ADVANCED SENSOR RESEARCH Open Access
www.advsensorres.com

Empirical Modification of Force Fields for the Development of Peptide-Based Gas Sensors

Thuc Anh Ngo, Tanju Yildirim, Meng-Qun Feng, Kosuke Minami, Kota Shiba, and Genki Yoshikawa**

Molecular dynamics models combined with computational approaches can be used as advanced screening techniques for finding highly efficient material-molecule interactions based on binding affinity, including in the development of gas sensors. However, most models are originally designed for liquid phase interactions, which do not align with gas sensing conditions, resulting in lower-than-expected performance. This study introduces an empirical modification method to adjust peptide interaction models for a gas phase, aiming to better accommodate the interaction between pentapeptides and target gas molecules. By adapting the weights of terms in the Gibbs free energy equation given in an empirical force field model, we demonstrate a significant increase in the absolute value of coefficient of determination (R_0^2) , from an average of 0.05 with conventional liquid phase models to 0.90 with proposed gas phase models. An empirical modification technique for gas phase interactions markedly enhances the prediction accuracy of models, facilitating the effective development of peptide-based gas sensors.

1. Introduction

One of the most important characteristics of a gas sensor is selectivity toward target molecules.[1,2] Among various materials, oligopeptides derived from the specific binding region of olfactory receptors have gained significant interest as a receptor material for gas sensors, which is due to their important role in attaining gas-specific recognition in the olfactory system.[3] To design and find effective oligopeptides with specific selectivity, the evaluation of binding affinity between a peptide and a target gas molecule is crucial.[4,5] Molecular docking is a computational technique aimed at predicting the affinity between a receptor and a ligand. Molecular docking has been applied to designing peptides as a material for gas sensors.[5–11] There are two major approaches to the engineering of pentapeptides; *i*) based on the interaction site of the olfactory proteins; and *ii*) comprehensive screening of short oligopeptides. For the former approach, oligopeptides with lengths ranging from 5 to 15 amino acids have been reported to recognize some specific gases such as trimethylamine, acetic acid, and butyric acid.[6,7,12–15] This method, however, has the limitation of finding actual interaction sites of the proteins due to the lack of accurate 3D protein structures. In contrast, the latter approach has been widely used for designing oligopeptides toward a wide range of gas species based on simulated binding affinities.[5,16,17] The binding affinity calculated based on Gibbs free energy (ΔG) between oligopeptides and gases has been utilized as a guideline for designing peptide-based materials for olfactory sensors.[17] Three tripeptides (WWW, WAW, and WHW) containing aromatic rings with the highest simulated binding affinities for aromatic compounds showed good performance in detecting explosive gases with aromatic structures. The limited accuracy toward some gases such as alcohols remains challenging using this approach.[8] Thus, further implementation of molecular docking to achieve better accuracy in predicting the interaction between pentapeptides and gases requires highly accurate docking calculations.

Several models exist for evaluating intermolecular energy landscapes in gas-peptide interactions, including the geometric model, first-principles calculations, and empirical force fields.[4,18] The geometric method using the GRAMM (Global Range Molecular Matching) docking program identified ligand-binding residues with the greatest exothermic interactions with chosen gas ligands.[6,7,18–20] However, the method requires multiple steps and biophysics knowledge for modeling structures and deriving peptide sequences based on interaction

T. A. Ngo, M.-Q. Feng, K. Minami, K. Shiba, G. Yoshikawa
Research Center for Macromolecules and Biomaterials
National Institute for Materials Science (NIMS)
1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan
E-mail: ngo.thucanh@nims.go.jp; yoshikawa.genki@nims.go.jp
T. A. Ngo, M.-Q. Feng, G. Yoshikawa
Materials Science and Engineering
Graduate School of Pure and Applied Science
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan
T. Yildirim
Faculty of Science and Engineering
Southern Cross University
East Lismore, NSW 2480, Australia

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adsr.202400122

DOI: 10.1002/adsr.202400122

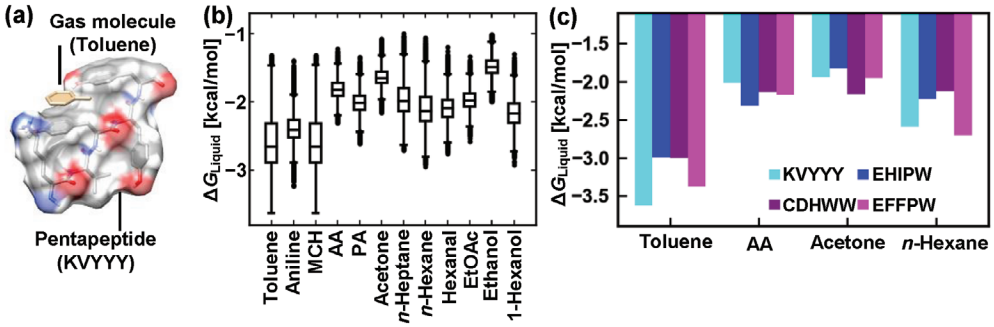

Figure 1. Evaluation of binding affinity between pentapeptide library and 12 gas molecules with AutoDock Vina. a) Typical working space between a pentapeptide and a gas molecule to compute binding energy (ΔG). b) Screening result of pentapeptide library (1.6 million candidates) toward different gas molecules using AutoDock Vina developed with liquid phase data shown using ΔG. It is noteworthy that the lower ΔG is, the higher the frequency shift (Δf) expected in QCM experiments. c) Predicted liquid phase binding energies of four selected pentapeptides (KVYYY, CDHWW, EHIPW, and EFFPW) toward four target gas molecules: toluene, acetic acid (AA), acetone, and *n*-hexane.

sites. The first-principles method is another way to evaluate the intermolecular energies between oligopeptides and target gases.[11,17] Even though the accuracy of the method is high, extensive computational work limits the length of a peptide in one simulation. Unlike physics-based models, empirical force fields rely on machine learning algorithms to calibrate experimental data and import results into simulation models, improve gas-peptide prediction accuracy, and reduce computational complexity. Empirical force fields such as *chemgauss* have been applied to design oligopeptides for specific biomarkers.[5,8] Peptides selected from the screening with *chemgauss4* show a good agreement between binding affinities and experiment values for various gases except for alcohols. AutoDock models are widely known empirical models in designing peptide-based materials for gas sensors.[9,10,17,21,22] Whilst the empirical model showed high performance in some gases that have aromatic structures, the model lacks accuracy for other gas species such as aldehydes.[9] Limitations of empirical models may be attributed to their original calibration being based purely on liquid phase data. Herein, we addressed this issue with gas phase data together with optimizing AutoDock Vina model.

