrm{sgm}\left(\sum_{k=1}^{N} w_{ki}\cdot\hat{I}_k\right) \text{ and } \hat{O} = \mathrm{sgm}\left(\sum_{i=1}^{M} w_{ij}\cdot\hat{H}_i\right) \tag{4}$$

where sgm(x) stands for the sigmoidal function $(1 + \mathrm{e}^{-x})^{-1}$. The output neuron is therefore given by

$$\hat{O} = \mathrm{sgm}\left(\sum_{i=1}^{M} w_{ij}\cdot\mathrm{sgm}\left(\sum_{k=1}^{N} w_{ki}\cdot\hat{I}_k\right)\right) \tag{5}$$

The adsorption energies from the DFT calculations serve as the baseline for optimizing the weighting factors to obtain the present 3D-QSPR model. This parameterization is completed with a gradient-based minimization of the error hypersurface F, which is given by the sum of the square differences between the DFT-calculated (D_j) and ANN-predicted (O_j) adsorption energies of N molecules in the training set: $F = \sum_{j=1}^{N}(D_j - O_j)^2$. The vector elements of $\hat{O}$, O_j's, correspond to the predicted adsorption energies of N

https://doi.org/10.1021/acsomega.3c07208
ACS Omega 2024, 9, 994−1000

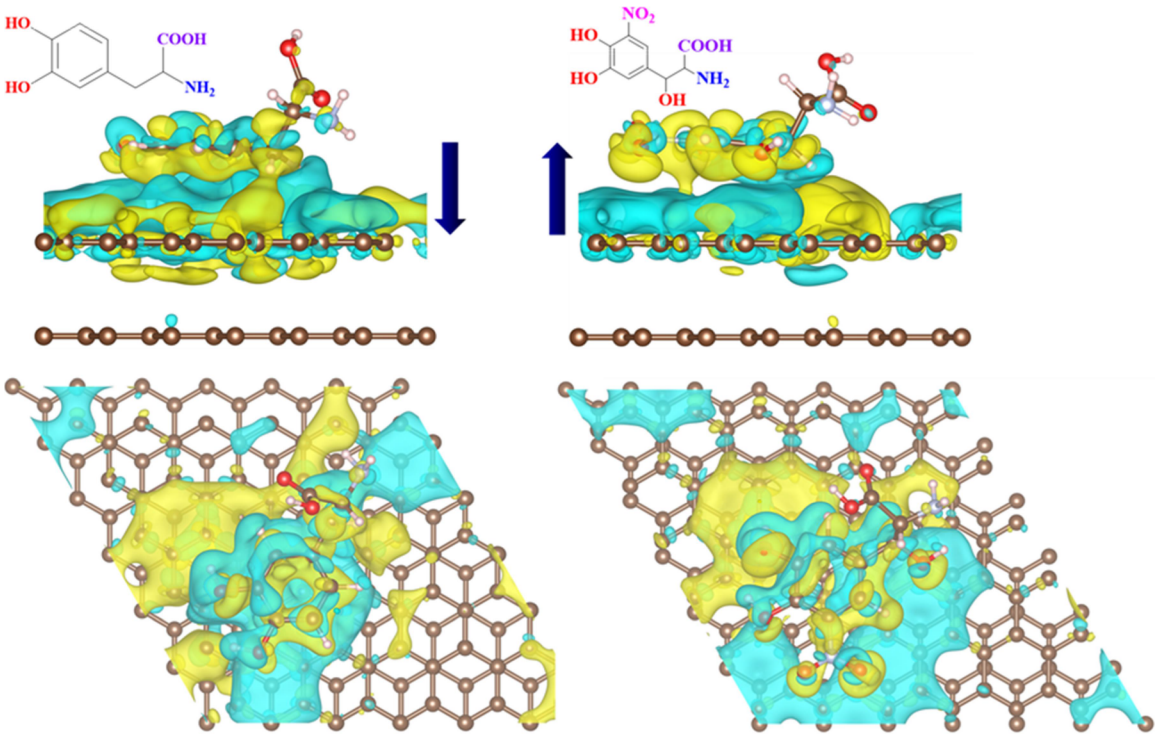


Figure 4. Charge density difference plots for DOPA (left) and Mol104 (right) adsorbed on the graphite surface. Drawn in yellow and cyan are the regions of negative charge (electron) accumulation and depletion, respectively. Side and top views are given in the top and bottom panels, respectively. A blue arrow signifies the direction of electron transfer, as determined by the Bader charge analysis conducted on each molecule before and after adsorption.

molecules in the training set. An F value < 10^{-4} is used as the criterion for the convergence of weighting parameters.

