Bulgarian Chemical Communications, Volume 53, Issue 2 (pp. 240 - 248) 2021 *DOI: 10.34049/bcc.53.2.5382*

Prediction of ^{1}H-NMR shifts with Ambit-HNMR software

N. T. Kochev[1*], S. H. Tsoneva[1], M. V. Frenkeva[1], N. G. Jeliazkova[2]

[1]*University of Plovdiv "Paisii Hilendarski", Department of Analytical Chemistry and Computer Chemistry, 24 Tzar Asen Str., 4000 Plovdiv, Bulgaria*
[2]*Ideaconsult Ltd, 4 A Kanchev Str., Sofia 1000, Bulgaria*

Received: February 11, 2021; Revised: March 29, 2021

We present a new software tool, Ambit-HNMR, for automatic calculation of ^{1}H-NMR chemical shifts of organic compounds. Ambit-HNMR is an open-source software, written in Java, part of the chemoinformatics platform Ambit. Ambit-HNMR software uses a CDK-based molecule presentation as a connection table where H atoms are implicitly treated. The chemical shifts for each H atom are calculated by means of so called H-atom environments which describe different types of resonance protons. Each H-atom environment consists of: (1) a SMARTS pattern defining a molecular substructure; (2) base H shift, Z_0; (3) atom positions {a_1, a_2, …, a_n} and topological distances within the structure; (4) a set of possible substituents {S_1, S_2, …, S_k} described by SMARTS patterns and (5) associated chemical shifts, {Z_{i1}, Z_{i2}, …, Z_{in}}, for each substituent, S_i, respectively for all atoms' positions of the H-atom environment (i=1, 2, …, k). For an input target structure, all H-atom environments and all substituents {S_{r1}, S_{r2}, … S_{rn},} attached to the corresponding environment positions are identified by means of substructure searching. The chemical shift for a given H atom is calculated by means of an additive scheme: $Z_0 + Z_{r,1} + Z_{r,2} + \cdots + Z_{rn,n}$. We created a knowledge base with H-atom environments for the main classes of organic compounds based on the additive schemes of Pretsch. Additionally, Ambit-HNMR calculates the H atom multiplicity using the information of molecule classes of equivalent atoms. Ambit-HNMR module is available for download as a Java library or as a command line application (https://doi.org/10.5281/zenodo.4506289). Software example usage and test results are presented.

Keywords: Ambit, ^{1}H-NMR, chemical shifts, additive scheme, software, open-source.

INTRODUCTION

NMR spectroscopy is a powerful analytical technique for structure elucidation and study of molecular structure and is the preferred method due to its advantages: non-destructive, fast, with minimal sample preparation. ^{1}H-NMR interpretation can be assisted or automated by computer-based calculation of chemical shifts. The chemical shifts for different types of H atoms can be estimated with the help of additivity rules using the shift values of increments for substituents in various positions.

The spectra interpretation, namely full assignment of chemical shifts, is a pivotal task in analyzing any molecular structures and one that can be achieved by a number of means: (1) a library search that matches an unknown spectrum to one contained in the particular library; (2) application of additive rules for each considered atom and comparing the calculated to the experimental chemical shifts or (3) using dedicated ^{1}H-NMR predictor software.

A number of such software tools exist, both open-source and commercial. NMRShiftDB is a NMR predictor which can be used free of cost *via* web interface. The site credits the work of Binev and Aires-de-Sousa, developers of the FCT-Universidade NOVA de Lisboa tool [1]. The SPINUS (Structure-based Predictions In NUclear magnetic resonance Spectroscopy) program, http://neural.dq.fct.unl.pt/spinus/, uses a feed-forward neural network (FFNN) system and a series of empirical proton descriptors. The prediction is then corrected by Associative Neural Network (ASNN) on the basis of observed errors for the k nearest neighbors.

Some of the most widely used commercial softwares in general, ChemDraw [2] and Chem3D, developed by PerkinElmer and part of the ChemOffice package [3], also offer ^{1}H-NMR and ^{13}C-NMR predictions as a fully integrated tool in their versions. Their main advantage is the easy to use interface, but the quality of the estimated shifts can sometimes be challenged. Both programs use GAMESS software for performing general quantum chemistry calculations, such as HF, DFT, GVB and others, that can then be subject to correlation correction [4]. ACD/Labs' NMR Predictor software [5] takes it further by allowing the prediction of solvent specific spectra for both ^{1}H-NMR and ^{13}C-NMR, predicting multiplicity, recognizing tautomers prior to predicting and even simulating exact experimental conditions (frequency, line width). A very useful feature is the user's access to the complete calculation protocol. ACD/Labs' NMR Predictor uses HOSE code and neural net algorithms

* To whom all correspondence should be sent:
E-mail: nick@uni-plovdiv.net

to predict the chemical shifts while taking into account stereochemistry. It also provides with the ability to train predictions with new experimental data [5].

Another commercially available software is NMRPredict [6] distributed by Modgraph Consultants Ltd. It predicts chemical shifts using two prediction methods: (1) additivity rules developed by Pretsch, as well as several stages of approximation and (2) predictions based on functional groups parametrized by Abraham and then automatically chooses "best" proton prediction. In the first case, chemical shifts are predicted for each proton with available additivity rules by assigning substructures following a hierarchical list. Thus the base value of the final shift is given. If there are no available data for a given substructure, e.g. ring system, it is to be disassembled so that a base value can be approximated. For each substructure the remaining part of the molecule is treated as substituents that contribute to the final shift by adding their increments to the base value. In the second case each functional group is identified and treated separately using CHARGE [7] by first generating 3D conformers and then predicting the proton spectra for all of them giving a weighted average spectrum.

