Journal of Biomolecular NMR, 6 (1995) 271–276
ESCOM

J-Bio NMR 308

Overexpression of myoglobin and assignment of its amide, C^{α} and C^{β} resonances

Patricia A. Jennings*, Martin J. Stone** and Peter E. Wright***

Department of Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037, U.S.A.

Received 20 June 1995
Accepted 10 August 1995

Keywords: Inclusion bodies; Heme protein; Heteronuclear NMR; Myoglobin

Summary

Sperm whale apomyoglobin was expressed to high levels on minimal media and isotopically labeled with ^{13}C and ^{15}N nuclei. The isotopically labeled apoprotein was purified to homogeneity in a single step by reversed-phase chromatography and reconstituted with hemin and carbon monoxide gas for NMR analysis. Sequence-specific backbone $^{1}H^{N}$, ^{15}N and $^{13}C^{\alpha}$ as well as side-chain $^{13}C^{\beta}$ resonance assignments have been made for over 90% of the amino acids in the carbon monoxide complex of the protein. Resonance assignments were made by analysis of a series of 3D triple resonance spectra measured on the uniformly labeled sample. These assignments will provide the basis for analyzing the effects of point site mutations on the structure, stability and dynamics of the protein in solution.

Introduction

The oxygen storage protein myoglobin (Mb) is one of the most widely studied systems for elucidating structure–function relationships. The protein from sperm whale has been the most studied variant, for historical as well as practical reasons. Until recently it was widely available in large quantities from natural sources and the first protein structure determined was that of this protein (Kendrew, 1958). Sperm whale Mb consists of a single polypeptide chain of 153 amino acids and the iron protoporphyrin IX (heme) prosthetic group. The globular protein is folded into eight α-helices, which pack to form the hydrophobic binding pocket for the heme (Phillips et al., 1990; Ösapay et al., 1994).

Ongoing studies have focused on understanding the dynamics of ligand binding (Austin et al., 1975; Gibson et al., 1986; Anasari et al., 1993) and the structural fluctuations that may facilitate binding (Case and Karplus, 1979; Elber and Karplus, 1990). The effects of amino acid replacements, both in and removed from the binding pocket, on ligand binding kinetics can be analyzed in terms of 'breathing' motions resulting from backbone and/or side-chain conformational changes (Case and Karplus, 1979; Kottalam and Case, 1988). Elucidation of the solution structure and dynamics of Mb will greatly advance the understanding of structure–function relationships in heme proteins. In a first step towards this goal, we have recently determined the ^{1}H and ^{15}N resonance assignments (Theriault et al., 1994) and the solution structure (Ösapay et al., 1994) for the carbon monoxide complex of sperm whale Mb. Further refinement of the solution structure, however, was hampered by expression levels of the holoprotein on minimal media, which were insufficiently high for economical labeling with both ^{13}C and ^{15}N isotopes.

We report here the expression, purification, reconstitution and ^{1}H, ^{15}N, C^{α} and C^{β} assignments of uniformly ^{13}C- and ^{15}N-labeled Mb. In an effort to overcome the limitations on expression levels associated with apoprotein instability (Chiu, 1992) and heme biosynthesis and incorporation, we targeted the apoprotein to inclusion bodies in *Escherichia coli* (*E. coli*). Expression levels on minimal media were increased from ~2 mg per liter of cell culture to ~200 mg per liter, enabling the economical uniform incorporation of ^{13}C and ^{15}N nuclei and making triple-resonance NMR studies feasible.

*Present address: Department of Chemistry and Biochemistry, University of California at San Diego, San Diego, CA 92093-0359, U.S.A.
**Present address: Department of Chemistry, Indiana University, Bloomington, IN 47405-4001, U.S.A.
***To whom correspondence should be addressed.

Materials and Methods

The vector pCKR102, containing the wild-type Mb gene under the control of the *tac-lac* double promoter (Springer and Sligar, 1987), was the generous gift of Dr. Steve Sligar. An NdeI restriction site was introduced at the 5' end of the gene by PCR techniques. The amplified NdeI-KpnI fragment, containing the wild-type gene, was subcloned into the NdeI-KpnI sites of doubly digested pET17b vector (Novagen Inc., Madison, WI). This vector allows for protein production under the control of the strong T7 promoter. The integrity of the Mb gene was verified by dideoxy sequencing (Sanger et al., 1979).

