

This item is the archived peer-reviewed author-version of:

Unveiling the interaction mechanisms of cold atmospheric plasma and amino acids by machine learning

Reference:

Chai Zhao-Nan, Wang Xu-Cheng, Yusupov Maksudbek, Zhang Yuan-Tao.- Unveiling the interaction mechanisms of cold atmospheric plasma and amino acids by machine learning

Plasma processes and polymers - ISSN 1612-8869 - Weinheim, Wiley-v c h verlag gmbh, (2024), p. 1-26

Full text (Publisher's DOI): <https://doi.org/10.1002/PPAP.202300230>

To cite this reference: <https://hdl.handle.net/10067/2055120151162165141>

Unveiling the interaction mechanisms of cold atmospheric plasma and amino acids by machine learning

Zhao–Nan Chai¹, Xu–Cheng Wang¹, Maksudbek Yusupov^{2–5}, and Yuan–Tao Zhang^{1*}

*1 School of Electrical Engineering, Shandong University,
Jinan, Shandong Province 250061, P. R. China.*

2 School of Engineering, New Uzbekistan University, Tashkent 100000, Uzbekistan

*3 Department of Information Technologies,
Tashkent International University of Education,
Tashkent, 100207, Uzbekistan and*

*4 Laboratory of Thermal Physics of Multiphase Systems,
Arifov Institute of Ion-Plasma and Laser Technologies,
Academy of Sciences of Uzbekistan, Tashkent 100125, Uzbekistan*

Abstract

Plasma medicine has attracted tremendous interests in a variety of medical conditions, ranging from wound healing to antimicrobial applications, even in cancer treatment by the interactions of cold atmospheric plasma (CAP) and various biological tissues directly or indirectly. The underlying mechanisms of CAP treatment are still poorly understood although the oxidative effects of CAP with amino acids, peptides, and proteins have been explored experimentally. In this study, the machine learning (ML) technology is introduced to efficiently unveil the interaction mechanisms of amino acids and ROS in seconds based on the data obtained from the reactive molecular dynamics (MD) simulations, which are performed to probe the interaction of five types of amino acids with various reactive oxygen species (ROS) on the timescale of hundreds of picoseconds but with the huge computational load of several days. The oxidative reactions typically start with H-abstraction, and the details of the breaking and formation of chemical bonds are revealed; the modification types, such as nitrosylation, hydroxylation, and carbonylation, can be observed. The dose effects of ROS are also investigated by varying the number of ROS in the simulation box, indicating the agreement with the experimental observation. To overcome the limits of timescales and the size of molecular system in reactive MD simulations, a deep neural network (DNN) with five hidden layers is constructed according to the reaction data, and employed to predict the type of oxidative modification and the probability of occurrence only in seconds as the dose of ROS varies. The well-trained DNN can effectively and accurately predict the oxidative processes and productions, which greatly improve the computational efficiency by almost ten orders of magnitude compared to the reactive MD simulation. This study shows the great potential of ML technology to efficiently unveil the underpinning mechanisms in plasmas medicine based on the data from reactive MD simulations or experimental measurements.

* ytzhang@sdu.edu.cn

I. INTRODUCTION

The utilization of CAP has gained significant attention for various applications due to its tremendous potential advantages, including being obtained from flexible facilitation with simplified operation and leading to the formation of abundant reactive oxygen and nitrogen species (RONS) [1–5], such as hydroxyl radicals, oxygen atoms and nitric oxide, which have been proved to contain extensive biological–oxidizing properties [6–10]. Plasma medicine is a novel field of biomedical application with multiple disciplinary intersections, including physics, biochemistry, and clinical medicine. With those attributes mentioned above, CAP exhibits therapeutic effects in various applications, including wound and ulcer healing [11–14], cancer therapy [15–19], dental applications [20], blood coagulation [21], disruption of the human hepatocyte cytoskeleton [22], and dermatology [23–32]. Furthermore, plasma medicine demonstrates excellent potential for surface modification of biomaterials, decontamination of medical devices, and disinfection of wound surfaces.

To explore the underlying mechanism of the action between CAP and biological organisms, insight into the interactions between RONS and biological functional biomolecules is crucial. Investigations have been performed to investigate the oxidation of CAP in the case of lipids, DNA, and proteins [33–42]. As the most abundant compound and the main undertaker of living organisms, proteins also dominate biological functions from the perspective of molecular biology. Learning about the interaction of RONS with proteins is necessary to deeply understand plasma biological mechanisms. Proteins are composed of 20 kinds of natural amino acids connected by peptide bonds, which are formed by dehydration condensation of two adjacent amino acids. The main difference between amino acids is the side chain structures, which can substantially impact the spatial structure and physical and chemical properties of proteins. Therefore, this paper focuses on the interaction between reactive species and several amino acids with different characteristics.

Much experimental research has revealed that RONS in plasma can induce the oxidation of amino acids and proteins. Takai et al. used lysozyme to study the chemical effects of atmospheric pressure low–temperature plasma on proteins in aqueous solutions. They first observed that plasma could lead to lysozyme inactivation [43] and hypothesized that the secondary structure of lysozyme was chemically modified under low–frequency plasma jets, which was actually the oxidative modification of amino side chains induced by reactive

particles such as OH radicals. Sharifian et al. reported that the carbonylation content of beef myofibrillar protein significantly increased while the free thiol content remarkably decreased during 10 minutes of dielectric barrier discharge plasma treatment, which indicated that the functional properties of myofibrillar protein were changed because of its oxidation caused by active species [44]. The research of chemical effects of CAP on the amino acids of Takai et al. showed that 14 of the 20 amino acids had oxidative modifications on their side chains, which were affected by chemical reactions between amino acids and plasma induced reactive species rather than chemical degradation caused by acidic conditions [33]. The study denoted that produced reactive particles react preferentially with sulfur- and aromatic ring-containing amino acids among all others. The hydroxylation and nitration of tyrosine (Tyr), phenylalanine (Phe), and tryptophan (Trp) were also observed by high-resolution mass spectrometer, as well as the sulfonation of cysteine (Cys) and sulfoxidation of methionine (Met), and ring-opening products of five-membered cyclic amino acids. In addition, Zhou et al. found oxidative modifications on amino acids side chains in their research of the interaction of plasma with amino acids and summarized oxidation types into hydroxylation, nitration, dehydrogenation, and dimerization [45].

Although the chemical oxidation of CAP has been proved experimentally, the underlying mechanisms of plasma medicine remained limited with independent experimental observations. Researchers suffer from the situation that the poor details of experimental results caused by the restrictions of experimental detection result in the incompleteness of fundamental understanding. However, numerical simulation tends to be a powerful instrument for experiment verification and supplements with the continuous progress of modern computer science and incredibly improved calculation rate. With the advantages of low-cost equipment and the ability to model molecular interactions, molecular simulation technology has attracted increasing attention in plasma medicine. The oxidative effects of plasma on amino acids have been studied theoretically and experimentally. However, a specific explanation of the inactivation mechanism of amino acids at atomic or molecular levels still remains to be proposed. The biological tissues are complex organisms with massive physiological and biological properties, which means all mechanisms in the whole process of interaction between plasma and biological tissues cannot be wholly explained by molecular simulations, such as the cell signaling pathway affected by plasma. In contrast, molecular simulations can deeply describe the first step of the interaction between plasma and biological tissues

at the molecular level, which actually is the active radicals reacting with and inactivating the biomolecules. Several simulation methods have been employed in the area of plasma medicine at the atomic or molecular scale, such as quantum mechanical calculation, density functional theory method, and molecular dynamics simulation [41, 42, 46, 47]. As one of the most popular methods of molecular dynamics simulation, the reactive molecular dynamics (MD) simulation can expose the processes of chemical bonds breaking and reforming. In this study, the reactive MD simulation is applied to explore the interaction process between active radicals and amino acids, while the drawbacks of low computational efficiency and huge simulation time consumption of reactive MD simulation have gradually emerged.

With the updation of calculating devices and the emergence of high-performance computing technology, such as GPU computing and parallel computing, the computational performance of current computing methods has been highly improved [48–50]. In recent years, the expanding applications of artificial intelligence (AI) across areas, especially in physics, emphasize its potential to address current challenges and explore novel prospects for plasma medicine [51–53]. The usage of AI in the field of fusion has developed for a while, achieving some outstanding research [54, 55]. However, the development of AI technology involving low-temperature plasmas, especially in plasma medicine, needs to be largely improved, with a couple of challenges and limitations [56, 57]. Machine learning (ML) is an essential branch of AI with the capability of data statistics and analysis, and it can automatically obtain the correspondence between tremendous amounts of data. The algorithm of ML is designed to identify patterns in data of a complex system to make predictions and decisions or discover the intrinsic relationship. Before feeding the training data, a transformation of row of ML algorithms is generally required to convert data into more informative representations or features, which is called feature extraction and has a critical impact on the success of ML. Deep learning (DL) is an essential branch of ML that can automatically learn features from massive data without being explicitly programmed. Thanks to the significant progress in large datasets and computing resources in recent years, the idea of complex system modeling through DL methods has achieved spectacular success across a wide range of applications, which also lays an objective foundation for studying plasma characteristics based on big data. Recent studies have shown that DL with multiple hidden layers has excellent results in the simulation and modeling of CAP [58–60]. In particular, several recently published review papers demonstrate advances in the application of AI technology in the field of plasma

medicine [51, 52, 61, 62]. In our previous studies, a deep neural network (DNN) was constructed as a surrogate model to investigate the discharge characteristics of atmospheric helium dielectric barrier discharges, showing excellent performance and incredible accuracy of discharge characteristics of DBDs [60]; the DNN was also applied to describe the discharge characteristics and plasma chemistry of CO₂ discharges, achieving accurate predictions with high computational efficiency [63]. However, there is a need for further investigation to apply DNN to the numerical simulation of plasma medicine to reveal the underlying mechanisms of interactions between amino acids and reactive radicals [64, 65].

In this paper, the computational efficiency of DNN is examined to study the oxidative modification of amino acids upon the impact of ROS produced in CAP with the data obtained by reactive MD simulation, exploring the further utilization of DL algorithms in the fields of plasma medicine. DL holds the capability to transform the numerical simulation of plasma medicine field and improve the effectiveness of reactive MD simulations.

Section II describes the numerical methods of reactive MD and the construction of DNN; the interactions of amino acids and ROS are profoundly explored in Section III, and the reaction pathways of the oxidative modifications of amino acids and the effects of different types and dosages of active species are presented in detail, respectively; the prediction data from DNN on the oxidative modifications of amino acids are discussed. The predictions show the potential of DNNs involving numerical simulation of plasma medicine with higher computational efficiency, and the training data can be collected by other methods except reactive MD simulation, such as experimental acquisition. In the final section, we summarize the oxidation of amino acids and discuss the tremendous potential of DNN in plasma medicine.

II. DESCRIPTION OF NUMERICAL METHODS

A. Reactive Molecular Dynamics simulation

Molecular dynamics simulation is a powerful tool to model natural molecular systems through statistical principles. There are several MD simulation studies devoted to the interaction of plasma with the material in plasma medicine at atomic/molecular scale [46, 47, 66, 67]. MD simulations regard specific species (atoms, ions, molecules) as the

object and consider the whole system as a particular characterized collection of particles. The underlying algorithm of MD simulations is the integration of motion equations of all involved particles in the system. In MD simulations, the movements of atoms are caused by the forces acting on them, which are obtained from quantum mechanical calculations and classical interatomic potential. Based on the capability of dissociation and formation of chemical bonds, MD simulations can be divided into nonreactive and reactive MD simulations, and both can achieve the specific characteristics of the system mentioned above at the atomic/molecular level. In nonreactive MD simulation, the molecules remain fixed in the system, which means that the breakage and formation of bonds are not involved in the simulation. The computational afford of the simulation is considerably decreased without the recalculation of bond order in each time step. In addition, with a less complicated functional form of force field, the nonreactive MD simulation is superior for an extensive system with $10^6 - 10^8$ atoms at a timescale of 0.1ns–10 μ s [68]. In contrast, the reactive MD simulations can explore the formation and breakage process of chemical bonds. Thus, the molecule connectivity needs to be recalculated at each time step of the equation integration. Due to the expensive computational burdens, the suitable space and time scale of reactive MD simulations are consequently reduced. Considering the complexity of the interatomic potential, the number of atoms shrinks to $10^4 - 10^6$, and the timescale decreases to 1ps–100ns, which is remarkably smaller than that of nonreactive MD simulations [68].

To explore the oxidative modification of amino acids upon the impact of reactive oxygen species (ROS), the reactive MD simulation is employed in this study, and the Reax force field (ReaxFF) [69] is applied as the interatomic interaction potential with high accuracy and low computational expense [66]. The applicability of ReaxFF has evolved from hydrocarbons to handling almost half of the elements of the periodic table. The ReaxFF can describe various multielement compounds and has been adopted to reveal the interaction of ROS with peptidoglycan, lipids [67], and DNA [70, 71] with outstanding achievement. The order and distance of chemical bonds are the basements of the force field because the connection of atoms is obtained from those two indispensable information. Upon the impact of the force field, the distance of atoms evolves at each time step of the reactive MD simulation, and the cleavage and formation of chemical bonds also occur during computation. The ReaxFF holds the capability to model non-bonded interactions like van der Waals and Coulomb forces and strong interactions such as covalent and ionic bonds. Eq. 1 shows the total

energy of the system upon the effect of ReaxFF as follows:

$$E_{\text{system}} = E_{\text{bond}} + E_{\text{over}} + E_{\text{under}} + E_{\text{val}} + E_{\text{Coulomb}} + E_{\text{vdw}} + E_{\text{conj}} + E_{\text{tors}} + E_{\text{pen}} \quad (1)$$

where E_{system} is the total potential energy of the system, consisting of several energy terms. E_{bond} on the right-hand side represents the chemical bond energy, determined by the bond order. E_{over} and E_{under} refer to the over- and under-coordination, respectively. The fourth term on the right-hand side denotes valence angles. E_{Coulomb} and E_{vdw} indicate the interaction energy of Coulomb and van der Waals, respectively. E_{conj} means conjugated systems, E_{tors} is torsion terms, and E_{pen} represents the occurrence of unfavorable configurations. A more detailed description of the formulation can be found in [69].

