CHEMISTRY

Rapidly determining the 3D structure of proteins by surface-enhanced Raman spectroscopy

Hao Ma[1,2], Sen Yan[1,2], Xinyu Lu[1,2], Yi-Fan Bao[1,2], Jia Liu[1], Langxing Liao[1], Kun Dai[3], Maofeng Cao[1,2], Xiaojiao Zhao[1,2], Hao Yan[1], Hai-Long Wang[2], Xiaohui Peng[1,2], Ningyu Chen[1,2], Huishu Feng[1,2], Lilin Zhu[1,2], Guangbao Yao[3], Chunhai Fan[3], De-Yin Wu[1], Binju Wang[1], Xiang Wang[1,2]*, Bin Ren[1,2]*

Despite great advances in protein structure analysis, label-free and ultrasensitive methods to obtain the natural and dynamic three-dimensional (3D) structures are still urgently needed. Surface-enhanced Raman spectroscopy (SERS) can be a good candidate, whereas the complexity originated from the interactions between the protein and the gradient surface electric field makes it extremely challenging to determine the protein structure. Here, we propose a deciphering strategy for accurate determination of 3D protein structure from experimental SERS spectra in seconds by simply summing SERS spectra of isolated amino acids in electric fields of different strength with their orientations in protein. The 3D protein structure can be reconstructed by comparing the experimental spectra obtained in a well-defined gap-mode SERS configuration with the simulated spectra. The gradient electric field endows SERS with a unique advantage to section biomolecules with atomic precision, which makes SERS a competent tool for monitoring biomolecular events under physiological conditions.

INTRODUCTION

The arrangement and combination of 20 amino acids create abundant proteins with complex structures and diverse functions. Accurate identification of proteins and deciphering of their structures are vital to drug development, clinical diagnosis, immunology, bioengineering, and so forth (*1*–*5*). Among various characterization techniques, traditional structural biology methods [nuclear magnetic resonance, cryo–electron microscopy (cryo-EM), and x-ray] can provide the accurate structural information for a protein of interest (POI). The developing trajectory of these techniques indicates that they require not only delicate experimental procedures but also deciphering strategies to enable a fast and accurate reconstruction of three-dimensional (3D) structure of a POI by comparing the calculated spectra or images of them with clear configurations with the experimental results (*6*, *7*). One of the examples is the maximum likelihood strategy used for single-particle analysis in cryo-EM, and cryo–transmission electron microscopy had not been widely used in structural biology until the deciphering strategies became matured (*8*). Although the above techniques have facilitated the identification of the structure-function relationship of POIs, the measurements still have to be performed either in high concentrations, at ultralow temperatures, or in crystalline form (*9*–*11*), which are vastly different from the natural environment of POI. It is therefore imperative to develop label-free and ultrasensitive methods that can obtain the natural and dynamic structural information of a POI in physiological environments, particularly at low concentration (*12*, *13*).

Surface-enhanced Raman spectroscopy (SERS) can provide fingerprint information of the molecule of interest as a vibrational spectroscopy and high sensitivity down to single molecule benefiting from the strongly enhanced electric field by the surface plasmon resonance effect of noble metal nanostructures (*14*, *15*). SERS is able to directly analyze biomolecules under physiological conditions (*16*, *17*). In particular, the local enhanced field exponentially decreases with the increasing distance from the nanostructure surface, forming a gradient field at the molecular scale (*18*). As a result, different functional groups or vibrational modes of surface species may experience substantially different field strengths, offering SERS a great potential to extract the atomic positions and determine the 3D structure of POI with a high precision, just like a molecular-scale computed tomography. Although SERS has been extensively applied in protein analysis (*19*–*22*), this advantage has not been fully taken because of the great challenges in obtaining reliable and reproducible SERS spectra. The gap-mode SERS offers a well-defined configuration for generating highly reproducible and reliable SERS spectra by placing small molecules in the gap of plasmonic nanoparticle (NP) and flat surface of noble metals (*23*–*26*). To make gap-mode SERS a powerful deciphering tool for protein structural determination (as illustrated in Fig. 1A), both the experimental and data interpretation issues need to be solved. The former includes the denaturation issue as a result of direct interactions between proteins and the metal surfaces and the diverse orientations of big proteins on the surface.

The latter requires an efficient strategy for interpreting SERS spectra of proteins. Although many elegant theoretical approaches have been proposed to simulate SERS spectra of simple molecules with the aid of quantum mechanics (QM) and machine learning (*27*–*34*), it is still challenging to accurately simulate Raman spectra for proteins with large molecular weight and flexible and dynamic structures under real physiological conditions (*35*). For

[1]State Key Laboratory of Physical Chemistry of Solid Surfaces, Collaborative Innovation Center of Chemistry for Energy Materials (i-ChEM), Department of Chemistry, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China. [2]Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM), Xiamen 361005, China. [3]School of Chemistry and Chemical Engineering, Frontiers Science Center for Transformative Molecules, National Center for Translational Medicine, Shanghai Jiao Tong University, Shanghai 200240, China.
*Corresponding author. Email: wangxiang@xmu.edu.cn (X.W.); bren@xmu.edu.cn (B.R.)

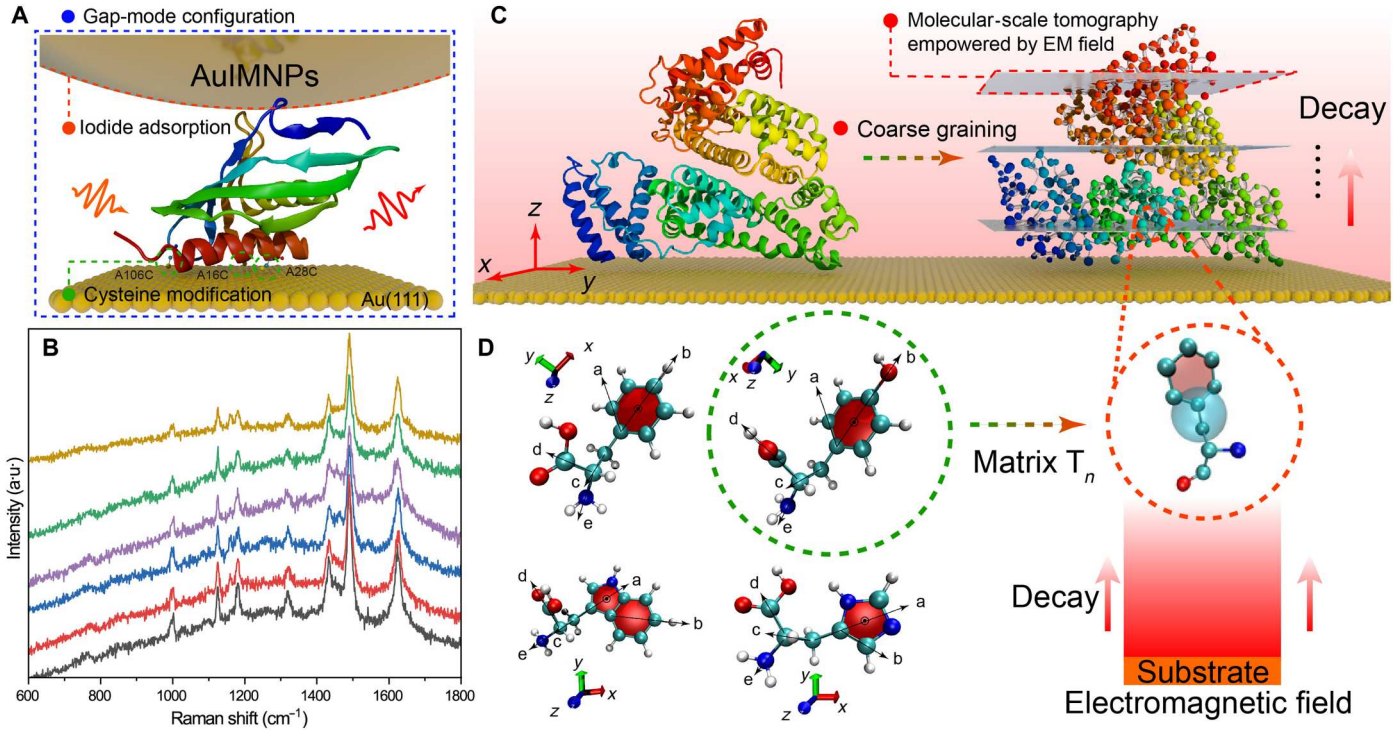

Fig. 1. Scheme of the proposed strategy for deciphering 3D protein structure. (**A**) Schematic diagram of protein in the nanocavity of the gap-mode configuration constructed by Au(111) surface and AuIMNPs, where every protein in the gap is subjected to the highly localized plasmonic field. Ketosteroid isomerase (KSI) was used as a model system (Protein Data Bank: 1buq), in which the positions of 16, 28, and 106 are replaced by cysteine. These cysteines can bind to the Au surface through thiol-Au interaction, allowing for orientation-controlled immobilization. (**B**) SERS spectra of KSI underneath six different NPs on a same single-crystal surface, indicating high reproducibility of the gap-mode configuration. a.u., arbitrary units. (**C**) On the basis of the coarse-grained model, the POI (bovine serum albumin as a model) on the surface can be sectioned into different layers, where each amino acid would experience different strengths in the decay electromagnetic field. The gray planes, indicating different field strengths, are similar to the sectioning plane in computed tomography but at the molecular scale, giving SERS the additional power to extract the amino acids position with atomic precision. (**D**) Every two vectors could define a bond rotation in amino acids. Thus, amino acids require at least three rotation matrices (T_n) describing the dynamic structure of amino acids, which would then be used to transform PDs under specific orientation to real orientation.

example, the most important parameter to determine the SERS intensity is the polarizability derivatives (PDs), which can be obtained by performing all-electron calculation. However, calculation of PDs of all vibrational modes of proteins requires enormous computation power scaling with N^3 (N is number of electrons) (*36*). The placement of PDs in the gradient local field further increases the computation costs, making it almost impossible to accomplish with the state-of-the-art computational resources.

