Runx1 binds as a dimeric complex to overlapping Runx1 sites within a palindromic element in the human GM-CSF enhancer

Bowers, Sarion R; Calero-Nieto, Fernando J; Valeaux, Stephanie; Fernandez-Fuentes, Narcis; Cockerill, Peter N

Published in:
Nucleic Acids Research

DOI:
10.1093/nar/gkq356

Publication date:
2010

Citation for published version (APA):
Bowers, S. R., Calero-Nieto, F. J., Valeaux, S., Fernandez-Fuentes, N., & Cockerill, P. N. (2010). Runx1 binds as a dimeric complex to overlapping Runx1 sites within a palindromic element in the human GM-CSF enhancer. *Nucleic Acids Research*, *38*(18), 6124-34. https://doi.org/10.1093/nar/gkq356

6124–6134 *Nucleic Acids Research, 2010, Vol. 38, No. 18* *Published online 18 May 2010*
doi:10.1093/nar/gkq356

Runx1 binds as a dimeric complex to overlapping Runx1 sites within a palindromic element in the human GM-CSF enhancer

Sarion R. Bowers[1], Fernando J. Calero-Nieto[1], Stephanie Valeaux[1], Narcis Fernandez-Fuentes[2] and Peter N. Cockerill[1,*]

[1]Experimental Haematology and [2]Experimental Therapeutics, Leeds Institute of Molecular Medicine, University of Leeds, St James's University Hospital, Leeds LS9 7TF, UK

Received December 3, 2009; Revised April 22, 2010; Accepted April 23, 2010

ABSTRACT

Runx1 is a developmentally regulated transcription factor that is essential for haemopoiesis. Runx1 can bind as a monomer to the core consensus sequence TGTGG, but binds more efficiently as a hetero-dimer together with the non-DNA binding protein CBFβ as a complex termed core binding factor (CBF). Here, we demonstrated that CBF can also assemble as a dimeric complex on two overlapping Runx1 sites within the palindromic sequence TGTGGCTGCCCA CA in the human granulocyte macrophage colony-stimulating factor enhancer. Furthermore, we demonstrated that binding of Runx1 to the enhancer is rigidly controlled at the level of chromatin accessibility, and is dependent upon prior induction of NFAT and AP-1, which disrupt a positioned nucleosome in this region. We employed *in vivo* footprinting to demonstrate that, upon activation of the enhancer, both sites are efficiently occupied. *In vitro* binding assays confirmed that two CBF complexes can bind this site simultaneously, and transfection assays demonstrated that both sites contribute significantly to enhancer function. Computer modelling based on the Runx1/CBFβ/DNA crystal structure further revealed that two molecules of CBF could potentially bind to this class of palindromic sequence as a dimeric complex in a conformation whereby both Runx1 and CBFβ within the two CBF complexes are closely aligned.

INTRODUCTION

Runx1 (also known as AML1, CBFα2 or PEBP2αB) is a Runt-domain transcription factor (1,2) that is essential for haemopoiesis (3,4). Runx1 is a member of a conserved family of closely related proteins that include mammalian Runx1, Runx2 and Runx3, and the *Drosophila* protein Runt (1,2). Runx1 binds to TGTGGNNN core sequences, typically TGTGGTTT or TGTGGTCA, as a heterodimer of Runx1 and CBFβ, termed core binding factor (CBF) (2,5,6). Although CBFβ does not directly contact DNA, it helps to stabilize Runx1 binding (2). While TGTGGT is the most commonly observed natural CBF-binding sequence, *in vitro* studies reveal that TGCGGT is also a high affinity, but much less often encountered, CBF-binding sequence (2,5,7).

Runx1 binds to DNA via the Runt domain, which shares some similarity with the Rel domain of NF-κB/Rel family proteins that also recognize GG core sequences (8,9). However, NF-κB binds as an obligate dimer that employs two Rel family proteins to bind to palindromic sequences such as GGGAAATTCCC (10). This is in contrast to Runx1 which interacts efficiently with single TGTGGNNN consensus motifs (6). The crystal structure of a Runx1-Runt domain/CBFβ complex bound to the DNA sequence TGCGGTTG has been determined, revealing contacts with both the bases and the phosphate

*To whom correspondence should be addressed. Tel: +44 113 3438639; Fax: +44 113 3438502; Email: p.n.cockerill@leeds.ac.uk
Present address:
Fernando J. Calero-Nieto, Department of Haematology, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Cambridge University, Hills Road, Cambridge CB2 0XY, UK.

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

backbone throughout this 8-bp sequence (11). Similar results were obtained by a parallel study of the Runx1 binding site TGTGGTTG (12). Runx1 clamps the phosphate backbone between the major and minor groove, while forming Rel-like base-specific contacts with the GG sequence in the major groove (11). In this way, Runx1 utilizes multiple protein:DNA contacts to enable efficient binding to single sites (11).

Runx1 synergizes with a variety of other factors to regulate composite elements. For example, Runx1 and Ets-1 interact directly and bind cooperatively to enhancers within the TCRα and TCRβ loci (13,14). Runx1 and C/EBP family proteins synergize in the activation of adjacent binding sites within the M-CSF receptor gene promoter (15). Runx1 and c-Myb synergize in the activation of adjacent sites in the TCR-δ enhancer, although in this instance synergy does not depend on cooperative binding (16). Runx1 also possesses an interaction domain within the C terminal region that mediates homodimerization, and this may promote binding to regulatory elements containing multiple Runx1 binding sites (17). The homodimerization domain appears to be distinct from the region of Runx1 required for interaction with other factors, such as Ets-1 (14). This domain is less structured than the Runt domain, and no detailed determination of its structure is available.

Runx1 shares some similarities with another class of Rel domain transcription factor, the NFAT family of proteins, which bind to GGAAANN consensus sequences. Like Runx1, NFAT uses multiple interactions to bind to DNA efficiently as a monomer. NFAT uses a single Rel domain to bind in the major groove to an essential GGA core sequence, and also has contacts in the minor groove along the next 4-bp downstream of the GGA core (18). NFAT typically functions and binds cooperatively together with other transcription factors such as AP-1 (18,19). However, NFAT family proteins can also bind as homodimers to NF-κB-like sequences conforming to the consensus sequence GGAAATTCC (18,20,21). This raises the possibility that Runx1, like NFAT, might also be able to bind to palindromic sequences containing two overlapping Runx1 binding sites. However, to date we are not aware of any reports of such examples.

The human granulocyte macrophage colony-stimulating factor (GM-CSF or CSF2) gene is a key target of regulation by Runx1 (22–24). GM-CSF is a pro-inflammatory cytokine produced by activated T cells and mast cells that is induced by stimuli that activate the immune system. The expression of the human GM-CSF gene is under the control of a promoter region located just upstream of the transcription start site and an inducible enhancer located 3-kb upstream, which are both activated by kinase and calcium signalling pathways that in T cells are activated via the T-cell receptor (TCR) (19,25–27). We and others have previously demonstrated the presence of functional Runx1 sites in both the GM-CSF promoter (22–24) and the inducible enhancer (22). GM-CSF gene expression can also be inhibited by the repressive AML1/ETO fusion protein found in *t*(8;21) chromosomal translocations in acute myeloid leukaemia (23,24). These translocations delete the homo-dimerization domain, and the addition of the ETO domain converts Runx1 from an activator to a repressor.

The GM-CSF enhancer is defined as a 717-bp BglII fragment of DNA that requires interactions with multiple transcription factor sites for function (25,28). Within the GM-CSF enhancer, Runx1 binds to the GM450 element within a region that is normally occluded by a positioned nucleosome (Figure 1A; 28,29). Activation of the enhancer by TCR signalling pathways results in recruitment of NFAT and AP-1 to sites located within this same nucleosome (N2) and to additional sites in an adjacent nucleosome (N1), leading to the rapid eviction of both nucleosomes (summarized in Figure 1A). This process results in the rapid creation of a DNase I hypersensitive site (DHS), a substantial increase in chromatin accessibility, and the recruitment of Runx1 to a site that was previously unoccupied. This mechanism thereby allows constitutively expressed factors such as Runx1 and Sp1 to bind in a highly inducible fashion.