3. RESULTS AND DISCUSSION

The adsorption energies E_{ads}s of the derivatives of DOPA range from 12.9 to 27.6 kcal/mol. Among these molecules, Mol17, Mol41, Mol95, and Mol104 (Supporting Information) have the highest E_{ads}s of 26.13, 24.33, 23.26, and 27.62 kcal/mol, respectively. The Mol104 is doubly substituted with $-NO_2$ and $-OH$ groups. Regardless of the position at the aromatic ring (R_1, R_2, or R_3), the NO_2 substitution gives rise to strong adsorption. The strong adsorption is attributed to the electron-withdrawing capability of $-NO_2$: the $-NO_2$ group pulls the electrons away from the aromatic ring of DOPA, making it more positive. This positive charge makes the DOPA derivative more attracted to the delocalized electron cloud on the graphite surface, enhancing the $\pi-\pi$ interaction between the DOPA derivative (a π-acceptor) and graphite (a π-donor).[47] Additionally, the electron-withdrawing capability of $-NO_2$ activates lateral benzene hydrogens. These activated hydrogens then engage in C—H$\cdots\pi$ interactions with the graphite surface, which enhances the adsorption strength. Another significant contribution to the strong adhesion is likely due to the attraction between the partially positive H atoms of the OH groups and the negative electric charges accumulated at the centers of the hexagonal rings of graphite. We have previously observed the existence of the electrostatic OH$-\pi$ interaction which attracts the OH groups toward the graphite surface.[30] Consequently, both $-$CH and $-$OH groups of the adsorbate are likely to interact with the π electrons on the underlying graphite surface.

The strong adsorption with the $-NO_2$ substitution is further examined by calculating the 3D distribution of the charge density difference $\Delta\rho$ defined as

$$\Delta\rho = \rho_{mol@surf} - \rho_{surf} - \rho_{mol} \quad (6)$$

where $\rho_{mol@surf}$, ρ_{surf}, and ρ_{mol} are the charge densities of the adsorbate@surface, surface, and adsorbate, respectively. Figure 4 illustrates the 3D distribution of $\Delta\rho$ for DOPA and Mol104. A negative charge accumulation (electron gain) and depletion (electron loss) are drawn as yellow and cyan isosurfaces, respectively. Interestingly, DOPA and Mol104 show opposite $\Delta\rho$s: a negative charge accumulates on the graphite surface with DOPA but on the molecule with Mol104.

In order to quantify the charge transfer at the interface, we conduct the Bader charge analysis[48] using the method developed by Henkelman et al.[49] The DOPA molecule has a net change in charge of $-0.0042|e|$ with adsorption, representing an electron transfer from the molecule to the surface. By contrast, Mol104 shows a net change in charge of $0.07421|e|$ upon adsorption, signifying an electron gain from the surface. Note that the amount of charge transfer for Mol104 is 17 times greater than that for DOPA. Therefore, a relatively large amount of electron transfer from the surface presumably gives rise to the strong adhesion found for Mol104. The electron density distribution in frontier molecular orbitals (FMOs) and molecular electrostatic potential surfaces (MEPS) of DOPA and Mol104 are shown in Figure S2.

In addition to the molecular insights into the adhesion, it is desired to predict the adsorption energy for a given molecule. This is achieved by using the QSPR model described above. The predictive capabilities of the present 3D-QSPR models are validated by checking the correlation between the DFT data

https://doi.org/10.1021/acsomega.3c07208
ACS Omega 2024, 9, 994−1000

(E_{ads}s) and the QSPR-predicted ones. The *squared Pearson correlation coefficients* of the training (R^2_{train}) and test (R^2_{test}) sets serve as yardsticks to measure the convergence and accuracy of the prediction models. These two statistical metric parameters are expressed as

$$R^2_{train} = 1 - \frac{\sum_{i=1}^{train}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{train}(y_i - \overline{y}_{train})^2} \text{ and } R^2_{test} = 1 - \frac{\sum_{i=1}^{test}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{test}(y_i - \overline{y}_{test})^2} \quad (7)$$

where $\overline{y}$ is the mean of the adsorption energies from DFT, while y_i and $\hat{y}_i$ are the adsorption energies of molecule i from DFT and QSPR models, respectively. The summations in R^2_{train} and R^2_{test} run over the molecules in the data set.

