Robbins, D. C., Shoelson, S. E., Rubenstein, A. H., & Tager, H. S. (1984) *J. Clin. Invest. 73*, 714–719.
Rupley, J. A., Renthal, R. D., & Praissman, M. (1976) *Biochim. Biophys. Acta 140*, 185.
Smith, G. D., Swenson, D. C., Dodson, G. G., & Reynolds, C. D. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7093–7097.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982), *J. Magn. Reson. 48*, 286–292.
Steiner, D. F. (1979) *Diabetes 27*, Suppl. 1, 145–148.
Steiner, D. F. (1984) *Harvey Lect. 78*, 191–228.
Steiner, D. F., & Oyer, P. E. (1967) *Proc. Natl. Acad. Sci. U.S.A. 57*, 473–480.
Steiner, D. F., Cunningham, D. D., Spigelman, L., & Aten, B. (1967) *Science 157*, 697–700.
Steiner, D. F., Cho, S., Oyer, P. E., Terris, S., Peterson, J. D., & Rubenstein, A. H. (1971) *J. Biol. Chem. 246*, 1365–1374.
Strickland, E. H., & Mercola, D. A. (1976) *Biochemistry 15*, 3875–3884.
Thim, L., Hansen, M. T., Norris, K., Hoegh, I., Boel, E., Forstrom, J., Ammerer, G., & Fil, N. P. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6766–6770.
Weiss, M. A., Eliason, J. A., & States, D. J. (1984) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6019–6023.
Weiss, M. A., Nguyen, D. T., Khait, I., Inouye, K., Frank, B. H., Beckage, M., O'Shea, E., Shoelson, S. E., Karplus, M., & Neuringer, L. J. (1989) *Biochemistry 28*, 9855–9873.
Williamson, K. L., & Williams, R. J. P. (1979) *Biochemistry 18*, 5966–5972.
Wodak, S. J., Alard, P., Delhaise, P., & Renngboog, S. C. (1984), *J. Mol. Biol. 151*, 317–322.
Wollmer, A., Fleischhauer, J., Strassburger, W., Thiele, H., Bradenbury, D., Dodson, G., & Mercola, D. (1979) *Biophys. J. 20*, 233–243.
Wood, S. P., Blundell, T. L., Wollmer, A., Lazarus, N. R., & Neville, R. W. J. (1975) *Eur. J. Biochem. 55*, 531–542.
Wuthrich, K., Wider, G., Wagner, G., & Braun, W. (1983) *J. Magn. Reson. 155*, 311.

Sequential ^{1}H NMR Assignments and Secondary Structure of Aponeocarzinostatin in Solution†

M. Lyndsay Remerowski,*,‡ Steffen J. Glaser,‡ Larry C. Sieker,§ T. S. Anantha Samy,∥ and Gary P. Drobny‡
Departments of Chemistry and of Biological Structure, University of Washington, Seattle, Washington 98195, and Department of Oncology, University of Miami, Miami, Florida 33101

Received March 26, 1990; Revised Manuscript Received June 13, 1990

ABSTRACT: Sequential assignments and secondary structural analysis have been accomplished for the 113-residue apoprotein of the antitumor drug neocarzinostatin (NCS) from *Streptomyces carzinostaticus*. A total of 98% of the main-chain and 77% of the side-chain resonances have been sequence specifically assigned by use of information from coherence transfer experiments and by sequential and interstrand NOEs. Because of the complexity of the NCS spectrum, several sequential assignment strategies were employed to complete the analysis. Apo-NCS consists of three antiparallel β-sheeted domains by NMR analysis. There is an extensive four-strand antiparallel β-sheet, and two two-stranded domains. One of the two-strand domains is contiguous, S72–N87, with chain reversal occurring through the region L77–R82. The other two-stranded domain has the section G16–A24 antiparallel with respect to the region S62–R70. This secondary structure is consistent with the crystal structure of holo-NCS at 2.8-Å resolution.

Neocarzinostatin (NCS)[1] is a protein antitumor drug first isolated from *Streptomyces carzinostaticus* culture medium 25 years ago (Ishida et al., 1965). NCS is perhaps the best studied of the *Streptomyces* anticancer agents. It has a relatively low toxicity (LD_{50} 10–30 mg/kg in mice and dogs) and broad in vivo activity against human leukemia, bladder cancer, liver metastasis, and sarcoma (Montgomery et al., 1981). Although NCS is 10–100 times less toxic than the *Streptomyces* antitumor antibiotics auromomycin (AUR) and actinoxanthin (AXN), the polypeptide sequence of NCS shows marked homologies with those of AUR and AXN suggesting that they are somehow related in structure and perhaps function (Goldberg et al., 1981).

A detailed molecular basis for the action of NCS is still lacking, but the last several years have seen an enormous expansion of information on the molecule. Holo-NCS has been separated into a polypeptide component (apo-NCS) of MW ca. 11 100 and a tightly bound chromophore (Chr, termed NCS-Chr) of composition $C_{35}H_{33}NO_{12}$ (MW 659). This non-protein chromophore has been found to consist of four

† This research was supported by NIH RO1 CA45643-01A1 and a NATO Fellowship from the Deutsche Akademische Austauschdienst (DAAD) to S.J.G.
‡ Department of Chemistry, University of Washington.
§ Department of Biological Structure, University of Washington.
∥ University of Miami.

[1] Abbreviations: NMR, nuclear magnetic resonance; NCS, neocarzinostatin; Chr, chromophore AUR, auromomycin; AXN, actinoxanthin; MCR, macromomycin; FABMS, fast atom bombardment mass spectroscopy; GCMS, gas chromatography–mass spectroscopy; MIR, multiple isomorphous replacement; 2D, two dimensional;EDTA, ethylenediaminetetraacetic acid; COSY, 2D correlated spectroscopy; DQF, double quantum filtered; TQF, triple quantum filtered; RELAY, 2D relayed coherence transfer spectroscopy; TOCSY, 2D total correlattion spectroscopy, NOE, nuclear Overhauser enhancement; NOESY, 2D NOE spectroscopy; TPPI, time-proportional phase incrementation; ppm, parts per million; d_{AB}, the NOE connectivity between protons A and B of the same (intraresidue) or different (interresidue) amino acids in a polypeptide (protons A and B are designated N for amide protons, α for $C^{\alpha}H$, β for $C^{\beta}H$, and δ for $C^{\delta}H$).

0006-2960/90/0429-8401$02.50/0

parts: 2-hydroxy-7-methoxy-5-methyl-1-naphthoate and 2,6-dideoxy-2-(methylamino)galactose linked to a $C_{15}H_8O_4$ substituent consisting of an ethylene cyclic carbonate group; a highly strained ether epoxide attached to a novel bicyclo-[7.3.0]dodecadyne system (Edo et al., 1985).

