Send Orders for Reprints to reprints@benthamscience.ae

*Drug Metabolism Letters*, 2018, *12*, 93-100

RESEARCH ARTICLE

# Using LC Retention Times in Organic Structure Determination: Drug Metabolite Identification

William L. Fitch[1,*], Cyrus Khojasteh[2], Ignacio Aliagas[3] and Kevin Johnson[2]

[1]*Department of Gastroenterology and Hepatology, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, CA 94305, USA;* [2]*Department of Drug Metabolism and Pharmacokinetics 1 DNA Way MS 412a, Genentech Inc., South San Francisco, CA 94080, USA;* [3]*Discovery Chemistry, 1 DNA Way, Genentech Inc., South San Francisco, CA 94080, USA*

**Abstract:** ***Background*:** There is a continued need for improvements in the efficiency of metabolite structure elucidation.

***Objective*:** We propose to take LC Retention Time (RT) into consideration during the process of structure determination.

***Methods*:** Herein, we develop a simple methodology that employs a Chromatographic Hydrophobicity Index (CHI) framework for standardizing LC conditions and introduce and utilize the concept of a predictable CHI change upon Phase 1 biotransformation ($CHI_{bt}$). Through the analysis of literature examples, we offer a Quantitative Structure-Retention Relationship (QSRR) for several types of biotransformation (especially hydroxylation) using physicochemical properties (clogP, hydrogen bonding).

***Results*:** The CHI system for retention indexing is shown to be practical and simple to implement. A database of $CHI_{bt}$ values has been created from re-incubation of 3 compounds and from analysis of an additional 17 datasets from the literature. Application of this database is illustrated.

***Conclusion*:** In our experience, this simple methodology allows complementing the discovery efforts that saves resources for in-depth characterization using NMR.

**ARTICLE HISTORY**

Received: March 10, 2018
Revised: June 13, 2018
Accepted: July 31, 2018

*DOI:*
*10.2174/1872312812666180802093347*

**Keywords:** Drug metabolite identification, liquid chromatography, mass spectrometry (MS), RT prediction, RT, CHI system.

## 1. INTRODUCTION

The lowering of intrinsic metabolic clearance is a key part of the optimization process that takes place during drug discovery. A low intrinsic metabolic clearance is one of the properties that could lead to a lower dose and an adequate circulating half-life that, in turn, allows for oral delivery of a drug. Metabolic stability optimization is conducted by in-depth analyses of the types of metabolic biotransformations a drug undergoes and the modification sites of the drug, investigations which are conducted primarily *via* liquid chromatography coupled to mass spectrometry (LC-MS) [1]. Recent advances of high-resolution MS allow for a very accurate method for the identification and characterization of highly complex organic mixtures. This methodology has been amenable to higher throughput applications for monitoring and quantifying a drug and its metabolites. Metabolites are detected and characterized based on their mass over charge ratio (*m/z*), and the interpretation of fragment ions are based on Collision-induced Dissociation (CID) spectra. This process assists in the determination of metabolite structures but often results in a number of possible positional isomers that have identical molecular ions and fragment ion spectra; this is true particularly for the addition of a single oxygen atom, the most frequently occurring biotransformation. Definitive proof of a structure often requires further characterization such as total synthesis and/or NMR. The latter requires metabolite generation followed by a laborious isolation. For marketed drugs, metabolites can be important for forensic or environmental reasons but are typically not available as reference standards. Here, we propose that standardized chromatographic behavior (*i.e.*, LC Retention Time (RT)) reflects certain physicochemical properties of a metabolite, that if considered together with MS information, can allow for greater insight into a metabolite's structure.

The prediction of RT from structure when authentic standards are unavailable is becoming more common in the fields of natural products, environmental contamination, metabolomics and forensics [2-7]. Firstly, new LC protocols have made RTs more reliable between runs, between labs and even with changing gradients and instruments [8, 9]. Annotating compound databases with RT has regained popularity, as 20 years ago this was commonplace before the advances in electrospray MS [10, 11]. Secondly, a maturing

*Address correspondence to this at the Department of Gastroenterology and Hepatology, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, California, 94305, USA; E-mail: bfitch@stanford.edu

1874-0758/18 $58.00+.00 

understanding of reverse phase LC retention mechanisms allows for reliable Quantitative Structure–Retention Relationship (QSRR) modeling between calculated molecular descriptors and experimental RTs of authentic compounds [7, 12]. The QSRR concept has been applied to drug metabolite identification, as is reported in the literature [13-17], and software packages include this type of modeling as a feature [12]. As an example, Herre and Pragst [15] proposed a biotransformation RT shift using data from C8 column chromatography to report 55 parameters for 29 biotransformations. Despite these advances, the QSRR concept has not really taken hold in the routine workflow of metabolite identification efforts in drug discovery.