The aim of this work is to adapt the empirical force fields by modifying the weights in the model of each gas based on gas-solid phase data for better correlations with sensing signals. Four pentapeptides, namely KVYYY, CDHWW, EIHPW, and EFFPW, were initially selected via the liquid phase model and used as reference data for evaluating the gas-peptide binding affinity via measuring the mass of adsorbed gas molecules on pentapeptides using a Quartz Crystal Microbalance (QCM). First, we adapted the empirical force fields considered in AutoDock Vina model for gas phase binding affinity by modifying the weights of the empirical force fields with the genetic algorithm (GA), which increases the predictability of the model from no correlation to strong correlation with the experimental results. Further improvement of the weights using the stepwise algorithm (SA) shows high linearity between the predicted normalized binding affinity and the sensor output signal. This adaptive approach demonstrates significant potential for enhancing the precision of empirical force fields for gas phase measurements, paving the way for more accurate and reliable applications of gas sensors in various fields.

2. Results and Discussion

2.1. Virtual Screening Results

The virtual screening was conducted to evaluate the binding properties of all possible pentapeptides bound to 12 different gases using the liquid phase AutoDock Vina model (**Figure 1a**). These gases include aromatic hydrocarbons, amines, acids, alkanes, ketones, aldehydes, esters, and alcohols. These chosen gases exhibit diverse physicochemical properties including varying molecular weights, solubility, and polarity. The first step is finding the most stable state of interaction in which a peptide and a gas molecule are placed in close contact inside a unit cell to compute the lowest binding energy. It should be noted that the lower binding affinity of the complex corresponds to the stronger interaction of the peptide and gas. Thus, the values of the binding affinity and the sensor frequency shift with the increased gas molecule absorption on peptide material are negatively correlated. The overall binding affinities of the pentapeptide library toward the 12 gases are diverse depending on the physicochemical characteristics of both peptides and gas molecules. For example, the gases containing phenyl groups, such as toluene and aniline, exhibit strong affinities toward pentapeptides because of π–π stacking (Figure 1b).[17,23] For the experimental verification, we chose four pentapeptides consisting of KVYYY, CDHWW, EHIPW, and EFFPW because of their lowest ΔG to toluene, acetic acid, acetone, and *n*-hexane, respectively (Figure 1c). To assess the overfitting and ensure the accuracy of the prediction, two pentapeptides with similarities in structural properties (DFIPW and RTYYY) were selected as test samples.

2.2. Liquid Phase Model Performance Assessment with Evaluation Metrics

To evaluate the performance of the phase model, a gas sensing array comprising of the six pentapeptides was tested with all

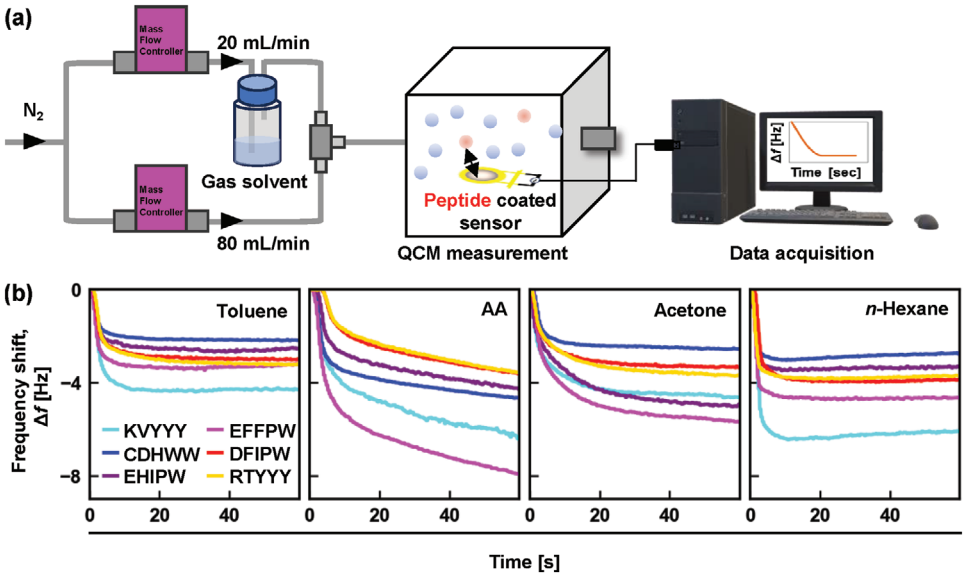

Figure 2. QCM responses of four selected pentapeptides based on comprehensive screening. a) Experiment setup b) QCM response of four pentapeptides toward toluene, acetic acid (AA), acetone, and *n*-hexane.

12 gases, including toluene, aniline, methylcyclohexane (MCH), acetic acid (AA), propionic acid (PA), acetone, *n*-hexane, *n*-heptane, hexanal, ethyl acetate (EtOAc), ethanol, and 1-hexanol (**Figure 2**a). The responses of each peptide-coated QCM sensor to four representative gases are shown in Figure 2b. The average absolute frequency shift ($|\Delta f|$) of replicate experiments is assumed as an experimental indicator for the affinity between the peptides and the gases (Figure S1, Supporting Information).

