CE_{50}: Quantifying Collision Induced Dissociation Energy for Small Molecule Characterization and Identification

Tzipporah M. Kertesz,[a] Lowell H. Hall,[b] Dennis W. Hill,[a] and David F. Grant[a]

[a] Department of Pharmaceutical Sciences, University of Connecticut, Storrs, Connecticut, USA
[b] Department of Chemistry, Eastern Nazarene College, Quincy, Massachusetts, USA

Survival yield analysis is routinely used in mass spectroscopy as a tool for assessing precursor ion stability and internal energy. Because ion internal energy and decomposition reaction rates are dependent on chemical structure, we reasoned that survival yield curves should be compound-specific and therefore useful for chemical identification. In this study, a quantitative approach for analyzing the correlation between survival yield and collision energy was developed and validated. This method is based on determining the collision energy (CE) at which the survival yield is 50% (CE_{50}) and, further, provides slope and intercept values for each survival yield curve. In initial experiments using a defined set of homologous compounds, we found that CE_{50} values were easily determined, quantitative, highly reproducible, and could discriminate between structural and even positional isomers. Further analysis demonstrated that CE_{50} values were independent of cone potential and orthogonal to compound mass. Experimentally determined CE_{50} values for a diverse set of 54 compounds were correlated to Molconn molecular structure descriptors. The resulting model yielded a statistically significant linear correlation between experimental and calculated CE_{50} values and identified several structural characteristics related to precursor ion stability and fragmentation mechanism. Thus, the CE_{50} is a promising method for compound identification and discrimination. (J Am Soc Mass Spectrom 2009, 20, 1759–1767) © 2009 American Society for Mass Spectrometry

Survival yield analysis was initially developed as a tool to quantify the distribution of precursor ion internal energies to explain fragmentation patterns that occur using mass spectrometry [1]. Survival yield has since been used as a method to correlate conditions in the mass spectrometer to the energetics of sample ions. These studies have developed a wide array of equations for understanding molecular decomposition in a mass spectrometer. The quasi-equilibrium theory of a unimolecular reaction indicates that the rate of molecular decomposition, as occurs in collision induced dissociation (CID), is dependent on the molecule's internal energy (E), activation energy (E_0), number of vibrational degrees of freedom (n), and the entropy of the reaction transition-state (ΔS^*). E_0, n, and ΔS^* are dependent on the structure of the molecule [2], whereas E is a function of the kinetic energy applied to the molecule in the collision cell. The fraction of a precursor molecule that survives a CID reaction (survival yield) depends on the reaction rate and the reaction time in the collision cell. In CID, transferring a portion of the kinetic energy of the accelerated precursor ion to internal energy by collisions with relatively stationary gas atoms increases the internal energy of a sample ion. The maximum energy available for absorption (E_{com}) by the precursor ion in the collision process is described by eq 1 [3]:

$$E_{com} = \frac{m_G * E_{kin}}{m_i + m_G} \quad (1)$$

where, E_{com} is the center of mass kinetic energy, m_G is the mass of the collision gas, E_{kin} is the kinetic energy of the sample ion, and m_i is the mass of the sample ion. The amount of energy available for conversion to internal energy is proportional to the sum of the initial internal energy of the precursor ion (E′), the internal energy of the collision gas (E_G), and the E_{com}. The fraction of this energy that is partitioned between the energy levels in the precursor ion (E) is given by eq 2; where D_i is the number of degrees of freedom of the sample molecule, D_G is the number of degrees of freedom of the collision gas, and D_T is the translational degrees of freedom [3].

Address reprint requests to Dr. D. F. Grant, Department of Pharmaceutical Sciences, University of Connecticut, 69 N. Eagleville Rd., UNIT 3092, Storrs, CT 06269-3092, USA. E-mail: david.grant@uconn.edu

1044-0305/09/$32.00
doi:10.1016/j.jasms.2009.06.002

Published online June 21, 2009
Received March 6, 2009
Revised May 28, 2009
Accepted June 2, 2009

$$E = \frac{(E^{'} + E_G + E_{com}) * D_i}{D_i + D_G + D_T} \quad (2)$$

All parameters affecting the internal energy of a sample ion in a CID experiment are either a property of the sample ion or collision gas, except E_{kin}. If the collision gas is held constant, then E_{kin} (as described by collision energy, CE) is an experimental parameter that can be varied to establish a correlation with the rate constant of a specific decomposition reaction. Equation 3 defines the decomposition reaction rate constant (k) in terms of the change in precursor ion intensity during a CID experiment,

$$\frac{I_{p,t}}{I_{p,o}} = e^{-kt} \quad (3)$$

Where t is the collision cell reaction time [4] and $I_{p,t}/I_{p,0}$ is the survival yield (SY). Rearrangement of eq 3 illustrates that the rate constant, k, at a given internal energy is equal to (ln(SY))/t. Since the reaction time in a specific collision cell is constant, the survival yield of a precursor ion at different collision energies correlates to the decomposition rate constant. For example, the collision energy that results in a survival yield of 0.5 reflects the amount of kinetic energy required to be transferred to internal energy of the precursor ion such that the reaction rate results in 50% decomposition during the reaction time as determined by the instrument collision cell geometry. Since the rate constant is dependent on the structure of the precursor ion, the collision energy at a survival yield of 50% should be dependent on the molecular structure of the precursor ion.

Recent work has used survival yield to analyze the fragmentation of peptides to differentiate energetics and fragmentation mechanisms caused by slight variations in peptide sequence [5–8]. The correlation between collision energy and the disruption of noncovalent interactions has been quantified in terms of CE_{50}, the collision energy that resulted in a 50% survival yield [9–12]. However, equations that describe the energetics of mass spectral fragmentation cannot be directly used to predict the collision energy required for the initiation or extent of fragmentation based on compound structure [4, 13–16]. Therefore, prediction of this energy for a given compound would require a quantitative structure–activity relationship approach to modeling by correlating structural descriptors to their associated experimental fragmentation energies based on empirical data. Unfortunately, current methods of survival yield analysis do not provide a rapid, quantitative method for assessing survival yield data for large numbers of compounds.