The combination of the above two challenges constitutes a substantial obstacle for SERS to be applied in protein structure determination. Here, we proposed a deciphering strategy on the basis of well-defined SERS detecting configuration and a separable PD-based algorithm with rotational coarse-grained model (SPARC) to bridge the experimental results and SERS theories for protein structure determination. On one hand, we constructed a gap-mode configuration with well-defined local gradient field on the basis of iodide-modified NPs over single-crystal surface. The strongly adsorbed monolayer iodides not only prevent strong chemical interactions between the protein and metal surfaces but also remove the impurities on the surface, enabling highly reproducible and reliable SERS spectra to be obtained. On the other hand, we innovatively introduced a coarse-grained model to isolate and obtain the PDs of each amino acid, as depicted in Fig. 1C. Following this divide-and-conquer strategy, the expensive QM and electrodynamics calculations of the complex interactions between protein and the electric field can be simplified into simple matrix computation of PDs, rotation matrix, and gradient local electric field. Such simplification enables a fast calculation of SERS spectra for proteins with the aid of full-atom molecular dynamics (MD) simulation (fig. S1). As a result, the protein structure can be determined in seconds with a high accuracy by analyzing the similarity between the calculated spectra for various configurations and the experimental spectrum. The feasibility of SPARC as a universal tool for scrutinizing interfacial molecular structures is verified with molecules of different sizes, ranging from small molecules to peptides and proteins, on surface. Such an accurate and cost-effective deciphering strategy disruptively enables rapid analysis of the 3D structures of complex biomolecules via SERS. It can be a promising tool to explore the dynamics of interfacial proteins in physiological environments.

RESULTS

Deciphering strategy

The first crucial step toward protein structure determination via SERS is the development of a delicate experimental technique that

can provide highly reproducible and reliable data encoding the information of the 3D structure of proteins. As mentioned before, the decay nature of SERS enables the protein to be sectioned at the molecular scale by the local gradient field (similar to computed tomography, as illustrated in Fig. 1C), allowing the SERS technique to extract the atomic positions of amino acids with high precision and thus reconstruct the protein structure. For this purpose, we proposed to borrow the gap-mode SERS configuration with some modifications to achieve a well-defined distribution of the gradient electromagnetic (EM) field, as illustrated in Fig. 1A. Instead of pure single-crystal surface, we modified the atomically flat Au (111) surface with cysteine (Cys), of which the POI was placed on top. Then, we placed iodide-modified Au NPs (AuIMNPs) over POIs. In this way, POIs are naturally trapped in the gap formed between AuNPs and the Au(111) surface. The distribution of the gradient EM field, including the decay constant of the plasmonic field, can be regulated by the size of AuNPs, the wavelength and incident angle of the excitation laser, as well as the size of POI (*37*, *38*). The AuIMNPs and Cys-modified Au(111) surface prevent the direct interaction of POI with them and can effectively avoid the protein denaturation issue. Iodide modification helps remove the reductants from the surface, which improves the reliability of SERS detection. Furthermore, the Cys layer helps the protein to be immobilized in a bioconjugated way with a defined orientation on the Au(111) surface, which can effectively remove the spectral averaging effect and enable us to obtain highly reproducible and reliable SERS spectra. The well-defined gap-mode configuration provides a unique EM field distribution, which encodes valuable 3D structural information in the experimentally obtained SERS spectra. This, in turn, empowers the great deciphering capability of SERS for determining the 3D structure of protein, which has not been achieved before. To demonstrate reliability and reproducibility of this gap-mode configuration, we engineered ketosteroid isomerase (KSI) protein with three Cys residues on one side (Fig. 1A and fig. S2), so that it can be adsorbed on an atomically flat Au(111) single-crystal surface with the designed orientation. A nanocavity of well-defined size was then constructed by placing 55-nm AuIMNPs on the Au(111) surface (see fig. S3 for the helium ion microscopy image of the sample). The iodide monolayer would prevent the direct interaction between KSI and AuIMNPs surface and help retain the native structure of KSI (*20*, *39*). This configuration provides highly reproducible SERS spectra as shown in Fig. 1B, which enables the reconstruction of KSI structure based on its SERS spectra as will be discussed later.

The next step is to decode the experimentally obtained protein SERS spectra and reconstruct its interfacial structure. For this purpose, it is crucial to obtain the accurate simulated SERS spectrum for the protein with specific structure, which requires an accurate description of the interaction of PDs and gradient local electric field. Basically, the scattering tensor (S_n) of a molecule that determines the SERS intensity can be described as follows

$$\{S_n^2\} = \sum_n |E_s^L \alpha'_n E_i^L|^2 \quad (1)$$

where α'_n is the PDs and $E_{s,i}^L$ represents the enhanced local fields along the incidence and scattering directions. SPARC uses a gradient enhanced local field exponentially decreasing with the increase of distance (r, C_α of amino acid in the corresponding layer), i.e., $E_{s,i}^L \approx r^{-3}$, which implies that the electric field can vary over the dimension of a molecule or vibrational mode (Fig. 1C). In addition, according to the surface selection rules of SERS, only the vibrational modes with a component (α'_{zz}) parallel to the electromagnetic (EM) field can be effectively enhanced, giving the orientation of amino acids. Eventually, the combination of distance-dependent field strength and the surface selection rules results in a unique spectrum for every conformation of protein on the surface, which makes it possible to determine 3D structure of proteins with SERS. Then, SPARC uses PDs of amino acids with a specific orientation to obtain α'_{zz}. The orientation of amino acids in a protein can be defined by several geometrical vectors, and it requires two vectors to define rotation matrices (T_n) to describe the dynamic structure of amino acids. The rotation matrix is the most important operation for accurately calculating interactions between proteins and gradient local electric fields (Fig. 1D and fig. S4), which can be calculated by the atomic coordinates of POI from MD result. By aligning the orientation of amino acids to their real orientations in POI via rotation transformation, the PDs of each amino acid can be obtained by rotating corresponding empirical PDs (more validations can be found in figs. S5 and S6 and table S1). Therefore, the SERS spectrum of protein ($S_{protein}$) can be the sum of spectra of m amino acids (S_{AA}) accumulated by n vibrational modes (S_n^2) and expressed as follows

$$S_{protein} \propto \sum_m S_{AA} = \sum_m \sum_n C_n\{S_n^2\} = \sum_m \sum_n \frac{\omega_s^4 M_s^2 M_i^2 F_p}{c^4(1 - e^{\hbar\omega_s/k_B T})} k_{scale} |E_s^L \alpha'_n(\Omega) E_i^L|^2 \quad (2)$$

where ω_s is the frequency of scattered light and M_i and M_s represent the enhancement factors for incident light and scattered light, respectively. Ω represents Euler angles that determine the molecular orientation relative to the local fields. F_p is the Purcell factor. c, $\hbar$, k_B, and T are the speed of light, reduced Planck constant, Boltzmann constant, and kelvin temperature, respectively (*18*). By using Eq. 2, SERS spectra can be acquired with a high calculation speed via performing matrix computation with an averaged scaling factor k_{scale} (more details can be found in the "Correction of algorithm and correlated coefficient" section), which was introduced to compensate for the overestimation coming from high-order polarizability (as indicated in fig. S7). This issue can be addressed by considering the interactions between the quantified EM field with the general integrated polarizability (α_n, including contribution of all orders of polarizability) obtained with time-dependent perturbation Hamiltonian (*33*). As a result, SERS spectra of small molecules can be accurately calculated with the optimized structure in a localized field by overlapped integrals for corresponding orbitals and EM field. However, it is too expensive to directly use this method to track dynamic protein structure and calculate its SERS spectra. All elements needed in the calculation are orbital-dependent. It means that we can further follow our idea to rotate the orbitals under the specific orientation and then calculate the overlapped integrals (see details in Materials and Methods). Note that it takes high-order polarizability into account and eradicates the so-called gauge problem, due to integrals for Gaussian functions in Cartesian coordinates. As a result, the calculation scarifies little speed but markedly improves

the accuracy by considering high-order terms of polarizability and EM field.

To testify the performance of SPARC, we revisited a simple, elegant, structurally well-defined system by Li *et al.* (*18*), which is composed of a set of viologen-based monolayers on Au(111) single-crystal surface with silica-coated Au NPs on the top, so-called gap-mode configuration (Fig. 2A). They demonstrated that the plasmonic field in the gap can be treated as a combination of fields from the Au(111) surface and the Au NPs. Thus, the distribution of the field in the longitudinal direction can be expressed as

$$E_z \propto Ae^{-\alpha z^2} + Be^{\beta(z-\gamma)^2} \quad (3)$$

where A (B) and α (β) are the maximum intensity and decay constant of a plasmonic field by the Au surface (the Au NPs), respectively. z is the distance between the Au(111) surface, and γ is the gap size of the nanocavity. The experimental SERS spectra for viologen of different alkane lengths are given in Fig. 2B. Because both conformation of viologen molecules and plasmonic field distribution are given, we can conveniently obtain the integral corrected PDs of viologen molecules and calculate their SERS spectra by SPARC. As shown in Fig. 2C, the calculation results obtained by SPARC agree well with both the experimental results (Fig. 2B) and reported theoretical results (Fig. 2D), indicating the reliability of our SPARC method.

