© 1991 Oxford University Press

Nucleic Acids Research, Vol. 19, No. 23 6603–6610

Interaction of DNA-binding proteins with a milk protein gene promoter in vitro: identification of a mammary gland-specific factor

Christine J.Watson, Katrina E.Gordon, Morag Robertson and A.John Clark
AFRC Institute of Animal Physiology and Genetics Research, Edinburgh Research Station, Roslin, Midlothian EH25 9PS, UK

Received July 8, 1991; Revised and Accepted October 30, 1991

EMBL accession no. X12817

ABSTRACT

The minimal 5′ regulatory region of the sheep β-lactoglobulin gene (BLG), as defined in transgenic mice, was used to identify nuclear factors which may be involved in milk protein gene expression in the lactating mammary gland. This 406bp promoter region was dissected into short, overlapping, double-stranded oligonucleotides to facilitate identification of the bound proteins. A variety of sites, for both known and previously undescribed DNA-binding proteins, are occupied *in vitro*. Some of these factors were investigated in detail. Two forms of nuclear factor I (NFI), which have different recognition site affinities, are present in nuclear extracts from lactating mammary gland and bind to at least 5 sites in this BLG control element. In addition, a factor (milk protein binding factor, MPBF) which is specific to extracts from both mouse and sheep lactating mammary gland binds to 3 BLG promoter sites and may be a milk protein gene transcription factor.

INTRODUCTION

Cellular differentiation is a process which is regulated by the differential expression of genes. Control of gene expression at the transcriptional level requires the sequence specific interaction of transcriptional activators and repressors with *cis*-acting DNA elements in gene promoters and enhancers. Both ubiquitous and cell-type specific DNA-binding proteins may be required for tissue-specific gene expression (reviewed in 1). Binding of such transcription factors to their cognate sites probably stimulates transcription by enhancing the formation of preinitiation complexes at the TATA/CAP region (2,3).

Many transcription factors belong to multigene families. Multiple variants of individual factors are generated by alternative splicing of primary transcripts and different post-translational modifications. Transcription is controlled by a complex interplay of multiple proteins, both within and between families and may be regulated by a variety of extracellular stimuli, including hormones. This can be achieved by the induction or modification of a transcription factor through a secondary signal transduction pathway or, as is the case for steroid hormones, by direct binding of the ligand-bound hormone receptor to its recognition site on DNA (4).

The mammary gland is a good model for studying the control of differentiation and gene expression. During gestation, the mammary gland undergoes differentiation to a lobuloalveolar structure (the secretory epithelial cells) with surrounding myoepithelial cells, in response to the hormones prolactin (in some species placental lactogen) estrogen and progesterone. By parturition, the fat pad is almost entirely replaced by the massive proliferation of secretory cells (5). Expression of most milk protein mRNAs is induced around midgestation, during the period of rapid proliferation and differentiation.

Milk protein gene expression is hormonally regulated, the major lactogenic hormone being prolactin although insulin and glucocorticoids are also required (6). Studying the regulation of milk protein gene expression in the mammary gland should provide insights into the mechanism of action of prolactin and its interplay with other lactogenic hormones. Identification of important mammary regulatory elements could enable the design of an optimal mammary promoter for targeting expression of foreign proteins to the mammary gland (7).

Work in this area has been hampered by the lack of mammary cell lines which accurately mimic differentiation and lactogenesis. Mouse mammary epithelial cell lines have been used to study the induction of β-casein gene expression in response to lactogenic hormones (8). However, not all milk protein genes appear to be expressed in such cells. Transgenic mice have therefore been used as an investigative tool for the elements which regulate gene expression in the mammary gland.

An ideal milk protein for such studies is β-lactoglobulin (BLG) which is the major whey protein in the milk of ruminants but is not present in mouse milk. Transgenic mice express a genomic BLG construct, carrying 4.3kb of 5′ flanking sequence, efficiently and specifically in the mammary gland (9,10). Essential sequences in the 5′ end were determined by a resection analysis in transgenic mice. The results clearly demonstrated that the region between −149 and −406 of the 5′ proximal control region are essential, though not necessarily sufficient, for high-level, tissue-specific expression (11).

In an attempt to identify the DNA-binding proteins and their cognate sites within this 406bp element which are involved in the transcriptional regulation of BLG, *in vitro* binding studies were carried out. In this paper we report the identification of a DNA-binding protein which appears to be specific to the lactating mammary gland and which we have called milk protein binding factor (MPBF). This factor binds to two high affinity sites and one low affinity site in the BLG minimal control region and may be a mammary gland-specific transcription factor which is involved in the regulation of milk protein genes. We also describe variant NFI-like activities which occupy multiple sites with different relative affinities in the BLG minimal promoter.

MATERIALS AND METHODS

Nuclear extracts

Nuclear extracts were prepared essentially as described by Dignam (12), with modifications. Tissues, from sheep or mice, were flash frozen in liquid N_2 and approximately 7g aliquots used for extract preparation. Frozen tissue was ground to a fine powder then suspended in a buffer A/NT/L (2:3:5) containing a cocktail of protease inhibitors. Dispersed cells were then homogenised by hand in a Dounce homogeniser and filtered through two layers of miracloth (Cambridge Bioscience). Nuclei were pelleted by centifugation at 2.5K for 5 min. in a Sorvall centrifuge, then resuspended in A/NT (1:1) and repelleted. Crude nuclei were extracted as in (12) and frozen in aliquots at −70°C. All steps were carried out at 4°C. Solution A : 0.6M sucrose, 120mM KCl, 15mM NaCl, 0.3mM spermine, 2mM spermidine, 28mM β-mercaptoethanol, 4mM EDTA, 2mM EGTA, 2mM DTT, 0.2% Triton X-100 and 10mM Tris.HCl pH 7.9. Solution NT : 15mM NaCl, 10mM Tris.HCl pH 7.9. Solution L : 10mM NaCl, 0.1% NP40 (v/v), 10mM Tris.HCl pH 7.9.

