JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

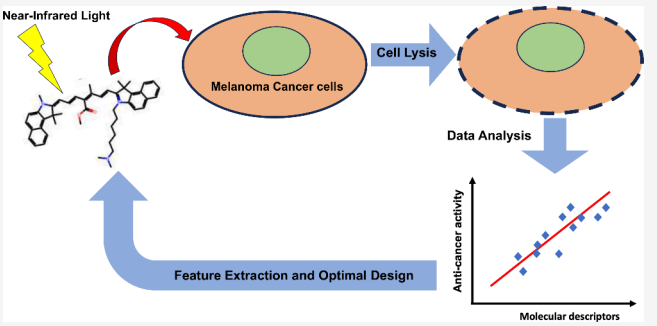
Chemoinformatics Insights on Molecular Jackhammers and Cancer Cells

Ciceron Ayala-Orozco, Hamid Teimouri, Angela Medvedeva, Bowen Li, Alex Lathem, Gang Li, Anatoly B. Kolomeisky,* and James M. Tour*

Cite This: J. Chem. Inf. Model. 2024, 64, 5570−5579 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: One of the most challenging tasks in modern medicine is to find novel efficient cancer therapeutic methods with minimal side effects. The recent discovery of several classes of organic molecules known as "molecular jackhammers" is a promising development in this direction. It is known that these molecules can directly target and eliminate cancer cells with no impact on healthy tissues. However, the underlying microscopic picture remains poorly understood. We present a study that utilizes theoretical analysis together with experimental measurements to clarify the microscopic aspects of jackhammers' anticancer activities. Our physical-chemical approach combines statistical analysis with chemoinformatics methods to design and optimize molecular jackhammers. By correlating specific physical-chemical properties of these molecules with their abilities to kill cancer cells, several important structural features are identified and discussed. Although our theoretical analysis enhances understanding of the molecular interactions of jackhammers, it also highlights the need for further research to comprehensively elucidate their mechanisms and to develop a robust physical-chemical framework for the rational design of targeted anticancer drugs.



INTRODUCTION

Despite significant progress in developing new cancer treatments, methods to reliably eradicate different types of tumors continue to be out of reach. Chemotherapy drugs and radiation therapy have dramatically increased cancer patient's survival rates,[1−5] but the collateral damage to healthy cells during their applications lead to dangerous side effects.[6−11] Considering the limitations of current cancer treatments, a new therapy that minimizes toxicity to healthy cells while effectively targeting and destroying cancer cells could offer significant benefits.

Recent experimental studies revealed that several classes of organic molecules based on aminocyanines, known as molecular jackhammers (MJHs), undergo whole-molecule concerted vibrations upon exposure to near-infrared (NIR) light.[12,13] Upon absorbing a photon in the NIR range, these molecules exhibit collective electron oscillations, also known as a molecular plasmon, that stimulate them to be expanded and contracted longitudinally and axially with large amplitudes.[14] The molecular plasmon is a collective electronic oscillation in the molecule that couples with the whole-molecule vibration in a concerted manner.[14] This property of concerted longitudinal and tangential motion is known as vibronic-driven action (VDA) since the electronic and vibrational oscillations are coupled. When the MJH associates with a cell membrane, experiments found that these vibrations might break the cellular membrane, eventually leading to cell death.[13] The ability to mechanically destroy cancer cells while having low toxicity to healthy cells is further strengthened by the relatively high penetration depth of the NIR light in human tissues compared to other waves of the electromagnetic spectrum.[15]

The unique properties of MJH in the selective eradication of cancer cells stimulated significant efforts to create a new anticancer therapy based on the action of jackhammers. However, our understanding of how MJH functions remains vague, slowing down the development of such a new method. The ability of a molecule to undergo coordinated vibronic motion depends heavily on the molecular structure, and the insertion or deletion of even one atom can drastically affect the mode.[16] Thus, quick examination and chemical intuition are frequently not sufficient to correctly predict the VDA of a molecular jackhammer, while the advanced computational methods like Density-Functional Theory (DFT) calculations needed to predict the existence of molecular plasmons are prohibitively time- and resource-intensive given a very large number of potential candidate molecules.[13]

Received: May 8, 2024
Revised: June 7, 2024
Accepted: June 25, 2024
Published: July 3, 2024

https://doi.org/10.1021/acs.jcim.4c00806
J. Chem. Inf. Model. 2024, 64, 5570−5579

The additional complexity in studying the mechanisms of action of MJH comes from the fact that some properties currently can only be evaluated in experiments. The VDA is closely associated with a quantity known as the Experimental Plasmonicity Index (EPI).[16] Not every organic molecule can behave as a molecular plasmon since specific chemical structure, size, and charge are required to allow the collective electron oscillations.[16] The EPI is a measurement that predicts and estimates the plasmonic character of an organic molecule. The higher the EPI, the higher the corresponding plasmonicity. Briefly, the EPI is obtained by measuring the optical response of its plasmon resonance as a function of the dielectric constant of the medium. The EPI is a predictive tool that can be utilized to obtain structure−activity relationships in MJH by examining their EPI values, VDA activity, and chemical structures. However, the EPI is only limited for structures that are measured experimentally and not for untested or theoretical structures. In addition, the EPI does not predict the toxicity of aminocyanines for healthy human cells. This means that the ratio of VDA to toxicity, known as the therapeutic index (TI), cannot be predicted from the EPI. The inability to predict TI based on structure alone inhibits the ability of synthetic chemists to identify and produce molecules which are likely to be effective for cancer therapy.

Our previous experimental studies allowed us to introduce several empirical rules on how to improve VDA activity. More specifically, we concentrated on VDA IC_{50}, which is the concentration required to permeabilize the cell membrane in 50% of the cell population.[16] Thus, a low-value VDA IC_{50} corresponds to a more potent jackhammer molecule. To find the molecules with higher potency, as explained above, we developed an empirical descriptor of the strength of VDA activity, the EPI.[16] It measures the ability of the MJH to polarize ("polarizability") in the solvent upon excitation with light.[16] It was found that EPI serves as a good predictor for the structure−activity relationships of plasmon-driven MJHs, with higher EPI values correlating with increased VDA efficacy in the mechanical disruption of the cancer cells.[16]

In this paper, we investigate the mechanisms of MJH by using a novel physical-chemical approach that is driven by a hypothesis that anticancer activities of MJH correlate with some specific physical-chemical properties of these molecules. We propose a theoretical framework supported by experimental observations that allows us to identify the most important physical-chemical features that correlate with the potency, toxicity, and EPI of different MJH molecules. Using correlation analysis, we determined key molecular features that exhibit strong correlations with the anticancer properties of jackhammers. Analyzing each of these selected features allows us to explain the connection between them and particular functional groups in the MJHs. Finally, we demonstrate that this physical-chemical method holds predictive power for synthesizing new therapeutically effective molecular jackhammers. This methodology has the potential to guide the rapid development of a minimally invasive and safe NIR therapy for cancer treatment and help bring this technology to maturity.