In this study, we developed a method for rapid quantification of survival yield data using CE_{50} to correlate the decomposition reaction kinetics to the structure of precursor ions. In addition, we have explored the use of molecular descriptors for predicting CE_{50}. Our results suggest that the CE_{50} may be useful as an orthogonal parameter for compound identification.

Experimental

Materials

The source and *m/z* values for the compounds used to determine CID mass spectral fragmentation profiles are listed in Supplementary Table 1, which can be found in the electronic version of this article. LC/MS grade methanol was purchased from Riedel-de Haen (Sigma-Aldrich, St. Louis, MO).

Instrumentation

Most samples were analyzed on a Micromass Q-TOF II (Beverly, MA); the Applied Biosystems Qstar Elite (Ontario, Canada) mass spectrometer was used as a comparison. The Q-TOF II was interfaced with a Z-Spray electrospray ionization source and the Qstar was interfaced with a TurboIon source. Unless otherwise noted, all analyses were performed in positive ion mode using a capillary potential of 3.1 kV on the Q-TOF II and 5.5 kV on the Qstar. The Q-TOF II source temperature was 120 °C and the desolvation gas (N_2) temperature was 150 °C. The Q-TOF II source cone gas (N_2) flow rate was 50 L/h and the desolvation gas flow rate was 450 L/h. The Qstar source gas (N_2) flow rate was 0.89 L/min and the curtain gas flow rate was 0.915 L/min. The Qstar declustering potential setting was 80, the declustering potential 2 setting was 15, and the focusing potential setting was 260. Compounds were introduced into either electrospray source by infusion with a Hamilton syringe pump (Reno, NV) at a flow rate of 5 μL/min. In the Q-TOF II, the protonated molecular ions of test compounds were generated at a cone potential of 20 V and isolated at unit resolution in the quadrupole analyzer. The Qstar was set to high-resolution to ensure unit resolution. The Q-TOF II used argon as the collision gas at a pressure of 10 psi and the Qstar used nitrogen with a CAD setting of 10.

Compound Analysis and Data Processing

Compounds were prepared at 1 μg/μL in methanol and analyzed individually by the described system to determine the CID spectrum at each collision energy. Samples were diluted if the intensity of the ion peaks indicated saturation of the micro-channel plate (MCP) detector on the mass spectrometer.

The CID spectrum of each test compound obtained at each collision energy was constructed from the sum of spectra generated over 20 scans. The resulting spectrum was smoothed and centered. A 1% threshold was employed to eliminate background noise.

The entire survival yield curve was determined by fragmenting the protonated molecular ions starting at a

collision energy of 5 eV and increasing in unit increments until the peak for the precursor ion was no longer detected. Five collision energies were analyzed in each run.

To determine the CE_{50} value without constructing the full survival yield curve, initial CID spectra of the protonated molecular ions were determined at 5, 10, 15, 20, and 30 eV in one analysis. If 50% or more of the molecular ion remained intact at 30 eV, another run was performed at 20, 30, 35, 40, and 50 eV. Twenty scans were co-added per processed spectrum. The spectrum for each collision energy was quantified to determine the experimental survival yield (eq 4) as the intensity of the precursor ion divided by the sum of the precursor and product ion intensities. The CE_{50} was approximated by the sigmoid survival yield curve (eq 5). The compound was then analyzed in a second run in which the approximate CE_{50} value was analyzed concurrently with two collision energies above and two energies below this value (in 2 eV increments). For example, if a compound was estimated to have a CE_{50} value of 20 eV, the accurate CE_{50} was determined using the energies 16, 18, 20, 22, and 24 eV. Each analysis was run in triplicate and averaged to determine the final CE_{50} values.

Structural Differentiation by CE_{50}

Six isomers with the molecular formula $C_9H_{11}NO_2$ were analyzed to determine the potential for differentiating compounds of the same molecular weight using CE_{50} analysis. Three of these analytes, 2-(dimethylamino)-benzoic acid, 3-(dimethylamino)benzoic acid, and 4-(dimethylamino)benzoic acid, were positional isomers.

Correlation between CE_{50} and E-State

CE_{50} values were determined for *N*-phenylglycine and 5 ortho-substituted derivatives. Two of these derivatives, (2-chloroanilino)acetic acid and (2-fluoroanilino)acetic acid, contained electron withdrawing groups expected to increase the electron density as well as the electronegativity of the heteroatoms on the glycine chain. Three of these derivatives, (2-ethylanilino)acetic acid and (2-methoxyanilino)acetic acid and (2-phenylanilino)acetic acid, contained electron donating groups expected to decrease the electronegativity of the heteroatoms on the glycine chain. E-state values for individual atoms within a compound were determined using E-Calc software [17]. The E-state values were used as surrogates of electronegativity to test the correlation between changes in electronegativity and CE_{50} values.

Modeling CE_{50} Data with Molecular Structure Descriptors

The polarity for the dataset of 55 compounds, as described by AlogPs, was determined using the online AlogPs applet (http://www.vcclab.org/lab/alogps/start.html) [18, 19]. Molconn software [20] was used to calculate a wide range of molecular descriptors for the set of 55 compounds for which we had determined experimental CE_{50} values. The set of structure descriptors that were evaluated consisted of E-state descriptors for atoms and bonds [17, 21–23], molecular connectivity descriptors representing ring structures and extended paths of atoms [24–27], and descriptors related to maximum and minimum electron accessibilities on hydrogen atoms and $-XH_n$ groups in the molecule [23, 27, 28]. A total of 39 structure descriptors were evaluated for multiple linear regression (MLR) analysis.