In this study, we demonstrated that the GM-CSF enhancer GM450 Runx1 binding site TGTGGCTGCCCACA actually has two overlapping TGTGGNNN Runx1-like elements where the TGTGG core sequences (underlined) exist 4 bp apart within this palindromic sequence. Furthermore, we obtained evidence that both sites become occupied in activated T cells and mast cells, and demonstrated that both sites contribute to enhancer function. Computer modelling predicted that the palindromic element is able to accommodate binding of two Runx1/CBFβ complexes closely aligned on the same face of the DNA helix to form multiple close associations along an axis perpendicular to the DNA.

MATERIALS AND METHODS

Cell culture

The CEM and Jurkat human T-cell lines were cultured in RPMI (Invitrogen) supplemented with 10% FCS, 4 g/l D-glucose, 1 mM sodium pyruvate, 1 × MEM essential amino acids (Invitrogen), 1 × MEM non-essential amino acids (Invitrogen), 100 U/ml penicillin and 100 μg/ml streptomycin. C42 human IL-3/GM-CSF transgenic mouse T lymphoblasts and mast cells were prepared and cultured as described previously (29,30).

In vivo footprinting analyses

Footprinting analyses were performed essentially as described (31,32). C42 transgenic mouse T lymphoblasts were either stimulated for 4 h with 20 ng/μl phorbol 12-myristate 13-acetate (PMA) and 1 μM calcium ionophore A23187, or left unstimulated, before treatment with 0.2% dimethyl sulphate (DMS) in phosphate buffered saline (PBS) for 5 min at room temperature. Specific DNA cleavage sites were detected using ligation-mediated (LM)-PCR as described (31,32). These analyses were performed with two sets of 3 nested primers designed to detect DNA cleavage sites on either the upper or the lower strand of the GM420/GM450 region of the human GM-CSF enhancer, plus the linker primer.

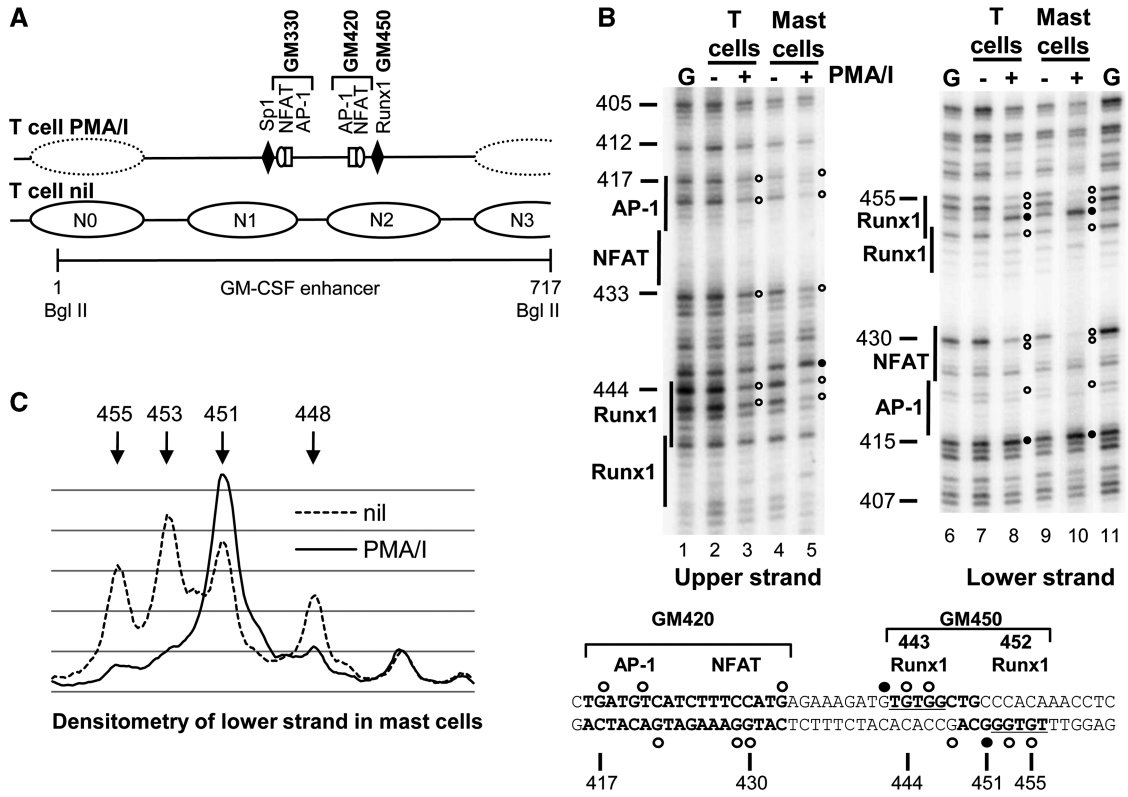

Figure 1. Transcription factors and nucleosomes associated with the human GM-CSF enhancer. (**A**) Map of the 717 bp BglII fragment of the enhancer showing the previously determined locations of positioned nucleosomes (N0–N3) and transcription factors bound to the GM-CSF enhancer in T cells before and after stimulation with PMA and calcium ionophore. The composite NFAT/AP-1 elements are termed GM330 and GM420, and the Runx1 element is termed GM450. The dotted ovals signify nucleosomes that become mobilized after stimulation which do not reside in fixed positions. (**B**) *In vivo* DMS footprints within the GM-CSF enhancer. Cultured C42 transgenic mouse T cells and mast cells were assayed before (−) and after (+) stimulation for 4 h with PMA and calcium ionophore. Either genomic DNA (G) or intact cells were treated with DMS, and the modified bases were cleaved with piperidine and identified by polyacrylamide gel electrophoresis of LM-PCR products. Open circles indicate protected bases, and filled circles indicate hyper-reactive bases. Bases are numbered relative to the upstream BglII site. (**C**) Quantification of the *in vivo* protection of the CBF site on the lower DNA strand. These densitometric traces of lanes 9 (nil) and 10 (PMA/I) in Figure 1B reveal DMS reactivity in mast cells before and after stimulation.

The three primers within each set are designated EB for the biotinylated primer used at the first stage, EP for the primer used at the second PCR stage and EL for those used in the final labelling stage. The 5′ primers used to detect cleavage on the lower strand (plus their annealing temperatures shown in brackets) were EB3F2, CACAGC CCCATCGGAGC (52°C); EP3F20, CTGAGTCAGCAT GGCTGGC (62°C); EL3F, GCATGGCTGGCTATCGG TTGACACTG (68°C). The 3′ primers used to detect cleavage on the upper strand (plus their annealing temperatures) were EB2R2, GCCCAAGTCAGCACAAAC (56°C); EP2R, GTCAGCACAAACAGGACAGAAATC (64°C); EL2R20, AGGACAGAAATCCATGGGTTTG GTGATG (64°C).

Electrophoretic mobility shift assays

To study binding of purified CBF, we obtained a sample of the purified recombinant Runx1/CBFβ complexes that had been used previously by Bravo *et al.* (11) to prepare crystals of CBF bound to DNA. These complexes contained just the His-tagged Runt domain region of Runx1 (residues 50–183), and residues 2–135 of CBFβ. To study binding of native CBF complexes we prepared nuclear extracts from unstimulated Jurkat T cells, essentially as described previously (25).

