CBF—A biophysical perspective

John H. Bushweller

Core binding factor (CBF) is a heterodimeric transcription factor consisting of a DNA-binding subunit (Runx, also referred to as CBFA, AML 1, PEBP2α) and a non-DNA-binding subunit (CBFB). Biophysical characterization of the two proteins and their interactions is providing a detailed understanding of this important transcription factor at the molecular level. Measurements of the relevant binding constants are helping to elucidate the mechanism of leukemogenesis associated with altered forms of these proteins. Determination of the 3D structures of CBFB and the DNA- and CBFB-binding domain of Runx, referred to as the Runt domain, are providing a structural basis for the functioning of the two proteins of CBF.

Key words: core binding factor / CBF / Runx / AML1 / leukemia / hematopoiesis / structure

© 2000 Academic Press

Introduction

Core binding factors (CBFs) are heterodimeric transcription factors consisting of a DNA-binding subunit (Runx) and a non-DNA-binding subunit (CBFB).1–4 All Runx subunits share a functional domain, referred to as the Runt domain, which is responsible for the DNA-binding as well as the CBFB-binding activities. As described elsewhere in this issue, CBFs are critical for a number of developmental processes in Drosophila as well as mammals. The mammalian RUNX2 gene is required for bone development5 and the RUNXI and CBFB genes have been demonstrated to be essential for the emergence of all fetal and blood cell lineages in the mammalian embryo.6,7 Chromosomal translocations involving the RUNXI and CBFB genes have been identified in a substantial number of myeloid and lymphocytic leukemias.8 The clear importance of CBF in normal development as well as in leukemia makes their biophysical characterization an essential component of an overall effort to understand the functioning of CBF both in its normal developmental role and in its role as an oncogene.

Stoichiometry and binding interactions

Of necessity, any biophysical characterization must begin with establishing the stoichiometry of the interactions between the proteins of interest as well as the binding constants involved. All of the equilibria involving the two proteins of CBF and DNA are illustrated in Figure 1. Both electromobility gel shift assays1,9 and more recent sedimentation equilibrium measurements10 have provided unequivocal proof for the 1:1:1 stoichiometry of interaction between the proteins of CBF and DNA. Based on this knowledge, more extensive binding studies have been carried out on the Runt domain of Runx1 and on CBFB.9–11 The results on CBFB demonstrated that an N-terminal 141-amino-acid form of CBFB was indistinguishable from the full-length protein in terms of its ability to bind the Runt domain of Runx1,11 consistent with the ability of this truncated form to rescue definitive hematopoiesis in CBFB knockout cells.12 The predominant protein resulting from the inv(16) associated with 12–15% of acute myeloid leukemias fuses the first 165 amino acids of CBFB to the coiled-coil region of smooth muscle myosin. These binding studies show that this fusion retains all of the determinants of CBFB necessary for binding to Runx. Subcellular localization studies of the inv(16) oncoprotein also confirm its ability to bind to Runx.13

A recent extensive EMSA and microcalorimetry
Figure 1. Schematic illustration of the four equilibria involving the two proteins of CBF and DNA. The proposed conformational change in Runx induced by CBFB is indicated as a change in the contacts made by Runx with the DNA upon CBFB binding.

Figure 2. (a) Ribbon representation of the solution structure of CBFB. The helices are colored red and yellow, the β-strands cyan and other segments gray. (b) Map of the chemical shift perturbation data on a ribbon representation of CBFB. Yellow indicates large chemical shift changes upon binding of Runx1–DNA to CBFB.

study of all the equilibria involving the two proteins of CBF and DNA\textsuperscript{10} has shown that the CBFB subunit enhances the DNA-binding affinity of the isolated Runt domain of Runx1 approximately sixfold. All of the studies described to this point were carried out utilizing the Runt domain of Runx, not full-length protein. Two studies on the binding behavior of longer forms of the Runx protein\textsuperscript{14,15} have shown the full-length proteins to have substantially reduced affinity for DNA relative to the isolated Runt domain. These data strongly suggest the presence of (an) inhibitory domain(s) present in the full-length protein outside of the Runt domain. This inhibition is relieved upon the binding of CBFB, providing a more definitive and substantive role for the CBFB subunit. Interestingly, the region of Runx implicated in this auto-inhibition is not present in most of the fusion proteins resulting from translocations involving the RUNXI gene. This implies that these oncprotein forms of Runx1 will have substantially higher DNA-binding affinities than their wildtype counterparts and may provide an explanation for their associated dysregulation. In addition, these fusion proteins also lack the region necessary for targeting of the protein to the nuclear matrix, a function which has also been shown to be essential for transcriptional regulation by Runx.\textsuperscript{16}

