Trends Phytochem. Res. 1(3) 2017 109-110

Trends in Phytochemical Research (TPR)

Journal Homepage: http://tpr.iau-shahrood.ac.ir

Guest Editorial

Automated structure elucidation of phytochemicals

Lutfun Nahar ✉ and Satyajit D. Sarker

Medicinal Chemistry and Natural Products Research Group, School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, James Parsons Building, Byrom Street, Liverpool L3 3AF, UK

Plants produce a variety of chemical compounds, and plants have been the main source of new chemical entities and novel chemical scaffolds or templates, unfolding new challenges for organic synthetic chemists to explore appropriate synthetic routes for their total synthesis. Because of the unique chemical diversity offered by plants, it is often a tedious and complicated process when it comes to structure elucidation of phytochemicals. Recent advances in spectroscopic techniques, particularly in NMR and MS methodologies, have provided various tools that assist phytochemists with the structure elucidation of known or new phytochemicals. However, spectroscopic data interpretation manually requires significant experience and expertise, knowledge, intellectual ability and patience; often the manual process can be quite time consuming and even be frustrating. Over the last several decades, especially with the phenomenal progress in computation and applications of artificial intelligence and various mathematical modelling, several automated spectral data interpretation and structure elucidation software have become available to the phytochemists. These automated tools, not necessarily have replaced human intelligence or efforts, but certainly have facilitated the process, and improved the accuracy of structure elucidation of phytochemicals.

In automated elucidation of the structure of an organic molecule or phytochemical from experimental data, two steps of data processing are necessary: a structure generator that will create proposals for the unknown molecular structure, and a filter that will validate these proposals usually by comparing easily derivable properties of the generated molecules with the corresponding experimental values and minimizing the deviation (Meiler and Will, 2001). Theoretically, it should be possible to elucidate the structure of an organic molecule from just one experimental quantity, if each molecule has its own characteristic experimental quantity and this value is obtainable without any experimental uncertainty. However, there are two more additional prerequisites: the experimental value can be calculated from the molecular structure with infinite precision, and infinite computational power is available.

Attempts for introducing computational approaches and algorithms for computer-aided structure elucidation (CASE) to automatically elucidate the structures of organic molecules started about half a century ago, and have gone through various developmental phases to come to a point now, when several commercially available automatic structure prediction tools and software are available to phytochemists. One of the earlier efforts in automated structure elucidation is evident in the publication by Funatsu et al. (1989), where they outlined the principles of an computer-assited automated structure elucidation system for organic compounds (CHEMICS), which relied on the way of structure generation method-the most probable structure could be generated by the automated analysis of spectral data (mainly NMR data) of an unknown using empirical and theoretical rules. They also highlighted its applications in natural products structure elucidation. CHEMICS was composed of four fundamental functional modules: data analysis, structure generator, stereo-generator and input of macro-component (partial structure). Later, Meiler and Will (2001) reported the protocol for automated structure elucidation of organic molecules from ^{13}C NMR spectral data using genetic algorithms and neural networks, and this method was applicable to phytochemicals as well. In their protocol, the structure of organic molecules could be optimized to meet experimental criteria, if a fast and accurate method for the prediction of the used physical or chemical features were available, e.g., ^{13}C NMR chemical shift, intensity, and multiplicity information from a ^{13}C NMR DEPT spectrum. Earlier, Meiler et al. (1999; 2000) demonstrated fast determination of ^{13}C NMR chemical

✉ Corresponding author: Lutfun Nahar
Tel: 0151-231-2622, Fax: :0151-231-2170
E-mail address: L.Nahar@ljmu.ac.uk; drnahar@live.co.uk

shifts using artificial neural networks, and there are several other related work on predictive spectroscopy, especially involving ^{13}C NMR chemical shifts, published during 1990-2010 (Furst et al., 1990; Pretsch et al., 1991; Anker and Jurs, 1992; Doucet et al., 1993; Hyami et al., 1993; Peng et al., 1994; Svozil et al., 1995; Munk, 1998; Meiler and Kock, 2004).

During this period, and following on from there until today, not only the 1D NMR spectroscopic techniques that were used in the prediction of structures of phytochemicals, but also various homo and heteronuclear 2D NMR correlation data were used (Blinov et al., 2001; Steinbeck, 2004; Golotvin et al., 2007; Kuhn et al., 2008; Elyashberg et al., 2009; Penchev et al., 2012; Jayaseelan and Steinbeck, 2014). Despite significant developments in this area, one of the existing key challenges is the unambiguous distinction between $2J_{CH}$ and $3J_{CH}$ in HMBC experiments and between $2J_{CC}$ and $3J_{CC}$ in ADEQUATE experiments, which will have to be incorporated into the CASE programmes to help reduce calculation time and generate a smaller number of solutions, and appropriate density functional theory (DFT) calculations should be used.