SAS regression software ('proc reg') with the 'selection = stepwise' procedure was used for preliminary analysis to determine which structure descriptors were likely to be most significant [29]. Further, if any pair of descriptors was intercorrelated ($r^2 > 0.60$), one of the pair of descriptors was eliminated. The CE_{50} model (eq 7) was developed using SAS regression software ('proc reg') with the 'selection = rsquare' procedure. This SAS procedure automatically examines all possible models for one to eight variables (structure descriptors) and identifies the model with the highest correlation. A maximum of eight variables was considered for the MLR model as a balance between using all the descriptors that explain the correlation and using the minimum number of statistical variables to retain statistical significance. The set of eight variables in the best model leads to a ratio of observations-to-variables (6.8) that satisfies the requirement range 5–10 commonly used in structure-property modeling. The SAS procedure insures that all other possible models with eight or fewer variables have inferior regression statistics. Finally, SAS regression software with the 'proc reg' module was used to analyze statistically the correlation between experimental and calculated CE_{50} values computed from the model (eq 7) and to determine the residual value for each compound [29].

Results and Discussion

In initial experiments, we determined complete survival yield curves for two compounds, tetracycline and prazosin, starting at a CE of 5 eV and increasing the CE in unit increments until the precursor ion was no longer detected. The spectrum for each CE was quantified as an experimental survival yield that was calculated as the ratio of the intensity of the precursor ion to the sum of intensities of the precursor and fragment ions (eq 4).

$$Experimental\ SY = \frac{IntensityP}{IntensityP + \sum (IntensityF)} \quad (4)$$

As expected, when the protonated molecular ion of each compound was exposed to increasing CE, the fraction of the molecular ion that survived decreased in a sigmoidal relationship (eq 5), where b is the slope of the linear segment and the intercept of the linear portion is the natural logarithm of c (Figure 1a).

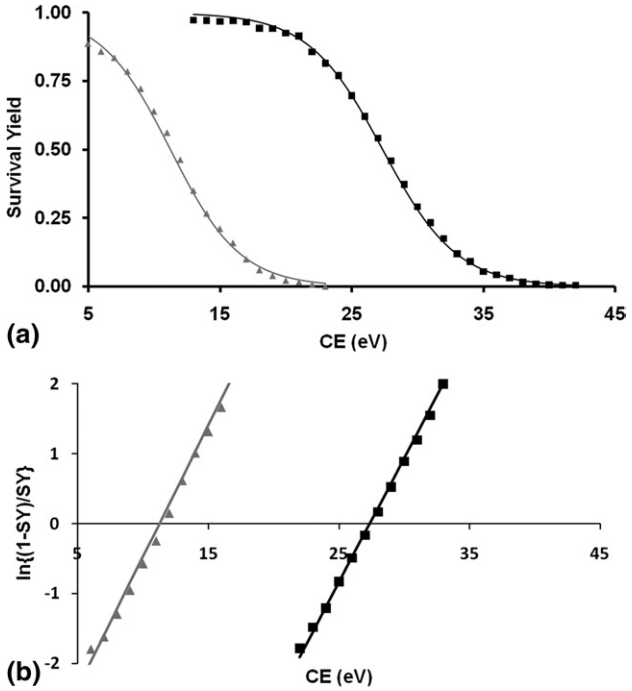


Figure 1. Survival yield curves. (**a**) Comparison of experimental survival yield data for prazosin (filled square) and tetracycline (filled triangle), with theoretical sigmoid survival yield curves for prazosin and tetracycline. Prazosin sigmoid survival yield = $1/(1 + 0.0000589e^{0.356*CE})$. Tetracycline sigmoid survival yield = $1/(1 + 0.0143766e^{0.378*CE})$. (**b**) Comparison of the linear survival yield for prazosin (filled square) and tetracycline (filled triangle). Prazosin linear curve ln[(1-SY)/SY] = ln(0.0000589) + 0.356 * CE [CE_{50} = 27.51 eV (±0.009 eV)]. Tetracycline linear curve ln[(1-SY)/SY] = ln(0.0143766) + 0.378 * CE [CE_{50} = 11.11 eV (± 0.038 eV)].

$$Sigmoidal\ SY = \frac{1}{1 + c * e^{-b \cdot CE}} \quad (5)$$

Equation 6 is the linear transformation of eq 5 and graphs of the SY data for tetracycline and prazosin in this form are illustrated in Figure 1b.

$$ln\left\{\frac{1 - SY}{SY}\right\} = ln(c) - b(CE) \quad (6)$$

Previous survival yield studies have been performed on in-source fragmentation data for analytes that fragment by a single pathway [4, 30–36]. The analytes in this study have the potential for multiple sites of protonation, charge migration before fragmentation in the collision cell, and multiple fragmentation pathways. The resultant sigmoid curve is therefore a composite of these reactions.

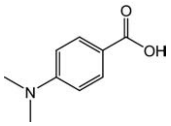
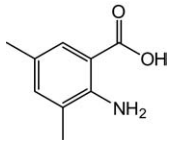
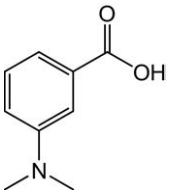
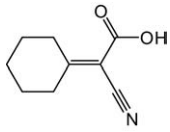
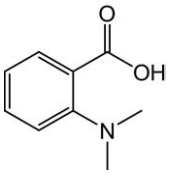
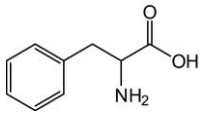
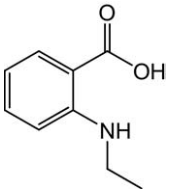
The point where 50% of the parent ion remains intact lies in the linear portion of the curve; this allows for the CE_{50} to be accurately calculated using fewer collision energies. An initial MS/MS experiment identifies an approximate CE_{50} and a second experiment using five collision energies bracketing the approximate CE_{50} is used to determine the final CE_{50} value. To further evaluate this approach, we compared two methods for determining the final CE_{50} value. In the first method, differences of single unit increments of collision energy were analyzed for five energies around the initial estimated CE_{50}. The second method employed differences of double unit increments of collision energy. The CE_{50} values determined for six compounds using the full sigmoid curve were used for comparison. The use of single unit increments resulted in an average CE_{50} standard deviation of 0.4 eV between the full curve and the subset, while double unit increments yielded an average deviation of 0.3 eV. Based on these results, we used double unit intervals for five points around the estimated CE_{50} to calculate the final experimental CE_{50} value.