Oligonucleotides and restriction fragments

Oligonucleotides were synthesized on an Applied Biosystems Model 381A DNA synthesiser or purchased from Oswel DNA service, Kings Buildings, Edinburgh. These were radiolabelled by phosphorylating the 5′ ends with [γ-^{32}P]ATP and polynucleotide kinase. Labelled, double-stranded oligonucleotides were separated from single strands and unincorporated label on 10% polyacrylamide gels, excised from the gel and eluted in high salt buffer (1.0M NaCl, 20mM Tris.HCl pH 7.5, 1mM EDTA). When restriction fragments of the BLG promoter were used as targets, these were labelled by filling-in recessed ends with [α-^{32}P]dCTP and/or [α-^{32}P]dATP and Klenow enzyme.

The DNA sequences of the various oligonucleotides which contain specific transcription factor binding sites are shown below (upper strands only):

NFI	5′-GATCTTTGGCTTGAAGCCAATA-3′
HIV-L	5′-GATCCGCGGAAAGTCCCTA-3′
HRE-BLG	5′-GATCCCAGGACACACCTGTCC-3′
SVE	5′-GATCTAGGGTGTCCAAAGTCCCG-3′
ERE	5′-GGTCANNNTGACC-3′
GRE	5′-GGTACANNNTGTTCT-3′

Electrophoretic mobility shift assays (EMSA)

For binding studies, 0.5ng end-labelled double-stranded DNA (10000−50000 cpm) was incubated with 2μg nuclear extract for 20 min at 20°C in a buffer containing 20mM Hepes pH 7.5, 1mM EDTA, 1mM DTT, 10% glycerol, 100mM NaCl, 0.05% NP40 and 1μg poly dI.dC in a 20μl reaction. Complexes and free DNA were resolved on non-denaturing polyacrylamide gels (6% unless indicated otherwise in the figures) with TAE (6mM Tris, 1mM EDTA, 7.5mM NaOAc pH 7.5) electophoresis buffer. Rabbit polyclonal antibodies, directed against the DNA-binding domain of NFI, were kindly provided by Dr. Ronald Hay (University of St. Andrews, Scotland). Protein concentrations were determined using the BCA protein assay system (Pierce).

RESULTS

Complex interactions between nuclear proteins and the BLG regulatory element

Gel mobility shift assays were initially performed on the 366bp fragment (Dp) of the BLG minimal promoter from −40 to −406 (Dpn I/Sph I fragment in figure 1). Nuclear extracts from virgin and lactating sheep mammary gland were compared with HeLa cell and mouse liver nuclear extracts for complex formation with this probe fragment (figure 1). Multiple binding interactions are seen with all tissue extracts. However, one complex is formed only with lactating mammary extract (arrowed M, figure 1). This suggests the existence of a mammary specific factor. Since a large number of different proteins bind to this DNA fragment, a detailed dissection of the element was undertaken in an effort to identify these proteins and their binding sites and to confirm the presence of a mammary gland-specific factor.

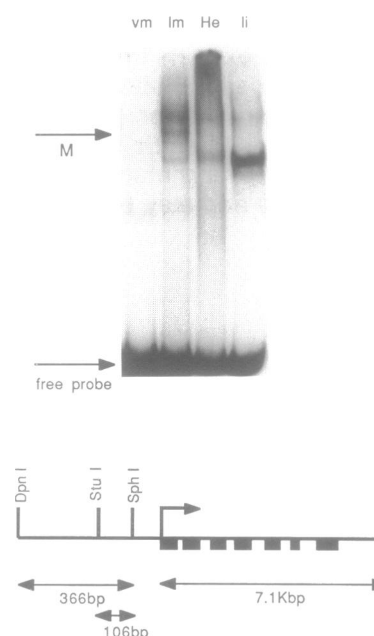

Figure 1. Gel retardation analysis using the Dpn I/Sph I restriction fragment of the BLG promoter as probe. Nuclear extracts (2μg protein) from virgin mouse mammary gland (lane vm), lactating mouse mammary gland (lane lm), mouse liver (lane li) and HeLa cells (lane He) were assayed. Following binding for 20 min at 20oC1, DNA-protein complexes and free DNA were separated on a 6% non-denaturing polyacrylamide gel. The upper arrow (M) indicates the mammary-specific factor. The structure of the BLG gene is illustrated schematically below-exons are represented by filled boxes. The 2 restriction fragments of the 5′ flanking sequence which were used in this study are indicated.

Oligonucleotide dissection of the BLG regulatory element

A series of short (35–49bp), overlapping, double-stranded oligonucleotides which span the region from −406 to −140 were synthesised (figure 2). Gel mobility shift assays were carried out with nuclear extracts from a variety of tissues using these oligonucleotides as probes. At least one complex was formed with each oligonucleotide. In an attempt to define common binding motifs this panel of oligonucleotides was utilised in a competitor analysis, in conjunction with another panel of double stranded oligonucleotides which contain the recognition sites of known transcription factors (see Materials and Methods). This allowed a preliminary identification of some of the binding proteins. Of these, nuclear factor I (NFI) and an apparently mammary-specific factor, which we have called MPBF (milk protein binding factor), were investigated further.

NFI sites in the BLG promoter

Gel shift assays using nuclear extracts from lactating sheep mammary gland with probe A1 shows two complexes which appear to be closely related and a faster migrating, less abundant complex (figure 3A). Formation of the doublet is competed by three of the other BLG panel oligonucleotides (D2, S2, and weakly by A2) indicating the presence of a similar recognition site in these sequences. This was confirmed by carrying out gelshifts with D2, A2 and S2 oligonucleotides as probes. A similar pattern of complex formation was observed with all three probes (data not shown). The nature of these complexes was suggested by the fact that they are also competed by a consensus NFI recognition motif as shown in figure 3A, lane NFI. The band of higher mobility is clearly unrelated to these NFI-like species since it is not competed by the same oligonucleotides. The appearance of two NFI complexes of similar mobility is most likely due to the presence of different forms of the protein. Shorter electophoresis times do not resolve these two species (see figure 4).