MATERIALS AND METHODS

Measurement of VDA Activity of Molecules Targeting A375 Human Melanoma Cells. The specific details of the methods have been already described before.[16] The LED lights for the activation of VDA were the 730 nm LED (model UHP-F-730) and 630 nm LED (model UHP-F-630) illumination systems obtained from Prizmatix, Israel. The list of different molecular jackhammers and their experimentally determined properties are assembled in Table 1.

Table 1. Summary of Anti-cancer Activity Data for Corresponding Molecular Jackhammers; VDA, Toxicity, and EPI Values Were Obtained from Reference 15

Molecule	VDA IC_{50} (μM)	Toxicity IC_{50} (μM)	TI	EPI
BL-204	0.12	0.6	5.0	4.6
GL-308-2	0.125	0.5	4.0	4.5
BL-141-2	0.15	0.75	5.0	2.5
BL-142	0.18	2.5	13.89	1.9
Cy7.5-amine	0.25	0.5	2.0	3.2
GL-291-2	0.27	1.5	5.56	1.8
BL-141-1	0.3	0.125	0.42	3.6
GL-297-2	0.75	1.0	1.33	N.A.
GL-286	1	2.875	2.88	N.A.
GL-176	1.2	0.25	0.21	2.6
Cy5.5-amine	2	2.0	1.0	1.1
GL-261-2	3	2.0	0.67	N.A.
Cy7-amine	3.5	2.0	0.57	2.4
GL-328-2	5	7.0	1.4	N.A.
Cy5-amine	8	6.2	0.8	0.9

Cell culture of A375 human melanoma. The A375 cell line was obtained from ATCC (CRL-1619). Cells were cultured in 10 cm polystyrene tissue culture dishes (Corning) containing DMEM with L-glutamine, 4.5 g/L glucose, and sodium pyruvate (Corning Inc. 10013CV) and supplemented with 10% FBS (Corning, 35010CV), 1X MEM vitamin solution (Gibco, 11120052), 1X MEM nonessential amino acid solution (Gibco, 11140050), and penicillin/streptomycin. Typically, 0.5−1 million cells are inoculated per dish and cultured for 2−3 days in an incubator at 37 °C and 5% CO_2. When confluency reached nearly 90%, cells were harvested and transferred into a new dish. For harvesting, cells are detached with 0.05% trypsin-EDTA (Gibco, 25-300-054).

Vibronic-driven action (VDA) activity. A375 cells were cultured as described before. The cells were inoculated at 2 million cells per dish (10 cm polystyrene tissue culture dishes) and cultured for 2 days. They were harvested using 0.05% trypsin−EDTA (Gibco, 25-300-054), and then the cells were prepared in a cell suspension containing 2×10^5 cells/mL in DMEM media with L-glutamine, 4.5 g/L glucose, and sodium pyruvate (Corning Inc. 10013CV) and supplemented with 10% FBS (Corning, 35010CV), 1X MEM vitamin solution (Gibco, 11120052), 1X MEM nonessential amino acid solution (Gibco, 11140050), and penicillin/streptomycin. 1 mL of this cell suspension containing 2×10^5 cells was used in each treatment. In a 1.5 mL Eppendorf tube, 1 μL of stock solution containing 2 mM Cy7.5-amine (or other cyanine molecule or other concentrations) in DMSO (Fisher, 99.7%) was placed in the bottom of the tube, then 1 mL of the cell suspension was added into the tube to get a final concentration of 2 μM of Cy7.5-amine containing 0.1% DMSO and 2×10^5 cells. The mixture was then incubated at 37 °C and 5% CO_2 for 30 min. Then, 1 μM DAPI was added to the cell suspension, and the cell suspension was transferred to a 35 mL polystyrene tissue culture dish.

The cells were immediately treated under NIR light at 730 nm and 80 mW/cm^2 or 630 nm LED for the Cy5-amine and

https://doi.org/10.1021/acs.jcim.4c00806
J. Chem. Inf. Model. 2024, 64, 5570−5579

Cy5.5-amine, over 10 min using LED light sources (Prizmatix, UHP-F-730 or UHP-F-630, Israel) that cover the entire dish. The instrument for flow cytometry analysis (SONY, MA900 Multi-Application Cell Sorter) was already set up and calibrated by the time the light treatment was finished. Therefore, as soon as the 10 min light treatment was completed, the cell suspension was rapidly transferred from the 35 mm dish to a flow cytometry tube, and the cells were analyzed for DAPI permeabilization and Cy7.5-amine binding. It took 30 s to load the sample and to start cell counting. The permeabilization of cells was measured as DAPI-positive cells and occurred immediately. The light intensity was measured using an Optical Power Meter from Thorlabs, sensor model S302C, and console model PM100D. The same protocol was repeated for all MJH molecules in the library.

Experimental Plasmonicity Index (EPI). The EPI has been described before.[16] It is an experimental method that measures the plasmonic character of each molecule based on the optical response of the molecule to the polarity of the solvent. Molecular plasmons are associated with electron oscillations in the excited molecule, which are highly polarizable by the polarity of the solvent. Therefore, changes in the dielectric constant (κ) of the solvent significantly affect the polarizability of the electrons and therefore should affect the resonance frequency. These changes in the resonance frequency are observed as a change in the absorption spectrum. To evaluate the EPI, each molecule was solubilized in various solvents at 2.7 μM, and the UV−vis spectrum was measured. We opted to choose a variety of solvents with a wide range of dielectric constants: isopropanol (κ = 19.92), ethanol (κ = 24.55), methanol (κ = 32.7), water (κ = 80.1), and DMSO (κ = 46.68). Then the dielectric constant was plotted against the absorption coefficients at the λ_{max} of the longitudinal molecular plasmon for each molecule.[13,16] A linear correlation was observed for the data points. The EPI is the slope of the linear correlation function multiplied by a factor of −1000 to obtain integer positive values since the slope is negative, namely, EPI = −1000 × slope.

Crystal Violet Viability Assay for Cell Toxicity. The crystal violet viability assay was used to measure cell viability/toxicity.[17] The working principle of this method is that the viable cells adhere to the surface of the cell culture dish while growing and remain attached through the standard cell culture conditions and during the staining step. In contrast, the dead cells do not adhere to the surface of the cell culture dish, do not grow, and detach easily during the manipulation steps of the assay.

A375 cells were harvested and counted and then diluted at 2 × 10^5 cells/mL to 1 mL of cell suspension. In a 1.5 mL Eppendorf tube, 1 μL of stock solution containing 2 mM Cy7.5-amine (or other cyanine molecule at different concentrations) in DMSO (Fisher, 99.7%) was placed in the bottom of the tube, then 1 mL of the cell suspension was added into the tube to obtain a final concentration of 2 μM of Cy7.5-amine containing 0.1% DMSO and 2 × 10^5 cells. Then, 100 μL of the cells (2 × 10^4 cells) were placed in a 96-well plate. Typically, 7 to 10 concentrations were prepared in quadruplicated samples in a 96-well plate. Then, the cells were cultured for 2 days at 37 °C and 5% CO_2. At the end of the incubation, the media was removed and the cells were washed with 150 μL of PBS once. The cells were stained with 100 μL of 0.5% w/v crystal violet solution in methanol/water (1:1) for 15 min. Then, the crystal violet was removed, and the excess crystal violet was washed away with water. The cells attached to the 96-well plate were dried at room temperature. Then, the crystal violet in each well was solubilized in 100 μL of 3.3% w/v acetic acid in water. The extracted crystal violet was quantified by its absorbance at 570 nm in a 96-well plate reader (Tecan model Infinite 200 PRO, Austria GmbH). The cell viability was calculated from the absorbance relative to the absorbance in the cells without any treatment; cells without treatment were normalized to 100% cell viability.

