# Molecular insights into PEBP2/CBFβ-SMMHC associated acute leukemia revealed from the structure of PEBP2/CBFβ

Michael Goger[1,2], Vineet Gupta[1,2], Woo-Young Kim[3], Katsuya Shigesada[3], Yoshiaki Ito[3] and Milton H. Werner[2]

*[1]These authors contributed equally to this work. [2]Laboratory of Molecular Biophysics, The Rockefeller University, 1230 York Avenue, Box 42, New York, NY 10021, USA. [3]Institute for Virus Research, Kyoto University, Japan.*

**PEBP2/CBF is a heterodimeric transcription factor essential for genetic regulation of hematopoiesis and osteogenesis. DNA binding by PEBP2/CBFα is accomplished by a highly conserved DNA binding domain, the Runt domain (RD), whose structure adopts an S-type immunoglobulin fold when bound to DNA. The supplementary subunit β enhances DNA binding by the RD *in vitro*, but its role in the control of gene expression has remained largely unknown *in vivo*. Chromosome 16 inversion creates a chimeric gene product fusing PEBP2/CBFβ to a portion of the smooth muscle myosin heavy chain (PEBP2/CBFβ-SMMHC) that is causally associated with the onset of acute myeloid leukemia in humans. The three-dimensional structure of PEBP2/CBFβ has been determined in solution and is shown to adopt a fold related to the β-barrel oligomer binding motif. Direct analysis of a 43.6kD ternary RD–β–DNA complex identifies the likely surface of β in contact with the RD. The structure of PEBP2/CBFβ enables a molecular understanding of the capacity of PEBP2/CBFβ-SMMHC to sequester PEBP2/CBFα in the cytoplasm and therefore provides a molecular basis for understanding leukemogenic transformation.**

The core structured domain of PEBP2/CBFβ was identified by limited proteolysis coupled to mass spectrometry to be comprised of the first 141 residues of the protein[1]. This core domain approximates the portion of β present in the PEBP2/CBFβ–SMMHC chimera[2]. The three-dimensional structure was derived from 1,853 restraints composed of 1,511 NOEs and 226 $\phi$, $\chi_1$, and $\chi_2$ dihedral angles and 116 $^3J_{NH\alpha}$ coupling constants. The structure was solved by multi-nuclear, multi-dimensional NMR employing standard techniques for assignment and structure refinement beginning with a fully extended polypeptide chain[3–5]. The coordinate precision for backbone and all heavy atoms was 0.61 Å and 1.2 Å root mean squared (r.m.s.) deviations, respectively for residues 8–140 in 25 structures. Structure quality factors were 80.7% of residues in the most favored region of the Ramachandran plot and a Prosa II Z score of -5.7 ± 0.2 (Table 1)[6–7].

The three-dimensional structure of PEBP2/CBFβ is a six-stranded β-barrel that is capped top and bottom by α- and/or $3_{10}$-helices (Fig. 1). The N-terminus of the protein contains two short helices (H1: Gln 8–Asn 14; H2: Glu 16–Leu 21) which set up a hydrophobic cluster involving four Phe residues (amino acids 12, 17, 18, 127). Following helix H1 and H2 are the six strands of anti-parallel sheet: $\beta_1$ (Glu 24–Thr 30), $\beta_2$ (Arg 52–Val 58), $\beta_3$ (Thr 62–Phe 68), $\beta_4$ (Lys 94–Ile 102), $\beta_5$ (Cys 107–Leu 116) and $\beta_6$ (Leu 119–Phe 127). Helices surround the ends of the barrel with α-helix H3 (His 37–Asp 50) and $3_{10}$-helix H4 (Arg 83–Arg 90) at the top and α-helix H5 (Glu 129–Gln 140) and the N-terminal helices at the bottom. This arrangement of the β-barrel and helices is reminiscent of the fold of an oligomer binding domain (OBD)[8], a heretofore unseen motif for a factor associated with gene regulation. A search of the protein data bank for proteins with homologous folds using the DALI[9] server identified the closest structural homolog as the C-terminal OBD of translation initiation factor 5A[10] with a Cα r.m.s. deviation of 3 Å over 51 out of 61