Double stranded oligonucleotides were labelled by end-filling with [α-^{32}P]-dCTP plus unlabelled dATP, dGTP and dTTP, and the labelled probes were purified by polyacrylamide gel electrophoresis. 6 µg of nuclear extract protein was incubated with 0.2 ng of labelled probe, and 3 µg of poly(dI.dC) in a 16 µl reaction, containing 10% glycerol, 20 mM HEPES, 30 mM KCl, 30 mM NaCl, 3 mM $MgCl_2$, 1% DTT, 0.1 mM PMSF, 5 µg/ml aprotinin, 5 µg/ml leupeptin, for 10 min at room temperature (~22°C) and 10 min on ice before analysis on 4% polyacrylamide gels run at 4°C. For super-shift assays 0.5 µl of Runx1 antibody (Santa Cruz C19 or Santa Cruz N20), or control IgG (Millipore 12-370) was incubated with the nuclear extract for 10 min at room temperature before addition to the above reaction. Competition assays were carried out by addition of competitor at stated concentrations in a 1 µl volume to nuclear extract and incubated for 10 min on ice before addition of 0.2 ng of labelled probe and further incubation for 10 min on ice. Electrophoretic mobility shift assays (EMSAs) of recombinant purified CBF were performed by incubating

0.25–2.0 pmol of CBF with 0.2 ng of labelled probe and 100 ng of poly(dI.dC) in a 16 µl reaction containing 10% glycerol, 20 mM HEPES pH7.9, 50 mM NaCl, 80 mM KCl, 3 mM $MgCl_2$, 1% DTT, 0.1 mM PMSF, 5 µg/ml aprotinin, 5 µg/ml leupeptin, 0.1 mg/ml BSA at room temperature for 10 min and 10 min on ice before analysis on 4% polyacrylamide gels run at 4°C.

The oligonucleotides used to prepare DNA probes had the following sequences, which include a 2 base 5′ overhang at each end:

TCR-δ Runx1, AGGCATGTGGTTTCCAACCGTT and TGAACGGTTGGAAACCACATGC;
Ideal, AGATGTGTGGTTAACCACAAAC and AGGTTTGTGGTTAACCACACAT;
Ideal∆1 AGATGTGTGGTTAAGGACAAAC and AGGTTTGTCCTTAACCACACAT.
Crystal, AGATGTGCGGTCGACCGCAAAC and AGGTTTGCGGTCGACCGCACAT;
WT GM450, AGATGTGTGGCTGCCCACAAAC and AGGTTTGTGGGCAGCCACACAT;
GM450∆443, AGATCTCTCACTGCCCACAAAC and AGGTTTGTGGGCAGTGAGAGAT;
GM450∆452, AGATGTGTGGCTGCGTAGAAAC and AGGTTTCTACGCAGCCACACAT;
GM450∆443/452, AGATCTCTCACTGCGTAGAAAC and AGGTTTCTACGCAGTGAGAGAT.

Autoradiography and densitometry

Images were collected on a BioRad Pharos FX Plus phosphorimager. Densitometry and band quantitation was performed using BioRad Quantity One software.

Plasmid construction

The previously described constructs pGM and pGM-GME contain the −627 to +28 region of the human GM-CSF promoter (GM627) in the absence or presence of a 717 bp Bgl II fragment of the human GM-CSF enhancer, respectively (33), inserted into the firefly luciferase reporter gene plasmid pXPG which includes highly efficient transcription termination and polyadenylation elements upstream of the enhancer (34). Site-directed mutagenesis was performed as previously described (28) to change the TGTGG core sequences located at positions 443 and 452 in the GM-CSF enhancer to TGTCC, as indicated in Figure 3. After mutagenesis was completed, the full sequence of the enhancer in isolated clones was determined, and if correct, was recloned into pGM so as to avoid any potential additional unanticipated mutations elsewhere in the plasmids.

The Renilla luciferase control vector pXRL-GME was created by excising just the firefly luciferase gene from the pXPG-based vector pGM-GME (33) by digestion with Xba I plus a partial digestion by HindIII, followed by insertion of a HindIII–Xba I of the Renilla luciferase gene excised from the control plasmid pRL-TK (Promega). This vector retains the same configuration of vector backbone, upstream SV40 polyA/termination sites, and GM-CSF enhancer and promoter as in pGM-GME, and is designed to respond in the same way upon transfection and stimulation.

Transient transfection and luciferase assays

Aliquots of 4.5×10^6 Jurkat T cells were transfected by electroporation with plasmid DNA purified by two rounds of CsCl gradient centrifugation. Cells were transfected with 5 µg of Firefly luciferase plasmid plus 1 µg of the control pXRL-GME Renilla luciferase plasmid. Cells were cultured for ∼21 h post-transfection, before stimulation with 20 ng/ml PMA and 2 µM calcium ionophore A23187 for 8 h. Cells were then washed in PBS and assayed for luciferase reporter gene activities using the Promega dual luciferase assay kit and a Berthold Mithras LB-940 microplate luminometer. Data were collected from at least two independently prepared clones of each construct, which were in each case found to behave in like fashion. A minimum of 6, and up to 21 independent transfections were performed for each construct.

Computational modelling of the dimeric form of Runx1-CBFβ-DNA ternary complex

An extended and detailed explanation of the computational modelling and refinement of complex can be found in the Supplementary Methods. Briefly, the Runt1-CBFb-DNA coordinates defined in Bravo *et al.* (11, 35) were superimposed over a model of guide B-DNA containing two Runx1 DNA binding sites generated using the nucleic acid builder tool (36) in the AmberTool package and the 21 bp GM-CSF enhancer sequence ATGTGTGGC TGCCCACAAAAC. The crystal structure of the human Runx1-CBFβ-DNA ternary complex (11) was used as template to model two Runx1–CBFβ complexes bound to the guide B-DNA using MODELLER (37). Sequences and templates were aligned so that DNA molecules of the template were aligned to the corresponding DNA binding sites in the guide B-DNA preserving the native Runt1–CBFβ–DNA interactions described in the crystal structure (11). The refinement of the model consisted of two rounds of energy minimization and a short molecular dynamic simulation of the solvated complex in order to ensure a reasonable stereochemical geometry, resolve steric hindrances, and assess the stability of the complex. The energy minimization and molecular dynamic simulation was performed in AMBER 10 (38) using the ff99SB force field (39), as follows: On the first round of minimization, protein and DNA atoms were kept restrained and solvent was relaxed. On the second round, restraints on DNA and protein atoms, with the exception of those mediating the interactions between protein and DNA, were lifted and system was further minimized. The temperature of the system was then raised gradually to reach 300K, while remaining restraints were also gradually removed during the heating process. Finally, the system was equilibrated for 0.5 ns and simulated for 4 ns. The final model was obtained after clustering of the entire molecular dynamic simulation, selecting a representative on each cluster, and assessing the quality of each representative using PROCHECK (40). The convergence of the simulation and stability of the complex was ensured by plotting the root mean square deviation and total energy as a function of the simulation time.

RESULTS

Runx1 binds to 2 overlapping sites within the GM-CSF enhancer *in vivo* upon stimulation

We previously defined the GM450 Runx1 binding site TG CCCACA that has a TGTGG core sequence located at position 452 within the human GM-CSF enhancer (22). This site forms specific Runx1/CBFβ complexes as previously defined with the aid of specific antibodies and DNA competitors, and it can be used to functionally replace the Runx1 site located within the GM–CSF promoter (22). To study the regulation of the human GM-CSF locus in defined populations of normal cells, we employed the transgenic mouse line C42 containing the intact human GM-CSF locus on a 130 kb BAC segment of DNA (30). To study Runx1 interactions with the GM-CSF enhancer inside live cells we employed DMS *in vivo* footprinting of T cells and mast cells prepared from C42 mice (29) (Figure 1B). With this assay we can identify sites of *in vivo* protein:DNA interactions by treating cells with dimethyl sulphate (DMS), which modifies any G bases that are not occupied by interacting factors. To this end, cells were treated with DMS before (−) or after (+) stimulation for 4 hours with 20 ng/ml PMA and 2 μM calcium ionophore A23187. As a control for protein-free DNA, we also treated purified genomic DNA with DMS (G). Modified bases were cleaved by incubation with piperidine, and the specific cleavage sites were detected by ligation-mediated PCR. Figure 1B displays the results obtained after analysis of sites on both the upper and the lower strands of DNA. The results are summarized below with protected bases marked by open circles, and hypersensitive bases marked with filled circles. Before stimulation, there was no evidence for any interactions between any factors and the enhancer, as the patterns were essentially the same as those obtained by treating genomic DNA with DMS. After stimulation, there was clear evidence of efficient binding of factors to the Runx1, NFAT and AP-1 sites within this region of the enhancer. Furthermore, the Runx1 site supported not just one, but two distinct sets of footprints. In addition to binding at the previously defined TGTGG motif at position 452 (lower strand, lanes 8 and 10), there was strong protection of the Runx1-like element TGTGGCT G located at position 443 (upper strand, lanes 3 and 5). There was also a hypersensitive reaction with a G at position 451 on the lower strand, which is symptomatic of an altered DNA conformation induced by factor binding.