Extraction of Molecular Descriptors. Our goal is to design molecular jackhammers with optimal anticancer properties. This includes higher potency (low VDA IC_{50} value) and least toxicity (high toxicity IC_{50} value). For this purpose, we have to identify which molecular features correlate most with the optimal therapeutic activity. A molecular descriptor, or molecular feature index, is the numerical output that comes from applying a mathematical procedure to translate the chemical data encapsulated in a molecule's symbolic representation.[18] Molecular descriptors are extensively used in drug discovery research to predict compound activity, toxicity, and pharmacokinetics.[19] They are crucial in QSAR (Quantitative Structure−Activity Relationship) modeling, which connects molecular structure with biological activity or physical-chemical properties.[20,21]

By utilizing the Mordred package[22] version 1.2.0, we extracted 1826 descriptors for each molecule provided in a recognized chemical file format named SMILES.[23] These features fall into three main categories. First, constitutional descriptors, which are the simplest and relate to the molecule's composition without considering the molecular connectivity or shape. They include information such as the count of specific types of atoms, molecular weight, the number of electrons in each atom, and the count of various functional groups. The second type of molecular descriptors is topological (or 2D), which are calculated based on the 2D structure of the molecule, represented as a graph. This category includes autocorrelation descriptors, which evaluate how different atomic properties (such as charge or electronegativity) are correlated with each other across the molecular graph. Other topological descriptors might include molecular connectivity indices, Wiener numbers, and Randić indices. And, the third group of molecular descriptors are physical-chemical features, which include a range of properties like lipophilicity, polarizability, refractivity, and electronic properties. A complete list of Mordred descriptors is provided in Table 3 of ref 22.

Representing molecules using numerical descriptors as points in a multidimensional space allows for evaluating the similarity of two molecules. There are several methods, including cosine similarity and Euclidean distance, for characterizing similarity in high-dimensional space.[24] This method is particularly useful in cheminformatics for tasks such as virtual screening, where the goal is to identify molecules with properties similar to a reference molecule.[25] Representing two arbitrary molecules A and B molecules as two vectors $\boldsymbol{A}$ = $A_1, A_2, ..., A_n$]and $\boldsymbol{B}$ = [$B_1, B_2, ..., B_n$], respectively, we can use cosine similarity formula to calculate the cosine of the angle Θ between vectors $\boldsymbol{A}$ and $\boldsymbol{B}$:

https://doi.org/10.1021/acs.jcim.4c00806
J. Chem. Inf. Model. 2024, 64, 5570−5579

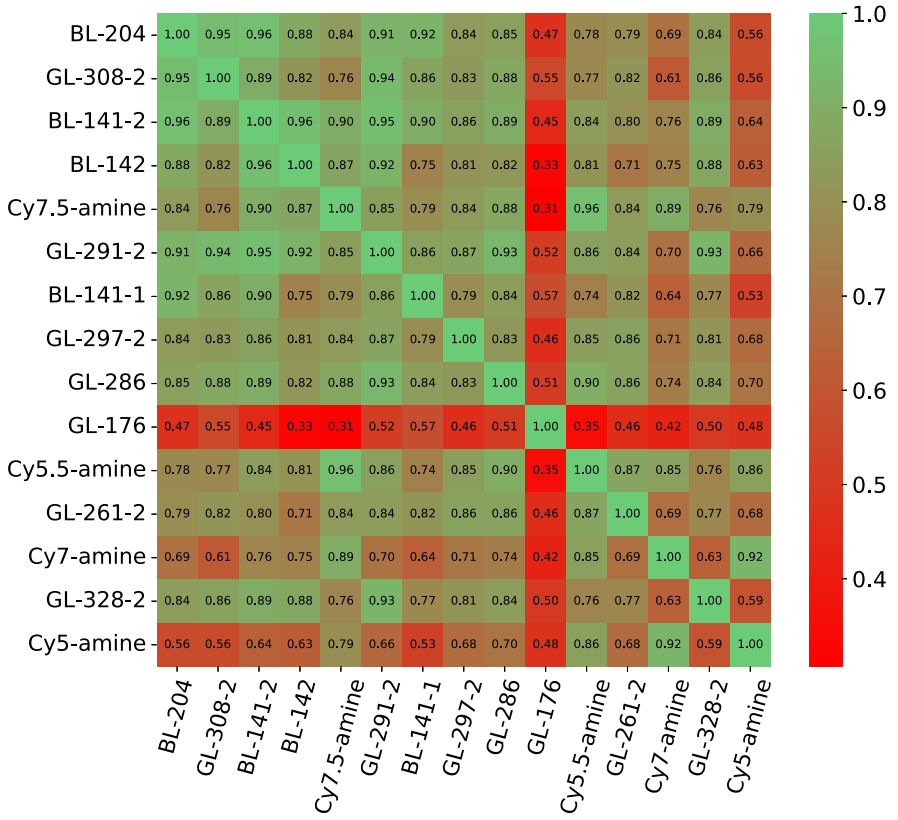


Figure 1. Cosine similarity (eq 1) of the molecular jackhammers as listed in Table 1. Each colored box represents the corresponding similarity of a pair of two different molecules.

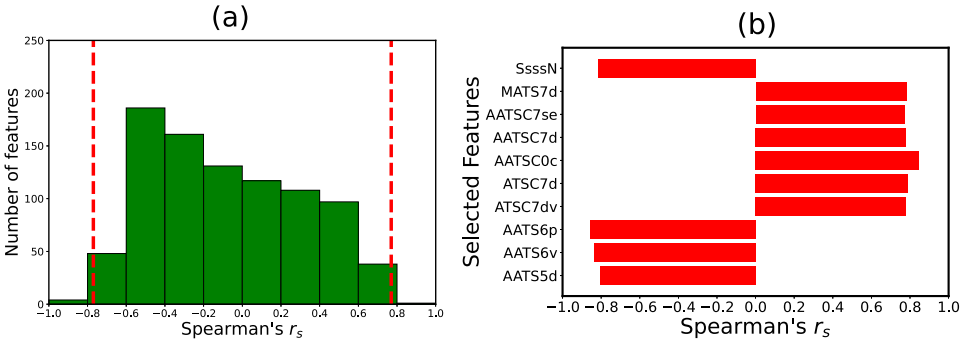


Figure 2. a) Distribution of Spearman correlation coefficients for all molecular descriptors that correlate with VDA IC_{50} of jackhammers. b) Relative importance of molecular descriptors, which highly correlated with the corresponding VDA IC_{50} values. A threshold value of $\rho = 0.77$ was used. Definitions of the selected features can be found in ref 22.

$$\text{Cosine Similarity}\,(\boldsymbol{A}, \boldsymbol{B}) = \cos(\Theta) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \quad (1)$$

Since the scales of physical-chemical features for each molecule are different, it is important to normalize them to have their values to be between 0 and 1. To normalize these quantities, we use $\hat{z} = \frac{(z - z_{\min})}{(z_{\max} - z_{\min})}$. With this scaling, since the physical-chemical features cannot be negative, the cosine similarity is bounded in [0,1]. When the vectors are completely aligned, the cosine similarity is 1, indicating identical descriptor profiles and, presumably, very similar molecular properties. If the vectors are orthogonal, the cosine similarity is 0, suggesting no similarity between the molecular descriptors of the two molecules. Consequently, the cosine similarity depends only on the angle between the vectors, and not on their magnitudes.