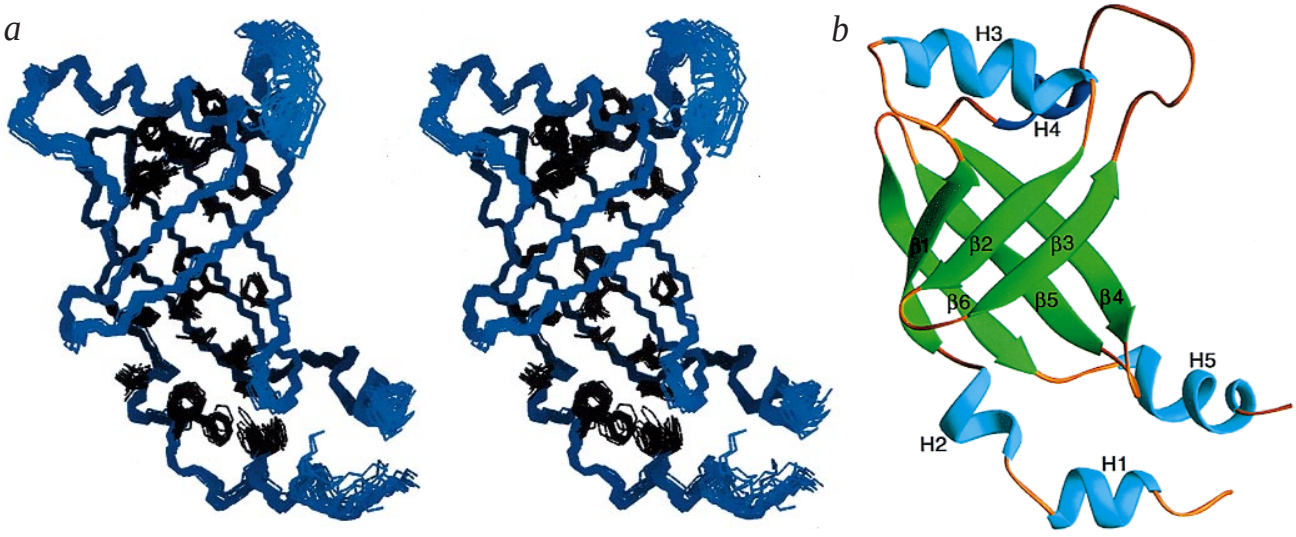


**Fig. 1** Three-dimensional structure of PEBP2/CBFβ. ***a***, Stereosuperposition of the final 25 simulated annealing structures. The backbone of the protein for residues 4–141 is shown in blue with the side chains making up the hydrophobic core of the protein shown in black (Phe 12, Phe 17, Phe 18, Leu 21, Ile 27, Tyr 29, Phe 44, Ile 57, Leu 64, Phe 68, Tyr 85, Val 95, Leu 97, Met 101, Leu 103, Val 108, Trp 110, Ile 114, Leu 125 and Phe 127). The relative orientation of helices H1 and H2 were not well-defined by the NMR data with few NOEs observed between them. ***b***, RIBBONS[32] representation of the PEBP2/CBFβ structure summarizing the α helices (cyan), $3_{10}$-helix (blue) and anti-parallel β sheets (green).

