

The Applicability of Molecular Descriptors for Designing an Electrospray Ionization Mass Spectrometry Compatible Library for Drug Discovery

Jennifer C. Copeland, Levi J. Zehr, Ronald L. Cerny and Robert Powers*

Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588-0304, USA

Abstract: Detecting a small molecular-weight compound by electrospray ionization mass spectrometry (ESI-MS) requires the compound to obtain a charge. Factors such as gas-phase proton affinities and analyte surface activity are correlated with a positive ESI-MS response, but unfortunately it is extremely challenging to predict from a chemical structure alone if a compound is likely to yield an observable molecular-ion peak in an ESI-MS spectrum. Thus, the design of a chemical library for an ESI-MS ligand-affinity screen is particularly daunting. Only 56.9% of the compounds from our FAST-NMR functional library [1] were detectable by ESI-MS. An analysis of ~1,600 molecular descriptors did not identify any correlation with a positive ESI-MS response that cannot be attributed to a skewed population distribution. Unfortunately, our results suggest that molecular descriptors are not a valuable approach for designing a chemical library for an MS-based ligand affinity screen.

Keywords: Chemical library, electrospray ionization, mass spectrometry, molecular descriptors, ligand-affinity screens, drug discovery.

INTRODUCTION

Designing a biologically relevant chemical library is an area of extensive interest in small molecule drug discovery [2-6]. High-throughput screening (HTS) assays routinely evaluate 10,000 to $\geq 100,000$ compounds per day for evidence of biological activity [7]. Pharmaceutical companies have developed and acquired large chemical libraries that may exceed a million compounds [8]. Optimizing these chemical libraries is expected to improve the success rate and efficiency of HTS, where only 100,000 to 400,000 compounds are routinely used per screen. In general, chemical diversity and “drug-like” characteristics are the primary features used to design an HTS library. Lipinski’s rule of five [9], which relates solubility, partition coefficient (logP), molecular weight, and the number of hydrogen bond donors and acceptors with known drug activity, is a standard approach to eliminate undesirable compounds from a chemical library [10]. Nevertheless, computational parameters used to characterize drug-like chemical properties continue to evolve [11], where there are additional concerns of being too restrictive to novel compound classes [6]. Alternatively, virtual screens are routinely used to predict likely active compounds to design a chemical library for HTS [12-14].

Filtering a chemical library based on drug-like characteristics or predicted biological activity enhances the probability that HTS chemical leads can be evolved into drug candidates. Similarly, designing a library that increases the structural diversity of the compounds within the library is expected to improve the efficiency of the assay [15-17]. An HTS assay has a higher likelihood of identifying novel chemical leads, if the chemical library has been designed to

maximize the coverage of chemical space within the practical constraints of library size [18, 19]. By simply removing chemically similar structures, more time and resources are made available to assay other unique compounds. Fragment-based chemical libraries are a recent approach to maximize structural diversity while simultaneously minimizing the size of the library [20-23]. Basically, a fragment-based library is composed of small-molecular weight compounds (≤ 300 Da) that are fragments of known drugs or biologically active molecules [1]. Since fragment-based chemical libraries are significantly smaller (~1,000 compounds) than traditional HTS libraries, it is common to use NMR [24-27] or mass spectrometry [28-34] ligand-affinity screens to identify chemical leads. While drug-like properties and structural diversity are important design considerations for NMR and MS fragment-based chemical libraries, other practical issues such as aqueous solubility, chemical stability, availability and cost also contribute to compound selection.

For electrospray ionization mass spectrometry (ESI-MS), a further consideration in library design is the ability of the compound to generate a detectable molecular species. Given a large, structurally diverse compound library and a constant ESI-MS screening protocol, it is a challenging premise to predict, *a priori*, which compounds in the library will yield a detectable ESI-MS signal. The ESI process is fundamentally dependent on the analyte’s ability to obtain a charge, which is related to factors such as: solution ionic state, adduct formation, gas-phase proton affinities and analyte surface activity [35]. ESI-MS response has been correlated with nonpolar surface area [36, 37], Gibbs free energy of transfer from nonpolar to polar solutions [37], and HPLC retention times [38]. Other factors such as basicity, acidity, salt concentration, and calculated logP have also been associated with response patterns in ESI-MS [39-41]. Nevertheless, these factors are only modestly beneficial to the design of a chemical library for an ESI-MS based ligand affinity screen. A desirable alternative would be a correlation between

*Address correspondence to this author at the Department of Chemistry, 722 Hamilton Hall, University of Nebraska-Lincoln, Lincoln, NE 68588-0304 USA; Tel: (402) 472-3039; Fax (402) 472-9402; E-mail: rpowers3@unl.edu

molecular descriptors and an ESI-MS response. Molecular descriptors are readily calculated from the chemical structure and provide an abundance of chemical and physical information about the compounds [42-48]. Molecular descriptors are generally correlated with structure-activity relationships, which are expected to include a compound's ability to obtain a charge.

Our chemical library of biologically relevant compounds [1] was designed for our FAST-NMR assay [49, 50]. The library is structurally and chemically diverse. We have evaluated the utility of this library for an affinity chromatography ESI-MS screen with the goal of improving the library design for MS. While 95% of the FAST-NMR library was (at least partially) soluble in an MS buffer, only 56.9% of the compounds yielded an effective ESI response. In an effort to better understand factors that may lead to a desirable MS signal, ~1,600 molecular descriptors for each compound from the FAST-NMR library was obtained from the online software E-Dragon [48, 51-53]. The molecular descriptors were used to identify relationships between an ESI-MS response and chemical structures. Any molecular descriptors that correlate with an observable ESI-MS response would beneficially contribute to the future design of chemical libraries for ESI-MS ligand-affinity screens.

METHODS AND MATERIALS

All solvents and chemicals were obtained from Sigma-Aldrich (St. Louis, MO) and used without further purification. Water was filtered and purified with a Barnstead NANOpure filtration system (Thermo Fisher Scientific Inc., Asheville, NC). All mass spectra were collected by ESI-MS with a Finnigan LCQ using a Pauli ion trap (Thermo Fisher Scientific Inc., West Palm Beach, FL). Prior to spectral collection, the LCQ was tuned with 50 μ M caffeine in acetonitrile/water (1:1 v/v).

Sample Preparation

10 mM stock solutions were made by dissolving the compounds in a 1:1 v/v 25 mM ammonium acetate buffer (pH 7.0)/acetonitrile solution. If required, heat and sonication were applied until the sample completely dissolved. 10 mM stock samples in sealed 1.5 mL sample tubes were partially submerged in a Branson 3510 ultrasonic sonicator (Danbury, CT) at 50° C for up to 30 min. Samples that didn't dissolve after 30 min were considered insoluble. The chemical compounds were then diluted to a final 50 μ M concentration in the ammonium acetate-acetonitrile solution and stored in sterile 15 mL polypropylene Falcon conical tubes (BD Biosciences, Mississauga, ON Canada). Effects of pH were tested by successively adding 100 μ L aliquots of formic acid to determine if changing the pH had an effect on ESI response.

Mass Spectrometry

Each compound was injected into the ESI-MS using a 250 μ L Hamilton glass syringe at a rate of 5-15 μ L/min. The scan range was dependent on the occasional appearance of larger contaminant peaks suppressing the peak of interest's intensity. When the peak of interest had a large relative intensity, the scan was started, and continued for sixty

seconds. The spray voltage was set to 4.00 kV, and the voltage applied to the heated capillary ranged from 33 to 42 V. These fluctuations were a result of tuning the LCQ to obtain the optimum caffeine standard response each day before data collection. Data were collected for one minute at 30 scans per minute with 4 microscans per scan.