The antibiotic and antitumor activity of NCS has been the subject of an extensive series of biological studies. NCS inhibits cell growth by inhibiting DNA replication, and this activity is correlated with DNA cleavage, both in vivo and in vitro (Beerman & Goldberg, 1974, 1977). Studies with agarose-immobilized NCS (Lazarus et al., 1977), as well as studies with resolved chromophore and apoprotein, reveal that the DNA-damaging activity resides exclusively in the non-protein chromophore (Kappen et al., 1980a; Suzuki et al., 1980); pure chromophore has at least as much biological activity toward cell growth and DNA scission as holo-NCS, while apo-NCS is without effect (Montgomery et al., 1981; Napier et al., 1980). Earlier in vivo studies with high levels of NCS suggested that the intact holoprotein penetrates into the interior of cells and into cell nuclei (Maeda et al., 1975; Sakamoto et al., 1979). A more recent model for the action of intact holo-NCS invokes a carrier or transport function for the polypeptide which stabilizes the labile chromophore, extruding it across the cell membrane into the cytoplasm. Such a role nicely explains the retained activity of agarose-bound NCS (Lazarus et al., 1977), but the earlier evidence of the entry of both ^{14}C- and fluorescein-labeled protein into cells cannot be dismissed either. It is also known that pure apoprotein blocks in vitro DNA cleavage by NCS-bound chromophore or pure chromophore in a dose-dependent manner (Napier et al., 1980) due to the extremely tight binding of chromophore by apoprotein with a K_d of ca. 1×10^{-10} M (Goldberg et al., 1981).

The amino acid sequence of NCS was first reported by Meienhofer and co-workers (Meienhofer et al., 1972; Maeda et al., 1974) using the Edman degradation procedure. The sequence was subsequently investigated and revised with the methods of FABMS and GCMS (Biemann, 1982; Gibson et al., 1984). The most recent revision, again utilizing mass spectrometric studies, appeared in 1986 (Hirayama et al., 1986). That sequence appears in Figure 6.

Before the application of 2D NMR to the study of proteins and nucleic acids, X-ray crystallography was the only method used to obtain detailed structural information on large biomolecules. Solution structures derived from NMR data offer an alternate view of the molecule, expanding the amount of information on a molecular system and allowing new insights based on comparisons between the structures in crystalline and solution states.

X-ray diffraction data at 3.5-Å resolution indicate that NCS is composed of a seven-strand antiparallel β-sandwich. The sandwich is formed by a three-strand β-sheet and a four-strand β-sheet. A smaller lobe portion of the molecule is composed of two loops oriented somewhat perpendicularly to the β-sandwich, forming a cleft between the sandwich and lobe portion of the molecule (Sieker, 1981).

The chromophore is located in the distinct cleft between the body of the β-sandwich and the lobe portion. This fact is verified by diffraction studies of DNA intercalator complexes in isomorphous crystals of apo-NCS where the intercalator is shown, by electron density difference maps, to reside in the putative binding site (Sieker, 1987).

When we began our NMR investigation of NCS, interpretation of the 2.8-Å map of holo-NCS was hindered by the lack of interpretable electron density in the region of the chromophore and the adjoining space where residues 39–45 ought to be . Three possible reasons were postulated to account for the situation: (a) The binding site may become disordered during the long-term crystallization process, possibly due to decomposition of the chromophore. As a result, the closely associated residue 39–45 segment may also become disordered from its response to changes in chromophore structure. (b) The cleft region is a binding site for heavy atoms in each of the heavy-atom derivatives. This may have caused the region to become nonisomorphous with the native crystal resulting in poor MIR phasing and indistinct density. (c) It is possible that the loop is flexible and has several configurations. One motivation for our investigation was the possibility that NMR spectroscopy could yield structural information about this region of the protein.

Other proteins derived from *Streptomyces* cultures are known to have antitumor activity and possess non-protein chromophores. AXN with 108 residues has about 40% sequence homology with NCS (Khoklov et al., 1976). Macromomycin (MCR) was isolated by Umezawa (Chimura et al., 1968) and was found later to be the apoprotein of AUR (Kappen et al., 1979). With 112 residues AUR has about 50% sequence homology with NCS and about 40% sequence homology with AXN (Samy et al., 1983). Results of X-ray studies confirm that MCR (Van Roey & Beerman, 1989), AXN (Pletnev et al., 1982) and NCS (Sieker and Samy, unpublished experiments) have similar overall structures, including a well-defined binding site cleft, as the degree of sequence homology suggests.

Both NCS-Chr and AUR-Chr act by causing single- or double-strand breaks in DNA, but they have key differences in the mechanisms of their activity (Kappen et al., 1979; Suzuki et al., 1980). The structure of the NCS-Chr is known (Edo et al., 1985; Myers et al., 1988), and its interactions with DNA have been studied by several research groups. The NCS-Chr naphthoic moiety intercalates into DNA via the minor groove (Povirk et al., 1981; Dasgupta & Goldberg, 1985) and causes strand scissions of DNA preferentially at thymidine or adenosine residues (Takeshita, 1981). The composition of the AUR-Chr is only partially elucidated; it is similar to but not identical with the chromophore of NCS (Kumada et al., 1983), and it exhibits a different site selectivity for DNA strand scission (Takeshita et al., 1981; Kappen et al., 1979; Suzuki et al., 1980). Despite the sequence homology between NCS and AUR, apo-NCS does not bind AUR-Chr well (Kappen et al., 1980b), nor does MCR stabilize the NCS-Chr properly (Kappen & Goldberg, 1980). Because of both the homology in structure and function and the variation in Chr structures and binding site composition, this family of proteins offers a unique opportunity for studies of structure–function relatioships and Chr–protein interactions. Continued investigations of these homologous systems are of interest because of their DNA-damaging activity. In particular, information about the Chr–protein interactions will advance research in the development of improved antitumor agents.

In this paper we report the sequential assignment and determination of the secondary structure of apo-NCS in solution. We compare our findings to the recently completed X-ray structure of holo-NCS at 2.8-Å resolution (Sieker and Samy, unpublished experiments). Our assignment of over 85% of the resonance lines in the spectrum of apo-NCS forms a solid basis for future NMR studies of the chromophore–protein interaction and detailed analysis of the solution structure of NCS.

Materials and Methods

Two apo-NCS preparations were used in our experiments:

(1) Lyophilized apo-NCS was purified from a crude preparation of holo-NCS supplied by Kayaku Co., Ltd., of Tokyo by a method described elsewhere (Samy et al., 1977). The chromophore was extracted by the procedure of Napier et al. (1979). (2) Purified lyophilized holo-NCS was a gift from Kayaku Co., Ltd. The chromophore was extracted with glacial acetic acid (Samy, unpublished experiments).

Several solution conditions were utilized over the course of our investigations. All apo-NCS solutions were brought up in 50 mM (monobasic) phosphate buffer and 10 mM EDTA. Solutions from NCS preparation 1 had a pH of 5.8 and concentrations from 1 to 5 mM. Spectra were recorded at temperatures at 15, 34, 40, and 52 °C. Solutions from NCS preparation 2 had a pH of 6.3 and concentrations of 5–8 mM. Spectra were recorded at 40 and 20 °C.

Solutions of apo-NCS were prepared in four different ways to produce protein in various states of proton–deuterium exchange (Redfield & Dobson, 1988). (1) Lyophilized protein was brought up in 90% H_2O/10% D_2O, producing a "nonexchanged" protein. Spectra were recorded at the temperatures given above. (2) Nonexchanged protein was lyophilized and then brought up in 99.98% D_2O to produce "partially exchanged" protein. Spectra were recorded at 15 °C. (3) Partially exchanged protein in 99.98% D_2O was heated to 50 °C for 1 h, giving the "fully exchanged" protein. Spectra were recorded at 40 and 37 °C. (4) Fully exchanged protein was lyophilized and then brought up in 90% H_2O/10% D_2O to give the "reverse-exchanged" protein. Spectra were recorded at 15 °C.

