ON-LINE COMPUTER-ASSISTED ANALYSIS OF 220 MHz NMR DATA OF PROTEIN IMIDAZOLE RESONANCES

J.S. Cohen, R.I. Shrager, M. McNeel, and A.N. Schechter

Division of Computer Research; and Laboratory of Chemical Biology, NIAMD; NIH, Bethesda, Maryland

Received May 27, 1970

The application of an interactive computer curve-fitting program to 220 MHz NMR spectra and titration data of protein imidazole resonances is described. The pH titration data for the two active site histidines of RNase require two theoretical curves each to obtain a satisfactory fit.

The application of NMR* data to the interpretation of protein phenomena is limited by the ability to resolve overlapping resonances. Computer curve-fitting can be a significant aid in analysing the data available from the NMR spectra of macromolecules (1). To date some curve simulation (2) and curve-fitting with analog curve generators (3) has been carried out, but these approaches are of limited applicability. The only published curve-fitting program for NMR spectra of which we are aware is DECOMP of Keller et al. (4) which is designed for sub-spectral analysis.

A flexible non-linear regression curve-fitting program (MODELAIDE) has been written by one of us (5). This is an extension of the Levenberg (6) and Marquardt (7) methods to include linear constraints by means of quadratic programming. Further, the program is designed to be interactive with the user at an on-line display console (IBM 2250 with an IBM 360/50 computer). We present here several applica-

*Abbreviations: NMR, nuclear magnetic resonance; RNase, bovine pancreatic ribonuclease A (Worthington); Nase, Staphylococcal nuclease; His, histidine; Gly, glycine; TMS, tetramethylsilane.

tions of these programming methods to the interpretation of protein NMR data.

A. Spectra: Digitized time-averaged 220 MHz NMR spectra of the downfield resonances of proteins were fitted to a sum of Lorentzian curves. Initially peak area, width at half-height and chemical shifts were estimated; the program was then allowed to alter these parameters until convergence was achieved. The fit was displayed, and, if satisfactory, recorded on an off-line CALCOMP plotter.

Curve-fitting, of all kinds, is subject to potential ambiguities. Two superimposed peaks of similar width cannot be uniquely resolved (8).

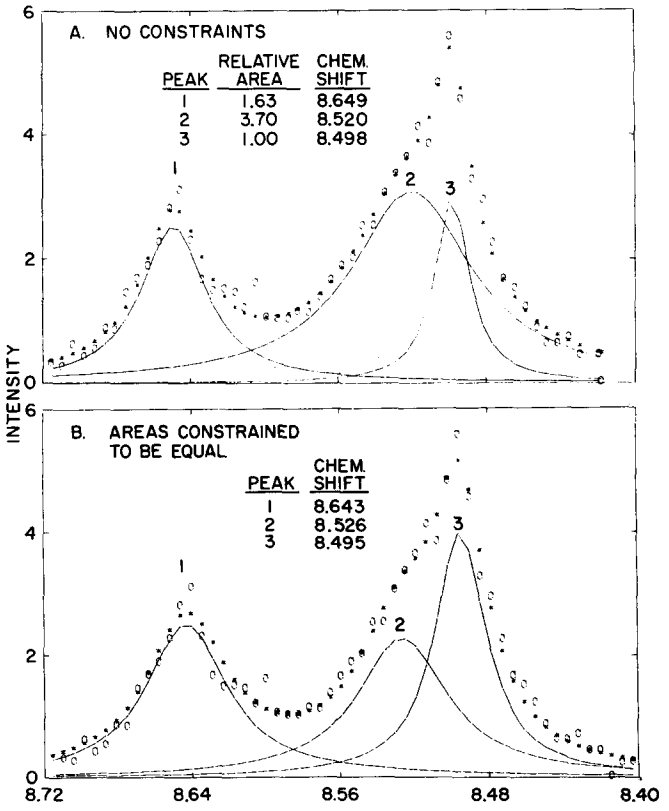

Fig. 1 Portion of a 220 MHz NMR Spectrum of RNase at pH 4.20 in 0.3 M NaCl-D_2O, fitted with Lorentzian curves with (B) and without (A) area constraints. These are CALCOMP plots; 0 = observed, * = calculated, ——— = components. Horizontal scale is chemical shift in ppm.

