Kitamura, S., & Sturtevant, J. M. (1989) *Biochemistry 28*, 3788-3792.
Matsumura, M., & Matthews, B. W. (1989) *Science 243*, 792-794.
Matthews, B. W., Nicholson, H., & Becktel, W. J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6663-6667.
Privalov, P. L. (1980) *Pure Appl. Chem. 52*, 479-497.
Remington, S. J., Anderson, W. F., Owen, J., Ten Eyck, L. F., Grainger, C. T., & Matthews, B. W. (1978) *J. Mol. Biol. 118*, 81-98.
Scheraga, H. A. (1978) *Pure Appl. Chem. 50*, 315.
Sturtevant, J. M. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2236-2240.
Suzuki, Y. (1989) *Proc. Jpn. Acad., Ser. B 65*, 146-148.
Tidor, B., & Karplus, M. (1991) *Biochemistry 30*, 3217-3228.
Weaver, L. H., & Matthews, B. W. (1987) *J. Mol. Biol. 193*, 189-199.

The Chemical Shift Index: A Fast and Simple Method for the Assignment of Protein Secondary Structure through NMR Spectroscopy†

D. S. Wishart,*,‡,§ B. D. Sykes,§ and F. M. Richards‡

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511, and MRC Group in Protein Structure and Function, Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

Received August 20, 1991; Revised Manuscript Received November 4, 1991

ABSTRACT: Previous studies by Wishart et al. [Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol.* (in press)] have demonstrated that ^{1}H NMR chemical shifts are strongly dependent on the character and nature of protein secondary structure. In particular, it has been found that the ^{1}H NMR chemical shift of the α-CH proton of all 20 naturally occurring amino acids experiences an upfield shift (with respect to the random coil value) when in a helical configuration and a comparable downfield shift when in a β-strand extended configuration. On the basis of these observations, a technique is described for rapidly and quantitatively determining the identity, extent, and location of secondary structural elements in proteins based on the simple inspection of the α-CH ^{1}H resonance assignments. A number of examples are provided to demonstrate both the simplicity and the accuracy of the technique. This new method is found to be almost as accurate as the more traditional NOE-based methods of determining secondary structure and could prove to be particularly useful in light of the recent development of sequential assignment techniques which are now almost NOE-independent [Ikura, M., Kay, L. E., & Bax, A. (1990) *Biochemistry 29*, 4659-4667]. We suggest that this new procedure should not necessarily be seen as a substitute to existing rigorous methods for secondary structure determination but, rather, should be viewed as a complement to these approaches.

For more than 20 years NMR spectroscopists have been attempting to apply chemical shift information to conformational problems of biological significance. Early efforts in this regard were first begun by Sternlicht and Wilson (1967) and Markley et al. (1967). Both groups were interested in studying the chemical shift tendencies of α-CH ^{1}H NMR[1] resonances in amino acid homopolymers, particularly with respect to the systematic changes in proton chemical shifts that were associated with helix formation and helix disruption in these compounds. Subsequent studies by Clayden and Williams (1982) and Dalgarno et al. (1983), based on accumulated data from naturally occurring proteins, suggested that reasonably strong conformationally dependent chemical shift tendencies existed in β-strands as well as in α-helices and that these trends were not confined to certain homopolymers or to selected solvent conditions. More recent work by Szilagyi and Jardetzky (1989) and Wishart et al. (1991) have confirmed these early observations by placing them on a more solid statistical basis. In fact, these workers have clearly demonstrated that a strong relationship exists between α-CH ^{1}H NMR chemical shifts and protein secondary structure for all 20 amino acids.

Some of these observations have already begun to be put to use. Pastore and Saudek (1990) have recently described a useful method for displaying chemical shift information which permits the qualitative identification of secondary structure in proteins. This procedure is based on plotting "smoothed" chemical shift differences (with respect to random coil values) versus protein sequence and using the resulting curve to approximate the location and identity of secondary structures. However, because of the qualitative nature of these plots it is often difficult to identify the exact limits as well as the true identity of all significant secondary structures in the protein of interest.