All NMR experiments were performed on either a home-built 500-MHz NMR spectrometer (Gladden and Drobny, unpublished results) or a Bruker AM-500 spectrometer. Phase-sensitive COSY (Aue et al., 1976), DQF- and TQF-COSY (Piantini et al., 1982; Shaka & Freeman, 1983), RELAY (Eich et al., 1982; Bax & Drobny, 1985), TOCSY (Braunschweiler & Ernst, 1983; Bax & Davis, 1985), NOESY (Jeener et al., 1979; Anil Kumar et al., 1980), and NOESY-TOCSY (Kessler et al., 1988; Basus & Scheek, 1988) experiments were acquired in hypercomplex (States et al., 1982) or TPPI (Drobny et al., 1979; Bodenhausen et al., 1980) mode with standard phase-cycling schemes. The water resonance was presaturated by selective irradiation for between 0.5 and 1.5 s, and the total relaxation delay was 2 s.

RELAY spectra were performed with mixing times of 36 ms (90% H_2O) and 26 and 30 ms (99.98% D_2O).

TOCSY spectra were performed in both H_2O and D_2O, a variety of mixing times (ranging between 42 and 87 ms) being utilized to optimize the coherence transfer between spins of interest (Remerowski et al., 1989). The MLEV-16 (Levitt et al., 1982) isotropic mixing pulse sequence was used with a radio-frequency field strength of typically 20 kHz.

NOESY spectra were recorded for several mixing times ranging from 50 to 150 ms, randomly varied by 10% (Macura et al., 1981).

The data were processed with the FTNMR software of Dr. Dennis Hare (Hare Research, Woodinville, WA). The spectra were apodized in both frequency dimensions by skewed sine-bell functions, shifted in the range of 10°–40° and having a skew factor typically of 0.8. The first t_1 experiment was multiplied by 0.5 to suppress t_1 ridges in the spectra (Otting et al., 1985).

Results and Discussion

Spin Resonance Assignments. Neocarzinostatin is a relatively large protein to assign by homonuclear 2D NMR. Because of its spectral complexity, we used the two-step sequential assignment method described by Wüthrich (1976), in which the *J*-correlated networks of cross-peaks are first assigned according to amino acid spin type. Several spin types can be uniquely assigned from the distinctive cross-peak patterns formed by their scalar-coupled spins, while others can only be assigned to more general classes by through-bond connectivities (Wüthrich, 1976). Sequential assignment methods utilize primarily the NH–C^{α}H, NH–C^{β}H, and C^{α}H–C^{β}H cross-peaks; therefore, some investigators (Englander & Wand, 1987) have sequentially assigned proteins without first doing spin-type assignments, an approach that was not feasible in this case.

A total of 58% of NCS is composed of uniquely identifiable amino acid spin systems, alanine (18), threonine (13), glycine (15), valine (12), leucine (6), lysine (1), and isoleucine (1). Approximately 85% of those proton resonances were identified before the sequential assignment began.

Of the uniquely identifiable spin systems, alanine (16/18), threonine (10/13), and glycine (14/15) were completely identified before sequential assignment by use of data from RELAY, COSY, and TQF-COSY experiments.

Distinguishing valine, leucine, and isoleucine spin systems can be difficult because of their structural similarities. Information from TOCSY experiments, in conjunction with COSY and RELAY data, can facilitate the conclusive identification of these residues. Weber et al. (1987) demonstrated a leucine resonance assignment method that relied on TOCSY data to assign the six leucine residue of apo-NCS. The utility of TOCSY data depends in part on the appearance of α-proton to δ-methyl proton cross-peaks in the spectrum. In the TOCSY spectrum alone, both C^{α}H–$C^{\gamma}H_3$ cross-peaks from valine and C^{α}H–$C^{\delta}H_3$ cross-peaks from leucine can appear. One distinguishing characteristic is that the valine C^{α}H–$C^{\gamma}H_3$ cross-peaks can also appear in RELAY spectra. In our D_2O RELAY spectra we found a cross-peak pair (at ω_1 = 4.25 ppm, ω_2 = 0.12 and −0.49 ppm) which had been identified by Weber et al. as leucine C^{α}H–$C^{\delta}H_3$ cross-peaks. Nine valines and five leucines were partially or wholly assigned before the sequential assignment was begun.

We made tentative assignments to the categories of serine, AMQX (aromatics, Asx, Cys), Glx (glutamine and glutamic acid), proline, or arginine using COSY, RELAY, and TOCSY data. AMQX-type spin systems were differentiated from serine on the basis of coupling pattern and chemical shift (Gross & Kalbitzer, 1988). We identified 10 of 11 serines before beginning the sequential assignment. TOCSY data were instrumental in the initial tentative identification separating AMQX-type and long side chain spin systems. TOCSY experiments were also extremely helpful in tracing entire spin systems at a later stage to confirm assignments made on the basis of sequential connectivities, as shown by the traced networks of long side chain spin systems in Figure 1.

Sequential Assignment Strategy. The next step in the sequential assignment process is the identification of COSY cross-peaks that have been assigned by spin type to a specific amino acid residue in the protein. This is accomplished by finding sequential $d_{\alpha N}$, d_{NN}, and $d_{\beta N}$ connectivities in the NOESY spectrum correlating protons on adjacent residues in the protein. Sequential connectivities are defined as NOE cross-peaks correlating protons on adjacent amino acid residues on the polypeptide or between protons on amino acids separated by as many as three residues in the primary sequence (Billeter et al., 1982; Wüthrich, 1986).

In order to sequentially assign a protein of this size, one must employ techniques that will resolve the many ambiguities

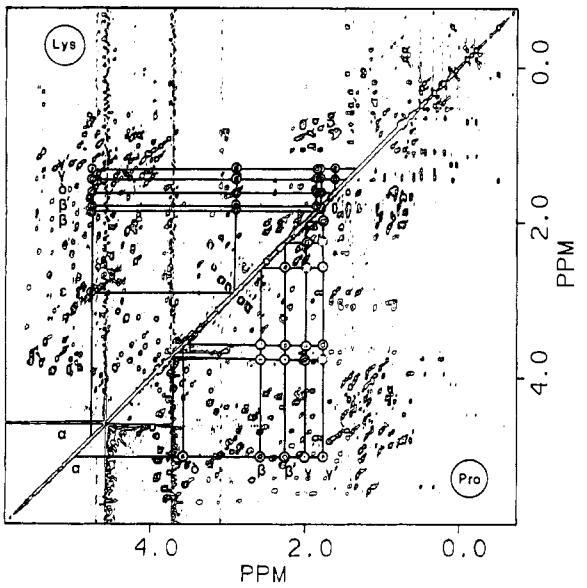

FIGURE 1: TOCSY spectrum of apo-NCS in 99.98% D_2O, with a mixing time of 70 ms, showing examples of long side chain spin systems fully assigned from the cross-peaks resulting from long-range coherence transfer. Residues P9 (below the diagonal) and K20 (above the diagonal) are traced out.

which arise from chemical shift redundancies among the cross-peaks. We used the following four techniques to aid the sequential assignment: (1) temperature variation, (2) NOE-TOCSY experiment (Kessler et al., 1988; Basus & Scheek, 1988), (3) solvent exchange (Redfield & Dobson, 1988), and (4) main chain directed methods (Englander & Wand, 1987).

