Overexpression, Purification, and Biophysical Characterization of the Heterodimerization Domain of the Core-binding Factor β Subunit*

(Received for publication, August 6, 1997, and in revised form, November 17, 1997)

Xuemei Huang, Barbara E. Crute‡§, Chaohong Sun§, Yen-Yee Tang‡, John J. Kelley III, Amy F. Lewis‡, Karli L. Hartman‡, Thomas M. Lau‡, Nancy A. Speck‡*, and John H. Bushweller**

From the Department of Chemistry, Dartmouth College, Hanover, New Hampshire 03755, the *Department of Biochemistry, Dartmouth Medical School, Hanover, New Hampshire 03755, and the ‡Department of Biochemistry, University of New Hampshire, Durham, New Hampshire 03824

Core-binding factors (CBF) are heteromeric transcription factors essential for several developmental processes, including hematopoiesis. CBFs contain a DNA-binding CBFα subunit and a non-DNA binding CBFβ subunit that increases the affinity of CBFα for DNA. We have developed a procedure for overexpressing and purifying full-length CBFβ as well as a truncated form containing the N-terminal 141 amino acids using a novel glutaredoxin fusion expression system. Substantial quantities of the CBFβ proteins can be produced in this manner allowing for their biophysical characterization. We show that the full-length and truncated forms of CBFβ bind to a CBFα-DNA complex with very similar affinities. Sedimentation equilibrium measurements show these proteins to be monomeric. Circular dichroism spectroscopy demonstrates that CBFβ is a mixed α/β protein and NMR spectroscopy shows that the truncated and full-length proteins are structurally similar and suitable for structure determination by NMR spectroscopy.

The core-binding factor β subunit (CBFβ)1 is the non-DNA binding subunit of the heterodimeric transcription factor complex called core binding factor (CBF) (1, 2). CBF binds to target DNA sites with the consensus sequence PyGPyGGT (3–6). CBF subunits are encoded by four genes in mammals. CBFA1, CBFA2 (AML1), and CBFA3 encode DNA-binding CBFα subunits, and CBFB encodes the common CBFβ subunit (1, 2, 7–12). The CBFα subunit can bind DNA in the absence of CBFβ in vitro but with 5–10-fold lower affinity than the CBFα/β complex (1, 2). CBFβ modulates the affinity of the CBFα subunit for DNA without establishing additional contacts on the DNA or changing the magnitude of DNA bending (2, 13).

The amino acid sequence of CBFβ yields few clues to its tertiary structure or the mechanism by which it modulates the affinity of the CBFα subunit for DNA. The heterodimerization domain in CBFβ has been localized to its N-terminal 135 amino acids, which corresponds to a region of significant homology between CBFβ and its two Drosophila homologs, Brother (Bro) and Big Brother (Bgb) (13, 14). A truncated CBFβ protein containing amino acids 1–141 (CBFβ(141)), which includes the region of homology to Bro and Bgb, stably associates with the CBFα subunit in vitro (1, 14, 15). Further truncation of the C terminus to amino acid 133 results in a protein that weakly associates with CBFα, and C-terminus truncation to amino acid 117 disrupts stable heterodimerization with CBFα (1, 15).

The CBFβ subunit is essential for the in vivo function of at least one of the CBFα subunits that were encoded by the CBFA2 gene (also known as the acute myeloid leukemia 1 or AML1 gene). Homozygous disruption of either the Cbfa2 or the Cbfβ genes in mice results in identical phenotypes: midgestation embryonic lethality accompanied by extensive hemorrhaging and a profound block at the fetal liver stage of hematopoiesis (15–19). In humans, chromosomal rearrangements that disrupt the CBFA2 and CBFB genes are associated with a variety of leukemias, including de novo acute myeloid leukemias (t(8;21)(q22;q22), inv(16)(p13;q22), (16;16), and del(16)(q22), acute lymphocytic leukemias (t(12;21)(p13;q22), acute lymphocytic leukemias (t(12;21)(p13;q22), and therapy-related leukemias and myelodysplasias t(3;21)(q26;q22) (10–11, 20–22). All of these translocations result in the synthesis of chimeric proteins, two of which have been directly demonstrated to block CBF function in a transdominant manner (23, 24).

The primary structures of CBFβ and its Drosophila homologs are not similar to those of any other proteins, and the mechanism by which CBFβ stabilizes the CBFα-DNA complex is unusual. The CBFβ subunit is an essential component of the CBF complex and is mutated in a significant percentage of human leukemias, making it both an interesting and important target for biophysical and structural analyses. In this study, we describe a novel system for expressing the CBFβ subunit in bacteria and a purification protocol with which we can obtain large amounts of homogeneous CBFβ protein. We confirm that a truncated form of CBFβ, CBFβ(141), demonstrated previously to contain an intact heterodimerization domain (1), binds to CBFα with an affinity very similar to that of the full-length CBFβ(187) protein. The isolated heterodimerization domain in CBFβ(141) also assumes a folded structure.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† These authors contributed equally to this work.

‡ A Leukemia Society of America Scholar. Supported by U. S. Public Health Service Grants R01 CA/GM 58343 and R01 CA75611. To whom correspondence should be addressed. Tel.: 603-650-1159; Fax: 603-650-1128.

