



Molecular architecture of CTCFL

Amy E. Campbell¹, Selena R. Martinez, JJ L. Miranda*

Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94158, USA

ARTICLE INFO

Article history:

Received 16 April 2010

Available online 8 May 2010

Keywords:

CTCFL
Chromatin
Transcription
Insulator
CTCF

ABSTRACT

The CTCF-like protein, CTCFL, is a DNA-binding factor that regulates the transcriptional program of mammalian male germ cells. CTCFL consists of eleven zinc fingers flanked by polypeptides of unknown structure and function. We determined that the C-terminal fragment predominantly consists of extended and unordered content. Computational analysis predicts that the N-terminal segment is also disordered. The molecular architecture of CTCFL may then be similar to that of its paralog, the CCCTC-binding factor, CTCF. We speculate that sequence divergence in the unstructured terminal segments results in differential recruitment of cofactors, perhaps defining the functional distinction between CTCF in somatic cells and CTCFL in the male germ line.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

The CTCF-like protein, referred to as CTCFL or BORIS, regulates transcription in the human male germ line. The CCCTC-binding factor, CTCF, organizes transcription and chromatin in the three-dimensional space of the nucleus [1,2]. While mammalian CTCF is ubiquitously expressed in somatic tissue, CTCFL is a paralog expressed only during male germ cell development [3,4]. Binding of CTCFL to differently methylated alleles suggests a role in regulating imprinting [5–7]. Indeed, evolution of testis-specific expression appears to correlate with the emergence of imprinting in therian mammals [4]. Displacement of CTCF by induced CTCFL in somatic cells perturbs gene expression [8–10]. Given the temporal and spatial separation of antagonistic functions, each paralog likely drives distinct transcriptional programs by using dissimilar mechanisms during development.

How do CTCF and CTCFL differentially regulate gene expression? Both proteins contain eleven central zinc fingers that are almost identical between the paralogs. The flanking polypeptide segments, which comprise about half of each protein, show very little sequence homology between CTCF and CTCFL [3,4]. One might therefore suspect that the divergent functions of each protein arise from differences in the terminal extensions. Our efforts to elucidate the structures of these polypeptides revealed that both fragments of CTCF are unstructured [11]. We have now learned that the C-terminal fragment of CTCFL is also unstructured.

Computational analysis predicts that the N-terminal segment may be disordered as well. We discuss the functional implications of a similar structural architecture on the functional differences between CTCF and CTCFL.

2. Materials and methods

2.1. Protein expression and purification

The terminal segments of CTCFL were defined as the polypeptides flanking the eleven zinc fingers predicted by a minimum consensus motif [12]. The gene fragment encoding the C-terminal segment was obtained by PCR amplification from human cDNA clone SC311151 (Origene). This amplicon contained an NdeI site, the sequence encoding amino acids 571–663, the sequence TGGAGCCACCCGAGTTCGAAAAA that encodes the Strep-tag WSHPPQFEK, a TAA stop codon, and a BamHI site, all of which was inserted between the NdeI and BamHI sites of pET-15b (EMD Biosciences) to generate pAEC38. Additional sequences flanking the gene fragment were included by incorporation in PCR primers. The construct encodes both an N-terminal His-tag and a C-terminal Strep-tag. Sequencing of the open reading frame (UCSF Genomics Core Facility) confirmed proper cloning of the reference sequence [3]. The C-terminal fragment of CTCFL was expressed in bacteria and purified with affinity and size-exclusion chromatography in the same manner as the C-terminal fragment of CTCF [11].

2.2. Biochemistry

Hydrodynamic radii were determined using size-exclusion chromatography [11]. Circular dichroism was measured with a

* Corresponding author.

E-mail address: miranda@cmp.ucsf.edu (JJ L. Miranda).

¹ Present addresses: Division of Hematology, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA.

J-715 spectropolarimeter (Jasco) equipped with a PTC-348WI Peltier temperature control system (Jasco). Purified protein was exchanged into 25 mM sodium phosphate, 250 mM NaF, 1 mM β -mercaptoethanol, pH 7.5 by repeated concentration and dilution in a centrifugal filter unit. Spectra were acquired with 10 μ M protein in a 1 mm cuvette (Hellma) at 20 °C. The spectrum of buffer alone was also acquired and subtracted from the spectrum of the protein solution. Fractional secondary structure content was estimated with the CONTIN/LL and CDSSTR algorithms of CDPro [13] using data from 200 to 240 nm compared to reference set SDP48.