One obstacle in RT prediction is that, unlike past experiences with gas chromatography [18], RTs for reverse phase LC tend to be less reproducible due to subtle effects of pH, underivatized silanols and mobile phase compositions; however, the robustness of RT assessment can be improved with RT indexing, for example using nitroalkanes [3], or amides [8] or drugs [11]. Valko's Chromatographic Hydrophobicity Index (CHI) [19] is attractive and uses language familiar to drug metabolism scientists. The CHI roughly approximates the percent acetonitrile that is required in a gradient for a solute to be equally distributed between stationary and mobile phases. This system converts RT to a CHI value, by calibrating a given linear gradient method with a set of drug-like CHI standards. CHI values are calculated using Equation 1, where a and b are constants.

$$RT=a(CHI)+b \quad (1)$$

As with other quantitative structure-property predictions, we assume functional group additivity and define a $CHI_{bt}$ as the change in chromatographic hydrophobicity due to a biotransformation.

$$CHI_{bt} = CHI_{metabolite} - CHI_{parent} \quad (2)$$

When conducting chromatographic separations of metabolites, bioanalytical chemists have largely settled on the use of reverse phase columns (mainly C18) with a combination of acetonitrile and water acidified with formic acid as the mobile phase. We, therefore, concentrated our efforts on the use of C18 columns in the determination of LC RT in this study. The theoretical basis of reverse phase LC retention has been recently reviewed [20]. Snyder and associates proposed a hydrophobic subtraction model for reverse phase LC selectivity [21]. Their theory relates the RTs of two compounds to their physical properties and to corresponding properties of the column, with hydrophobicity as the main retention mechanism in reverse phase, especially in C18 columns. Secondary factors are steric interactions (which prevent insertion of a solute into the hydrocarbon chains of the stationary phase), hydrogen bonding (between donor and acceptor in the solute as well as between donor and acceptor in the stationary phase) and ion exchange interactions between a charged solute and the column.

The model developed by Snyder *et al*. is used primarily for comparing LC column behaviors, and not solute behavior. In addition, their predictions of solute RT are more descriptive than quantitative. However, the selectivity parameters from different columns show the similarities shared by many of our favorite columns (Table **1**). As can be seen in Table **1**, the values of hydrophobicity are one to two orders of magnitude larger than most of the other parameters, indicating the importance of the hydrophobic component in chromatographic separation. The ion exchange term is relevant mainly for cationic solutes (bases, quaternary amines) interacting with ionized silanols from the column. Working at an acidic pH has become the preferred LC-MS method (pH 2.8; 0.1% formic acid) because silanols remain neutralized, thus minimizing ion exchange effects. For most type B silica C18 columns, the ion exchange term is relatively small and unchanged by metabolism; however, this term must be considered when a drug's basic site is affected by metabolism (*N*-oxidation, *N*-acetylation). The last column of Table **1** shows the parameters for a type A silica column. Literature data from such a column may be very different in terms of both hydrogen bonding and ion exchange and will not be included in this analysis.

## 2. EXPERIMENTAL SECTION

### 2.1. Measurement of $CHI_{bt}$

Triclocarban, imipramine and atorvastatin were incubated with human liver microsomes using a standard proto-

**Table 1. Selectivity parameters used in the hydrophobic subtraction model for various C18 reverse phase HPLC columns [21, 45].**

| Column Type | Hydrophobicity | Steric Interaction | Column Acceptor Hydrogen Bonding | Column Donor Hydrogen Bonding | Ion Exchange (pH 2.8) |
|---|---|---|---|---|---|
| Zorbax Eclipse Plus C18 | 1.03 | 0.007 | –0.072 | –0.02 | –0.004 |
| Luna C18 | 1.018 | –0.02 | 0.07 | 0.008 | –0.36 |
| Polaris C18A | 0.929 | 0.007 | –0.227 | 0.062 | 0.149 |
| ACE C18 | 1 | 0.026 | –0.095 | –0.006 | 0.143 |
| Hypersil Gold C18 | 0.881 | –0.002 | –0.017 | 0.036 | 0.162 |
| Inertsil ODS-2 | 0.994 | 0.032 | –0.045 | –0.005 | –0.116 |
| Xterra MS C18 | 0.985 | 0.012 | –0.141 | –0.014 | 0.133 |
| Zorbax C18 (type A silica) | 1.089 | 0.055 | 0.474 | 0.060 | 1.489 |

**Table 2. List of compounds studied. New data was generated for the first 3. Literature data was utilized for the latter 17.**

| | | |
|---|---|---|
| Atorvastatin | Buspirone | Glibenclimide |
| Imipramine | ZLN005 | Terfenadine |
| Triclocarban | Rivoglitazone | Cisapride |
| RO9237 | Clemizole | Diltiazem |
| Loperamide | Axitinib | Thioridazine |
| Pioglitizone | Quetipine | Dextromethorphan |
| 4-Methoxy-alpha PVP | Nefazodone | |

col. The metabolites were identified by LC/MS/MS on a Thermo Fisher Orbitrap Velos and comparison of MS/MS spectra and RT to literature values. The Valko CHI standards were run with the same chromatography. All RT data was pasted in an Excel (2007) spreadsheet (Microsoft, WA; Supplemental Data, S-1) along with reference information, structures and details of each chromatographic method. Equation 1 was used with the literature CHI values for standards at pH 2.6 to calculate slope and intercept for conversion of each metabolite's RT to a CHI value.

### 2.2. Theory for Production of a QSRR for Estimation of $CHI_{bt}$ from Literature RT Data

Equation 3 is used to relate CHI values to the octanol–water partition coefficient (logP) [22] for neutral molecules. Pharmaceutical research departments use this concept to estimate hydrophobicity in a high throughput laboratory using LC gradients [19, 22-25].

$$\log P=0.047(CHI^{n})+0.36(HBC)-1.1 \quad (3)$$

HBC is the hydrogen bond donor count, obtained by summing up all the NH and OH groups in a molecule and the superscript n shows that this only applies for neutrals. An increase in the hydrogen bond acidity of the solute decreases LC retention, whereas hydrogen bond acidity does not practically affect logP. This is because reversed phase C18 stationary phases have much less hydrogen bonding than do acetonitrile–water mobile phases, and octanol has almost the same level of hydrogen bonding as water [26]. Thus, a correction term for solute acidity is needed to get a valid relationship between an LC retention parameter and logP.

