Journal of Biomolecular NMR, 4 (1994) 741–760
ESCOM

J-Bio NMR 212

^{1}H and ^{15}N resonance assignments and secondary structure of cellular retinoic acid-binding protein with and without bound ligand

Josep Rizo, Zhi-Ping Liu and Lila M. Gierasch*

Department of Pharmacology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75235-9041, U.S.A.

Received 2 May 1994
Accepted 21 June 1994

Keywords: Cellular retinoic acid binding protein; Multidimensional NMR; Resonance assignment; Secondary structure; Selective labeling

SUMMARY

Sequence-specific assignments for the ^{1}H and ^{15}N backbone resonances of cellular retinoic acid-binding protein (CRABP), with and without the bound ligand, have been obtained. Most of the side-chain resonances of both apo- and holo-CRABP have also been assigned. The assignments have been obtained using two-dimensional homonuclear and heteronuclear NMR data, and three-dimensional ^{1}H-^{15}N TOCSY-HMQC and NOESY-HMQC experiments. The secondary structure, deduced from nuclear Overhauser effects, amide H/D exchange rates and H^{α} chemical shifts, is analogous in both forms of the protein and is completely consistent with a model of CRABP that had been constructed by homology with the crystal structure of myelin P2 protein [Zhang et al. (1992) *Protein Struct. Funct. Genet.*, **13**, 87–99]. This model comprises two five-stranded β-sheets that form a sandwich or β-clam structure, and a short N-terminal helix–turn–helix motif that closes the binding cavity between the two sheets. Comparison of the data obtained for apo- and holo-CRABP indicates that a region around the C-terminus of the second helix is much more flexible in the apo-protein. Our data provide experimental evidence for the hypothesis that the ligand-binding mechanism of CRABP, and of other homologous proteins that bind hydrophobic ligands in the cytoplasm, involves opening of a portal to allow entry of the ligand into the cavity.

INTRODUCTION

Cellular retinoic acid-binding protein (CRABP) (136 residues, 15.5 kDa) is a member of a family of small intracellular proteins that bind hydrophobic ligands such as fatty acids, lipids and retinoids (for reviews, see Chytil and Ong, 1987; Sweetser et al., 1987; Banaszak et al., 1994).

*To whom correspondence should be addressed at: Department of Chemistry, University of Massachusetts, Amherst, MA 01003-4510, U.S.A.

Supplementary material available from the authors: one table with proton and nitrogen chemical shifts of apo-CRABP.

Retinoic acid (RA), the endogenous ligand of CRABP, is an active metabolite of vitamin A (retinol) that mediates many vitamin A-dependent processes (Napoli et al., 1991). RA has profound effects on growth, morphogenesis and differentiation of both normal cells and different types of tumor cells (Willhite et al., 1989; Hoffman, 1990; Amos and Lotan, 1991; Maden, 1994). RA and other retinoids have been shown to inhibit tumorigenesis in experimental animals (Moon et al., 1983). The action of RA is exerted through nuclear receptors (De Luca, 1991; Lotan and Clifford, 1991), which act as transcriptional regulators. CRABP and other retinoid-binding proteins are believed to solubilize and stabilize their ligands, to transport them to the nucleus and to make them nontoxic to cells (Wolf, 1991). Although the role of CRABP in cell differentiation is still not clear, the differential distribution of the protein in developing chick limb buds has led to the proposal that CRABP and the isoform known as CRABPII may act by regulating the effective concentration of RA for appropriate spatial modulation of gene transcription (Maden et al., 1988; Blomhoff et al., 1990).

The family of intracellular hydrophobic ligand-binding proteins related to CRABP includes 16 additional members, with homologies ranging from 15% to 81% (Banaszak et al., 1994). The crystal structures of eight of these proteins, including those of cellular retinol-binding protein (CRBP) (Cowan et al., 1993) and CRBPII (Winter et al., 1993), have been determined by X-ray diffraction. All the structures determined have analogous β-barrel architectures, consisting of two five-stranded β-sheets and a short N-terminal helix–turn–helix motif. The two β-sheets are disposed face to face, forming a sandwich structure commonly referred to as a β-clam. The conservation of the backbone structure is very remarkable, with rms deviations between the positions of the α-carbons ranging from 0.63 to 2.38 Å for proteins with homologies between 17% and 67% (Banaszak et al., 1994). Ligand binding occurs in a large cavity in the protein interior and has a very small effect on the structure of the proteins. The solution structure of heart fatty acid-binding protein has been solved by NMR spectroscopy and is very similar to the corresponding crystal structure, but some differences were observed in a flexible region where the ligand presumably enters the binding pocket (Lassen et al., 1993). The β-clam fold observed in the family of proteins referred to above may occur more widely in nature, since the sequence of the *N*-methyl-D-aspartate receptor contains a domain homologous to these proteins (Petrou et al., 1993).

The structures of CRABP and CRABPII in crystals or in solution have not yet been determined. Based on the crystal structure of myelin P2 protein (43% homologous to CRABP), we constructed a 3D model of CRABP to rationalize mutagenesis studies on the contribution of specific amino acid residues to the interaction with RA (Zhang et al., 1992). Mutation of Arg^{111} or Arg^{131} or both to glutamine (the residue found in the homologous positions of CRBP and CRBPII) resulted in a substantially lower affinity for RA, but no binding of retinol was observed. In analogous experiments, mutation of Gln^{109} or Gln^{129} to arginine in CRBPII produced a decrease in affinity for retinol, but no increased binding of RA was observed (Cheng et al., 1991). These results show that, although the arginine (glutamine) residues in positions 111 and 131 (109

Abbreviations: CRABP, cellular retinoic acid-binding protein; CRBP, cellular retinol-binding protein; 2D, 3D, two-, three-dimensional; DQF-COSY, double-quantum filtered COSY; FID, free induction decay; HMQC, heteronuclear multiple-quantum coherence; HSQC, heteronuclear single-quantum coherence; NOESY, nuclear Overhauser enhancement spectroscopy; RA, retinoic acid; Rms, root mean square; TOCSY, total correlation spectroscopy; Tris, tris(hydroxymethyl)aminomethane.

and 129) of CRABP (CRBPII) are important for ligand-binding specificity, the specificity depends also on additional structural factors.

In order to elucidate the structural details that determine the binding specificity of CRABP for RA, we have undertaken the determination of the solution structure of CRABP, with and without bound ligand, using multidimensional NMR methods. These studies are also aimed at comparing the dynamic properties of the apo- and holo-forms of CRABP to understand the mechanisms of ligand binding and release, which may depend critically on the flexibility of part of the structure. Understanding the binding interaction between CRABP and its ligand will aid in the design of analogs of RA with altered relative affinities for CRABP and the RA nuclear receptors. Such analogs may help to clarify the function of CRABP and could have clinical applications in the regulation of RA-dependent processes. CRABP is also a very interesting model to study the folding mechanism of single-domain proteins containing mostly β-sheet structure (Liu et al., 1994).

Here we describe the assignment of the ^{1}H and ^{15}N backbone resonances, as well as the secondary structure, of the apo- and holo-forms of CRABP. Assignments were also obtained for most side-chain protons of both forms. The secondary structure is analogous in both forms and corresponds to that observed in the crystal and solution structures of other proteins of the family (see above). A preliminary comparison between the NMR data and the conformational model of CRABP built from the crystal structure of myelin P2 shows that the only observed inconsistencies correspond to residues that were substituted or inserted in the modeling process. The most significant differences between the NMR data of apo- and holo-CRABP are observed in a region near the C-terminus of the second helix, which appears to be more flexible in the ligand-free form of the protein and may act as a hinge to allow ligand entry into the binding cavity.

MATERIALS AND METHODS

Sample preparation

All deuterated or ^{15}N-labeled compounds were purchased from Cambridge Isotope Laboratories. Recombinant CRABP (with or without uniform ^{15}N-labeling) was obtained as described previously, by overexpression in *Escherichia coli* strain BL21(DE3) in LB-CB or in M9 minimal medium containing $^{15}NH_4Cl$, followed by purification from inclusion bodies (Liu et al., 1994). Typical yields of either labeled or unlabeled protein were 15 mg/l cell culture. Selectively labeled samples of CRABP were obtained by growing 1 l of cells in M9 minimal medium with NH_4Cl as the sole nitrogen source, and adding either 140 mg of [^{15}N]leucine, 110 mg of [^{15}N]lysine or 80 mg of [^{15}N]D,L-threonine before induction with isopropyl-β-D-thiogalactopyranoside. The sample with reverse selective labeling of arginine residues was prepared in an analogous manner, but using $^{15}NH_4Cl$ and 150 mg of nonlabeled arginine.

Samples for NMR spectroscopy contained 1.5–2.5 mM protein in 0.6 ml of the appropriate buffer, except for the selectively labeled samples, which contained 0.5–1.0 mM protein. All samples were prepared at 4 °C and the buffers contained 0.02% sodium azide. For experiments in water without RA, the protein was dialyzed against 12 mM Tris-d_{11} at pH 7.5 and concentrated by ultracentrifugation to 0.54 ml, after which 0.06 ml of D_2O were added. Samples with RA were prepared in 40 mM acetic acid-d_4 at pH 3.8, and contained 1.5 mole of RA per mole of protein. To prepare such samples, RA dissolved in dimethylsulfoxide-d_6 or acetone-d_6 was added to a diluted (ca. 30 μM) protein solution in 12 mM Tris-d_{11} at pH 7.5, keeping the final volume of

organic solvent below 1%; the pH was then brought to 3.8 by fast addition of 200 mM acetic acid-d_4, and the protein solution was concentrated. For experiments in D_2O, protein solutions were dialyzed against water, concentrated to ca. 0.5 mM, lyophilized and redissolved in 0.6 ml of either 12 mM Tris-d_{11}/D_2O at pD 7.5 (uncorrected) or 40 mM acetic acid-d_4/D_2O at pD 3.8 (uncorrected); for samples of holo-CRABP, RA was added before concentrating the protein solution. Samples to measure the protection of amide protons against deuterium exchange were prepared freshly before starting acquisition of the data, with the exception that the pD was adjusted to 6.4 (uncorrected) in the experiment without RA.