§ Supported by U. S. Public Health Service Grants K02 AI01481 and R29 AI39536 from the NIAID, National Institutes of Health. To whom correspondence should be addressed. Tel.: 603-646-1567; Fax: 603-646-1567.

** Supported by U. S. Public Health Service Grants K02 AI01481 and R29 AI39536 from the NIAID, National Institutes of Health. To whom correspondence should be addressed. Tel.: 603-650-1159; Fax: 603-650-1128.

The abbreviations used are: CBFβ, core-binding factor β subunit; CBFβ(141), the 143-amino acid protein containing residues Gly-Ser followed by the N-terminal 141 amino acids of murine CBFβ; CBFβ(187), the 189-amino acid protein containing residues Gly-Ser followed by amino acids encoding the 187-amino acid isoform of CBFβ; DTT, dithiothreitol; Grx, E. coli glutaredoxin-1; HSQC, heteronuclear single quantum coherence; NH, protein backbone amide NH; PAGE, polyacrylamide gel electrophoresis; MALDI, matrix-assisted laser desorption ionization.

2480 This paper is available on line at http://www.jbc.org
Purification and Characterization of CBFβ

indistinguishable from that which it assumes in the context of the full-length CBFβ(187) protein. We also demonstrate that both the full-length CBFβ(187) and truncated CBFβ(141) proteins are monomeric and contain a mixture of α-helical and β-strand secondary structural elements.

EXPERIMENTAL PROCEDURES

**Plasmid Construction**—The pGRXCBF141 plasmid encoding the glutaredoxin-CBFβ(141) fusion protein was constructed in the following manner. DNA sequence corresponding to Escherichia coli glutaredoxin(1–187) was one-reading frame reaction amplified and pooled, and the pMG524-GRX (25) containing the following primers: 5′-CGGATCCCTGGCAC-3′ (GRX) and 5′-GGGAGATCCTTAAAAGGA-3′ (AS-GRX) was used. The sense primer (GRX) contains the recognition site for EcoRI at its 5′ end. The GRX expression cassette contains a restriction site for BamHI, followed by sequences encoding a cleavage site for the protease enterokinase (DDDSDK). DNA encoding CBFβ(141) was amplified from the plasmid CBFβ(pG215) (2) using the primers 5′-AGCGGATCCATCCGC-3′ and 5′-GGCCGACTTCTACGTCAGTTGTGCTATAATGCTATCCTC-3′ (AS-Bam). A restriction site for BamHI was included at the 5′ end of the sense primer (S-Bam141), and a HindIII site followed by a translational stop codon was added to the HindIII site of the antisense primer (AS-Bam141). The polymerase chain reaction amplified fragments were digested with the appropriate restriction enzymes and subcloned between the EcoRI and HindIII sites of pMG524. The resultant plasmid was transformed into the bacterial strain N99cl for the purpose of DNA characterization and sequencing and into strain N4830 or AR858 for protein expression.

The pGRXCBF187 plasmid encoding the glutaredoxin protein fused to CBFβ(1–187) was generated by replacing the BamHI-HindIII fragment in pGRXCBF141 that encodes CBFβ(141) with a BamHI-HindIII fragment containing the open reading frame of CBFβ(1–187), which was excised from a previously described plasmid containing CBFβ(1–187) subcloned into pBluescript SK+ (2). A cleavage site for Factor Xa was subsequently introduced into this plasmid. Complementary oligonucleotides encoding a Factor Xa cleavage site flanked by 5′-GATC ATGCACTGAGGTCG-3′ and 5′-GATCCACGACCTTGCATG-3′ were annealed, phosphorylated, and subcloned into the BamHI site, between the enterokinase cleavage site and the open reading frame of CBFβ(187). The insert in all plasmids was confirmed by sequencing. These plasmids are available upon request to the corresponding authors.

**Expression of CBFβ(141) and CBFβ(187)**—*E. coli* AR858 were transformed with the pGRXCBF141 plasmid, and a single colony was used to inoculate 10 ml of terrific broth (Sigma) containing 100 μg/ml ampicillin. The culture was used to inoculate 1 liter of terrific broth (100 μg/ml ampicillin) and grown at 37 °C with shaking at 200 rpm. After overnight shaking at 200 rpm at 37 °C, the culture was used to inoculate 1 liter of terrific broth containing 100 μg/ml ampicillin. After overnight shaking at 200 rpm at 29 °C, the bacterial cell pellet was washed with 0.8 ml/min. The active fractions were again identified by SDS-PAGE, pooled, and concentrated to 4 ml by ultrafiltration. The concentrated protein was loaded onto a 2.5 × 10-cm column of Q-Sepharose S-100 in 25 mM potassium phosphate (pH 6.5), 0.1 mM EDTA, 0.1% NaN₃, and eluted at 0.8 ml/min. The CBFβ(187) fractions were identified by SDS-PAGE, pooled, and concentrated to 4 ml by ultrafiltration. The concentrated protein was loaded onto a 2.5 × 113-cm column of Sephacryl S-100 (Pharmacia) in 25 mM Tris-Cl (pH 7.5), 1 mM EDTA, 75 mM NaCl, and eluted with a 500-ml linear gradient of 25–500 mM NaCl in the same buffer at 0.8 ml/min. The final purified CBFβ was analyzed by SDS-PAGE and stained with Coomassie blue.

