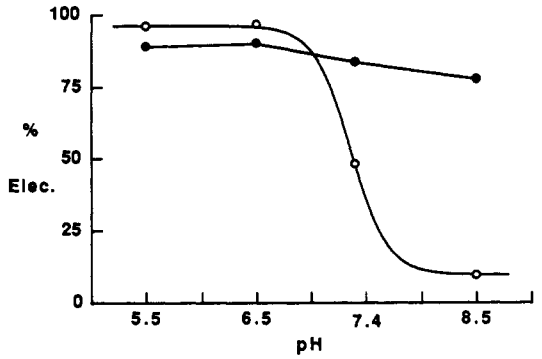
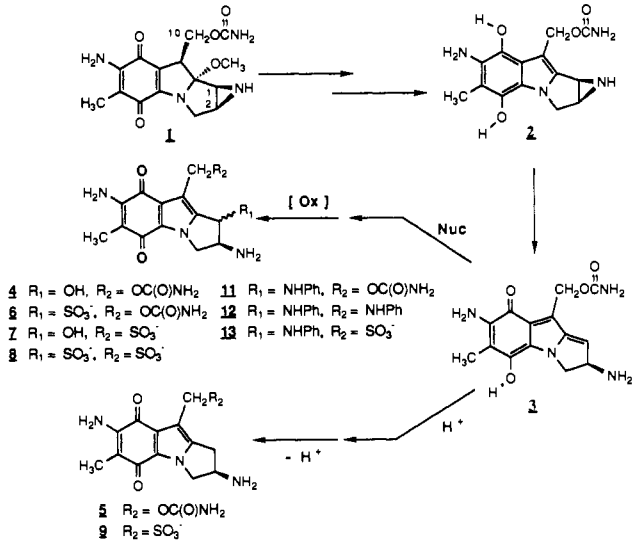


Figure 1. Percent C(1)-electrophilic mitosene products generated as a function of pH. The percent mitosene C(1)-electrophilic products generated with limiting amounts (0.2 equiv) of $Na_2S_2O_4$ (method A) was computed using the formula %(**5** + **9**)/%(**4** + **5** + **6** + **7** + **9**) (○); and with excess (1.2–2.0 equiv) $Na_2S_2O_4$ (method B), using the formula %(**5** + **9**)/%(**4** + **5** + **6** + **7** + **8** + **9**) (●). The reactions were run in duplicate, and the data was obtained using HPLC protocol 1.[10a]

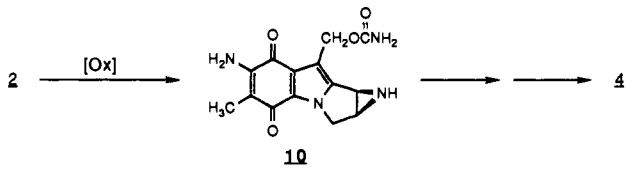
Scheme I. Previously Proposed Pathway for the Reduction of **1**[a]



[a] Reference 6a.

reaction mixture. The two procedures provided dramatically different pH–product profiles. When we limited the amounts of reductant, we observed the previously noted[6] change from predominantly electrophilic (**5**) to nucleophilic (**4**) compounds with increasing pH, whereas with excess $Na_2S_2O_4$ the C(1)-electrophilic adducts **5** and **9** were the major products (77.2–89.9%) at all pH values in the absence of external nucleophiles.[11]

To determine the origin of the product dependency as a function of pH with method A we monitored the reactions using a second HPLC protocol, which employed a more "basic" eluant system.[10b] At pH 8.5, we saw that *anti*- and *syn*-**10**[12] and **4** accounted for 24.9% and 62.4%, respectively, of the product profile after exclusion of unreacted mitomycin C (64.8%). Reducing the reaction time from 6 min to 20 s and then adding first oxygen and then aniline (10 equiv) led to the formation of **11** and **4** in a 2.9:1 ratio (combined product yield after exclusion of **1**: 87.2%) and decreased consumption of mitomycin C (16.9%). These results demonstrated that, in basic solutions where reduction of **1** is incomplete, the acid-promoted aziridine ring opening of **2** to **3** was slow enough to permit reoxidation of this species to **10**.[13] Compound **10** furnishes **4** rapidly in water (pH < 7.0).[12]