To assess the accuracy of the liquid phase model in predicting the gas response of the peptide-coated sensors with binding affinities, we compare the normalized Gibbs free energy (ΔG_{norm}) with the absolute change in frequency ($|\Delta f|$) in terms of Pearson's correlation coefficient (R), Spearman's rank correlation coefficient (ρ), and the coefficient of determination through the origin (R_0^2) as described in the Experimental Section. The prediction results based on the liquid phase AutoDock Vina agree with the experimental results in the case of toluene with an R value of −0.9. This finding aligns with the previous screening results of triple pentapeptides, where aromatic amino acids were preferred for enhancing binding affinities with aromatic ring-containing molecules.[17] It is, however, less effective for other gases, particularly acids, alcohols, and ketones (**Figure 3**; Figures S1 and S2, Supporting Information). The inverse relationship between the binding affinities and the response of peptide-coated sensors with the positive values of R clearly illustrates the limitation of the liquid phase model in the cases of acetic acid, propionic acid, and ethanol. The earlier attempt to use the empirical force fields for choosing peptides as the sensor receptor also resulted in a poor performance for short alkyl chain alcohols.[21] The inconsistency between the predicted binding affinities obtained from the liquid phase model and the experimental results stems from the weights that were empirically determined solely based on the liquid phase data. To improve the efficiency of this liquid phase AutoDock Vina model for the gas-solid phase, it is necessary to scale the physical terms with appropriate weights using gas phase data.

2.3. Concept Demonstration with GA

We demonstrate the approach to the development of the gas phase model by scaling each physical term with experimental data and multiplying it with optimal weights. The gas phase model was developed for each gas, in which all pentapeptides were exposed to the same vapors with the same conditions, resulting in 12 corresponding models. GA is employed to adjust weights to achieve optimal agreement in the trend between the binding affinities and experimental results. In this modified approach, a group of a certain number of different solutions with six weight values each was initialized. In this study, a group of twenty solutions was employed to balance the sufficient diversity of each group and the feasible computation. Each weight value was randomly set between 0 and 1. After the molecular dynamic simulation, each solution was evaluated to have a score (□) considering R_i, ρ_i, and R_{0i}^2 for determining the performance of the model with the updated weights. New solutions were created by crossing over the weight values from their parent sets. To keep solutions diverse and avoid the local minimum, some randomness was introduced by changing some values in these new solutions. New sets were ranked based on how well they performed in the S_i for generating the next groups. This cycle of selection, crossover, introducing randomness, and evaluation was repeated for 1000 cycles. The optimal weights entail iteratively decreasing the score (S_i) in Equation (4), with the optimal weights emerging from the most successful weights of the preceding generation. The final weights for all 12 gases after 1000 generations with the minimum S_i are shown in **Table 1**.

By applying GA to the weight optimization process, good R and ρ are quickly achieved with average values of −0.83 and −0.87,

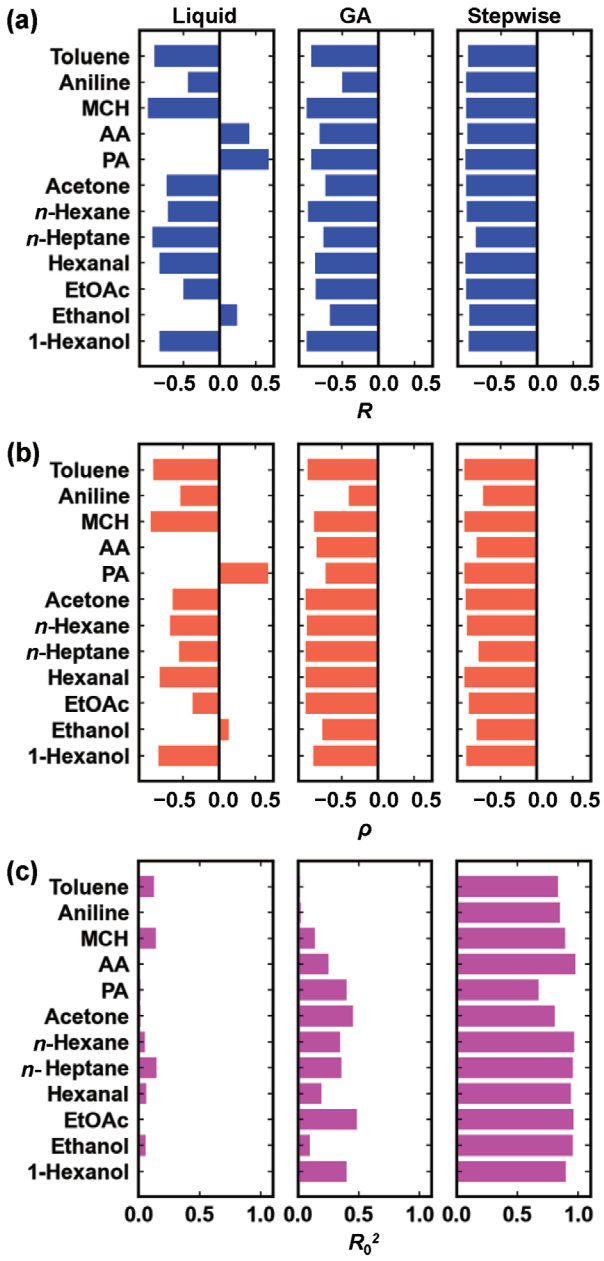

Figure 3. Performance comparisons of the AutoDock Vina, developed using liquid phase data and gas phase data, with optimal weights obtained with GA and SA. a) Pearson's correlation coefficient (R_i) implies the linearity between ΔG and $|\Delta f|$. A value closer to −1 indicates high linearity. b) Spearman's rank correlation coefficient (ρ_i) implies the correct ranking predicted binding affinity of one gas molecule to four different peptides. A value closer to −1 indicates high selectivity. c) Coefficient of determination through the origin (R_{0i}^2) implies the linearity between ΔG and $|\Delta f|$ considering gas-solid phase interaction. An R_0^2 value closer to 1 indicates greater linearity with zero-intercept models.

respectively (**Figures** 3a,b and 4). In the cases of acetic acid, propionic acid, and ethanol, all of which previously show the inverse relationships between the binding affinities and the experimental results, Pearson's correlation coefficients increase considerably, reaching almost −1 for acetic acid (Figure 3a,b). Considering the

Table 1. GA optimized weights for 12 gas molecules.