The potential for using CE_{50} to discriminate among compounds with similar structures was tested by analyzing seven isomers (including three positional isomers) with the molecular formula $C_9H_{11}NO_2$. The CE_{50} values ranged from 8.24 to 16.52 and the three positional isomers had significantly different CE_{50} values (Table 1).

A possible explanation for the difference in CE_{50} values for these three positional isomers is the difference in electron withdrawing and donating resonance effects between the ortho, meta, and para positions. To test this hypothesis, *N*-phenylglycine and a group of five ortho-substituted derivatives were analyzed. Two of these derivatives, 2-chloro and 2-fluoro, contain electron withdrawing groups that should increase the electronegativity of the nitrogen and oxygen atoms. The three other derivatives, 2-phenyl, 2-ethyl, and 2-methoxy, contain electron donating groups that should decrease the electronegativity of the nitrogen and oxygen atoms. The two derivatives containing electron withdrawing groups had CE_{50} values less than *N*-phenylglycine. The three derivatives containing electron donating groups had CE_{50} values greater than N-phenylglycine (Table 2). There was a significant correlation between electronegativity (quantified by the E-state) of either oxygen ($P < 0.01$) or the nitrogen ($P < 0.02$) and CE_{50}. However, the electronegativity of the carbons on the glycine chain showed no correlation with CE_{50} ($P > 0.1$) suggesting the importance of heteroatom electronegativity on CE_{50}. The predominant fragment formed for these derivatives was the same as that for *N*-phenylglycine (loss of 46 Da) as described by the proposed scheme in Table 2. Thus, these data are consistent with a concept implicating the importance of electronegativity in determining fragmentation rates and mechanisms [13, 37, 38].

CE_{50} values were then determined for a diverse set of 55 compounds ionized in positive ion mode (Supplementary Table 2). Compounds were chosen to cover a wide range of molecular weights, polarities, functional groups, and CE_{50} values (Supplementary Figure 1). The monoisotopic molecular weights of these 55 test compounds were not statistically correlated with CE_{50} ($P = 0.12$, Figure 2). The reproducibility of CE_{50} values determined on the same day had an average standard

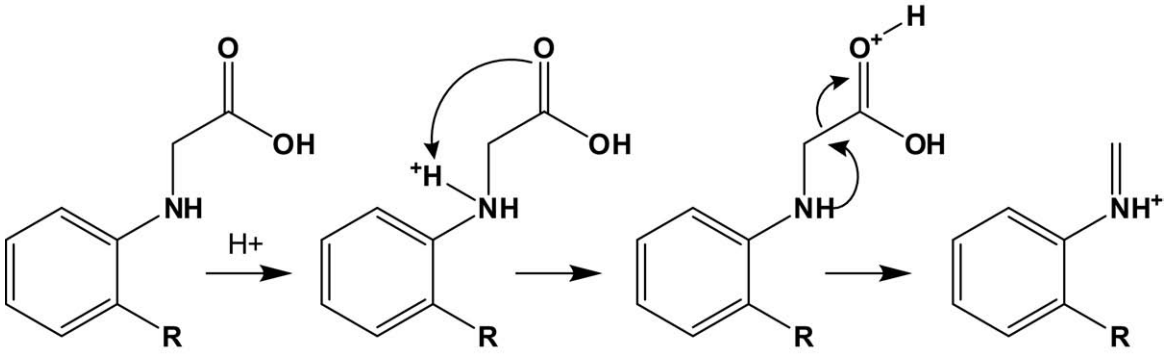
Table 1. CE_{50} values for $C_9H_{11}NO_2$ isomers

Structure	CE_{50} (eV)	Standard deviation	Structure	CE_{50} (eV)	Standard deviation
	15.60	0.062		8.24	0.135
	16.52	0.068		11.93	0.072
	11.49	0.026		8.60	0.012
	10.14	0.013			

deviation of 0.04 eV and a maximum standard deviation of 0.49 eV. Four compounds analyzed repeatedly over an 8 mo period had a maximum standard deviation of 0.66 eV.

The results in Table 2 suggest the importance of electronegativity in determining CE_{50}. Therefore, the use of E-state descriptors was evaluated for creating a statistically significant model of CE_{50} using the 55 compounds listed in Supplementary Table 2. The approach used in modeling the CE_{50} data were based on structure information representation as previously described for biological properties (serum protein bind-

Table 2. Proposed fragmentation mechanism and CE_{50} values for N-phenylglycine derivatives



R group	$CE_{50} \pm SD$ (eV)	N E-state	O E-state	O(H) E-state	R group	$CE_{50} \pm SD$ (eV)	N E-state	O E-state	O(H) E-state
F	8.23 ± 0.005	2.44	10.1	8.28	OCH_3	9.28 ± 0.016	2.74	10.27	8.44
Cl	8.51 ± 0.016	2.68	10.18	8.36	CH_2CH_3	10.19 ± 0.014	2.86	10.31	8.48
H	9.00 ± 0.015	2.74	10.11	8.31	Ph	10.87 ± 0.003	2.91	10.56	8.67

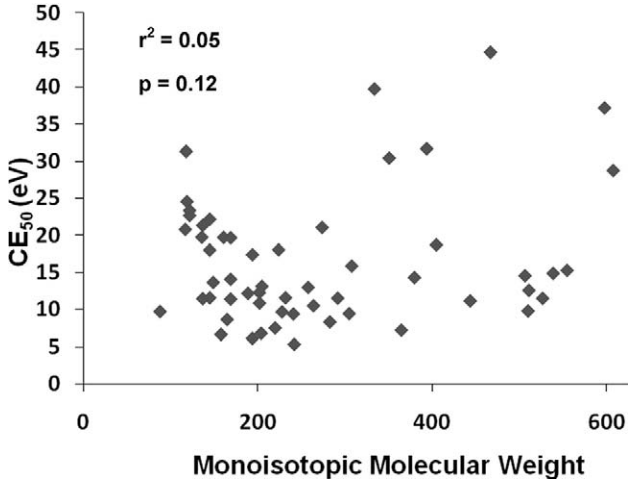


Figure 2. Plot of monoisotopic molecular weight versus CE_{50} for 55 compound dataset.

ing, blood–brain barrier partitioning, and human intestinal absorption) as well as chemical properties such as solubility and chromatographic retention index [23, 28, 39–44].