Since the chemical shifts of the exchangeable backbone amide protons change with temperature, we collected matched COSY and NOESY data sets at four temperatures (52, 40, 34, and 15 °C) over a range of 37 deg. Relative changes in the fingerprint cross-peak positions over that temperature range eliminated a great number of ambiguities in the assignment of d_{NN} and $d_{\alpha N}$ connectivities. Cross-peaks are also lost at the water resonance frequency in the ω_2 dimension from bleaching and from the baseline perturbation caused by the large water signal. Cross-peaks lost at one temperature would appear at another temperature because of the shift in the water resonance. This was especially important to the success of the sequential assignment of this heavily β-sheeted protein, since most of the sequential connectivities are of the $d_{\alpha N}$ type and therefore appear in that section of the spectrum.

Another way to eliminate ambiguities is through use of a relayed NOESY type experiment (Wagner, 1984), where the mixing process consists of both an incoherent and a coherent step of magnetization transfer, resulting in cross-peaks from a two-step magnetization transfer from NH_i and NH_{i-1}. We accomplished the coherent transfer step by isotropic mixing, which yields in-phase NOE-TOCSY cross-peaks, in order to increase the sensitivity of this experiment as independently described elsewhere (Kessler et al., 1988; Basus & Scheek, 1988). Because of its heavily β-sheeted structure, d_{NN} connectivities are sparse in the NCS spectrum; therefore the NH_i–NH_{i-1} NOE-TOCSY cross-peaks appear in a less densely populated portion of the spectrum. Our NOE-TOCSY experiments yielded 13 strong NOE-TOCSY peaks as shown in Figure 2.

In addition to providing information about possible hydrogen bonding, D_2O–H_2O solvent exchange spectra provide spectral simplification in the amide cross-peak regions. We recorded matched COSY and NOESY spectra for the exchanged and reverse-exchanged cases (cf. Materials and Methods) of amide

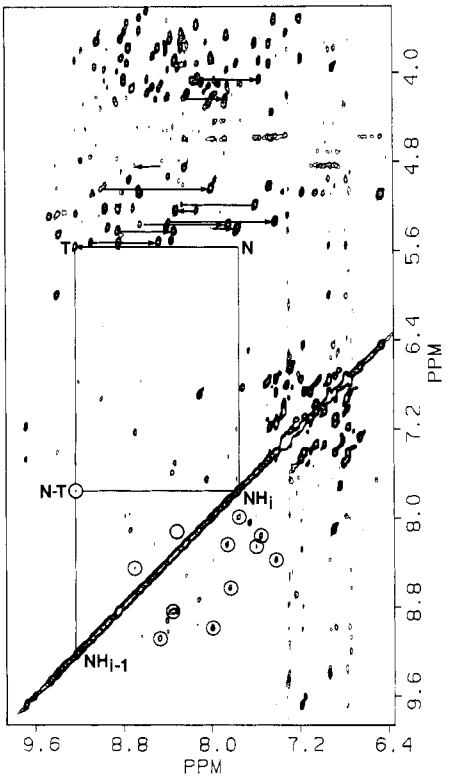

FIGURE 2: NOESY-TOCSY spectrum of a 3 mM solution of apo-NCS in 90% H_2O/10% D_2O at 40 °C. The water resonance was presaturated by selective irradiation for 1 s prior to the first 90° pulse and during the NOE mixing time of 150 ms. The isotropic mixing time was 45 ms, corresponding to 58 MLEV-16 cycles with an rf field strength of 20.8 kHz.

proton exchange. The exchanged COSY spectrum recorded in D_2O showed 35 residual $C^{\alpha}H$–NH cross-peaks, most with persistent sequential $d_{\alpha N}$ connectivities appearing in the companion NOESY spectrum. The residues with strong cross-peaks in the exchanged spectrum are defined as having slowly exchanging amides. Residues with cross-peaks that did not appear (or appeared weakly) in the reverse-exchanged spectrum and also did not appear (or appeared weakly) in the exchanged spectrum are defined as having an intermediate rate of amide exchange. This distinction is made in Figure 6, but not in Figure 3.

Main chain directed patterns of NOEs showing interstrand connectivities between strands of antiparallel β-sheet (Englander & Wand, 1987) were used to both extend and confirm sequential assignments, and also as evidence of extended β-sheet secondary structure. Figure 3 illustrates all the interstrand connectivities found with this method.

Sequential Assignment. All fingerprint cross-peaks with the exceptions of A1, A2, and D15 appeared in one or more of our spectra, at one or more of the temperatures used in recording the spectra, thereby facilitating the sequential assignment. Figure 4 shows the fingerprint section of a COSY spectrum of NCS recorded at 40 °C.

Despite the absence of their fingerprint cross-peaks, we were able to obtain sequential connectivities and resonance assignments for D15 and tentatively for A2. Tentative assignments of the α-protons and the β-methyl protons of A2 were made, on the basis of α–δ proton NOEs connecting the δ-protons of P3 to a COSY cross-peak having typical intensity and chemical shift of an alanine $C^{\alpha}H$–$C^{\beta}H_3$ cross-peaks. The α- and β-protons of D15 were discovered from a main chain directed connectivity pattern. These resonance assignments and all others for NCS are available as supplementary material