Demonstration of the deciphering strategy

To faithfully demonstrate the feasibility of our strategy for deciphering protein structure, we used the self-assembled peptides (YSATFTTY with acylated N and amidated C termini, named MoSBP1) on MoS_2 surface as a model system, which shows well-defined interfacial structure characterized by atomic force microscopy (AFM) (*40*). If the spectrum reproduced with known structure from AFM is well aligned with the spectrum by tip-enhanced Raman spectroscopy (TERS), then the feasibility of our strategy can be further confirmed. For this purpose, we transferred a monolayer of MoS_2 on the Au surface (MoS_2-Au; Fig. 3A), and the thin MoS_2 layer allows effective plasmonic coupling between the Au surface and Au TERS tip to obtain a strong TERS signal of the peptide. When incubated with the MoS_2-Au substrates, MoSBP1 assembled into an elongated layer with three equivalent directions (with 60° angle) as shown in Fig. 3B, which is consistent with the phenomenon observed by Chen *et al.* (*40*). The layer is 0.7 nm in height and consists of parallel rows with a periodicity of ca. 4 nm on MoS_2-Au substrates (e.g., framed region in Fig. 3B). According to

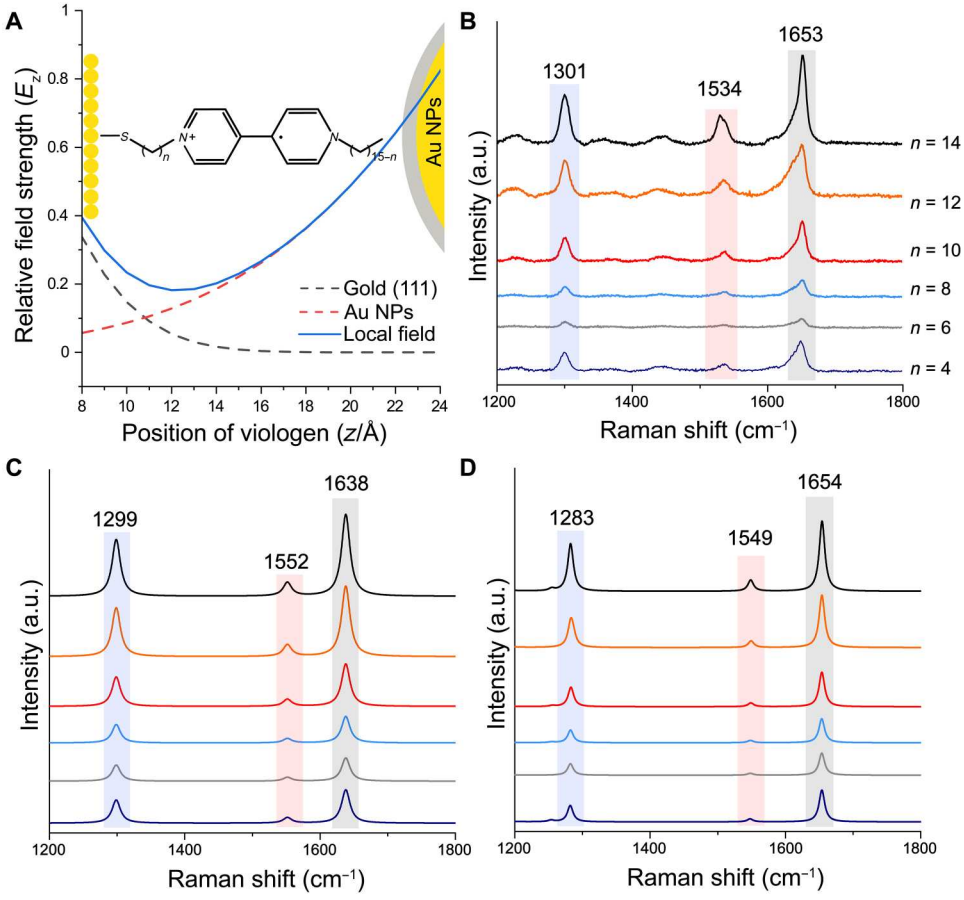

Fig. 2. Testifying the performance of SPARC. (**A**) Schematic diagram of the proposed system by Li *et al.* as well as local electric field distribution for viologen in the gap. Surface-enhanced Raman spectra of viologen with different alkane lengths (n) to the Au(111) surface in the gap of silica coating AuNPs and Au(111) surfaces. (**B**) Experimental SERS spectra. (**C**) Simulated SERS spectra by SPARC. (**D**) Simulated SERS spectra in the literature. (B and D) Reproduced with original data by courtesy of Li *et al.* (*18*).

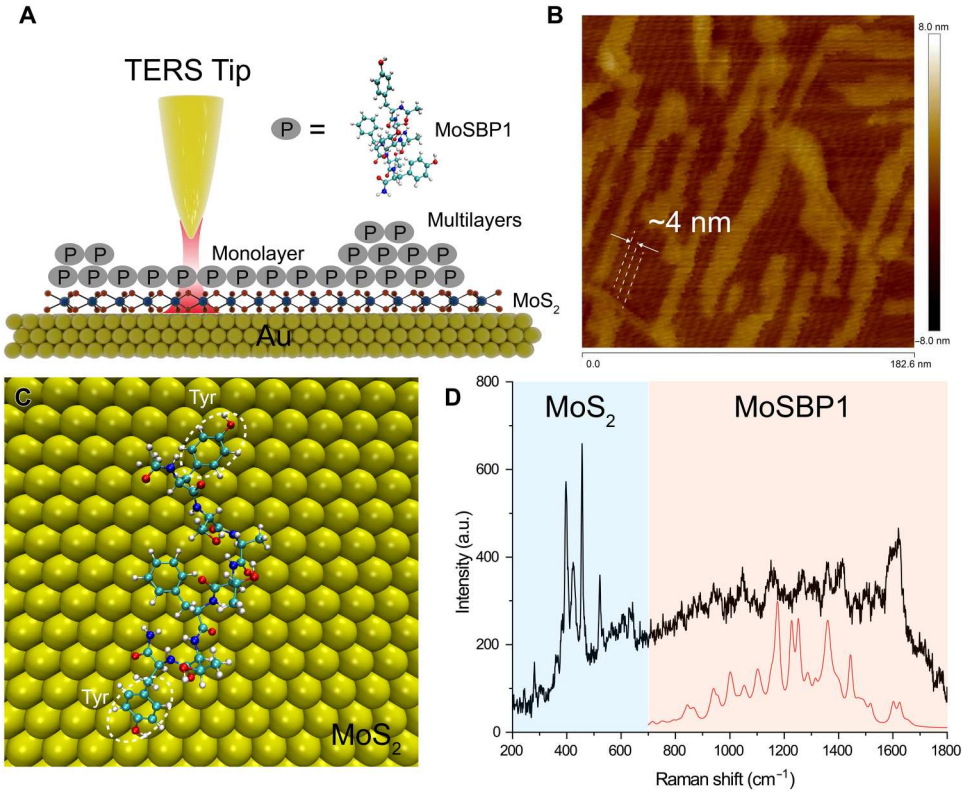

Fig. 3. Demonstration of the deciphering strategy by TERS. (**A**) Schematic diagram of TERS detection of the self-assembled peptide on MoS_2. (**B**) AFM image of the sample, showing a full monolayer of order structure of peptides in the bottom layer and some multilayer features on top. (**C**) Interfacial repeating unit of peptide on MoS_2 in the literature (*40*). (**D**) Simulated TERS spectrum (red) by SPARC and experimental TERS spectrum of the monolayer of peptide (black). The peaks in the light blue region are from MoS_2, and the light red region is from peptides.

Heinz and colleagues, the two tyrosines (Tyr) at the two ends of the peptide are flat-lying on the MoS_2-Au, where the hydrogen bonding between two peptides in an end-to-end way via the Tyr forms the building block of the order assembled structure. These two peptides determine the packing direction(*41*). In addition, there are several multilayer domains (with a height of 1.4 nm) of MoSBP1 over the monolayer structure, which has not been reported before. Thus, we used SPARC with the known conformation (as indicated in Fig. 3C) to calculate the TERS spectrum of the monolayer of MoSBP1 on MoS_2-Au substrates, as the red spectrum shown in Fig. 3D. Note that we used the same approach as outlined in Eq. 3 to calculate the electromagnetic (EM) field along the z direction. The experimental TERS spectrum (black in Fig. 3D) shows several intense peaks in the range of 300 to 700 cm^{-1} contributed by MoS_2 (*42*), suggesting a strong enhancement in the gap between the Au tip and substrate. The peptide contributes the peaks in the spectral region from 700 to 1800 cm^{-1}. Because of the flat-lying adsorption of Tyr and Phe, their in-plane vibrational modes (e.g., 830 cm^{-1} for Tyr and 1000 cm^{-1} for Phe) show low intensity. We evaluate the similarity of our simulated spectra and experimental spectra by Spearman's rank correlation coefficients (r_s; see Materials and Methods for detail). Despite the low intensity and signal-to-noise ratio of the experimental spectrum, our simulated spectrum of monolayer still reveals high similarity (r_s = 0.66, normally r_s > 0.6 indicates a good theoretical prediction) (*43*) to the experimental one, which demonstrates the feasibility of SPARC for deciphering protein structure with known sequence.