Gas molecules	w_{g1}	w_{g2}	$w_{rep.}$	w_{HP}	w_{HB}	$w_{rot.}$
Toluene	−0.4	−0.2	1.8	−0.8	−0.8	−2.7
Aniline	0.0	−0.4	−0.2	−1.0	−0.4	−0.8
MCH	−0.4	0.0	−1.9	−0.7	−0.4	7.3
AA	−1.0	−0.2	0.7	−0.1	−0.2	−1.6
PA	−0.6	−0.2	−0.8	−0.4	−0.1	0.1
Acetone	−0.4	−0.3	5.0	−0.4	0.0	0.9
n-Hexane	0.0	0.0	−0.2	−1.0	−0.4	−1.0
n-Heptane	−0.9	−0.6	−0.2	−0.6	−0.4	−0.9
Hexanal	−0.1	−0.2	−0.7	−0.8	−0.2	−0.4
EtOAc	−1.0	−0.5	−1.3	−0.8	0.0	−0.2
Ethanol	−0.9	−0.6	−0.7	−0.7	−0.8	−0.5
1-Hexanol	−0.7	−0.1	−0.3	−0.3	−0.2	4.9

R_0^2, a notable increase with the GA optimized gas phase models is observed, particularly for EtOAc and 1-hexanol. The R_0^2 values, however, remain low for other gases, such as toluene, aniline, and ethanol. GA optimization method only provides R_0^2 values with less than 0.5 (Figure 3c; Figure S3, Supporting Information). This limitation of GA-optimized models is possibly due to the alterations in specific weight values within the GA functions, influencing the overall performance (Figure S4, Supporting Information). The GA optimization method mostly modifies the rotation and repulsion weights as shown in Figures S5–S8, Supporting Information. A correlation test among weight values across 1000 generations and S_i, however, suggests that variations in weight values of rotation and repulsion terms do not significantly influence the overall performance of the model in some instances (Figure S4, Supporting Information). For example, in the case of MCH, the weight value of the $gauss_1$ term largely affects the improvement of S_i but is not modified intensively further in the optimization process. Since $gauss_1$ represents the short-range repulsive interactions, the results imply that two Gaussian terms may be critical components of steric interactions between gas molecules and peptides in the gas phase, affecting sensor performance. The strength of intermolecular forces—including both steric and electrostatic interactions—is largely influenced by the electron distribution within the molecules. Charge density plays a crucial role in determining this electron distribution, which leads to variations in the strength of the steric interactions. A previous study involving first-principles molecular modeling of sensing material selection demonstrated that changes in charge density are correlated with the sensor's response to different gases.[21] Therefore, we decided to comprehensively analyse all the changing directions of the weights including the Gaussian terms using the SA method.

2.4. Further Investigation with SA Optimization Method for all Weight Changing Directions

SA optimization method systematically explores a wide range of weight values from −1.0 to 0.0 with a step increment of 0.2. Each solution was evaluated to have the S_i as explained in the previous section. All combinations were ranked according to S_i, and

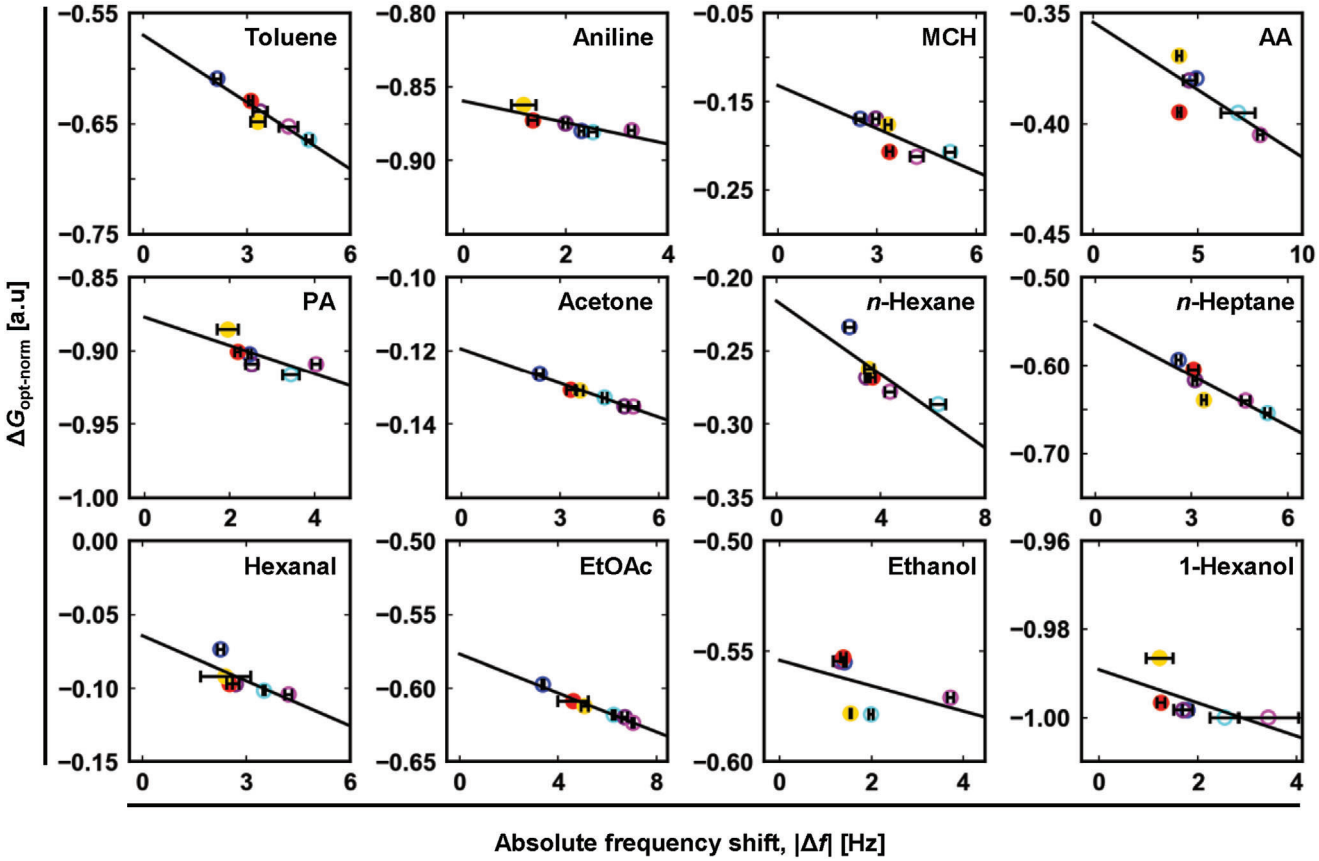

Figure 4. Parity plots of absolute frequency shift versus the normalized optimized binding energy using GA. The binding energy is normalized to its maximum value. The training dataset, which includes four pentapeptides (KVYYY, cyan; CDHWW, blue; EHIPW, purple; and EFFPW, magenta), is shown using open circles. The validating dataset, which includes pentapeptides DFIPW (red) and RTYYY (yellow), is shown with filled circles. Circles represent the average values calculated from three measurements, with error bars indicating the corresponding standard deviations. Black lines show the results of a linear regression.