Equation 4 replaces logP with clogP (calculated logP) and solves for $CHI^{n}$.

$$CHI^{n}=21.3\text{clogP}-7.67HBC+23.4 \quad (4)$$

Most drugs are charged, however, and we must, therefore, consider the effect of charge on hydrophobicity and on ion exchange properties. At pH 2.8 (the pH of 0.1% formic acid), most carboxylic acids are protonated hence not charged, and therefore neutral (especially at a high percent of acetonitrile [27]), but amines will be protonated, and therefore positively charged. Predicting the hydrophobicity of charged molecules is very complex. The RT of a charged molecule is sensitive to temperature, counterions, pH and mobile phase composition. To take into account the presence of charged molecules, we introduce a compound specific variable, $F_p$, to include those charged molecule stationary phase interactions that are not incorporated into HBC or clogP or are miscalculated by the simplified equation 4.

$$CHI_p=CHI^{n}_p+F_p \quad (5)$$

Again we assume additivity and define an $F_{bt}$ so that the

$$CHI_m=CHI^{n}_m+F_p+F_{bt} \quad (6)$$

For the time being, we will not focus on Phase 2 conjugation reactions such as sulfonation, glucuronidation and glutathione adduction that introduce new charge states to a molecule. We will also ignore major dealkylations or amide cleavages, which dramatically alter a molecule's weight and/or charge state. With these limitations, the parent molecule will often mimic the charge-related attributes of the metabolite, so that for some metabolites, the $F_{bt}$ value of the metabolite will be zero. Substituting equation 5 back into equation 1, yields equation 7.

$$RT_p=a(CHI^{n}_p+F_p)+b \quad (7)$$

If one or more metabolites,$m_1$, have $F_{bt}=0$, then equation 8 will hold.

$$RT_{m1}=a(CHI^{n}_{m1}+F_p)+b \quad (8)$$

We can take the fixed Fp out of the parenthesis and combine it with b to make a new constant b`$=aF_p+b$. generating equation 9 for the parent, 10 for the well-behaved metabolite and 11 for the general case

$$RT_p=a(CHI^{n}_p)+b' \quad (9)$$

$$RT_{m1}=a(CHI^{n}_{m1})+b' \quad (10)$$

$$RT_m=a(CHI^{n}_m+F_{bt})+b' \quad (11)$$

### 2.3. Detailed Procedure for Applying the QSRR to Literature RT Data

For each of the 20 examples (the 3 compound datatsets from section 2.1 and 17 literature datasets; all 20 compounds shown in Table **2**), the structure was redrawn in MoKa (Molecular Discovery; version 2.6.4) [28], and clogP was calculated and added to the spreadsheet. HBC was determined by manual inspection as the total number of NH and OH bonds. $CHI^{n}$ was calculated using equation 4.

In most cases, the parent molecule and a metabolite with unchanged nitrogen atoms and no new major intramolecular hydrogen bonds provide at least two reference points to establish a linear correlation between RT and $CHI^{n}$ and gener-

ate a reference fitting line. The slope, a, and y-intercept of the linear correlation, b`, were then used to calculate the $CHI^n+F_{bt}$ for each molecule with equation 11. Cases in which no metabolite had $F_{bt}=0$ were not considered. Finally substituting in equations 5 and 6 generates equation 12. The unknown $F_p$ values are eliminated giving equation 13 as the method we used to estimate $CHI_{bt}$.

$$CHI_{bt}=(CHI^n_m+F_p+F_{bt})-(CHI^n_p+F_p) \quad (12)$$

$$CHI_{bt}=(CHI^n_m+F_{bt})-(CHI^n_p) \quad (13)$$

Many metabolism projects resulted in multiple $CHI_{bt}$ results for the same biotransformation. These were averaged in each spreadsheet.

## 3. RESULTS AND DISCUSSION

### 3.1. Measurement of $CHI_{bt}$

A single dedicated laboratory can achieve highly reproducible LC RTs [14]; however, interlaboratory reproducibility of RT is not so simple [29], demanding scrupulous attention to mobile phase, temperature, LC pump dead volume, if not identical column manufacture. We chose to evaluate the CHI retention index system to produce reproducible $CHI_{bt}$'s. We ran the 10 recommended standards and found that 3 coeluted (colchicine, 8-phenyltheophylline and acetophenone) at pH 2.8. We replaced the former two with benzamide [19, 22] to get better coverage. After conversion of RT to CHI for parent and identified metabolite the calculation of $CHI_{bt}$ is obvious. We chose 3 molecules (atorvastatin, imipramine and triclocarban) to incubate with microsomes in-house to generate metabolites and placed the results in the Supplemental Excel file sheets 2-4. Once enough molecules have been tested a table of $CHI_{bt}$ values such as Table **3** can be used to estimate where in a chromatogram a predicted biotransformation product would elute. Note that of the three molecules we tested, triclocarban yields a metabolite with a CHI value of 150, well outside the linear portion of the CHI calibration curve. It might be valuable to extend the set of CHI reference standards to cover this range, but such an intensely hydrophobic drug metabolite is rare.