Mestrelab offers a similar product, Mestrenova [8], using HOSE code, that starts at a given hydrogen/carbon atom and tries to find the environment one bond away in its database. This process is repeated for every hydrogen/carbon atom until reaching the boundary of the molecule or coming across environment not represented in the database. Gaussian software [9] has by far the widest array of modeling methods including HF, DFT, MP2, CCSD, etc. The accuracy of the predicted shifts depends on the basis set used and the application of adequate solvent correction. An alternative free option is ORCA *ab initio* program with its variety of methods ranging from semi-empirical methods and density functional theory to correlated single- and multi-reference wave-function based methods [10]. Downloading and running these or other similar programs may be viewed as necessary, especially when using your own confidential data, but with the growing numbers and popularity of Web browser applications it may not be actually requisite. AMBIT chemoinformatics platform is one such open-source platform for predictive models [11].

We present Ambit-HNMR, a new software tool for automatic calculation of [1]H-NMR chemical shifts of organic compounds. It is an open-source tool and is part of the chemoinformatics module Ambit-GCM [12] which facilitates interoperability with external software packages, thus making running predictions and sharing online resources much easier without installing any additional software. We have implemented a knowledge base with H-atom environments for the main classes of hydrocarbon compounds based on the additive schemes of Pretsch [13] and the knowledge base is being updated with new rules.

Software characteristics and architecture

Ambit-HNMR is implemented in object-oriented programing language Java. It is an open-source, OS independent software module distributed under LGPL license [14]. Ambit-HNMR is an extension of Ambit-GCM [12] (previously developed by us software tool) part of the chemoinformatics platform Ambit [11] where it is integrated as a separate module (ambit2-groupcontribution.nmr). Ambit integration allows the usage of plenty of chemoinformatics functionalities from other Ambit modules [11] developed by our group, as well as utilities from external open-source resources. The full capability of Ambit-HNMR could be accessed when using it as a Java library with APIs. In addition, we have developed a console tool, Ambit-HNMR, available as a command-line interface (CLI) Java application, as well as a GUI application. Detailed information about Ambit software platform, as well as the source code of the Java library (ambit2-groupcontribution.nmr) is available at http://ambit.sourceforge.net/. Executable *.jar files with the latest Ambit-HNMR knowledge base can be downloaded from the Zenodo repository: https://doi.org/10.5281/zenodo.4506289. Ambit-HNMR implementation includes the following basic components:

(1) *Data input/output utilities*. Ambit modules are developed on top of the CDK (Chemistry Development Kit) [15, 16] library. CDK provides input and output for basic structure presentation formats thus Ambit-HNMR supports most popular chemoinformatics formats: SMILES [17, 18] and InChI [19] linear notations, CML chemical format, MOL/SDF file formats, CSV and TXT file formats. The software configuration can be done from command-line interface options. The variety of supported file formats allows easy integration of our tool with other software applications.

(2) *Structural information management*. The basic structure management in Ambit-HNMR is developed on top of CDK classes. The chemical structure representation relies on the CDK Java class AtomContainer which implements the molecular connection table (CT).