We describe a new method for secondary structure determination based on chemical shift propensity that was developed quite independently of Pastore and Saudek's work. This particular technique can be used for the precise identification of protein secondary structure from chemical shift information alone. It is fast, simple, and accurate and can be used either

† Financial support by the MRC Group in Protein Structure and Function and by the National Institutes of Health (NIH Grant GM-22778) is gratefully acknowledged.
* To whom correspondence should be addressed.
‡ Yale University.
§ University of Alberta.

[1] Abbreviations: BPTI, bovine pancreatic trypsin inhibitor; CSI, chemical shift index; DSS, 2,2-dimethyl-2-silapentane-5-sulfonic acid; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; TMS, tetramethylsilane; TSP, sodium 3-(trimethylsilyl)propionate.

0006-2960/92/0431-1647$03.00/0

Table I: Summary of Peptides and Proteins Used in This Analysis

protein (no. of residues)	source	conditions	reference
α-bungarotoxin (74)	*Bungarus multicinctus*	pH 4.0, 35 °C	Basus (1988)
α-neurotoxin (60)	*Dryopteris polylepsis*	pH 4.2, 36 °C	Labhardt (1988)
anaphylotoxin C5a (75)	bovine	pH 2.3, 10 °C	Zuiderwig (1988)
antennapedia homeo (68)	*Drosophila melanogaster*	pH 4.3, 20 °C	Billeter (1990)
arc repressor (53)	phage P22	pH 5.4, 50 °C	Breg (1989)
histidine protein (85)	*Escherichia coli*	pH 6.5, 30 °C	Wittekind (1990)
histone H5 (77)	chicken	pH 3.7, 25 °C	Zarbock (1986)
interleukin 8 (72)	human	pH 5.2, 40 °C	Clore (1989)
lac Repressor (51)	*E. coli*	pH 6.9, 18 °C	Zuiderwig (1983)
plastocyanin (99)	spinach	pH 7.3, 35 °C	Driscoll (1987)
thioredoxin (108)	*E. coli*	pH 5.7, 35 °C	Dyson (1989)
trypsin inhibitor (58)	bovine	pH 4.6, 36 °C	Wagner (1987)

by itself or in conjunction with other NMR-based methods to determine protein secondary structure. All that is required is that the α-CH ^{1}H NMR assignments of the protein of interest be known.

Materials and Methods

Chemical Shift and Secondary Structure Assignments. Chemical shift data were collected entirely from published sources. All proton chemical shift values were referenced to either DSS, TSP, or TMS. No corrections were introduced for temperature, pH, or possible intrinsic shift differences between these three standards.

Table I provides a detailed list of the 12 proteins selected for this study. A more comprehensive list of all 75 proteins used to develop and test the procedure appears in Wishart et al. (1991). Included in Table I are the number of residues present, the conditions under which the data were collected, and the reference from which the sequential assignments were obtained. Only proteins dissolved in aqueous conditions were included in this work.

Secondary structure was assigned to one of three categories: α-helix (H), β-strand (B), or coil (C).[2] Assignments were generally made according to the secondary structures indicated in the original source. Where possible, these were confirmed by secondary structure assignments obtained from published crystal structures of the identical protein or from the coordinates of strongly homologous proteins. Discrepancies were usually resolved in favor of the proposed NMR structure.

It is important to remember that secondary structures are often notoriously difficult to define, particularly at their initiation and termination points. As a result, the secondary structure assignments made for this or any other study of a similar nature should not be considered to be absolute or immutable—they are only "best guess" estimates.

Identification of Secondary Structure Using Chemical Shifts. The method described below provides a technique for rapidly and accurately determining the type and location of secondary structures in proteins from proton chemical shift information alone. Only the α-CH ^{1}H NMR chemical shifts for the protein of interest need be known or at least partially known in order to apply this technique. No information regarding NOE intensities, NOE buildup curves, amide exchange rates, *J* couplings, or any additional chemical shift information is required.