NMR spectroscopy

All NMR spectra were recorded on a Varian VXR-500S spectrometer, operating at a proton frequency of 500 MHz and a ^{15}N frequency of 50.65 MHz. Data processing was performed on a Silicon Graphics 4D/25 Personal Iris workstation using the program Felix (Biosym Technologies, Inc.). All data were obtained at 25 °C, except for the experiments to measure deuterium exchange rates in apo-CRABP, which were performed at 5 °C. Proton chemical shifts were referenced to the position of the water signal at 4.76 ppm with respect to 3-(trimethylsilyl)propionic acid (at 25 °C). A 200 mM solution of $^{15}NH_4Cl$ in 1 M HCl was used as external reference (24.93 ppm) for nitrogen chemical shifts.

All 2D and 3D experiments were acquired in the phase-sensitive mode, with quadrature detection achieved by the hypercomplex method (States et al., 1982). A 1 s relaxation delay was used in all cases. Two-dimensional DQF-COSY (Rance et al., 1983), TOCSY (Davis and Bax, 1985) and NOESY (Jeener et al., 1979; Kumar et al., 1981) spectra were acquired with 6000 or 6800 Hz spectral widths in both dimensions, using low-power presaturation for suppression of the water resonance. The mixing times ranged from 30 to 70 ms in TOCSY experiments and from 100 to 150 ms in NOESY experiments. A WALTZ17 mixing sequence (Bax, 1989) was used for TOCSY experiments, with delays before and after the 180° pulses to compensate for the development of negative ROE cross peaks (CLEAN-TOCSY; Griesinger et al., 1988); the delays were set to twice the length of a 90° pulse. In general, 2×256 FIDs of 1024 complex points each (64 scans per FID) were acquired. Data sets were apodized using Gaussian or sine-bell (shifted by 10–70°) functions. A zero-order baseline correction was applied after Fourier transformation in the t_2 dimension, and linear prediction of the first point of each FID along the t_1 dimension (Marion and Bax, 1989) was used to obtain flat baseplanes. Zero-filling and Fourier transformation along the t_1 dimension yielded matrices of 1024×1024 real points for TOCSY and NOESY, or 4096×1024 real points for DQF-COSY.

^1H-^{15}N HSQC experiments (Bodenhausen and Ruben, 1980) were recorded with spectral widths of 7600 Hz in the proton dimension and 1360–1800 Hz (apo-CRABP) or 1414–1800 Hz (holo-CRABP) in the nitrogen dimension. A 2 ms spin-lock purge pulse was used to suppress the water resonance (Messerle et al., 1989), which was further removed after data acquisition with a low-frequency filter in the time domain (Marion et al., 1989a). Data sets obtained in water consisted of 2×150 FIDs of 768 complex points each (32 scans per FID), and were processed in a similar manner to that described above to obtain matrices of 512×512 real points after discarding the aliphatic part of the data in the ω_2 dimension. To measure deuterium exchange rates of the amide protons, a series of HSQC spectra was acquired using analogous parameters, except that 2×100 FIDs were obtained. For the first HSQC spectrum after sample preparation, 16 scans per

FID were averaged (total acquisition time 1 h); ensuing HSQC spectra were acquired with 32 scans per FID (total acquisition time 2 h).

Three-dimensional ^{1}H-^{15}N TOCSY-HMQC and NOESY-HMQC data (Marion et al., 1989b,c; Zuiderweg and Fesik, 1989) were acquired for both apo- and holo-CRABP, using pulse sequences adapted for solvent suppression with two spin-lock purge pulses (of 0.5 and 2 ms duration) (Messerle et al., 1989). The spectral widths along the ω_3, ω_2 and ω_1 dimensions were 6300, 1360 and 6300 Hz, respectively, for apo-CRABP, and 6800, 1414 and 6300 Hz, respectively, for holo-CRABP. In all cases, 768 complex t_3, 29 complex t_2 and 110 complex t_1 data points were obtained, with 16 scans per FID (64 scans per hypercomplex t_1,t_2 increment), for a total acquisition time of ca. 3 days. The mixing times used were 40 and 130 ms for the TOCSY-HMQC and NOESY-HMQC spectra, respectively. Isotropic mixing in the TOCSY-HMQC experiments was achieved as described above. A low-frequency filter (Marion et al., 1989a) was applied before data processing. The apodization functions used were a Gaussian for the t_3 domain, a 45°-shifted sine bell added to a 100 Hz line-broadening exponential for the t_2 domain (Fesik and Zuiderweg, 1990), and a 70°-shifted sine bell for the t_1 domain. Zero-order baseline corrections were applied after Fourier transformation along the t_3 and t_2 dimensions, and the first point of each FID along the t_2 and t_1 domains was calculated by linear prediction (Marion and Bax, 1989). Zero-filling and removal of the aliphatic part of the spectrum in the ω_3 dimension yielded matrices consisting of $512 \times 64 \times 512$ real points. We also acquired 2D TOCSY-HMQC and NOESY-HMQC data sets to obtain the positions of intraresidue and sequential H^N/H^α connectivities with higher precision. For these experiments, 2×256 FIDs (64 scans per FID) were obtained, and zero-filling was used to yield matrices of 512×1024 real points.

RESULTS AND DISCUSSION

Experimental conditions for NMR experiments

Acquisition of high-quality NMR data for both apo- and holo-CRABP is hindered by the tendency of both forms to self-associate. Our initial efforts to perform structural characterization by NMR were directed to the apo-protein. The isoelectric point of apo-CRABP is 4.76 (Fiorella and Napoli, 1991), and thus the protein has a high tendency to aggregate in solutions with pH near this value. The protein denatures below pH 3.5 (Liu et al., 1994) and the denaturation temperature is ca. 30 °C at pH 3.8. At pH 3.8–4.0, 0.5 mM protein concentrations can be reached, but aggregation occurs at temperatures above 5 °C, and the NMR data obtained at this temperature and protein concentration had poor sensitivity. Above the isoelectric point, millimolar solutions of apo-CRABP are only stable at 25 °C if the pH is 7.5 or higher. Increasing the ionic strength or raising the temperature favored the aggregation of the protein at all pH values tested. Addition of an organic solvent, such as acetonitrile-d_3 up to 15% (v/v), did not improve the quality of the NMR data. Based on these observations, we chose to perform the NMR analysis of apo-CRABP in 12 mM Tris-d_{11} at pH 7.5 and 25 °C. No aggregation was observed for several weeks under these conditions, although the observed linewidths suggest that a small degree of self-association may still occur. An expansion of the aliphatic region of a TOCSY spectrum obtained in D_2O under analogous conditions is shown in Fig. 1. Although cross peaks were found for most side chains, no connectivities were observed for several residues, all of which correspond to central parts of the molecule according to the structural model described below. In TOCSY

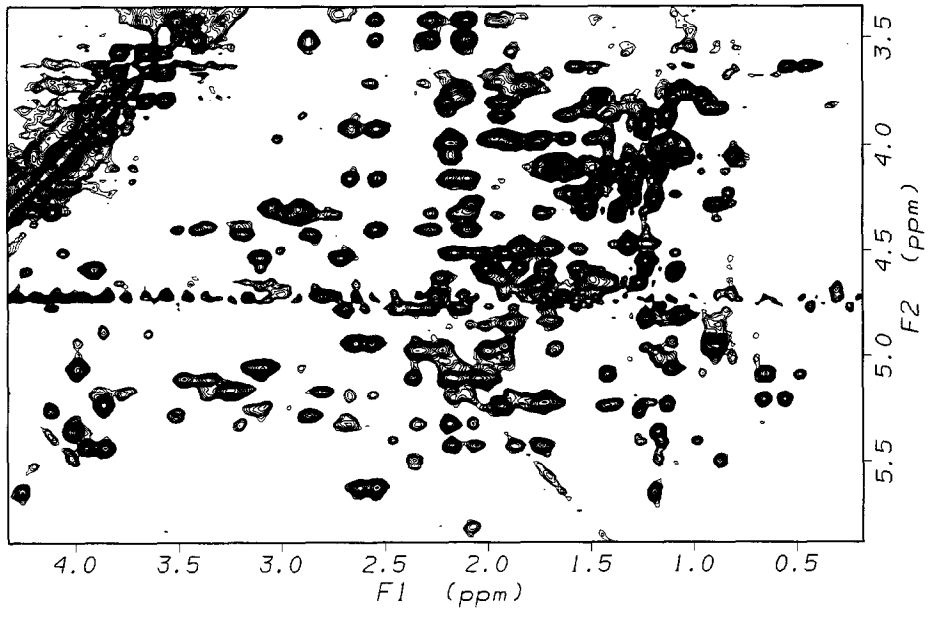


Fig. 1. Expansion of a 2D TOCSY spectrum of apo-CRABP in D_2O (2 mM protein, 12 mM Tris-d_{11}, pD 7.5 (uncorrected), 25 °C, 50 ms mixing time).

spectra obtained in H_2O, relayed connectivities between amide and side-chain protons were only observed for the more mobile parts of the molecule. The pH used (7.5) also caused broadening of some amide proton resonances due to exchange with the solvent, and a few amide signals were completely absent from the spectrum. As described below, these problems hampered the sequential assignment of the NMR spectrum of apo-CRABP, and complete assignment could only be obtained with the help of selectively ^{15}N-labeled samples of apo-CRABP and by comparison with the assignments obtained for holo-CRABP.

Binding of RA to CRABP stabilizes the structure of the protein, and holo-CRABP has a denaturation temperature 17 °C higher than apo-CRABP at pH 7.0 (Zhang et al., 1992). Addition of the ligand did not improve the behavior of the protein at pH values above the isoelectric point. However, 1.5 mM solutions of holo-CRABP are stable for weeks at pH 3.8 and 25 °C, due to the stabilization produced by the ligand. We chose these conditions for NMR analysis of holo-CRABP, since the linewidths are slightly narrower than at pH 7.5, and there is no broadening or loss of amide resonances due to exchange with the solvent.