2.3. Computational analysis

Disordered regions of CTCFL were predicted by examining the protein primary sequence. Analysis was performed with the PONDR VL-XT algorithm [14].

3. Results and discussion

3.1. Biophysical properties of the CTCFL C-terminal fragment

The hydrodynamic radius of the CTCFL C-terminal fragment is larger than that expected for a globular protein. The recombinant C-terminal polypeptide can be purified from bacteria as a single migrating band on an SDS-PAGE gel (Fig. 1A) and as a single eluting peak during size-exclusion chromatography (Fig. 1B). Because of previous experience with unstructured fragments of CTCF, we tested this CTCFL polypeptide for behavior that suggests the lack of a folded domain. We therefore measured hydrodynamic proper-

ties with size-exclusion chromatography (Fig. 1B). Unstructured proteins have larger Stokes radii than compact folds because extended conformations generate more drag in solution. The CTCFL C-terminal fragment yields a Stokes radius of 26.2 ± 0.3 Å ($N = 6$). A 13 kDa protein is expected to migrate with a radius of 19 Å if globular and 32–34 Å if unfolded [15]. We excluded oligomerization as a cause of increased drag; the molar mass of the purified polypeptide as observed by multi-angle light scattering coupled with size-exclusion chromatography is close to that expected for a monomer (data not shown). A larger than expected hydrodynamic radius suggests that the C-terminal fragment could be unstructured, but determination of the amount of secondary structure is needed to verify this deduction.

The C-terminal fragment of CTCFL is predominantly unordered. We estimated secondary structure content by measuring circular dichroism. The far UV spectrum (Fig. 2) contains a strong minimum at ~ 199 nm and is very similar to that of random coils, which contains a minimum at ~ 197 nm [16]. α helices or β strands yield distinct minima above 200 nm, none of which are seen here. We then calculated the fractional secondary structure content by comparison to proteins of known structures [13]. The CONTIN/LL algorithm estimated $4 \pm 0\%$ helix, $12 \pm 2\%$ strand, $6 \pm 1\%$ turns, and $78 \pm 2\%$ unordered content ($N = 3$). Analysis with the CDSSTR algorithm calculated similar results, estimating $3 \pm 0\%$ helix, $15 \pm 2\%$ strand, $10 \pm 1\%$ turns, and $71 \pm 4\%$ unordered content ($N = 3$). We note that while determining the exact unordered content of polypeptides suffers from multiple computational challenges, the rough estimates are informatively accurate [17,18]. Our observations can be interpreted as concluding that the C-terminal fragment consists mostly of unstructured content.

3.2. Sequence analysis of the CTCFL N-terminal segment

Computational analysis predicts that the N-terminal segment of CTCFL may be disordered. We could not purify native preparations of the N-terminal fragment sufficiently well behaved for biochemical experiments. Although we would have preferred to obtain experimental data, we consequently turned to computational methods to identify unstructured portions of CTCFL. Sequence analysis trained on known ordered and disordered polypeptides [14] predicts that a significant portion of the N-terminal segment may be unstructured (Fig. 3). Correct identification of disorder in the rest of CTCFL gives us confidence in this analysis. The C-terminal segment is predicted to be unstructured as experimentally observed. The probability of disorder in each zinc finger is appropriately low, but much higher in the linkers between some domains. Moreover, this same algorithm also correctly identifies the ordered and disordered regions of CTCF (data not shown). If one accepts the computational prediction, then both terminal segments of CTCFL may be unstructured.

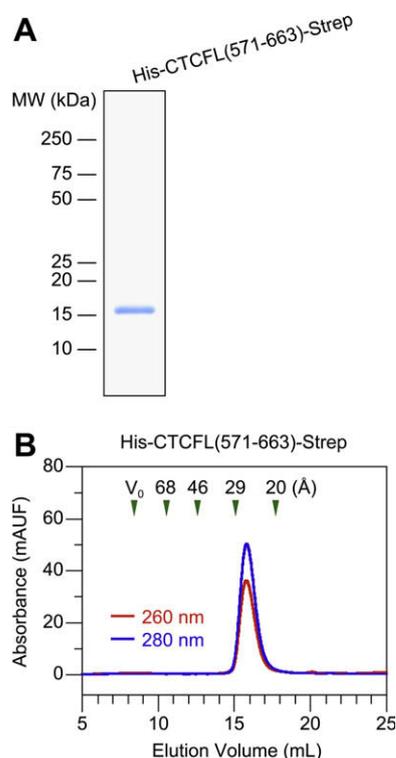


Fig. 1. Purification of the CTCFL C-terminal fragment. (A) Purified protein resolved on a 10–20% SDS-PAGE gel. Positions of molecular mass standards are indicated on the left. (B) Size-exclusion chromatography. Red and blue lines represent UV absorbance at 260 and 280 nm, respectively. V_0 denotes the excluded void volume. Green triangles mark elution times of standards with known Stokes radii.