Particularly in the mast cells, it was evident that there was very strong protection of both Runx1 sites (lanes 5 and 10), indicating that both sites are likely to be stably occupied simultaneously in a substantial proportion of cells. Quantitation of a densitometric analysis of the data in lane 10 (for the lower strand) in stimulated mast cells (Figure 1C) indicated that only about 16% of all G bases at position 455 within the motif at position 452 remained accessible. A similar densitometric analysis of lane 5 (the upper strand) indicated that only 34% of G bases at positions 444 and 446 within the motif at position 443 remained accessible after stimulation (data not shown). Based on this, we can calculate that 66% of all motifs at position 443 on the upper strand become occupied. If only 16% of motifs at position 452 on the lower strand remain free, then this implies that a minimum of 50% (and maximum of 66%) of all GM450 palindromic Runx1 sites must be stably occupied at both sites in stimulated mast cells.

The GM-CSF enhancer Runx1 element can recruit two CBF complexes at the same time

We used EMSAs to study the *in vitro* binding of recombinant CBF complexes to the Runx1 sites (Figure 2A and B). For this purpose, we obtained a sample of the original purified recombinant CBF complex that had been used previously by Bravo *et al* to determine the crystal structure of the CBF/DNA complex (11). This recombinant CBF contained the His-tagged Runt domain region of Runx1 (residues 50–183), and the region of CBFβ (residues 2–135) that mediates interactions with Runx1. With 0.25 pmol of CBF (lane 1), two different mobility complexes were observed with the wild type probe (WT GM450). Although we have been unable to determine why two different mobility complexes are detected, it is likely that that they both contain a single molecule of Runx1, because both complexes were still detected when either the position 443 or 452 Runx1 sites were mutated individually (lanes 5 and 8). These complexes were both regarded as being specific for the TGTGG motifs, because no binding was detected when both Runx1 sites were mutated simultaneously (lanes 11–13). It is possible that the lower band represents a Runx1–DNA complex, from which CBFβ has dissociated, but we have been unable to confirm this.

When amounts of up to 2 pmol of recombinant CBF were added to the intact GM450 Runx1 probe, increasing amounts of a third lower mobility complex were observed (lanes 2–4). This upper complex was assumed to be a dimeric CBF complex containing two Runx1/CBFβ heterodimers, because it was not detected when either individual Runx1 site was mutated (lanes 5–10, Figure 2B).

In one published study, the structure of the CBF/DNA complex was determined using the DNA sequence TGCG GTTG. To determine whether we could use this structure as a basis for constructing a computer model of a dimeric complex, we designed the palindrome TGCGGTCGACC GCA that contains two overlapping TGCGGTCG motifs on opposite strands in the same arrangement as the two Runx1 sites in the GM-CSF enhancer (designated as Crystal in Figure 2A). In EMSAs, the crystal palindrome sequence was also capable of assembling a dimeric CBF complex, slightly more efficiently than the GM450 sequence (compare the 0.25 pmol EMSAs in lanes 1 and 14 in Figure 2B).

We next studied the binding of intact natural CBF complexes to the GM450 and crystal palindromic Runx1 sites using Jurkat human T cell nuclear extracts in EMSAs (Figure 2C). However, because neither of these palindromes conforms to the preferred TGTGGTCA or TGT GGTTT consensus, we also designed an additional palindromic probe containing the sequence TGTGGTTAACC ACA that matches the ideal consensus TGTGGTTA on