### 3.2. Prediction of Structure for Unknown Metabolites with Known $CHI_{bt}$

Many labs will not have the CHI system or wish to evaluate literature RT data, so we developed a QSRR. To generate an initial table of data we chose an additional 17

**Table 3. Changes to CHI for the compounds used in the study following different biotransformations (n is the number of occurrences in the supplemental file).**

| Biotransformation | Average $CHI_{bt}$ | Standard Deviation | n |
|---|---|---|---|
| *N*-Oxidation tertiary amine | 4 | 3 | 8 |
| -secondary amine | 32 | | 1 |
| Aromatic oxidation | -17 | 4 | 8 |
| -with HB | 4 | 10 | 2 |
| -heteroaromatic | -7 | 1 | 2 |
| Aliphatic oxidation | -30 | 6 | 6 |
| -with HB | -15 | 5 | 5 |
| *N*-demethylation- basic amine | -2 | 2 | 5 |
| -amide | -15 | | 1 |
| Aromatic O-demethylation | -18 | 1 | 3 |
| Aliphatic O-demethylation with HB | -8 | | 1 |
| Alcohol oxidation to aldehyde | 12 | | 1 |
| Alcohol oxidation to ketone | 6 | 5 | 2 |
| Alcohol oxidation to acid | -1 | 6 | 4 |
| Sulfide oxidation to sulfoxide | -36 | 25 | 3 |
| Sulfoxide oxidation to sulfone | 17 | 10 | 1 |
| Acid dehydration to lactone | 8 | | 1 |
| Primary amine oxidation to alcohol | 53 | | 1 |
| Cyclic amine oxidation to iminium | 9 | 23 | 2 |
| Cyclic amine oxidation to lactam | 29 | 31 | 2 |

molecules from the literature to test our RT prediction method. These literature reports utilized reverse phase LC with acetonitrile and acidic mobile phase and most undergo mono-oxidation biotransformations. The need for better methods of exactly specifying the site of mono-oxidation biotransformation has been pointed out in a recent review [30]. The complete results for all compounds are presented in the Supplemental Material, tabs 2-21, but the process is illustrated here with atorvastatin.

### 3.3. Atorvastatin

Fig. (**1**) illustrates the structure determination process. It shows the extracted m/z 575.255 for the mono-hydroxylated metabolites plotted to a $CHI_{bt}$ axis. The colored bars take the average $CHI_{bt}$ for the important carbon mono-oxidations and set a bandwidth at one standard deviation. MS/MS had shown that the main peaks are results of hydroxylation of the acylaniline function. Comparison to the bars indicates that the early eluter must be the non-hydrogen bonded *meta* or *para* isomer while the late eluter must be the hydrogen bonded *ortho* isomer. In solution, atorvastatin is present in both free acid and lactone forms, so a total of six major peaks (supplemental file, Table **2**) are observed in the LC-MS chromatogram following human liver microsome incubation [31]. The parent lactone is readily identified. The identification of the two hydroxylated lactones is simple in that they elute in the same relative pattern as the unlactonized forms. The effects of dual biotransformations that occur on separate portions of a drug, for example, the lactone formation and aromatic hydroxylation of atorvastatin, are additive. Importantly, the difference in LC RT values allows for the assignment of regiochemistry as *ortho* or *meta/para* in this case.

The application of the QSRR prediction is illustrated in the supplemental file, Table **2**. The initial correlation between RT and $CHI^n$ is fairly poor (first graph, $r^2$=0.49). Both the *ortho* hydroxylation and the lactonization effect intramolecular H-bonding, not reflected in clogP or the nominal HBC count. Excluding the *ortho*-hydroxy point and the 3 lactones allows a two point "corrected line" (middle graph). Applying this new slope and intercept to the RT values for the others allows for the calculation of corrected $CHI^n+F_{bt}$ (column g). $CHI_{bt}$ is then calculated by subtracting the column g value for the metabolite from that of the appropriate precursor.

When atorvastatin and metabolites were analyzed inhouse with the CHI standards the $CHI_{bt}$ results were quite comparable to the predicted results (ortho-hydroxylation, predicted -2.9, observed -3.6; para hydroxylation predicted -16.4 observed -20.6; lactonization, predicted 6.2, observed 7.7). Similarly, comparable results were observed for imipramine and triclocarban.

All the changes in CHI due to metabolism in the compounds used in this study are collected in Table **1** of the supplemental file. Table **3** summarizes the average $CHI_{bt}$ for each biotransformation along with the standard deviations and n values. Where we had measured $CHI_{bt}$ results and predicted results, Table **3** shows only the measured results. The effects of different functional group biotransformations on CHI are discussed below.

### 3.4. Aliphatic C-oxidation

Aliphatic C-oxidation is typically distinguished from aromatic C-oxidation by the loss of water fragment ion in the CID spectrum. The $CHI_{bt}$ is typically more negative for aliphatic C-oxidation than for aromatic C-oxidation or N-oxidation. However, subtle differences in chemical structure cannot be easily predicted. A literature report [32] describes two side-chain hydroxylated derivatives of pioglitizone with similar CID spectra, and so synthesis and chemical derivatization were required to differentiate between the two isomers. If only one isomer was present, identifying it by RT prediction would be impossible since the prediction error