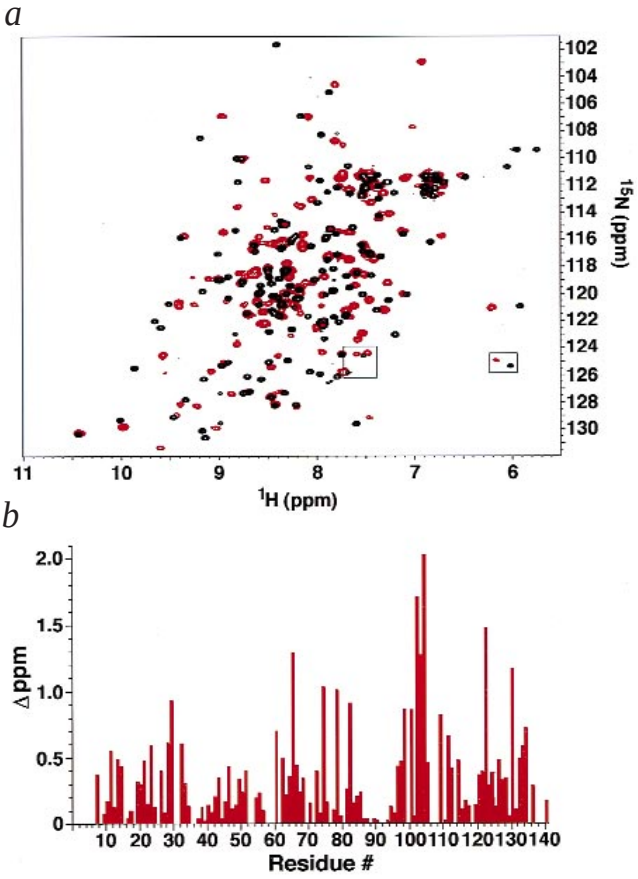


**Fig. 2** Chemical shift perturbation mapping of the PEBP2/CBFβ interaction surface for the PEBP2/CBFα Runt domain. ***a***, Superposition of $^{15}$N-$^{1}$H HSQC spectra of free (18 kDa, black) and ternary complex bound (43.6 kDa, red) PEBP2/CBFβ. The spectra were collected under identical conditions from a 1 mM sample of free or bound protein at 30 °C. Boxes indicate the Arg-Nε side chain resonances which are aliased into the spectrum from ~85 p.p.m. ***b***, Histogram of total $^{1}$H and $^{15}$N chemical shift changes per residues between free and bound NMR spectra in (*a*) (see Methods for a description of this representation). The assignments were accomplished by pattern matching the two spectra and by comparison of $^{15}$N-edited NOESY spectra of PEBP2/CBFβ in the free and bound state. Approximately 80% of the assignments were completed for the backbone amide protons and nitrogens. To identify the interaction surface for the RD–DNA complex, a threshold was chosen as ≥0.5 p.p.m. total change. The identified surface includes residues Gly 61–Gln 67, Lys 98–Trp 110 and Gln 133–Glu 135. Three additional amino acids, Gln 74, Gln 79 and Arg 83, also show significantly perturbed backbone signals in the ternary complex.

residues. The topology of PEBP2/CBFβ observed here, however, differs somewhat from the canonical OBD fold in that $\beta_4$ does not cross $\beta_3$ to fully close the barrel. In addition, the polarity of β-strands 4–6 is opposite to that observed in the canonical OBD motif[8], further distinguishing the fold of PEBP2/CBFβ from a typical OB fold. OBD motifs in biology occur in a wide variety of proteins, ranging from those secreted by bacterial pathogens such as Verotoxin 1B[11] to DNA enzymes such as staphylococcal nuclease[12]. This protein motif is also found in RNA interacting proteins such as the anticodon binding domain of yeast tRNA$^{Asp}$-synthetase[13], and the C-terminal domain of translation initiation factor 5A[10]. In nearly all of these examples the OBD motif is responsible for the binding of small molecules, either oligosaccharides or oligonucleotides, which participate in the activity of the protein. There is no evidence to suggest that PEBP2/CBF's gene activating properties *in vivo* are dependent on small molecule cofactors, although *in vitro*, DNA binding by the heterodimer is sensitive to reducing agents and NADPH[14]. The structure of PEBP2/CBFβ described herein is very similar to that of Huang *et al*. who independently identified the same core domain of the protein and determined its three-dimensional structure in solution by multi-nuclear NMR[15,16].

