QUANTITATIVE ANALYSIS OF NUCLEIC ACIDS, PROTEINS, AND VIRUSES BY RAMAN BAND DECONVOLUTION

GEORGE J. THOMAS, JR.
Department of Chemistry, Southeastern Massachusetts University, North Dartmouth, Massachusetts 02747

DAVID A. AGARD
Department of Biochemistry and Biophysics, University of California School of Medicine, San Francisco, California 94143

ABSTRACT A constrained, iterative Fourier deconvolution method is employed to enhance the resolution of Raman spectra of biological molecules for quantitative assessment of macromolecular secondary structures and hydrogen isotope exchange kinetics. In an application to the Pf1 filamentous bacterial virus, it is shown that the Raman amide I band contains no component other than that due to α-helix, indicating the virtual 100% helicity of coat proteins in the native virion. Comparative analysis of the amide I band of six filamentous phages (fd, If1, IKe, Pf1, Xf, and Pf3), all at the same experimental conditions, indicates that the subunit helix-percentage ranges from a high of 100% in Pf1 to a low of 71% in Xf. Deconvolution of amide I of Pf3 at elevated temperatures, for which an α-to-β transition was previously reported (Thomas, G. J., Jr., and L. A. Day, 1981, *Proc. Natl. Acad. Sci. USA.*, 78:2962–2966), allows quantitative evaluation of the contributions of both α-helix and β-strand conformations to the structure of the thermally perturbed viral coat protein. Weak Raman lines of viral DNA bases and coat protein side chains, which are poorly resolved instrumentally, are also distinguished for all viruses by the deconvolution procedure. Application to the carbon-8 hydrogen isotope exchange reaction of a purine constituent of transfer RNA permits accurate determination of the exchange rate constant, which is in agreement with calculations based upon curve-fitting methods.

INTRODUCTION

Curve-fitting techniques, factor analysis, and difference spectroscopy are among the most commonly used methods for analysis of multicomponent bands in vibrational spectra. (For a review, see Laane, 1983.) More recently, deconvolution techniques have provided a convenient and powerful alternative to improving experimental resolution of intrinsically overlapped bands in infrared and Raman spectra (King and Bower, 1980; Kauppinen et al., 1981). Applications for NMR spectroscopy (Ferrige and Lindon, 1978) and for analysis of overlapping bands in gel electrophoretograms (Agard et al., 1981) have also been described.

In this paper, we illustrate the use of a constrained, iterative deconvolution procedure for the quantitative analysis of certain nucleic acid and protein Raman bands in spectra of viruses and related model compounds. Although the Raman spectrum of a virus or nucleoprotein typically contains a substantial number of well-resolved Raman lines (Thomas, 1976), it is sometimes desirable to improve the information content of the spectrum by resolution of multicomponent or overlapped bands. Such improved resolution may be difficult or impossible to achieve instrumentally because of the inherent band shapes and half-widths ordinarily encountered in condensed phase vibrational spectra. However, it is possible to increase the degree to which component lines can be distinguished using the Fourier deconvolution approach (Agard et al., 1981; Kauppinen et al., 1981).

The applications described here include: (*a*) monitoring the isotopic hydrogen exchange at the 8C ring position of a purine nucleotide (a constituent of transfer RNA) and calculation of the exchange rate constant by deconvolution of the partially overlapped Raman bands of 8C-H and 8C-D isomers (Ferreira and Thomas, 1981); (*b*) deconvolution of the intense amide I Raman band from coat protein of the filamentous bacterial virus, Pf1, to deduce the positions of overlapped weaker bands of DNA and amino acid side chains; (*c*) deconvolution of the broad

This paper is part XXVII in the series Raman Spectral Studies of Nucleic Acids. Paper XXVI in this series is Benevides, J. M., A. H. -J. Wang, G. A. van der Marel, J. H. van Boom, A. Rich, and G. J. Thomas, Jr., 1984, *Nucleic Acids Research*, 12:5913–5925.

Address all correspondence to Dr. G. J. Thomas, Jr.

amide I envelope in the Raman spectrum of the filamentous bacterial virus, Pf3, following a transition of viral coat protein from α-helix to β-strand conformation (Thomas and Day, 1981); and (*d*) quantitative estimation of subunit secondary structure in six filamentous viruses, representing both class I (fd, If1, and IKe) and class II (Pf1, Xf, and Pf3) symmetries (Thomas et al., 1983). The results obtained here permit realistic estimates of the secondary structures of the subunits in native and perturbed states of the filamentous phages.