The E-state value for an atom is derived from its valence state electronegativity and its local topology [17, 23]. The E-state value for an atom or group represents the electron accessibility for that atom or group. In this way, an atom with high electronegativity that is also a terminal atom/group (i.e., $-NH_2$, $-OH$, $-F$, $=O$) has a high E-state value. On the other hand, an atom or group with low electronegativity or one that is more topologically crowded (i.e., $-O-$, $>N-$, $-S-$) has a lower E-State value. Highly buried atoms/groups (i.e., $>C<$, S in $-SO_2-$) have very low E-state values. The success of these descriptors in representing intermolecular interactions is reflected in the development of excellent models for boiling point and critical temperature for diverse sets of organic compounds [45, 46].

The model with the highest r^2 and lowest standard error includes eight molecular descriptors from the starting set of thirty-nine:

$$\begin{aligned} CE_{50} = {} & 26.8 - 1.31 * \text{sumI} + 15.0 * \text{xvp4} - 15.5 * \text{dxvp3} \\ & - 1.26 * \text{SssCH2} + 4.60 * \text{SallNp} \\ & + 0.0866 * \text{SCar} - 1.41 * \text{eaC2C3s} \\ & + 1.33 * \text{eaC2N2a} \end{aligned} \quad (7)$$

where: sumI is the sum of the E-state intrinsic values for all atoms [17]; xvp4 is the sum of the extent of branching for groups of four consecutive atoms [23, 25, 26, 47]; dxvp3 is the sum of the extent of branching for groups of three consecutive atoms and is normalized to be independent of molecular size [23, 24, 26, 42, 48]; SssCH2 is the sum of E-state values for all methylene groups ($-CH_2-$) [23, 48]; SallNp is the sum of E-State values for all quaternary nitrogen atoms [23, 48]; SCar is the sum of E-State values for all carboxylic acid functional groups [23, 48]; eaC2C3s is the sum of bond E-State values for all aromatic carbon-carbon bonds [23, 48]; eaC2N2a is the sum of bond E-state values for all unsubstituted carbon-nitrogen bonds in aromatic rings [23, 48].

$$r^2 = 0.810,\ \text{standard error} = 3.8,\ F\ \text{value} = 24.0,$$
$$P < 0.0001,\ n = 54.$$

Figure 3 presents the plot of the CE_{50} values calculated from eq 7 versus the experimental CE_{50} values. As can be seen, the calculated CE_{50} and experimental CE_{50} values are statistically correlated ($P < 0.0001$, $r^2 = 0.81$). The compound name, measured CE_{50} value, calculated CE_{50} value (eq 7), and residual value (residual = experimental – calculated) are listed in Supplementary Table 2 in order of increasing CE_{50}.

Detailed information explaining the meaning of the descriptors used in the model (eq 7) has been previously published [17, 23, 28, 41–48]. The descriptor that contributes most (50%) to the computed value of CE_{50} is sumI. SumI is equal to the sum of all E-state intrinsic values and is also equal to the sum of atom E-state values in the molecule. Thus, the more electronegative and the more electron-accessible the atoms, the larger is the value of sumI. As indicated by the negative coefficient in eq 7, for larger values of sumI, the CE_{50} value is lower. The importance of sumI further supports the influence of electronegativity seen with the *N*-phenylglycine derivatives. The percent contribution for each descriptor to the model is listed in Supplementary Table 3. Of the eight descriptors identified, the three most important (sumI, xvp4, and dxvp3) account for 93.7% of the contribution to the model. The different variables either increase or decrease CE_{50} according to the sign of their coefficients in eq 7. This combination of descriptors in the model describes the structural features most related to the survival of the parent ion in the collision

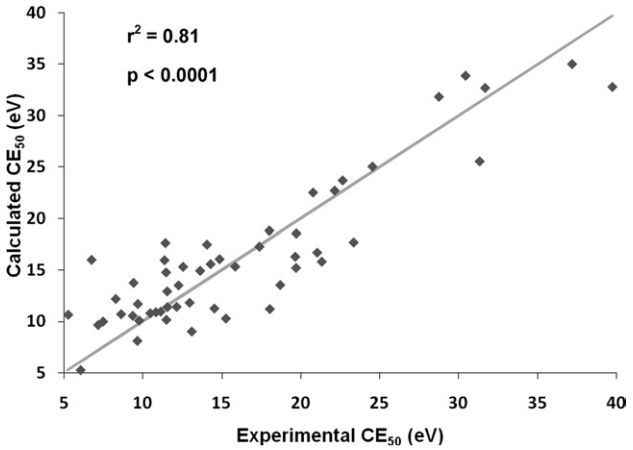


Figure 3. Plot of CE_{50} values calculated from eq 7 versus CE_{50} experimental values.

cell. The three most important descriptors (sumI, xvp4, and dxvp3) represent features associated with the entire molecule since every atom contributes to the descriptor. The five other structure descriptors represent more localized features. These eight descriptors appear to be consistent with mass spectral fragmentation; however, the relevance of these descriptors in predicting CE_{50} awaits development of a predictive model based on a larger dataset.

One compound, buprenorphine (last entry, Supplementary Table 2), possesses the largest experimental CE_{50} value (44.67 eV) in this dataset and is considered to be a singleton. Its CE_{50} value is 4.95 eV greater than the next largest value (strychnine, 39.72 eV); that difference represents 12.6% of the data range. Thus, buprenorphine possessed excessive leverage on the regression and was not used in the final analysis.