The only biochemical activity identified for β is its ability to enhance the DNA binding affinity of the RD *in vitro*[15,17–19]. A possible mechanism by which DNA binding affinity could be enhanced by β was suggested by the identification of a contact surface on the RD for β which placed β proximal to one of the putative DNA binding loops of the RD and thereby also proximal to the DNA[1]. The proximity of β to the DNA binding loop and DNA could either stabilize the conformation of a DNA contacting loop and/or could promote non-specific contacts between β and the DNA. To explore the origin of DNA-binding enhancement by β further, the interaction surface for a Runt domain–DNA complex was identified by direct analysis of $^{15}$N and $^{1}$H chemical shift changes between free β and β bound in a 43.6 kDa ternary complex with the RD and DNA (Fig. 2). The region of β in contact with the RD–DNA complex forms a convex surface comprised of residues Gly 61–Gln 67, Lys 98–Val 106 and Gln 133–Glu 135 (Figs 2,3). Gln 74, Gln 79 and Arg 83, which reside in the loop between $\beta_3$ and H4, display perturbed backbone resonances in the ternary complex even though they are located distantly from the main contact surface for the RD (Fig. 3). Closer examination of the PEBP2/CBFβ structure and the electrostatic surface also reveals a cluster of arginine residues on one end of the barrel opposite the N- and C-terminal helices (Fig. 4). These observations and the completion of the tertiary structure of β permit an expansion of the original hypothesis regarding the effect of β on RD DNA binding[1]. If β were oriented in the heterodimer such that helices H3/H4 were down and helices H1, H2 and H5 were up, this would place Gln 74, Gln 79 and the arginine cluster on the same face of the heterodimer as the DNA binding loops of the RD (Fig. 3). Glutamine and arginine are frequent residues utilized by transcription factors to form contacts with the DNA phosphodiester backbone. If some of the Gln and/or Arg side chains were now buried at the heterodimer/DNA interface, it would be expected that those residues should display perturbed signals in the NMR spectrum. Thus, it is possible that some of the residues that display perturbed signals in the ternary complex and are away from the main contact region for the RD may be located near the protein–DNA interface (Fig. 3). From this experiment alone, we cannot rule out the possibility that Gln 74, Gln 79 and Arg 83 are being buried at the heterodimer interface in lieu of the DNA interface. The insolubility of the RD in the absence of the DNA precludes the investigation of the heterodimer interface alone.

OB domains frequently bind their ligands using a group of extended loops on one side of the barrel, with one of these loops proximal to the amino end of the capping helices[8]. The extended loop between $\beta_3$ and $\beta_4$ (residues Ala 71–Thr 80) may be equivalent to one of the oligonucleotide binding loops identified in other proteins using this motif[8]. The hypothesis that residues in this loop may be positioned near the DNA in the ternary complex is generally consistent with the region of the motif that commonly binds oligonucleotides, although the