Figure 5 displays the linear correlation diagrams for the adsorption energies from DFT versus those from the QSPR

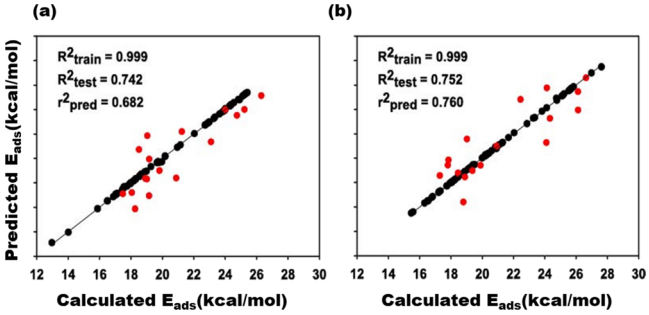


Figure 5. Linear correlation diagram between the DFT-calculated and QSPR-predicted adsorption energies for subsets 1 (a) and 2 (b). Indicated as black and red circles are the molecules in the training and test sets, respectively.

model. The R^2_{train} values close to 1 for both subsets confirm that the present QSPR model is optimized successfully. This implies that the parameterization of the ANN model converges well without regard to the MW range in the training set. By contrast, the R^2_{test} parameters significantly decrease to 0.742 and 0.752 for subsets 1 and 2, respectively. These relatively low R^2_{test} values in both subsets may be attributed to the wide range of adsorption energies: the difference between the maximum and minimum adsorption energies amounts to 65 and 59% of the averages in subsets 1 and 2, respectively. Such a large variation in the adsorption energy in turn gives an overtraining in the optimization of 3 of the D-QSPR model as exemplified by a significant difference (0.25 on average) between R^2_{train} and R^2_{test} values. However, the predictive capability is likely to be enhanced by augmenting the present data set with more molecules.

To further assess the performance of the present QSPR model, we also measure the *external predictivity parameter* (r^2_{pred}) widely used to quantify the predictive capability of a QSPR model.[50,51] This parameter is expressed as

$$r^2_{pred} = 1 - \frac{\sum_{i=1}^{test}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{test}(y_i - \overline{y}_{train})^2} \quad (8)$$

where y_i and $\hat{y}_i$ denote the DFT and QSPR data in the test set, respectively, while $\overline{y}_{train}$ is the mean value of the data for the training set. The r^2_{pred} parameter would be meritorious over the R^2_{test} value because the data in the training set are also reflected in assessing a prediction model as well as those in the test set. As shown in Figure 5, the r^2_{pred} parameters for subsets 1 and 2 amount to 0.682 and 0.760, respectively. Both the r^2_{pred} values exceed the threshold (0.6) for the qualification as a statistical prediction model.[50] This supports the reliability of the QSPR model. It is also noteworthy that the difference between the r^2_{pred} and R^2_{test} values is negligible in subset 2, implying that the training and test sets are prepared reasonably well for subset 2 which comprises relatively heavy molecules. On the other hand, the predictive capability seems to be influenced greatly by the selection of training and test sets in subset 1 as indicated in a relatively large difference (0.06) between r^2_{pred} and R^2_{test} (Figure 5). The relatively low predictive capability in subset 1 can be understood by noting its wider MW range than that in subset 2, which is likely to induce increased errors in the 3D structural alignment. In this regard, the present 3D-QSPR model is expected to be more accurate by further subdividing molecules in subset 1 according to MW.

Real adhesives commonly take polymeric forms. Presumably, the polymers of the present adhesive molecules are more effective in adhesion because the polymer chains can establish large contact areas with a surface and can be cross-linked to give a more durable adhesion. The polymers based on the present adhesive molecules might be structurally and chemically different from their constituting monomers. Nevertheless, as seen from the prior success of PDA, the adhesive capability of a monomer should be largely intact in the resulting polymer.

The present study focuses on the adhesion on a nonpolar surface where the adsorbate interacts with the surface through a van der Waals or $\pi-\pi$ interaction. In the case of a polar surface, however, the molecule−surface interaction will be electrostatic or hydrogen bonding in nature. The two hydroxyl groups of DOPA will play an essential role in the electrostatic or hydrogen bonding interaction with a surface.[52,53] It is left as future work to extend the present QSPR model to include hydrophilic surfaces such as metal oxides and silica.