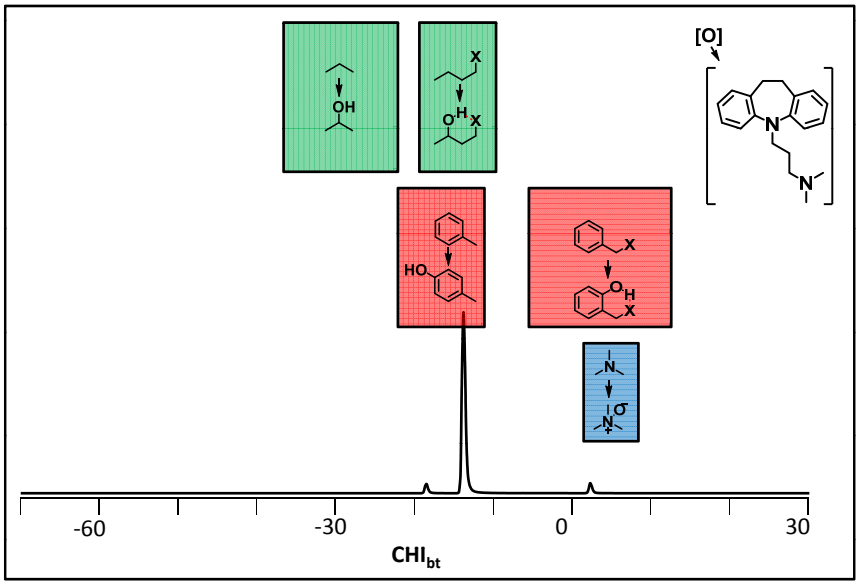


**Fig. (1).** The extracted m/z 575.255 for the mono-hydroxylated metabolites of atorvastatin plotted to a $CHI_{bt}$ axis. The colored bars take the average $CHI_{bt}$ for the important carbon mono-oxidations and set a bandwidth at 1 standard deviation. Comparison to the bars indicates that the early eluter must be the non-hydrogen bonded *meta* or *para* isomer while the late eluter must be the hydrogen bonded *ortho* isomer.

would be greater than the difference in RT. In this case, clogP cannot distinguish the isomers which show a slightly lower hydrophobicity for the primary alcohol *vs*. the secondary. This difference in hydrophobicity is reflected in a lower RT for the primary alcohol. The example of pioglitazone (and the following example, glibenclimide) demonstrates a general rule: with two closely related hydroxylation isomers, the one with the new hydroxyl that is furthest from the unchanged polar part of the parent drug will elute the earliest. This observation is intuitive in that the new hydroxyl group would disrupt the insertion of the more hydrophobic part of the metabolite into the C18 chains of the stationary phase.

Glibenclimide contains a mono acylamino substituted cyclohexyl ring which was extensively oxidized. The 4-trans isomer showed the largest $CHI_{bt}$ while the values for the 4-cis,3-trans and 3-cis isomers were intermediate. The 2-trans isomer had the lowest $CHI_{bt}$, likely due to increased intramolecular hydrogen bonding. These results are not reflected in clogP and HBC. Using a complex QSRR combining descriptive parameters and Canvas fingerprints Falchi *et al*. [14] could accurately predict the relative retention of 2-, 3- and 4-hydroxycyclohexylamines but could not distinguish 4-cis *vs*. 4-trans isomers because they only used 2D descriptors. Herre and Pragst [15] describe 4 isomers of hydroxytestosterone which are well separated on LC but have comparable predicted hydrophobicities. These stereochemical issues require modeling of minimized conformational structures to calculate their hydrophobicities more accurately. Such techniques are not typically available for high throughput applications by non-specialists.

Buspirone is metabolized to four mono-hydroxylated isomers. The *N*-oxidized metabolite has the longest RT, while the shortest RT is the result of a remote aliphatic oxidation. The two intermediate isomers are due to an oxidation of a pyrimidine and an aliphatic oxidation next to a carbonyl, which creates an intramolecular hydrogen bond.

### 3.5. Aromatic C-oxidation

Determining the regiochemistry of aromatic C-oxidation is not typically possible by MS/MS. Ion mobility MS has been used for distinguishing regioisomeric metabolites [33, 34], but it is not widely available. As demonstrated herein for atorvastatin and triclocarban, LC retention is a simple alternative that can often identify the aromatic substitution of a mono-hydroxylated metabolite. In particular, *ortho*-oxidation will yield a much longer RT than *meta*- or *para*-oxidations if hydrogen bonding is possible. *Ortho*-oxidation leading to intramolecular hydrogen bonding decreased hydrophobicity in many other substituted phenols (nitro, methoxy, catechol, acetophenone, benzamide) [35-38]. Relative retention of the isomers of acetaminophen [39] also follows the "furthest from the unchanged polar part of the parent drug rule" with the relative retention in the order *para*<*meta*<*ortho*.

We report two examples of heteroaromatic oxidation, a benzimidazole and a pyrimidine. These often occur with aldehyde oxidase involvement in metabolic clearance. These cases are hard to predict and possibly not well modeled by clogP. But in general, this should decrease CHI less than phenyl oxidation. As with aliphatic oxidation, there appears to be a general correlation between CHI and the distance of the new hydroxyl group to the next most polar part of the molecule. Greater adoption of the simple rules that relate RT to metabolite structure might obviate the need for compound synthesis to prove the structures of all mono-hydroxylated aromatic metabolites.

### 3.6. Oxidation of Amines

Conversion of basic amines to hydroxylamines or *N*-oxides typically causes an increase in retention at pH 2.8 [40]. As illustrated with nine of our examples, this increased RT distinguishes them from most C-oxidations. The related aliphatic oxidation of the carbon next to a basic cyclic amine, creating an aminol, leads to a minor change in CHI.

### 3.7. Oxidation of Sulfur Compounds

Our table does not contain many sulfur oxidations but the $CHI_{bt}$ for conversion of a sulfide to a sulfoxide is the largest we have observed. Subsequent oxidation to the sulfone has a smaller positive CHI effect.