Three-dimensional structure of CBFB

Recent structural work on the heterodimerization domain of CBFB\textsuperscript{17,18} and on the Runt domain of Runx\textsuperscript{1}\textsuperscript{9,20} have provided a much more detailed picture of the functioning of the two proteins of CBF. The structure of CBFB [Figure 2(a)] revealed it to have a novel α/β fold that consists of two non-autonomously folded regions that are separated by a flexible loop. The N-terminal region starts with α-helices 1 and 2 which lead into β-strand 1. β-strand 1 is followed by a short 3\textsubscript{10} helix and subsequently α-helix 3. α-helix 3 is followed by β-strands 2 and 3 with a type 1 β-turn between them. The first three β-strands combine to form a three-stranded anti-parallel β-sheet. The N-terminal region is connected via a loop region that was shown by NMR relaxation measurements to be flexible in solution. The C-terminal region begins with β-strands 4, 5 and 6, which also combine to form a three-stranded anti-parallel β-sheet and the final C-terminal α-helix 4. There are limited contacts between the two sheets that make it one continuous β-sheet.

Binding site for Runx on CBFB

The mapping of binding sites for ligands, be they proteins, nucleic acids or small molecules, can be identified readily using an NMR method referred to as chemical shift perturbation analysis. Addition, in this case, of a Runx Runt domain–DNA complex to CBFB results in the formation of a ternary complex. Because the chemical shifts, or frequencies, of the
nuclei in an NMR experiment are very sensitive to their local environment, comparison of the backbone NH chemical shifts of CBFB in the presence and absence of the Runt domain–DNA complex can identify those amide NH moieties that display substantial perturbations upon binding. All of the amino acids save proline have an associated amide NH signal in the NMR experiment, thus this provides a high-resolution map of the perturbations. The observed changes may result from regions of CBFB in direct contact with the Runt domain–DNA complex or they may result from conformational changes associated with that binding. This has proven to be an extremely powerful method for the elucidation of binding sites on proteins and nucleic acids. Such a study has been carried out in two laboratories\textsuperscript{17, 18} for the identification of the Runx Runt domain binding site on CBFB, with very similar results. The results of one such study are shown in Figure 2(b).\textsuperscript{17} The associated perturbations map to a very localized region on the structure of CBFB that includes \(\beta\)-strand 3 and the loops leading into and out of \(\beta\)-strand 3 as well as \(\beta\)-strands 4 and 5 and their associated connecting loop. We have also observed perturbations for two residues at the N-terminal end of the protein. Mutagenesis of residues in this region of CBFB confirms the importance of this region for interaction with Runx (Tang and Speck, unpublished results).

Three-dimensional structure of the Runx1 Runt domain

Structural studies of the Runt domain of Runx1 have been carried out using Runt domain–DNA complexes because of the poor behavior of the isolated protein in solution. As shown in Figure 3(a), the protein has been shown to be a \(\beta\)-protein with no \(\alpha\)-helical content whatsoever, as was predicted from an earlier circular dichroism (CD) study.\textsuperscript{9} The fold of the protein consists of seven clearly discernible \(\beta\)-strands with strands 1, 2 and 5 combining to form one anti-parallel \(\beta\)-sheet and strands 3, 4, 6 and 7 combining to form a second anti-parallel \(\beta\)-sheet. The two sheets are packed against one another at a shallow angle to yield a \(\beta\)-sandwich arrangement. Two short \(\beta\)-sheet interactions were also identified, involving L117/R138-R135/F136 and I168/T169/H78/W79, which, because of their proximity to the DNA (see below), have been suggested to be functionally relevant. The fold identified for the Runt domain belongs to the classic immunoglobulin (Ig) fold, specifically the s-type Ig fold best exemplified by the structure of the fibronectin repeat element. Interestingly, this fold has been identified in the DNA-binding domains of a number of other critical mammalian transcription factors including NF-\(\kappa\)B, NFAT, p53, STAT-1 and the T-domain. All of the DNA-binding domains of these proteins share the s-type Ig fold, thus the Runt domain belongs to a family of structurally related s-type Ig fold, DNA-binding domains. In addition to the overall fold being shared among these proteins, a number of these proteins also display homologous short \(\beta\)-sheet interactions like those observed in the Runt domain.\textsuperscript{19} For all of these structurally related proteins, binding to DNA is mediated by the loop regions that connect the various \(\beta\)-strands. For all the previously characterized members of this family, DNA binding is mediated by loops located either at the ‘top’ or ‘bottom’ of the fold as depicted in Figure 3, but not by both ends.

