Cloning and Characterization of the Rat Bone Sialoprotein (BSP) Gene Promoter

By

Jack Jun Li

Faculty of Dentistry
University of Toronto

A thesis submitted in conformity with the requirements
for the Degree of Doctor of Philosophy at the
University of Toronto

© Copyright by Jack Jun Li 1998
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-35220-X
To my beloved grandmother, Madam Gu Feng-Zhi
ABSTRACT

Cloning and Characterization of the Rat Bone Sialoprotein (BSP) Gene Promoter
Jack Jun Li, Ph.D., 1998
Faculty of Dentistry, University of Toronto

Bone sialoprotein (BSP), a phosphorylated and sulphated glycoprotein that is essentially specific to mineralized connective tissues, has been implicated in hydroxyapatite nucleation. The expression of BSP gene is developmentally regulated in accordance with tissue mineralization and is further modulated by osteotropic hormones such as 1,25-dihydroxyvitamin D₃, which suppresses its expression, and glucocorticoids, which stimulate its expression. To elucidate mechanisms of BSP gene regulation at the transcriptional level, the promoter of the rat BSP gene was cloned and shown to be functional in transient transfection assays. The rat BSP promoter, comprising ~3 kb of the 5'-flanking region of the gene, contains a unique inverted TATA box overlapped by a vitamin D₃ response element (VDRE) sequence, a homeobox-binding site, a novel 18-nucleotide palindrome and a number of other consensus cis-acting DNA elements that may contribute to tissue-specific and developmentally regulated transcription. Based on the highly conserved sequence of the proximal promoter between the rat and human BSP genes, a 370 bp sequence was designated the “BSP box”. The inverted TATA box was found to drive downstream transcription and the TATA-binding protein (TBP) shown to be required for this function using DNase I footprinting, site-directed mutagenesis, and in vitro transcription analyses. Thus the direction of transcription is not determined by the asymmetric nature of a TATA box. Gel shift analysis showed that the vitamin D₃ receptor (VDR) binds specifically to the BSP VDRE with a strong preference for homodimer formation. However, a truncated VDR that includes the DNA binding domain of the VDR failed to interact with the BSP VDRE, suggesting that complete configuration of the receptor protein is required for DNA binding. That the putative VDRE
overlaps the inverted TATA box indicates a novel mechanism for gene suppression by vitamin D₃, whereby the VDR competes with the TBP for occupancy of the TATA/VDRE element and prevents the assembly of the pre-initiation complex which is required for the initiation of transcription. These studies were the first to clone and characterize a BSP promoter and have identified novel mechanisms of transcriptional regulation through a unique inverted TATA/VDRE element.

Key words:

BSP, Bone Sialoprotein, Promoter, Transcription, Transcription Initiation, Transcriptional Regulation, TATA box, TATA-box Binding Protein (TBP), TFIID, Vitamin D Response Element (VDRE), Vitamin D₃ Receptor (VDR)
ACKNOWLEDGMENTS

"Never clone alone!" - These words mean a lot even beyond the molecular world.

The production of this thesis involves not only a long time since I started my joint Ph.D./D.D.S. program but also the effort of a number of professors, fellow students, and technical personnel. Without their stimulatory advice, arguments, and assistance, the completion of this work would not have been possible. Thus, the list of those deserving my appreciation is very long. I would apologize since I could not include them all into this single page.

I would like to express my sincere thanks to my supervisor, Dr. Jaro Sodek, for his generosity to give me the opportunity to work in his laboratory. In particular, his stimulatory guidance, unbelievable patience, and immeasurable support in all aspects of my studies are unforgettable learning experiences I had never gone through before. I especially enjoyed the free, frank, yet focused and fruitful talks and debates with him over the years.

I also would like to thank Dr. Neil Miyamoto at the Ontario Cancer Institute who generously allow me to work in his lab during my first phase of Ph.D. training so that I got a chance to broaden my knowledge, to be exposed to the cutting-edge research environment, and to experience the thrills and disappointments inherent in research. The initial helpful advice from Drs. J. Aubin, V. Giguere, L. Freedman, and J. Greenblatt are also appreciated. They generously provided the rat BSP cDNA, luciferase plasmid, expression vectors for the vitamin D3 and glucocorticoid receptors, recombinant VDR and TBP used in my studies. I also grateful to Kamling Yao, Richard Kim, Yori Ogata, and Harry Moe for their excellent technical assistance.

My wife and son have patiently accompanied me through this long march and offered me the joy of life. I am greatly indebted to them. Our special family friends, Miriam and Phil Irwin, who are not scientists and have never visited my research lab, are also deserve a special note here. They made this new world my home. It is all these folks who remind me constantly that there is a greater world outside the PCR microtubes and the humane side of osteoblasts.

I also would like to acknowledge my parents, a doctor and a scientist, for providing me an intellectually challenging environment and for demonstrating their constant encouragement of my academic endeavors. It has taken so long for me to complete the final draft that my grandmother, Madam Gu Feng-Zhi, was not able to witness this accomplishment. Although she had never been in a school, she always held a high hope for me to get a higher education and her sound philosophy about life has fostered me since I was an infant. This thesis is dedicated to her.

Special thanks should go to the I’Anson Foundation, who provided me with an I’Anson Fellowship to support me in the combined Ph.D./D.D.S. program, and to MRC (Canada) for supporting my research through a Group Grant.
PUBLICATIONS DURING THE PH.D. PROGRAM

Refereed Articles:


Book Chapters:

Abstracts and Presentations:


## AWARDS DURING THE PH.D. PROGRAM

<table>
<thead>
<tr>
<th>Year</th>
<th>Award</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992-97</td>
<td>I'Anson Fellowship</td>
</tr>
<tr>
<td>1995</td>
<td>International Association for Dental Research (IADR) Travel Award</td>
</tr>
<tr>
<td>1995</td>
<td>First Prize for the Best Student Dental Research Project and Presentation, Faculty of Dentistry, University of Toronto</td>
</tr>
<tr>
<td>1994</td>
<td>Canadian Association for Dental Research (CADR) Student Research Award</td>
</tr>
<tr>
<td>1993</td>
<td>International Association for Dental Research (IADR) Travel Award</td>
</tr>
<tr>
<td>1992</td>
<td>Canadian Association for Dental Research (CADR) Student Research Award</td>
</tr>
<tr>
<td>1992</td>
<td>UpJohn Achievement Award: Grand prize (Graduate, Postgraduate and Undergraduate) for the Best Student Dental Research Project and Presentation, Faculty of Dentistry, University of Toronto</td>
</tr>
<tr>
<td>1990-92</td>
<td>University of Toronto Open Fellowship</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>PUBLICATIONS DURING THE PH.D. PROGRAM</td>
<td>vi</td>
</tr>
<tr>
<td>AWARDS DURING THE PH.D. PROGRAM</td>
<td>ix</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xiv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xvi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xviii</td>
</tr>
</tbody>
</table>

## Chapter I Literature Review

A. Bone: A Mineralized Connective Tissue
   A.1. Development and Maturation of Bone                                2
   A.2. Cells Involved in Bone Formation and Resorption                 3
   A.3. Extracellular Matrix (ECM) Proteins of Bone                      5

B. Extracellular Signals that Regulate the Lineage Development of Osteoblast
   B.1. Glucocorticoids                                                 12
   B.2. Vitamin D₃                                                       13
   B.3. Parathyroid Hormone (PTH) and PTH-related Protein (PTHrP)       14
   B.4. Transforming Growth Factor-β (TGF-β) Family                    15
   B.5. Other Hormones, Growth Factors, and Cytokines                  16

C. Expression and Regulation of Eukaryotic Genes
   C.1. Structure of Eukaryotic Genes                                    17
   C.2. Transcription                                                    18
   C.3. Post-transcription Regulation                                    19

D. Transcriptional Regulation: the Ultimate Target of Signal Transduction and the Initial Step of Gene Expression
   D.1. Transcription Factors and cis-acting Elements                    20
   D.2. Protein Structure and DNA Recognition                            21
   D.3. General Transcription Factors (GTFs)                             25
   D.4. TATA-Binding Protein (TBP): a Universal Transcription Factor     27
   D.5. TATA Box: a Central Role in Transcriptional Control              30
 CHAPTER II RATIONALE, OBJECTIVES, AND HYPOTHESES ........................................... 39

CHAPTER III CLONING OF THE RAT BONE SIALOPROTEIN (BSP) GENE PROMOTER ................................................................. 42
Summary ........................................................................................................ 43
Introduction ............................................................................................... 44
Materials and Methods ............................................................................ 45
Genomic Library Screening ......................................................................... 45
Genomic Mapping and DNA Sequencing .................................................... 45
Comparison of the Rat and Human BSP Gene Promoters ........................... 48
Primer Extension Analysis ........................................................................... 48
CAT Plasmid Construction ........................................................................ 48
DNA Transfection and CAT Analysis ......................................................... 48
Results and Discussion .............................................................................. 49

CHAPTER IV AN INVERTED TATA BOX DIRECTS DOWNSTREAM TRANSCRIPTION OF THE BSP GENE ............................................... 62
Summary ...................................................................................................... 63
Introduction ............................................................................................... 64
Materials and Methods ............................................................................. 66
DNase Footprinting ..................................................................................... 66
Site-directed Mutagenesis .......................................................................... 66
CAT Plasmid Construction and the CAT Assays .......................................... 69
In Vitro Transcription ............................................................................... 69
Results ....................................................................................................... 70
TATA-binding Protein Binds to the Inverted TATA Box ............................... 70
Disruption of Binding of TBP to the Inverted TATA Element by Site-Directed Mutagenesis ................................................................. 71
Requirement of the Inverted TATA Box in Directing Downstream Expression of the BSP Gene ................................................................. 71
TBP is Required for Initiation of the BSP Gene Promoter in Transcription...... 74
Discussion

Chapter V Repression of the Rat BSP Gene Expression by Vitamin D3 through a Receptor Element Overlapping the Inverted TATA Box .................. 85
Summary ................................................................. 86
Introduction ............................................................. 87
Materials and Methods ............................................. 88
    Construction of Recombinant Plasmid ......................... 88
    Cell Culture and Co-transfection ............................... 88
    Analysis of Luciferase Activity ................................. 89
    Preparation of Endogenous and Recombinant VDR ............... 89
    Electrophoretic Gel Mobility Shift Assays ...................... 89
Results ........................................................................ 90
    Molecular Organization of Various VDREs ....................... 90
    DNA Elements that Mediates Suppression of BSP Transcription
        by Vitamin D3 ....................................................... 92
    Recognition of the rBSP VDRE (but not the TATA Box Element)
        by Natural VDR ................................................... 92
    Binding of Both Natural and Recombinant VDRs to the rBSP VDRE .... 95
    Binding Specificity of the Natural and Recombinant VDRs to the
        rBSP VDRE .......................................................... 96
Discussion .................................................................... 101

Chapter VI General Discussion and Future Directions ............... 106
1. Cloning of the BSP Promoter ........................................ 108
2. Tissue-specific Transcription of the BSP Gene .................. 111
3. The Inverted TATA Box Provides a Unique Model to Explore Many
    Fundamental Issues in Transcription Initiation .................. 116
4. The BSP VDRE Indicates a Novel Mechanism for Vitamin D3
    to Repress Gene Transcription ....................................... 119
5. Characterization of Upstream Enhancers in the Rat BSP Promoter .... 122
Chapter VII  Concluding Remarks ................................................................. 125

BIBLIOGRAPHY ................................................................. 129
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALP</td>
<td>alkaline phosphatase</td>
</tr>
<tr>
<td>α-MEM</td>
<td>α-minimal essential medium</td>
</tr>
<tr>
<td>BMP</td>
<td>bone morphogenetic protein</td>
</tr>
<tr>
<td>bp</td>
<td>basepair</td>
</tr>
<tr>
<td>BSP</td>
<td>bone sialoprotein</td>
</tr>
<tr>
<td>Ca²⁺</td>
<td>calcium ion</td>
</tr>
<tr>
<td>CAT</td>
<td>chloramphenicol acetyltransferase</td>
</tr>
<tr>
<td>CHX</td>
<td>cycloheximide</td>
</tr>
<tr>
<td>DRB</td>
<td>5-6-dichloro-1-β-D-ribofuranosyl benzimidazole</td>
</tr>
<tr>
<td>DSP</td>
<td>dentin sialoprotein</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EGF</td>
<td>epidermal growth factor</td>
</tr>
<tr>
<td>FBS</td>
<td>fetal bovine serum</td>
</tr>
<tr>
<td>FGF</td>
<td>fibroblast growth factor</td>
</tr>
<tr>
<td>GR</td>
<td>glucocorticoid receptor</td>
</tr>
<tr>
<td>GRE</td>
<td>glucocorticoid response element</td>
</tr>
<tr>
<td>HA</td>
<td>hydroxyapatite</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase</td>
</tr>
<tr>
<td>LSC</td>
<td>liquid scintillation counting</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>OC</td>
<td>osteocalcin</td>
</tr>
<tr>
<td>OPN</td>
<td>osteopontin</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PIC</td>
<td>pre-initiation complex</td>
</tr>
<tr>
<td>Pol II</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>RC cell</td>
<td>rat calvarial cells</td>
</tr>
<tr>
<td>RGD</td>
<td>arginine-glycine-aspartate</td>
</tr>
<tr>
<td>ROS 17/2.8 cells</td>
<td>rat osteosarcoma 17/2.8 cells</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SPARC</td>
<td>secreted protein acidic rich in cysteine</td>
</tr>
<tr>
<td>TAF</td>
<td>TBP-associated factors</td>
</tr>
<tr>
<td>TBP</td>
<td>TATA-box binding protein</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>TBS</td>
<td>Tris-HCl-buffered saline</td>
</tr>
<tr>
<td>TE</td>
<td>Tris-HCl/EDTA buffer</td>
</tr>
<tr>
<td>TFIID</td>
<td>transcription factor IID</td>
</tr>
<tr>
<td>TGF</td>
<td>transforming growth factor</td>
</tr>
<tr>
<td>t.l.c.</td>
<td>thin-layer chromatography</td>
</tr>
<tr>
<td>TPA</td>
<td>12-O-tetradecanoylphorbol 13-acetate</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
<tr>
<td>VDR</td>
<td>vitamin D receptor</td>
</tr>
<tr>
<td>VDRE</td>
<td>vitamin D response element</td>
</tr>
<tr>
<td>vitamin D$_3$</td>
<td>1,25-dihydroxyvitamin D$_3$</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Chapter I

Figure 1.1 Initiation of Transcription and Activation of Basal Transcription Machinery...... 31
Figure 1.2 Regulation of Gene Transcription by Steroid Hormones............................ 38

Chapter III

Figure 3.1 cDNA Probes Generated by PCR for Screening a Rat Genomic DNA Library.... 46
Figure 3.2 Restriction Enzyme Map of the 5'-Flanking Region of the Rat BSP Gene....... 51
Figure 3.3 Partial Nucleotide Sequence of the Promoter Region of the Rat BSP Gene.... 52
Figure 3.4 Extended Nucleotide Sequence of the Promoter Region of the Rat BSP Gene .......................................................... 53
Figure 3.5 Sequence Comparison of the 5'-flanking Regions of the Rat (R) and Human (H)BSP Genes............................................................ 54
Figure 3.6 Determination of the Transcription Start Site of the Rat BSP Promoter........ 55
Figure 3.7 Promoter Activity Analysis of the BSP-CAT Chimeric Constructs.............. 59
Figure 3.8 Down-regulation of Rat BSP Gene Promoter by 1,25-(OH)2-vitamin D3....... 61

Chapter IV

Figure 4.1 Footprint of the TATA-binding Protein on the Rat BSP Promoter Sequence.......................... 67
Figure 4.2 Nucleotide Sequencing of Point Mutations in the Rat BSP TATA Box........ 68
Figure 4.3 DNase Footprinting Experiments Demonstrating the Binding of Recombinant Human TBP to the Inverted TATA Element in the BSP Gene.Promoter............ 72
Figure 4.4 Effect of Point Mutations on the Binding of TBP to the Inverted TATA Box................................................................. 73
Figure 4.5 Effect of Point Mutations on the Rat BSP Promoter Activity.................... 75
Figure 4.6 In vitro Transcription Analysis of the Rat BSP Promoter Encompassing Wild-type or Mutated Forms of the Inverted TATA Box........................................ 76

xvi
Figure 4.7  Reconstitution of *In Vitro* Transcription Directed by the Rat BSP Promoter with Recombinant Human TBP.......................... 77

Figure 4.8  Frequency of Nucleotide Substitutions in Functional Eukaryotic TATA Boxes.................................................... 80

Figure 4.9  Model Depicting Possible Interactions between the TBP and the Inverted TATA Box.................................................. 84

Chapter V

Figure 5.1  Characterization of the DNA Sequences that Confer the Repressive Effect of Vitamin D₃ by Transient Transfection Assays.............................................. 93

Figure 5.2  Gel Mobility Shift Analysis of the Natural VDR Binding to the rBSP VDRE.................................................................. 94

Figure 5.3  Gel Mobility Shift Analysis of the Binding Specificity of the Natural VDR from MG63 Cells to the Rat BSP VDRE...................................................... 97

Figure 5.4  Gel Mobility Shift Analyses of the Truncated Form of Recombinant VDR (hVDRF) binding to VDREs.......................................................... 98

Figure 5.5  Gel Mobility Shift Analyses of the Binding of the Recombinant hVDR to VDREs.................................................................................. 99

Figure 5.6  Gel Mobility Shift Analyses of the Binding Specificity of Recombinant hVDR to the Rat BSP VDRE.................................................. 100

Figure 5.7  Model of Vitamin D₃-mediated Suppression of the rBSP Gene Transcription....................................................... 103
**LIST OF TABLES**

**Table 1.1** Major Non-collagenous Proteins of Mineralized Bone........................................ 7

**Table 3.1** Positive Clones from the Rat Genomic Library Screening..................................... 47

**Table 5.1** Comparison of the Putative rBSP VDRE with the Consensus and
Other Identified VDREs........................................................................................................ 91
Chapter I

Literature Review
A. Bone: A Mineralized Connective Tissue

Bone is a specialized connective tissue that, together with cartilage, forms the skeletal system. As such, it has important primary functions in the support and protection of internal organs such as the brain and lungs, and provides an attachment for muscles required for locomotory and breathing functions. Also, it forms the framework of craniofacial complex, including alveolar processes that support teeth for mastication. In addition, bone houses the haematopoietic system within marrow spaces and provides a dynamic reservoir for Ca²⁺, P; and other ions in the body. As in other connective tissues, the fundamental constituents of bone are the cells and extracellular matrix. Characteristically, the mineralized tissues form the majority of the bone structure. Bone tissue is formed by the synthesis of a collagen matrix by osteoblastic cells, and the deposition of mineral ions within and between fibrils of type I collagen. The mineral exists as small, irregular crystals of hydroxyapatite [Ca₁₀(PO)₄(OH)₂]. Mineralization, or more specifically calcification, is a unique characteristic of bone, cementum and dentine, which does not normally occur in other collagen-containing connective tissues such as skin chorium, tendons, sclera and periodontal ligament.

A.1. Development and Maturation of Bone

Connective tissues, including the bone and cartilage of the skeleton, are derived from undifferentiated mesenchyme. Bone formation begins as a condensation of mesenchymal cells, which differentiate into osteoblastic cells that produce bone matrix molecules. Descriptive embryology and anatomy distinguish two types of bone development: intramembranous and endochondral. In the process of intramembranous ossification, which occurs in the formation of the flat bones of the skull, mesenchymal precursor cells first condense then differentiate directly into osteoblastic cells. The flat bones, known as lamellar bones, grow by the addition of new bone in layers to the outer surfaces and also at osteogenic fronts, which generate the suture lines. Osteoblastic differentiation within the sutures and subsequent bone deposition are intimately
coordinated with the growth of the skull and the timing of suture closure. In a similar manner to intramembranous ossification, radial growth of cortical bone in long bones is generated by committed osteoblasts in the periosteum. In the process of endochondral ossification, undifferentiated mesenchyme condenses initially to form a core of differentiated chondroblasts which produce cartilage templates that initiate long bone formation and subsequent growth. The cartilage templates are largely replaced by bone which forms after vascular invasion. Osteoblasts and osteoclasts, brought in by the vasculature, function coordinately to gradually replace the cartilage scaffold with a trabecular bone matrix, within which the bone marrow cavity is formed. Longitudinal bone growth takes place through endochondral ossification in the growth plates located at the epiphyses of long bones. In the epiphyseal plates cartilage cells undergo hypertrophy and the surrounding matrix mineralizes. Subsequently, the mineralized cartilage is removed by osteoclasts and replaced by new trabecular bone. Ultimately, almost all remaining cartilage is replaced by bone except at the articular surfaces of the joints (reviewed in Cormack, 1987).

A.2. Cells Involved in Bone Formation and Resorption

The formation and growth of the skeletal tissues involves the regulated differentiation of specialized cells of the osteogenic and chondrogenic, as well as osteoclastic, lineages. These cells function inter-dependently in a temporo-spatial manner to generate the different types of bones. The major cell types that contribute to skeleton formation and homeostasis are described below:

Osteoblasts, Osteocytes and Lining Cells: Osteoblasts are the cells responsible for bone formation. Preosteoblasts, the committed progenitor cells, are believed to originate from local mesenchymal stem cells residing in the stromal tissues of the bone marrow and in other connective tissue compartments, such as the periosteal layer of bone (Owen, 1985). These osteogenic precursors, upon receiving appropriate signals, undergo proliferation and differentiate into mature osteoblasts. The osteoblasts form as clusters of cuboidal cells along the bone surface where they produce a number of extracellular matrix proteins, including type I collagen, which forms the bone tissue. The plasma membrane of the osteoblast is characteristically rich in alkaline
phosphatase (ALP) which is attached to the cell surface and can elevate \([P_i]\) by hydrolysis of organic phosphates. Some ALP can be cleaved from the cell surface and is incorporated into the bone matrix. Osteoblasts characteristically express receptors for parathyroid hormone, estrogens, vitamin D_3, and glucocorticoids, which mediate the signalling pathways that regulate cell activities (Heersche and Aubin 1986; Rodan and Noda, 1991; Aubin et al. 1992, 1993). Upon completion of their synthetic activity some osteoblasts become surrounded by the bone tissue matrix, becoming osteocytes which retain cellular contact with other osteocytes and osteoblasts through cellular processes that exist within an intricate network of canals known as canaliculi. Although osteocytes no longer produce bone matrix, they may have an important role in maintaining bone homeostasis. While some osteoblasts undergo apoptosis, those that remain on the bone surface after matrix formation has terminated change in morphology, becoming flattened, and exist as lining cells which cover the mineralized bone surface.

**Chondroblasts and Chondrocytes:** Chondroblasts, which are the cells responsible for cartilage formation, are also of mesenchymal origin, and are probably derived from a common osteoblast precursor. Under the influence of a series of extracellular signals, committed pre-chondroblasts differentiate into chondroblasts which produce a cartilage matrix which is characterized by an extensive network of type II collagen and large proteoglycan aggregates, that comprise chondroitin sulphate (aggrecan) subunits and hyaluronan. The chondroblasts become embedded in the cartilage as chondrocytes. However, in the epiphyses, chondroblastic cells follow a defined pattern of proliferation which is followed by hypertrophy, associated with the expression of type X collagen and matrix calcification, and eventually by re-differentiation into osteogenic cells, or cell death (Erlebacher et al., 1995; Roach, 1996). Most cartilage in the skeletal system serves as a template for organogenesis and is subsequently replaced by bone and bone marrow.

**Osteoclasts:** Osteoclasts are the cells responsible for bone and mineralized cartilage resorption. They are large multinucleated cells usually found in contact with calcified bone or
cartilage surfaces within a lacuna (Howship's lacunae) which is formed as a result of their own resorptive activity. Osteoclasts are of hematopoietic origin, and most likely differentiate from the mononuclear-phagocyte lineage during the pro-monocyte stage (Kahn et al. 1975, Walker, 1975). However, the osteoclast membrane is devoid of several macrophage markers. Receptors for calcitonin and estrogen, but not for parathyroid hormone (PTH) nor vitamin D₃, are expressed by these cells. Calcitonin is a potent inhibitor of osteoclastic bone resorption, acting through multiple mechanisms, whereas PTH and 1,25-vitamin D₃ are potent stimulators of osteoclasts, that appear to act through osteoblastic (lining cells) to stimulate osteoclast progenitors to differentiate and fuse into mature osteoclasts (Roodman et al. 1985, McSheehy and Chambers, 1986).

Although cells comprise less than 2% of the total bone volume, they dictate the development and turnover of the skeleton throughout life. The activities and differentiation of osteoblasts and osteoclasts are closely coordinated, not only during bone development and growth, but also during adult life as bone undergoes continuous remodelling. Imbalance between bone formation and resorption as a result of abnormal remodelling results in osteoporosis (reduced bone mass) or osteopetrosis (increased bone mass), and in many other bone diseases.

A.3. Extracellular Matrix (ECM) Proteins of Bone

The ECM of bone is composed of dense collagen fibres embedded in a matrix that is rich in glycoproteins and proteoglycans. While type I collagen fibres form the framework of bone, the non-collagenous proteins also have important structural roles and are involved in regulation of bone remodelling. Some of the ECM proteins are believed to be involved in the nucleation and regulated growth of the hydroxyapatite (HA) crystals that are associated with collagen fibrils, while others may play roles in proliferation and maturation of bone cells. In addition, some of the ECM proteins appear to be chemotactic for a variety of cells, and can induce cell attachment and spreading through Arg-Gly-Asp (RGD) amino acid motifs. There are also proteins in bone that are synthesized in other tissues, circulate in the blood and tissue fluids, and are harboured in the bone
matrix because of their affinity for HA.

**Collagen:** Collagen represents the major protein in bone and cartilage matrices, and plays an important role in determining the size, shape, and strength of these tissues. Various types of collagen are expressed at defined times and locations during bone development: type I in bone and dentin, type II in cartilage, and type X in hypertrophic cartilage (Tilstra and Byers, 1994). In bone, the collagenous matrix is formed almost entirely of type I collagen which represents ~70% of the organic constituents and ~90% of the proteins. The collagen is assembled into fibrils and fibres within and between which crystals of hydroxyapatite are formed. The intertwined collagen fibres in woven bone contrast the more regular arrangement of fibres in lamellar bone. A small amount (~3%) of type V collagen is also found in bone (reviewed in Sodek *et al.*, 1991). Collagen has been shown to be capable of heterogenous (epitactic) nucleation of hydroxyapatite, but the induction of HA nucleation by collagen is slow (Mergenhagen *et al.* 1960). Therefore, interest has been focused on collagen-associated non-collagenous proteins which have an ability to initiate the mineralization process, particularly those which are produced specifically by bone-forming cells.