RESULTS AND DISCUSSIONS

In Figure 1, we present the results of cosine similarity calculations for all pairs of molecules given in Table 1. For two molecules, BL-204 and BL-141-2, with identical TI of 5, the cosine similarity is equal to 0.96. Likewise, we obtain a significantly low cosine similarity when we compare BL-204 and BL-141-2 with two molecules with low TI, GL-176 (TI = 0.21) and Cy5-amine (TI = 1). Moreover, cosine similarity sets GL-176 (with the lowest TI = 0.21) apart from other molecules (red cross in Figure 1). These results suggest that

https://doi.org/10.1021/acs.jcim.4c00806
J. Chem. Inf. Model. 2024, 64, 5570−5579

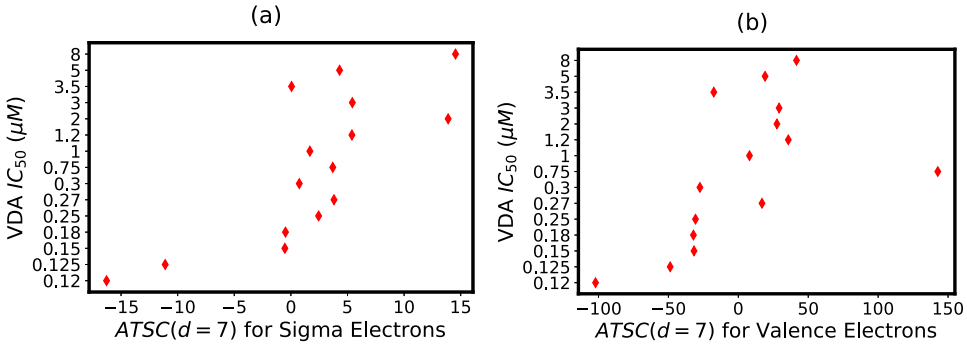


Figure 3. Elucidating the correlations between VDA IC_{50} of molecular jackhammers and two selected molecular descriptors. a) Centered Moreau−Broto autocorrelation (as defined in eq 4) of lag 7 [$ATSC(d = 7)$] weighted by sigma electrons with Spearman's correlation $r_s = 0.79$ and $p = 0.0005$. This descriptor is labeled as ATSC7d in Mordred. b) Centered Moreau−Broto autocorrelation of lag 7 [$ATSC(d = 7)$] weighted by valence electrons with Spearman's correlation $r_s = 0.78$ and $p = 0.0006$. This descriptor is labeled as ATSC7dv in Mordred.[22]

structural similarity might have a predictive power for the therapeutic index of molecular jackhammers.

Current efforts in drug discovery predominantly rely on structure−activity relationships (SAR), where the biological effects of a potential drug are predicted based on the known activities of similar compounds.[26] This approach employs a quantitative structure−activity relationship (QSAR), which assumes the following relationship,

$$\text{Biological Activity} = F(\text{Molecular Descriptors}) \tag{2}$$

Predicting structure−activity relationships with acceptable accuracy requires a huge amount of data. In our case, since we have only a few tested molecules (see Table 1), we limit our analysis to identifying the strongest correlations that may warrant further study. To elucidate such correlations, we utilize Spearman's rank correlation coefficient, which is defined as Pearson's correlation coefficient between the ranks of anticancer activity (including potency, toxicity, EPI, and TI), $R(A)$, and ranks of molecular descriptor, $R(MD)$,

$$r_s = \frac{\langle R(A)R(MD)\rangle - \langle R(A)\rangle\langle R(MD)\rangle}{\sqrt{\langle R^2(A)\rangle - \langle R(A)\rangle^2}\sqrt{\langle R^2(MD)\rangle - \langle R(MD)\rangle^2}} \tag{3}$$

The Spearman's rank correlation coefficient is a tool used to assess how effectively a monotonic function can capture the relationship between two variables. Consequently, it serves as a practical method for analyzing the correlations between two quantities. We computed the Spearman's rank correlations between VDA IC_{50} of molecules and each of their descriptors. The corresponding histograms of correlation coefficients are shown in Figure 2a. Correlation coefficients range from −1 to 1. We are specifically interested in features that are strongly correlated with vibronic-driven activity (VDA IC_{50}). Thus, we set three conditions for the selection of important molecular descriptors. First, we choose an arbitrary threshold ρ for Spearman's correlation, such that $|r_s| \geq \rho$. Second, to ensure that correlations are statistically significant, we consider only p-values of less than 0.005. The third condition is that the minimum and maximum values of the selected molecular descriptor must correspond to the minimum and maximum values of VDA IC_{50}, respectively. In other words, the relationship between VDA IC_{50} and the selected features are expected to be either monotonically increasing or decreasing.

The selected features, which correspond to a threshold of 0.77, are presented in Figure 2b. A similar analysis was performed for toxicity IC_{50} and EPI (see the Supporting Information). Now let us provide a physical-chemical interpretation of the selected features in Figure 2b. Among those features, ATSC stands for Autocorrelation of Topological Structure Centered, and AATSC represents Averaged Autocorrelation of Topological Structure Centered, with their definitions provided as follows. For a molecule with N atoms, the centered Moreau−Broto Autocorrelation of Topological Structure (ATSC) measures the correlation between physical-chemical properties of atom i and atom $i + d$ (according to the SMILES structure),[22]

$$ATSC(d) = \sum_{i=1}^{N-d} (P_i - \overline{P})\cdot(P_{i+d} - \overline{P}) \tag{4}$$

where P_i and P_{i+d} are physical-chemical properties of atom i and atom $i + d$, respectively. Also, $\overline{P} = \frac{1}{N}\sum_{i=1}^{N} P_i$ is the average physical-chemical property of the molecule. Since this method involves centering the property P values by subtracting the mean, the resultant autocorrelation values can be both positive and negative. The parameter d is the topological distance over which autocorrelation is calculated. The unit of topological distance is not a physical length but rather a unitless count of the number of bonds between two atoms. It is a purely mathematical concept used in graph theory and cheminformatics to describe the connectivity of a molecule. Alternatively, one can define the averaged centered Moreau−Broto autocorrelation for a specific distance d, $AATSC(d)$,[22]

$$AATSC(d) = \frac{1}{N-d}\sum_{i=1}^{N-d} (P_i - \overline{P})(P_{i+d} - \overline{P}) \tag{5}$$