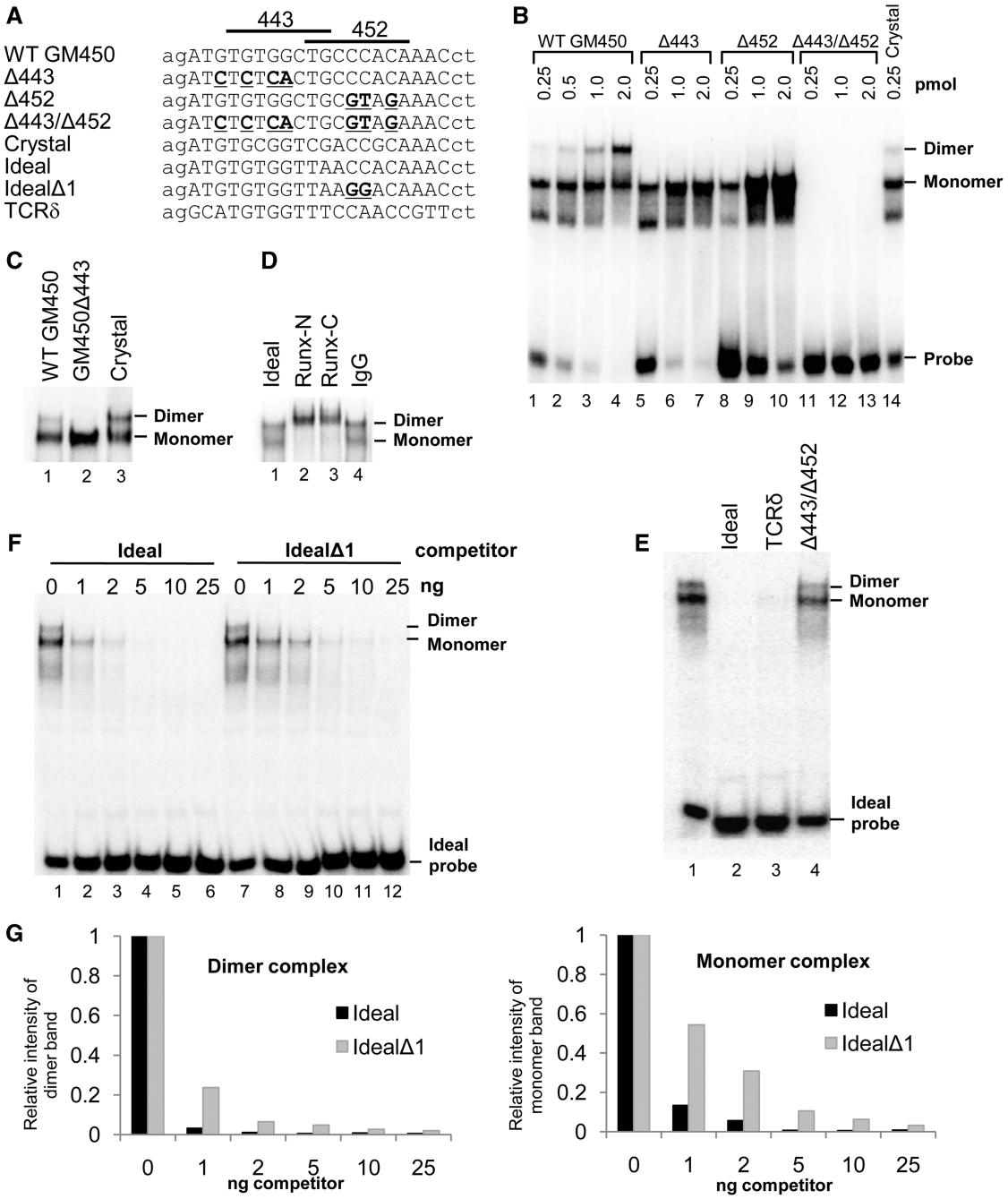

Figure 2. EMSAs of the GM450 Runx1 binding site. (**A**) Sequences of the wild-type and mutated GM450 Runx1-binding probes, plus the Ideal, the IdealΔ1 and the crystal structure-based palindrome sequence used in (**B–G**), and the TCRδ single Runx1 site. The Crystal sequence is a palindrome derived from the DNA sequence used to prepare crystals by Bravo *et al.* (11). The ideal sequence is a palindrome derived from the ideal Runx1 consensus binding site, and the IdealΔ1 sequence has the second Runx1 site in this element mutated. The 8-bp regions spanned by the two Runx1 binding sites are indicated by bars above the sequence, and the mutated GG core elements are shown as underlined in bold. Bases in lower case represent 2 base 5′ overhangs at each end that become double stranded when probes are labelled. (B) EMSAs with increasing concentrations recombinant Runx1/CBFβ complexes binding to the wild-type and mutated GM450 sequences and the Crystal sequence. Bands corresponding to a monomer and a dimer of Runx1 are marked. (C and D) EMSAs with 6 μg Jurkat cell nuclear extract with (C) the wild-type GM450 (lane 1), the mutated GM450Δ443 (lane 2) and the crystal (lane 3) sequences and (D) the ideal sequence (lane 1) in the presence or absence of an antibody against either the N-terminal (Runx–N) (lane 2) or C-terminal (Runx-C) (lane 3) of Runx1, or control IgG (lane 4). (E) Competition assay with 6 μg of Jurkat nuclear extract bound to labelled Ideal probe (lane 1) and competed with 25 ng of either Ideal (lane 2), TCRδ (lane 3) or GM450Δ443/Δ452 (lane 4) unlabelled oligonucleotides. (F) EMSAs with 6 μg of Jurkat cell nuclear extract bound to the Ideal probe (lanes 1 and 7) and competed with increasing amounts of either unlabelled ideal (lanes 2–6) or IdealΔ1 (lanes 8–12) oligonucleotides. (G) Densitometric quantification of the EMSAs in (F) showing the proportion of the monomeric and dimeric complexes remaining in the presence of competing Ideal or mutated IdealΔ1 oligonucleotides.

both strands (designated as Ideal in Figures 2A). Both the GM450 and Crystal probes supported the binding of what appeared to be monomeric and dimeric CBF complexes (Figure 2C). This conclusion was supported by the fact that a mutation of the 443 site led to disappearance of the lower mobility complex.

The Ideal palindromic Runx1 consensus probe also formed two distinct complexes that were assumed to be CBF monomer and dimer complexes (lane 1 in Figure 2D and E). The designated CBF monomer and dimer complexes were both confirmed as specific Runx1 complexes because they (i) were both super-shifted by two different Runx1 antibodies (lanes 2 and 3, Figure 2D), but not by control IgG (lane 4, Figure 2D), and (ii) were both inhibited by oligonucleotides containing either the Ideal Runx1 palindrome or a single Runx1 site from the TCRδ locus (lanes 2 and 3, Figure 2E), but not by GM450 oligonucleotides in which both Runx1 sites had been mutated (Δ443/Δ452, lane 4, Figure 2E). Taken together, the *in vitro* and *in vivo* binding studies both support the conclusion that intact Runx1/CBFβ complexes can efficiently assemble as a dimeric complex on palindromic Runx1 binding sites. Note that in this structure the Runx1 Rel-like domains contact palindromic GG elements in a conformation, and at a spacing (i.e. GGNNNNCC) similar to that that observed for the Rel domains within NFAT dimers and NF-κB complexes, which typically bind to GGNNNNNCC elements.

Although the assembly of CBF dimers on DNA is not as cooperative as either NFAT or NF-κB proteins (20), we attempted to address the issue of cooperativity by testing different probes as competitors (Figure 2E and F). To this end, we studied the binding of native CBF complexes to the Ideal palindromic probe in the presence of oligonucleotides containing either the Ideal palindromic sequence (lanes 2–6, Figure 2F) or the Ideal sequence with the second Runx1 site mutated (IdealΔ1, lanes 8–12, Figure 2F). The percentage of the monomer and dimer complexes remaining in the presence of increasing amounts of competitor was calculated by densitometry and plotted in Figure 2G. This analysis revealed that the binding properties of two Runx1 sites in one sequence are not just additive, but that a dimeric element is a substantially stronger competitor than a similar element containing a single binding site. This raises the possibility that CBF dimer formation may be assisted by interactions between Runx1 and/or CBFβ.

Both TGTGG motifs contribute to GM-CSF enhancer function

To study the functions of the two TGTGG motifs at positions 443 and 452 in the GM-CSF enhancer, we performed transient transfection assays in the CEM and Jurkat T cell lines. These assays used the previously defined luciferase plasmid pGM-GME (26) containing a 717 bp Bgl II fragment of the human GM-CSF enhancer inserted upstream of a −627 to +28 segment of the GM-CSF promoter in the luciferase vector pXPG (41). This plasmid was assayed in parallel with derivatives of pGM-GME containing mutations in either one or both of the position 443 and 452 TGTGG motifs, plus the plasmid pGM containing just the GM-CSF promoter (26). To control for both transfection and stimulation efficiency at the same time, we co-transfected cells with the plasmid pXRL-GME. This plasmid was created from pGM-GME by excising the firefly luciferase gene and replacing it with the Renilla luciferase gene.

Firefly and Renilla luciferase plasmids were co-transfected into cells that were then cultured for ∼21 h, before stimulating the cells for 8 hours with 20 ng/ml PMA and 2 μM calcium ionophore A23187, and harvesting of extracts for dual luciferase assays. We first confirmed that there was no unwanted cross-contamination of activities between firefly and Renilla luciferase assays by transfecting Jurkat cells with each plasmid alone and performing dual luciferase assays (data not shown). We next demonstrated that both the pGM-GME and pXRL-GME plasmids were induced by a factor of ∼100-fold upon stimulation (data not shown) and showed that inclusion of the enhancer increased pGM promoter activity by about 5 to 10 fold in CEM and Jurkat cells (Figure 3). The introduction of single GG to CC mutations within the position 443 or 452 motifs led to a ∼3-fold reduction of activity in CEM cells and to a ∼2-fold reduction of activity in Jurkat cells (Figure 3). The double mutation of both the position 443 and 452 TGTGG motifs essentially abolished enhancer activity in CEM cells, and reduced activity of the enhancer in Jurkat cells to ∼2-fold the activity of the promoter alone (Figure 3). Hence, we have confirmed that both Runx1 sites are required for efficient inducible enhancer function.

Computational modelling of a dimeric CBF complex on a palindromic sequence

To determine (i) how overlapping Runx1 motifs might accommodate two CBF complexes, and (ii) whether dimer formation is potentially supported by interactions between the two CBF complexes, we generated a computer model of the dimer incorporating the Crystal palindrome defined in Figure 2. For this purpose, we mapped the Runx1-CBFβ coordinates determined by Bravo *et al* for the single Runx1–CBFβ–DNA complex (11) onto the two Runx1 motifs within GM-CSF enhancer sequence ATGTGTGGCTGCCCACAAAAC. The core of the DNA binding regions, and in particular the GG core elements, were aligned to those in the palindromic sequence in order to preserve the geometry of the interaction between Runx1 and DNA. Two major assumptions were made when modelling the structure of the dimeric CBF complex: first, that each individual CBF complex will recognize the DNA binding region comparably to the monomeric form as reported by Bravo *et al*., and second, that each individual CBF complex would not undergo major conformation changes when forming the dimeric CBF complex.