The propensity of reductively activated (i.e., 1.2–2.0 equiv of $Na_2S_2O_4$) mitomycin C to undergo C(1)-electrophilic substitution transformations in the presence of a competing nucleophile was assessed by incorporating excess aniline (4 equiv) in the original reaction mixture. Under these conditions, substantial amounts of C(1)-anilino nucleophilic products (**11**, **12**, **13**) were observed at pH 5.5 (69.9%), whereas the C(1)-electrophilic adduct **9**[8] predominated at pH 8.5 (70.8%). These results supported an earlier finding that mitomycin C–DNA cross-linking processes were promoted at lower pH values.[3]

These collective findings document that reductively activated mitomycin C in water in the absence of external nucleophiles undergoes principally electrophilic substitution processes. Previous observations[6] reporting the primary formation of *cis*- and *trans*-1-hydroxy-2,7-diaminomitosenes (**4**) at pH 7 and above are now principally, although not exclusively, attributed to the hydrolysis of the *oxidized* 7-aminoaziridinomitosene (**10**).[12] The decreased reactivity of **2** under mildly alkaline conditions raises the intriguing possibility that the effects of the medium or cellular constituents might influence drug activation and bonding by permitting **2** sufficient lifetime to translocate from the site of reduction[14] to the cellular nuclei before DNA bonding.

Acknowledgment. This work was funded by NIH Grant RO1CA29756 and the Robert A. Welch Foundation (E-607). We gratefully acknowledge Drs. A. M. Casazza and W. Rose of the Bristol-Myers Squibb Laboratories for their generous gift of **1**.

(11) $Na_2S_2O_4$ reduction (1.5–2.0 equiv) of **4** under comparable conditions gave only **7**.[8]

(12) Compound **10** exists as a mixture of anti and syn isomers. These conformers differ in the orientation of the aziridine N–H proton in relation to the pyrrolo[1,2-*a*]indole ring system. Han, I.; Kohn, H. *J. Org. Chem.* **1991**, *56*, 4648. Kinetic studies performed in buffered methanol gave the following solvolysis $t_{1/2}$ values for **10**: "pH" 7.0, 3 min; "pH" 8.5, 228 min.

(13) Oxidation of **2** may have occurred by disproportionation with **1** or by reaction with adventitious amounts of oxygen.

(14) Keyes, S. R.; Rockwell, S.; Sartorelli, A. C. *Cancer Res.* **1984**, *45*, 3642.

Site-Specific Isotopic Labeling of Proteins for NMR Studies

Jonathan A. Ellman, Brian F. Volkman, David Mendel, Peter G. Schultz,*,† and David E. Wemmer*

Department of Chemistry
University of California
Berkeley, California 94720
Center for Advanced Materials
Materials Science Division and
Laboratory of Chemical Biodynamics
Lawrence Berkeley Laboratory
Berkeley, California 94720

Received June 22, 1992

NMR spectroscopy is a sensitive, site-specific probe of biomolecular structure. For relatively small proteins and peptides, the ^1H resonances can be assigned using the sequential method.[1] However, there are many cases, especially larger proteins, in which the spectra are too complex for complete, systematic resonance assignments. In some cases, assignments can be made by selective isotopic labeling (e.g., uniform incorporation of a ^{13}C-labeled amino acid) in conjunction with site-directed mutagenesis or ^{13}C,^{15}N double labeling of adjacent amino acids.[2,3] However,

* Author to whom correspondence should be addressed.
† Department of Chemistry, University of California, Berkeley, CA 94720.

(1) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley and Sons: New York, 1986.

