compound II the Fe-containing protein ν_{FeO} is at $\sim$788 cm^{-1}, consistent with a Fe=O bond trans to a strongly H-bonded imidazole ligand.[15b] At pH values below 8.5 the frequency shifts down $\sim$13 cm^{-1}. This effect has been attributed to H-bonding from a protonatable distal residue, probably histidine.[15c] Thus in compound II the distal proton resides on an adjacent protein residue when the M-O bond contains Fe^{IV}, but on the oxo ligand (perhaps with a reverse H-bond interaction to the same distal ligand) when it contains Mn^{IV}. The 623-cm^{-1} Mn-HRP II RR peak is unaltered between pH 7 and 10, indicating the distal proton to be much more tightly held in the Mn protein.

Finally, it is worth remarking on the implications of the present results with respect to the reactivity patterns of manganyl and ferryl porphyrins. Oxomanganese(V) porphyrins and oxoiron(IV) porphyrin cation radicals are powerful oxidants capable of alkane hydroxylation and olefin epoxidation even at low temperatures.[8a,17] Other things being equal, manganese(III) should be easier to oxidize than iron(III). Indeed, manganyl(IV) porphyrins are relatively stable; in this study they have been characterized at room temperature. Ferryl(IV) porphyrins have limited stability, however, and have been characterized only at low temperature.[2] To the extent that the π-bond order is lower for Mn^{IV}=O than for Fe^{IV}=O, the oxygen atom of the former should both be less electrophilic and have less unpaired electron density than the latter. Consistent with this notion we find that electrochemically generated ferryl(IV) porphyrins *do* epoxidize olefins at room temperature but with an uncharacteristic *loss* of olefin configuration,[34] just as manganese(IV) porphyrins do.[8b]

This weakness of the Mn^{IV}=O bond may help account for why nature has chosen manganese for the O_2 generating apparatus of green plants. The water oxidation center associated with the chloroplast photosystem contains a cluster of manganese ions whose oxidation level is intimately associated with O_2 production.[35] It is believed that the Mn ions shuttle between the Mn^{III} and Mn^{IV} oxidation levels.[36] While the Mn ions are not imbedded in porphyrins, the present inference about the weakness of the Mn^{IV}=O bond is not limited to porphyrins. Rather it is associated with the electronic properties of the d^3 ions; oxidation of Mn^{III}-bound water, accompanied by proton transfer, should generally lead to weak Mn^{IV}=O bond formation. Two Mn^{IV}=O units in the proper orientation could readily generate Mn^{III}-O-O-Mn^{III}, and then O_2 via electron transfer, perhaps from other Mn ions in the cluster.[37]

Acknowledgment. This work was supported by U.S. Department of Energy Grant DE-AC02-81ER10861 (to T.G.S.) and National Science Foundation Grant CHE 8706310 (to J.T.G.). We thank Prof. Harold Goff for helpful discussions.

Registry No. **1**, 85939-49-7; **2**, 104025-51-6; **3**, 108535-08-6; **4**, 104025-52-7; **6**, 114300-80-0; $ClMn^{III}OEP$, 28265-17-0; $ClMn^{III}TPP$, 32195-55-4; $ClMn^{III}TMPyP$, 97057-20-0; $(AcO)Mn^{III}TMPyP$, 114324-35-5; $(Mn^{IV}O)TPP$, 76077-77-5; $[(HO)_2Mn^{III}TMPyP]^-$, 89043-43-6.

(33) (a) Ferrante, R. F.; Wilkerson, J. L.; Graham, W. R. M.; Weltner, W. *J. Chem. Phys.* **1977**, *67*, 5904. (b) Richens, D. T.; Sawyer, D. T. *J. Am. Chem. Soc.* **1979**, *101*, 3681-3683. (c) Morishima, I.; Ogawa, S. *Biochemistry* **1978**, *17*, 4384-4388.

(34) Groves, J. T.; Stern, M. K., presented at the Third Chemical Congress of North America, Toronto, June 1988.

(35) Dismukes, G. C.; Siderer, Y. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 274-278.

(36) (a) Dismukes, G. C.; Fems, K.; Watnick, P. *Photobiochem. Photobiophys.* **1982**, *3*, 243-256. (b) Dismukes, G. C. *Photochem. Photobiol.* **1986**, *43*, 99-115.

(37) (a) Renger, G. In *Photosynthetic Oxygen Evolution*; Metzner, H., Ed.; Academic: London, 1978; pp 239-248. (b) Radner, R.; Olligner, O. *FEBS Lett.* **1986**, *195*, 285-289.