BL21 (DE3) cells were transformed with the pet17b vector containing the gene for Mb, and grown on minimal media supplemented with ^{15}N-ammonium chloride and ^{13}C-labeled glucose. The growth medium for isotopic labeling consisted of M9 salts supplemented with 2 g/l ^{15}N-ammonium sulfate and 2 g/l ^{13}C-glucose, 1 μM $FeCl_3$, 25 μM $ZnSO_4$, 10 mg/l thiamine, 10 ml/l Basal Eagle vitamin mixture (BRL laboratories, Grand Island, NY) and 200 μg/ml ampicillin. Bacterial cells were grown to an OD_{600} of 1.0 at 37 °C prior to induction of directed protein synthesis with IPTG and a temperature shift to 42 °C. Cells were harvested by centrifugation 5 h after induction. Isotopic labeling was carried out on freshly transformed cells, in order to maximize purified protein yields.

Mb was produced as the apoprotein (apoMb) in inclusion bodies in BL21 (DE3) cells. Cells were resuspended in opening buffer, containing 50 mM Tris-HCl, 50 mM dipotassium ethylene diamine tetraacetic acid and 10 mM β-mercaptoethanol. Cells were lysed with hen egg white lysozyme (Sigma, St. Louis, MO) on ice at 4 °C with repeated sonication. Inclusion bodies containing Mb were recovered by centrifugation of the lysis solution at 10 000 rpm for 30 min. Inclusion bodies were repeatedly washed with opening buffer and collected by centrifugation. Solubilization of the inclusion bodies was accomplished by resuspension in 10% sodium dodecyl sulfate (SDS) solution. Resuspended inclusion bodies were diluted with 0.1% trifluoroacetic acid (TFA) solution to a final SDS concentration of 5% prior to HPLC purification.

Protein purification was accomplished by reversed-phase HPLC (Waters, Bedford, MA) on a C_4 (Vydac, Hesperia, CA) column. The apoprotein was eluted with a linear gradient of 36–60% acetonitrile containing 0.1% TFA. Fractions containing apoMb were verified by mass spectrometric analysis and pooled. The protein was lyophilized to dryness prior to reconstitution with heme. ApoMb was >98% pure from this single chromatographic step, as confirmed by SDS–polyacrylamide gel electrophoresis.

The holoprotein was reconstituted in the following manner. The lyophilized apoprotein was solubilized in buffer containing 10 mM sodium acetate, 6 M urea at pH 6.1 and 4 °C. Refolding was initiated by rapid 7.5-fold dilution of the protein in urea into 10 mM sodium acetate buffer at pH 6.1 and 4 °C. The protein is fully folded within 5 s under these conditions (Jennings and Wright, 1993). Bovine hemin (Sigma) was dissolved in 0.1 M sodium hydroxide and added immediately, under stirring at 4 °C, to the refolded apoprotein in a 1.2-fold molar excess. The reconstituted holoprotein was concentrated to ~2 mM with an Amicon concentrator. Exchange into potassium phosphate buffer (0.05 M, pH 5.6) and removal of excess hemin were accomplished by applying the holoprotein to a PD-10 column (Pharmacia Biotech, Uppsala, Sweden) equilibrated with phosphate buffer. The carbon monoxide complex of myoglobin was prepared by reduction of CO-saturated protein solution with a threefold molar excess of freshly prepared sodium dithionite in 0.1 M NaOH, followed by removal of excess reductant by successive concentration/dilution of the protein with CO-saturated buffer in an Amicon concentrator. The NMR sample was prepared in phosphate buffer (0.05 M, pH 5.6) in 90% H_2O/10% D_2O solution. The sample was incubated at 35 °C for 24 h prior to NMR analysis. Equilibration is necessary to obtain a single heme isomer.

All NMR experiments were performed at 35 °C on a Bruker AMX-500 spectrometer equipped with a three-channel interface. ^{1}H chemical shifts were referenced to the water signal at 4.6516 ppm, while ^{15}N and ^{13}C were referenced using an indirect method to NH_3(l) and TMS, respectively, each at 0 ppm (Bax and Subramanian, 1986). In all experiments, the ^{1}H carrier was placed at the H_2O frequency during all pulses and shifted to 7.8 ppm for acquisition, while the ^{15}N carrier was placed at 120.1 ppm. Spectral widths of 5000 and 1066 Hz were employed for ^{1}H and ^{15}N, respectively. Frequency discrimination in the indirect dimensions was achieved using the TPPI-States method (Marion et al., 1989).