Buffers and Solubility

Several buffers were tested to determine optimal solubility and ESI-MS compatibility. Optimization of the FAST-NMR library is a prerequisite for a planned affinity chromatography ESI-MS ligand-affinity assay using human serum albumin (HSA). So a buffer must be compatible with the FAST-NMR library, the mass spectrometer and protein stability in order to obtain reliable MS reference spectra. An aqueous ammonium acetate buffer was chosen because of its general utility with protein samples [54]. Ammonium acetate is also a volatile buffer, making it an acceptable solvent for mass spectrometry. Dimethyl sulfoxide (DMSO) is routinely used to solubilize chemical libraries for water-based assays [55], but is not compatible with polyetheretherketone (PEEK) HPLC tubing. DMSO causes PEEK tubing to swell and would cause inconsistencies in the HPLC performance. Instead, methanol was considered for its organic solvent properties. Similarly, acetonitrile was chosen since it is commonly used in HPLC for protein purification. Several mobile phase mixture combinations (with varying amounts of methanol, acetonitrile, water and ammonium acetate) were explored to maximize the ESI-MS response for the largest number of compounds. The mixture ratio corresponds to the post-column solvent composition, just prior to entering the mass spectrometer. These will be adjusted from the 100% ammonium acetate buffer with a post-column, low-pressure pump with an 80/20 mixture of acetonitrile and methanol, making the resultant mobile phase for the mass spectrometer 50/40/10 ammonium acetate/ acetonitrile/ methanol, to coordinate with the buffer conditions of the reference spectra.

Data Analysis

Each compound from the FAST-NMR chemical library was classified based on the observed solubility, ESI-MS response and chemical class. The compounds were assigned to a chemical class by manual inspection of each compound's chemical structure.

A compound's solubility was manually ranked on a scale of 1 to 0 based on the amount of time the compound required to completely dissolve under various experimental conditions. Compounds that dissolved at room temperature with only sample vortexing were given the best score of 1. Compounds that required 5, 10, and 30 minutes of sonication at 50° C were assigned a score of 0.8, 0.6, and 0.4 respectively. Compounds that did not appear to completely dissolve, but still generated an ESI-MS response were assigned a score of 0.2. Compounds that did not dissolve and lacked an ESI-MS response were assigned a score of 0.0.

Since observing a detectable signal in a MS spectrum is the primary requirement for a MS-based ligand affinity screen, the ESI-MS response was ranked using a simple binary scale. The ESI-MS response was determined by

simply recording the absolute intensity of the most intense peaks that correspond to the correct mass of either the parent compound or an obvious fragment based on a rapid structural analysis. If the apparent signal-to-noise ratio of the most intense molecular species peak is ≥ 2 , the compound was determined to be detectable by ESI-MS under our experimental conditions and was assigned a score of 1. All other compounds were given a score of 0.

The 355 compounds in the soluble library were clustered into thirty general chemical classes (Fig. 2). A histogram plot indicates the total number of compounds in each class, along with the percentage of compounds within the class that received an ESI-MS response score of 1.

The online software E-Dragon (VCClabs, <http://www.vcclab.org/lab/edragon/>) was used to calculate approximately 1,600 molecular descriptors for each soluble compound in the FAST-NMR library [48, 51-53]. The molecular descriptors were selected based on properties related to small molecule analysis. In order to facilitate the analysis of the E-Dragon molecular descriptors, the statistical program SIMCA-P (UMETRICS, Kinnelon, NJ) was used to identify molecular descriptors correlated with an ESI-MS response. A partial least squares (PLS) regression was calculated based on each compound's set of molecular descriptors and the ESI-MS response score. The loading plot from the PLS analysis identified the molecular descriptors with the largest contribution to the data variability (Fig. 1). Molecular descriptors with relatively large positive (upper right circle) and negative (lower left circle) PC1 and PC2 coefficients are

correlated or anti-correlated, respectively, with the experimentally observed ESI-MS response. The PLS loading plot suggests 59 molecular descriptors are correlated with ESI-MS response and 16 are anti-correlated.

Each molecular descriptor identified by the PLS loading plot was analyzed in detail. Simply, the compounds from the FAST-NMR library were binned based on the molecular descriptor score. The percentage of compounds in each bin that received an ESI-MS response score of 1 was calculated and the results plotted as a histogram using Microsoft Excel (Figs. 3-5). A histogram that displayed an increase in the percentage of compounds with a positive ESI-MS response as a function of the molecular descriptor score would further support a correlation between the molecular descriptor and an ESI-MS response. The process was repeated for a total of 75 molecular descriptors, where the bin sizes varied for each histogram and was dependent on an approximate equal distribution of the number of compounds per bin. Also, the bins were adjusted based on descriptor parameters, i.e. the bins were restricted to whole numbers if the descriptors were based on discrete whole numbers.

RESULTS AND DISCUSSION

Structural diversity and drug-like characteristics are typical issues that dominate discussions of library design, but practical considerations such as solubility, stability and ease of replenishing the compound are commonly the deciding factors for the inclusion of a specific compound in a library. Unfortunately, these factors are not readily predictable and