**Purification of CBFβ(187)**—All operations were carried out at 4 °C. The CBFβ(187) fusion protein was purified in the same manner as described above for the Grx-CBFβ(141) fusion protein. Following fractionation on DEAE-Sephaloc and Sephacryl S-100, the resulting homogeneous Grx-CBFβ(187) fusion protein was cleaved by treatment with 3.5 units of Factor Xa (Pharmacia) in 25 mM Tris-Cl (pH 7.5), 1 mM EDTA, 75 mM NaCl, 1 mM CaCl₂, and 0.1% DTT. Secondary cleavage at another Factor Xa-sensitive site was minimized by carrying out the reaction for 6 h at 22 °C. The cleavage reaction was halted by addition of 4 mM Pefabloc with a subsequent 40-min incubation at 37 °C. The cleaved protein was exchanged into a lower ionic strength buffer (25 mM NaCl) by ultrafiltration, loaded onto a 2.5 × 10-cm column of Q-Sepharose in 25 mM Tris-Cl (pH 7.5), 1 mM EDTA, 25 mM NaCl, 1 mM DTT, and eluted with a 500-ml linear gradient from 25 to 500 mM NaCl in the same buffer at 0.8 ml/min. The CBFβ(187) fractions were identified by SDS-PAGE, pooled, and concentrated to 4 ml by ultrafiltration. The concentrated protein was loaded onto a 2.5 × 113-cm column of Sephacryl S-100 (Pharmacia) in 25 mM Tris (pH 7.5), 1 mM EDTA, 1 mM DTT, 75 mM NaCl and eluted with the same buffer to separate the full-length protein from the product of the secondary cleavage reaction.

**SDS-Polyacrylamide and Isoelectric Focusing Gel Electrophoresis**—SDS-polyacrylamide gel electrophoresis was carried out using 15% Ready Gels from Bio-Rad with Coomassie staining as described by the manufacturer. Isoelectric focusing gel electrophoresis was carried out on a Pharmacia PhastSystem using IEF 3–9 gels, and Coomassie staining was performed as described by the manufacturer.

**The Molar Extinction Coefficient**—The molar extinction coefficients for the CBF(1141) and CBF(187) proteins were calculated on the basis of the tryptophan, tyrosine, and cystine content of the proteins as described by Pace et al. (27). A value of 18,500 M⁻¹ cm⁻¹ was calculated for both proteins.

**MALDI Mass Spectrometry**—MALDI mass spectrometry was carried out by Dr. Gary Siuzdak at the mass spectrometry laboratory at the Beckman Center for Chemical Sciences, Scripps Research Institute. Samples were desalted into 20 mM ammonium acetate (pH 7), 0.5 mM DTT prior to analysis.

**N-terminal Sequencing**—To confirm the site of cleavage by enterokinase, N-terminal sequencing of the first five residues was carried out at the Biotechnology Resource Laboratory located at Yale University. A sequence of Xaa-Xaa-Met-Pro-Arg was obtained showing that cleavage had occurred at the correct site; however, the sequence of the first two amino acids was not determined with certainty.

**Measurement of Dissociation Constants**—The affinities of CBFβ(141) and CBFβ(187) for CBFβ(141) and CBFβ(187) proteins were measured by electrophoretic mobility shift assay. The source of CBFβ was an 41–214-amino acid fragment from the murine CBFβ protein encompassing the DNA-binding Runt domain, expressed in bacteria and purified as described previously (28). Basically, binding conditions were chosen such that the concentration of DNA was 10-fold above the Kᵣ of the CBFβ(2) Runt domain for the DNA site (4 × 10⁻⁸ M DNA), and the concentration of active CBFβ(2) Runt domain was 10-fold below the Kᵣ of CBFβ for the CBFβ(2)-DNA complex ([5 × 10⁻¹⁰ M CBFβ(2)]. A fixed amount of DNA and the Runt domain, along with various amounts of CBFβ protein, were used in binding reactions, and bands were resolved by electrophoretic mobility shift assay. Assay conditions and methods for quantitation were the same as described previously (15). The data are plotted as the percentages of CBFβ(2):DNA complex versus the concentration of CBFβ, and the Kᵣ is defined as the concentration of CBFβ at 50% saturation (Kaleidagraph, Synergy Software). 100% ternary complex was defined as the point of saturation, and 0%
ternary complex is taken from the background at the position in the absence of added CBFβ.

Sedimentation Equilibrium—High speed (29) sedimentation equilibrium experiments were conducted at 20 °C in a Beckman XL-1 ultracentrifuge at rotor speeds of 15,000, 20,000, 25,000, and 30,000 rpm (CBFβ(141)) or 10,000, 15,000, 20,000, and 25,000 rpm (CBFβ(187)) using absorbance detection. For each sample, three cell loading concentrations were examined, from 0.25 to ~1 mg/ml in 25 mM Tris-Cl (pH 7.5), 150 mM KCl, 1 mM EDTA. Data were fit using NONLIN (30). Partial specific volumes and solvent density (1.00960 g/ml) were calculated as described (31). For CBFβ(141), the calculated partial specific volume (0.7225 ml/g) was adjusted for the peptide charge (~3 at pH 7.5) by decreasing the molar volume 25 ml/mol charge to 0.7185 ml/g. An additional uncertainty of 0.004 ml/g in the partial specific volume was included when calculating the molecular weights (32, 33). For CBFβ(187), the calculated partial specific volume (0.7161 ml/g) also was reduced for its expected charge (~4 at pH 7.5) by 0.004 ml/g, and the same additional uncertainty was included in the molecular weight calculations.