Differences between Human and Porcine Insulin Investigated by Linear Prediction Carbon-13 NMR

Jens J. Led,*,† Henrik Gesmar,† Kim R. Hejnaes,‡ and Finn B. Hansen‡

Contribution from the Department of Chemical Physics, University of Copenhagen, The H. C. Ørsted Institute, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark, and Nordisk Gentofte, Niels Steensensvej 1, DK-2820 Gentofte, Denmark. *Received September 21, 1987*

Abstract: The linear prediction method was used to analyze the ^{13}C NMR free induction decay signals of a series of insulins and insulin derivatives. The higher resolution provided by this method, as compared with the conventional Fourier transformation approach, revealed differences between the ^{13}C NMR spectra of human and porcine insulins. This may indicate differences between the structures of these closely related proteins beyond the difference between their primary structures. Such differences have not been observed previously either in solution or in the crystal phase.

The function of biologically active proteins depends on even small differences and changes in their solution structures. This emphasizes the importance of methods by which such structural variations can be detected. In general, NMR spectroscopy is suitable for that purpose. Thus, by a variety of techniques,[1,2] information about structural and kinetic details of proteins has been obtained from the ^1H NMR signals of characteristic groups in the molecules. Furthermore, with the introduction of two-dimensional (2D) experiments, NMR spectroscopy has become the first technique by which the overall three-dimensional structure of small proteins in solution can be determined.[3-5] So far, these techniques have almost exclusively been based upon the Fourier transform (FT) of the time-domain NMR signal—or the free induction decay (FID). Recently it was demonstrated[6] that linear prediction (LP) analyses of the ^1H and ^{13}C FID's of peptides and small proteins are feasible, yielding quantitative spectral infor-

† The H. C. Ørsted Institute.
‡ Nordisk Gentofte.

(1) Jardetsky, O.; Roberts, G. C. K. *NMR in Molecular Biology*; Academic: New York, 1981.

(2) Fox, R. O.; Evans, P. A.; Dobson, C. M. *Nature (London)* **1986**, *320*, 192-194.

(3) Wagner, G.; Wüthrich, K. *J. Mol. Biol.* **1982**, *155*, 347-366.

(4) Zuiderweg, E. R. P.; Scheek, R. M.; Boelens, R.; van Gunsteren, W. F.; Kaptein, R. *Biochimie* **1985**, *67*, 707-715.

(5) Kline, A. D.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1986**, *189*, 377-382.

(6) Gesmar, H.; Led, J. J. *J. Magn. Reson.* **1988**, *76*, 183-192.

0002-7863/88/1510-4165$01.50/0

mation that is significantly better than what can be obtained from the FT, in terms of both resolution and sensitivity, although at the cost of relatively heavy computations. In the study reported here, this approach has been extended and used to investigate differences between porcine and semisynthetic human insulin, as well as to demonstrate the accordance between the structures of semisynthetic and pancreatic human insulin, on the basis of their LP ^{13}C NMR spectra.

Theoretical Considerations

In the LP analysis it is assumed that the FID signal consists of p exponentially damped sinusoids, i.e. the nth data point of the FID is given by eq 1, where ω_j, I_j, R_{2j}, and φ_j are the frequency,

$$y_n = \sum_{j=1}^{p} I_j e^{-R_{2j}\tau n} \cos(\omega_j \tau n + \varphi_j) \quad (1)$$

the intensity, the damping factor, and the phase, respectively, of the jth sinusoid, while τ is the time interval between the sampled points. The basis of the LP analysis is the relation (2) that applies

$$y_n = \sum_{i=1}^{2p} b_i y_{n+i} \quad (2)$$

to a set of exponential data given by eq 1; i.e., each data point can be expressed as a linear combination of the $2p$ following data points.[7] The backward prediction coefficients, b_i, in eq 2 are independent of n.

Primarily, the LP analysis yields the four parameters ω_j, R_{2j}, I_j, and φ_j, corresponding to each one of the p sinusoids.[6] This result could *in principle* be achieved by a direct fitting of eq 1 to the experimental FID signal by a least-squares (LSQ) procedure. However, since eq 1 is nonlinear, this would require extensive iterations as well as an estimation of reasonable initial values of the $4p$ parameters, which makes this solution impossible *in practice*, if the FID signal consists of more than a few sinusoids. The LP method overcomes this problem by breaking it down to the solution of two sets of linear equations and one n-degree polynomial equation. Thus, the LP method concentrates the nonlinearity of the problem in a polynomial equation that, furthermore, can be solved by a direct search method with no initial values needed.[6] Also, unlike the FT approach the LP method can distinguish between signal and noise,[7] just as it allows an estimation of the uncertainty of the obtained parameters.[6,8]