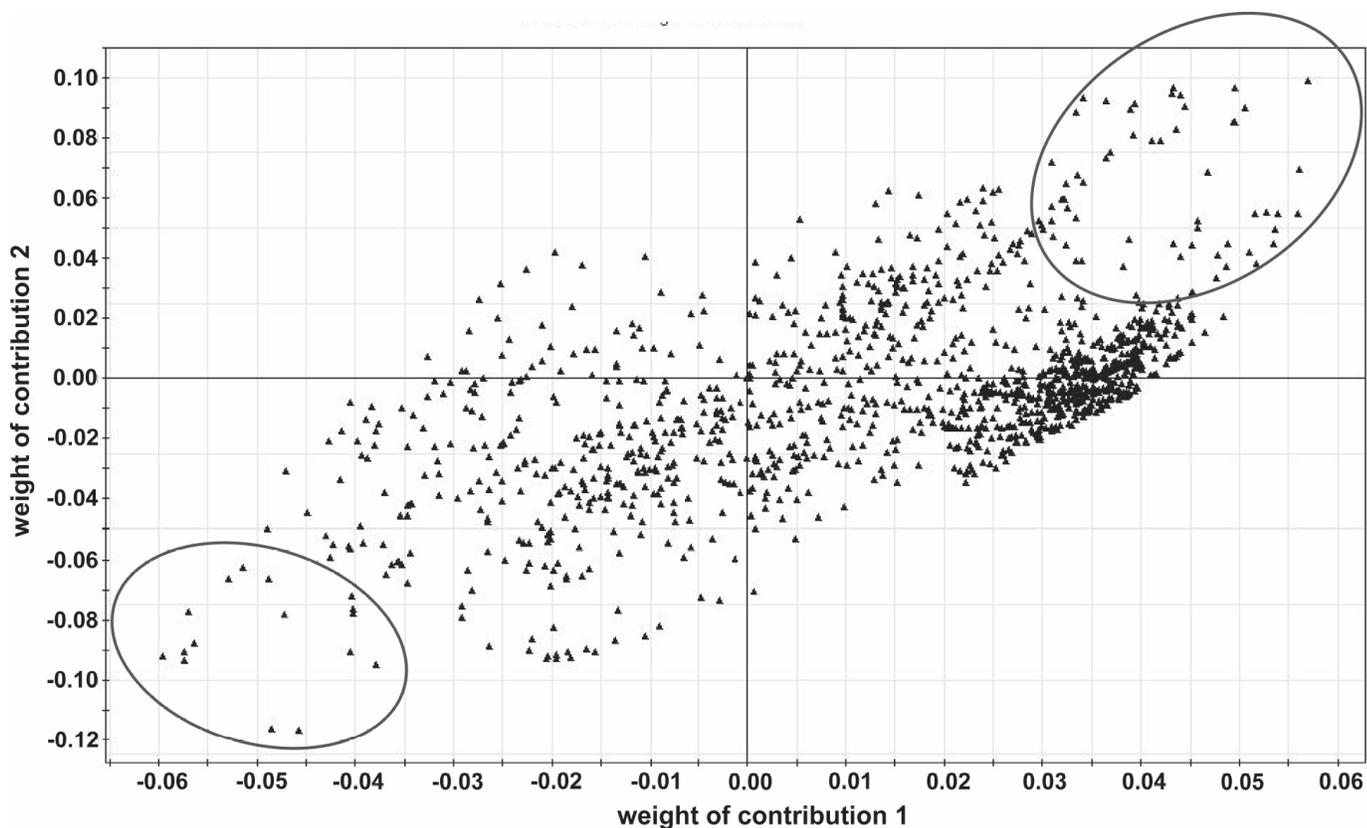


Fig. (1). PLS loading plot of ~1,600 molecular descriptors contribution to a detectable signal in an ESI-MS spectrum, where molecular descriptors at each axial extreme (circled) were chosen for further investigation.

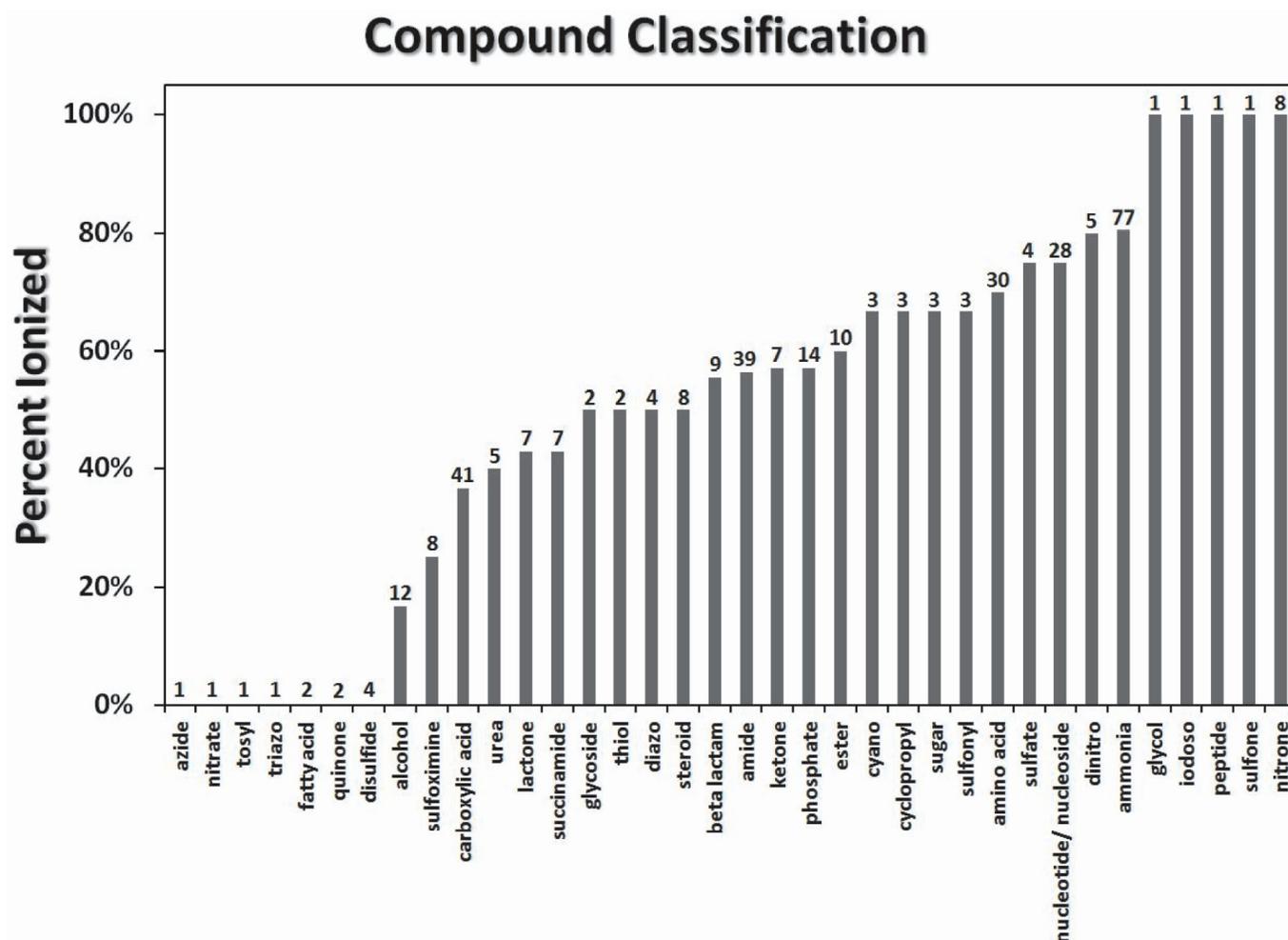


Fig. (2). The soluble compounds within the FAST-NMR chemical library were assigned to one of thirty different chemical classes. The histogram plot indicates the number of compounds within each chemical classification and the percentage of compounds within each class that yielded a detectable signal in an ESI-MS spectrum.

require experimental verification. MS-based screens have an additional unique concern in regards to library design: does a compound generate a detectable molecular species peak? The FAST-NMR library was selected based on functional diversity [1]. All the compounds in the library have known biological activity that involves distinct proteins or protein classes. The compounds correspond to co-factors, metabolites, substrates, carbohydrates, fatty acids, protein inhibitors, and known drugs, among other classes. Furthermore, the compounds were selected based on experimentally determined aqueous solubility (100 μ M), long term stability (> 6 months), cost (~ \$32/compound) and commercial availability. Nevertheless, attempting to use the FAST-NMR library for an affinity chromatography-ESI-MS ligand affinity screen proved to be an unexpected challenge.

While the FAST-NMR library is generally water soluble, using only an aqueous buffer in an ESI-MS ligand affinity screen is not feasible. Unlike NMR, the MS solvent must also be volatile. Common ESI-MS solvents, methanol, water, and acetonitrile (ACN), were used to test compound solubility. Only 76% of the FAST-NMR library was soluble in a solvent composed of an approximately 50:50:0.1% (v/v) mixture of methanol, water and ACN. Conversely, 95% of

the library was soluble in a 50/50 (v/v) mixture of ACN and 25 mM aqueous ammonium acetate buffer. After dissolving these soluble compounds (most of them from powder) into the 50/50 (v/v) ACN/25 mM aqueous ammonium acetate buffer, only about half of the compounds had a discernible ESI-MS response. 0.1% (v/v) of formic acid was added to all samples that failed to generate a molecular species peak in the ESI-MS. An ESI-MS spectrum was then re-acquired in the positive-ion mode, but only a slight increase in a detectable response was achieved. In all, only 56.9% of the soluble compounds, or 202 compounds in total, produced an observable ESI-MS spectrum.

