Computational Methods to Predict the Regioselectivity of Electrophilic Aromatic Substitution Reactions of Heteroaromatic Systems

Monika Kruszyk,†§ Mikkel Jessing,† Jesper Langgaard Kristensen,§ and Morten Jørgensen*†

†Discovery Chemistry and DMPK, Neuroscience Drug Discovery, H. Lundbeck A/S, Ottiliavej 9, 2500 Valby, Denmark
§Department of Drug Design and Pharmacology, Faculty of Health and Medical Science, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen Ø, Denmark

ABSTRACT: The validity of calculated NMR shifts to predict the outcome of electrophilic aromatic substitution reactions on different heterocyclic compounds has been examined. Based on an analysis of >130 literature examples, it was found that the lowest predicted $^{13}$C and/or $^1$H chemical shift of a heterocycle correlates qualitatively with the regiochemical outcome of halogenation reactions in >80% of the investigated cases. In the remaining cases, the site of electrophilic aromatic substitution can be explained by the calculated HOMO orbitals obtained using density functional theory. Using a combination of these two methods, the accuracy increases to >95%.

1. INTRODUCTION

Heterocyclic compounds are important building blocks used in many fields of chemistry, including drug discovery and material science. Their halogenated derivatives are widely utilized as substrates in reactions such as C–C bond forming reactions (e.g., Pd-catalyzed cross-coupling) and as valuable cores in medicinal chemistry structure–activity relationship (SAR) explorations. There are several ways to introduce halogen substituents into heterocyclic rings, with electrophilic aromatic substitution (EAS) being one of the most widely applied, especially for electron-rich substrates.

The general mechanism of EAS (Scheme 1) involves the formation of a π-complex between the substrate and the electrophile.

Scheme 1. General Mechanism of EAS

Next, in the rate-determining step, the Wheland σ-complex is generated. In order to restore the aromaticity, this highly reactive complex is quickly deprotonated to afford the final substitution product. The relative reactivity and the regiochemical outcome of EAS on benzene derivatives is common textbook knowledge. Conversely, the regioselectivity of EAS on aromatic heterocycles is in many cases not straightforward to predict, particularly for molecules which contain multiple heterocyclic rings.

The ability to predict and understand reactivity to guide chemical synthesis has been a long-sought, as yet elusive, objective in the field of organic synthesis. To predict the site of EAS, it is important to determine the most nucleophilic center of the substrate. Early in the 1950s, Fukui et al. introduced the concept of frontier molecular orbitals theory as a tool to predict the reactivity. The concepts based on the hard and soft acids and bases principle (HSAB), activation hardness, reactive hybrid orbitals (RHO), local nucleophilicity, and Fukui indices as well as average local ionization energies have also been applied for aromatic systems. Despite being quite successful, these methods generally require specialized software and time-consuming calculations. The use of computers to aid in the planning process was pioneered by Corey as early as 1969. Four decades later Grzybowski published the related computational tool Chematica with a much higher number of known transformations than Corey’s LHASA. Despite the obvious potential, this concept has not had the expected impact on organic chemistry, and it is not broadly available to the scientific community yet.

Our approach in the present study was to develop and validate a very simple method, that can be readily used by synthetic chemists, to plan the synthesis with more confidence using readily available software. The method is based on the predicted NMR shifts, as these values are the result of electronic effects in the molecule.

$^1$H NMR chemical shift values have previously been used by Handy et al. for predicting regioselectivity in the coupling reactions of polyhaloaromatic heterocycles. Attempts to correlate...
the site of EAS with the chemical shifts have been undertaken for a series of substituted benzenes.\textsuperscript{18–20} NMR predictions can easily be accessed directly from the chemical structure using a range of computer programs commonly used by the chemists of different fields. In this work, we used ChemDraw, as the program is widely available. The procedure for estimating the NMR spectra by ChemDraw applies the linear additive rules, which were introduced in the early 1960s for the simple alkanes and were subsequently expanded to other classes of organic compounds.\textsuperscript{21,22} First, the program determines key substructures of a molecule, for which the base NMR shifts are given. The base values are then modified by the increments, which depend on the nature of a substituent. For the cases that the increment for a substituent cannot be determined, ChemDraw replaces it with the smaller units with the same neighboring atoms. In principle the lowest (upfield) $^{13}$C and/or $^1$H chemical shift ($\delta$) within the aromatic region would indicate the most electron-rich carbon and hence most reactive in EAS. However, the examples of aniline and fluorobenzene indicate the weak side of the predictions based only on the calculated NMR shifts. In contrast to the modest difference in chemical shifts between these two compounds (Figure 1), reactivity toward electrophilic halogenation is obviously higher for aniline than for fluorobenzene.\textsuperscript{23}

Therefore, unsubstituted and deactivated benzenes are not considered as the competing rings for EAS in our analysis, and we limit our predictions to the heterocyclic cores, unless an

