Structure of recombinant rat UBF by electron image analysis and homology modelling

Keith J. Neil, Ross A. Ridsdale, Brenda Rutherford, Laura Taylor, Dawn E. Larson, Marija Glibetic, Lawrence I. Rothblum and George Harauz*

Department of Molecular Biology and Genetics, University of Guelph, Guelph, Ontario N1G 2W1, Canada, 1Molecular and Structural Biology, Ontario Cancer Institute, 610 University Avenue, Toronto, Ontario M5G 2M9, Canada and 2Weis Center for Research, Geisinger Clinic, 100 North Academy Avenue, Danville, PA 17822-2618, USA

Received January 3, 1996; Revised and Accepted March 4, 1996

ABSTRACT

We have studied the structure of recombinant rat UBF (rrUBF), an RNA polymerase I transcription factor, by electron microscopy and image analysis of single particles contrasted with methylamine tungstate. Recombinant rat UBF appeared to be a flat, U-shaped protein with a central region of low density. In the dominant projections, 2-fold mirror symmetry was seen, consistent with the dimerization properties of this molecule, and of dimensions in agreement with the length of DNA that rat UBF protects in footprinting studies. Electron microscopy of various rrUBF–DNA complexes confirmed that our recombinant protein was fully able to bind the 45S rDNA promoter, and that it caused substantial bends in the DNA. Upon extended incubation in a droplet covered by a lipid monolayer at the liquid–air interface, rrUBF formed long filamentous arrays with a railway track appearance. This structure was interpreted to consist of overlapping rrUBF dimers 3.5 nm apart, which value would represent the thickness of the protein. Our results show rrUBF to interact with and bend the promoter DNA into a roughly 10 nm diameter superhelix. Based on all these electron microscopical results, an atomic structure was predicted by homology modelling of the HMG fingers, and connected by energy minimized intervening segments.

INTRODUCTION

The synthesis of ribosomal RNA (rRNA) in vertebrates depends on the efficient initiation of rRNA gene transcription by RNA polymerase I and its associated transcription factors (1–3). Before RNA polymerase I can bind to the promoter region, there must be interactions amongst a species specific factor SL1 containing the TATA box binding protein (TBP), along with other TBP-associated factors (TAFs), the upstream binding factor (UBF), the core promoter element (CPE), and the upstream promoter element (UPE) (1–8). UBF has been implicated in the recognition of the promoter region, binding both to the CPE and UPE. An extended DNase I footprint occurs on the rRNA promoter when UBF and SL1 are present together, implying that there is a cooperation between UBF and the binding of one or two complexes of SL1 (7–11). In some cases UBF is a requirement for transcription initiation but instances where transcription initiation can occur without UBF have been described (8,12). One model of 45S rRNA transcription initiation has the UBF dimer binding to the upstream promoter element and bending the DNA further, bringing the CPE and UPE into closer proximity (13,14). The DNA is already intrinsically bent in this region (15). Subsequently, SL1, RNA polymerase I, and ancillary transcription factors such as factor 1c, the core promoter binding factor (CPBF), and factors E1BF/KU all bind, resulting in formation of a stable transcription initiation complex with potential subsequent nucleosomal disruption (16). The role of UBF in the committed complex has yet to be precisely defined (17).

The cDNAs for UBF have been cloned from a variety of species including human, mouse, rat and Xenopus laevis (18–21). The high sequence conservation amongst these species suggests that there is a large degree of structural similarity between the different proteins. There are three distinct regions to UBF. The N-terminal region, consisting of the first 100 or so amino acids, makes up the dimerization domain which is similar to the helix–turn–helix dimerization domains found in RNA polymerase II transcription factors (22). Approximately half of the mammalian UBF sequence consists of four tandem repeats of an 80 amino acid segment. They are the highly basic DNA binding domain referred to as the HMG-box. This region has high homology to the DNA binding domain in the non-histone chromatin associated protein HMG1 (18). Finally, at the C-terminal end of the UBF protein is a highly acidic region which is involved in transcription initiation, but which is not required for DNA binding or nucleolar localization (23–25). The C-terminal tail also contains serine residues that can be phosphorylated by casein kinase II. Unphosphorylated UBF cannot stimulate transcription, and additional cellular protein kinases are required for growth-dependent UBF phosphorylation (26,27). Purified rat, mouse and human UBF fractions contain two proteins, of 97 and 95 kDa, designated UBF1 and UBF2, respectively. UBF2 has a 37 amino acid deletion in the second of the four HMG-boxes (20).