Circular Dichroism Spectroscopy—Circular dichroism spectra were collected at 20 °C on a Jasco 715 spectrometer calibrated using 10-camphorsulfonic acid. Mean amide Δε values were calculated using the known protein sequence. The protein solutions were dialyzed extensively against 25 mM potassium phosphate (pH 7.5), 0.1 mM EDTA, 0.1% sodium azide, 5 mM DTT. 15N-1H HSQC spectra were recorded using the gradient sensitivity-enhanced HSQC sequence (36). The number of complex points and acquisition times for these experiments were 15N (F1), 128 points, 79.3 ms and 1H (F2), 1024 points, 157 ms.

For measurement of 15N T1 and T2 values, the pulse sequences of Farrow et al. (37) optimized for minimal saturation of water were used. A recycle delay of 1.0 s was used between acquisitions to ensure sufficient recovery of NH magnetization (38). 15N T1 values were measured from spectra recorded with eight different durations of the delay: 44.4, 111.1, 222.2, 333.2, 555.4, 777.5, 999.7, and 1444.0 ms for CBFβ(141) and 44.4, 111.1, 222.2, 333.2, 444.4, 555.4, 666.5, 888.6, and 1333.0 ms for CBFβ(187). 15N T2 values were measured from spectra recorded with delays corresponding to the duration of the CPMG sequence: 16.3, 32.6, 48.9, 65.2, 81.4, 97.7, 130.3, and 162.9 ms for CBFβ(141) and 16.3, 32.6, 48.9, 65.2, 81.4, 97.7, and 114.0 ms for CBFβ(187).

To generate pure absorptive 2D line shapes, the N- and P-type signals recorded for each t1 point were stored separately to carry out the necessary addition and subtraction of free induction decay and 90 ° phase correction as described previously (36). The necessary data manipulations were carried out using software written in-house. All other data processing was carried out using the program PROSA (39).

The intensities of the peaks in the two-dimensional spectra were analyzed by measuring the peak heights using the integration routine in the program XEASY (40). The uncertainties in the measured peak intensities were set equal to the root-mean-square base-line noise of the spectra. The T1 and T2 values were determined by fitting the measured peak heights to the following two-parameter function:

$$I(t) = I_0 \exp \left(-\frac{t}{T_1/2}\right)$$

(Eq. 1)

in which I(t) is the intensity at time t and I0 is the intensity at time zero. T1 and T2 values and uncertainties were determined by nonlinear least squares fitting of the experimental data points to the monoexponential decay given in Equation 1 using the Levenberg-Marquardt algorithm (41) to minimize the $\chi^2$ goodness of fit parameter. The goodness-of-fit of the data to Equation 1 was evaluated by comparison of the calculated $\chi^2$ value with tabulated values of $\chi^2$ at 95% confidence level as described previously (42). An initial estimate of the correlation time for CBFβ(141) and CBFβ(187) were obtained from a 10% trimmed mean of the T1/T2 ratio (43) for those backbone amides with T1 and T2 values deemed adequate by the goodness-of-fit analysis described above.

RESULTS

Construction of Glutaredoxin Fusion Protein Vector—The biophysical and structural studies we wish to carry out to characterize CBFβ require large quantities of homogeneous protein. Attempts to overexpress CBFβ in a number of different vectors including glutathione S-transferase fusion vectors met with limited success. To overcome these difficulties and provide the milligram quantities of CBFβ necessary for biophysical and structural studies, we have engineered a novel fusion vector based on E. coli glutaredoxin. Based on the high level of overexpression obtained for E. coli glutaredoxin (25), we reasoned that a fusion vector with CBFβ fused to the C-terminal end of glutaredoxin should also give very high levels of expression. Fig. 1A illustrates the construct we have employed in which the coding sequence for E. coli glutaredoxin-1 has been fused to that of CBFβ(141) via a linker containing the DDDDK recognition sequence for enterokinase. The DDDDK sequence permits the release of CBFβ(141) from the fusion protein via proteolysis with enterokinase. Two additional amino acids are retained at the N-terminal end (Gly-Ser) due to the BamHI
Restriction site used for subcloning purposes. A similar fusion vector has been constructed with the related protein thioredoxin (44); however, this vector employs tryptophan for induction, making it unsuitable for labeling proteins with $^{13}$C and $^{15}$N for NMR spectroscopy. Our Grx fusion vector employs a heat-inducible induction based on the heat stability of a mutant $\lambda$ repressor (25), thus avoiding these difficulties.

For reasons described below, it was necessary to replace the enterokinase site in the vector utilized for expression of CBF$\beta$ (141). This alternate vector was constructed by subcloning a double stranded oligonucleotide encoding a Factor Xa cleavage site (IEGR) between the cleavage site for enterokinase and the CBF$\beta$ (141) open reading frame (Fig. 1B).