However, when peaks are overlapping and an inflexion (or shoulder) exists a useful fit may be made. This is illustrated in Fig. 1, both with and without the constraint of equal area, for three imidazole C2-H resonances of RNase. The use of the constraint (Fig. 1B) is more physically meaningful in this case because the areas were found to be equal at lower and higher pH values where the peaks are resolved, and the three resonances each arise from a single proton. Such results of curve-fitting have been found useful in determining chemical shift and peak width as a function of pH.

A further possible ambiguity in curve-fitting arises in the choice of a base-line. This is particularly important when dealing with time-averaged spectra of proteins due to the HDO peak, the large number of C-H resonances, and the partially exchanged N-H resonances in the downfield region (9). Fig. 2 shows the downfield region of Nase spectra fitted with two choices of base-line. A single large resonance upfield (Fig. 2A) from the region in question gives a fit in which the four imidazole resonances (H 1-4) are essentially equal in area, as expected. In Fig. 2B an alternative choice of base-line requires one less peak, but results in a much reduced relative area for H-4. More extensive analysis of this region of spectra of Nase is reported elsewhere (1).

Because of these and other possible complications, it is important that curve-fitting be interactive with the user. The system described here is fully applicable to NMR spectra of small molecules, containing sharp resonances. In such cases some of the problems described above may not arise, although better results may be obtained when dealing with broader protein resonances than with sharp resonances containing fewer points.

B. Titration Data: The C2-H resonances of imidazole side chains are known to shift with pH, and to produce a titration curve reflecting

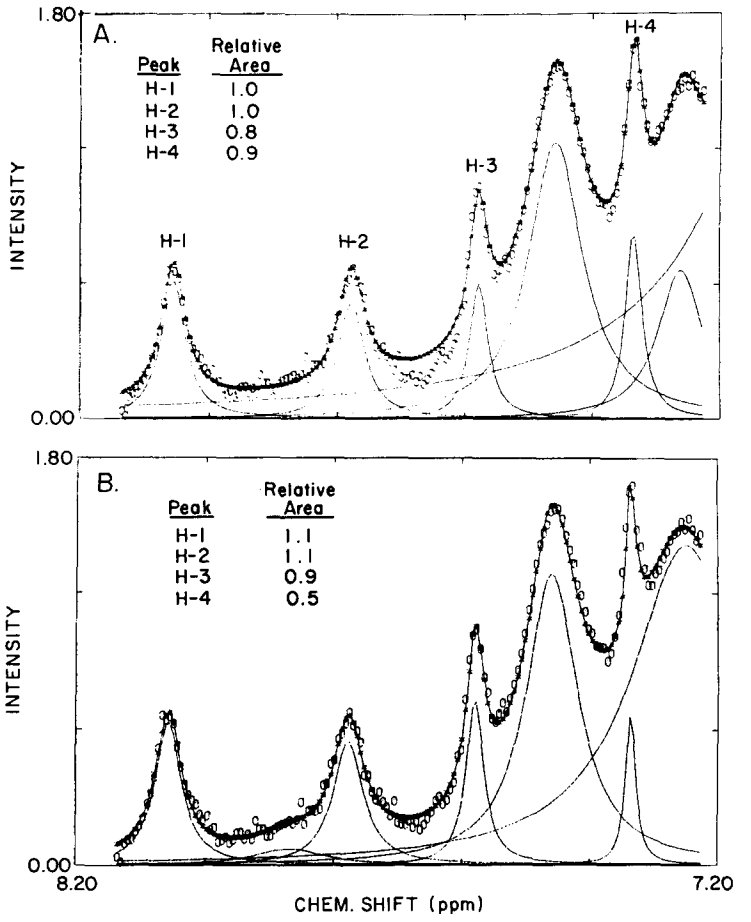

Fig. 2 220 MHz spectrum of Nase at pH 6.39 in 0.1 M NaCl-D_2O fitted with multiple Lorentzian curves, but with two different baselines.