Table 1. Five-Membered Heterocycles\textsuperscript{44}

\begin{table}
\begin{tabular}{|c|c|c|c|c|}
\hline
1, 100\% (NBS)\textsuperscript{44} & 2, 85\% (Br)\textsuperscript{35} & 3, 82\% (Br)\textsuperscript{36} & 4, 74\% (SO$_2$Cl)\textsuperscript{37} & 5, 89\% (NBS)\textsuperscript{36} \\
\hline
$^1$H & $^{13}$C & $^{1}$H & $^{13}$C & $^{1}$H & $^{13}$C \\
\hline
F & NH$_2$ & 115.0 & 6.67 & 122.4 & 7.14 \\
124.9 & 7.10 & 123.8 & 7.52 & 132.8 & 7.52 \\
& & 7.32 & 6.75 & 7.32 & 6.75 \\
\hline
\end{tabular}
\end{table}

Figure 1. Predicted NMR shifts for fluorobenzene and aniline. The $^1$H and $^{13}$C values are represented in green and red, respectively.

The predicted NMR values in green and red refer to $^1$H and $^{13}$C shifts, respectively. The green and red circles point to the most reactive site according to the predictions, while the arrow indicates the site of the actual reaction. For the $^1$H spectrum, the variation of 0.1 ppm is accepted, and for $^{13}$C spectrum, the variation of 5 ppm relative to the lowest predicted value is accepted.
electron-rich benzene is present in the molecule. We employed the highest occupied molecular orbital (HOMO) descriptions and displayed them graphically onto the electron density isosurface. The obtained models depict the distribution of the electron density in the HOMO. This can be calculated using a number of quantum chemistry software packages, but not typically performed by synthetic organic chemists. These analyses require significantly more time compared to the NMR predictions. While ChemDraw gives the output in less than a second, the time needed for quantum chemical calculations is in the range from a few to tens of minutes, depending on the complexity of the molecule and the software. Efforts to guide synthesis based on similar calculations have been reported previously.\textsuperscript{24} Herein, we report a review of 50 literature examples (see Supporting Information for additional 80 examples from the literature) of EAS and compare the regiochemical outcome with the chemical shifts predicted by ChemDraw supported by the HOMO maps to provide a set of guidelines to predict the regioselectivity of EAS in molecules containing one or more heteroaromatic rings, either in fused systems or as distinct substructures within the same molecule.

2. METHODS

The $^1$C and/or $^1$H chemical shift predictions have been calculated using ChemDraw from PerkinElmer, version currently available at the time the predictions were performed (14.0 and higher). The $^1$H and $^1$C values

Table 2. Six-Membered Heterocycles

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The predicted NMR values in green and red refer to $^1$H and $^1$C shifts, respectively. The green and red circles point to the most reactive site according to the predictions, while the arrow indicates the site of the actual reaction. For the $^1$H spectrum, the variation of 0.1 ppm is accepted, and for $^1$C spectrum, the variation of 5 ppm relative to the lowest predicted value is accepted.
are represented in green and red, respectively. For the $^1$H spectrum, the variation of 0.1 ppm is accepted, and for $^{13}$C spectrum, the variation of 5 ppm relative to the lowest predicted value is accepted. The circles point to the most reactive site according to the predictions, while the arrow indicates the site of the actual reaction. For simplicity, the values for inactivated and/or deactivated benzenes are omitted. The calculations were obtained using density functional theory (DFT) with B3LYP density functional model performed by the Jaguar in the gas phase (Jaguar, Version 8.9, Schrödinger, LLC, New York, NY, 2015).25−27 The geometries were optimized using the 6-31G** basis set.28−33 The HOMO orbitals were mapped onto electron density isosurface and are considered as the determining factor for the regioselectivity. On the graphical models, red suggests the site of the potential reaction.

**Figure 3.** The predicted NMR values, the HOMO orbitals and HOMO mapped onto electron density isosurface of 2-phenyloxazole 43.

The predicted NMR values in green and red refer to $^1$H and $^{13}$C shifts, respectively. The green and red circles point to the most reactive site according to the predictions, while the arrow indicates the site of the actual reaction. For the $^1$H spectrum, the variation of 0.1 ppm is accepted, and for $^{13}$C spectrum, the variation of 5 ppm relative to the lowest predicted value is accepted.

### 3. RESULTS AND DISCUSSION

#### 3.1. Five-Membered Heterocycles.
Electron-rich five-membered heterocycles are common substrates for EAS.3 As shown in Table 1, a range of 1,2- and 1,3-azoles react with the electrophiles at the carbon characterized by the lowest $^{13}$C and/or $^1$H chemical shift, which is C4 for 1,2-azoles (pyrazoles 1−5, isoxazoles 6−9, isothiazoles 10−13) and imidazoles (14−17) and C5 for thiazoles (18−19).