**Non-collagenous Proteins (NCPs):** Non-collagenous proteins, including glycoproteins and proteoglycans, comprise 10-15% of the total bone proteins. To date, the major NCPs which are present in bone tissues have been characterized (summarized in Table 1.1; Heinegärd *et al.*, 1989; Sodek, 1991; Delmas and Malaval, 1993; Young *et al.*, 1993). However, the specific functions of these proteins have yet to be clearly elucidated.

The major proteoglycans (PGs) isolated from bone, CS-PG I (biglycan), CS-PG II (decorin), and CS-PG III, are small in size and all contain chondroitin sulphate (CS) as the side chains (reviewed in Sodek *et al.*, 1991). While CS-PG I and II are also present in non-mineralized tissues, CS-PG III appears to be more specific to bone (Goldberg *et al.*, 1988). CS-PG I and II may influence the assembly and packing of collagen fibrils (Ruoslanti, 1988; Fisher *et al.*, 1989), suggesting that they have an important role in organizing the collagenous matrix of bone. CS-PG III associates with HA rapidly. Therefore, it could influence the crystal growth and stability
(Sodek et al., 1991; 1992a). Recently it was found that decorin can bind to TGF-β, implying that PGs can modulate the effect of TGF-β, thereby influencing cell proliferation and differentiation in a variety of tissues, including bone (Yamaguchi et al., 1990; Takeuchi et al., 1994).

### Table 1.1 Major Non-collagenous Proteins of Mineralized Bone

<table>
<thead>
<tr>
<th></th>
<th>Size (kDa)</th>
<th>SO₄</th>
<th>PO₄</th>
<th>Properties and Suggested Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proteoglycans</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS-PG I (Biglycan)</td>
<td>-200</td>
<td>++</td>
<td>-</td>
<td>Binds to collagen, modulates fibril formation, growth factor binding, cell differentiation?</td>
</tr>
<tr>
<td>Core protein</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS-PG II (Decorin)</td>
<td>-120</td>
<td>++</td>
<td>-</td>
<td>Binds to collagen?, inhibition of mineralization? regulation of fibrillogenesis?</td>
</tr>
<tr>
<td>Core protein</td>
<td>45, 47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS-PG III</td>
<td>-110</td>
<td>++</td>
<td>-</td>
<td>Binds to HA, regulates growth and properties of crystals?</td>
</tr>
<tr>
<td>Core protein</td>
<td>30-38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sialoproteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bone Sialoprotein</td>
<td>34</td>
<td>++</td>
<td>+++</td>
<td>Binds to HA, nucleation of HA, marker of mineralization, cell adhesion.</td>
</tr>
<tr>
<td>(BSP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteopontin (OPN)</td>
<td>32</td>
<td>+</td>
<td>+++</td>
<td>Binds to HA, regulates crystal growth? cell adhesion and migration.</td>
</tr>
<tr>
<td>Osteocalcin (OCN)</td>
<td>6</td>
<td>-</td>
<td>-</td>
<td>Binds to HA, osteoblast-specific, marker of bone turnover, regulates mineralization and bone resorption?</td>
</tr>
<tr>
<td>Osteonectin/SPARC</td>
<td>30</td>
<td>-</td>
<td>+/-</td>
<td>Binds to HA, cell-substrate and cell-cell interactions, cell proliferation, tissue remodelling?</td>
</tr>
</tbody>
</table>

---

7
Osteonectin (also called SPARC protein) represents up to 25% of the NCPs in the bones of larger mammals (Termine et al., 1981). Although it was originally thought to be specific to bone, osteonectin is produced by a variety of cell types including fibroblasts and endothelial cells (reviewed in Young et al., 1993). While it was described as phosphoprotein in bovine bone, several studies have failed to observe any incorporation of $^{32}$PO$_4$ into rat or pig osteonectin (Sodek et al., 1991). Thus, osteonectin is unlikely involved in the mineralization process, although it could affect later growth and dissolution of the crystals due to its ability to bind to HA and collagen. More recent studies indicate that osteonectin may play a role in developmental processes, wound healing, and tissue remodelling (Young et al., 1993; Lane and Sage, 1994).

Osteocalcin is probably the most intensively studied NCP in bone although its precise function remains unknown. Also known as bone γ-carboxyglutamic acid (gla) protein, osteocalcin is a small 5.8 kDa protein that represents 1% to 2% of the total rat bone proteins. Studies in the past few years have demonstrated that osteocalcin is specific to mineralized connective tissues (Mark et al., 1988), although the related mRNA has been detected in platelets and megakaryocytes (Thiede et al., 1994). While osteocalcin is produced by osteoblasts and osteocytes, its expression is initiated after bone formation has occurred (Yoon et al., 1987; Yao et al., 1994). It has been shown that osteocalcin is involved in osteoclastic resorption (Glowacki et al., 1989), and osteocalcin in serum is commonly used as a measurement of bone remodelling activity.

Phosphoproteins are of particular importance in view of their perceived efficacy in initiating and promoting mineralization (Endo and Glimcher, 1989, Glimcher, 1989). During the last decade, two major phosphoproteins enriched in sialic acid have been characterized from bone: osteopontin (OPN, or bone sialoprotein I, or 44-kDa phosphoprotein) and bone sialoprotein (BSP, or bone sialoprotein II) (Franzén and Heinegärd, 1985; Butler et al., 1989; Fisher et al., 1990; Denhardt and Guo 1993). Another sialic acid containing protein, BAG-75 (bone acidic glycoprotein), whose N-terminal sequence is distinct from both OPN and BSP, has also been identified from rat bone (Gorski et al., 1990). Both OPN and BSP are similar in size and chemical
properties. Each protein is glycosylated, phosphorylated and sulfated, contains polyacidic amino acid segments and an RGD motif, and binds to HA strongly. Whereas BSP is synthesized almost exclusively by cells actively forming mineralizing tissue (Heinegärd, 1988; Oldberg et al., 1988a; Fisher et al., 1990; Bianco et al., 1991; Chen et al., 1991), OPN can be produced by a variety of cell types and is widely distributed in the body (Butler, 1989; Sodek et al., 1992a; Denhardt and Guo, 1993; Young et al., 1993). As the focus of this dissertation, BSP is described in more detail in next section.

A.4. Bone Sialoprotein (BSP): Properties and Expression

BSP comprises 8-12% of the total NCPs in bone and cementum (Fisher et al., 1983; Fisher et al., 1987; Fisher et al., 1990) with substantially lower amounts (~1%) in dentin (Fujisawa et al., 1993). To date, BSP has been purified from human (Fisher et al., 1987), steer (Fisher et al., 1983; Franzén and Heinegärd, 1985), rabbit (Kinne and Fisher, 1987), and pig (Zhang et al., 1990) bones. The primary sequence of BSP was first determined from a rat cDNA (Oldberg et al., 1988a). Subsequently, the cDNAs encoding human (Fisher et al., 1990), porcine (Shapiro et al., 1993), bovine (Chenu et al., 1993), and murine (Young et al., 1994) BSPs were cloned. The genomic locus of the human BSP gene is localized to the long arm of chromosome 4 (Fisher et al., 1990; Kerr et al., 1993), close to genes coding for a number of bone, dentin and enamel matrix proteins, such as OPN, dentin matrix protein (DMP1), dentin specific sialoprotein (DSPP) which codes for dentin sialoprotein (DSP) and dentin phosphoprotein (DPP, phosphoryn) (MacDougall, 1997). The nascent BSP, which has a molecular weight of 33-34 kDa, undergoes extensive post-translational modification including phosphorylation, sulfation and glycosylation, such that the secreted protein has a Mr of 60-80K on SDS-PAGE. The high carbohydrate content of BSP, reported to represent ~50% of the molecular weight, includes 12% sialic acid, 7% glucosamine, and 6% galactosamine (Fisher et al., 1983; 1987).

Immunohistochemical staining has shown that BSP is present throughout the lamellae of human (Bianco et al., 1991; Ingram et al., 1993) and porcine (Chen et al., 1991b) cortical and
trabecular bone and is also found in cement lines and osteoid (Roach, 1994). In addition, BSP has also been specifically localized to the normal and healing rat periodontium (Lekic et al., 1996a, 1996b). BSP is synthesized by differentiated osteoblasts (Chen et al., 1991a; Bianco et al., 1991), cementoblasts and odontoblasts at sites of de novo mineralized tissue formation (Chen et al., 1991a; 1992; 1993). Contrasting osteocalcin, which is expressed at a later stage of bone formation (Chen et al., 1992; Yao et al., 1994), BSP is present in early mature osteoblasts and osteocytes. Similarly, an analysis of cementum formation has shown that the BSP gene is activated with the differentiation of cementoblasts and high levels of expression are associated with the early, rapid formation of the mineralizing cementum matrix (MacNeil et al., 1995; 1996). The expression of BSP is also developmentally regulated. Northern hybridization analysis of the temporal expression of BSP in rat calvariae and tibiae has revealed that the induction of the BSP gene coincides with the initial formation of a mineralized matrix in both membranous and endochondral bones, and that maximal levels of BSP mRNA are attained during embryonic bone formation (Chen et al., 1992). In situ hybridization has further demonstrated that BSP mRNA is highly expressed by newly-formed cuboidal osteoblastic cells actively forming bone matrix on the ectocranial surface of the embryonic rat cranium. Maximal expression occurs in day 21 rat embryos, and declines thereafter, with relatively low levels of expression in adult bone. In contrast, cells lining the endocranial surface of fetal rat calvariae show little expression of BSP (Chen et al., 1991a; 1992). In long bones strong expression of BSP is localized to the primary spongiosa and later (~14 days post-natal) in secondary centres of ossification in the articular cartilage (Chen et al., 1992; Shen et al., 1995). The only non-skeletal sites where BSP has been detected are the trophoblastic cells of the human placenta (Bianco et al., 1991) and platelets, in which BSP is thought to be endocytosed from the serum (Chenu and Delmas, 1992).

Several stretches of "polyglutamic acid" residues in the BSP, together with the sialic acid residues and phosphate groups, likely contribute to the high affinity of BSP for hydroxyapatite (Oldberg et al., 1988a; 1988b; Fisher et al., 1990). Moreover, BSP has been shown to bind to collagen (Fujisawa and Kuboki, 1992; Fujisawa et al., 1995) and to be present in the ECM of bone
immediately prior to mineralization (Nagata et al., 1991; Kasugai et al., 1992). Importantly, BSP has also been localized to loci of early crystal formation in developing bone (Sodek et al., 1992a; Bianco et al., 1993; Chen et al., 1994). Consequently, it has been suggested that BSP could act as a nucleator, inducing crystal formation at the surface of collagen fibers (Nagata et al., 1991; Sodek et al., 1992a). That BSP can nucleate hydroxyapatite crystal formation from steady-state, physiological concentrations of Ca\(^{2+}\) and P\(_i\) has recently been demonstrated (Hunter and Goldberg, 1993) and the polyglutamic acid regions implicated in the nucleation process (Hunter et al., 1996). In contrast, OPN, despite having similar properties to BSP, inhibits hydroxyapatite crystal formation and growth under the same conditions (Hunter et al., 1994; Goldberg et al., 1995).

BSP has an Arg-Gly-Asp (RGD) cell attachment motif near the C-terminus of the molecule and a membrane receptor for the RGD motif of BSP has been identified in ROS 17/2.8 cells as the integrin receptor \(\alpha_v\beta_3\), which is identical to the vitronectin receptor (Oldberg et al., 1988a; 1988b). This receptor has also been identified on the surfaces of normal osteoblasts (Prince et al., 1991) as well as osteoclasts (Miyauchi et al., 1991; Ross et al., 1993). Although the exact function of this RGD sequence in BSP has not been elucidated, its importance is indicated from experiments in which the RGD motif mediates both cell attachment and signalling activities. Thus, BSP can mediate attachment and spreading of fibroblasts on tissue culture dishes (Somerman et al., 1988) and osteoclast attachment on glass surfaces (Grano et al., 1995; Horton et al., 1995; Flores et al., 1996). BSP, and BSP peptides containing the RGD motif, can generate calcium transients in osteoclasts (Paniccia et al., 1993; 1995) and increase their bone-resorptive activity dose-dependently through the \(\alpha_v\beta_3\) integrin (Chenu et al., 1994; Raynal et al., 1996). In recent studies the expression of BSP has been reported in breast adenocarcinomas and its expression was found to correlate with a poor prognosis and with the metastasis of tumour cells to bone (Bellahcène et al., 1996a; 1996b). The bipartite cell and hydroxyapatite binding moieties of BSP are thought to be able to target the metastatic cells to mineralized surfaces of bone.

The expression of the BSP gene is highly modulated by osteotropic hormones and
cytokines. Studies in our laboratory have shown that glucocorticoids (Kasugai et al., 1991; Yao et al., 1994) and BMP-7/OP-1 (Li et al., 1996) induce BSP expression in association with the differentiation of pre-osteoblasts into osteoblasts and the formation of bone in vitro. In addition, glucocorticoids, but not BMP-7 (Li et al., 1996), will further stimulate the expression of BSP in osteoblastic cells (Oldberg et al., 1989, Sodek et al., 1994). However, 1,25-dihydroxyvitamin D3 suppresses osteoblast differentiation (Ishida et al., 1993) and suppresses BSP gene expression (Oldberg et al., 1989) in accordance with the inhibition of bone formation in vitro.

B. Extracellular Signals that Regulate the Differentiation of Osteoblasts

Osteoblast differentiation involves the sequential activation of many genes including those that encode ECM proteins such as BSP, osteopontin, osteocalcin, osteonectin, proteoglycans, and collagen (Rodan and Noda, 1991; Lian and Stein, 1992; Liu et al., 1994). During bone formation and remodelling, a myriad of systemic and local factors, such as vitamin D3, glucocorticoids, retinoid acid, PTH, and TGF-β family members, direct cell differentiation and regulate matrix production (Canalis et al., 1991; Young et al., 1993). However, the molecular mechanisms by which these signals regulate osteoblast differentiation and ECM protein synthesis are still unclear.

B.1. Glucocorticoids

Glucocorticoids are recognized as important regulators of growth and development in the formation of many tissues and organs including bone (Cutroneo et al., 1986). However, the effects of glucocorticoids on bone are complex. While chronic administration of this hormone frequently leads to decreased bone formation (Gennari, 1985) and increased bone resorption (DeFranco et al., 1992) in vivo, studies in vitro have demonstrated that glucocorticoids stimulate bone formation (Maniatopoulos et al., 1988; Bellows et al. 1990). The glucocorticoid-stimulated bone formation in vitro might result from a selective stimulation of proliferation and differentiation of osteoblast precursors (Tenenbaum and Heersche, 1985; McCulloch and Tenenbaum, 1986) and
from an increase of bone matrix formation by osteoblastic cells (Yao et al., 1994). Despite an understanding of glucocorticoid actions in general, the molecular mechanisms involved in the regulation of osteoblastic differentiation have not been elucidated.

The cellular receptor for glucocorticoids belongs to the superfamily of transactivating “zinc-finger” proteins as detailed in section D.2 of this chapter. Upon binding with ligand, the receptor proteins bind to cognate glucocorticoid response elements present within the promoters of responsive genes (Beato, 1989). Interaction of the glucocorticoid receptor (GR) with various response elements on different genes can either promote or suppress transcription directly. In addition, glucocorticoid receptor binding can also relieve gene repression in a “hit and run” mechanism in which the receptor protein de-stabilizes nucleosome structure (Rigaud, 1991). In contrast, genes can be repressed by glucocorticoids, where the GRE overlaps binding sites for activating factors and the GR blocks the activator by steric hindrance (Akerblom et al., 1988; Sakai et al., 1988; Oro et al., 1988; Drouin et al., 1989). Thus, glucocorticoids are known to have dual functions in regulating transcription. Activation and repression of gene expression by glucocorticoids is also seen with bone matrix proteins. For example, while glucocorticoids stimulate BSP (Oldberg et al., 1988a; Kasugai et al., 1991) and OPN (Oldberg et al., 1986; Kasugai et al., 1991) expression, they suppress osteocalcin (Morrison et al., 1989) expression.

B.2. Vitamin D3

This osteotropic hormone is an important regulator of calcium and bone homeostasis and is a potent stimulator of differentiation for cells of the hemopoietic lineage, including osteoclasts. In vitamin D3 deficiency excessive osteoid is formed but mineralization is impaired and bone resorption is increased, resulting in malformation of bones, characteristic of rickets. A recent study has also shown an association between various mutations in vitamin D receptor alleles and osteoporosis (Morrison et al., 1994), but how changes in receptor expression lead to the cellular changes resulting in osteoporosis remains unclear. If chronic treatment of osteoblast cultures is initiated during the proliferation period, cellular proliferation is inhibited and bone nodule
formation is decreased (Ishida et al. 1993). In osteoblasts, vitamin D₃ down-regulates the expression of many bone matrix genes, including collagen type I and BSP, while the expression of osteocalcin, osteopontin, and alkaline phosphatase are up-regulated (Rodan and Noda, 1991). Transcriptional regulation of these genes involves a specific nuclear hormone receptor belonging to the class of transactivating “zinc-finger” proteins described for glucocorticoids (see also section D.2, this chapter). The vitamin D₃-activated receptor binds to its cognate DNA sequence, the vitamin D response element (VDRE), and can regulate gene transcription positively or negatively.

**B.3. Parathyroid Hormone (PTH) and PTH-Related Protein (PTHrP)**

It has long been known that PTH has remarkable effects on the proliferation, development and activities of osteogenic cells, including osteoblasts, osteoclasts and also chondroblasts. Continuous presence of PTH in cultures of rat calvarial cells can suppress bone nodule formation (Bellows et al. 1990). The proliferation and differentiation of osteoblastic progenitors, as well as osteoblastic activities can be decreased by PTH. In contrast, continuous application of PTH in bone marrow cultures stimulates the proliferation and differentiation of osteoclast progenitors and increases the activity of osteoclasts, resulting in extensive bone resorptive activity (reviewed by Heersche et al. 1994). However, the action of PTH on osteoclasts may be modulated indirectly via intercellular signals generated by osteoblasts which have receptors for PTH, whereas osteoclasts do not (Rodan and Martin 1981).

Direct actions of PTH on osteoblasts include an inhibition of type I collagen synthesis (Scott et al., 1988) and ALP activity (Majestka and Rodan, 1982), and stimulation of osteocalcin (Sutherland et al., 1994) and collagenase production (Partridge et al., 1987). It is now well established that cellular responses to PTH are mediated through multiple second messenger signals, such as Ca²⁺, cyclic adenosine monophosphate (cAMP), and inositol-1,4,5-trisphosphate (reviewed by Coleman et al. 1994). However, very little is known about how these signals are transduced to regulate specific bone-related genes. Although most of the effects of PTH appear to suppress bone formation and promote bone resorption, anabolic effects of PTH have also been
documented, which depend on the dose of PTH and stage of differentiation of the target cells (Canalis et al., 1994).

PTHrP is also important in co-ordinating the rate of chondrocyte maturation in the hypertrophic zone of the epiphyseal plate. In mice with an inactivated PTHrP gene, the reduced proliferation and premature differentiation of chondrocytes in the epiphyseal plate leads to a greatly increased rate of endochondral ossification, resulting in a skeleton that is almost completely ossified by the time of birth (Amizuka et al., 1994; Karaplis et al., 1994).

**B.4. Transforming Growth Factor-β (TGF-β) Family**

Members of the TGF-β superfamily play a fundamental role in regulating embryogenesis, development and tissue homeostasis. A subfamily of these cytokines, including TGF-βs, activins, bone morphogenetic proteins (BMPs), and growth and differentiation factors (GDFs), act sequentially in skeletal pattern formation and development (Lyons et al., 1989). TGF-β itself may regulate differentiation of committed osteoprogenitor cells, as suggested by its ability to induce cartilage and bone formation when administered under the periosteum (Joyce et al., 1990), but not when injected into nonskeletal sites (Centrella et al., 1994). Although TGF-β can stimulate chondrocytic differentiation, whether it can modulate osteoblastic differentiation is unclear due to differential responses in various cell systems (reviewed by Centrella et al., 1994). TGF-β suppresses osteoblastic differentiation in cell culture (Antosz et al., 1989) and is a powerful antagonist of the osteo-inductive cytokine BMP-7 (Cheifetz et al., 1996). While TGF-β stimulates collagen, fibronectin, SPARC/osteonectin (Wrana et al., 1988; 1991), OPN, and BSP (Sodek and Overall, 1988) expression, it suppresses the expression of ALP and collagenolytic enzymes. Despite extensive study of TGF-β and its family members, there is a paucity of information on how TGF-βs regulates gene transcription. However, a TGF-β1 activation element (TAE) has been identified in the mouse α2(I) collagen (Rossi et al., 1988) and rat α1(I) collagen (Ritzenthaler et al., 1991) genes. A TGF-β1 inhibitory element (TIE) has also been identified in the rat transin/stromelysin gene (Kerr et al., 1990), and in the rat osteocalcin gene an AP-1/CRE site has
been identified as a target of TGF-β1 suppression (Banerjee et al., 1996).

BMPs are potent osteo-inductive agents. Several BMPs are thought to regulate the early commitment of mesenchymal cells to the osteogenic and chondrogenic lineages, as suggested by their ability to induce ectopic bone and cartilage formation at non-skeletal sites such as muscle (Reddi, 1981; Wozney et al.; 1988; Rosen and Thies, 1992; Kingsley, 1994). They can stimulate osteoblastic differentiation of both uncommitted and committed mesenchymal cells in vitro and they can induce cells other than osteoblasts to express osteoblastic markers (Vukicevic et al., 1989; Yamaguchi et al., 1991; Katagiri et al., 1994). It is possible that different members of the BMP family could operate in different parts of skeleton (Cheryl et al., 1994). However, the molecular basis of bone induction and the precise roles of different BMPs are not known. Studies of BMP effects on bone cells in vitro have shown that ALP expression is stimulated (Vukicevic et al., 1989; Tokuwa et al.; 1990, Hiraki et al., 1991; Zhou et al., 1993), contrasting to the activity of TGF-β1 (Wrana et al., 1988). BMP-3 has also been shown to stimulate collagen and non-collagenous production and to increase the cAMP response to PTH in bone cells (Vukicevic et al., 1990). The expression of BSP is also stimulated by BMPs (Sodek et al., 1994). However, these effects are indirect and are associated with a stimulation of osteogenic differentiation (Li et al., 1996).

B.5. Other Hormones, Growth Factors, and Cytokines

Many other systemic and local factors have also been shown to play various roles during mesenchymal differentiation and skeletal development. Skeletal growth ultimately ceases when the growth plate is entirely replaced with bone, a process that requires estrogen in both male and females (Federman, 1994; Smith et al., 1994). The rate of skeletal growth that occurs until epiphyseal closure is regulated by growth hormone (GH) which acts through the expression of insulin-like growth factor 1 (IGF-1). Reduced GH levels in humans and inactivated IGF-1 in mice results in dwarfism (Baker et al., 1993). That mechanical tension is also important in determining local bone architecture is evident from experiments in which mice with impaired muscle
development, resulting from an inactivation of the myf-5 and myoD genes, have impaired trabecular bone formation (Rudnicki et al., 1993).

Among the important signalling agents in bone development are the retinoids which mediate a broad range of patterning events in embryogenesis, and the vertebrate hedgehog protein which seems to be involved in determining skeletal element identity across the limbs. These signals appear to be instrumental in determining distinct anatomical features (Riddle et al. 1993). The hedgehog protein can induce fibroblast growth factors (FGFs) which provide signals that promote the limb bud elongation and the subsequent organization of bones along the length of the limb ending with fingers or toes (Niswander et al. 1993; Laufer et al., 1994). In addition, hedgehog protein can induce the local expression of BMP2 and hox-13. These inductive interactions may regulate limb skeletal patterning by establishing the expression patterns of homeobox genes in the limb bud mesenchyme, which give cells a positional address (Dolle et al. 1993). It is the interpretation of this positional information that leads to the appropriate development of particular bones. The challenge remains to elucidate how these signals are transduced downstream to regulate cellular activities and the associated gene expression. Bone matrix genes that involve the formation of the skeletal tissue are the ultimate targets for these signals.

C. Expression and Regulation of Eukaryotic Genes

Almost all cells in the human body contain the complete genome, comprising ~150,000 genes that form the blueprint for a human being. The intricate processes that direct a fertilized egg to give rise to a myriad of cell types, which are subjected to different developmental pathways, involve the activation and inactivation of discrete genes in a precise temporo-spatial pattern. The development of skeleton is no exception, although our understanding of the mechanisms that selectively activate genes is at an early stage. However, after more than three decades of exploration, the processes involved in the expression of genes in eukaryotic cells have been defined and some understanding of mechanisms that regulate gene transcription has evolved.
C.1. Structure of Eukaryotic Genes

A typical eukaryotic gene can be divided into several functionally distinct regions: the exons, introns, and the regulatory regions. The exons are the coding regions for mRNA which in turn specifies the sequence of amino acids. Introns are intervening DNA sequences which separate the exons. While introns and exons are transcribed into heteronuclear (hn) RNA, the introns must be spliced out from hnRNA to form the mRNA. Although the exact function of introns are not clear, they may play an important role in regulating gene expression and allow the formation of multiple mRNA transcripts through alternate splicing. The regulatory regions, which direct where and when a particular gene should be expressed, can be classified in at least two categories: a core promoter and upstream (or downstream) regulatory elements. The core promoter resides within a stretch of nucleotides (~100 bp) located immediately upstream of the coding region. A typical core promoter includes an AT-rich region designated as TATA box which has been thought to be position and orientation dependent, and one or more sequence elements of 8 to 12 bp designated as an upstream promoter element (UPE). The upstream (or downstream) regulatory elements, termed enhancers (or silencers), can be located anywhere in the gene and function in an orientation independent manner (reviewed by Maniatis et al., 1987).

C.2. Transcription

Cellular growth, differentiation and behaviour all depend upon the expression of selected genes. Gene transcription is the first step for the generation of a protein. To initiate transcription, RNA polymerase must attach to the core promoter. However, since RNA polymerase is unable to recognize the site of transcription, other proteins must assemble on the DNA into a tight complex, termed the pre-initiation complex (PIC), to guide RNA polymerase to transcribe DNA into mRNA. The PIC constituents have been combined in vitro to yield a fully operational transcription unit (Sawadogo and Sentenac, 1990; Tjian and Maniatis, 1995). The formation of PIC and its role in transcription will be reviewed further in a later section.
Transcription can essentially be classified as one of two types, based on in vitro transcription experiments. Basal level transcription is initiated and maintained by RNA polymerase with the help of PIC, at a low and invariant rate. The rate of basal transcription may be so low that it might not be meaningful in vivo. Activated transcription may be the only one that is functional in vivo (Tjian and Maniatis, 1995). In eukaryotic organisms, three nuclear RNA polymerases are responsible for the synthesis of ribosomal (Pol I), messenger (Pol II), transfer (Pol III), and small nuclear (Pol II or Pol III) RNAs. These three RNA polymerases work at structurally distinct class I, II, and III promoters, respectively, using distinct sets of basic transcription factors.