B. Structure of amino acids and simulation setup

As the most widely distributed biological macromolecule in living organisms, proteins are composed of amino acids connected by peptide bonds. Proteins are the most vital substances of life phenomena and the essential material to express biological genetic traits. Due to the complicated molecular structure, proteins need to be modified and processed to various degrees to perform desired biological functions. Post-translational modifications (PTMs) are mainly used to regulate protein function by adding small chemical molecule groups to amino acid side chains of proteins or modifying existing groups [44]. PTMs bear an essential role in the regulation of protein structure and function, which can adjust to the physiological state of cells and the external environment, thus regulating the signaling pathways of cells [39, 72, 73]. Proteins are especially susceptible to oxidative modifications considering the structure of proteins and the presence of containing amino acid residues like methionine, cysteine, lysine, proline, tyrosine, histidine and tryptophan [44]. Various research has shown that conformational variation occurs in proteins upon the impact of plasma, demonstrating the oxidative modification effect of plasma treatment [74–76]. The oxidative modifications of proteins upon plasma treatment seem qualitatively in line with the post-translational modifications of proteins in living organisms, both of which can significantly influence the function and lifetime of proteins.

The oxidative modification of proteins mainly occurs at the side chain of the amino acids,

which have a significant impact on the spatial structure and physical and chemical properties of proteins. Thus, this study focuses on the oxidation of amino acids, the elementary units of proteins. The connection of amino acids are peptide bonds constructed by the dehydration condensation between the amino and carboxyl groups of two adjacent amino acids, resulting in the formation of peptides and proteins. According to the different side chain structures, 20 kinds of amino acids can be divided into sulfur-containing amino acids, aromatic amino acids, polar amino acids, carbon chain amino acids, and five-membered cyclic amino acids. This article takes methionine (Met), phenylalanine (Phe), cysteine (Cys), glutamic acid (Glu), and proline (Pro) as representatives, and the schematic representation of the structures of these amino acids is illustrated in Fig. 1. Five amino acids are randomly placed in the simulation box, a cube with an edge length of 30 \AA , to observe all the possible oxidative modifications upon the impact of ROS, and the breakage and formation of chemical bonds are carefully investigated. The number of adopted ROS (OH , O , O_3 and H_2O_2) gradually increases from 10 to 50 while the simulation box keeps at the same. Fig. 2 shows an example of five amino acids under the interaction with 30 oxygen atoms in the simulation box. For each simulation, different numbers of the ROS are randomly placed in the box with a number much larger than that of the amino acids. The statistical nature of molecular encounters is addressed by random sampling. It is worthy to be noted that the ROS density applied in the simulation is far more than that generated by plasma in the experiments [28, 77–79]. Periodic boundary conditions are employed in all three dimensions to emulate a larger system in our simulations. Essentially, the simulation box is replicated infinitely in all directions, creating an infinite periodic arrangement. This approach allows us to study the global properties of the entire system by simulating a single box. Once the periodic boundary conditions are established, if any particles exit the box, identical particles can return to the box from the opposite interface in the simulation of particle movement. Furthermore, the atoms located at the boundaries of the simulation box can interact with the atoms on the opposite side of the box. The ReaxFF is applied as the force field to simulate the oxidation of amino acids at the atomic scale. The entire system is equilibrated at room temperature for 50 ps in the canonical ensemble using a Berendsen thermostat. The thermostat coupling parameter is set to 0.1 ps. In order to ensure accurate integration of the equation of motion and facilitate smooth progression of the reactions, the time step is limited to 0.1 fs [71]. The total calculation time for each simulation is set to 300 ps, resulting in 3 million iterations. During this time, all

expected types of reactions should take place, leading to the final simulated results. After conducting numerous tests, it has been determined that this duration is sufficiently long to observe the majority of reactions. Each interaction is simulated using 10 independent runs to gather comprehensive statistical data on the processes that occur. It should be noted that the simulation statistics are influenced by the number of calculations performed and by the fact that this type of simulation primarily aims to provide molecular-level insights due to the limited time and length scale.

C. Data Preprocessing and structuring for DNN

Thanks to the improvement in computation performance and the availability of large datasets, AI based on data-driven science has attracted growing attention. Data-driven science is recognized as the fourth paradigm of discovery, and the previous three paradigms are empirical or experimental (Galileo Galilei), theoretical (Isaac Newton), and computational (not a single person), respectively [80]. Data-driven science is a transformational technical with fundamental differences from the previous three paradigms. Most notably, it aims to create a “learning machine” in which AI makes fully automated scientific discovery, taking human intelligence out of the discovery process or performing tasks without explicit programming. Computationally expensive operators in the reactive MD simulation being used to investigate the interaction of amino acids with plasma-generated ROS can significantly impact the ability to simulate the oxidative reactions sufficiently. By replacing these expensive operators with less computational consumption surrogate models, the overall performance of the simulation can be greatly enhanced. By applying machine learning (ML) algorithms that employ a plenty of data to train a DNN, the surrogate model is used as a replacement for the expensive operator in reactive MD simulation.

Data-driven surrogate models have been developed to integrate ML algorithms with plasma fluid models [59, 60, 63, 81, 82] and kinetic models [83]. The amount of available data, training methods, and network architecture can significantly impact the quality of the resulting ML surrogate models. The availability of a massive corpus of data can ensure the development and application of ML surrogate models. Deep learning (DL) is a subset of ML that employs deep neural networks to explore complicated correlations in data. ML methods with multiple parameters and complex mathematical structures need vast data to reveal the

most informative features or representations. The characteristics of data from simulation or experiments should be carefully considered to incorporate the physics knowledge into the DL surrogate models [52].

An open reaction database (ORD) to support machine learning has been suggested to provide a structured data format for chemical reaction data, and the related studies such as reaction prediction, chemical synthesis planning have been discussed [84]. As the representation of the data is crucial and the premise for ML [85], a reaction description with a width equal to the number of starting materials in the simulation box and bits representing the dosages present in a given reaction mixture with a number, similar to the one-hot coding, is adopted in this work. Example vector representations for the oxidative reactions consisting of five amino acids and four types of ROS are shown in Fig. 3 (a). This representation of the vector space eliminates the inclusion of chemical structure, and the DNN is able to directly recognize the generated vectors without prior knowledge of reactivity and chemical structure.

To further explore the vector representation of this work, a recently described Suzuki–Miyaura reaction space [86] (see Fig. 3 (b)) is also investigated and incorporated to predict the yield of products. The vector is divided into two parts: reactants, including the amino acids and ROS, and products of corresponding oxidative modifications. In addition, specific categories are differentiated by region; the reactant part is further divided into five amino acids and four ROS, and the product part is also classified into five zones corresponding to the type of oxidative modification for each amino acid, respectively. The numbers in different regions of the vector have different meanings: for the amino acid part of the reactant, 1 and 0 indicate the presence or absence of the corresponding amino acid. Considering that the dose of ROS is far beyond the scope of DNNs and the data need to be normalized, the number corresponding to ROS is $0.1 \sim 0.5$ after normalization, referring to the actual value of active radicals with $10 \sim 50$. The amount of oxidative modification types is much more than the reactions, and each position illustrates the probability of the corresponding product according to statistics, which can be approximately considered as its yield. More details of product information are described in detail in Fig. 3 (b).

D. Construction of DNN

With the ability to learn complex chemical reactivity patterns from oxidative modification data becomes increasingly apparent, data-driven ML surrogate models are constructed and employed for synthetic route planning, reaction condition recommendations, and even prediction of significant products for untested reactions. With the templating mechanism for chemical data set enumeration suitable for machine learning described in detail in the previous subsection, the oxidation data obtained from the chemical reactions simulated by Reactive MD is recorded in a sample database. To leverage the full potential of the collected simulation data, a DNN is employed to explore the hypothesis of a surrogate ML model to predict the type of oxidative modification and the corresponding probability of occurrence. After data structuring, all the details of the chemical oxidative modification reaction, such as the structures of reactants and products and the dosage of reactive radicals, expressed in unstructured text format in the original database, can be alternated into vector representation and recognized by the DNN. The simplified architecture of the DNN, shown in Fig. 4, is adopted in this work with full connection mode and back-propagation algorithm. The framework of DNN consists of three parts: the input layer, the hidden layer, and the output layer. The input layer located on the leftmost layer of Fig. 4 is used to import the structured data or conditions with input neurons, as in this case, nine input neurons are involved, representing five amino acids and four active species. The rightmost or output layer contains the output neurons, which are divided into five parts referring to the oxidative modification types illustrated in Fig. 3 for achieving the prediction of products or classification corresponding to the input data. The intermediate part is called the processing or hidden layer, which contains single or multiple layers and adjustable numbers of neurons according to the specific application. Each neuron in hidden layers is connected to neurons in the former and latter layers, but neurons in the same layer do not communicate with each other. The repetition of this process is referred to as the free-forward transmission algorithm for data, and the specific formulation of the equation is expressed as:

$$x' = \sigma(\omega x + \beta) \tag{2}$$

$$\begin{bmatrix} x'_1 \\ x'_2 \\ \dots \\ x'_j \\ \dots \\ x'_m \end{bmatrix} = \sigma \left(\begin{bmatrix} w_{11} & w_{12} & \dots & w_{1k} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2k} & \dots & w_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ w_{j1} & w_{j2} & \dots & w_{jk} & \dots & w_{jn} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ w_{m1} & w_{m2} & \dots & w_{mk} & \dots & w_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_k \\ \dots \\ x_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_j \\ \dots \\ b_m \end{bmatrix} \right) \tag{3}$$

Note that all parameters shown in Eq. 2 are stored as lists of matrices, where x indicates data tensors transmitted from the former layer. ω represents a matrix whose element w_{jk} is the weight for the j -th neuron of the current layer connecting the k -th neuron in the former layer. β suggests a vector with b_j denoting the biases of the j -th neuron in the current layer. The matrix representation is described in Eq.3, where it is assumed that the previous layer contains n neurons and the number of neurons in the current layer is m . x' denotes the output vector of the present layer, which is also the input data for the next hidden layer. On the basis of adjusting transformation of data by weight and bias, the activation function $\sigma(a)$ is able to determine the trend of data conversion. Various forms of activation functions have been developed and explored with the development of artificial intelligence. Currently, only one activation function is considered for each hidden layer, and the primary common forms of activation functions are Sigmoid, Tanh, Relu, Elu, Mish, Softmax. Taking the sigmoid function as an example, the equation is defined by:

$$\sigma_{\text{sigmoid}}(z) = \frac{1}{1 + e^{-z}} \tag{4}$$

Where z is expressed as $z = w \cdot x + b$, as z is monotonically transformed from a large negative number to a large positive number, the value of $\sigma_{\text{sigmoid}}(z)$ is also monotonically changed from zero to one. The linearity guarantees choosing the small changes of Δw and Δb in weights and biases to achieve the desired change in the output. Thus, the algorithm of sigmoid neurons has much of the same qualitative behavior as perceptrons.

The DNN can recognize the vector of chemical interaction data generated by the one-hot encoding. It is necessary to construct an appropriate DNN architecture to describe the efficiency and effectiveness of the DNN surrogate model in predicting the products and yields

of amino acids interacting with different doses of ROS with currently available data. The DNN employed here to analyze this information, leveraging its prior utilization in our earlier research to investigate discharge characteristics [60, 63]. The outcomes of the current study reaffirm the effectiveness of this neural network. DNN construction requires detailed consideration of various parameters, such as the selection of the number of hidden layers and neurons in each layer, the sequence of activation functions, and the setting of the loss function and learning rate. Since the ML algorithms have only recently been introduced to the study of low-temperature plasmas, to our knowledge, there are currently no unified principles to determine and optimize the structure of neural networks by adjusting the activation functions, hidden layers, neurons, and other parameters, which often requires a case-by-case analysis based on the characteristics of data from the simulation or experiments. After conducting thorough adjustments to the number of hidden layers, neurons per layer, and activation functions in the DNN, the finalized structure consists of five hidden layers, each with 128 neurons. The arrangement order of the activation functions is also essential in determining the optimal DNN structure. The results suggest that the tanh–relu–relu–relu–sigmoid structure (shown in Fig. 4) has the lowest simulation error and is considered the final sequence of activation function in the hidden layers. At last, with the number of iterations increasing to 10,000, the simulation errors are further reduced, and a desirable yield prediction of oxidative products is obtained from trained DNNs. In addition, MSE is employed as the loss function, and the learning rate is locked at 0.0001 during the above experimentations, for the effect of these two parameters is out of consideration in this study.

To test the effectiveness of the constructed DNN, the test set containing data outside the training set is necessary. With a sufficient amount of data, the data is generally randomly partitioned into the training set, validation set, and test set in the ratio of 6 : 3 : 1. The training set is used to tune the weights and biases in the DNN framework to achieve the desired prediction. The role of the validation set is to verify the prediction effectiveness of the trained DNN by inputting data outside the training set, and the validation set is merged within the training set when the amount of data is lacking, or no validation is required. The test set is capable of testing and evaluating the effectiveness and applicability of DNNs where its data are outside the training and validation sets. The demonstration of DNN performance also requires the utilization of the test data.

In this work, approximately 50 days were required to obtain the chemical reaction data of five amino acids interacting with several dosages of four ROS by reactive MD simulation. A total of 100 sets of data were collected as the training data for the DNN. To facilitate the analysis and discussion of the results, one set of data given the five dosage effects ranging from 10 to 50 is taken as the test set, and the training and testing datasets are partitioned in a ratio of 4 : 1. Therefore, a total of five groups of DNNs are trained, and the discussion of each group is given in detail below. It is worth noting that after several hours of training, the DNN takes only one second to predict the oxidative product with relatively accuracy (over 80%) , significantly reducing the computational consumption compared to the traditional reactive MD methods. In addition, the experimental data can also be performed as the training dataset to train DNN, and with the continuous acquisition of data and gradually supplemented into the training dataset, the prediction performance of the DNNs will be further enhanced [56]. Naturally, we plan to further optimize the present DNN and explore alternative types of neural networks in the next work.

III. RESULTS AND DISCUSSION

A. Effect of interaction of OH radical with amino acids

The main difference between amino acids is the structure of the side chains. According to the different side chains, five kinds of amino acids considered in this work can be classified into sulfur-containing amino acids (Met, Cys), aromatic amino acids (Phe), polar amino acids (Glu), and five-membered cyclic amino acids (Pro), respectively. The carbon chain amino acids are not involved in this study for poor oxidative modifications under the effect of ROS. In this part, the interaction of five amino acids with OH radical is described in detail to unravel the reaction pathway. The cleavage and formation of chemical bonds are carefully analyzed and the oxidative products under the impact of OH radical are summarized.