Resonance assignment

Assignment of ^1H and ^{15}N resonances was perforemd in an analogous way for apo- and holo-CRABP. We will present the data obtained for the holo-protein, and will point out particularities of the assignment of apo-CRABP and problems encountered due to the fast exchange of some amide protons with the solvent. The process followed to obtain the assignments is similar to that described for other proteins with relatively large resonance linewidths (see, for instance, Marion et al., 1989c; Driscoll et al., 1990; Stockman et al., 1992). We used primarily 2D TOCSY and NOESY data obtained in D_2O on nonlabeled protein samples, and ^1H-^{15}N 2D HSQC, 3D

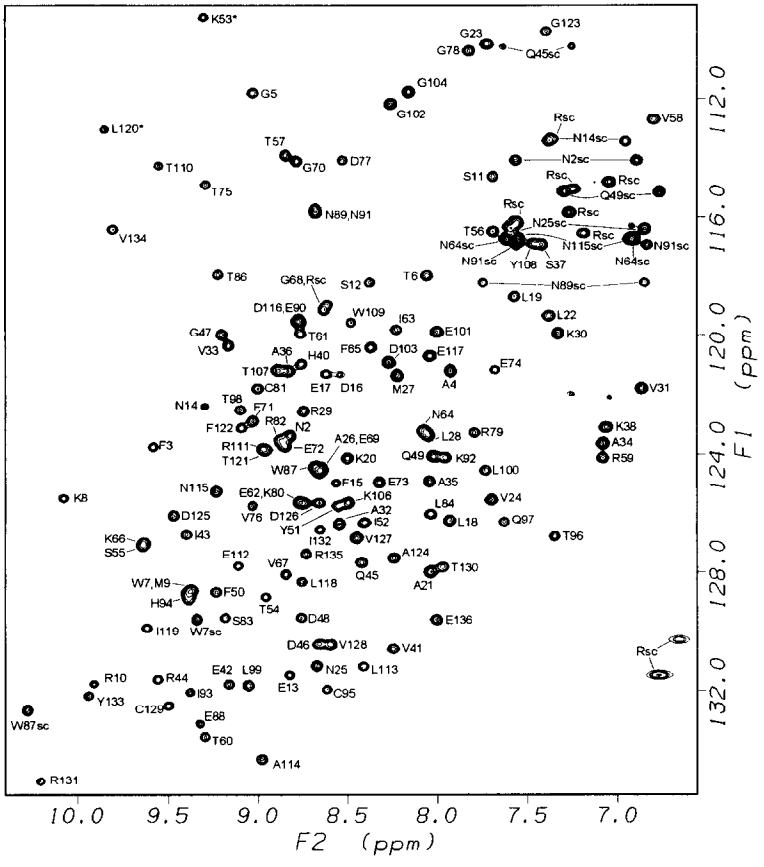


Fig. 2. Amide region of a 2D ^{1}H-^{15}N HSQC spectrum of holo-CRABP in 90% H_2O/10% D_2O (1.5 mM protein, 40 mM acetic acid-d_4, pH 3.8, 25 °C). The assignments of all cross peaks corresponding to backbone amide protons and to asparagine, glutamine and tryptophan side chains are indicated (note that the cross peak corresponding to the Trp^{109} side chain falls outside the region shown). Arginine side-chain cross peaks have been labeled Rsc.

TOCSY-HMQC and 3D NOESY-HMQC data obtained in H_2O on uniformly ^{15}N-labeled samples. The use of spin-lock purge pulses for solvent suppression in the heteronuclear experiments (Messerle et al., 1989) was extremely useful to avoid bleaching of cross peaks with chemical shifts near that of the water resonance, and to minimize saturation of amide protons in fast exchange with the solvent. A complete count of N^{α}/H^{N} cross peaks in the HSQC data (Fig. 2), and of H^{N}/H^{α} cross peaks in the 3D TOCSY-HMQC data, was obtained for holo-CRABP. For apo-CRABP, 126 cross peaks corresponding to backbone amide groups could be identified in the HSCQ data (the protein contains 132 non-proline residues); H^{N}/H^{α} cross peaks were identified for 120 residues in the 3D TOCSY-HMQC data, and 3D NOESY-HMQC data were observed for 124 residues, although these data were of poor quality for about 15 residues.

The assignment was initiated by identifying as many spin systems as possible in the 2D and 3D TOCSY data. Since the 3D data have poor digital resolution, the chemical shifts of the cross peaks were determined whenever possible from the corresponding cross peaks in 2D TOCSY-HMQC data. Although relayed connectivities between the amide and the side-chain protons were

only observed for about 60% and 40% of the residues of holo-CRABP and apo-CRABP, respectively, assignment of additional side-chain protons to their corresponding amide resonances could be obtained by comparison of the H^{α} chemical shifts of H^{N}/H^{α} cross peaks in the 3D TOCSY-HMQC data with those of H^{α}/side-chain cross peaks in the 2D TOCSY data. Intraresidue H^{N}/side-chain NOEs in the 3D NOESY data were helpful to solve ambiguities due to degeneracies in the H^{α} chemical shifts. Spin systems for many residues with short side chains were identified in this way, but very little information was obtained for residues with long side chains (lysine, arginine, leucine, isoleucine).

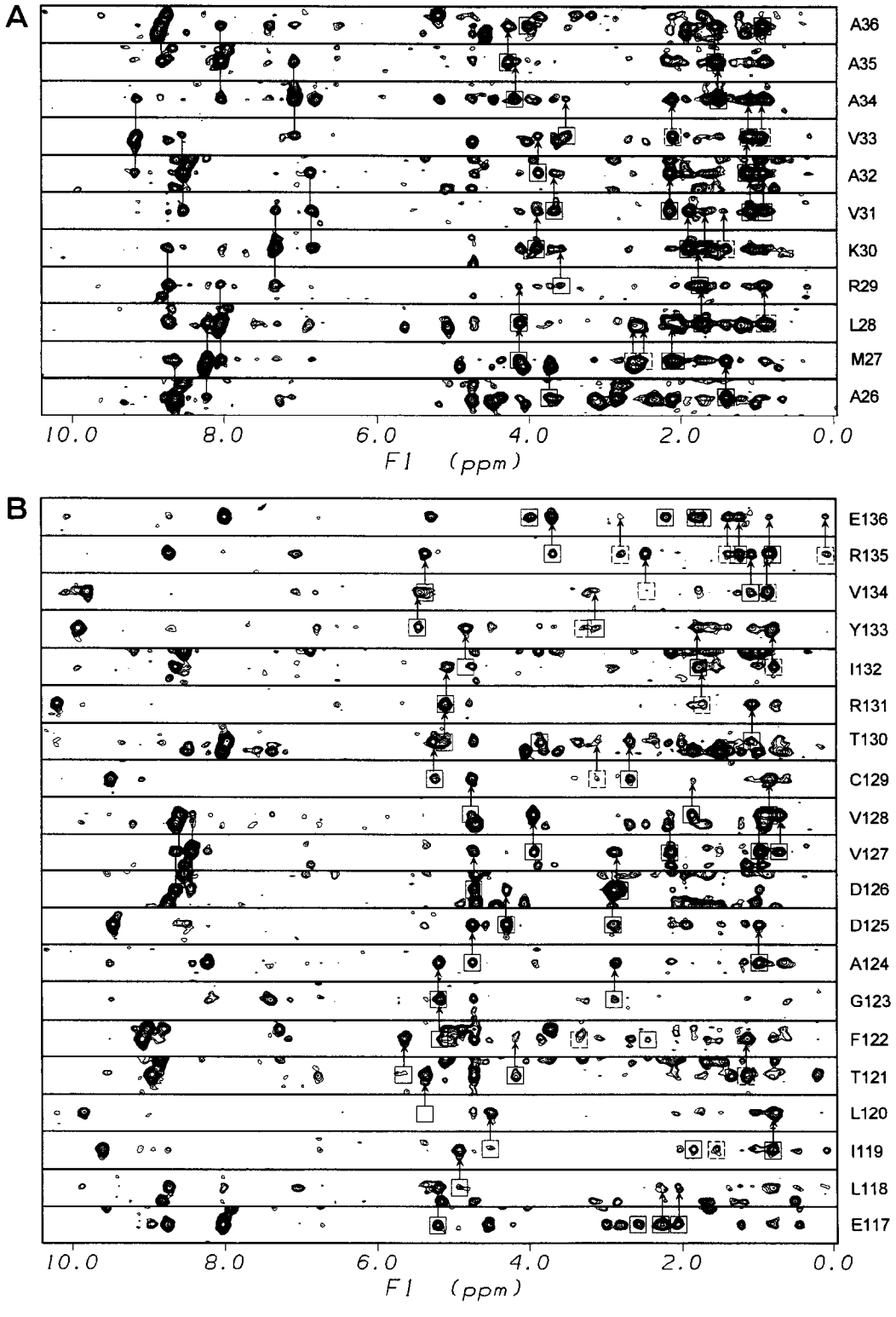
In order to obtain sequential assignments of backbone ^{1}H and ^{15}N resonances, a list of NOEs extracted from the 3D NOESY-HMQC data was compiled for each amide proton (identified as a cross peak in the HSQC data), and 2D (ω_1,ω_3) strips along the ω_1 dimension, taken at the chemical shift of the corresponding ^{15}N in ω_2, were reorganized according to the amino acid sequence as the assignment progressed. Two α-helical regions in holo-CRABP, comprising residues Phe^{15}–Gly^{23} and Ala^{26}–Ala^{36}, were easily assigned due to the presence of strong sequential $d_{NN}(i,i+1)$ connectivities, and because consecutive amide protons in these regions had many NOEs in common. Selected strips of the 3D NOESY-HMQC data, illustrating the assignment of the second helix, are shown in Fig. 3A. The same helical regions could be assigned for the apoprotein, but, due to fast exchange of amide protons with the solvent (see below), the data were of very poor quality at the edges of the helices (Phe^{15}–Glu^{17}, Met^{27} and Ala^{35}–Ala^{36}; no Ala^{26} N^{α}/H^{N} cross peak was observed in the HSQC data).