C.3. Post-transcriptional Modification

Although the factors that regulate the speed and efficiency of processing mRNA and transporting the transcripts through the nuclear membranes to the cytoplasm have not been well characterized, regulation by nuclear post-transcriptional processes is not uncommon. Regulation of the transcript processing in the nucleus may involve: capping at the 5'-terminus of hnRNA, removal of introns by splicing, and generation of 3'-ends coupled with polyadenylation. This is followed by the export of the transcripts, which may involve specific proteins (Krug, 1993). The stability and half-life of mRNA are not only influenced by these procedures of mRNA production, but also by mRNA degradation. Therefore, measurements of the steady-state mRNA content in the cells does not indicate speed or efficiency of transcription.

D. Transcriptional Regulation: the Ultimate Target of Signal Transduction and the Initial Step of Gene Expression

Since transcription is an essential step in protein synthesis, considerable attention has been aimed at understanding, and ultimately manipulating, the biochemical machinery that regulates the transcription process.
D.1. Transcription Factors and *cis*-acting Elements

In eukaryotes, transcription is governed by DNA sequences comprising several functional classes, generally referred to as *cis*-acting elements. These DNA sequences include the core promoter element and enhancers. The core promoter element usually contains a "TATA box" element (consensus TATAAA) at -25 to -30 which serves as a docking site for RNA Pol II, and controls mRNA transcription to be initiated at nucleotide +1, the cap site (Breathnach and Chambon, 1981; Sawadogo and Sentenac, 1990). Many TATA-less promoters contain an initiator element (consensus PyPyA+1NT/APyPy) overlapping the transcription start site (Zawel and Reinberg, 1992). In the absence of a TATA box, the initiator may specify the start site of transcription. The lack of a TATA box or initiator may lead to a weakened promoter. The upstream promoter elements and enhancers play important roles in regulating the rate at which RNA Pol II initiates new rounds of transcription from the core promoter. These *cis*-acting elements direct the action of two classes of transcription factors: general transcription factors, which are essential for initiation and are sufficient to direct a basal level transcription from core promoters, and regulatory factors, which are not required for initiation but are required to mediate the action of upstream promoter elements and enhancers (Maniatis *et al.*, 1987; Johnson and McKnight, 1989). It is such combinations of different *cis*-acting elements and the amounts of their cognate *trans*-acting factors that confer gene responses to a range of signals necessary for cell growth, differentiation and responses to regulatory signals, infectious agents and physiological stress. Perturbations in the normal operation of these assemblies can alter growth and developmental process that eventually lead to various developmental diseases.

Enhancers are defined as *cis*-acting elements that augment promoter activity from a distance in a position- and orientation-independent manner. By analogy, silencers are *cis*-acting elements that repress promoter activity (Brand *et al.*, 1985). Enhancers and silencers can be in the 5' flanking region of a gene or downstream of the protein-coding region, in introns or in exons including the protein-coding regions (Maniatis *et al.*, 1987; Frenkel *et al.*, 1994). Enhancers and
silencers serve as docking sites for transcription factors, called activators and repressors respectively, which carry stimulatory or repressive messages to RNA polymerase to regulate transcriptional activity (Mitchell and Tjian, 1989; Lewin, 1990; Ptashne and Gann, 1990).

Another group of proteins called “co-activators”, which do not bind to DNA directly, serve as adaptors to relay messages from activators to the basal factors in cases where the activators cannot interact directly with the basal transcription machinery (Tjian and Maniatis, 1994).

D.2. Protein Structure and DNA Recognition

In recent years, biochemical and biophysical methods have provided significant insights into understanding DNA-binding properties and the function of transcription factors. Observations from sequence comparisons and X-ray crystallographic studies indicate that many DNA-binding proteins can be grouped into classes which use common structural motifs for DNA sequence recognition. The major families are summarized here to illustrate the general principles of protein-DNA interaction.

*Helix-turn-Helix (HTH) Transcription Factors:* The HTH structure was the first DNA-recognition motif discovered, and structures now have been determined for many HTH proteins and protein-DNA complexes, including the *E. coli* catabolite activator protein (CAP), λ repressor, Lac repressor, and the ETS domain. All the HTH protein structures share a number of features (Pabo and Sauer, 1992; Nelson, 1995). The classical HTH structure consists of a three-helix bundle, with the second and third helices angled ~120°. The third helix, known as the recognition helix, fits into the major groove of DNA close to the edges of the base pairs, while the second helix lies above the major groove contacting the DNA backbone. Each helix has an extensive network of hydrogen bonds between the protein and the DNA. The first helix includes an additional domain that regulates transcriptional activity, allowing dimer formation or cAMP attachment. Although each of these proteins uses the HTH unit in a generally similar way, there are differences in the precise arrangement of the HTH unit relative to the DNA (Harrison and
Homeodomain Proteins: The homeodomain is a DNA-binding motif that is present in a large family of eukaryotic regulatory proteins such as *engrailed, fushi tarazu, oct-1, POU* etc. The homeodomain also contains a HTH motif. However, unlike the isolated HTH unit, the 60-amino acid homeodomain forms a stable, folded structure which can bind DNA (Kinssinger *et al.*, 1990; Pabo and Sauer, 1992). The homeodomain contains an extended N-terminal arm and three α-helices. Helix 1 and helix 2 pack against each other in an antiparallel arrangement, while helix 3 is roughly perpendicular to the first two helices. The main contacts with DNA are made by residues in helix 3, which fits into the major groove making a number of contacts with the DNA backbone, and by residues in the extended N-terminal arm, which fits into the minor groove. Helices 1 and 2 span the major groove, allowing only two contacts with DNA for each helix.

Zinc Finger Proteins: The list of transcription factors in this family is long, ranging from the TFIIA, characterized in *Xenopus*, to many proto-oncogenes. Proteins in this family usually contain tandem repeats of a "zinc finger" motif, a sequence that has a set of cysteines and/or histidines within a short region of polypeptide chain (Rhodes and Klug, 1993). Each zinc finger contains an antiparallel β-sheet and an α-helix. Two cysteines, which are near the turn in the β-sheet region, and two histidines, which are in the α-helix, coordinate a central zinc ion and hold these secondary structures together to form a compact globular domain. Each zinc finger binds in the major groove of DNA and partially wraps around the double helix, making base contacts with a three-base-pair subsite. Neighbouring fingers are aligned in a way that reflects the helical pitch of DNA and the three-base pair periodicity of the subsites (Pavletich and Pabo, 1991).

Steroid Hormone Receptors: Steroid and related hormones, like glucocorticoids, thyroid hormone, estrogen, retinoic acid, and the seco-steroid vitamin D₃, are uptaken by cells and exert their regulatory effects by binding to intracellular cognate receptors. Upon binding to their specific ligands, these hormone receptors influence gene transcription via interaction with specific
DNA elements, the hormone-response elements (HREs), which include the glucocorticoid-response element (GRE), the thyroid hormone-response element (TRE), the estrogen-response element (ERE), the retinoic acid-response element (RARE/RXRE) and the vitamin D-response element (VDRE) (Beato, 1989; Urnesono et al., 1991). Members of this protein receptor family have separate domains for ligand binding, DNA binding and transcriptional regulation, and also belong to a superfamily of "zinc finger" transcription factors as described above (Green and Chambon, 1986; Evans, 1988; O'Malley and Conneely, 1990). However, the sequence pattern seen in the steroid receptors is quite different from those found in the conventional zinc finger, and it is clear that the steroid hormone receptors form a distinct structural motif (Frankel and Pabo, 1988; Hard et al., 1990; Schwabe et al., 1990).

The structure of the DNA-binding domain of the glucocorticoid receptor and its interaction with its cognate receptor DNA sequence has been determined crystallographically (Luisi et al., 1991). The DNA binding region of the steroid receptor is characterized by a pair of "zinc fingers", stabilized by tetrahedral co-ordination of zinc through two sets of four cysteine disulphide bridges that have an amphipathic α-helix at the carboxy-terminus of each domain (Freedman et al., 1988; Schwabe et al., 1990; Luisi et al., 1991). The crystal structure of the glucocorticoid receptor shows that the peptide binds DNA as a dimer (Luisi et al., 1991), even though NMR studies have shown that the peptide exists as a monomer in solution. The two helix fingers of the glucocorticoid receptor are roughly perpendicular each other. The first helix finger of each subunit fits into the major groove, and side chains from this helix make contacts with the edges of the base pairs. The second helix finger provides phosphate contacts with the DNA backbone and provides the dimerization interface.

**Helix-Loop-Helix (HLH) and Leucine Zipper Proteins:** HLH and leucine zipper proteins provide unique examples that illustrate the important role of dimer formation in the regulation of gene expression. For instance, the transcription factor AP-1 consists of a FOS protein and a JUN protein. The FOS/JUN family of proteins generally contain two distinct
subdomains: the leucine-rich zipper region mediating dimerization and a basic region contacting the DNA (Kerppola and Curran, 1991). The zipper sequences are characterized by a heptad repeat of leucines over a region of 30-40 residues which form two parallel α-helices in a coiled-coil arrangement. The basic region is an extension of the HLH motif that is rich in arginine and lysine. Prior to binding to DNA, a pair of the HLH peptides form a dimer, shaped like a Y, with the zipper region as the stem and the basic region as arms which extend along the major groove of the DNA (reviewed by Nelson, 1995).

**β-Sheet Motif Proteins:** This family originally included prokaryotic regulatory proteins MetJ, Arc and Mnt, and later was joined by the high mobility group (HMG) proteins from *Drosophila*, such as HMG-1 and SRY-1. Although MetJ is much larger than Arc, both proteins form dimers in solution and their structures are homologous in a region that contains a β-strand and two α-helices. In the protein dimers, the β-strands pair to form an antiparallel β-sheet, while the α-helical regions pack against the sheet and against each other to stabilize the dimer (Pabo and Sauer, 1992). MetJ binds as tetramer to an 18-base-pair DNA site. In this complex, a dimer binds to each half-site, and each half-site contains a two-fold symmetry axis that is complementary to the two-fold axis of the β-sheet. The β-sheet fills the major groove, and side chains on the exposed face of the sheet contact the base pairs (Somers and Philips, 1992). Given the limited sequence identity among members of this family, identification of members of this group based upon sequence searches is difficult and more structures are needed to fully appreciate the boundaries of this motif (Pabo and Sauer, 1992; Nelson, 1995).

**Other Classes:** While many DNA-binding proteins can be described in families, some important factors have unique features. For example, the structure of the tumor suppressor gene, p53, does not contain any standard DNA-binding motif (Cho et al., 1994). DNA binding is mediated through contacts in the major groove by a helix and loop of a loop-sheet-helix motif, as well as by a contact in the minor groove from an arginine positioned by another loop. This unusual type of DNA-binding surface is susceptible to single amino acid substitutions which can
disrupt the function of p53. The TATA-binding protein (TBP), the universal transcription factor, provides another example of a unique protein-DNA interaction as revealed by x-ray crystallography (Nikolov, 1992). TBP, which has intramolecular symmetry, straddles the minor groove like a saddle, causing DNA to change configuration drastically and promote transcription. In addition, the upper surface of the molecular saddle allows multiple interactions with other regulatory proteins to control gene transcription (detailed in section D.4. of this chapter).

D.3. General Transcription Factors (GTFs)

Initiation of gene transcription by RNA polymerase II requires a number of GTFs to form a specific multiprotein complex, the Pre-initiation Complex (PIC), which interacts with the core promoter near the transcriptional start site. During the past decade, considerable effort has been devoted to understand how RNA Pol II recognizes and initiates transcription and how this process is regulated. A number of GTFs have been identified and many of the genes have been cloned (Sawadogo and Sentenac, 1990; Conaway and Conaway, 1993; Buratowski, 1994; Zawel and Reinberg, 1995). The polypeptide components of RNA Pol II and the GTFs are generally well conserved in structure and function among eukaryotes from yeast to human, allowing lower eukaryotes to be used as powerful genetic and biochemical models to investigate the complicated structure-function relationships. The GTFs that guide RNA polymerase II to bind to the core promoter region and to initiate transcription are called basal factors. At least seven of those basal factors have been characterized: transcription factors II A, IIB, IID, IIE, IIF, IIH, and IIJ (reviewed by Conaway and Conaway, 1993). As a result, it is possible now to reconstitute the basal transcription machinery with well-defined components in vitro.

TFIIA is composed of three subunits of 37 kDa (α), 19 kDa (β) and 13 kDa (γ). The precise function of TFIIA in basal transcription is still controversial (Conaway and Conaway, 1993), although it has long been known that TFIIA can stabilize the binding of TFIID to the core promoter (Davison et al., 1983; Buratowski et al., 1989). Nevertheless, TFIIA can bind to DNA
and significantly enhance basal transcription and also mediate transcriptional activation by upstream activators, probably by binding to both TFIID and the activator (Maldonado and Reinberg, 1995).

TFIIB plays at least two pivotal roles in basal transcription although it does not bind to DNA itself (Maldonado and Reinberg, 1995). First, it serves as a bridge between promoter-bound TFIID and RNA Pol II. Second, the interaction between TFIIB and RNA Pol II seems to be critical in determining the transcription start site, a notion supported by genetic experiments in which mutations in the yeast gene encoding TFIIB (SUA7) shifted the transcription initiation site of several yeast genes studied in vivo (Pinto et al., 1992). Biochemical evidence for this interaction and its importance in determining the start site was reported later (Li et al., 1994).

Among the GTFs, TFIID is the first and key protein complex that binds to the TATA box in the core promoter and facilitates assembly of the complete basal transcriptional machinery (Buratowski et al., 1989; Greenblatt, 1991; Pugh and Tjian, 1992). Native TFIID is a multiunit protein complex comprising the TATA-binding protein (TBP) and several TBP-Associated Factor (TAFs) that form an elongated molecule of ~750 kDa (Dynlacht et al., 1991). The fact that TBP can bind some activators or repressors suggested TBP might be able to modulate upstream regulatory signals directly (Stringer et al., 1991). However, although TBP is sufficient for TATA element recognition and subsequent basal transcription, the activation of basal transcription requires native TFIID, which consists of TAFs in addition to TBP (Tjian and Maniatis, 1994). Since TBP is small, it is unlikely to provide binding sites for all activators and repressors. In fact, many activators and repressors do not bind to TBP directly. TAFs, therefore, are believed to convey molecular signals from the multitude of activators and repressors to the basal transcriptional machinery, at the same time providing species- and cell-specific regulation (Dynlacht et al., 1991; Brou et al., 1993; Chen et al., 1994). Certain TAFs may also mediate interaction between TFIID and other PIC components (Choy and Green, 1993). To date, eight TAFs from Drosophila TFIID have been purified, their cDNAs cloned and recombinant proteins produced that can be reconstituted for in vitro transcription (Tjian and Maniatis, 1994). However, the pattern of protein-
protein interactions and the function of most TAFs have yet to be determined.

TFIIF is apparently a heterodimer of RAP30/RAP70, initially identified by Greenblatt and colleagues (refs in Conaway and Conaway, 1993). Although this factor does not appear to bind to DNA, it acts in concert with TFIIB to promote selective binding of RNA Pol II to the PIC. In addition, TFIIF has been shown to prevent non-selective binding of RNA Pol II to DNA. All these functions of TFIIF in eukaryotes are likely mechanisms that are utilized by σ factors in bacteria (Killeen and Greenblatt, 1992).

The perceived roles of the multi-unit TFIIF are complicated and include promoter clearance during transcription, nucleotide excision repair (NER), and perhaps cell cycle progression (Seroz et al., 1995, and references therein). In fact, TFIIF is the only GTF known to have enzymatic activity, including an ATPase, a helicase and a kinase, specific for the carboxy-terminal domain (CTD) of the largest subunit of RNA Pol II (the CTD-kinase), which are all required during the entire initiation and elongation of the transcription process. Moreover, NER, one of the main pathways for repair of damaged DNA, can be greatly facilitated by transcription. Dysfunction of TFIIF could result in a large number of genetic disorders, such as xeroderma pigmentosum, Cockayne's syndrome and trichothiodystrophy (Broughton et al., 1994). The spectrum of these diseases includes unrelated symptoms (such as photosensitivity, dental caries, brittle hairs, ichthyosis and neurodemyelination) which cannot be accounted for entirely by deficiencies in the NER process, but rather may reflect the multiple functionality of TFIIF, in which a subtle defect in one of its subunits will result in a series of reactions that direct transcription and/or some steps of the cell cycle (Seroz et al., 1995).

D.4. TATA-box Binding Protein: a Universal Transcription Factor

The TATA-box Binding Protein (TBP) is the only subunit in the native TFIID complex that truly interacts with the TATA element in the core promoter. Complementary DNAs encoding TBP have been cloned from yeast, insects, plants, and mammals, ranging in size from 22 kDa
(Arabidopsis) to 38 kDa (Drosophila and human) (reviewed in Conaway and Conaway, 1993). Comparison of the deduced amino acid sequences of TBPs reveals that they are composed of a highly conserved C-terminal domain and an N-terminal domain that varies in length and sequence (Greenblatt, 1991). In fact, TBP is the most highly conserved eukaryotic transcription factor, with its 180-amino-acid C-terminal showing greater than 80% sequence identity in a wide variety of species from plants to primates. Included within the C-terminal domain are two ~60-amino acid direct repeats interrupted by a lysine-rich spacer region of ~40 amino acids (Greenblatt, 1991; Conaway and Conaway, 1993). This C-terminal part is believed to be the functional domain because its presence is crucial for TATA box recognition and basal transcription in vitro (Horikoshi et al., 1990; Reddy and Hahn, 1991; Yamamoto et al., 1992), and in vivo (Cormack et al., 1991; Gill and Tjian, 1991; Reddy and Hahn, 1991).

Despite its small size, TBP carries out an impressive array of functions in transcription. First, TBP alone binds with high affinity to the TATA box (Conaway and Conaway, 1993, and references therein). Second, TBP interacts with associated factors (TAFs) to form distinct multiprotein complexes, SL1, TFIID, and TFIIIB that, respectively, are specific for transcription by RNA Pol I, Pol II, and Pol III (reviewed by Cormack and Struhl, 1993; Reinberg, 1995). Third, for most RNA Pol II promoters, specific binding of TBP to the TATA element initiates the assembly of an active transcription complex (reviewed in Sawadogo and Sentenac, 1990). In the course of this assembly process, promoter-bound TBP interacts with TFIIA and TFIIIB. Fourth, TBP can interact in vitro with some transcription activators and repressors (Stringer et al., 1990), indicating that it may be a relevant target of regulatory proteins in vivo. Finally, TBP is a subunit of the SNAP complex, which binds specifically to the promoter of small nuclear protein gene, allowing RNA Pol I and Pol II to transcribe from this promoter (Sadowski et al., 1993).

TBP differs from many other sequence-specific DNA-binding proteins in several respects (Conaway and Conaway, 1993 and references therein). First, TBP binds to and dissociates from TATA elements (see next section) extremely slowly, and is heat-sensitive. Second, whereas many
DNA-binding proteins bind as dimers to the major groove, TBP binds as a monomer to the minor groove using its bipartite repeat. Third, although TBP is most active in transcription at promoters containing the consensus TATA box (TATAAA), it can also bind to a variety of non-consensus TATA elements (Hahn et al., 1989, Singer et al., 1990), and even some TATA-less promoters at a location around -30 bp (Wiley et al., 1992), but with lower affinity. Lastly, TBP cannot be divided into distinct DNA-binding and transcriptional activation domains, since deletion of amino acids anywhere within the entire 180-amino-acid C-terminal region abolishes both promoter binding and transcription activities.

As mentioned earlier, TBP is an intramolecular dimer that binds DNA as a saddle which consists of a curved 10-stranded, antiparallel β-sheet, with four α-helices lying on its upper surface. Structural, biochemical, and mutational analysis have indicated that the concave underside of the saddle binds to the minor groove of DNA, whereas the α-helices and the convex surface of the saddle are likely to bind to other regulatory proteins. The co-crystal structure of the TBP-TATA complex reveals a dramatic and unprecedented distortion in the DNA helix (Y. Kim et al., 1993; J.L. Kim et al., 1993). Binding causes a sharp (90°) kink at the end of the TATA element, severe unwinding and a compensating superhelical twist, and strong bending of DNA toward the major groove. The incoming and outgoing double helices of the DNA are sharply angled (100°) and markedly displaced (by 18 Å) with the shallow and very wide minor groove of the TATA element interacting with most of the entire undersurface of the TBP saddle. This unique conformational change in DNA induced by TBP binding is important for transcription initiation. First, the remarkable distortion at the TATA element could be propagated downstream, causing further unwinding and perhaps strand separation around the transcription initiation site. Second, DNA bending might bring other trans-acting factors into closer proximity or promote the correct stereochemistry required for transcription. Finally, TBP bends the TATA element in a direction opposite to that preferred in nucleosomal DNA, suggesting a simple mechanism for the observed competition between histones and TBP, an important rate-determining step in transcription (Klug, 1993).
D.5. The TATA Box: a Central Role in Transcriptional Control

The TATA box was initially discovered in prokaryotic cells as the “Pribnow box”. Its importance in eukaryotes was then revealed by comparison of several eukaryotic genes in a search for eukaryotic counterparts of this AT-rich region (reviewed by Breathnach and Chambon, 1981). Centred about 25-30 bp upstream from the transcription start site, the TATA box has a consensus form 5'-TATAAA-3', which has been derived from a number of eukaryotic genes. Although many variants have been discovered, the crucial positions in the TATA box are highly conserved. The TATA box has the following fundamental characteristics involved in controlling gene transcription:

1). The TATA box is the only signalling DNA sequence which is recognized by TBP. It is, therefore, a central DNA element that directs the binding of TFIID to the promoter of eukaryotic genes, which is required for the subsequent formation of a pre-initiation complex on the core promoter, as illustrated in Fig. 1.1 (Buratowski et al., 1989; Pugh and Tjian, 1992).

2). Because its distance from the transcription initiation site is relatively constant, the TATA box is believed to determine the start site of transcription. This hypothesis was initially based on the observations that, while TATA-containing genes have fewer and more definite start sites, TATA-less genes generated more scattered initiation sites of transcription (Dynan 1986; Maniatis et al., 1987). Recent studies have indicated that the constant distance between the TATA box and transcription start site may be bridged by TFIIB, which can interact with TFIID bound to the TATA box at one end and RNA Pol II bound to the initiation site at the other end, thereby accurately setting the transcription start site (Pinto et al., 1992; Li et al., 1994).

3). Since the TATA is the central site for TFIID binding, and TFIID is the key link between upstream activator/repressors and basal transcription apparatus, the TATA box can be regarded as a convergent point that mediates the activities of upstream activator elements (see Fig. 1.1; Maniatis et al., 1987; Sawadago and Sentenac, 1990; Pugh and Tjian, 1992).
Figure 1.1 - Initiation of Transcription and Activation of the Basal Transcription Machinery. (a). General transcription factors, TFIIA, B, D, G, E, F, assemble at a RNA polymerase II promoter to form a pre-initiation complex (PIC) and initiate basal transcription. Transcription activators bind to specific cis-acting elements and then target the basal transcription machinery to activate transcription. (b). A schematic illustration of a TFIID complex. Only the TBP subunit in the complex can bind to a TATA box and is required for basal transcription. Both TBP and TAFs, and sometimes co-activators, are required for activator-responsive transcription. (Figure adapted from Pugh and Tjian, 1992).
4). Heterogeneous TATA boxes may function as a selective element to determine tissue-specific gene expression because they can recruit distinctive TFIID complexes to the core promoter (Wefald et al., 1990; Pugh and Tjian, 1992; Brou et al., 1993). At least a few TAFs have been demonstrated to modulate cell-type specific transcription (reviewed in Tjian and Maniatis, 1994).

5). The TATA box is thought to be capable of orientating TBP and other PIC factors thereby defining the direction of transcription because of the inherently asymmetrical nature of most TATA sequences (5'-TATAAA-3'), together with the asymmetrical nature of the TBP (Greenblatt, 1991, 1992; Ready and Hahn, 1991; Philips; 1993; Klug, 1993). Moreover, unlike enhancers and silencers which are orientation- and position-independent, the core promoter and the TATA box encompassed within can apparently only function in one orientation (Dynan, 1986; Maniatis et al., 1987). For example, it has been shown that conversion of a consensus TATA box into an inverted sequence (5'-TTTATA-3') is associated with the loss of down-stream gene transcription activity (Nagawa and Fink, 1985; Ruden et al., 1988). However, some experiments in which the TATA sequence was inverted artificially have implied that the orientation of a TATA box might not determine the direction of transcription, although the transcription activity was significantly decreased in that case (O'Shea-Greenfield and Smale, 1992; Xu et al., 1992). Consequently, the importance of the TATA box orientation in determination of the TBP/TATA polarity and the direction of transcription is still ambiguous.

D.6. Assembly of the Preinitiation Complex (PIC)

The initiation stage of transcription is a major phase for regulation of gene expression. Although substantial evidence indicates that the GTFs are required for transcription of most, if not all, core promoters, most of this evidence is based on studies of their roles in initiation from simple, TATA box-containing promoters. Two models of PIC formation on TATA-containing promoters have been described: an ordered multistep model, and a holoenzyme assembly model. In the multistep model, the GTFs and RNA Pol II assemble on the promoter DNA through protein-
DNA and protein-protein interactions in a highly ordered fashion (reviewed in Zawel and Reinberg, 1995). The first step is the binding of TFIID to the TATA element, followed by the binding of TFIIB, which results in the formation of the TFIID/B complex intermediate. This serves as a bridge for the delivery of RNA Pol II by TFIIF, which is followed by the sequential association of TFIIE, TFIIH, and TFIJ. This model of PIC assembly was derived from studies using purified components from yeast to human cells (Buratowski et al., 1989). The holoenzyme assembly model is derived from the study of yeast promoters (Koleske and Young, 1994). In the yeast Saccharomyces cerevisiae, a fraction of cellular RNA Pol II is tightly associated with a few other GTFs like TFIIB, TFIIF, TFIIH etc., forming the holoenzyme complex. This complex is able to accurately initiate transcription when supplemented with TBP and TFIIE. Thus, the pre-assembled holoenzyme can be recruited to promoters in one step after the binding of TBP to the TATA box. However, this has not been documented in higher eukaryotes. Consequently, the multistep model is favoured by most investigators for eukaryotic transcription.