the best set of the six weight values was determined. The final weights for all 12 gases after 46656 (= 6^6) iterations are shown in **Table 2**. As a result, the multidirectional changes in the weights derived from the SA optimization method exhibit substantial enhancements in the overall model performances.

Table 2. SA optimized weights for 12 gas molecules.

Gas molecules	w_{g1}	w_{g2}	$w_{rep.}$	w_{HP}	w_{HB}	$w_{rot.}$
Toluene	−0.2	−1.0	−0.4	−0.6	−0.6	−0.2
Aniline	−1.0	−0.8	−1.0	−0.6	−1.0	−1.0
MCH	−0.4	−1.0	−0.8	0.0	−0.2	−0.4
AA	−0.2	−0.4	0.0	0.0	−1.0	−0.2
PA	0.0	−0.8	−1.0	−1.0	−1.0	0.0
Acetone	−0.6	−0.8	−0.4	−0.4	−1.0	−0.6
n-Hexane	−0.8	−0.2	−1.0	−1.0	−1.0	−0.8
n-Heptane	−1.0	−0.4	−0.4	−1.0	−1.0	−1.0
Hexanal	−0.2	−0.6	−0.2	0.0	−1.0	−0.2
EtOAc	0.0	−0.2	−0.8	−0.6	−1.0	0.0
Ethanol	−0.2	−0.8	−0.2	−0.8	−1.0	−0.2
1-Hexanol	−1.0	−0.4	−0.4	−1.0	−1.0	−1.0

In comparison to the liquid phase model and the GA-optimized gas model, the SA-optimized gas models greatly improve R and ρ (**Figures** 3a,b and 5). An average R value of −0.97 demonstrates the adequate ability of the models to predict the peptide behavior with gases. Furthermore, the optimized weights from the SA method resolve the issue of assessing polar gases using the AutoDock Vina model, especially in the case of aniline. For instance, the dominant effect of π–π stacking in the liquid phase and GA optimized gas phase models results in the same best affinities toward KVYYY for both toluene and aniline despite the difference in the polarity of the gas molecules. The SA optimized gas phase model accurately predicts EFFPW as the most affinitive peptide toward aniline as observed in the experiments. The SA optimized gas phase models further achieve the R_0^2 values of above 0.9 in many gases (Figure 3c). The improvement of the SA optimized gas phase models in discriminating peptides that gave similar gas responses can be attributed to the consideration of R_0^2 in the weight optimization process. Furthermore, the SA optimized models can accurately predict which pentapeptide has a higher response than the other one for all 12 gases in the validating peptide set (DFIPW and RTYYY), whilst the liquid phase model can predict only for toluene, acetic acid, propionic acid, and *n*-heptane. The success of the SA optimized model can