There is room for improvement in the present QSPR model. The statistical accuracy might be improved by adding the molecular structures and E_{ads} values to the present data set. An improved method for reducing the dimensionality of 3D descriptors other than PCA would be helpful for enhancing the predictive capability because PCA is known to be sensitive to gross errors in the data set.[54] A further study is needed to address the possibility of these new techniques.

The present DFT calculations and QSPR models provide insights into the molecular origin of the adsorption. Further experimental studies on the interactions between a DOPA derivative and graphite might be very helpful: for example, scanning tunneling microscopy could be used to image the molecular mechanisms of adsorption for DOPA molecules, to measure the binding energy of DOPA molecules to graphite, and to study the dynamics of DOPA self-assembly on graphite.[55]

The present study focuses on the adsorption of a single molecule. In a real adhesive material, however, many molecules might adhere to a surface simultaneously, for example, through a self-assembly process or in a polymeric form. In that case, intermolecular interactions between adsorbates (monomers) might play an important role in the adhesion, presumably giving rise to an adhesion stronger than that of a single adsorbate.[56] A study of the coadsorption of multiple molecules and the role of intermolecular interaction is left as future work.

https://doi.org/10.1021/acsomega.3c07208
ACS Omega 2024, 9, 994−1000

4. CONCLUSIONS

Motivated by the remarkably water-resistant and universal adhesivity of marine mussels, extensive efforts have been made to synthesize adhesive molecules containing analogues of DOPA, an essential amino acid largely contained in mussel foot proteins. Prior research and applications along this line have focused on polydopamines. In this study, we expand the chemical space of mussel-inspired adhesive by designing diverse derivatives of DOPA. By constructing a data set of 170 derivatives of DOPA with chemically distinct substituents, we find that a strong electron-withdrawing group ($-NO_2$) intensifies the adsorption onto a graphite surface. The strong adsorption with $-NO_2$ substitution is related to the electron transfer at the interface by calculating the 3D distribution of charge density difference plots and Bader charge analysis. We also derived a QSPR model that can predict the adsorption energy of a given molecule. The present QSPR model utilizes an ANN model and the 3D ESP of a molecule as a descriptor. The 3D ESP is accurately calculated using first-principles methods. The present quantum mechanical 3D descriptor contains more accurate information on the molecular structures and properties, compared to prior classical descriptors. The present QSPR model accurately predicts the adsorption energy from DFT simulations and has the potential to be further improved by adding more molecules to the data set. Our findings might serve as fundamental guidelines for developing a novel adhesive derived from DOPA.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c07208.

Molecular structures of 170 derivatives of DOPA, their adsorption energies on graphite surface calculated by using DFT, and electron density distribution in the frontier molecular orbitals (FMOs) and molecular electrostatic potentials (MEPs) of DOPA and Mol104 (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Hwangseo Park − *Department of Bioscience and Biotechnology, Sejong University, Seoul 05006, Republic of Korea;* orcid.org/0000-0001-5806-2472; Phone: +82-2-3408-3766; Email: hspark@sejong.ac.kr

Joonkyung Jang − *Department of Nanoenergy Engineering, Pusan National University, Busan 46241, Republic of Korea;* orcid.org/0000-0001-9028-0605; Phone: +82-51-510-3928; Email: jkjang@pusan.ac.kr

Authors

Srimai Vuppala − *Department of Nanoenergy Engineering, Pusan National University, Busan 46241, Republic of Korea*

Ramesh Kumar Chitumalla − *Department of Nanoenergy Engineering, Pusan National University, Busan 46241, Republic of Korea;* orcid.org/0000-0002-9523-7056

Seyong Choi − *Department of Nanoenergy Engineering, Pusan National University, Busan 46241, Republic of Korea*

Taeho Kim − *Department of Bioscience and Biotechnology, Sejong University, Seoul 05006, Republic of Korea*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c07208

Author Contributions

[§]S.V. and R.K.C. have contributed equally to this work and share the first authorship.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was supported by the National Research Grants funded by the Korean Government (RS-2023-00220748, 2022R1I1A1A01071131, and 2021R1I1A1A01061036). This research was also supported by Korea Planning & Evaluation Institute of Industrial Technology (TO230005, Development of gas hydrate extinguishing agent and fire suppression technology for inaccessible fires).