Constant-time 3D HNCA and HN(CO)CA experiments were performed using published pulse sequences (Grzesiek and Bax, 1992a). The ^{13}C carrier was placed at 53.8 ppm. ^{13}CO pulses were generated at 177 ppm using a PTS160 frequency synthesizer and amplified using a M3205A pulse amplifier (American Microwave Technology, Brea, CA). The $^{13}C^{\alpha}$ spectral width was 3111 Hz. A total of 1024, 27, and 32 complex points were acquired in the ^{1}H, ^{15}N, and ^{13}C dimensions, respectively, resulting in acquisition times of 204.8, 24.8, and 10.0 ms, respectively. The HNCA and HN(CO)CA experiments were acquired using 256 and 192 scans per hypercomplex (t_1,t_2) increment, with total experiment times of 109 and 85 h, respectively.

Constant-time 3D CBCA(CO)NH and CBCANH experiments were performed using published pulse sequences (Grzesiek and Bax, 1992a,b), modified in the following ways. (1) Bloch–Siegert phase error compensation in the

CBCA(CO)NH experiment was achieved by insertion of an additional $^{13}C^{\alpha}$ 180° pulse immediately before the second ^{13}CO 90° pulse, while the phase of this 90° pulse was x. (2) The ^{13}C carrier was shifted to the center of the C^{α} region (54 ppm) for the final aliphatic ^{13}C 90° and 180° pulses of the CBCANH experiment; while this modification gives improved excitation profiles for these pulses, it may result in suboptimal sensitivity due to loss of phase coherence. (3) Decoupling of protons from the transverse nuclei during ^{13}C and ^{15}N chemical shift constant-time evolution periods was accomplished using ^{1}H 180° pulses instead of DIPSI-2 sequences. (4) All C^{α} 180° pulses were shaped using a profile described by the center lobe of a (sin x)/x function. The $C^{\alpha\beta}$ carrier was placed at 40.0 ppm and the $C^{\alpha\beta}$ spectral width was 7289 Hz. ^{13}CO pulses were applied as shifted laminar pulses (Patt, 1992) with the center of the excitation profile at 177 ppm and were shaped in the time domain using a profile described by the center lobe of a (sin x)/x function. For the CBCA(CO)NH experiment, 1024, 29, and 47 complex points were acquired in the ^{1}H, ^{15}N, and ^{13}C dimensions, respectively, giving acquisition times of 204.8, 26.7, and 6.4 ms, respectively. For the CBCANH experiment, 1024, 25, and 46 complex points were acquired in the ^{1}H, ^{15}N, and ^{13}C dimensions, respectively, giving acquisition times of 204.8, 23.0, and 6.2 ms, respectively. The CBCA(CO)NH and CBCANH experiments were acquired using 128 and 256 scans per hypercomplex (t_1,t_2) increment, with total experiment times of 90 and 149 h, respectively.

Spectra were processed on an Iris Indigo computer using FELIX 2.05 software (Hare Research). A line broadening of 10 Hz was applied in ω_3, and cosine-bell weighting functions were used in ω_1 and ω_2. The residual water signal was suppressed by application of a low-pass filter and polynomial base-line correction was employed.

Results and Discussion

Detailed analysis of the structure and dynamics of proteins in solution is facilitated by overexpression of the protein in *E. coli*. Although the previous expression system for sperm whale Mb had been optimized for codon usage and allowed for overexpression of the target protein on rich media, the yield of labeled Mb from minimal media did not allow for cost-effective uniform labeling of the protein with ^{13}C and ^{15}N nuclei. We took advantage of our extensive experience with the apoprotein in order to optimize our expression levels. Protein induction with IPTG, coupled with a temperature shift to 42 °C, both increased production levels and deposited the expressed protein into inclusion bodies in *E. coli*. The apoprotein was thus stabilized against proteolysis in vivo. The deposition of protein into inclusion bodies offers the distinct advantage of facilitating purification, *if* conditions for the reversible folding are known.