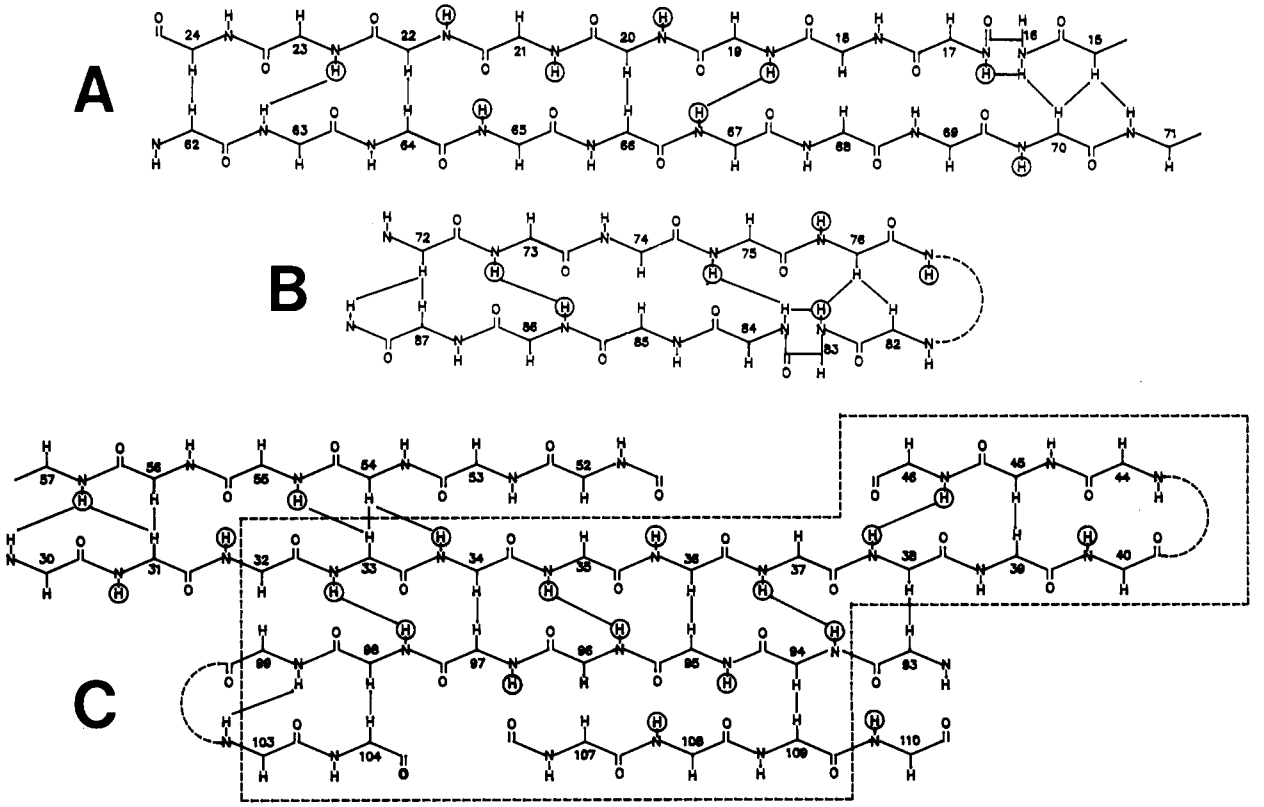

FIGURE 3: The three antiparallel β-sheet domains of apo-NCS seen by 2D NMR. Interstrand d_{NN}, $d_{\alpha\alpha}$, and $d_{\alpha N}$ NOEs are indicated by solid lines. Amide protons with a slow or intermediate hydrogen exchange rate (see text) are circled. (A) Two strands of a three-strand β-sheet face of the seven-strand β-sandwich seen by X-ray crystallography. (B) One loop of the lobe portion of apo-NCS. (C) The four-strand face of the β-sandwich forming one side of the cleft and the residues 38–46, the other loop of the lobe portion of apo-NCS, which forms the other side of the cleft. The framed residues are considered possible candidates for association with the chromophore.

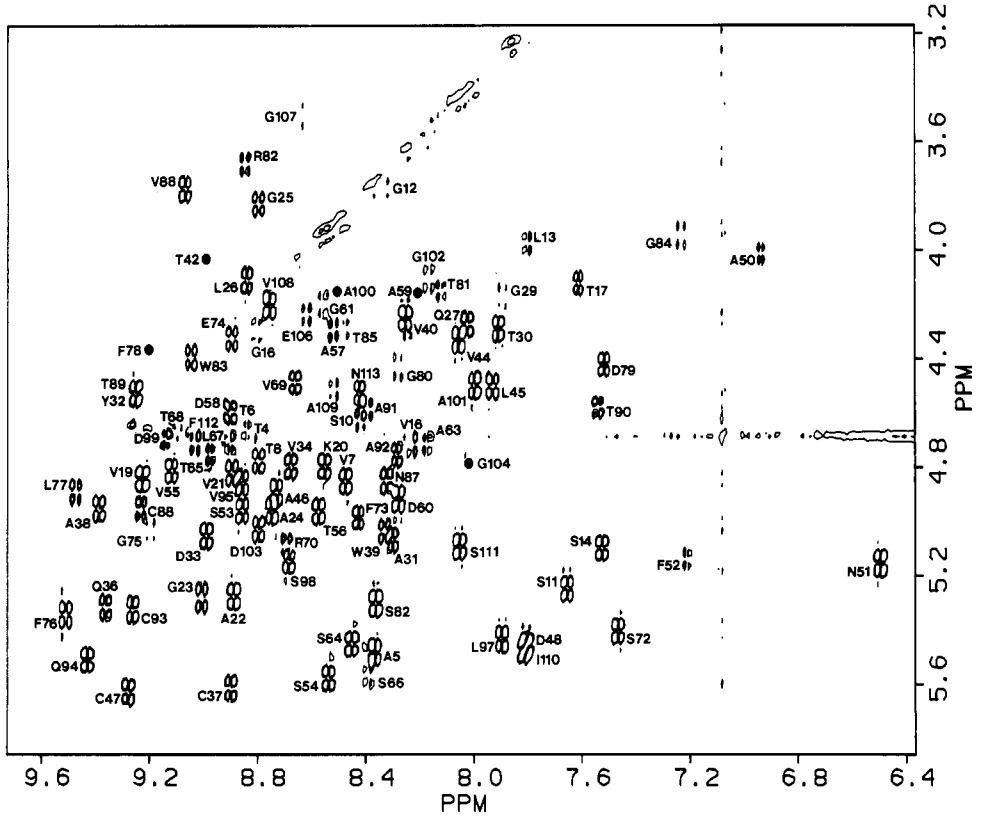

FIGURE 4: COSY fingerprint region of apo-NCS recorded at 40 °C. Cross-peaks represented by circles appear at lower contour levels than shown in this figure. The R71 fingerprint cross-peak is not seen here since the C$^{\alpha}$H of R71 is upfield of the range shown in this figure. The fingerprint cross-peaks of G96, G35, G43, A28, and N41 do not appear in this spectrum, but all have appeared in other COSY and/or RELAY spectra. The NH–C$^{\alpha}$H cross-peaks for A1, A2, and D15 appeared in no spectrum. See text for details.

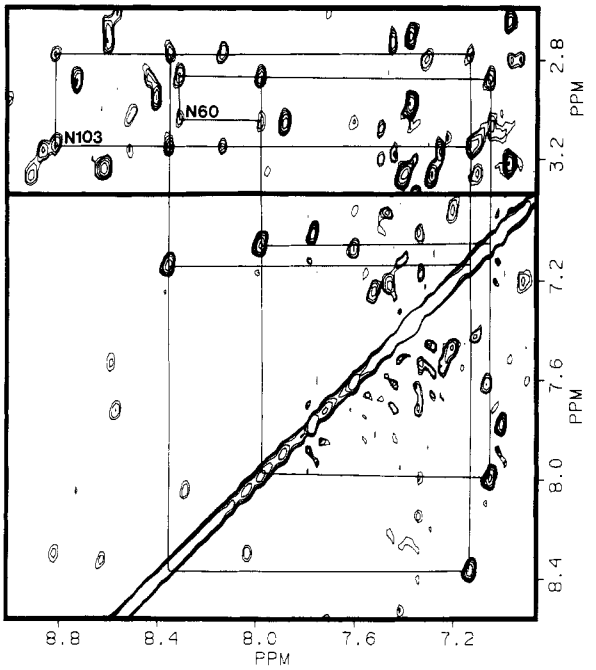

FIGURE 5: NOESY diagram showing the connections of the Asn(60) amide side-chain resonances to its $C^{\beta}H_2$ protons. Similar connectivities are seen for residue 103, tentatively identifying it as an Asn, not Asp as given in the primary sequence.

(see paragraph at end of paper regarding supplementary material).

One result of the sequential assignment was the discovery of a possible error in the primary sequence. The primary sequence of NCS has been revised twice since its first publication in 1972 by Meienhofer and co-workers. The most recent revision (Hirayama et al., 1986) changed Asp(60) to Asn(60), a finding which is corroborated by our data showing NOE connectivities between an amide proton pair and the β-protons of residue 60 (Figure 5). Figure 5 shows similar amide to β-proton NOEs for residue 103, which is still identified as an aspartate residue in the primary sequence. Tertiary structural information from X-ray analysis and secondary structural evidence from NMR show that residue 103 is not in the vicinity of any residues identified as having amide-containing side chains. These data suggest the possibility that Asp(103) could be Asn(103).