* To whom correspondence should be addressed
Crosslinking studies have shown that rat UBF (rUBF) exists as a dimer in solution (23), with the N-terminal region and HMG-box 2 being important for dimerization. Based on NMR spectroscopy, the solution structure of an HMG-box can be described as an arrowhead formed from three alpha helices, representing a novel DNA binding domain (24,25).

Footprinting studies and Southwestern blots have identified the HMG-boxes closest to the N-terminal end as being the dominant DNA binding domains, with the first HMG-box being the most significant to specific DNA binding (18,22–24). These N-terminal HMG-boxes are also more closely related to the original HMGI motif. The HMG-boxes alone can bind non-specifically, but preferentially to four-way junction DNA (28). The C-terminal HMG-boxes and the acidic tail are required for the extended footprint on the rDNA gene promoter in the presence of SL1, implicating them in protein–protein contact between UBF and SL1 (10). HMG-boxes from lymphoid enhancer factor 1, SRY (a transcription factor encoded by the sex-determining region of the human X chromosome) and HMGI have been shown to induce bends up to 130° in DNA. X.laevis UBF (xUBF) has been found to introduce supercoils into plasmid DNA containing core promoter sequences (29–34).

In this paper, transmission electron microscopy and image analysis of preparations of rUBF as single particles and as filaments have yielded a model of rUBF structure to low resolution, and with electron spectroscopic images of rUBF–DNA complexes allowed definition of its interaction with the rRNA gene promoter.

**MATERIALS AND METHODS**

**Protein and nucleic acids preparation**

The mRNA for rat UBF1 was cloned, and an N-terminal His-tagged form without the acidic tail was overexpressed in E.coli. This His-tagged recombinant rUBF was purified by affinity chromatography on a Ni column according to the manufacturer’s instructions (Qiagen). Proteins were dialysed against a buffer containing 20 mM HEPES–KOH pH 7.9, 100 mM KCl, 5 mM MgCl₂, 0.2 mM EDTA, 20% glycerol, 0.5 mM DTT and 0.5 mM PMSF (buffer A), and stored at −70°C until use. The high degree of homogeneity of the rrUBF preparation was confirmed by SDS–PAGE with both Coomassie Blue and silver staining.

Two plasmids were used, p2.0 and p8.5, both containing fragments of the rat 45S rRNA gene promoter (Fig. 1) (35,36). Plasmid p2.0 contains 170 bp of non-transcribed spacer (NTS) upstream of the initiation site, and 1829 bp of the external transcribed spacer (ETS), cloned into pBluescript. Plasmid p8.5 contains 3686 bp of the NTS and 4812 bp of ETS and part of the 18S rRNA gene. Plasmid p2.0 was purified by cesium chloride gradient centrifugation. A 2000 bp Sall fragment containing the rat 45S ribosomal RNA gene promoter was cut from this plasmid, purified on a 1% agarose gel excised from the gel, and the DNA extracted by electroelution and ethanol precipitated in a dry ice–ethanol bath. DNA was verified free of agarose by electron microscopy. The p8.5 plasmid was used as the template from which to amplify, by the polymerase chain reaction (PCR), a 445 bp region containing 230 bp of the NTS and 215 bp of the ETS. The 445 bp PCR product was purified by high pressure liquid chromatography.

**Electron microscopy and image analysis of rrUBF**

Purified rrUBF at 120 µg/ml was adsorbed to an ultrathin carbon film on a 400 mesh Cu or Ni grid for 2 min in buffer A, or in a low salt buffer obtained by concentrating the above sample through a 10 000 Da cutoff ultrafiltration device (Millipore Ultrafree MC 10 000), and diluting with 20 mM HEPES–KOH pH 7.9, 0.5 mM EDTA and 10 mM KCl. Positive staining was then attained by addition of 0.25% methylamine tungstate to the carbon film for 1 min, washing twice with double-distilled water (ddH₂O), and air drying. The experiments were repeated with buffer alone as a negative control. Samples were micrographed at a nominal magnification of 50 000× on a JEOL 100-CX operating at 80 kV. Each specimen area was not pre-illuminated prior to being micrographed, to minimize the total electron dose.