Consistent with our results, there does not appear to be any correlation between the pKa of a compound and the pH of the solution [35]. Instead, obtaining a charge has been postulated to depend on the partitioning of the compound to the droplet surface [56] and the relative pKa of the compound and the solvent [35]. While these details are valuable for understanding the process of an analyte obtaining a charge by ESI, applying this information to assist in the design of an MS library is not particularly straightforward without experimental data. As an illustration, psoralen, leflunomide, carisoprodol, estriol, p-aminohippuric

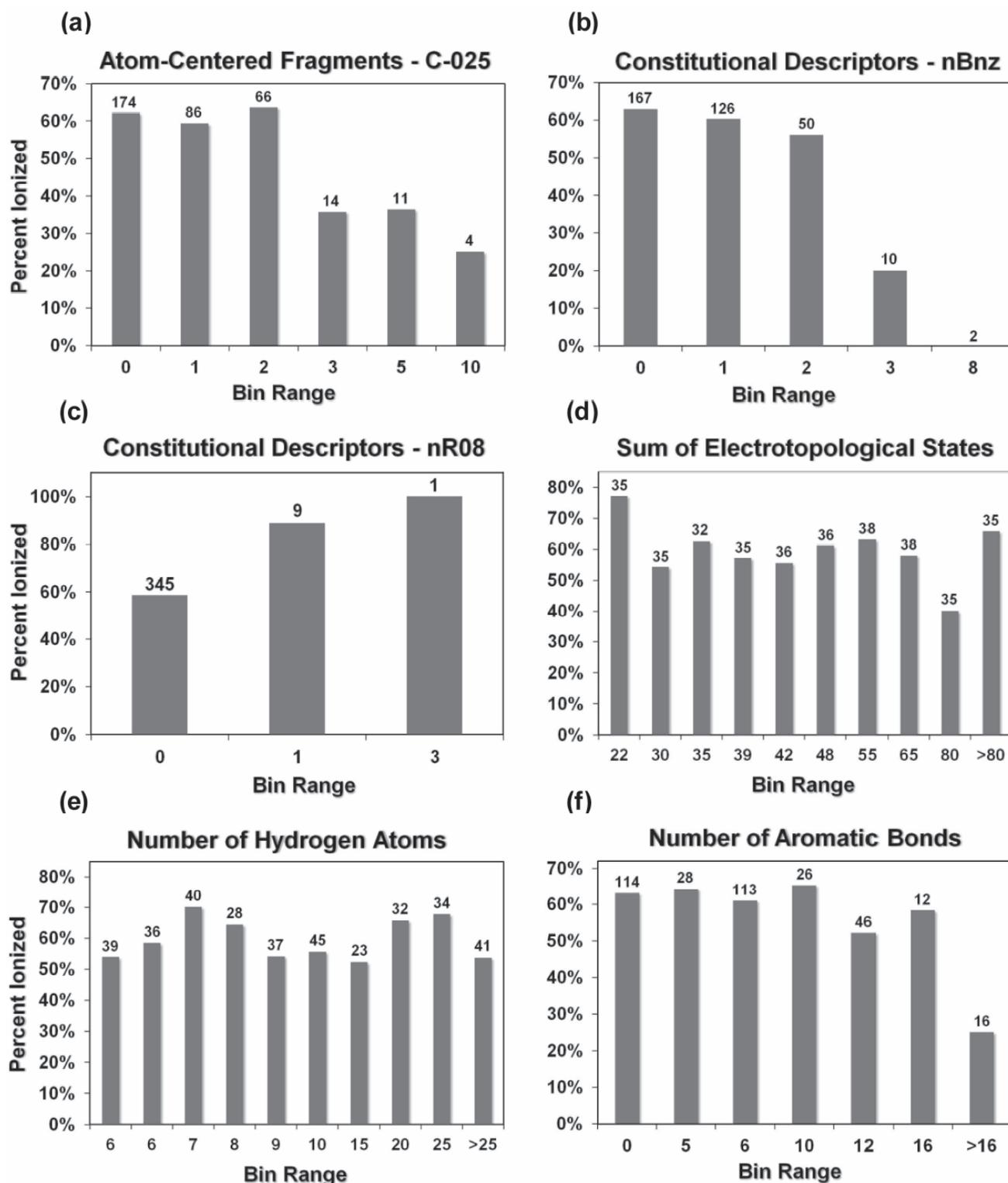


Fig. (3) Histogram plots where the x-axis corresponds to the bin range for the E-Dragon molecular descriptor and the y-axis is the percentage of compounds with a detectable signal in an ESI-MS spectrum within each bin. Each bar is labeled with the total number of compounds within the bin. Molecular descriptors corresponding to (a) the number of R--CR--R atom-centered fragments per bin, (b) the number of benzene-like rings (nBnz), (c) the number of 8-membered rings (nR08), (d) the Kier-Hall electrotopological state of the compound, (e) the number of Hydrogen atoms, and (f) the number of aromatic rings per compound.

acid, 1-octanol, ethylenediaminetetraacetic acid (EDTA), flutamide, and 1-phenyl-1-cyclopropanecarboxylic acid were all initially unresponsive to ESI-MS and were titrated with 0.1% formic acid. But, only psoralen, p-aminohippuric acid,

and EDTA generated a positive ESI-MS response after the addition of formic acid (Table 1). Unfortunately, there are no obvious structural features among this short-list of compounds that readily explains the variable ESI-response.