The sulfur atoms in sulfur-containing amino acids make them most susceptible to oxidation, which means that the thioether group (R–S–R) of Met is an excellent oxidative target. With the addition of one or two oxygen atoms under the interaction with OH radicals, the thioether group of Met is oxidized to form the structure of $-S=O$ or $-S(=O)_2$. In addition, the oxidative modifications of the methyl group at the end of the thioether group

start with H-abstraction and the formation of an alkyl radical, which is unstable and can further convert to an alcohol group ($-\text{CH}_2\text{OH}$) under the effect of OH radicals. Cys is another sulfur-containing amino acid, the sulfhydryl (or thiol) group ($-\text{SH}$) of which is also a centralized site of oxidative modification. The bond dissociation energy of the S-H bond in amino acids is 363 kJ/mol, which is the lowest compared with the O-H bond (459 kJ/mol), N-H bond (386 kJ/mol), and C-H bond (411 kJ/mol) [45]. The bond dissociation energy is released from bond formation or absorbed by bond breakage. With higher bond dissociation energy, the chemical bond will be more stable. Thus, the oxidative modifications of Cys begin with the H-abstraction of the sulfhydryl group. The reaction mechanism involved can be explained by the fact that OH facilitates the dissociation of the S-H bond, leading to the generation of an unsaturated site that becomes susceptible to OH-addition. Additionally, the free hydroxyl group continues to oxidize both the newly formed alcohol group and the remaining S-H bond. As a result, the generated unsaturated sites undergo further reactions, eventually leading to the formation of the $-\text{S}=\text{O}$ bond. With the further addition of two or three oxygen atoms on the sulfhydryl group, the structure of $-\text{SO}_{2-3}$ is composed, which is called sulfonation. The simulation results show that the sulfonated Cys is the main oxidative product, which is qualitatively in line with the conclusion of experimental studies of amino acids or proteins [33, 45, 87]. Under the interaction with OH radicals, the oxidative modifications of Phe start with the H-abstraction in the carboxyl group. With further oxidation, the H-abstraction occurs in the phenol group on the side chain of Phe, which plays an essential role in signal transduction. The oxidation of the phenol structure will change the enzymatic activities of target proteins in response to cytokines, growth factors, and hormones, thus affecting the cell signaling process [45]. The side chain of Glu in polar amino acids contains two methylene groups and one carboxyl group, which is relatively stable with low reactivity. The oxidative reaction between Glu and the OH radical usually starts with the abstraction of H atoms in the carboxyl group. The H-abstraction then occurs in the amino group in Glu. The processes of oxidative modifications of Pro in five-membered cyclic amino acids are basically similar to Glu. The H-abstraction reaction mainly focuses on the carboxyl and amino group of Pro under the impact of OH radicals. In addition, the biological function of Pro may be influenced by the destruction of the five-membered ring.

B. Effect of interaction of other ROS with amino acids

In addition to OH radicals, the oxidative effect of other ROS, such as O atoms, O₃ molecules, and H₂O₂ molecules is also investigated in this study. Compared with OH radicals, the types of oxidative modifications of amino acids are different under the action of O atoms, O₃ molecules and H₂O₂ molecules, which indicates that the species of ROS is an important factor affecting the oxidative modification of amino acids.

The oxidative modifications of Met in the sulfur-containing amino acids are demonstrated in Fig. 5. Fig. 5 (b) and Fig. 5 (c) describe the oxidation of the thioether group with the addition of one or two O atoms and the formation of $-S=O$ or $-S(=O)_2$. Upon the impact of O atoms, another oxidation of the thioether group is shown in Fig. 5 (d) with the absorption of three O atoms to form the structure of $-SO(=O)_2$. Some other modifications occur under the interaction with O atoms, such as the dissociation of C-S bonds. The methyl group on the terminal of the side chain converts to CO₂ or COH, and the sulfur-containing group detaches from Met and forms SO₂ or SO₃. The oxidative reactions between Met and O₃ molecules bring out some new products like the H-abstraction of the amino group and even the addition of an O atom. The sulfur-containing groups are oxidized with the absorption of $-OH$ upon the impact of H₂O₂ molecules, while the other oxidative modifications are similar to the products of OH radicals. Another sulfur-containing amino acid mentioned in this study, i.e., Cys has similar oxidative modification types to Met. Upon the interaction with O atoms, O₃ molecules and H₂O₂ molecules, the oxidative products of Cys can be summarized as the detachment of the sulfhydryl group to form SO₂ or SO₃, the H-abstraction and O-addition of amino group, and the dehydrogenation of sulfhydryl group with addition of $-OH$. The main differences between the oxidative modification of Phe under the impact of O atoms and OH radicals are the H-abstraction of methylene and the formation of CO₂ from the detachment of the carboxyl group. Under the impact of O₃ molecules, H-abstraction may occur at other sites on the benzene ring of Phe. The oxidation effect of H₂O₂ molecules on Phe is weaker than other ROS, and H-abstraction only occurs on the carboxyl group. The simulation results show that the carboxyl group on the side chain of Glu is the main reaction site. The H atom of the carboxyl group is easily detached in the oxidative modification of Glu with each kind of ROS because the polarity of the O atom on the carboxyl group is relatively large, which is conducive to the departure

of hydrogen atom as the proton, leading to the rupture of O–H bond quickly. Upon the impact of O atoms and O₃ molecules, the H atom in the amino group is also abstracted. Particular reactivity of Glu to O atoms is also observed in the simulation. The unsaturated C=C bond is formed due to the abstraction of H atoms from the methylene group at the distal carboxyl end of the Glu side chain, which leads to the dissociation of the C–C bond connecting the carboxyl group and methylene group. Finally, the carboxyl group detaches from the Glu side chain to form CO₂. The reaction of the carboxyl group detaching to CO₂ is called the decarboxylation reaction, and the decarboxylation reaction on polar amino acid side chains containing the carboxyl group was also observed in the study of protein oxidative modification by Xu et al. [40]. The oxidative modification of Pro under the interaction with other ROS has little change with OH radicals. Only the decarboxylation reaction occurs at the carboxyl group of Pro under the action of O₃ molecules.

C. Oxidation products

The previous two sections introduced the oxidative modifications of amino acids under the impact of OH radicals and compare with the effects of different kinds of ROS (O atoms, O₃ molecules and H₂O₂ molecules) on five amino acids. Under the impact of plasma, the oxidative modifications of amino acids cause variations in polypeptide and protein conformation, which will affect the biological function of the protein to a certain extent. In this part, all oxidative modification types of five kinds of amino acid molecules interacting with ROS are summarized, which is helpful in revealing the mechanism of action of the oxidative modification of proteins at the molecular level.

In the interaction process with ROS, dehydrogenation occurs first on the side chain of amino acids, including H–abstraction on the site of the S atom, C atom, N atom, and O atom. The dehydrogenation reaction on (or near) the S atom indicates that the thioether group of Met and the sulfhydryl group of the Cys are significant targets of H–abstraction. The abstraction of the H atom on the C atom commonly occurs in all the five amino acids considered in this work. The dehydrogenation on the N atom refers to the cleavage of amino groups of five amino acids and N–H bond on the five–membered heterocyclic ring of Pro. The reaction site of H–abstraction on O atoms includes the carboxyl and hydroxyl groups of almost all five amino acids.

The free radicals formed after H–abstraction reaction continue to participate in oxidation under the impact of ROS, and the addition of O atom occurs on S atom, N atom and C atom. The addition reaction of O atoms to the S atom is called sulfonation. The sulfonic acid group is generated after sulfonation of the sulfhydryl of Cys, and the sulfonyl group is obtained from the oxidation of the thioether of Met. The sulfur-containing groups can be further oxidized upon the effect of ROS to form SO_2 or SO_3 . Two main addition reactions on the N atom are observed in the simulation. One is the addition of $-\text{OH}$ to the N atom, called hydroxylation, and sometimes double hydroxylation occurs with two hydroxyl groups absorbed by the N atom. Another additional reaction refers to the formation of $\text{N}=\text{O}$ structure, called nitrosylation, which occurs less frequently. The addition reactions on the C atom are hydroxylation and carbonylation. Hydroxylation refers to the absorption of $-\text{OH}$ on the methyl radical or methylene radical after the dehydrogenation reaction, forming an alcohol group. However, the carbonylation modifications are subdivided into the formation of ketone and aldehyde groups on the side chains of amino acids.

In addition to the H–abstraction and O–addition reactions discussed above, several oxidative modifications are illustrated as follows: In five–membered cyclic amino acids, ring–opening products occur upon the impact of ROS. The detachment of carboxyl groups of five kinds of amino acids can be observed under the effect of ROS, which is called the decarboxylation reaction. The oxidative modification types of these amino acids obtained from the simulations are summarized in Table I. It can be found that the leading oxidative site of ROS is the side chain of the five kinds of amino acid molecules. Sebastian et al. [88] analyzed the chemical modification of two peptides with plasma treatment by nanoflow liquid chromatography–mass spectrometry in their experiments, and 17 different modifications were detected in the product spectrum. The experimental results show that the addition of O atoms is the most common modification, which is consistent with the absorption reaction of O atoms observed in the calculation. In addition, dehydrogenation, sulfonation, hydroxylation, carbonylation, nitrosylation, ring cleavage, and other oxidative modification types summarized in Table I, except decarboxylation, can be found in the product spectrum. The high consistency between the calculation and experimental results confirms the reliability of the reactive MD simulations. The method of reactive MD simulation also provides new insight into the molecular mechanism of oxidative modification of amino acids from another perspective.

D. Dose effects

The dose effect of ROS is considered in studying the oxidative reaction between ROS and amino acid molecules. The dosage of ROS is investigated by adjusting the number of ROS in the simulation box. According to the simulation results, it can be concluded that the dose of ROS is the crucial factor affecting the oxidative modification. The biological effect of plasma also depends on the therapeutic dose in cells or biological tissues in clinical applications. Thus, to better explore the application of plasma in biomedicine, the plasma dose should be carefully investigated [89]. However, the plasma dosage is difficult to define because of its complex characteristics and diverse plasma-generated sources. The biological effects of plasma are the synergistic effects of ROS, reactive nitrogen particles (RNS), electrons, ions, charged and neutral particles, electromagnetic radiation, and heat effect. Some groups regard plasma processing time as dose. However, the processing duration is not the only effect on plasma dose, which is also affected by background gas, plasma generating source, discharge frequency, electrode interval, and other factors. In addition to the processing duration, Gidon et al. [90] proposed to express it by the energy deposited into the plasma in each region, but different background gases have different effects under the same power and processing time. Recently, Lu et al. [91] proposed the concept of Equivalent Total Oxidation Potential (ETOP) to evaluate the biological effects of RONS in plasma, which refers to the definition of plasma dose. Although there is no official definition of plasma dose, the dose-effect is crucial for applying plasma in clinical medicine, therefore the dose effect of ROS is studied in this part. Different concentrations of ROS are simulated by gradually increasing the number of active radicals in the simulation chamber to explore the dose effect on the amino acids. Here, taking Met upon the impact of O atoms as an example, the specific oxidative modification of Met is deeply investigated, with the number of O atoms increasing from 10 to 50.

Under the effect of 10 O atoms, the H-abstraction usually occurs on the methyl and carboxyl group of the Met, while the methyl group subsequently absorbs a hydroxyl in some cases. The thioether group on the side chain of Met is oxidized with the addition of one or two O atoms and the formation of $-S=O$ or $-S(=O)_2$. At low doses of O atoms, the predominant reaction involves the dehydrogenation of the thioether group and the addition of O atoms. Notably, the probability of one O atom addition is highest, reaching 80%. Upon the

interaction with 20 O atoms, it is found that the reactions are initiated by the H-abstraction of the methyl group, which subsequently leads to the detachment of the methyl group in the side chain of Met. The dissociation of the S-H bond as well as the formation of structure of $-\text{SO}(=\text{O})_2$ occurs in the thioether group of Met. The interactions primarily located at the H-abstraction of the thioether group and the carboxyl group, as well as the O-addition of the thioether group. The addition of two O atoms on the thioether group exists the highest probability, reaching a value of 80%. It is worth noting that O atoms can break structurally important C-C and C-S bonds under the effect of 30 O atoms. This subsequently triggers a cascade of bond dissociation events, leading to the formation of CO_2 or COH , SO_2 or SO_3 , respectively. Indeed, through sequential interactions, high-dose exposure can generate modification types that are not observable at low doses of ROS. As depicted in Fig. 6, the oxidation modification products depend on the number of oxygen atoms. When 40 O atoms are placed in the simulation box, the oxidative reactions of impinging O atoms become more intense with a higher probability of oxidative outputs, like the H-abstraction and further fragment of the methyl and the thioether groups. In addition to the H-abstraction reaction, the presence of a substantial quantity of O atoms induces oxidation and bond breakage of C and S atoms, with both the probability and proportion increasing. These oxidation modification products are not observed at low doses of O atoms. Similar to dosage 40, the oxidative modifications of Met lead to more violent reactions upon the impact of 50 O atoms. The concentration O atoms are highly reactive and can easily abstract H in carboxyl groups, which can also lead to the dissociation of the whole methyl groups to CO_2 , as well as the fragment of the thioether groups to SO_3 . Thus, with the increase of the number of oxygen atoms, the oxidative reactions of Met upon the impact of the O atom become increasingly intense, indicating that the dose of ROS is a crucial factor for the oxidative modification of amino acids. Additionally, the trend of oxidative modifications of Met under the action of increasing O atoms from 10 to 50 is also investigated and illustrated in Fig. 6. It is clear that the proportion of oxygen on the thioether group of Met continuously rises with the increase of O atoms from 10 to 50, which leads to the detachment of the thioether group at a high concentration of O atoms. The proportion of dissociation of C-H bonds on the carboxyl group rising to 100 % as the increasing of O atoms (see reaction mechanisms M1-3 and M10,11 in Fig. 6). The H-abstraction reaction of the methyl group occurs in the simulation of 10 O atoms participating. Then, the addition of O atoms to the methyl

group is observed as the concentration of O atoms increases, which subsequently triggers the formation of COH, the dissociation of the C–S bond, and the formation of CO₂. The simulation results indicate that the oxidation products of Met become more abundant with the increase of the dosage of O atoms within a specific range.