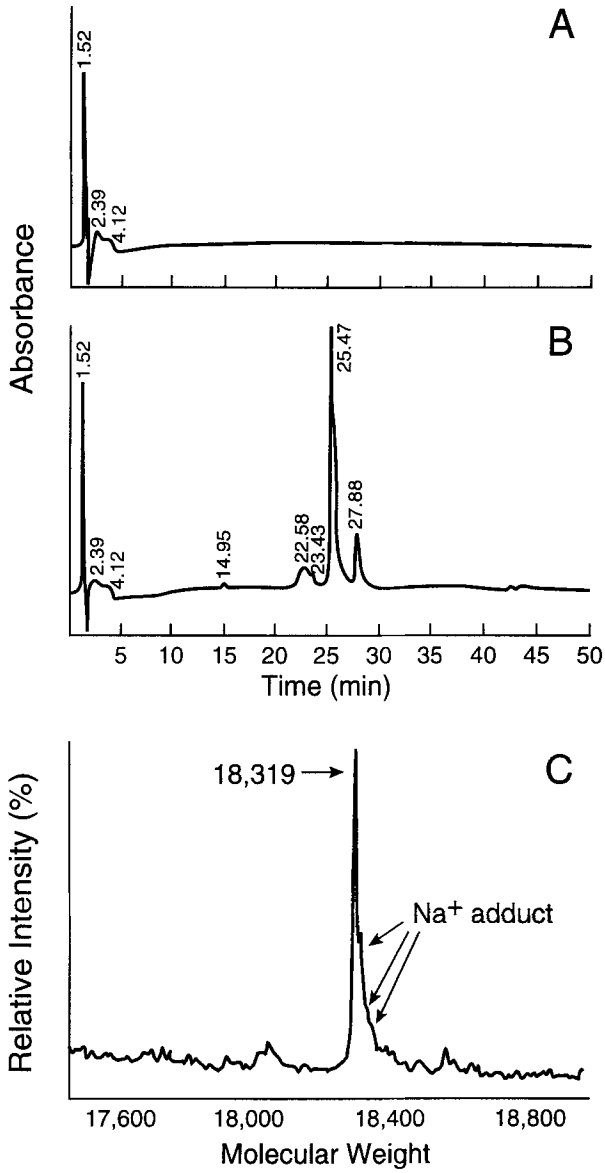

Fig. 1. (A) Analytical reversed-phase HPLC of 5% SDS in 0.1% TFA solution. A 15-µl sample of the solution was injected on a reversed-phase C_4 column and eluted with a linear gradient of 36–60% acetonitrile containing 0.1% TFA at a flow rate of 1.0 ml/min. Effluent was monitored by measuring the absorption at 214 nm. (B) Analytical reversed-phase HPLC of the inclusion bodies, solubilized in SDS solution. The elution conditions were as described above. ApoMb has a retention time of 25.47 min under these conditions. (C) Electrospray mass spectrum of purified labeled apoMb. The observed mass is indicated. The expected molecular weight for fully enriched material (99.5 atom % ^{15}N and 99 atom % ^{13}C) is 18 318 Da.

The inclusion bodies containing apoMb were especially resistant to solubilization with the standard urea and guanidine salt chaotropic agents. Therefore, we resorted to solubilization with the detergent SDS; very recently, we have found that solubilization can also be achieved using a mixture of acetonitrile and guanidine hydrochloride. The HPLC chromatograms of 5% SDS and SDS-solubilized inclusion bodies are shown in Figs. 1A and B. SDS does not appreciably bind to the C_4 column and is easily

removed from the solubilized protein by this technique. The complete removal of SDS from the protein was confirmed by the absence of additional aliphatic resonances in 1D ^{1}H spectra of purified MbCO (data not shown). Figure 1B illustrates the high concentration of apoMb contained in the pellet. The retention time of apoMb is not affected by the presence of SDS in the sample (data not shown). The target protein as well as any contaminating proteases are unfolded in SDS, minimizing protein loss due to proteolysis during solubilization and purification. This method of purification is advantageous, because it relies on the properties of the unfolded protein. Thus, disruptive mutations, which might result in reduced yields of soluble protein because of loss of heme binding and subsequent degradation in vivo, can be effectively studied in vitro due to the stabilization of the apoprotein in inclusion bodies. Our extensive experience with refolding apoMb (Jennings and Wright, 1993; Jennings et al., 1994) makes this single-step purification protocol especially useful.