Mapping of the DNA-binding site on the Runx1 Runt domain

In order to identify the region of the Runt domain involved in DNA binding, nuclear Overhauser (NOE) spectroscopy was utilized. This type of NMR experiment allows one to observe protons that are close to one another in space, less than ca 5 Å in practice. By utilizing a sample containing fully deuterated Runt domain, i.e. no observable protons save the protein amide NH groups, bound to unlabeled DNA, one is able to identify NOEs from the protein NH groups to the DNA. The results of this study\textsuperscript{19} are shown in Figure 3(b). In addition, a substantial number of point mutations have been made in the Runt domain by several labs.\textsuperscript{9, 21–25} Those that have been shown to affect either DNA or CBFB binding are also shown in Figure 3(b). As Figure 3(b) shows, DNA binding is mediated by the loop regions of the protein. There is very good concurrence in the location of the observed NOEs and the locations of the point mutations affecting DNA binding, confirming this characterization. Unlike the other members of the structural family the Runt domain appears to belong to, all the loops appear to be involved in DNA binding, that is, loops at the ‘top’ and ‘bottom’ of the structure are involved in DNA recognition. The Runt domain may be unique among this group in this respect. This may possibly be necessary for the DNA bending that the Runt domain has been shown to effect.\textsuperscript{24}
Mapping of the CBFB-binding site on the Runx1 Runt domain

Chemical shift perturbation mapping studies have, as described above for CBFB, been employed to identify the CBFB-binding site on the Runx1 Runt domain. The results of one of these studies are illustrated in Figure 3(c). Both studies have identified one face of the protein, that of the four-stranded β-sheet, as the location for the binding of CBFB. The region thus identified is in good agreement with the mutagenesis data shown in Figure 3(b) in that those point mutations in the Runt domain that affect CBFB binding are localized to the same region as identified by the chemical shift perturbation analysis. The proximity of the CBFB-binding site on the Runt domain to regions of the Runt domain involved in DNA binding provides a rationale for the ability of CBFB to alter the DNA-binding affinity of the Runt domain. Although a clear delineation of this will require structural characterization of a ternary complex, which is not yet available, two pieces of data tend to support the latter mechanism. First, circular dichroism studies of the interaction between the Runt domain of Runx1 and CBFB show a clear conformational change in one or both proteins upon interaction either on or off of the DNA. Second, one of the chemical shift perturbation studies has identified dramatic chemical shift changes for a specific region of the Runt domain located in close proximity to the DNA which was not assigned in the other study. The magnitude of the observed changes and the dramatic alterations observed in the Cα chemical shifts, which are a strong indicator of backbone conformation, strongly support a conformational change in the Runt domain associated with this interaction (illustrated schematically in Figure 1).

Other structural studies

The only other structural characterization of the CBF proteins reported thus far has been a crystal structure of the nuclear matrix targeting signal (NMTS) of Runx1 fused to glutathione-S-transferase. The results of this study indicated that part of the peptide...
adopts a defined tertiary structure described as two finger-like loops connected by a U-shaped peptide chain. These results provide a meaningful framework to examine the determinants for nuclear localization in the NMTS of Runx proteins.

Future directions

Regulation of transcription in mammals has been shown to be a complex process involving the interplay and interaction of multiple proteins at a particular site on the DNA. Recent progress in structurally characterizing a number of ternary complexes with two proteins bound to DNA has provided detailed atomic-resolution pictures of how this interplay is mediated at the molecular level.27 In these cases, the interactions with the various partner proteins have proven to be essential to the in vivo function. CBF has been shown to interact with a number of other proteins including Ets proteins (particularly Ets-1), C/EBP, TLE, PU.1, ALY, Myb, Ear2 and p300.28–35

In the case of the Runx /Ets-1 interaction, the cooperative nature of the DNA binding of Runx with Ets-1 provides a clear functional role for the interaction between the two.28 Future biophysical studies to elucidate the mechanism of this myriad of protein–protein interactions and particularly their structural basis will clearly be an area for fertile future work on CBF.

Acknowledgements

We wish to thank Dr Nancy Speck for many hours of fruitful discussions. JHB is supported by grants from the United States Public Health Service R29 AI39536, K02 AI10481 and R01 AI42097 from the Institute for Allergy and Infectious Disease.

Note added in proof

Very recently, a crystal structure of a CBFβ-Runt domain complex has also been reported by Warren et al. (2000).35

References

α Runt domain is a member of a family of structurally and functionally related Ig fold DNA binding domains. Structure 7:1247–1256