Electron micrographs of positively stained molecules of rrUBF imaged in the bright field mode were visually scanned to select distinct particles of recurring appearance and size. Particles of size considerably greater or smaller than that expected for rrUBF (~10 nm diameter) were considered to be aggregates or debris and were not selected. Putative rrUBF particles were digitized and stored in arrays of 128 by 128 pixels (picture elements), where one pixel corresponded to an area of 0.125 nm by 0.125 nm at the object level. A total of 287 distinct particles were selected.

Single particle analysis of digitized images was performed in the framework of the IMAGIC image processing system as described elsewhere in detail (37–41), and summarized here. The single macromolecular images were pre-treated by band-pass filtering to suppress noise, normalized to have a common variance, and put into register with one another by a multi-refer- ence alignment. We used both multi-variate statistical and visual classification to subdivide this set of 287 images into eight major classes according to criteria of size and shape. Averaging was done using the 20 images that gave the highest correlation.
Electron microscopy of rrUBF–DNA complexes

The 2000 bp SalI restriction fragment, with the 45S rRNA promoter positioned closer to one end, was used to form rrUBF–DNA complexes. The binding reaction was performed at room temperature for 15 min in buffer A, at a ratio of two rrUBF molecules per DNA fragment (0.05 µg/µl DNA: 0.1 µg/µl rrUBF) in a total volume of 20 µl. Following a 15 min incubation in the presence of 0.1% glutaraldehyde, the complexes were spread for microscopy on ultrathin carbon films either with the DMP-30 microdroplet technique (42) or directly onto carbon coated grids recently exposed to a tungsten light source (43). Formation of complexes with 445 bp DNA was performed in the same manner, except that 0.01% glutaraldehyde was used, along with a brief exposure (1 min) to short-wave ultraviolet light (254 nm) at the end of the 15 min incubation.

Electron spectroscopic imaging (ESI) of unstained complexes was performed on a Zeiss EM902 equipped with a Castaing–Henry energy filter at a nominal magnification of 30 000x. ESI used scattered electrons with energy losses within a narrow range, and was ‘tuned’ to phosphorus by imaging at ±8 eV loss. These 150 eV loss images thus presented the DNA with high contrast without the need for exogenous heavy metal stains or shadowing (33). All microscopy was performed with the specimen side up, facing the electron source. Measurements of the position of binding of rrUBF, the complex size, and the bend angle effected in the DNA by protein binding were performed on images captured on an Alpha Innotech 1000 (Alpha Innotech Corp., San Leandro, CA) gel scanner. A program was written in Microsoft Visual Basic, running under Microsoft Windows, which allowed interactive delineation of the path of DNA or the periphery of the complex using the mouse, and subsequent calculation of distances, areas and bend angles. A total of 297 images of the rrUBF: 445 bp DNA complexes was thus analysed.

Two-dimensional crystallization on lipid monolayers

A lipid monolayer approach was utilized to attempt to form planar arrays of rrUBF (44–46). A drop of size 10 µl of rrUBF at a concentration of 0.2 µg/µl was placed in a 1 mm deep by 3 mm diameter well in a block of Teflon. Then ±1 µl 7:3 phosphatidylycholine–phosphatidylserine (Avanti Polar Lipids, Alabama) solution in chloroform–hexane was touched to the surface to form a lipid monolayer at the liquid–air interface. The droplet was incubated in a humid chamber under argon, at 4°C for times ranging from 30 min to overnight. Arrays were then picked up onto ultrathin carbon coated 400 mesh Ni grids, contrasted with 2% uranyl acetate, and observed at a nominal magnification of 50 000x as before.

Computational models of rrUBF

With the known molecular weight of the rrUBF dimer, the average density of proteins, the size and shape constraints of the rrUBF dimer determined by electron image analysis of single particles and the appearance of the linear filaments, and the length of the upstream binding element protected by UBF in footprinting assays, we hypothesised that we could create a three-dimensional model of the surface of the rrUBF dimer as a first step towards more detailed predictions. The symbolic mathematical language MAPLE V was used in this task (47).