0002-7863/92/1514-7959$03.00/0 © 1992 American Chemical Society

in many large proteins, protein complexes, and unfolded proteins, resonance overlap and broadening prevents assignments. The ability to synthesize proteins with unnatural amino acids, beyond those specified by the genetic code, makes it possible to isotopically label a single amino acid residue in a protein.[4] We report here the use of this approach to generate a T4 lysozyme (T4L) mutant containing a unique ^{13}C-labeled alanine, for which ^{13}C-filtered proton spectra were obtained in *both* the native and denatured states. This general methodology should be applicable to a variety of NMR measurements in large proteins and protein complexes.

T4 lysozyme is a well-characterized enzyme that has served as a model for studies of protein stability and catalytic mechanism.[5] Its NMR resonances have been assigned through isotope-assisted methods by McIntosh et al.[6] In addition, a number of unnatural amino acids have been substituted at positions 20, 82, and 133 of T4L in order to probe the effects of side-chain and backbone structure on protein stability and catalytic activity.[4a-d] As a test of our ability to site-specifically incorporate isotopic labels into proteins, [^{13}C]alanine was substituted for Ala 82, a surface residue at a break between two α-helices.[7] The mutant protein was generated as previously described[4b] by in vitro suppression of an Ala 82 → TAG nonsense mutation (encoded on the plasmid pT4LA82am)[8] with a chemically aminoacylated suppressor tRNA.[9-11] [^{13}C]Ala was incorporated with approximately 28% suppression efficiency at this site. The isotopically labeled protein was purified by a combination of anion- and cation-exchange chromatography[4b] and judged to be >90% homogeneous by polyacrylamide gel electrophoresis with silver staining. Approximately 150 μg of purified [^{13}C]Ala 82 T4L was obtained from eight 5-mL in vitro reactions to give 30 μL of a 30 μM NMR sample. Control in vitro experiments in the absence of suppressor tRNA and with unacylated suppressor tRNA afforded less than 1% full length T4L.

The upfield portion of the ^1H NMR spectrum[12] of in vitro synthesized T4L is shown in Figure 1A. The distribution and frequencies of upfield-shifted methyl resonances are exactly as expected on the basis of previous ^1H NMR studies of this protein. The ^{13}C-filtered spectrum was obtained by standard methods[13] and is shown in Figure 1B. In contrast to the ^1H T4L spectrum, the C_α proton and methyl resonances are immediately apparent in the filtered spectrum. The chemical shifts agree with those obtained from earlier studies of wild-type T4L.[6] The protein was then denatured by heating to 60 °C. The proton spectrum of the unfolded protein is shown in Figure 1C, indicating the collapse of shift dispersion due to loss of secondary structure and tertiary interactions. For a protein of this size (18.7 kDa), it would normally be impossible to assign the resonances of the denatured state. Figure 1D shows the ^{13}C-filtered spectrum of the upfield region, with the CH_3 resonance clearly visible at the normal random coil shift for alanine. Unfortunately, the random coil α proton chemical shift[14] is the same as that for water at this temperature and could not be detected under the residual HOD peak. The relatively small chemical shift change of the methyl group is consistent with the fact that the Ala 82 methyl group is essentially fully exposed in both the folded and unfolded states.