Deciphering SERS spectra of peptides

After verifying the reliability of the strategy, it will be interesting to take a step forward to reconstruct the structure of more complex biomolecules from SERS spectra. We assembled a heptapeptide (CAFKAAH; Fig. 4A) with Cys termination on the Au surface via Au—S bond, followed by assembly with 140 nm AuIMNPs. In this way, a nanocavity with a well-defined size was formed to produce SERS signal of peptide in a gap-mode way (Fig. 4A). SPARC was then used to calculate the SERS spectra of POI to decipher protein structure from the experimental SERS spectra. To obtain possible orientations of POIs, we used the GolP-CHARMM force field, which has been accepted as a powerful force field to describe the reasonable protein structure at the interface to simulate the adsorption of peptides on Au (*44*–*46*). The heptapeptide is initially placed at least 1 nm away from the surface. After 100 ns, residues like Phe and Cys are successively attached to the Au surface, as shown in Fig. 4B. We performed 35 independent MD simulations (ca. 4 μs) to fully sample possible orientations. According to the free energy of landscapes shown in fig. S8, we extracted the dominant conformations of the peptide. The root mean square deviation (RMSD) between conformations is illustrated in fig. S9A, which also indicates marked structural variations. Experimentally, we combined a dark-field scattering microscope to locate single NPs

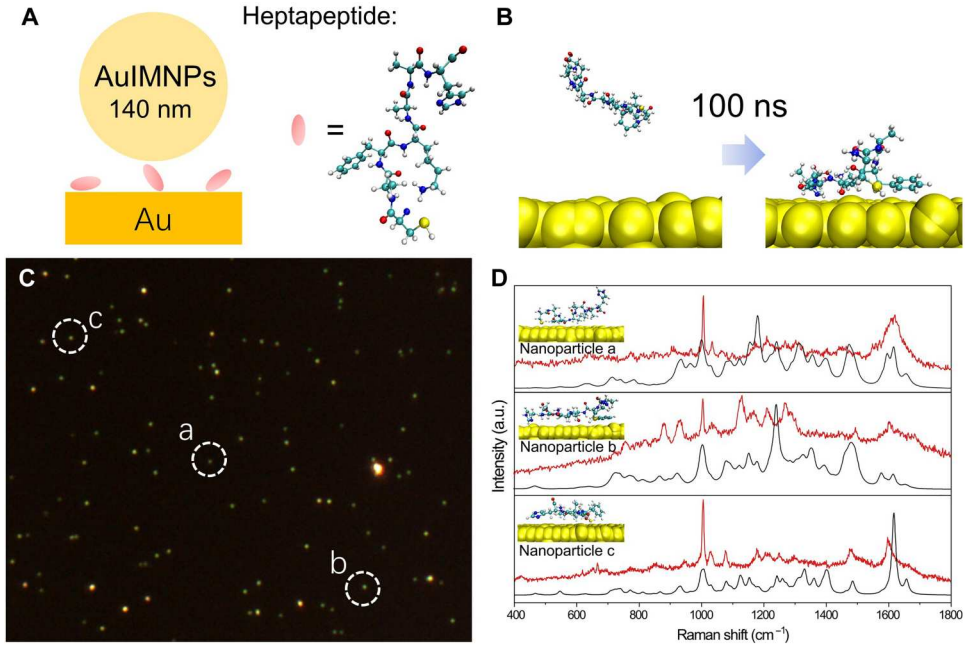

Fig. 4. MD simulation and experimental and theoretical SERS spectra of peptides by SPARC. (**A**) Schematic diagram of single NP gap-mode SERS. (**B**) An example of the initial configuration and final configuration of the heptapeptide on a Au surface. (**C**) Dark-field scattering image of the sample. (**D**) Theoretical SERS spectra by SPARC (black spectrum) in comparison with experimental SERS spectra (red spectrum) obtained from the NPs as indicated in (C) with white dashed circles. Insets are corresponding configuration scrutinized by our strategy.

(Fig. 4C) and acquired the SERS spectra from these NPs. We then compared the similarity of our simulated spectra (based on 40 dominant conformations; fig. S9B) to the experimental spectra of three single NPs in Fig. 4C by Spearman's rank correlation coefficients (r_s) and identified corresponding conformations accordingly. As the peptides right underneath the NPs have less freedom to move, we envisage that one specific orientation may dominate in the final spectra as shown in Fig. 4D. The resulting correlation coefficients are 0.85, 0.81, and 0.77, respectively, indicating good theoretical predictions. In return, the peptide structure can be scrutinized from experimental SERS spectra, as illustrated in Fig. 4D.

Reconstructing protein structure based on its SERS spectra

The above shows the results on peptides, and it will be more exciting if proteins with much more complex structure can also be directly deciphered only with SERS spectra. In Fig. 1A, we have shown that we were able to obtain the reproducible spectra of KSI, and we can now use the SPARC to extract its 3D structure. To obtain structural information from the experimental spectrum of this well-defined system, we need to get the atomic coordinates of amino acids and EM field distribution. The atomic coordinates were obtained by over 50 MD simulations (ca. 5 μs) to fully sample all possible stable conformations. The free energy landscapes are shown in fig. S10, from which we extracted the representative orientations of KSI and showed four representative ones in Fig. 5A. We used the same approach as in Eq. 3 to obtain the EM field along the z direction (Fig. 5C). The simulated spectra are given in Fig. 5B, and different orientations result in different spectral features. The correlation coefficient and theoretical spectra of all conformations are illustrated in fig. S11 (A and B). It is interesting to find that conformation 2 in Fig. 5A shows the best correlation with the experimental spectrum of KSI (with r_s = 0.75). This conformation agrees well with the intended orientation that would arise by Cys modifications.

From the above result, we can find that our strategy has great potential to obtain the 3D structure of proteins at the surface from its SERS spectrum. Furthermore, the gradient electric field endows SERS with a unique advantage to section protein with atomic precision to reconstruct the 3D protein structure. Hence, with known atomic coordinates and electric field distribution, the SERS spectrum of the POI will be uniquely determined and vice versa. We can also foresee that more structural information can be obtained by tuning the gradient electric field, which promotes spectra deconvolution with enlarged experimental data points. This was demonstrated through SERS measurement of the same system under various laser wavelengths, as depicted in fig. S12. The SERS spectra of KSI displayed various spectral features at different wavelengths, possibly due to the presence of different local electric field distributions within the gap (fig. S13). Because the tensor nature of polarizability and decay of EM involved in SPARC are the most fundamental and universal features of SERS, SPARC should be a general spectrum interpreter that is capable of interpreting SERS spectra of not only protein but also other biomolecules. We further extended SPARC to interpret SERS spectra of DNA (fig. S14). The DNA spectra obtained by SPARC reveal a high correlation with experimental results in the literature (see fig. S15), indicating the potential to analyze the interfacial structure of nuclei acid as well as other biomolecules.

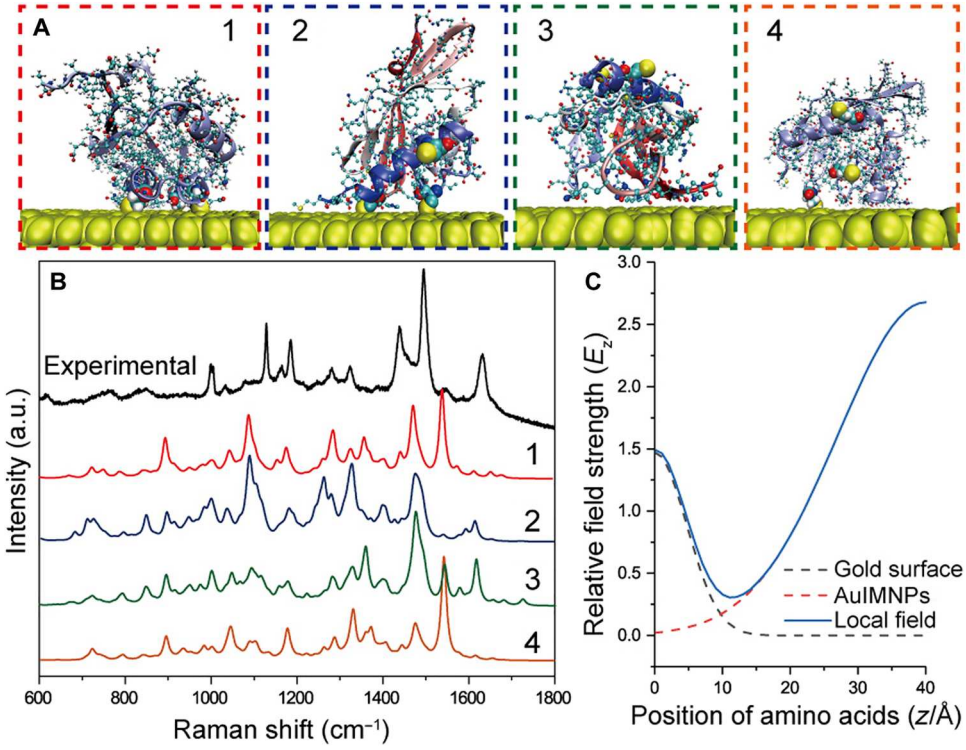

Fig. 5. Scrutinization of the KSI structure from the gap-mode SERS spectrum. (**A**) Four representative orientations of KSI obtained by MD simulation. (**B**) Experimental gap-mode SERS spectrum of KSI in comparison with theoretical SERS spectra of four representative conformations calculated by SPARC. (**C**) Local EM field from contributions of Au surface and AuIMNPs.

DISCUSSION

In summary, we have reported an accurate and cost-effective deciphering strategy for rapidly reconstructing the 3D protein structure of POI based on SERS. By constructing well-defined detecting configuration, we obtained the reliable and reproducible SERS spectra of protein in the gradient EM field, which contains information of the amino acids position with atomic precision. SPARC can simplify the conformation of proteins at the interface as a spatial arrangement and combination of amino acids. As a result, SERS spectra can be obtained by simple matrix computation of PDs, rotation matrix, and gradient local electric field. Moreover, calculation accuracy can be further improved by introducing interaction Hamiltonian with integrated polarizability. The predicted SERS spectra by SPARC show a high similarity to the experimental spectra of protein in well-defined systems, which, in turn, enable us to reconstruct the 3D structure of proteins with known sequences from the obtained SERS spectra. Because the description of PDs and EM fields are the most fundamental features of SERS, SPARC can be a universal SERS interpreter for any other complex biomolecules, which has also been verified in the DNA system. Our method will open a door for monitoring biomolecular events under physiological conditions, which are important to applications, such as biocatalysis, drug development, and clinical diagnosis. Both experimental and theoretical efforts need to make to address the great challenge. In the experimental aspect, one may tune the EM field to profile protein molecules so that abundant and effective experimental data can be obtained. In the theoretical aspect, more efficient sampling methods should be developed to sample all possible orientations without relying on the less efficient traditional MD simulation. Combined with machine learning, SPARC will be able to consider the coupling effects between amino acids and the influence of the chemical environment in a fast and cost-effective way that is impossible with state-of-the-art QM/ED methods. At that stage, we can foresee that SPARC can be efficient enough to scrutinize the structure of unknown protein in a short time.