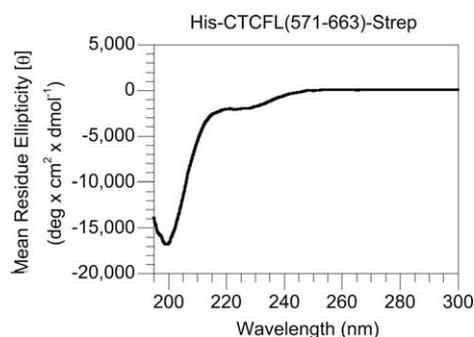


Fig. 2. Circular dichroism of the CTCFL C-terminal fragment. Far UV spectrum.

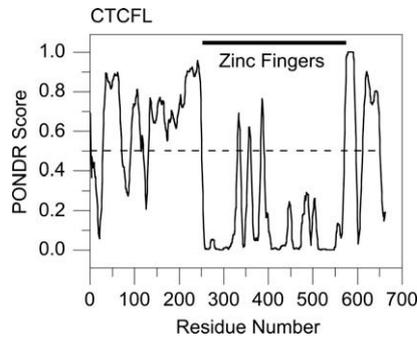


Fig. 3. Computational prediction of disordered regions in CTCFL. Analysis with the PONDNR VL-XT algorithm. Scores above the threshold value of 0.5, shown as a dashed line, predict disorder at a given position. A solid bar indicates the position of the zinc finger domains.

3.3. Functional implications of CTCFL molecular architecture

The unstructured terminal segments of CTCF and CTCFL could recruit different binding partners to chromosomes. With CTCF, we found that no domains, defined biochemically as autonomously folding units of stable secondary structure, exist in the terminal extensions [11]. Such an architecture limits possible functions of the terminal segments to that observed with other unstructured polypeptides, which is predominantly molecular recognition [19]. With CTCFL, the C-terminal fragment consists mostly of unordered content, and the N-terminal segment is predicted to be disordered. CTCF and CTCFL thus appear to have a similar molecular architecture, eleven zinc fingers flanked by unstructured extensions. The zinc fingers are strongly conserved between paralogs, but the terminal segments are significantly diverged [3,4]. Lack of sequence homology suggests that if the unstructured extensions of CTCF and CTCFL function in molecular recognition as we surmise, then perhaps each paralog binds different proteins to mediate distinct effects on chromatin and transcription. Although both factors may bind the same DNA sequence, the two could recruit dissimilar activities to each site. We hypothesize that differential assembly of regulatory proteins by the unstructured terminal segments of CTCF and CTCFL could explain how each paralog directs specific genetic programs in somatic cells and the male germ line.

Acknowledgments

We thank Meghan M. Holdorf for insightful discussions and Brian K. Shoichet for use of the circular dichroism spectropolarimeter. Our research is supported by an institutional grant to JJ L. Miranda from the UCSF Fellows Program, which is funded in part by the UCSF Program for Breakthrough Biomedical Research and the Sandler Foundation.