METHODS

Deconvolution

The present method uses the modified Jansson-VanCittert iterative deconvolution scheme (Jansson et al., 1970) developed by Agard and Stroud (Agard et al., 1981). This method has been used for the quantitative deconvolution of profiles derived from one- and two-dimensional gel electrophoretograms (Agard et al., 1981) and in two- (Agard and Sedat, 1983) and three-dimensional (Agard and Stroud, 1982) image enhancement problems.

Deconvolution and curve-fitting approaches offer different and complementary strengths for resolution enhancement (Laane, 1983). The major advantage of deconvolution for present applications is that the number of components in the complex band shape need not be known beforehand; nor is it necessary to provide starting guesses for peak amplitudes or locations. Although we have not quantitated in detail the differences between deconvolution and curve fitting in computation time or accuracy, preliminary studies (Williams, 1983; as well as present results) suggest that they are comparable.

To obtain maximum enhancement with the deconvolution method, it is necessary that within the region of interest the band broadening be described by a single, accurately known, lineshape function. Under these conditions a five-to-tenfold resolution enhancement can be realized (Agard et al., 1981; Kauppinen et al., 1981). However, in all applications so far encountered, such maximal resolution is not required. By choosing a symmetric lineshape function that is slightly narrower than any of the observed peaks, it is possible to uniformly sharpen all of the peaks in the spectral range of interest, independent of the true bandwidths. Any inherent asymmetry, as well as residual broadening, will remain in the deconvolved peaks. Throughout this operation the correct area under each peak is maintained, which allows for an accurate quantitative analysis of the results. As shown previously (Agard et al., 1981) and discussed below, the constrained deconvolution procedure appears to be generally superior to nonconstrained methods in accuracy, degree of enhancement, and noise rejection.

Although this approach is described in detail elsewhere (Agard et al., 1981), the salient features will be summarized here. The starting assumption is that there exists a true spectrum that, when broadened by a single lineshape function, will produce the observed data. The goal of the deconvolution procedure is to reverse this process. Thereby, the true spectrum is extracted from the observed data and a knowledge of the lineshape function. Unfortunately, unless additional information is provided, the inversion process is not unique, i.e., there may be more than one solution. This problem is further compounded by the presence of experimental noise in the observed data. We have found that the addition of a simple physical constraint, namely, that the true spectrum must be everywhere non-negative, can effectively solve the problems of uniqueness and noise sensitivity. Introducing this constraint requires that the deconvolved spectrum be obtained in an iterative manner. At each iteration, the current estimate to the true spectrum is broadened by convolution with the lineshape function and compared with the observed data. The differences are then used to generate a new, more accurate estimate of the true spectrum. The iterations are stopped when the total difference falls below some initially specified value. The total difference, E, between the observed spectrum $O(x)$ and the blurred guess $O'(x)$ is defined as

$$E = \Sigma[|O(x) - O'(x)|]/[|O(x)|]. \quad (1)$$

For deconvolution of the Raman bands we have employed a Lorentzian-Gaussian (*LG*) lineshape function, defined by

$$LG = N[\Gamma/(x - u)^2 + (\Gamma/2)^2](\exp\{-1/2[(x - u)/\sigma]^2\}), \quad (2)$$

where Γ and 2σ are, respectively, the Lorentzian and Gaussian band half-widths (full width at half maximum), $x - u$ is the band center, and N is a normalization constant. The actual half-width of the product function was determined empirically. The *LG* product function is superior to either the Gaussian function alone, which poorly represents the Raman band contour near the band center, or the Lorentzian function alone, which contains undesirable asymptotic behavior. Other acceptable lineshape functions for the present applications are linear combinations of L and G, which are similar to the Voigt function (Laane, 1983). Since the intrinsic half-widths of the vibrational Raman lines of proteins and nucleic acids are typically 20 cm^{-1} or higher, i.e., much greater than the instrument slit width employed here (8 cm^{-1}), the slit function does not significantly influence the deconvolution results.