A comparison of the sequences of A1, A2, D2 and S2 with the NFI consensus is shown in figure 3B. NF1 is known to bind to its recognition site, which is the palindromic sequence TGGC/A(N)$_5$GCCAA, as a dimer. However, it will bind to half palindromic sites, albeit with much lower affinity (13), and interestingly all four of the BLG oligonucleotides contain only half sites. NFI consists of a family of site-specific DNA-binding proteins. Heterogeneous forms of NFI can be generated from

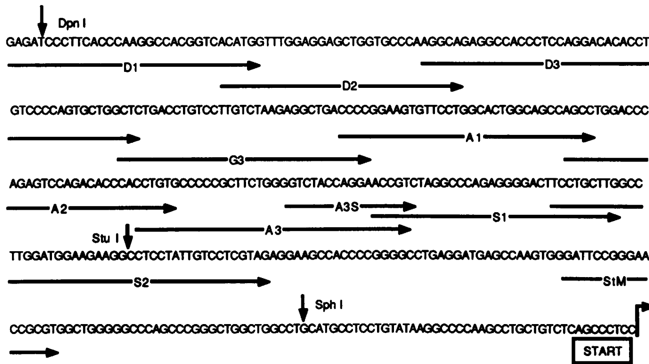

Figure 2. BLG oligonucleotides used in this study. Only the upper strand (5′-3′) of the BLG promoter/regulatory element is shown. Oligonucleotides are represented by labelled arrows below the appropriate sequence. Relevant restriction sites are indicated. Each double-stranded oligonucleotide has a Bam HI (GATC) overhang at either end for ligation and labelling purposes. The BLG sequence is as published, except for 3 corrections (37).

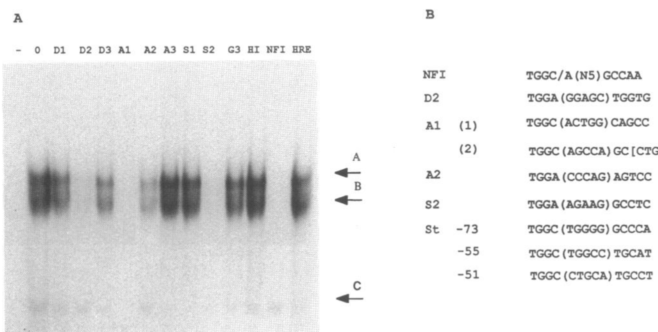

Figure 3. Analysis of NFI binding sites. **(A)** EMSA with probe A1 and 2μg crude lactating sheep nuclear extract. Cold competitor oligonucleotides (100ng) were preincubated with the extract for 15 min. on ice before adding probe. Competitors are indicated above each lane. −: no extract added, 0: no competitor added. The oligonucleotide sequences are shown in figure 2 and materials and methods. **(B)** Comparison of putative NFI binding sites in the BLG promoter oligonucleotides with the consensus NFI palindromic motif. The free probe has run off the bottom of the gel.

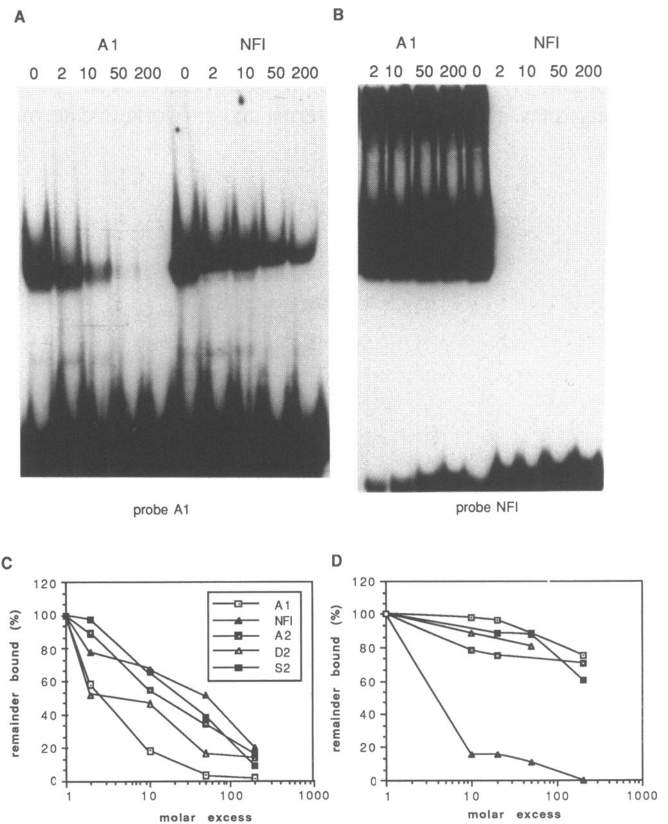

Figure 4. Binding affinities of the various NFI sites. **(A)** Competitor bandshift with A1 target sequences and increasing amounts of NFI and AI. The molar excess of cold competitor added is indicated above each lane, taking account of the different sizes of the oligonucleotides. **(B)** Competitor titration as in (A), using NFI as target. **(C)** Quantitative analysis of percentage residual binding to probe A1 versus the molar excess of the competitor oligonucleotides D2, A2, S2, A1 and NFI. Complexes were quantitated by densitometry scanning using a Shimatzu dual wavelength flying spot scanner. **(D)** A similar quantitative analysis to (C) with NFI as target. The key is the same as in C.

the transcription of multiple genes (14), by alternative splicing (15) or by different post-translational modifications (16). It is possible that the variable binding affinities observed results from the presence of several different heterodimeric forms of NFI in the mammary gland.