D.7. Transcription Activation and Tissue-Specific Gene Expression

Earlier analysis of eukaryotic gene regulation has generally focused on positive cis-acting elements and trans-acting factors, and a number of positive transcriptional components have been characterized (Ptashne, 1988; Dynan, 1989). Two general models for the regulation of genes expressed in multiple cell types can be considered (Maniatis et al., 1987). In one model, expression of a single gene is controlled by distinct trans-acting factors, the tissue-specific factors are present in different tissues. In an alternative model, common trans-acting factors present in all cell types and such factors interact with their respective enhancers in an identical manner in each cell type. However, although some genes may contain identical enhancers, no two genes share precisely the same elements in combination with adjacent sequences. This unique arrangement may explain why cells are able to control transcription of every gene individually.

Although many of the components involved in positive regulation have been identified, it is still not clear exactly how the regulatory signals are transmitted to the transcription initiation
complex. The current consensus, based on indirect evidence, is that transcription activators can directly interact with the basal factors and loop out the intervening DNA (see Fig. 1.1; Ptashne, 1988; Müller et al. 1989). In addition to TFIID, other GTFs like TFIIB and TFIIH are also potential targets for upstream transcription activators (Tjian and Maniatis, 1994). Trans-acting factors, therefore, usually contain two functional domains: a DNA binding domain which binds to the enhancer, and an activation domain which interacts with basal transcription apparatus via an acidic-, glutamine-, or proline-rich region. In many cases, intermediary factors (termed co-activators) are required to connect activators to the initiation complex (Ptashne and Gann, 1990; Lewin, 1990) and TAFs, as discussed earlier, are well-documented as example of co-activators (reviewed by Tjian and Maniatis, 1994). DNA structure distorted by PIC formation and activator binding, therefore, enables RNA polymerase to begin its journey along the activated gene. The multiple, co-operative interactions among the transcription factors could increase the rate of PIC formation and/or stabilize an otherwise unproductive complex. For instance, the amino- and carboxy-terminal domains of TFIIIB are folded together by an intermolecular interaction. The binding of VP16 to TFIIIB disrupts this molecular interaction, exposing the binding site for TFIIF, TBP, and RNA Pol II, favouring the PIC formation (Roberts and Green, 1994). It was shown recently that TFIIH interacts directly with activator VP16 and the tumour suppressor p53 (Xiao et al., 1994), which may facilitate promoter clearance by RNA Pol II (Goodrich and Tjian, 1994).

D.8. Transcription Repression and Tissue-Specific Gene Expression

How repressors convey negative regulatory signals is less well understood even though silencing specific genes during growth and development is as important as activating of other genes. A good example of gene silencing is provided by the tumour-suppressor gene product p53, whose function is to repress a number of genes that have potential to promote abnormal growth of cells. Mutations in p53 protein render the loss of its suppressive function on many target genes, leading to tumour formation (reviewed in Hansen and Oren, 1997).

Negative regulation of gene expression may involve multiple mechanisms. Repression of
gene transcription can be achieved through a passive mechanism in which negative transcription factors compete with transcriptional activators to bind to DNA. For example, the human factor NF-E binds to the CCAAT box of the fetal γ-globin gene and inhibits binding of a positive factor CP1 thereby preventing expression of the gene in the adult (Superti-Furga et al., 1988). Intriguingly, positive factors can also compete with activators to repress transcription. For instance, during *Drosophila* development, induction of the segment polarity gene *engrailed* (*en*) by homeodomain gene products of *zen-related* (*z2*), *fushi tarazu* (*ftz*) or *paired* (*prd*) is inhibited by the products of the *en* or *eve* (*even-skipped*) genes, which all recognize similar or identical DNA sequences (Levine and Manley, 1989; Ohkuma et al., 1990a; Müller and Bienz, 1992). In addition, gene repression can also be achieved via an active mechanism in which the negative transcription factors "silence" gene transcription at non-overlapping sites (Biggin and Tjian, 1989; Jaynes and O'Farrell, 1991). In the development of T lymphocytes, a silencer within the first intron of the CD4 genes has been identified (Sawada et al., 1994). This silencer is required to selectively extinguish CD4 gene expression in CD4-CD8+ T cells, but not in CD4+CD8- nor CD4+CD8+ T cells, allowing the maturation of the CD4-CD8+ cytotoxic cells. However, in some cases, DNA binding is not necessarily required to block the effect of an activator. Thus, the co-activator which transmits a positive signal from a DNA-bound activator to the transcription initiation complex is another target for repression. Repressors may also bind to co-activators and inhibit transcription by preventing activators from coupling to their usual attachment sites on co-activators (Renkawitz, 1990). As exemplified in yeast, the repressor GAL80, which cannot bind to DNA, binds the activator GAL4 and blocks its *trans*-activating ability (Ma and Ptashne, 1988). Other factors, such as *polycomb* (*Pc*) group products, provide long-range gene silencing by acting at distance from the transcriptional initiation site to alter chromatin structure (Paro, 1990; Bienz, 1992).
D.9. Signal Transduction and Transcriptional Regulation

Transcription of a gene is modulated by a number of extracellular signals such as hormones, growth factors, cytokines, infectious agents and stress. To activate or repress transcription, transcription factors must be located in the nucleus, bind DNA, and interact with the basal transcription apparatus. There are two general signal transduction pathways: one that utilizes cytoplasmic or nuclear receptors (intracellular receptors) which interact directly with DNA upon activation by ligands to regulate transcription of target genes, and a second that utilizes membrane-bound receptors which are activated at the cell surface by their ligands and transmit their signals downstream via second messengers. The first pathway is typified by the signalling via steroid and related hormones as depicted in Fig.1.2. The second pathway generally involves reversible phosphorylations, triggered by growth factors, cytokines, and peptide hormones, that ultimately target transcriptional factors (Hunter and Karin, 1992; Hill and Treisman, 1995).

The activation of transcription factors by phosphorylation can occur at the cell membrane, in the cytoplasm, or in the nucleus. An example of the activation occurring at the cell membrane is the JAK/STAT pathway, which is involved in the transcriptional activation of many cytokine- and growth factor-inducible genes (Fu and Zhang, 1993; Ihle et al., 1994). It is known that the STATs are in the cytoplasm prior to receptor activation, and are transiently associated with their cognate transmembrane receptors. As ligand binds to the receptor tyrosine kinase, the STAT protein is phosphorylated, dimerizes, and subsequently leaves the receptor and migrates into the nucleus (Darnell et al., 1994). STAT activation by cytokine receptors lacking intrinsic kinase activity involves specific members of the JAK (Janus kinase) family that are associated with the cytokine receptors. Activation in the cytoplasm is illustrated by the transcription factor NF-κB, an important factor mediating the immune response in B cells. NF-κB is activated by many agents, such as IL-1, IL-2, TNFα, LPS, and viral infection, which induce acute responses. NF-κB is retained in the cytoplasm by binding to its inhibition factor IκB. Phosphorylation of IκB by activated kinase leads the loss of its ability to bind to NF-κB. Dissociation of NF-κB from IκB
leads the translocation of NF-κB into the nucleus and binding to DNA element (Beg and Baldwin, 1993).

The MAP kinase (MAPK) pathways provide a good example of the activation of transcription factors in the nucleus (Marshall, 1994). The MAPKs are a family of protein kinases including ERKs. Activation of receptor tyrosine kinase activates a signalling cascade involving transient formation of ras-GTP and activation of raf kinase at the membrane, followed by sequential activation of ERK1/ERK2 which enter the nucleus. The ras/ERK kinase pathway provides a common route by which signals from different growth factor receptors converge at a major regulatory element, the serum response element (SRE), which is present in the promoters of many serum-regulated genes such as c-fos. The major phosphorylation targets in the pathway are the TCF proteins, Elk-1 or SAP-1 that are crucial members of the complex bound to the SRE (Gille et al., 1992; Treisman, 1994). Studies of the mammalian JUN transcription factor has led to the identification of the JNK/SAPKs (JUN N-terminal/stress-activated protein kinases) as a family of MAPKs distinct from the ERKs (Marshall, 1994; Hill and Treisman, 1995). Transcriptional activation by JUN requires phosphorylation at S/T-P motifs and can be potentiated by the RAS and SRC oncoproteins (Binetruy et al., 1991). While the two S/T-P motifs that are phosphorylated in JUN are poor substrates for the ERKs, they are efficiently phosphorylated by the JNKs/SAPKs. Factors that strongly stimulate JUN phosphorylation, such as ultraviolet light, strongly activate the JNKs/SAPKs, but activate the ERKs only weakly.
Figure 1.2 - Regulation of Gene Transcription by Steroid Hormones, Using Vitamin D₃ as a Model Example. (1). Vitamin D₃ uptake by the cell; (2). Binding of the vitamin D₃ ligand to its receptor (VDR) in the nucleus; (3). Binding of VDR to the vitamin D₃-response element (VDRE); (4). Gene transcription and mRNA processing; (5). Protein synthesis and secretion.
Chapter II

Rationale, Objectives, and Hypotheses
Elucidating the molecular events associated with cellular differentiation is fundamental to understanding the potential of precursor cells to form a specific connective tissue. While a great deal of information is available on the molecular events involved in the differentiation of many cell types such as haematopoietic and myogenic cells and adipocytes, our knowledge of osteoblast differentiation is more limited. Also, despite considerable progress that has been made towards an understanding of the general mechanisms involved in gene transcription, little of this knowledge has been applied to bone-related genes. A major difficulty in this regard has been the paucity of cellular markers and the lack of well-characterized, homogeneous populations of cells that reproducibly differentiate into osteoblasts. Since the expression of the BSP and osteocalcin is highly restricted to mineralized tissues, these genes are especially valuable as phenotypic markers in osteogenic differentiation. Indeed osteocalcin has been studied extensively in this regard. In comparison with osteocalcin, BSP has received little attention. However, based on the temporospatial expression of BSP, which correlates with the initiation of mineralization in bone, and its unique physico-chemical properties indicating that it may act as a nucleator of hydroxyapatite crystal formation, considerable interest has been generated in this protein.

Since transcriptional control of tissue-specific genes is an integral part of the regulated development of the osteoblastic phenotype, an understanding of the transcriptional regulation of the BSP gene can provide important insights into osteoblastic differentiation and the regulation of tissue mineralization. My initial objective therefore was to clone the rat BSP gene promoter and to identify components of the promoter that could be involved in the initiation and further regulation of BSP transcription in a tissue-specific and developmental manner. Following the isolation of the rat BSP gene promoter, an inverted TATA sequence overlapped by a VDRE-like sequence were identified in the proximal promoter. I thus hypothesized that:
1) The immediate promoter region could direct tissue specific expression of the BSP gene.

2) The inverted TATA box could be functional and that it would be capable of directing downstream transcription.

3) The VDRE-like sequence may be involved in mediating the suppression of BSP gene transcription by vitamin D3.

To test these hypotheses, studies were aimed at examining the functionality of the proximal promoter with respect to tissue specificity and determining the ability of the TATA/VDRE elements to initiate and modulate BSP gene transcription. These studies are described in Chapters III-V.
Chapter III

Cloning of the Rat BSP Gene Promoter*


The nucleotide sequence of the rat BSP promoter in this report is available in the GenBank Nucleotide Sequence Database under accession no. L06562. All the experiments in this chapter were performed by J.J. Li except Fig. 3.4, in which the extended nucleotide sequence upstream of nt. -801 was determined by Y. Ogata. The human BSP promoter was cloned subsequently by R.H. Kim et al. (1994) and the nucleotide sequence was obtained in the GenBank Nucleotide Sequence Database under accession no. L06823.
SUMMARY

To study the transcriptional regulation of the rat bone sialoprotein (BSP) gene, the nucleotide sequence of a ~1kb HindIII/KpnI sub-fragment from a genomic clone containing the 5' flanking sequence, exon 1, and part of intron 1 was determined and the transcription start site defined. This region includes a unique inverted TATA element (nt -24 to -19), an inverted CCAAT box, a homeobox-binding site, a putative vitamin D₃ response element (VDRE) sequence overlapping the inverted TATA sequence and a novel 18-nt palindrome that may control the tissue-specific transcription of the BSP gene. Comparison of the rat BSP promoter with a later cloned human BSP promoter (Kim et al., 1994) has revealed that 71% of the nucleotides have been conserved. Higher similarity between the rat and human BSP promoters is demonstrated in the proximal promoter region up to -370 bp, which includes the inverted TATA and CCAAT boxes, NFkB, CRE, AP-1 and the homeobox-binding sites. Consequently, this highly conserved proximal promoter region has been termed the "BSP" box. The shortest promoter sequence capable of directing bacterial chloramphenicol acetyltransferase (CAT) reporter gene expression included the inverted TATA element and the inverted CCAAT box. However, the promoter activity was down-regulated by vitamin D₃, indicating that the unique VDRE-like sequence overlapping the TATA element is functional. Thus, the rat BSP gene promoter is characterized by novel cis-acting elements that may be involved in hormone and tissue specific regulation of transcription. This study represents the first cloning and characterization of a functional BSP gene promoter.
INTRODUCTION

Bone sialoprotein (BSP) is a highly sulphated and phosphorylated glycoprotein present almost exclusively in the mineralized connective tissue matrices of bone, dentine and cementum (Fisher et al., 1987, 1990; Zhang et al., 1990; Sodek et al., 1992a). The cDNA of rat BSP has been cloned and found to encode a 33,600-Da secreted protein of 320 residues (Olberg et al., 1988). The protein is characterized by its ability to bind strongly to hydroxyapatite, presumably through 2-3 stretches of polyglutamic acid, and to mediate cell attachment through an RGD motif (Oldberg et al., 1988; Somerman et al., 1988; Helfrich et al., 1992). In rat bone, BSP is specifically expressed by fully-differentiated osteoblasts with especially high levels of expression observed at sites of de novo bone formation (Chen et al., 1991a; 1991b; Sodek et al., 1992a). Moreover, the expression of the BSP gene is developmentally regulated (Chen et al., 1991). The expression of the BSP gene by adult rat bone marrow cells in vitro is induced by the synthetic glucocorticoid dexamethasone in association with bone formation (Kasugai et al., 1991). In contrast, BSP gene expression is down-regulated by the osteotropic hormone 1,25-vitamin D₃ (Oldberg et al., 1989), which promotes bone remodelling. Consequently, BSP is believed to function in the initial formation of the mineralized bone matrix (Sodek et al., 1992b).

Since the elucidation of the molecular mechanisms that regulate the tissue specific expression of BSP gene can provide important insights into the understanding of bone and tooth development, as well as metabolic and congenital diseases of mineralized tissues, I have isolated a genomic clone that includes the promoter region of the rat BSP gene. Characterization of the promoter has revealed novel regulatory elements that may control the tissue-specific and developmental expression of this protein. This study describes the first characterization of a BSP gene promoter.
MATERIALS AND METHODS

Genomic Library Screening

cDNA probes used for genomic library screening were generated by polymerase chain reaction (PCR), using a rat BSP cDNA clone as template. Specific regions of rat BSP cDNA were amplified, with EcoRI tails added to the primers, cloned into pT7T3 18U, and radiolabelled with [a-32P]dCTP by oligolabelling (T7QuickPrime™ Kit, Pharmacia), as showing in Fig. 3.1. A rat genomic library prepared from rat testis DNA and assembled into λ phage DASH II vector (Stratagene) was screened by plaque hybridization to the full-length cDNA probe, including the 5'-UTR and entire BSP coding sequence (Oldberg et al., 1988), according to the supplier's (Stratagene) instructions. Positive clones were further selected using cDNA probes derived from a 5'-fragment and a 3'-fragment (nucleotides -70 to +56 and +829 to +960 relative to the ATG start codon, respectively) (Fig. 3.1 and Table 3.1).

Genomic Mapping and DNA Sequencing

From a clone (λ4A) encoding the 5' region of the rat BSP gene a ~5 kb XbaI fragment was digested with various restriction enzymes, and the sub-fragments separated by agarose gel, and directly hybridized with the 21-mer [32P]-end-labelled oligonucleotide (RBSP5.1) (“In-Gel Hybridization”) (Fig. 3.1 to 3.3) to provide a genomic map (Ausubel et al., 1989). Relevant DNA fragments were excised with the appropriate restriction enzymes, and subcloned into pT7T3 18U. Sequencing was performed on both DNA strands using the Sanger dideoxy chain termination technique (T7 Sequencing Kit™, Pharmacia).

Comparison of the Rat and Human BSP Gene Promoters

The nucleotide sequences were analyzed in the Transcription Factor Database using the University of Wisconsin Genetics Computer Group (GCG) Program. The nucleotide sequences of
1. Designation of Primer:  
   - **EcoRI**
   - RBSP5.1 (Sense): 5'-ggccggaattcTGGAGAACAATCCGCACT-3'  
   - RBSP3.1 (Antisense): 5'-ggccggaattcATCGAGAAAGCACTCGCCAT-3'  
   - RBSP5.2 (Sense): 5'-ggccggaattcGCGTATGACGAG-3'  
   - RBSP3.2 (Antisense): 5'-ggccggaattcCTGGCATAGGATCTATCGCCT-3'  

2. PCR using BSP cDNA as template and paired primers indicated in the maps.
3. Digest the PCR products with EcoRI and clone into phagemid pT7T3 18U.
4. Nucleotide sequencing to confirm identity of PCR products.
5. Label with [α-³²P]dCTP by random-primer

**Figure 3.1 - cDNA Probes Generated by PCR for Screening a Rat Genomic DNA Library.** A rat genomic library was screened by plaque hybridization with the full-length cDNA probe, including the 5'-UTR and entire BSP coding sequence. Positive clones were further selected using cDNA probes derived from the 5'- and 3'-regions of the rat BSP cDNA (nucleotides -70 to +56 and +829 to +960 relative to the ATG start codon, respectively). Only the positive clones which hybridized with the 5'-probe, but not the 3'-probe, were subjected for further analysis (see Table 3.1).
Table 3.1  Positive Clones from the Rat Genomic Library Screening

<table>
<thead>
<tr>
<th>ID</th>
<th>Full-length Probe</th>
<th>5' - Probe</th>
<th>3' - Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ2A</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>λ4A</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>λ10A</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>λ12A</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>λ14A</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>λ20A</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>λ21A</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 3.1 - Positive Clones from the Rat Genomic Library Screening. Seven positive clones were obtained from the screening of the rat genomic library with the full-length rat BSP cDNA probe. Only the positive clones which hybridized with the 5'-probe, but not the 3'-probe, λ2A and λ4A, were subjected to further analysis.
the rat and human BSP gene promoters were then compared for homology and alignment by the BESTFIT option of the University of Wisconsin Genetics Computer Group (GCG) program.

**Primer Extension Analysis**

A 20-mer $^{32}$P-end-labelled oligonucleotide (50,000 dpm per reaction) complementary to nucleotides 37 to 56 downstream of the ATG start codon (RBSP3.1) (Fig. 3.1; Oldberg, 1988) was hybridized overnight to 10 µg of total fetal rat bone RNA at 30°C. Nascent cDNA was synthesized by AMV reverse transcriptase according to Frederickson et al. (1989) and the $^{32}$P-labelled cDNA generated was analyzed on an 8% polyacrylamide-8 M urea gel.

**CAT Plasmid Construction**

The transient expression vector pCAT-Basic (Promega) containing the bacterial chloramphenicol acetyltransferase (CAT) reporter gene, but lacking a promoter and enhancers, was used to test for BSP gene promoter activity. Using the isolated rat BSP genomic fragment as template, 5' -deletion mutants were generated by PCR with HindIII or SalI tails added to the primers, and cloned into the HindIII/SalI sites of the pCAT-Basic vector. The identities of the deletion mutants were confirmed by nucleotide sequencing on both DNA strands.

**DNA Transfection and CAT Analysis**

The osteoblast cell lines, ROS17/2.8 and RCJ3.1, were grown in α-minimal essential medium containing 10% (v/v) fetal bovine serum and seeded 24 h prior to transfection at a density of 4 x 10⁵ cells per 100-mm dish. Transfections were performed by the calcium phosphate precipitation technique (Sambrook et al., 1986; Ausubel et al., 1989), using 10 µg of BSP/CAT fusion plasmid DNA for each dish. Normalization of CAT enzyme activity for transfection efficiency was determined by co-transfection with 2 µg of pCH110 (Pharmacia) containing the $lacZ$ gene driven by the SV40 promoter. Cells were harvested at ~60 h after transfection, and cell extracts were prepared by repeated freeze-thawing. Some transfected cells were treated with 100
nM 1,25-vitamin D₃ for 60 h after transfection to test the responsiveness of the BSP gene promoter. Either liquid scintillation counting (LSC) or thin layer chromatography (TLC) were employed to determine the CAT activities (Promega's instructions).

RESULTS AND DISCUSSION

To characterize the cis-acting sequences involved in regulation of the transcription of the rat BSP gene, a rat genomic library was screened for clones hybridizing with a full-length rat BSP probe. Seven independent λ clones were obtained from the initial screen of 1 x 10⁶ phage plaques (Table 3.1). One clone (λ4A) which hybridized to a 5'-end cDNA probe, but not to a 3'-end probe, was digested with Xba I. A ~5 kb Xba I fragment from λ4A, recognized by the 21-mer oligonucleotide (RBSP5.1), was subcloned into the Xba I site of the pT7T3 18U vector (clone pGRBSPX31) and a restriction map of the insert determined using the following enzymes: Hind III, HincII, KpnI, SacI, Sal I and XbaI (Fig.3.2). Since this insert was found to include ~3 kb of 5' flanking region, a ~1 kb Hind III/KpnI fragment from pGRBSPX31 that hybridized to oligonucleotide RBSP5.1 was subcloned for further analysis (clone pGRBSHK7). The nucleotide sequence of the insert of pGRBSHK7 was determined for both DNA strands (Fig. 3.3) and numbered relative to the transcription start site (see below). The nucleotide sequence of the insert of pGRBSHK7 contained a 63 bp region which is identical to the 5'-UTR of the rat BSP cDNA except for one G-T mismatch (+19 from transcription start site) (Oldberg et al., 1988), and includes a consensus splice donor site (AG:GTAAGGT), defining the 3' boundary of exon 1. Further upstream, the nucleotide sequence was subsequently determined up to nt -2992 to include the entire 5'-flanking region of the rat BSP promoter (Fig. 3.4).

An examination of the 5'-flanking sequences of rat BSP gene revealed a number of sequences resembling consensus eukaryotic cis-acting elements (Fig. 3.3 to 3.5): an inverted TATA element (TTTATA) at -24 to -19, an inverted CCAAT box (ATTGG) at -50 to -46, a cAMP response element (CRE) (TGACGTCTCG) at -75 to -66, a putative NFκB binding site
(CGGATTTTCT) at -102 to -93, a putative AP-1 binding site (TTATTCA) at -143 to -137, a homeobox-binding site (TCAATAAAT) at -194 to -185, and a putative AP-2 site (GGCAGCCC) at -447 to -440. A sequence (AGGGTTatAGGTCA) overlapping the inverted TATA element that resembles a vitamin D₃ response element (VDRE) and a novel 18-nucleotide perfect palindrome was located at -453 to -436 of the rat BSP promoter. A glucocorticoid receptor element (GRE) was located ~1 kb upstream from the transcription start sites of both rat and human BSP promoters and is overlapped with a AP-1 site in the rat promoter and with a RARE-like sequence in the human promoter.

The human BSP promoter was also cloned and was demonstrated to be functional in directing CAT reporter gene expression in osteoblastic cells (Kim et al., 1994). A comparison of the promoter and the first exon of the rat BSP gene with the corresponding region of the human BSP gene revealed a high degree (71%) of nucleotide identity (Fig. 3.5). The similarity is even higher (75%) in the region proximal to the transcription start site (up to nt -370). Notably, the inverted TATA element, together with the CCAAT box, NFκB site, CRE, homeobox-binding site, and AP-1 site, are conserved in the rat and human BSP gene promoters. Consequently, this 370 bp proximal promoter region has been termed the “BSP Box” (Fig. 3.5).

The transcription start site of the rat BSP gene was determined initially by primer extension mapping. Total fetal rat calvarial RNA, shown previously to express an abundant amount of BSP transcripts (Chen et al., 1991a; 1991b), was used for primer extension analysis with total fetal rat liver RNA and yeast tRNA as controls. Precise localization of initiation sites of BSP transcripts were determined by comparison of the primer extension products with an adjacent Sanger sequencing ladder. Three start sites were found (Fig. 3.6), the major start site being designated in the nucleotide sequence as +1 (Fig. 3.3). The position of the major transcription start site was subsequently confirmed by RNase protection mapping.
Figure 3.2 - Restriction Enzyme Map of the 5'-Flanking Region of the Rat BSP Gene. A ~5 kb XbaI fragment from genomic clone λ4A that includes ~3 kb of 5'-flanking region was subcloned and designated as pGRBPX31. The position of the first exon is represented as an open box with restriction sites as indicated. The transcription start site, as determined in Fig. 3.6, is indicated as +1. The position of hybridization probe RBSP5.1 used for in-gel Southern blotting is also illustrated.
Figure 3.3 - Partial Nucleotide Sequence of the Promoter Region of the Rat BSP Gene. A -1 kb Hind III/Kpn I fragment, as illustrated in Fig. 3.2, was subcloned and sequenced. The major transcription start site (↓), determined as shown in Fig. 3.6, is numerically designated as +1. Minor start sites are also presented (↑). The splice donor (S.D.) site is indicated, and the partial 5'-UTR sequence of rat BSP cDNA (Oldberg, et al., 1988) is shown in italics for comparison. The site of probe RBSP5.1, used for in-gel Southern blotting (Fig. 3.1), is indicated by a broken underline. Consensus sequences for various cis-acting elements discussed in the text are also depicted. The positions of the 5' ends of the chimeric gene constructs used in transient transfection studies are indicated by the arrowheads, whereas the 3' end is indicated at +60.
Figure 3.4 - Extended Sequence of the Rat BSP Gene Promoter. The nucleotide sequence of the upstream promoter region of the rat BSP gene is shown from nts +259 to -2992. Sequences conforming to GRE-half-sites (underlined) and AP-1 (broken line box) sites, as indicated, were identified from a transcription factor binding site data base search. Also shown are the regions of the promoter included in the chimeric (pLUC) constructs in which luciferase was used as a reporter gene in the transient transfection assays (adapted from Ogata et al., 1995).
Figure 3.5 - Sequence Comparison of the 5'-Flanking Regions of the Rat (R) and Human (H) BSP Genes. The highly conserved “BSP Box” comprising the proximal promoter region up to nt -370 includes the inverted TATA box overlapped by a VDRE, an inverted CCAAT box, and putative binding sites for cAMP (CRE), NFkB, JUN/FOS (AP-1), and Ftz/E (Hox) homeobox proteins. A GRE is located -1 kb upstream from the transcription start site and is overlapped with a AP-1 site in the rat BSP promoter and with a RARE-like sequence in the human BSP promoter. The rat promoter also has a unique 18 nt inverted repeat (IR).
Figure 3.6 - Determination of the Transcription Start Site of the Rat BSP Promoter. Primer-extended products obtained using RBSP3.1 as a primer were co-electrophoresed with Sanger dideoxy sequencing reaction products to determine the precise start sites. The longest and major product was 154 bp and, therefore, designated as +1 as shown in Fig. 3.3. Minor start sites were evident at 151 bp and 150 bp. Fetal rat calvarial RNA (lane 1), was used as a template for the BSP transcripts. Negative controls are fetal rat liver RNA (lane 2) and yeast tRNA (lane 3).
The close proximity (within 5 base pairs) of the major and minor transcription start sites (Fig. 3.6) supports the existence of a TATA box that positions the RNA polymerase II for accurate transcription initiation (Dynan, 1986; Maniatis et al., 1987). Notably, a TATA-like sequence was located at nts -24 to -19 in the rat BSP gene promoter. However, this putative TATA box is in a reversed orientation on the complementary strand (inverted). The TATA box is believed to determine the transcription start site, direct the formation of a pre-initiation complex, and mediate the activity of upstream activator elements (Breathnach and Chambon, 1981; Maniatis et al., 1987; Sawadago and Sentenac, 1990; Pugh and Tjian; 1992). In addition, the TATA box is thought to define the direction of transcription because of the inherently asymmetrical nature of most TATA sequences (Greenblatt, 1991). However, that the TATA box determines the direction of transcription is controversial. By using synthetic promoters and an in vitro transcription system O'Shea-Greenfield and Smale (1992) have shown that both normal and inverted TATA sequences can drive transcription from similar initiation sites and in the same direction, although transcription from an inverted TATA box is reduced by 7-fold. In addition, a functional inverted TATA box located 15 nucleotides downstream of the transcription start site has been described in the adenovirus IVa2 promoter (Carcamo et al., 1990). Should the putative TATA element in the rat BSP promoter prove to be functional in directing downstream transcription, it would support the contention that the orientation of transcription is not dependent upon the orientation of TATA box, and also would be the first perfectly inverted TATA element located at the position of a TATA box in a eukaryotic gene. Moreover, since the spatial and temporal expression of the BSP gene is highly restricted, and heterogeneous TATA boxes may function as a selective element to determine tissue-specific gene expression by interaction with distinctive TFIID complexes (Wefald et al., 1990; Pugh and Tjian, 1992), it was of interest to investigate the significance of the inverted TATA sequence in the BSP gene.