Purification of CBF$\beta$ (141) and CBF$\beta$ (187)—Utilizing the pGRXCBF$\beta$ (141) fusion vector, we obtain high levels of expression (Fig. 2, lane 3) of a protein that is virtually completely soluble (Fig. 2, lanes 4 and 5). Fig. 2 illustrates a typical purification of CBF$\beta$ (141) starting with cells from a 1-liter bacterial culture. After lysis of the cells and clarification by centrifugation, the Grx-CBF$\beta$ (141) fusion protein can be purified close to homogeneity in two steps via ion-exchange chromatography on Q-Sepharose and size exclusion chromatography on Sephacryl S-100 (Fig. 2, lanes 6 and 7). The fusion protein-containing fractions were readily identified using the 2-hydroxyethyl disulfide assay specific for glutaredoxin (26), pooled, and concentrated by ultrafiltration.

Treatment of the Grx-CBF$\beta$ (141) fusion protein with enterokinase and subsequent ion-exchange chromatography on Q-Sepharose (Fig. 2, lanes 8 and 9) yields significant quantities of homogeneous CBF$\beta$ (141) protein (typically 20 mg from a 1-liter bacterial culture) as judged by SDS-PAGE as well as by isoelectric focusing gel electrophoresis. The conditions employed for the enterokinase cleavage of the Grx-CBF$\beta$ (141) fusion protein were optimized to obtain the highest yield of protein and utilize the minimum quantity of the somewhat costly enterokinase. Preliminary circular dichroism spectra of CBF$\beta$ (141) versus temperature showed that the protein begins to unfold above 30 °C (data not shown), so the temperature employed for the cleavage reaction was lowered from 37 to 30 °C, resulting in a substantial improvement in yield but requiring a longer reaction time. Optimization of the amount of enterokinase resulted in a final ratio of 90 units of enterokinase/µg of fusion protein/ml resulting in a total of 5000 units of enterokinase being employed for a typical purification from a 1-liter bacterial culture.

To confirm the N-terminal sequence and size of the resulting protein, both N-terminal sequencing and MALDI mass spectrometry were carried out on CBF$\beta$ (141). N-terminal sequencing and MALDI mass spectrometry yielded a molecular mass of 16.73 kDa, in good agreement with the calculated value of 16.73 kDa for the 143-amino acid protein.

Overexpression and purification of CBF$\beta$ (187) was initially attempted utilizing the same construct used for the CBF$\beta$ (141) protein, namely Grx fused to CBF$\beta$ (187) with a linker containing an enterokinase site. Again, high levels of expression of the resulting fusion protein were obtained, and the Grx-CBF$\beta$ (187) protein could readily be purified utilizing the same two-step protocol employed for CBF$\beta$ (141). However, unlike the CBF$\beta$ (141) fusion protein, treatment of the CBF$\beta$ (187) fusion protein with enterokinase resulted in proteolytic degradation of CBF$\beta$ (187). To characterize the site of this cleavage, MALDI mass spectrometry of the cleaved protein was carried out. Assuming an intact N terminus based on the behavior of the CBF$\beta$ (141) protein in the enterokinase cleavage reaction, the observed molecular masses (17.4 and 18.6 kDa) indicated cleavage after residues 147 and 156 of CBF$\beta$ (187). The sequences of amino acids leading up through residues 147 and 156, respectively, are Phe-Glu-Glu-Ala-Arg and Glu-Phe-Glu-Asp-Arg, which apparently resemble closely enough the Asp-Asp-Asp-Lys recognition sequence for enterokinase such that cleavage also occurs at these positions.

To overcome this secondary proteolysis, a Factor Xa cleavage site was inserted between the enterokinase cleavage site and the CBF$\beta$ (187) coding sequence. Again, high levels of fusion protein are obtained and can be purified via the same two-step protocol used previously. Cleavage with Factor Xa minimizes secondary cleavage of the protein (Fig. 3); however, there continues to be some secondary cleavage (Fig. 3, lane 4) indicative of a mobile protease-accessible region in the tail of the protein. For this reason, cleavage reactions with Factor Xa cannot be allowed to go to completion, thus reducing the overall yield of CBF$\beta$ (187) obtained. Ion-exchange chromatography on Q-Sepharose is employed as for CBF$\beta$ (141) followed by size exclusion chromatography on Sephacryl S-100 to remove the residual degraded CBF$\beta$ (187) from the full-length protein. The resulting CBF$\beta$ (187) is homogeneous as judged by both SDS-PAGE (Fig. 3, lane 5) as well as isoelectric focusing gel electro-
Purification and Characterization of CBFβ

CBFβ Binding Properties of CBFβ(141) and Full-length CBFβ(187)—Previous studies have demonstrated that the first 141 amino acids of CBFβ were sufficient for interaction with CBFα (1, 15, 16). To determine whether CBFβ(141) and the full-length CBFβ(187) have similar affinities for CBFα, we performed electrophoretic mobility shift assays to measure the $K_d$ of both proteins for a CBFα-DNA complex. The source of CBFα used in these experiments consists of a 175-amino acid fragment containing the DNA binding Runt domain from the murine CBFα2 (AML1) protein, which was expressed and purified as described previously (28). The supershift of the Runt domain-DNA complex associated with binding of CBFβ has been employed to obtain equilibrium dissociation constants for the binding of CBFβ to a Runt domain-DNA complex (15). Fig. 4 shows gel shift data obtained from these measurements, and the data employed for the calculation of the $K_d$ values of (4.5 ± 0.5) × 10⁻⁸ M and (1.8 ± 0.1) × 10⁻⁸ M (measured at 4 °C) were obtained for the binding of CBFβ(141) and CBFβ(187) to a Runt domain-DNA complex, respectively. These values are very similar, demonstrating that virtually all the determinants necessary for binding to the Runt domain reside in the N-terminal 141 amino acids. This agrees well with the observed homology between mammalian CBFβ and its Drosophila homologs where there is a high degree of homology in the N-terminal 157 amino acids but very low homology in the C-terminal region of the proteins. The exact function of the C-terminal region is unknown at this time.