Alternative forms of NFI with different recognition site affinities

The possible existence of variant NFI forms in the mammary gland was investigated further by a competitor titration analysis (figure 4). Binding to oligonucleotide A1 was reduced by almost 90% by a 50-fold molar excess of cold A1 (figure 4A). In contrast, a 50-fold molar excess of cold NFI competed binding by just 50%. In the reciprocal experiment, the NFI consensus motif was used as target (figure 4B). Here, NFI is clearly a much more efficient competitor than A1 (10×better). These data, coupled with the different pattern of complexes formed with each probe, support the notion that there are at least two forms of NFI in the mammary gland, one of which has higher affinity for a consensus NF1 site and another which preferentially binds to an element within the BLG promoter region, A1. Competitor titration experiments, using D2, A2, S2, A1 and NFI as competitors, with A1 and NFI as targets were carried out and the results presented in figure 4 (C and D). The relative binding affinities for each probe (at a 100-fold molar excess) are: A1 > S2 > D2 > A2 > NFI (for A1) and NF1 > S2 > A2 > D2 > A1 (for NFI). The reason for this hierarchy of binding affinities is not entirely clear from a simple examination of the putative binding sites. However, Meisterernst and co-workers determined the equilibrium binding constants of NFI to mutated sites (17). Changes to the GCCAA half of the site have a greater influence on binding the closer they are to the centre of the palindrome. On this basis, BLG site S2 should be the best competitor of NFI and indeed this is the case (figure 4D). None of these five bases are conserved in the other sites.

Downstream from these NFI sites, a palindromic binding motif occurs at −55 which also forms an NFI-like complex with sheep mammary nuclear extracts and is competed by the consensus NFI (data not shown and figure 7B).

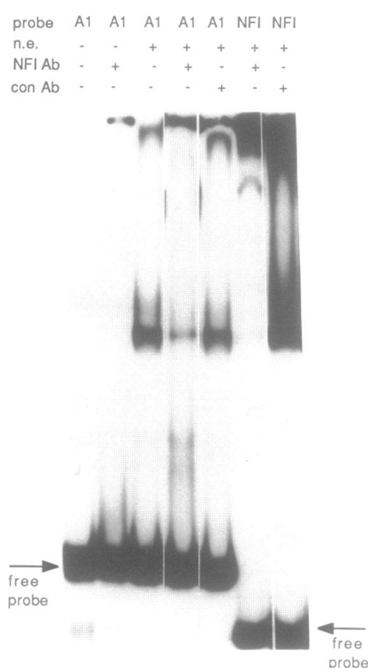

Figure 5. Complex formation in the presence of specific antibodies. Polyclonal antibodies directed against the DNA-binding domain of human NFI were raised in rabbits. Either 1μλ of the NFI antiserum (8-fold dilution, NFI Ab) or control rabbit antiserum (8-fold dilution, con Ab) was added to a standard binding reaction with 2μg of crude sheep lactating nuclear extract (n.e.) and incubated on ice prior to the addition of either A1 or NFI probe sequences. Antibody/NFI complexes will be unable to bind to the target due to blocking of the NFI DNA-binding domain.

Inhibition of binding by an anti-NFI antibody

The NFI-like nature of the factor which recognises the A1 motif was confirmed by an antibody binding experiment coupled with gel shift analysis. Rabbit polyclonal antiserum, raised against the DNA-binding domain of human NFI (expressed in insect cells and kindly provided by Dr. R.T. Hay) was incubated with sheep mammary nuclear proteins then a standard gel shift assay performed (figure 5). Complex formation on A1 was substantially reduced in the presence of the NF1 antibody but unaffected by control antibody. Likewise, the NFI antiserum, but not the control, reduced binding to the NFI motif. These results strongly suggest that the complexes formed with oligonucleotide A1 are due to NFI family members.

A mammary-specific factor

The putative mammary-specific factor which binds fragment Dp was studied in more detail by identifying the members of the panel of BLG oligonucleotides to which it binds. Oligonucleotide A3 contains a binding site for a factor which is restricted to the lactating mammary gland. Binding is competed by A1 (data not shown). A similar factor binds to a sequence within the restriction fragment St (106bp see figure 1) in addition to other factors including NFI and a steroid hormone receptor-like factor (figure 6A). A sequence comparison of oligonucleotides A1 and A3 and

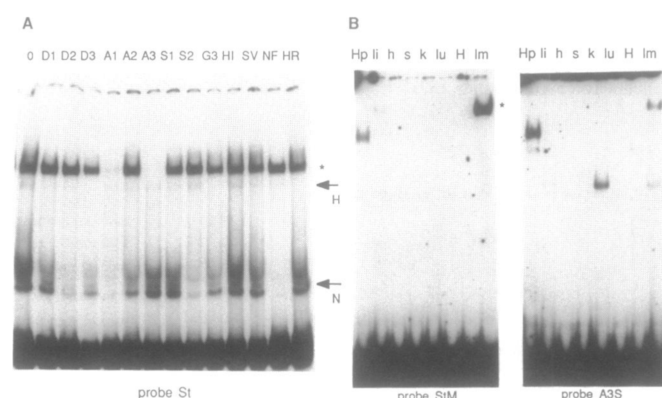

Figure 6. Multiple binding sites for a mammary gland-specific factor. (**A**) EMSA with the 106bp restriction fragment St (Stu I to Sph I, see figure 1) as probe. Competitors are indicated above each lane. − 0: no competitor added, D1 to G3: BLG panel oligonucleotides, HI: HIV (NFκB motif), SV: SVE (mutant NFκB), NF: nuclear factor I consensus, HR: HRE-BLG (hormone response element in BLG promoter, in oligonucleotide D3). (**B**) Tissue specificity of complex formation. Nuclear extracts from a variety of tissues were incubated with the 17bp oligonucleotides A3S and StM and complexes separated on 6% polyacrylamide gels. Tissues are indicated above each lane and are: Hp: partially purified HeLa cell, li: liver, h: heart, s: spleen, k: kidney, lu: lung, H: crude HeLa cell, lm: lactating mammary. The mammary-specific factor is indicated.