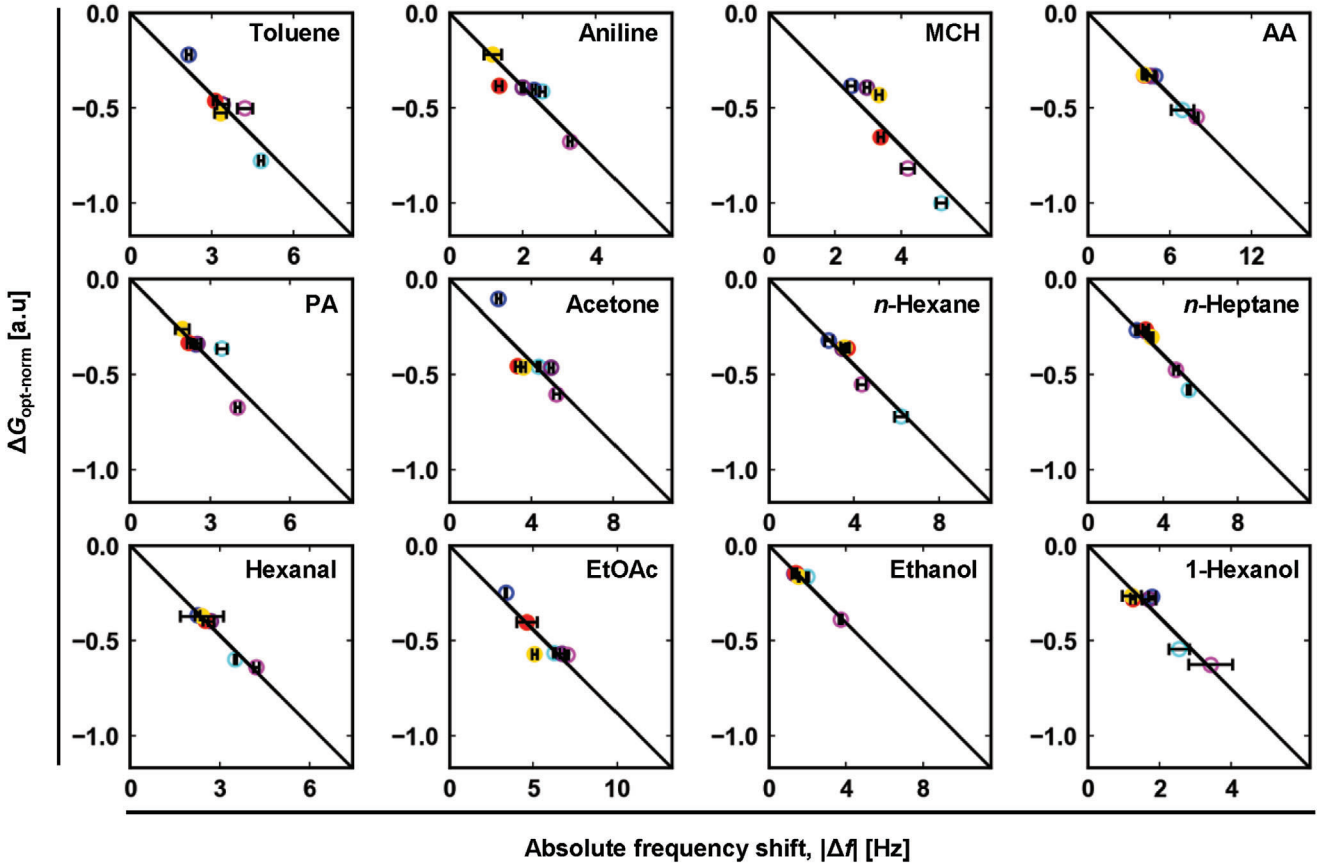

Figure 5. Parity plots of absolute frequency shift versus the normalization values of optimized binding energy SA. The binding energy is normalized to its maximum value. The training dataset, which includes four pentapeptides (KVYYY, cyan; CDHWW, blue; EHIPW, purple; and EFFPW, magenta), is shown using open circles. The validating dataset, which includes pentapeptides DFIPW (red) and RTYYY (yellow), is shown with filled circles. Circles represent the average values calculated from three measurements, with error bars indicating the corresponding standard deviations. Black lines show the results of a linear regression.

be attributed to the improvement of two Gaussian terms (w_{g1} and w_{g2}) describing the steric interaction and the hydrogen bonding terms. All models increase the effects of hydrogen bonding by decreasing the value of the hydrogen bonding weight, in most cases to the minimum value of −1. Overall, the implementation of the SA optimization method improves all three metrics, leading to the best agreement in trends between binding affinities and experimental results.

3. Conclusion

This study introduces empirical modification techniques for the adaptation of the AutoDock Vina model based on gas phase data, significantly enhancing the accuracy of predicting the gas responses of pentapeptides coated on QCM sensors. We also propose the coefficient of determination through the origin (R_0^2) as an evaluation metric to support the weight optimization process with respect to the nature of the gas sensing phenomenon. Despite the significant improvements made by tuning weights, the gas phase model still falls short in predicting absolute values of ΔG and the behavior of aniline and ethanol for peptide-coated sensors that have similarities in the gas responses. These limitations of the current gas phase AutoDock Vina models could be attributed to the constraints of their ability to capture the interactions with small molecules, such as gas molecules. These problems could be addressed with further comprehensive optimization of the model by modifying the energetic terms based on the parameters involved in addition to the tuning of their weight values. By addressing these factors, future models may be able to predict molecular interactions more accurately in gas phase environments.

Our work has improved the ability of the AutoDock Vina model to predict gas response of peptide-coated sensors by addressing gas phase limitations and optimizing weight values. This study paves the way for the efficient application of empirical force fields, as in this case, AutoDock Vina, in engineering peptide materials with the desired selectivity toward gas molecules for the further development of gas sensors.

4. Experimental Section

Material Preparations: Synthetic pentapeptides (KVYYY, CDHWW, EHIPW, EFFPW, RTYYY, and DFIPW) were purchased from Cosmo Bio Co., Ltd. with 80% purity. Toluene, aniline, methylcyclohexane (MCH), acetic acid (AA), propionic acid (PA), acetone, *n*-hexane, *n*-heptane, hex-

anal, ethyl acetate (EtOAc), ethanol, and 1-hexanol were purchased from Sigma–Aldrich, Fujifilm Wako Pure Chemical Corporation, and Kanto Chemical Co., Ltd. MilliQ water (Millipore) was used for preparing the peptide stock solution.

QCM measurements: An AT-cut quartz crystal resonator (QA-A9M-AU(M), SEIKO EG&G Co., Ltd.) was used as the sensing element in which pentapeptides can be coated onto the top surface. Each peptide (1 mg) was dissolved in 1 mL of distilled water to make a stock solution. To deposit each peptide on the QCM, the QCM was cleaned with O_2 plasma and immersed in the peptide solution for 48 h at room temperature. The peptide-coated QCMs were stored at room temperature in the dark until use.

The QCMs were placed in a Teflon chamber, which was connected to two gas lines: an inlet and an outlet. The inlet was connected to a gas system, which consisted of two mass flow controllers, a mixing chamber, a purging line, and a sampling line with a vial (40 mL) containing a solvent liquid (5 mL). The sensing setup was placed in an incubator and the temperature was maintained at 25 °C. Before measuring the QCM output, pure nitrogen gas was introduced into the QCM chamber at the flow rate of 50 mL min^{-1} for 5 min to remove any contaminants. Nitrogen gas, which was one of the inert gases, was bubbled into a pure solvent, and the resulting headspace vapor was directed into the QCM chamber for 1 min followed by 1 min of purging with nitrogen and repeated five times. The concentration of vapors was precisely controlled at 20% (i.e., 20 mL min^{-1} of sample vapors mixed with 80 mL min^{-1} of nitrogen) with two mass flow controllers. The total flow rate was 100 mL min^{-1} for both sampling and purging cycles. Data was recorded using a Crystal oscillator measurement system (QCM922A, SEIKO EG&G Co., Ltd.) in the form of frequency shift (Δf).