■ REFERENCES

(1) Waite, J. H.; Andersen, N. H.; Jewhurst, S.; Sun, C. Mussel Adhesion: Finding the Tricks Worth Mimicking. *J. Adhes.* **2005**, *81*, 297−317.

(2) Benedict, C. V.; Waite, J. H. Location and Analysis of Byssal Structural Proteins of Mytilus Edulis. *J. Morphol.* **1986**, *189*, 171−81.

(3) Zhao, H.; Waite, J. H. Linking Adhesive and Structural Proteins in the Attachment Plaque of Mytilus Californianus*. *J. Biol. Chem.* **2006**, *281*, 26150−26158.

(4) Gerhards, M.; Perl, W.; Schumm, S.; Henrichs, U.; Jacoby, C.; Kleinermanns, K. Structure and Vibrations of Catechol and Catechol·H2o(D2o) in the S and S1 State. *The J. Chem. Phys.* **1996**, *104*, 9362−9375.

(5) Waite, J. H. Nature's Underwater Adhesive Specialist. *Int. J. Adhes. Adhes.* **1987**, *7*, 9−14.

(6) Faure, E.; Falentin-Daudre, C.; Jerome, C.; Lyskawa, J.; Fournier, D.; Woisel, P.; Detrembleur, C. Catechols as versatile platforms in polymer chemistry. *Prog. Polym. Sci.* **2013**, *38* (1), 236−270.

(7) Holl, S. M.; Hansen, D.; Waite, J. H.; Schaefer, J. Solid-State Nmr Analysis of Cross-Linking in Mussel Protein Glue. *Arch. Biochem. Biophys.* **1993**, *302* (1), 255−258.

(8) Burzio, L. A.; Waite, J. H. Cross-linking in adhesive quinoproteins: Studies with model decapeptides. *Biochemistry-Us.* **2000**, *39* (36), 11147−11153.

(9) Anderson, T. H.; Yu, J.; Estrada, A.; Hammer, M. U.; Waite, J. H.; Israelachvili, J. N. The Contribution of Dopa to Substrate−Peptide Adhesion and Internal Cohesion of Mussel-Inspired Synthetic Peptide Films. *Adv. Funct. Mater.* **2010**, *20*, 4196−4205.

(10) Lyu, Q. H.; Hsueh, N.; Chai, C. L. L. The Chemistry of Bioinspired Catechol(amine)-Based Coatings. *Acs Biomater. Sci. Eng.* **2019**, *5* (6), 2708−2724.

(11) Lu, Q.; Danner, E.; Waite, J. H.; Israelachvili, J. N.; Zeng, H.; Hwang, D. S. Adhesion of Mussel Foot Proteins to Different Substrate Surfaces. *J. R. Soc. Interface.* **2013**, *10*, No. 20120759.

(12) Hwang, D. S.; Zeng, H.; Lu, Q.; Israelachvili, J.; Waite, J. H. Adhesion Mechanism in a Dopa-Deficient Foot Protein from Green Mussels. *Soft Matter.* **2012**, *8*, 5640−5648.

(13) Lu, Q.; Hwang, D. S.; Liu, Y.; Zeng, H. Molecular Interactions of Mussel Protective Coating Protein, Mcfp-1, from Mytilus Californianus. *Biomaterials.* **2012**, *33*, 1903−11.

(14) Salonen, L. M.; Ellermann, M.; Diederich, F. Aromatic Rings in Chemical and Biological Recognition: Energetics and Structures. *Angew. Chem., Int. Ed. Engl.* **2011**, *50*, 4808−42.

(15) Schlaich, C.; Li, M. J.; Cheng, C.; Donskyi, I. S.; Yu, L. X.; Song, G.; Osorio, E.; Wei, Q.; Haag, R. Mussel-Inspired Polymer-Based Universal Spray Coating for Surface Modification: Fast Fabrication of Antibacterial and Super hydrophobic Surface Coatings. *Adv. Mater. Interfaces.* **2018**, *5* (5), No. 1701254.

(16) Lynge, M. E.; Schattling, P.; Stadler, B. Recent Developments in Poly (Dopamine)-Based Coatings for Biomedical Applications. *Nanomed.* **2015**, *10*, 2725−2742.