fragment St revealed the presence of a short conserved motif in all three elements, which could be the binding site for the mammary factor. Two 17bp double-stranded oligonucleotides, containing the potential factor binding sites from oligonucleotide A3 and fragment St (A3S and StM respectively, figure 7C) were synthesised and binding analysed by gel shift. In total, 7 different tissue nuclear extracts were analysed (fig. 6B). Two interesting results emerge from this experiment. The complex marked with an asterisk is clearly mammary specific. It also binds to the StM motif with higher affinity. We have called this factor(s) milk protein binding factor (MPBF). In non-mammary tissue extracts, a number of complexes of different mobility to MPBF are formed with probes A3S and StM, there being distinct differences between these targets for complex formation. Competitor gel shift analysis suggests that some of these may be hormone receptors (HR, data not shown). Recognition by multiple, often unrelated factors is a common feature of transcription factor sites.

Using the panel of BLG oligonucleotides as competitors for the restriction fragment Dp, and mouse lactating mammary nuclear extract, the faint band which was suggested to be mammary specific (see figure 1) is indeed competed by the MPBF site-containing oligonucleotides A3 and A1 (data not shown).

MPBF is present in both sheep and mouse mammary extracts

These experiments demonstrate that both sheep and mouse lactating mammary gland extracts contain MPBF. This was confirmed by DNA binding competition experiments using the motifs A3S, StM, and A1 as competitors and StM as probe with both lactating sheep and mouse mammary nuclear extracts (figure 7A). Complex formation is qualitatively and quantitatively the same for the sheep and mouse extracts. Binding is most effectively competed by StM itself, A3S competes slightly less efficiently and A1 is a poor competitor. When A3S is used as a probe, the most efficient competitor is again StM (data not shown). Binding was quantified by densitometry scanning of the complexes and the results shown in figure 7B. This allows the relative affinities of the recognition sites to be determined. StM is 5-fold higher than A3S which is in turn 8-fold higher than A1. The conserved site (C), which is a 10bp highly conserved motif (9 out of 10 base match between A3S and StM, figure 7C) does not compete for binding. This is not due to melting of the strands of this short oligonucleotide since, when the binding assay is carried out at 14°C, which is 20° below the theoretical melting temperature, the conserved site still does not compete (data not shown). A mutant double-stranded oligonucleotide, with the central GG changed to CC does not compete for binding (figure 7C and data not shown), suggesting that this dinucleotide may be important for factor binding. Competition for A1 site occupancy between MPBF and NFI is a distinct possibility since their binding sites overlap. However, removal of NFI from its cognate site by competition, does not appear to increase binding of MPBF (figure 3A).

Potential MPBF sites in other milk protein promoters

A number of matches to the MPBF recognition site were found in the 5′ flanking regions of a variety of milk protein genes from different species (figure 8). All abundantly expressed milk proteins have at least one putative MPBF recognition motif

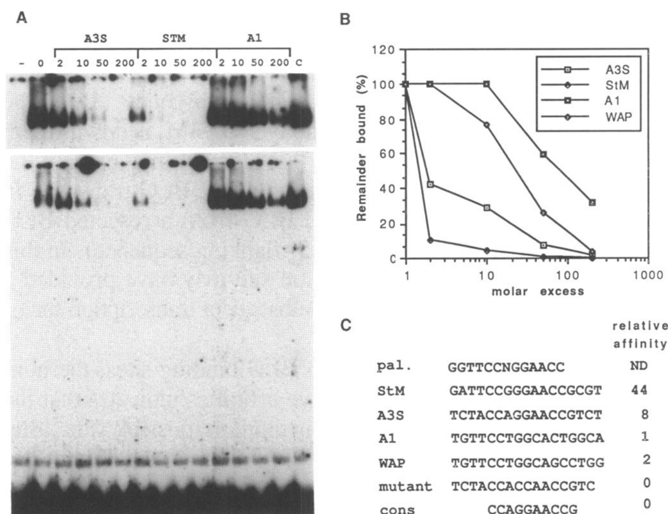

Figure 7. Comparison of mammary factor (MPBF) binding site affinities. **(A)** Increasing molar concentrations of oligonucleotides A3S, StM or A1 were added to binding reactions with either sheep (upper panel) or mouse (lower panel) lactating mammary nuclear extracts (2μg) and radiolabelled StM target. The molar excess of competitor is indicated above each lane and C is the core sequence oligonucleotide (200-fold excess). **(B)** Quantitative analysis of relative binding site affinities to sequence StM. WAP is a17bp element from the whey acidic protein promoter which has homology to the mammary factor binding site. **(C)** Sequence comparison of the MPBF sites used in this analysis and their relative binding affinities as determined by GMSA.

A

MPBF motifs in milk gene promoter regions.