To test promoter activity, I initially used an osteoblastic cell line, UMR106-01, that expresses BSP gene constitutively (Midura et al., 1990) for transient transfection. However, the
transfection efficiency for UMR106-01 cells was insufficient to test for promoter activity. Consequently, five BSP/CAT fusion constructs were transiently transfected into another osteoblastic cell line, ROS17/2.8, and examined for CAT enzyme activity (Fig. 3.7). While variable promoter activities have been demonstrated using different lengths of the upstream regions, the shortest sequence (-116 to +60) capable of directing high level (10.6 times relative to the vector control) CAT expression included the inverted TATA element, the inverted CCAAT box, and the CRE. Similar results were found when the same constructs were transfected into a non-transformed rat bone cell line, RCJ3.1. The existence of a CCAAT box can ensure optimum promoter activity and the CCAAT box is operative in either orientation (Maniatis et al., 1987; Sawadag and Sentenac, 1990). Moreover, the CCAAT element is also believed to be a main target for regulation of specific genes, as exemplified by the expression of multiple CCAAT-transcription factors (CTF) in a tissue-specific and development-specific pattern (Santoro et al., 1988), while the CCAAT-enhancer binding protein (C/EBP) can act as a differentiation switch for execution of a specialized phenotype (Umek et al., 1991). Notably, a 10 bp consensus homeobox-binding site (TCAATTTAAAT) that is recognized by the engrailed (en) and the fushi tarazu (ftz) proteins and a unique 18-nucleotide palindrome which could also be involved in cell specific expression of BSP are found within this region. In this regard homeobox proteins are thought to be key regulators of embryonic development (Levine and Hoey, 1988) and the same homeobox-binding site is conserved in the human BSP gene promoter at the corresponding position. The sequence upstream of the CCAAT box appears to contain a negative regulatory element(s), since inclusion of the DNA sequences upstream of -116 in the construct effectively suppressed transcription in ROS17/2.8 cells (Fig. 3.7). Inspection of this region revealed a sequence (CACCTCT at nts -408 to -402) that closely resembles the suppressing element (CACCTCC) found in the rat collagen II gene (Savagner et al., 1990). Of interest, similar suppression sequences are also present in the promoter of the porcine osteopontin gene (Zhang et al., 1992).

To demonstrate promoter activity and to identify sites involved in cell-specific expression of BSP gene, chimeric constructs encompassing different regions of the rat BSP promoter up to nt
-2992, designated pLUC3 to pLUC13 (Fig. 3.4), were ligated to a luciferase reporter gene and transient transfection assays using both osteoblastic cells (ROS 17/2.8) and non-bone cells (NRK 49F). When the transcription activities of these constructs were compared, no marked differences were evident between BSP-producing (ROS 17/2.8) and non-producing (NRK 49F) cells (Sodek et al., 1994). However, since only a small amount (3%) of the transfected DNA is actively transcribed in the transiently transfection system, and it is recognized that the transfected DNA may not be regulated in the same manner as the endogenous gene (Sippel et al., 1992), studies on tissue-specific expression need to be extended to other systems such as the transgenic mouse model.

A vitamin D₃ response element (VDRE)-like sequence was found overlapping the putative inverted TATA box, located at nts -28 to -14. This sequence (AGGGTTtatAGGTCA) is very similar to the consensus VDRE (AGGTCAnnnAGGTCA) (Umesono et al., 1991). Notably, in the promoter of the human osteocalcin gene the glucocorticoid receptor binds to a sequence overlapping the TATA box and provides a mechanism for suppressing osteocalcin transcription by glucocorticoids (Strömstedt et al., 1991). To study whether the VDRE-like sequence in the rat BSP gene promoter is involved in the down-regulation of BSP gene transcription by vitamin D₃, transiently transfected ROS17/2.8 cells were treated with 100 nM vitamin D₃. Compared with vehicle control treatment (ethanol), vitamin D₃ reduced the promoter activity by 28.4% (Fig. 3.8). These findings indicate that the binding of the vitamin D₃ receptor protein may block transcription by interfering with TFIID binding to the inverted TATA element and that upstream sequences may be required either to stabilize the binding of the vitamin D receptor protein, or to provide sites for co-regulator(s) to bind.
Figure 3.7 - Promoter Activity Analysis of the BSP-CAT Chimeric Constructs. Five fragments of the rat BSP gene 5'-flanking region were generated by PCR using specific oligonucleotide primers with Hind III or Sal I tails (shown in Fig. 3.3), cloned in the 5' to 3' orientation between the Hind III and Sal I sites of pCAT-Basic plasmid, and designated pCAT1 to pCAT5. The identities of the promoter fragments were verified by nucleotide sequencing. The bar values represent means ± standard deviations from seven independent transfection experiments, each performed with replicates as measured by CAT assays and normalized to β-galactosidase activity. pCAT-Control (pCATC), which has both the promoter and enhancer of SV40, was used as a positive control, whereas pCAT-Basic (pCATB), which lacks a promoter and enhancer, was used as negative control. The relative levels of CAT expression to pCATB from each construct are illustrated on the right.
In summary, in this study I have isolated a functional rat BSP gene promoter characterized by a putative inverted TATA box and an inverted CCAAT box, with a VDRE-like sequence overlapping the inverted TATA element. The existence of an inverted TATA element suggests that the orientation of a TATA box may not determine the direction of transcription, while the unique VDRE-like element indicates a novel mechanism through which vitamin D₃ can suppress gene transcription. Upstream elements including a conserved homeobox-binding site and a novel 18 nucleotide palindrome may be involved in the cell-specific expression of the BSP gene. The proximal promoter region (up to nt -370) is highly conserved between the rat and human BSP promoters, and has been termed the "BSP Box".
Figure 3.8 - Down-regulation of the Rat BSP Gene Promoter by 1,25(OH)2-vitamin D3.
ROS17/2.8 cells were transfected with pCAT3 (as in Fig. 3.3) and cultured in the presence (lane D+) or absence
(lane D-) of 1,25(OH)2-vitamin D3. pCAT-Basic (lane B) transfection was used as a negative control, and
pCAT-Control (lane C) transfection as a positive control. The data shown represents a typical result of TLC
CAT assays obtained from three independent transfection experiments, each performed with replicates.
Acetylated and unacetylated forms of [14C]-chloramphenicol were separated on silica plates and exposed to X-ray
film. The pCAT3 promoter activity is suppressed by 28.4% ± 1.4% (mean ± SE) as measured by liquid
scintillation. Neither 1,25(OH)2-vitamin D3 nor ethanol vehicle effected expression of the control vectors (not
shown).
Chapter IV

An Inverted TATA Box Directs Downstream Transcription of the Bone Sialoprotein (BSP) Gene*

*Part of this chapter was published: J.J. Li, R.H. Kim, and J. Sodek. (1995). Biochem. J. 289:310:33-40. The original article in the publication was written by J.J. Li. All the experiments were performed by J.J. Li except those shown in Figs. 4.6 and 4.7., which were contributed by R.H. Kim.
SUMMARY

The orientation of the TATA box is thought to direct downstream transcription of eukaryotic genes by RNA polymerase II. However, the putative TATA box in the promoter of the Bone Sialoprotein (BSP) gene, which codes for a tissue-specific and developmentally regulated bone matrix protein, is inverted (5'-TTTATA-3') relative to the consensus TATA box sequence (5'-TATAAA-3') and is overlapped by a vitamin D₃-response element. Here I show that the inverted TATA sequence in the BSP gene binds to recombinant TATA-Binding Protein (TBP) with an affinity similar to that observed with the consensus TATA box, while site-directed point mutations in the inverted TATA sequence (mutating TTTATA into TCTCTA) abrogate both TBP binding and BSP promoter activity. However, when the inverted TATA sequence is changed to a canonical TATAAA, the TBP binding properties and the BSP promoter activity are retained. In addition, it was found that TBP is required to reconstitute in vitro transcription driven by the BSP promoter. These studies, which have revealed a naturally occurring inverted TATA box that can bind TBP and direct downstream transcription, demonstrate that the orientation of the TATA box does not determine the direction of transcription in higher eukaryotic genes. Consequently, the inverted TATA box that is conserved in the human, rat, and mouse BSP gene promoters provides an excellent in vivo model to investigate the polarity of the transcription factor IID-DNA complex and its relation to downstream transcription.
INTRODUCTION

The transcription of eukaryotic genes is highly regulated by interactions between a variety of *trans*-acting factors and their cognate *cis*-acting elements. The binding of the TFIID complex to the TATA box is crucial for the initiation of transcription and for controlling gene expression (reviewed in Sawadogo and Sentenac, 1990; Greenblatt, 1991). Even in genes lacking a recognizable TATA box, binding of TFIID to a region approximately -30 bp upstream from the transcription start site is required to initiate transcription by RNA polymerase II (Pugh and Tjian, 1992 and references therein; Wiley et al., 1992). Included in the TFIID complex is the universal transcription factor TBP, which specifically recognizes the TATA box, and the TBP-Associated Factors (TAFs) (Dynlacht et al., 1991; Pugh and Tjian, 1992; Takada et al. 1992). Thus, the TATA box provides a specific site that directs TFIID, RNA polymerase II, and other auxiliary factors to assemble the pre-initiation complex (PIC) which promotes the synthesis of mRNA at the cap site, and mediates the activity of upstream activator elements (Breathnach and Chambon, 1981; Maniatis et al., 1987; Buratowski et al., 1989; Stringer et al., 1990; Ptashne and Gann, 1990; Brou et al., 1993).

Since most TATA sequences found so far are asymmetrical (5'-TATAAA-3' as a canonical form) (Breathnach and Chambon, 1981; Hahn et al., 1989; Reddy and Hahn, 1991), the TATA box is thought to be capable of orientating TBP and other PIC factors therefore defining the direction of transcription (Greenblatt, 1991, 1992; Reddy and Hahn, 1991; Phillips, 1993). In support of this contention, it has been shown that conversion of a consensus TATA box into an inverted sequence (5'-TTTATA-3') is associated with a loss of down-stream gene transcription activity (Nagawa and Fink, 1985; Ruden et al. 1988). However, some experiments in which the TATA sequence was inverted artificially implied that the orientation of a TATA box might not determine the direction of transcription (O'Shea-Greenfield and Smale, 1992; Xu et al., 1992). Consequently, the importance of the TATA box orientation in the determination of the TBP/TATA
polarity and the direction of transcription is still ambiguous.

Bone Sialoprotein (BSP) has been characterized as a major non-collagenous protein mainly found in the extracellular matrices of bone, cementum and dentine whose expression is developmentally regulated (Sodek et al., 1992a, 1992b and references therein). In addition to its specific expression by fully-differentiated osteoblast at sites of de novo bone formation, BSP is able to nucleate hydroxyapatite crystal formation from steady-state, physiological concentrations of calcium and inorganic phosphate (Hunter and Goldberg, 1993), supporting the contention that BSP functions in the initial formation of the mineralized bone tissue (Sodek et al., 1992b). To study the transcriptional regulation of BSP gene expression, I cloned the rat BSP gene promoter (Chapter III; Li and Sodek, 1993). Intriguingly, a perfect inverted TATA sequence (5'-TITATA-3') is located at nts -24 to -19 bp upstream of the transcription initiation site in the rat BSP gene promoter. Subsequent studies have shown that this inverted TATA sequence is conserved in the human and mouse BSP gene promoters (Kerr et al., 1993; Kim et al., 1994; Gupta and Aubin, unpublished work), implying its importance in controlling BSP gene expression. Here, I show that this inverted TATA element is a functional TATA box that is required for directing downstream expression of this eukaryotic gene both in vitro and in vivo. This study provides the first characterization of a naturally occurring inverted TATA box in higher eukaryotic genes and demonstrates that the specific orientation of the TATA box does not determine the direction of transcription.
MATERIALS AND METHODS

DNase Footprinting

To prepare DNA probes for DNase I footprinting experiments, plasmid pCAT3 (Chapter III) was digested with either HindIII or SalI. Either the HindIII end (for the coding strand) or the SalI end (for the non-coding strand) was first labeled with [γ-32P]-ATP by T4 polynucleotide kinase. The end-labeled 176 bp DNA fragments (position -116 to +60 of the rat BSP promoter, see Fig. 4.1) were then digested with secondary restriction enzymes, SalI or HindIII, and purified by PAGE. A specific radioactivity of ~2 x 10^7 cpm/μg was usually obtained. About 5 fmol (~0.5 ng) of [32P]-labelled DNA probe was used in each footprinting reaction. Footprinting experiments were performed as described (Horikoshi et al., 1989), with slight modification, in a reaction mixture (50 ml) containing 12 mM Tris-HCl (pH 7.9), 40 mM HEPES (pH 8.4), 60 mM KCl, 12% (vol/vol) glycerol, and 8 mM MgCl₂. After incubating end-labeled DNA probes with various amounts of proteins at 25°C for 45 min, DNase I digestions were done at room temperature (22°C) for 2 min. Products were analyzed on a 10% urea-polyacrylamide sequencing gel. TBP-DNA binding affinity was determined according to Ausubel et al. (1992). The radiograph density over footprint sites was digitized and quantified by two-dimensional densitometry of autoradiograms using an Macintosh Apple One Scanner system. Dissociation constants (kd) were calculated from data calibrated to internal standards and expressed using the Langmuir isotherm.

Site-directed Mutagenesis

Site-directed point mutations were generated by polymerase chain reactions (PCR) according to Landt et al. (1990) with modifications. Briefly, two separate reactions were performed in the first round amplification. One reaction contained 1.0 ng of plasmid pCAT3 mixed with a 5' flanking primer, antisense mutant primer, plus the necessary PCR components.
Figure 4.1 - Footprint of the TATA-binding Protein (TBP) on the Rat BSP Promoter Sequence. Nucleotide sequences upstream to nt -116 of the immediate promoter region of the BSP gene including the inverted CCAAT and TATA boxes are shown. A 20 bp region (bracketed) encompassing the inverted TATA box (rectangled) is protected by human TBP in the DNase I footprinting experiments (Fig. 4.3). The putative VDRE overlapping the inverted TATA box is shadowed.
Figure 4.2 - Nucleotide Sequencing of Point Mutations in the Rat BSP TATA Box. Site-directed mutagenesis by PCR was used to generate mutated forms of the inverted TATA box in constructs of the rat BSP gene proximal promoter that were linked to a bacterial CAT reporter gene. (a). The wild type TATA box sequence in pCAT3 (see Figs. 3.3 and 4.1). (b). The inverted TATA box with point mutations, changing 5'-TCTATA-3' into 5'-TCTCTTA-3' (p3ΔT-CAT). (c). Changing 5'-TTTTATA-3' into 5'-TATATAA-3' to produce a canonical TATA box (p3NT-CAT).
Another reaction contained identical components except for a 3'-flanking primer and the sense mutant primer. After 20 cycles of amplification, one mutant fragment with the mutation at the 3' end and another mutant fragment with the mutation at the homologous 5' end were generated. Those two PCR products were then separated on a low melting point agarose gel, sliced, combined, and melted. A small volume (10 µl of ~200 µl) was taken to perform the second round PCR with the 5' flanking primer and the 3' flanking primer used in the first round PCR. The final PCR fragment containing the desired point mutation was subcloned into pCAT-Basic vector. The designated mutations and fidelity of PCR were confirmed by dideoxy sequencing (Fig. 4.2).

**Chloramphenicol Acetyltransferase (CAT) Plasmid Construction and the CAT Assays**

Transient transfections and CAT assays were conducted as described previously (Chapter III). The transient expression vector pCAT-Basic (Promega) containing the bacterial CAT reporter gene, but lacking a promoter and enhancers, was used to test for BSP gene promoter activity. The wild type and mutated rat BSP promoters (nts -116 to +60) were cloned into the HindIII/SalI sites of the pCAT-Basic vector. The osteoblastic cell line, ROS17/2.8, was grown in α-minimal essential medium containing 10% (v/v) fetal bovine serum and seeded 24 h prior to transfection at a density of 4 x 10^5 cells per 100-mm dish. Transfections were performed by the calcium phosphate precipitation technique (Ausubel et al., 1992), using 10 µg of BSP/CAT fusion plasmid DNA for each dish. Normalization of CAT enzyme activity for transfection efficiency was determined by co-transfection with 3 µg of pSV-β-Galactosidase Control vector (Promega) containing the lacZ gene driven by the SV40 promoter. Cells were harvested at 48 h after transfection, and cell extracts were prepared by repeated freeze-thawing. Liquid scintillation countings (LSC) was employed to determine the CAT activities (Promega's instructions).

**In Vitro Transcription**

The templates for the *in vitro* transcription reaction were prepared by PCR based on wild-
type or mutated BSP/CAT fusion plasmids described above. One 5'-primer was derived from the rat BSP promoter (nts -116 to -107) and two 3'-primers were derived from sequences within the CAT reporter gene to produce templates which could generate transcripts of 171 bp or 311 bp in size. The reaction was carried out according to the protocol described previously (Farnham and Schimke, 1986; Innis et al., 1991) with slight modifications. Briefly, each reaction was performed in 25 µl of 8.8 mM HEPES, pH 7.9 at 25 °C, containing 8.8% glycerol, 44 mM KCl, 0.088 mM EDTA, 0.22 mM DTT, 10 mM MgCl₂, and 200 ng (8 mg/ml) of DNA template. The reaction was started by mixing the HeLa cell nuclear extract (Promega) and DNA and incubating for 15 min at 30°C. Transcription was initiated by adding 200 mM each of ATP, CTP, and UTP, and 16 mM of GTP and 10 µCi of [α-32P] GTP (3000 Ci/mmol) and incubated at 30°C for a further 45 min. The reaction was terminated by adding 175 µl of 0.3 M Tris-HCl, pH 7.5 at 25°C, 0.3 M sodium acetate, 0.5% SDS, 2 mM EDTA, and 3 mg/ml tRNA. RNA was extracted with phenol/chloroform and precipitated with ethanol. The pellet was washed once with 70% ethanol and dissolved in 10 µl of loading dye and each reaction analyzed on a 5% urea-polyacrylamide denaturing gel. In the reconstitution assay, the HeLa cell nuclear extract was heated at 45°C for 15 min to inactivate the endogenous TBP (Nakajima et al., 1988). To this heat-treated nuclear extract either BSA, or varying amounts of TBP, were added and the in vitro transcription reactions were carried out as described above.

**Results**

**TATA-Binding Protein (TBP) Binds to the Inverted TATA Box**

Since the binding of TBP to the TATA box is a preliminary step in the formation of the PIC complex that is required for basal transcription of a class II gene promoter, I used DNase I footpointing to assess the ability of TBP to bind to the inverted TATA element. Recombinant human TBP was demonstrated to bind to a ~20 base pair region covering the inverted TATA element on both DNA strands (Fig. 4.3A and 4.3B), similar to its interaction with consensus TATA box sequences (Nakajima et al., 1988; Hahn et al., 1989; Hoey et al., 1990; Lee et al.,
Moreover, unlike many other non-consensus TATA elements that interact with TBP with lower affinity, the binding affinity between TBP and the inverted TATA box ($kd = 2.0 \times 10^{-9} \text{ M}$) was as high as observed with the consensus TATA box ($kd = -2 \times 10^{-9} \text{ M}$) (Hahn et al., 1989).

**Disruption of Binding of TBP to the Inverted TATA Element by Site-directed Mutagenesis**

The inverted TATA elements in the rat and human BSP gene promoters were located at the position corresponding to a TATA box in polymerase II-transcribed genes (about -30). To investigate if this inverted TATA element was critical for TBP binding, I created point mutations in the inverted TATA element, either changing 5'-TTTATA-3' into 5'-CTCTTA-3' (p3AT-CAT, mutant) to inactivate potential TBP binding, or changing 5'-TTTATA-3' into 5'-TATAAA-3' (p3NT-CAT, "normal" TATA box) to produce a canonical TATA box (Fig. 4.2). Two site-directed point mutations in the inverted TATA sequence (mutating 5'-TTTATA-3' into 5'-CTCTTA-3') abrogated TBP binding (Fig. 4.4). However, conversion of the inverted TATA sequence into a canonical TATA box (5'-TATAAA-3') did not affect TBP binding ($kd = 2.2 \times 10^{-9} \text{ M}$) to the template significantly (Fig. 4.4). Therefore, the inverted TATA element is crucial for the TBP-DNA interaction in the BSP gene.

**Requirement of the Inverted TATA Box in Directing Downstream Expression of the BSP Gene**

In a previous study I demonstrated basal promoter activity in the rat BSP gene, measured by bacterial chloramphenicol acetyltransferase (CAT) reporter gene expression, within a sequence nts -116 to +60, which encompasses the inverted TATA element and an inverted CCAAT box (shown in Fig. 3.7, Chapter III; sequences in Fig. 4.1). To determine the importance of the inverted TATA sequence in downstream transcription, the chimeric constructs with the mutated
Figure 4.3 - DNase Footprinting Experiments Demonstrating the Binding of Recombinant Human TBP (hTBP) to the Inverted TATA Element in the Rat BSP Gene Promoter.

(a). Footprint of hTBP on the coding strand. (b). Footprint of hTBP on the non-coding strand. DNA coordinates including the position of the inverted TATA box are included in a and b. The nucleotide position was deduced from G+A sequencing reactions (lane 1 in a and b). BSA was used as a control (lane 2 in a and b). Labelled DNA probes were incubated with increasing amounts of hTBP. Amounts of hTBP used in lane 3 to 7 in a: 1/16 FPU (=1.25 ng), 1/8 FPU (=2.5 ng), 1/4 FPU (=5 ng), 1/2 FPU (=10 ng), 1 FPU (=20 ng). Amounts of hTBP used in lane 3 to 9 in b: 1/64 FPU (=0.125 ng), 1/32 FPU (=0.625 ng), 1/16 FPU (=1.25 ng), 1/8 FPU (=2.5 ng), 1/4 FPU (=5 ng), 1/2 FPU (=10 ng), 1 FPU (=20 ng).
Figure 4.4 - Effect of Point Mutations on the Binding of hTBP to the Inverted TATA Box. Recombinant hTBP is shown to bind to the wild type inverted TATA sequence (TTTATA in pCAT3) (lanes 2 to 7), as well as to the canonical TATA sequence (T\DeltaATA\AA\ in p3NT-CAT), which was created by mutagenesis (lanes 14 to 19), with similar binding affinity. Two point mutations in the inverted TATA element (T\underline{C}T\underline{T}TA in p3\DeltaT-CAT) abrogated the TBP-DNA interaction (lanes 8 to 13). The nucleotide position was deduced from G+A sequencing reactions (lanes 1). Labelled DNA probes were incubated with BSA controls (lanes 2,8,14) and increasing amounts of hTBP: 1/16 FPU (=1.25 ng) (lanes 3,9,15); 1/8 FPU (=2.5 ng) (lanes 4,10,16); 1/4 FPU (=5 ng) (lanes 5,11,17); 1/2 FPU (=10 ng) (lanes 6,12,18); 1 FPU (=20 ng) (lanes 7,13,19).
TATA element described above were transiently transfected into the osteoblastic cell line ROS17/2.8 and promoter activity measured by CAT assays. In accordance with the TBP-DNA binding data, mutating 5'-TTTATA-3' into 5'-TCTCTA-3' resulted in a marked reduction (by 60-70%) of promoter activity (Fig. 4.5). However, converting the inverted TATA sequence into a canonical TATA box did not alter promoter activity significantly (Fig. 4.5).

These differences in in vivo promoter activity shown in transfection assays were also demonstrated by "run-off" in vitro transcription analysis (Fig. 4.6a and 4.6b). When two DNA templates containing the rat BSP promoter with different "run-off" transcript sites were incubated with HeLa cell nuclear extract in the presence of NTPs and [α-32P] GTP, radiolabelled "run-off" transcripts, corresponding in size (171 nucleotide and 311 nucleotide) to the products expected for downstream transcription, were identified (Fig. 4.6a and 4.6b). As in the CAT assays (Fig. 4.5) pCAT3 and p3NT-CAT showed similar transcription activity in this system whereas no transcription activity was observed with the p3∆T-CAT mutant.