### 3.8. *N,O*-Demethylation

Demethylation is a very common biotransformation pathway. The removal of an *N*-methyl group from a basic amine causes little change to its RT. These metabolites should be much more polar than the parent according to clogP and HBC. Perhaps the loss of the methyl group increases access to charged silanols with a subsequent increase in ion exchange interaction. Loss of a methyl group from an amide (loperamide) or a methyl ether (4-methoxy-α-PVP, rivoglitazone, diltiazam) has the expected decrease in LC retention due to new polar functional interactions with the solvent. However, O-demethylation of cisapride creates a new intramolecular hydrogen bond and a smaller $CHI_{bt}$.

### 3.9. Other Biotransformation Reactions

Other biotransformation reactions are not the focus of this paper- their structures will typically be obvious from the mass change or yield to MS/MS analysis. Biotransformation which involves loss of the positively charged amino group will be especially hard to model but will likely lead to an increase in retention time.

## 4. DISCUSSION

Note that there is considerable variability in the $CHI_{bt}$ values for a particular biotransformation but some general conclusions can be made:

- Aliphatic oxidation most remote from the other polar functionalities causes the greatest loss of CHI.
- Aromatic oxidation is generally a smaller CHI effect than aliphatic.
- Nitrogen atom oxidation typically increases CHI.
- Biotransformation which introduces an intramolecular hydrogen bond results in a smaller change in CHI than that predicted by clogP.

- The CHI effects of multiple biotransformations are additive.

The graphical abstract takes the average $CHI_{bt}$ for the important mono-oxidations and sets a bandwidth at one standard deviation. The extracted ion chromatogram for the mono-oxidation metabolites of imipramine is overlaid. There is 1 dominant and two minor metabolites. These have $CHI_{bt}$ values of 2.3,-14.1 and -18.9 (supplemental Excel file, worksheet 4). Inspection of the color coded $CHI_{bt}$ regions would indicate that the first is very likely N-oxidation while the latter two could be hydrogen bonded aliphatic or aromatic oxidation. These hypotheses can then be taken into the evaluation of the CID spectra [41], simplifying the interpretations.

The prediction of RT would benefit from further improvements to QSRR modeling [24]. The simple rules described here can supplement common metabolite identification methods to substantiate the results and minimize the need to prove a compound's structure by scale-up incubation plus NMR analysis or chemical synthesis.

## CONCLUSION

Tools for predicting the MS/MS [42] spectra and RT behavior of drug metabolites will allow for the creation of a database of possible structures. A similar development occurred in proteomics, where the process of identifying peptides by comparing theoretically generated predicted MS/MS spectra and RTs obviated the need for unavailable reference standards. The prediction of LC RT can greatly aid in structure determination in related fields such as untargeted metabolomics, natural products, forensics and environmental analysis [43]. The prediction of electrospray LC-MS/MS sensitivity [44] when reference standards are unavailable is daunting, but knowledge of the percent acetonitrile at elution (one interpretation of the CHI value) could assist such a calculation.

Recent developments with retention indexing schemes raise the hope of interlaboratory reproducibility of reverse phase retention behavior. True interlaboratory reproducible reverse phase behavior could supply compound identification criteria useful for cases in which species have identical mass spectral behavior. Much more effort is needed in defining the protocols for measuring reverse phase retention reproducibly and for predicting RT from a structure. One advantage we have in this work is that an additional standard, the parent drug, is present in all samples, thus giving more confidence in the reproducibility of the $CHI_{bt}$ values. We hope to extend this approach to other areas (such as identification of glucuronidation position) where mass spectrometry cannot always provide sufficient structure information. We have herein proposed a paradigm in which a metabolite's mass is first defined *via* high-resolution MS, followed by a prediction of LC RTs for the possible structures formed to assist in the determination of the exact metabolite structure.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

## HUMAN AND ANIMAL RIGHTS

No Animals/Humans were used for this research.

## CONSENT FOR PUBLICATION

Not applicable.

## CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

## ACKNOWLEDGEMENTS

Thanks to Agilent, Santa Clara Labs for providing the LC-MS instrumentation used for loperamide metabolite identification.

## SUPPORTING INFORMATION

Excel file of LC retention predictions (22 worksheets).

Supplementary material is available on the publisher's web site along with the published article.

## REFERENCES

[1] Wen, B.; Zhu, M. Applications of mass spectrometry in drug metabolism: 50 years of progress. *Drug Metab. Rev.,* **2015**, *47*, 71-87.

[2] Eugster, P.J.; Boccard, J.; Debrus, B.; Breant, L.; Wolfender, J.; Martel, S.; Carrupt, P. Retention time prediction for dereplication of natural products (CxHyOz) in LC-MS metabolite profiling. *Phytochemistry,* **2014**, *108*, 196-207.

[3] Hall, L.M.; Hill, D.W.; Menikarachchi, L.C.; Chen, M.; Hall, L.H.; Grant, D.F. Optimizing artificial neural network models for metabolomics and systems biology: An example using HPLC retention index data. *Bioanalysis,* **2015**, *7*, 939-955.

[4] Bade, R.; Bijlsma, L.; Miller, T.H.; Barron, L.P.; Sancho, J.V.; Hernandez, F. Suspect screening of large numbers of emerging contaminants in environmental waters using artificial neural networks for chromatographic retention time prediction and high resolution mass spectrometry data analysis. *Sci. Total Environ.,* **2015**, *538*, 934-941.