In ref 6 the application of the LP method was extended to the complicated time-domain NMR signals of peptides and small proteins. To ensure sufficient spectral resolution, this application requires linear LSQ calculations of very large orders and the solution of very high degree polynomial equations. As shown,[6] the numerical stability problem often associated with calculations of this order can be overcome by using the Cholesky decomposition in the LSQ calculations and the direct-search method[9] for solving the high-degree polynomial equations. Even so, the solution of the latter equations is possible only because the roots, corresponding to the resonances of high-resolution NMR FID's, are very close to the unit circle in the complex plane.[6]

Experimental Section

The insulins were all highly purified samples. The des-Ala^{B30} porcine insulin was prepared enzymatically from porcine insulin as described previously.[10] Human insulin was obtained[11] by substituting the B30 Ala of porcine insulin with Thr (semisynthetic insulin) or from human pancreas.

The ^{13}C NMR FID's were recorded with 8K data points and a sweep width of 17 241 Hz, using a Bruker HX 270 NMR spectrometer operating at a ^{13}C frequency of 67.9 MHz. The FID's consist of 12 000 scans obtained with a 60° pulse of 12 μs and a delay of 6 s between scans. All NMR samples contained 50 mg/mL of insulin in water (90% H_2O/10% D_2O) at pH 7.8.

The LP calculations were carried out on an RC 8000 minicomputer with 0.03×10^6 floating-point operations/s. The applied, optimized algorithms[6] were written in Algol and Assembler code.

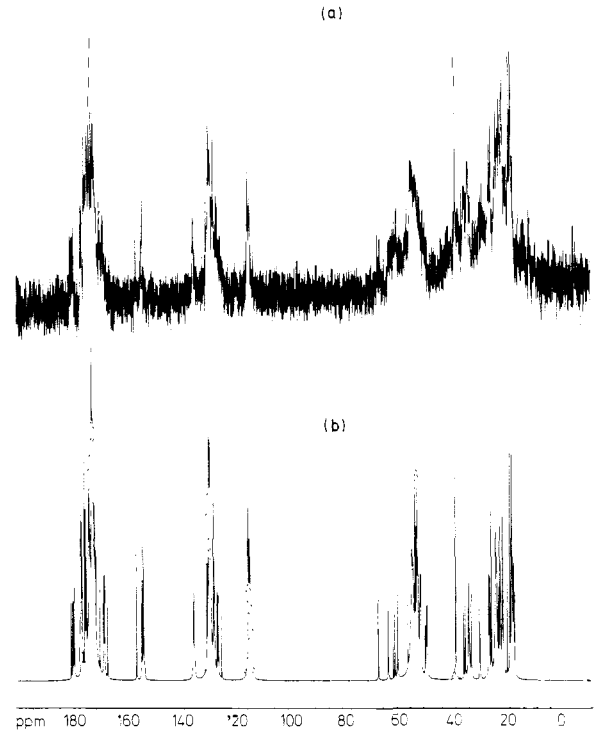

Figure 1. (a) ^{13}C FT NMR and (b) ^{13}C LP NMR absorption spectra of semisynthetic human insulin. The spectra were derived from the same FID. No digital filtering was applied to the FID. The LP spectrum was calculated from the spectral parameters obtained in an analysis with 2400 backward linear prediction coefficients. The 2400 order calculations took 90 h, on the RC 8000 minicomputer. Signals smaller than 3 times their standard deviations, and two signals entirely out of phase with the adjacent signals, were deleted, leaving 85 lines in the LP spectrum. The artifacts are probably due to deviations from purely exponential decays, as discussed previously.[6] In the spectrum displayed, all phases are set to zero.

Results and Discussion

The ^{13}C FID of semisynthetic human insulin was analyzed by the described LP procedure. The frequency spectrum calculated from the $4p$ parameters obtained in the LP analysis is shown in Figure 1 together with the conventional FT spectrum, derived from the same experimental FID. Clearly, most of the signals found by the LP analysis are indicated in the FT spectrum. But unlike the FT, the LP analysis gives quantitative measures of all the resolved signals in terms of values of the $4p$ spectral parameters. In particular ω_j and R_{2j} are of interest as potential sources of information about differences between the structures and the dynamics of the two insulins. However, in the LP spectrum in Figure 1b most of the resolved signals are still superpositions, corresponding to carbon atoms in slightly different environments. Hence, a reliable physical interpretation of the derived parameters can be given for only a few of the signals.