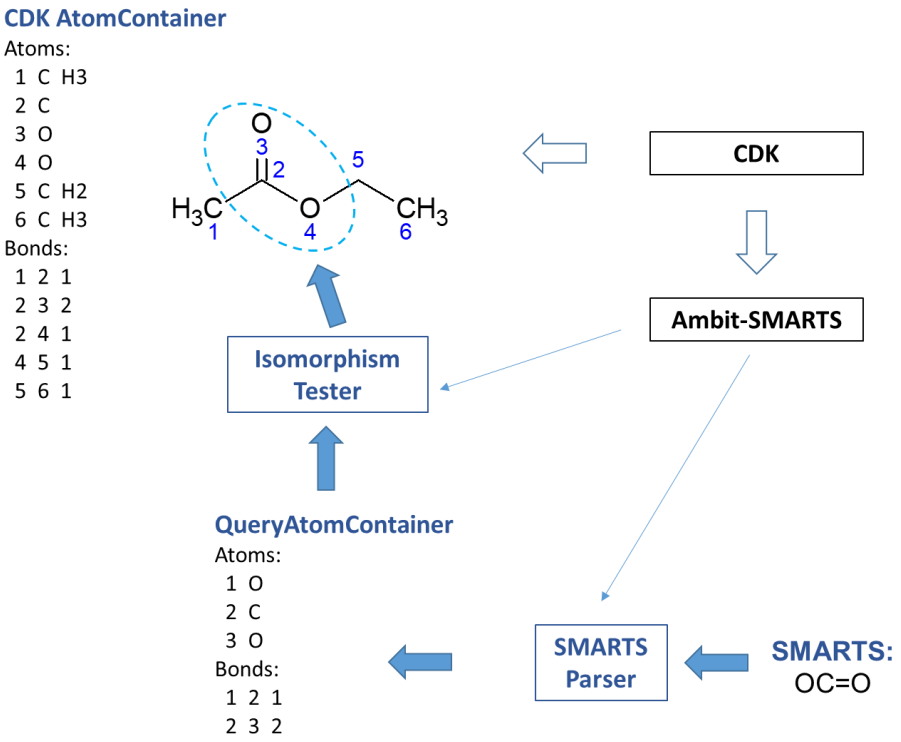


Figure 1. Chemical object management in Ambit-HNMR

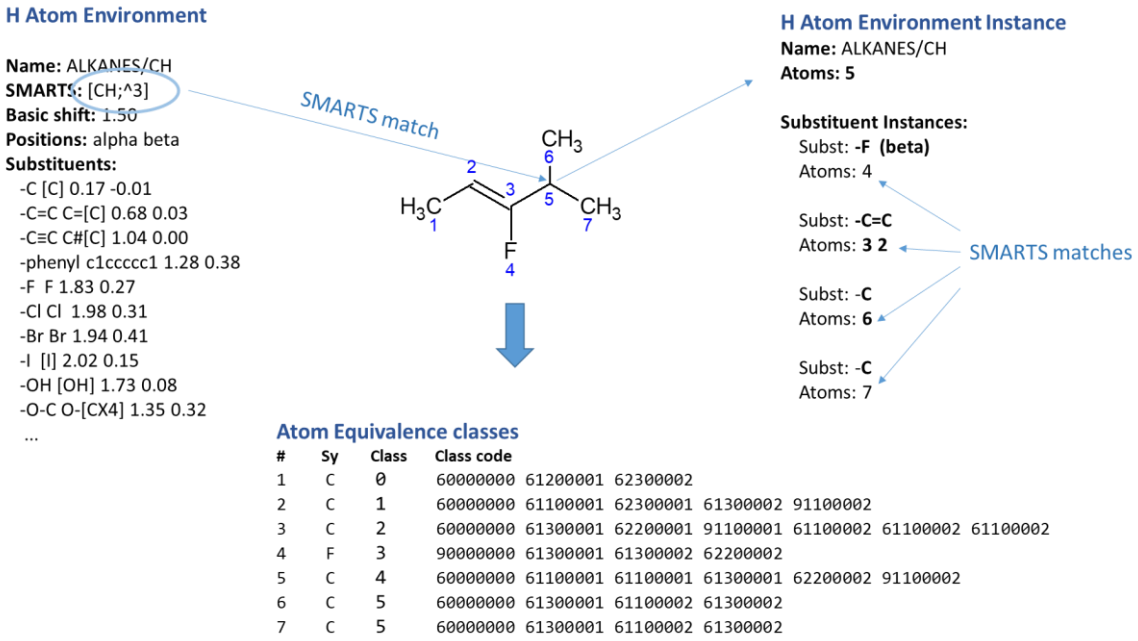


Figure 2. H Atom Environment and atom equivalence classes management

Ambit-HNMR processing requires handling of the two basic types of chemical objects shown in Figure 1: chemical structures (basic CT) and chemical fragments or substructures (represented by means of specialized CT) plus dedicated data structures for H atom environments and classes of equivalent atoms. H atom environment management includes multiple SMARTS linear notations used for definition of the heavy atoms associated to a given proton (H atom) and SMARTS definitions for corresponding substituents (see Figure 2). Atom equivalency within a molecule is determined by Ambit-SMARTS [20] module using specially designed atomic codes also illustrated in Figure 2.

(3) *Substructure searching.* Ambit-SMARTS [20] module is also developed by our group and implements the key substructure search algorithm (see in Figure 1, Isomorphism Tester is applied to

match OC=O against the target molecule). It is used for finding all matchings (called instances) of a given H atom environment and the matchings (instances) of substituents. Figure 2 illustrates the H atom environment for tertiary sp3 carbon atoms and matching of an instance at atom 5 with four substituents at three alpha and one beta positions at atoms {3, 2}, {6}, {7} and {4} respectively.

(4) *Chemical shift calculation*. This is the most crucial software component. The calculation of H-shifts is based on exhaustive searching of all possible instances of all H atom environments from the Ambit-HNMR knowledge base described in following sections.

HNMR shifts calculation algorithm

Ambit-HNMR algorithm for H-shift calculation includes the following main stages:

(i) loading of a knowledge base with H-shift rules formalized as H atom environments (see example in Figure 2 and format syntax example below);

(ii) target molecule input using one of the popular molecular formats (e.g. SMILES linear notation) and conversion into the internal CDK representation;

(iii) calculation of topological distance matrix of the target molecule;

(iv) finding of all groups mappings for all SMARTS definitions from the knowledge base;

(v) determination of all H Atom Environment Instances;

(vi) finding all substituents for each H Atom Environment Instance;

(vii) calculation of H-shifts using additive schemes;

(viii) determination of atom equivalence classes;

(ix) calculation of multiplicity.

The basic workflow of Ambit-HNMR is summarized in Figure 3 illustrating stages (i), (ii), (v), (vi) and (vii).