The frequent presence in turns and loops of $d_{NN}(i,i+1)$ and $d_{\beta N}(i,i+1)$ connectivities, in addition to $d_{\alpha N}(i,i+1)$ NOEs, facilitated also the assignment of resonances corresponding to these regions and to contiguous residues at the edges of β-strands. Residues Gly^{47}–Phe^{50}, Asn^{64}–Phe^{71}, Thr^{75}–Gly^{78}, Thr^{86}–Lys^{92}, Leu^{100}–Gly^{104} and Gly^{123}–Val^{128} were thus easily assigned for both forms of CRABP. Residues Thr^{56}–Thr^{61} and Ala^{114}–Glu^{117} of holo-CRABP were also easily assigned, but could not be identified initially for apo-CRABP due to the presence of fast-exchanging amide protons in these regions.

The NOE patterns observed for more than 50% of the amide protons of CRABP are characteristic of β-sheets, with strong $d_{\alpha N}(i,i+1)$ interactions, weak or absent intraresidue $d_{N\alpha}$ connectivities, and few NOEs common to consecutive amide protons. Despite the good dispersion of H^{α} chemical shifts observed in the spectra of CRABP (see, for instance, Fig. 1), there is still considerable chemical-shift degeneracy, which is compounded by the low digital resolution of the 3D NOESY-HMQC data. Assignment was also hindered by the lack of residue-type information from the TOCSY data for many of the amide protons in β-sheet residues, in particular for apo-CRABP. These problems were solved by recording HSQC spectra on selectively ^{15}N-labeled CRABP samples.

→

Fig. 3. Composite of (ω_1,ω_3) strips from a 3D ^{1}H-^{15}N NOESY-HMQC spectrum of holo-CRABP, illustrating the assignment of (A) the second helix; and (B) the two C-terminal β-strands (conditions as in Fig. 2; 130 ms mixing time). The strips were taken at the ω_2 chemical shift of the $^{15}N^{\alpha}$ atom of the corresponding residue. Boxes indicate cross peaks that were also observed in the 3D TOCSY-HMQC data, and dashed boxes indicate cross peaks that were not observed in the 3D ^{1}H-^{15}N TOCSY-HMQC data, but could be identified as intraresidue NOEs. Arrows point from intraresidue NOEs to sequential $d_{\alpha N}(i,i+1)$ and $d_{sidechainN}(i,i+1)$ connectivities. Lines connect diagonal cross peaks to sequential $d_{NN}(i,i+1)$ and $d_{NN}(i,i-1)$ NOEs.


A
A36
A35
A34
V33
A32
V31
K30
R29
L28
M27
A26
10.0
8.0
6.0
4.0
2.0
0.0
F1 (ppm)
B
E136
R135
V134
Y133
I132
R131
T130
C129
V128
V127
D126
D125
A124
G123
F122
T121
L120
I119
L118
E117
10.0
8.0
6.0
4.0
2.0
0.0
F1 (ppm)

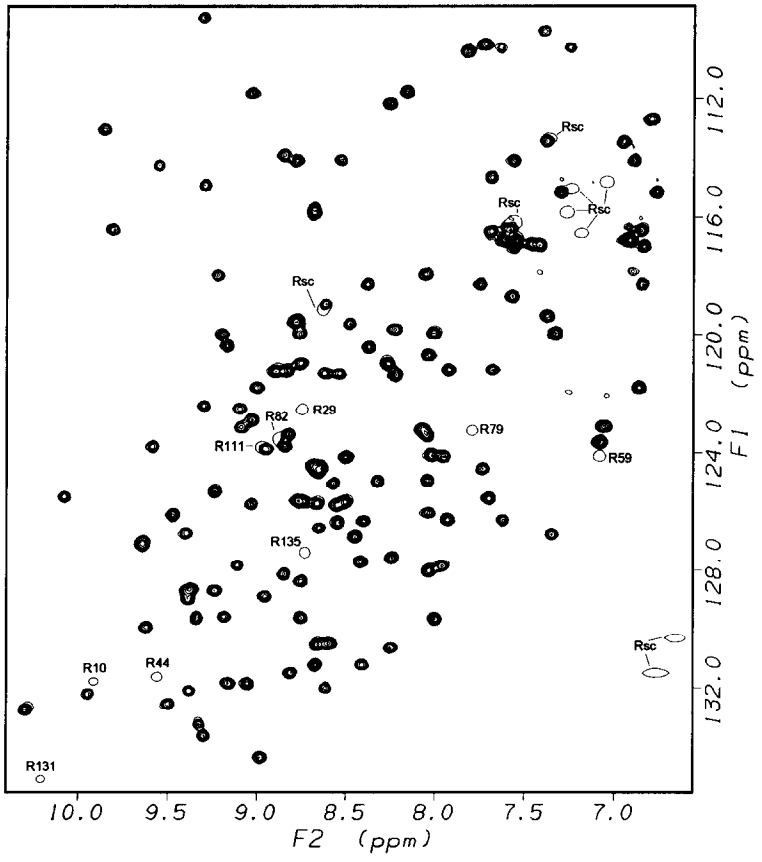


Fig. 4. Superposition of 2D ^{1}H-^{15}N HSQC spectra corresponding to holo-CRABP (single contours) and holo-CRABP with reverse selective labeling of arginine residues (multiple contours) (conditions as in Fig. 2). The sample with reverse selective labeling was obtained by growing cells in the presence of $^{15}NH_4Cl$ and nonlabeled arginine.

Considering which residues were more abundant in the β-sheets, we chose to prepare three selectively labeled samples, growing cells in the presence of ^{15}N-lysine, ^{15}N-D,L-threonine or ^{15}N-leucine. With the prototrophic cell line used, full scrambling of threonine to glycine occurs; leucine scrambles fully to valine and isoleucine, and partially to alanine (Muchmore et al., 1989). We also prepared a sample with 'reverse selective labeling' of arginine residues, i.e., the cells were grown in the presence of $^{15}NH_4Cl$ and nonlabeled arginine so that all residues were ^{15}N-labeled, except for arginine. We recorded HSQC spectra in the absence of RA for the four samples, and in the presence of RA for the samples obtained with ^{15}N-D,L-threonine and with arginine. In Fig. 4, the HSQC spectrum obtained for 'reverse arginine-labeled' holo-CRABP (multiple contours) is shown superimposed with that of uniformly ^{15}N-labeled holo-CRABP (single contours). The 'reverse selective-labeling' method yields basically the same information as that obtained by selectively labeling an amino acid residue, and is particularly useful when the ^{15}N-labeled amino acid is expensive or difficult to obtain commercially.

With the information obtained from the selective-labeling experiments and that compiled earlier from the TOCSY data, it was straightforward to obtain complete assignments for the

TABLE 1
CHEMICAL SHIFTS FOR ASSIGNED ^{1}H AND ^{15}N RESONANCES OF THE RETINOIC ACID-BOUND FORM OF CRABP (PPM)[a]

Residue	N^{α}	H^{N}	H^{α}	H^{β}	Others
Pro[1]			4.40	2.51,2.27	H^{γ} 2.08*; H^{δ} 3.50,3.40
Asn[2]	123.39	8.82	5.04	3.09,3.06	H^{γ} 7.56,6.89; N^{δ} 114.10
Phe[3]	123.77	9.58	4.40	3.34,3.18	H^{δ} 7.38; H^{ε} 7.21; H^{ζ} 7.36
Ala[4]	121.23	7.93	4.05	1.66	
Gly[5]	111.83	9.03	4.33,3.89		
Thr[6]	118.03	8.06	5.30	4.05	H^{γ} 1.25
Trp[7]	128.69	9.37	5.28	2.96,2.66	$H^{\delta 1}$ 6.78; $H^{\zeta 2}$ 7.13; $H^{\eta 2}$ 7.03; $H^{\zeta 3}$ 6.74; $H^{\varepsilon 3}$ 7.19; $H^{\varepsilon 1}$ 9.36; $N^{\varepsilon 1}$ 129.62
Lys[8]	125.52	10.07	5.20	1.95,1.72	H^{γ} 1.40
Met[9]	128.69	9.37	3.86	2.05	
Arg[10]	131.80	9.91	4.53	1.67	H^{γ} 1.38; H^{δ} 3.15*
Ser[11]	114.65	7.69	4.72	3.86,3.70	
Ser[12]	118.26	8.38	5.07	3.85,3.61	
Glu[13]	131.50	8.82	4.90	2.18,2.05	H^{γ} 2.47,2.44
Asn[14]	122.43	9.30	5.00	3.38,3.33	H^{δ} 7.38,6.96; N^{δ} 113.45
Phe[15]	125.00	8.57	4.49	2.88,2.81	H^{δ} 7.19; H^{ε} 6.59; H^{ζ} 6.74
Asp[16]	121.34	8.55	3.88	2.66,2.54	
Glu[17]	121.32	8.63	3.67	1.93,1.78	H^{γ} 2.57,2.39
Leu[18]	126.28	7.93	3.71	2.08	
Leu[19]	118.72	7.57	3.58	1.66	H^{δ} 0.98,0.72
Lys[20]	124.15	8.51	3.82	1.88	H^{γ} 1.36; H^{δ} 1.57; H^{ε} 2.87*
Ala[21]	128.03	8.04	4.08	1.50	
Leu[22]	119.37	7.38	3.98	1.62	
Gly[23]	110.17	7.72	4.06,3.62		
Val[24]	125.57	7.70	3.83	1.46	H^{γ} 0.94,0.89
Asn[25]	131.20	8.67	4.39	3.18,2.84	H^{δ} 7.59,6.86; N^{δ} 116.43
Ala[26]	124.56	8.66	3.74	1.43	
Met[27]	121.39	8.23	4.16	2.15,2.07	H^{γ} 2.66,2.54
Leu[28]	123.42	8.05	4.13	1.71	H^{δ} 0.96,0.86
Arg[29]	122.57	8.75	3.62	1.82	
Lys[30]	119.97	7.33	3.92	1.93,1.70	H^{γ} 1.43*; H^{δ} 1.61*; H^{ε} 2.98*
Val[31]	121.81	6.87	3.70	2.17	H^{γ} 1.10,0.95
Ala[32]	126.39	8.55	3.90	1.17	
Val[33]	120.39	9.16	3.54	2.14	H^{γ} 1.10,0.98
Ala[34]	123.66	7.08	4.20	1.52	
Ala[35]	124.95	8.05	4.30	1.58	
Ala[36]	121.23	8.84	4.02	0.96	
Ser[37]	116.98	7.42	4.25	4.04*	
Lys[38]	123.12	7.07	4.83	1.99,1.67	H^{γ} 1.46,1.38
Pro[39]			5.17		
His[40]	121.01	8.76	5.27	3.39,3.21	
Val[41]	130.65	8.24	5.38	1.77	H^{γ} 0.80,0.56
Glu[42]	131.83	9.16	5.29	2.29,2.09	H^{γ} 2.35*
Ile[43]	126.75	9.40	5.53	2.32	$H^{\gamma 2}$ 0.84; $H^{\delta 1}$ 0.53
Arg[44]	131.66	9.55	4.82	1.82	
Gln[45]	127.71	8.42	4.70		H^{γ} 1.73; H^{ε} 7.63,7.26; N^{ε} 110.26