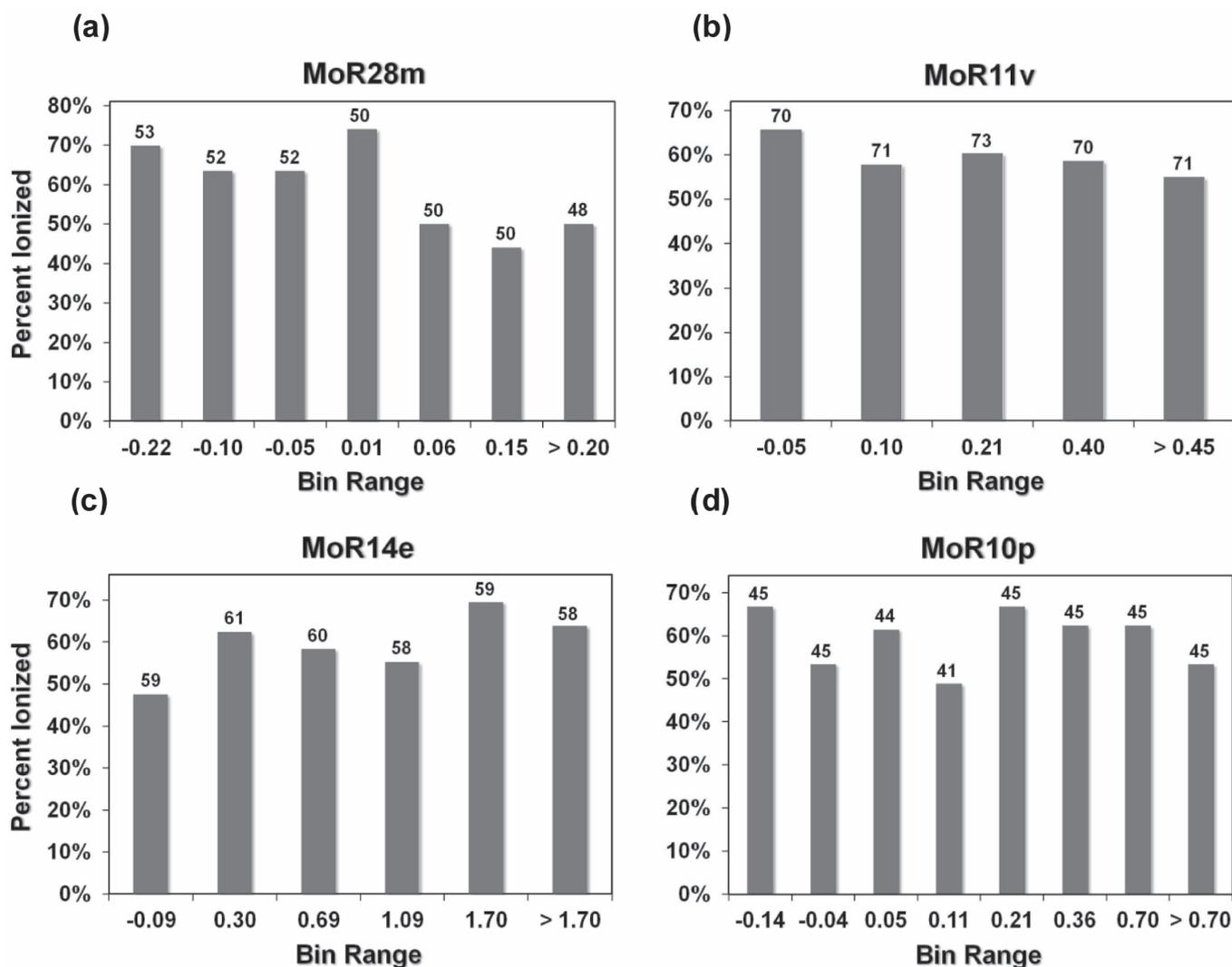


Fig (4). Histogram plots of 3D Molecular Representation of Structures based on Electron diffraction (MoRSE) molecular descriptors. The x-axis is the empirical MoRSE values calculated by E-Dragon. The y-axis corresponds to the percentage of compounds with a detectable signal in an ESI-MS spectrum. Each bar is labeled with the total number of compounds within the bin. Molecular descriptors corresponding to (a) MoRSE value at signal 28 weighted by mass (Mor26m), (b) MoRSE descriptor weighted by van der Waals volume from signal 11 (Mor11v), (c) MoRSE descriptor weighted by Sanderson electronegativities (Mor14e), and (d) a MoRSE descriptor weighted by polarizability (Mor10p).

Conversely, when all the soluble compounds from the FAST-NMR library were manually classified into general chemical classes, some distinct correlations with an ESI-MS response were observed (Fig. 1). Several of these chemical categories were found to have a high probability of an ESI-MS response, while other classes had no compounds that generated a detectable peak in the MS spectrum. Some chemical categories with a positive ESI-MS response included: ketones, nitrones, sulfones and esters. The double bonded oxygen in this chemical class can form resonance structures that enable the compound to accept a charge. Nucleic acid derivatives and sugars are also adequate responders due to the flexible ring nature of the sugar molecule. Amphoteric molecules, such as amino acids prove to be good responders as well. Some compound types that gave poor response are carboxylic acids, alcohols, lactones and urea-containing molecules. These trends between chemical classification and ESI-MS response are beneficial for designing an MS screening library. But, there are

important caveats to the data since the apparent high or low ESI-MS response for some classes may be misleading because of the low number of members. A 100% or 0% percent response for a chemical class that has only a few compounds is not particularly meaningful.

Alternatively, molecular descriptors provide a more robust and unbiased approach to establish a correlation between chemical structures and an ESI-MS response. E-Dragon [48, 51-53] was used to calculate ~1,600 molecular descriptors for each soluble compound in the FAST-NMR library. A PLS analysis was used to identify a relationship between any of the individual molecular descriptors and each compound's ESI-MS response. The PLS loading plot (Fig. 1) identified the molecular descriptors that made the largest contribution to principle components and preferentially explain the variation in the data. The two circles on the PLS loading plot highlight the molecular descriptors that are either highly correlated (positive coefficients) or anti-

