An evaluation of least-squares fits to COSY spectra as a means of estimating proton–proton coupling constants. I. Simulated test problems

Ju-Xing Yang and Timothy F. Havel*

Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, U.S.A.

Received 24 November 1993
Accepted 6 July 1994

Keywords: 2D NMR; Simulation of spectra; COSY; Correlated spectroscopy; Proton–proton coupling constants; Scalar coupling; Parameter estimation

SUMMARY

A computational method is described that takes an initial estimate of the chemical shifts, line widths and scalar coupling constants for the protons in a molecule, and refines this estimate so as to improve the least-squares fit between an experimental COSY spectrum and the spectrum simulated from these parameters in the weak-coupling approximation. In order to evaluate the potential of such refinements for estimating these parameters from COSY experiments, the method has been applied to a large number of sample problems which were themselves simulated from standard conformations of the amino acids, along with 25 near-native conformations of the protein bovine pancreatic trypsin inhibitor. The results of this evaluation show that: (i) if the chemical shifts are known to within ca. 0.01 ppm and no noise or artifacts are present in the data, the method is capable of recovering the correct coupling constants, starting from essentially arbitrary values, to within 0.1 Hz in almost all cases. (ii) Although the precision of these estimates of the coupling constants is degraded by the limited resolution, noise and artifacts present in most experimental spectra, the large majority of coupling constants can still be recovered to within 1.0 Hz; the local minimum problem is not made significantly worse by such defects in the data. (iii) The method assigns an 'effective' line width to all the resonances, and in the process can resolve overlapping cross peaks. (iv) The method is not capable of determining the chemical shifts a priori, due to the presence of numerous local minima in the least-squares residual as a function of these parameters.

INTRODUCTION

Numerous procedures have been developed for estimating scalar coupling constants from homonuclear COSY-like experiments. These can be divided into roughly three classes. In the first class, special pulse sequences are used to make it relatively easy to measure the coupling constants.

*To whom correspondence should be addressed.

0925-2738/$ 10.00 © 1994 ESCOM Science Publishers B.V.
from the fine structures of the cross peaks. For example, E.COSY (Griesinger et al., 1986) and z-COSY (Oschkinat et al., 1986) experiments yield spectra with simplified in-phase multiplet patterns for all the cross peaks. In the second class, data processing methods are used to extract the coupling constants from the spectra. Examples here include the DISCO method (Kessler et al., 1985), a related method involving NOESY spectra (Ludvigsen et al., 1991), and elimination of the peak width between absorptive and dispersive components (Kim and Prestegard, 1989). In the third class, simulations of various sorts are used to find values for the coupling constants that reproduce the observed fine structures. These methods have usually been developed to work in combination with other experiments and/or data processing techniques, for example TOCSY (Titman and Keeler, 1990), E.COSY (Smith et al., 1991), NOESY (Szyperski et al., 1992), and the 'J-doubling' method (Jones et al., 1993). Of course, if the molecule of interest can be isotopically labeled many heteronuclear methods are also applicable, but in this paper we limit ourselves to homonuclear experiments.

Once one has the correct coupling constants, density matrix or product operator calculations (Ernst et al., 1987) can be used to simulate the experimental spectra and thereby confirm them. Alternatively, one can carry out the simulation in the frequency domain, by diagonalizing the Hamiltonian directly (Widmer and Wüthrich, 1986). In principle, it should also be possible to refine the coupling constants so that the spectra obtained in this way agree with the experimental spectra, and perhaps even to automatically extract the coupling constants from the data. This has successfully been done with the E.COSY experiment (Mádi and Ernst, 1988), but to date the method has been too computationally demanding to be applied to entire spectra of large molecules such as proteins.

In this paper we describe and validate a simplified and computationally more efficient version of the frequency domain approach. This purely phenomenological procedure simulates 2D COSY spectra directly from the coupling constants, chemical shifts and line widths, without attempting to reproduce the full quantum-mechanical evolution of the spin systems involved. The efficiency of the procedure is greatly improved by using a novel matrix decomposition of 2D NMR spectra (Have1 et al., 1994), which also makes it relatively easy to compute the derivative of the spectrum with respect to these parameters. The availability of these derivatives, in turn, makes it possible to minimize the sum over all points of the squared differences between the observed and simulated spectra with respect to the parameters, using a standard conjugate gradient algorithm. This algorithm has the advantage of requiring computer storage that grows only linearly with the number of spins, and is computationally quite efficient.

One obvious defect of the procedure for the purpose of estimating the coupling constants is that, as currently implemented, it uses the weak-coupling approximation for all pairs of spins. This is, of course, true for most existing methods of estimating homonuclear coupling constants. Our primary goal in this paper, however, is simply to evaluate the severity of three other, potentially more serious, obstacles to the procedure. The first of these stems from the fact that functions measuring the difference between observed and calculated spectra might possess numerous local minima, so that it could be very difficult to find reasonably good fits by minimizing them. The second lies in the fact that least-squares methods are only assured of being able to alleviate the effects of random errors in the data, whereas actual 2D NMR spectra, particularly of macromolecules in aqueous solution, are afflicted by numerous artifacts. Finally, the least-squares residual may change slowly with changes in the coupling constants and other parameters,