$ATSC(d)$ is the sum of topological distances between all pairs of atoms in a molecule that are d bonds apart. Thus, $N - d$ is simply the count of these pairs of atoms. With these definitions, the relationship between the two autocorrelation functions reads as

$$AATSC(d) = \frac{ATSC(d)}{N-d} \tag{6}$$

In Figure 3, we plotted VDA IC_{50} of MJHs versus two autocorrelation functions that characterize the number of sigma electrons (Figure 3a) and the number of valence electrons (Figure 3b). The importance of $ATSC(d = 7)$ for sigma electrons lies in the fact that molecules with the lowest potency ($IC_{50} = 8$ μM) and the molecule with the highest

https://doi.org/10.1021/acs.jcim.4c00806
J. Chem. Inf. Model. 2024, 64, 5570−5579

potency (IC_{50} = 0.12 μM) correspond to the maximum and minimum values of this descriptor, respectively (Figure 3a and Figure 4). A positive $ATSC(d = 7)$ value indicates that the

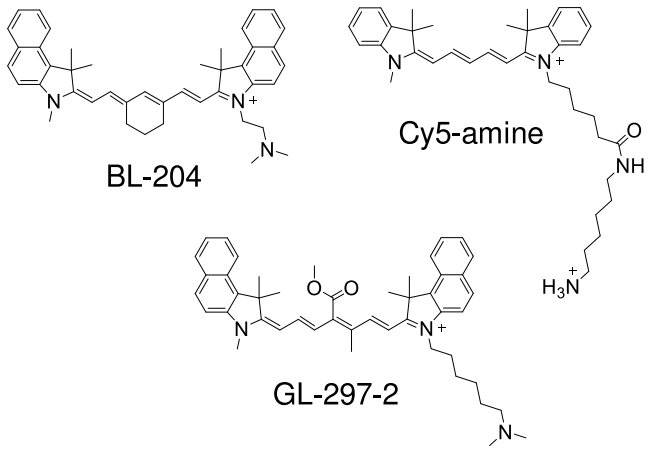


Figure 4. Chemical structures of jackhammer molecules with maximum and minimum potency and EPI values. (top left) BL-204 with IC_{50} = 0.12 μM and EPI = 4.6; (top right) Cy5-amine with IC_{50} = 8 μM and EPI = 0.9; (bottom) GL-297-2 with IC_{50} = 0.75 μM.

numbers of sigma electrons for atoms at a topological distance of 7 bonds tend to be uniformly distributed. In contrast, a negative autocorrelation suggests that atoms at the distance of 7 bonds tend to have very different numbers of sigma electrons. The specific distance d = 7 reveals unique aspects of electron distribution in MJH that might not be evident when considering atoms closer together or further apart. Similar arguments can be given for the dependence of VDA IC_{50} on the autocorrelation of valence electrons (Figure 3b). We elucidated the correlations between $ATSC(d = 7)$ of sigma electrons versus $ATSC(d = 7)$ of valence electrons (see Figure S4 in the SI). Pearson's correlations of 0.68 and a p-value of 0.005 are obtained. This indicates a moderate to strong positive linear relationship between autocorrelations of sigma electrons and autocorrelations of valence electrons.

It is important to note that while sigma electrons, a subset of valence electrons, are involved in forming strong sigma bonds, which provide structural stability to molecules, the rest of the valence electrons (including pi electrons) can participate in various types of weaker bonds and interactions. These interactions are significant in molecular absorption and emission properties, which are influenced by the electronic transitions involving these electrons. For the VDA, both the strict bonding provided by sigma electrons and the more dynamic behavior of other valence electrons are critical. The latter can be influenced by NIR light, leading to induced bond vibrations and subsequent light-driven vibronic activities, which are central to the phenomena observed and reported in our study.

While molecular potency, which is measured by VDA IC_{50}, is an important feature that characterizes the light-driven activity of molecular jackhammers against cancer cells, it is also crucial to consider the toxicity of these molecules. Toxicity is simply defined as the half maximal inhibitory concentration (IC_{50}) of molecular jackhammers in the absence of light-driven activity. Thus, the therapeutic index (TI) is the ratio of toxicity IC_{50} and VDA IC_{50} values. Results of correlation analysis for toxicity IC_{50} are presented in the Supporting Information (see Figure S2). To analyze the toxicity, we selected two important molecular descriptors: the average centered Moreau−Broto autocorrelation for van der Waals volume and the ionization potential of atoms at a distance of 3. The dependence of toxicity (IC_{50}) versus these descriptors is presented in Figure 5.

To understand the relationship between the molecular toxicity and the distribution patterns of van der Waals volumes and ionization potentials in the atoms in MJH, we can present the following arguments. The van der Waals volume is the volume occupied by a molecule and is one of the most fundamental properties of the drug structure controlling biological activity. The molecular size and shape, which are important for the drug−receptor interaction, are generally determined by the van der Waals volume of the molecule. Among many other molecular descriptors, it has been long recognized that the toxicity behavior of drug molecules is significantly correlated with the van der Waals volume.[27,28] This is in agreement with the findings of our theoretical method, which predicts high correlations of the toxicity and the van der Waals volume. The ionization potential was also identified to have a high correlation with the toxicity. This is in agreement with the literature observations.[29] One can argue that since the ionization potential has been recognized as one of the fundamental parameters that define the redox properties

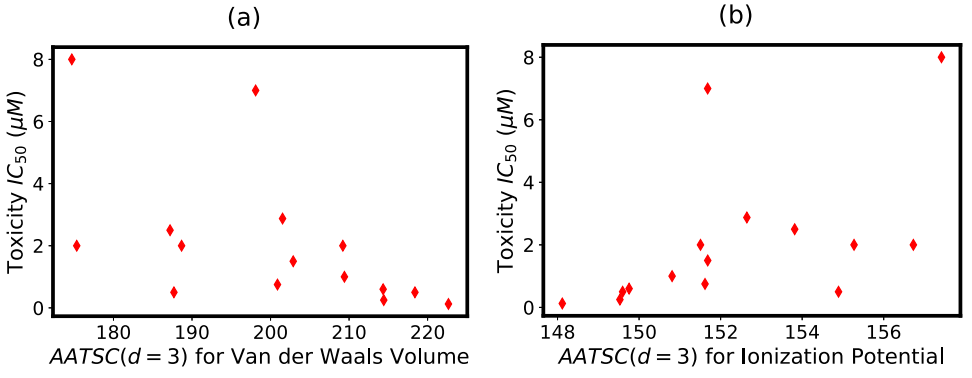


Figure 5. Elucidating the correlations between toxicity of the molecular jackhammers and selected molecular descriptors. a) Averaged Moreau−Broto autocorrelation (as defined in eq 5) of lag 3 [$AATSC(d = 3)$] weighted by van der Waals volume with Spearman's correlation r_s = −0.7 and p = 0.0035. This descriptor is labeled as AATS3v in Mordred. b) Centered Moreau−Broto autocorrelation of lag 3 [$AATSC(d = 3)$] weighted by ionization potential with Spearman's correlation r_s = 0.7 and p = 0.0036. This descriptor is labeled as AATS3i in Mordred.[22]

https://doi.org/10.1021/acs.jcim.4c00806
J. Chem. Inf. Model. 2024, 64, 5570−5579