The determination of polypeptide secondary structure by NMR is based on characteristic types of NOE connectivities correlating the amide protons, α-protons, and β-protons of the polypeptide (Wüthrich et al., 1984; Wüthrich, 1986). The sequential assignment of apo-NCS, summarized in Figure 6, shows several stretches in the primary sequence connected by sequential $d_{\alpha N}$ connectivities indicative of an extensively β-sheeted secondary structure. The $d_{\alpha N}$ connectivities are broken by single or small stretches of d_{NN} and $d_{\beta N}$ NOEs indicating light turns or bends. These turn regions do not manifest the medium-range (protons within one or two residues of each other) $d_{\alpha N}$ and $d_{\beta N}$ connectivities which are characteristic of helices. Only the region from P9 to D15 deviates from regular secondary structural features, perhaps indicating a small "bubble" or wide looping turn.

There are only three breaks in the sequential assignment, giving four unbroken stretches of connectivity ranging in length from 13 to 64 residues. The first break in sequential connectivities occurs between residues S14 and D15. Although no amide resonance was assigned for D15, there is no NOE with either the amide proton (in ω_1) or α-proton (in ω_2) chemical shift of S14 to suggest a $d_{\alpha N}$ or d_{NN} connection to

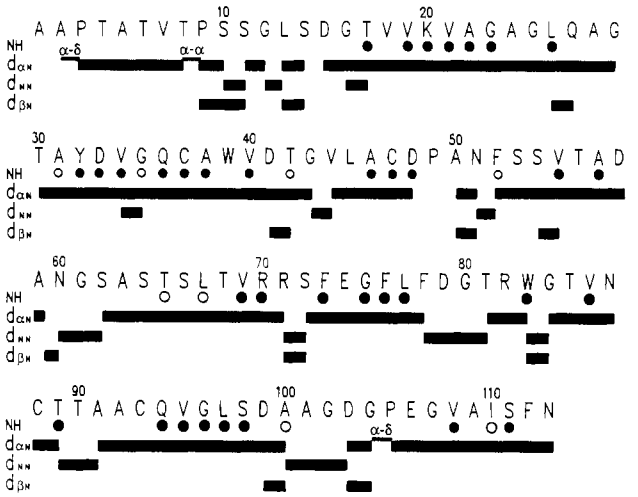

FIGURE 6: Amino acid sequence of apo-NCS with a summary of amide proton exchange rates and NOE connectivities found by sequential assignment procedures. Filled circles beneath sequence elements indicate slowly exchanging amide protons for those residues; open circles indicate an intermediate exchange rate (see text). The thin bars labeled α–α and α–δ refer to $d_{\alpha\alpha}$ and $d_{\alpha\delta}$ connectivities, respectively, which sequentially connect Pro residues to their nearest (N-terminal side) neighbors.

any amide proton. There is another break between residues G29 and T30. Their amide chemical shifts are redundant except at 15 °C. At that temperature, a $d_{\alpha N}$ connectivity could be visible, but none appears in that spectrum. A d_{NN} connectivity would still be lost in the diagonal. The last break occurs at P49, for which we have not yet confidently assigned either resonances or sequential connectivities.

NCS Has Three Structural Domains As Seen by NMR. Holo-NCS and the homologous proteins macromomycin (MCR) and actinoxanthin (AXN) have all been studied by X-ray crystallography. NCS and MCR have refined crystal structures of 2.8 Å (Sieker and Samy, unpublished experiments) and 1.6 Å (Van Roey & Beerman, 1989), respectively, and the AXN crystal structure has been partially refined to 2.5 Å (Pletnev et al., 1982). The tertiary structures of all three proteins are essentially the same. That structure is shown in the schematic "ribbon" representation of apo-NCS in Figure 7 (Sieker and Ramanadham, unpublished experiments). The major portion of these proteins is composed of a well-defined seven-stranded antiparallel β-sheet structure that forms a β-sandwich where one "face" is a three-strand β-sheet and the other is a four-strand β-sheet. The rest of the protein is a lobe composed of two loops, one consisting of residues 37–47 and the other consisting of residues 72–87. The two loops partially overlap at the base of the U-shaped cleft where the chromophore resides (Sieker and Samy, unpublished experiments). The cleft is formed by the lobe on one side and the four-strand "face" of the β-sandwich on the other as seen in Figure 7.

On the basis of our data, NCS can be divided into the three antiparallel β-sheeted structural domains depicted in Figure 3. These structural domains and the additional information from sequential assignment in Figure 6 correspond to the three-dimensional crystallographic interpretation.

Three-Strand Antiparallel β-Sheet. In the crystal structure of NCS there is a triple-stranded antiparallel β-sheet as seen in Figure 7. Two of the strands are the same as those seen in Figure 3A. The third strand is composed of residues T4–T8, which run antiparallel to residues K20–A24. We found no interstrand NOEs between those two strands, but the sequential connectivities of the T4–T8 stretch indicate a β-sheet secondary structure. Also, the amide protons of residues 20

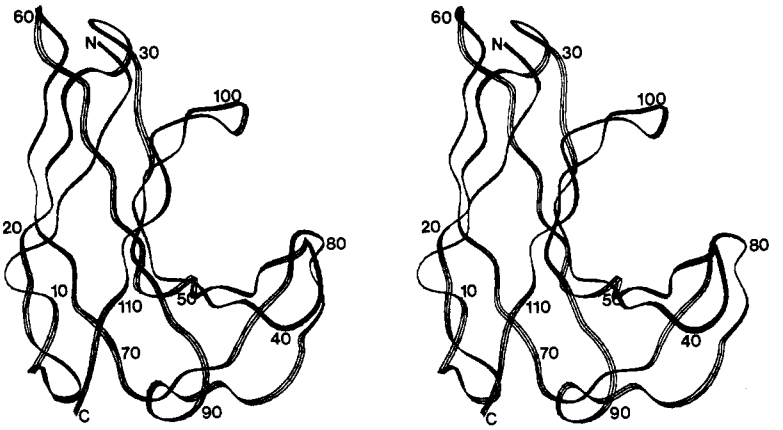

FIGURE 7: Stereo ribbon picture of the holo-NCS backbone by X-ray crystallography (Sieker and Ramanadham, unpublished experiments). N and C indicate the amino and carboxyl termini, respectively.

and 22, which would be oriented toward the T4–T8 strand, are persistent in D_2O, indicating the possibility of H-bonding to atoms of β-strand T4–T8.

The β-sheet structure of T4–T8 is then interrupted by P9, which has a cis conformation on the basis of both the evidence of the crystal structure (Sieker and Samy, unpublished experiments) and the $d_{\alpha\alpha}$ type of sequential connectivity between T8 and P9 (Wüthrich, 1986). X-ray analysis shows that region S10–S14 forms a wide loop to bring the backbone around so that section K20–A24 runs antiparallel to section T4–T8 as mentioned earlier. That looping structure appears as the irregular pattern of $d_{\alpha N}$, $d_{\beta N}$, and $d_{\alpha\alpha}$ connectivities seen in the sequential assignment of those residues. The irregularity is illustrated in Figure 3A by the pattern of connectivities of G16 to T17 and R70, at the end of the loop.