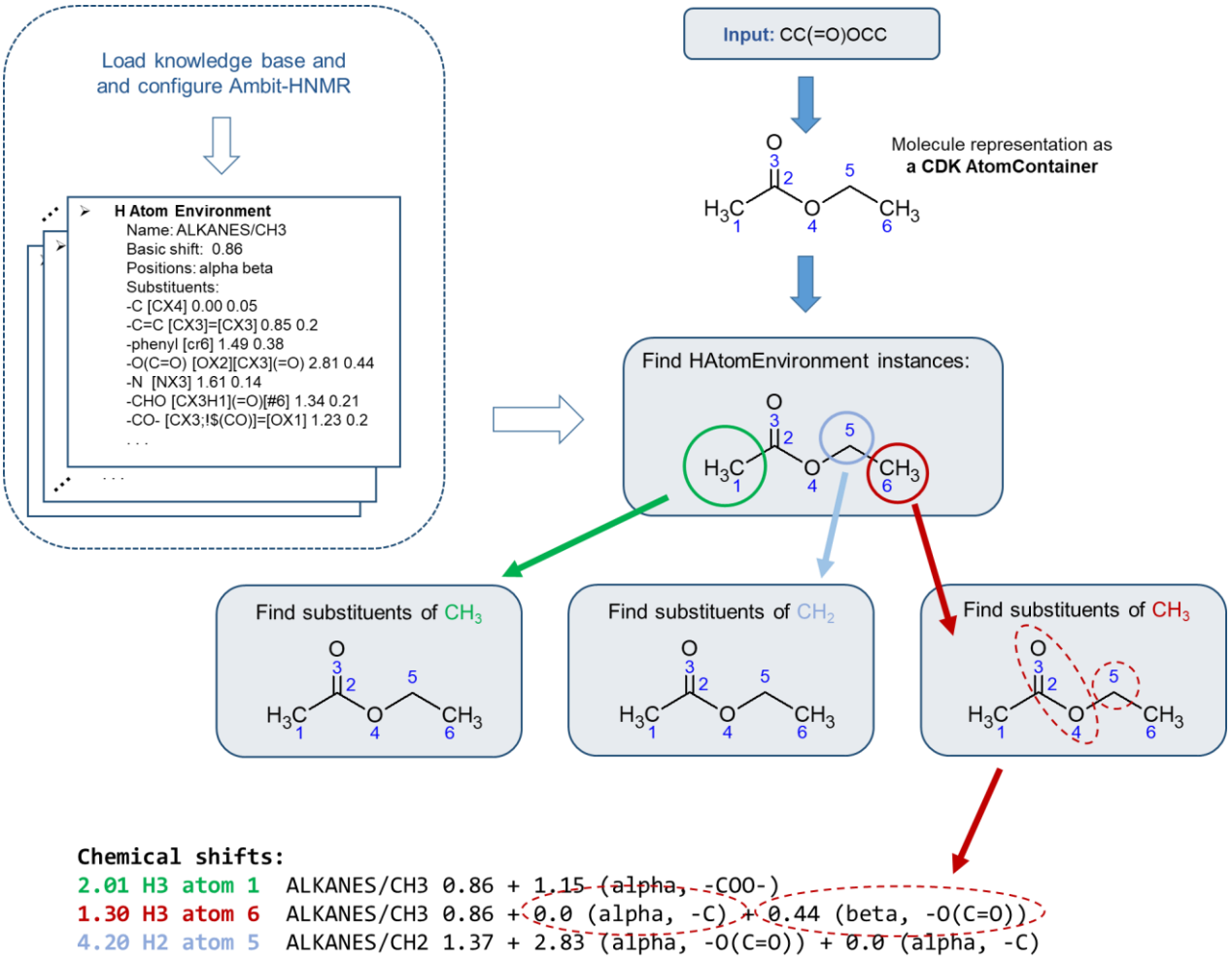


Figure 3. Flow chart of the Ambit-HNMR H-shift calculation algorithm

Ambit-HNMR knowledge base is configurable and stored in an external text file. It contains additivity rules described as H atom environments with the following syntax:

```
$$H_ATOM_ENVIRONMENT=
$$NAME= ALKANES/CH2
$$SMARTS= [CH2;^3]
$$BASIC_SHIFT= 1.37
$$IMPLICIT_H_ATOMS_NUMBER= 2
$$SHIFT_DESIGNATIONS= alpha beta
$$SHIFT_ASSOCIATION                    =
SUBSTITUENT_POSITION
$$SUBSTITUENT_POS_ATOM_INDICES= 1 1
$$POSITION_DISTANCES = 1 2
$$SUBST= -C [C] 0.00 -0.04
$$SUBST= -C=C C=[C] 0.63 0.00
$$SUBST= -C≡C C#[C] 0.70 0.13
$$SUBST= -phenyl c1ccccc1 1.22 0.29
```

H-atom environment consists of: (1) a SMARTS pattern defining a molecular substructure or an atom used for topological identification of the H atoms (e.g. [CH2;^3] defines a secondary carbon); (2) base H shift, Z_0 (key word BASIC_SHIFT) is the value used for unsubstituted substructure/atom; (3) possible substituents atom positions $\{a_1, a_2, \ldots, a_n\}$, given as atomic indices within the defined substructure and associated topological distances from the substituents to the substructure; (4) a set of possible substituents, $\{S_1, S_2, \ldots, S_k\}$ described by SMARTS patterns; and (5) associated chemical shift contributions, $\{Z_{i1}, Z_{i2}, \ldots, Z_{in}\}$, for each substituent, S_i, respectively for all atoms' positions and distances of the H-atom environment (i=1, 2, …, k). Points (4) and (5) are given by means of a list of substituents with additive contributions for the chemical shifts described in the form:

```
$$SUBST=       <name>       <smarts>
<contribution 1> <contribution 2> ...
```

For the example given above, row "$$SUBST= -phenyl c1ccccc1 1.22 0.29" means that benzene ring will have additive contribution 1.22 ppm in alpha position and 0.29 ppm in beta position added to the basic chemical shift, 1.37 ppm of H atoms of CH_2 carbon atom.