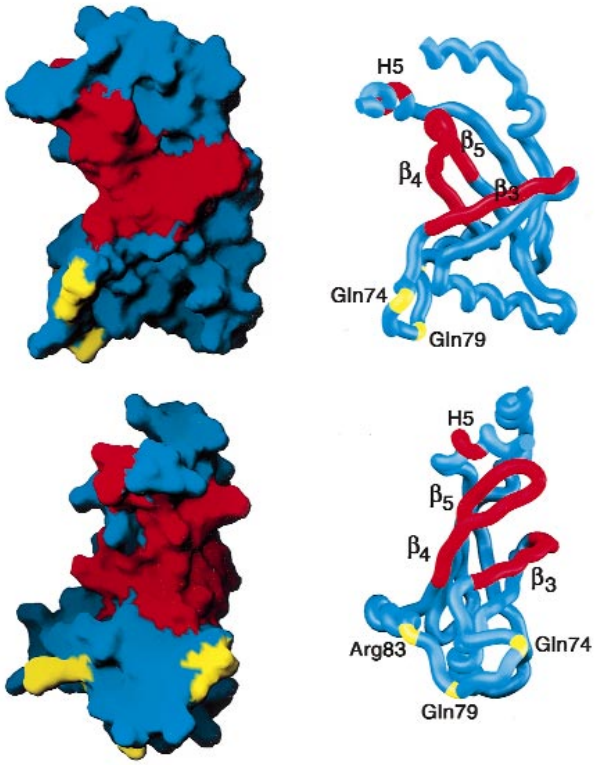


**Fig. 3** Binding surface of PEBP2/CBFβ for the Runt domain. Two views of the molecular surface are shown left. The bottom view is rotated 90° relative to the top view. Both views are 180° rotated top-to-bottom relative to Fig. 1. The proposed interaction surface for the Runt domain is shown in red. The specific segments identified to form the Runt domain binding surface are: $\beta_3$ (residues Gly 61–Gln 67), portions of $\beta_4$–$\beta_5$ (residues Lys 98–Trp 110) and part of helix H5 (residues Gln 133–Glu 135). Shown in yellow are Gln 74, Gln 79 and Arg 83, which display significant changes in chemical shift and are implicated to be proximal to or possibly contacting the DNA in the ternary complex (see text).

details of such an interaction in this case are expected to be different relative to the conserved mode of interaction observed for the canonical OB fold[8]. PEBP2/CBFβ does not bind DNA on its own and does not appear to significantly expand the footprint of the RD in ethylation interference experiments[18,20]. Our hypothesis is not inconsistent with these observations as only a small segment of the β subunit is suggested to be located proximal to the DNA.

Truncation of the C-terminus of β to Glu135 (Δ135) does not disrupt stable heterodimerization as measured by electrophoretic mobility supershift experiments[19]. This truncated β is also competent at transcriptional transactivation[21], suggesting that the structure of the core domain of PEBP2/CBFβ reported herein contains the full functional complement of protein activity. Truncations beyond Glu 135, however, result in a complete loss of heterodimerization and transcriptional stimulation[19]. These observations are consistent with the map of the heterodimerization surface of β which suggests involvement of helix H5 in the RD interaction up to Glu 135. Helix H1 had previously been implicated to form part of the interaction surface for the RD by deletion analysis[20,22]. Truncation of this helix would be expected to destabilize the protein conformation due to disruption of the phenylalanine cluster formed by residues 12, 17, 18 and 127. In turn, this destabilization would also be expected to impact the RD interaction surface that is present in this region (Fig. 2).

The structure and RD interaction surface of PEBP2/CBFβ also provides a molecular explanation for the competency of the leukemogenic fusion protein PEBP2/CBFβ-SMMHC in heterodimerization with the RD. This fusion product is causally linked with the onset of acute myeloid leukemia resulting from chromosome 16 inversion[2]. The chimeric protein produced contains the first 165 amino acids of PEBP2/CBFβ fused in frame with a C-terminal portion of the smooth muscle myosin heavy chain (SMMHC)[2]. Recent evidence indicates that this fusion product sequesters the α subunit in the cytoplasm and associates with cytoskeletal structures[21,22]. Sequestration of the α subunit in the cytoplasm makes it unavailable for the activation of gene expression for genes dependent on PEBP2/CBF in hematopoiesis. The observation that the interaction surface for α is contained within a contiguous surface of β that does not extend beyond amino acid 135 explains how the fusion product accomplishes this task. The ability of PEBP2/CBFβ-SMMHC to bind α[21,22] and retain the chimeric heterodimer in the cytoplasm also explains the failure of PEBP2/CBF-dependent transcriptional activation by the chimera[21]. It appears that PEBP2/CBFβ-SMMHC heterodimerizes with PEBP2/CBFα and sequesters the protein in