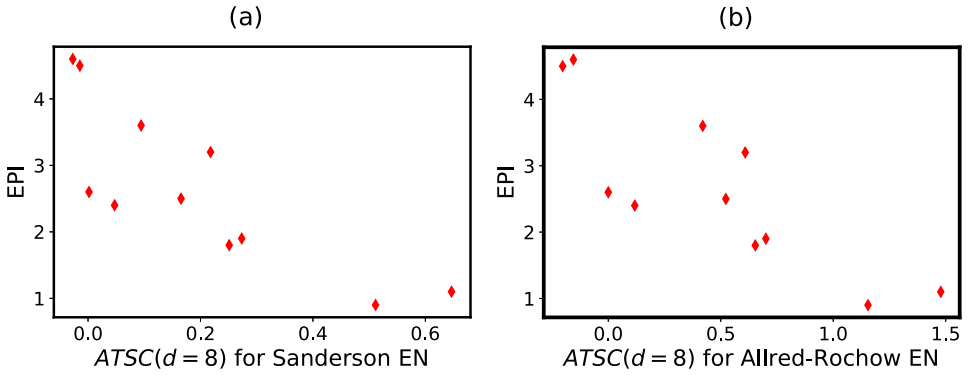


Figure 6. Elucidating the correlations between EPI of the molecular jackhammers and two selected molecular descriptors. a) Centered Moreau−Broto autocorrelation (as defined in eq 4) of lag 8 $[ATSC(d = 8)]$ weighted by Sanderson electronegativity with Spearman's correlation $r_s = -0.86$ and $p = 0.00061$. This descriptor is labeled as ATSC8se in Mordred. b) Centered Moreau−Broto autocorrelation of lag 8 $[ATSC(d = 8)]$ weighted by Allred−Rocow electronegativity with Spearman's correlation $r_s = -0.85$ and $p = 0.00081$. This descriptor is labeled as ATSC8 are in Mordred.[22]

of a drug, it clearly governs the enzymatic metabolic activation and the toxicity of a molecule.[29]

Another important characteristic of molecular jackhammers is the EPI, which directly measures the vibronic activity of a molecule exposed to NIR light. It is important to note that EPI is directly proportional to the potency of molecules, i.e., it is inversely proportional to VDA IC_{50}. Using our theoretical method, we analyzed correlations between EPI of molecular jackhammers and molecular descriptors, and the corresponding results are in the Supporting Information. Two of the important descriptors that correlate with EPI are $ATSC(d = 8)$ weighted by Sanderson electronegativity and $ATSC(d = 8)$ weighted by Allred−Rocow electronegativity (Figure 6). Electronegativity, which is a measure of an atom's ability to attract and hold onto electrons, plays a crucial role in determining a molecule's electronic structure, and thus its optical properties. Unlike other physical properties such as charge and mass, electronegativity cannot be directly measured. Since its quantification relies on theoretical models, there are various definitions for the electronegativity of atoms. Sanderson's electronegativity focuses on the equalization of electron densities within a molecule, emphasizing the average electron density in the valence shell.[30] Allred−Rochow electronegativity is calculated based on the effective nuclear charge acting on valence electrons and the distance to the nucleus, offering a more electrostatic perspective.[31] As a result, these scales yield different numerical values and are used for different applications in Chemistry. Each molecule has unique NIR-active modes depending on its structure and composition. The distribution of electronegativity of atoms across the molecule plays an important role in defining these modes. Thus, molecules with nonuniform electronegativity distributions might have complex absorption spectra and can exhibit specific behaviors under NIR light.

In Figure 4, structures of two different molecular jackhammers with maximum (BL-204) and minimum (Cy5-amine) potency are illustrated. One can see that there are significant topological differences between these molecules. Cy5-amine features a lengthy arm predominantly composed of single bonds, whereas BL-204 contains more ring structures characterized by alternating single and double bonds, thereby enhancing conjugation. It seems that more electronically conjugated molecules possess stronger anticancer abilities, and the correlations with the distributions of electronegativities, as discussed above, reflect this observation.

It is also important to discuss how the selected molecular features can aid in designing new molecular jackhammers with optimal anticancer activities. Because of the strong relationship between EPI and VDA, future synthesis of molecular jackhammers ought to be guided by the likely EPI of candidate molecules. Molecular jackhammers share a common overall structure consisting of a central chain, side functional groups, and arms extending from the two nitrogens at the end of the central chain, but each of the major features can be one of several different functional groups. New molecules could be proposed algorithmically, and features strongly correlated with EPI could be checked before any synthesis is done. For example, the ATSC of lag 8 weighted by Sanderson electronegativity could be checked for a candidate molecule very similar to one already tested, but with an indole replaced by a benzoindole (effectively adding another benzene ring to one end of the molecule). If the descriptor in question is lower, then the correlation in Figure 6 indicates that the candidate molecule is likely to have a higher EPI (and therefore VDA) than the molecule it was based on. Such automated screening can be enacted on hundreds of thousands of potential molecules, with the most promising molecules being synthesized and tested. Additional results could be fed into the analysis of this paper, leading to more and more accurate predictions. Synthesizing new chemicals requires many hours of specialized work, so the ability to rank candidate molecules with even a moderate degree of confidence could save countless hours of trial and error in the laboratory.

SUMMARY AND CONCLUSIONS

In this study, we presented a new idea that the physical-chemical features of molecular jackhammers correlate with their therapeutic anticancer potency and toxicity. This idea was realized by performing a statistical correlation analysis that highlighted the critical role of molecular descriptors in understanding the underlying mechanisms. The application of the Mordred software framework for extracting molecular descriptors has enabled us to obtain a comprehensive analysis, revealing strong correlations between specific molecular features and the activities of MJHs, such as VDA, toxicity, and the EPI.

Our analysis revealed that only a few molecular descriptors highly correlate with VDA potency of molecular jackhammers. Specifically, it was found that distribution of both sigma electrons and valence electrons of atoms throughout each

https://doi.org/10.1021/acs.jcim.4c00806
J. Chem. Inf. Model. 2024, 64, 5570−5579

molecule might have strong connections with the vibronic-driven activity. Autocorrelations of sigma and valence electrons highlight aspects of the molecular structures related to the electronic delocalizations in single (sigma) bonds and double bonds (sigma and pi). The distribution pattern of sigma electrons and valence electrons, as quantified by $ATSC(d = 7)$, are primarily involved in forming the molecule's basic bonding framework. This distribution can influence the molecule's overall electronic environment, which is crucial for absorbing NIR light and undergoing subsequent electronic transitions. Moreover, the distribution of sigma electrons and valence electrons affects the molecule's electron density, polarity, and other electronic properties, which in turn can impact how efficiently it can undergo vibronic transitions when exposed to NIR light. A molecule with a specific pattern of sigma and valence electron distribution, as indicated by its $ATSC(d = 7)$ value, might be more or less effective at absorbing NIR light and undergoing efficient vibronic transitions. For instance, a more uniform distribution of sigma electrons (a higher positive $ATSC(d = 7)$ value) might lead to different vibronic activity compared to a molecule with an uneven distribution (indicated by a lower or negative $ATSC(d = 7)$ value). Molecules that more effectively absorb NIR light and undergo beneficial vibronic transitions may be more potent in their anticancer effects. Thus, the $ATSC(d = 7)$ descriptor could indirectly give insights into which molecules might be more effective in NIR-light-activated anticancer therapies.