More simple and clearcut is the result in Figure 2a of an LP analysis of the difference between the FID's of semisynthetic human and porcine insulin. The LP analysis is particularly useful here compared with the FT method, since it allows the observation of difference signals between even very closely spaced signals, due to its higher resolution and sensitivity. In the FT approach this can be achieved only at the expense of a significant decrease in the signal to noise ratio.

In the spectrum in Figure 2a only signals reflecting differences between the two insulins remain. Among these, the carbon signals

(7) Kay, S. M.; Marple, S. L. *Proc. IEEE* **1981**, *69*, 1380–1419.
(8) Barkhuijsen, H.; De Beer, R.; van Ormondt, D. *J. Magn. Reson.* **1986**, *67*, 371–375.
(9) Svejgaard, B. *BIT* **1967**, *7*, 240–246.
(10) Schmitt, E. W.; Gattner, H. G. *Hoppe-Seyler's Z. Physiol. Chem.* **1978**, *359*, 799–802.
(11) Morihara, K.; Oka, T.; Tsuzuki, H. *Nature (London)* **1979**, *280*, 412–413.

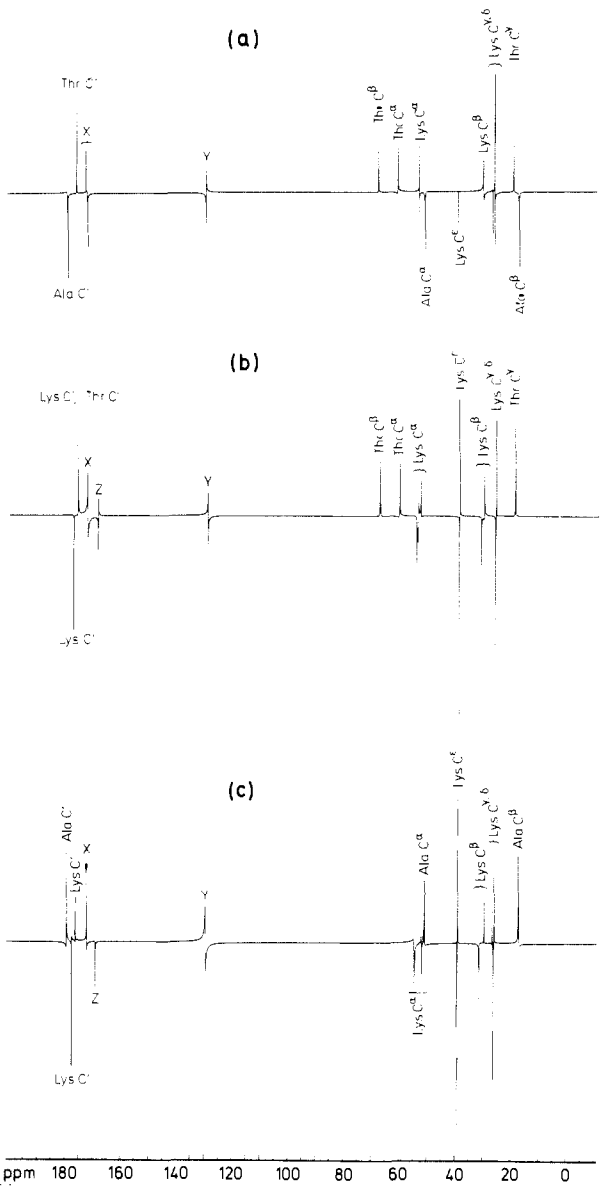

Figure 2. LP NMR absorption spectra obtained from LP analyses of the difference between the ^{13}C FID's of (a) human and porcine insulin, (b) human and des-Ala^{B30} porcine insulin, and (c) porcine and des-Ala^{B30} porcine insulin. The FID's were recorded under identical experimental conditions (see text). The details of the LP analyses were as described in the legend of Figure 1. The *x*, *y*, and *z* signals indicate as yet unidentified differences between the solution structures; see the text. The difference spectra were phase-corrected with the phase parameters from the FT spectra.