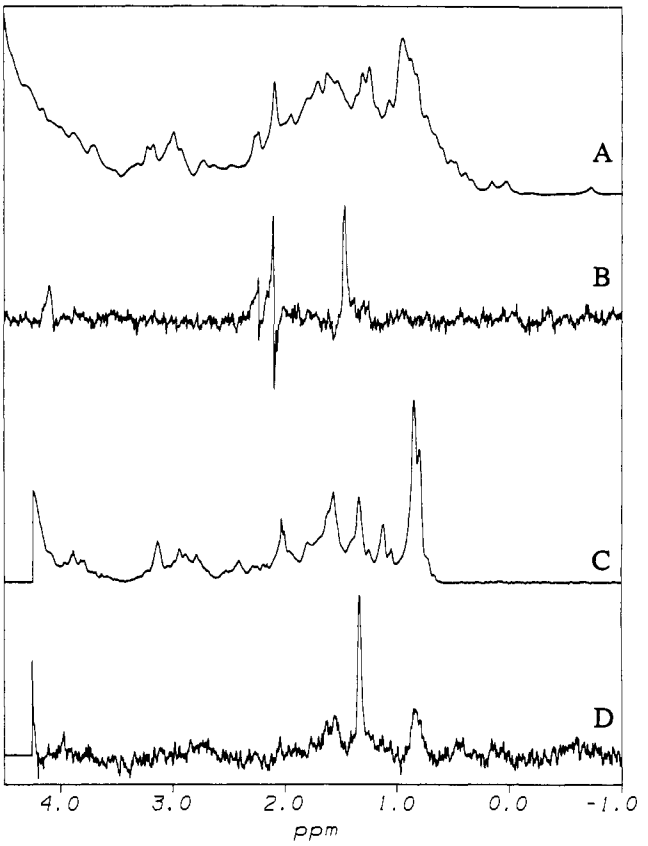


Figure 1. (A) ^1H NMR spectrum of in vitro synthesized [^{13}C]Ala 82 T4 lysozyme. (B) ^{13}C-filtered ^1H spectrum of the same sample of T4L. The α-H and CH_3 resonances occur at 4.10 and 1.47 ppm, respectively. There are artifacts arising from incomplete suppression of sharp contaminant peaks at 2.23 and 2.10 ppm. (C) ^1H NMR spectrum of denatured in vitro synthesized [^{13}C]Ala 82 T4L at 60 °C. (D) ^{13}C-filtered ^1H spectrum of denatured T4L at 60 °C; the CH_3 resonance is at 1.34 ppm.

The general methodology described here should make possible a variety of detailed NMR studies of larger proteins, including determination of chemical shifts, pK_a values, and relaxation parameters for individual amino acids in both the native and denatured states. In addition, it will be possible to determine and compare denaturation curves from various regions of a protein, making it possible to detect domainwise folding. Further optimization of the in vitro synthesis system may yield larger quantities of protein sufficient for X-ray diffraction analysis.

(2) (a) McIntosh, L. P.; Dahlquist, F. W. *Q. Rev. Biophys.* **1990**, *23*, 1. (b) Otting, G.; Wüthrich, K. *Q. Rev. Biophys.* **1990**, *23*, 39.

(3) (a) Kainosho, M.; Tsuji, T. *Biochemistry* **1982**, *21*, 6273. (b) LeMaster, D. M.; Richards, F. M. *Biochemistry* **1985**, *24*, 7263. (c) Bachovchin, W. M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *25*, 7751. (d) Kleanthous, C.; Wemmer, D. E.; Schachman, H. *J. Biol. Chem.* **1988**, *263*, 13062.

(4) (a) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182-188. (b) Ellman, J. A.; Mendel, D.; Schultz, P. G. *Science* **1992**, *255*, 197-200. (c) Mendel, D.; et al. *Ibid.* **1992**, *256*, 1798. (d) Mendel, D.; Ellman, J. A.; Schultz, P. G. *J. Am. Chem. Soc.* **1991**, *113*, 2758. (e) Bain, J. D.; et al. *Biochemistry* **1991**, *30*, 5411. (f) Bain, J. D.; et al. *J. Am. Chem. Soc.* **1989**, *111*, 8013.

(5) (a) Tsugita, A. *The Enzymes*; Boyer, P. D., Ed.; Academic Press: New York, 1971; Vol. 3, pp 343-411. (b) Eriksson, A. E.; et al. *Science* **1992**, *255*, 178 and references cited therein.

(6) McIntosh, L. P.; Wand, A. J.; Lowry, D. F.; Redfield, A. G.; Dahlquist, F. W. *Biochemistry* **1990**, *29*, 6341.

(7) Matthews, B. W.; Nicholson, H.; Becktel, W. J. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6663.

