CHINESE JOURNAL OF ANALYTICAL CHEMISTRY
Volume 34, Issue 3, March 2006
Online English edition of the Chinese language journal

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Cite this article as: Chin J Anal Chem, 2006, 34(3), 329–332.

RESEARCH PAPER

Simulation of ^{13}C Nuclear Magnetic Resonance Spectra for Derivatives of Bases and Nucleotides

Liang Guizhao[1,2,3], Mei Hu[1,3], Zhou Yuan[1,2,3], Zhou Peng[2], Li Zhiliang[1,2,*]

[1]Key Laboratory of Biomechanics and Tissue Engineering at Chongqing University, Chongqing 400044, China
[2]College of Chemistry and Chemical Engineering, Chongqing University, Chongqing 400044, China
[3]College of Bioengineering, Chongqing University, Chongqing 400044, China

Abstract: A quantitative structure spectroscopy relationship (QSSR) model of ^{13}C nuclear magnetic resonance (NMR) in 630 carbon atoms of 81 derivatives of bases and nucleotides has been developed using atomic electronegativity interaction vector (AEIV) and atomic hybrid state index (AHSI) combined with γ calibration. The prediction correlation coefficient (R) value of the QSSR model based on multiple linear regression analysis was 0.970. The stability and prediction capacity of the QSSR model have been tested using the leave-one-out and leave-group (molecular)-out cross-validation methodology. The correlation coefficients *R* obtained were 0.969 and 0.969, respectively. Excellent results were obtained by successfully predicting the correlation between the chemical shifts and the structural parameters for three series of derivatives of bases and nucleotides. The correlation coefficients *R* were 0.969, 0.921, and 0.884, which showed that the predictive potential of the proposed models was quite robust.

Key Words: Atomic electronegativity interaction vector (VAEI); Atomic hybrid state index (AHSI); γ Calibration; ^{13}C Nuclear Magnetic Resonance (NMR) spectroscopy; Quantitative structure spectroscopy relationship (QSSR)

1 Introduction

Simulation of nuclear magnetic resonance (NMR) spectra using quantitative structure spectroscopy relationship (QSSR) that was established by the computational method according to the molecular structure can help scientists in structural analysis. It has been reported that several measures can be used for spectra simulation[1], particularly, the measure to develop a mathematical model with a key idea for selecting appropriate descriptors to rationally express the molecular structures is often used. The usual methods of expressing carbon-13(^{13}C) are Randic[2] index, Wiener[3,4] index, molecular connectivity[3] index, and the Padmakar–Ivan[5] parameter calculation, etc. The atomic distance-edge vector for expression of the base molecules was advocated by our group[6] to simulate the chemical shift of the ^{13}C NMR spectra and satisfying results were obtained. The atomic electronegativity interaction vector (AEIV) and the atomic hybrid state index (AHSI) [7] were further proposed to predict the chemical shifts for ^{13}C NMR of 20 natural amino acids. The results indicated that AEIV and AHSI could well express the environment and the hybrid state of carbons. The microenvironments and hybrid states for 630 carbons of bases and nucleosides and their derivatives were then expressed using AEIV and AHSI while combining with γ calibration, and the QSSR model developed using multiple linear regression was employed to predict ^{13}C chemical shifts of three kinds of derivatives from pyrimidine and purine.

2 Principles and methods

2.1 Atomic electronegativity interaction vector, γ calibration, and atomic hybrid state index

Chemical shift is correlated with the chemical microenvironment and the hybrid state of the atom itself. For chemical microenvironments, in this study, atoms were divided into five categories according to families of elementary periods because they were correlated with

* Corresponding author. Email: zlli2662@163.com; Tel: +86 23-65106677
This work was supported by Chongqing Basic and Applied Fund [01-3-6], State Chunhui Project Fund (SCPF, 98-3-8) and Chongqing University Innovation Fund for graduate students [CUIF, 06-01-A]
Received 26 January 2005; accepted 20 September 2005

electronegativity and six squares of bond length peripheral atoms[8], and atoms with different chemical properties had different effects on the target atoms. The computational method of electronegativity for five different atoms against the target atoms was then achieved. Atomic electronegativity based on the Pauling [9] scale was defined as the ratio of the atomic elecronegativity to that of the carbon atom, for example, the relative electronegativity of the oxygen atom was 3.44/2.55=1.349. Similarly, the relative bond length between two atoms was defined as the ratio of the length of the bond investigated to the length of the C–C bond[9], for example, the relative bond length of C–O was 0.122 nm/0.154 nm = 0.792. Thus, the atomic elctronegativity and bond length wereunified, and the atomic electronegativity interaction vector (AEIV) was proposed and calculated as follows:

$$v_{ik} = \sum_{j\in k, j\neq i}^{all(j)} \frac{\chi_j}{d_{i,j}^{\ 6}} \quad (1\leq k\leq 5)$$