Spectral noise, which can be seriously detrimental to the generation of an accurately deconvoluted profile, was minimized by either Savitzky-Golay smoothing in the frequency domain (Savitzky and Golay, 1964) or truncation of the interferogram in the Fourier domain (Kauppinen et al., 1982).

The lineshape half-width for deconvolution is estimated arbitrarily and selection of an optimum value is somewhat subjective. Objective criteria have been proposed for the simpler case where the non-negativity constraint is not used (Kauppinen et al., 1981). In that case the correct bandwidth was chosen to be intermediate between severe underestimation and overestimation following repeated trials. We too seek to choose such intermediate values. If the lineshape is too narrow, insufficient enhancement takes place; when it is too broad, the algorithm can never completely converge. We have imposed the minimum requirement of 95% replication of the original intensity profile in the reconvoluted spectrum (i.e., $E \leq 0.05$ in Eq. 1) as the criterion for not overestimating the deconvolution bandwidth.

Experimental Raman Spectroscopy

Previous publications from the laboratory of the Southeastern Massachusetts University (North Dartmouth, MA) group give full details of sample handling procedures and Raman instrumentation employed to study isotopic hydrogen exchange of inosine nucleotides (Ferreira and Thomas, 1981; Thomas and Ferreira, 1982) and structures of the filamentous bacterial viruses Pf3 and Pf1 (Thomas and Day, 1981; Thomas et al., 1983). Filamentous phage were generously provided by Dr. Loren A. Day of the Public Health Research Institute of the City of New York.

Computations

The deconvolution programs were developed in the laboratory of the University of California at San Francisco (San Francisco, CA) group in collaboration with Professor Robert M. Stroud (Agard et al., 1981). All computations were carried out on a VAX 11/780 or DEC-20 (Digital Equipment Corp., Marlboro, MA). FORTRAN programs are available on request.

APPLICATIONS AND DISCUSSION

Quantitative Analysis of Deuterium Exchange of the 8C-H Group in Inosine-5'-Monophosphate (5'rIMP)

Each of the first nine panels of Fig. 1 compares the observed Raman spectrum of 5'rIMP (recorded in the