Nowadays, there are several commercially available structure prediction software available for elucidating structure of phytochemicals and other organic molecules-known or unknown. One of the well-known automated structure elucidation systems is offered by the ACD-LABS (http://www.acdlabs.com/download/publ/2004/enc04/autostruceluc.pdf), and is called ACD/Structure Elucidator-StrucEluc, which is quite popular among the phytochemists and other natural products researchers. Bruker offers a suite of products, known as Complete Molecular Confidence-CMC, for automated characterization of small molecules and refined small molecule analysis (https://www.bruker.com/products/mr/nmr/nmr-software/nmr-software/complete-molecular-confidence.html). A list of various other software for automated structure elucidation using NMR spectroscopy is available at http://www.spincore.com/nmrinfo/software_s.shtml. There is no doubt, that the tremendous development in the field of computation and its applications in automating structure elucidation and structure prediction of phytochemicals have enhanced the quality and output of phytochemical research, whether it is plant metabolomics, or phytochemical drug discovery.

References

Anker, L.S., Jurs, P.C., 1992. Prediction of carbon-13 nuclear magnetic resonance chemical shifts by artificial neural networks. Anal. Chem. 64(10), 1157-1164.

Blinov, K.A., Elyashberg, M.E., Molodtsov, S.G., Williams, A.J., Martirosian, E.R., 2001. An expert system for automated structure elucidation utilizing ^{1}H-^{1}H, ^{13}C-^{1}H and ^{15}N-^{1}H 2D NMR correlations. Fresenius J. Anal. Chem. 369(7-8), 709–714.

Doucet, J.P., Panaye, A., Feuilleaubois, E., Ladd, P., 1993. Neural networks and ^{13}C NMR shift prediction. J. Chem. Inf. Comput. Sci. 33(3), 320-324.

Elyashberg, M., Blinov, K., Molodtsov, S., Smurnyy, Y., Williamn, A.J., Churanova, T., 2009. Computer-assisted methods for molecular structure elucidation: realizing a spectroscopist's dream. J. Cheminformatics 1, 3.

Funatsu, K., Susuta, Y., Sasaki, S-I., 1989. Application of the automated structure elucidation system (CHEMICS) to the chemistry of natural products. Pure & Appl. Chem. 61(3), 609-612.

Furst, A., Pretsch, E., Robien, W., 1990, Comprehensive parameter set for the prediction of the ^{13}C NMR chemical shifts of sp^3-hybridized carbon atoms in organic compounds. Anal. Chim. Acta 233(1), 213-222.

Golotvin, S.S., Vodopianov, E., Pol, R., Lefebvre, B.A., William, A.J., Rutkowske, R.D., Spitzer, T.D., 2007. Automated structure verification based on a combination of 1D ^{1}H NMR and 2D ^{1}H-^{13}C HSQC spectra. Magn. Reson. Chem. 45(10), 803-813.

Hyami, K-I., Funatsu, K., Sasaki, S-I., 1993. Improvement in the structure generation of the automated structure elucidation system for organic compounds, CHEMICS. Bunseki Kagaku 42(6), 369-374.

Jayaseelan, K.V., Steinbeck, C., 2014. Building blocks for automated elucidation of metabolites: natural product-likeness for candidate ranking. BMC Bioinformatics 15, 234.

Kuhn, S., Egert, B., Neumann, S., Steinbeck, C., 2008. Building blocks for automated elucidation of metabolites: Machine-learning methods for NMR prediction. BMC Bioinformatics 9, 400.

Meiler, J., Meusinger, R., Will, M., 1999. Neural network prediction of ^{13}C NMR chemical shifts of substituted benzenes. Monatsh. Chem./Chem. Monthly 130(9), 1089-1095.

Meiler, J., Meusinger, R., Will, M., 2000. Fast determination of ^{13}C NMR chemical shifts using artificial neural networks. J. Chem. Inf. Comput. Sci. 40(5), 1169-1176.

Meiler, J., Will, M., 2001. Automated structure elucidation of organic molecules from ^{13}C NMR spectra using genetic algorithms and neural networks. J. Chem. Inf. Comput. Sci. 41(6), 1535-1546.

Meiler, J., Kock, M., 2004. Novel methods of automated structure elucidation based on ^{13}C NMR spectroscopy. Magn. Reson. Chem. 42(12), 1042-1045.

Munk, E.M., 1998. Computer-based structure determination: Then and now. J. Chem. Inf. Comput. Sci. 38(6), 997-1009.

Penchev, P.N., Schulz, K-P., Munk, M.E., 2012. INFERCNMR: A ^{13}C NMR interpretive library search system. J. Chem. Inf. Model. 52(6), 1513-1528.

Peng, C., Yuan, S.G., Zheng, C.Z., Hui, Y.Z., Wu, H.M., Ma, K., Han, X.W., 1994. Application of expert system CISOC-SES to the structure elucidation of complex natural products. J. Chem. Inf. Comput. Sci. 34(4), 814-819.

Pretsch, E., Furst, A., Robien, W., 1991. Parameter set for the prediction of the ^{13}C NMR chemical shifts of sp^2- and sp-hybridized carbon atoms in organic compounds. Anal. Chim. Acta 248(2), 415-428.

Steinbeck, C., 2004. Recent advances in automated structure elucidation of natural products. Nat. Prod. Rep. 21(4) 512-518.

Svozil, D., Pospichal, J., Kvasnicka, V., 1995. Neural network prediction of carbon-13 NMR chemical shifts of alkanes. J. Chem. Inf. Comput. Sci. 35(5), 924-928.