References

- [1] V.V. Lobanekov, R.H. Nicolas, V.V. Adler, H. Paterson, E.M. Klenova, A.V. Polotskaja, G.H. Goodwin, A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene, *Oncogene* 5 (1990) 1743–1753.
- [2] J.Q. Ling, T. Li, J.F. Hu, T.H. Vu, H.L. Chen, X.W. Qiu, A.M. Cherry, A.R. Hoffman, CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1, *Science* 312 (2006) 269–272.
- [3] D.I. Loukinov, E. Pugacheva, S. Vatolin, S.D. Pack, H. Moon, I. Chernukhin, P. Mannan, E. Larsson, C. Kanduri, A.A. Vostrov, H. Cui, E.L. Niemitz, J.E. Rasko, F.M. Docquier, M. Kistler, J.J. Breen, Z. Zhuang, W.W. Quitschke, R. Renkawitz, E.M. Klenova, A.P. Feinberg, R. Ohlsson, H.C. Morse 3rd, V.V. Lobanekov, BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma, *Proc. Natl. Acad. Sci. USA* 99 (2002) 6806–6811.
- [4] T.A. Hore, J.E. Deakin, J.A. Marshall Graves, The evolution of epigenetic regulators CTCF and BORIS/CTCF in amniotes, *PLoS Genet.* 4 (2008) e1000169.
- [5] P. Jelinic, J.C. Stehle, P. Shaw, The testis-specific factor CTCFL cooperates with the protein methyltransferase PRMT7 in H19 imprinting control region methylation, *PLoS Biol.* 4 (2006) e355.
- [6] P. Nguyen, G. Bar-Sela, L. Sun, K.S. Bisht, H. Cui, E. Kohn, A.P. Feinberg, D. Gius, BAT3 and SET1A form a complex with CTCFL/BORIS to modulate H3K4 histone dimethylation and gene expression, *Mol. Cell. Biol.* 28 (2008) 6720–6729.
- [7] P. Nguyen, H. Cui, K.S. Bisht, L. Sun, K. Patel, R.S. Lee, H. Kugoh, M. Oshimura, A.P. Feinberg, D. Gius, CTCFL/BORIS is a methylation-independent DNA-binding protein that preferentially binds to the paternal H19 differentially methylated region, *Cancer Res.* 68 (2008) 5546–5551.
- [8] J.A. Hong, Y. Kang, Z. Abdullaev, P.T. Flanagan, S.D. Pack, M.R. Fischette, M.T. Adnani, D.I. Loukinov, S. Vatolin, J.I. Risinger, M. Custer, G.A. Chen, M. Zhao, D.M. Nguyen, J.C. Barrett, V.V. Lobanekov, D.S. Schrupp, Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells, *Cancer Res.* 65 (2005) 7763–7774.
- [9] S. Vatolin, Z. Abdullaev, S.D. Pack, P.T. Flanagan, M. Custer, D.I. Loukinov, E. Pugacheva, J.A. Hong, H. Morse 3rd, D.S. Schrupp, J.I. Risinger, J.C. Barrett, V.V. Lobanekov, Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and derepression of MAGE-A1 and reactivation of other cancer-testis genes, *Cancer Res.* 65 (2005) 7751–7762.
- [10] L. Sun, L. Huang, P. Nguyen, K.S. Bisht, G. Bar-Sela, A.S. Ho, C.M. Bradbury, W. Yu, H. Cui, S. Lee, J.B. Trepel, A.P. Feinberg, D. Gius, DNA methyltransferase 1 and 3B activate BAG-1 expression via recruitment of CTCFL/BORIS and modulation of promoter histone methylation, *Cancer Res.* 68 (2008) 2726–2735.
- [11] S.R. Martinez, J.L. Miranda, CTCF terminal segments are unstructured, *Protein Sci.* 19 (2010) 1110–1116.
- [12] S.C. Harrison, A structural taxonomy of DNA-binding domains, *Nature* 353 (1991) 715–719.
- [13] N. Sreerama, R.W. Woody, Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set, *Anal. Biochem.* 287 (2000) 252–260.
- [14] P. Romero, Z. Obradovic, X. Li, E.C. Garner, C.J. Brown, A.K. Dunker, Sequence complexity of disordered protein, *Proteins* 42 (2001) 38–48.
- [15] V.N. Uversky, Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule, *Biochemistry* 32 (1993) 13288–13298.
- [16] N. Greenfield, G.D. Fasman, Computed circular dichroism spectra for the evaluation of protein conformation, *Biochemistry* 8 (1969) 4108–4116.
- [17] N. Sreerama, S.Y. Venyaminov, R.W. Woody, Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis, *Anal. Biochem.* 287 (2000) 243–251.
- [18] S. Venyaminov, I.A. Baikalov, Z.M. Shen, C.S. Wu, J.T. Yang, Circular dichroic analysis of denatured proteins: inclusion of denatured proteins in the reference set, *Anal. Biochem.* 214 (1993) 17–24.
- [19] A.K. Dunker, C.J. Brown, J.D. Lawson, L.M. Jakoucheva, Z. Obradovic, Intrinsic disorder and protein function, *Biochemistry* 41 (2002) 6573–6582.