**Table 1 Structural statistics[1]**

| | <SA> |
|---|---|
| R.m.s. deviations from experimental distance restraints (Å)[2] | |
| All (1,511) | 0.057 ± 0.003 |
| Sequential (\|i - j\| = 1) (377) | 0.043 ± 0.006 |
| Short range (1 < \|i - j\| ≤ 5) (266) | 0.088 ± 0.019 |
| Long range (\|i - j\|) > 5) (421) | 0.062 ± 0.004 |
| Intraresidue (355) | 0.025 ± 0.006 |
| Hydrogen bonds (92) | 0.109 ± 0.005 |
| R.m.s. deviations from experimental | |
| Dihedral restraints (°) (226) | 0.48 ± 0.11 |
| $^3J_{NH\alpha}$ coupling constants (Hz) (116) | 1.07 ± 0.03 |
| Deviations from idealized covalent geometry | |
| Bonds (Å) | 0.0053 ± 0.0003 |
| Angles (°) | 0.70 ± 0.03 |
| Impropers (°) | 0.68 ± 0.04 |
| Coordinate precision[3] | |
| Backbone (residues 8–140) | 0.61 ± 0.06 |
| All non-hydrogen atoms (residues 8–140) | 1.20 ± 0.07 |
| Quality factors[4] | |
| % residues in most favorable Ramachandran (3,450) | 80.7% |
| Prosa II Z score | -5.7 ± 0.2 |

[1]R.m.s. deviations are calculated relative to the mean coordinates <SA> for the family of 25 simulated annealing structures excluding residues 1–3.
[2]No restraints between protons separated by three bonds were utilized (~600 NOEs).
[3]The precision of the coordinates is defined as the average atomic r.m.s. difference between the 25 individual simulated annealing structures and the mean coordinates <SA> for residues 8–140.
[4]PROCHECK_NMR[6] and ProsaII[7] were used to assess the overall quality of the structures for residues 4–140.

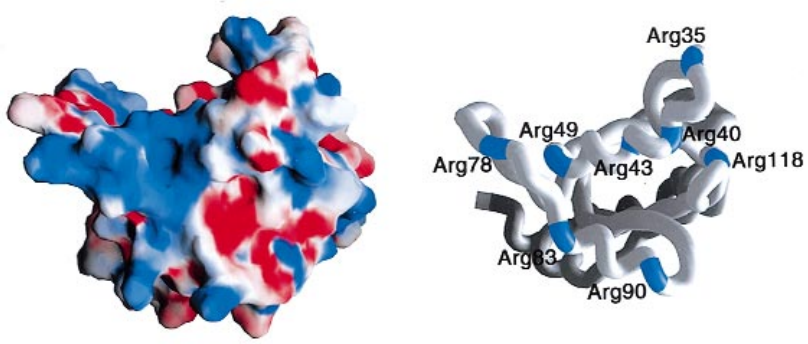


**Fig. 4** Arginine cluster of PEBP2/CBFβ. The electrostatic surface[33] (±6kT, left) identifies a distinctly positive surface (blue) formed by arginines 49, 78 and 83 on one end of the barrel. The worm representation (right) identifies the position of the arginine residues on the β surface. This surface coincides with that suggested by chemical shift perturbation mapping to possibly be near the DNA binding loops of the Runt domain and/or the DNA in the PEBP2/CBF heterodimer (see text).

the cytoplasm[21,22], leaving *bona fide* PEBP2/CBF unavailable for nuclear translocation and transcriptional activation. This provides a molecular explanation for the initiation of events leading to leukemogenic transformation of cells harboring the PEBP2/CBFβ-SMMHC chimera.