---

the protonation of the imidazole nitrogen atoms (10-14). In the simplest case this curve is described by the Henderson-Hasselbach equation:

$$pH = pK + \log [B]/[HB] \qquad (1)$$

Chemical shift may be considered to be a linear function of the concentration of each species and the equation may be solved for the observed chemical shift at any pH (1).

Curve-fitting of the titration data is of significant value in determining pK values precisely and in testing the fit of the data

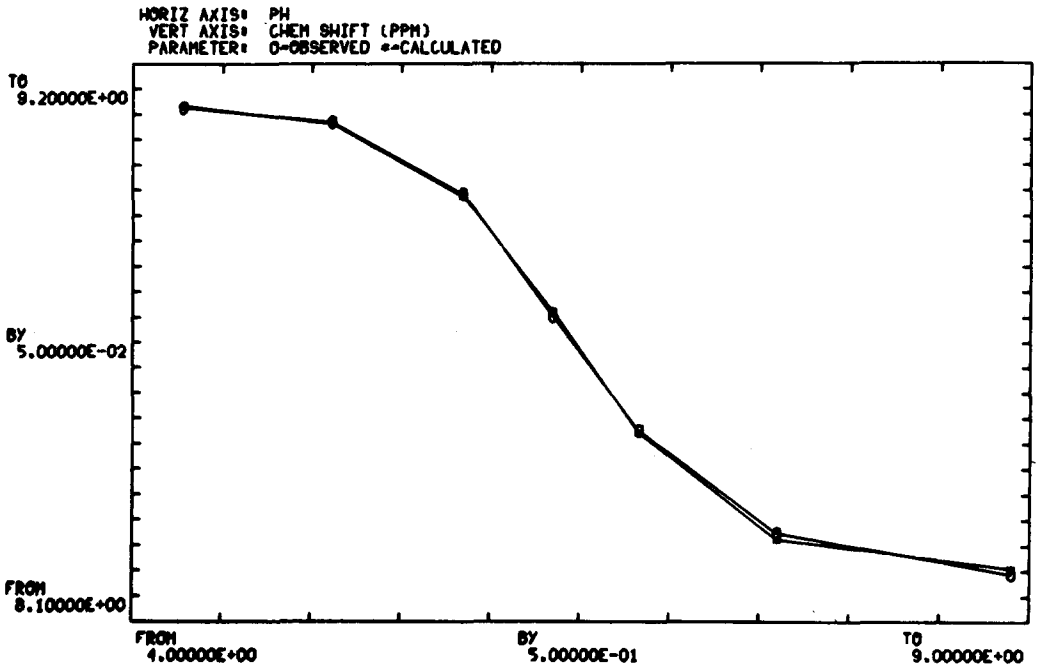

Fig. 3 Titration curve of the imidazole C2-H resonance of Gly-His-Gly in 0.1 M NaCl-D_2O at 45°. Data obtained at 100 MHz. O = observed, * = calculated.

---

to the Henderson-Hasselbach model. Excellent results were obtained, as shown in Fig. 3, for the model peptide Gly-His-Gly, with a calculated apparent pK of 6.48. Previous data obtained at 100 MHz for the single histidine residue of human lysozyme (13), and at 220 MHz for the four histidine residues of staphylococcal nuclease (1) all give equally good fits to the theoretical curves. On the other hand, we consistently find that a single theoretical curve does not describe well our 220 MHz data for the His-119 of RNase (Fig. 4). Roberts et al. (12) in their studies at 100 MHz noted that this curve was somewhat asymmetric at low pH*, although they interpreted their data

---

*We do not find the use of a Linderstrøm-Lang-type addition to equation (1) used by these authors (12) to be necessary.