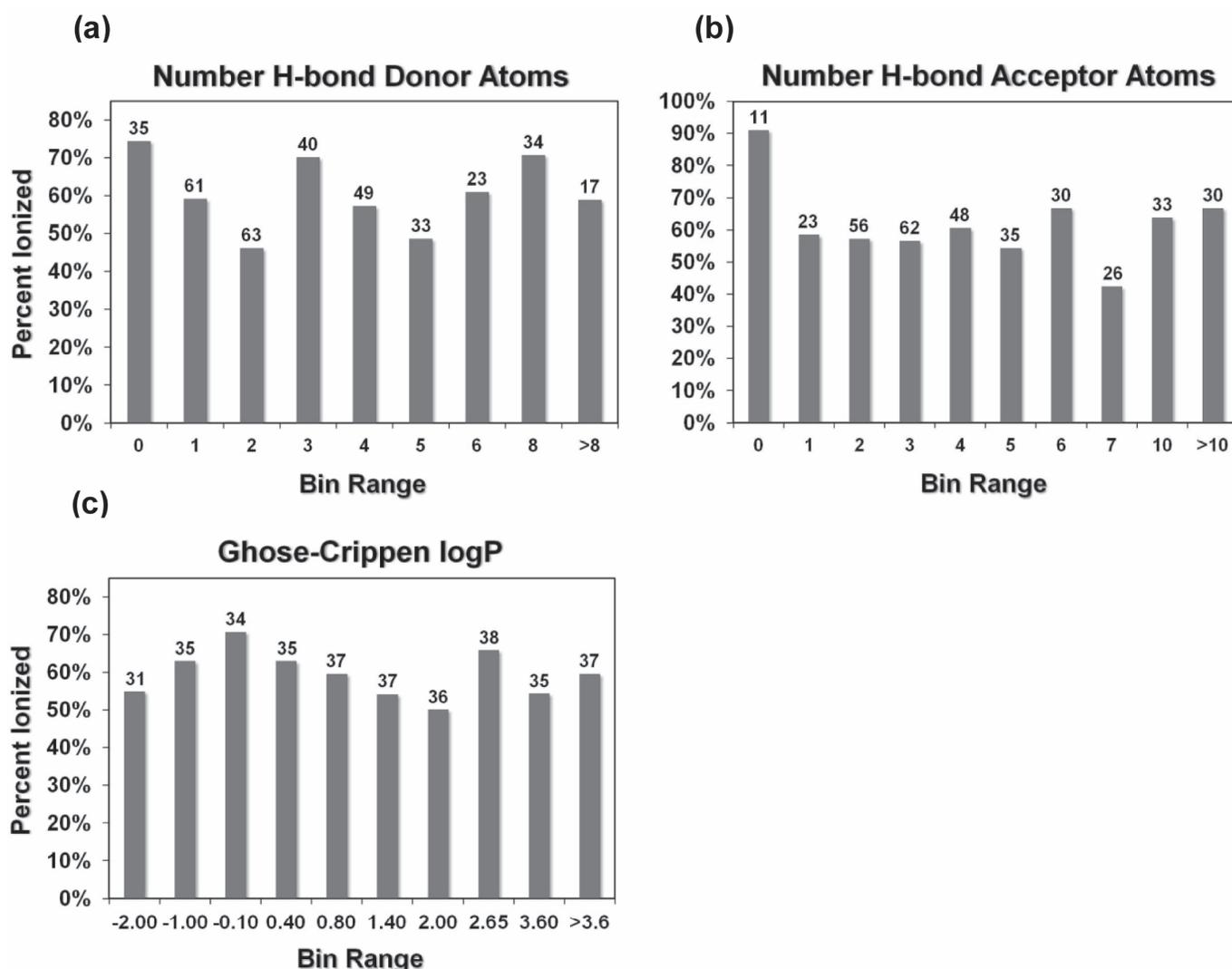


Fig (5). Histogram plots where the x-axis corresponds to the bin range for the E-Dragon molecular descriptor and the y-axis is the percentage of compounds with a detectable signal in an ESI-MS spectrum within each bin. Each bar is labeled with the total number of compounds within the bin. Molecular descriptors corresponding to (a) the number of hydrogen-bond donor atoms per bin, (b) the number of hydrogen-bond acceptor atoms per bin, and (c) the Ghose-Crippen value per bin [64].

correlated (negative coefficients) with an ESI-MS response. A total of 75 molecular descriptors were identified for further analysis.

Histograms (Figs. 3-5) were generated for the molecular descriptors that were obvious outliers from the PLS-DA scatter plot in order to identify a correlated trend. The percentage of compounds yielding a positive ESI-MS response in each bin was plotted against the molecular descriptor score (in bins). Trends were identified if the percentage of compounds yielding an ESI-MS response increased or decreased uniformly as a function of the molecular descriptor score. For example, as the number of benzene-like rings increases from 0 to 8, the percentage of compounds with a positive ESI-MS response decreases from 61 to 0 percent (Fig. 3b). To be clear, a compound that contains zero to two benzene-like rings had a 56.0% likelihood of yielding a positive ESI-MS response. Conversely, the likelihood that a compound would have a detectable ESI-MS molecular ion peak dropped to less than 20% for compounds with three or more rings. While this

does indicate that adding more benzene-like rings to a compound would decrease the likeliness of ionization in ESI, the lack of these rings provides only a slightly higher-than-average rate of ionization (62.9%) despite the depolarizing nature of a benzene-like system. Also, the perceived difference may be skewed by the relatively low population of structures with three or more benzene-like rings.

Based on prior correlations between ESI-MS responses with nonpolar surface area [36, 37], Gibbs free energy of transfer from nonpolar to polar solutions [37], basicity, acidity, and calculated logP [39-41]; a trend would be expected to be found for certain molecular descriptors such as Morse values, mean electrotopological state, logP and the number of hydrogen bond donating or accepting atoms. MoRSE descriptors are part of the 3D-MoRSE (Molecule Representation of Structures based on Electron diffraction) class of descriptors designed to represent a chemical structure for QSAR calculations [57]. Basically, the intensity of scattered radiation is calculated from the three-dimensional atomic coordinates of a molecule for 32

Table 1. ESI-MS Response After Adding Formic Acid to the Ionization Buffer

Name	Structure	pKa	ESI-MS Response	Name	Structure	pKa	ESI-MS Response
EDTA		2.0, 2.7, 6.2, 10.3	Yes	estriol		10.4	No
psoralen			Yes	flutamide		13.1	No
<i>p</i> -aminohippuric acid		3.6	Yes	leflunomide		10.8	No
octanol			No	carisoprodol		4.2	No
1-phenyl-1-cyclopropane carboxylic acid		4.3	No				

different scattering angles. The intensity can be unweighted or weighted by atomic masses, atomic van der Waals volumes, electronegativities or atomic polarizabilities. The goal of 3D-MoRSE is to transform the molecular structure into a set of variables that maintain the structure-property relationships. As a result, 3D-MoRSE molecular descriptors have been previously shown to be correlated with the structural details of infrared spectra [57] and mass spectra [58]. Thus, a correlation with an ESI-MS response is a reasonable expectation. Unfortunately, no apparent trend was observed between a detectable ESI-MS molecular-ion peak and the 3D-MoRSE molecular descriptors (Fig. 4).

The mean electrotopological state molecular descriptor calculated by E-Dragon is the average of the Kier-Hall electrotopological state for each non-hydrogen atom in the structure [59]. Briefly, the electrotopological state describes the electronic and topological characteristics for each atom in a structure and the electronic influence by all the other atoms in the molecule on that atom. Each atom type is assigned an intrinsic state based on Kier-Hall electronegativity that is modified by the number of σ bonds, number of hydrogen atoms, number of electrons in π orbitals, and number of lone pair electrons. The intrinsic value (I_i) is large for electronegative atoms and decreases with increasing number of σ bonds. The electrotopological state (S_i) is calculated by summing the difference between

each pair of intrinsic values weighted by the distance ($r_{i,j}$) between the atoms:

$$S_i = I_i + \sum_{i,j} (I_i - I_j) / r_{i,j}^2 \quad (1)$$

The electrotopological state has been correlated with ^{17}O and ^{13}C NMR spectra [60, 61] consistent with capturing the intrinsic electronegativity of an atom within the context of a molecular structure. Thus, observing a correlation between an ESI-MS response, the ability of a molecule to accept a charge, and the overall electronegative characteristics of a compound is not unexpected, especially since positive results have been previously observed with polar and ionic analytes [62]. Recently, a correlation was observed between basic analytes and ESI-MS response [63]. Similarly, an ESI-MS response would be expected to be correlated with the number of hydrogen atoms (Fig. 3e), number of hydrogen bond donating or accepting atoms (Fig. 5a, b), logP (Fig. 5c) [64], or similar descriptors. Unfortunately a clear correlation between ESI-MS response and molecular descriptors related to electronegative characteristics was not observed (Fig. 3d). In fact, a convincing correlation between an ESI-MS response and any molecular descriptor was not observed (Figs. 3-5). For example, the atom-centered fragments and the constitutional descriptors, like the number of benzene-like rings descriptor, potentially exhibit a minor correlation (Fig. 3a-c). But, the population difference between the bins

are highly skewed, where the percent ionization in a low populated bin could change dramatically if the number of compounds increased by a factor of ten or more, to be on par with the highly populated bins. In effect, these perceived differences may simply be attributed to a skewed population distribution for a particular molecular descriptor. It is important to note, that for most of the populated bins, the percentage of ionized compounds was typically close to the overall average response rate of 56.9%. This result was independent of the molecular descriptor and suggests that any perceived correlation is not significant. Fundamentally, our analysis implies that any relationship between a molecular descriptor and an ESI-MS response is purely coincidental.

CONCLUSION

There are a number of inherent challenges in the creation and design of a diverse ESI-MS chemical library for drug discovery. Paramount to this process is the identification of compounds that are able to accept a charge and generate a detectable molecular species. Here, we examined the application of molecular descriptors, which are readily determined from chemical structures, as a means to predict the relative response in electrospray ionization mass spectrometry. We did not find a satisfactory correlation between calculated molecular descriptors and an ESI-MS response. Unfortunately, our results suggest that molecular descriptors are not an effective approach to design a chemical library for MS-based ligand affinity screens.