The default knowledge base contains rules based on the Pretsch additives schemes. The piece of default knowledge base shown above describes an additive scheme for secondary sp3 carbon atom, CH_2, in the form:

$$\delta[CH_2] = 1.37 + \sum_i Z_{\alpha_i} + \sum_j Z_{\beta_j},$$

where Z_{α_i} are the contributions to the chemical shift of the substituents at alpha position (topological distance 1 from the carbon atom) and respectively Z_{β_j} are the contributions of substituents at beta positions (topological distance 2). The default knowledge base includes H atom environments for basic atoms of hydrocarbon chains (e.g. carbons with different numbers of H atoms and different hybridizations) with a rich set of possible substituents: alkyl, phenyl, halogens, chemical groups and fragments containing oxygen, nitrogen and sulfur atoms, etc. The full list of the possible substituents can be seen in the knowledge base file available at the Zenodo repository (https://doi.org/10.5281/zenodo.4506289). In the current knowledge base version, we have also included H atom environments for mono substituted benzene rings and pyridines. The user can enrich the knowledge base with additional rules or, if needed, completely replace the default rules. Ambit-HNMR calculates multiplicity for all predicted H-shifts using the atom equivalence classes. Atom equivalence is determined by calculating so called atomic codes which include information about a predefined number of topological layers of an atom (default number of layers is 2). The atomic code is composed of pieces of information for each atom in each topological environment encoding: atom element, bond type, topological degree, charge, isotope and topological layer number (i.e. distance from the center atom). All atoms having the same atomic code are considered as equivalent. For example, in Figure 2, atoms 6 and 7 (the terminal methyl groups) have exactly the same atomic code, "60000000 61300001 61100002 61300002", hence they are identified as equivalent atoms (class 5). The multiplicity of a particular H atom is determined by counting the number of H atoms, $\{n_1, n_2, \ldots, n_s\}$, for each neighbor atom equivalence class. Then multiplicity is $(n_1+1) \times (n_2+1) \times \ldots \times (n_s+1)$.

RESULTS AND DISCUSSION

Ambit-HNMR software usage

Ambit-HNMR software version 1.2 is available as a command-line interface application with following options:

```
HNMR shifts predict
usage:
ambit2.groupcontribution.cli.HNMRPred
ictCli
   -c,--config <config>         HNMR
database configuration file
   -e,--explanation <on|off>    Switch
on/off H shift calculation
```

```
explanation. Default explanation is
on
    -h,--help                    Shows
this help info
    -i,--input <input>           Input
molecule file
    -l,--log <on|off>            Switch
on/off log printing. Default log is
                                 off
    -p,--multiplicity <on|off>   Switch
on/off H multiplicity feature.

Default is on
    -s,--smiles <smiles>         Input
molecule smiles
```

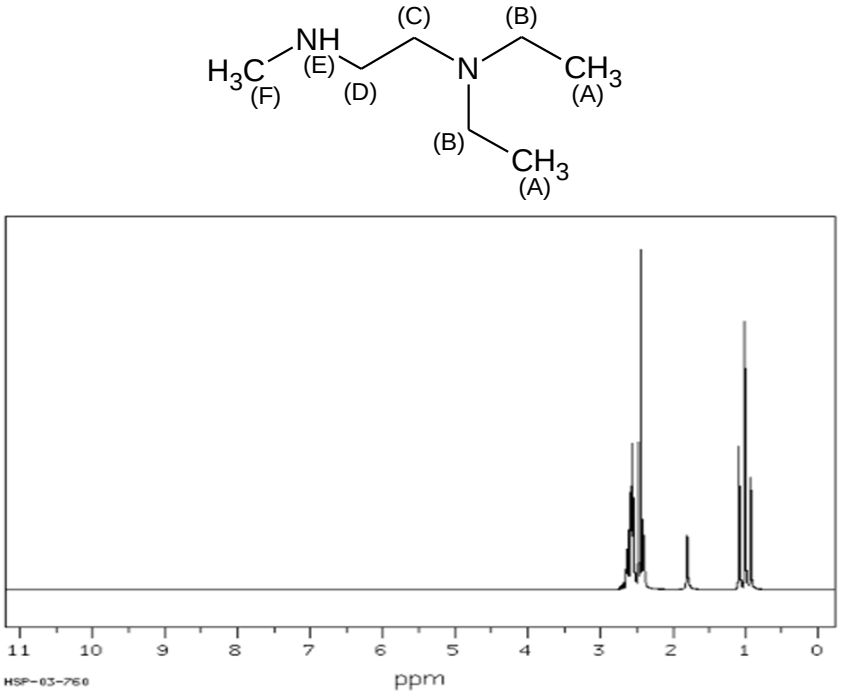
Example of Ambit-HNMR application for the molecule of ethyl acetate inputted as a SMILES notation from the command line is given below:

```
java -jar ambit-hnmr.jar -s CC(=O)OCC
```

```
  Using default HNMR database: ./hnmr-
knowledgebase.txt
  Input smiles: CC(=O)OCC
  2.01  H3  atom  1  multiplicity
ALKANES/CH3 0.86 + 1.15 (alpha, -COO-)
  1.30  H3  atom  6  multiplicity  3
ALKANES/CH3  0.86  +  0.0  (alpha,  -
C_alkyl) + 0.44 (beta, -O(C=O))
  4.20  H2  atom  5  multiplicity  4
ALKANES/CH2  1.37  +  2.83  (alpha,  -
O(C=O)) + 0.0 (alpha, -C_alkyl)
```

Ambit-HNMR can also be applied in a batch mode for a set of molecules specified by means of *-i* option. Executable *.jar file, detailed documentation and more usage examples are available at: https://doi.org/10.5281/zenodo.4506289.