Moreover, we found that the pattern of electronegativities of atoms strongly correlate with the EPI. The ability of a molecule to absorb NIR light is primarily determined by its electronic structure, specifically the energy gap between its electronic states. The distribution of electronegativities in a molecule affects its electronic structure by influencing the energy levels of molecular orbitals. Different electronegativities can lead to variations in orbital energies, potentially creating electronic states that are accessible with NIR light. Electronegativity differences between atoms in a molecule lead to the formation of dipole moments. These dipole moments are crucial for the interaction of the molecule with light and support the formation of molecular plasmon, a collective electronic oscillation that couples with a whole molecule vibrational mode in MJH. In NIR spectroscopy, for instance, transitions between different vibrational states are often dipole-allowed, meaning they require a change in the dipole moment of the molecule. The nonuniform distribution of electronegativity affects these moments and thus increases the likelihood of such transitions.

While our theoretical-experimental approach is able to determine specific molecular properties of jackhammers that correlate most with their anticancer abilities, it is important to consider the limitations of this study. Our analysis was performed on a small number of data. Advanced statistical and machine learning methods are needed to establish a stronger relationship between molecular descriptors and anticancer activities of molecular jackhammers. Given the small data set, methods such as linear regression for modeling and principal component analysis (PCA) for dimensionality reduction can be particularly useful.[32] Additionally, leave-one-out cross-validation (LOOCV) can also be utilized for model validation.[33] Future studies should continue to develop and refine chemoinformatics methods to further improve the robustness of the results and design of effective chemical compounds including but not limited to MJHs by systematically considering various molecular features in relation to specific metrics of efficacy (toxicity, VDA, and EPI in this case). It is particularly essential to develop robust methods for small data sets. Moreover, additional analyses are needed for addressing certain aspects of cancer such as drug resistance and tumor heterogeneity. Yet the proposed framework can guide researchers on this path.

The presented framework holds significant promise for the rational development of new cancer therapies and the enhancement of existing anticancer molecules investigated in this study. By leveraging chemoinformatics methods, we can systematically analyze molecular descriptors and uncover critical structural features associated with therapeutic potency and toxicity. This knowledge enables the rational design and optimization of MJHs based on physical-chemical properties, leading to the development of more targeted and efficacious cancer therapeutics. For instance, insights gained from our analysis, such as the correlation between specific molecular features and activities of MJHs like VDA and toxicity, can guide the synthetic modifications of existing molecules to enhance their therapeutic efficacy while minimizing adverse effects. Additionally, the identification of key structural properties, such as the distribution of sigma and valence electrons, offers valuable guidance for the design of new compounds with improved NIR light absorption and efficient vibronic transitions, thereby enhancing their anticancer effects. By applying our physical-chemical methodology iteratively, in conjunction with experimental validations, researchers can sequentially refine and optimize anticancer molecules in the most rational way.

■ ASSOCIATED CONTENT

Data Availability Statement

The data obtained in this work and the in-house scripts are available on GitHub at the following URL: https://github.com/hamid-teimouri/Chemoinformatics_molecular_jackhammers/tree/main

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.4c00806.

SMILES representation of molecular jackhammers; Chemical structures of molecular jackhammers; Distribution of Spearman correlation coefficients for all descriptors that correlated with toxicity IC_{50} of molecular jackhammers; Selected important features that correlate with toxicity IC_{50}; Distribution of Spearman correlation coefficients for all descriptors that correlated with EPI of molecular jackhammers; Selected important features that correlate with EPI; The relationship between ATSC7d of sigma electrons and ATSC7d of valence electrons. (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Anatoly B. Kolomeisky − *Department of Chemistry, Center for Theoretical Biological Physics, Department of Chemistry, and Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0001-5677-6690; Email: tolya@rice.edu

James M. Tour − *Department of Chemistry, Department of Physics and Astronomy, Department of Materials Science and NanoEngineering, Smalley-Curl Institute, and Rice Advanced*

https://doi.org/10.1021/acs.jcim.4c00806
J. Chem. Inf. Model. 2024, 64, 5570−5579

Materials Institute, Rice University, Houston, Texas 77005, United States; orcid.org/0000-0002-8479-9328; Email: tour@rice.edu

Authors

Ciceron Ayala-Orozco − *Department of Chemistry, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0002-2574-0860

Hamid Teimouri − *Department of Chemistry and Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States*

Angela Medvedeva − *Department of Chemistry and Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States*

Bowen Li − *Department of Chemistry, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0003-4359-1712

Alex Lathem − *Department of Chemistry, Rice University, Houston, Texas 77005, United States*

Gang Li − *Department of Chemistry, Rice University, Houston, Texas 77005, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.4c00806

Author Contributions

Ciceron Ayala-Orozco and Hamid Teimouri contributed equally to this work. C.A.O, H.T., A.M., A.B.K., and J.M.T. designed the research. C.A.O, B.L., A.L., and G.L. performed experiments. H.T. and A.M. performed data analysis. C.A.O, H.T., A.M., A.B.K., and J.M.T. wrote the article. All authors reviewed the article.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.B.K. acknowledges the support from the Welch Foundation (C-1559), the NIH (R01 HL157714-02 and R01 GM 148537-02), the NSF (CHE-2246878), and the Center for Theoretical Biological Physics sponsored by the NSF (PHY-2019745). J.M.T. acknowledges support from the Discovery Institute and the Welch Foundation (C-2017-20220330).

■ REFERENCES

(1) Kocakaya, S.; Beier, C. P.; Beier, D. Chemotherapy increases long-term survival in patients with adult medulloblastomaa literature-based meta-analysis. *Neuro-oncology* **2016**, *18*, 408−416.

(2) Kopetz, S.; Chang, G. J.; Overman, M. J.; Eng, C.; Sargent, D. J.; Larson, D. W.; Grothey, A.; Vauthey, J.-N.; Nagorney, D. M.; McWilliams, R. R. Improved survival in metastatic colorectal cancer is associated with adoption of hepatic resection and improved chemotherapy. *Journal of clinical oncology* **2009**, *27*, 3677.

(3) Scher, H. I.; Fizazi, K.; Saad, F.; Taplin, M.-E.; Sternberg, C. N.; Miller, K.; De Wit, R.; Mulders, P.; Chi, K. N.; Shore, N. D.; et al. Increased survival with enzalutamide in prostate cancer after chemotherapy. *New England Journal of Medicine* **2012**, *367*, 1187−1197.