Recently, the structures of individual HMG-boxes from rat and hamster have been solved using solution NMR (24,25). We obtained the coordinates of the HMG-box from one of these studies (25), and predicted the coordinates of the HMG-boxes within the rrUBF using multiple sequence alignment and homology modelling functions of the INSIGHT II molecular modelling software package (Biosym Corporation, Parsippany, NJ).

RESULTS AND DISCUSSION

Electron microscopy, image analysis and solid modelling of purified rrUBF

Our first approach to analysing the structure of rrUBF was bright-field transmission electron microscopy and single particle analysis of highly purified preparations. Figure 2 shows both fields of view and enlargements of individual molecules of the
rrUBF spread on the thin carbon support, contrasted by basic methylamine tungstate stain. The molecules were ∼10 nm in diameter, and were well separated from each other. Observation of the buffer alone (without protein) demonstrated a negligible amount of particulate material due to precipitants in the buffer or stain. This negative control discounted the possibility that these images were artefacts, and indicated that the recurring motifs of the expected size range that were observed were indeed rrUBF.

The enlargements in Figure 2b show a number of the individual rrUBF molecules that were digitized and used for image analysis. These samples demonstrate the typically observed recurring views, some toroidal- or U-shaped, with a central depression and with a suggestion of 2-fold mirror symmetry. The independent averages of characteristic projections obtained by image analysis are shown in Figure 3. Two-fold symmetry is evident especially in Figures 3a and b, consistent with rrUBF in dimeric form. The rrUBF dimers have dimensions of approximately 8 × 6 nm², and appear as a closed toroidal- or an open U-shaped molecule, with a central area of low density. Based on these initial results, a surface model for rrUBF was generated using Maple V (47). This model was essentially a torus of internal radius r, bend radius R, truncated to 7/8 of a turn (Fig. 4). The model shown here is the result of refinement using also the appearance of the rrUBF filaments (below).

The toroidal- or U-shaped molecule with a central low density core has dimensions that would allow a 10 nm solenoid of DNA to wrap around the rrUBF complex as a partial superhelix, bringing the CPE and the UPE into close proximity for interaction with SL1. The extended footprint observed when full length UBF is incubated with SL1 indicates an interaction between the DNA bound UBF and the TBP and TAFs that comprise SL1 (7,8). A 10 nm solenoid would bring both CPE and UPE in line in close geometric position, allowing the SL1 DNA contact to be made with the CPE and UPE. It is known that rUBF protects ∼77 bp of DNA in DNAse I footprinting assays, a linear distance of 23 nm (8). The circumference of the toroidal shape of the dimensions above would be ∼25 nm, closely corresponding to the footprinting data.

Complexes of rrUBF–DNA

In order to verify that the purified rrUBF indeed bound DNA specifically at the upstream promoter element, rrUBF was incubated with the 2000 bp fragment or 445 bp fragment of DNA and visualized. Figures 5 and 6 show several rrUBF–DNA complexes imaged electron spectroscopically having neither been stained with a heavy metal salt nor rotary shadowed with platinum–palladium to increase contrast. The rrUBF molecule was usually seen to bind the DNA fragment preferentially at one
end. This observation suggests that rrUBF is correctly bound to the promoter region. Particles ~10 nm in size are interpreted to represent the dimeric form of rrUBF, while those structures roughly 20 nm in diameter or greater probably represent multimeric rrUBF complexes. Frequently the DNA undergoes a sharp bend, sometimes looping back on itself. This bend angle usually measures 90–180° but would actually be more if the plasmid DNA is fully wrapped around the rrUBF molecule. The mean length of the 445 bp DNA complexed with rrUBF was 129 ± 18 nm (± standard deviation), and of naked DNA it was 149 ± 10 nm (Fig. 7). The measured length of bound DNA was thus shorter than for naked DNA by roughly 20 nm or 60 bp. A one tailed t-test for means shows this difference to be statistically significant at a level >99%. Within the limits of measurement error, these values are equivalent to the amount of DNA protected in footprinting studies, and to one supercoil turn of DNA as described above. The complex diameter was found to lie most often in the range 6–18 nm (Fig. 7b). Those complexes >10 nm in size could represent higher-order polymerisation.