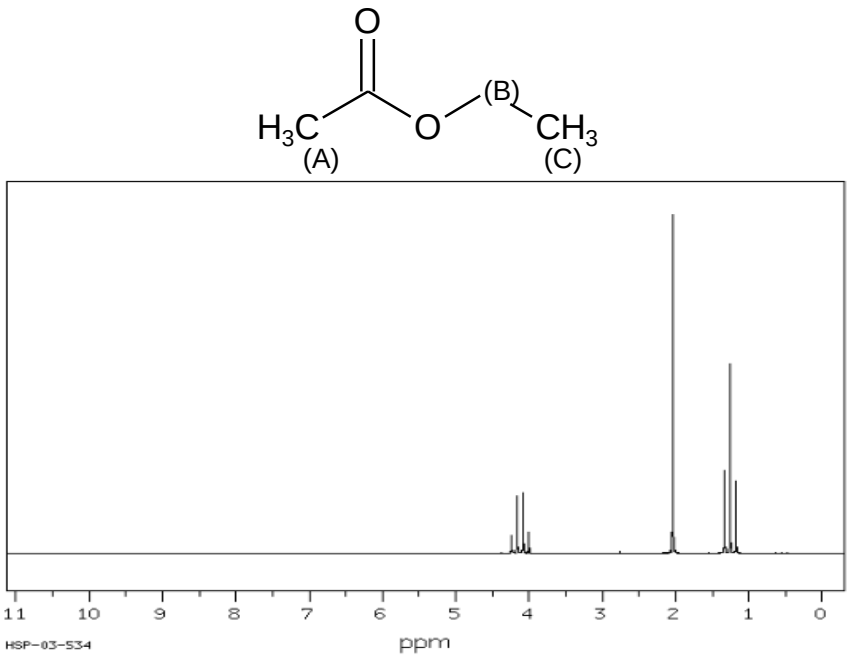
We present a comparison of Ambit-HNMR calculated chemical shifts with the experimental [1]H-NMR spectra of organic compounds from the public database SDBS [21] and four different software tools for prediction of [1]H-NMR chemical shifts.



Protons	Shifts [1]H-NMR spectrum (ppm)	Ambit-HNMR (ppm)	Diff (ppm)	Chem Draw (ppm)	Diff. (ppm)	nmrdb.org (ppm)	Diff. (ppm)	Mestre nova (ppm)	Diff. (ppm)	Gaussian (ppm)	Diff. (ppm)
H(A, A)	1.01	1.00	0.01	1.00	0.01	0.96	0.05	1.09	0.08	1.01	0.00
H(B, B)	2.51	2.61	0.10	2.40	0.11	2.63	0.12	2.45	0.06	2.10	0.41
H(C)	2.54	2.83	0.29	2.48	0.06	2.83	0.29	2.59	0.05	2.07	0.47
H(D)	2.62	2.87	0.25	2.65	0.03	2.8	0.18	2.72	0.10	2.35	0.27
H(E)	1.81							1.14	0.67	0.43	1.38
H(F)	2.47	2.52	0.05	2.47	0.00	2.31	0.16	2.56	0.09	2.34	0.13

Figure 4. H-NMR spectrum of N,N-diethyl-N'-methylethylenediamine and comparison with the calculated shifts from Ambit-HNMR, ChemDraw, nmrdb.org, Mestrenova and Gaussian

Figure 4 shows the structure and spectrum of N, N-diethyl-N'-methylethylenediamine (measurement conditions: 89.56 MHz, solution 0.04 ml: 0.5 ml $CDCl_3$) and comparison of the experimental and predicted chemical shifts. The structure of N,N-diethyl-N'-methylethylenediamine is non-cyclic with secondary and tertiary amino groups in it. There are two methyl (A) and two methylene (B) groups that are chemically equivalent with each other that have the same chemical shifts in the ^{1}H-NMR spectrum. According to the positions of the H atoms in the structure of the compound and the calculated distances it is seen in the spectrum that there is a certain number of multiplets. The comparison between spectrum and Ambit-HNMR shows that chemical shift differences are with deviation less than 0.3 ppm. Figure 4 also shows a comparison between chemical shifts predicted with Mestrenova and Gaussian and their difference in ppm from the experimental ^{1}H NMR spectrum. Mestrenova simulates full spectra with their chemical shifts and multiplicity with the help of the NMRPredict Desktop.



Protons	Shifts ^{1}H-NMR spectrum (ppm)	Ambit-HNMR (ppm)	Diff. (ppm)	Chem Draw (ppm)	Diff. (ppm)	nmrdb.org (ppm)	Diff. (ppm)	Mestre nova (ppm)	Diff. (ppm)	Gaussian (ppm)	Diff. (ppm)
H(A)	2.04	2.01	0.03	2.01	0.03	2.04	0.00	2.05	0.01	2.06	0.02
H(C)	1.26	1.30	0.04	1.30	0.04	1.16	0.10	1.37	0.11	1.38	0.12
H(B)	4.12	4.20	0.08	4.12	0.00	4.10	0.02	4.16	0.04	3.80	0.32

Figure 5. HNMR spectrum of ethyl acetate and comparison with the calculated shifts from Ambit-HNMR, ChemDraw, nmrdb.org, Mestrenova and Gaussian

The calculations were quick and fairly accurate for all protons except for the one in the NH group. For the purpose of comparing the capabilities of Gaussian 98, Revision A.7, the geometries of all molecules (N,N-diethyl-N'-methylethylenediamine and ethyl acetate and styrene, presented below) were optimized on B3LYP (Becke three-parameter Lee–Yang–Parr) exchange-correlation functional with 6-31g basis set with subsequent calculations carried out on HF with 6-311+g (2d, p) basis set for the ^{1}H NMR spectra prediction.