The middle strand of the three-strand β-sheet seen by X-ray, approximately T17–A24, forms an antiparallel β-sheet with strand three, residues S62–V69. These two strands are depicted in Figure 3A. The secondary β-sheet character of the protein continues beyond A24; residues G25–V40 extend in a virtually unbroken stretch of β-sheet secondary structure as defined by the $d_{\alpha N}$ type of sequential connectivity. In the crystal structure the region from G25 to T30 is seen to be going through a gradual chain reversal; the turning is not tight enough to cause many breaks in the β-sheet secondary structure, but there is one break in $d_{\alpha N}$ connectivities between G29 and T30. Residues G25 and Q27–A31 also exhibit intraresidue $d_{\alpha N}$ NOEs, which can indicate a gentle twist or turn of the β-sheet that brings the amide protons and α-protons of the same residue close enough to manifest an NOE cross-peak (Klevit et al., 1986).

Four-Strand Antiparallel β-Sheet. Most of the residues that make up the four-stranded β-sheet, illustrated in Figure 3C, form the cleft of the molecule where the chromophore resides. The tertiary backbone structure schematically depicted in Figure 7 shows the U-shaped conformation of the cleft region. Residues that are considered possible candidates for association with the chromophore are framed in Figure 3C. This NMR-defined domain contains the four-strand β-sheet which forms one side of the U-shaped chromophore binding cleft, as well as one of the loops, residues 37–47, involved in formation of the other side of the cleft. The second loop of the lobe portion of NCS is detailed in Figure 3B.

Regions C47–N51 and G104–E106 do not appear in the four-strand β-sheet section depicted in Figure 3C because we found no interstrand NOEs to place them relative to the other residues. Both sequences contain proline residues, which often result in secondary structural kinks because of the constraints they place on backbone flexibility. Region G104–E106 manifests β-sheet $d_{\alpha N}$ connectivities, except for the G104–P105 $d_{\alpha\delta}$ connectivity indicating a trans conformation for that proline, so the structure in that stretch is essentially β-sheet. The region from C47 to N51 is a "kink" in the crystal structure; it is clear from the schematic representation in Figure 3 that five residues could not form a regular antiparallel β-sheet in a space that is opposite one occupied by only two residues on the adjacent strand. Our sequential assignments do not define the proline conformation, but d_{NN} and $d_{\beta N}$ NOEs indicate the presence of a turn or bend in residues A50–F52. The crystal structure shows a very sharp kink at that proline, as shown in Figure 7 near residue 50.

Loop C37–C47. This loop was of interest because initial X-ray studies of holo-NCS at 3.5-Å resolution (Sieker, 1981) found poor density in that region of the protein. One possible condition that could account for those data is that the protein is flexible in that portion of the molecule to accommodate the chromophore; therefore, crystals form with a variety of conformational states, producing disorder in the electron density map for the loop (Sieker and Samy, unpublished experiments).

Our 2D NMR studies have shown the region to be very well-defined on the NMR time scale. All of the side-chain as well as the main-chain resonances of those residues have been assigned. Additionally, the secondary structural and interstrand connectivities indicate a well-defined contiguous antiparallel β-sheet formed by the residues between C37 and C47 (which form a disulfide bridge). The presence of a tight turn is manifested by a single d_{NN} NOE connecting G43 and V44.

Our results were corroborated by the eventual resolution of this area as the X-ray interpretation progressed; the initial problem of a lack of interpretable density in that region was probably not due to the presence of multiple conformations of that loop in the crystal (Sieker and Ramanadham, unpublished experiments).

The aromatic side group of Trp(39) extends toward the cleft, as evidenced by a strong NOE connecting H7 of the Trp(39) indole ring to the backbone amide of Val(44), suggesting that this side group could be important in the chromophore binding.

Loop S72–N87. This loop domain defined by NMR is detailed in Figure 3B; it forms part of the lobe portion of the molecule. NOE connectivities shown in Figures 3B and 6 indicate the presence of a contiguous antiparallel β-sheet with chain reversal occurring in the region F78–T81. There is one irregularity occurring at W83; a β-bulge (Richardson et al., 1978) places the bulky aromatic moiety of the tryptophan away from the lobe. Biochemical studies (Samy et al., 1974) indicate

that this moiety is solvent exposed and not critical for biological activity of the chromophore. Side chains of residues L77 and F78 in this loop extend into the cleft and may be involved with the chromophore (Sieker and Samy, unpublished experiments). There is evidence of contacts between the two loops of the lobe from both NMR and crystallography, which may be expanded as the respective investigations progress. By NMR a d_{NN} NOE connects Cys(47) to Phe(76), and X-ray evidence shows association of Phe(73) to the disulfide bridge between Cys(37) and Cys(47).

Conclusions. The sequential assignment of apo-NCS has progressed to the point where 85% of the proton resonances have been identified; they are available as supplementary material. Only residues A1 and P49 have no resonance assignments or sequential connectivities. The secondary structure of NCS is virtually all β-sheet with several turn regions, but no helical structure. Three antiparallel β-sheeted structural domains are seen by NMR, and our secondary structure in solution is consistent with the X-ray crystal structure of holo-NCS at 2.8-Å resolution. The loop consisting of residues C37–C47, which is probably involved in chromophore binding, has a well-defined antiparallel β-sheet secondary structure as characterized by the sequential $d_{\alpha N}$ connectivities, interstrand NOE connectivities, and slowly exchanging amide protons. There is a tight turn from residues D41 to V44 indicated by a d_{NN} connecting G43 to V44.

Many questions concerning the structure of the holo-NCS and the interactions between the protein and its chromophore remain. How NCS-Chr is stabilized by the protein is unknown. Even when bound to the protein, the chromophore is sensitive to heat, oxygen, and visible and ultraviolet radiation; the extent to which the Chr is stabilized by the polypeptide under various conditions is unexplored. The possibility remains that the reason for the ill-defined electron density in the region of the cleft of holo-NCS is that the labile chromophore has degraded as the result of either the crystallization procedure or the data collection procedure, which would prevent detailed study of the protein–chromophore interaction by X-ray crystallography. NMR studies of the holoprotein could yield specific interactions of protein residues involved in chromophore binding and information on conditions required for the stability of the bound chromophore in solution.

Our sequential assignment of apo-NCS and assignment of the majority of the proton resonances are the foundation for our continued study of this protein–chromophore system. Future plans include determination of the solution structures of apo- and holo-NCS and a comparative study of bound chromophore to that of the free chromophore determined by Edo et al. (1985) and Myers et al. (1988).

Acknowledgments

We thank Dr. M. Ramanadham and Prof. Rachel Klevit for helpful discussions, Dr. Paul Weber for some early data collection, and Kayaku Co., Ltd., of Tokyo for providing us with holo-NCS.

Supplementary Material Available

Table I listing the aponeocarzinostatin ^{1}H NMR resonance assignments at 40 °C and pH 5.3 (10 pages). Ordering information is given on any current masthead page.

Registry No. Apo-NCS, 101359-79-9.

References

Anil Kumar, Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–6.

Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229–2246.

Basus, V. J., & Scheek, R. M. (1988) *Biochemistry 27*, 2772–2775.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.

Bax, A., & Drobny, G. P. (1985) *J. Magn. Reson. 61*, 306–320.

Beerman, T. A., & Goldberg, I. H. (1974) *Biochem. Biophys. Res. Commun. 59*, 1254–1261.