13bp palindrome		GGTTCC(N)GGAACC
StM	- 93	GATTCC(G)GGAACC
A3S	- 210	TCTACC(A)GGAACC
A1	- 278	TGTTCC(T)GGCACT
bovine BLG	- 92	GATTCG(G)GGAACC
guineapig α-lac	- 470	TGTTCC(A)TGAATG
	- 300	TGTTCC(T)AGAACA
	- 77	TCTTCC(T)GGATGT
human α-lac	- 450	TGATCC(T)GGGACT
	- 280	GCTTCC(C)AGAACC
	- 60	GGTACC(A)GAGCCC
bovine α-lac	- 491	TGTTCT(T)GGAACT
bovine α-s1-cas	- 346	TGTCCC(A)AGAATT
	- 24	ATAGCT(T)GGAAGC
α-s2-cas	- 335	GGTTCT(G)TGGTAC
β-cas	-1540	GGCCTT(G)GCAACC
	- 955	CCTTCC(A)TGAATA
rat α-cas	- 251	TGATCC(A)GCAACA
β-cas	- 99	ACTTCT(T)GGAATT
γ-cas	- 415	GGTTCC(T)GCATGC
	- 150	ACTCCC(T)AGAATC
mouse WAP	- 150	TGTTCC(T)GGCAGC
	- 23	GCATCC(C)AGACAC

B

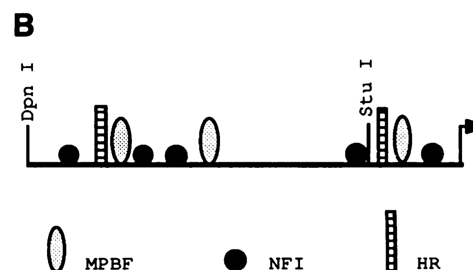

Figure 8. (A) Sequence comparison of putative MPBF sites in the promoters and 5′ flanking sequences of milk protein genes from a variety of species. Sequences were searched with the 13bp perfect palindromic sequence, GGTT-CCNGGAACC, using the Wisconsin FIND programme and allowing up to 5 mismatches. **(B)** Schematic repesentation of the sites occupied *in vitro* in the BLG minimal promoter by NFI, the mammary-specific factor, MPBF and a hormone receptor, HR.

(allowing up to 5 mismatches). It is therefore possible that this mammary factor has a role to play in regulating the expression of milk proteins. The putative binding sites are divergent and their binding potential will have to be determined *in vitro* by gel shift assay. So far, we have investigated the potential site at −150 in the mouse WAP promoter was investigated by synthesising a 17bp double-stranded oligonucleotide which was used as a competitor in EMSA with StM as probe (figure 7B and C). This competes with an efficiency intermediate between A3S and A1. The WAP and A1 sites are identical apart from a 2bp mismatch at the end of the right half of the palindrome (positions 12 and 13). Since WAP competes 2-fold more efficiently than A1, and has a conserved C at position 13 (which is changed to a T in A1), this may indicate that this C residue is important for binding. It is interesting to note that oligonucleotide A1 contains a sequence element which is very similar to a motif in the mouse mammary tumour virus (MMTV) promoter that has been shown to be essential for transcription induction by NFI and GR (18). This sequence (TGTTCCTGGCA) overlaps the MPBF site and therefore proteins bound at this site may compete with the mammary factor for occupancy of its cognate site. Since other milk protein promoters, notably α-lactalbumin, also contain a similar element, this sequence may also be involved in the regulation of milk protein genes.

DISCUSSION

An understanding of the mechanisms which regulate gene transcription in the mammary gland is of major interest for biological, oncological and even biotechnological reasons. We have used a milk protein gene promoter as a model to study the factors which are involved in regulating gene expression in the mammary gland during lactation. The ovine milk protein gene β-lactoglobulin (BLG) was selected for this study since earlier work (11) demonstrated that the 406bp proximal promoter region of BLG contained the elements necessary for high level, mammary gland-specific expression in transgenic mice. By dissecting this promoter into short, overlapping oligonucleotides, which were used in EMSA, we have identified at least 2 factors which occupy multiple sites in the BLG promoter region *in vitro*.

Multiple NFI sites in the BLG promoter region

A number of sites which bind NFI-like factors *in vitro* have been identified in this study. The NFI family of proteins has been shown to be involved in the regulation of a number of diverse genes (19) in addition to playing a role in the replication of adenovirus DNA (20). Some forms of NFI have been shown to be restricted to certain tissues (21,22) and a recent analysis of NFI in human and murine cell lines by EMSA showed that there are cell line-specific differences in both the forms and amounts of NFI binding activity (23). Recently, it has been shown that multiple NFI proteins from bovine brain bind the proenkephalin enhancer (24) and cloning of mouse brain NFI cDNAs demonstrated that the myelin basic promoter is activated by a particular form of NFI (25). In this work, we clearly demonstrate that the lactating mammary gland has at least 2 types of NFI activity with distinctly different mobility patterns and recognition site affinities (figure 4). Further analysis will be necessary to determine the composition of these complexes. It is possible that a mammary gland-specific form of NFI is involved in the regulation of BLG. Isolation of NFI cDNAs from a lactating mammary gland library is currently underway to investigate this possibility.

Since at least 5 NFI sites have been identified in the BLG promoter, it is likely that this factor plays a regulatory role in BLG transciption. DNase I footprinting experiments on the rat α-lactalbumin promoter, using nuclear extracts from lactating rat mammary glands, showed that the region from −125 to −85 is protected by bound NFI (26). This region is part of the so-called 'milk box' (27), a highly conserved element in the promoters of many milk protein genes. In BLG, the 'milk box' (−133 to −172) is poorly conserved. However, it also binds NFI *in vitro* (oligonucleotide S2, figure 3). Confirmation of the functional significance of this factor awaits *in vivo* footprinting and mutational analysis of cognate sites in milk protein gene regulatory regions. In this context, it should be noted that mutation of an NFI site in the MMTV-LTR abolishes glucocorticoid-induced transcription (18).

Multimerization of NFI sites in viral enhancers does not activate gene expression (28,29) and a single NFI site is a weak activator of a chimeric promoter (30). However, NFI interacts synergistically with the ER in transcriptional activation via a common factor (31). It is therefore likely that a combination of NFI with other regulatory factors is required for efficient transcriptional activation of BLG.