(8) Plasmid pHSe54,97.TA encodes a cysteine-free T4L behind a twin *tac* promoter and was kindly provided by L. McIntosh of the Dahlquist group at the University of Oregon. The cysteine-free version of T4L is virtually identical to the wild-type in structure and activity but does not suffer oxidative degradation (Perry, L. J.; Wetzel, R. *Biochemistry* **1986**, *25*, 733; *Protein Eng.* **1987**, *1*, 101).

(9) Robertson, S. A.; Ellman, J. A.; Schultz, P. G. *J. Am. Chem. Soc.* **1991**, *113*, 2722.

(10) (a) Ellman, J. A.; Mendel, D.; Noren, C. J.; Anthony-Cahill, S.; Schultz, P. G. *Methods Enzymol.* **1991**, *202*, 301. (b) Noren, C. J.; et al. *Nucleic Acids Res.* **1990**, *18*, 83.

(11) Large scale in vitro protein synthesis reactions were carried out as described in ref 10a with the following modifications: (i) Ligation reactions of aminoacyl pdCpA to tRNA(-CA) required less T4 RNA ligase (25%) with reaction times extended to 40 min. After precipitation of the aminoacyl tRNA with 250 mM NaOAc, pH 4.5, and 3 volumes of ethanol, the tRNA pellet was dissolved in 1 mM KOAc, pH 4.5, and reprecipitated before a final 70% ethanol rinse. (ii) In order to achieve complete deprotection on a large scale, it was necessary to stir the aminoacyl tRNA solution and to extend the irradiation time to 20 min. (iii) Suppressor tRNA can be purified from in vitro transcription reactions by phenol extraction, ethanol precipitation, and dialysis. (iv) Each 5-mL in vitro synthesis reaction contained 1.5 mL of S-30, 1.62 mg of acyl tRNA, 1.25 mL of low molecular weight mixture, and 500 μg of plasmid.

(12) Data were acquired at 600 MHz on a Bruker AMX spectrometer. Presaturation was used to suppress residual HOD, 2-20K scans were acquired with a recycle delay of 2 s, and the temperature was 25 or 60 °C.

(13) Griffey, R. H.; Redfield, A. G. *A. Rev. Biophys.* **1987**, *19*, 51.

(14) Bundi, A.; Wüthrich, K. *Biopolymers* **1979**, *18*, 285.

Acknowledgment. We acknowledge support by the Director, Office of Energy Research, Office of Basic Energy Sciences, Division of Material Sciences, Division of Energy Biosciences (P.G.S.), and Office of Health and Environmental Research, Health Effects Research Division (D.E.W.), of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098, and a grant from the Lucille P. Markey Charitable Trust. J.A.E. was supported by an NSF postdoctoral fellowship (CHE-8907488), and D.M. by an American Cancer Society postdoctoral fellowship (PF-4014A).

Additions and Corrections

Electron Transfer vs Polar Mechanisms. Transition-State Structures and Properties for Reactions of a Cation Radical and a Nucleophile [*J. Am. Chem. Soc.* **1991**, *113*, 9890]. JEOUNG KI CHO and SASON SHAIK*

The $\Delta S^{\ddagger}$ value for the ET process in Table I should be −28.7 eu. The relevant discussion in the text was based upon this value and therefore there are no changes in the conclusions of the paper.

Does the Mechanism of Symmetric Methyl Transfer to Water from Water Differ from That for Transfer to Water from Other Leaving Groups? [*J. Am. Chem. Soc.* **1989**, *111*, 792]. JOSEPH L. KURZ* and MICHAEL W. DANIELS

Recent measurements show that the values reported in this communication for the isotope effect and rate constants are not correct. We (H. S. Yathirajan and Joseph L. Kurz) plan to submit an article providing details after we have completed our investigation of these rates and isotope effects.

Computer Software Reviews

PEAKFIT. Version 3.01. Jandel Scientific: 2591 Kerner Blvd., San Rafael, CA 94901. List price $595.00.