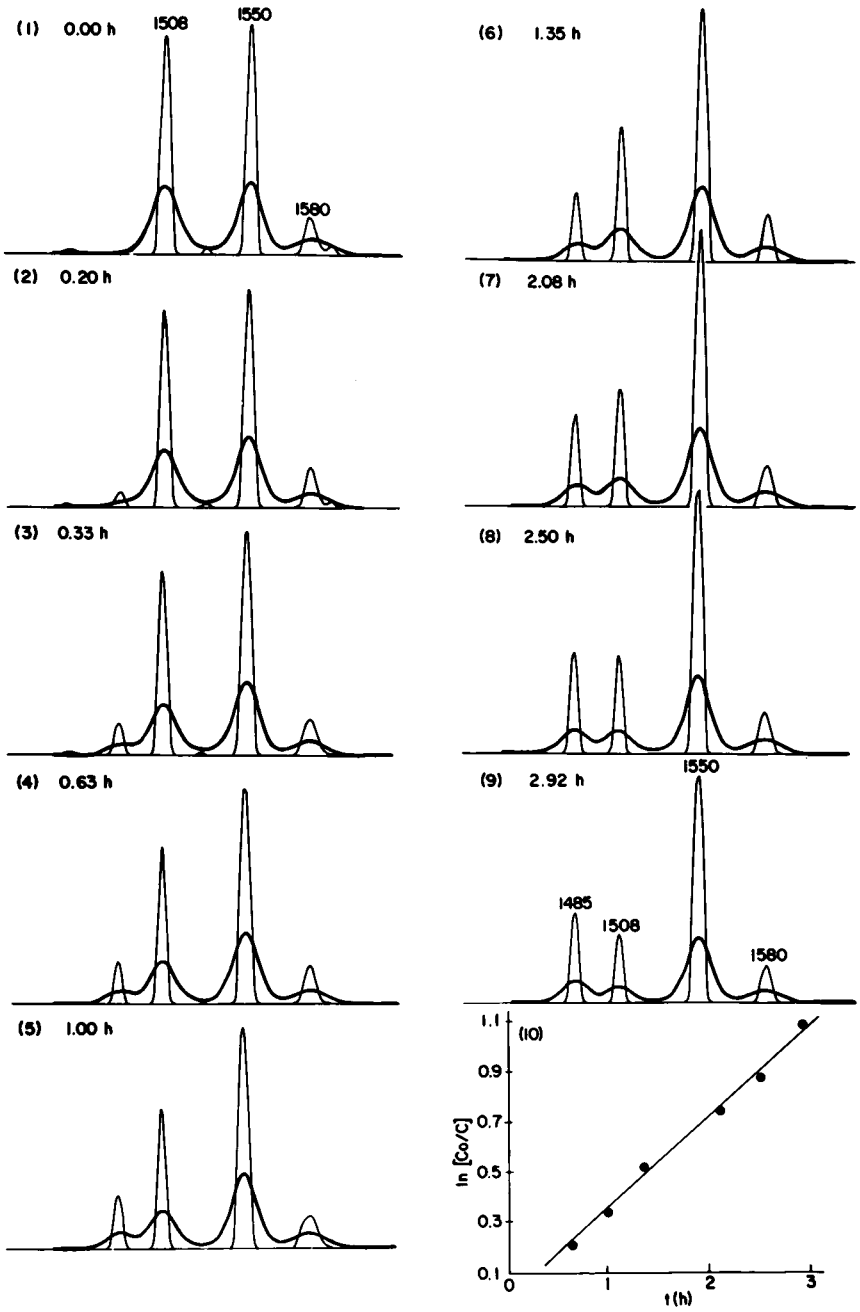

FIGURE 1 Panels labeled *1–9* show the observed (heavy line) and deconvoluted spectra (light line) of 5'rIMP at successive stages of deuterium exchange of the inosinic 8C-H group. The deconvolution bandwidth was 12 cm^{-1}. Each panel lists the time of exchange (t) at 80°C. *10* shows the plot of $\ln[Co/C]$ vs. t, the slope of which is the exchange rate constant, $k = 0.365\ h^{-1}$.

region 1,450–1,600 cm^{-1}) with its deconvoluted spectrum at a given stage in the 8C-H exchange process. In this sequence, we have employed a deconvolution bandwidth of 12 cm^{-1}. The most intense band in the spectrum, centered at 1,550 cm^{-1}, and the much weaker band at 1,580 cm^{-1}, are both due to vibrational modes of the hypoxanthine ring and are nearly unaffected by deuterium exchange. The former serves as a convenient intensity standard. The band centered at 1,508 cm^{-1} is due to a hypoxanthine ring vibration, which involves participation of the 8C-H in-plane bending motion. It is, therefore, sensitive to deuterium exchange and is gradually replaced by the corresponding ring mode (1,485 cm^{-1}) of the 8C-D isomer as exchange progresses (Ferreira and Thomas, 1981).

The deconvolution of Fig. 1 achieves excellent resolution of the 1,485 cm^{-1} shoulder from the dominant 1,508 cm^{-1} band even during the early stages of hydrogen isotope exchange. The semilogarithmic plot of intensity ratio vs. time of exchange, which yields the exchange rate constant (Ferreira and Thomas, 1981), is strictly linear as shown in

the lower right panel of Fig. 1. The rate constant computed from these data is 0.365 h^{-1}, which is within 2% of the value computed by the curve-fitting methods described previously (Thomas and Livramento, 1975).

Deconvolution of the Amide I Band of Filamentous Virus Pf1

Fig. 2 compares the observed and deconvoluted Raman bands of Pf1 virus in the region of the intense amide I protein mode, which dominates the 1,500–1,800-cm^{-1} interval of the spectrum. The experimental curve, taken from Fig. 1 of Thomas et al. (1983), represents the average of six scans from which solvent contributions have been subtracted.

Deconvolution clearly identifies the very intense 1,651 cm^{-1} amide I band of the protein α-helix conformation and reveals as well the absence of amide I bands between 1,655 and 1,675 cm^{-1} that could be assigned to either β-strand or irregular conformations. The deconvolution confirms the uniformity of α-helical secondary structure of the coat protein in native Pf1 virus. The weak band resolved at 1,678 cm^{-1} is assigned to the carbonyl group stretching modes of the bases of viral DNA and glutamine side chains of viral coat protein. The remaining weak bands resolved at 1,577 cm^{-1} and at 1,598 and 1,618 cm^{-1}, are assigned, respectively, to aromatic ring modes of the DNA purines and to tyrosines of the viral protein (Prescott et al., 1984; Thomas et al., 1983).