MATERIALS AND METHODS

Calculation of amino acids and their PDs

The geometries of amino acids were optimized by density functional theory (DFT) calculations using the basis set B3LYP and adapted the triple splitting valence basis set of 6-311+G*. The structures of amino acids used can be found in fig. S4. All calculations were performed by the Gaussian 09 package (*47*). The scale factor for vibrational frequency is 0.98 obtained using the peak of phenylalanine at 1000 cm^{-1}. The calculations of nucleic acids are performed at the same level without the scale factor. To obtain PDs of amino acids, we need to transform the atomic PD tensor in the Cartesian coordinate (PD_C) to the PD tensor in the normal coordinate frame (PD_N) according to the following expression (*48*)

$$PD_N = PD_C A L_S$$

The Raman scattering factor S_m is calculated from the isotropic polarizability tensor derivative (α'_m) and the anisotropic PD (β'_m) of

the mth vibrational mode

$$S_m = g_m[45\alpha'^2_m + 7\beta'^2_m]$$

where

$$\alpha'^2_m = \frac{1}{3}(\tilde{\alpha}'_{xx,m} + \tilde{\alpha}'_{yy,m} + \tilde{\alpha}'_{zz,m})$$

and

$$\beta'^2_m = \frac{1}{2}[(\tilde{\alpha}'_{xx,m} - \tilde{\alpha}'_{yy,m})^2 + (\tilde{\alpha}'_{yy,m} - \tilde{\alpha}'_{zz,m})^2 + (\tilde{\alpha}'_{zz,m} - \tilde{\alpha}'_{xx,m})^2 + 3(\alpha'^2_{xy,m} + \alpha'^2_{xz,m} + \alpha'^2_{yz,m})]$$

After transformation, we can obtain the PDs of each vibrational mode of every amino acid and nucleic acid, which are saved as empirical files to be invoked. To testify the effect of protonation states and side chain on PDs, we took the peak of phenylalanine at 1000 cm^{-1} as an example, owing to its high sensitivity to the chemical environment change. The five structures are illustrated in fig. S5. We aligned the benzene ring of phenylalanine with the same orientation and compared the PDs values of five structures. As depicted in table S1, the PD value change is modest, and the SD is low. This result indicates that we can obtain reasonable values of PDs based on isolated amino acids. A dipeptide (Ace-phe-ala-nme) with 50 conformations from SPICE dataset was selected to confirm the ability of SPARC for differentiating between different structures (*49*). The structures and corresponding RMSD are illustrated in fig. S6 (A and B). For DFT-simulated results, only α_{zz} component of PDs was considered to produce the SERS spectra of dipeptide. For SPARC, we ignored the functional group of the dipeptide (i.e., Ace and nme) and calculated the SERS spectra on the basis of two amino acids. As shown in fig. S6C, we showed theoretical spectra of eight conformations of dipeptide simulated by DFT (black curve) and SPARC (red dotted curve). It can be found that the spectra of dipeptide with different orientations and structures can be differentiated. In comparison with DFT-simulated spectra, most of spectral features of the dipeptide can be observed in the spectra simulated by SPARC, which demonstrates the feasibility of SPARC to differentiate between different structures.

MD simulation

The 3D structure of proteins was obtained by homologous modeling using the I-TASSER server, in which full-length atomic models were constructed by iterative template-based fragment assembly simulations (*50*, *51*). Last, five representative protein structures were obtained from the I-TASSER server, for later MD simulation. GROMACS (*52*) (2019.5) was used for the MD simulation with CHARMM force field (*53*) and GolP force field (*53*). A periodic boundary was used in the simulation, where the protein was immersed in a cuboid box of water. The simplified point charge model TIP3P (*54*) was used to represent water molecules. In the simulation, we first used the steepest descent approach to minimize the energy of the system with a convergence criterion of 1000 kJ mol^{-1} nm^{-1} and a step of 0.01 nm. Energy minimization was followed by 100 ns of NVT MD simulation. To visualize the free energy landscape, the anaeig module of GROMACS was used to analyze the trajectories of backbone and project two principle components as horizontal and vertical axes in figs. S8 and S10.

Finite-element method simulation

A commercial finite-element method simulation software (COMSOL Multiphysics package 6.0) was used to model the near-field enhancement of Au NPs on Au film. The perfectly matched layers were used to simulate an open boundary. The permittivity values of Au were from Johnson and Christy (*55*). The medium over this structure was set to be air. The diameter and size of Au NPs were set on the basis of the helium ion microscopy image (fig. S3).

Algorithm based on Python

Figure S1 schematically shows the processes of SPARC. GROMACS (gmx) was used for MD calculation to obtain the reasonable and dominant conformation of POI on the surface. Multiwfn, a powerful wave function analysis software, was used for calculating SERS spectra (*56*). Thus, the algorithm has three functions: (i) read coordinates from results provided by gmx, which is based on Bio.PDB.PDBParser; (ii) invoke the parameters to calculate the rotation matrix and calculate the PDs of POI in the real conformation; and (iii) calculate the empirical electromagnetic (EM) field according to coordinates. The SERS activity of each mode can be calculated, and Multiwfn reads the output files. Note that the EM field equation used for gap-mode simulation is adopted from Li *et al.* (*18*). The local field distribution of the nanocavity constructed by Au(111) and AuIMNPs in Fig. 5C follows the expression of $E_z \propto 1.469e^{-0.023z^2} + 2.679e^{-0.003(z-\gamma)^2}$, with the gap size of the nanocavity (γ) of the size of protein (40 Å). We proposed two methods for accurate calculation of SERS spectra in SPARC. The first method is to use Eq. 2, which is simple matrix computations of empirical PDs, rotation matrix, and EM field by Numpy library. The scaling factors (k_{scale}) are introduced to SERS intensity of each mode to compensate the overestimation to save the calculation resources, which is suitable for calculation of large molecules, like proteins. Another more accurate but time-consuming method is to use Eq. 4 (see the next section for details), in which the overlapped integrals are used for calculating SERS intensity of each mode, which can be used for calculation of small molecules, like viologen. The use of Eq. 2 can also obtain similar results as in Eq. 4 for small molecules as depicted in fig. S7. Then, the SERS activity of all vibrational modes obtained by the above methods can then be merged into one file with three columns (frequency shift, intensity, and width), which can be read by Multiwfn to produce SERS spectrum with Lorentz spectral line.

Correction of algorithm and correlated coefficient

The theoretical SERS spectrum was calculated by SPARC (fig. S7B) and was compared with the theoretical result from Li *et al.* (*18*) (fig. S7A). It is obvious that SPARC overestimates the intensity of molecules at a high EM field (when the distance of the molecule to the substrate is less than 1.5 nm), which may be due to the gauge problem (*33*). To correct this problem, we positioned every amino acid at a different distance (r) to the surface (fig. S7C) and calculated their SERS spectra using SPARC and gauge-invariant method developed by Duan *et al.* (*33*). Then, we fitted the intensity of spectra with exponential functions ($k = ax^b$). It should be noted that every amino acid under different orientations has its k function due to a unique electronic structure used in the interaction Hamiltonian. As shown in fig. S7D, scaling amino acids separately provides more accurate polarizability. However, it also implies that it

is necessary to scale all amino acids with different orientations to fit their k functions to calculate the dynamic SERS protein spectra, which requires enormous computational resources. Instead, we propose to introduce an averaged factor k_{scale} with all the obtained k for all amino acids to avoid the accumulated scaling process. After scaling this algorithm with a general scaling factor (k_{scale} = 0.1 for r < 1 nm), we obtain a theoretical result highly correlated with their result in the literature (Fig. 2D) (*18*). The SERS spectrum of protein ($S_{protein}$) can be expressed as follows

$$S_{protein} = \sum_m \sum_n k_{scale} C_n \{S_n^2\} = \sum_m \sum_n k_{scale} C_n |E_s^L \alpha'_n(\Omega) E_i^L|^2 \propto \sum_m \sum_n C_n |\alpha_n|^2 \quad (4)$$

$$|\alpha_n|^2 = \sum_r \left[\frac{\langle \phi_{g(\Omega)} | \hat{V}_\rho^\dagger | \phi_{r(\Omega)} \rangle \langle \phi_{r(\Omega)} | \hat{V}_\sigma | \phi_{g(\Omega)} \rangle}{\Delta E_{rg} - \omega} + \frac{\langle \phi_{g(\Omega)} | \hat{V}_\sigma | \phi_{r(\Omega)} \rangle \langle \phi_{r(\Omega)} | \hat{V}_\rho^\dagger | \phi_{g(\Omega)} \rangle}{\Delta E_{rg} + \omega} \right]$$

where $\phi_{g(\Omega)}$ and $\phi_{r(\Omega)}$ are electronic ground and resonant excited states with respect to the orientation after rotation (Ω), respectively. ΔE_{rg} represents the vertical excitation energy between $\phi_{g(\Omega)}$ and $\phi_{r(\Omega)}$. $\hat{V}_\rho^\dagger$ and $\hat{V}_\sigma$ represent the time-dependent operator. Following our idea to integrate Duan′s method into SPARC, all we need is to rotate the orbitals of the molecule. To this end, we extracted molecular orbital (MO) coefficients (C) of amino acids under standard orientation and rotated the orbitals by the Wigner D matrix

$$D^l_{m',m}(\alpha,\beta,\gamma) = e^{-im'\alpha} d^l_{m',m}(\beta) e^{-im\gamma}$$

where α, β, and γ are Euler′s angle of rotation, which can be calculated from the rotation matrix (T). m, m', and l are determined by shell type. The reduced Wigner matrices can be written as

$$d^l_{m',m}(\beta) = \sqrt{(l+m')!(l-m')!(l+m)!(l-m)!} \times \sum_s \frac{(-1)^{m'-m+s} \cos^{\frac{\beta}{2}}{}^{2l+m-m'-2s} \sin^{\frac{\beta}{2}}{}^{m'-m+2s}}{(l+m-s)!s!(m'-m+s)!(l-m'-s)!}$$

In this way, we can construct a diagonal matrix (M), whose matrix elements are in the order of shells. Then, we can obtain $\phi_{g(\Omega)}$ and $\phi_{r(\Omega)}$ easily with simple matrix multiplication (i.e., C' = M.T @ C @ M).