Sedimentation Equilibrium—To assess the solution oligomerization state of CBFβ(141) and CBFβ(187), sedimentation equilibrium measurements have been carried out. Both CBFβ(141) and CBFβ(187) sediment as single ideal species with apparent molecular weights of 16,600 ± 1200 (root mean square of the fit = 0.008 Å) and 25,700 ± 1000 (root mean square of the fit = 0.0011 Å). These values are in good agreement with the calculated molecular masses of the proteins (16.7 and 22.1 kDa); thus we conclude that both proteins are monomeric under these conditions. Because the two proteins do not aggregate to any observable extent up to the concentrations employed for the sedimentation equilibrium measurements, both proteins should be good candidates for structure determination by NMR spectroscopy.

CD Analysis—The far ultraviolet CD spectra for the two proteins have been analyzed to obtain estimates of the percentages of various secondary structures (Fig. 5) using the variable selection procedure of Johnson (34) and Manavalan and Johnson (35). Both CBFβ(141) and CBFβ(187) are mixed α/β proteins with CBFβ(141) having 21% α-helix (30 residues) and 27% β-sheet (39 residues) content, whereas CBFβ(187) displays 34% α-helix (~64 residues) and 21% β-sheet (~39 residues) content. The remainder of the residues in both proteins are partitioned between turn and random coil conformations with both proteins having a similar number of residues in random coil conformations, but CBFβ(187) having a larger number of residues in turn conformations. Based on the similar numbers of residues in β-sheet conformations, the similar biochemical activity, and the very similar NMR spectra of the two proteins (see below), it appears that the N-terminal 141 amino acids retain the same conformation in both CBFβ(141) and the full-length protein. Based on this, the increase in the number of α-helical residues (64 for CBFβ(187) versus 30 for CBFβ(141)) must correspond to residues in the C-terminal portion of CBFβ(187). This implies that this nonconserved C-terminal tail is almost completely α-helical.

NMR Spectroscopy—15N-1H gradient sensitivity-enhanced HSQC (36) spectra of 15N-labeled CBFβ(141) and CBFβ(187) have been recorded to assess the structural similarity of the...
two proteins and to assess their suitability for structure determination by NMR spectroscopy. Fig. 6 shows the $^{15}$N-$^1$H HSQC spectra of the two proteins. Both proteins could readily be concentrated to the $1\text{mM}$ concentration necessary for NMR spectroscopy and remained soluble for extended periods. Both proteins give well dispersed spectra characteristic of mixed $\alpha/\beta$ proteins. The spectra show a striking resemblance to one another in almost all regions, providing additional confirmation of the retention of the conformation of the N-terminal 141 amino acids in both proteins. In particular, the residues with downfield NH frequencies that generally correspond to amino acids in extended conformations in $\beta$-sheets display virtual identity to one another.

To confirm that the two proteins remain monomeric at the high concentrations necessary for NMR spectroscopy, we have recorded $^{15}$N $T_1$ and $T_2$ relaxation data on both $^{15}$N-labeled samples. Assuming reasonably isotropic rotation of the protein, quite accurate estimates of the overall correlation time can be obtained from the ratio of the $^{15}$N $T_1$ and $T_2$ values (43). Fig. 7 shows the values of this ratio for 121 of the resolved NH cross-peaks observed in the HSQC spectrum of CBF$\beta$(141) and 91 of the resolved NH cross-peaks observed in the HSQC spectrum of CBF$\beta$(187). Based on these measurements, we obtain overall correlation times of $11.43 \pm 0.03$ ns for CBF$\beta$(141) and $16.9 \pm 0.4$ ns for CBF$\beta$(187). The value for CBF$\beta$(141) is in good agreement with values obtained for a number of monomeric proteins of similar size (43, 47–50), thus we conclude CBF$\beta$(141) is also monomeric at the high concentrations employed for NMR spectroscopy. As the overall correlation time of an isotropically tumbling molecule should be proportional to its volume (51, 52) and by inference its mass, the increase in the value observed for CBF$\beta$(187) over CBF$\beta$(141) is consistent with its increase in mass. We therefore conclude that CBF$\beta$(187) also remains monomeric at these high concentrations.