Deconvolution of the Complex Amide-I Band of the Filamentous Virus Pf3

The filamentous bacterial virus Pf3 undergoes a gradual structure transition in which the coat protein conformation is largely converted from α-helix to β-strand as the temperature is increased from 20° to 60°C (Thomas and Day, 1981). Quantitative estimates of the percentages of α and β structures that coexist at various stages of the transition

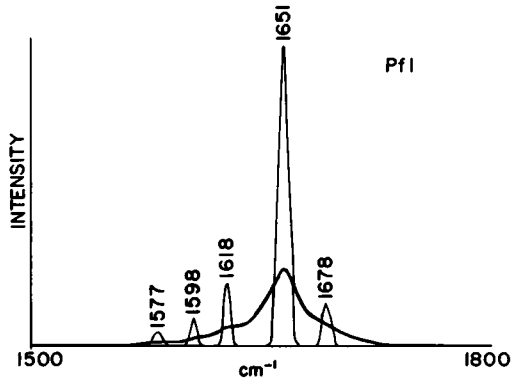

FIGURE 2 Comparison of observed (heavy line) and deconvoluted spectra (light line) of native Pf1 filamentous virus in the region of the intense amide I Raman band. The deconvolution bandwidth was 22 cm^{-1}. The experimental data are from Thomas et al., 1983.

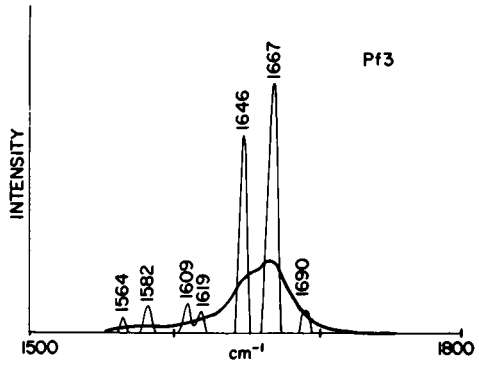

FIGURE 3 Comparison of observed (heavy line) and deconvoluted spectra (light line) of Pf3 filamentous virus in the region of amide I Raman scattering. The deconvolution bandwidth was 22 cm^{-1}. The native virus structure was perturbed by heating the sample at 60°C for 2 h to promote the α- to β-structure transition. (See Thomas and Day, 1981.)

cannot be made from the unrefined Raman data because of extensive overlap of the respective amide I bands. However, it is possible to obtain quantitative estimates by Fourier deconvolution of the complex band envelope as shown in Fig. 3. Here amide I components are clearly resolved at 1,646 cm^{-1} (α-helix structure) and 1,667 cm^{-1} (β-strand structure). We note that irregular protein conformations can be excluded from consideration because of the characteristic amide III Raman scattering of Pf3 (Thomas et al., 1983) not shown in Fig. 3. The weaker bands resolved in the deconvoluted spectrum are assigned by analogy with model compounds as follows: 1,564, tryptophan; 1,582, DNA purines and phenylalanine; 1,609, phenylalanine; 1,619, tryptophan; 1,690 cm^{-1}, glutamine (Thomas et al., 1983). The integrated intensities of the two amide I bands indicate a 60: 40% ratio of β to α structure in Pf3 at the conditions of Fig. 3.

Comparison of Subunit Secondary Structures by Deconvolution of Raman Amide I Bands of fd, If1, IKe, Pf1, Xf, and Pf3 Viruses

Application of the above methods to the amide I band of each of the native filamentous viruses for which Raman data are available gives the results of Table I. Deconvolution indicates that the percentage of protein helix varies considerably among the class II viruses (Pf1, Xf, and Pf3), but is relatively invariant among the class I viruses (fd, If1 and Ike). These findings are qualitatively consistent with the secondary structures expected from protein prediction algorithms (Finer-Moore et al., 1984) and are also consistent with semiquantitative estimates of helix content from amide I half-width measurements (Thomas et al., 1983). Deconvolution provides a reliable basis for inference that the order of decreasing helicity of the viral subunits is as follows: Pf1 > IKe > fd ~ If1 > Pf3 > Xf. Further, we find that the 29% nonhelical regions of Xf subunits are struc-