From the statistical analysis of the calculated results, it can be concluded that with the continuous increase of ROS dose, the changes in amino acid structure will significantly alter protein conformation, which will seriously affect the function of the protein and even inactivate the protein upon the impact of high concentration of ROS. The simulation conclusions are also in line with experimental data of the CAP effect on the protein molecule. The corresponding relationship between plasma dose and effect was also concluded by Graves [5]. The author reported that low-dose plasma can stimulate cell REDOX signals and promote cell proliferation and differentiation. A moderate dose leads to apoptosis, and high doses can directly cause cell necrosis. The ROS generated by plasma, called exogenous ROS, are the same as the ROS existing in living organisms. Intracellular ROS plays an essential role in cell signaling and balance maintenance and is a critical factor in inducing oxidative stress. Thus, under the plasma treatment, the ROS concentration around cells may exceed the limit of REDOX equilibrium, which leads to programmed cell death (apoptosis), and excessive ROS even cause more serious cell necrosis. Considering simulation and experimental results, the dose effect of plasma must be treated with caution before plasma medicine formally enters clinical applications.

In this section, the oxidative modifications of amino acids upon the impact of plasma-generated ROS are investigated by reaction MD simulation, and the oxidation reaction process between amino acids and ROS, the final oxidation products, as well as the kinds and dose effect of ROS are also provided at the molecular level. The modifications in the molecular structure of amino acid side chains will lead to the alternates in protein function, thus interfering with cell signaling processes. This work is of great significance for understanding the mechanisms of oxidative modification of amino acids and, consequently, protein inactivation under the influence of plasma treatment. However, some limitations of reactive MD simulation methods with ReaxFF potential are exposed during the calculation process. Due to the low computational efficiency of reactive MD simulations performed at microscopic time scales, the simulation results remain at the nanosecond level with a considerable amount of computational resources consumed. This limitation indicates that the reactive MD sim-

ulation takes immense computation costs and can not afford the simulation of long-lived active radicals even by performance computers. At the spatial scale, the simulation box volume restricts the molecular size and relative molecular mass of biological functional macromolecules. Therefore, exploring the effects of plasma on biological tissues based on reactive MD simulations still requires continuous development.

E. Prediction data from DNN

The discussion of predicted data by DNN is divided into five parts based on the given dose, and each group shows ten prediction results (two corresponding to each amino acid), seen Figs. 7–11. In these figures, the the horizontal coordinate of each subplot corresponds to the type of oxidative modification shown in Fig. 3, and the vertical coordinate is the predicted probability (percentage of data statistics corresponding type of modifications, while the DNN prediction may exceed 100% due to prediction error). The red straight lines represent the actual results, and the blue dashed lines are the predicted values. The heading of the subfigure indicates the name of the amino acid, as well as the type and dose of ROS that interact with this amino acid (e.g., Met–H₂O₂–10), and several significant data values are also recorded in the graphs. Although the current dataset of around 100 instances is still not very large, it takes two to three months to complete the calculation and collect the final data, which further demonstrates that the computational burden of reactive MD simulations is very heavy. This also highlights the necessity to introduce the new computational strategies in the study of plasma medicine. Furthermore, a single DNN using the entire collected data as the training set is also trained, which can rapidly provide almost any prediction of in-between-sample values. However, it should be noted that the predictions generated by this full-data trained DNN have not undergone rigorous validation and are therefore not discussed in detail here. For those interested in exploring this aspect further, detailed information and access to the open-source code are available.

In the first group of simulations, data with the ROS dose of 10 serves as the test set and 20–50 doses are used to train the DNN. The prediction results of oxidative products and corresponding probabilities of amino acids under the interaction with 10–dosage of ROS are shown in Fig. 7. The simulation results show that the mean square error between the predicted and actual values is 0.0037, with a high accuracy of over 80%, even though the test

set is outside the range of training data. Since abundant zeros in the structured vector are misleading as to the actual errors, which contain only the products corresponding to each reactants, the actual errors are recalculated manually, ignoring those of other oxidation modifications out of consideration. The error for each amino acid-ROS interaction is calculated only for the mean square error of the corresponding oxidation modification products. The recalculated mean squared error is more authentic and reasonable, closely aligning with the subplots presented in the results. With the most abundant type of oxidative modification products, the prediction errors of Met upon the impact of H_2O_2 and O are 0.0122 and 0.0248, respectively. Under the effect of H_2O_2 , the predicted possibility of the H-abstraction of the carboxyl group (C_6) and the absorption of one O atom on the thioether group (R-S-R) of Met (C_1) have a significant deviation from the actual value. The major deviations occur in the addition of one or two O atoms on the thioether group of Met (C_1 , C_2) upon the impact of O atoms, both of which are caused by the corresponding probability of training set different from 10-dose. The predicted results of Phe are basically consistent with the actual values under the interaction with O_3 , with an error of 0.0062, and only the H-abstraction on the methylene group (D_{11}) is significantly different. Upon the interaction with OH atoms, only two types of oxidative modifications of Phe are produced (D_1 , D_2), and the predictions have several deviations with an error of 0.0297. As another sulfur-containing amino acid, the prediction of oxidative modifications of Cys agrees with the actual products, but some deviations exist in the probability values. The main differences located at the H-abstraction and O-addition on the sulfhydryl group (E_3) and the H-abstraction on the carboxyl group (E_9) upon the impact of O_3 , while the latter is also the most significant aberration upon the impact of O. The prediction errors of Cys are 0.0111 and 0.0082, respectively. Compared with the previous results, the DNN has an excellent prediction effect for Glu and Pro with fewer oxidatively modified products. Significantly, under the impact of OH and H_2O_2 , the error between the predicted value and the actual value of Glu are 0.0018 and 0.0009, respectively. In the product prediction of Pro, the H-abstraction of amino and carboxyl group (G_1 , G_2) show prediction errors upon the impact of O_3 and O, with a value of 0.0218 and 0.0045.

Then, 20-dosage is selected as the test set and the other doses of ROS as the training set, and the prediction results are shown in Fig. 8. Compared with the 10-dosage group, the error between the predicted and actual values is 0.0028, with a higher accuracy. Under the

interaction with H_2O_2 , four modified products of Met are generated, and the prediction errors mainly occur in the absorption of an O atom on the thioether group (C_1) and H-abstraction of the carboxyl group (C_6) with a value of 0.0037. The prediction results of most of the oxidative products of Met under the impact of O_3 are accurate with an error of 0.0082, but the deviation is significant in the addition of two O atoms on the thioether group (C_2) and the dissociation of C-S bond to form SO_3 (C_{19}). The former deviation is caused by the fact that only at a specific 20 doses with high product probability, the dose below 20 can not form sulfoxide groups due to poor oxidative effect, while the dose over 20 with strong oxidation will lead to the other modifications such as sulfonyl groups and even the detachment of S atom. Thus, the prediction of C_2 is much lower compared with actual value. The fluctuation of the corresponding product causes the latter deviation. The oxidation predictions of Phe are different upon the impact of O_3 and OH with errors of 0.0104 and 0.0072. The prediction results of the H-abstraction of the methylene group of Phe (D_{11}) under the effect of O_3 has a remarkable deviation, while the H-abstraction of the amino group (D_9) under the effect of OH radicals has a glaring miscalculation. The predicted value of amino dehydrogenation is zero due to such oxidative modification only exists in 20 doses of OH in the data set, and the training set lacks corresponding data. Therefore, the DNN is incapable of predicting oxidation types with data lacking. The predicted errors of Cys under the effect of OH and O are 0.0123 and 0.0158, respectively. For the prediction of Cys upon the interaction with OH, there are some errors on several major products (E_1 , E_2 , E_3 , E_4 , and E_9). Under the interaction with O, the prediction deviation mainly focuses on the addition of two O atoms with H-abstraction on the sulfhydryl group (E_3) and the fragment of S atom (E_{13}). The results of Glu and Pro are relatively accurate. Under the impact of O and H_2O_2 , the prediction errors between the predicted value and the actual value of Glu are 0.0094 and 0.0019, while the errors of Pro are 0.0055 and 0.0201 under the effect of O_3 and O, respectively. The main error in the prediction focuses on the H-abstraction of the carboxyl group of Pro (G_2).

The testing group of 30-dosage ROS is located at the middle value of the training data range (10–50), which obtains the most accurate prediction with an overall error of 0.0019. As shown in Fig. 9, the oxidative modifications of Met upon the impact of OH are basically in line with H_2O_2 , and both have relatively precise predictions with errors of 0.0011 and 0.0026, respectively. Under the impact of H_2O_2 , the prediction errors mainly focus on the

O–addition of the thioether group (C_1 , C_2), the H–abstraction of the carboxyl group (C_6) and these modifications on the amino acids (C_{21}) of Met. Upon the interaction with 30 OH radicals, the main oxidative modifications of Met can be accurately predicted (C_1 , C_2 , and C_6), and part of the errors exist in the types of oxidative modification that did not actually occur (C_8 , C_{10} , and C_{12}). The oxidative products of Phe change particularly under the effect of O_3 and OH. Under the impact of O_3 , Phe has various oxidation modifications, while the prediction errors of three main modifications are relatively acceptable, and the prediction of the detachment of the carboxyl group to CO_2 (D_{13}) has a significant deviation. The total error of interaction between Phe and 30 O_3 molecules is 0.0099. In contrast, only three oxidation modifications occur in the interaction with OH radicals. The differences between the predicted and actual values are negligible, and the main deviation is concentrated on the prediction of oxidative modifications that did not occur (D_8 , D_9) and has an error of 0.0038. The oxidative prediction errors of Cys under the action of O_3 and OH are 0.0143 and 0.0043. The predicted deviations mainly reflect in the detachment of the sulfhydryl group to form SO_2 (E_{14}) under the impact of O_3 and in the H–abstraction and O–absorption (E_3) of the sulfhydryl group to form the sulfinic acids group, respectively. The prediction of Glu is in excellent agreement with the actual value, primarily upon the impact of O atoms, and the prediction error is only 0.0002. Under the effect of H_2O_2 , the error is slightly higher with a value of 0.0005. The prediction deviation mainly occurs in the fragment of the amino group (F_3). Under the impact of H_2O_2 and O, the prediction deviations of Pro are 0.0036 and 0.0051, respectively. The predicted error mainly focuses on the H–abstraction of the carboxyl group (G_2).

As the dose of ROS increases, the oxidative reaction approaches the critical value, and the oxidation modifications tend to be stable. Fig. 10 demonstrates the prediction of 40–dosage ROS with an overall predicted error of 0.0027, and the predicted accuracy is only inferior to the 30–dosage group. The figure indicates that the prediction of oxidative modifications of Met under the impact of O_3 and O has a high degree of accuracy with errors of 0.0221 and 0.0082. Upon the impact of O_3 , the considerable errors focus on the detachment of the methyl and the thioether groups of Met (C_{12} , C_{13} , C_{18} , and C_{19}). Under the effect of O, the main oxidative modifications are roughly predicted, but the predictions of absorption of two or three O atoms on the thioether group and the fragment of the thioether group (C_{18} and C_{19}) have a specific deviation. The types of oxidative modifications of Phe under the effect

of 40 O₃ molecules and OH radicals are consistent with the 30–dosage group and have errors of 0.0272 and 0.0011, respectively. It is worth noting that the influences of 30–dosage and 40–dosage to each other, such as the H–abstraction of the carboxyl, benzene, and amino groups (D₂, D₅, and D₉) of Phe under the interaction with 40 O₃ molecules. Only four types of oxidative modifications of Phe occur under the impact of OH, and the prediction is precise except for the OH-addition (D₈). The predicted probabilities of Cys under the interaction with OH radicals have little deviation from the actual value, with a prediction error of 0.0009. However, the predicted deviations in the absorption of three O atoms on the sulfhydryl group of Cys (E₄), even detachment of the sulfhydryl group (E₁₅) upon the interaction with 40 O atoms, show a significant error of 0.0159. The predicted modifications of Glu upon the effect of O atoms show little deviation from the actual value, and the error is only 0.0002. However, few deviations occur in the prediction under the impact of H₂O₂, especially in predicting the H–abstraction of the carboxyl groups (F₁) and the detachment of the amino group (F₃). The primary reason is that this modification only exists in 50 doses, affecting the current training set. In addition, the 50–dosage group is also influenced during testing, and the prediction of the corresponding product is also biased. The predictions of Pro upon the impact of OH and H₂O₂ are also presented here, with errors of 0.0006 and 0.0087, and the main deviation focuses on the H–abstraction from carboxyl group (G₈).

Finally, the oxidative modification predictions of 50–dosage of ROS are illustrated in Fig. 11. Although this group is also outside the scope of the corresponding training set, the oxidative modifications tend to be saturated, and the predicted results are more accurate than the 10–dosage group. First, the modified predictions of Met upon the impact of OH and O are displayed with errors of 0.0057 and 0.0169. Under the effect of OH, the prediction errors of Met mainly lie in the dehydrogenation of the carboxyl group and the absorption of two O atoms on the thioether group (C₂ and C₆). The DNN also misjudges the probabilities of H–abstraction and O–addition of the methyl group and the H–abstraction of the amino group (C₁₀ and C₂₀). Upon the interaction with 50 O atoms, the prediction errors mainly focus on the addition of three oxygen atoms in the thioether group, the fragmentation of the methyl group, and the thioether group (C₃, C₁₃, C₁₈, and C₁₉). The figure also presents the oxidative modifications of Phe upon the impact of OH and H₂O₂, and the prediction errors are 0.0034 and 0.0068, respectively. The former slightly differs from the actual value for the H–abstraction of the carboxyl group (D₂). In contrast, the latter has only two oxidative

modified products (D1 and D2), with significant deviation predictions. The deviation between the predicted and actual value of Cys under the impact of OH is acceptable with an error of 0.0017, and the deviation mainly focuses on the H–abstraction and the absorption of three O atoms in the sulfhydryl group (E_4). However, under the interaction with 50 O_3 molecules, there are significant errors in the detachment of the sulfhydryl group of Cys and the oxidative modifications of the amino group (E_{14-18}), with a total error of 0.0295. For the oxidative modifications of Glu, the predicted results upon the impact of O_3 and H_2O_2 are discussed. Under the effect of O_3 , all three oxidative modifications (i.e., F1, F2 and F4) have predicted deviations and the prediction error is 0.0313. The reasons for the deviation are consistent with the results discussed in the previous paragraph. Under the action of H_2O_2 , there is an inevitable error in the prediction of the H–abstraction of the carboxyl group of Glu (F_1), and the error value is 0.0109. Under the impact of OH radicals, the prediction of the oxidative modifications of Pro is relatively accurate, while under the effect of O atoms, the prediction has a specific deviation on the H–abstraction of the carboxyl group of Pro (G_2). The errors in the interaction of Pro with OH and O are 0.0041 and 0.0031. Thus, the DNN surrogate model holds the capability of rapidly predicting oxidative modifications of amino acids in only seconds, and an accuracy of over 80%, which indicates that DNN has tremendous potential value to effectively predict the processes of chemical breaking and formation based on the data from reactive MD simulations.