Table II: Chemical Shift Values of α-Protons Used in the Determination of Secondary Structure

residue	α-^{1} H range (ppm)	residue	α-^{1} H range (ppm)
Ala	4.35 ± 0.10	Met	4.52 ± 0.10
Cys	4.65 ± 0.10	Asn	4.75 ± 0.10
Asp	4.76 ± 0.10	Pro	4.44 ± 0.10
Glu	4.29 ± 0.10	Gln	4.37 ± 0.10
Phe	4.66 ± 0.10	Arg	4.38 ± 0.10
Gly	3.97 ± 0.10	Ser	4.50 ± 0.10
His	4.63 ± 0.10	Thr	4.35 ± 0.10
Ile	3.95 ± 0.10	Val	3.95 ± 0.10
Lys	4.36 ± 0.10	Trp	4.70 ± 0.10
Leu	4.17 ± 0.10	Tyr	4.60 ± 0.10

The method is almost entirely empirical in nature although it is strongly based on the observation of the conformational chemical shift tendencies of α-protons which have been recently noted (Wishart et al., 1991). The technique is essentially a two-stage digital filtration process. In the first stage, a chemical shift index (a ternary index having values −1, 0, and 1) is assigned to all identifiable residues on the basis of their α-proton chemical shift values. In the second stage, secondary structures are delineated and subsequently identified on the basis of the values and "densities" of these chemical shift indices. The exact protocol is as follows:

(1) For the protein of interest, obtain the sequential assignment of backbone α-protons using standard 2-D or 3-D NMR techniques.

(2) Using the chemical shift values in Table II as a guide, carry out the following procedure for each residue in the protein: (a) If the α-proton chemical shift is greater than the range given in Table II for that residue, mark a "1" beside it. (b) If the α-proton chemical shift is less than the range given in Table II for that residue, mark a "−1" beside it. (c) If the α-proton chemical shift is within the given range in Table II for that residue, mark a "0" beside it.

The above procedure defines the chemical shift index for each residue in the protein. Using these chemical shift indices, we proceed to identify the secondary structures as follows:

(3) Any "dense" grouping of four or more "−1's" not interrupted by a "1" is a helix. Any "dense" grouping of three or more "1's" not interrupted by a "−1" is a β-strand. All other regions are designated as coil.

(4) A local "density" of nonzero chemical shift indices which exceeds 70% is required when defining regions of helical or extended structure. The local density may be measured over a window of four or five residues. A minimum of three consecutive "1's" is needed to define a β-strand, and a minimum of four (not necessarily consecutive) "−1's" is needed to define a helix. All remaining regions not identified as either helix or β-strand or those regions where the local density of "1's" and "−1's" falls below 70% are defined as "coil".

(5) Termination points (at either end) of helices or β-strands can often be recognized by the first appearance of chemical shift indices that are opposite in magnitude to those of the corresponding secondary structure. In cases where this does not occur, the first appearance of two consecutive zero-valued chemical shift indices marks the termination point.

As is evident from the description given above, the method is completely objective and, therefore, easily computerizable. However, given that the procedure was originally devised with

[2] The term "coil" often implies random coil, which is quite inappropriate despite its wide use. In this paper it is taken to be all structures (loop, random coils, turns, etc.) which are neither α-helix nor β-strand.