Molecular Docking Analysis: The peptide libraries and ligands were subjected to molecular docking using AutoDock Vina.[24] Structures of pentapeptides were constructed by PyMol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.). Each peptide and gas molecule conformation were visualized and checked to guarantee their accuracy in terms of hydrogen atoms and bond orders. The energies of all molecules were minimized based on a chemistry forcefield function named MMFF94s integrated with AutoGrid4. Possible stable conformers were precalculated to consider the molecule's flexibility. In docking, a rectangular search space of 20 Å × 20 Å × 20 Å was adapted, enclosing both a pentapeptide and a corresponding gas molecule. The flexibility of both partners was considered indirectly by employing nine different conformations in ensemble docking results. A docking solution with the lowest root mean square deviation (RMSD) was selected as the best conformation between a pentapeptide and a gas molecule to calculate the intermolecular free energy.[25] Based on the scoring function of Vina, ΔG was predicted as the sum of distance-dependent atom pair interactions, $f(d)$:

$$\Delta G = \sum_{\text{atom pair}} \frac{f_{\text{inter}}\left(d_{\text{inter}}\right)}{1 + w_{\text{rot}} N_{\text{rot}}} + f_{\text{inter}}\left(d_{\text{intra}}\right), \tag{1}$$

where d_{inter} and d_{intra} are the intermolecular and intramolecular surface distance of each atom pair, N_{rot} is the number of active rotatable bonds, and w_{rot} is the associated weight. Each atom pair interacts through steric interactions (i.e., gauss_1, gauss_2, and repulsion) and could be hydrophobic interaction and non-directional hydrogen bonding (H_{bond}). The atom pair interactions, $f(d)$ is given by

$$f(d) = w_{g1}\text{gauss}_1(d) + w_{g2}\text{gauss}_2(d) + w_{\text{rep}}\text{repulsion}(d) + w_{\text{HP}}\text{hydrophobic}(d) + w_{\text{HB}}H_{\text{bond}}(d), \tag{2}$$

where w_{g1}, w_{g2}, $w_{\text{rep.}}$, w_{HP}, and w_{HB} are associated weights for each term. Detailed scoring function can be found in the original paper by Trott and Olson, 2010.[21] For virtual screening, all possible pentapeptides with 12 gas molecules including toluene, aniline, MCH, AA, PA, acetone, *n*-hexane, *n*-heptane, hexanal, EtOAc, ethanol, and 1-hexanol were analyzed and the corresponding binding affinity was estimated using the reported weights for liquid phase as shown in Table S1, Supporting Information.[25] For the gas phase model, the binding affinity was estimated using the optimized weights determined using optimization algorithms described in the following section.

Evaluation Methods: To assess the performance of the obtained set of weights as a scoring function for the solid-gas phase, a score of a scoring function was evaluated for the *i*th gas molecule (S_i) based on three metrics: namely, Pearson's coefficient correlation, Spearman's rank correlation coefficient, and coefficient of determination through the origin. These three metrics were the direct measurements of linearity between normalized binding affinity ΔG_{norm} calculated by diving ΔG to the maximum absolute value of the system, $|\Delta G|_{\text{max}}$, and the absolute experimental QCM frequency shift, $|\Delta f|$, for each gas molecule. Pearson's coefficient correlation (R) assesses the linear correlation between normalized values of the predicted binding affinities and the experimental results.[26] Pearson's correlation coefficient of the *i*th gas molecule (R_i) was calculated based on ΔG_{norm} and, $|\Delta f|$ for each scoring function. Spearman's rank correlation coefficient assesses the ability of a scoring function to correctly rank the predicted binding affinity of one gas molecule to four different pentapeptides.[26] Spearman's rank correlation coefficient of the *i*th gas molecule (ρ_i) was calculated based on ΔG_{norm} and $|\Delta f|$ for each scoring function. To consider the nature of the solid-gas phase interaction, an additional parameter known as the coefficient of determination through the origin was introduced for evaluating the model's accuracy, indicating that no binding state between a pentapeptide and a gas molecule leads to no signal response toward the gas molecule. The coefficient of determination through the origin evaluates the linear correlation to the linear regression given by $\hat{y}_{ij} = a_i x_{ij} + b_i$, where x_{ij} is the absolute experimental frequency shift, $|\Delta f|$, $\hat{y}_{ij}$ is the normalization of estimated binding affinity, and a_i and b_i are the slope and the intercept of a linear regression between predicted binding affinity and experimental frequency shift. If ΔG_{norm} was zero, meaning no binding affinity, the gas molecules do not adsorb on the peptide, resulting in a zero value of $|\Delta f|$. Thus, the intercept b_i in this study was set to zero. The coefficient of determination through the origin was determined between predicted binding affinity and linear regression:

$$R_{0i}^2 = 1 - \frac{\sum_{j=1}^{n}\left(y_{ij} - \hat{y}_{ij}\right)^2}{\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_{ij}\right)^2}, \tag{3}$$

where y_{ij} and $\bar{y}_{ij}$ are the predicted binding affinity and the mean of the predicted binding affinity. It should be noted that the range of R_i is from −1 to 1, while R_{0i}^2 in this study is always in the range from 0 to 1. Based on the three metrics, the score, S_i for a scoring function is defined as

$$S_i = (1 + R_i) + (1 + \rho_i) + (1 - R_{0i}^2). \tag{4}$$

The optimization algorithms, described in the following section, minimize the S_i evaluated for each scoring function based on the set of weights.

Optimization Algorithms: In the AutoDock Vina models adapted to the solid-gas phase, the associated weights of each term in Equations (1) and (2) were optimized for each gas species. As an initial proof of concept, it was started with the same magnitude for all weights in the preliminary evaluation of the ability of AutoDock Vina for obtaining good agreement between predicted binding properties and the response of peptide-coated sensors. Two optimization algorithms: GA[27] and comprehensive screening via SA were applied.

GA was used as the first approach to obtaining optimal linear weights for a quick demonstration. In this modified approach, a group of twenty different solutions with six weight values each was initialized. Each weight value was randomly set between 0 and 1. After molecular dynamics simulation and evaluation with the above metrics, the top ten solutions were selected as parent solutions for the next generation based on their S_i as explained in the previous section. New solutions were generated by performing crossovers on the weight values from the parent solutions. This process involved selecting two parent solutions and combining their weight values to create offspring solutions for the next generation. To maintain

diversity among the solutions and to avoid convergence to a local minimum, randomness was introduced by modifying one weight value in this new solution. Specifically, a random value between −1 and 1 was added to a weight value, effectively implementing mutation. The new solution sets were then evaluated based on their performance using S_i. These new solutions were ranked according to their score, and the best performing solutions were selected to form new groups for subsequent iterations. This iterative cycle of selection, crossover, introduction of randomness (mutation), and evaluation was repeated for 1000 cycles. Extensive iterations allowed the algorithm to thoroughly explore the solution space and increased the likelihood of finding the global optimal solution rather than setting on a local optimum.