Spearman′s rank correlation coefficient was used to evaluate the performance of SPARC (*43*). The equation of Spearman′s rank correlation coefficient is shown as follows

$$r_s = \frac{\mathrm{cov}[R(x), R(y)]}{\sigma_{R(x)} \sigma_{R(y)}}$$

where σ and cov denote the SD and covariance, respectively. $R(x)$ and $R(y)$ correspond to the ranked Raman intensities of experiment spectra and predicted spectra, respectively. Normally, a high Spearman coefficient, >0.6, is required as a threshold for a good theoretical prediction, in addition to reasonable orientation and relative peak intensity.

Purification of protein

Primer STAR Max DNA Polymerase was purchased from Takara (Shiga, Japan). The plasmid mini kit I was provided from Omega Bio-Tek (USA). *Escherichia coli* BL21 (DE3) competent cell, DMT chemically competent cell, DMT enzyme, isopropyl-β-D-thiogalactopyranoside (IPTG), and kanamycin were purchased from TransGen (Beijing, China). SDS–polyacrylamide gel electrophoresis (PAGE) was purchased from GenScript Biotech Corp. (Nanjing, China). Modified Bradford Protein Assay Kit and LB were purchased from Sangon Biotech (Shanghai, China).

Using *E. coli* as the expression host, the amino acid sequence of KSI from *Pseudomonas putida* (UniProtKB: P00947) was optimized and then synthesized by GenScript Biotech Corp. (Nanjing, China). The gene of KSI was ligated with pET-28a (+) under the restriction sites Nde I and Xho I to construct the recombinant plasmid pET-28a (+)–KSI, which was accompanied by His_6-tagged in the N terminus of KSI. Mutagenic forward primers and reverse primers (table S1) were used to construct a site-directed mutation plasmid that was called pET-28a (+)–KSI–A16C/A28C/A106C. The plasmids were lastly transformed into *E. coli* BL21(DE3).

The enzymes were expressed from recombinant plasmids containing the KSI or mutation gene under the control of the T7 promoter. Typically, cultures of recombinant *E. coli* BL21 (DE3) in LB medium containing kanamycin (40 mg/liter) were grown at 37°C (200 rpm) until optical density at 600 nm reached 0.6 to 0.8 and then induced with 0.2 mM IPTG at 25°C for 16 hours. After induction, cells were harvested by centrifugation at 7000 rpm for 5 min at 4°C, the supernatant was discarded, and the cell pellet was resuspended with phosphate-buffered saline (PBS) buffer. Then, the cells were disrupted by ultrasonication and removed by centrifugation at 9000 rpm for 20 min under 4°C. Obtained crude cell extract was then added to His-Trap column (His-Trap HP 5 ml, GE Healthcare Corp., Piscataway, NJ, USA) that has been pre-equilibrated with binding buffer [20 mM sodium phosphate, 0.5 M NaCl, and 20 mM imidazole (pH 7.4)]. The enzyme-modified column was further equilibrated with binding buffer, followed by elution with buffer containing 20 mM sodium phosphate, 0.5 M NaCl, and 0.5 M imidazole (pH 7.4). The enzymes were filled in dialysis membranes (MD44, 3500D, Solarbio) and dialyzed with PBS buffer overnight. The purity of the obtained enzymes was tested by 10% (w/v) SDS-PAGE, as shown in fig. S2. All the protein concentrations were determined using the modified Bradford Protein Assay Kit with bovine serum albumin as a standard. Amino acid sequence of KSI (Protein Data Bank: 1BUQ): MNTPEHMTAVVQRYVAALNAGDLDGIVALFADDATVEDPVGSEPRSGTAAIREFYANSLKLPLAVELTQEVRAVANEAAFAFTVSFEYQGRKTVVAPIDHFRFNGAGKVVSMRALFGEKNIHAGA. Gene sequence of KSI: ATGAATACTCCCGAACACATGACAGCTGTAGTGCAGCGTTACGTGGCGGCTTTGAATGCAGGTGATCTGGATGGTATTGTAGCGCTCTTCGCCGACGACGCGACCGTGGAGGACCCGGTTGGTAGCGAACCGCGTAGCGGCACCGCAGCGATTCGCGAGTTCTACGCGAACAGCCTGAAGCTGCCACTGGCGGTGGAGCTGACGCAAGAAGTTCGCGCGGTTGCCAATGAAGCTGCGTTCGCTTTCACTGTCTCCTTTGAGTATCAGGGTCGTAAAACCGTCGTGGCTCCGATCGACCACTTTCGTTTTAACGGTGCCGGCAAAGTTGTTTCTATGAGAGCGTTGTTTGGCGAAAAGAACATCCATGCAGGCGCATGA.

Raman instrument

SERS spectra were obtained on a confocal Raman microscope (Renishaw Invia, UK) equipped with multiple lasers. Three wavelengths, including 632.8, 691, and 785 nm, were used as the

excitation source in this work. The objective was of ×50 magnification and 0.85 numerical aperture to provide a high incident and collection angle. The laser power used was 0.4 mW, and the typical exposure time for each measurement was 10 s without accumulation unless otherwise stated.

Gap-mode SERS experiments

The Au NPs (with a diameter of 55 and 140 nm) AuIMNPs were prepared according to our previous methods (*39*, *57*). A high-quality Au(111) single-crystal surface was prepared by flame annealing according to the Clavilier method (*58*). Then, the fresh Au(111) single crystal was immersed in a KSI (or DNA origami) solution overnight to form monolayer biomolecules. Afterward, it was dried under a nitrogen gas flow. AuIMNPs were cleaned by centrifugation and then dropped on the single-crystal surface to form a well-defined gap-mode configuration after evaporation of water (fig. S3). In this way, high-quality and highly reproducible gap-mode SERS signals can be obtained.

AFM imaging

Newly prepared MoS_2-Au was immersed in 3 μM MoSBP1 aqueous solution for 1 hour. Afterward, the sample was dried with nitrogen gas. The AFM images were obtained on Dimension FastScan (Bruker). The images were captured using silicon nitride probes (SCANASYST-FLUID+, Bruker) in peak-force mode at room temperature under the protection of argon atmosphere. The preparation of MoS_2-Au was based on the previous method (*59*). MoS_2 samples were purchased from Nanjing MKNANO Tech. Co. Ltd. Au(111) substrates were bought from Phasis, which were 200-nm-thick films on mica. Instead of thermal release tape, we used polydimethylsiloxane tape to perform mechanical exfoliation.

TERS measurement

AFM-TERS measurements were performed on a LabRAM HR Evo Nano system (HORIBA Scientific) with side illumination. A 100× objective with a long working distance and a numerical aperture of 0.7 (Mitutoyo) was used for both excitation and collection. A 632.8-nm laser was used for TERS measurements with a p-polarization. The laser power used was 0.1 mW, and the exposure time was 1 s. Au-coated AFM TERS probes used in the TERS measurements were produced by electrochemical deposition of Au on commercial Si AFM tips (ACCESS SNC, Applied Nanostructures Inc.) (*60*). TERS spectra were obtained in contact mode.

Interpreting SERS spectra of DNA

Following a similar idea, we simplified the DNA or RNA structure into the coarse-grained model with several vectors as shown in fig. S14A. With the given structure of DNA, the algorithm can calculate the SERS spectra according to Eq. 2. Furthermore, unexpectedly, up to now, there are still no calculation results of deoxyadenylic monophosphate (dAMP), deoxyguanosine monophosphate (dGMP), deoxycytidine monophosphate (dCMP), and deoxythymidine monophosphate (dTMP) in the literature. We performed DFT calculation of their Raman spectra as well as that of A, G, C, and T. It can be found in fig. S14 (B to E) that peaks at about 800 cm^{-1} emerged, which are assigned to O-P-O stretching vibration (*61*). The ring breathing mode of C and T may also appear at this position; therefore, the overlaps should be carefully considered. To demonstrate its feasibility in predicting SERS spectra of DNA, we first revisited a well-defined system by Halas and colleagues as shown in fig. S15A (*62*). The correlation coefficient of the two spectra in fig. S15 (A and B) is 0.68, which indicates good performance of SPARC. We then constructed another well-defined system by orientation-controlled immobilization of DNA origami as indicated in fig. S15C. The DNA-origami contains a series of tetrahedral DNA nanostructures (TDNs) with three thiol groups on vertices to anchor firmly to the Au(111) surface (*63*). Following this method, we used TDN-7 (whose edges contain 7 base pairs giving a length of the edge to be 2.4 nm) to construct this system. As shown in fig. S15D, the averaged spectrum reveals a good agreement with the theoretical spectrum by SPARC, showing a Spearman's rank correlation coefficient of 0.64.