The electrospray mass spectrum of the purified doubly labeled apoMb is presented in Fig. 1C. The data indicate complete incorporation of the labels (see figure legend) and, along with SDS–PAGE analysis, demonstrate the high purity of the sample obtained with this simple protocol. The UV-absorption and proton NMR spectra of the refolded and reconstituted carbon monoxide complex of the recombinant material are identical to those of the natural sperm whale protein (data not shown). The reconstituted carbon monoxide complex of recombinant Mb was used for all NMR experiments described here.

Sequence-specific NMR chemical shift assignments of MbCO were obtained by comparison of the intrarresidue NH to C^{α}/C^{β} connectivities observed in HNCA and CBCANH spectra with the interresidue NH to C^{α}/C^{β} connectivities relayed through the intervening carbonyl carbon in the HN(CO)CA and CBCA(CO)NH experiments. A series of such connectivities for residues Asp^{122} to Met^{131} is presented in Fig. 2. The sensitivity of the CBCANH experiment was significantly lower than that of the CBCA(CO)NH experiment, as expected due to the longer period during which C^{α} magnetization is transverse in the former. Thus, although most residues exhibited intraresidue cross peaks to C^{α} and C^{β}, relayed through the $^{1}J_{C\alpha N}$ coupling in the CBCANH experiment, very few residues exhibited corresponding interresidue cross peaks, relayed through the slightly smaller $^{2}J_{C\alpha N}$ coupling. Nevertheless, the spectra were of sufficient quality to allow assignment of amide ^{15}N and ^{1}H resonances for 143 residues (97% of the backbone secondary amides), C^{α} resonances for 148 (97% of all) residues, and C^{β} resonances for 128 (90% of the non-glycine) residues. Assignments of $^{1}H^{N}$, $^{15}N^{H}$, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ resonances have been deposited in the BioMagRes Data Bank (University of Wisconsin, Madison, WI). Previously, amide $^{1}H^{N}$ and $^{15}N^{H}$ assignments have been determined based on 3D ^{15}N-correlated

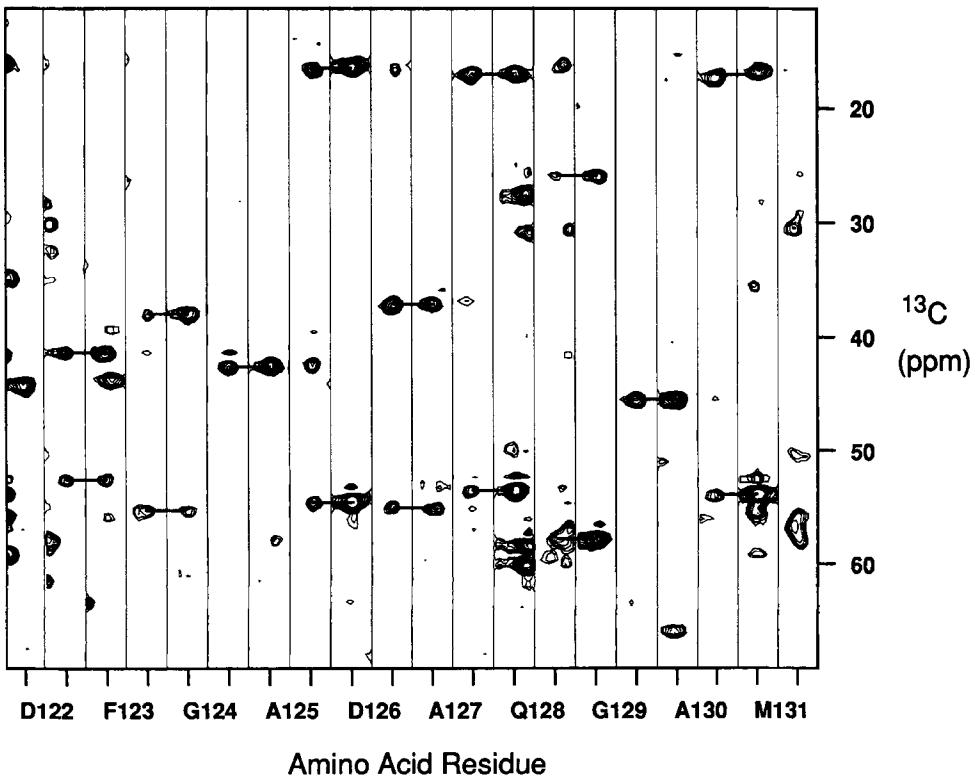

Fig. 2. Strips taken from the CBCA(CO)NH (left strip) and CBCANH (right strip) spectra at the amide ^{15}N and ^{1}H chemical shifts of residues 122–131 in MbCO. Sequential connectivities are indicated as horizontal lines.