However, when C5 of thiazole is substituted and the heterocyclic ring is activated with an electron donating group (EDG), the bromination takes place at C4, which exhibits a relatively high $^{13}$C but still the lowest $^1$H chemical shift. The regiochemical outcome is consistent with the relative magnitudes of orbital coefficients in the HOMO (Figure 2). Interestingly, the graphical orbital description distinguishes which of the two activated aromatic rings undergoes EAS with the higher probability.

#### 3.2. Six-Membered Heterocycles.
The reactivity of nonprotonated pyridine is comparable to that of nitrobenzene.3 Therefore, it is not surprising that electrophilic aromatic substitution is generally not an useful reaction for simple, unsubstituted six-membered aromatic heterocycle. However, if activated by an EDG, pyridines undergo the EAS easily, and unsurprisingly, the regioselectivity is dictated by the ability of a substituent to direct ortho and para. The predicted NMR shifts from ChemDraw generally correlate with the observed EAS regioselectivity for 2- and 4-substituted pyridines (21−24). The upfield signals correspond to the ortho- and/or para-position (Table 2). The same trend is followed for diazines, such as pyrimidines (25−29), pyrazines (30−31), and pyridazines (32−34).

#### 3.3. Other Examples.
Fused systems follow the same guidelines as the parent heterocycles (Table 3). As expected, the five-membered components are generally more reactive than their six-membered counterparts, and this is indeed represented by the upfield $^{13}$C and/or $^1$H signals. However, it should be noted that the regiochemical outcome of EAS on molecules containing a 1,2,4-triazole (41) moiety does not correlate well with the predictions. The same observation was made for isolated triazoles. This might be due a different reaction mechanism;74 therefore, examples of 1,2,4-triazoles are excluded from our investigations. The NMR-based method also fails for 1,3-oxazoles. This class of compounds does not undergo EAS easily unless activated by an EDG. Reactivity toward the electrophilic substitution follows the order C5 > C4 > C2.75,76 This difference in regioselectivity between C5 and C4 does not correlate with the predicted NMR shift trend, as it points to C4 being the more reactive site. The susceptibility of C5 to electrophilic attack is supported by the quantum chemical calculations which show that the HOMO...
is concentrated on the C5 (Figure 3). However, if C5 is substituted, the electrophile reacts at C4.

Another problem that arises for oxazoles is the accuracy of the NMR spectrum predicted by ChemDraw. Having analyzed a range of 2-substituted oxazoles, we noticed that regardless of the substituent character, ChemDraw gives the same output for both the $^1$H and $^{13}$C spectra (Figure 4). In such cases one must be aware of the software imperfections and, if possible, base predictions on the experimental spectra for the actual substrates or suitable model systems.

The chemical shift trend is not followed by 3-aminopyridines and some pyrazines (Figure 5A,B, for more see SI), but again the regioselectivity is explained by the of HOMO coefficients represented by the graphical models.

4. CONCLUSIONS

In our literature based search of over 130 examples, we found that the regioselectivity of the halogenation reactions correlates with the NMR predictions for almost 85% of the investigated structures. We found that some aromatic heterocycles are predicted better than others, but combined with the HOMO models obtained using DFT calculations, the accuracy level increases to more than 95%. We suggest to use Figure 6 to choose the most suitable method for predicting the site of EAS, depending on the nature of the aromatic heterocyclic core present in the molecule.

The evident drawbacks of the NMR-based method includes a quite large deviation between the predicted and the actual values for some heterocyclic cores. Moreover, the method considers neither steric factors nor the nature of the electrophile. Consequently, NMR is not able to differentiate between ortho and para positions and does not predict the high para-selectivity often observed. However, we believe that this very simple and fast method is of practical value for synthetic organic chemists and may be used prospectively to prioritize
between different synthetic routes and for late-stage functionalization of complex molecules.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.joc.6b00584.

Additional literature examples of EAS reactions and their regiochemical outcome correlation with the predicted chemical shift values and the HOMO maps, Cartesian coordinates, the number of imaginary frequencies and computed total energies of the optimized structures (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: mojj@lundbeck.com.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work was financially supported by The Innovation Fund Denmark (Ph.D. scholarship to M.M.K., grant 4135-00085B) and H. Lundbeck A/S. The authors would like to thank Dr. Ana Negri Martinez for the introduction to DFT calculations and Dr. Guiying Li for the translation of refs 18 and 19 from Chinese into English.

REFERENCES

(61) Jones, R. M.; Semple, G.; Fioravanti, B.; Pereira, G.; Calderon, I.; Uy, J.; Duvvuri, K.; Choi, J. S. K.; Xiong, Y.; Dave, V. 1,2,3-Trisubstituted aryl and heteroaryl derivatives as modulators of metabolism and the prophylaxis and the treatment of disorders related thereto such as diabetes and hyperglycemia. WO 2004065380 A8, October 21, 2004.