## Methods

**NMR spectroscopy.** The core structured domain of PEBP2/CBFβ was identified, subcloned, overexpressed and purified as described[1]. Samples of $^{13}C$ and/or $^{15}N$-enriched protein were concentrated to 1 mM in 10 mM sodium phosphate, 1 mM DTT and assignments conducted at 30 °C using standard techniques on Bruker DMX 500 and DMX 600 spectrometers[3,4]. $^{3}J_{NH\alpha}$, $^{3}J_{NH\beta}$, $^{3}J_{C\gamma N}$, $^{3}J_{C\gamma CO}$ coupling constants were measured by quantitative J correlation spectroscopy[23]. $^{15}N$-edited and $^{13}C$-edited three and four dimensional NOE spectroscopy was conducted with mixing times of 110 ms ($^{15}N$) and 120 ms ($^{13}C$), respectively. Secondary structure elements were identified from a combination of secondary $^{13}C\alpha$ and $^{13}C\beta$ shifts using the chemical shift index[23–25] as well as the pattern and intensity of NOEs observed in $^{15}N$-edited NOESY[26].

**Heterodimerization.** The ternary complex of core PEBP2/CBFβ–RD–DNA was formed as described[1] with only core PEBP2/CBFβ enriched with $^{15}N$ to selectively observe its chemical shifts. The ternary complex was analyzed by chemical shift perturbation mapping to identify the interface with the RD[1,27]. Assignments of the $^{15}N$ and $^{1}H$ backbone chemical shifts in the ternary complex were accomplished by pattern matching to the $^{15}N$-$^{1}H$-HSQC spectrum of free β in combination with a comparision of NOE patterns in free and bound $^{15}N$-edited 3D-NOESY. The total change in chemical shift was determined by treating the HSQC spectrum as a coordinate grid in which the position of each peak was defined by the $^{1}H$ and $^{15}N$ chemical shift. The change in the location of a peak was then calculated by measuring the distance between the initial (free) and final (bound) position of a peak in p.p.m. units. Treating the data in this manner avoids the use of artificial scaling factors and allows measurement of the change in a magnetic field-independent manner. Changes in the chemical shift coordinate of greater than 0.5 p.p.m. were deemed significant as these changes stood out above those observed for the majority of peaks in the spectrum.

**Structure determination.** NOEs within the protein were grouped into four distance ranges as described[1,5,27]. Distances involving methyl groups, aromatic ring protons and non-stereospecifically assigned methylene protons were represented as a $(\Sigma r^{-6})^{-1/6}$ sum[28]. $\phi$, $\chi_1$ and $\chi_2$ angles were derived from $^{3}J$ coupling constants and qualitative analysis of heteronuclear NOEs as previously described[1,5,23,27]. Protein backbone hydrogen bonding restraints ($r_{NH\text{-}O}$ = 1.5–2.8 Å, $r_{N\text{-}O}$ = 2.4–3.5 Å) within areas of regular secondary structure were introduced during the final stages of refinement. The minimum ranges employed for $\phi$, $\chi_1$ and $\chi_2$ torsion angle restraints were ±30°. The structures were calculated with the program X-PLOR-3.843 (ref. 29) adapted to incorporate pseudo-potentials for $^{3}J_{NH\alpha}$ coupling constants[30] and a conformational database potential[31] employing a protocol as described[5]. There were no hydrogen-bonding, electrostatic or 6–12 Lennard-Jones empirical potential energy terms in the target function. Structure quality was assessed with PROCHECK_NMR[6] and Prosa II[7] (Table 1).

**Coordinates**. The coordinates of the 25 structures of PEBP2/CBFβ have been deposited in the Protein Data Bank (1CL3).

**Acknowledgments**

The authors wish to acknowledge many stimulating discussions with T. Nagata, L. Glaser, J. Hill, A. Kim, D. Sorce, A. Seth and M. Osato. This work was supported in part by generous grants from the Sidney Kimmel Cancer Foundation, the Alexander and Alexandrine Sinsheimer Foundation and the New York Community Trust to M.H.W., a postdoctoral fellowship from the Leukemia Research Foundation to V.G. and grants from the Ministry of Education, Science, Culture and Sports of Japan to Y.I. and K.S.

Correspondence should be addressed to M.H.W. *email: mwerner@portugal.rockfeller.edu*

Received 29 March, 1999; accepted 14 May, 1999.