**DISCUSSION**

The core binding factor proteins constitute a small family of transcription factors containing a DNA-binding $\alpha$ subunit and a non-DNA-binding $\beta$ subunit (53). The Runt domain of CBF$\alpha$ is responsible for both DNA binding and heterodimerization with CBF$\beta$. CBF$\beta$ binds to the Runt domain and is believed to induce a conformational change that results in an increase in the affinity of the Runt domain for DNA of approximately 6-fold (1–2). The three-dimensional structures of CBF$\beta$ and the Runt domain are unknown, and the primary sequences do not show any homology to other DNA-binding domains or protein-protein interaction domains. Knockouts of the genes encoding CBF$\alpha$2 and CBF$\beta$ in mice are embryonic lethal and lead to a complete blockage in definitive hematopoiesis (15–19), indicating that CBF plays a critical role in hematopoietic development. Mutations in the genes encoding CBF$\alpha$2 and CBF$\beta$ are associated with a large number of leukemias (15, 23). Biophysical characterization and structure determination of these proteins and their interactions with each other as well as with DNA should greatly facilitate the development of novel therapeutic strategies to treat the leukemias associated with the variant forms of these proteins.

We have developed a procedure for the high level overexpression and purification of full-length CBF$\beta$, CBF$\beta$(187), as well as a truncated form containing only the N-terminal 141 amino acids, CBF$\beta$(141), using a novel protein fusion system employing *E. coli* glutaredoxin as the fusion partner. The use of the glutaredoxin protein as the fusion partner results in very high levels of expression of the proteins in a soluble form. Cleavage of the desired CBF$\beta$ product from the fusion protein has been effected via either an enterokinase or Factor Xa site in the fusion protein. Use of Factor Xa was necessary for the full-

---

2 B. E. Crute, X. Huang, J. J. Kelley, N. A. Speck, and J. H. Bushweller, manuscript in preparation.
length protein because of the presence of two secondary sites of cleavage for enterokinase in the C-terminal portion of CBFβ(187). These sites have been identified by MALDI mass spectrometry of the cleaved products and do show a similarity to the Asp-Asp-Asp-Asp-Lys recognition sequence of enterokinase. The use of glutaredoxin fusions for overexpression may be generally applicable to many other proteins.

We have utilized a number of different methods to characterize the biophysical properties of CBFβ(141) and CBFβ(187). Electromobility gel shift assays have been utilized to measure the binding of both proteins to demonstrate their structural similarity, and sedimentation equilibrium NMR measurements have been conducted on the two proteins to demonstrate their structural similarity and establish their suitability for structure determination by NMR spectroscopy.