The structure, HNMR spectrum and predicted H-shifts of ethyl acetate are shown in Figure 5. The compound has two methyl (A, C) groups that are not chemically equivalent with each other and have different chemical shifts in the H-NMR spectrum as it is shown in Figure 5. According to the positions of the H atoms in the structure and the topological distances, the spectrum has a triplet and a quadruplet, which are correctly calculated by Ambit-HNMR (see the multiplicity in the console output for the molecule of ethyl acetate shown previously in this section). Chemical shifts comparison shows that the differences are with deviation less than 0.1 ppm.

Figure 6 shows the GUI (graphical use interface) application with the result screen for the molecule of styrene.

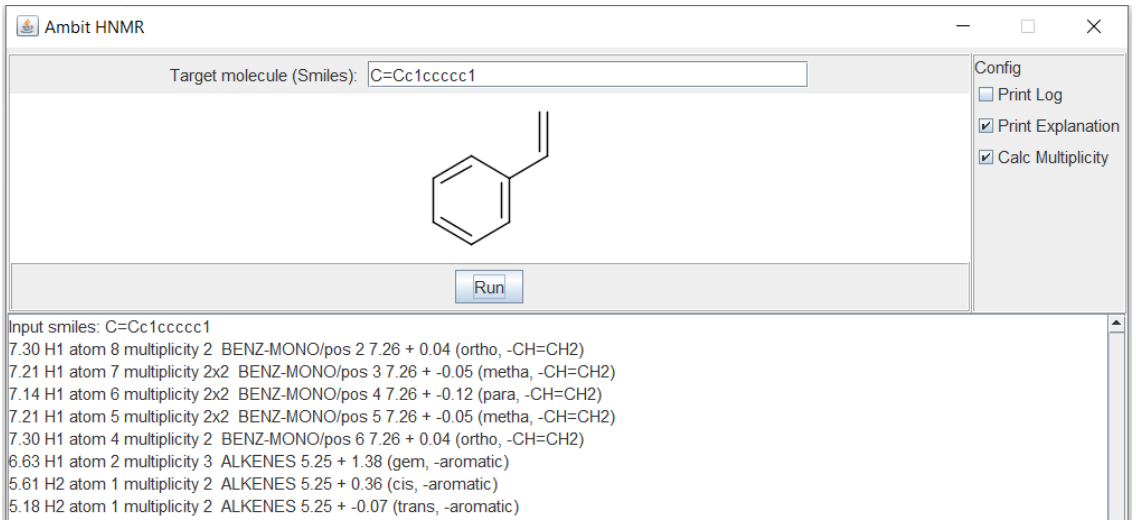
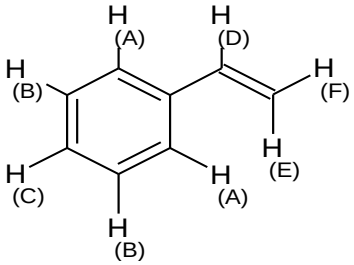
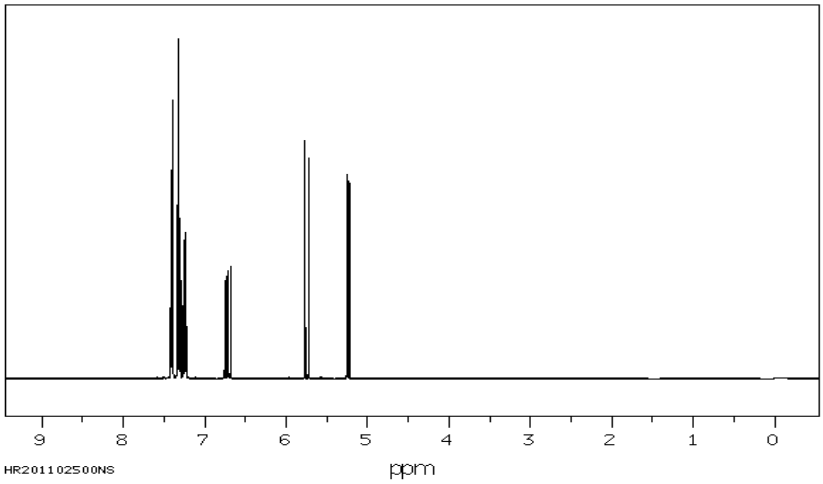


Figure 6. Ambit-HNMR-GUI screen with the calculated shifts for the molecule of styrene.