The raw structural model of the dimeric CBF complex presented some minor steric clashes between the two CBF complexes. In particular, some atoms of residues located in the interface between the CBFβ subunits and the βE′-F

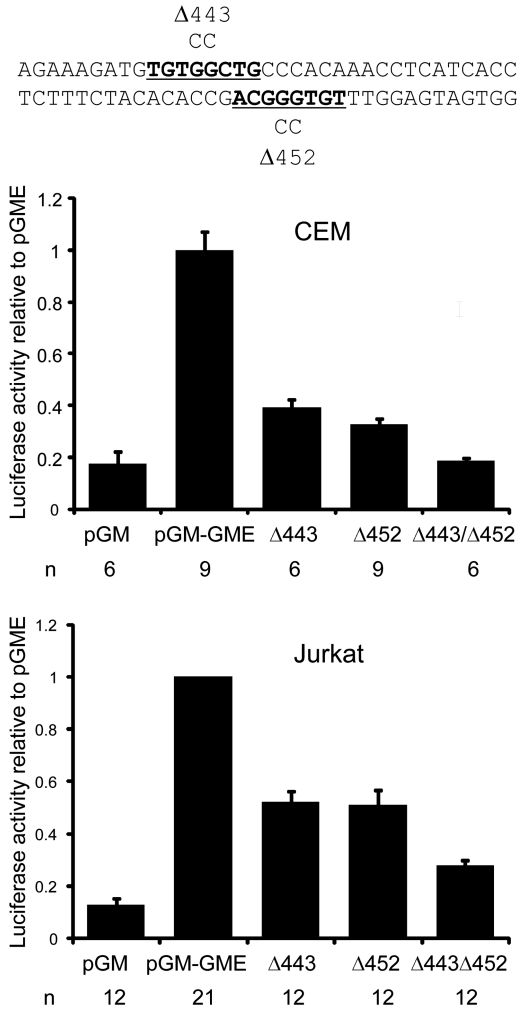

Figure 3. Transient transfection assays of GM-CSF enhancer function. Firefly luciferase reporter gene plasmids were transfected into CEM and Jurkat T cells. Cells were assayed after stimulation for 8 h with PMA and calcium ionophore. Plasmids contained either, just the GM-CSF promoter (pGM) or the promoter plus the upstream enhancer (pGM-GME), with or without specific mutations in either or both of the TGTGG motifs located at positions 443 and 452. Plasmids were cotransfected with the Renilla luciferase control plasmid pXRL-GME. The data represent the averages of data obtained from 3 to 6 independent transfections of each of 2 to 4 independent plasmid preparations for each construct. The number of independent transfections is indicated below each column (*n*). Error bars indicate the standard error.

loop [as reported in Bravo *et al.* paper (11)] were at an atomic distance inferior to the sum of van der Waals radii (data not shown). The reasons for this are because (i) during the modelling of the complex, the DNA molecule containing the palindromic sequences was modelled as straight B-DNA configuration, and was kept rigid, and (ii) the interactions between the Runx1 domains and DNA was maintained in the same conformation as in the crystal structure (11). However, when the complex was subjected to an energy minimization step and a short dynamic simulation, i.e. protein and DNA atoms were allowed to relax and change conformation, these steric clashes were largely resolved. PROCHECK did not report any major stereochemical problems in the minimized complex. Furthermore, the model of the complex remained stable throughout the simulation and the total energy of the complex converged to a minimum. From the simulation, we were able to derive several clusters of structures, and select representative examples of each. From these, we selected the structure of the complex that provided the best G-factor in PROCHECK, and this is presented in Supplementary Data File 1.

As represented in the model of this structure in Figure 4A, two CBF complexes can comfortably bind simultaneously to the DNA binding sites without any major steric impediment. The two GG core sequences of the DNA motifs are separated by one half turn of helix and exist in opposite orientations. This means that each Runx1 domain is rotated 180 degrees relative to the other with respect to its interaction with the DNA binding site (Figures 4A and B). The atomic analysis of this theoretical complex reveals that atomic interactions between the C terminal regions of the Runt domains and their DNA binding sites are similar to the ones reported in the crystal structure, including Arg80, Arg142, Thr 169, Asp 171, Arg174 and Arg177 [residue numbers as reported in Bravo *et al.* paper (11); (Figure 4B)]. The model also suggests that the dimer may be maintained by stabilizing interactions between the two molecules of CBFβ (Figure 4C). The proposed structural model predicts favourable interactions between anti-parallel domains of CBFβ that allow optimal electrostatic pairing between polar residues (Figure 4D).

DISCUSSION

It was previously assumed that Runx1 binding to DNA principally involved the interaction of a single Runt domain with one isolated TGYGG core sequence (5,6). However, we have made the novel observation that two molecules of Runx1 can assemble together on palindromic sequences containing two closely spaced inverted repeats of a TGYGG core sequence separated by 4 bp. In this arrangement, there is a 2 bp overlap between the 8 bp TGYGGNNN DNA sequences contacted by the individual Runx1 complexes. Runx1 bound most efficiently as a dimeric complex to DNA elements conforming to the consensus TGTGGTTAACCACA or TGTGGTCGACCAC A. In this configuration, each of the repeated motifs closely matches the TGTGGTTT or TGTGGTCA consensus sequences that represent the most commonly encountered Runx1 binding motifs. Binding of CBF to the Ideal consensus palindromic Runx1 site TGTGG TTAACCACA was more efficient than to the GM450 element, which lacks the preferred TGYGGT motifs.

This work is to our knowledge the first study that reports the potential dimerization of CBF complexes on DNA in a conformation similar to that observed for dimers of the Rel homology domains of NFAT and NF-κB family proteins. We are only aware of one other report of Runx1 binding to a palindromic element (42). However, in this instance, only one of the two TGCGG

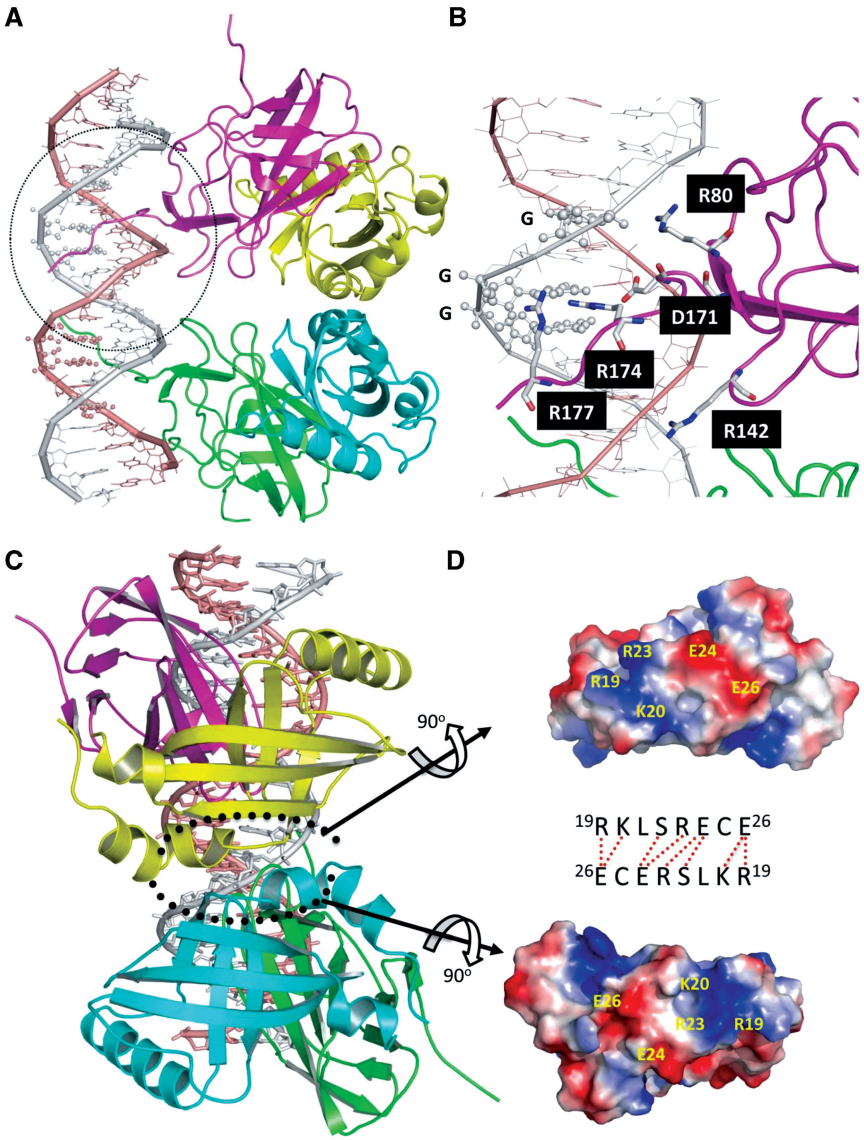

Figure 4. Structural model of the dimeric form of CBF complex bound to the palindromic DNA sequence ATGTGTGGCTGCCCACAAAAC, predicted from the crystal structure coordinates published by Bravo *et al.* for a single CBF:DNA complex. (**A**) Ribbon representation of the complex from a lateral view using DNA as the reference point. Runx1 domains are depicted in magenta and green, and CBFβ domains in yellow and cyan. (**B**) Close-up view of the DNA binding site encircled in (A). Proteins and DNA are shown in ribbon and residues in direct contact with DNA labelled in black boxes. The G bases within the core consensus TGYGG are labelled on the strand contacted by the indicated amino acids. (C) Ribbon representation of the complex from a frontal view using CBFβ as a the reference. (**D**) Surface potentials calculated for the interaction surfaces of the CBFβ domains. Positive charge is shown as blue, and negative charge is shown as red. Residues that potentially mediate the interactions between the two CBFβ domains are labelled in yellow. The predicted polar interactions between the antiparallel sequences are indicated by the dashed lines between the linear representation of the sequences. In this view, the interacting surfaces depicted within bracketed regions in (C) have been rotated 90°.