The statistical information presented in eq 7 indicates that a significant model was developed with 54 compounds and eight variables. The correlation coefficient ($r^2 = 0.810$), standard error ($s = 3.8$), and F ratio ($F = 24.0$, $P < 0.0001$) clearly indicate statistical significance. The distribution of residual values is also reasonable; only one residual value (D-tryptophan, residual −9.25) is greater than twice the standard error and none are greater than three times the standard error. No trends appear in the distribution of residuals; the distribution appears random. A functional predictive model would require a standard error closer to the experimental standard error. A model with a higher correlation coefficient and lower standard error will likely be produced with a larger dataset. The development of a predictive model using our methods was demonstrated previously in a model developed for predicting chromatographic retention index values [39]. That model was based on a much larger dataset ($n = 498$, 394 in training set), permitting the use of a greater number of structure descriptors and also allowing the use of a nonlinear artificial neural network analysis. Predictive models have been developed using these methods for a range of data types [41–47]. Thus, although the present model is not useful for predictive purposes, its development clearly suggests that modeling of a larger dataset could be predictive [23, 28, 39, 41–44]. Furthermore, the anticipated improved model may also assist in identifying further structural features important in determining precursor ion stability and fragmentation mechanisms [23, 28, 42–46]. In addition, our model (eq 7) was created based on the structures of the unprotonated precursors. Since fragmentation is initiated from protonated ions, a model based on the protonated precursors may prove to be more accurate. However, many of the compounds analyzed here contain multiple sites of protonation preventing the use of a single charged structure in the model. More extensive work will be required to prepare a model based on all potential protonated species.

Beyond improving the current model, compound identification can be aided by the use of additional compound descriptors. The survival yield curve generates two additional quantitative values: slope (b) and intercept [ln (c)]. However, from the 55 compound dataset we found that both the slope and the intercept were statistically correlated to CE_{50} ($P < 0.001$). This correlation suggests that using either slope or intercept as an additional compound descriptor for compound identification might not be beneficial.

A predictive model would be most useful if CE_{50} studies are transferable between instruments. Previous studies have shown that survival yield data are dependent on source settings such as cone potential and capillary voltage [4, 30, 33, 34, 36]. These differences change the mean internal energy of the precursor in the source, and thereby have the potential to change the energy requirements to induce fragmentation. If all of these parameters have significant impacts on CE_{50}, reproducing results between instruments might be difficult. The effect of cone potential on CE_{50} was tested for three compounds. Although the CE_{50} values changed slightly with cone potential, there was no consistent trend and the change in CE_{50} never exceeded 1 eV over a cone potential range of 75 V (Figure 4). This rather surprising result suggests that factors that determine protonation in the ionization source do not have a large impact on CE_{50}.

In addition to the ionization source, conditions in the collision cell may have the potential to impact CE_{50}. One commonly encountered variable is the type of collision gas used in the collision cell. Ten compounds were analyzed on the Micromass Q-TOF II using either argon or nitrogen as the collision gas at a head pressure of 10 psi. All of the CE_{50} values obtained by fragmentation using nitrogen were higher than those using argon (Supplementary Table 4). The nitrogen to argon CE_{50} ratio had a mean value of 1.35, with a standard deviation of 0.058 and a coefficient of variation of 4%. These results are consistent with what was expected based on eq 1, which predicts that a higher mass collision gas will provide more energy to an analyte during a collision. Therefore, a lower collision energy would be required to induce fragmentation with a higher mass collision gas.

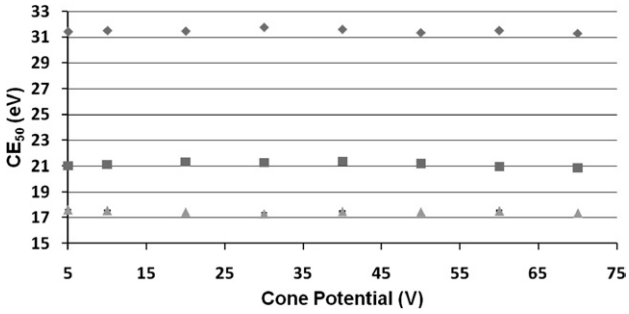


Figure 4. Effect of cone potential on CE_{50} for (filled diamond) benzimidazole, (filled square) 1-methylnicotinamide, and (filled triangle) caffeine. Error bars show standard deviation of three replicates.

These same 10 compounds were also analyzed on an Applied Biosystems Qstar Elite, another quadrupole-time of flight hybrid instrument with an electrospray ionization source. This instrument typically uses nitrogen as its collision gas, so if the collision gas is the only factor controlling CE_{50}, the CE_{50} values might be expected to be similar between the two instruments. Preliminary results yielded similar CE_{50} values, when fragmenting with nitrogen, between the instruments for eight of the 10 compounds investigated; however two of the compounds had significantly different CE_{50} values on the two instruments. This suggests that factors other than collision gas may contribute to CE_{50} values. These factors might include the geometry of the source and/or the collision cell [30]. As with other commonly used tools, such as retention time, a proper conversion factor based on the analysis of a group of standards may be required to compare CE_{50} values obtained on different instruments.

Conclusions

The work presented here describes the use of collision induced dissociation as a quantitative method for correlating fragmentation energy (CE_{50}) to precursor ion structure. We found that CE_{50} values could be quickly determined and were highly reproducible over time. Differences in heteroatom electronegativity were shown to influence CE_{50} and provide an example of how this approach can be used to understand CID fragmentation mechanisms. Although CE_{50} values obtained on one mass spectrometer are not necessarily transferable to other instruments, CE_{50} values determined on any mass spectrometer can be used for analyte comparison.