1. Nagata, T. *et al. Nature Struct. Biol.* **6**, 615–619 (1999).
2. Look, A. T. *Science* **278**, 1059–1064 (1997).
3. Bax, A. & Grzesiek, S. *Acc. Chem. Res.* **26**, 131–138 (1993).
4. Clore, G.M. & Gronenborn, A.M. *Prot. Sci.* **3**, 372–390 (1994).
5. Omichinski, J.G., Pedone, P. V., Felsenfeld, G., Gronenborn, A.M. & Clore, G.M. *Nature Struct. Biol.* **4,** 122–132 (1997).
6. Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R. & Thornton, J.M. *J. Biomol. NMR* **8**, 477–486 (1996).
7. Sippl, M.J. *Proteins* **17**, 355–362 (1993).
8 Murzin, A.G. *EMBO J.* **12**, 861–867 (1993).
9. Holm, L. & Sander, C.*J. Mol. Biol.* **233**, 123–138 (1993).
10. Peat, T.S., Newman, J., Waldo, G., Berendzen, J. & Terwilliger, T.C. *Structure* **6**, 1207–1214 (1998).
11. Stein, P.E., Boodhoo, A., Tyrrell, G.J., Brunton, J.L. & Read, R.J. *Nature* **355**, 748–750 (1992).
12. Hynes, T. R. & Fox, R.O. *Proteins* **10**, 92–105 (1991).
13. Ruff, M. *et al. Science* **252**, 1682–1689 (1991).
14. Akamatsu, Y. *et al. J. Biol. Chem.* **272**, 14497–14500 (1997).
15. Huang, X. *et al. J. Biol. Chem.* **273**, 2480–2487 (1998).
16. Huang, X., Peng, J. W., Speck, N. A. & Bushweller, J. H. *Nature Struct. Biol.* **6**, 624–627 (1999).
17. Ogawa, E. *et al. Virol.* **194**, 314–331 (1993).
18. Wang, S. *et al. Mol. Cell. Biol.* **13**, 3324–3339 (1993).
19. Kagoshima, H., Akamatsu, Y., Ito, Y. & Shigesada, K. *J. Biol. Chem.* **271**, 33074–33082 (1996).
20. Golling, G., Li, L., Pepling, M., Stebbins, M. & Gergen, J.P. *Mol. Cell. Biol.* **16**, 932–942 (1996).
21. Kanno, Y., Kanno, T., Sakakura, C., Bae, S-C. & Ito Y. *Mol. Cell. Biol.* **18**, 4252–4261 (1998).
22. Adya, N., Stacy, T., Speck, N. A. & Liu, P. P. *Mol. Cell. Biol.* **18**, 7432–7443 (1998).
23. Bax, A. *et al. Methods Enzymol.* **239**, 79–105 (1994).
24. Spera, S. & Bax, A.*J. Am. Chem. Soc.* **113**, 5490–5492 (1991).
25. Wishart, D. S. & Sykes, B.D. *J. Biomol. NMR* **4**, 171–180 (1994).
26. Wüthrich, K. *NMR of proteins and nucleic acids* (Wiley, New York; 1986).
27. Groft, C. M., Uljon, S.N., Wang, R. & Werner, M.H. *Proc. Natl. Acad. Sci. USA* **95**, 9117–9122 (1998).
28. Nilges, M. *Prot. Struct. Funct. Genet.* **17**, 295–309 (1993).
29. Brünger, A. T. X-PLOR Manual, Version 3.1 (Yale University Press, New Haven, Conecticut; 1992).
30. Garrett, D.S. *et al. J. Magn. Reson. (B)* **104**, 99–103 (1994).
31. Kuszewski, J., Gronenborn, A.M., Clore, G.M. *J. Magn. Reson.* **125**, 171–177 (1997).
32. Carson, M. *J. Mol. Graphics* **5**, 103–106 (1987).
33. Nicholls, A., Sharp, K. & Honig, B. *Proteins* **11,** 281–296 (1991).