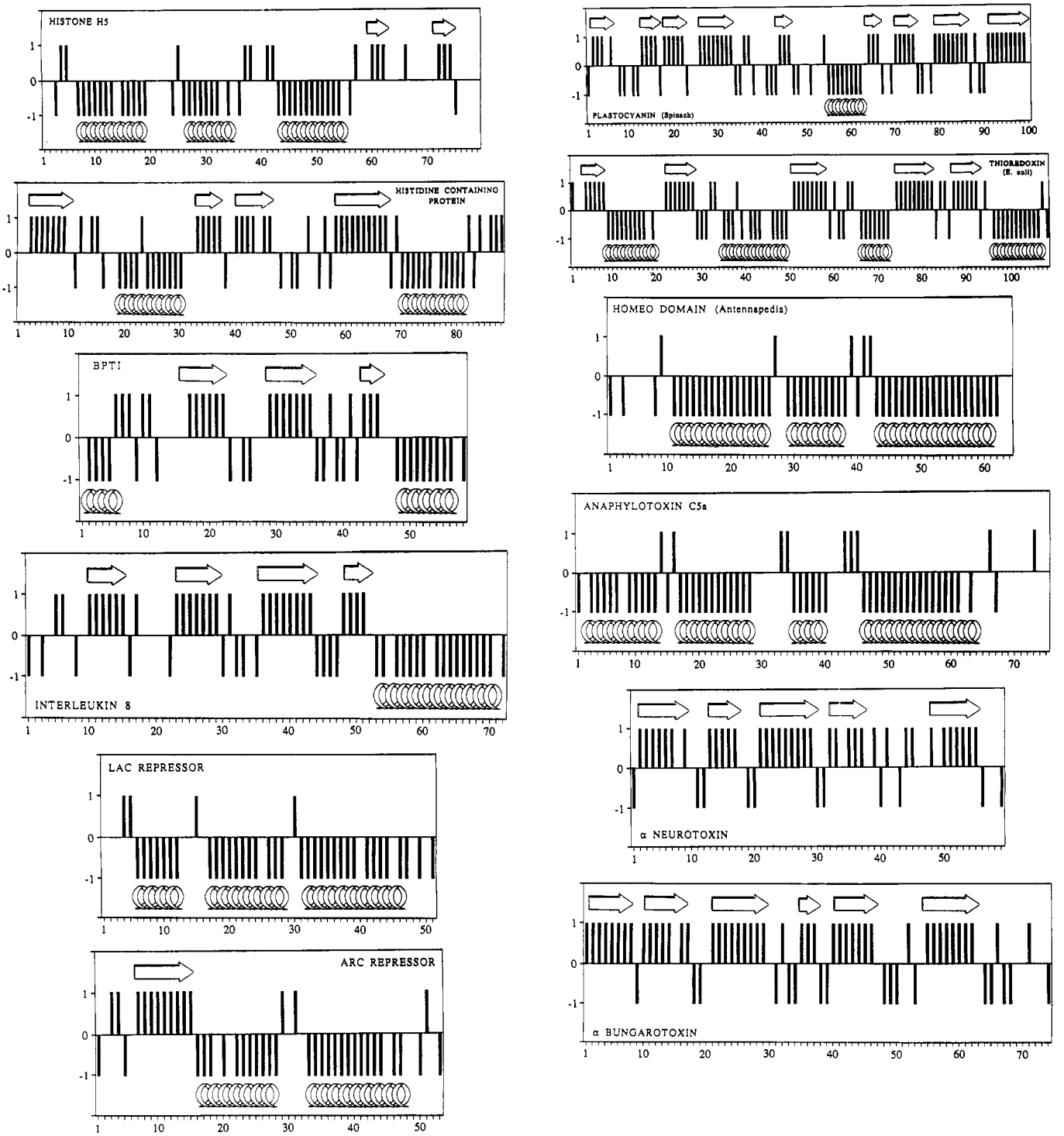

FIGURE 1: Chemical shift index of α-protons as plotted for 12 selected proteins. Proteins from α, β, α/β, and mixed folding classes have been chosen to demonstrate the general applicability of the procedure. Chemical shift values were taken from current literature sources under experimental conditions listed in Table I.

simplicity in mind, it is hoped that one should never require more than a pencil and a piece of paper in order to complete a secondary structure assignment. Indeed, with a little practice, one can easily assign and characterize a medium-sized protein (80 residues), by hand, in a little less than a minute.