TABLE 1 (continued)

Residue	N^{α}	H^{N}	H^{α}	H^{β}	Others
Asp^{46}	130.49	8.66	4.76	2.75,2.50	
Gly^{47}	120.00	9.21	4.04,3.64		
Asp^{48}	129.56	8.76	5.12	3.25,2.94	
Gln^{49}	124.10	8.02	4.74	2.03	H^{γ} 2.19; H^{ε} 7.29,6.77; N^{ε} 115.16
Phe^{50}	128.71	9.24	4.72		H^{δ} 6.74; H^{ε} 6.97; H^{ζ} 6.59
Tyr^{51}	125.77	8.55	5.37	3.04,2.65	H^{δ} 6.95; H^{ε} 6.59
Ile^{52}	126.34	8.41	4.71		
Lys^{53}	137.19	9.30	5.23	1.89	
Thr^{54}	128.88	8.96	4.87	3.92	H^{γ} 0.98
Ser^{55}	127.10	9.64	5.36	3.87*	
Thr^{56}	116.53	7.69	4.93	4.37	H^{γ} 1.27
Thr^{57}	113.94	8.85	4.05	4.27	H^{γ} 1.36
Val^{58}	112.74	6.80	4.49	2.21	H^{γ} 0.89,0.86
Arg^{59}	124.15	7.08	4.48	1.49	H^{γ} 1.18*
Thr^{60}	133.60	9.30	5.37	3.96	H^{γ} 1.12
Thr^{61}	119.98	8.76	4.71	4.08	H^{γ} 1.02
Glu^{62}	125.68	8.76	4.97	2.14,1.97	
Ile^{63}	119.84	8.23	5.07	1.96	$H^{\gamma 1}$ 1.24
Asn^{64}	123.32	8.07	5.65	2.63,2.55	H^{δ} 7.62,6.92; N^{δ} 116.77
Phe^{65}	120.44	8.37	4.70		H^{δ} 6.22; H^{ε} 6.45; H^{ζ} 6.12
Lys^{66}	127.10	9.64	5.20	2.05,1.65	H^{γ} 1.29,1.22
Val^{67}	128.12	8.85	3.12	1.75	H^{γ} 0.53,−0.15
Gly^{68}	118.91	8.61	4.54,3.64		
Glu^{69}	124.56	8.66	4.79	2.35,2.15	H^{γ} 2.45*
Gly^{70}	114.16	8.79	4.90,3.77		
Phe^{71}	122.92	9.03	5.13	3.73,3.32	H^{δ} 7.29
Glu^{72}	123.80	8.85	5.09	2.17,2.06	H^{γ} 2.51,2.37
Glu^{73}	124.97	8.33	4.82		
Glu^{74}	121.20	7.68	5.39	2.01,1.70	H^{γ} 2.11
Thr^{75}	114.92	9.30	4.18	4.59	H^{γ} 1.07
Val^{76}	125.77	9.03	3.71	1.49	H^{γ} 0.56,0.42
Asp^{77}	114.10	8.53	4.45	2.77*	
Gly^{78}	110.42	7.83	4.00,3.59		
Arg^{79}	123.31	7.79	4.44	1.85,1.69	H^{γ} 1.53,1.34
Lys^{80}	125.68	8.76	4.41		
Cys^{81}	121.83	9.00	5.16	2.67,2.16	
Arg^{82}	123.60	8.87	4.69	1.72,1.59	
Ser^{83}	129.59	9.18	5.10	2.85,1.87	
Leu^{84}	126.06	8.03	5.03	1.87	H^{δ} 1.18
Pro^{85}			5.03		
Thr^{86}	118.01	9.22	4.50	4.07	H^{γ} 1.04
Trp^{87}	124.46	8.69	5.23	3.25,3.08	$H^{\delta 1}$ 7.28; $H^{\zeta 2}$ 8.08; $H^{\eta 2}$ 7.03; $H^{\zeta 3}$ 7.39; $H^{\varepsilon 3}$ 7.67; $H^{\varepsilon 1}$ 10.28; $N^{\varepsilon 1}$ 132.68
Glu^{88}	133.14	9.33	4.30	1.96,1.80	H^{γ} 2.31*
Asn^{89}	115.82	8.69	4.61	3.19,3.09	H^{δ} 7.75,6.85; N^{δ} 118.26
Glu^{90}	119.57	8.78	3.79	1.98,1.91	H^{γ} 2.31*
Asn^{91}	115.82	8.69	5.22	3.45,2.89	H^{δ} 7.56,6.84; N^{δ} 117.10
Lys^{92}	124.13	7.96	5.65	1.82	H^{γ} 1.65

TABLE 1 (continued)

Residue	N^{α}	H^{N}	H^{α}	H^{β}	Others
Ile^{93}	132.10	9.39	4.63	1.37	$H^{\gamma 1}$ 0.25; $H^{\gamma 2}$ −0.25; $H^{\delta 1}$−0.36
His^{94}	129.01	9.39	4.65		
Cys^{95}	131.99	8.61	4.73		
Thr^{96}	126.80	7.35	4.19	4.08	H^{γ} 1.23
Gln^{97}	126.31	7.63	5.03		H^{γ} 2.60; H^{ε} 7.42,6.90; N^{ε} 117.88
Thr^{98}	122.54	9.10	4.56	3.87	H^{γ} 1.21
Leu^{99}	131.86	9.06	4.06	1.60	H^{δ} 0.74
Leu^{100}	124.56	7.73	4.18	1.57,1.46	H^{δ} 0.81*
Glu^{101}	119.94	8.00	4.52	2.08,1.91	H^{γ} 2.31*
Gly^{102}	112.19	8.26	4.09,3.73		
Asp^{103}	120.93	8.27	4.95	2.74,2.66	
Gly^{104}	111.81	8.17	4.13,3.97		
Pro^{105}			4.59		
Lys^{106}	125.68	8.50	4.59	2.12,1.96	H^{γ} 1.48; H^{δ} 1.69
Thr^{107}	121.21	8.88	5.38	4.56	H^{γ} 1.16
Tyr^{108}	116.95	7.47	4.98	3.08,2.65	H^{δ} 6.57; H^{ε} 6.73
Trp^{109}	119.62	8.48	5.76	3.27,3.10	$H^{\delta 1}$ 7.15; $H^{\zeta 2}$ 6.60; $H^{\eta 2}$ 6.73; $H^{\zeta 3}$ 6.96; $H^{\varepsilon 3}$ 6.91; $H^{\varepsilon 1}$ 12.00; $N^{\varepsilon 1}$ 143.09
Thr^{110}	114.29	9.55	5.45	4.19	H^{γ} 1.20
Arg^{111}	123.86	8.97	5.36	1.71	H^{γ} 1.39
Glu^{112}	127.84	9.11	5.77	2.07*	H^{γ} 2.24*
Leu^{113}	131.23	8.42	5.00	1.40,0.98	H^{δ} 0.52,0.49
Ala^{114}	134.34	8.98	4.60	1.23	
Asn^{115}	125.30	9.24	4.24	3.00,2.87	H^{δ} 7.55,6.92; N^{δ} 116.77
Asp^{116}	119.57	8.78	4.59	3.07,2.87	
Glu^{117}	120.71	8.05	5.27	2.29,2.09	H^{γ} 2.61*
Leu^{118}	128.36	8.76	4.94		
Ile^{119}	129.92	9.62	4.53	1.86	$H^{\gamma 1}$ 1.57; $H^{\gamma 2}$ 0.78
Leu^{120}	140.99	9.85	5.38		
Thr^{121}	123.86	8.97	5.68	4.21	H^{γ} 1.16
Phe^{122}	123.15	9.09	5.20	3.33,2.52	H^{δ} 7.18; H^{ε} 7.05; H^{ζ} 7.27
Gly^{123}	109.73	7.39	5.22,2.88		
Ala^{124}	127.57	8.24	4.76	1.02	
Asp^{125}	126.12	9.47	4.32	2.95*	
Asp^{126}	125.68	8.66	4.73	2.83*	
Val^{127}	126.86	8.45	3.98	2.18	H^{γ} 0.99,0.73
Val^{128}	130.49	8.60	4.76	1.87	H^{γ} 0.86*
Cys^{129}	132.54	9.50	5.28	3.12,2.69	
Thr^{130}	127.87	7.98	5.12	3.86	H^{γ} 1.09
Arg^{131}	135.08	10.20	5.12	1.75	
Ile^{132}	126.58	8.66	4.85	1.81	$H^{\gamma 2}$ 0.82
Tyr^{133}	132.21	9.94	5.50	3.20,3.08	H^{δ} 6.82; H^{ε} 6.50
Val^{134}	116.45	9.80	5.40	2.50	H^{γ} 1.09,0.90
Arg^{135}	127.43	8.73	3.71	1.37,1.23	H^{γ} 0.80,0.17; H^{δ} 2.82
Glu^{136}	129.62	8.00	4.02	1.88,1.78	H^{γ} 2.27*