The choice of a chemical property for the characterization and identification of unknown compounds by mass spectrometry requires consideration of several factors. The molecular characteristic needs to have sufficient resolution to differentiate among compounds having the same molecular formula and very similar structures. The analysis needs to be easily performed, highly reproducible, and, to aid in database searches, the chemical property should be predictable using appropriate computational models. Comparing chemical structures to experimental CE_{50} for an eclectic mix of 54 compounds yielded a statistically significant multivariable linear regression model based on E-state descriptors. Thus, our results suggest that the CE_{50} could aid in our understanding of fragmentation mechanisms and can be used as an orthogonal chemical descriptor for compound identification.

Acknowledgments

The authors acknowledge support for this work by a grant from Pfizer Inc., Groton, CT, the University of Connecticut Foundation, and by an American Foundation for Pharmaceutical Education predoctoral fellowship. The authors thank Chad M. Kormos and Dr. Nicholas Leadbeater, Chemistry Department, University of Connecticut, for the (2-phenylanilino)acetic acid.

Appendix A Supplementary Material

Supplementary material associated with this article may be found in the online version at doi:10.1016/j.jasms.2009.06.002.

References

1. Kenttämaa, H. I.; Cooks, R. G. Internal energy distributions acquired through collisional activation at low and high energies. *Int. J. Mass Spectrom. Ion Processes* **1985,** *64,* 79–83.
2. Sleno, L.; Volmer, D. A. Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.* **2004,** *39,* 1091–1112.
3. Drahos, L.; Vékey, K. MassKinetics: A theoretical model of mass spectra incorporating physical processes, reaction kinetics, and mathematical descriptions. *J. Mass Spectrom.* **2001,** *36,* 237–263.
4. Gabelica, V.; De Pauw, E. Internal energy and fragmentation of ions produced in electrospray sources. *Mass Spectrom. Rev.* **2005,** *24,* 566–587.
5. de Maaijer-Gielbert, J.; Gu, C.; Somogyi, A.; Wysocki, V. H.; Kistemaker, P. G.; Weeding, T. L. Surface-induced dissociation of singly and multiply protonated polypropylenamine dendrimers. *J. Am. Soc. Mass Spectrom.* **1999,** *10,* 414–422.
6. Dongre, A. R.; Somogyi, A.; Wysocki, V. H. Surface-induced dissociation: An effective tool to probe structure, energetics, and fragmentation mechanisms of protonated peptides. *J. Mass Spectrom.* **1996,** *31,* 339–350.
7. Guo, X.; Duursma, M. C.; Kistemaker, P. G.; Nibbering, N. M.; Vekey, K.; Drahos, L.; Heeren, R. M. Manipulating internal energy of protonated biomolecules in electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *J. Mass Spectrom.* **2003,** *38,* 597–606.
8. Vekey, K.; Somogyi, A.; Wysocki, V. H. Average activation energies of low-energy fragmentation processes of protonated peptides determined by a new approach. *Rapid Commun. Mass Spectrom.* **1996,** *10,* 911–918.
9. Bovet, C.; Wortmann, A.; Eiler, S.; Granger, F.; Ruff, M.; Gerrits, B.; Moras, D.; Zenobi, R. Estrogen receptor-ligand complexes measured by chip-based nanoelectrospray mass spectrometry: an approach for the screening of endocrine disruptors. *Protein Sci.* **2007,** *16,* 938–946.
10. Danikiewicz, W.; Tarnowski, P.; Bienkowski, T.; Jurczak, J. Estimation of the noncovalent bond dissociation energies of the gas-phase complexes of macrocyclic polyethers with alkali metal cations using an electrospray ionization/triple quadrupole mass spectrometer. *Polym. J. Chem.* **2004,** *78,* 699–709.
11. Rosu, F.; Nguyen, C. H.; De Pauw, E.; Gabelica, V. Ligand binding mode to duplex and triplex DNA assessed by combining electrospray tandem mass spectrometry and molecular modeling. *J. Am. Soc. Mass Spectrom.* **2007,** *18,* 1052–1062.
12. Rosu, F.; Pirotte, S.; Pauw, E. D.; Gabelica, V. Positive and negative ion mode ESI-MS and MS/MS for studying drug-DNA complexes. *Int. J. Mass Spectrom.* **2006,** *253,* 156–171.
13. Marcus, R. A. Unimolecular dissociations and free radical recombination reactions. *J. Chem. Phys.* **1952,** *20,* 359–364.
14. McLafferty, F. W.; Turecek, F. Theory of Unimolecular Ion Decompositions. In *Interpretation of Mass Spectra,* 4th ed.; University Science Books: Sausalito, 1993; pp 115–134.
15. Morgan, D. G.; Bursey, M. M. Linear energy correlation in the low-energy tandem mass spectra of protonated tripeptides Xxx-Gly-Gly but failure for Gly-Xxx-Gly. *J. Mass Spectrom.* **1995,** *30,* 290–295.
16. Vékey, K.; Internal energy effects in mass spectrometry. *J. Mass Spectrom.* **1996,** *31,* 445–463.
17. Hall, L. H.; Kier, L. B. Molecular Structure Description: The Electrotopological State; Academic Press: San Diego, 1999; p 286.
18. 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,* 453–463.
19. VCCLAB, Virtual Computational Chemistry Laboratory http://www.vcclab.org, 2005.
20. Hall Associates Consulting. *Molconn-Z*, **3.5**; 2 Davis Street, Quincy MA 02170.
21. Hall, L. H.; Kier, L. B. The Electrotopological State: Structure Modeling for QSAR and Database Analysis. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J.; Balaban, A. T., Eds. Gordon and Breach: Reading, UK, 1999; pp 491–462.
22. Hall, L. H.; Kier, L. B. The E-state as the basis for molecular structure space definition and structure similarity. *J. Chem. Inf. Comput. Sci.* **2000,** *40,* 784–791.
23. Hall, L. H.; Kier, L. B.; Hall, L. M. Electrotopological State Indices to Assess Molecular and ADMET Properties. In *Comprehensive Medicinal*