The proliferation of and utility of varied forms of spectroscopy coupled with the now general availability of desktop computers makes sophisticated data analysis a real part of many projects. In many cases, deconvolution of overlapping signals through the use of nonlinear equations and curve fitting is an application of computers which potentially provides useful qualitative and quantitative information. Peakfit by Jandel Scientific is an IBM PC-based software program designed as an aid to researchers with an interest in using nonlinear equations and curve-fitting procedures in data analysis.

Peakfit constructs and evaluates components of a peak outside the boundaries of the instrumental resolution. For example, we used the review copy to resolve overlapping peaks from NMR spectra, IR spectra, and XPS (X-ray photoelectron) spectra. The main requirement is simply that the spectral data be available in a digital format, most commonly in an ASCII format for importation into the program. The program features rapid, high quality, nonlinear curve fit procedures. The program offers an excellent graphical/numerical review for a critical evaluation of the fit. A variety of different curve-fit procedures are available either as standard analytic functions or as user modified functions.

Peakfit is written for IBM compatible computers. The program is operated in a windows-like environment using pull-down menus with either a mouse or keyboard commands. It is compatible with windows but does not run as a windows application. A hard disk is required and the program uses about 1.9 Mb of hard disk. The program supports VGA, EGA, MCGA, Hercules (Monochrome), ATT, IBM 8514, or IBM PC-3270 graphics. A math-coprocessor is highly recommended. Without a coprocessor, the program is too slow to be useful. We tested the program on various brands of 80286, 80386, and 80486 computers and experienced no major compatibility problems. This program runs as a DOS application under Windows.

Using the program revolves around four steps. First, the X–Y data table is either entered or imported from another source. This is followed by the users choice of functions in such a way that the sum of the functions approximates the data. The program then performs an iterative curve fit to obtain the best fit. The user then evaluates the fit graphically and numerically. Finally the output is sent to a printer or as a file according to the users choice.

Data input/output is versatile. Besides ASCII import, you can automatically read in files from several sorts of spreadsheets. Manual data entry is also possible. Calculation options include a variety of arithmetic functions which can be used to modify these values in a manner like that used with common spreadsheets. Portions of the data can be selected and ignored to facilitate analysis. Data imported as an ASCII file can include a maximum of 32 000 points with 5 MB of Virtual Memory (can be any combination of expended, extended, RAM, or hard disk memory). With 640K of RAM and 256K of Virtual Memory (can be simply hard disk space), a maximum of 3000 points can be included in the data set.

The next step is to approximate the data to a function. Several application families such as spectroscopic, chemical-pharmacological, waveform, and statistical applications ease the selection of the proper function. Multiple functions can be used to fit a given set of data points. A known parameter can be fixed during the process of curve fitting. User-defined functions further broaden the utility of the standard built-in functions and provide for added flexibility. Choice of a particular function is followed by curve fitting. Curve fitting occurs with a visual update after each iteration as the program searches for a best fit. Curve fitting can be aborted if desired to change one or more parameters. At the users discretion, the visual updates can be suppressed to speed up the calculations though speed was not a problem on the machines we used.

The curve fitting is followed by a graphical review displaying the original data, the constituent functions, and the sum curve. A graphical display of the residuals gives a better idea about the closeness of the fit. Apart from the graphical review, a numerical review enables one to analyze the fit for each individual point. In cases of unsatisfactory fit, alternative functions can be chosen to improve the fit or parameters can be adjusted and curve fitting used again.

The curve-fitted data can be printed to most dot-matrix printers, HP LaserJet printers, HPGL plotters, etc. The fitted data can be exported to other programs such as Lotus-123 and SigmaPlot 4.0 or exported to HPGL or metafile format. Postscript output is also possible.

A manual is provided which is comprehensive and takes the user through a total tour of all the functions and capacities of this powerful software. A chemist with some computer experience should be able to sit at the keyboard with the manual and would feel at home with the software in a few hours. The powerful program stimulates further research and challenges the user to import more complicated problems to