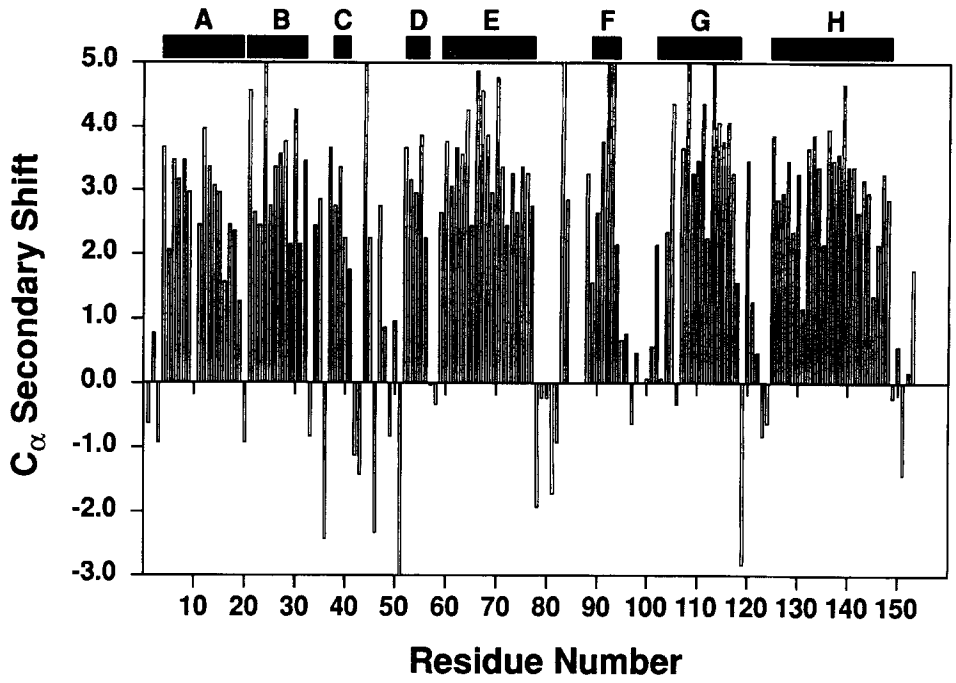

Fig. 3. C^α secondary shifts plotted as a function of residue number for MbCO. The positions of the eight α-helices predicted by the chemical shift index method (see text for details) are indicated at the top of the figure by grey boxes. The secondary shifts for off-scale residues are 7.0 (His^{24}), −4.03 (Thr^{51}), 5.37 (Glu^{83}), 9.57 (Ser^{92}), 6.27 (His^{93}), and 5.87 (Ser^{108}).

experiments (Theriault et al., 1994). Using triple-resonance experiments, we have confirmed the reported amide assignments for all residues except Glu^{85} and Leu^{86}, which were not observed in our spectra, and we have identified the amide assignments for two additional residues (Glu^{105} and Gly^{121}), which were not observed previously.

The C^α chemical shift is a reliable predictor of secondary structure in proteins. In particular, Wishart et al. (1991) have found that downfield shifts relative to reference or 'random coil' values indicate the presence of α-helix, while upfield shifts indicate the presence of β-sheet. The C^α secondary shifts (observed chemical shifts minus reference chemical shifts) of MbCO are plotted in Fig. 3. These secondary shifts were used to determine the chemical shift indices for assigned C^α resonances and hence the secondary structure elements (Wishart and Sykes, 1994). This method indicates that MbCO contains eight α-helices in the following positions: residues 4–19 (helix A), residues 21–32 (helix B), residues 38–41 (helix C), residues 52–56 (helix D), residues 59–77 (helix E), residues 89–94 (helix F), residues 102–118 (helix G), and residues 125–148 (helix H). The initiation residue of helix F cannot be determined precisely, since the C^α resonances of Glu^{85}, Leu^{86} and Lys^{87} were not assigned in the present study. Otherwise, the helix positions determined by the secondary shift method are the same (within one residue) as those indicated by X-ray crystallography (Phillips et al., 1990) and by NMR structure determination (Ösapay et al., 1994). The only exception is helix B, which was found previously to terminate at residue 35 in both the NMR and X-ray structures; the negative secondary shift of Phe^{33} observed in the present study implies the termination of helix B at this point.