IV. CONCLUSION

In this paper, the reactive MD simulation with the ReaxFF potential is employed to provide dataset on the oxidative mechanisms of four kinds of ROS (O atoms, OH radicals, O_3 and H_2O_2 molecules) interacting with five types of amino acids (Met, Phe, Cys, Pro, Glu). A DNN is constructed by introducing the new data structure to represent the reactions and trained by the reaction data from reactive MD simulations, then performed to investigate the prediction of oxidative modifications of five amino acids with various ROS and corresponding dosages. The simulation results and prediction data indicated that the oxidative reactions are usually initiated by H-abstraction, and the oxidative modifications, such as the detachment of amino and carboxyl groups, the insertion of functional groups (aldehyde and alcohol groups), and the fragmentation of S atoms in sulfur-containing amino

acids, can be observed from the simulation data with the increase of ROS dose and variation of active radicals, which agree well with the experimental observation qualitatively. By analyzing plenty of reaction data, the chemical structures can be alternated into vector representations through proper data processing. Based on the collocated data, the structure tensors are imported into a DNN model to efficiently predict the oxidative modifications and corresponding probabilities under the impact of ROS. The results revealed that the well-trained DNN model is capable of rapidly predicting oxidative modifications of amino acids in seconds of computational time and with an accuracy of over 80%, indicating the tremendous potential in chemical prediction. The present simulation results provide a deep insight into the interactions of CAP with amino acids, which could promote the integration and development of plasma pharmacy and plasma medicine combined with DNN. This study demonstrates that DNNs can play a complementary and enhancing role, optimizing reactive MD simulations to significantly improve the computational efficiency, and suggesting the great potential of ML technology in the exploration of plasma medicine.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Nos. 11975142 and 12375201).

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request, and the Python code of DNN used in this study can be accessed via <https://github.com/SDU-HV-Plasma/Plasma-medicine.git>

-
- [1] B. R. Locke, M. Sato, P. Sunka, M. R. Hoffmann, and J. S. Chang, *Industrial & Engineering Chemistry Research* **45**, 882 (2006).
- [2] M. Sahni and B. R. Locke, *Industrial & Engineering Chemistry Research* **45**, 5819 (2006).
- [3] R. Burlica, K. Y. Shih, and B. R. Locke, *Industrial & Engineering Chemistry Research* **49**, 6342 (2010).
- [4] A. Tani, Y. Ono, S. Fukui, S. Ikawa, and K. Kitano, *Applied Physics Letters* **100**, 254103 (2012).
- [5] D. B. Graves, *Physics of Plasmas* **21** (2014).
- [6] P. Bruggeman, F. Iza, D. Lauwers, and Y. A. Gonzalvo, *Journal of Physics D: Applied Physics* **43**, 012003 (2009).
- [7] P. Bruggeman and C. Leys, *Journal of Physics D: Applied Physics* **42**, 053001 (2009).
- [8] S. Ikawa, K. Kitano, and S. Hamaguchi, *Plasma Processes and Polymers* **7**, 33 (2010).
- [9] D. X. Liu, M. Z. Rong, X. H. Wang, F. Iza, M. G. Kong, and P. Bruggeman, *Plasma Processes and Polymers* **7**, 846 (2010).
- [10] M. Sahni and B. R. Locke, *Plasma Processes and Polymers* **3**, 668 (2006).
- [11] S. Emmert, F. Brehmer, H. Hänßle, A. Helmke, N. Mertens, R. Ahmed, D. Simon, D. Wandke, W. Maus-Friedrichs, and G. Däschlein, *Clinical Plasma Medicine* **1**, 24 (2013).
- [12] J. Heinlin, G. Isbary, W. Stolz, G. Morfill, M. Landthaler, T. Shimizu, B. Steffes, T. Nosenko, J. L. Zimmermann, and S. Karrer, *Journal of the European Academy of Dermatology and Venereology* **25**, 1 (2011).
- [13] G. Isbary, G. Morfill, H. Schmidt, M. Georgi, K. Ramrath, J. Heinlin, S. Karrer, M. Landthaler, T. Shimizu, and B. Steffes, *British Journal of Dermatology* **163**, 78 (2010).
- [14] G. Isbary, J. Heinlin, T. Shimizu, J. Zimmermann, G. Morfill, H.-U. Schmidt, R. Monetti, B. Steffes, W. Bunk, and Y. Li, *British Journal of Dermatology* **167**, 404 (2012).
- [15] T. Sato, M. Yokoyama, and K. Johkura, *Journal of Physics D: Applied Physics* **44**, 372001 (2011).
- [16] C. H. Kim, S. Kwon, J. H. Bahn, K. Lee, S. I. Jun, P. D. Rack, and S. J. Baek, *Applied Physics Letters* **96**, 243701 (2010).

- [17] S. J. Kim, T. H. Chung, S. H. Bae, and S. H. Leem, *Applied Physics Letters* **97**, 023702 (2010).
- [18] K. Kim, J. D. Choi, Y. C. Hong, G. Kim, E. J. Noh, J.-S. Lee, and S. S. Yang, *Applied Physics Letters* **98**, 073701 (2011).
- [19] N. Barezzi and M. Laroussi, *Plasma Processes and Polymers* **10**, 1039 (2013).
- [20] H. Yamazaki, T. Ohshima, Y. Tsubota, H. Yamaguchi, J. A. Jayawardena, and Y. Nishimura, *Dental Materials Journal* **30**, 384 (2011).
- [21] G. Fridman, M. Peddinghaus, M. Balasubramanian, H. Ayan, A. Fridman, A. Gutsol, and A. Brooks, *Plasma Chemistry and Plasma Processing* **26**, 425 (2006).
- [22] B. Gweon, D. Kim, D. B. Kim, H. Jung, W. Choe, and J. H. Shin, *Applied Physics Letters* **96**, 101501 (2010).
- [23] S. Perni, G. Shama, J. Hobman, P. Lund, C. Kershaw, G. Hidalgo-Arroyo, C. Penn, X. T. Deng, J. L. Walsh, and M. G. Kong, *Applied Physics Letters* **90**, 073902 (2007).
- [24] H. Feng, R. Wang, P. Sun, H. Wu, Q. Liu, J. Fang, W. Zhu, F. Li, and J. Zhang, *Applied Physics Letters* **97**, 131501 (2010).
- [25] R. Bussiahn, R. Brandenburg, T. Gerling, E. Kindel, H. Lange, N. Lembke, K.-D. Weltmann, T. Von Woedtke, and T. Kocher, *Applied Physics Letters* **96**, 143701 (2010).
- [26] J. F. Kolb, A. A. H. Mohamed, R. O. Price, R. J. Swanson, A. Bowman, R. L. Chiavarini, M. Stacey, and K. H. Schoenbach, *Applied Physics Letters* **92**, 241501 (2008).
- [27] X. Lu, T. Ye, Y. Cao, Z. Sun, Q. Xiong, Z. Tang, Z. Xiong, J. Hu, Z. Jiang, and Y. Pan, *Journal of Applied Physics* **104**, 053309 (2008).
- [28] X. Zhang, J. Huang, X. Liu, L. Peng, L. Guo, G. Lv, W. Chen, K. Feng, and S. Yang, *Journal of Applied Physics* **105**, 063302 (2009).
- [29] A. Majumdar, R. K. Singh, G. J. Palm, and R. Hippler, *Journal of Applied Physics* **106**, 084701 (2009).
- [30] R. Burlica, R. G. Grim, K. Y. Shih, D. Balkwill, and B. R. Locke, *Plasma Processes and Polymers* **7**, 640 (2010).
- [31] E. Takai, S. Ikawa, K. Kitano, J. Kuwabara, and K. Shiraki, *Journal of Physics D: Applied Physics* **46**, 295402 (2013).
- [32] M. Nagatsu, Y. Zhao, I. Motrescu, R. Mizutani, Y. Fujioka, and A. Ogino, *Plasma Processes and Polymers* **9**, 590 (2012).

- [33] E. Takai, T. Kitamura, J. Kuwabara, S. Ikawa, S. Yoshizawa, K. Shiraki, H. Kawasaki, R. Arakawa, and K. Kitano, *Journal of Physics D: Applied Physics* **47**, 285403 (2014).
- [34] N. Khosravian, B. Kamaraj, E. C. Neyts, and A. Bogaerts, *Scientific Reports* **6**, 1 (2016).
- [35] M. Dizdaroglu and P. Jaruga, *Free Radical Research* **46**, 382 (2012).
- [36] J. A. Reisz, N. Bansal, J. Qian, W. Zhao, and C. M. Furdai, *Antioxidants & Redox Signaling* **21**, 260 (2014).
- [37] C. C. W. Verlackt, E. Neyts, T. Jacob, D. Fantauzzi, M. Golkaram, Y. K. Shin, A. C. T. Van Duin, and A. Bogaerts, *New Journal of Physics* **17**, 103005 (2015).
- [38] J. Van der Paal, C. Verlackt, M. Yusupov, E. Neyts, and A. Bogaerts, *Journal of Physics D: Applied Physics* **48**, 155202 (2015).
- [39] C. L. Hawkins and M. J. Davies, *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1504**, 196 (2001).
- [40] G. Xu and M. R. Chance, *Chemical Reviews* **107**, 3514 (2007).
- [41] Y.-H. Ding, X.-L. Wang, S.-Q. Tian, S.-R. Li, L. Li, Q.-X. Li, T. Zhao, and Y.-T. Zhang, *Plasma Processes and Polymers* **20**, 2200148 (2023).
- [42] J.-S. Guo, S.-Q. Tian, and Y.-T. Zhang, *Physics of Plasmas* **30** (2023).
- [43] E. Takai, K. Kitano, J. Kuwabara, and K. Shiraki, *Plasma Processes and Polymers* **9**, 77 (2012).
- [44] A. Sharifian, N. Soltanizadeh, and R. Abbaszadeh, *Innovative Food science & Emerging Technologies* **54**, 1 (2019).
- [45] R. Zhou, R. Zhou, J. Zhuang, Z. Zong, X. Zhang, D. Liu, K. Bazaka, and K. Ostrikov, *PloS One* **11**, e0155584 (2016).
- [46] M. Yusupov, A. Bogaerts, S. Huygh, R. Snoeckx, A. C. T. Van Duin, and E. C. Neyts, *The Journal of Physical Chemistry C* **117**, 5993 (2013).
- [47] T. Zhao, L. Shi, Y. T. Zhang, L. Zou, and L. Zhang, *Physics of Plasmas* **24**, 103518 (2017).
- [48] A. R. Brodtkorb, T. R. Hagen, and M. L. Sætra, *Journal of Parallel and Distributed Computing* **73**, 4 (2013).
- [49] M. Pandey, M. Fernandez, F. Gentile, O. Isayev, A. Tropsha, A. C. Stern, and A. Cherkasov, *Nature Machine Intelligence* **4**, 211 (2022).
- [50] A. Eklund, P. Dufort, D. Forsberg, and S. M. LaConte, *Medical Image Analysis* **17**, 1073 (2013).

- [51] U. K. Ercan, G. D. Özdemir, M. A. Özdemir, and O. Güren, *Plasma Processes and Polymers* , e2300066 (2023).
- [52] R. Anirudh, R. Archibald, M. S. Asif, M. M. Becker, S. Benkadda, P.-T. Bremer, R. H. Budé, C.-S. Chang, L. Chen, R. Churchill, *et al.*, arXiv preprint arXiv:2205.15832 (2022).
- [53] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, *Reviews of Modern Physics* **91**, 045002 (2019).
- [54] A. Piccione, J. W. Berkery, S. Sabbagh, and Y. Andreopoulos, *Nuclear Fusion* **60**, 046033 (2020).
- [55] Y. Fu, D. Eldon, K. Erickson, K. Kleijwegt, L. Lupin-Jimenez, M. D. Boyer, N. Eidietis, N. Barbour, O. Izacard, and E. Kolemen, *Physics of Plasmas* **27**, 022501 (2020).
- [56] A. Mesbah and D. B. Graves, *Journal of Physics D: Applied Physics* **52**, 30LT02 (2019).
- [57] P. Daoutidis, J. H. Lee, S. Rangarajan, L. Chiang, B. Gopaluni, A. M. Schweidtmann, I. Harjunkoski, M. Mercangöz, A. Mesbah, F. Boukouvala, *et al.*, *Computers & Chemical Engineering* , 108523 (2023).
- [58] J. Han, A. Jentzen, and W. E, *Proceedings of the National Academy of Sciences* **115**, 8505 (2018).
- [59] Y. T. Zhang, S. H. Gao, and F. Ai, *Frontiers in Physics* **11**, 50 (2023).
- [60] Y. T. Zhang, S. H. Gao, and Y. Y. Zhu, *Journal of Applied Physics* **133** (2023).
- [61] V. D. Ramaswamy and M. Keidar, *Applied Sciences* **14**, 355 (2023).
- [62] V. R. Devi and M. Keidar, (2023).
- [63] X. C. Wang and Y. T. Zhang, *Journal of Applied Physics* **133** (2023).
- [64] M. Ghasemtarei, T. Ghorbi, M. Yusupov, Y. Zhang, T. Zhao, P. Shali, and A. Bogaerts, *Biomolecules* **13**, 1371 (2023).
- [65] Y. Li, S. Tan, D. Liu, and Y. Zhang, *Plasma Processes and Polymers* , e2300119.
- [66] A. Bogaerts, M. Yusupov, J. Van der Paal, C. C. W. Verlackt, and E. C. Neyts, *Plasma Processes and Polymers* **11**, 1156 (2014).
- [67] J. Van der Paal, S. Aernouts, A. C. Van Duin, E. C. Neyts, and A. Bogaerts, *Journal of Physics D: Applied Physics* **46**, 395201 (2013).
- [68] A. Bogaerts, N. Khosravian, J. Van der Paal, C. C. Verlackt, M. Yusupov, B. Kamaraj, and E. C. Neyts, *Journal of Physics D: Applied Physics* **49**, 054002 (2015).