Beerman, T. A., & Goldberg, I. H. (1977) *Biochim. Biophys. Acta 475*, 281–293.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535–542.

Biemann, K. (1982) *Int. J. Mass Spectrom. Ion Phys. 45*, 183–194.

Billeter, M., Braun, W., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 321–346.

Bodenhausen, G., Vold, R. L., & Vold, R. R. (1980) *J. Magn. Reson. 37*, 93–106.

Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.

Chimura, H. M., Ishizuka, M., Hamada, M., Hori, S., Kimura, K., Iwanaga, J., Takeuchi, T., & Umezawa, H. (1968) *J. Antibiot. 21*, 44–49.

Dasgupta, D., & Goldberg, I. H. (1985) *Biochemistry 24*, 6913–6920.

Drobny, G. P., Pines, A., Sinton, S., Weitekamp, D. P., & Wemmer, D. E. (1979) *Symp. Faraday Soc. 13*, 49–55.

Edo, K., Mizugaki, M., Koide, Y., Seto, H., Furihata, K., Otake, N., & Ishida, N. (1985) *Tetrahedron Lett. 26*, 331–334.

Eich, G., Bodenhausen, G., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 3732–3733.

Englander, W. S., & Wand, J. A. (1987) *Biochemistry 26*, 5953–5958.

Gibson, B. W., Herlihy, W. C., Samy, T. S. A., Hahm, K.-S., Maeda, H., Meienhofer, J., & Biemann, K. (1984) *J. Biol. Chem. 259*, 10801–10806.

Goldberg, I. H., Hatayama, T., Kappen, L. S., Napier, M. A., & Povirk, L. F. (1981) in *Molecular Actions and Targets for Cancer Chemotherapeutic Agents*, pp 163–191, Academic Press, New York.

Gross, K.-H., & Kalbitzer, H. R. (1988) *J. Magn. Reson. 76*, 87–99.

Hirayama, K., Ando, T., Takahashi, R., & Murai, A. (1986) *Bull. Chem. Soc. Jpn. 59*, 1371–1378.

Ishida, N., Miyazaki, K., Kumagi, K., & Rikimaru, M. (1965) *J. Antibiot. A18*, 68–76.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.

Kappen, L. S., & Goldberg, I. H. (1980) *Biochemistry 19*, 4786–4795.

Kappen, L. S., Goldberg, I. H., & Samy, T. S. A. (1979) *Biochemistry 18*, 5123–5127.

Kappen, L. S., Napier, M. A., & Goldberg, I. H. (1980a) *Proc. Natl. Acad. Sci. U.S.A. 77*, 1970–1974.

Kappen, L. S., Napier, M. A., Goldberg, I. H., & Samy, T. S. A. (1980b) *Biochemistry 19*, 4780–4785.

Kessler, H., Gemmecker, G., & Steuernagel, S. (1988) *Angew. Chem., Int. Ed. Engl. 27*, 564–566.

Kholov, A. S., Reshetov, P. D., Chupova, L. A., Cherches, B. Z., Zhigis, L. H., & Stoyachenko, I. A. (1976) *J. Antibiot. 29*, 1026–1034.

Klevit, R. E., Drobny, G. P., & Waygood, E. B. (1986) *Biochemistry 25*, 7760–7769.

Kumada, Y., Miwa, T., Naoi, N., Watenabe, K., Naganawa, H., Takita, T., Umezawa, H., Nakamura, H., & Iitaka, Y. (1983) *J. Antibiot. 36*, 200–202.
Lazarus, H., Raso, V., & Samy, T. S. A. (1977) *Cancer Res. 37*, 3736.
Levitt, M., Freeman, R., & Frenkiel, T. (1982) *J. Magn. Reson. 47*, 328–330.
Macura, S., Huang, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259–281.
Maeda, H., Glaser, C. B., Kuromizu, K., & Meienhofer, J. (1974) *Arch. Biochem. Biophys. 164*, 379–385.
Maeda, H., Aikawa, S., & Yamashita, A. (1975) *Cancer Res. 35*, 554–559.
Meienhofer, J., Maeda, H., Glaser, C. B., Czomboz, J., & Kuromizu, K. (1972) *Science 178*, 875–876.
Montgomery, R., Shepherd, V. L., & Vandrẽ, D. D. (1981) in *Antitumor Compounds of Natural Origin: Chemistry and Biochemistry, Vol. 1* (Aszalos, A., Ed.) pp 79–122, CRC Press, Inc., Boca Raton, FL.
Myers, A. G., Proteau, P. J., & Handel, T. M. (1988) *J. Am. Chem. Soc. 110*, 7212–7214.
Napier, M. A., Holmquist, B., Strydom, D. J., & Goldberg, I. H. (1979) *Biochem. Biophys. Res. Commun. 89*, 635–642.
Napier, M. A., Kappen, L. S., & Goldberg, I. H. (1980) *Biochemistry 19*, 1767–1773.
Otting, G., Widmer, H., Wagner, G., & Wüthrich, K. (1985) *J. Magn. Reson. 66*, 187–193.
Piantini, U., Sørensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800–6801.
Pletnev, V. Z., Kuzin, A. P., Trakhanov, S. D., & Kostetsky, P. V. (1982) *Biopolymers 21*, 287–300.
Povirk, L. F., Dattagupta, N., Warf, B. C., & Goldberg, I. H. (1981) *Biochemistry 20*, 4007–4017.
Redfield, C., & Dobson, C. M. (1988) *Biochemistry 27*, 122–136.
Remerowski, M. L., Glaser, S. J., & Drobny, G. P. (1989) *Mol. Phys. 68*, 1191–1218.
Richardson, J. S., Getzoff, E. D., & Richardson, D. C. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 2574–2578.
Sakamoto, S., Maeda, H., & Ogata, J. (1979) *Experiments 35* (9), 1233–1235.
Samy, T. S. A., Hu, J.-M., Meienhofer, J., Lazarus, H., & Johnson, R. K. (1977) *J. Natl. Cancer Inst. 58*, 1765–1768.
Samy, T. S. A., Hahn, K. S., Modest, E. J., Lampman, G. W., Keutmann, H. T., Umezawa, H., Herlihy, W. C., Gibson, B. W., Carr, S. A., & Biemann, K. (1983) *J. Biol. Chem. 258*, 183–191.
Shaka, A. J., & Freeman, R. (1983) *J. Magn. Reson. 51*, 169–173.
Sieker, L. C. (1981) X-Ray Diffraction Analysis of the Antitumor Protein Neocarzinostatin: Results at 3.5 Å Resolution, Ph.D. Thesis, University of Washington, Seattle, WA.
Sieker, L. C. (1987) International Meeting on Biochemistry, August, Beijing, China, Abstract B-05-0083.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
Suzuki, H., Miura, K., Kumada, Y., Takeuchi, T., & Tanaka, N. (1980) *Biochem. Biophys. Res. Commun. 94*, 255–261.
Takeshita, M., Kappen, L. S., Grollman, A. P., Eisenberg, M., & Goldberg, I. H. (1981) *Biochemistry 20*, 7599–7606.
Van Roey, P., & Beerman, T. A. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 6587–6591.
Wagner, G. (1984) *J. Magn. Reson. 57*, 497–505.
Weber, P. L., Sieker, L. C., Samy, T. S. A., Reid, B. R., & Drobny, G. P. (1987) *J. Am. Chem. Soc. 109*, 5842–5844.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.