**TBP is Required for Initiation of the BSP Gene Promoter in Transcription**

In vitro transcription experiments were also performed to determine whether TBP is absolutely required in the initiation of transcription driven by the BSP promoter. The TBP was first inactivated by heating the HeLa cell nuclear extract at 45°C for 15 min (Nakajima et al., 1988). The nuclear extract containing heat-inactivated TBP was not able to initiate transcription when incubated with wild type rat BSP promoter/CAT construct (pCAT3) (Fig. 4.7). However, the in vitro transcription could be rescued by adding recombinant human TBP to the heat-treated nuclear extract (Fig. 4.7).
Figure 4.5 - Effect of Point Mutations on the Rat BSP Promoter Activity. The bar values represent means±S.D. from ten independent transfection experiments, each performed with replicates as measured by liquid scintillation counting CAT assays and normalized to β-galactosidase activity. pCAT-Control (pCATC), which has both the promoter and enhancer of SV40, was used as a positive control, whereas pCAT-Basic (pCATB), which lacks the promoter and enhancer, was used as negative control. The relative levels of CAT expression to pCAT3, the wide type promoter sequence, from each construct are illustrated on the right. p3ΔT-CAT represents a mutated promoter changing 5'-TTATA-3' into 5'-TCTCTA-3', while p3NT-CAT represents a mutated promoter inverting the 5'-TTATA-3' sequence to the canonical 5'-TATAAA-3' sequence.
Figure 4.6 - In Vitro Transcription Analysis of the Rat BSP Promoter Encompassing Wide-type or Mutated Forms of the Inverted TATA Box. Two rat BSP promoter templates were used in an in vitro transcription experiment to generate an 171 bp transcript (a) and a 311 bp transcript (b), as indicated by arrows. In both a and b, lane 1: molecular weight markers, lane 2: mRNA transcribed from wild type rat BSP promoter, lane 3: mRNA transcribed from mutated rat BSP promoter with a canonical TATA box (changing 5'-TATAA-3' into 5'-TATA-3'), lane 4: mRNA transcribed from rat BSP promoter with point mutations in the inverted TATA element (changing 5'-TTATA-3' into 5'-TTCTA-3'). This experiment was performed by R.H. Kim.
Figure 4.7 - Reconstitution of *In Vitro* Transcription Directed by the Rat BSP Promoter with Recombinant Human TBP. Lane 1: molecular weight markers, lane 2: 171 bp transcript generated by a nuclear extract from the HeLa cells (indicated by arrow), lane 3: *in vitro* transcription of the rat BSP promoter using a heat-treated HeLa nuclear extract, lane 4: *in vitro* transcription of the rat BSP promoter using heat-treated HeLa nuclear extract, plus 20 ng of bovine serum albumin, lanes 5-7: *in vitro* transcription of the rat BSP promoter using a heat-treated HeLa nuclear extract plus 10, 20 and 40 ng of recombinant human TBP, respectively. This experiment was performed by R.H. Kim.
DISCUSSION

Although several variants of the classical TATA sequence have been observed, most of the sequences are asymmetrical (Breathnach and Chambon, 1981; Hahn et al., 1989; Reddy and Hahn, 1991), implying that the orientation of the TATA box may be important for directing downstream transcription (Reddy and Hahn, 1991; Greenblatt, 1991, 1992; Nikolov et al., 1993; Phillips, 1993). The nucleotide sequence in the TATA box of the BSP gene is inverted relative to the consensus TATA box, yet it is still functional and crucial in driving downstream transcription of the BSP gene. Characterization of the naturally occurring inverted TATA box in a higher eukaryotic gene in this study provides conclusive evidence that the orientation of transcription is not dependent upon the orientation of TATA box.

Previously, an apparently inverted TATA-box sequence (5'-TTTGTA-3') has been reported in the adenovirus IVa2 promoter (Carcamo et al., Kasai et al. 1992). However, in the IVa2 gene, the "TATA" box is located 15 nucleotides downstream of the transcription start site. Although synthetic promoters with both canonical and inverted TATA sequences have been shown to drive transcription from similar initiation sites and in the same direction when used in an in vitro transcription system, transcription activity from the artificially inverted TATA sequence is greatly reduced (Carcamo et al. 1991; O'Shea-Greenfield and Smale 1992). In the BSP gene promoter, transcription activity is comparable between the wild-type inverted TATA sequences and the mutated canonical TATA box both in vitro and in vivo. As in all other RNA polymerase II-transcribed genes, either with or without a TATA box, the activity of the BSP promoter containing a perfect inverted TATA box requires the participation of TBP. Thus, it is apparent that TATA boxes, similar to enhancers and silencers (Maniatis et al., 1987), are operational in either orientation.

The concept that the orientation of the TATA box directs the polarity of transcription has been based on the observation that most TATA boxes discovered so far are asymmetric in
nucleotide sequences (Reddy and Hahn, 1991; Greenblatt, 1991, 1992; Phillips, 1993). Additionally, inversion of a TATA box artificially results in loss of down-stream transcription activity (Nagawa and Fink, 1985; Ruden et al., 1988). The recent elucidation of the tertiary structure of TBP from the plant Arabidopsis thaliana (Nikolov et al., 1992) has revealed that TBP resembles a molecular saddle which could sit astride the duplex DNA, protecting approximately 10 bp of DNA sequence. Although the two functional domains of TBP are symmetrical topologically, the amino acids forming the surfaces that interact with the DNA are different (Nikolov et al., 1992), as predicted by previous mutagenesis studies of TBP (Reddy and Hahn, 1991; Strubin and Struhl, 1992; Yamamoto et al., 1992). Consequently, it is likely that TBP interacts with the TATA element in a directional manner (Reddy and Hahn, 1991; Nikolov et al., 1992; Greenblatt, 1992). More recently, the three-dimensional structure of plant and yeast TBPs complexed with the typical TATA boxes were determined by X-ray crystallography (Kim et al., 1993; Kim et al., 1993). In both cases, the TBPs interacted with the TATA boxes with the same polarity, and the TBPs bound to the minor groove of DNA as predicted by earlier studies (Lee et al., 1991; Starr and Hawley, 1991). Based on those data, Kim et al. (1993) and Klug (1993) hypothesized that "the preferred orientation of TBP/TATA binding may be derived from asymmetry in the deformability of the TATA element". However, the possibility that the same polarity of TBP/TATA binding in both cases of co-crystallization might be achieved due to random process still can not be excluded. My study indicates that the questions of TBP/TATA binding polarity and the transcription directionality are more complicated than previously thought. It is of interest to note that the initiator element (Inr) has also been suggested to be a determinant for the direction of transcription in case of the absence of a TATA box (O'Shea-Greenfield and Smale 1992). However, it is unlikely that the initiation of the BSP transcription utilizes the mechanisms of the initiator element because of the following findings: 1). The rat BSP promoter does not contain an identifiable initiator sequence; 2). TBP binds strongly to the inverted TATA-box of the rat BSP promoter; 3). TBP is absolutely required for the initiation of BSP gene transcription.
Figure 4.8 - Frequency of Nucleotide Substitutions in Functional Eukaryotic TATA Boxes. The frequency of nucleotides occurring in each position of the consensus TATA box (upper case) together with the flanking nucleotides (lower case) is shown as a percentage obtained from 60 different genes (Darnell et al., 1986). Below the percentage frequency of the occurrence of the nucleotides present in the BSP TATA box is shown for both the upper strand "inverted" TATA sequence and the complementary lower strand sequence.
The three-dimensional structure of the TBP from A. thaliana complexed with a 14 bp oligonucleotide bearing the AdMLP TATA box has recently been refined at 19 nm (1.9 Å) resolution by Kim and Burley (1994). Model building based on their newly refined co-crystal structures predicted that substitution of a G-C base pair at position 2 reduces TATA box efficiency to 1% whereas a T-A substitution at this position would be tolerated. Inversion of the AdMLP TATA box generates an inverted TATA box with T-A to A-T substitutions at positions 2, 4, and 6, and inversion of the yeast CYI TATA box generates A-T to T-A substitutions at positions 2 and 7. Model building studies suggested that these inverted TATA box do not generate obviously unfavourable interactions with TBP. These studies may therefore explain my observations that inverted TATA box of the BSP gene maintains a comparable activity with that of a canonical TATA box. Of note, the inverted TATA box of the BSP gene differs from the above-mentioned inverted TATA box in that the T-A substitutions are located at position 2 and 5 (Fig. 4.8) (Sodek et al., 1996). Although the binding polarity could be due to asymmetry in the deformability of the two domains of TBP and the two halves of the recognition site of the TATA box (Kim et al., 1993; Kim and Burley, 1994), how TBP actually binds to the inverted TATA box and how the direction of transcription is achieved is yet to be resolved. Thus it would be of interest to co-crystallize TBP with the functional inverted TATA box of the BSP gene. The inverted TATA box in the BSP gene promoter therefore can provide an excellent model to facilitate studies of the polarity of the TFIID-DNA complex and its relation to downstream transcription.

The ability of an asymmetrical TBP molecule to bind to an inverted TATA box and to form a functional PIC, thereby directing downstream transcription, has important implications in elucidating the binding interactions between the promoter, TBP and PIC components. Two possible models are presented (Fig. 4.9) to explain the functional implications with respect to the binding of TBP to an inverted TATA box. First, TBP binding to the 5'-TTTATA-3' sequence on the upper strand of the inverted TATA box could be modulated by the flanking sequences of the TATA box or by the auxiliary factors (eg. TAFs) as depicted in Fig. 4.9a., and the downstream
transcription is thus determined. In this model, the binding of TBP and its orientation would be the same as in the consensus TATA box and variant TATA-box sequences, and perhaps also in TATA-less promoters. As an example, nucleotides flanking a conserved TAAT core are capable of dictating the DNA binding of murine homeodomain proteins (Catron et al., 1993). Secondly, TBP could bind to the 5'-TATAAA-3' sequence on the lower strand of the inverted TATA box. The positioning of the polymerase II for downstream transcription might thus be determined by the auxiliary factors such as the TAFs (Fig. 4.9b), in which case protein-DNA interactions other than those involved with TBP binding might be anticipated. In this regard there is evidence indicating that the direction of transcription might depend on the relative positions of the TATA box and other upstream elements such as the CCAAT box (O'Shea-Greenfield and Smale 1992; Xu et al., 1992). Notably, an inverted CCAAT box that is required for BSP promoter activity is retained within the highly conserved immediate promoter region of the rat and human BSP (Chapter III).

The functional significance of an inverted TATA box in the BSP gene promoter besides controlling basal transcription still remains to be investigated. Variant TATA boxes have been suggested to function as a selective element to determine tissue-specific gene expression by interaction with distinctive TFIID complexes (Wefald et al., 1990; Pugh and Tjian, 1992; Brou et al., 1993). As a variant form of TATA box, it is possible that the inverted TATA box may play an important role in controlling the spatial and temporal expression of the BSP gene during bone formation and bone development. It is conceivable that the variant sequences of the TATA boxes might be important for the recognition of certain transcription factors that compete with TBP for binding in this region and thereby regulate gene expression. Moreover, since the binding of TBP to the TATA box induces the DNA to bend (Kim et al., 1993; Kim et al., 1993), variant nucleotide(s) in the TATA boxes might be able to dictate the degree of DNA bending and, as such, provide a means of fine tuning the stimulation or repression of transcription via TBP/specific trans-
factor interactions. As it is recognized that TBP is able to interact with a number of transcription factors (Brou et al., 1993; Greenblatt, 1992 and references therein), it is also conceivable that the specificity of such interactions might depend on the specific changes of TBP/DNA conformation upon the binding of TBP to various types of TATA boxes.
Figure 4.9 - Model Depicting Possible Interactions between the TBP and the Inverted TATA Box. In (a) the flanking DNA sequence of the inverted TATA box (5'-TTTATA-3') or the auxiliary factors (e.g., TAFs) determines TBP binding to the upper strand, and polymerase II is thus orientated for downstream transcription. In (b) the TBP binds to the 5'-TATAAA-3' sequence on the lower strand, but TAFs or other auxiliary factors may orientate polymerase II for downstream transcription.
Chapter V

Repression of the Rat BSP Gene Expression by Vitamin D₃ through a Receptor Element Overlapping the Inverted TATA Box*

*Part of this chapter is published: J.J. Li., R.H. Kim, Q. Zhang, Y. Ogata, and J. Sodek. (1997). Eur. J. Oral Sci., in press. The original article in the publication was written by J.J. Li. Y. Ogata assisted with the experiment in Fig. 5.1. Experiment in Fig. 5.3 were performed by Q. Zhang; Experiment in Figs. 5.4 and 5.6 were performed by R.H. Kim.
SUMMARY

BSP is a mineralized tissue-specific protein that is highly expressed during the initial formation of bone and cementum. Expression of BSP is suppressed by the osteotropic hormones, parathyroid hormone and 1,25-dihydroxyvitamin D₃, which regulate bone remodelling. In previous studies, a vitamin D-response element (VDRE) was identified that is integrated with a novel inverted TATA box in the rat BSP promoter which mediates the suppression of BSP transcription (Kim et al., 1996). Although the nucleotide sequences of VDREs in different genes conform to a direct (hexamer) repeat, spaced by three nucleotides, the precise sequences are unique for each VDRE. To determine whether the nucleotide differences in the VDRE influence VDR binding, interactions of VDR proteins with various VDREs have been compared using gel mobility shift analysis. Both natural and recombinant VDRs bound to rat BSP and both mouse and porcine osteopontin (OPN) VDRE oligonucleotides in a concentration-dependent manner with a strong preference for dimer formation, whereas equal amounts of dimer and monomer were bound to the human osteocalcin VDRE. However, whereas a truncated VDR comprising the DNA binding domain alone bound the mouse osteopontin VDRE it failed to interact with the porcine OPN and rat BSP VDREs. VDR binding to the BSP was sequence specific, as shown by mutagenesis analysis, and could be abolished by heat and VDR antibody. Thus, repression of the rat BSP gene expression by vitamin D₃ is achieved through a VDRE overlapping the inverted TATA box, indicating a novel mechanism of gene repression by vitamin D₃, whereby the VDR can compete with the TBP and subsequently prevent the formation of pre-initiation complex. These studies also demonstrate that subtle differences in the nucleotide sequence of VDREs affect VDR binding which mediates the vitamin D₃ response.
INTRODUCTION

The formation of bone and cementum is an integral part of tooth and periodontal tissue development. Although the precise mechanism for bone formation has not been revealed, a number of lines of evidence have indicated that bone sialoprotein (BSP), a major non-collagenous glycoprotein, may play an important role in the initial formation of the mineralized matrices of bone, cementum and dentine. Studies of BSP expression have demonstrated that BSP is a tissue-specific (reviewed in Sodek et al., 1992) and developmentally regulated protein that is synthesized by differentiated osteoblasts, cementoblasts and odontoblasts at sites of de novo mineralized tissue formation (Chen et al., 1992). It has been demonstrated that BSP is able to nucleate hydroxyapatite crystal formation from steady-state, physiological concentration of Ca$^{2+}$ and Pi (Hunter and Goldberg, 1993).

The expression of BSP gene is regulated by osteotropic hormones and cytokines that modulate mineralized tissue formation. For instance, glucocorticoids have been shown to induce BSP expression (Oldberg et al., 1989) in association with the promotion of bone formation in vitro (Kasugai et al., 1991). In contrast, vitamin D$_3$ can suppress BSP gene expression (Oldberg et al., 1989) in accordance with its ability to inhibit bone formation in vitro (Ishida et al., 1993). Steroid and steroid-related hormones, including glucocorticoids and vitamin D$_3$, exert their regulatory effects by binding to cognate receptors which belong to a superfamily of "zinc finger" transcription factors (Evans, 1988). Upon binding to the specific ligands, these hormone receptors influence gene transcription via interaction with specific DNA elements such as the glucocorticoid-response element (GRE) and the vitamin D-response element (VDRE) (Umesono et al., 1991).

It is evident that the elucidation of the molecular mechanisms that regulate the expression of BSP gene by osteotropic hormones can provide important insights into the understanding of bone formation and tooth development. To decipher the DNA sequences and transcription factors that control and regulate the expression of BSP gene, the rat and human BSP gene promoters were
isolated (Chapter III; Li and Sodek, 1993; Kim et al., 1994) and I have shown that a novel inverted TATA box is crucial for BSP gene expression (Chapter IV; Li et al., 1995). The DNA sequences that are involved in the regulation of BSP gene expression by glucocorticoids and vitamin D₃, i.e. GRE and VDRE have also been identified (Sodek et al., 1994, Ogata et al., 1995; Kim et al., 1996). The present study provides further characterization of the VDRE in the BSP gene and its molecular mechanisms interacting with VDR.

MATERIALS AND METHODS

Construction of Recombinant Plasmids

Initially five deletion mutants between nts -801 to +60 relative to the transcription start site of the 5'-flanking region of the rat BSP gene were generated by polymerase chain reactions (PCR) and cloned into a CAT expression vector, designated as pCAT 1 to 5 (Fig. 3.7, Chapter III). These deletion mutants were then cut and ligated into the multi-cloning region of an expression vector which contains the firefly luciferase reporter gene but has no promoter and enhancer, designated as pLUC 1 to 5, respectively. Additional deletion mutants extending to -3 kb were prepared according to appropriate restriction enzyme sites, or by exonuclease III digestion from the 5' end, and cloned into the same luciferase expression vector, designated as pLUC 6 to 13 as described previously (Ogata et al., 1995). The recombinant plasmids were purified by using the Qiagen Plasmid Kit (Qiagen Inc.).

Cell Culture and Co-transfection

The rat osteoblastic cell line ROS17/2.8 was maintained in culture as described in Chapter III. Exponentially growing ROS 17/2.8 cells, at 30%-50% confluence, were transfected with 1 µg of a BSP/luciferase fusion plasmid and 0.5 µg of human VDR expression vector using the DEAE-Dextran method (Ogata et al., 1995). After 3 h, the DNA-DEAE-Dextran complex was added to the culture medium and the cells were then shocked by exposure to 10% DMSO in PBS buffer for 2 min at room temperature (22°C). The dishes were then washed and fresh medium containing
10^{-8}\text{M} \text{ of vit. D}_3 \text{ was added and incubation resumed for } 48 \text{ h prior to harvesting.}

**Analysis of Luciferase Activity**

The luciferase assay was carried out according to the Promega protocol, using a Berthold Lumat LB-9501 luminometer (Fisher Scientific, Toronto) to determine the luciferase activity. The luciferase activity was normalized for transfection efficiency by analyzing a co-expressed CAT enzyme using a CAT-ELISA Kit (Boehringer Mannheim) following the manufacturer's instructions.

**Preparation of Endogenous and Recombinant VDRs**

To obtain the natural VDR, crude nuclear extracts were prepared from human bone cells (MG63) after treatment with 10^{-8}\text{M} \text{ vitamin D}_3. Recombinant human VDR (hVDR) and a truncated human VDR containing the DNA binding domain (VDRF), prepared as described previously (Towers et al., 1993), were generously provided by Dr. L.P. Freedman (Sloan Kettering Cancer Institute, New York).

**Electrophoretic Gel Mobility Shift Assays**

The ability of the rat BSP (rBSP) VDRE sequence to recognize the VDR protein was determined by Gel Mobility Shift (GMS) assays. The complementary strands of 25-mer (or 21-mer) oligonucleotides, encompassing the rBSP VDRE and VDREs from porcine osteopontin (OPN) and human osteocalcin (OC), were annealed by mixing the two strands in 1:1 molar ratio, heated to 60°C for 2 min and slowly cooling to room temperature (22°C) for 30 min. The respective duplex oligonucleotides (3.5 pmol) were end-labeled with [\gamma^{32}\text{P}]\text{ATP} (3000 Ci/mmol) using T4 polynucleotide kinase (Pharmacia). Nuclear extracts from the human bone cell line MG63, or recombinant hVDR, were incubated with 0.035 pmol of the labelled, double-stranded DNA probe at room temperature for 45 min (Towers et al., 1993). Competition experiments were performed by including various unlabelled oligonucleotides, or specific antibodies, in the binding reactions.
A monoclonal antibody against the C-terminal of the DNA binding zinc finger domain of the VDR was purchased from Chemicon Intl., Inc., CA. Protein-DNA complexes in the reaction mixture were resolved by electrophoresis on a 6% non-denaturing acrylamide gel.

**RESULTS**

*Molecular Organization of Various VDREs*

Recent studies have characterized a number of VDREs in the human (Ozono *et al.*, 1990) and rat (Demay *et al.*, 1990) osteocalcin genes, in the mouse (Noda *et al.*, 1990) and porcine (Zhang *et al.*, 1992) osteopontin genes, in the rat calbindin-D9K (Darwish and DeLuca, 1992) and mouse calbindin-D28K (Gill and Christako, 1993) genes, in the human parathyroid hormone (PTH) gene (Demay *et al.*, 1992), in the rat 25-hydroxyvitamin D3 24-hydroxylase gene (Ohyama *et al.*, 1994), and in the interleukin-2 (IL-2) gene (Alroy *et al.*, 1995). Among these only VDREs in the PTH and IL-2 genes function as repressor elements. A consensus VDRE, which contains a direct repeat of two six-nucleotide half sites with a three-nucleotide spacer (AGGTCAnnnAGGTCA) (Umesono *et al.*, 1991), has been derived from comparisons of a number of these VDREs (Table 5.1). A computer search of the 3 kb upstream sequence of rat BSP promoter (Chapter III), using the consensus sequence, identified a putative VDRE (AGGGTTtatAGGTCA) in the proximal promoter region (-28 to -14) in which 12 out of 15 nucleotides were matched (80% identity). Intriguingly, this motif overlapped the functional inverted TATA box (Li *et al.*, 1995). While the 3' half-site of this putative rat BSP (rBSP) VDRE conforms to the consensus VDRE the last three nucleotides in the 5' half-site are divergent.
<table>
<thead>
<tr>
<th>Genes</th>
<th>Position</th>
<th>Requirement of</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enhancers</strong></td>
<td></td>
<td>NAF</td>
</tr>
<tr>
<td>Consensus VDRE</td>
<td>A G G T C A n n n A G G T C A</td>
<td>(-499/-485)</td>
</tr>
<tr>
<td>Human osteocalcin (hOC)</td>
<td>G G G T G A a c g G G G G C A</td>
<td>(-460/-446)</td>
</tr>
<tr>
<td>Rat osteocalcin (rOC)</td>
<td>G G G T G A a t g A G G A C A</td>
<td>(-489/-475)</td>
</tr>
<tr>
<td>Mouse osteopontin (mOPN)</td>
<td>G G T T C A c g a G G T T C A</td>
<td>(-757/-743)</td>
</tr>
<tr>
<td>Rat calbindin-D9k (rCaBP)</td>
<td>G G G T G T c g g A A G C C C</td>
<td>(-489/-475)</td>
</tr>
<tr>
<td>Mouse calbindin-D28k (mCaBP)</td>
<td>G G G G A T g t g A G G C A</td>
<td>(-198/-182)</td>
</tr>
<tr>
<td>Porcine osteopontin (pOPN) (i)</td>
<td>G G G T C A t a t G G T T C A</td>
<td>(-2259/-2245)</td>
</tr>
<tr>
<td>Porcine osteopontin (pOPN) (ii)</td>
<td>G G C T A A t t a G C A G C A</td>
<td>(-845/-831)</td>
</tr>
<tr>
<td>Rat 24-hydroxylase (rHDL) (i)</td>
<td>A G G T G A g t g A G G C G</td>
<td>(-137/-151, AS)</td>
</tr>
<tr>
<td>Rat 24-hydroxylase (rHDL) (ii)</td>
<td>G G T T C A g c g G G T G C G</td>
<td>(-245/-259, AS)</td>
</tr>
<tr>
<td>Repressors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human parathyroid hormone (hPTH)</td>
<td>G G T T C A a a g C A G A C A</td>
<td>(-108/-122, AS)</td>
</tr>
<tr>
<td>Rat bone sialoprotein (rBSP)</td>
<td>A G G G T T t a t A G G T C A</td>
<td>(-28/-14)</td>
</tr>
</tbody>
</table>

Table 5.1 - Comparison of the Putative rBSP VDRE with the Consensus and Other Identified VDREs.

The consensus VDRE was suggested (Umesono et al., 1991) as a direct repeat with 3-nt spacer (DR3). The identified VDREs are from previous publications: hOC (Ozono et al., 1990); rOC (Demay et al., 1990); mOPN (Noda et al., 1990); rCaBP (Darwish et al., 1992); pOPN (Zhang et al., 1992); hPTH (Demny et al., 1992); mCaBP (Gill and Christakos, 1993); rHDL (Ohyama et al., 1994); and IL-2 (Alroy et al., 1995). The rBSP VDRE is shown overlapping the inverted TATA box (underlined). AS, antisense; n/a, not available.
DNA Element that Mediates Suppression of BSP Transcription by Vitamin D₃

To characterize this putative VDRE, a series of deletion mutants were prepared and chimeric constructs extending from nts -2992 to +60 (Fig. 5.1; also see Fig. 3.4, Chapter III) relative to the transcription start site were transfected into ROS17/2.8 cells. Following treatment by vitamin D₃, the promoter activities were measured by determining the expression of the luciferase reporter gene (Fig. 5.1). These results revealed that the repressive effects of vitamin D₃ on BSP transcription are mediated through DNA sequences between nts -116 to +60, since a 50–60% reduction of promoter activity was observed in constructs including this region of the promoter.