[a] Data obtained at 25 °C and pH 3.8. The accuracy is ± 0.02 ppm for proton chemical shifts and ± 0.1 ppm for nitrogen chemical shifts. In cases where only one resonance has been identified for diastereotopic protons in methylene groups or diastereotopic methyl groups in isopropyl groups, an asterisk indicates certainty that the diastereotopic protons or methyl groups have identical chemical shifts.

backbone ^{1}H and ^{15}N resonances of the β-sheet residues of both holo- and apo-CRABP. Strips from the 3D NOESY-HMQC data, illustrating the sequential assignment of the two C-terminal β-strands of holo-CRABP, are shown in Fig. 3B. After assigning the β-sheet residues, comparison with the data obtained for holo-CRABP helped obtain assignments for some of the residues in the Thr[56]–Thr[61] and Ala[114]–Glu[117] loops of apo-CRABP. No N^{α}/H^{N} cross peak is observed for the amide groups of Thr[57], Thr[60] and Asn[115] in the HSQC data obtained at pH 7.5 and 25 °C, but cross peaks for Thr[60] and Asn[115] could be observed in the data obtained at pH 3.8 and 5 °C on 0.5 mM apo-CRABP (see above).

The resonances of many side chains were identified during the sequential assignment process. Additional assignments were obtained a posteriori by revising all the data acquired, in particular the 2D data obtained in D_2O. Intraresidue NOE connectivities were used to assign ^{1}H and ^{15}N resonances corresponding to the side chains of asparagine, glutamine and aromatic residues. Complete assignment of most aromatic side chains was obtained from TOCSY and NOESY data. DQF-COSY data were of poor quality due to the large resonance linewidths, but helped solving ambiguities in the assignment of aromatic and long side chains.

The process summarized above yielded assignments for all ^{1}H and ^{15}N backbone resonances of holo-CRABP, as well as full ^{1}H assignments for 69% of the side chains and partial ^{1}H assignments for 23% of the side chains (Table 1). For apo-CRABP, we obtained assignments of all ^{1}H and ^{15}N backbone resonances except for four amide groups and two H^{α} protons; 69% of the side chains were fully assigned and 19% partially (Supplementary material).

Secondary structure

The sequential and short-range ($|i-j| \leq 4$) NOEs involving H^{N}, H^{α} and H^{β} protons observed in the 3D NOESY-HMQC data corresponding to holo-CRABP are summarized in Fig. 5. The observation of consecutive $d_{NN}(i,i+1)$ and of numerous (i,i+3) and (i,i+4) connectivities in the segments Phe[15]–Gly[23] and Ala[26]–Ala[36] of holo-CRABP shows that these segments adopt an α-helical conformation. The NOE patterns indicate that the rest of the molecule consists mostly of turns and β-strands. A total of 93 interstrand NOEs (summarized in Fig. 6) were observed in the 2D NOESY and 3D NOESY-HMQC data of holo-CRABP. These data show unequivocally the presence of 10 strands of antiparallel β-sheet (one of them with a β-bulge), connected by reverse turns or by loops. The assignment and nomenclature used to identify the strands is indicated in Fig. 5 and their relative topology is shown in Fig. 6. Further characterization of the secondary structure of holo-CRABP was obtained from amide deuterium exchange rates, measured from the evolution of cross-peak intensities in a series of ^{1}H-^{15}N HSQC spectra obtained in D_2O, and from Wishart indexes, calculated from the upfield or downfield shifts of H^{α} protons with respect to random-coil values (Wishart et al., 1992) (Fig. 5). The deuterium exchange rates and the Wishart indexes are fully consistent with the secondary structure deduced from the NOE data. We also recorded an HMQC-J spectrum (Kay and Bax, 1990) to obtain $^{3}J_{HN\alpha}$ coupling constants. The low sensitivity of the data when processed with strong resolution enhancement did not allow accurate measurements, but the data were completely consistent with the secondary structure assignment (small couplings for residues in α-helices and large couplings for β-sheet residues).

The patterns of Wishart indexes and of sequential and short-range NOEs observed for apo-CRABP (not shown) are very similar to those of the holo-form. Similar patterns of interstrand interactions were also observed; differences between the two forms correspond to weak H^{N}/H^{α}

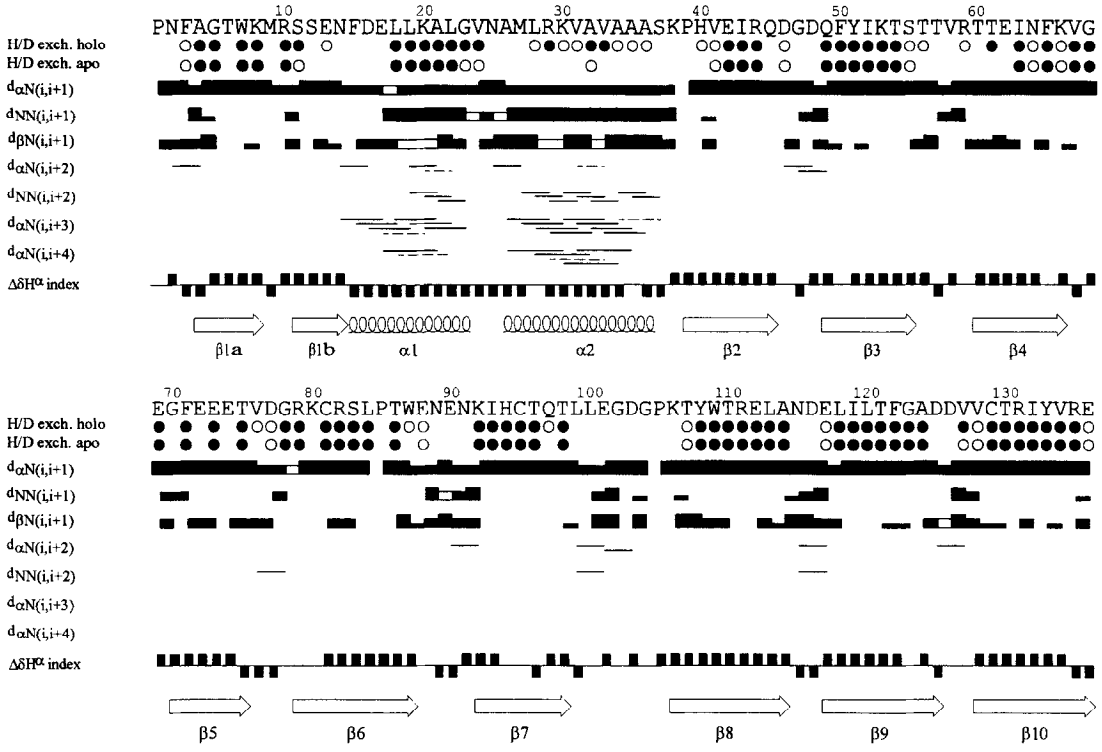


Fig. 5. Summary of sequential and short-range ($|i-j| \leq 4$) NOEs involving H^N, H^α and H^β protons, amide H/D exchange data, Wishart indexes ($\Delta\delta H^\alpha$ index) and secondary structure of holo-CRABP. Amide H/D exchange data for apo-CRABP are also shown. The size of the boxes for the sequential NOEs reflects the intensities of the corresponding cross peaks. Open boxes indicate NOEs whose presence could not be ascertained due to resonance degeneracy. All (i,i+2), (i,i+3) and (i,i+4) NOEs are weak, and those whose presence cannot be ascertained because of overlap with sequential connectivities are indicated by broken lines. The H/D exchange data for holo-CRABP were acquired at pD 3.8 (uncorrected) and 25 °C; solid circles indicate amide groups that remain partially protonated after 25 h of exchange, and open circles indicate amide protons that remain partially protonated after 2 h of exchange, but are fully deuterated after 25 h. For apo-CRABP, we measured H/D exchange rates at pD 6.4 (uncorrected) and 5 °C; solid circles indicate amide groups with exchange rates smaller than 10^{-6} min^{-1} and open circles indicate amide groups with exchange rates between 10^{-2} and 10^{-6} min^{-1}. Wishart indexes (Wishart et al., 1992) are indicated by solid boxes (above the central line for index +1, and below the line for index −1).

interstrand NOEs and are not concentrated in any particular region (this is illustrated in Fig. 6, where dotted arrows correspond to interactions observed in apo-CRABP that were not detected in holo-CRABP). Thus, such differences must be attributed to the low signal-to-noise ratios of these NOEs, rather than to real structural differences. The most significant differences in the NOE data of the apo-form correspond to the two helical segments, where fewer (i,i+3) and (i,i+4) interactions are observed, particularly at the termini. These observations are indicative of helix fraying at these termini.