Supplementary Materials

This PDF file includes:

Figs. S1 to S15
Tables S1 and S2

REFERENCES AND NOTES

1. P. K. Panda, M. N. Arul, P. Patel, S. K. Verma, W. Luo, H.-G. Rubahn, Y. K. Mishra, M. Suar, R. Ahuja, Structure-based drug designing and immunoinformatics approach for SARS-CoV-2. *Sci. Adv.* **6**, eabb8097 (2020).
2. J. P. K. Bravo, M.-S. Liu, G. N. Hibshman, T. L. Dangerfield, K. Jung, R. S. McCool, K. A. Johnson, D. W. Taylor, Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature* **603**, 343–347 (2022).
3. L. Cao, B. Coventry, I. Goreshnik, B. Huang, W. Sheffler, J. S. Park, K. M. Jude, I. Marković, R. U. Kadam, K. H. G. Verschueren, K. Verstraete, S. T. R. Walsh, N. Bennett, A. Phal, A. Yang, L. Kozodoy, M. DeWitt, L. Picton, L. Miller, E.-M. Strauch, N. D. DeBouver, A. Pires, A. K. Bera, S. Halabiya, B. Hammerson, W. Yang, S. Bernard, L. Stewart, I. A. Wilson, H. Ruohola-Baker, J. Schlessinger, S. Lee, S. N. Savvides, K. C. Garcia, D. Baker, Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022).
4. H. Chen, O. Simoska, K. Lim, M. Grattieri, M. Yuan, F. Dong, Y. S. Lee, K. Beaver, S. Weliwatte, E. M. Gaffney, S. D. Minteer, Fundamentals, applications, and future directions of Bioelectrocatalysis. *Chem. Rev.* **120**, 12903–12993 (2020).
5. F. Sesterhenn, C. Yang, J. Bonet, J. T. Cramer, X. Wen, Y. Wang, C.-I. Chiang, L. A. Abriata, I. Kucharska, G. Castoro, S. S. Vollers, M. Galloux, E. Dheilly, S. Rosset, P. Corthésy, S. Georgeon, M. Villard, C.-A. Richard, D. Descamps, T. Delgado, E. Oricchio, M.-A. Rameix-Welti, V. Más, S. Ervin, J.-F. Eléouët, S. Riffault, J. T. Bates, J.-P. Julien, Y. Li, T. Jardetzky, T. Krey, B. E. Correia, De novo protein design enables the precise induction of RSV-neutralizing antibodies. *Science* **368**, (2020).
6. Y. Lu, J. Liu, L. Zhu, B. Zhang, J. He, 3D reconstruction from cryo-EM projection images using two spherical embeddings. *Commun. Biol.* **5**, 304 (2022).
7. T. Herrmann, P. Güntert, K. Wüthrich, Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR* **24**, 171–189 (2002).
8. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
9. J. B. Stiller, R. Otten, D. Häussinger, P. S. Rieder, D. L. Theobald, D. Kern, Structure determination of high-energy states in a dynamic protein ensemble. *Nature* **603**, 528–535 (2022).
10. E. Lerner, T. Cordes, A. Ingargiola, Y. Alhadid, S. Chung, X. Michalet, S. Weiss, Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer. *Science* **359**, eaan1133 (2018).
11. L. Zheng, N. Liu, X. Gao, W. Zhu, K. Liu, C. Wu, R. Yan, J. Zhang, X. Gao, Y. Yao, B. Deng, J. Xu, Y. Lu, Z. Liu, M. Li, X. Wei, H.-W. Wang, H. Peng, Uniform thin ice on ultraflat graphene for high-resolution cryo-EM. *Nat. Methods* **20**, 123–130 (2023).
12. D. Buhrke, P. Hildebrandt, Probing structure and reaction dynamics of proteins using time-resolved resonance raman spectroscopy. *Chem. Rev.* **120**, 3577–3630 (2020).
13. K. Henzler-Wildman, D. Kern, Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
14. J. Langer, D. Jimenez de Aberasturi, J. Aizpurua, R. A. Alvarez-Puebla, B. Auguié, J. J. Baumberg, G. C. Bazan, S. E. J. Bell, A. Boisen, A. G. Brolo, J. Choo, D. Cialla-May, V. Deckert, L. Fabris, K. Faulds, F. J. García de Abajo, R. Goodacre, D. Graham, A. J. Haes, C. L. Haynes, C.

Huck, T. Itoh, M. Käll, J. Kneipp, N. A. Kotov, H. Kuang, E. C. Le Ru, H. K. Lee, J.-F. Li, X. Y. Ling, S. A. Maier, T. Mayerhöfer, M. Moskovits, K. Murakoshi, J.-M. Nam, S. Nie, Y. Ozaki, I. Pastoriza-Santos, J. Perez-Juste, J. Popp, A. Pucci, S. Reich, B. Ren, G. C. Schatz, T. Shegai, S. Schlücker, L.-L. Tay, K. G. Thomas, Z.-Q. Tian, R. P. Van Duyne, T. Vo-Dinh, Y. Wang, K. A. Willets, C. Xu, H. Xu, Y. Xu, Y. S. Yamamoto, B. Zhao, L. M. Liz-Marzán, Present and future of surface-enhanced raman scattering. *ACS Nano* **14**, 28–117 (2020).