Recognition of the rBSP VDRE (but not the TATA Box Element) by Natural VDR

To test if the rBSP VDRE was able to interact with native VDR, nuclear extracts prepared from MG63 cells were incubated with a [³²P]-labelled double-stranded oligonucleotide, representing the rBSP VDRE sequence, and analyzed by gel mobility shifts. A major and a faster-migrating minor retardation band, corresponding to dimer and monomer forms, respectively (see below), of the VDR were observed to bind to the rBSP VDRE sequence (Fig. 5.2A). Since the rBSP VDRE includes the inverted TATA-box that is able to interact with the TATA-box Binding Protein (TBP) (Chapter IV; Li et al., 1995), the core component of universal transcription factor TFIID, an oligonucleotide containing a consensus TATA box was tested for binding activity. However, no interaction with the MG63 nuclear extract proteins was observed (Fig. 5.2A) indicating that the protein causing the shift with the rBSP oligonucleotide recognizes the VDRE rather than the TATA element of the rBSP gene.
Figure 5.1 - Characterization of the DNA Sequences that Confer the Repressive Effect of Vitamin D3 by Transient Transfection Assays. Chimeric constructs pLUC3 to pLUC13 that include various lengths of the rat BSP promoter sequence (as shown) and a luciferase reporter gene were used in transient transfection assays to analyze for vitamin D3 effects on gene transcription. Numbers on the left of each deletion mutant represent the 5'-end of the promoter sequence relative to the transcription start site, all of which encompass the putative VDRE overlapping the inverted TATA box. Numbers on the right represent % suppression of luciferase activity in the rBSP promoter constructs by vitamin D3. This experiment was assisted by Y. Ogata.
Figure 5.2 Gel Mobility Shift Analysis of Native VDR Binding to the rBSP VDRE.
a). The VDR from MG63 cells recognizes the rBSP VDRE but not the TATA box. Synthetic oligonucleotides containing the rBSP VDRE or the consensus TATA box (TFIIID oligo) were used as $^{32}$P-labelled probes for gel mobility shift assays with nuclear extracts of MG63 cells. Lane 1, rBSP VDRE probe alone; Lane 2, rBSP VDRE probe plus MG63 extract; Lane 3; TFIID oligo probe plus MG63 extract.
b). Binding of VDR from the MG63 cells to the VDREs in the hOC, pOPN and rBSP gene promoters. Synthetic oligonucleotides containing various VDREs (see Table 5.1) were used as $^{32}$P-labelled probes to perform gel mobility shift assays with either MG63 cell nuclear extract or the DNA binding domain of the recombinant hVDR (VDRF). A). hOC VDRE; B). pOPN VDRE; and C). rBSP VDRE. Lane 1, no protein; Lane 2, MG63 cell nuclear extract; Lane 3, 4, 5 with 50 ng, 150 ng, 300 ng of VDRF, respectively. Two species of bound proteins, as well as the unbound probes, are indicated by arrows.
To further characterize the nuclear proteins of the MG63 cells that interact with the rBSP VDRE, the properties of the MG63 nuclear extract protein that bound to the rBSP VDRE were compared to known VDREs in the human osteocalcin (hOC) (Ozono et al., 1990) and the porcine osteopontin (pOPN) genes (Zhang et al., 1992). Gel mobility shift analysis revealed that two species of MG63 nuclear proteins, corresponding to those formed with the rBSP VDRE, were formed with both the hOC and the upstream (5') pOPN VDREs (Fig. 5.2B). Hence, these two distinct species of MG63 nuclear proteins appear to represent the monomer and dimer forms of natural VDR. To confirm that the protein component in the MG63 nuclear extracts that bind to the VDREs in rBSP, pOPN and hOC genes is actually the VDR, the ability of a VDR antibody to block the protein-DNA interaction were tested. As demonstrated in Fig. 5.3A, an anti-VDR monoclonal antibody blocked the protein-DNA interaction. In contrast, an unrelated monoclonal antibody raised against rat BSP protein, had no effect on the binding confirming that the endogenous VDR in the MG63 nuclear extracts binds to the rBSP VDRE. Of interest, the endogenous VDR lost its VDRE binding ability when the MG63 nuclear extracts were heated at 45°C for 15 min prior to performing binding assay (Fig. 5.3A), suggesting that a heat-labile component in the MG63 nuclear protein is required to facilitate the interaction between the VDR and the rBSP and pOPN VDRE.

The effects of VDR antibody and heating on the interactions of the MG63 VDR with the 5' pOPN (Fig 5.3B) and hOC (Fig 5.3C) VDREs were also studied. In both cases results comparable to those obtained with the rBSP VDRE were obtained.

**Binding of both Natural and Recombinant VDRs to the rBSP VDRE**

While the MG63 VDR binds to the hOC VDRE producing similar amounts of dimer and monomer, binding to the pOPN and rBSP VDREs occurred largely in dimer form. To determine whether these patterns of binding were consistent, studies were also conducted with purified recombinant hVDR protein. Similar to the natural VDR from the MG63 cells, the recombinant
hVDR bound to the rBSP VDRE and the pOPN VDRE predominantly as the dimer form, and the binding of recombinant hVDR to these VDRE sequences was concentration-dependent (Fig. 5.5A & 5.5B). Notably both the 5' upstream (nts -2259 to -2245) and a more recently recognized 3' downstream (nts -845 to -831) VDRE in the pOPN gene showed strong interaction with the hVDR protein (Fig. 5.5A).

To gain further insights on the interaction of the VDR with the rBSP VDRE, a truncated form of recombinant hVDR that contains the DNA zinc-finger binding domain (VDRF) was used for GMS analysis. Although the truncated hVDR (VDRF) was able to bind to the mouse osteopontin VDRE with high affinity (Fig. 5.4; Towers et al., 1993), under the same binding conditions specific interactions of the VDRF with the VDREs of the rBSP, pOPN, and hOC genes were not observed (Fig. 5.2B; Fig. 5.4; Fig.5.5B).

**Binding Specificity of the Natural and Recombinant VDRs to the rBSP VDRE**

A series of GMS experiments were performed to test if the binding of the MG63 VDR to the rBSP VDRE is sequence-specific. As demonstrated in GMS competition experiments, unlabeled hOC, pOPN, and rBSP VDRE oligonucleotides were capable of competing with the [32P]-labelled rBSP VDRE for binding to the native VDR (Fig. 5.3). Similarly, the unlabelled rBSP VDRE, hOC VDRE and pOPN VDRE were capable of competing with the rBSP VDRE for binding to the recombinant hVDR (Fig. 5.3; Fig. 5.6). In contrast, non-specific competitors, including DNA sequences for transcription factors AP-1, A-P2, CREB, CTF/NF1, OCT1, as reported previously (Kim et al., 1996) failed to compete with the rBSP VDRE for VDR binding. Of note, the 3'-half-site of the rBSP VDRE could compete with the complete element for VDR binding effectively whereas the 5'-half-site of the VDRE showed much less competitive capacity (Fig. 5.6).
Figure 5.3 - Gel Mobility Shift Analysis of the Binding Specificity of the Natural VDR from MG63 Cells.

a). Binding specificity to the rBSP VDRE. [32P]-labelled synthetic oligonucleotide containing the rBSP VDRE was used as a probe to study the interaction with the endogenous VDR from MG63 cells. Various competitors, including rBSP VDRE (Lane 3), hOC VDRE (Lane 4), or pOPN VDRE (Lane 5), at 100-fold excess, were added to the binding reactions. Both monoclonal antibody against the VDR (αVDR Ab) (Lanes 6 and 7) and monoclonal antibody against BSP protein (αBSP Ab) (Lane 8) were tested for neutralization of the VDR binding ability to the rBSP VDRE. MG63 nuclear extracts were also heated at 45°C for 15 min prior to gel mobility assays (Lane 9). b). Binding specificity to the pOPN VDRE. [32P]-labelled synthetic oligonucleotide containing the pOPN VDRE was used to study the interaction with the natural VDR from MG63 cells. Competitors include rBSP VDRE (Lane 3), monoclonal antibody against the VDR (αVDR Ab) (Lanes 4 and 5) and monoclonal antibody against BSP protein (αBSP Ab) (Lane 6). MG63 nuclear extracts were also heated at 45°C for 15 min prior to gel mobility shift assays to test the effect of heat on the VDR-DNA interaction (Lane 9). c). Binding specificity to the hOC VDRE. [32P]-labelled synthetic oligonucleotide containing the hOC VDRE was used as probe to study the interaction with the endogenous VDR from MG63 cells. Competitors include hOC VDRE (Lane 3), monoclonal antibody against the VDR (αVDR Ab) (Lanes 4 and 5). This experiment was performed by Q. Zhang.
Figure 5.4 - Gel Mobility Shift Analyses of the Truncated form of Recombinant VDR (hVDRF) Binding to VDREs. $^{32}$P-label labelled synthetic oligonucleotides containing the mOPN VDRE(A), rBSP VDRE (B), 5'pOPN VDRE (C), 3'pOPN VDRE (D), or hOC VDRE (E) were used as probes to perform gel mobility shift assays with the recombinant hVDRF that contains only the DNA binding domain. Lane 1 to 3, increasing amount of the VDRF were added into each reaction mixture, respectively. This experiment was performed by R.H. Kim.
Figure 5.5 - Gel Mobility Shift Analyses of Recombinant VDR Binding to VDREs.

a). Recombinant human VDR binds to both upstream and downstream pOPN VDREs. $^{32}$P-labelled synthetic oligonucleotides containing the pOPN VDREs located at nts -2259 to -2245 (A) or at nts -845 to -831 (B) (see Table 5.1) were used as probes to perform gel mobility shift assays with recombinant hVDR. Lanes 1 to 4: 0 ng, 50 ng, 100 ng, and 200 ng of recombinant hVDR, respectively. b). Recombinant human VDR, but not the VDRF, binds to the rBSP and the pOPN VDREs. $^{32}$P-labelled synthetic oligonucleotides containing the rBSP VDRE (A) or the pOPN VDRE (B) were used as probes to perform gel mobility shift assays with recombinant hVDR. Lanes 1 to 4: 50 ng, 100 ng, 300 ng, 600 ng of recombinant hVDR, respectively. Lanes 5 to 8: 50 ng, 100 ng, 300 ng, 600 ng of the DNA binding domain of the recombinant hVDR (VDRF), respectively.
Figure 5.6 - Gel Mobility Shift Analysis of the Binding Specificity of the Recombinant hVDR to the rBSP VDRE. $[^{32}P]$-labelled synthetic oligonucleotide containing the rBSP VDRE was used as a probe to study the interaction with the recombinant hVDR. Various competitors, including the specific competitors; rBSP VDRE (Lane 2), 5' half-site of the rBSP VDRE (Lane 3), 3' half-site of the rBSP VDRE (Lane 4), hOC VDRE (Lane 5), and pOPN VDRE (Lane 6), at a 50-fold excess were added to the binding reactions. This experiment was performed by R.H. Kim.
DISCUSSION

In a previous study it was reported that the rat BSP promoter activity could be suppressed approximately 28% by vitamin D$_3$ in the absence of co-transfected VDR expression vector in transient transfection analysis (Li and Sodek, 1993). By co-expressing hVDR with BSP/Luciferase fusion gene, 50~60% of suppression of promoter activity was obtained, indicating that the repressive effect of vitamin D$_3$ on this gene is mediated by a DNA segment between nt -116 to +60 in the rat BSP promoter (Fig. 5.1). Synthetic oligonucleotides encompassing the rBSP VDRE within this region bind to both native and recombinant VDRs with a strong preference for dimer formation (Figs. 5.2 to 5.5). Furthermore, the binding of the VDR to the rBSP VDRE is sequence-specific and antibody-sensitive as demonstrated by GMS competition studies (Figs 5.2 to 5.6). Similar to the rBSP VDRE, the native and recombinant VDRs bind to both pOPN VDREs preferably in a dimer form. In contrast, both native and recombinant VDRs bind to the hOC VDRE with approximately equal amounts of dimer and monomer.

Previously I demonstrated that the unique inverted TATA box is required to drive downstream transcription of the rat BSP gene (Chapter IV; Li et al., 1995). That the VDRE overlaps the inverted TATA box indicates a novel mechanism for gene suppression by vitamin D$_3$, although the presence of a steroid hormone receptor element overlapping a TATA box is not unique since in the human osteocalcin gene a GRE overlaps the TATA element and is believed to mediate the suppression effect on osteocalcin gene expression by glucocorticoids (Strömstedt et al., 1991). This mechanism, which may represent a general mechanism of steroid hormone action on bone cells during bone development and remodelling, supports the concept that repressive signals can directly affect the core element of transcription initiation, namely the TATA box region. Thus, a homeodomain protein (engrailed, En) that is important in regulating temporal and spatial gene expression during Drosophila development has been shown to repress in vitro transcription of the Drosophila Hsp 70 promoter by competing with the TATA-box binding protein (Ohkuma et al.,
1990). Also, more recently the repression of IL-2 gene transcription by vitamin D₃ has been shown to occur through a VDRE overlapping NFAT-1 (nuclear factor of activated T cells) element; the VDR competitively knocking out NFAT (Alroy et al., 1995). Because the binding of TBP to the inverted TATA box is a crucial step for the transcription of the rat BSP gene (Chapter III), it is conceivable that the VDR, following activation by vitamin D₃, could compete with TBP for occupancy of the TATA/VDRE element thereby preventing the assembly of the pre-initiation complex required for the initiation of transcription as depicted in Fig. 5.7.

In contrast to most identified VDREs which accommodate the VDR complexed with an accessory protein as heterodimer (Ohyama et al., 1994), the rBSP VDRE can bind to purified recombinant hVDR largely as homodimer in the absence of vitamin D₃ in vitro (Fig. 5.5). However, the functional relevance of this homodimeric binding is unknown. Carlberg et al. (1993) have suggested that both homodimeric and heterodimeric VDR complexes may be involved in vitamin D₃-induced function and that the nature of dimerization could be dependent on the specific arrangement of the DNA target site. The accessory proteins that can heterodimerize with VDR to bind to the human osteocalcin VDRE have been identified as a heat-labile nuclear accessory protein (NAF) (Liao et al., 1990; MacDonald et al., 1991), or a co-regulator retinoid X receptor b (RXRb) (Carlberg et al., 1993). In this study, a heat-labile protein from MG63 cells apparently contributed to the binding of the endogenous VDR to the rBSP and pOPN VDREs (Fig. 5.3). It remains to be investigated whether or not this heat-labile protein is related to the identified NAF, RXRb, or is a novel accessory protein. That complete repression of the rat BSP transcription by vitamin D₃ has not been achieved by co-expression of the hVDR could imply that a co-regulator may be required for full vitamin D₃ action (Fig. 5.1). Transfections in cells lacking RXR background, such as yeast (Saccharomyces cerevisiae) or RXR-null mammalian cell lines, or employment of RXR-null mouse strains may provide valuable insights to elucidate the contribution of retinoic acid receptors in BSP gene regulation by vitamin D₃.
Figure 5.7 - Model of Vitamin D$_3$-mediated Suppression of rBSP Gene Transcription.

(a) The binding of TBP, the core protein of the TFIID that recruits the pre-initiation complex, to the inverted TATA box is a crucial step for the transcription of the rat BSP gene (Chapter IV; Li et al., 1995). (b) Subsequent to vitamin D$_3$ signalling, the VDR could compete with TBP for occupancy of the TATA/VDRE element, thereby preventing the assembly of the pre-initiation complex that is required for the initiation of transcription.
Notably, while the DNA binding domain of VDR binds to the mouse osteopontin VDRE with high affinity, its interaction with the rBSP, hOC, and pOPN VDREs under the same conditions is poor (Figs 5.2; 5.4; 5.5). The fact that the VDRF alone was not able to bind to the rBSP and pOPN VDREs suggests that a full configuration of the VDR is necessary to complex with another VDR molecule to form a homodimer of VDR, or to complex with an accessory protein to form a heterodimer on either VDRE. In studies of vitamin D3 repression of IL-2 transcription, it was shown that DNA binding by VDR is necessary but not sufficient to mediate IL-2 suppression. Only the VDR-RXR heterodimer was capable of stable binding to the NFAT-1 element causing the abrogation of NFATp/AP1 complex formation that results in IL-2 repression (Alroy et al., 1995). Recently, it was demonstrated that the C-terminal region of the hVDR is essential to heterodimerize with an auxiliary factor that is required for high affinity DNA binding, and the functional domains involving vitamin D3 ligand binding and hetero-dimerization have been mapped to two separate segments (Nakajima et al., 1994).

Although members of the steroid receptor superfamily share highly conserved DNA binding domains and their cognate response elements have a high degree of similarity, a wide range of biological effects can be achieved as a result of highly specific DNA-protein interaction. Therefore, the selectivity of protein-DNA recognition may be dictated by subtle differences in the hormone response elements together with specific protein conformations. While GRE and ERE are often palindromic DNA sequences, the VDRE is organized as direct repeats (Umesono et al., 1991) and no currently identified VDREs are identical (Table 5.1). Notably, another repressor VDRE found in the human parathyroid hormone gene is inverted (i.e. 5' to 3' on the antisense strand) (Demay et al., 1992).
In summary, this study has demonstrated that the repression of the rat BSP gene transcription by vitamin D₃ is achieved through a VDRE overlapping the inverted TATA box, indentifying a novel mechanism of gene repression by vitamin D₃, whereby the VDR can compete with the TBP and subsequently prevent the formation of pre-initiation complex that is required for gene transcription. Subtle differences in the nucleotide sequence of VDREs can effect VDR binding which may be important for the modulation of the vitamin D₃ response in different genes.
Chapter VI

General Discussion and Future Directions
My studies have described the first cloning and characterization of a BSP promoter which has been used to study potential regulatory elements that control the expression of the BSP gene in a tissue- and developmental-specific manner. A particularly interesting feature of the promoter is the presence of a unique inverted TATA box which is overlapped by a VDRE. I have shown that the inverted TATA box in the BSP gene is functional and, contrasting prevailing beliefs, have demonstrated that the orientation of a TATA box is not required to determine the direction of downstream transcription. Notably, this is the first demonstration of a perfectly inverted, functional TATA box in a eukaryotic gene. The VDRE overlapping the inverted TATA box is also functional, revealing a unique mechanism for the suppression of gene transcription by steroids, in which VDR appears to compete with the TBP for occupancy of the VDRE/TATA site. Additionally, my studies have shown that subtle differences in VDRE sequences can affect VDR binding, indicating a potential mechanism through which the responsiveness of a variety of genes to vitamin D3 can be specifically modulated.

Although the importance of BSP in bone formation and development is indicated by its highly regulated expression during bone development, the molecular mechanisms that govern its unique expression were virtually unknown when my studies were initiated. An array of hormones and cytokines are involved in regulation of bone formation and development and the ultimate target for controlling gene expression in response to extracellular signals is the transcription machinery. While our knowledge on the transcription of genes involved in skeleton formation is limited in comparison with many other tissues, the regulation of osteocalcin gene expression has provided valuable information. Osteocalcin, like BSP, is tissue-specific and its expression is regulated by numerous extracellular signals. The osteocalcin gene promoter contains a series of consensus regulatory elements including responsive elements for vitamin D (Kerner et al. 1989, Lian et al. 1989, Demay et al. 1990, Markose et al 1990; Darwish and DeLuca, 1993), vitamin A (Schüle et al., 1990a), estrogen (Aronow et al. 1990), glucocorticoids (Akerblom et al 1988, Morrison et al. 1989, Schepmoes et al. 1991, Schüle et al. 1990a, Strömstedt et al. 1991), retinoic acid
(Nishimoto et al., 1987, Schüle et al. 1990b), γ-interferon (Nanes et al. 1990), and TNFα (Li and Stashenko, 1993). In addition, two novel elements (OSE1 and OSE2) within the proximal promoter of the mouse osteocalcin gene have been implicated in tissue-specific expression of osteocalcin (Ducy and Karsenty, 1995).

Since the elucidation of the molecular mechanisms that regulate the tissue-specific expression of the BSP gene can also provide important insights into the understanding of bone and tooth development, as well as metabolic and congenital diseases of mineralized tissues, the initial objective of my research was to clone the BSP promoter and identify the regulatory elements of the BSP gene.

1. Cloning of the BSP Promoter.

The isolation of a genomic clone that includes the promoter region of the rat BSP gene (Chapter III) was the first step towards the study of the molecular mechanisms that regulate BSP gene transcription (Li and Sodek, 1993). Sequence analysis of the promoter revealed a number of putative regulatory elements that could control the tissue-specific and developmental expression of this protein. A comparison of the promoter region, extending from nt -801 to the first exon in the rat BSP gene, with the corresponding region of the human BSP gene, which was subsequently cloned (Kerr et al., 1993; Kim et al., 1994), has revealed a high degree of nucleotide identity which is also similar to the mouse BSP promoter (Gupta and Aubin, unpublished). Within a 370 bp region of the proximal promoter 75% of the nucleotides have been conserved in the rat and human BSP promoter. Consequently, this region has been named the “BSP box” (Sodek et al., 1996). Notably, included in this “BSP box” is the inverted TATA element, together with the inverted CCAAT box, and NFκB, CRE, homeobox-binding, and AP-1 sites. However, the recently cloned chicken BSP promoter is distinct from the mammalian BSP promoters in that the inverted TATA box is not present and the inverted CCAAT box is replaced by SP1 element, although a TATA-like element is located at nt -21 (Yang and Gertenfeld 1997).
The shortest promoter sequence capable of directing the CAT or luciferase reporter gene expression in the transient transfection system was found to include the inverted TATA and CCAAT boxes, which matches the definition of a "core promoter" (up to nt -116 relative to the transcription start site). In addition, variable promoter activities were demonstrated with different lengths of constructs up to --3 kb upstream of the promoter region (Sodek et al., 1994). Within the sequence upstream of -116 bp a negative regulatory element(s) is apparent, since inclusion of the DNA sequences upstream of -116 bp in the construct suppresses transcription in ROS17/2.8 cells (Chapter III; Li and Sodek, 1993).

Within the rat BSP gene promoter a number of novel cis-acting elements that may be involved in hormonal and tissue specific regulation of transcription were identified. For example, heterogeneous TATA boxes, such as the inverted TATA box in the BSP promoter, can function as a selective element to determine tissue-specific gene expression by interaction with distinctive TFIID complexes. Also, the 18 bp palindrome, which has no similarity with any other identified cis-acting elements, may be a potential DNA element that could interact with a tissue-specific transcription factor. Moreover, the CCAAT element could also be a target of tissue-specific regulation as indicated in the regulation of adipocyte differentiation (Santoro et al., 1988; Umek et al., 1991). The CCAAT box has been shown to be the target of serum factors and also for the proto-oncogene src. Notably, the marked increase in transcriptional activity that I observed in a transient transfection assay with the rat BSP promoter constructs which included the inverted CCAAT element (Li and Sodek, 1993), has subsequently been found to be dependent upon serum concentration. Recent studies (Kim et al., unpublished) have demonstrated that the inverted CCAAT box, which appears to be required for serum factor effects, is also the target of v-src regulation. Transfection of src/- cells with a v-src expression factor stimulates transcription driven by a short BSP promoter containing the inverted CCAAT element by 3-4 fold. The src appears to exert its effects through the RAS signalling pathway and the NF-Y transcription factor was the only nuclear complex identified that bound specifically to the inverted CCAAT motif.
Homeobox binding-sites are also of interest in the regulation of gene transcription associated with cellular growth and differentiation, since, historically, homeobox proteins are thought to be key regulators of embryonic development (Levine and Hoey, 1988). The unique expression pattern of the Hox-8 gene suggests its role in specifying tooth initiation and shape (Mackenzie et al., 1992). Genetic analysis demonstrated that Msx-2/Hox8.1 gene product existed within proliferative mesenchymal cells of cranial sutures and the homeobox protein may regulate cranial suture ossification in the developing skull (Jabs et al., 1993). Also, msx1 and msx2 mRNAs are localized at the sutures of osteogenic fronts and at various other sites during craniofacial development of the mouse. Inactivation of msx1 gene results in mice with cleft palate due to failure of the palatal shelves to grow towards each other and fuse, suggesting that this class of homeobox genes plays an important role in normal growth of intramembranous bones at osteogenic fronts (Satokata and Maas, 1994). Of note, mutations in the msx2 gene are found in patients with Boston-type craniosynostosis (Jabs et al., 1993). Remarkably, the expression of both msx1 and msx2 has been shown, at least in developing teeth, to be controlled by BMP4 which has potent cartilage and bone inducing activity (Vainio et al., 1993). As shown by Towler et al. (1994; 1995), the protein encoded by Msx2/Hox8.1 gene can bind to the hox-site in the rat osteocalcin promoter and is required, not only for basal transcription but possibly also for differentiation stage-specific transcription of the osteocalcin gene. Therefore, the family of transcription factors encoded by homeobox genes may play a critical role in skeletal pattern formation and bone development (Morgan and Tabin, 1993; Krumlauf, 1994). Consequently, further study of the role of the hox-site in the BSP promoter may shed light on our understanding of BSP expression during craniofacial and tooth development. Thus, future experiments should include a functional characterization of nuclear protein(s) that bind to the BSP homeobox and mutational analysis of the BSP homeobox site to confirm the role of this element in developmental processes involving BSP expression.
2. Tissue-specific Transcription of the BSP Gene.

Since BSP expression exhibits a clear tissue-specific and developmental stage-specific pattern (reviewed in Chapter I), it was of particular interest to investigate the mechanisms by which BSP gene expression is regulated at the transcription level. To address this question, initial studies were performed using transient transfection assay in various cell types. Although transient transfection assays have proven to be a valuable means of elucidating cell-type specificity of signal transduction and transcriptional regulation, preliminary mapping of a specific DNA element involved in the specificity of BSP expression has not been successful. No clear differences were observed in transcriptional activity of individual constructs spanning ~3 kb of the rat BSP promoter when osteoblastic cells, such as ROS 17/2.8 and UMR 106-06 cells, which constitutively express BSP, were compared with normal pre-osteoblastic cells and kidney fibroblast NRK cells which do not express BSP (Sodek et al., 1994). Similar results have also been observed using constructs of the human BSP promoter (Kerr et al., 1997).

Since transiently transfected promoter constructs are removed from normal chromatin configuration and are not in the context of normal nuclear matrix, it was thought that the BSP promoter constructs may not be regulated in the same way as the endogenous gene. Therefore, transgenic mouse lines were generated in which a BSP-luciferase construct encompassing 2.7 kb of the rat BSP promoter was incorporated into the mouse genome (Chen et al., 1996). The transgene was expressed at consistently high levels in bone tissues with negligible activities in various other organs including kidney, liver, stomach, intestine and spleen, indicating that tissue-specific expression is primarily mediated through this region of promoter. However, absolute specificity was not achieved since variable expression of the transgene was also observed in brain and skin in four transgenic lines. In bone tissues the differentiated osteoblasts expressed the highest levels of the transgene, which is consistent with the pattern of endogenous BSP gene expression. Furthermore, consistent with the expression pattern of endogenous BSP gene, temporal analyses revealed the highest transgene expression in neonatal bones, with expression
decreasing markedly with subsequent growth and development. Thus, some cis-acting element(s) regulating tissue and developmental expression of BSP gene could be located within the ~3 kb DNA sequence of the promoter. However, other regions of the BSP gene, including introns, should also be investigated for a comprehensive analysis of elements that could regulate BSP gene expression in a temporospatially specific manner. Of note, the expression of type II collagen gene in chondrocytes is controlled by both positive and negative elements involving an enhancer in the first intron and a silencer in the promoter region to ensure that the collagen II gene is only expressed in appropriate cells (Savagner et al., 1990).

That cell-type specific cis-acting elements and their cognate trans-acting factors are able to control tissue-specific transcription of particular gene has been clearly demonstrated for myogenic genes. Here, factors of the muscle-specific Myo-D family, which belongs to bHLH family of transcription factors, are required for the differentiation of skeletal myoblasts and can induce several cell types to express a skeletal muscle phenotype (reviewed by Olson and Klein, 1994). In a “knock out” transgenic mice, Myo-D or myf-5 have been shown to be crucial in the formation of skeletal muscle (Rudnicki et al., 1993). Although no specific bHLH genes have been found in osteoblasts, there is indirect evidence of their existence. The expression of Id, an antagonist of bHLH protein binding to the promoter of osteocalcin gene, is down-regulated during osteoblast differentiation (Ogata and Noda, 1991; Kawaguchi et al., 1992), and overexpression of Id inhibits osteoblastic differentiation (Murray et al., 1992). Therefore, further study is necessary to identify the DNA elements, if any, that would direct to the specific expression of BSP gene in bone tissue.