RESULTS

Figure 1 provides a graphical demonstration of the procedure as applied to a selected set of 12 proteins. The secondary structure, as determined by X-ray crystallography or by more elaborate NMR measurements, is indicated with the usual schematic ribbon representation. Inspection of these diagrams should indicate that the method is both accurate and consistent in determining the type and the location of secondary structural elements. These chemical shift index (CSI) plots should also assist in clarifying any ambiguities in the rules described above. They may also be used as references or examples when attempting to apply this procedure on new proteins or if one desires to encode the CSI protocol on a computer.

It is not hard to see how the method could be improved upon, particularly if one wishes to completely computerize the procedure (perhaps as part of a more comprehensive automated assignment package). Indeed, the "noise" from most of these CSI plots could be easily eliminated with the addition

of a second filtering step such as median data sieve (Bangham, 1988). However, these esthetic improvements would take away from the simplicity of the procedure and its original intention as a "pencil and paper" technique.

It is important to note, however, that the CSI method is not completely foolproof nor is it always correct (an accuracy of 90–95% is more typical). In the course of testing this procedure on more than 50 proteins, the inevitable ambiguity or exception would sometimes arise. Caveats, modifications, or addenda to the general rules stated above could be helpful in clarifying some potential problems. Some of these are (1) The procedure works best for proteins where assignments have been carried out between pH 3.0 and 8.0 and temperatures between 15 and 50 °C. (2) Application of the CSI method to proteins with paramagnetic centers can lead to some misleading results, particularly among residues located very close to the paramagnetic core. (3) When determining the CSI for glycine, one should use the average value of the two nonequivalent α-protons as the glycine chemical shift. (4) The procedure is not intended to wholly replace existing rigorous methods for secondary structure determination in NMR. Rather, it should be viewed as a complement to these methods. Its primary appeal should be the speed and ease with which it can be applied.

Discussion and Conclusions

The chemical shift index method for determining secondary structure essentially resembles a digital filtration technique wherein widely varying values (measured chemical shifts) are filtered using some preassigned threshold (Table II) to yield a series of digital "−1's", "0's", and "1's". As is typical of most digital filtration techniques, this produces a visually pleasing and/or a more easily interpretable description of chemical shifts than might otherwise be attained using conventional data-smoothing techniques. Furthermore, this digital approach permits the easy delineation of boundaries and the simple classification of structures which make the technique both objective and accurate.

The success of the CSI protocol is very much based on the proper choice of reference chemical shifts. It is, therefore, important to note that the chemical shift values presented in Table II are generally ±0.10 ppm about the random coil chemical shifts originally quoted by Wüthrich (1986). The fact that a range or chemical shift "bandwidth" was chosen as opposed to a single cutoff value was due, in part, to the consideration of temperature, pH, and calibration variations in α-proton chemical shifts which often vary from sample to sample or from lab to lab. This approach also helps to make the CSI technique very robust, allowing it to be applied to spectra collected under a rather wide, if not unusual, variety of conditions.

It is also worth noting that during the course of developing and testing this procedure it was found necessary to change the range and hence the "expected" random coil chemical shift values for some residues. This was originally done in order to improve the accuracy and general applicability of the technique. As it happened, small changes in the random coil values of three aliphatic residues (in particular Ile, Leu, and Val) produced significant improvements in the overall accuracy. As a result, the "expected" random coil chemical shift values for these three residues were reduced by about 0.20 ppm over those quoted by Wüthrich (1986). Interestingly, these changes were later found to be quite comparable to the reduced random coil chemical shift values for these residues which were recently calculated in a newly updated protein sequence/chemical shift database (Wishart et al., 1991).

While the CSI method was originally developed for the quick identification of α-helices and β-strands, it appears likely that it may be applied toward the identification of other types of secondary structures as well. Preliminary indications are that certain types of reverse or β-turns are often manifested by the appearance of coupled "1, −1" or "−1, 1" CSI patterns in "coiled" regions or at the ends of β-strands. While these patterns may prove to be difficult to distinguish from other chemical shift "noise", it is hoped that refinements to the general CSI procedure or to the values in Table II may eventually improve the method to the point that β-turns could be identified with the same reliability as helices and β-strands. Work is currently in progress to attempt to add such a feature to the general CSI protocol.