As the second approach to determining the optimal weight values, a comprehensive screening of all possible combinations of six weight values ranging from 0 to −1 was conducted with a step size of 0.2. As each of the six weights could have six possible values, this resulted in over 40,000 possible combinations in total. Each combination represented a potential solution and was evaluated using S_i. After evaluating all combinations, all solutions were ranked according to their performance scores. Rankings were then used to determine the best solution for the six weights. The comprehensive search method ensured a wide range of weight values to enhance the probability of identifying the global optimum.

Statistical Analysis: All statistical analyses were performed using Python 3.7.11 with the following libraries: pandas (v1.3.5) for data manipulation, NumPy (v1.21.6) for numerical computations, SciPy for statistical tests (v1.7.3), and matplotlib (v3.1.1) for data visualization. Experimental data sets were collected from at least three independent replicates and were expressed as means ± standard deviations.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

T.A.N. and M.-Q.F thank NIMS Joint Graduate School Program, NIMS. The calculations in this study were partially performed on the Numerical Materials Simulator at the NIMS. This work was financially supported partially by the Public/Private R&D Investment Strategic Expansion Program (PRISM), Cabinet Office, Japan.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

empirical force fields, gas phase model, gas sensor, parameter optimization, volatile organic compounds

Received: August 22, 2024
Revised: October 18, 2024
Published online: November 29, 2024

[1] J. Vessman, R. I. Stefan, J. F. van Staden, K. Danzer, W. Lindner, D. T. Burns, A. Fajgelj, H. Müller, *Pure Appl. Chem.* **2001**, *73*, 1381.
[2] P. Barik, M. Pradhan, *Analyst* **2022**, *147*, 1024.
[3] J. Wang, E. J. Murphy, J. C. Nix, D. N. M. Jones, *Sci. Rep.* **2020**, *10*, 3300.
[4] A. J. M. Barbosa, A. R. Oliveira, A. C. A. Roque, *Trends Biotechnol.* **2018**, *36*, 1244.
[5] M. Mascini, D. Pizzoni, G. Perez, E. Chiarappa, C. Di Natale, P. Pittia, D. Compagnone, *Biosens. Bioelectron.* **2017**, *93*, 161.
[6] T. Z. Wu, Y. R. Lo, *J. Biotechnol.* **2000**, *80*, 63.
[7] T. Z. Wu, Y. R. Lo, E. C. Chan, *Biosens. Bioelectron.* **2001**, *16*, 945.
[8] D. Pizzoni, M. Mascini, V. Lanzone, M. Del Carlo, C. Di Natale, D. Compagnone, *Biosens. Bioelectron.* **2014**, *52*, 247.
[9] T. Wasilewski, B. Szulczyński, M. Wojciechowski, W. Kamysz, J. Gębicki, *Sensors* **2019**, *19*, 4284.
[10] T. Wasilewski, B. Szulczyński, M. Wojciechowski, W. Kamysz, J. Gębicki, *Microchem. J.* **2020**, *154*, 104509.
[11] M. Blanco, M. C. McAlpine, J. R. Heath, *Computational Methods for Sensor Material Selection* (Eds: M. A. Ryan, A. V. Shevade, C. J. Taylor, M. L. Homer, M. Blanco, J. R. Stetter), Springer US, New York, NY, USA **2010**, pp. 135–148.
[12] Z. Kuang, S. N. Kim, W. J. Crookes-Goodson, B. L. Farmer, R. R. Naik, *ACS Nano* **2010**, *4*, 452.
[13] H.-H. Lu, Y. K. Rao, T.-Z. Wu, Y.-M. Tzeng, *Sens. Actuators, B* **2009**, *137*, 741.
[14] S. Panigrahi, S. Sankaran, S. Mallik, B. Gaddam, A. A. Hanson, *Mater. Sci. Eng. C Mater. Biol. Appl.* **2012**, *32*, 1307.
[15] M. Son, D. Kim, J. Kang, J. H. Lim, S. H. Lee, H. J. Ko, S. Hong, T. H. Park, *Anal. Chem.* **2016**, *88*, 11283.
[16] S. Ju, K.-Y. Lee, S.-J. Min, Y. K. Yoo, K. S. Hwang, S. K. Kim, H. Yi, *Sci. Rep.* **2015**, *5*, 9196.
[17] J. Park, J.-M. Lee, H. Chun, Y. Lee, S. J. Hong, H. Jung, Y.-J. Kim, W.-G. Kim, V. Devaraj, E. J. Choi, J.-W. Oh, B. Han, *Biosens. Bioelectron.* **2021**, *177*, 112979.
[18] T. Wasilewski, J. Gębicki, W. Kamysz, *Sens. Actuators, B* **2018**, *257*, 511.
[19] E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A. A. Friesem, C. Aflalo, I. A. Vakser, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 2195.
[20] A. Singh, M. M. Copeland, P. J. Kundrotas, I. A. Vakser, *Methods Mol. Biol.* **2024**, *2714*, 101.
[21] O. Trott, A. J. Olson, *J. Comput. Chem.* **2010**, *31*, 455.
[22] J. Eberhardt, D. Santos-Martins, A. F. Tillack, S. Forli, *J. Chem. Inf. Model.* **2021**, *61*, 3891.
[23] M. Brylinski, *Chem. Biol. Drug Des.* **2018**, *91*, 380.
[24] R. Quiroga, M. A. Villarreal, *PLoS One* **2016**, *11*, e0155183.
[25] E. W. Bell, Y. Zhang, *J. Cheminform.* **2019**, *11*, 40.
[26] C. Wang, Y. Zhang, *J. Comput. Chem.* **2017**, *38*, 169.
[27] C. S. de Magalhães, D. M. Almeida, H. J. C. Barbosa, L. E. Dardenne, *Inf. Sci.* **2014**, *289*, 206.