**REFERENCES**

1. Ogawa, E., Inuzuka, M., Maruyama, M., Satake, M., Naito-Fujimoto M., Ito, Y., and Shigesada, K. (1993) *Virology* 194, 314–331.
2. Wang, S., Wang, Q., Crute, B. E., Melnikova, I. N., Keller, S. R., and Speck, N. A. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 5394–5398.
3. Kamachi, Y., Ogawa, E., Asano, M., Ishida, S., Murakami, Y., Satake, M., Ito, Y., and Shigesada, K. (1990) *J. Virol.* 64, 4808–4819.
4. Melnikova, I. N., Crute, B. E., Wang, S., and Speck, N. A. (1993) *J. Virol.* 67, 2408–2411.
5. Thernell, A., Hallberg, B., and Grundstrom, T. (1991) *J. Virol.* 65, 42–50.
6. Meyers, S., Downing, J. R., and Hiebert, S. W. (1993) *Mol. Cell. Biol.* 13, 6356–6365.
7. Bae, S. C., Takahashi, E., Zhang, Y. W., Ogawa, E., Shigesada, K., Namba, Y., Satake, M., and Ito, Y. (1995) *Gene (Amst.)* 159, 245–248.
8. Bae, S. C., Yamaguchi-Iwai, Y., Ogawa, E., Maruyama, M., Inuzuka, M., Kagoshima, H., Shigesada, K., Satake, M., and Ito, Y. (1993) *Oncogene* 8, 809–814.
9. Levanon, D., Negreanu, V., Bernstein, Y., Bar-Am, I., Avivi, L., and Groner, Y. (1994) *Genomics* 23, 425–432.
10. Liu, P., Tarle, S. A., Claxton, D. F., Marlton, P., Freedman, M., Siciliano, M. J., and Collins, F. S. (1993) *Science* 261, 1041–1044.
11. Miyoshi, H., Shizuku, K., Koro, T., Maseki, N, Kaneko, Y., and Ohki, M. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 10431–10434.
12. Ogawa, E., Maruyama, M., Kagoshima, H., Inuzuka, M., Lu, J., Satake, M., Shigesada, K., and Ito, Y. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 6597–6602.
13. Golling, G., Li, L. H., Pepling, M., Stiehns, M., and Gergen, J. P. (1996) *Mol. Cell. Biol.* 16, 932–942.
14. Kagoshima, H., Akamatsu, Y., Ito, Y., and Shigesada, K. (1996) *J. Biol. Chem.* 271, 33074–33082.
15. Wang, Q., Stacy, T., Miller, J. D., Lewis, A. F., Huang, X., Bushwell, J. H., Boris, J.-C., Alt, F. W., Ryan, G., Liu, P. P., Wynshaw-Boris, A., Binder, M., Marin-Padilla, M., Sharpe, A., and Speck, N. A. (1996) *Cell* 87, 697–708.
16. Niki, M., Okada, H., Takano, H., Kuno, J., Tani, K., Hibino, H., Satake, M., Ito, Y., Satake, M., and Noda, T. (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 5697–5702.
17. Okuda, T., van Deursen, J., Hiebert, S. W., Grosfeld, G., and Downing, J. R. (1996) *Cell* 84, 321–330.
18. Saraki, K., Yagi, H., Bronson, R. T., Tomimaga, K., Matsumani, T., Deguchi, K., Tan, Y., Kishimoto, T., and Komori, T. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 12359–12363.
19. Wang, Q., Stacy, T., Binder, M., Marin-Padilla, M., Sharpe, A. H., and Speck, N. A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 3444–3449.
20. Golub, T. R., Barker, G. F., Chohlander, S. K., Hiebert, S. E., Ward, D. C., Bray-Ward, P., Morgan, E., Raimondi, S. C., Rowley, J. D., and Gilliland, D. G. (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 4917–4921.
21. Nuijifora, G., Buey, C. R., Erickson, D., Dukhin, H. A., and Rowley, J. D. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 7784–7788.
22. Romana, S. P., Mauchauflle, M., Le Coniat, M., Chamukow, I., Le Paslier, D., Berger, R., and Bernad, J. F. (1995) *Blood* 85, 3602–3607.
23. Yergeau, D. A., Hetherington, C. J., Wang, Q., Zhang, P., Sharpe, A. H., Binder, M., Marin-Padilla, M., Speck, N. A., and Zhang, D.-E. (1997) *Nat. Genet.* 15, 303–306.
24. Castilla, L. H., Wijmenga, C., Wang, Q., Stacy, T., Speck, N. A., Eckhaus, M., Marin-Padilla, M., Collins, F. S., Wynshaw-Boris, A., and Liu, P. P. (1997) *Cell* 87, 687–696.
25. Bjornberg, O., and Holmgren, A. (1991) *Protein Expression Purif.* 2, 277–296.
26. Holmgren, A. (1979) *J. Biol. Chem.* 254, 3664–3671.
27. Pace, N. C., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) *Protein Sci.* 4, 2411–2425.
28. Crute, B. E., Lewis, A. F., Wu, Z., Bushwell, J. H., and Speck, N. A. (1996) *J. Biol. Chem.* 271, 26251–26260.
29. Yphantis, D. A. (1964) *Biochemistry* 3, 297–317.
30. Johnson, M. L., Correa, J. J., Yphantis, D. A., and Halvorson, H. R. (1981) *Biophys. J.* 36, 575–588.
31. Laue, T. M. (1992) *Short Column Sedimentation Equilibrium Analysis for Rapid Characterization of Macromolecules in Solution*, Technical Information Rep. 835 Beckman Instruments Inc., Palo Alto, CA.
32. Starvosnik, M. A., Blackwell, T. K., Laue, T. M., Weintraub, H., and Klevit, R. E. (1992) *Biochemistry* 31, 9891–9903.
33. Khurasko, D. P. (1989) *Biophys. Chem.* 34, 115–125.
34. Johnson, W. C. (1990) *Proteins Struct. Funct. Genet.* 23, 84–96.
35. Mavarakas, D., and Zervas, D. N. (1993) *Anal. Biochem.* 201, 619–629.
36. Farrow, N. A., Bhutani, R., Singher, A. U., Pascall, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry* 33, 5984–6003.
37. Sklenar, V., Torchia, D. A., and Bax, A. (1987) *J. Magn. Reson.* 73, 375–379.
38. Guntert, P., Dotsch, V., Wider, G., and Wuthrich, K. (1992) *J. Biomol. NMR* 2, 219–229.
39. Bartels, C., Xia, T.-H., Billeter, M., Guntert, P., and Wuthrich, K. (1995) *J. Biomol. NMR* 5, 1–10.
40. Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1986) *Numerical Recipes*, pp. 683–688, Cambridge University Press, Cambridge
42. Palmer, A. G., III, Rance, M., and Wright, P. E. (1991) *J. Am. Chem. Soc.* **113**, 4371–4380
43. Kay, L. E., Torchia, D. A., and Bax, A. (1989) *Biochemistry* **28**, 8972–8979
44. LaVallie, E. R., DiBlasio, E. A., Kvacic, S., Grant, K. L., Shendel, P. F., and McCoy, J. M. (1993) *Bio/Technology* **11**, 187–193
45. Williamson, M. P. (1990) *Biopolymers* **29**, 1423–1431
46. Liu, P. P., Hajra, A., Wijmenga, C., and Collins, F. S. (1995) *Blood* **85**, 2289–2302
47. Mandel, A. M., Akke, M., and Palmer, A. G., III (1995) *J. Mol. Biol.* **246**, 144–163
48. Clore, G. M., Drissoll, P. C., Wingfield, P. T., and Gronenborn, A. M. (1990) *Biochemistry* **29**, 7387–7401
49. Gruberger, B. L., Gronenborn, A. M., and Clore, G. M. (1993) *J. Mol. Biol.* **230**, 364–372
50. Powers, R., Clore, G. M., Stahl, S. J., Wingfield, P. T., and Gronenborn, A. (1992) *Biochemistry* **31**, 9150–9157
51. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry*, Part 2, p. 563, Freeman, San Francisco
52. Venable, R. M., and Pastor, R. W. (1988) *Biopolymers* **27**, 1001–1014
53. Speck, N. A., and Stacy, T. (1995) *Crit. Rev. Eukaryotic Gene Expr.* **5**, 337–364