A mammary gland-specific factor (MPBF)

One such factor could be the novel DNA-binding protein which we have identified (figures 1 and 6) and termed MPBF (for milk protein binding factor). This factor is abundant in nuclear extracts from both lactating sheep and mouse mammary gland but does not appear to be present in six other control tissue extracts leading us to speculate that it may be a mammary gland-specific transcription factor with an essential role in the regulation of milk protein gene expression. One low affinity site and two high affinity sites have been shown to bind this factor *in vitro*. Two of these motifs lie within the −406 to −166 essential region of the BLG control region, removal of which drammatically reduces the frequency of expression of BLG transgenes (11). The promoter proximal, and highest affinity site, StM, is downstream from this region. The presence of this element may explain the high level, tissue-restricted expression of the BLG transgene in the single mouse which expressed BLG from a resected BLG construct (carrying only 146bp of 5′-flanking sequence). In this animal, the chromosomal integration site may have provided a favourable chromatin structure for subsequent transcription factor binding (11).

A sequence comparison of the 3 BLG binding sites, the non-binding site in A2, and their relative affinities indicates that the highest affinity MPBF recognition motif is possibly the 13bp palindromic sequence GATTCCNGGAACC. A putative optimal site is indicated in figure 7C. Compared to this, site StM has a single base mismatch, and sites A3S and A1 three mismatched bases. A3S contains a 'perfect' half site while A1 does not. The relatively large difference in binding affinity of these two sites, coupled with the fact that the consensus half site alone does not compete for binding, suggests that a complete, albeit imperfect, palindromic site is required for binding and that important protein/DNA contact points are mutated in the A1 motif (ie. the A at position 10 or the C at position 13). Confirmation of this is underway and will require a protein/DNA contact point analysis by methylation interference coupled with a detailed mutational analysis of the binding site.

This binding site is similar to those for steroid hormone receptors. The estrogen responsive element (ERE) is a 13bp palindromic motif with the consensus sequence GGTCANNNTG-ACC. This is mismatched in 5 out of 10 positions with the StM element and 3 out of 6 of the G residues which are in contact with the ER (32). There is also some sequence similarity to the glucocorticoid responsive element consensus (GRE), with 4 out of 6 matches to a half site (4). Competition analysis between consensus ERE and GRE oligonucleotides and StM for binding of the mammary complex demonstrates that the MPBF is not an ER or a GR (data not shown). It is, therefore, possible that this factor is a novel member of the steroid hormone receptor superfamily.

Occupancy of low affinity hormone response elements, but not high affinity sites, is enhanced by the presence of multiple sites. Binding of a progesterone receptor dimer to one GRE/PRE site increased the binding affinity for a second site 100-fold (33) and two ERE which failed to bind receptor individually were functional when placed together (34). This provides a mechanism for ensuring site-filling at physiological concentrations of activator while eliminating binding to 'random' weak sites. Binding of MPBF to the high affinity StM element may therefore synergistically enhance binding to the upstream elements A3S and A1.

The putative palindromic nature of the MPBF binding site raises the possibility that a dimer will be the functional entity which binds to DNA. The MPBF complex could therefore be either homo- or heterodimeric. In some transcription factor families, dimerisation is mediated by a leucine zipper structure. In the case of the AP1-binding family (*fos/jun*), changes in the redox state of the environment modulate binding activity (35). Preliminary experiments indicate that MPBF-binding is unaffected by the addition of dithiothreitol (DTT) to the binding reaction (data not shown). This would suggest that MPBF does not form dimers through a cysteine-containing leucine zipper.

MPBF-binding to a synthetic WAP promoter fragment, the existence of multiple, putative MPBF motifs in other milk protein genes (figure 8A) and the presence of MPBF in both sheep and mouse mammary gland tissue supports the idea of a general milk protein gene regulatory factor. Using a rat WAP promoter fragment, Lubon and Henninghausen (36) described the binding of a factor present in lactating rat mammary glands and MCF-7 cells (a mammary epithelial cell line) but absent in rat liver and HeLa cells. They localised the binding site for the mammary-specific factor the the region between −175 and −88 but did not further define the recognition motif or complex. The mobility of their mammary-specific complex by EMSA and the sequence of the binding domain do not correlate with the factor we have described here. Recently, Lefebvre and co-workers (37) identified a regulatory element in the MMTV LTR which binds a nuclear factor (MP4) that appears to be present only in mammary cell lines. The binding domain for this factor does not contain the MPBF recognition site.

Binding motifs occupied *in vitro* on the BLG minimal control element

The factors NFI and MPBF bind *in vitro* to elements within the 406bp control region of the BLG gene. The relative positions of their cognate sites are illustrated schematically in figure 8B. It is not known at present which of these sites are occupied *in vivo*. The proximity of NFI sites to the MPBF motifs suggests a functional interaction, and in the case of the A1 motifs where they overlap, a competition between them. It is also possible that some of these factors and sites may be involved in repression of the BLG gene in tissues other than the mammary gland. Binding sites for other transcription factors, including NFκB, ER and AP1, have also been identified in the BLG minimal promoter (CJW, unpublished observations).

The functional significance of the binding factors and sites that we have described remains to be determined. The role of the MPBF *in vivo* is presently being investigated by site-directed mutagenesis of its cognate sites. The 406bp BLG promoter/control region, with individually mutated sites, linked to a BLG genomic construct will be analysed for expression in transgenic mice. Transient co-transfection experiments in HeLa cells with NFI, MPBF and hormone receptor expression vectors and various BLG promoter constructs should provide some insights into the significance of these factors *in vivo*.

The molecular mechanisms which regulate milk protein gene expression in the mammary gland during gestation and lactation are at present unknown. The work described in this paper provides a basis for investigating the roles and likely interactions of specific transcription factors and lactogenic hormones in the mammary gland in the transcriptional control of milk protein genes.