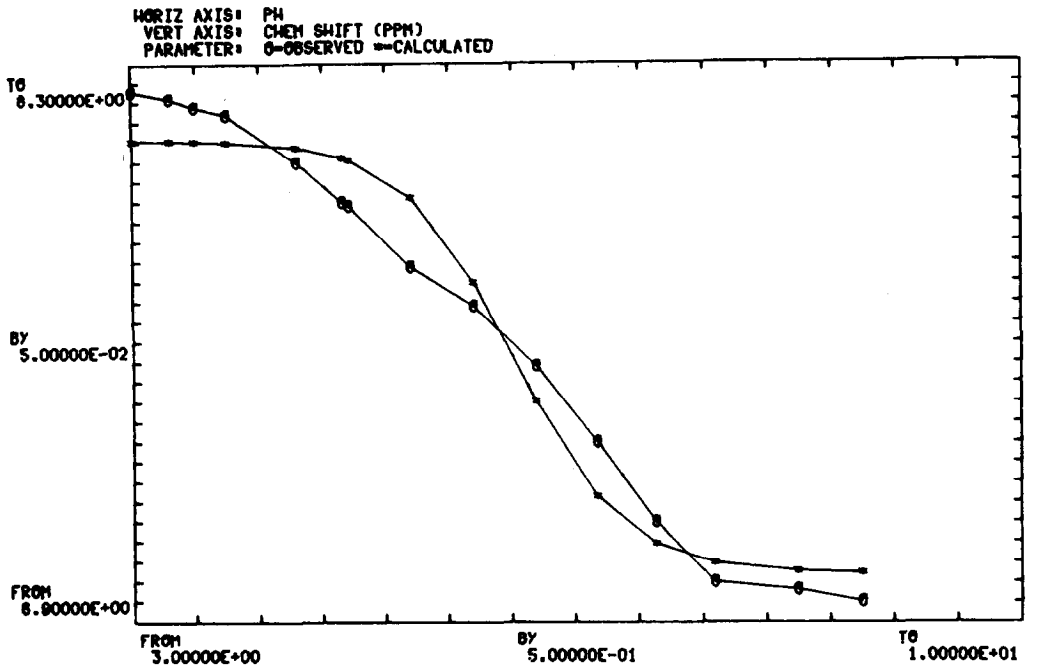

Fig. 4 Titration data for H-3 (His-119) of RNase in 0.1 M NaCl-D_2O, fitted with one theoretical titration curve (this is the best fit possible). 0 = observed, * = calculated.

as single transitions. Ruterjans and Witzel (14), conversely, have noted at 100 MHz double inflexions in both the His-12 and 119 titration curves. They suggest that this indicates hydrogen-bonding between these histidine residues in the active site, which may be of significance for the mechanism of action of the enzyme.

We find good fits to the titration data for His-119 at 220 MHz only if we use two theoretical curves, as shown in Fig. 5. The two calculated apparent pK values in this case are 4.57 and 6.68. Using a different method of calibration[+] a very similar result was obtained, giving pK's 4.54 and 6.71 respectively.

[+]External calibration: a) a Wilmad insert of TMS, and b) 6% TMS in CCl_4 substituted for the sample.