Treatment with the bifunctional protein–protein crosslinker glutaraldehyde is a common procedure for preparation of macromolecular complexes for EM although caution is urged in its application (49–59). The purpose of this fixation step is to keep intact the complex being studied as it adsorbs to the support film and is subjected to strong surface tension forces. Indeed, stringent fixation (~16 h at 4°C of 0.1% glutaraldehyde) is stated to be necessary for any study of chromatin; without this step, only unravelled DNA filaments are seen (54,57). For these reasons, then, we included glutaraldehyde fixation in our preparation protocol. Under certain conditions, e.g. after depurination of DNA, this chemical can also be used to crosslink DNA to protein (60). However, glutaraldehyde is almost never used in in vitro protein–DNA crosslinking studies since other agents, e.g. formaldehyde and ultraviolet irradiation are far more effective in this regard. Thus, given the wealth of published experience with glutaraldehyde as an EM fixative, it is reasonable to assume that this agent did not induce either any extra protein binding to the plasmid, nor any spurious looping due to the formation of protein–DNA crosslinks, to the extent that we did use it.

**Lipid monolayer crystallization**

Some soluble proteins have been induced to form two-dimensional crystals by their interaction with a mixed lipid monolayer formed on an air–water interface (44–46). In the case of proteins such as the RNA polymerases, hydrostatic and electrostatic interactions with the lipid monolayer, diffusion of the lipids within the monolayer, and subsequent protein–protein interactions, led to the formation of a planar crystal (45,46). The success of this step is dependent upon parameters such as the purity of the sample, composition of buffers, kinds of lipid mixtures, lipid–protein ratios, presence of detergents, and temperatures and times for crystallization.

We examined the formation of ordered arrays of rrUBF on lipid monolayers under conditions similar to that reported for RNA polymerase II, with incubation times ranging from 0.5 to 24 h (45). A limited number of experiments were performed using specially engineered Ni-chelator lipids, with essentially the same...
Figure 7. (a) Upstream, downstream and combined length distributions of the rrUBF–445 bp DNA complexes, and the length distribution of several uncomplexed fragments. The total length of the complexed DNA, as measured from the sum of the upstream and downstream lengths, and the diameter of the protein complex, is remarkably less than the length of the uncomplexed fragments. This length difference can be shown to be statistically significant. This result suggests that the DNA is wrapped around the rrUBF complex, and is not simply bound to it. The greater frequency of shorter length upstream fragments is consistent with complex formation in the vicinity of the promoter.

(b) Diameter distribution of the rrUBF:445 bp DNA complex. Most complexes are between 6 and 18 nm in diameter. The smaller ones are interpreted as representing monomers of rrUBF, whilst the larger ones probably represent dimers and bigger multimers.

It was found that rrUBF formed one-dimensional filaments rather than two-dimensional arrays under these conditions (Fig. 8). These filaments have a width of ~7 nm, and the repeat along the filament axis is ~3.5 nm. These dimensions concur with those of the UBF dimer inferred from the single particle work above, if we consider the fiber to be formed of molecules stacked one on top of the other. In some cases multiple filaments came together as a superfilament, and a ribbon-like appearance was evident. This is the first time UBF has been reported to form such ordered arrangements, although the formation of filaments with purified transcription factors is not in itself a novel discovery (reviewed in 44). For example, E.coli RNA polymerase and the REV protein of human immunodeficiency virus 1 have been shown to form ordered helical polymers at high concentrations in solution (62,63). The formation of these filaments may help in solving the structure of UBF by cryo- transmission electron microscopy, since suitable large and ordered 3D crystals of this protein may be difficult or impossible to obtain. Future higher resolution structural studies will focus on extending the degree of order and the sizes of these arrays, and imaging in vitreous ice.