- [69] A. C. Van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard, *The Journal of Physical Chemistry A* **105**, 9396 (2001).
- [70] R. M. Abolfath, P. Biswas, R. Rajnarayanan, T. Brabec, R. Kodym, and L. Papiez, *The Journal of Physical Chemistry A* **116**, 3940 (2012).
- [71] R. M. Abolfath, A. C. T. Van Duin, and T. Brabec, *The Journal of Physical Chemistry A* **115**, 11045 (2011).
- [72] A. L. Santos and A. B. Lindner, *Oxidative Medicine and Cellular Longevity* **2017** (2017).
- [73] T. M. Karve and A. K. Cheema, *Journal of Amino Acids* **2011** (2011).
- [74] X. Yu, S. Huang, C. Nie, Q. Deng, Y. Zhai, and R. Shen, *Journal of Food Science* **85**, 2010 (2020).
- [75] C. C. W. Verlackt, W. Van Boxem, D. Dewaele, F. Lemiere, F. Sobott, J. Benedikt, E. C. Neyts, and A. Bogaerts, *The Journal of Physical Chemistry C* **121**, 5787 (2017).
- [76] S. Wenske, J. W. Lackmann, S. Bekeschus, K. D. Weltmann, T. Von Woedtke, and K. Wende, *Biointerphases* **15**, 061008 (2020).
- [77] D. Ellerweg, J. Benedikt, A. von Keudell, N. Knake, and V. Schulz-von der Gathen, *New Journal of Physics* **12**, 013021 (2010).
- [78] Y. T. Zhang, Y. Y. Chi, and J. He, *Plasma Processes and Polymers* **11**, 639 (2014).
- [79] Y. T. Zhang and J. He, *Physics of Plasmas* **20**, 013502 (2013).
- [80] T. Hey, *The fourth paradigm* (United States of America., 2009).
- [81] C. Ma, B. Zhu, X.-Q. Xu, and W. Wang, *Physics of Plasmas* **27** (2020).
- [82] R. Maulik, N. A. Garland, J. W. Burby, X.-Z. Tang, and P. Balaprakash, *Physics of Plasmas* **27** (2020).
- [83] Z.-B. Liu, X.-C. Wang, and Y.-T. Zhang, *IEEE Transactions on Plasma Science* (2023).
- [84] S. M. Kearnes, M. R. Maser, M. Wlekinski, A. Kast, A. G. Doyle, S. D. Dreher, J. M. Hawkins, K. F. Jensen, and C. W. Coley, *Journal of the American Chemical Society* **143**, 18820 (2021).
- [85] Y. Bengio, A. Courville, and P. Vincent, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**, 1798 (2013).
- [86] D. Perera, J. W. Tucker, S. Brahmabhatt, C. J. Helal, A. Chong, W. Farrell, P. Richardson, and N. W. Sach, *Science* **359**, 429 (2018).
- [87] L. B. Poole, *Free Radical Biology and Medicine* **80**, 148 (2015).

- [88] S. Wenske, J.-W. Lackmann, L. M. Busch, S. Bekeschus, T. von Woedtke, and K. Wende, *Journal of Applied Physics* **129**, 193305 (2021).
- [89] A. B. Shekhter, A. V. Pekshev, A. B. Vagapov, A. V. Butenko, A. L. Fayzullin, T. G. Rudenko, N. A. Sharapov, N. B. Serejnikova, and V. N. Vasilets, *Clinical Plasma Medicine* **19**, 100101 (2020).
- [90] D. Gidon, D. B. Graves, and A. Mesbah, *Plasma Sources Science and Technology* **28**, 025006 (2019).
- [91] H. Cheng, J. Xu, X. Li, D. Liu, and X. Lu, *Physics of Plasmas* **27** (2020).

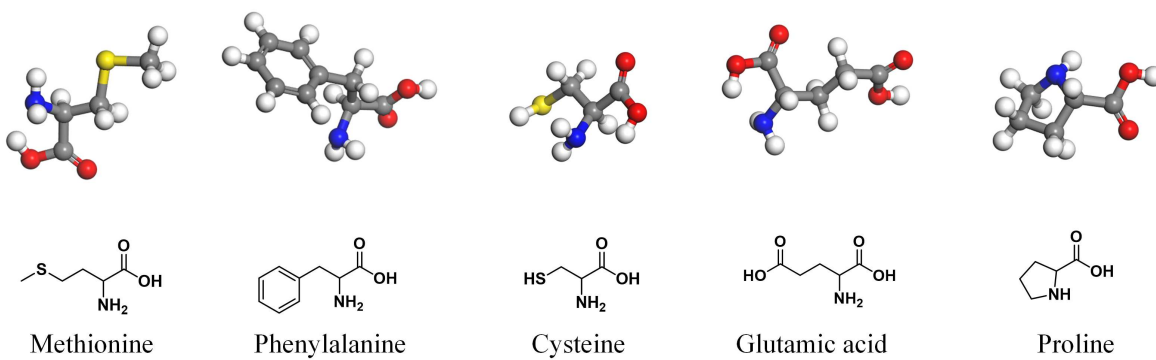


FIG. 1. Schematic representation (3D and 2D) of the structures of five model amino acids. Atoms of H, C, N, O and S are represented by white, gray, blue, red and yellow spheres, respectively.

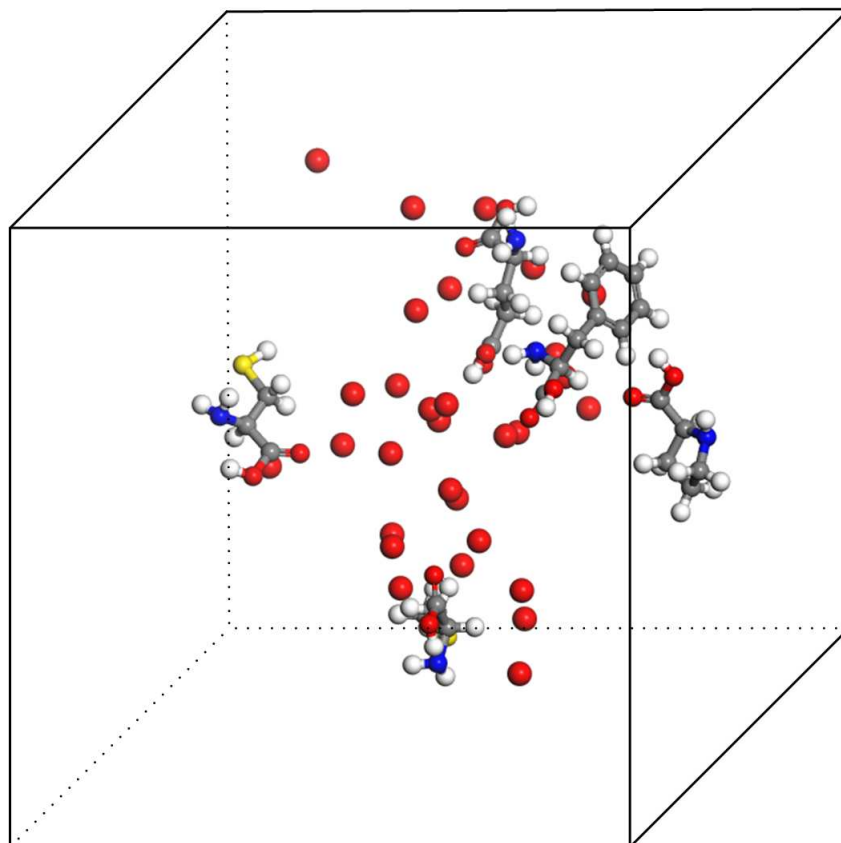


FIG. 2. Example of simulation box consisting of five amino acids (i.e., Met, Cys, Phe, Glu and Phe) and 30 O atoms with periodic boundary conditions applied in all three directions.

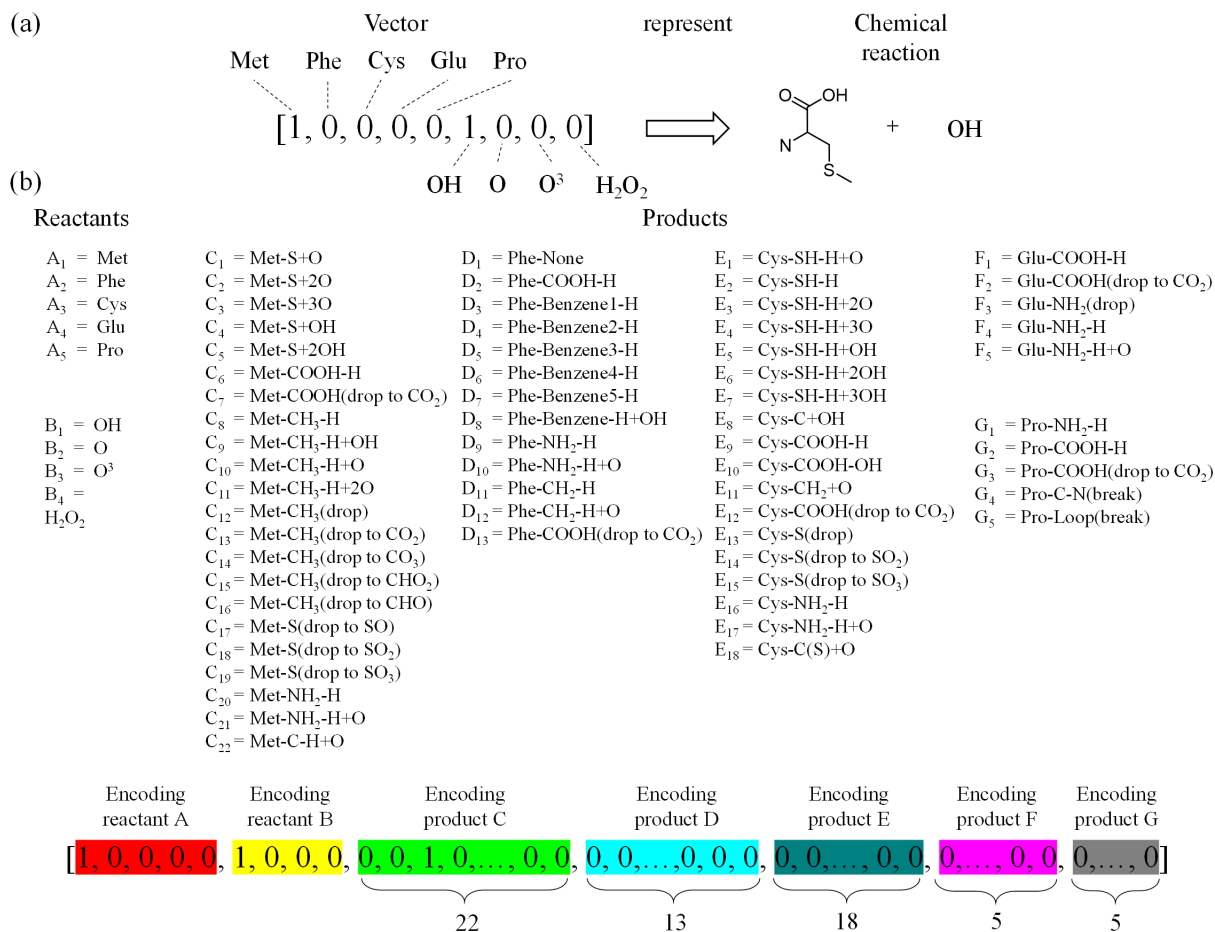


FIG. 3. (a) The example of reaction space representation of five amino acids and four ROS using vectors and (b) exploring the Suzuki–Miyaura reaction space using our surrogate machine learning models. Shown are the identity of reactants and products, and the vector representation of the oxidative reaction. The template of the product consists of three parts: the first part indicates the type of reactant, the second part represents the site or group of oxidative modification, and the third part denotes the type of oxidative modification, such as H–abstraction, O–addition, or abscission of the group.