In addition to its general utility toward interpreting ^{1}H resonances, the CSI method is also applicable to ^{13}C resonances. Reasonably good estimates of secondary structure have been made for both BPTI and calmodulin using either α-carbon or carbonyl carbon chemical shift values (data not shown). The technique is hindered, however, by the lack of good random coil values for ^{13}C chemical shifts and by the difficulty in finding a large enough sample to test and thereby optimize the technique. The situation may improve in the new few years as more carbon chemical shift data becomes available.

The CSI protocol may prove to be particularly useful in the areas of automated and NOE-independent (Ikura et al., 1990) assignment techniques—two approaches which are becoming increasingly popular among NMR spectroscopists today. Both of these "automated" procedures typically generate long lists of sequential assignments without providing the user with much, if any, explicit information on NOE connectivities or secondary structure. As a result, one must either acquire additional spectra or reinterpret existing data to determine the existence or location of putative secondary structural elements. By implementing the CSI algorithm with any one of these computerized procedures, it would be quite simple to assign putative or potential secondary structural elements to the protein of interest. These "estimates" could then serve as excellent starting points in analyzing subsequent NOESY data or in obtaining more refined structural measurements. Such an approach could offer a tremendous saving in time and greatly add to the utility of any existing automated assignment routine.

While it may be tempting to use the CSI protocol on its own, we must caution the reader that the CSI procedure, as it presently stands, is not intended to wholly replace existing rigorous methods for secondary structure determination in NMR. Instead, we hope the CSI technique will eventually find its place as a quick and easily implemented complement to existing structural assignment strategies and, in so doing, make the NMR spectral analysis of protein structure just a little easier and perhaps a little more complete.

Acknowledgments

We thank Colleen Wilson for her help in preparation of the manuscript.

References

Basus, V. J., Billeter, M., Love, R. A., Stroud, R. M., & Kuntz, I. D. (1988) *Biochemistry 27*, 2763–2768.

Bangham, J. A. (1988) *Anal. Biochem. 174*, 142–145.

Billeter, M., Qian, Y. Q., Otting, G., Muller, M., Gehring, W. J., & Wüthrich, K. (1990) *J. Mol. Biol. 214*, 183–197.

Breg, J. N., Boelens, R., George, A. V. E., & Kaptein, R. (1989) *Biochemistry 28*, 9826–9833.

Clayden, N. J., & Williams, R. J. P. (1982) *J. Magn. Reson. 49*, 383-396.
Clore, G. M., Appella, E., Yamada, M., Matsushima, J., & Gronenborn, A. M. (1989) *J. Biol. Chem. 264*, 18907-18911.
Dalgarno, D. C., Levine, B. A., & Williams, R. J. P. (1983) *Biosci. Rep. 3*, 443-452.
Driscoll, P. C., Hill, A. O., & Redfield, C. (1987) *Eur. J. Biochem. 170*, 279-292.
Dyson, H. J., Holmgren, A., & Wright, P. E. (1989) *Biochemistry 28*, 7074-7087.
Ikura, M., Kay, L. E., & Bax, A. (1990) *Biochemistry 29*, 4659-4667.
Labhardt, A. M., Hunziker-Kwik, E. H., & Wüthrich, K. (1988) *Eur. J. Biochem. 177*, 295-305.
Markley, J. L., Meadows, D. H., & Jardetzky, O. (1967) *J. Mol. Biol. 27*, 25-35.
Pastore, A., & Saudek, V. (1990) *J. Magn. Reson. 90*, 165-176.
Sternlicht, H., & Wilson, D. (1967) *Biochemistry 6*, 2881-2892.
Szilagyi, L., & Jardetzky, O. (1989) *J. Magn. Reson. 83*, 441-449.
Wagner, G., Braun, W., Havel, T. E., Shaumann, T., Go, N., & Wüthrich, K. (1987) *J. Mol. Biol. 196*, 611-639.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol.* (in press).
Wittekind, M., Reizer, J., & Klevit, R. E. (1990) *Biochemistry 29*, 7191-7200.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Zarbock, J., Clore, G. M., & Gronenborn, A. M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7628-7632.
Zuiderwig, E. R. P., Kaptein, R., & Wüthrich, K. (1983) *Eur. J. Biochem. 137*, 279-292.
Zuiderwig, E. R. P., Millison, K. W., Henkin, J., & Carter, G. W. (1988) *Biochemistry 27*, 3568-3580.