Chemistry, 2nd ed.; Triggle, D. J.; Taylor, J. B., Eds. Elsevier: Oxford, UK, 2007; Vol. V, pp 555–576.
24. Hall, L. H.; Kier, L. B. Issues in representation of molecular structure the development of molecular connectivity. *J. Mol. Graph. Model.* **2001,** *20,* 4–18.
25. Kier, L. B.; Hall, L. H. Molecular Connectivity in Chemistry and Drug Research; Academic Press, Inc.: New York, 1976; pp 55–61, 66.
26. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; Research Studies Press Ltd.: Hertfordshire, England and John Wiley and Sons: New York, 1986.
27. Kier, L. B.; Hall, L. H. The prediction of ADMET properties using structure information representations. *Chem. Biodivers* **2005,** *2,* 1428–1437.
28. Votano, J. R.; Parham, M.; Hall, L. H.; Kier, L. B.; Hall, L. M. Prediction of aqueous solubility based on large datasets using several QSPR models utilizing topological structure representation. *Chem. Biodivers* **2004,** *1,* 1829–1841.
29. *SAS*, SAS Institute: Cary, NC 27513.
30. Collette, C.; De Pauw, E. Calibration of the internal energy distribution of ions produced by electrospray. *Rapid Commun. Mass Spectrom.* **1998,** *12,* 165–170.
31. Collette, C.; Drahos, L.; De Pauw, E.; Vékey, K. Comparison of the internal energy distributions of ions produced by different electrospray sources. *Rapid Commun. Mass Spectrom.* **1998,** *12,* 1673–1678.
32. Drahos, L.; Heeren, R. M.; Collette, C.; De Pauw, E.; Vekey, K. Thermal energy distribution observed in electrospray ionization. *J. Mass Spectrom.* **1999,** *34,* 1373–1379.
33. Gabelica, V.; De Pauw, E.; Karas, M. Influence of the capillary temperature and the source pressure on the internal energy distribution of electrosprayed ions. *Int. J. Mass Spectrom.* **2004,** *231,* 189–195.
34. Hampton, C. Y.; Silvestri, C. J.; Forbes, T. P.; Varady, M. J.; Meacham, J. M.; Fedorov, A. G.; Degertekin, F. L.; Fernández, F. M. Comparison of the internal energy deposition of venturi-assisted electrospray ionization and a venturi-assisted array of micromachined ultrasonic electrosprays (AMUSE). *J. Am. Soc. Mass Spectrom.* **2008,** *19,* 1320–1329.
35. Luo, G.; Marginean, I.; Vertes, A. Internal energy of ions generated by matrix-assisted laser desorption/ionization. *Anal. Chem.* **2002,** *74,* 6185–6190.
36. Nefliu, M.; Smith, J. N.; Venter, A.; Cooks, R. G. Internal energy. Distributions in desorption electrospray ionization (DESI). *J. Am. Soc. Mass Spectrom.* **2008,** *19,* 420–427.
37. Rosenstock, H. M.; Wallenstein, M. B.; Wahrhaftig, A. L.; Eyring, H. Absolute rate. Theory for isolated systems and the mass spectra of polyatomic molecules. *Proc. Natl. Acad. Sci. U.S.A.* **1952,** *38,* 667–678.
38. Smith, R. M. *Understanding Mass Spectra: A Basic Approach*, 2nd ed.; John Wiley and Sons, Inc.: Hoboken, 2004; p 392.
39. Albaugh, D. R.; Hall, L. M.; Hill, D. W.; Kertesz, T. M.; Parham, M.; Hall, L. H.; Grant, D. F. Prediction of HPLC retention index using artificial neural networks and I-group E-state indices. *J. Chem. Inf. Model.* **2009,** *49,* 788–799.
40. Hall, L. H.; Hall, L. M. QSAR modeling based on structure-information for properties of interest in human health. *SAR QSAR Environ. Res.* **2005,** *16,* 13–41.
41. Hall, L. M.; Hall, L. H.; Kier, L. B. Modeling drug albumin binding affinity with e-state topological structure representation. *J. Chem. Inf. Comput. Sci.* **2003,** *43,* 2120–2128.
42. Rose, K.; Hall, L. H.; Kier, L. B. Modeling blood-brain barrier partitioning using the electrotopological state. *J. Chem. Inf. Comput. Sci.* **2002,** *42,* 651–666.
43. Votano, J. R.; Parham, M.; Hall, L. H.; Kier, L. B. New predictors for several ADME/Tox properties: Aqueous solubility, human oral absorption, and Ames genotoxicity using topological descriptors. *Mol. Divers* **2004,** *8,* 379–391.
44. Votano, J. R.; Parham, M.; Hall, L. M.; Hall, L. H.; Kier, L. B.; Oloff, S.; Tropsha, A. QSAR modeling of human serum protein binding with several modeling techniques utilizing structure-information representation. *J. Med. Chem.* **2006,** *49,* 7169–7181.
45. Hall, L. H.; Story, C. T. Boiling point and critical temperature of a heterogeneous data set: QSAR with atom type electrotopological state indices using artificial neural networks. *J. Chem. Inf. Comput. Sci.* **1996,** *36,* 1004–1014.
46. Hall, L. H.; Story, C. T. Boiling point of a set of alkanes, alcohols, and chloroalkanes: QSAR with atom type electrotopological state indices using artificial neural networks. *SAR QSAR Environ. Res.* **1997,** *6,* 139–161.
47. Hall, L. M.; Hall, L. H.; Kier, L. B. QSAR modeling of beta-lactam binding to human serum proteins. *J. Comput. Aided. Mol. Des.* **2003,** *17,* 103–118.
48. Hall, L. H.; Kier, L. B.; Hall, L. M. Topological Quantitative Structure–Activity Relationship Applications: Structure Information Representation in Drug Discovery. In *Comprehensive Medicinal Chemistry*, 2nd ed.; Taylor, J. B.; Triggle, D. J., Eds. Elsevier: Oxford, UK, 2007; Vol. 4, pp 537–574.