The osteocalcin-specific elements (OSE1 and OSE2) in the mouse osteocalcin promoter have been implicated in tissue-specific expression of osteocalcin gene (Ducy and Karsenty, 1995). Most recently, the transcription factor that binds to OSE2 in the mouse osteocalcin promoter has been cloned and characterized as a “runt domain” protein, CBFA1 (Ducy et al., 1997), which is essential to osteogenic differentiation and osteogenesis, since knocking-out the CBFA1 gene in mice results in the complete lack of bone and extremely low expression of genes associated with
the osteoblastic phenotype (Komori et al., 1997; Otto et al., 1997). Conversely, forced expression of CBFA1 in non-osteogenic cells induces the expression of osteoblastic phenotype associated with the production of bone proteins including osteocalcin and BSP. However, based on previous transfection assays (Sodek et al., 1994), tissue-specific expression is not evident in rat BSP promoter constructs that include putative CBFA binding sites in the rat BSP gene, although tissue specificity has been demonstrated in a reporter construct that includes nts -1224 to +24 of the chicken BSP promoter encompassing a CBFA-1 binding site (Yang and Gerstenfeld, 1997).

Of note, ubiquitous elements are also able to regulate cell-type specific transcription in many cases. For example, the CCAAT element is believed to be a target for regulation of specific genes, as exemplified by the expression of multiple CCAAT-transcription factors (CTF) in a tissue-specific and development-specific pattern (Santoro et al., 1988), while the CCAAT-enhancer binding protein (C/EBP) can act as a differentiation switch for the specialized phenotype of adipocytes (Umek et al., 1991). That ubiquitous DNA elements and transcription factors can determine tissue-specific transcription may be explained by the fact that no ubiquitous cis-acting elements are aligned in identical combination with other DNA sequences, and sterospecificity generated by multiprotein complexes is important in transcription regulation (Tjian and Maniatis, 1994). Thus, through interactions with other transcription factors, ubiquitous factors should be able to assemble a unique functional structure that is cell-type specific. Consequently, ubiquitous cis-acting elements in the BSP promoter, like the CCAAT box, NF1 elements, and hox-binding sites, need to be studied carefully in the context of tissue-specific expression.

An increasing number of observations indicate that the mechanisms by which the core promoter governs transcriptional activity may be gene specific. As indicated in Chapter I, variant TATA boxes have been suggested to function as a selective element to determine tissue-specific gene expression as in the case of a heterogenous TATA element of the muscle-specific myoglobin gene (Wefald et al., 1990). It has been demonstrated that synergistic interactions between heterologous upstream activation elements and specific TATA sequences can confer cell-type
specificity for myoglobin transcription (Grayson et al., 1995). Distinctive TFIID complexes composed of specific TAFs are likely responsible for selecting different core promoter elements including TATA sequences (Pugh and Tjian, 1992; Chiang et al., 1993; Chen et al., 1994; Goodrich and Tjian, 1994; Jacq et al., 1994; Verrijzer et al., 1995). Consistent with this hypothesis, recent studies have shown that different core promoters may require distinct basal GTFs for efficient transcription initiation (Pavin et al., 1992; Tyree et al., 1993; Pavin and Sharp, 1993; Holstege et al., 1995). Further, Das et al. (1995), using an elegant in vivo experiment model, have demonstrated that changing the arrangement of core promoter elements can alter the response of RNA Pol II to upstream activation by the same activator.

While gene-specific activators have been studied intensively in an effort to elucidate the mechanisms of cell-type specific expression, the role of transcriptional suppression has not received enough attention. Possible reasons for the requirement of mechanisms to inhibit gene transcription include: 1). Negative regulatory elements may be responsible for complete gene inactivation in tissues where the particular gene should be silent; 2). In those tissues where a gene is inducible it may be important to turn transcription off rapidly once the inducing stimulus is removed; 3). The presence of negative elements in addition to positive elements allows gene activity to be adjusted according to multiple signals, since transcription efficiency is dependent on the combined effects of all regulatory elements (Renkawitz, 1990). It should be noted that gene suppression can be achieved through multiple mechanisms, such as nucleosome binding and DNA methylation, in addition to transcriptional repression by silencers or regulatory elements in a gene.

It is of interest to note that mammalian BSP gene expression can be repressed by extracellular signals such as vitamin D3 (Chapter V) and PTH (Li and Sodek, unpublished data), although studies of chicken BSP promoter have shown PTH stimulation through the cAMP/PKA pathway (Yang and Gerstenfeld, 1996). The discrepancy may reflect the complex nature of PTH in both anabolic and catabolic effects on bone remodelling that involve different signalling pathways depending upon target cells at different differentiation stages (cAMP/PKA and
Ca\textsuperscript{2+}/PKC) (reviewed in Chapter I). Although it is not known why BSP is down regulated, it could be speculated that it is important that BSP is not expressed in an inappropriate cell type. Even in bone cells, it is conceivable that no more than a defined amount of nucleator such as BSP is required during normal bone development and formation so that connective tissue would not be “mineralized abnormally”. Thus, BSP can be detected in certain abnormal tissue other than mineralized connective tissue, such as breast tumours. Examples of gene repression are increasing dramatically in the literature, and a true understanding of the different mechanisms that cells use to carry out this important process in tissue-specific gene expression should be forthcoming. Most recently, a YY1 binding motif was found in the first intron of the human BSP gene which has silencer effect (Kerr et al., 1997). It was shown that the YY1 element in the human BSP gene may contribute to elevated transcriptional activity of the human BSP promoter construct in UM106-01 bone cells, while this element can suppress the promoter activity in rat skin fibroblast cells. Similarly, a portion of the tissue-specific regulation of the chicken BSP gene expression was found to be controlled by a silencer element between nts -1244 to -620 in the chicken BSP promoter (Yang and Gerstenfeld, 1997). Of note, elements closely resembling the type II collagen silencer were found in the first intron and the promoter region (nts -408 to -402) in the rat BSP gene (Chapter III; Li and Sodek, 1993) as well as in the chicken BSP promoter (Yang and Gerstenfeld, 1997), although their functions have yet to be characterized.

Isolation of the rat BSP promoter has provided a unique model to study tissue-specific transcription of bone-related genes. The mechanism by which BSP gene expression is induced, stimulated, and repressed by various hormones, growth factors, and cytokines remains a challenging field to explore. However, with the techniques currently available, some of which are discussed below, it should be possible to address some of these questions in detail. It is anticipated that functional dissection of the regulatory regions of BSP gene will provide significant insights into understanding the molecular mechanisms of bone-specific gene expression during osteoblast differentiation.
3. The inverted TATA box provides a unique model to explore many fundamental issues in transcription initiation.

The TATA-like element in rat BSP gene promoter was chosen as the first cis-acting element to study, primarily because of the following considerations. First, in most genes, a TATA element is the central element controlling gene transcription and is the ultimate target for upstream signal transduction. Elucidating the functionality of the putative TATA element in the BSP gene was, therefore, necessary for subsequent characterization of other regulatory elements and to fully understand the regulatory mechanisms that control BSP gene transcription. Second, the inverted TATA box provides a unique model to explore many fundamentally important issues involving transcription initiation. Third, variations in TATA sequences could be involved in tissue-specific gene expression. It was of interest, therefore, to determine if this unique inverted TATA box contributes to the bone-specific expression of the BSP gene.

The most extensively characterized transcription factor that recognizes the core promoter sequence probably is TBP, the central subunit of the TFIID complex (reviewed by Hernandez, 1993; Maniatis and Tjian, 1994; Chapter I). The functions attributed to TFIID, such as promoter recognition, coactivator activity, and RNA Pol II recruitment, all point to a pivotal role for TFIID in receiving, integrating, and relaying a mosaic of molecular signals imparted by the enhancer-binding proteins that regulate gene-specific transcription. Despite intensive work during the past decade that have revealed important structure-function relationships, several important issues have not been resolved. For example, the generality of these structure-function relationships is uncertain, since previous data are derived mostly from studies of the adenovirus IV promoter which contains a canonical TATA box (TATAAA). Thus, it is still not clear how the GTFs are utilized in promoters that contain a non-consensus TATA box or no TATA box for transcription initiation. Furthermore, it is unclear why many TAF proteins are required to facilitate the TBP function. The ability of an asymmetrical TBP molecule to bind to an inverted TATA box and to form a functional PIC (Chapter IV; Li et al., 1995) has important implications in elucidating the
binding interactions between the promoter, TBP and PIC components, since it is a mirror image of the canonical TATA box. The inverted TATA box in the BSP gene, therefore, could be useful in attempt to answer some general questions concerned in the transcription initiation.

There are two inter-related questions concerning the initiation of transcription utilizing GTFs: (1) How do GTFs select the inverted TATA box (as well as non-consensus TATA boxes or TATA-less promoters) to initiate transcription? (2) what factors really determine the polarity of transcription? To date, most studies of basal transcription have used a small set of TATA-containing promoters. There is evidence that non-consensus promoters may use alternative pathways for PIC assembly. One pathway may involve recognition of an initiator element, defined as sequences spanning a transcription start site that can function autonomously as a core element in the absence of a TATA box. However, it is unlikely that the initiation of the BSP transcription utilizes the mechanisms of the initiator element, because the rat BSP promoter does not contain an identifiable initiator sequence and the inverted TATA box is critically involved BSP gene transcription (Chapter IV; Li et al., 1995). Thus, how TBP selects and use the inverted TATA box is an intriguing question.

I have proposed two models to explain the functional implications with respect to the binding of TBP to an inverted TATA box (Chapter IV; Li et al., 1995). One possibility is that TBP could orientate itself, bind to the upper strand of the inverted TATA box, and recruit other GTFs and RNA Pol II for downstream transcription just like the regular canonical TATA box (Fig. 4.9a). In this model, the flanking DNA sequence of the inverted TATA box might be important in dictating TBP binding to the upper strand, while the auxiliary factors (eg. TAFs) could stabilize its binding. This hypothesis is supported by the computer generated model in which the binding of TBP to an inverted TATA sequence is not much weaker than to a canonical one (Kim and Burley, 1994) and TFIIIB may help TBP define its orientation of the DNA (Nikolov et al., 1995). This model may explain the binding of TBP to variant TATA-box sequences, and perhaps also to TATA-less promoters. Alternatively, TBP could bind to the 5'-TATAAA-3' sequence on the
lower strand of the inverted TATA box. The positioning of the RNA Pol II for downstream transcription would then be determined by TAFs or other auxiliary factors (Fig. 4.9b), in which case protein-DNA interactions other than those involved with TBP binding, or protein-protein interaction might be anticipated. To further elucidate this structure-function relationship, various mutated TBPs with different DNA binding affinities could be tested in footprinting experiment and \textit{in vitro} transcription analysis to provide clear insights into TBP-TATA interactions. The definite solution would be obtained by dissection of the PIC required for initiation of BSP transcription and the co-crystalization of the PIC proteins with the inverted TATA box. Recent co-crystalization of the TFIIIB-TBP-TATA complex (Nikolov \textit{et al}., 1995) and the TFIIA-TBP-TATA complex (Tan \textit{et al}., 1996) has already shed light on this research direction.

Another question is: does the inverted TATA box have a role in the specific regulation in BSP expression in bone cells at a defined developmental stage? Are there any "tissue-specific" TAFs involved in the specific transcription of BSP? Variant TATA boxes have been suggested to function as selective elements to determine tissue-specific gene expression by interaction with distinctive TFIID complexes (Pugh and Tjian, 1992) and certain TAFs have been shown to have promoter-specificity (Chen \textit{et al}., 1994; Goodrich and Tjian, 1994; Jacq \textit{et al}., 1994). Moreover, since the binding of TBP to the TATA box induces the DNA to bend, variant nucleotide(s) in the TATA boxes might be able to dictate the degree of DNA bending, and as such provide a means of fine tuning in stimulation or repression of transcription via TBP/specific trans-factor interactions. In this regard, the transcriptional efficiency of BSP promoters carrying either the natural inverted TATA box or the mutated "normal" TATA box could be tested by transient transfection of different cell types and, in the future, introduced into transgenic mice to verify these results.

In my studies, I have demonstrated that the inverted TATA-box in the BSP gene is functional and crucial for directing downstream transcription. Thus, it can be concluded that the orientation of a TATA element is not a determinant for orientating transcription in eukaryotic genes. The discovery of a naturally occurring inverted TATA box in the BSP gene provides an excellent
model which can be used to answer several fundamental questions described above.

4. The BSP VDRE Indicates a Novel Mechanism for Vitamin D₃ to Repress Gene Transcription.

While vitamin D₃ can induce expression of a number of genes, it suppresses BSP gene expression at both basal and stimulated levels (Oldberg et al., 1989) in accordance with its ability to inhibit bone formation in vitro (Ishida et al., 1993). The majority of VDREs characterized so far are cis-acting elements which have enhancer properties, i.e., they mediate gene activation by vitamin D₃. The only reported VDREs that have a silencer effect are the VDRE in the PTH gene and the VDRE in the interleukin-2 (Table 5.1). A unique VDRE in the rat BSP gene that mediates repressive signal from vitamin D₃ was identified and described in Chapter V. Importantly, the rat BSP VDRE indicates a novel mechanism for vitamin D₃ repressing gene transcription. The VDR appears to compete with TBP for occupancy of the TATA/VDRE element, following a vitamin D₃ signal, thereby preventing the assembly of pre-initiation complex that is required for the initiation of transcription (Fig. 5.7). This mechanism supports the concept that regulatory signals can directly affect the core promoter element. Together with the GRE overlapping the TATA element in the osteocalcin gene which is responsible for the suppressive effect of glucocorticoids on osteocalcin transcription (Morrison et al., 1989; Strömstedt et al., 1991), these observations may represent a general mechanism of steroid hormone action on bone cells during bone development and remodelling.

In comparison with other VDREs that have been characterized, some novel characteristics were revealed for the rat BSP VDRE. However, to completely elucidate the mechanisms whereby vitamin D₃ represses the transcription of BSP gene, a few important questions have to be addressed.

First, direct evidence for competitive binding between VDR and TBP in the BSP promoter has yet to be established. Attempts to study this competition in an in vitro transcription system...
using the inverted TATA box and crude Hela nuclear extracts have not been successful (Kim et al., 1996), apparently due to differences in binding conditions required for the VDR and TFIID complex. Also, since the binding of TFIID to the TATA sequence is much more stable than the TBP binding alone due to the assistance of TAFs (Verrijzer et al., 1993; Verrijzer et al., 1995), competitive binding to the inverted TATA box needs to be examined using purified recombinant TBP and VDR. The functional significance of the competitive binding can then be tested by in vitro transcription assays using purified components. In fact, most recently, it was demonstrated that the GR and TBP bound to the overlapping GRE/TATA elements in the osteocalcin gene in a mutually exclusive manner, and the competitive binding of the GR is dependent on dimerization (Meyer et al., 1997). It is also possible that a VDR with higher binding affinity (perhaps in an in vivo form) has to be present in order to compete the TFIID complex. Thus, it should be determined in what form the VDR can bind to the BSP VDRE with maximal affinity, ie., as a homodimer or heterodimer, and/or in the presence of ligand.

Consequently, a second issue is to identify the nature of the functional complex which bind to the VDRE of the BSP gene. More specifically, is a homodimer or heterodimer of the VDR required for suppression in vivo? What is the function of the vitamin D3 ligand? If a heterodimer is required, what other protein interacts with the VDR?

While a sub-class of the steroid hormone receptor superfamily, including receptors for glucocorticoids and estrogen, preferably form homodimers on palindromic DNA sequences (Kumar and Chambon, 1988; Tsai et al., 1988, Wrange et al., 1989), another sub-class of the steroid hormone receptor superfamily, including receptors for vitamin D3, retinoic acid and thyroid hormone, recognizes a DNA element organized as direct repeats of hexanucleotide sequences in heterodimer form (Umesono et al., 1991; Kliewer et al., 1992a), with RXRs being most common partner (Kliewer et al., 1992b). However, in contrast to most identified VDREs that accommodate VDR complexed with an accessory protein as a heterodimer, the rBSP VDRE can bind to purified recombinant VDR largely as homodimer in the absence of vitamin D3. Recently, Carlberg and co-
workers (1993) suggested that both homodimeric and heterodimeric VDR complexes may be involved in vitamin D₃-induced functions and that the nature of dimerization could be dependent on the specific arrangement of DNA of the target site. My studies have also revealed that variation in the consensus VDRE sequence can impact a VDR binding characteristics (Chapter V; Li et al., 1997). Thus, while the DNA binding domain of VDR (VDRF) binds to the mouse osteopontin VDRE with high affinity, its interaction with the hOC and rBSP VDREs is poor, suggesting that a full configuration of the VDR is necessary to complex with another protein to form a homodimer or heterodimer on either VDRE.

Although members of the steroid receptor superfamily share highly conserved DNA binding domains and their cognate response elements have a high degree of similarity, a wide range of biological effects can be achieved as a result of highly-specific DNA-protein interaction. The surprisingly high selectivity of protein-DNA recognition may therefore be dictated by subtle differences in the hormone response elements together with specific protein conformation. In this regard, none of the identified VDREs are identical (Table 5.1).

Attempts to completely abrogate VDR binding to DNA and the vitamin D₃-mediated repression by point mutations have not been successful (Kim et al., 1996), indicating that the mechanisms of the repression are more complex than what was anticipated. Although the effects of mutagenesis on the VDR binding and function were not remarkable, the changes in the VDR binding to the mutated rBSP VDRE did correlate with the changes in the responsiveness of the rBSP promoter to vitamin D₃ (Kim et al., 1996). Kerner et al. (1989) have demonstrated that multiple copies of VDREs are more efficient than a single copy in mediating vitamin D₃ signals. Therefore, the VDRE-like half-sites flanking the VDRE/TATA element might also contribute to the VDR recognition and function for the rat BSP promoter, acting as multiple copies of VDRE to facilitate the binding of VDR. However, extensive mutational analysis, DNase footprinting, and methylation interference studies are necessary in order to decipher which nucleotide(s) within or adjacent to this putative VDRE dictate the VDR binding and function.

Cloning of the BSP promoter has facilitated studies of BSP gene regulation through upstream regulatory cis-acting elements including the GRE/AP-1 site (Ogata et al., 1995; Yamauchi et al., 1996) and a TAE site that mediates the effect of TGF-β (Ogata et al., 1997).

Upon the completion of the proliferation stage, the osteogenic progenitor cells mature into osteoblasts, at which stage, BSP is expressed. Glucocorticoids have been shown to induce BSP expression in association with osteogenic differentiation (Yao et al., 1994) and to stimulate BSP expression in osteoblasts (Oldberg et al., 1989; Kasugai et al., 1991). Analysis of the mechanisms of glucocorticoid-induced BSP expression in osteoblasts has demonstrated the involvement of both direct and indirect pathways. Dexamethasone stimulates BSP expression <6-fold in ROS 17/2.8 cells, most of the stimulation being mediated by a nuclear post-transcriptional mechanism. That there is also a direct effect of glucocorticoid hormones on BSP gene transcription is indicated from an increase in BSP mRNA that was not blocked by cycloheximide and a comparable increase in transcription observed in nuclear run-on assays. A similar modest, but reproducible, <1.5-fold increase in transcription was observed with rat BSP promoter constructs that included the upstream sequence of pLUC6, nts -938 to -899. Gel mobility shift assays have identified a putative glucocorticoid response unit (GRU), comprising three GRE half-sites and a putative AP-1 site, located between nts -906 to -931 of the rat BSP gene promoter (Fig. 3.4; Ogata et al., 1995). That dexamethasone exerts its effect on BSP transcription through this site is supported by the observation that the suppression of BSP transcription through the AP-1 site in this region is abrogated by dexamethasone, since either the c-JUN or GR could displace its counterpart from the GRE/AP1 site, indicating that the abrogation of AP-1-mediated gene suppression by glucocorticoids involves competitive binding (Yamauchi et al., 1996).

A steroid regulatory site, analogue to the GRE/AP-1 element structure in the rat BSP promoter, is present in the human osteocalcin promoter where a VDRE is overlapped by an AP-1
binding site. The binding of FOS-JUN protein complex on the AP-1 site in the VDRE can suppress both basal level and vitamin D-enhanced transcription of the osteocalcin gene (Owen et al. 1990b; Lian and Stein 1992). Since a reciprocal relationship between AP-1 activity and osteocalcin gene expression is observed in osteoblastic cells (Owen et al. 1990a), a coordinate occupancy of the FOS-JUN protein complex at the VDRE can account for the absence of osteocalcin gene expression in proliferating osteoblasts. Similarly, the absence of BSP expression in proliferating osteogenic cells can be related to AP-1 mediated suppression (Yamauchi et al., 1996).

Another possible explanation of the modest stimulation of BSP achieved by the GRE identified in the rat BSP promoter is that additional GREs may exist in unidentified regions and the synergistic action of all GREs is required to modulate the stimulative effect of glucocorticoids on BSP transcription. Thus, recent studies on glucocorticoid regulation of the rat osteocalcin gene have indicated that multiple GREs located in the distal and proximal promoter are required to mediate the repressive effect of glucocorticoid on osteocalcin gene transcription (Heinrichs et al., 1993; Aslam et al., 1995). Therefore, further sequence analysis to search for additional GRE(s) and studies of the combinatorial effects of all possible GREs in the rat BSP gene is necessary to comprehensively assess the mechanisms whereby BSP gene expression is up-regulated by glucocorticoids.

In other studies, it has been shown that stimulation of BSP gene expression by TGF-β in osteoblastic cells also involves in both direct and indirect pathways (Ogata et al., 1997). A TGF-β activation element (TAE) which overlaps an AP-2 site was identified at nt ~500 in the rat BSP promoter (Fig. 3.3; Chapter III). This TAE is similar, but not identical, to the NF-1 binding site characterized in the mouse α2(I) collagen (Rossi et al., 1988) and rat α1(I) collagen (Ritzenthaler et al., 1991; 1993) genes. Constructs that include the BSP TAE located within nts -801 to -426 of the rat BSP promoter sequence (Fig. 3.3; Chapter III) were found to enhance transcriptional activity ~1.8-fold in osteoblast cells treated with TGF-β1, a result consistent with a nuclear "run-on" transcription analysis. The functionality of the BSP TAE was shown by an increased binding
of a nuclear protein from TGF-β1-stimulated cells in gel mobility shift assays and from the attenuation of TGF-β1-induced luciferase activity when cells were co-transfected with a double-stranded TAE oligonucleotide.
Chapter VII

Concluding Remarks
Skeleton formation involves multi-step processes, which include patterning of skeleton elements, commitment of mesenchymal cells to osteogenic and chondrogenic lineage cells, and terminal differentiation of precursor cells into osteoblasts and chondrocytes. Osteoblast differentiation and matrix protein production are carefully regulated by hormones and growth factors that target individual genes through complex signalling pathways. The signalling pathways converge on promoter elements that control the transcription of specific genes. By selecting a bone-specific gene, such as BSP, that is expressed in a well-defined temporo-spatial manner for detailed study, it was hoped that a part of this complex regulatory system can be dissected. Cloning of the promoter of the rat BSP gene has provided a unique opportunity to characterize molecular mechanisms that are involved in controlling the expression of this gene and to generate insights into the broader regulation of osteogenic differentiation. The following conclusions can be drawn from my studies:

1. This study describes the first cloning and characterization of a BSP gene promoter, which is a pre-requisite for studies of molecular mechanisms that regulate BSP gene expression at the transcription level.

2. The rat and human BSP promoters are characterized by a highly-conserved proximal promoter region, termed the “BSP box”, that extends upstream from the transcription start site to nt -370. Within this BSP box an inverted TATA box overlapped by a VDRE and an inverted CCAAT box, together with NFkB, CRE, AP-1, and homeobox protein binding sites are all conserved. The rat BSP promoter also contains a novel 18 nucleotide palindrome. While the inverted TATA box and the inverted CCAAT box are required for basal transcription activity of the BSP promoter, other cis-acting elements may be involved in the cell-type specific and temporo-spatial specific expression of the BSP gene.

3. Transcription driven by the rat BSP promoter is directed by a novel inverted TATA box which is conserved in the human BSP promoter. This is the first, perfectly inverted TATA box
characterized in a eukaryotic gene. The existence of an inverted TATA element indicates that the orientation of a TATA box does not determine the direction of transcription, contrasting prevailing beliefs. This unique inverted TATA box can help address many unresolved issues involving the assembly of the pre-initiation complex.

4. The characterization of the VDRE-like sequence overlapping the inverted TATA box indicates a novel mechanism for gene suppression by vitamin D₃. Vitamin D₃ appears to suppress BSP transcription through a competition between the VDR and TBP for occupancy of the TATA/VDRE element; binding of the VDR preventing the assembly of the pre-initiation complex that is required for the initiation of transcription of the BSP gene.

5. The vitamin D₃ receptor binds to this rBSP VDRE with a strong preference for homodimer formation, whereas the VDR binds to other identified VDREs in both heterodimer and monomer forms. Homodimer binding of the VDR to the rBSP VDRE was demonstrated using recombinant VDR protein. A truncated VDR, that includes the DNA binding domain of the VDR, interacts with the rBSP VDRE very weakly, suggesting that complete configuration of the receptor protein is required for binding and functional activity in the BSP gene.

6. Subtle differences in the nucleotide sequence of VDREs have been shown to affect VDR binding, which mediates the vitamin D₃ response. No two VDREs identified to date are identical despite their similarity to the consensus DNA sequence. Thus, while similarity has greatly facilitated our understanding of transcription factors and cis-acting elements, variation undoubtedly enriches our knowledge of specificity.
While I have been able to identify and characterize several important cis-acting elements that regulate the transcription of rat BSP gene, many more regulatory elements and their cognate binding transcription factors still remain to be characterized. Given the complexity of transcriptional regulation, it is not surprising that many difficulties were encountered, which will be a challenge for future studies. Future studies of transcriptional regulation and signal transduction in the regulation of BSP gene expression will undoubtedly provide a better understanding of the role of BSP in bone formation and skeletal development. This thesis has laid a foundation for these future studies.
Bibliography


Fisher LW, Termine JD, and Young MF. (1989). Deduced protein sequence of bone small proteoglycan I (biglycan) shows homology with proteoglycan (decorin) and several non-connective tissue proteins in a variety of species. J Biol Chem 264:4571-4576.


Forrest SM, Ng KW, Findlay DM, Michelangeli VP. Livesey SA, Partridge NC, Zajac JD, and Martin TJ. (1985). Characterization of an osteoblast-like clonal cell line which responds to both parathyroid hormone and calcitonin. Calcif Tiss Int 37: 51-56