Proton Nuclear Overhauser Effect Study of the Heme Active Site Structure of Chloroperoxidase†

L. B. Dugad, X. Wang, C.-C. Wang, G. S. Lukat, and H. M. Goff*
Department of Chemistry, University of Iowa, Iowa City, Iowa 52242
Received July 2, 1991; Revised Manuscript Received October 23, 1991

ABSTRACT: Chloroperoxidase, a glycoprotein from the mold *Caldariomyces fumago*, has been investigated in its ferric low-spin cyanide-ligated form through use of nuclear Overhauser effect (NOE) spectroscopy to provide information on the heme pocket electronic/molecular structure. Spin-lattice relaxation times for the hyperfine-shifted heme resonances were found to be three times less than those in horseradish peroxidase. This must reflect a slower electronic relaxation rate for chloroperoxidase than for horseradish peroxidase as a consequence of axial ligation of cysteine in the former versus histidine in the latter enzyme. Isoenzymes A_1 and A_2 of chloroperoxidase show the largest chemical shift differences near the heme propionate on the basis of NOE measurements. This suggests that the primary structure differences for the two isoenzymes are communicated to the heme group through the ring propionate substituents. A downfield peak has been detected in chloroperoxidase with chemical shift, T_1, and line width characteristics similar to those of the Cε-H proton of the distal histidine in horseradish peroxidase. This finding is in agreement with a previous suggestion for a distal histidine residue. The NOE pattern and T_1's of the peaks in the 0.0 to −5.0 ppm upfield region are consistent with the presence of an arginine amino acid residue in the heme pocket near either the 1-CH_3 or 3-CH_3 group. Existence of catalytically important distal histidine and arginine amino acid residues in chloroperoxidase shows it to be structurally similar to peroxidases rather than to the often compared monooxygenase, cytochrome P-450. This result supports the earlier conclusions of Sono et al. [Sono, M., Dawson, J. H., Hall, K., & Hager, L. P. (1986) *Biochemistry 25*, 347-356].

Chloroperoxidase (CPO)[1] has been the subject of several investigations because of its versatile catalytic properties typical of peroxidases, catalases, and oxygenases (Hewson & Hager, 1979; Dawson & Sono, 1987; Dawson, 1988). In addition, CPO catalyzes halogenation (except fluorination) of substrates in the presence of halide ions and hydrogen peroxide (Hager et al., 1966).

Chloroperoxidase is a glycoprotein (42 kDa) with a heme prosthetic group (Figure 1A) secreted by the mold *Caldariomyces fumago* (Morris & Hager, 1966). The axial ligand to the heme iron atom in chloroperoxidase, unlike other known heme peroxidases, is a cysteine sulfur atom (Bangcharoenpaurpong et al., 1986; Blanke & Hager, 1988). Various spectral properties of chloroperoxidase are similar to those of the monooxygenase cytochrome P-450, presumably due to presence of a cysteine axial ligand to the heme in both enzymes (Dawson & Sono, 1987; Dawson, 1988). The high-resolution X-ray crystal structure of cyt P-450cam is available (Poulos

† This work was supported by NIH Grant GM 28831.

[1] Abbreviations: CPO, chloroperoxidase; HRP, horseradish peroxidase; CCP, cytochrome *c* peroxidase; CPOCN, CCPCN, and HRPCN, cyanide-ligated ferric low-spin complexes of CPO, CCP, and HRP, respectively; cyt P-450, cytochrome P-450; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect.

0006-2960/92/0431-1651$03.00/0