(17) Batul, R.; Tamanna, T.; Khaliq, A.; Yu, A. Recent Progress in the Biomedical Applications of Polydopamine Nanostructures. *Biomater. Sci.* **2017**, *5*, 1204−1229.
(18) Cheng, W.; Zeng, X.; Chen, H.; Li, Z.; Zeng, W.; Mei, L.; Zhao, Y. Versatile Polydopamine Platforms: Synthesis and Promising Applications for Surface Modification and Advanced Nanomedicine. *ACS Nano*. **2019**, *13*, 8537−8565.
(19) Flouda, P.; Shah, S. A.; Lagoudas, D. C.; Green, M. J.; Lutkenhaus, J. L. Highly Multifunctional Dopamine-Functionalized Reduced Graphene Oxide Supercapacitors. *Matter*. **2019**, *1*, 1532−1546.
(20) Yang, M.; Choi, J.; Kim, S.-K.; Braun, P. V. Flexible Binder-Free Cus/Polydopamine-Coated Carbon Cloth for High Voltage Supercapacitors. *Energy Technol.* **2018**, *6*, 1852−1858.
(21) Yang, L.; Gu, B.; Chen, Z.; Yue, Y.; Wang, W.; Zhang, H.; Liu, X.; Ren, S.; Yang, W.; Li, Y. Synthetic Biopigment Supercapacitors. *ACS Appl. Mater. Interfaces.* **2019**, *11*, 30360−30367.
(22) Bie, Y.; Yang, J.; Liu, X.; Wang, J.; Nuli, Y.; Lu, W. Polydopamine Wrapping Silicon Cross-Linked with Polyacrylic Acid as High-Performance Anode for Lithium-Ion Batteries. *ACS Appl. Mater. Interfaces.* **2016**, *8*, 2899−2904.
(23) Wu, Z.-S.; Parvez, K.; Li, S.; Yang, S.; Liu, Z.; Liu, S.; Feng, X.; Müllen, K. Alternating Stacked Graphene-Conducting Polymer Compact Films with Ultrahigh Areal and Volumetric Capacitances for High-Energy Micro-Supercapacitors. *Adv. Mater.* **2015**, *27*, 4054−4061.
(24) Becke, A. D. Density-Functional Thermochemistry. Iv. A New Dynamical Correlation Functional and Implications for Exact-Exchange Mixing. *J. Chem. Phys.* **1996**, *104*, 1040−1046.
(25) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H. *Gaussian 16*. Gaussian, Inc.: Wallingford, CT, 2016.
(26) Furthmüller, J.; Hafner, J.; Kresse, G. Ab Initio Calculation of the Structural and Electronic Properties of Carbon and Boron Nitride Using Ultrasoft Pseudopotentials. *Phys. Rev. B Condens. Matter.* **1994**, *50*, 15606−15622.
(27) Kresse, G.; Furthmüller, J. Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Comput. Mater. Sci.* **1996**, *6*, 15−50.
(28) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B Condens. Matter.* **1996**, *54*, 11169−11186.
(29) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. 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.
(30) Chitumalla, R. K.; Kim, K.; Gao, X.; Jang, J. A Density Functional Theory Study on the Underwater Adhesion of Catechol onto a Graphite Surface. *Phys. Chem. Chem. Phys.* **2021**, *23*, 1031−1037.
(31) Novoselov, K. S.; Geim, A. K.; Morozov, S. V.; Jiang, D.; Katsnelson, M. I.; Grigorieva, I. V.; Dubonos, S. V.; Firsov, A. A. Two-Dimensional Gas of Massless Dirac Fermions in Graphene. *Nature.* **2005**, *438*, 197−200.
(32) Taychatanapat, T.; Watanabe, K.; Taniguchi, T.; Jarillo-Herrero, P. Quantum Hall Effect and Landau-Level Crossing of Dirac Fermions in Trilayer Graphene. *Nat. Phys.* **2011**, *7*, 621−625.
(33) Kumar, A.; Escoffier, W.; Poumirol, J. M.; Faugeras, C.; Arovas, D. P.; Fogler, M. M.; Guinea, F.; Roche, S.; Goiran, M.; Raquet, B. Integer Quantum Hall Effect in Trilayer Graphene. *Phys. Rev. Lett.* **2011**, *107*, No. 126806.
(34) Hasegawa, M.; Nishidate, K. Semi-empirical Approach to the Energetics of Interlayer Binding in Graphite. *Phys. Rev. B.* **2004**, *70*, No. 205431.
(35) Tan, P. H.; Han, W. P.; Zhao, W. J.; Wu, Z. H.; Chang, K.; Wang, H.; Wang, Y. F.; Bonini, N.; Marzari, N.; Pugno, N.; et al. The Shear Mode of Multilayer Graphene. *Nat. Mater.* **2012**, *11*, 294−300.