To study the protection of amide protons in apo-CRABP, we performed deuterium exchange experiments at pD 6.4 and 5 °C, monitored by HSQC spectra (most HSQC cross peaks observed under these conditions can be assigned by correlation to the assignments obtained at pH 7.5 and 25 °C). In Fig. 5, the protection patterns are compared with those observed for the holo-protein

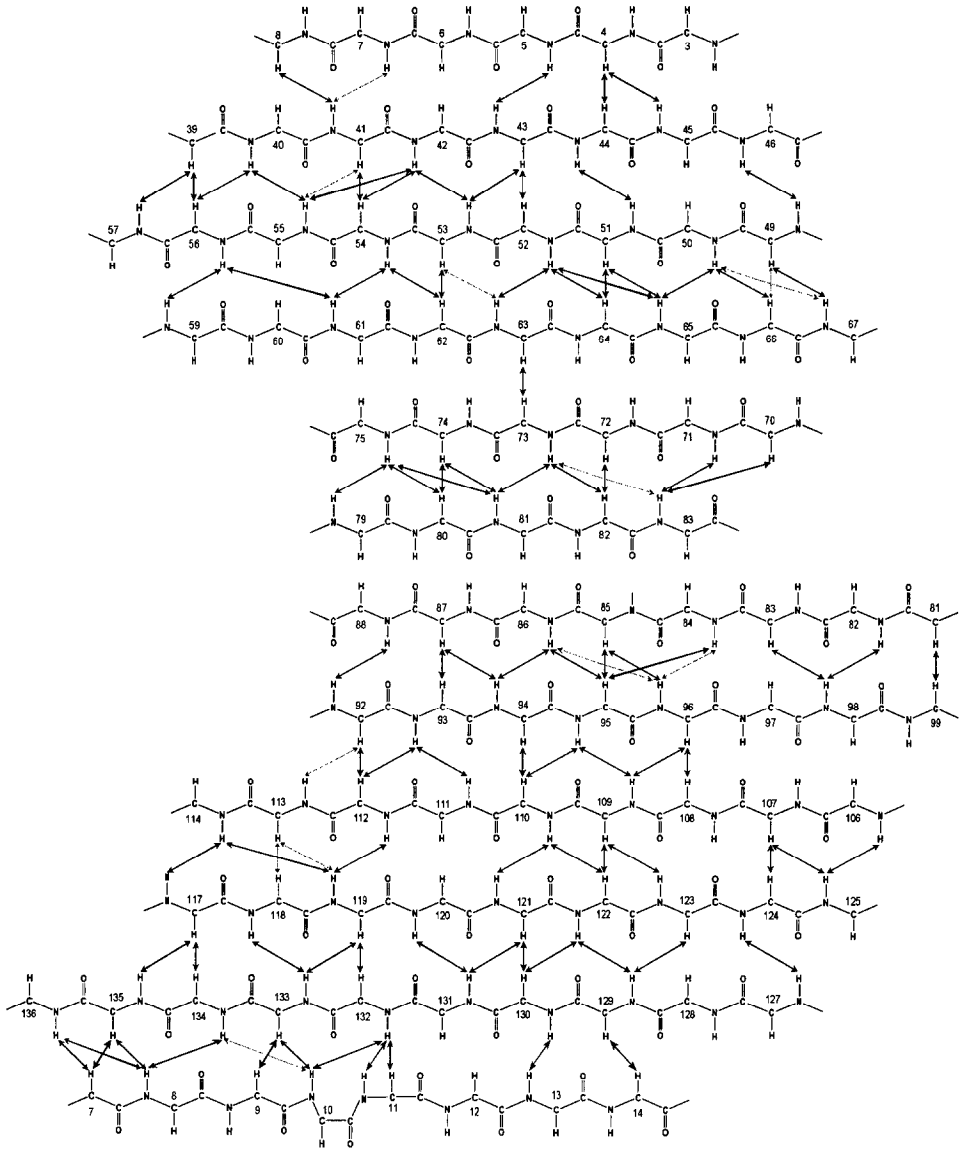


Fig. 6. Schematic representation of the relative topology of the β-strands in apo- and holo-CRABP. Interstrand NOEs observed for the holo-protein are indicated by solid arrows. Arrows with dotted lines indicate interstrand NOEs observed for apo-CRABP that were not observed for holo-CRABP. Most amide protons in the β-strands have slow H/D exchange rates in both forms of the protein and are presumably hydrogen bonded (hydrogen bonds are not indicated).

at pD 3.8 and 25 °C. The patterns are very similar in both forms of the protein, with the exception of that observed for helix α2. The low or absent protection of amide protons in this helix observed for apo-CRABP indicates that the helix has a significantly lower stability than in the holo-protein. Helix α1 also appears to be somewhat less stable in apo-CRABP. Such helix instability caused the low quality of the 3D NMR data observed for residues at the edges of the helices in apo-CRABP (see above), and is also reflected in the chemical shifts of the H^{α} protons of both

helices, which have smaller upfield shifts in apo-CRABP than in holo-CRABP (see Fig. 7; note that the upfield shifts of H^{α} protons with respect to random-coil values can be used as a measure of α-helicity (Rizo et al., 1993)). The correlation of these observations with ligand binding is discussed below.

Three-dimensional structural model of CRABP

Figure 8 shows a ribbon diagram, representing the 3D model of the structure of CRABP that we constructed based on the crystal structure of myelin P2 protein (Zhang et al., 1992). The NMR data described above show that the basic architecture predicted by the model is correct. There is a complete coincidence between the model and the elements of secondary structure deduced from the NMR data (Fig. 5). In addition, most interstrand interactions and amide protection patterns observed (Figs. 5 and 6) are consistent with the model. Thus, we can conclude that the structure of CRABP is analogous to those of other homologous hydrophobic ligand-binding proteins (Banaszak et al., 1994). The bulk of this structure is formed by two β-sheets, containing five strands each, all disposed in an antiparallel arrangement. The two sheets form a sandwich or β-clam structure, with the strands of each sheet oriented in perpendicular directions (Fig. 8). In addition, the structure contains an N-terminal helix–turn–helix motif that seals the binding cavity between the two sheets.

Hydrogen bonding is maximized in most of the CRABP structure, with the two sheets forming indeed a continuum of hydrogen-bonding interactions. Out of 132 backbone amide protons, 94 are weakly or strongly protected from the solvent in the holo-form (Fig. 5). The most N-terminal strand, β1, is mainly part of the N-terminal sheet, but forms a bulge in residue 10 so that the last residues of the strand (labeled β1b in Figs. 5 and 8) are part of the C-terminal sheet. Note that there are interstrand interactions between the C-terminal strand and residues Trp^{7} to Asn^{14} (Fig. 6). The hydrogen-bonding network is only interrupted between strands β4 and β5 of the N-terminal sheet, which are somewhat distant from each other as in all other homologous structures studied so far (Banaszak et al., 1994). Only one NOE between backbone protons of these strands is observed (Fig. 6) and the amide protons in the interface of the two strands (Glu^{62}, Asn^{64}, Lys^{66}, Gly^{70}, Glu^{72} and Glu^{74}) have weak or no protection from the solvent (Fig. 5). On the other side, strand β5 forms hydrogen bonds and has several NOE interactions with the N-terminal residues of the first strand of the C-terminal sheet, strand β6 (See Figs. 6 and 8). Thus, residues Arg^{79}–Arg^{82} can be considered part of the N-terminal sheet, much in the same way that residues Ser^{11}–Asn^{14} are part of the C-terminal sheet. The latter sheet does not have any gap between strands and, in the model, has a more standard antiparallel β-sheet geometry than the N-terminal sheet.

All inconsistencies between the NMR data and the structural model of CRABP detected so far correspond to residues that were inserted or to side chains that were replaced in the modeling process. For instance, the loop formed by the segment Leu^{99}–Lys^{106}, which contains three more residues than the corresponding sequence in myelin P2 protein, protrudes outwards in the model, but the NMR data suggest that this loop is somewhat folded over the loop formed by residues Ala^{124}–Val^{127}. Several NOEs involving aromatic protons also reveal the presence of side chains with wrong orientation in the model. Thus, a detailed analysis of the binding mechanism of RA to CRABP and of the interactions that define binding specificity will only be possible upon determination of the full solution structure of apo- and holo-CRABP, which is underway in our

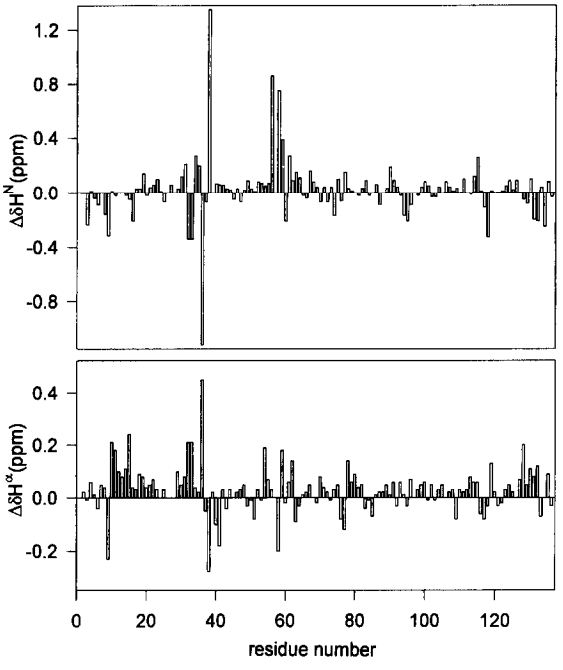


Fig. 7. Bar diagrams representing differences in the chemical shifts of H^N (above) and H^α (below) protons between apo- and holo-CRABP ($\Delta\delta$ = δ(apo) − δ(holo)).

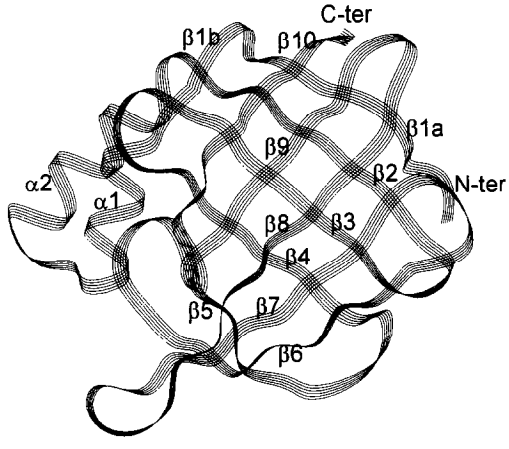


Fig. 8. Ribbon diagram representing the structural model of CRABP built by homology with the crystal structure of myelin P2 protein (Zhang et al., 1992). The positions of the N- and C-termini, as well as the α-helices and β-strands, are indicated.

laboratory. However, comparison of the data described above for the apo- and holo-forms already gives interesting clues on dynamic factors that may be crucial for ligand binding (see below).