[5] Barron, L.P.; McEneff, G.L. Gradient liquid chromatographic retention time prediction for suspect screening applications: A critical assessment of a generalised artificial neural network-based approach across 10 multi-residue reversed-phase analytical methods. *Talanta,* **2016**, *147*, 261-270.

[6] Tyteca, E.; Veuthey, J.; Desmet, G.; Guillarme, D.; Fekete, S. Computer assisted liquid chromatographic method development for the separation of therapeutic proteins. *Analyst,* **2016**, *141*, 5488-5501.

[7] Tyteca, E.; Talebi, M.; Amos, R.; Park, S.H.; Taraji, M.; Wen, Y.; Szucs, R.; Pohl, C.A.; Dolan, J.W.; Haddad, P.R. Towards a chromatographic similarity index to establish localized Quantitative Structure-Retention Models for retention prediction: Use of retention factor ratio. *J. Chromatogr. A.,* **2017**, *1486*, 50-58.

[8] Abate-Pella, D.; Freund, D.M.; Ma, Y.; Simon-Manso, Y.; Hollender, J.; Broeckling, C.D.; Huhman, D.V.; Krokhin, O.V.; Stoll, D.R.; Hegeman, A.D.; Kind, T.; Fiehn, O.; Schymanski, E.L.; Prenni, J.E.; Sumner, L.W.; Boswell, P.G. Retention projection enables accurate calculation of liquid chromatographic retention times across labs and methods. *J. Chromatogr. A.,* **2015**, *1412*, 43-51.

[9] Stanstrup, J.; Neumann, S.; Vrhovsek, U. PredRet: Prediction of retention time by direct mapping between multiple chromatographic systems. *Anal. Chem.,* **2015**, *87*, 9421-9428.

[10] Bogusz, M.; Hill, D.W.; Rehorek, A. Comparability of RP-HPLC retention indices of drugs in three databases. *J. Liq. Chromatogr. Relat. Technol.,* **1996**, *19*, 1291-1316.

[11] Elliott, S.P.; Hale, K.A. Applications of an HPLC-DAD drug-screening system based on retention indices and UV spectra. *J. Anal. Toxicol.,* **1998**, *22*, 279-289.

[12] McEachran, A.D.; Mansouri, K.; Newton, S.R.; Beverly, B.E.J.; Sobus, J.R.; Williams, A.J.A comparison of three liquid chromatography (LC) retention time prediction models. *Talanta,* **2018**, *182*, 371-379.

[13] Noorizadeh, H.; Sobhan-Ardakani, S.; Raoofi, F.; Noorizadeh, M.; Mortazavi, S.S.; Ahmadi, T.; Pournajafi, K. Application of artificial neural network to predict the retention time of drug metabolites in two-dimensional liquid chromatography. *Drug Test. Anal.,* **2013**, *5*, 315-319.

[14] Falchi, F.; Bertozzi, S.M.; Ottonello, G.; Ruda, G.F.; Colombano, G.; Fiorelli, C.; Martucci, C.; Bertorelli, R.; Scarpelli, R.; Cavalli, A.; Bandiera, T.; Armirotti, A. Kernel-Based, partial least squares quantitative structure-retention relationship model for UPLC retention time prediction: A useful tool for metabolite identification. *Anal. Chem.,* **2016**, *88*, 9510-9517.

[15] Herre, S.; Pragst, F. Shift of the high-performance liquid chromatographic retention times of metabolites in relation to the original drug on an RP8 column with acidic mobile phase. *J. Chromatogr. B: Biomed. Sci. Appl.,* **1997**, *692*, 111-126.

[16] Baker, J.K. Estimation of high-performance liquid chromatographic retention indexes of glucuronide metabolites. *J. Liq. Chromatogr.,* **1981**, *4*, 271-278.

[17] Valko, K.; Szabo, G.; Rohricht, J.; Jemnitz, K.; Darvas, F. Prediction of retention of metabolites in high-performance liquid chromatography by an expert system approach. *J. Chromatogr.,* **1989**, *485*, 349-363.

[18] d'Acampora Z.B.; Bicchi, C.; Dugo, P.; Rubiolo, P.; Dugo, G.; Mondello, L. Linear retention indices in gas chromatographic analysis: a review. *Flavour Fragrance J.,* **2008**, *23*, 297-314.

[19] Valko, K.; Bevan, C.; Reynolds, D. Chromatographic hydrophobicity index by fast-gradient RP-HPLC: A high-throughput alternative to Log P/log D. *Anal. Chem.,* **1997**, *69*, 2022-2029.

[20] Sykora, D.; Vozka, J.; Tesarova, E. Chromatographic methods enabling the characterization of stationary phases and retention prediction in high-performance liquid chromatography and supercritical fluid chromatography. *J. Sep. Sci.,* **2016**, *39*, 115-131.

[21] Snyder, L.R.; Dolan, J.W.; Carr, P.W. The hydrophobic-subtraction model of reversed-phase column selectivity. *J. Chromatogr. A.,* **2004**, *1060*, 77-116.

[22] Valko, K.; Du, C.M.; Bevan, C.; Reynolds, D.P.; Abraham, M.H. Rapid method for the estimation of octanol/water partition coefficient (Log Poct) from gradient RP-HPLC retention and a hydrogen bond acidity term (Î£Î±2H). *Curr. Med. Chem.,* **2001**, *8*, 1137-1146.

[23] Valko, K.L. Lipophilicity and biomimetic properties measured by HPLC to support drug discovery. *J. Pharm. Biomed. Anal.,* **2016**, *130*, 35-54.