ACKNOWLEDGMENTS

Aid in classification of the compounds was provided by Dr. Charles Schiaffo, University of Minnesota, and we would like to thank him for his effort. This work was supported by Award Number R21AI081154 from the National Institute of Allergy and Infectious Diseases Nebraska, and by the Tobacco Settlement Biomedical Research Development Fund. Research was performed in facilities renovated with support from the NIH under Grant RR015468-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Allergy and Infectious Diseases.

CONFLICT OF INTEREST:

The author(s) confirm that this article content has no conflicts of interest.

REFERENCES

- [1] Mercier, K. A.; Germer, K.; Powers, R. Design and characterization of a functional library for NMR screening against novel protein targets. *Comb. Chem. High Throughput Screen.*, **2006**, *9* (7), 515-534.
- [2] Villar, H. O.; Hansen, M. R. Design of chemical libraries for screening. *Expert Opin. Drug Discovery*, **2009**, *4* (12), 1215-1220.
- [3] Egan, W. J.; Walters, W. P.; Murcko, M. A. Guiding molecules towards drug-likeness. *Curr. Opin. Drug Discovery Dev.*, **2002**, *5* (4), 540-549.
- [4] Matter, H.; Baringhaus, K.-H.; Naumann, T.; Klabunde, T.; Pirard, B. Computational approaches towards the rational design of drug-like compound libraries. *Comb. Chem. High Throughput Screen.*, **2001**, *4* (6), 453-475.

- [5] Mitchell, T.; Showell, G. A. Design strategies for building drug-like chemical libraries. *Curr. Opin. Drug Discovery Dev.*, **2001**, *4* (3), 314-318.
- [6] Walters, W. P.; Murcko, M. A. Prediction of 'drug-likeness'. *Adv. Drug Delivery Rev.*, **2002**, *54* (3), 255-271.
- [7] Hill, D. C. Trends in development of high-throughput screening technologies for rapid discovery of novel drugs. *Curr. Opin. Drug Discovery Dev.*, **1998**, *1* (1), 92-97.
- [8] Razvi, E. Emerging Themes and Quantitative Metrics in High-Throughput Screening Space. *Drug & Market Development*, **2003**, *14*.
- [9] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.*, **2001**, *46*(1-3), 3-26.
- [10] Chuprina, A.; Lukin, O.; Demoiseaux, R.; Buzko, A.; Shivanyuk, A. Drug- and Lead-likeness, Target Class, and Molecular Diversity Analysis of 7.9 Million Commercially Available Organic Compounds Provided by 29 Suppliers. *J. Chem. Inf. Model.*, **50** (4), 470-479.
- [11] Muresan, S.; Sadowski, J. Properties guiding drug- and lead-likeness. *Methods Princ. Med. Chem.*, **2008**, *37* (Molecular Drug Properties), 441-461.
- [12] Guido, R. V. C.; Oliva, G.; Andricopulo, A. D. Virtual screening and its integration with modern drug design technologies. *Curr. Med. Chem.*, **2008**, *15* (1), 37-46.
- [13] Koepfen, H. Virtual screening - what does it give us? *Curr. Opin. Drug Discovery Dev.*, **2009**, *12*(3), 397-407.
- [14] Pellicchia, M. Fragment-based drug discovery takes a virtual turn. *Nat. Chem. Biol.*, **2009**, *5* (5), 274-275.
- [15] Agrafiotis, D. K.; Lobanov, V. S.; Rassokhin, D. N.; Izrailev, S. The measurement of molecular diversity. *Methods Princ. Med. Chem.*, **2000**, *10* (Virtual Screening for Bioactive Molecules), 265-300.
- [16] Spellmeyer, D. C.; Grootenhuys, P. D. J. Recent developments in molecular diversity. Computational approaches to combinatorial chemistry. *Annu. Rep. Med. Chem.*, **1999**, *34*, 287-296.
- [17] Willett, P. Chemoinformatics - similarity and diversity in chemical libraries. *Curr. Opin. Biotechnol.*, **2000**, *11*(1), 85-88.
- [18] Medina-Franco, J. L.; Martinez-Mayorga, K.; Giulianotti, M. A.; Houghten, R. A.; Pinilla, C. Visualization of the chemical space in drug discovery. *Curr. Comput.-Aided Drug Des.*, **2008**, *4*(4), 322-333.
- [19] Trigg, D. J. The chemist as astronaut: Searching for biologically useful space in the chemical universe. *Biochem. Pharmacol.*, **2009**, *78*(3), 217-223.
- [20] Bailey, D.; Boyd, S.; England, P.; de Esch, I. Fragment-based drug design. *Innovations Pharm. Technol.*, **2008**, *25*, 18-22.
- [21] Brewer, M.; Ichihara, O.; Kirchoff, C.; Schade, M.; Whittaker, M., Assembling a fragment library. In *Fragment-Based Drug Discovery* ed.; Zartler, E. R.; Shapiro, M. J., John Wiley & Sons Ltd.: Chichester, UK, **2008**; pp 39-62.
- [22] Lepre, C. Fragment-based drug discovery using the SHAPES method. *Expert Opin. Drug Discovery*, **2007**, *2* (12), 1555-1566.
- [23] Schulz, M. N.; Hubbard, R. E. Recent progress in fragment-based lead discovery. *Curr. Opin. Pharmacol.*, **2009**, *9*(5), 615-621.
- [24] Powers, R. Advances in nuclear magnetic resonance for drug discovery. *Expert Opin. Drug Discovery*, **2009**, *4*(10), 1077-1098.
- [25] Dalvit, C. NMR methods in fragment screening: Theory and a comparison with other biophysical techniques. *Drug Discovery Today*, **2009**, *14* (21/22), 1051-1057.
- [26] Schultz, J., Practical aspects of using NMR in fragment-based screening. In *Fragment-Based Drug Discovery* ed.; Zartler, E. R.; Shapiro, M. J., John Wiley & Sons Ltd.: Chichester, UK, **2008**; pp 63-98.
- [27] Mercier, K. A.; Shortridge, M. D.; Powers, R. A multi-step NMR screen for the identification and evaluation of chemical leads for drug discovery. *Comb. Chem. High Throughput Screening*, **2009**, *12* (3), 285-295.
- [28] Powers, R.; Siegel, M. M., Applications of Nuclear Magnetic Resonance and Mass Spectrometry to Anticancer Drug Discovery. In *Novel Anticancer Agents: Strategies for Discovery and Clinical Testing*, First ed.; Adjei, A. A.; Buolamwini, J. K., Elsevier Science: New York, NY, **2006**; pp 107-190.
- [29] Moy, F. J.; Haraki, K.; Mobilio, D.; Walker, G.; Tabei, K.; Tong, H.; Siegel, M. M.; Powers, R. MS/NMR: A Structure-Based