TABLE I
SECONDARY STRUCTURES OF FILAMENTOUS VIRUS SUBUNITS BY DECONVOLUTION OF THE RAMAN AMIDE I BAND

Virus	Observed amide I*	Deconvoluted amide I‡		Percent helix§
fd	1648 ± 1	1648 (0.92)	1667 (0.08)	92 ± 2
If1	1650 ± 1	1651 (0.90)	1667 (0.10)	90 ± 2
IKe	1650 ± 1	1652 (0.93)	1670 (0.07)	93 ± 2
Pf1	1650 ± 1	1651 (1.00)	— —	100 ± 2
Xf	1652 ± 1	1646 (0.71)	1659 (0.29)	71 ± 2
Pf3	1648 ± 1	1645 (0.82)	1666 (0.18)	82 ± 2

*Frequencies of the observed amide I band center are measured to an accuracy of ±1 cm^{-1}. From data of Thomas et al., 1983.
‡Deconvoluted peaks are determined to a precision of ±0.5 cm^{-1} or less. Integrated peak areas are believed accurate to ±2%.
§The uncertainties reflect the estimated limits of error in peak area measurements. Each deconvolution employed a Gaussian-Lorentzian product function of 22 cm^{-1} bandwidth and a convergence limit of 5%, i.e., $E = 0.05$. (See text.) The same results were obtained using a 1:1 Gaussian-Lorentzian sum function of the same half-width.

turally different from the nonhelical domains (18%) of Pf3 subunits. The latter appear to be β-stranded by virtue of the positions of amide III (1,240 cm^{-1}; Thomas et al., 1983) and amide I (1,666 cm^{-1}; Fig.3, above) components. For Xf, the amide I (1659 cm^{-1}) and amide III frequencies (1,257 cm^{-1}) together suggest a departure from conventional α, β, and disordered protein structures upon which the Raman assignments are based (Williams, 1983). Further study of model compounds will be required to better characterize this Xf protein conformation.

CONCLUSIONS AND SUMMARY

The significant conclusions derived from deconvolution of Raman spectra of filamentous viruses are: (*a*) 100% α-helicity of the Pf1 viral subunit and lesser helix content of other class II and class I virus subunits at physiological conditions, (*b*) predominant conversion of the Pf3 subunit from α-helix to β-strand at elevated temperatures, (*c*) indication that the substantial nonhelical regions of Pf3 are β-stranded, while those of Xf are not, and (*d*) indication of small amounts of β-structure in the class I virus subunits.

We have illustrated the use of a constrained, iterative Fourier deconvolution procedure to separate overlapping Raman bands in spectra of native and structurally perturbed viruses for analysis of protein and nucleic acid constituents. The deconvolution method permits several-fold improvement of spectral resolution and allows quantitative treatment of the Raman data. Weak shoulders that cannot be instrumentally resolved from nearby strong Raman lines in the unrefined experimental data are clearly distinguished in the deconvoluted spectra.

The methods employed here are applicable also to the analysis of conformational structure in individual proteins and nucleic acids. In particular, deconvolution provides a convenient alternative to curve fitting of the amide I band for assessment of protein secondary structure (Williams, 1983), and to difference Raman spectroscopy for the quantitation of nucleic acid secondary structures (Benevides and Thomas, 1983).

To facilitate the use of powerful deconvolution techniques for biological applications of infrared and Raman spectroscopy, workable programs are required for smaller (spectrometer-dedicated) computers. We are currently developing BASIC and FORTRAN versions of the deconvolution programs for implementation on laboratory microcomputers.

We thank L. A. Day, Public Health Research Institute of the City of New York (New York, NY), for filamentous viruses; B. Prescott and J. M. Benevides, Southeastern Massachusetts University (North Dartmouth, MA), for assistance in recording spectra; and J. Finer-Moore and R. M. Stroud, University of California San Francisco (San Francisco, CA), for assistance in use of computer facilities.

This work was supported by grants AI11855 and AI18758 (to G. Thomas), and GM31627 (to D. Agard) from the National Institutes of Health. D. Agard is a Searle Scholar.