Protons	Shifts [1]H-NMR spectrum (ppm)	Ambit-HNMR (ppm)	Diff. (ppm)	Chem Draw (ppm)	Diff. (ppm)	nmrdb.org (ppm)	Diff. (ppm)	Mestre nova (ppm)	Diff. (ppm)	Gaussian (ppm)	Diff. (ppm)
H(A, A)	7.40	7.30	0.10	7.30	0.10	7.40	0.00	7.29	0.11	7.95	0.55
H(B, B)	7.32	7.21	0.11	7.21	0.11	7.33	0.01	7.23	0.09	7.70	0.38
H(C)	7.25	7.14	0.11	7.14	0.11	7.26	0.01	7.11	0.14	7.62	0.37
H(D)	6.72	6.63	0.09	6.63	0.09	6.82	0.10	6.63	0.09	6.98	0.26
H(E)	5.74	5.61	0.13	5.61	0.13	5.84	0.10	5.59	0.15	6.20	0.46
H(F)	5.24	5.18	0.06	5.18	0.06	5.13	0.11	5.30	0.06	5.39	0.15

Figure 7. HNMR spectrum of styrene and the calculated shifts from Ambit-HNMR, ChemDraw, nmrdb.org, Mestrenova and Gaussian.

Full comparison of the experimental H-NMR spectrum shifts and the calculated chemical shifts from Ambit-HNMR tool and four other testing software tools is given in Figure 7. For the molecule of styrene, the average chemical shift difference of Ambit-HNMR is 0.1 ppm. The corresponding

difference averages of the other tools are: ChemDraw - 0.1 ppm, nmrdb.org - 0.55, Mestrenova - 0.11, and Gaussian – 0.36, showing that the Ambit-HNMR performance is comparable with the considered test software tools.

A major challenge of the additive scheme methods is the problems of missing fragments. i.e. the cases of organic compounds containing fragments which have no additive contributions in the knowledge base (e.g. the missing secondary amine in Figure 4). That is why continuous improvement of the knowledge base is required. In this context, one advantage of the Ambit-HNMR software is that the knowledge base rules are not hardly encoded but are configurable and externally stored so that the user can update, enrich or replace the default rules.

CONCLUSION

A new software tool, Ambit-HNMR, for prediction of [1]H-NMR chemical shifts is developed. The software uses a configurable knowledge base, which can be modified by the user. The command line application can be applied for organic compounds inputted directly from the console as SMILES linear notations or in a batch mode for an input molecular file. Software performance is demonstrated with examples. Ambit-HNMR source code is available at http://ambit.sourceforge.net/ and the software can be easily integrated as part of a bigger scientific workflow. Executable jar file with the latest software version and knowledge base, additional usage examples and full documentation is present at: https://doi.org/10.5281/zenodo.4506289. Our team is continuing the improvement of default knowledge base rules and new releases will be available at the Zenodo repository.

Acknowledgement: *This work was supported by the University of Plovdiv Scientific Fund under project MU19-HF-003.*

REFERENCES

1. Y. Binev, M. M. Marques, J. Aires-de-Sousa, *J. Chem. Inf. Model.*, **47**(6), 2089 (2007).
2. N. Mills, *J. Am. Chem. Soc.*, **128** (41), 13649 (2006).
3. A. Zielesny, *J. Chem. Inf. Model.*, **45** (5), 1474 (2005**).**
4. M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, *J. Comput. Chem.*, **14**, 1347 (1993).
5. E. E. Kwan, *J. Chem. Inf. Model.*, **52** (7), 1898 (2012).
6. J. Abraham. M. Mobli, Modelling [1]H NMR Spectra of Organic Compounds: Theory, Applications and NMR Prediction Software, John Wiley & Sons Ltd, 2008.
7. R. J. Abraham, M. Mobli, *Spectroscopy Europe*, **16**(4), (2004).
8. T. Claridge, M. Nova, *J. Chem. Inf. Model.*, **49**(4), 1136 (2009).
9. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian 16, Revision C.01, Gaussian, Inc., Wallingford, CT, 2016.
10. F. Neese, F. Wennmohs, U. Becker, C. Riplinger, *J. Chem. Phys*. **152**, 224108 (2020).
11. N. Jeliazkova, V. Jeliazkov, *J. Cheminform*., **3**(1), 18 (2011).
12. N. Kochev, V. Paskaleva, O. Pukalov, N. Jeliazkova, *Molecular Informatics*, **38**(8-9), e1800138 (2019).
13. E. Pretsch, P. Buhlmann, M. Badertscher, Structure Determination of Organic Compounds, Springer, 2009.
14. LGPL, GNU Lesser General Public License, https://www.gnu.org/licenses/lgpl-3.0.en.html, last accessed on 04.02.2021.
15. E. L. Willighagen, J. W. Mayfield, J. Alvarsson, A. Berg, L. Carlsson, N. Jeliazkova, S. Kuhn, T. Pluskal, M. Rojas-Chertó, O. Spjuth, G. Torrance, C. T. Evelo, R. Guha, C. Steinbeck, *J. Cheminform.*, **9**(1), 33 (2017).
16. Chemistry Development Kit, https://cdk.github.io/ accessed 04.02.2021.
17. D. Weininger, *J. Chem. Inf. Comput. Sci.*, **28**(1), 31 1988.
18. https://www.daylight.com/dayhtml/doc/theory/theory.smiles, Daylight Theory: SMILES, accessed 04.02.2021.
19. S. Heller, A. McNaught, I. Pletnev, S. Stein, D. Tchekhovskoi, *J. Cheminform.*, **7**(23), 2 (2015).
20. N. Jeliazkova, N. Kochev, *Molecular Informatics*, **30**(8), 707 (2011).
21. Spectral Database for Organic Compounds, SDBS, https://sdbs.db.aist.go.jp/sdbs/cgi-bin/cre_index.cgi, accessed 04.02.2021.