motifs present within the sequence TGCGGAGACCGCA actually contributed to Runx1 binding (underlined), and there was no evidence for the assembly of dimeric CBF complexes. It has also been reported that the t(8;21) chromosomal translocation product Runx1-Eto is able to bind to direct repeats of Runx1 consensus motifs as found in the consensus sequence TGYGGTTN$_{(0–13)}$TGC GGT (42). In this situation, the mechanism of dimerization is likely to be different to that showed in our study since in that report it was proposed that oligomerization was mediated by Eto and not by Runx1.

As mentioned above, Runx1 is a distant relative of the Rel and NFAT families of transcription factors, and it shares with Rel domain factors the ability to recognize GG core sequences. Runx1 is also a distant relative of p53 which similarly binds as a dimer to palindromic sequences such as GGGCATGCCC (43) which have a conformation resembling both the NF-κB consensus and the dimeric Runx1 site described here. While Runx1 resembles NF-κB and NFAT with respect to its ability to bind simultaneously to both copies of an inverted repeat of the GG motifs, there are significant differences in the nature of the

intervening sequences that they recognize. Whereas NFAT (18,20,21) and NF-κB (10) each bind as dimers to sequences resembling GGAAATTCC or GGAGACT CC, Runx1 assembles as a dimer on sequences resembling GGTTAACC. Runx1 also differs substantially from NF-κB with regard to the fact that NF-κB is not able to use a single Rel domain to bind to a single GG motif. In this respect, Runx1 has more in common with NFAT than with NF-κB. Runx1 and NFAT both frequently cooperate with other classes of transcription factor, but still retain the ability to bind to DNA efficiently to a single GG core element via a single Rel-like DNA-binding domain.

Based on the *in vitro* binding of purified truncated proteins, there was no evidence for cooperativity in the binding of the recombinant CBF complexes that contain just the Runt domain of Runx1. Nevertheless, it was apparent from the *in vivo* footprinting that both of the Runx1 sites within the GM450 element were efficiently occupied simultaneously inside activated cells at about half of all sites, raising the possibility of cooperativity in the assembly of the dimeric complex. Furthermore, assembly of dimeric Runx1 complexes appeared to be more efficient with full length naturally occurring Runx1/CBFβ complexes present in nuclear extracts than with the truncated recombinant Runx1/CBFβ. Additional evidence for cooperativity was provided by competition studies showing that a dimeric site was a substantially stronger binding site than an equivalent single binding site. This may be because Runx1 also contains a homodimerization domain at the C terminus that is absent in the recombinant Runx1 (17). It has been previously established that this interaction domain promotes Runx1:Runx1 interactions between separated Runx1 sites. This same domain could also potentially help to stabilize the dimeric complex and support cooperative binding at palindromic Runx1 sites. The structure of the C terminal region of Runx1 is unknown, partly because much of it is unstructured, and partly due to technical difficulties associated with preparing full-length recombinant Runx1. However, from the proposed model structure of the dimer it is evident that the C terminal portions of the Runt domains cross over the DNA helices on opposite sides of the DNA. Hence, this allows for the possibility that the CBF dimer encircles the DNA by forming close CBFβ:CBFβ interactions on one side, and Runx1 C terminal interactions on the opposite side of the helix. If the C-terminal region does contain unstructured domains, it is likely to be flexible enough to accomplish this. It is also likely that within the nucleus, the dimeric CBF complex exists as a higher order complex with multiple other regulatory factors that would act to further stabilize such a complex. Further genome-wide analyses will be required to determine how often Runx1 sites within the genome do in fact function as dimeric sites.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Alan Warren for advice and for providing purified recombinant Runx1/CBFβ protein complexes. We thank Jenny Barton for assistance and advice with Runx1 binding assays.

FUNDING

Yorkshire Cancer Research, Leukaemia and Lymphoma Research; Biotechnology and Biological Sciences Research Council. Funding for open access charge: Research grants.

Conflict of interest statement. None declared.

REFERENCES

1. Kagoshima,H., Shigesada,K., Satake,M., Ito,Y., Miyoshi,H., Ohki,M., Pepling,M. and Gergen,P. (1993) The Runt domain identifies a new family of heteromeric transcriptional regulators. *Trends Genet*., **9**, 338–341.
2. Speck,N.A. and Terryl,S. (1995) A new transcription factor family associated with human leukemias. *Crit. Rev. Eukaryot. Gene Expr*., **5**, 337–364.
3. Lacaud,G., Gore,L., Kennedy,M., Kouskoff,V., Kingsley,P., Hogan,C., Carlsson,L., Speck,N., Palis,J. and Keller,G. (2002) Runx1 is essential for hematopoietic commitment at the hemangioblast stage of development in vitro. *Blood*, **100**, 458–466.
4. Okuda,T., van Deursen,J., Hiebert,S.W., Grosveld,G. and Downing,J.R. (1996) AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell*, **84**, 321–330.
5. Melnikova,I.N., Crute,B.E., Wang,S. and Speck,N.A. (1993) Sequence specificity of the core-binding factor. *J. Virol*., **67**, 2408–2411.
6. Golemis,E.A., Speck,N.A. and Hopkins,N. (1990) Alignment of U3 region sequences of mammalian type C viruses: identification of highly conserved motifs and implications for enhancer design. *J. Virol*., **64**, 534–542.
7. Thornell,A., Hallberg,B. and Grundstrom,T. (1991) Binding of SL3-3 enhancer factor 1 transcriptional activators to viral and chromosomal enhancer sequences. *J. Virol*., **65**, 42–50.
8. Berardi,M.J., Sun,C., Zehr,M., Abildgaard,F., Peng,J., Speck,N.A. and Bushweller,J.H. (1999) The Ig fold of the core binding factor alpha Runt domain is a member of a family of structurally and functionally related Ig-fold DNA-binding domains. *Structure*, **7**, 1247–1256.
9. Nagata,T., Gupta,V., Sorce,D., Kim,W.Y., Sali,A., Chait,B.T., Shigesada,K., Ito,Y. and Werner,M.H. (1999) Immunoglobulin motif DNA recognition and heterodimerization of the PEBP2/CBF Runt domain. *Nat. Struct. Biol*., **6**, 615–619.
10. Kunsch,C., Ruben,S.M. and Rosen,C.A. (1992) Selection of optimal kappa B/Rel DNA-binding motifs: interaction of both subunits of NF-kappa B with DNA is required for transcriptional activation. *Mol. Cell Biol*., **12**, 4412–4421.
11. Bravo,J., Li,Z., Speck,N.A. and Warren,A.J. (2001) The leukemia-associated AML1 (Runx1)—CBF beta complex functions as a DNA-induced molecular clamp. *Nat. Struct. Biol*., **8**, 371–378.
12. Tahirov,T.H., Inoue-Bungo,T., Morii,H., Fujikawa,A., Sasaki,M., Kimura,K., Shiina,M., Sato,K., Kumasaka,T., Yamamoto,M. *et al.* (2001) Structural analyses of DNA recognition by the AML1/Runx-1 Runt domain and its allosteric control by CBFbeta. *Cell*, **104**, 755–767.
13. Giese,K., Kingsley,C., Kirshner,J.R. and Grosschedl,R. (1995) Assembly and function of a TCR alpha enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes Dev*., **9**, 995–1008.
14. Kim,W.Y., Sieweke,M., Ogawa,E., Wee,H.J., Englmeier,U., Graf,T. and Ito,Y. (1999) Mutual activation of Ets-1 and AML1