15. W. Fang, S. Jia, J. Chao, L. Wang, X. Duan, H. Liu, Q. Li, X. Zuo, L. Wang, L. Wang, N. Liu, C. Fan, Quantizing single-molecule surface-enhanced Raman scattering with DNA origami metamolecules. *Sci. Adv.* **5**, eaau4506 (2019).
16. C. Zong, M. Xu, L.-J. Xu, T. Wei, X. Ma, X.-S. Zheng, R. Hu, B. Ren, Surface-enhanced Raman spectroscopy for bioanalysis: Reliability and challenges. *Chem. Rev.* **118**, 4946–4980 (2018).
17. E. Feng, T. Zheng, X. He, J. Chen, Y. Tian, A novel ternary heterostructure with dramatic SERS activity for evaluation of PD-L1 expression at the single-cell level. *Sci. Adv.* **4**, eaau3494 (2018).
18. C. Y. Li, S. Duan, B. Y. Wen, S. B. Li, M. Kathiresan, L. Q. Xie, S. Chen, J. R. Anema, B. W. Mao, Y. Luo, Z. Q. Tian, J. F. Li, Observation of inhomogeneous plasmonic field distribution in a nanocavity. *Nat. Nanotechnol.* **15**, 922–926 (2020).
19. X. X. Han, G. G. Huang, B. Zhao, Y. Ozaki, Label-free highly sensitive detection of proteins in aqueous solutions using surface-enhanced raman scattering. *Anal. Chem.* **81**, 3329–3333 (2009).
20. L.-J. Xu, C. Zong, X.-S. Zheng, P. Hu, J.-M. Feng, B. Ren, Label-free detection of native proteins by surface-enhanced Raman spectroscopy using iodide-modified nanoparticles. *Anal. Chem.* **86**, 2238–2245 (2014).
21. H. J. Park, S. Cho, M. Kim, Y. S. Jung, Carboxylic acid-functionalized, graphitic layer-coated three-dimensional sers substrate for label-free analysis of alzheimer's disease biomarkers. *Nano Lett.* **20**, 2576–2584 (2020).
22. X. Dai, W. Fu, H. Chi, V. S. D. Mesias, H. Zhu, C. W. Leung, W. Liu, J. Huang, Optical tweezers-controlled hotspot for sensitive and reproducible surface-enhanced Raman spectroscopy characterization of native protein structures. *Nat. Commun.* **12**, 1292 (2021).
23. K. Ikeda, N. Fujimoto, H. Uehara, K. Uosaki, Raman scattering of aryl isocyanide monolayers on atomically flat Au(111) single crystal surfaces enhanced by gap-mode plasmon excitation. *Chem. Phys. Lett.* **460**, 205–208 (2008).
24. J. F. Li, Y. F. Huang, Y. Ding, Z. L. Yang, S. B. Li, X. S. Zhou, F. R. Fan, W. Zhang, Z. Y. Zhou, D. Y. Wu, B. Ren, Z. L. Wang, Z. Q. Tian, Shell-isolated nanoparticle-enhanced Raman spectroscopy. *Nature* **464**, 392–395 (2010).
25. Y.-H. Wang, S. Zheng, W.-M. Yang, R.-Y. Zhou, Q.-F. He, P. Radjenovic, J.-C. Dong, S. Li, J. Zheng, Z.-L. Yang, G. Attard, F. Pan, Z.-Q. Tian, J.-F. Li, In situ Raman spectroscopy reveals the structure and dissociation of interfacial water. *Nature* **600**, 81–85 (2021).
26. F. Benz, M. K. Schmidt, A. Dreismann, R. Chikkaraddy, Y. Zhang, A. Demetriadou, C. Carnegie, H. Ohadi, B. de Nijs, R. Esteban, J. Aizpurua, J. J. Baumberg, Single-molecule optomechanics in "picocavities". *Science* **354**, 726–729 (2016).
27. M. Banik, P. Z. El-Khoury, A. Nag, A. Rodriguez-Perez, N. Guarrottxena, G. C. Bazan, V. A. Apkarian, Surface-enhanced raman trajectories on a nano-dumbbell: Transition from field to charge transfer plasmons as the spheres fuse. *ACS Nano* **6**, 10343–10354 (2012).
28. D. V. Chulhai, L. Jensen, Determining molecular orientation with surface-enhanced raman scattering using inhomogenous electric fields. *J. Phys. Chem. C* **117**, 19622–19631 (2013).
29. E. Aprà, A. Bhattarai, P. Z. El-Khoury, Gauging molecular orientation through time domain simulations of surface-enhanced Raman scattering. *J. Phys. Chem. A* **123**, 7142–7147 (2019).
30. L. Jensen, C. M. Aikens, G. C. Schatz, Electronic structure methods for studying surface-enhanced Raman scattering. *Chem. Soc. Rev.* **37**, 1061–1073 (2008).
31. J. Mullin, G. C. Schatz, Combined linear response quantum mechanics and classical electrodynamics (QM/ED) method for the calculation of surface-enhanced Raman spectra. *J. Phys. Chem. A* **116**, 1931–1938 (2012).
32. S. Duan, G. Tian, Y. Ji, J. Shao, Z. Dong, Y. Luo, Theoretical modeling of plasmon-enhanced Raman images of a single molecule with subnanometer resolution. *J. Am. Chem. Soc.* **137**, 9515–9518 (2015).
33. S. Duan, G. Tian, Z. Xie, Y. Luo, Gauge invariant theory for super high resolution Raman images. *J. Chem. Phys.* **146**, 194106 (2017).
34. W. Hu, S. Ye, Y. Zhang, T. Li, G. Zhang, Y. Luo, S. Mukamel, J. Jiang, Machine learning protocol for surface-enhanced Raman spectroscopy. *J. Phys. Chem. Lett.* **10**, 6026–6031 (2019).
35. P. Bouř, J. Sopková, L. Bednárová, P. Maloň, T. A. Keiderling, Transfer of molecular property tensors in cartesian coordinates: A new algorithm for simulation of vibrational spectra. *J. Comput. Chem.* **18**, 646–659 (1997).
36. W. Yang, Direct calculation of electron density in density-functional theory. *Phys. Rev. Lett.* **66**, 1438–1441 (1991).
37. A. Xomalis, X. Zheng, A. Demetriadou, A. Martínez, R. Chikkaraddy, J. J. Baumberg, Interfering plasmons in coupled nanoresonators to boost light localization and SERS. *Nano Lett.* **21**, 2512–2518 (2021).
38. X. Wang, M. Li, L. Meng, K. Lin, J. Feng, T. Huang, Z. Yang, B. Ren, Probing the location of hot spots by surface-enhanced Raman spectroscopy: Toward uniform substrates. *ACS Nano* **8**, 528–536 (2014).
39. L.-J. Xu, Z.-C. Lei, J. Li, C. Zong, C. J. Yang, B. Ren, Label-free surface-enhanced Raman spectroscopy detection of DNA with single-base sensitivity. *J. Am. Chem. Soc.* **137**, 5149–5154 (2015).
40. J. Chen, E. Zhu, J. Liu, S. Zhang, Z. Lin, X. Duan, H. Heinz, Y. Huang, J. J. De Yoreo, Building two-dimensional materials one row at a time: Avoiding the nucleation barrier. *Science* **362**, 1135–1139 (2018).
41. J. Liu, J. Zeng, C. Zhu, J. Miao, Y. Huang, H. Heinz, Interpretable molecular models for molybdenum disulfide and insight into selective peptide recognition. *Chem. Sci.* **11**, 8708–8722 (2020).
42. T.-X. Huang, X. Cong, S.-S. Wu, K.-Q. Lin, X. Yao, Y.-H. He, J.-B. Wu, Y.-F. Bao, S.-C. Huang, X. Wang, P.-H. Tan, B. Ren, Probing the edge-related properties of atomically thin MoS2 at nanoscale. *Nat. Commun.* **10**, 5544 (2019).
43. S. Ye, K. Zhong, J. Zhang, W. Hu, J. D. Hirst, G. Zhang, S. Mukamel, J. Jiang, A machine learning protocol for predicting protein infrared spectra. *J. Am. Chem. Soc.* **142**, 19071–19077 (2020).
44. L. B. Wright, P. M. Rodger, S. Corni, T. R. Walsh, GolP-CHARMM: First- principles based force fields for the interaction of proteins with Au(111) and Au(100). *J. Chem. Theory. Comput.* **9**, 1616–1630 (2013).
45. F. Iori, R. Di Felice, E. Molinari, S. Corni, GolP: An atomistic force-field to describe the interaction of proteins with Au(111) surfaces in water. *J. Comput. Chem.* **30**, 1465–1476 (2009).
46. G. Brancolini, D. B. Kokh, L. Calzolai, R. C. Wade, S. Corni, Docking of ubiquitin to gold nanoparticles. *ACS Nano* **6**, 9863–9878 (2012).
47. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox. (Wallingford, CT, 2009), vol. Gaussian 09 Rev. D.01.
48. R. Pang, L.-J. Yu, M. Zhang, Z.-Q. Tian, D.-Y. Wu, DFT study of hydrogen-bonding interaction, solvation effect, and electric-field effect on Raman spectra of hydrated proton. *J. Phys. Chem. A* **120**, 8273–8284 (2016).
49. P. Eastman, P. K. Behara, D. L. Dotson, R. Galvelis, J. E. Herr, J. T. Horton, Y. Mao, J. D. Chodera, B. P. Pritchard, Y. Wang, G. De Fabritiis, T. E. Markland, SPICE, a dataset of drug-like molecules and peptides for training machine learning potentials. *Sci. Data* **10**, 11 (2023).
50. J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER suite: Protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
51. A. Roy, A. Kucukural, Y. Zhang, I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
52. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).
53. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
54. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
55. P. B. Johnson, R. W. Christy, Optical constants of the noble metals. *Phys. Rev. B* **6**, 4370–4379 (1972).
56. T. Lu, F. Chen, Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* **33**, 580–592 (2012).
57. J. F. Li, X. D. Tian, S. B. Li, J. R. Anema, Z. L. Yang, Y. Ding, Y. F. Wu, Y. M. Zeng, Q. Z. Chen, B. Ren, Z. L. Wang, Z. Q. Tian, Surface analysis using shell-isolated nanoparticle-enhanced Raman spectroscopy. *Nat. Protoc.* **8**, 52–65 (2013).

58. J. Clavilier, R. Faure, G. Guinet, R. Durand, Preparation of monocrystalline Pt microelectrodes and electrochemical study of the plane surfaces cut in the direction of the {111} and {110} planes. *J. Electroanal. Chem. Interfacial Electrochem.* **107**, 205–209 (1980).
59. F. Wu, Z. Liu, N. Hawthorne, M. Chandross, Q. Moore, N. Argibay, J. F. Curry, J. D. Batteas, Formation of coherent 1H–1T heterostructures in single-layer MoS_2 on Au(111). *ACS Nano* **14**, 16939–16950 (2020).
60. L.-K. Yang, T.-X. Huang, Z.-C. Zeng, M.-H. Li, X. Wang, F.-Z. Yang, B. Ren, Rational fabrication of a gold-coated AFM TERS tip by pulsed electrodeposition. *Nanoscale* **7**, 18225–18231 (2015).
61. E. Garcia-Rico, R. A. Alvarez-Puebla, L. Guerrini, Direct surface-enhanced Raman scattering (SERS) spectroscopy of nucleic acids: From fundamental studies to real-life applications. *Chem. Soc. Rev.* **47**, 4909–4923 (2018).
62. A. Barhoumi, D. Zhang, N. J. Halas, Correlation of molecular orientation and packing density in a dsDNA self-assembled monolayer observable with surface-enhanced Raman spectroscopy. *J. Am. Chem. Soc.* **130**, 14040–14041 (2008).
63. M. Lin, J. Wang, G. Zhou, J. Wang, N. Wu, J. Lu, J. Gao, X. Chen, J. Shi, X. Zuo, C. Fan, Programmable engineering of a biosensing interface with tetrahedral DNA nanostructures for ultrasensitive DNA detection. *Angew. Chem. Int. Ed.* **54**, 2151–2155 (2015).

Acknowledgments: We thank J. Li and S. Duan for helpful discussion and kindness in providing the original data to reproduce Fig. 2 (B and D). We also thank H. Heinz for providing their MD simulation results about MoSBP1 on MoS_2. **Funding:** This work was supported by the National Natural Science Foundation (grant nos. 22021001, 22227802, 22104125, 22032004, 21633005, 21790354, and 92061118) of China, China postdoctoral science foundation (2021M691870), and the Fundamental Research Funds for the Central Universities (20720220018). **Author contributions:** H.M., X.W., and B.R. conceived the idea and designed the experiments. H.M. and S.Y. performed the experiments, COMSOL simulations, and conducted the data analysis. H.M. and X.L. developed the Python code of SPARC. H.M. and D.-Y.W. performed DFT calculations and calculated PDs. H.M., J.L., and B.W. performed MD simulations. M.C., X.Z., Y.-F.B., and X.P. prepared the MoS_2/Au sample. H.Y. and H.-L.W. conducted the AFM mapping of MoSBP1/MoS_2/Au. L.L. prepared the engineered KSI. Y.-F.B., N.C., L.Z., and H.F. prepared Au(111) single-crystal surface and gold NPs. K.D., G.Y., and C.F. prepared the sample of DNA origami. H.M., X.W., D.-Y.W., and B.R. supervised the project. All authors participated in the result discussion and the manuscript preparation. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. A full version of SPARC is available from the Zenodo repository at https://doi.org/10.5281/zenodo.8261122. In addition, a light version without overlapped integrals function can be obtained from the GitHub repository at https://github.com/Oaham725/SPARC.

Submitted 20 March 2023
Accepted 23 October 2023
Published 22 November 2023
10.1126/sciadv.adh8362

ScienceAdvances

Rapidly determining the 3D structure of proteins by surface-enhanced Raman spectroscopy

Hao Ma, Sen Yan, Xinyu Lu, Yi-Fan Bao, Jia Liu, Langxing Liao, Kun Dai, Maofeng Cao, Xiaojiao Zhao, Hao Yan, Hai-Long Wang, Xiaohui Peng, Ningyu Chen, Huishu Feng, Lilin Zhu, Guangbao Yao, Chunhai Fan, De-Yin Wu, Binju Wang, Xiang Wang, and Bin Ren

Sci. Adv. **9** (47), eadh8362. DOI: 10.1126/sciadv.adh8362

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.