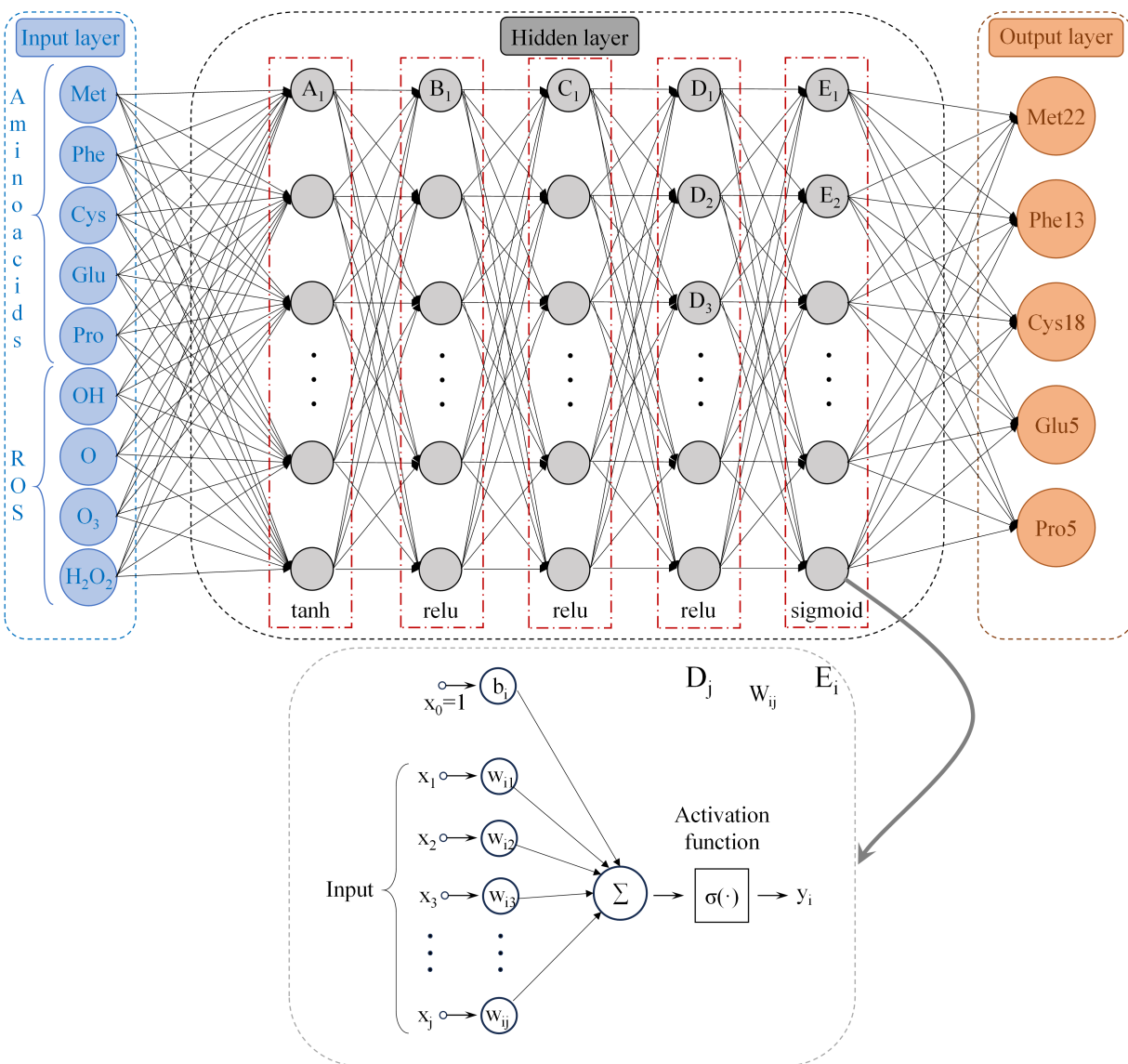


FIG. 4. Framework of deep neural network constructed based on the characteristics of amino acids oxidation reaction with five hidden layer and the operational logic of each neuron of hidden layer.

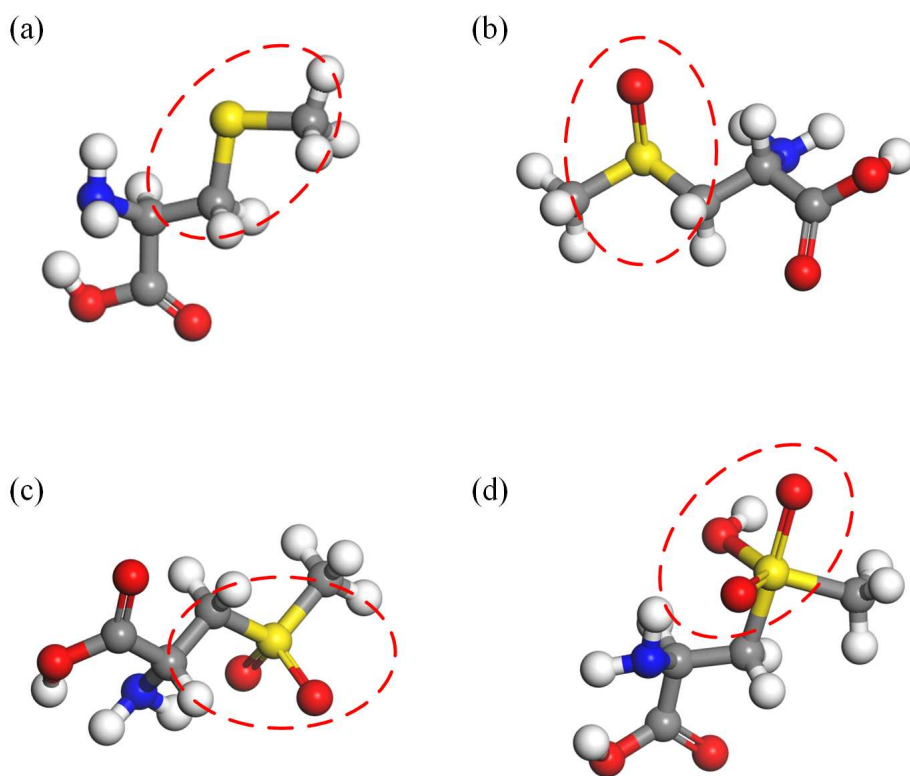


FIG. 5. The O-addition in the thioether group of Met upon the impact of O atoms. The structure of Met is shown (a), and the addition of one, two, and three O atoms in the thioether group of Met to form the structure of $-S = O$, $-S(=O)_2$, and $-SO(=O)_2$ are illustrated in (b), (c), and (d).

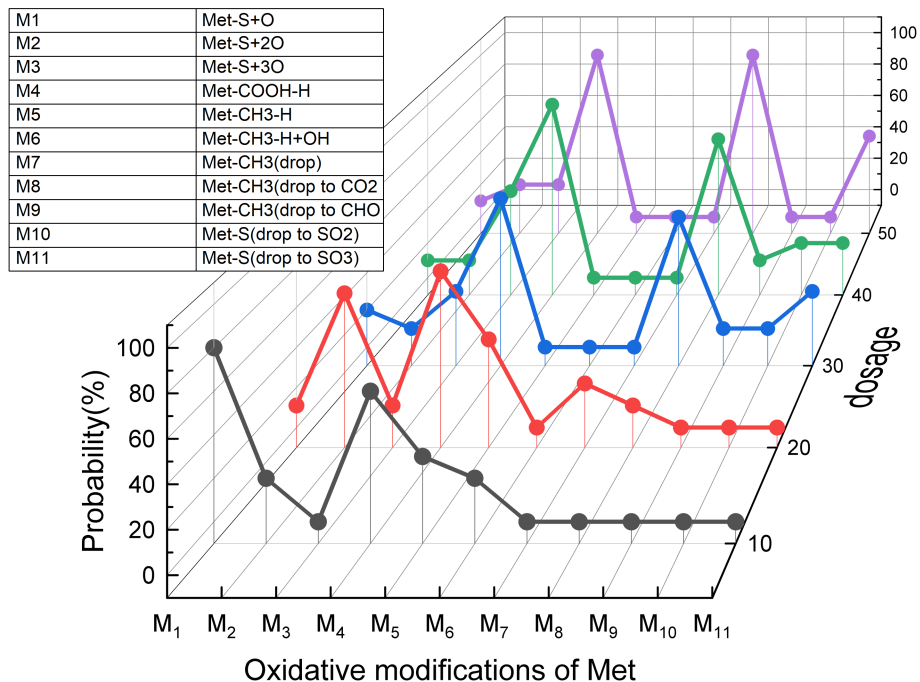


FIG. 6. Variation in interaction of Met with O atoms while the dosage ranging from 10 to 50. The x-axis indicates the types of oxidative modifications of Met, the y-axis denotes the dosage of O atoms, and the z-axis refers to the probability (yield).

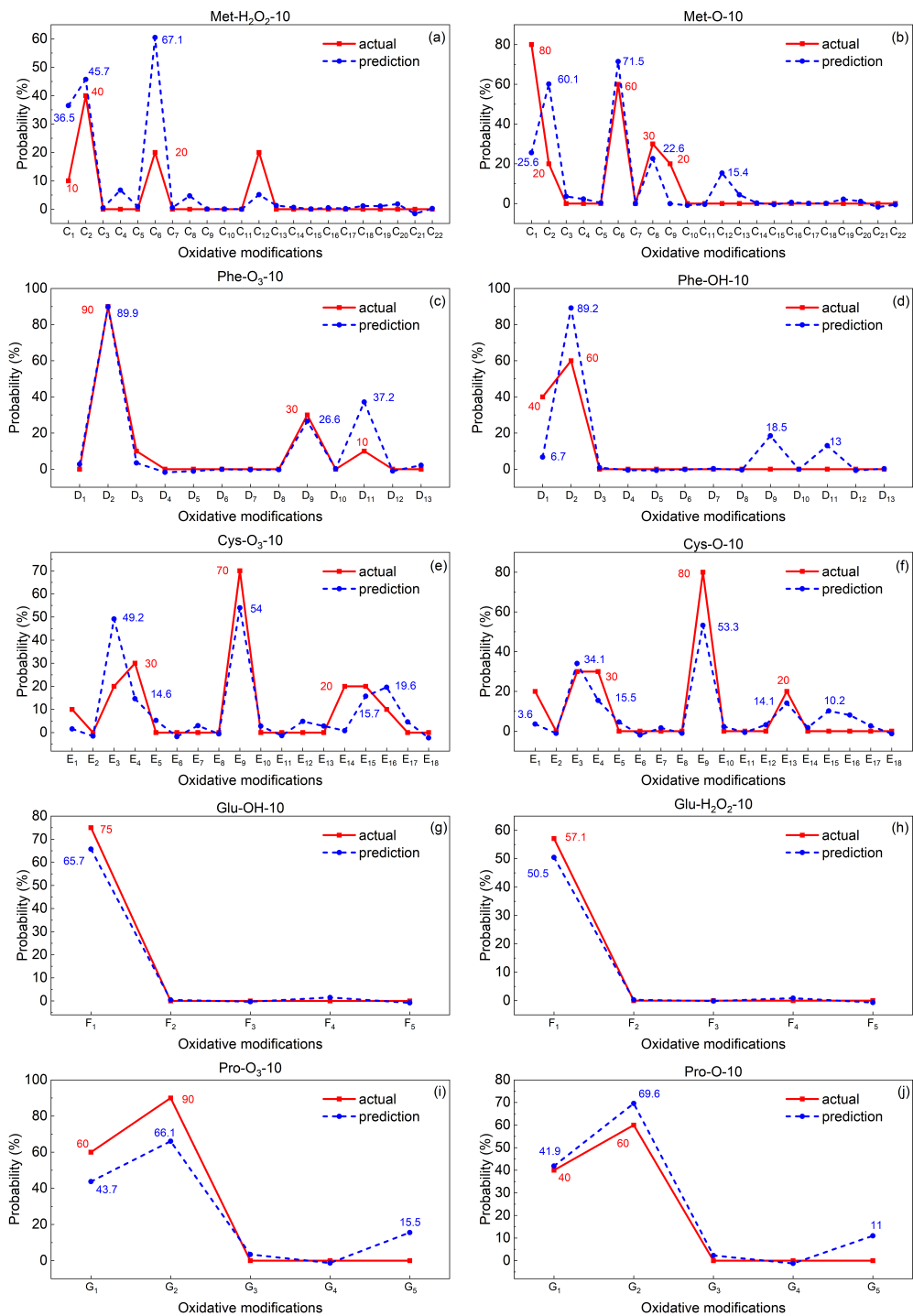


FIG. 7. Diagram of partial prediction results of 10 doses of ROS interacting with various amino acids. The x-axis indicates the types of oxidative modification corresponding to the products displayed in Fig. 3, respectively, and the y-axis denotes the probability (yield). (a) depicts the interaction of Met with H₂O₂, labeled at the top of the subfigure. The rest of (b-j) are consistent with (a). The same applies to the figures below.

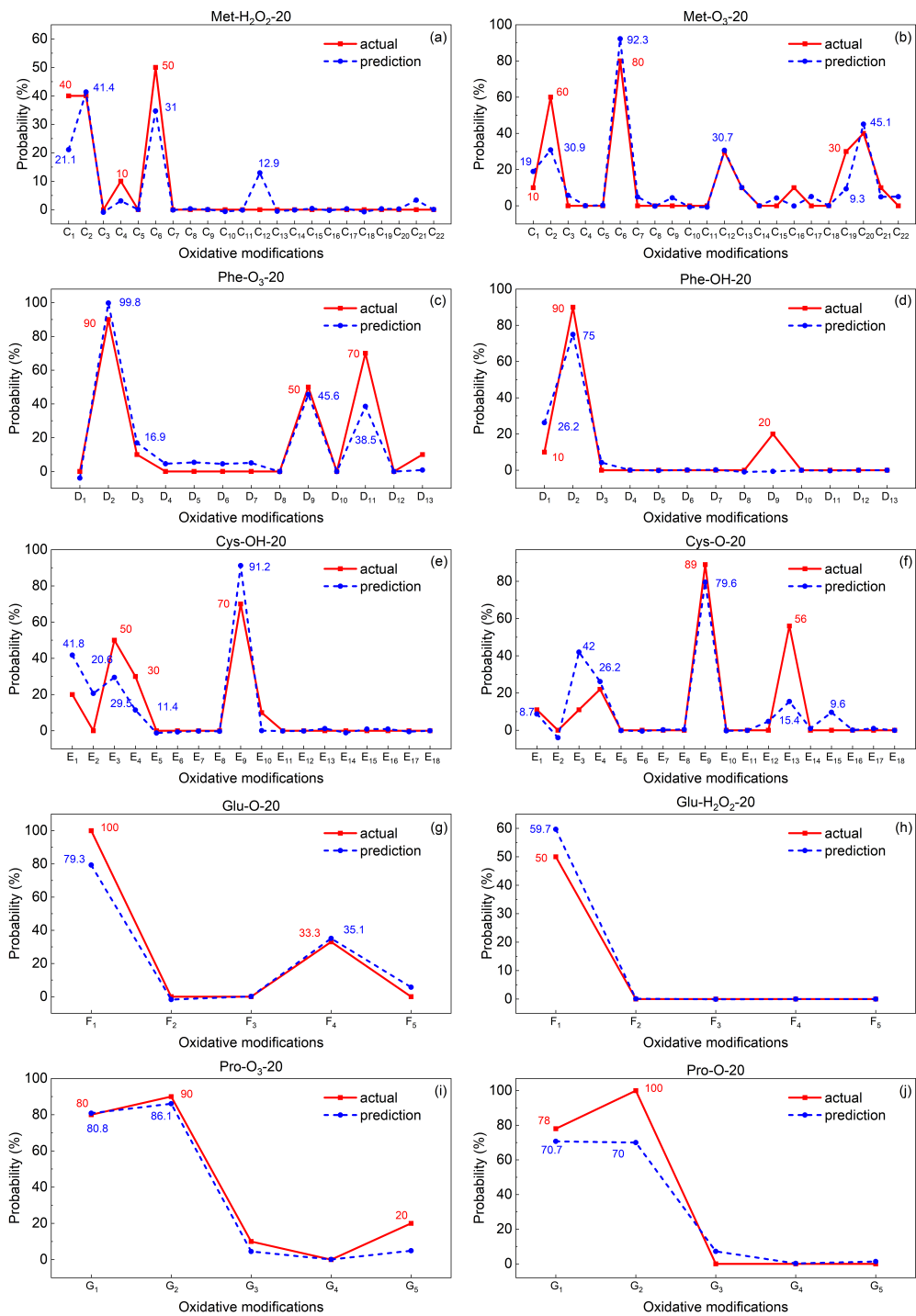


FIG. 8. Diagram of partial prediction results of 20 doses of ROS interacting with various amino acids.