The structures of the apo- and holo-forms have only been determined for three proteins homologous to CRABP, all in the crystal state (Banaszak et al., 1994). Ligand binding produces very few changes in the structure of these proteins and occurs in a large cavity between the two helices and strands β2–β5 and β7–β10, which is almost sealed from the exterior of the protein. Ligand entry probably occurs through a small opening between the two helices and the loops joining strands β3−β4 and β5–β6. Based on calculations of the electrostatic potential of the model structure of CRABP, we proposed that entry of RA into the binding cavity may be guided by a positive potential, extending from the cavity to the entry site (Zhang et al., 1992).

Although comparison of the NMR data obtained for apo- and holo-CRABP is hindered by the fact that the data for the two forms were obtained under different conditions, the similarity between the NMR parameters observed for most of the residues in both forms of the molecule indicates that the distinctions observed may be meaningful, in particular if there are numerous differences in several types of NMR parameters concentrated in a region of the molecule. The NOE data observed for apo- and holo-CRABP indicate that the overall architecture and the secondary structure of the protein do not change upon ligand binding. As explained above, the NOE data, the deuterium exchange rates and the H^α chemical shifts indicate that the two helices, in particular α2, are less stable in the apo-protein than in the holo-protein. Significant differences in deuterium exchange rates are also observed in the loop connecting strands β3 and β4 (residues Thr^{56}–Thr^{61}), where better protection of the amide protons is observed for the holo-protein (Fig. 5). Compari-

son of the chemical shifts of the H^N and H^α protons (Fig. 7) shows that the major differences correspond, in addition to the helical residues, to the segments Arg^{10}–Asn^{14} (strand β1b), Ser^{37}–Val^{41} (loop connecting α2 and β2) and Thr^{56}–Glu^{62} (loop β3–β4). These residues are all located in the same region of the protein (Fig. 8), i.e., in the upper part of the ligand-binding cavity. Most of the amide protons that have weak or absent cross peaks in the ^1H-^{15}N HSQC spectrum of apo-CRABP at pH 7.5 and 25 °C, due to fast exchange with the solvent, belong to this region of the molecule. All these results suggest strongly that this region is more flexible in apo-CRABP than in holo-CRABP. The observation that limited digestion with chymotrypsin cleaves primarily at the C-terminus of helix α2 in apo-CRABP, but that no cleavage is observed in holo-CRABP under the same conditions (Liu, 1993), also agrees with this conclusion. Based on the observation that the denaturation temperature of apo-CRABP is 17 °C lower than that of holo-CRABP, we suggested that flexibility in the apo-protein may be necessary to allow ligand entry into the cavity (Zhang et al., 1992), and crystallographic data on proteins homologous to CRABP have led to the proposal that a small portal opens to allow ligand entry into the binding cavity. Our NMR data provide experimental support for these hypotheses and suggest that the region around the C-terminus of helix α2 may act as a flexible hinge to facilitate opening of the portal.

CONCLUSIONS

We have assigned most of the ^1H and ^{15}N resonances, as well as the secondary structure, of apo-and holo-CRABP. Both forms of CRABP have a β-clam structure analogous to other homologous intracellular hydrophobic ligand-binding proteins. To the best of our knowledge, this is the first time that the structures in solution of both the apo- and holo-forms of a protein of this family have been studied by multidimensional NMR methods. The analysis of both forms of CRABP in solution is particularly important because, in addition to the structural information that can be obtained at atomic resolution, dynamic effects that may be critical for the mechanism of ligand binding can be investigated. The most significant differences in the data obtained for apo- and holo-CRABP are found in the two helices and in the region surrounding the C-terminus of the second helix. The data suggest that this region may be flexible in the apo-protein to facilitate access of RA to the binding cavity.

We are currently using the NMR data described above to obtain the complete 3D structure in solution of both forms of CRABP. Since the NMR data for the holo-form were obtained at pH 3.8 to facilitate the complete assignment of amide resonances, in the future we will also obtain data on holo-CRABP at pH 7.5 to assure that differences with the data observed for the apo-form are only due to ligand binding. Heteronuclear NMR methods will be used to study the interaction with the ligand and the dynamics of apo- and holo-CRABP. The fact that a high percentage of amide protons are protected from the solvent in the structure of CRABP makes it a very good model to study the folding of predominantly β-sheet proteins, and we are currently studying its folding mechanism using pulsed-labeling H/D exchange experiments.

ACKNOWLEDGEMENTS

We thank Robert Barnes for writing the computer routine for digitally filtering the solvent resonance. This work was supported in part by a grant from the National Institutes of Health (GM27616). Support from the Robert A. Welch Foundation is gratefully acknowledged.

REFERENCES

Amos, B. and Lotan, R. (1991) *Methods Enzymol.*, **190**, 217–225.

Banaszak, L., Winter, N., Xu, Z., Bernlohr, D.A., Cowan, S. and Jones, T.A. (1994) *Adv. Protein Chem.*, **45**, 89–151.

Bax, A. (1989) *Methods Enzymol.*, **176**, 151–168.

Blomhoff, R., Green, M.H., Berg, T. and Norum, K.R. (1990) *Science*, **250**, 399–404.

Bodenhausen, G. and Ruben, D.J. (1980) *Chem. Phys. Lett.*, **69**, 185–189.

Cheng, L., Qian, S.-j., Rotschild, C., d'Avignon, A., Lefkowith, J.B., Gordon, J.I. and Li, E. (1991) *J. Biol. Chem.*, **266**, 24404–24412.

Chytil, F. and Ong, D.E. (1987) *Annu. Rev. Nutr.*, **7**, 321–335.

Cowan, S.W., Newcomer, M.E. and Jones, A.T. (1993) *J. Mol. Biol.*, **230**, 1225–1246.

Davis, D.G. and Bax, A. (1985) *J. Am. Chem. Soc.*, **107**, 2820–2821.

De Luca, L.M. (1991) *FASEB J.*, **5**, 2924–2933.

Driscoll, P.C., Clore, G.M., Marion, D., Wingfield, P.T. and Gronenborn, A.M. (1990) *Biochemistry*, **29**, 3542–3556.

Fesik, S.W. and Zuiderweg, R.P. (1990) *Q. Rev. Biophys.*, **23**, 97–131.

Fiorella, P.D. and Napoli, J.L. (1991) *J. Biol. Chem.*, **266**, 16572–16579.

Griesinger, C., Otting, G., Wüthrich, K. and Ernst, R.R. (1988) *J. Am. Chem. Soc.*, **110**, 7870–7872.

Hoffman, M. (1990) *Science*, **250**, 372–373.

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

Kay, L.E. and Bax, A. (1990) *J. Magn. Reson.*, **86**, 110–126.

Kumar, A., Wagner, G., Ernst, R.R. and Wüthrich, K. (1981) *J. Am. Chem. Soc.*, **103**, 3654–3658.

Lassen, D., Lücke, C., Kromminga, A., Lezius, A., Spener, F. and Rüterjans, H. (1993) *Mol. Cell. Biochem.*, **123**, 15–22.

Liu, Z.-P. (1993) Ph.D. Thesis, University of Texas Southwestern Medical Center, Dallas, TX.

Liu, Z.-P., Rizo, J. and Gierasch, L.M. (1994) *Biochemistry*, **33**, 134–142.

Lotan, R. and Clifford, J.L. (1991) *Biomed. Pharmacother.*, **45**, 145–156.

Maden, M. (1994) *Curr. Biol.*, **4**, 281–284.

Maden, M., Ong, D.E., Summerbell, D. and Chytil, F. (1988) *Nature*, **335**, 733–735.

Marion, D. and Bax, A. (1989) *J. Magn. Reson.*, **83**, 205–211.

Marion, D., Ikura, M. and Bax, A. (1989a) *J. Magn. Reson.*, **84**, 425–430.

Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. and Bax, A. (1989b) *J. Am. Chem. Soc.*, **111**, 1515–1517.

Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989c) *Biochemistry*, **28**, 6150–6156.

Messerle, B.A., Wider, G., Otting, G., Weber, C. and Wüthrich, K. (1989) *J. Magn. Reson.*, **85**, 608–613.

Moon, R.C., McCormik, D.L. and Mehta, R.G. (1983) *Cancer Res.*, **43**, 2469S–2475S.

Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson, D.E. and Dahlquist, F.W. (1989) *Methods Enzymol.*, **177**, 44–73.

Napoli, J.L., Posch, K.P., Fiorella, P.D. and Boernman, M.H.E.M. (1991) *Biomed. Pharmacother.*, **45**, 131–143.

Petrou, S., Ordway, R.W., Singer, J.J. and Walsh Jr., J.V. (1993) *Trends Biochem. Sci.*, **18**, 41–42.

Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **117**, 479–485.

Rizo, J., Blanco, F.J., Kobe, B., Bruch, M.D. and Gierasch, L.M. (1993) *Biochemistry*, **32**, 4881–4894.

States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) *J. Magn. Reson.*, **48**, 286–292.

Stockman, B.J., Nirmala, N.R., Wagner, G., Delcamp, T.J., De Yarman, M.T. and Freisheim, J.H. (1992) *Biochemistry*, **31**, 218–229.

Sweetser, D.A., Heuckeroth, R.O. and Gordon, J.I. (1987) *Annu. Rev. Nutr.*, **7**, 337–359.

Willhite, C.C., Wier, P.J. and Berry, D.L. (1989) *CRC Crit. Rev. Toxicol.*, **20**, 113–135.

Winter, N.S., Bratt, J.M. and Banaszak, L.J. (1993) *J. Mol. Biol.*, **230**, 1247–1259.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) *Biochemistry*, **31**, 1647–1651.

Wolf, D.P. (1991) *Nutr. Rev.*, **49**, 1–12.

Zhang, J., Liu, Z.-P., Jones, T.A., Gierasch, L.M. and Sambrook, J.F. (1992) *Protein Struct. Funct. Genet.*, **13**, 87–99.

Zuiderweg, E.R.P. and Fesik, S.W. (1989) *Biochemistry*, **28**, 2387–2391.