Here, v_{ik} is the AEIV descriptor for the No. i atom, j represents all atoms belonging to k type ($j \neq i$); X is the atomic relative electronegativity, d_{ij} is the relative distance between the i and j atoms, i.e. the sum of the relative bond length of the minimum connecting path for the i and j atoms. The γ calibration [8] was also an important factor related to the chemical shift, and hence it was considered in this study. For the characterization of the hybrid state of the atom itself, the atomic hybrid state index (AHSI) was derived from the atomic intrinsic state (I) proposed by Hall [10] *et al.* and was introduced and modified. AHSI was calculated as follows:

$$\mathrm{AHSI} = \sqrt{v}\cdot\left(\left(2/n\right)^2\delta_{\sigma+\pi}+1\right)/2\delta_{\sigma}$$

Here, v is the number of electrons in the valence shell of that atom; n represents the principal quantum number; $\delta_{\sigma+\pi}$ is the total number of electrons forming σ and π bonds; and δ_{σ} is the number of electrons forming σ bonds.

The differences from primary definition by Hall *et al*. were that AHSI included the coefficient of $(v)^{1/2}/2$ and did not detract the electrons forming the bond with hydrogen during calculation of $\delta_{\sigma+\pi}$ and δ_{σ}, and showed that AHSI emphasized on hybrid types[7].

2.2 Development of the linear QSSR model including seven parameters

The 630 ^{13}C NMR chemical shifts of 81 molecules contained common five bases, eight nucleosides and their derivatives from the literature [11–14]. These atoms were numbered, and nonhydrogen atomic number, the atomic attribution, and the connection relation in every molecule were fed into a computer. The atomic attribution[15] represented the atomic characteristics and the chemical bond property of the nonhydrogen atoms. The AEIV descriptors and the γ calibration were calculated by using AEIVMP.exe, which was programmed using the C^{++} language. The models about the chemical shifts and the AEIV (v_1–v_5), γ(v_6) and AHSI (v_7) descriptors were constructed based on multiple linear regression (MLR). The internal predictive ability of the model was evaluated using the leave one out (LOO) and leave-molecule-out (LG(M)O) cross-validations.

2.3 Simulation of ^{13}C NMR for three groups of base and nucleoside and their derivatives

An excellent model must have a higher estimation capability for not only internal samples but also external samples. Three groups of base and nucleoside and their derivatives as test samples were selected from references [16–18] to predict their chemical shifts using the model, and furthermore, their external predictive ability was evaluated.

3 Results and discussion

3.1 Evaluation for QSSR model

81 molecules containing common five bases, eight nucleosides and their derivatives were evaluated. It was also noted that 10 molecules (Table 1) in the literature [13] exhibited molecular tautomerism, and thereby the presence of isomers results in misinterpretation of the evaluation system. The 81 molecules containing 2 isomers were then treated as training set to develop M1 and M2 with seven parameters, respectively (Table 2).

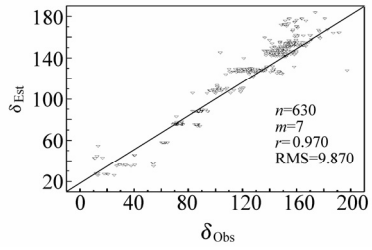

Fig. 1 Plot of prediction versus observation

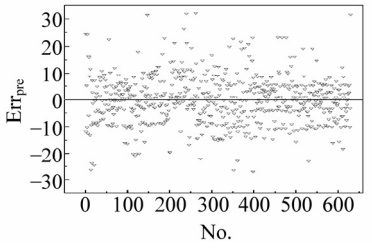

Fig. 2 Distribution of estimated errors

As shown in Table 2, the simulative result of M1 was superior to that of M2. 86 Carbons of 10 molecules or their isomers were picked, and the remaining 544 carbons were used to develop the model to investigate how 10 molecules or their isomers influence the simulative result of the models, and the chemical shifts of 86 carbons were predicted. R of 0.986 and 0.985, RMS of 6.211 ppm and 6.280 ppm were obtained by M1 and M2, respectively. The results indicated