Received for publication 30 January 1984 and in final form 15 June 1984.

REFERENCES

Agard, D. A., and J. W. Sedat. 1983. Three-dimensional architecture of a polytene nucleus. *Nature (Lond.)*. 302:676–681.

Agard, D. A., and R. M. Stroud. 1982. Linking regions between helices in bacteriorhodopsin revealed. *Biophys. J.* 37:589–602.

Agard, D. A., R. A. Steinberg, and R. M. Stroud. 1981. Quantitative analysis of electrophoretograms: a mathematical approach to superresolution. *Anal. Biochem.* 111:257–268.

Benevides, J. M., and G. J. Thomas, Jr. 1983. Characterization of DNA structures by Raman spectroscopy: high-salt and low-salt forms of double helical poly(dG-dC) in H_2O and D_2O solutions and application to B, Z and A-DNA. *Nucleic Acids Research.* 11:5747–5761.

Ferreira, S. A., and G. J. Thomas, Jr. 1981. Kinetics of hydrogen-deuterium exchange and kinetic isotope effects in inosine-5′-monophosphate and inosine-3′5′-monophosphate determined by laser Raman spectroscopy. *J. Raman Spectrosc.* 11:508–514.

Ferrige, A. G., and J. C. Lindon. 1978. Resolution enhancement in FT-NMR through the use of a double exponential function. *J. Magn. Res.* 31:337–340.

Finer-Moore, J., R. M. Stroud, B. Prescott, and G. J. Thomas, Jr. 1984. Subunit secondary structure in filamentous viruses: predictions and observations. *J. Biomolec. Struct. Dynam.* 2:93–100.

Jansson, P. A., R. H. Hunt, and E. K. Plyler, E. K. 1970. Resolution enhancement of spectra. *J. Opt. Soc. Am.* 60:596–599.

Kauppinen, J. K., D. J. Moffatt, H. H. Mantsch, and D. G. Cameron. 1981. Fourier self-deconvolution: a method for resolving intrinsically overlapped bands. *Appl. Spectrosc.* 35:271–276.

Kauppinen, J. K., D. J. Moffatt, H. H. Mantsch, and D. G. Cameron. 1982. Smoothing of spectral data in the Fourier domain. *Appl. Opt.* 21:1866–1872.

King, J., and D. I. Bower. 1980. Use of Fourier transforms in the resolution of the C-Cl stretching region of the Raman spectrum of polyvinyl chloride. *Proc. Int. Conf. Raman Spectrosc., 7th, Ottawa, Canada.* 242–243.

Laane, J. 1983. Raman difference spectroscopy. *In* Vibrational Spectra and Structure. J. R. Durig, editor. Elsevier/North Holland, Amsterdam. 12:405–467.

Prescott, B., W. Steinmetz, and G. J. Thomas, Jr. 1984. Characterization of DNA structures by laser Raman spectroscopy. *Biopolymers.* 23:235–256.

Savitzky, A., and M. J. E. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36:1627–1639.

Thomas, G. J., Jr., and L. A. Day. 1981. Conformational transitions in Pf3 and their implications for the structure and assembly of filamentous bacterial viruses. *Proc. Natl. Acad. Sci. USA.* 78:2962–2966.

Thomas, G. J., Jr., and S. A. Ferreira. 1982. Isotopic hydrogen exchange in the dinucleoside monophosphate IpI. J. *Raman Spectrosc.* 12:122–124.

Thomas, G. J., Jr., and J. Livramento. 1975. Kinetics of hydrogen-deuterium exchange in adenosine-5'-monophosphate, adenosine-3':5'-monophosphate and poly(riboadenylic) acid determined by laser Raman spectroscopy. *Biochemistry.* 14:5210–5218.

Thomas, G. J., Jr., B. Prescott, and L. A. Day. 1983. Structure similarity, difference and variability in the filamentous viruses fd, If1, IKe, Pf1, Xf and Pf3. Investigation by laser Raman spectroscopy. *J. Mol. Biol.* 165:321–356.

Thomas, G. J., Jr. 1976. Raman Spectroscopy and virus research. *Appl. Spectrosc.* 30:483–494.

Williams, R. W. 1983. Estimation of protein secondary structure from the laser Raman amide I spectrum. *J. Mol. Biol.* 166:581–603.