- Approach for Discovering Protein Ligands and for Drug Design by Coupling Size Exclusion Chromatography, Mass Spectrometry, and Nuclear Magnetic Resonance Spectroscopy. *Anal. Chem.*, **2001**, *73* (3), 571-581.
- [30] Annis, D. A.; Nickbarg, E.; Yang, X.; Ziebell, M. R.; Whitehurst, C. E. Affinity selection-mass spectrometry screening techniques for small molecule drug discovery. *Curr. Opin. Chem. Biol.*, **2007**, *11* (5), 518-526.
- [31] Cloutier, T. E.; Comess, K. M. Library screening using ultrafiltration and mass spectrometry. *Methods Princ. Med. Chem.*, **2007**, *36* (Mass Spectrometry in Medicinal Chemistry), 157-183.
- [32] de Boer, A. R.; Lingeman, H.; Niessen, W. M. A.; Irth, H. Mass spectrometry-based biochemical assays for enzyme-inhibitor screening. *TrAC, Trends Anal. Chem.*, **2007**, *26* (9), 867-883.
- [33] Deng, G.; Sanyal, G. Applications of mass spectrometry in early stages of target based drug discovery. *J. Pharm. Biomed. Anal.*, **2006**, *40* (3), 528-538.
- [34] Liu, H. C.; Liu, R. H.; Ho, H. O.; Lin, D. L. Development of an information-rich LC-MS/MS database for the analysis of drugs in postmortem specimens. *Anal. Chem.*, **2009**, *81* (21), 9002-9011.
- [35] Cech, N. B.; Enke, C. G. Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom. Rev.*, **2002**, *20* (6), 362-387.
- [36] Karplus, P. A. Hydrophobicity regained. *Protein Sci.*, **1997**, *6* (6), 1302-1307.
- [37] Cech, N. B.; Enke, C. G. Relating Electrospray Ionization Response to Nonpolar Character of Small Peptides. *Anal. Chem.*, **2000**, *72* (13), 2717-2723.
- [38] Cech, N. B.; Krone, J. R.; Enke, C. G. Predicting electrospray response from chromatographic retention time. *Anal. Chem.*, **2001**, *73* (2), 208-213.
- [39] Constantopoulos, T. L.; Jackson, G. S.; Enke, C. G. Effects of salt concentration on analyte response using electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.*, **1999**, *10* (7), 625-634.
- [40] Ehrmann, B. M.; Henriksen, T.; Cech, N. B. Relative importance of basicity in the gas phase and in solution for determining selectivity in electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.*, **2008**, *19* (5), 719-728.
- [41] Henriksen, T.; Juhler, R. K.; Svensmark, B.; Cech, N. B. The relative influences of acidity and polarity on responsiveness of small organic molecules to analysis with negative ion electrospray ionization mass spectrometry (ESI-MS). *J. Am. Soc. Mass Spectrom.*, **2005**, *16* (4), 446-455.
- [42] Bravi, G.; Gancia, E.; Green, D. V. S.; Hann, M. M. Modelling structure-activity relationships. *Methods Princ. Med. Chem.*, **2000**, *10* (Virtual Screening for Bioactive Molecules), 81-116.
- [43] Helguera, A. M.; Combes, R. D.; Gonzalez, M. P.; Cordeiro, M. N. D. S. Applications of 2D descriptors in drug design: a DRAGON tale. *Curr. Top. Med. Chem. (Sharjah, United Arab Emirates)*, **2008**, *8* (18) 1628-1655.
- [44] Katritzky, A. R.; Fara, D. C.; Petrukhin, R. O.; Tatham, D. B.; Maran, U.; Lomaka, A.; Karelson, M. The present utility and future potential for medicinal chemistry of QSAR/QSPR with whole molecule descriptors. *Curr. Top. Med. Chem. (Hilversum, Neth.)*, **2002**, *2* (12), 1333-1356.
- [45] Prabhakar, Y. S.; Gupta, M. K. Chemical structure indices in in silico molecular design. *Sci. Pharm.*, **2008**, *76* (2), 101-132.
- [46] Sutter, J. M.; Jurs, P. C. Selection of molecular descriptors for quantitative structure-activity relationships. *Data Handl. Sci. Technol.*, **1995**, *15*, 111-132.
- [47] Yang, G.-F.; Huang, X. Development of quantitative structure-activity relationships and its application in rational drug design. *Curr. Pharm. Des.*, **2006**, *12* (35), 4601-4611.
- [48] Consonni, V.; Todeschini, R.; Pavan, M.; Gramatica, P. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. *J. Chem. Inf. Comput. Sci.*, **2002**, *42* (3), 693-705.
- [49] Powers, R.; Copeland, J.; Mercier, K. Application of FAST-NMR in Drug Discovery. *Drug Discovery Today*, **2008**, *13* (3-4), 172-179.
- [50] Mercier, K. A.; Baran, M.; Ramanathan, V.; Revesz, P.; Xiao, R.; Montelione, G. T.; Powers, R. FAST-NMR: Functional Annotation Screening Technology Using NMR Spectroscopy. *J. Amer. Chem. Soc.*, **2006**, *128* (47), 15292-15299.
- [51] Tetko, I. V. Computing chemistry on the web. *Drug Discov Today*, **2005**, *10* (22), 1497-1500.
- [52] Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V. Virtual computational chemistry laboratory--design and description. *J. Comput. Aided Mol. Des.*, **2005**, *19* (6), 453-463.
- [53] VCCLAB Virtual Computational Chemistry Laboratory. <http://www.vcclab.org>
- [54] Ugwu, S. O.; Apte, S. P. The effect of buffers on protein conformational stability. *Pharm. Technol.*, **2004**, *28*, (3), 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108.
- [55] Di, L.; Kerns, E. H. Solubility issues in early discovery and HTS. *Biotechnol.: Pharm. Aspects*, **2007**, *6* (Solvent Systems and Their Selection in Pharmaceutics and Biopharmaceutics), 111-136.
- [56] Du, L.; White, R. L. Improved partition equilibrium model for predicting analyte response in electrospray ionization mass spectrometry. *J. Mass Spectrom.*, **2009**, *44* (2), 222-229.
- [57] Schuur, J. H.; Selzer, P.; Gasteiger, J. The Coding of the Three-Dimensional Structure of Molecules by Molecular Transforms and Its Application to Structure-Spectra Correlations and Studies of Biological Activity. *J. Chem. Inf. Comput. Sci.*, **1996**, *36* (2), 334-344.
- [58] Scsibrany, H.; Karlovits, M.; Demuth, W.; Muller, F.; Varmuza, K. Clustering and similarity of chemical structures represented by binary substructure descriptors. *Chemom. Intell. Lab. Syst.*, **2003**, *67* (2), 95-108.
- [59] Hall, L. H.; Kier, L. B.; Brown, B. B. Molecular Similarity Based on Novel Atom-Type Electrotopological State Indices. *J. Chem. Inf. Comput. Sci.*, **1995**, *35* (6), 1074-1080.
- [60] Kier, L. B.; Hall, L. H. An electrotopological-state index for atoms in molecules. *Pharm. Res.*, **1990**, *7* (8), 801-807.
- [61] Beger, R. D.; Wilkes, J. G. Developing ¹³C NMR quantitative spectrometric data-activity relationship (QSDAR) models of steroid binding to the corticosteroid binding globulin. *J. Comput.-Aided Mol. Des.*, **2001**, *15* (7), 659-669.
- [62] Nguyen, H. P.; Schug, K. A. The advantages of ESI-MS detection in conjunction with HILIC mode separations: fundamentals and applications. *J. Sep. Sci.*, **2008**, *31* (9), 1465-1480.
- [63] Ehrmann, B. M.; Henriksen, T.; Cech, N. B. Relative importance of basicity in the gas phase and in solution for determining selectivity in electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.*, **2008**, *19* (5), 719-728.