Molecular modelling

Finally, based on the cumulative electron microscopical results, a molecular model was generated using the INSIGHT II molecular modelling package (Fig. 9). The structure of the HMG-domain from hamster HMG1 protein was applied to the four HMG-box domains of the UBF DNA binding domain (25). The rrUBF HMG-boxes were predicted by multiple sequence alignment based on this sequence. Then the coordinates for the hamster HMG-box were applied to rrUBF by homology modelling with INSIGHT II. The rrUBF HMG-boxes were oriented to one another based on their positions on a supercoiled DNA strand. The degree of supercoiling was designed to correspond to the outer surface of the toroid model. The distance of boxes from each other on the DNA was taken from DNA fingerprinting studies (22,23). The structures of the intervening sequences were produced by loop generation with energy minimization. The acidic tail is not part of this model, as we could not reliably predict its structure.

The ‘top’ view of the protein in Figure 9 corresponds to the images in Figures 2 and 3 interpreted to be the dimer views and is essentially of a toroidal- or U-shaped macromolecule. Another view shows the dimer edge-on with the lobes of the toroid being regions of high density. To accommodate the HMG-boxes in this structure, they would be tandemly arranged in the same orientation facing the DNA on the exterior surface. It can be proposed that the HMG-boxes at the top of the structure interact initially with the DNA, and lead to the bending of DNA as successive boxes bind. If UBF binds the DNA and causes positive supercoils in the enhancer DNA as proposed in (34), then the binding of UBF to chromatin would help displace nucleosomes.
from negatively supercoiled DNA in the promoter region, thereby favouring SL1 binding and subsequent transcription initiation complex formation. The proposed structure of rUBF is also another example of ‘holy’ (non-globular) soluble proteins such as ribonuclease inhibitor (64) and lytic transglycosylase (65).

CONCLUSIONS

We have deduced a low resolution structure of the important 45S rDNA transcription factor UBF from rat. In dimeric form, recombinant rat UBF is a flat, toroidal-shaped protein which has the potential of being flexible. We have also shown the rUBF dimer to bind the 45S rDNA gene promoter, and to induce bends into the DNA. The images are consistent with the protein interacting with the DNA minor groove. We have also shown rUBF to form filamentous arrays at high concentrations on lipid monolayers. This observation suggests the rUBF dimer is capable of forming multimers, by an overlapping association between its exterior surfaces. These results, and knowledge of the tandem repeat of the HMG-boxes in rUBF, suggested a model of the rUBF dimer which has the DNA of the promoter region wrapped around the outer edge of the toroidal shape, bringing the CPE and UPE in close proximity. This geometry would facilitate the interaction of SL1 with rUBF, and the binding of SL1 to the UPE and CPE, and would be consistent with the extended footprint observed when human UBF and SL1 are footprinted on the interaction of SL1 with rUBF, and the binding of SL1 to the CPE and UPE in close proximity. This geometry would facilitate the interaction of SL1 with rUBF, and the binding of SL1 to the CPE and UPE, and would be consistent with the extended footprint observed when human UBF and SL1 are footprinted on the human promoter. Moreover, this structure would be consistent with the mechanism by which HMG proteins bind to DNA.

ACKNOWLEDGEMENTS

This was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada to G.H., and by National Institutes of Health grants GM46991 and HL46738 to L.I.R. We are grateful to Mr Philip Chang for assistance with and use of their Zeiss EM902 for micrographs, and to Mr Yew-Meng Heng and Dr Peter Ottens-meyer for assistance with the HMG modelling, Dr Aled Edwards for HL46738 to L.I.R. We are grateful to Mr Philip Chang for and by National Institutes of Health grants GM46991 and

REFERENCES


Figure 9. Models of the HMG-domains and of rUBF generated using INSIGHT II software. (a, b) Space-filling and ball-and-stick atomic models of the HMG-domain from hamster HMGI protein, based on structural information obtained from D.G. Norman (25). The yellow, blue and green regions represent the first, second and third alpha helices, respectively. The latter alpha helix interacts with the DNA minor groove. The red region is the N-terminal end. (c) Ribbon representation of the INSIGHT II generated models of the HMG-domains from rUBF, showing the three alpha helices, and their putative interaction with the DNA minor groove. (d, e) INSIGHT II generated models of rUBF oriented with respect to DNA. The red and white dashed line represents the position of the DNA minor groove. The DNA is represented here as a single strand since it was not possible both to generate and supercoil a full DNA double strand model with the software available to us. Here, we used the results of Putnam et al. (34) in choosing to orient the UBF dimer such that it would bend DNA into positive supercoils upon binding.