ACKNOWLEDGEMENTS

We should like to thank Ron Hay for the anti-NFI rabbit antiserum and the HIV and SVE oligonucleotides, and Ian Armitt for oligonucleotide synthesis. We are particularly grateful to Ron Hay for helpful suggestions and comments and would also like to thank Bruce Whitelaw for critical reading of the manuscript. This work was supported by a grant from the Transgenic Animal Programme of the AFRC.

REFERENCES

1. Mitchell,J.P. and Tijan,R. (1989) *Science*, **245**, 371−378.
2. Horikoshi,M., Hai,T., Lin,Y-S., Green,M.R. and Roeder,R.G. (1988) *Cell*, **54**, 1033−1042.
3. Workman,J., Taylor, I.C.A. and Kingston,R.E. (1991) *Cell*, **64**, 533−544.
4. Beato,M. (1989) *Cell*, **56**, 335−344.
5. Borellini,F. and Oka,T. (1989) *Environmental Health Pers.*, **80**, 85−99.
6. Topper,J. and Freeman,S. (1985) *Phys. Rev.*, **60**, 1049−1106.
7. Clark,A.J., Simons,J.P., Wilmut,I., and Lathe,R. (1987) *Trends in Biotechnology*, **5**, 20−24.
8. Doppler,W., Groner,B., and Ball,R.K. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 104−108.
9. Simons,J.P., McClenaghan,M., and Clark,A.J. (1987) *Nature*, **328**, 530−532.
10. Harris,S., McClenaghan,M., Simons,J.P., Ali,S. and Clark,A.J. (1991) *Dev. Gen.* **12** 299−307.
11. Whitelaw,C.B.A., Harris,S., McClenaghan, M., Simons,J.P., and Clark,A.J. (1991). manuscript submitted.
12. Dignam,J.D., Lebovitz,R.M., and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475−1489.
13. de Vries,E., van Driel,W., van den Heuvel,S.J.L., and van der Vliet,P.C. (1987) *EMBO J.*, **6**, 161−168.
14. Gil,G., SmithJ.R., Goldstein,J.L., Slaughter,C.A., Orth,K., Brown,M.S., and Osborne,T. (1988) *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 8963−8967.
15. Santoro,C., Mermod,N., Andrews, P.C. and Tijan,R. (1988) Nature **334**, 218−224.
16. Jackson,S.P. and Tijan,R. (1988) *Cell*, **55**, 125−133.
17. Meisternst,M., Gander,I., Rogge,L. and Winnacker, E-L. (1988) *Nuc. Acids Res.* **16** 4419−4435.
18. Bruggemeier,U., Rogge,L., Winnaker,E.L., and Beato,M. (1990) *EMBO J.*, **9**, 2233−2239.
19. Graves,R.A., Tontonoz,P., Ross,S.R., and Spiegelman,B.M. (1991) *Genes & Dev.*, **5**, 428−437.

20. Nagata,K., Guggenheimer,R.A., and Hurwitz,J. (1983) *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 6177–6181.
21. Paonessa,G., Gounari,R., Frank,R. and Cortese,R. (1988) *EMBO J.*, **7**, 3115–3123.
22. Lichtsteiner,S., Wuarin,J., and Schibler,U. (1987) *Cell*, **51**, 963–973.
23. Goyal,N., Knox,J., and Gronostajski,R.M. (1990) *Mol.and Cell. Biol.*, **10**, 1041–1048.
24. Chu,H.-M., Fischer,W.H., Osborne,T.F., and Comb,M.J. (1991) *Nucleic Acids Res.*, **19**, 2721–2728.
25. Inoue,T., Tamura,T., Furuichi,T., and Mikoshiba,K., (1990) *J. Biol. Chem.* **265**, 19065–19070.
26. Lubon,H. and Hennighausen,L. (1988) *Biochem. J.*, **256**, 391–396.
27. Hall,L., Emery,D.C., Davies,M.S., Parker,D., and Craig,R.K. (1987) *Biochem. J.*, **242**, 735–742.
28. Gloss,B., Yeo,G.M., Meisterenst,M., Rogge,L., Winnaker,E.L., and Bernard,H.U. (1989). *Nucleic Acids Res.*, **17**, 3519–3533.
29. Nowock,J., Borgmeyer,U., Puschel,A.W., Rupp,R.A.W., and Sippel,A.E. (1985) *Nucleic Acids Res.*, **13**, 2045–2061.
30. Knox,J.J., Rebstein,P.J., Manoukian,A., and Gronostajski,R.M. (1991) *Mol. & Cell. Biol.*, **11**, 2946–2951.
31. Martinez,E., Dusserre,Y., Wahli,W., and Mermod,,N. (1991) *Mol. & Cell. Biol.*, **11**, 2937–2945.
32. Kumar,V. and Chambon, P. (1988) *Cell*, **55**, 145–156.
33. Tsai,S.Y., Tsai,M.-J., and O'Malley,B.W. (1989) *Cell*, **57**, 443–448.
34. Martinez,E., Givel,F., and Wahli,W. (1987) *EMBO J.*, **6**, 3719–3727.
35. Frame,M.C., Wilkie,N.M., Darling,A.J., Chudleigh,A., Pintzas,A., Lang,J.C., and Gillespie,D.A.F. (1991) *Oncogene* in press.
36. Lubon,H. and Hennighausen,L. (1987) *Nucleic Acids Res.*, **15**, 2103–2121.
37. Lefebvre,P., Berard,D.S., Cordingley,M.G., and Hager,G.L. (1991) *Mol. & Cell. Biol.*, **11**, 2529–2537.
38. Harris,S., Ali,S., Anderson,S., Archibald,A., and Clark,A.J. (1988) *Nucleic Acids Res.*, **16**, 10379.