(4) Koshy, M.; Malik, R.; Weichselbaum, R. R.; Sher, D. J. Increasing radiation therapy dose is associated with improved survival in patients undergoing stereotactic body radiation therapy for stage I non-small-cell lung cancer. *International Journal of Radiation Oncology Biology Physics* **2015**, *91*, 344−350.

(5) Shridhar, R.; Almhanna, K.; Hoffe, S. E.; Fulp, W.; Weber, J.; Chuong, M. D.; Meredith, K. L. Increased survival associated with surgery and radiation therapy in metastatic gastric cancer: a Surveillance, Epidemiology, and End Results database analysis. *Cancer* **2013**, *119*, 1636−1642.

(6) Bentzen, S. M. Preventing or reducing late side effects of radiation therapy: radiobiology meets molecular pathology. *Nature Reviews Cancer* **2006**, *6*, 702−713.

(7) Deng, X.; Shao, Z.; Zhao, Y. Solutions to the drawbacks of photothermal and photodynamic cancer therapy. *Advanced Science* **2021**, *8*, 2002504.

(8) Jensen, K. E.; Soril, L. J.; Stelfox, H. T.; Clement, F. M.; Lin, Y.; Marshall, D. A. Side effects associated with the use of intensity-modulated radiation therapy in breast cancer patients undergoing adjuvant radiation therapy: a systematic review and meta-analysis. *Journal of medical imaging and radiation sciences* **2017**, *48*, 402−413.

(9) Love, R. R.; Leventhal, H.; Easterling, D. V.; Nerenz, D. R. Side effects and emotional distress during cancer chemotherapy. *Cancer* **1989**, *63*, 604−612.

(10) Monsuez, J.-J.; Charniot, J.-C.; Vignat, N.; Artigou, J.-Y. Cardiac side-effects of cancer chemotherapy. *International journal of cardiology* **2010**, *144*, 3−15.

(11) van den Boogaard, W. M.; Komninos, D. S.; Vermeij, W. P. Chemotherapy side-effects: not all DNA damage is equal. *Cancers* **2022**, *14*, 627.

(12) Mustroph, H.; Towns, A. Fine Structure in Electronic Spectra of Cyanine Dyes: Are Sub-Bands Largely Determined by a Dominant Vibration or a Collection of Singly Excited Vibrations? *ChemPhysChem* **2018**, *19*, 1016−1023.

(13) Ayala-Orozco, C.; Galvez-Aranda, D.; Corona, A.; Seminario, J. M.; Rangel, R.; Myers, J. N.; Tour, J. M. Molecular jackhammers eradicate cancer cells by vibronic-driven action. *Nat. Chem.* **2024**, *16*, 456−465.

(14) Cui, Y.; Lauchner, A.; Manjavacas, A.; García De Abajo, F. J.; Halas, N. J.; Nordlander, P. Molecular Plasmon-phonon Coupling. *Nano Lett.* **2016**, *16*, 6390−6395.

(15) Weissleder, R. A clearer vision for in vivo imaging. *Nature biotechnology* **2001**, *19*, 316−317.

(16) Ayala-Orozco, C.; Li, G.; Li, B.; Vardanyan, V.; Kolomeisky, A. B.; Tour, J. M. How to Build Plasmon-Driven Molecular Jackhammers that Disassemble Cell Membranes and Cytoskeletons in Cancer. *Adv. Mater.* **2024**, *36*, 2309910.

(17) Feoktistova, M.; Geserick, P.; Leverkus, M. Crystal Violet Assay for Determining Viability of Cultured Cells. *Cold Spring Harb Protoc.* **2016**, *4*, 343−346.

(18) Todeschini, R.; Consonni, V. *Molecular descriptors for chemoinformatics: volume I: alphabetical listing/volume II: appendices, references*; John Wiley & Sons, 2009.

(19) Mamada, H.; Takahashi, M.; Ogino, M.; Nomura, Y.; Uesawa, Y. Predictive Models Based on Molecular Images and Molecular Descriptors for Drug Screening. *ACS omega* **2023**, *8*, 37186−37195.

(20) Rodríguez-Pérez, R.; Miljković, F.; Bajorath, J. Machine Learning in Chemoinformatics and Medicinal Chemistry. Annual review of biomedical data science. *Annual review of biomedical data science* **2022**, *5*, 43−65.

(21) Mitchell, J. B. Machine learning methods in chemoinformatics. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014**, *4*, 468−481.

(22) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: a molecular descriptor calculator. *Journal of cheminformatics* **2018**, *10*, 1−14.

(23) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences* **1988**, *28*, 31−36.

(24) Saracco, B. H. Data Science and Predictive Analytics: Biomedical and Health Applications Using R. *Journal of the Medical Library Association* **2020**, *108*, 334−334.

(25) Safizadeh, H.; Simpkins, S. W.; Nelson, J.; Li, S. C.; Piotrowski, J. S.; Yoshimura, M.; Yashiroda, Y.; Hirano, H.; Osada, H.; Yoshida, M.; et al. Improving Measures of Chemical Structural Similarity Using Machine Learning on Chemical-Genetic Interactions. *J. Chem. Inf. Model.* **2021**, *61*, 4156−4172.

(26) Liew, S. K.; Malagobadan, S.; Arshad, N. M.; Nagoor, N. H. A review of the structure-activity relationship of natural and synthetic antimetastatic compounds. *Biomolecules* **2020**, *10*, 138.
(27) Chavan, S.; Nicholls, I. A.; Karlsson, B. C.; Rosengren, A. M.; Ballabio, D.; Consonni, V.; Todeschini, R. Towards Global QSAR Model Building for Acute Toxicity: Munro Database Case Study. *Int. J. Mol. Sci.* **2014**, *15*, 18162−18174.
(28) Lei, T.; Li, Y.; Song, Y.; Li, D.; Sun, H.; Hou, T. ADMET evaluation in drug discovery: 15. Accurate prediction of rat oral acute toxicity using relevance vector machine and consensus modeling. *Journal of Cheminformatics* **2016**, *8*, 6.
(29) Xue, W.; Zapien, D.; Warshawsky, D. Ionization potentials and metabolic activations of carbazole and acridine derivatives. *Chem. Res. Toxicol.* **1999**, *12*, 1234−1239.
(30) Sanderson, R. T. Electronegativity and bond energy. *J. Am. Chem. Soc.* **1983**, *105*, 2259−2261.
(31) Allred, A. L.; Rochow, E. G. A scale of electronegativity based on electrostatic force. *Journal of Inorganic and Nuclear Chemistry* **1958**, *5*, 264−268.
(32) James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An introduction to statistical learning*; Springer, 2013; Vol. *112*.
(33) Arlot, S.; Celisse, A. A survey of cross-validation procedures for model selection. *Stat. Surv* **2010**, *4*, 40−79.

https://doi.org/10.1021/acs.jcim.4c00806
J. Chem. Inf. Model. 2024, 64, 5570−5579