[24] Subirats, X.; Roses, M.; Bosch, E. High-throughput log Po/w determination from UHPLC measurements: Revisiting the chromatographic hydrophobicity index. *J. Pharm. Biomed. Anal.,* **2016**, *127*, 26-31.

[25] Camurri, G.; Zaramella, A. High-Throughput liquid chromatography/mass spectrometry method for the determination of the chromatographic hydrophobicity index. *Anal. Chem.,* **2001**, *73*, 3716-3722.

[26] Roses, M.; Bosch, E.; Rafols, C.; Fuguet, E. Chromatographic Hydrophobicity Index (CHI). *Adv. Chromatogr.,* **2012**, *50*, 377-414.

[27] Suu, A.; Jalukse, L.; Liigand, J.; Kruve, A.; Himmel, D.; Krossing, I.; Roses, M.; Leito, I. Unified pH values of liquid chromatography mobile phases. *Anal. Chem.,* **2015**, *87*, 2623-2630.

[28] Milletti, F.; Storchi, G.; Cruciani, G. New and original pKa prediction method using Grid Molecular Interaction fields. *J. Chem. Inf. Mod.,* **2007**, *47*, 2172-2181.

[29] Boswell, P.G.; Abate-Pella, D.; Hewitt, J.T. Calculation of retention time tolerance windows with absolute confidence from shared liquid chromatographic retention data. *J. Chromatogr. A.,* **2015**, *1412*, 52-58.

[30] Ponting, D.J.; Murray, E.; Long, A. Quantifying confidence in the reporting of metabolic biotransformations. *Drug Discov. Today,* **2017**, *22*, 970-975.

[31] Park, J.; Kim, K.; Bae, S. K.; Moon, B.; Liu, K.; Shin, J. Contribution of cytochrome P450 3A4 and 3A5 to the metabolism of atorvastatin. *Xenobiotica,* **2008**, *38*, 1240-1251.

[32] Liu, D.Q.; Hop, C.E.C.A. Strategies for characterization of drug metabolites using liquid chromatography-tandem mass spectrometry in conjunction with chemical derivatization and on-line H/D exchange approaches. *J. Pharm. Biomed. Anal.,* **2005**, *37*, 1-18.

[33] Reading, E.; Munoz-Muriedas, J.; Roberts, A.D.; Dear, G.J.; Robinson, C.V.; Beaumont, C. Elucidation of drug metabolite structural isomers using molecular modeling coupled with ion mobility mass spectrometry. *Anal. Chem.,* **2016**, *88*, 2273-2280.

[34] Shimizu, A.; Chiba, M. Ion mobility spectrometry-mass spectrometry analysis for the site of aromatic hydroxylation. *Drug Metab. Dispos.,* **2013**, *41*, 1295-1299.

[35] Goetz, G.H.; Farrell, W.; Shalaeva, M.; Sciabola, S.; Anderson, D.; Yan, J.; Philippe, L.; Shapiro, M.J. High throughput method for the indirect detection of intramolecular hydrogen bonding. *J. Med. Chem.,* **2014**, *57*, 2920-2929.

[36] Kuhn, B.; Mohr, P.; Stahl, M. Intramolecular hydrogen bonding in medicinal chemistry. *J. Med. Chem.,* **2010**, *53*, 2601-2611.

[37] Clark, M.R.; Garcia-Roura, L.E.; Clark, C.R. Intramolecular hydrogen bonding effects on the reversed-phase retention of substituted acetophenones. *J. Liq. Chromatogr.,* **1988**, *11*, 3213-3221.

[38] Holmes, H.L.; Lough, C.E. Effect of intramolecular hydrogen bonding on partition coefficients. *U. S. NTIS, AD Rep.* **1976**, 24.

[39] Yamauchi, S. Retention indexes of phenols for internal standards in reversed-phase high-performance liquid chromatography. Application to retention prediction and selectivities of mobile phases and packing materials. *J. Chromatogr.,* **1993**, *635*, 61-70.

[40] Qi, X.; Wu, B.; Cheng, Y.; Qu, H. Simultaneous characterization of pyrrolizidine alkaloids and N-oxides in Gynura segetum by liquid chromatography/ion trap mass spectrometry. *Rapid Commun. Mass Spectrom.,* **2009**, *23*, 291-302.

[41] Van Breemen, R.B.; Nikolic, D.; Bolton, J.L. Metabolic screening using online ultrafiltration mass spectrometry. *Drug Metab. Dispos.,* **1998**, *26*, 85-90.

[42] Wright, P.; Alex, A.; Pullen, F. Predicting collision-induced dissociation spectra: Semi-empirical calculations as a rapid and effective tool in software-aided mass spectral interpretation. *Rapid Commun. Mass Spectrom.,* **2014**, *28*, 1127-1143.

[43] Fitch, W.L. A Manual approach to the CASMI 2014 challenges and a review of the isomer problem. *Curr. Metabolomics,* **2017**, *5*, 40-44.

[44] Broeckling, C.D.; Ganna, A.; Layer, M.; Brown, K.; Sutton, B.; Ingelsson, E.; Peers, G.; Prenni, J.E. Annotation from the MS1-spectrum and time predictions approach to LC-MS metabolomics data. *Anal. Chem.,* **2016**, *88*, 9226-9234.

[45] Carr, P.W.; Dolan, J.W.; Neue, U.D.; Snyder, L.R. Contributions to reversed-phase column selectivity. I. Steric interaction. *J. Chromatogr. A.,* **2011**, *1218*, 1724-1742.