DNA binding by direct interaction of their autoinhibitory domains. *EMBO J.*, **18**, 1609–1620.
15. Zhang,D.E., Hetherington,C.J., Meyers,S., Rhoades,K.L., Larson,C.J., Chen,H.M., Hiebert,S.W. and Tenen,D.G. (1996) CC AAT enhancer-binding protein (C/EBP) and AML1 (CBF alpha2) synergistically activate the macrophage colony-stimulating factor receptor promoter. *Mol. Cell Biol.*, **16**, 1231–1240.
16. Hernandez-Munain,C. and Krangel,M.S. (1995) c-Myb and core-binding factor/PEBP2 display functional synergy but bind independently to adjacent sites in the T-cell receptor delta enhancer. *Mol. Cell Biol.*, **15**, 3090–3099.
17. Li,D., Sinha,K.K., Hay,M.A., Rinaldi,C.R., Saunthararajah,Y. and Nucifora,G. (2007) RUNX1-RUNX1 homodimerization modulates RUNX1 activity and function. *J. Biol. Chem.*, **282**, 13542–13551.
18. Hogan,P.G., Chen,L., Nardone,J. and Rao,A. (2003) Transcriptional regulation by calcium, calcineurin, and NFAT. *Genes Dev.*, **17**, 2205–2232.
19. Cockerill,G.W., Bert,A.G., Ryan,G.R., Gamble,J.R., Vadas,M.A. and Cockerill,P.N. (1995) Regulation of granulocyte-macrophage colony-stimulating factor and E-selectin expression in endothelial cells by cyclosporin A and the T-cell transcription factor NFAT. *Blood*, **86**, 2689–2698.
20. Falvo,J.V., Lin,C.H., Tsytsykova,A.V., Hwang,P.K., Thanos,D., Goldfeld,A.E. and Maniatis,T. (2008) A dimer-specific function of the transcription factor NFATp. *Proc. Natl Acad. Sci. USA*, **105**, 19637–19642.
21. Shang,C., Attema,J., Cakouros,D., Cockerill,P.N. and Shannon,M.F. (1999) Nuclear factor of activated T cells contributes to the function of the CD28 response region of the granulocyte macrophage-colony stimulating factor promoter. *Int. Immunol.*, **11**, 1945–1956.
22. Cockerill,P.N., Osborne,C.S., Bert,A.G. and Grotto,R.J. (1996) Regulation of GM-CSF gene transcription by core-binding factor. *Cell Growth Differ.*, **7**, 917–922.
23. Frank,R., Zhang,J., Uchida,H., Meyers,S., Hiebert,S.W. and Nimer,S.D. (1995) The AML1/ETO fusion protein blocks transactivation of the GM-CSF promoter by AML1B. *Oncogene*, **11**, 2667–2674.
24. Takahashi,A., Satake,M., Yamaguchi-Iwai,Y., Bae,S.C., Lu,J., Maruyama,M., Zhang,Y.W., Oka,H., Arai,N., Arai,K. *et al.* (1995) Positive and negative regulation of granulocyte-macrophage colony-stimulating factor promoter activity by AML1-related transcription factor, PEBP2. *Blood*, **86**, 607–616.
25. Cockerill,P.N., Shannon,M.F., Bert,A.G., Ryan,G.R. and Vadas,M.A. (1993) The granulocyte-macrophage colony-stimulating factor/interleukin 3 locus is regulated by an inducible cyclosporin A-sensitive enhancer. *Proc. Natl Acad. Sci. USA*, **90**, 2466–2470.
26. Cockerill,P.N., Bert,A.G., Roberts,D. and Vadas,M.A. (1999) The human GM-CSF gene is autonomously regulated in vivo by an inducible tissue-specific enhancer. *Proc. Natl Acad. Sci. USA*, **96**, 15097–15102.
27. Cockerill,P.N. (2004) Mechanisms of transcriptional regulation of the human IL-3/GM-CSF locus by inducible tissue-specific promoters and enhancers. *Crit. Rev. Immunol.*, **24**, 385–408.
28. Johnson,B.V., Bert,A.G., Ryan,G.R., Condina,A. and Cockerill,P.N. (2004) GM-CSF enhancer activation requires cooperation between NFAT and AP-1 elements and is associated with extensive nucleosome reorganization. *Mol. Cell Biol.*, **24**, 7914–7930.
29. Bert,A.G., Johnson,B.V., Baxter,E.W. and Cockerill,P.N. (2007) A modular enhancer is differentially regulated by GATA and NFAT elements that direct different tissue-specific patterns of nucleosome positioning and inducible chromatin remodeling. *Mol. Cell Biol.*, **27**, 2870–2885.
30. Mirabella,F., Baxter,E.W., Boissinot,M., James,S.R. and Cockerill,P.N. (2010) The human IL-3/granulocyte-macrophage colony-stimulating factor locus is epigenetically silent in immature thymocytes and is progressively activated during T cell development. *J. Immunol.*, **184**, 3043–3054.
31. Tagoh,H., Cockerill,P.N. and Bonifer,C. (2006) In vivo genomic footprinting using LM-PCR methods. *Methods Mol. Biol.*, **325**, 285–314.
32. Bowers,S.R., Mirabella,F., Calero-Nieto,F.J., Valeaux,S., Hadjur,S., Baxter,E.W., Merkenschlager,M. and Cockerill,P.N. (2009) A conserved insulator that recruits CTCF and cohesin exists between the closely related but divergently regulated interleukin-3 and granulocyte-macrophage colony-stimulating factor genes. *Mol. Cell Biol.*, **29**, 1682–1693.
33. Cockerill,P.N., Bert,A.G., Roberts,D. and Vadas,M.A. (1999) The human granulocyte-macrophage colony-stimulating factor gene is autonomously regulated in vivo by an inducible tissue-specific enhancer. *Proc. Natl Acad. Sci. USA*, **96**, 15097–15102.
34. Bert,A.G., Burrows,J., Hawwari,A., Vadas,M.A. and Cockerill,P.N. (2000) Reconstitution of T cell-specific transcription directed by composite NFAT/Oct elements. *J. Immunol.*, **165**, 5646–5655.
35. Warren,A.J., Bravo,J., Williams,R.L. and Rabbitts,T.H. (2000) Structural basis for the heterodimeric interaction between the acute leukaemia-associated transcription factors AML1 and CBFbeta. *EMBO J.*, **19**, 3004–3015.
36. Macke,T.J. and Case,D.A. (1998) *Modeling Unusual Nucleic Acid Structures*. American Chemistry Society, Washington DC.
37. Sali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779.
38. Case,D.A., Darden,T.E., Cheatham,T.E. III, Simmerling,C., Wang,J., Duke,R.E., Luo,R., Crowley,M., Ross,C., Walker,W. *et al.* (2008) *AMBER 10*. University of California, San Francisco.
39. Hornak,V., Abel,R., Okur,A., Strockbine,B., Roitberg,A. and Simmerling,C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, **65**, 712–725.
40. Laskowski,R.A., Rullmannn,J.A., MacArthur,M.W., Kaptein,R. and Thornton,J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
41. Bert,A.G., Burrows,J., Osborne,C.S. and Cockerill,P.N. (2000) Generation of an improved luciferase reporter gene plasmid that employs a novel mechanism for high-copy replication. *Plasmid*, **44**, 173–182.
42. Okumura,A.J., Peterson,L.F., Okumura,F., Boyapati,A. and Zhang,D.E. (2008) t(8;21)(q22;q22) Fusion proteins preferentially bind to duplicated AML1/RUNX1 DNA-binding sequences to differentially regulate gene expression. *Blood*, **112**, 1392–1401.
43. Kitayner,M., Rozenberg,H., Kessler,N., Rabinovich,D., Shaulov,L., Haran,T.E. and Shakked,Z. (2006) Structural basis of DNA recognition by p53 tetramers. *Mol. Cell*, **22**, 741–753.