(36) Choi, H.; Kang, H.; Chung, K. C.; Park, H. Development and application of a comprehensive machine learning program for predicting molecular biochemical and pharmaceutical properties. *Phys. Chem. Chem. Phys.* **2019**, *21*, 5189−5199.
(37) Kim, T.; You, B. H.; Han, S.; Shin, H. C.; Chung, K. C.; Park, H. Quantum artificial neural network approach to derive a highly predictive 3D-QSAR model for blood-brain barrier passage. *Int. J. Mol. Sci.* **2021**, *22*, 10995.
(38) Cramer, R. D.; Wendt, B. Template CoMFA: The 3D-QSAR grail? *J. Chem. Inf. Model.* **2014**, *54*, 660−671.
(39) Song, M.; Clark, M. Development and evaluation of an *in silico* model for hERG binding. *J. Chem. Inf. Model.* **2006**, *46*, 392−400.
(40) Liu, R.; Sun, H.; So, S. S. Development of quantitative structure−property relationship models for early ADME evaluation in drug discovery. 2. Blood-brain barrier penetration. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1623−1632.
(41) Yershova, A.; Jain, S.; Lavalle, S. M.; Mitchell, J. C. Generating uniform incremental grids on SO (3) using the Hopf fibration. *Int. J. Robot. Res.* **2010**, *29*, 801−812.
(42) Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S. PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins* **2006**, *65*, 392−406.
(43) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* **2014**, *57*, 4977−5010.
(44) Buslaev, P.; Gordeliy, V.; Grudinin, S.; Gushchin, I. Principal component analysis of lipid molecule conformational changes in molecular dynamics simulations. *J. Chem. Theory Comput.* **2016**, *12*, 1019−1028.
(45) Maisuradze, G. G.; Liwo, A.; Scheraga, H. A. Principal component analysis for protein folding dynamics. *J. Mol. Biol.* **2009**, *385*, 312−329.
(46) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533−536.
(47) Roy, J.; Ghosh, S.; Ojha, P. K.; Roy, K. Predictive Quantitative Structure−Property Relationship (Qspr) Modeling for Adsorption of Organic Pollutants by Carbon Nanotubes (Cnts). *Environ. Sci. Nano.* **2019**, *6*, 224−247.
(48) Bader, R. F. W. A Quantum Theory of Molecular Structure and Its Applications. *Chem. Rev.* **1991**, *91*, 893−928.
(49) Henkelman, G.; Arnaldsson, A.; Jónsson, H. A Fast and Robust Algorithm for Bader Decomposition of Charge Density. *Comput. Mater. Sci.* **2006**, *36*, 354−360.
(50) Tropsha, A.; Gramatica, P.; Gombar, V. K. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* **2003**, *22*, 69−77.
(51) Gramatica, P.; Sangion, A. A historical excursus on the statistical validation parameters for QSAR models: A clarification concerning metrics and terminology. *J. Chem. Inf. Model.* **2016**, *56*, 1127−1131.
(52) Mian, S. A.; Saha, L. C.; Jang, J.; Wang, L.; Gao, X.; Nagase, S. Density Functional Theory Study of Catechol Adhesion on Silica Surfaces. *J. Phys. Chem. C* **2010**, *114*, 20793−20800.
(53) Mian, S. A.; Gao, X.; Nagase, S.; Jang, J. Adsorption of Catechol on a Wet Silica Surface: Density Functional Theory Study. *Theor. Chem. Acc.* **2011**, *130*, 333−339.
(54) Jolliffe, I. T.; Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. A Math. Phys. Eng. Sci.* **2016**, *374*, No. 20150202.
(55) Fisher, A. J.; Blöchl, P. E. Adsorption and scanning-tunneling-microscope imaging of benzene on graphite and MoS_2. *Phys. Rev. Lett.* **1993**, *70*, 3263−3266.
(56) Matin, M. A.; Chitumalla, R. K.; Lim, M.; Gao, X.; Jang, J. Density Functional Theory Study on the Cross-Linking of Mussel Adhesive Proteins. *J. Phys. Chem. B* **2015**, *119*, 5496−5504.

https://doi.org/10.1021/acsomega.3c07208
ACS Omega 2024, 9, 994−1000