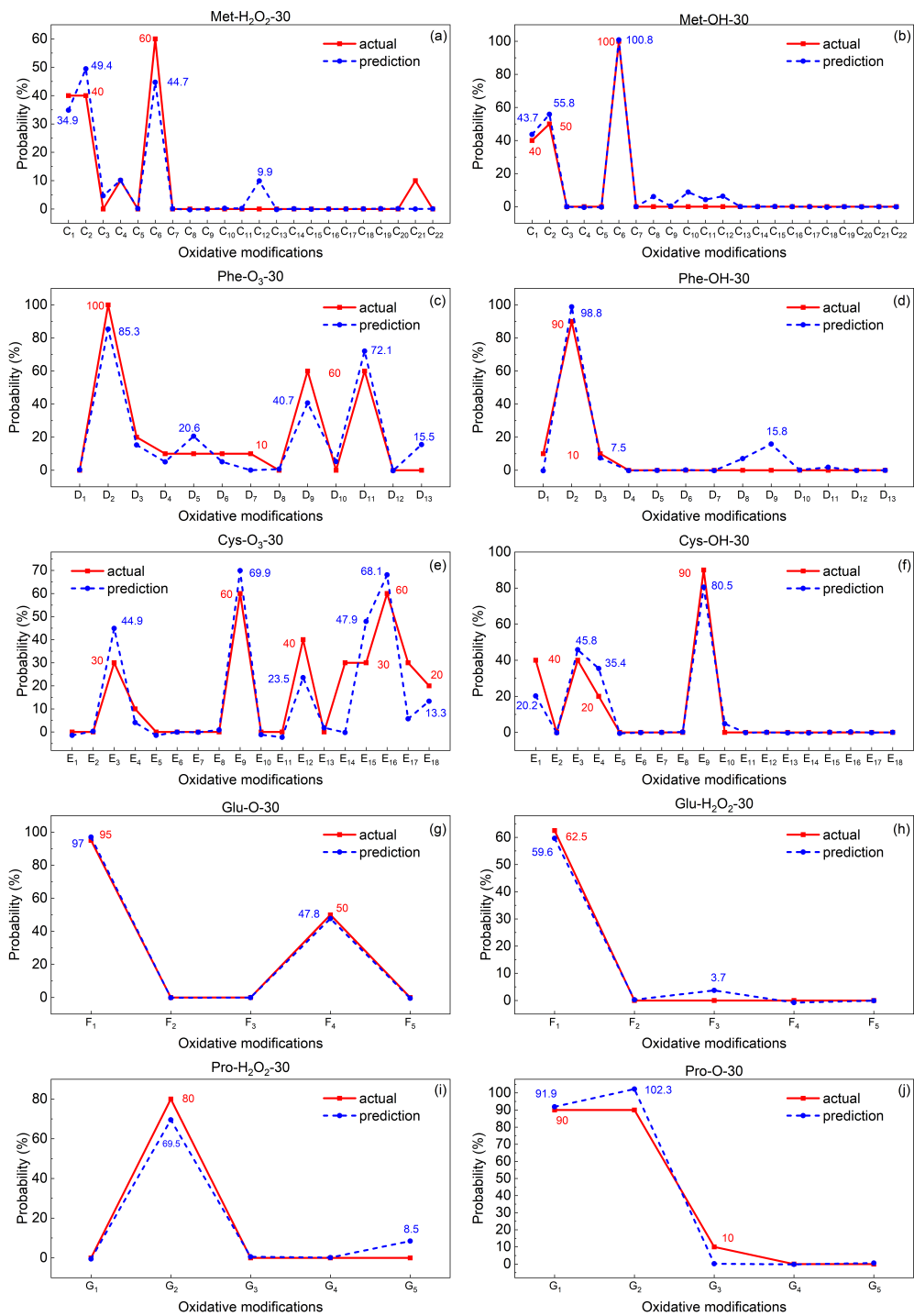


FIG. 9. Diagram of partial prediction results of 30 doses of ROS interacting with various amino acids.

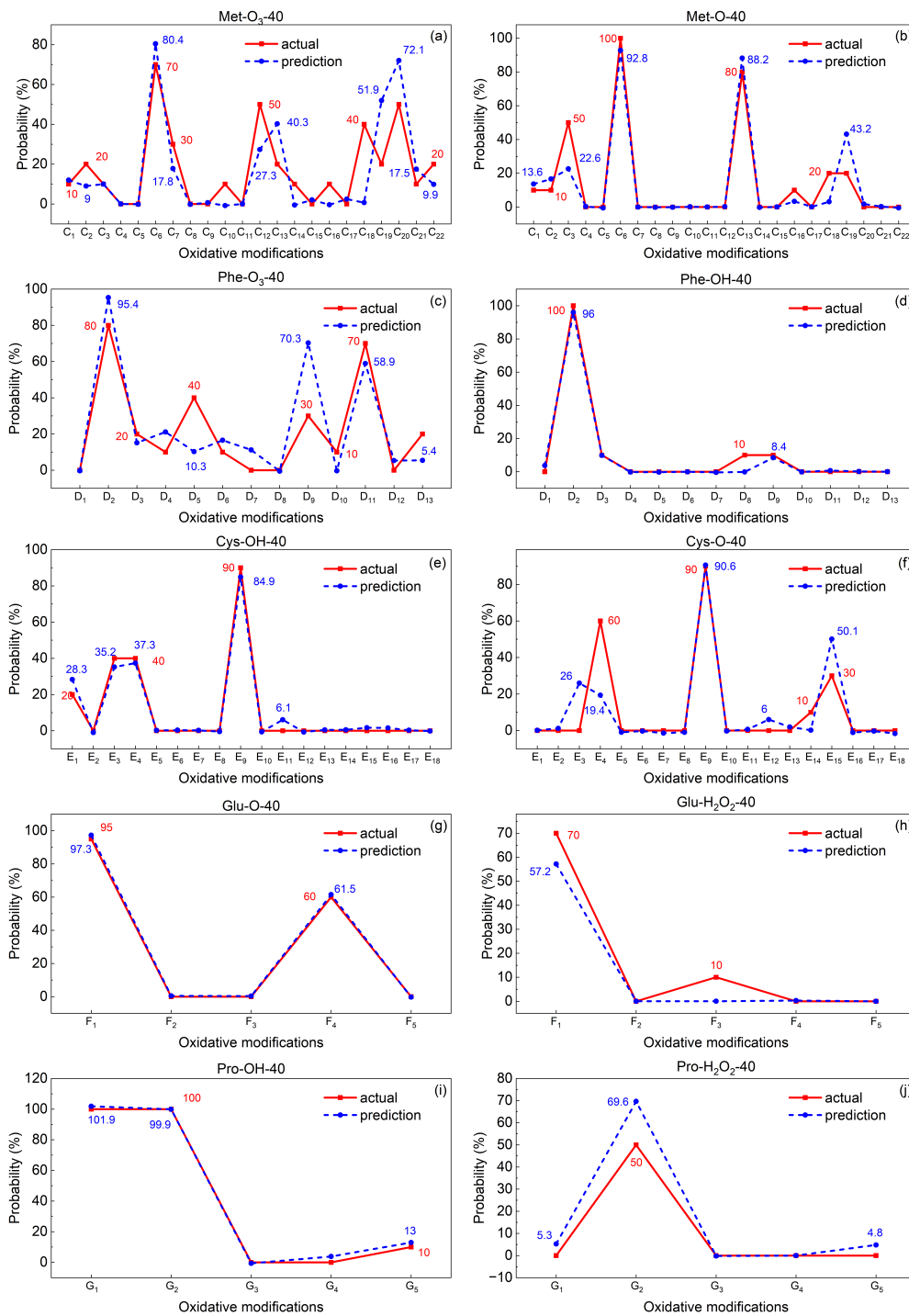


FIG. 10. Diagram of partial prediction results of 40 doses of ROS interacting with various amino acids.

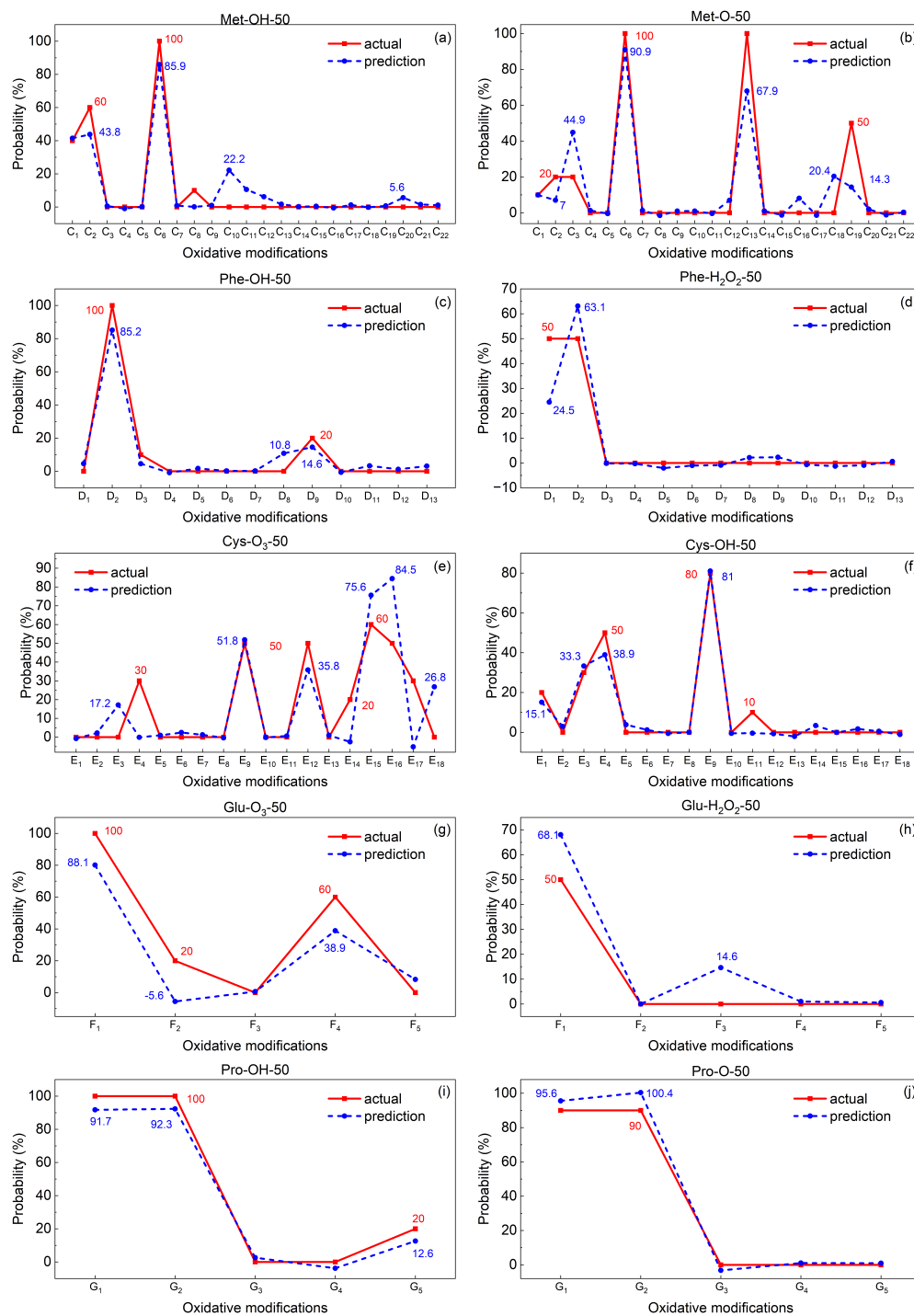
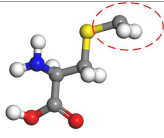
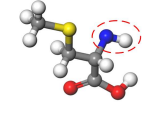
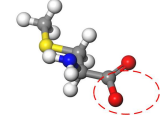
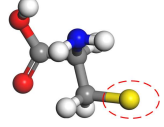
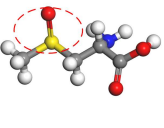
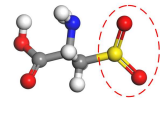
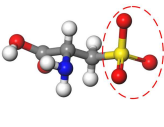
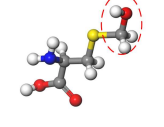
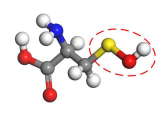
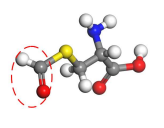
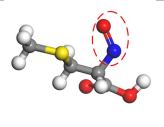
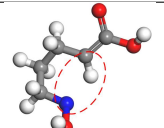
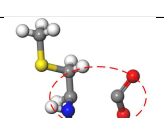


FIG. 11. Diagram of partial prediction results of 50 doses of ROS interacting with various amino acids.

TABLE I. Types of plasma-induced oxidative modifications in amino acids

| Oxidative modifications | Elemental composition | Figure examples | Involved amino acids |
|-------------------------|-----------------------|--|----------------------|
| Dehydrogenation | C-H |  | Met Phe Cys |
| | N-H |  | Met Phe Cys Glu Pro |
| | O-H |  | Met Phe Cys Glu Pro |
| | S-H |  | Cys |
| Sulfonation | S+O |  | Met |
| | S+2O |  | Met Cys |
| | S+3O |  | Met Cys |
| Hydroxylation | C+OH |  | Met Phe Cys |
| | S+OH |  | Met Cys |
| Carbonylation | C+O |  | Met Phe Cys |
| Nitrosylation | N+O |  | Met Phe Cys Glu |
| Ring cleavage | - |  | Pro |
| | 47 |  | |