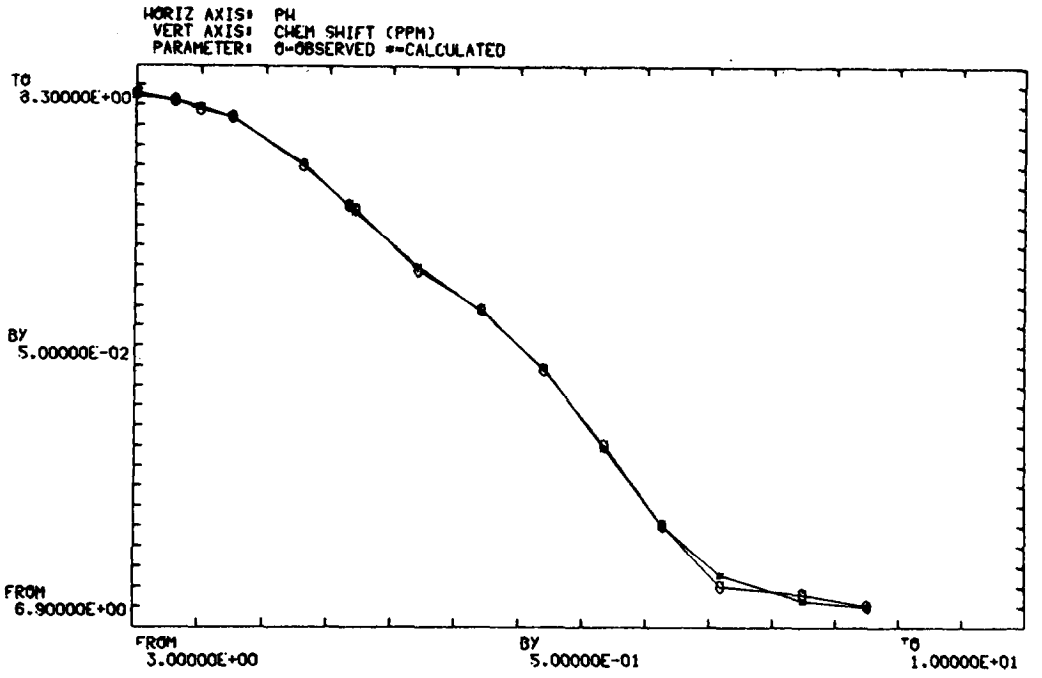

Fig. 5 Titration data as in Fig. 3, but fitted with two theoretical titration curves. 0 = observed, * = calculated.

The titration curves of His-12 and 119 of RNase cross each other at several pH values (as illustrated in the overlapping peaks of Fig. 1). We have found curve-fitting to be useful in providing a more objective criterion in establishing the true continuity at these points (1). However, in the case of RNase the concatenation is fairly obvious visually, and our interpretation of the continuities is in agreement with those of Meadows et al. (11) and Ruterjans and Witzel (14). With this assumed continuity the data for His-105 can be fitted with a single curve with pK 6.75, but that of His-12 also requires two curves with pK's of about 4.8 and 7.0.

These results indicate that His-12 and 119 are either, a) interacting with each other, b) interacting with one or two adjacent groups, or c) are both responding to a local pH-dependent conformational

change. For reasons to be presented later, we favor the first explanation, and are pursuing its implications for the mechanism of action of ribonuclease (15) utilizing substrate analogs with phosphonate bonds (16).

Experimental details are as described elsewhere (1).

References

1. J.S. Cohen, R.I. Shrager, M. McNeel, and A.N. Schechter, submitted for publication.
2. I. Putter, A. Barreto, J.L. Markley, and O. Jardetzky, Proc. Natl. Acad. Sci., 64, 1396 (1969).
3. C.C. McDonald and W.D. Phillips, J. Am. Chem. Soc., 91, 1513 (1969).
4. W.D. Keller, T.R. Lusebrink and C.H. Sederholm, J. Chem. Phys., 44, 782 (1966).
5. R.I. Shrager, Computer Graphics (SIGGRAPH-ACM) 3, No. 3, p. 17 (1969); J. Assoc. Comp. Mach., in press.
6. K. Levenberg, Quart. Applied Maths. No. 2, p. 164 (1944).
7. D.W. Marquardt, SIAM Journal, 2, 431 (1963).
8. J.W. Perram, J. Chem. Phys., 49, 4245 (1968).
9. J.D. Glickson, C.C. McDonald and W.D. Phillips, Biochem. Biophys. Res. Comm. 35, 492, (1969).
10. C.C. McDonald and W.D. Phillips, J. Am. Chem. Soc. 85, 3736 (1963).
11. D.H. Meadows, O. Jardetzky, R.M. Epand, H.H. Ruterjans, and M.A. Scheraga, Proc. Natl. Acad. Sci., 60, 766 (1968).
12. G.C.K. Roberts, D.H. Meadows and O. Jardetzky, Biochemistry, 8, 2053 (1968).
13. J.S. Cohen, Nature, 223, 43 (1969).
14. H. Ruterjans and H. Witzel, Europ. J. Biochem., 9, 118 (1969).
15. G.C.K. Roberts, E.A. Dennis, D.H. Meadows, J.S. Cohen and O. Jardetzky, Proc. Natl. Acad. Sci.,62, 1151 (1969) and references therein.
16. N.P. Damodarran, G. Jones and J. Moffatt, unpublished results.