Conclusions

We have demonstrated a simple method for the production in *E. coli* and subsequent purification of the protease-sensitive protein apoMb. This method is likely to be applicable to a wide variety of proteins and may facilitate detailed NMR structural analyses. The CO complex of the reconstituted holoprotein was used to determine nearly complete backbone and C^β resonance assignments using 3D triple-resonance NMR techniques. Analysis of the C^α secondary shifts allowed independent prediction of the locations of the eight α-helices, thus verifying the utility of this method for the rapid identification of secondary structure elements in proteins.

Acknowledgements

We thank Linda Tennant for expert technical assistance and Drs. Mark Rance, John Cavanagh, Wolfgang Jahnke, Ad Bax, and Shohei Koide for many helpful discussions. This work was done during the tenure of an NIH postdoctoral fellowship (P.A.J.) and a research Fellowship (M.J.S.) from the American Heart Association, California Affiliate. This work was supported by Grant No. DK34909 from the National Institutes of Health.

References

Anasari, A., Jones, C.M., Henry, E.R., Hofrichter, J. and Eaton, W.A. (1993) *Biophys. J.*, **64**, 852–868.

Austin, R.H., Beeson, K.W., Eisenstein, L., Frauenfelder, H. and Gunsalus, I.C. (1975) *Biochemistry*, **14**, 5355–5373.

Bax, A. and Subramanian, S. (1986) *J. Magn. Reson.*, **67**, 565–569.

Case, D.A. and Karplus, M. (1979) *J. Mol. Biol.*, **132**, 343–368.

Chiu, M.L. (1992) Ph.D. Thesis, University of Illinois, Champaign-Urbana, IL.

Elber, R. and Karplus, M. (1990) *J. Am. Chem. Soc.*, **112**, 9161–9175.

Gibson, Q.H., Olson, J.S., McKinnie, R.E. and Rohlfs, R.J. (1986) *J. Biol. Chem.*, **261**, 10228–10239.

Grzesiek, S. and Bax, A. (1992a) *J. Magn. Reson.*, **96**, 432–440.

Grzesiek, S. and Bax, A. (1992b) *J. Am. Chem. Soc.*, **114**, 6291–6293.

Jennings, P.A. and Wright, P.E. (1993) *Science*, **262**, 892–896.

Jennings, P.A., Dyson, H.J. and Wright, P.E. (1994) In *Statistical Mechanics, Protein Structure, and Protein Substrate Interactions* (Ed., Doniach, S.), Plenum Press, New York, NY, pp. 7–18.

Kendrew, J.C. (1958) *Nature*, **181**, 662–666.

Kottalam, J. and Case, D.A. (1988) *J. Am. Chem. Soc.*, **110**, 7690–7697.

Marion, D., Ikura, M., Tschudin, R. and Bax, A. (1989) *J. Magn. Reson.*, **85**, 393–399.

Ösapay, K., Theriault, Y., Wright, P.E. and Case, D.A. (1994) *J. Mol. Biol.*, **244**, 183–197.

Patt, S.L. (1992) *J. Magn. Reson.*, **96**, 94–102.

Phillips Jr., G.N., Arduini, R.M., Springer, B.A. and Sligar, S.G. (1990) *Protein Struct. Funct. Genet.*, **7**, 358–365.

Sanger, F., Nicklen, F. and Coulson, A.R. (1979) *Proc. Natl. Acad. Sci. USA*, **74**, 5460–5463.

Springer, B.A. and Sligar, S.G. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8961–8965.

Theriault, Y., Pochapsky, T.C., Chiu, M., Sligar, S. and Wright, P.E. (1994) *J. Biomol. NMR*, **4**, 491–504.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.

Wishart, D.S. and Sykes, B.D. (1994) *J. Biomol. NMR*, **4**, 171–180.