of B30 Thr and B30 Ala, corresponding to the only difference between the primary structures of human and porcine insulin, are immediately identified on the basis of their chemical shifts.[12] In addition, a series of difference signals in the region of 25–53 ppm appears. According to their positions these signals can be assigned to minor differences in the chemical shifts of the α and side-chain carbons of B29 Lys in the two insulins. This assignment is supported by the LP analysis of the difference between the FID's of human insulin and des-Ala^{B30} porcine insulin shown in Figure 2b. Since the B29 Lys is C-terminal in the latter derivative, the positions of the carbon signal from this residue will shift relative to those from B29 Lys in human insulin.[13] A similar result is obtained from an LP analysis of the difference between the FID's of porcine insulin and the des-Ala^{B30} derivative, as shown in Figure 2c. Thus, the LP analyses show spectral differences between human and porcine insulin involving B29 Lys. This may reflect differences between their secondary and tertiary structures around B29, although a pure environmental effect, caused merely by the dissimilarity between the adjacent amino acid residue in the two cases, could give rise to the observed changes in chemical shift. A similar difference between the structures of the two homologous hormones has not been observed in the crystal phase.[14] Another discrepancy between the two insulins is indicated by the difference signals at ca. 129.4 ppm (*y*) and ca. 172.4 ppm (*x*) in Figure 2a. The same discrepancy together with a difference signal at 168.8 ppm (*z*) appear in both of the difference spectra with des-Ala^{B30} insulin in Figure 2b,c. Although no specific assignments of these difference signals can be made at the moment, they emphasize the sensitivity of the LP approach presented here. Tentatively, the signal at 129.4 ppm in all three difference spectra could agree with the two C^{δ} carbons in a Tyr or a Phe residue (nonterminal), while the signal at 172.4 ppm may be assigned to the carbonyl carbons of the same two residues.[12] A possible candidate is the A19 Tyr, which, in the crystal structure, forms a hydrogen bond with the B25 Phe amide group through its carbonyl group.[15] Moreover, the length of this hydrogen bond varies among insulins from different species.[16] A variation of this type between the solution structures of human and porcine insulin, and of human or porcine insulin and the des-Ala^{B30} derivative, could explain the difference signals at 129.4 and 172.4 ppm. Likewise, the chemical shift (168.8 ppm) of the *z* signal could arise from a difference between the environments of the carbonyl carbon of one of the two N-terminal amino acids, i.e. A1 Gly and B1 Phe, in the insulins and the des-Ala^{B30} derivative, respectively.[12,13] It should be emphasized, however, that the assignments of the *x*, *y*, and *z* signals made here are merely suggestions. Unequivocal assignments and interpretations must await the results of more extensive NMR studies.

The LP analyses also indicate a larger mobility of the C-terminal residue in human insulin than in porcine insulin. Thus, the line widths of the $^{13}C^{\beta}$ and $^{13}C^{\gamma}$ signals of B30 Thr of the former insulin are about a factor of 2 smaller than that of the $^{13}C^{\beta}$ signal of B30 Ala of the latter. By the same token the mobility of B29 Lys seems larger in porcine insulin than in human insulin. Thus, the negative difference signal corresponding to C^{ϵ} of this residue is due entirely to a smaller line width in the porcine insulin.

For comparison, an LP analysis was performed on the difference between the FID's of semisynthetic human insulin and pancreatic human insulin. In this case no difference signals were found within the uncertainty of the method, indicating identical solution structures of these two insulins.

Conclusion

In conclusion we have found that LP is superior to FT in the analysis of complex NMR spectra of peptides and small proteins, by providing considerably more detailed and quantitative information about differences in structures and dynamics. The price to be paid for this improvement is a substantial increase in the computing time. However, the development in computing speed will, undoubtedly, reduce this time consumption considerably.

Acknowledgment. This work was supported by the Danish Technical Research Council, j.no. 16-3922.H, and the Danish Natural Science Research Council, j.no. 11-3742 and 11-5916. H.G. acknowledges a predoctoral fellowship from the Carlsberg Foundation. Free access to the RC 8000 computer at the H. C. Ørsted Institute is also acknowledged.

Registry No. Human insulin, 11061-68-0; porcine insulin, 12584-58-6; des-Ala^{30} porcine insulin, 39416-73-4; insulin, 9004-10-8.

(12) Richarz, R.; Wüthrich, K. *Biopolymers* **1978**, *17*, 2133–2141.
(13) Christl, M.; Roberts, J. D. *J. Am. Chem. Soc.* **1972**, *94*, 4565–4573.
(14) Hodgkin, D. C.; Dodson, E. J.; Dodson, G. G.; Reynolds, C. D. *Biochem. Soc. Trans.* **1983**, *11*, 411–417.
(15) Blundell, T.; Dodson, G. G.; Hodgkin, D. C.; Mercola, D. *Adv. Protein Chem.* **1972**, *26*, 279–402.
(16) Dodson, E. J.; Dodson, G. G.; Hubbard, R. E.; Reynolds, C. D. *Biopolymers* **1983**, *22*, 281–291.