COMPUTATIONALLY INEXPENSIVE GIAO ^{13}C NMR CALCULATIONS COUPLED WITH ARTIFICIAL NEURAL NETWORK PATTERN RECOGNITION AS A NEW STRATEGY TO AVOID CHASING MOLECULES THAT WERE NEVER THERE

Ariel M. Sarotti

Instituto de Química Rosario (CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario. Suipacha 531, (2000) Rosario, Argentina.

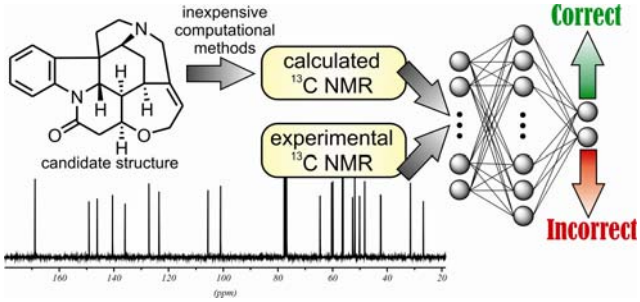
sarotti@iquir-conicet.gov.ar

Keywords: GIAO NMR calculations, artificial neural networks, structural elucidation

Abstract Speech

Structure elucidation relies heavily on modern NMR spectroscopy. Nevertheless, despite a steady advancement in complex multidimensional NMR experiments and more powerful spectrometers has been achieved over the course of the last few decades, several examples of structural and stereochemical misassignments are still appearing in the literature.[1] As pointed out by Nicolaou, incorrectly assigned natural products complicate the assessment of biosynthetic schemes and can also have profound consequences both in terms of time and money if a research group is willing to venture into their total synthesis.[1a] The accurate calculation of NMR chemical shifts with quantum chemical methods has laid a solid foundation to solve, at least in part, some of these problems.[2] The typical approach involves the calculation of the chemical shifts for all the candidate structures to identify the molecular arrangement that best matches the experimental data.[2]

In this work, GIAO NMR chemical shift calculations coupled with trained artificial neural networks (ANNs) have been shown to provide a powerful strategy for simple, rapid and reliable identification of structural misassignments of organic compounds using only one set of both computational and experimental data. The geometry optimization, usually the most time-consuming step in the overall procedure, was carried out using computationally inexpensive methods (MM+, AM1 or HF/3-21G) and the NMR shielding constants at the affordable mPW1PW91/6-31G(d) level of theory. As low quality NMR prediction is typically obtained with such protocols, the decision making was foreseen as a problem of pattern recognition. Thus, given a set of statistical parameters computed after correlation between experimental and calculated chemical shifts the classification was done using the knowledge derived from trained ANNs. The training process was carried out with a set of 200 molecules chosen to provide a wide array of chemical functionalities and molecular complexity, and the results were validated with a set of 26 natural products that had been incorrectly assigned along with their 26 revised structures. The high prediction effectiveness observed makes this method a suitable test for rapid identification of structural misassignments, preventing not only the publication of wrong structures but also avoiding the consequences of such a mistake.[3]



REFERENCES

[1] (a) Nicolaou, K. C.; Snyder, S. A. *Angew. Chem. Int. Ed.* **2005**, *44*, 1012. (b) Suyama, T. L.; Gerwick, W. H.; McPhail, K. L. *Bioorg. Med. Chem.* **2011**, *19*, 6675.

[2] (a) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. *Chem. Rev.* **2012**, *112*, 1839-1862. (b) Sarotti, A. M.; Pellegrinet, S. C. *J. Org. Chem.* **2012**, *77*, 6059. (c) Sarotti, A. M.; Pellegrinet, S. C. *J. Org. Chem.* **2009**, *74*, 7254.

[3] Sarotti, A. M. *Org. Biomol. Chem.* **2013**, *11*, 4847.